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 non-negative 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 non-negative 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—Hyperspectral unmixing, robust matrix factorization, sparse representation, blind source separation, unsupervised unmixing.

I. INTRODUCTION

HYPERSPECTRAL imaging acquires precise spectral information of 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 regions on the Earth’s surface and minimal light scattering is observed in the scene [3]. This work 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], [7], [8], [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 (ICE) [18], Minimum Volume Simplex Analysis [19] and Sparsity promoting ICE [20] 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] has shown the effectiveness of sparse regression methods [24], [23], [25] for hyperspectral unmixing by approaching the problem in a supervised 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, e.g. [22]. 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 [53]. It is well known that high coherence of the dictionary can lead to poor performance of the sparse regression framework [54].

Hyperspectral unmixing can be readily formulated as a Non-negative 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 non-convexity 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, e.g. [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] also proposed a manifold learning and sparse NMF based hyperspectral unmixing approach that considers local space information for the improved performance. Later, Liu 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 sparsity and 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 form 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 work, 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, 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 work, we also propose a non-negative variant of the Subspace Pursuit [44] algorithm, that is known for its computational efficiency [22]. The proposed Non-negative Subspace Pursuit is exploited in our approach for efficient matrix factorization. The endmembers computed by our approach are non-negative 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 work focuses on the Linear Mixing Model (LMM) [36], that represents a pixel $y \in \mathbb{R}^m$ of a hyperspectral image as:

$$y = \Phi \alpha + \epsilon,$$

where $\Phi \in \mathbb{R}^{m \times K}$ contains the endmembers as its columns, $\alpha \in \mathbb{R}^K$ encodes their fractional abundances and $\epsilon \in \mathbb{R}^m$ represents the error, considered as additive Gaussian noise. Under this model, the coefficients $\alpha_{i} \in \{1, \ldots, K\}$ of $\alpha$ must satisfy two constrains [3]: (1) $\forall i, \alpha_{i} \geq 0$, i.e. Abundance Non-negativity 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 non-negative quantities that add up to 1.

B. Unmixing as Constrained Matrix Factorization

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

$$Y = \Phi A + E,$$

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

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

where, $\alpha_{i,j}$ is the coefficient of $A$ at index $(i, j)$, $\alpha_j$ denotes the $j^{th}$ column of $A$ and $||.||_1$ computes the $\ell_1$-norm.

As $\Phi$ contains the endmembers of the complete image and $\alpha_j$ corresponds to the $j^{th}$ pixel only, we can expect $\alpha_j$ to be generally sparse. However, introducing an explicit sparsity
constraint over $\alpha_j$ can come with an additional computational cost. It is a well-known fact that sparseness can be imposed over a representation ($\alpha_j$, in this case) by restricting its $\ell_1$-norm \cite{63, 17}. Since the ASC in \cite{3} is doing exactly the same thing, a form of sparseness is naturally exhibited by $\alpha_j$. Therefore, we do not impose additional sparsity constraint in \cite{3}. Previously, Chen et al. \cite{63}, also reported the sparseness-inducing effect of restricting the sum of coefficients of a non-negative representation vector to 1.

The above mentioned sparseness is of the form where many coefficients of the vector shrink towards zero. From that perspective, if an abundance vector $\alpha_j$ does not follow the ASC, additionally imposing ASC will sparsify the vector by shrinking its coefficients when $\sum \alpha_{i,j} > 1$. When $\sum \alpha_{i,j} = 1$, the representation remains unaltered, and in the case of $\phi$ dictionary atoms have bounded energy, resulting in substantial note that this analysis applies to practical conditions where the cases. Therefore, we can argue that the solutions computed coefficients must still remain very small in both of the latter dimensionality of the representation vector, most of the vector instead of shrinking them. Nevertheless, due to the high sensitivity is of the form where many coefficients of the vector shrink towards zero. From that perspective, if an abundance vector $\alpha_j$ does not follow the ASC, additionally imposing ASC will sparsify the vector by shrinking its coefficients when $\sum \alpha_{i,j} > 1$. When $\sum \alpha_{i,j} = 1$, the representation remains unaltered, and in the case of $\phi$ dictionary atoms have bounded energy, resulting in substantial note that this analysis applies to practical conditions where the cases. Therefore, we can argue that the solutions computed coefficients must still remain very small in both of the latter dimensionality of the representation vector, most of the vector instead of shrinking them. Nevertheless, due to the high

Since spectra are non-negative quantities, we can also force the coefficients $\phi_{h,i}$ of $\Phi$ to be non-negative, resulting in:

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

Estimating $\Phi$ and $A$ that satisfy \cite{4}, is a constrained matrix factorization problem. A matrix $\Phi$ obtained by solving that problem, would generally approximate $Y$’s endmembers well. Nevertheless, some of the computed endmembers in $\Phi$ 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 non-convexity, which also makes their physical interpretation hard. In order to mitigate this issue, we reformulate the model in \cite{4} as follows.

We force the columns $\phi_{i,\{1,...,K\}}$ of $\Phi$ to be non-negative linear combinations of the image pixels themselves. Moreover, we force each $\phi_i$ to only use at most $k$ pixels in its construction, where $k$ is a small positive integer. Concretely, our model becomes:

$$Y = Y \Xi A + E \quad \text{s.t.} \quad \forall i, j, \xi_{j,i}, \alpha_{i,j} \geq 0; \quad ||\alpha_j||_1 = 1; \quad ||\xi_i||_0 \leq k, \quad \text{(5)}$$

where $\Xi \in \mathbb{R}^{n \times K}$ is a matrix with coefficients $\xi_{j,i}$ and $\xi_i$ as its $i^{\text{th}}$ column. The symbol $||.||_0$ denotes the $\ell_0$ pseudo-norm, that counts the number of non-zero coefficients in a vector. Since both $Y$ and $\Xi$ are non-negative, the endmembers in $\Phi = Y \Xi$ also remain non-negative in the above formulation.

At this point, it is worth mentioning that Ambikapathi et al. \cite{67} observed an important geometric property of hyperspectral data, that convex hull of pixels lies strictly inside the convex hull of the endmembers. Considering this, the model in \cite{5} may appear restrictive because constructing endmembers as non-negative 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 \cite{67}. In \cite{5}, we always consider $E$ to be a non-zero matrix, which also qualifies the scope of this work 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 \cite{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 \cite{V} and \cite{W}.

The model in \cite{5} records the relationship between the endmembers and the pixels in the matrix $\Xi$. Hence, it is also able to explicate any pixels that it considers to be the pure endmembers\cite{2}. For such pixels, $||\xi_i||_0$ becomes exactly 1. This characteristic of the model is certainly desirable, which also justifies the use of the $\ell_0$-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 $\Xi$ and $A$ under \cite{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 Eq. \cite{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 $\Xi$ and $A$ under \cite{5} can be formulated as:

$$\min_{\Xi, A} ||Y - Y \Xi A||_F^2 \quad \text{s.t.} \quad \forall i, j, \xi_{j,i}, \alpha_{i,j} \geq 0; \quad ||\alpha_j||_1 = 1; \quad ||\xi_i||_0 \leq k, \quad \text{(6)}$$

where $||.||_F$ denotes the Frobenius norm of a matrix. In \cite{5}, 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 the objective function to the following:

$$\min_{\Xi, A} \left\{ ||(Y - Y \Xi A)\Delta^{-\frac{1}{2}}||_F^2 + ||\Delta||_1 \right\} \quad \text{s.t.} \quad \forall i, j, ||\alpha_j||_1 = 1; \quad \xi_{j,i}, \alpha_{i,j} \geq 0; \quad ||\xi_i||_0 \leq k; \quad \delta_j \geq \varepsilon, \quad \text{(7)}$$

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

\footnote{We 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 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 $\Xi$ and $A$ to be fixed. In that case, (7) reduces to the following optimization problem:

$$\min_{\delta_j} \left\{ \frac{||y_j - Y\Xi\alpha_j||_2^2}{\delta_j} + \delta_j \right\} \text{ s.t. } \delta_j \geq \varepsilon. \tag{8}$$

By differentiating the expression in the brackets w.r.t. $\delta_j$ and equating it to zero, we can show that $h = ||y_j - Y\Xi\alpha_j||_2$ becomes the minimizer $\delta_j^*$, which represents a linear penalty. However, this penalty becomes applicable only when $h \geq \varepsilon$. Otherwise, we must choose $\delta_j^* = \varepsilon$ because $\varepsilon$ 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 $\left\{ \frac{h^2}{\varepsilon^2} + \varepsilon \right\}$. 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 non-convex. Nevertheless, it has the following desirable characteristics. (a) With fixed $\Xi$ and $\Delta$, the optimization of $A$ becomes convex. (b) By fixing $\Xi$ and $A$, we get a closed-form solution for the optimal $\Delta$. (c) With known $A$ and $\Delta$, $\Xi$ can be computed by solving a constrained sparse optimization problem. These properties are further explained below. We propose Algorithm 1 that exploits these properties in employing a Block Coordinate Descent (BCD) scheme [37] 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 $A$. By fixing $\Xi$ and $\Delta$, and replacing $Y\Xi$ by $\Phi$ in (7), we can estimate $A$ as follows:

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

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 work, we implement the SUnSAL algorithm provided by Bioucas-Dias and Figueiredo [25] to solve this problem. In Algorithm 1 this computation is carried out on line ‘2’.

To compute $\Delta$, we separately estimate each of its diagonal entries $\delta_j$, by solving for the objective function in (8). From the discussion in Section II-A it is clear that an optimal $\delta_j^*$ can be directly computed by choosing the larger value among $\varepsilon$ and $||y_j - \Phi\alpha_j||_2$. This procedure is performed in line ‘3-6’ of the algorithm. It is worth mentioning that although the computation is element-wise, it exactly emulates solving (7) for the complete matrix $\Delta$ (with fixed $\Xi$ and $A$) due to the special construction of $\Delta$, which gives us a closed-form solution of the matrix.

**Algorithm 1 Robust Constrained Matrix Factorization**

**Input:** Data $Y \in \mathbb{R}^{m \times n}$ normalized in $\ell_2$-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, $\Phi = Y\Xi, \Delta \in \mathbb{R}^{n \times n}$ as an identity matrix, and $q = 1$.

**Iterations:**

1: for $q = 1$ to $Q$ do
2: $A = \text{argmin}_A ||Y - \Phi A||_F^2$ 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: $\delta_j^* = \max(\varepsilon, ||y_j - \Phi\alpha_j||_2)$
5: $\delta_j \leftarrow \delta_j^*$
6: end for
7: $\Gamma = Y - \Phi A$
8: for $i = 1$ to $K$ do
9: $\theta = \Delta^{-\frac{1}{2}}\alpha_i^T$
10: $\psi = \Gamma\theta + Y\xi_i$
11: $\xi_i^* = \text{argmin}_{\xi_i} ||\psi - Y\xi_i||_F^2$ s.t. $\forall j, \xi_{i,j} \geq 0; ||\xi_i||_0 \leq k$
(Solve using Algorithm 2)
12: $\Gamma = \Gamma + Y(\xi_i - \xi_i^*)\alpha_i$
13: $\xi_i \leftarrow \xi_i^*$
14: end for
15: $\Phi = Y\Xi$
16: end for

**Output:**

17: Endmember matrix $\Phi$, Abundance matrix $A$.

To compute $\Xi$, we can write the optimization objective in (7) as:

$$\min_{\Xi} ||(Y - Y\Xi A)\Delta^{-\frac{1}{2}}||_F^2,$$

s.t. $\forall i, j, \xi_{i,j} \geq 0; ||\xi_i||_0 \leq k. \tag{10}$$

Notice that, we do not disregard $\Delta$ 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 $\Delta$ 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 $\Xi$. This can be done by solving the following problem:

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

where $\alpha_i \in \mathbb{R}^{1 \times n}$ denotes the $i^{th}$ row of $A$ and $\xi_i^*$ is an updated version of $\xi_i$ that would minimize the loss. By changing the input argument from $\xi_i$ to $\xi_i^*$, $\Xi$ also becomes a constant in (11). Exploiting the fixed matrices, we can further modify our optimization objective as

$$\forall i, \min_{\xi_i} \left\{ \frac{1}{2} ||(Y\Xi A)\Delta^{-\frac{1}{2}}\alpha_i^T \right\}_{\alpha_i} \left( Y\Xi A\Delta^{-\frac{1}{2}}\alpha_i^T + Y\xi_i \right)^2 \tag{12},$$

where the superscript ‘$T$’ signifies the transpose operation. Let us denote the expression within the braces by $\psi$ to finally
It arrives at the following form of (10):

\[ \forall i, \min \frac{||\psi - Y\xi_i^*||_2^2}{\xi_i^*} \]

s.t. \( \forall j, \xi_{j,i}^* \geq 0; ||\xi_i^*||_0 \leq k \). \hspace{1cm} (13)

Above, is an objective function of a sparse optimization problem, with a non-negativity constraint over the sparse codes \( \xi_i^* \). 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 \( \Xi \) are given in lines ‘7’ - ‘14’ of Algorithm \( [1] \). In line ‘7’, we compute the residual matrix \( \Gamma \in \mathbb{R}^{m \times n} \) outside the for-loop that iterates over the columns of \( \Xi \). In line ‘12’, we update \( \Gamma \) to account for the change in the residual caused by the newly computed \( \xi_i^* \), before updating \( \xi_i^* \) in line ‘13’ of the algorithm.

**Algorithm 2: Non-negative Subspace Pursuit**

**Input:** Sparsity level \( k \), data \( \psi \), dictionary \( Y \).

**Initialize:** \( S^0 \leftarrow \) Indices of the \( k \) largest coefficients of \( \frac{Y^T\psi}{||Y\psi||_2} \).

\( \beta = \arg\min_{\beta} ||\psi - Y\beta||_2^2 \) s.t. \( \{\text{supp}\{\beta\} \subseteq S^0; \forall z, \beta_z \geq 0, \}

\( r^0 = \psi - Y\beta \).

**Iterations:**

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

It is worth mentioning that Algorithm \( [2] \) differs from SP \( [44] \) in two core operations, namely; (a) identifying the sets \( S^0 \) and \( \hat{S}^i \), and in (b) computing the vectors \( \beta \) and \( \xi^* \). In SP, \( S^0 \) has been previously proposed. The Non-negative Subspace Pursuit (NSP) is given in Algorithm \( [2] \). We note that a non-negative variant of Orthogonal Matching Pursuit algorithm (OMP) \( [23] \) has been proposed by Bruckstein et al. \( [60] \) that 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 \( Y \), such that the input signal \( \psi \) (or its best approximation) lies in the positive orthant of this subspace. Following the conventions of the sparse representation literature, below, we refer to \( Y \) 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 \( \psi \). These indices are recorded in a set \( S^0 \). Then, we compute the orthogonal projection of \( \psi \) onto the positive orthant of the atoms indexed in \( S^0 \). In Algorithm \( [2] \) the operator ‘\( \{\cdot\} \)’ used for this purpose, indicates the support, i.e. the indices of the non-zero coefficients, of a vector. We then compute the residue vector \( r^0 \) by subtracting from \( \psi \) its computed projection (i.e. \( Y\beta \)).

The algorithm performs at most \( k \) iterations. In the \( i \)-th iteration, it augments the set \( S^{i-1} \) with \( k \) more indices to compute a set \( \hat{S}^i \). The newly added indices correspond to the atoms subtending the smallest angles with the residue vector \( r^{i-1} \). Using \( \hat{S}^i \), the \( k \)-best atoms are identified in \( S^i \). These atoms maximally contribute to the non-negative least squares approximation of \( \psi \), when only the atoms in \( S^i \) are used as the basis. This procedure is stated in lines ‘3-4’ of the algorithm.

Next, \( \psi \) is approximated in the positive orthant of the column space of the atoms indexed in \( S^i \). This results in a potential solution \( \xi^* \), that is used to compute the new residue vector \( r^i \) (lines ‘5-6’). If the new residue is zero or it does not improve in the current iteration, the iterations stop and \( \xi^* \) becomes the solution. However, if the residue increases as a result of the \( i \)-th iteration, \( S^{i-1} \) is used to re-compute the sparse codes \( \xi^* \), as mentioned in line ‘11-13’ of the algorithm.

We emphasize on one important characteristic of NSP. That is, it reconstructs \( \psi \) using a non-negative linear combination of ‘at-most’ \( k \) pixels indexed in \( S \). Bruckstein et al. \( [60] \) showed that the non-negativity in such reconstructions naturally leads to the sparsest combinations of the signals. For NSP, this means that the sparsity of \( \xi^* \) is upper bounded by \( k \). If \( \psi \) 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 \( \Xi \) 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 non-negativity imposes sparsity by forcing the unrequired coefficients of \( \xi^* \) to zero \( [66] \). However, for a larger \( k \), a larger subspace would have to be parsed for computing the solution, that can result in more computations. Therefore, smaller \( k \) values are preferred for our approach.
records the indices of the $k$ coefficients of the vector $s_0 = Y\psi$ that have the largest magnitudes. Similarly, $\hat{S}^i$ is computed by augmenting $S^{i-1}$ with the indices of largest magnitude entries in $s_i = Y\hat{r}^{i-1}$. Such constructions of the sets are governed by the fact that SP also allows negative coefficients in $\beta$ (and $\xi$). We place a non-negativity constraint over $\beta$ (and $\xi$). Therefore, we allow only those atoms in sets $S^0$ and $\hat{S}^i$ that have positive correlations with $\psi$ and $\hat{r}^{i-1}$, 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 Lab’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 $P_{25}$. Details of the materials corresponding to $P_{25}$ are provided in Table IV in the Appendix. The table contains the information regarding the type, class, subclass and particle size of each material. We randomly selected the endmembers in $P_{25}$ 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 [33], [22]. For $P_{25}$, the value of $\mu = 0.9998$. However, it should be noted that $\mu$ is only a worst-case measure and the mean absolute inner product for each possible pair of normalized endmembers in $P_{25}$ 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 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 [22], [16], 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 10,000 mixed pixels, using a set $P_{20} \subset P_{25}$ that contained 20 randomly chosen endmembers. The final results were computed by averaging the performance on 10 images. Each pixel $y_j \in \mathbb{R}^m$ 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 [33]. 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 ($\sigma^2_h$) of the matrix was computed as:

$$
\sigma^2_h = \frac{\sum_{j=1}^{n} ||y_j||_2^2}{n \times 10^{\text{SNR}} \sigma \sqrt{2\pi}} \exp^{(h-h)^2/(2\sigma^2)} \quad \forall h \in \{1, ..., m\}, \quad (14)
$$

where $\eta$ 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 ($\theta_{avg}$) 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 $\mathcal{P} = \{\varphi^*_p | \varphi^*_p \text{ is a pure endmember}\}$ to generate a synthetic image,
then $\theta_{avg}$ is defined as:

$$\theta_{avg} = \frac{180}{|P|} \sum_{q=1}^{|P|} \arccos \left( \frac{\varphi^*_q \varphi_q}{||\varphi_q||_2 ||\varphi^*_q||_2} \right),$$

where $|.|$ denotes the cardinality of the set and $\varphi_q$ is the estimated endmember, that best matches the $q^{th}$ true endmember. Notice that the computed angle is in degrees.

The metric $\theta_{avg}$ only evaluates the soundness of the extracted endmembers. To evaluate the estimated fractional abundances, we used the Root Mean Squared Error (RMSE), defined as:

$$\text{RMSE} = \sqrt{\frac{|A - A^*|_F^2}{|P| \times n}},$$

where $A^* \in \mathbb{R}^{|P| \times n}$ 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 $A$ accordingly.

### C. Benchmarking

To benchmark, we mainly compare our approach to 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 non-negative 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 non-negative 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 large number of pixels, this can result in considerable computational advantage over conventional Archetypal Analysis. In our experiments, we used the Archetypal Analysis enhancement of Chen et al. [63] that also accounts for outliers. To explicate the effect of robustness against outliers in our approach, we also provide the results of the non-robust variant of our approach, denoted as CMF. This variant is implemented by removing the lines ‘3-6’ and the matrix $\Delta$ from line ‘9’ in Algorithm [1].

![Image](image.png)

**TABLE I: Average spectral angle ($\theta_{avg}$):** For noisy cases, the Signal to Noise Ratio is 30dB.

<table>
<thead>
<tr>
<th>Method</th>
<th>White noise</th>
<th>Correlated noise</th>
<th>Noise free</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVMAX [15]</td>
<td>10.35 ± 1.56</td>
<td>9.34 ± 0.37</td>
<td>--</td>
</tr>
<tr>
<td>ODL [61]</td>
<td>10.55 ± 1.64</td>
<td>8.99 ± 0.79</td>
<td>3.54 ± 0.21</td>
</tr>
<tr>
<td>K-SVD [60]</td>
<td>12.85 ± 0.25</td>
<td>13.86 ± 0.70</td>
<td>7.40 ± 1.39</td>
</tr>
<tr>
<td>ODL-NN [62]</td>
<td>10.65 ± 0.33</td>
<td>8.70 ± 0.58</td>
<td>7.70 ± 0.51</td>
</tr>
<tr>
<td>Archetypal [63]</td>
<td>5.49 ± 0.36</td>
<td>4.95 ± 0.33</td>
<td>4.24 ± 0.48</td>
</tr>
<tr>
<td>CMF</td>
<td>5.53 ± 0.20</td>
<td>5.11 ± 0.25</td>
<td>4.11 ± 0.21</td>
</tr>
<tr>
<td>RCMF</td>
<td>5.19 ± 0.22</td>
<td>4.55 ± 0.59</td>
<td>3.97 ± 0.22</td>
</tr>
</tbody>
</table>

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 SUrSal implementation [25] (as used by our approach) for computing the abundances. For VCA and AVMAX, we set the implementation to solve a fully constrained least squares problem. For SVMAX, we used the constrained version of SUrSal 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 [1] and [11]. In Table [1], 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 SNR = 30dB. The corresponding RMSEs of the computed fractional abundance matrices are reported in Table [11]. 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 $P_{25}$, 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 $\epsilon = 10^{-10}$. 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 [1] 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 non-robust 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 RCMF. However, it happens only for the noise-free case. For practical settings, RCMF outperforms SVMAX with a...
TABLE II: RMSE of reconstructed abundance matrices: For noisy cases, the SNR is 30dB.

<table>
<thead>
<tr>
<th>Method</th>
<th>White noise</th>
<th>Correlated noise</th>
<th>Noise free</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCA [11]</td>
<td>0.130 ± 8.8 × 10^{-3}</td>
<td>0.128 ± 4.08 × 10^{-3}</td>
<td>--</td>
</tr>
<tr>
<td>AVMAX [13]</td>
<td>0.129 ± 4.3 × 10^{-3}</td>
<td>0.127 ± 6.53 × 10^{-3}</td>
<td>--</td>
</tr>
<tr>
<td>SVMAX [15]</td>
<td>0.141 ± 1.3 × 10^{-2}</td>
<td>0.132 ± 7.9 × 10^{-3}</td>
<td>0.0088 ± 8.08 × 10^{-3}</td>
</tr>
<tr>
<td>K-SVD [60]</td>
<td>2.075 ± 0.16</td>
<td>2.888 ± 1.419</td>
<td>6.06 ± 2.63</td>
</tr>
<tr>
<td>ODL [61]</td>
<td>1.506 ± 0.16</td>
<td>1.786 ± 0.200</td>
<td>88.98 ± 70.23</td>
</tr>
<tr>
<td>ODL-NN [62]</td>
<td>1.050 ± 0.07</td>
<td>1.261 ± 1.159</td>
<td>1.49 ± 0.05</td>
</tr>
<tr>
<td>Archetypal [63]</td>
<td>0.103 ± 2.2 × 10^{-3}</td>
<td>0.103 ± 6.43 × 10^{-3}</td>
<td>0.105 ± 4.00 × 10^{-3}</td>
</tr>
<tr>
<td>CMF</td>
<td>0.101 ± 0.8 × 10^{-3}</td>
<td>0.100 ± 0.010</td>
<td>0.106 ± 2.34 × 10^{-3}</td>
</tr>
<tr>
<td>RCMF</td>
<td>0.097 ± 3.5 × 10^{-3}</td>
<td>0.095 ± 6.0 × 10^{-3}</td>
<td>0.104 ± 1.10 × 10^{-2}</td>
</tr>
</tbody>
</table>

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

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 non-negativity 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 non-negative in our experiments, this was not the case for their computed abundances, unless the non-negativity was explicitly imposed. Due to better recovery of the endmembers by SVMAX [15] in the noise-free settings, 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 is any pixel, which is a practical possibility. We used a high SNR of 100dB, 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 non-existing additive noise.

Even in the presence of 3% outliers for such data the $\theta_{avg}$ 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 $\theta_{avg}$ 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 worse performance. The differences between the performances of the proposed approach and archetypal analysis are significant in these experiments which highlights better suitability of our approach for unmixing as compared to 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 $\theta_{avg}$...
We extracted 200 approaches against different levels of noise in the data. For Archetypal Analysis, we used 20 pure spectra. This happened due to the presence of outliers, which were not considered in the noise-free settings of Tables I and II. 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.

In Fig. 7 and 8, we plot the \( \theta_{\text{avg}} \) 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 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 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 @ 3.4GHz and 8 GB RAM. In the table, the mean computation time of CMF exceeds RCMF because the later 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 to the proposed approach, 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.

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 is 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 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.
TABLE III: Mean computation time in seconds, for unmixing one thousand pixels.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18.09</td>
<td>18.76</td>
<td>18.81</td>
<td>17.82</td>
<td>3.27</td>
<td>3.18</td>
<td>3.20</td>
<td>18.13</td>
<td>17.67</td>
</tr>
</tbody>
</table>

Fig. 9: Fractional abundance analysis of the AVIRIS data: The first row shows the classification maps by the Tricorder software. Results of RCMF are displayed in the second row. The third row shows the result of VCA endmembers used for solving a fully constrained least squares problem and the fourth row shows CSUnSAL [25] results with ASTER library spectra.

In Fig. 9 we display the fractional abundance maps computed by the proposed RCMF (second row) for the selected sub-cube. 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 algorithm<sup>4</sup> 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 a fully constrained least squares 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]. 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

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 $70 \times 70$ patch of the image. Some spectral distortions are observable in this plot by VCA.

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 the figure, 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 $70 \times 70$ patch of the image was used in the experiment. According to the mineral map computed by Tetracorder [69], this patch contains Kaolinite as mixtures 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.

VI. CONCLUSION

We proposed a novel matrix factorization approach for linear hyperspectral unmixing. In addition to accounting for the non-negativity of the endmembers and the physical constraints over their abundances, the proposed approach forces the extracted endmembers to be sparse non-negative 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.

APPENDIX

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

REFERENCES

<table>
<thead>
<tr>
<th>Sr. #</th>
<th>Type</th>
<th>Class</th>
<th>Subclass</th>
<th>Particle size</th>
<th>Sr. #</th>
<th>Type</th>
<th>Class</th>
<th>Subclass</th>
<th>Particle size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mineral</td>
<td>Arsenate</td>
<td>None</td>
<td>Coarse</td>
<td>14</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Inosilicate</td>
<td>Fine</td>
</tr>
<tr>
<td>2</td>
<td>Mineral</td>
<td>Carbonate</td>
<td>None</td>
<td>Fine</td>
<td>15</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Phyllosilicate</td>
<td>Coarse</td>
</tr>
<tr>
<td>3</td>
<td>Mineral</td>
<td>Oxide</td>
<td>None</td>
<td>Coarse</td>
<td>16</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Phyllosilicate</td>
<td>Medium</td>
</tr>
<tr>
<td>4</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Inosilicate</td>
<td>Coarse</td>
<td>17</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Tectosilicate</td>
<td>Fine</td>
</tr>
<tr>
<td>5</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Nesosilicate</td>
<td>Fine</td>
<td>18</td>
<td>Mineral</td>
<td>Sulfate</td>
<td>None</td>
<td>Medium</td>
</tr>
<tr>
<td>6</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Phyllosilicate</td>
<td>Fine</td>
<td>19</td>
<td>Soil</td>
<td>Inceptisol</td>
<td>Halplumbrept</td>
<td>Coarse</td>
</tr>
<tr>
<td>7</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Tectosilicate</td>
<td>Coarse</td>
<td>20</td>
<td>Rock</td>
<td>Igneous</td>
<td>Intermediate</td>
<td>Solid</td>
</tr>
<tr>
<td>8</td>
<td>Mineral</td>
<td>Silicate</td>
<td>Tectosilicate</td>
<td>Medium</td>
<td>21</td>
<td>Rock</td>
<td>Sedimentary</td>
<td>Sandstone</td>
<td>Solid</td>
</tr>
<tr>
<td>9</td>
<td>Mineral</td>
<td>Sulfide</td>
<td>None</td>
<td>Coarse</td>
<td>22</td>
<td>Rock</td>
<td>Igneous</td>
<td>Felsic</td>
<td>Solid</td>
</tr>
<tr>
<td>10</td>
<td>Mineral</td>
<td>Borate</td>
<td>None</td>
<td>Medium</td>
<td>23</td>
<td>Rock</td>
<td>Igneous</td>
<td>Ultranite</td>
<td>Fine</td>
</tr>
<tr>
<td>11</td>
<td>Mineral</td>
<td>Element</td>
<td>None</td>
<td>Coarse</td>
<td>24</td>
<td>Rock</td>
<td>Metamorphic</td>
<td>Marble</td>
<td>Coarse</td>
</tr>
<tr>
<td>12</td>
<td>Mineral</td>
<td>Oxide</td>
<td>None</td>
<td>Medium</td>
<td>25</td>
<td>Rock</td>
<td>Sedimentary</td>
<td>Limestone</td>
<td>Fine</td>
</tr>
<tr>
<td>13</td>
<td>Mineral</td>
<td>Halide</td>
<td>None</td>
<td>Medium</td>
<td>26</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

These details are directly taken from the ASTER library.


Naveed Akhtar is a Research Fellow at The College of Engineering and Computer Science, Australian National University. He completed his PhD in Computer Vision in 2017 from The University of Western Australia (UWA) under the supervision of Prof. Ajmal Mian. Previously, he received the Bachelor of Engineering degree with distinction in Avionics from the College of Aeronautical Engineering, National University of Sciences and Technology (NUST), Pakistan, in 2007 and M.Sc. degree with distinction in Autonomous Systems from Hochschule Bonn-Rhein-Sieg (HBRS), Sankt Augustin, Germany, in 2012. He has served as a Research Assistant at Research Institute for Microwaves and Millimeter-waves Studies, NUST, Pakistan, from 2007 to 2009 and as a Research Associate at the Department of Computer Science at HBRS, Germany in 2012. His current research is focused on hyperspectral image analysis, with emphasis on sparse representation based techniques.

Ajmal Mian completed his PhD from The University of Western Australia in 2006 with distinction and received the Australasian Distinguished Doctoral Dissertation Award from Computing Research and Education Association of Australasia. He received the prestigious Australian Postdoctoral 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-Chancellors Mid-Career Research Award in 2014. He has secured seven Australian Research Council grants and one National Health and Medical Research Council grant with a total funding of over $3 Million. He is currently with the School of Computer Science and Software Engineering at The University of Western Australia and is 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, 3D shape analysis, hyperspectral image analysis, pattern recognition, and multimodal biometrics.