

RCMF: Robust Constrained Matrix Factorization for Hyperspectral Unmixing

Naveed Akhtar and Ajmal Mian

Abstract—We propose a constrained matrix factorization approach for linear unmixing of hyperspectral data. Our approach factorizes a hyperspectral cube into its constituent endmembers and their fractional abundances such that the endmembers are sparse nonnegative linear combinations of the observed spectra themselves. The association between the extracted endmembers and the observed spectra is explicitly noted for physical interpretability. To ensure reliable unmixing, we make the matrix factorization procedure robust to outliers in the observed spectra. Our approach simultaneously computes the endmembers and their abundances in an efficient and unsupervised manner. The extracted endmembers are nonnegative quantities, whereas their abundances additionally follow the sum-to-one constraint. We thoroughly evaluate our approach using synthetic data with white and correlated noise as well as real hyperspectral data. Experimental results establish the effectiveness of our approach.

Index Terms—Blind source separation, hyperspectral unmixing, robust matrix factorization, sparse representation, unsupervised unmixing.

I. INTRODUCTION

HYPERSPECTRAL imaging acquires precise spectral information about the scene radiance that is exploited for efficient earth exploration in remote sensing. Nevertheless, contemporary hyperspectral imaging lacks in spatial resolution [1], [2], causing a pixel of a remotely sensed image to generally correspond to a large area on the ground (see Fig. 1). This causes the spectra sensed at a pixel to be a mixture of reflectances of different materials present in that area. Moreover, multiple scatterings of light and the presence of intimate material mixtures on the ground also result in mixing of the sensed material spectra [3]. Identifying materials on the earth's surface by extracting their pure spectral signatures (*endmembers*) and computing their proportions (*fractional abundances*) in a hyperspectral pixel are the two fundamental tasks handled by *hyperspectral unmixing*.

To unmix a pixel, it is common to model it as a linear combination of its constituent endmembers [4], [5]. Such a modeling is effective when materials occur in spatially distinct

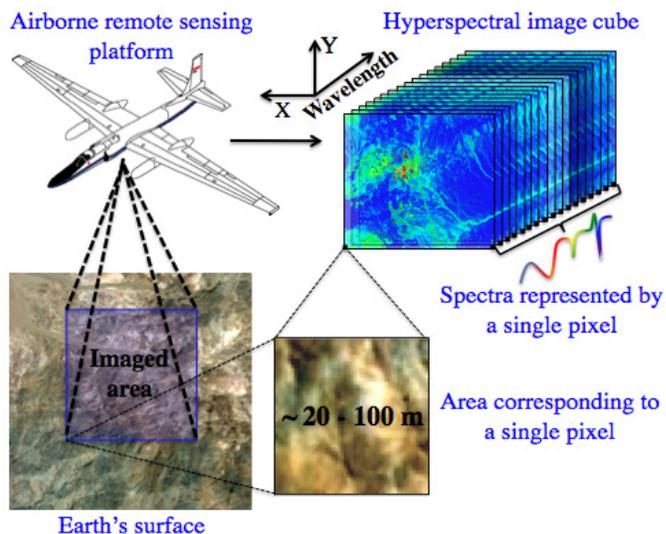


Fig. 1. Hyperspectral images comprise hundreds of spectral channels (fewer channels are shown for illustration), but a pixel usually corresponds to a large area on the ground, making it a mixture of reflectance spectra of multiple materials.

regions on the earth's surface and minimal light scattering is observed in the scene [3]. This paper also focuses on the linear mixing model (LMM) of the spectra [36]. Under LMM, the convex geometry of the observed spectra is often exploited by the unmixing approaches [6]–[9]. These approaches identify the endmembers as the vertices of a simplex, formed by the convex hull of the endmembers. These techniques assume the presence of at least one pure endmember pixel in the image. The simplex growing algorithm [10], vector component analysis [11], N-FINDR [12], pixel purity index [13], iterative error analysis [14], and successive volume maximization [15] are some classic examples in this direction.

In practice, pure pixels are not always present [16], [17] for each endmember in a hyperspectral image. Hence, approaches like iterative constrained endmembers (ICEs), minimum volume simplex analysis, and sparsity promoting ICE tend to generate the endmembers from the image itself. Nevertheless, these approaches are computationally expensive [21] and they do not perform well in highly mixed scenarios [16]. Formulating hyperspectral unmixing as a statistical inference problem under the Bayesian framework improves performance for such scenarios [3]. However, the computational complexity of Bayesian methods generally remains prohibitive [22].

More recently, Iordache *et al.* [16] have shown the effectiveness of sparse regression methods [23]–[25] for hyperspectral unmixing by approaching the problem in a supervised

Manuscript received August 6, 2016; revised November 20, 2016 and February 7, 2017; accepted February 8, 2017. This work was supported by the Australian Research Council project under Grant DP110102399.

N. Akhtar is with the College of Engineering and Computer Science, The Australian National University, Canberra, ACT 2601, Australia (e-mail: naveed.akhtar@anu.edu.au).

A. Mian is with the School of Computer Science and Software Engineering, The University of Western Australia, Crawley, WA 6009, Australia (e-mail: ajmal.mian@uwa.edu.au).

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.2017.2669991

manner. Their approach assumes prior knowledge of the potential endmembers in an image, and represents a mixed pixel as a sparse linear combination of those endmembers. The success of this framework has led to numerous efforts in tailoring sparse regression algorithms for sparse unmixing (see [22], [26]–[32]). Although useful, the sparse unmixing framework relies on the exhaustiveness of the dictionary comprising the potential endmembers in the image. Whereas larger dictionaries are required to ensure correct identification of the endmembers, increasing the dictionary size generally results in increasing its coherence [33]. It is well known that high coherence of the dictionary can lead to poor performance of the sparse regression framework [34].

Hyperspectral unmixing can be readily formulated as a nonnegative matrix factorization (NMF) problem [47], widely solved for blind source separation [48]. A major advantage of NMF over sparse unmixing [16] is that it does not assume the potential endmembers to be known *a priori*. Nevertheless, the nonconvexity of the problem usually results in solutions that are only locally optimal. In general, this issue is resolved by adding more constraints to the problem, giving rise to a variety of constrained NMF approaches (see [49]–[52]). However, most of these approaches underperform for hyperspectral unmixing, as they are not originally proposed for this purpose [59]. Pauca *et al.* [53] first imposed a spectral smoothness constraint over NMF for spectral data analysis. However, as in sparse unmixing, their approach requires *a priori* known library of endmembers for effective unmixing. Miao and Qi [54] used a minimum volume constraint with NMF for extracting the endmembers. Nevertheless, their technique necessarily requires a data dimensionality reduction step for unmixing, which can result in the loss of useful information.

Jian and Qian [55] incorporated a piecewise smoothness constraint into NMF for spectral unmixing. Lu *et al.* [56] proposed a manifold learning and sparse NMF-based hyperspectral unmixing approach that considers local space information for the improved performance. Later, Lu *et al.* [57] also proposed a structure constrained sparse NMF method. Qian *et al.* [58] constrained NMF with an $L_{1/2}$ -sparsity constraint for hyperspectral unmixing. Similarly, Yuan *et al.* [59] introduced a substance dependence constraint into NMF to make the factorization procedure more stable. The aforementioned matrix factorization-based approaches generally unmix hyperspectral data well. Nevertheless, they suffer from a common problem that they can also result in *artificial* endmember spectral signatures that do not associate with real materials in the scene. These spectra are formed as by-products of the unmixing approaches themselves, but the approaches fail to provide any useful physical interpretation of these spectra.

In this paper, we propose a robust constrained matrix factorization approach for hyperspectral unmixing that mitigates the problem of artificial endmember spectra. To keep the association between the extracted endmember spectra and the materials in the scene, our approach constrains the spectral signatures to be linear combinations of the image pixels themselves. In this manner, the presence of any pure material

pixel in the image is also readily exploited by our approach. Similarly, it is also able to effectively utilize the pixels with dominant endmembers because extracting the endmember spectra from these pixels is easier than to compute them as pure mathematical outcomes. Our approach explicitly notes the contributions of the pixels to the extracted endmember spectra. This helps in physical interpretability of the extracted spectra. Since the endmember spectra are constructed using the observed data, we make the approach robust to any outliers present in the data. To achieve our objectives, we reformulate the unmixing problem to incorporate the said physical interpretability of the endmembers and systematically derive an efficient optimization algorithm to solve it. The algorithm simultaneously extracts the endmember spectral signatures and computes their abundances.

In this paper, we also propose a nonnegative variant of the subspace pursuit (SP) algorithm, which is known for its computational efficiency [22]. The proposed nonnegative SP is exploited in our approach for efficient matrix factorization. The endmembers computed by our approach are nonnegative values and their estimated abundances additionally follow the well-known sum-to-one constraint [3]. We show the effectiveness of our approach on synthetic and real hyperspectral data.

II. PROBLEM FORMULATION

A. Linear Mixing Model

This paper focuses on the LMM [36], which represents a pixel $\mathbf{y} \in \mathbb{R}^m$ of a hyperspectral image as

$$\mathbf{y} = \Phi\boldsymbol{\alpha} + \boldsymbol{\epsilon} \quad (1)$$

where $\Phi \in \mathbb{R}^{m \times K}$ contains the endmembers as its columns, $\boldsymbol{\alpha} \in \mathbb{R}^K$ encodes their fractional abundances, and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ represents the error, considered as additive Gaussian noise. Under this model, the coefficients $\alpha_{i \in \{1, \dots, K\}}$ of $\boldsymbol{\alpha}$ must satisfy two constraints [3]: 1) $\forall i, \alpha_i \geq 0$, i.e., *abundance nonnegativity constraint* (ANC) and 2) $\sum_{i=1}^K \alpha_i = 1$, i.e., *abundance sum-to-one constraint* (ASC). These constraints signify the fact that the proportions of the endmembers in a pixel are nonnegative quantities that add up to 1.

B. Unmixing as Constrained Matrix Factorization

Let $\mathbf{Y} \in \mathbb{R}^{m \times n}$ be the matrix formed by arranging the n pixels of a hyperspectral image as its columns. Assuming that Φ now contains all the endmembers in the whole image, we can compactly write LMM as follows:

$$\mathbf{Y} = \Phi\mathbf{A} + \mathbf{E} \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{K \times n}$ and $\mathbf{E} \in \mathbb{R}^{m \times n}$ are the abundance matrix and the noise matrix, respectively. In this paper, both Φ and \mathbf{A} are considered to be unknown, making unmixing a blind source separation problem. Incorporating ANC and ASC in (2) results in

$$\mathbf{Y} = \Phi\mathbf{A} + \mathbf{E} \quad \text{s.t. } \forall i, j, \alpha_{i,j} \geq 0; \|\boldsymbol{\alpha}_j\|_1 = 1 \quad (3)$$

where, $\alpha_{i,j}$ is the coefficient of \mathbf{A} at index (i, j) , $\boldsymbol{\alpha}_j$ denotes the j th column of \mathbf{A} and $\|\cdot\|_1$ computes the ℓ_1 -norm.

As Φ contains the endmembers of the complete image and α_j corresponds to the j th pixel only, we can expect α_j to be generally sparse. However, introducing an explicit sparsity constraint over α_j can come with an additional computational cost. It is a well-known fact that sparseness can be imposed over a representation (α_j in this case) by restricting its ℓ_1 -norm [17], [68]. Since the ASC in (3) is doing exactly the same thing, a form of sparseness is naturally exhibited by α_j . Therefore, we do not impose additional sparsity constraint in (3). Previously, Chen *et al.* [63] also reported the sparseness-inducing effect of restricting the sum of coefficients of a nonnegative representation vector to 1.

The above-mentioned sparseness is of the form where many coefficients of the vector shrink toward zero. From that perspective, if an abundance vector $\tilde{\alpha}_j$ does not follow the ASC, additionally imposing ASC will sparsify the vector by shrinking its coefficients when $\sum_i \tilde{\alpha}_{i,j} > 1$. When $\sum_i \tilde{\alpha}_{i,j} = 1$, the representation remains unaltered, and in the case of $\sum_i \tilde{\alpha}_{i,j} < 1$ the ASC may actually expand a few coefficients instead of shrinking them. Nevertheless, due to the high dimensionality of the representation vector, most of the vector coefficients must still remain very small in both of the latter cases. Therefore, we can argue that the solutions computed under ASC and ANC already exhibit a form of sparseness. We note that this analysis applies to practical conditions where the dictionary atoms have bounded energy, resulting in substantial weights of the representation coefficients.

Since spectra are nonnegative quantities, we can also force the coefficients $\varphi_{h,i}$ of Φ to be nonnegative, resulting in

$$\mathbf{Y} = \Phi \mathbf{A} + \mathbf{E} \quad \text{s.t. } \forall i, j, \alpha_{i,j}, \varphi_{h,i} \geq 0; \quad \|\alpha_j\|_1 = 1. \quad (4)$$

Estimating Φ and \mathbf{A} that satisfy (4), is a constrained matrix factorization problem.¹ A matrix Φ obtained by solving that problem, would generally approximate \mathbf{Y} 's endmembers well. Nevertheless, some of the computed endmembers in Φ could also be artificial. Artificial endmembers do not belong to any real material in the scene, but exist only in the solution space of the problem due to its inherent nonconvexity, which also makes their physical interpretation hard. In order to mitigate this issue, we reformulate the model in (4) as follows.

We force the columns $\varphi_{i \in \{1, \dots, K\}}$ of Φ to be nonnegative linear combinations of the image pixels themselves. Moreover, we force each φ_i to only use *at most* k pixels in its construction, where k is a small positive integer. Concretely, our model becomes

$$\begin{aligned} \mathbf{Y} &= \mathbf{Y} \Xi \mathbf{A} + \mathbf{E} \\ \text{s.t. } \forall i, j, \zeta_{j,i}, \alpha_{i,j} &\geq 0; \quad \|\alpha_j\|_1 = 1 \\ \|\xi_i\|_o &\leq k \end{aligned} \quad (5)$$

where $\Xi \in \mathbb{R}^{n \times K}$ is a matrix with coefficients $\zeta_{j,i}$ and ξ_i as its i th column. The symbol $\|\cdot\|_o$ denotes the ℓ_o pseudonorm that counts the number of nonzero coefficients in a vector. Since both \mathbf{Y} and Ξ are nonnegative, the endmembers in $\Phi = \mathbf{Y} \Xi$ also remain nonnegative in the above formulation.

¹We intentionally use a broader term than *NMF* because the problem contains more than just nonnegativity constraints.

At this point, it is worth mentioning that Ambikapathi *et al.* [67] observed an important geometric property of hyperspectral data, in that convex-hull of pixels lies strictly inside the convex hull of the endmembers. Considering this, the model in (5) may appear restrictive because constructing endmembers as nonnegative combinations of the pixels seems improbable under this observation. However, this is not true. Note that the observation is valid only in noise-free settings [67]. In (5), we always consider \mathbf{E} to be a nonzero matrix, which also qualifies the scope of this paper to the practical noisy scenarios. In noisy conditions, the convex hull of the observed pixels is not necessarily bounded by the convex hull of the endmembers. Not to mention, we do not restrict $\|\xi_i\|_1$ to 1 in (5), which would be required to define the convex hull of the pixels. At practical noise levels, our model accurately reconstructs the endmembers, which is verified by the unmixing accuracy of our approach in Sections IV and V.

The model in (5) records the relationship between the endmembers and the pixels in the matrix Ξ . Hence, it is also able to explicate any pixels that it considers to be the pure endmembers.² For such pixels, $\|\xi_i\|_o$ becomes exactly 1. This characteristic of the model is certainly desirable, which also justifies the use of the ℓ_o -sparsity constraint. However, there one subtle issue remains. Since the model requires the endmembers to be constructed from the pixels themselves, an approach for computing Ξ and \mathbf{A} under (5) must ensure that the computations are robust to outliers among the pixels.

III. PROPOSED APPROACH

We propose a constrained matrix factorization approach for hyperspectral unmixing that computes the endmembers and their abundances according to the model in (5). For reliable unmixing, the approach is also kept robust to any outliers among the pixels. We first describe the objective function for our approach and then explain its optimization procedure.

A. Objective Function

The objective function for computing Ξ and \mathbf{A} under (5) can be formulated as

$$\begin{aligned} \min_{\Xi, \mathbf{A}} \quad & \|\mathbf{Y} - \mathbf{Y} \Xi \mathbf{A}\|_F^2 \\ \text{s.t. } \forall i, j, \zeta_{j,i}, \alpha_{i,j} &\geq 0; \quad \|\alpha_j\|_1 = 1; \quad \|\xi_i\|_o \leq k \end{aligned} \quad (6)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. In (6), the cost associated with the reconstruction of a pixel is quadratic. Generally, a large reconstruction error can be expected for an outlier because outliers do not follow the same distribution as the actual data. This makes the quadratic penalty for the outliers to be too strict. Hence, to render the optimization procedure less sensitive to the outliers, we force the penalty for the larger errors to become linear. This is done by modifying

²We do not claim that such pixels would necessarily be pure endmembers, as it may only be the case that they are linearly inseparable. However, explicit identification of such pixels can be useful in further scrutiny, if desired.

the objective function to the following:

$$\begin{aligned} & \min_{\Xi, \mathbf{A}, \Delta} \left\{ \|(\mathbf{Y} - \mathbf{Y}\Xi\mathbf{A})\Delta^{-\frac{1}{2}}\|_F^2 + \|\Delta\|_1 \right\} \\ & \text{s.t. } \forall i, j, \|\alpha_j\|_1 = 1; \quad \zeta_{j,i}, \alpha_{i,j} \geq 0; \quad \|\xi_i\|_o \leq k; \quad \delta_j \geq \varepsilon \end{aligned} \quad (7)$$

where $\Delta \in \mathbb{R}^{n \times n}$ is a diagonal matrix with strictly positive diagonal entries, the j th of which is denoted as δ_j ; and ε is a scalar constant.

The above modified function is an outlier robust form of (6). To show this, let us focus on the minimization of the cost associated with the j th pixel only, considering both Ξ and \mathbf{A} to be fixed. In that case, (7) reduces to the following optimization problem:

$$\min_{\delta_j} \left\{ \frac{\|\mathbf{y}_j - \mathbf{Y}\Xi\alpha_j\|_2^2}{\delta_j} + \delta_j \right\} \quad \text{s.t. } \delta_j \geq \varepsilon. \quad (8)$$

By differentiating the expression in the brackets *with respect to* δ_j and equating it to zero, we can show that $h = \|\mathbf{y}_j - \mathbf{Y}\Xi\alpha_j\|_2$ becomes the minimizer δ_j^* , which represents a linear penalty. However, this penalty becomes applicable only when $h \geq \varepsilon$. Otherwise, we must choose $\delta_j^* = \varepsilon$ because ε is the closest value to the minimizer that is allowed by the outer constraint ($\delta_j \geq \varepsilon$). In that case, the penalty takes the quadratic form $\{(h^2/\varepsilon) + \varepsilon\}$. Note that, this transformation of the penalty in our objective function is in line with the transformation of the penalty in robust linear regression [40] under the widely known Huber loss function [35].

B. Optimization Algorithm

The optimization problem in (7) is nonconvex. Nevertheless, it has the following desirable characteristics. (a) With fixed Ξ and Δ , the optimization of \mathbf{A} becomes convex. (b) By fixing Ξ and \mathbf{A} , we get a closed-form solution for the optimal Δ . (c) With known \mathbf{A} and Δ , Ξ can be computed by solving a constrained sparse optimization problem. These properties are further explained below. We propose Algorithm 1 that exploits these properties by employing a block-coordinate descent (BCD) scheme to solve the optimization problem (7). The BCD scheme guarantees the algorithm to asymptotically converge to a stationary point in the solution space [37].

To understand the aforementioned properties of the problem and their exploitation in Algorithm 1, let us first focus on the computation of \mathbf{A} . By fixing Ξ and Δ , and replacing $\mathbf{Y}\Xi$ by Φ in (7), we can estimate \mathbf{A} as follows:

$$\min_{\mathbf{A}} \|\mathbf{Y} - \Phi\mathbf{A}\|_F^2 \quad \text{s.t. } \forall i, j, \|\alpha_j\|_1 = 1; \quad \alpha_{i,j} \geq 0. \quad (9)$$

This is a fully constrained least squares (FCLS) problem that can be solved using many existing techniques, e.g., active sets method [38], alternating direction method of multipliers (ADMM) [39]. In this paper, we set the implementation of SUNSAL algorithm provided by Bioucas-Dias and Figueiredo [25] to solve this problem. In Algorithm 1, this computation is carried out in line “2.”

To compute Δ , we separately estimate each of its diagonal entries δ_j , by solving for the objective function in (8). From the discussion in Section III-A, it is clear that an

Algorithm 1 Robust Constrained Matrix Factorization

Input: Data $\mathbf{Y} \in \mathbb{R}^{m \times n}$ normalized in ℓ_2 -norm, number of endmembers K , sparsity level k , number of iterations Q .

Initialize: $\Xi \in \mathbb{R}^{n \times K}$ as a binary matrix with 1 appearing randomly in each column only once, $\Phi = \mathbf{Y}\Xi$, $\Delta \in \mathbb{R}^{n \times n}$ as an identity matrix, and $q = 1$.

Iterations:

- 1: **for** $q = 1$ to Q **do**
- 2: $\mathbf{A} = \text{argmin}_{\mathbf{A}} \|\mathbf{Y} - \Phi\mathbf{A}\|_F^2$ s.t. $\forall i, j, \|\alpha_j\|_1 = 1; \alpha_{i,j} \geq 0$ (Solve FCLS using [25])
- 3: **for** $j = 1$ to n **do**
- 4: $\delta_j^* = \max(\varepsilon, \|\mathbf{y}_j - \Phi\alpha_j\|_2)$
- 5: $\delta_j \leftarrow \delta_j^*$
- 6: **end for**
- 7: $\Gamma = \mathbf{Y} - \Phi\mathbf{A}$
- 8: **for** $i = 1$ to K **do**
- 9: $\varrho = \Delta^{-\frac{1}{2}}\alpha^{i\top}$
- 10: $\psi = \frac{\Gamma\varrho}{\alpha^i\varrho} + \mathbf{Y}\xi_i$
- 11: $\xi_i^* = \text{argmin}_{\xi_i^*} \|\psi - \mathbf{Y}\xi_i^*\|_2^2$ s.t. $\forall j, \zeta_{j,i}^* \geq 0; \|\xi_i^*\|_o \leq k$ (Solve using Algorithm 2)
- 12: $\Gamma = \Gamma + \mathbf{Y}(\xi_i - \xi_i^*)\alpha^i$
- 13: $\xi_i \leftarrow \xi_i^*$
- 14: **end for**
- 15: $\Phi = \mathbf{Y}\Xi$
- 16: **end for**

Output:

- 17: Endmember matrix Φ , Abundance matrix \mathbf{A} .
-

optimal δ_j^* can be directly computed by choosing the larger value among ε and $\|\mathbf{y}_j - \Phi\alpha_j\|_2$. This procedure is performed in line “3–6” of the algorithm. It is worth mentioning that although the computation is elementwise, it exactly emulates solving (7) for the complete matrix Δ (with fixed Ξ and \mathbf{A}) due to the special construction of Δ , which gives us a closed-form solution of the matrix.

To compute Ξ , we can write the optimization objective in (7) as

$$\begin{aligned} & \min_{\Xi} \|\mathbf{Y} - \mathbf{Y}\Xi\mathbf{A}\Delta^{-\frac{1}{2}}\|_F^2 \\ & \text{s.t. } \forall i, j, \zeta_{j,i} \geq 0; \quad \|\xi_i\|_o \leq k. \end{aligned} \quad (10)$$

Note that, we do not disregard Δ in (10). This is because we intend to reformulate the objective function by changing its input argument, which will be made clear shortly. In that case, rescaling of the new argument by Δ must also be taken into consideration for the correct reformulation.

Let us briefly ignore the outer constraints in (10). We can minimize the quadratic loss in the remaining objective by separately minimizing the cost incurred by each column of Ξ . This can be done by solving the following problem:

$$\forall i, \min_{\xi_i^*} \|\mathbf{Y}\Delta^{-\frac{1}{2}} - \mathbf{Y}(\Xi\mathbf{A} - \xi_i\alpha^i + \xi_i^*\alpha^i)\Delta^{-\frac{1}{2}}\|_F^2 \quad (11)$$

where $\alpha^i \in \mathbb{R}^{1 \times n}$ denotes the i th row of \mathbf{A} and ξ_i^* is an updated version of ξ_i that would minimize the loss. By changing the input argument from ξ_i to ξ_i^* , Ξ also becomes a

Algorithm 2 Nonnegative Subspace Pursuit**Input:** Sparsity level k , data $\boldsymbol{\psi}$, dictionary \mathbf{Y} .**Initialize:** $\mathcal{S}^0 = \{\text{Indices of the } k \text{ largest coefficients of } \frac{\mathbf{Y}^T \boldsymbol{\psi}}{\|\boldsymbol{\psi}\|_2}\}$,
 $\boldsymbol{\beta} = \operatorname{argmin}_{\boldsymbol{\beta}} \|\boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\beta}\|_2^2$ s.t. $\operatorname{supp}\{\boldsymbol{\beta}\} \subseteq \mathcal{S}^0; \forall z, \beta_z \geq 0$,
 $\mathbf{r}^0 = \boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\beta}$.**Iterations:**

- 1: **for** $i = 1$ to k **do**
 - 2: $\widehat{\mathcal{S}}^i = \mathcal{S}^{i-1} \cup \{\text{Indices of the } k \text{ largest coefficients of } \mathbf{Y}^T \mathbf{r}^{i-1} / \|\mathbf{r}^{i-1}\|_2\}$.
 - 3: $\boldsymbol{\beta} = \operatorname{argmin}_{\boldsymbol{\beta}} \|\boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\beta}\|_2^2$ s.t. $\operatorname{supp}\{\boldsymbol{\beta}\} \subseteq \widehat{\mathcal{S}}^i; \forall z, \beta_z \geq 0$.
 - 4: $\mathcal{S}^i = \{\text{Indices of the largest } k \text{ coefficients of } \boldsymbol{\beta}\}$.
 - 5: $\boldsymbol{\xi}^* = \operatorname{argmin}_{\boldsymbol{\xi}^*} \|\boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\xi}^*\|_2^2$ s.t. $\operatorname{supp}\{\boldsymbol{\xi}^*\} \subseteq \mathcal{S}^i; \forall z, \xi_z^* \geq 0$.
 - 6: $\mathbf{r}^i = \boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\xi}^*$.
 - 7: **if** $\|\mathbf{r}^i\|_2 = 0$ **or** $\|\mathbf{r}^i\|_2 \geq \|\mathbf{r}^{i-1}\|_2$ **then**
 - 8: **break**.
 - 9: **end if**
 - 10: **end for**
 - 11: **if** $\|\mathbf{r}^i\|_2 > \|\mathbf{r}^{i-1}\|_2$ **then**
 - 12: $\boldsymbol{\xi}^* = \operatorname{argmin}_{\boldsymbol{\xi}^*} \|\boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\xi}^*\|_2^2$ s.t. $\operatorname{supp}\{\boldsymbol{\xi}^*\} \subseteq \mathcal{S}^{i-1}; \forall z, \xi_z^* \geq 0$.
 - 13: **end if**
- Output:**
14: Sparse codes $\boldsymbol{\xi}^*$.

constant in (11). Exploiting the fixed matrices, we can further modify our optimization objective as

$$\forall i, \min_{\boldsymbol{\xi}_i^*} \left\| \left\{ \frac{(\mathbf{Y} - \mathbf{Y}\boldsymbol{\Xi}\mathbf{A})\boldsymbol{\Delta}^{-\frac{1}{2}}\boldsymbol{\alpha}^{i\top}}{\boldsymbol{\alpha}^i \boldsymbol{\Delta}^{-\frac{1}{2}}\boldsymbol{\alpha}^{i\top}} + \mathbf{Y}\boldsymbol{\xi}_i \right\} - \mathbf{Y}\boldsymbol{\xi}_i^* \right\|_2^2 \quad (12)$$

where the superscript “ \top ” signifies the transpose operation. Let us denote the expression within the braces by $\boldsymbol{\psi}$ to finally arrive at the following form of (10):

$$\forall i, \min_{\boldsymbol{\xi}_i^*} \|\boldsymbol{\psi} - \mathbf{Y}\boldsymbol{\xi}_i^*\|_2^2 \quad \text{s.t. } \forall j, \xi_{j,i}^* \geq 0; \|\boldsymbol{\xi}_i^*\|_0 \leq k. \quad (13)$$

Above is an objective function of a sparse optimization problem, with a nonnegativity constraint over the sparse codes $\boldsymbol{\xi}_i^*$. As mentioned in line “11” of the algorithm, we use Algorithm 2 to solve this problem. For the sake of continuity, we momentarily defer the discussion on Algorithm 2 to the next paragraph. The computation and the update procedure for the matrix $\boldsymbol{\Xi}$ are given in lines “7–14” of Algorithm 1. In line “7,” we compute the residue matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times n}$ outside the for-loop that iterates over the columns of $\boldsymbol{\Xi}$. In line “12,” we update $\boldsymbol{\Gamma}$ to account for the change in the residue caused by the newly computed $\boldsymbol{\xi}_i^*$, before updating $\boldsymbol{\xi}_i$ in line “13” of the algorithm.

It is possible to use or extend existing algorithms [22], [23], [41]–[44] to solve (13). These algorithms employ a common optimization strategy, known as the *greedy pursuit* in the sparse representation literature [34]. A detailed discussion of this strategy and the above referred algorithms can be found in our previous work [22]. In that work, a comprehensive analysis of these algorithms revealed that the SP [44] can solve (13) very efficiently, however, without accounting for the nonnegativity constraint. Hence, in this paper,

we extend SP to additionally incorporate the nonnegativity constraint and use it to solve (13). The nonnegative variant of SP has analogous computational advantages over the nonnegative variants of the existing algorithms. To the best of our knowledge, such an extension of SP has not been previously proposed. The nonnegative SP (NSP) is given in Algorithm 2. We note that a nonnegative variant of orthogonal matching pursuit algorithm (OMP) [23] has been proposed by Brukenstein *et al.* [66], which can also be used to solve (13). However, compared to that algorithm, NSP shows more robustness against the local optimality of the greedy pursuit strategy, just as SP is more robust than OMP in this regard.

In NSP, we iteratively identify a subspace of *at-most* k columns of \mathbf{Y} , such that the input signal $\boldsymbol{\psi}$ (or its best approximation) lies in the positive orthant of this subspace. Following the conventions of the sparse representation literature, below, we refer to \mathbf{Y} as the *dictionary* and to its columns as the *atoms*. To initialize the dictionary, we first identify the indices of k atoms, having the smallest angles with $\boldsymbol{\psi}$. These indices are recorded in a set \mathcal{S}^0 . Then, we compute the orthogonal projection of $\boldsymbol{\psi}$ onto the positive orthant of the atoms indexed in \mathcal{S}^0 . In Algorithm 2, the operator “ $\operatorname{supp}\{\cdot\}$ ” used for this purpose, indicates the *support*, i.e., the indices of the nonzero coefficients, of a vector. We then compute the residue vector \mathbf{r}^0 by subtracting from $\boldsymbol{\psi}$ its computed projection (i.e., $\mathbf{Y}\boldsymbol{\beta}$).

The algorithm performs at most k iterations. In the i th iteration, it augments the set \mathcal{S}^{i-1} with k more indices to compute a set $\widehat{\mathcal{S}}^i$. The newly added indices correspond to the atoms subtending the smallest angles with the residue vector \mathbf{r}^{i-1} . Using $\widehat{\mathcal{S}}^i$, the k -best atoms are identified in \mathcal{S}^i . These atoms maximally contribute to the nonnegative least squares approximation of $\boldsymbol{\psi}$, when only the atoms in $\widehat{\mathcal{S}}^i$ are used as the basis. This procedure is stated in lines “3–4” of the algorithm. Next, $\boldsymbol{\psi}$ is approximated in the positive orthant of the column space of the atoms indexed in \mathcal{S}^i . This results in a potential solution $\boldsymbol{\xi}^*$, that is used to compute the new residue vector \mathbf{r}^i (lines “5–6”). If the new residue is zero or it does not improve in the current iteration, the iterations stop and $\boldsymbol{\xi}^*$ becomes the solution. However, if the residue increases as a result of the i th iteration, \mathcal{S}^{i-1} is used to recompute the sparse codes $\boldsymbol{\xi}^*$, as mentioned in line “11–13” of the algorithm.

We emphasize one important characteristic of NSP. That is, it reconstructs $\boldsymbol{\psi}$ using a nonnegative linear combination of “at-most” k pixels indexed in \mathcal{S} . Bruckstein *et al.* [66] showed that the nonnegativity in such reconstructions naturally leads to the sparsest combinations of the signals. For NSP, this means that the sparsity of $\boldsymbol{\xi}^*$ is upper bounded by k . If $\boldsymbol{\psi}$ can be reconstructed with fewer pixels, the algorithm automatically uses that sparsity level, to the end that it uses only one pixel if it belongs to a pure endmember. Thus, the pixels considered as pure endmembers are explicitly recorded by $\boldsymbol{\Xi}$ in our approach. It is easy to see that NSP can potentially compute the exact same solutions under different values of k , if these values are above a certain threshold. This is because the nonnegativity imposes sparsity by forcing the unrequired coefficients of $\boldsymbol{\xi}^*$ to zero [66]. However, for a larger k , a larger subspace would have to be parsed for computing the solution,

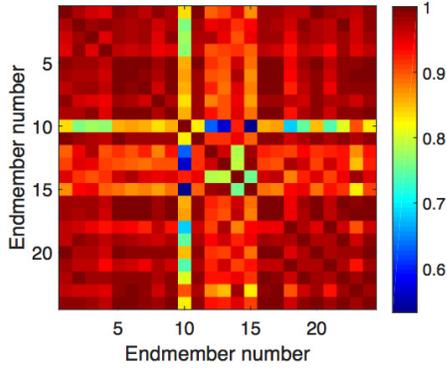


Fig. 2. Coherence matrix for \mathcal{P}_{25} . Each coefficient of the matrix shows the value of an absolute inner product between two normalized endmembers. The diagonal entries are self-dot products.

that can result in more computations. Therefore, smaller k values are preferred for our approach.

It is worth mentioning that Algorithm 2 differs from SP [44] in two core operations, namely: 1) identifying the sets \mathcal{S}^0 and $\widehat{\mathcal{S}}^i$ and 2) computing the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\xi}^*$. In SP, \mathcal{S}^0 records the indices of the k coefficients of the vector $\mathbf{s}_0 = \mathbf{Y}\boldsymbol{\psi}$ that have the largest *magnitudes*. Similarly, $\widehat{\mathcal{S}}^i$ is computed by augmenting \mathcal{S}^{i-1} with the indices of largest *magnitude* entries in $\mathbf{s}_i = \mathbf{Y}\mathbf{r}^{i-1}$. Such constructions of the sets are governed by the fact that SP also allows negative coefficients in $\boldsymbol{\beta}$ (and $\boldsymbol{\xi}^*$). We place a nonnegativity constraint over $\boldsymbol{\beta}$ (and $\boldsymbol{\xi}^*$). Therefore, we allow only those atoms in sets \mathcal{S}^0 and $\widehat{\mathcal{S}}^i$ that have positive correlations with $\boldsymbol{\psi}$ and \mathbf{r}^{i-1} , respectively.

IV. EXPERIMENTS WITH SYNTHETIC DATA

To quantitatively analyze the performance of our approach we first experiment with synthetically mixed endmembers.

A. Data

To generate the synthetic data, we used the NASA Jet Propulsion Laboratory's Advanced Space-borne Thermal Emission and Reflectance Radiometer (ASTER) library (<http://speclib.jpl.nasa.gov>). We selected a set of 25 spectra from the library that were considered as pure endmembers in our experiments. Henceforth, we denote this set as \mathcal{P}_{25} . Details of the materials corresponding to \mathcal{P}_{25} are provided in Table IV in the Appendix. Table IV contains information regarding the type, class, subclass, and particle size of each material. We randomly selected the endmembers in \mathcal{P}_{25} such that their mutual coherence [33] was higher than 0.9995. Mutual coherence ($0 \leq \mu \leq 1$) is defined as the maximum absolute inner product between any two distinct normalized endmembers in the set. It is well established that the accuracy of matrix factorization approaches can be adversely affected by the large mutual coherence of the sources generating the data [22], [33]. For \mathcal{P}_{25} , the value of $\mu = 0.9998$. However, it should be noted that μ is only a worst case measure and the mean absolute inner product for each possible pair of normalized endmembers in \mathcal{P}_{25} is 0.9428. Fig. 2 provides an image of the coherence matrix for the used endmembers. Each element of the matrix shows the absolute inner product of

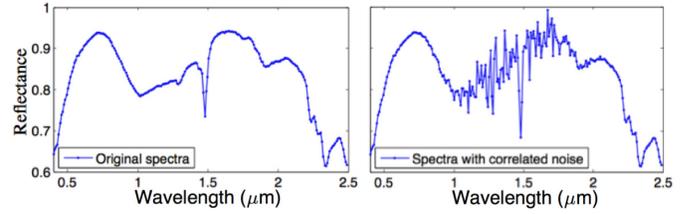


Fig. 3. Illustration of additive correlated noise at SNR = 30 dB, with $\eta = 18$.

one endmember with another. The diagonal entries are the self-dot products. Whereas high mutual coherence of the used endmembers makes our experiments challenging, there generally remains a discernible difference between the used endmembers that can be distinguished from the modeling error. To follow a general experimental protocol [16], [22], the ASTER library spectra were used after resampling at the sampling wavelengths of the NASA's AVIRIS sensor [45].

In a single experiment, we simulated a synthetic hyperspectral image with 10000 mixed pixels, using a set $\mathcal{P}_{20} \subset \mathcal{P}_{25}$ that contained 20 randomly chosen endmembers. The final results were computed by averaging the performance on 10 images. Each pixel $\mathbf{y}_j \in \mathbb{R}^m$ was constructed by mixing $2 \leq p \leq 5$ pure endmembers such that their fractional abundances followed a Dirichlet distribution. These settings are motivated by the common knowledge that the number of materials in a typical remote sensing scene is 20 or less [16], and that the number of mixed materials for a pixel is usually of the order of 4 to 5 [36]. Existence of a pure pixel gives our approach an additional advantage, as it then directly selects that pixel as an endmember. Nevertheless, pure pixels are not always present in real data. Therefore, we evaluate our approach for a more challenging scenario and keep $p > 1$ to ensure the absence of pure endmember pixels. To simulate the outliers, we corrupted 3% of the pixels of each image by randomly saturating 50% of their channels. This is done by replacing the reflectance values by 1 at those channels.

We have considered spectral mixing with both white and correlated additive noises. To simulate the white noise we used the MATLAB's inbuilt `awgn` function, with measured signal power. We followed Bioucas-Dias and Nascimento [46] to add the correlated noise to the data. We defined the diagonal noise correlation matrix such that its entries formed a Gaussian shape. A diagonal entry (σ_h^2) of the matrix was computed as

$$\sigma_h^2 = \frac{\sum_{j=1}^n \|\mathbf{y}_j\|_2^2}{n \times 10 \left(\frac{\text{SNR}}{10}\right) \eta \sqrt{2\pi}} \exp\left(-\frac{(h-\frac{n}{2})^2}{2\eta^2}\right) \quad \forall h \in \{1, \dots, m\} \quad (14)$$

where η controls the variance of the bell curve. In our experiments we fixed $\eta = 18$. In Fig. 3, we illustrate the correlated noise added to a mixed pixel. For further theoretical details on generation of the correlated noise, we refer to [46].

B. Evaluation Metrics

To evaluate the performance, we compute the *average spectral angle* (θ_{avg}) in \mathbb{R}^m between our estimates of the endmembers and the actual endmembers used to generate the synthetic data. Suppose we used a set $\mathcal{P} = \{\boldsymbol{\varphi}_q^* | \boldsymbol{\varphi}_q^* \text{ is a pure endmember}\}$ to generate a synthetic image,

then θ_{avg} is defined as

$$\theta_{\text{avg}} = \frac{180}{\pi |\mathcal{P}|} \sum_{q=1}^{|\mathcal{P}|} \arccos \frac{\boldsymbol{\varphi}_q^T \boldsymbol{\varphi}_q^*}{\|\boldsymbol{\varphi}_q\|_2 \|\boldsymbol{\varphi}_q^*\|_2} \quad (15)$$

where $|\cdot|$ denotes the cardinality of the set and $\boldsymbol{\varphi}_q$ is the estimated endmember, that best matches³ the q th true endmember. Notice that the computed angle is in degrees.

The metric θ_{avg} only evaluates the soundness of the extracted endmembers. To evaluate the estimated fractional abundances, we used the root mean squared error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{\|\mathbf{A} - \mathbf{A}^*\|_F^2}{|\mathcal{P}| \times n}} \quad (16)$$

where $\mathbf{A}^* \in \mathbb{R}^{|\mathcal{P}| \times n}$ is the actual fractional abundance matrix, used to generate the image. We compute the RMSE after matching the computed endmembers with the actual endmembers and re-arranging the computed matrix \mathbf{A} accordingly.

C. Benchmarking

To benchmark, we mainly compare our approach with the popular constrained matrix factorization methods. We also report the results of commonly used endmember extraction algorithms, combined with abundance estimation methods. We used the author-provided implementations for all the approaches. Among the sparse matrix factorization techniques, we compare our results with K-SVD [60] and the Online Dictionary Learning (ODL) [61]. Note that, the sparsity constrained matrix factorization has already established its effectiveness in hyperspectral unmixing [5], [55], [58], [64], and K-SVD and ODL are the state-of-the-art approaches. We additionally compare our results with a variant of ODL that computes nonnegative endmembers and fractional abundances [62]. We refer to this approach as ODL-NN. We also include the archetypal analysis-based matrix factorization [65] in our comparisons. Similar to our approach, the archetypal analysis constructs the endmembers as nonnegative linear combinations of the observed pixels. However, unlike our approach, it also assumes an endmember to lie in a simplex whose vertices are formed by the observed pixels. This assumption is restrictive and it is generally not followed by hyperspectral data. Another major difference between Archetypal Analysis and our approach is that the latter provides an explicit control over the maximum number of pixels to be used in constructing an endmember. In the presence of a large number of pixels, this can result in considerable computational advantage over conventional Archetypal Analysis. In our experiments, we used the archetypal analysis enhancement in [63], which also accounts for outliers. To explicate the effect of robustness against outliers in our approach, we also provide the results of the nonrobust variant of our approach, denoted as CMF. This variant is implemented

³Matching is done by minimizing the angles between the estimated and the true endmembers, such that, for each true endmember, there is only one match and each estimated endmember is used only once in the process.

TABLE I

AVERAGE SPECTRAL ANGLE (θ_{avg}): FOR NOISY CASES, SNR IS 30 dB

Method	White noise	Correlated noise	Noise free
VCA [11]	10.13 ± 1.22	9.25 ± 0.58	—
AVMAX [15]	10.35 ± 1.50	9.34 ± 0.57	—
SVMAX [15]	10.55 ± 1.63	8.99 ± 0.79	3.54 ± 0.21
K-SVD[60]	12.85 ± 0.25	13.86 ± 0.70	7.40 ± 1.39
ODL[61]	10.65 ± 0.33	8.79 ± 0.58	7.70 ± 0.51
ODL-NN[62]	10.63 ± 0.29	8.79 ± 0.58	7.70 ± 0.51
Archetypal[63]	5.49 ± 0.36	4.95 ± 0.33	4.24 ± 0.48
CMF	5.35 ± 0.20	5.11 ± 0.25	4.11 ± 0.21
RCMF	5.19 ± 0.22	4.55 ± 0.59	3.97 ± 0.22

by removing the lines “3–6” and the matrix \mathbf{A} from line “9” in Algorithm 1.

Among the endmember extraction techniques, we compare the performance of our approach with the commonly used methods, known as vector component analysis (VCA) [11], successive volume maximization (SVMAX) [15], and its variant, alternating volume maximization (AVMAX) [15]. The endmembers extracted by these approaches were used with abundance estimation methods to perform the unmixing. We used the SUnSAL implementation [25] (as used by our approach) for computing the abundances. For VCA and AVMAX, we set the implementation to solve an FCLS problem. For SVMAX, we used the constrained version of SUnSAL that additionally imposes sparsity on the abundance vectors along ANC and ASC. In our experiments, these combinations resulted in the best performance of the endmember extraction-based unmixing.

D. Results

We summarize the results of our experiments in Tables I and II. In Table I, we report the *average spectral angles* between the ground truth and the computed endmembers. These values were computed when the data contained white noise, correlated noise and no noise at all. For the noisy cases, we used signal-to-noise ratio (SNR) = 30 dB. The corresponding RMSEs of the computed fractional abundance matrices are reported in Table II. In our experiments, all the approaches required the value of the total number of endmembers “ K ” as an input. Matching the cardinality of the set \mathcal{P}_{25} , we chose $K = 25$ for each technique. All the remaining parameters of the approaches were carefully optimized on a separate cross-validation data set. For our approach, we used the sparsity level $k = 5$ and $\varepsilon = 10^{-10}$. For K-SVD, we used 10 as the sparsity threshold. The regularization constant in ODL and ODL-NN was fixed to 10^{-6} . We refer to the original works for details on the significance of these parameters.

From Table I, we can see that the proposed RCMF is able to recover the endmembers very accurately, both in the presence and absence of noise. Compared to RCMF, its nonrobust variant CMF generally underperforms in these settings. Nevertheless, its performance remains acceptable. When there is no noise AVMAX [15] and VCA [11] are unable to converge. This happens because of absence of pure pixels in the data. On the other hand, SVMAX [15] is able to recover the endmembers with an accuracy higher than

TABLE II
RMSE OF RECONSTRUCTED ABUNDANCE MATRICES: FOR NOISY CASES, SNR IS 30 dB

Method	White noise	Correlated noise	Noise free
VCA [11]	$0.130 \pm 8.8 \times 10^{-3}$	$0.128 \pm 4.08 \times 10^{-3}$	—
AVMAX [15]	$0.129 \pm 4.3 \times 10^{-3}$	$0.127 \pm 6.53 \times 10^{-3}$	—
SVMAX [15]	$0.141 \pm 1.1 \times 10^{-2}$	$0.132 \pm 7.9 \times 10^{-3}$	$0.088 \pm 8.08 \times 10^{-3}$
K-SVD[60]	2.075 ± 0.16	2.388 ± 1.419	6.06 ± 2.63
ODL[61]	1.506 ± 0.16	1.786 ± 0.200	88.98 ± 70.23
ODL-NN[62]	1.059 ± 0.07	1.261 ± 1.159	1.49 ± 0.05
Archetypal[63]	$0.103 \pm 2.2 \times 10^{-3}$	$0.103 \pm 6.43 \times 10^{-3}$	$0.105 \pm 4.00 \times 10^{-3}$
CMF	$0.101 \pm 0.8 \times 10^{-3}$	0.100 ± 0.010	$0.106 \pm 2.34 \times 10^{-3}$
RCMF	$0.097 \pm 3.5 \times 10^{-5}$	$0.095 \pm 6.6 \times 10^{-3}$	$0.104 \pm 1.10 \times 10^{-2}$

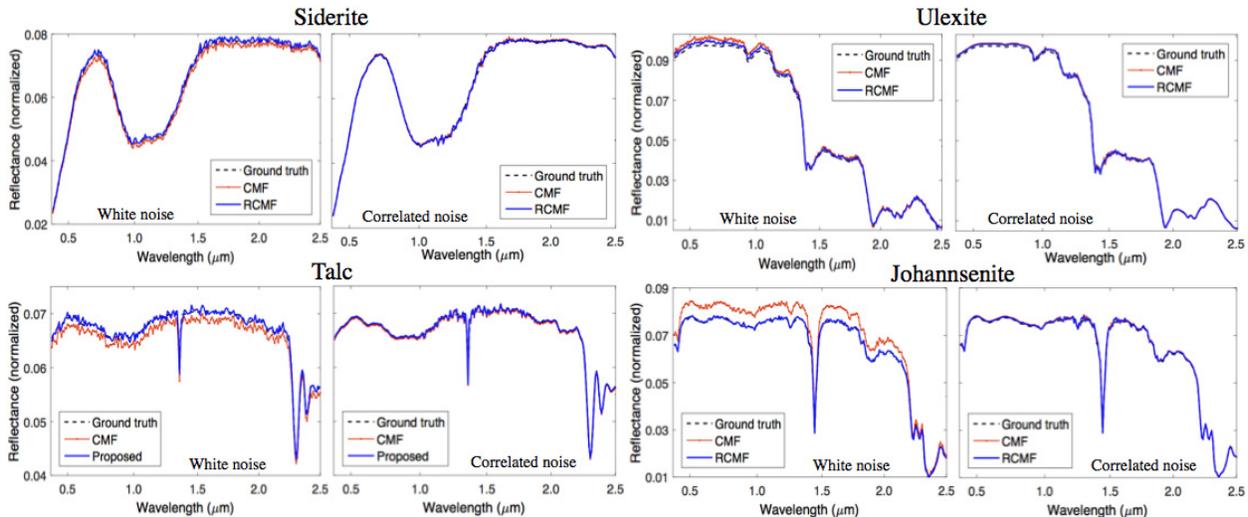


Fig. 4. Examples of the endmembers recovered by the proposed approach at 35 dB SNR.

RCMF. However, it happens only for the noise-free case. For practical settings, RCMF outperforms SVMAX with a significant margin. In Fig. 4, we show a few representative examples of the endmembers recovered by RCMF and CMF along the ground truth.

Due to the better recovery of the endmembers, RCMF is also able to compute their fractional abundances accurately. This is evident from Table II. While computing the RMSEs, we disregarded the abundances related to the pixels representing the outliers. Therefore, the difference between the results of CMF and RCMF is not significant. Nevertheless, RCMF still outperforms CMF because of its better endmember recovery. One interesting observation in Table II is that the additional nonnegativity constraint in ODL proved to be particularly effective in computing the abundances. Whereas the endmembers recovered by K-SVD and ODL were generally observed to be nonnegative in our experiments, this was not the case for their computed abundances, unless the nonnegativity was explicitly imposed. Due to better recovery of the endmembers by SVMAX [15] in the noise-free settings, the constrained version of SUnSAL [25] resulted in the best abundance estimation when used with SVMAX in the absence of noise.

From the tables it may appear that the proposed approach achieves only a slight improvement over Archetypal Analysis. However, this is not true in general. Whereas Archetypal Analysis shows comparable performance with our approach

for the used experimental setup, there are also scenarios where it significantly underperforms due to its restrictive model. To demonstrate that we conducted further experiments in which image pixels were constructed using 2 to 3 endmembers, with Dirichlet samples for abundances. However, in all the pixels, the abundance of any endmember was upper bounded by 0.8. We did this to suppress dominance of a single endmember in any pixel, which is a practical possibility. We used a high SNR of 100 dB, which is also a practical value for modern remote sensing hyperspectral instruments. As can be imagined, under these conditions the actual endmembers would generally lie well outside the simplex of the pixels because of the absence of strongly dominant (and pure) pixels in the data, and virtually nonexistent additive noise.

Even in the presence of 3% outliers for such data the θ_{avg} for RCMF was 4.69 ± 0.27 , whereas this value was 5.47 ± 0.48 for the robust archetypal analysis. The RMSE values were 0.098 ± 0.0074 and 0.121 ± 0.0077 for RCMF and the archetypal analysis, respectively. In the absence of outliers, the θ_{avg} for CMF and archetypal analysis were 4.00 ± 0.33 and 5.02 ± 0.35 respectively. Similarly, the RMSE of abundances were 0.096 ± 0.0076 and 0.108 ± 0.0087 for CMF and the archetypal analysis. For the later cases we did not use the robust version of the archetypal analysis for a fair comparison with CMF, as robust archetypal analysis resulted in a worse performance. The differences between the performances of

TABLE III
MEAN COMPUTATION TIME IN SECONDS, FOR UNMIXING 1000 PIXELS

K-SVD [60]	ODL [61]	ODL-NN [62]	Archetypal [63]	VCA [11]	AVMAX [15]	SVMAX [15]	CMF	RCMF
18.09	18.76	18.81	17.82	3.27	3.18	3.20	18.13	17.67

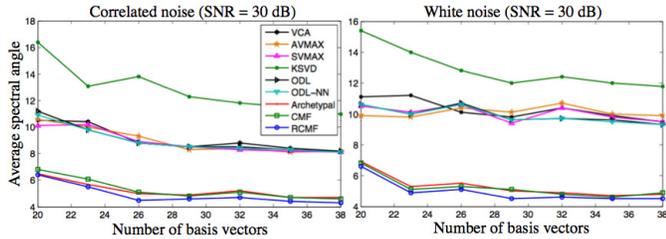


Fig. 5. Average spectral angle " θ_{avg} " as a function of the number of extracted endmembers " K ." The images are created using 20 pure spectra.

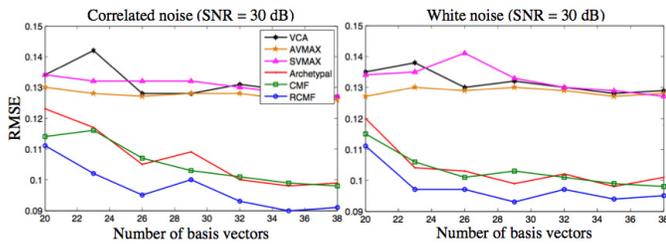


Fig. 6. RMSE as a function of the number of extracted endmembers " K ." The images are created with 20 pure spectra.

the proposed approach and archetypal analysis are significant in these experiments, which highlights better suitability of our approach for unmixing compared with the conventional archetypal analysis.

Generally, the total number of endmembers present in a scene is not known *a priori*. Therefore, we also evaluated the performance of approaches by varying the number endmembers allowed to be extracted by them. That is, by changing the number of the basis vectors allowed to be learned in the matrix factorization procedure. In Fig. 5, we plot θ_{avg} against this number. It can be seen that the approaches are generally able to extract the endmembers better by allowing more basis vectors. The value of θ_{avg} remains very small for our approach throughout the plots. Note that, we used 20 endmembers to generate a synthetic image. The combination of endmember extraction algorithm and abundance estimation algorithms performed somewhat similar to ODL [61] and ODL-NN, which use ℓ_1 -sparsity constraint. The legend only shows the names of the endmember extraction algorithms. Since the endmembers computed by ODL were generally nonnegative, ODL-NN did not result in improved θ_{avg} . The performance of archetypal analysis [63] is similar to that of our approach due to the similarities in the underlying assumptions of the algorithms.

The RMSE values of the fractional abundances for the above experiments are plotted in Fig. 6. For clarity, we do not plot the results of ODL, ODL-NN, and KSVD in the figure because they were not comparable to the other results, as can be verified from Table II. From Fig. 6, a consistent performance of the proposed approach is evident.

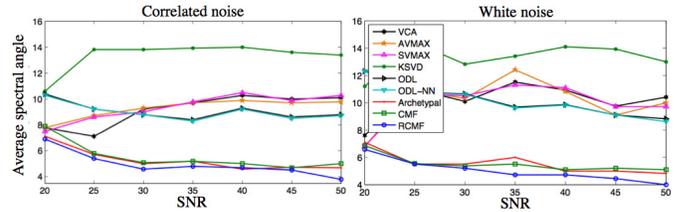


Fig. 7. Average spectral angle " θ_{avg} " as a function of SNR. In the experiments, $K = 25$.

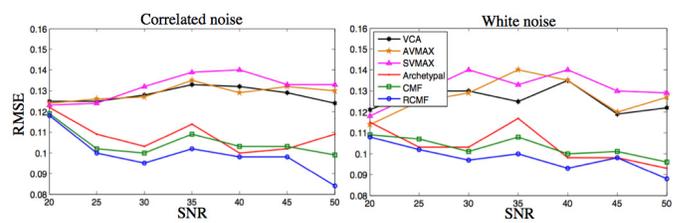


Fig. 8. RMSE as a function of SNR. In the experiments, $K = 25$.

In Figs. 7 and 8 we plot the θ_{avg} and RMSE values for the approaches against different levels of noise in the data. We extracted 25 basis vectors with each approach in these experiments. The plots clearly show that the proposed approach performs reasonably well even for very low SNR, and the performance generally improves with higher SNR. In Fig. 8, the RMSE values for KSVD, ODL, and ODL-NN are not included because they were not comparable to the other results. Despite the good results of SVMAX in Tables I and II, its performance did not improve too much at high SNR values. This happened due to the presence of outliers, which were not considered in the noise-free settings of Tables I and II.

In all the experiments, we used 300 iterations of K-SVD, ODL, and ODL-NN. For archetypal analysis, we used 200 iterations. Our algorithms generally required less than 100 iterations for convergence. Nevertheless, we used $Q = 100$ iterations in our experiments. The total number of iterations for each algorithm was decided with the help of cross-validation data. The mean computation time for all the experiments in Tables I and II is reported in Table III. The time is computed on a desktop computer with Intel Core i7-2600 CPU at 3.4GHz and 8 GB RAM. In the table, the mean computation time of CMF exceeds RCMF because the latter often reaches the breaking condition of NSP in line "7" of Algorithm 2 in fewer iterations. In our opinion, this is a result of considering robustness against the outliers in RCMF. For the endmember extraction algorithms, the time is provided for the complete process of extracting the endmembers and computing the abundances. Compared with the proposed approach,

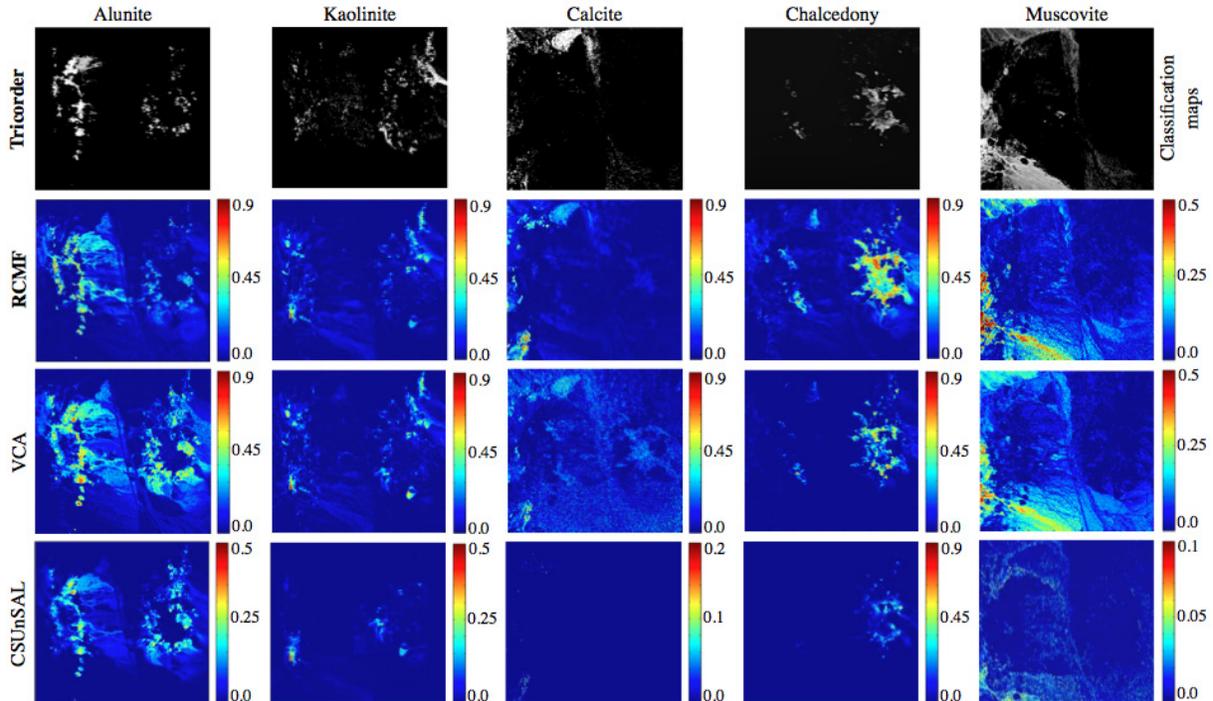


Fig. 9. Fractional abundance analysis of the AVIRIS data. (First row) Classification maps by the Tricorder software. (Second row) Results of RCMF. (Third row) Result of VCA endmembers used for solving an FCLS problem. (Fourth row) CSUnSAL [25] results with ASTER library spectra.

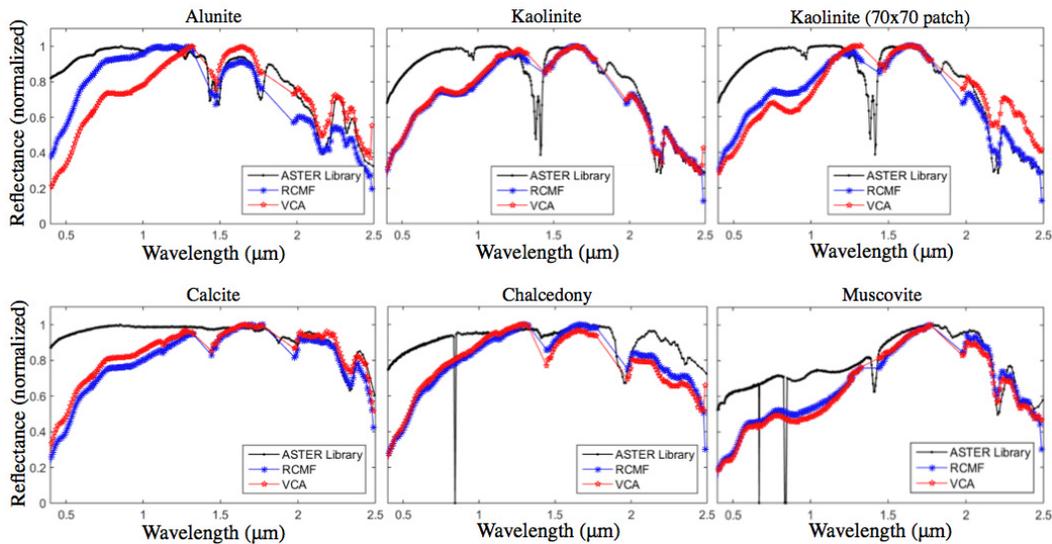


Fig. 10. Extracted endmembers from the AVIRIS image. Spectra from the ASTER Library are also provided for reference. For kaolinite, an additional plot is given that was computed using only the top right 70×70 patch of the image. Some spectral distortions are observable in this plot by VCA.

these timings are better. Nevertheless, our approach does not assume presence of pure endmembers in the image and performs significantly better than these algorithms in practical conditions.

In Table IV, we provide details of the minerals used to create the synthetic data in our experiments in Section IV.

V. EXPERIMENTS WITH REAL DATA

We analyze the NASA's AVIRIS Cuprite data (http://aviris.jpl.nasa.gov/data/free_data.html) with our

approach to establish its effectiveness on a real-world unmixing problem. Since the quantitative evaluation of the results is not possible because of the unavailability of the ground truth, we present our results for qualitative analysis, following [16], [22]. The said data are collected by the AVIRIS sensor [45] over a region of Cuprite mines in Nevada. The region is well studied for its geological properties that makes its hyperspectral image a suitable benchmark. The analyzed image is a $512 \times 512 \times 224$ cube, acquired in the wavelength range 370–2500 nm. To process

TABLE IV

DETAILS OF THE MATERIALS CORRESPONDING TO EACH SPECTRA IN \mathcal{P}_{25} . THESE DETAILS ARE DIRECTLY TAKEN FROM THE ASTER LIBRARY

Sr. #	Type	Class	Subclass	Particle size	Sr. #	Type	Class	Subclass	Particle size
1	Mineral	Arsenate	None	Coarse	14	Mineral	Silicate	Inosilicate	Fine
2	Mineral	Carbonate	None	Fine	15	Mineral	Silicate	Phyllosilicate	Coarse
3	Mineral	Oxide	None	Coarse	16	Mineral	Silicate	Phyllosilicate	Medium
4	Mineral	Silicate	Inosilicate	Coarse	17	Mineral	Silicate	Tectosilicate	Fine
5	Mineral	Silicate	Nesosilicate	Fine	18	Mineral	Sulfate	None	Medium
6	Mineral	Silicate	Phyllosilicate	Fine	19	Soil	Inceptisol	Haplumbrept	Coarse
7	Mineral	Silicate	Tectosilicate	Coarse	20	Rock	Igneous	Intermediate	Solid
8	Mineral	Silicate	Tectosilicate	Medium	21	Rock	Sedimentary	Sandstone	Solid
9	Mineral	Sulfide	None	Coarse	22	Rock	Igneous	Felsic	Solid
10	Mineral	Borate	None	Medium	23	Rock	Igneous	Ultramafic	Fine
11	Mineral	Element	None	Coarse	24	Rock	Metamorphic	Marble	Coarse
12	Mineral	Oxide	None	Medium	25	Rock	Sedimentary	Limestone	Fine
13	Mineral	Halide	None	Medium					

the cube, we first removed its 36 channels corresponding to the wavelengths 370, 380, 1330 to 1430, 1780 to 1970, 2490, and 2500 nm. This is a common protocol [16], [22] to avoid the low SNR and water absorption bands in the analysis. Then, we selected a $350 \times 350 \times 188$ sub-cube, where discernible spatial patterns of multiple minerals were present.

In Fig. 9, we display the fractional abundance maps computed by the proposed RCMF (second row) for the selected subcube. The maps are of five different minerals that were commonly analyzed by recent unmixing approaches [16], [22]. These minerals show clear spatial patterns in the analyzed region. For reference, we also provide the mineral classification maps in the first row of the figure. These maps were computed by the USGS Tricorder algorithm⁴ in year 1995. Although, they were computed two years prior to the acquisition of the analyzed hyperspectral image, these maps provide a good reference for the abundance maps because we can expect high abundance values at the pixels classified by these maps as the pure minerals [22]. We can also expect high proportions of the same minerals in the nearby regions of the pure pixels. It is clear from the figure that RCMF generally assigns large fractional abundances to the correct regions. The spatial patterns of the computed abundances clearly match the classification maps. For comparison, we also provide the abundance maps computed by solving an FCLS problem using the endmembers extracted by VCA [11]. These maps are shown in the third row of the figure. Very similar maps resulted when we used the CSUnSAL algorithm [25] for computing the abundances with VCA endmembers. Those maps are not included to avoid redundancy. In the fourth row, we provide the sparse unmixing results of CSUnSAL [25], when ASTER library spectra were used as endmembers. Note that, sparse unmixing is the state-of-the-art *supervised* unmixing framework and the accuracy of CSUnSAL is well established for this framework [16], [22], [31]. Nevertheless, the spatial patterns of the abundance maps computed by RCMF visually appear better than those computed by CSUnSAL. Interestingly, the maps resulting from VCA endmembers are very close to those computed by our approach. We observed similar resemblance

in the maps of other minerals as well for the two approaches. This happens because our approach also solves a fully constraint least squares problem in line 2 of Algorithm 1 to compute the abundances. When RCMF learns the endmembers that are similar to those extracted by VCA, their abundances computed (simultaneously) by RCMF naturally resemble to those computed by the abundance estimation methods used in conjunction with VCA (or other endmember extraction algorithms).

In Fig. 10, we show the spectra of the minerals recovered by RCMF and VCA. For reference, we also plot the spectra from the ASTER library. The spectra are shown as normalized reflectance values plotted against the wavelengths. Note that, the ASTER library spectra are only provided as a reference and should not be considered as the ground truth because they are measured in laboratory conditions, whereas the spectra computed by RCMF and VCA were measured in real-world settings. From Fig. 10, we can see that the spectra learned by RCMF are generally close to the endmembers extracted by VCA. However, it should be noted that VCA assumes availability of pure pixels in the image whereas our approach does not make this assumption. To note the effects of this assumption we also plot the computed spectra of kaolinite when only the top-right 70×70 patch of the image was used in the experiment. According to the mineral map computed by Tetracorder [69], this patch contains kaolinite as a mixture with alunite and/or other minerals (and it is possibly present in disordered form). The computed spectra are also shown in Fig. 10 for VCA and RCMF. It can be seen that the proposed approach is able to preserve the spectra much better than VCA when only the said patch is used instead of the complete image. This happens because, in contrast to VCA, our approach does not assume availability of pure pixels of each endmember in every image.

Comparing the spectra computed by RCMF and VCA with those available in the ASTER library, we can say that both the algorithms are generally able to preserve the important features of the spectra accurately. There is a consistent difference between the computed spectra and the laboratory measured spectra for the smaller wavelengths in each plot because the channels of the AVIRIS image corresponding to these wavelengths have lower intensity values. The parameter

⁴<http://speclab.cr.usgs.gov/PAPERS/tricorder.1995/tricorder.1995.html>

settings for RCMF and VCA for this experiment is the same as in Section IV-D. For CSUnSAL, we used the optimized parameter values and a dictionary created by the ASTER library. We refer to our previous work [22] for details of the parameter values and the dictionary formation.

VI. CONCLUSION

We proposed a novel matrix factorization approach for linear hyperspectral unmixing. In addition to accounting for the nonnegativity of the endmembers and the physical constraints over their abundances, the proposed approach forces the extracted endmembers to be sparse nonnegative combinations of the observed pixels. The association between the pixels and the endmembers is explicitly noted by our approach. Moreover, our approach also incorporates robustness against any possible outliers among the pixels. That makes our results more reliable. We systematically designed an efficient matrix factorization algorithm for our approach. Our experiments with synthetic hyperspectral images corrupted by white and correlated noise quantitatively establish the effectiveness of our approach. The proposed approach also shows promising qualitative results on the real hyperspectral data.

REFERENCES

- [1] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3631–3640.
- [2] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3586–3594.
- [3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [4] J. M. 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.
- [5] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 219–228, Jan. 2012.
- [6] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, vol. 1, 1993, pp. 11–14.
- [7] T. H. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.
- [8] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.
- [9] A. Plaza and C. I. Chang, "Impact of initialization on design of endmember extraction algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3397–3407, Nov. 2006.
- [10] C. I. Chang, C. C. Wu, W. Liu, and Y. C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.
- [11] J. M. P. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [12] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE's Int. Symp. Opt. Sci., Eng., Instrum.*, Oct. 1999, pp. 266–275.
- [13] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.
- [14] R. A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *Proc. 21st Can. Symp. Remote Sens.*, Jan. 1999, pp. 21–24.
- [15] T. H. Chan, W. K. Ambikapathi, and C. Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.
- [16] 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.
- [17] N. Akhtar, F. Shafait, and A. Mian, "Repeated constrained sparse coding with partial dictionaries for hyperspectral unmixing," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2014, pp. 953–960.
- [18] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernt, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.
- [19] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. Geosci. Remote Sens. Symp.*, Jul. 2008, pp. III-250–III-253.
- [20] A. Zare and P. Gader, "Sparsity promoting iterated constrained end-member detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.
- [21] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Grenoble, France, Aug. 2009, pp. 1–4.
- [22] N. Akhtar, F. Shafait, and A. Mian, "Futuristic greedy approach to sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2157–2174, Apr. 2015.
- [23] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
- [24] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.
- [25] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd WHISPERS*, Jun. 2010, pp. 1–4.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] N. Akhtar, F. Shafait, and A. Mian, "SUNGP: A greedy sparse approximation algorithm for hyperspectral unmixing," in *Proc. 22nd Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2014, pp. 3726–3731.
- [30] M.-D. Iordache, J. M. 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.
- [31] 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.
- [32] Y. Xu, F. Faming, and G. Zhang, "Similarity-guided and ℓ_p -regularized sparse unmixing of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2311–2315, Nov. 2015.
- [33] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [34] M. Elad, *Sparse Redundant Representation: From Theory to Application Signal Image Processing*. New York, NY, USA: Springer-Verlag, 2010.
- [35] J. P. Huber, "Robust estimation of a location parameter," *Ann. Statist.*, vol. 53, no. 1, pp. 73–101, 1964.
- [36] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [37] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
- [38] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 2006.
- [39] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, Apr. 1992.
- [40] R. Andersen, "Modern methods for robust regression," in *Sage Uni Paper Series on Quantitative Applications in Social Sciences*. Los Angeles, CA, USA: SAGE Publications, 2008, pp. 152–607.
- [41] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

- [42] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
- [43] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.
- [44] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [45] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
- [46] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [47] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.
- [48] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, Jul. 1991.
- [49] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.
- [50] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Online nonnegative matrix factorization with robust stochastic approximation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1087–1099, Jul. 2012.
- [51] B. Gao, W. L. Woo, and S. S. Dlay, "Variational regularized 2-D nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 703–716, May 2012.
- [52] W. Ren, G. Li, D. Tu, and L. Jia, "Nonnegative matrix factorization with regularizations," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 153–164, Mar. 2014.
- [53] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.
- [54] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
- [55] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.
- [56] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [57] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2746–2758, May 2014.
- [58] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.
- [59] Y. Yuan, M. Fu, and X. Lu, "Substance dependence constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2975–2986, Jun. 2015.
- [60] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing over-complete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [61] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary Learning for sparse coding," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 689–696.
- [62] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
- [63] Y. Chen, J. Mairal, and Z. Harchaoui, "Fast and robust archetypal analysis for representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1478–1485.
- [64] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 963–978, Sep. 2011.
- [65] A. Cutler and L. Breiman, "Archetypal analysis," *Technometrics*, vol. 36, no. 4, pp. 338–347, 1994.
- [66] A. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Nov. 2008.
- [67] A. Ambikapathi, T. H. Chan, C. Y. Chi, and K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using p-norm-based pure pixel identification algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2753–2769, May 2013.
- [68] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [69] G. A. Swayze *et al.*, "Mapping advanced argillic alteration at cuprite, Nevada, using imaging spectroscopy," *Econ. Geol.*, vol. 109, no. 5, pp. 1179–1221, 2014.



Naveed Akhtar received the B.Eng. degree (with distinction) in avionics from the College of Aeronautical Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2007, the M.Sc. degree (with distinction) in autonomous systems from Hochschule Bonn-Rhein-Sieg (HBRS), Sankt Augustin, Germany, in 2012, and the Ph.D. degree in computer vision from The University of Western Australia, Perth, WA, Australia, in 2017, under the supervision of Prof. A. Mian.

He is a Research Fellow at the College of Engineering and Computer Science, Australian National University, Canberra, ACT, Australia. He was a Research Assistant at the Research Institute for Microwaves and Millimeter-waves Studies, NUST, from 2007 to 2009, and a Research Associate at the Department of Computer Science, HBRS, in 2012. His research interests include hyperspectral image analysis, with an emphasis on sparse representation-based techniques.



Ajmal Mian received the Ph.D. degree (with distinction) from The University of Western Australia, Perth, WA, Australia, in 2006.

He is currently with the School of Computer Science and Software Engineering at The University of Western Australia.

Dr. Mian received the Australasian Distinguished Doctoral Dissertation Award from the Computing Research and Education Association of Australasia. He received the prestigious Australian Post-Doctoral and Australian Research Fellowships in 2008 and 2011, respectively. He received the UWA Outstanding Young Investigator Award in 2011, the West Australian Early Career Scientist of the Year Award in 2012, and the Vice-Chancellor's Mid-Career Research Award in 2014. He has been a Guest Editor of Pattern Recognition, Computer Vision and Image Understanding, and Image and Vision Computing journals. His research interests include computer vision, machine learning, 3-D shape analysis, hyperspectral image analysis, pattern recognition, and multimodal biometrics.