Hyperspectral Unmixing with Rare Endmembers via Minimax Nonnegative Matrix Factorization

Timothy Marrinan and Nicolas Gillis
University of Mons, Belgium

***Abstract*—Hyperspectral images are used for ground-cover classification because many materials can be identified by their spectral signature, even in images with low spatial resolution. Pixels in such an image are often modeled as a convex combination of vectors, called endmembers, that correspond to the reflectance of a material to different wavelengths of light. This is the so-called linear mixing model. Since reflectance is inherently nonnegative, the task of unmixing hyperspectral pixels can be posed as a low-rank nonnegative matrix factorization (NMF) problem, where the data matrix is decomposed into the product of the estimated endmembers and their abundances in the scene. The standard NMF problem then minimizes the residual of the decomposition. Thus, using NMF works well when materials are present in similar amounts, but if some materials are under-represented, they may be missed with this formulation. Alternatively, we propose a novel hyperspectral unmixing model using a collection of NMF subproblems solved for patches sampled from the original image. The endmembers are estimated jointly, such that the the maximum residual across all patches is minimized. In this paper we estimate the solution to the patch-based minimax NMF model, and show that it can estimate rare endmembers with superior accuracy.**

***Index Terms*—nonnegative matrix factorization, hyperspectral unmixing, minimax, approximate subgradient, low-rank**

I. Introduction

Hyperspectral images (HSI) are 3-mode tensors that measure the reflectance (or radiance) of a scene with respect to a large number of narrow wavelength-bands of light. Due to the high spectral resolution of modern imaging systems, it is often possible to distinguish ground-cover materials in an HSI even when spatial resolution is quite low. Pixels in such an image can contain one or more materials depending on scale, so they are commonly modeled by a linear mixing model. Let $X \in \mathbb{R}_+^{b\times p}$ be a nonnegative matrix corresponding to a raster-scanned hyperspectral image with b wavelengths measured at each of p pixels. If the image contains r unique materials, the linear mixing model that describes the matrix X is written as

$$X = WH \tag{1}$$

for $W \in \mathbb{R}_+^{b\times r}$ and $H \in \mathbb{R}_+^{r\times p}$ with $H^T\mathbf{1}_r \leq \mathbf{1}_r$ where $\mathbf{1}_r \in \mathbb{R}^r$ is the vector of all 1's. Note that we have relaxed the sum-to-one constraints (that is, $H^T\mathbf{1}_r = \mathbf{1}_r$) on the abundances to allow for different illuminations across the pixels of the image.

One significant arm of research in the remote sensing community focuses on identifying the constituent materials in an observed HSI by 'unmixing' the data matrix to reveal the spectral signatures for the materials present in the image and the weight of their contribution to each pixel. This blind hyperspectral unmixing (HU) problem has the following typical assumptions: W has full column rank, H has full row rank, and r is known.

If we view the origin and the columns of W as the vertices of an r-dimensional simplex, then in the noiseless case, the data points are contained in this simplex. One technique to perform the hyperspectral unmixing and recover this simplex is nonnegative matrix factorization (NMF), which solves

$$\min_{W\in\mathbb{R}_+^{b\times r} H\in\mathbb{R}_+^{r\times p}} d(X, WH), \tag{2}$$

where $d : \mathbb{R}^{b\times p} \times \mathbb{R}^{b\times p} \to \mathbb{R}$ is some distance measure. The squared Frobenius norm is often chosen for the distance function, $d(A, B) = \|A - B\|_F^2 = \sum_{i=1}^{b}\sum_{j=1}^{p} |A(i, j) - B(i, j)|^2$, because it makes the optimization problem computationally easier to tackle (in particular, it is a convex nonnegative least-squares problem in W for H fixed, and vice versa), and is the maximum likelihood estimator for Gaussian noise. However, two issues confound this approach to HU, first, Problem (2) is ill-posed so the solution is not unique, and second, if the image is dominated by a subset of the materials, finding r endmembers with the best least-squares fit does not always lead to the complete set of materials in the presence of noise.

The first issue, the uniqueness and identifiability of NMF, is an active area of research in its own right. Under some appropriate conditions on the data, it has been shown that the solution can be uniquely identified by searching for a solution such that the simplex defined by the columns of W has the minimum volume [1]–[4], which is referred to as minimum volume NMF (min-vol NMF).

The second issue, improving the estimation of rare endmembers, is the topic of this paper. It has garnered less attention, but research suggests that the number of groundtruth endmembers in some popular datasets is much higher than typically reported [5]. Existing approaches to improve their estimation require a rough identification of the pixels that contain the rare materials, and include techniques such as separately processing pixels with abundant materials and rare materials [6], or using bootstrap resampling on pixels containing rare materials to increase their impact on unmixing [7].

In contrast, this paper proposes a novel method for blind hyperspectral unmixing with rare materials that requires no pre-identification of pixels containing those materials. We use

This work is supported by the Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlanderen (FWO) under EOS Project no O005318F-RG47, and by the European Research Council (ERC starting grant no 679515).

efficient existing methods to solve a variant of min-vol NMF, and improve the estimation of the rare endmembers relative to standard least-squares techniques. The main contributions of this paper are three-fold: (i) the least-squares model of the data is reformulated by partitioning the image into smaller regions and minimizing the maximum least-squares fit across all the patches, (ii) a heuristic solution is proposed for the minimax problem based on an approximate subgradient algorithm applied to the dual, and (iii) we illustrate the effectiveness of the proposed approach on realistic data sets.

II. Background

Given a nonnegative matrix $X \in \mathbb{R}_+^{b\times p}$ and a factorization rank r, NMF attempts to decompose X into the product of two nonnegative matrices, $X = WH$, where $W \in \mathbb{R}_+^{b\times r}$ and $H \in \mathbb{R}_+^{r\times p}$. The decomposition problem can be written as in (2). The constraints $W,\ H \geq 0$ mean that the elements of W and H are nonnegative. Solving Problem (2) exactly is NP-hard in general [8], and the solution is not unique [4]. A common variant of the NMF problem that makes it well-posed is to search for a solution such that the simplex defined by W has the minimum volume. This approach to NMF has roots in the hyperspectral image processing community, see for example [9]–[11]. With this modification and a *sufficiently scattered* condition on the data, which essentially amounts to points being distributed on all facets of the convex hull spanned by W, the minimum volume solution to Problem (2) is unique and identifiable up to permutation and scaling of the columns of W and the rows of H [1]–[3]. The standard approach to solve min-vol NMF is to penalize the volume of simplex defined by the columns of W in the objective function:

$$\min_{\substack{W,H_i\geq 0\\ H_i^T\mathbf{1}_r\leq\mathbf{1}_r}} d(X,WH)+\beta v(W), \tag{3}$$

where $v:\mathbb{R}^{b\times r}\to\mathbb{R}$ measures the volume of W and $\beta>0$ is a penalty parameter. The additional constraint, $H^T\mathbf{1}_r\leq\mathbf{1}_r$, enforces the rows of X to be within the convex hull of the columns of W and the origin. Thus one possibility for the function v is the volume of this convex hull,

$$\mathrm{vol}\big(\mathrm{conv}\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r,\mathbf{0}\}\big)=\frac{1}{r!}\sqrt{\det(W^TW)}. \tag{4}$$

In this work, however, the simplex volume is measured by

$$v(W)\doteq\log\det(W^TW+\delta I) \tag{5}$$

for some $\delta>0$. Adding the small diagonal weighting, δ, to the Gram matrix improves the conditioning of W^TW, and the $\log\det(\cdot)$ function has a tight convex upper bound which leads to more efficient optimization algorithms. It has been empirically shown to provide good results as a volume regularizer [11]–[13].

III. Minimax NMF and Proposed Method

Suppose that a hyperspectral image is partitioned into n patches of m pixels, and raster-scanned to create $X_i\in\mathbb{R}_+^{b\times m}$ for $i=1,\ldots,n$. We can perform spectral unmixing with an abundance matrix, $H_i\in\mathbb{R}_+^{r_i\times m}$, for each of these n patches and a single endmember matrix, $W\in\mathbb{R}_+^{b\times r}$, that is jointly estimated for all patches. That is, for $i=1,\ldots,n$ we solve

$$\min_{\substack{W,H_i\geq 0\\ H_i^T\mathbf{1}_r\leq\mathbf{1}_r}} \|X_i-WH_i\|_F^2+\beta v(W). \tag{6}$$

The overall accuracy of the reconstruction can be assessed by some function of the n costs described by Equation (6). One possibility is the sum of the residuals,

$$\min_{\substack{W,H_i\geq 0\\ H_i^T\mathbf{1}_r\leq\mathbf{1}_r}} \sum_{i=1}^{n}\|X_i-WH_i\|_F^2+\beta v(W), \tag{7}$$

but this is equivalent to the original problem with pixel indices appropriately permuted to match the partitioning because

$$\sum_{i=1}^{N}\|X_i-WH_i\|_F^2=\|X-WH\|_F^2 \tag{8}$$

where $X=[X_1^TX_2^T\cdots X_n^T]^T$ and $H=[H_1^TH_2^T\cdots H_n^T]^T$.

Alternatively, we could measure the error as the maximum of (6) across all patches. That is, we could solve

$$\min_{\substack{W,H_i\geq 0\\ H_i^T\mathbf{1}_r\leq\mathbf{1}_r}} \max_{1\leq i\leq n} \|X_i-WH_i\|_F^2+\beta v(W). \tag{9}$$

By minimizing the maximum patch-residual, the decomposition will have to approximate each patch with similar reconstruction error, including atypical patches containing the endmembers present in small proportions. Thus for HU, the W that minimizes Equation (9) will provide a better fit for rare endmembers as desired.

a) Dual Problem: We will apply a subgradient method to the dual of (9), but for convenience of terminology we will consider the negative of the cost function and maximize its minimum value. By adding an auxiliary variable, τ, (9) can be reformulated as

$$\begin{aligned}\max_{\tau\in\mathbb{R},W,\{H_i\}_{i=1}^n}\ & \tau\\ \text{s.t. }& \tau+\|X_i-WH_i\|_F^2+\beta v(W)\leq 0,\\ & H_i^T\mathbf{1}_r\leq\mathbf{1}_r,\ \text{and } W,\ H_i\geq 0 \text{ for } 1\leq i\leq n.\end{aligned} \tag{10}$$

Let $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_n]^T$ be a vector of Lagrange multipliers associated with the first n inequality constraints in (10). Then the Lagrangian is

$$\begin{aligned}\mathcal{L}(W,\{H_i\}_{i=1}^n,\tau,\boldsymbol{\lambda}) &= \tau-\sum_{i\leq n}\lambda_i\tau\\ &\quad-\sum_{i\leq n}\lambda_i\left(\|X_i-WH_i\|_F^2+\beta v(W)\right)\\ &\quad\text{s.t. } H_i^T\mathbf{1}_r\leq\mathbf{1}_r,\ \text{and } \boldsymbol{\lambda},\ W,\ H_i\geq 0.\end{aligned} \tag{11}$$

The dual cost function, $q(\boldsymbol{\lambda})$, is then found by taking the supremum of (11) over W, H_i, and τ. The supremum over τ yields $q(\boldsymbol{\lambda})=+\infty$ unless $\boldsymbol{\lambda}^T\mathbf{1}_n=1$, in which case the first term is zero and the $\log\det$-term can be moved outside the

sum. Thus the dual cost simplifies to

$$q(\boldsymbol{\lambda}) = \sup_{\substack{W, H_i \geq 0 \\ H_i^T \mathbf{1}_r \leq \mathbf{1}_r}} -\beta v(W) - \sum_i \lambda_i \|X_i - WH_i\|_F^2, \tag{12}$$

and the dual problem is given by

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}_+^n} q(\boldsymbol{\lambda}) \text{ s.t. } \boldsymbol{\lambda}^T \mathbf{1}_n = 1. \tag{13}$$

b) Approximate Subgradient: We propose a heuristic to minimize Problem (13) based on the subgradient method [14]. Recall that a vector $\mathbf{g} \in \mathbb{R}^n$ is a subgradient of $q : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x} \in \text{dom } q$ if for all $\mathbf{z} \in \text{dom } q$,

$$q(\mathbf{z}) \geq q(\mathbf{x}) + \mathbf{g}^T (\mathbf{z} - \mathbf{x}). \tag{14}$$

To minimize q in Problem (13), the subgradient method uses the iteration

$$\boldsymbol{\lambda}^{(t+1)} = \Pi \left(\boldsymbol{\lambda}^{(t)} - \alpha^{(t)} \mathbf{g}^{(t)} \right), \tag{15}$$

where $\Pi : \mathbb{R}^n \to \mathbb{R}^n$ is the projection onto the unit simplex that keeps $\boldsymbol{\lambda}^{(t+1)}$ dual-feasible, and $\alpha^{(t)} > 0$ is a step size like $\alpha = a/t$ for $a > 0$. Other step sizes that also guarantee convergence can be found in textbooks such as [15].

At step t, suppose the supremum of Equation (12) is achievable for $\boldsymbol{\lambda}^{(t)}$, and that $W^{(t)}$ and $\{H_i^{(t)}\}_{i=1}^n$ are the matrices which achieve that maximum. Classical subgradient theory says that the vector $\mathbf{g}^{(t)} \in \mathbb{R}^n$ defined as

$$\mathbf{g}^{(t)} = -\left[\|X_1 - W^{(t)} H_1^{(t)}\|_F^2, \ldots, \|X_n - W^{(t)} H_n^{(t)}\|_F^2 \right]^T \tag{16}$$

is a subgradient to $q(\boldsymbol{\lambda}^{(t)})$ [15]. However, the NMF subproblem represented by $q(\boldsymbol{\lambda}^{(t)})$ is NP-hard, and thus computing $W^{(t)}$ and $\{H_i^{(t)}\}_{i=1}^n$ exactly is intractable.

Instead we use a fixed number of iterations of the min-vol NMF method of [12] to find approximate solutions to the supremum in Equation (12). This leads to an approximate subgradient, $\tilde{\mathbf{g}}$, and hence the heuristic nature of our method. Algorithm 1 summarizes our proposed approach. We initialize our algorithm using a pure-pixel search algorithm, namely the successive nonnegative projection algorithm (SNPA) from [16], as in min-vol NMF from [12].

Algorithm 1 Minimax min-vol NMF via dual subgradient

input Data: $X \in \mathbb{R}_+^{b \times p}$, window size: m, rank: r, step size: $a > 0$, volume penalty: $\tilde{\beta} > 0$, eigenvalue weight: $\delta > 0$, max iterations: *maxiter*, max subproblem iterations: *inneriter*

output W^*, $\{H_i^*\}_{i=1}^n$

$W, H \leftarrow$ approx. solution to $\|X - WH\|_F^2$ with SNPA [16]
$\{X_i \in \mathbb{R}_+^{b \times m}\}_{i=1}^n \leftarrow$ raster-scanned windows of X
$\{H_i \in \mathbb{R}_+^{r \times m}\}_{i=1}^n \leftarrow$ raster-scanned windows of H
$\boldsymbol{\lambda} \leftarrow [1/n, \ldots, 1/n]^T \in \mathbb{R}^n$
$\beta \leftarrow \tilde{\beta} \frac{\|X - WH\|_F^2}{\log\det(W^T W + \delta I_r)}$
$\tilde{\mathbf{g}} \leftarrow -\left[\|X_1 - WH_1\|_F^2, \ldots, \|X_n - WH_n\|_F^2 \right]^T$
$f(W, \{H_i\}_{i=1}^n) \leftarrow \min_{i \leq n} \{-\|X_i - WH_i\|_F^2\} - \beta v(W)$
$q(\boldsymbol{\lambda}) \leftarrow \boldsymbol{\lambda}^T \mathbf{g} - \beta v(W)$
$W^*, \{H_i^*\}_{i=1}^n \leftarrow W, \{H_i\}_{i=1}^n$
for $t = 1, \ldots, maxiter$ **do**
  $\alpha \leftarrow a/t$
  $\boldsymbol{\lambda} \leftarrow \Pi(\boldsymbol{\lambda} - \alpha \tilde{\mathbf{g}})$
  $X_\lambda \leftarrow \left[\sqrt{\lambda_1} X_1^T, \sqrt{\lambda_2} X_2^T, \ldots, \sqrt{\lambda_n} X_n^T \right]^T$
  for $p = 1, \ldots, inneriter$ **do** % Min-vol NMF [12]
    $H_\lambda \leftarrow \left[\sqrt{\lambda_1} H_1^T, \sqrt{\lambda_2} H_2^T, \ldots, \sqrt{\lambda_n} H_n^T \right]^T$
    $W \leftarrow \arg\max_{W \geq 0} -\|X_\lambda - WH_\lambda\|_F^2 - \beta v(W)$
    $\{H_i\}_{i=1}^n \leftarrow \arg\max_{H_i \geq 0,\ H_i^T \mathbf{1}_r \leq \mathbf{1}_r} -\|X_i - WH_i\|_F^2$
  $\tilde{\mathbf{g}} \leftarrow -\left[\|X_1 - WH_1\|_F^2, \ldots, \|X_n - WH_n\|_F^2 \right]^T$
  $f(W, \{H_i\}_{i=1}^n) \leftarrow \min_{i \leq n} \{-\|X_i - WH_i\|_F^2\} - \beta v(W)$
  $q(\boldsymbol{\lambda}) \leftarrow \boldsymbol{\lambda}^T \mathbf{g} - \beta v(W)$
  if $f(W, \{H_i\}_{i=1}^n) > f(W^*, \{H_i^*\}_{i=1}^n)$ **then**
    $W^*, \{H_i^*\}_{i=1}^n \leftarrow W, \{H_i\}_{i=1}^n$

IV. Numerical Experiments

The synthetic data for the experiments, $X \in \mathbb{R}^{b \times p}$, is created using the linear mixing model, so that the ith sample is

$$\mathbf{x}_i = W\mathbf{h}_i + \eta, \tag{17}$$

where $\eta \sim \mathcal{N}(\mathbf{0}, \sigma_N^2)$ is additive white Gaussian noise. The p data samples can be thought of as the raster-scan of an image with spatial dimensions $\sqrt{p} \times \sqrt{p}$. The last columns of W corresponds to rare endmembers that are present in a small proportion of the samples. More precisely, each column of H is distributed with a Dirichlet distribution using parameter $[0.05, \ldots, 0.05]^T \in \mathbb{R}^r$ for samples with the rare endmembers[1], and $[0.05, \ldots, 0.05, 0, \ldots, 0]^T \in \mathbb{R}^r$ for those without. Samples containing the rare endmembers are chosen by randomly selecting a rectangular region in the image for each rare endmember such that the area of the rectangle equals the specified number of samples. We use a small value for the Dirichlet parameter in order to satisfy, with high probability, the sufficiently scattered condition necessary for identifiability. Additionally, any column of H whose maximum element is greater than 0.8 is resampled to avoid pixels that are too pure. This makes the scenario rather challenging; see Figure 1 for an illustration with $r = 4$, and a single rare endmember only present in 1% of the samples.

a) Illustrative Example: The first experiment illustrates the difference between applying min-vol NMF to the traditional least-squares objective function using [12], and the minimax objective function via Algorithm 1. The endmembers

[1]Since we use the Dirichlet distribution that generates columns of H that are sparse, rare endmembers are not present in all the samples with an non-zero Dirichlet parameter for that endmember.

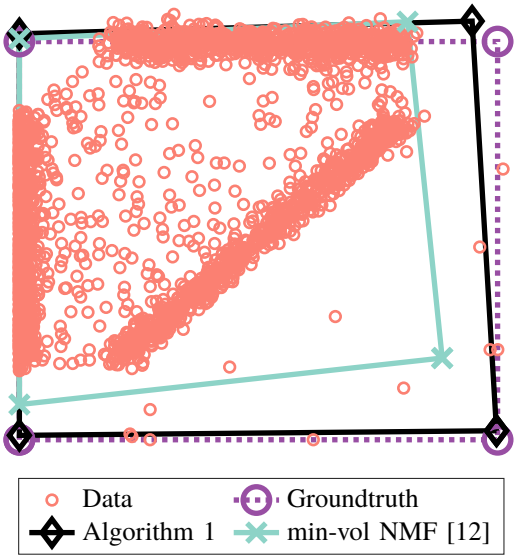

Fig. 1: True and estimated endmembers for the illustrative example projected onto the canonical basis vectors, $\mathbf{e}_2$ and $\mathbf{e}_3$.

are the columns of the matrix

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \tag{18}$$

with 2500 samples drawn according to Equation (17) with $\sigma_N^2 = 0.001$, and 1% of samples containing the rare endmember which is the last column of W. The penalty parameter $\tilde{\beta}$ must balance the data fidelity term with the minimum volume term. Each patch contains less samples than the entire data set so the parameter is chosen as $\tilde{\beta} = 10^{-3}$ for Algorithm 1 and $\tilde{\beta} = 0.1$ for min-vol NMF (as recommended in [12]). The eigenvalue weight is selected as $\delta = 0.1$ for both methods. The step size for the proposed method is $a = \frac{2}{\min_i \|X_i\|_F^2}$, and we perform each method for $maxiter = 200$ and $inneriter = 20$. Patches are 10×10 squares that tessellate the 50×50 image.

Figure 1 shows a projection of the data and results onto the 2-dimensional subspace spanned by the 2nd and 3rd canonical basis vectors, $\mathbf{e}_2$ and $\mathbf{e}_3$. The pink circles are the data samples generated from noisy convex combinations of the groundtruth endmembers (purple circles). The estimated endmembers using [12] are demarcated by turquoise x's and the estimated endmembers from the proposed method are the black diamonds. We observe the effect of the minimax reformulation clearly in this illustration. The endmember in the bottom right corner is the rare element, and although samples exist on all facets of the simplex, the least-squares estimate is further from the true endmember than the estimate from the proposed method. The relative error of reconstructing the endmembers, $\frac{\|W - W_{\text{est}}\|_F}{\|W\|_F}$ (after a proper permutation), is 17.1% for [12] and 4.8% for the proposed method.

The solid lines in Figure 2 show the effect of noise on the accuracy of estimating endmembers (left) and reconstructing the data (right) averaged over 20 runs. Min-vol NMF leads, on average, to lower global reconstruction error. This is expected since it minimizes this criterion. However, for low noise levels, both methods actually have almost the same reconstruction error (Figure 2, right). Moreover, the proposed method outperforms min-vol NMF in terms of estimating the endmembers when the noise level is sufficiently low, with an average improvement of about 10% (Figure 2, left).

b) Synthetic HSI Data with Groundtruth: The second set of experiments demonstrates the utility of the proposed method on more realistic synthetic data. Data is generated from $r = 10$ endmembers selected from the USGS Spectral Library [17] to create synthetic hyperspectral images. The spectral signatures of the endmembers used are shown in Figure 3, however the results are comparable for any collection signatures randomly selected from the library. The noisy bands 1-2, 221-224, and the water absorption bands 104-113, 148-167, have been removed from the data leaving 188 channels. The two spectra denoted by dashed lines in Figure 3, 'Chabazite HS193.3B' and 'Nepheline HS19.3', correspond to the rare endmembers. They have nonzero Dirichlet parameters for 2% of samples. Using the model in Equation (17), 2500 samples are generated to create $X \in \mathbb{R}^{188 \times 2500}$. Patches and hyperparameters are selected in the same way as in the first experiment.

This experiment is more challenging due to the number of endmembers, the additional rare endmember, the magnitudes of the endmembers, (min=25, max=151.5), and their spectral similarity (the minimum angle between two spectra is 0.06 radians). The relative error for this experiment is shown by the dashed lines in Figure 2. Even with these hurdles, we observe a similar behavior as in the first experiment: the difference in global reconstruction error is minor, Algorithm 1 reports 1% lower error for the smallest noise variance (Figure 2, right), but Algorithm 1 allows to recover the endmembers more accurately when the noise variance is low (Figure 2, left). The proposed method reduces the error of the endmember matrix of >13% on average at the lowest noise levels.

V. Conclusion

We proposed a novel heuristic for solving minimax min-vol NMF to improve the estimation of rare endmembers in blind hyperspectral unmixing. The method requires no pre-identification of pixels containing atypical materials. As the minimax paradigm fits the worst examples from the data set, it is key that the data be free from outliers before processing. Using a minimax strategy such as the one presented here, it is possible to trim outliers by removing the support set (the patches with non-zero dual variables) after the subgradient algorithm has converged and repeating the process [18], however some inliers may be rejected as well without more sophisticated methods.

References

[1] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2306–2320, 2015.

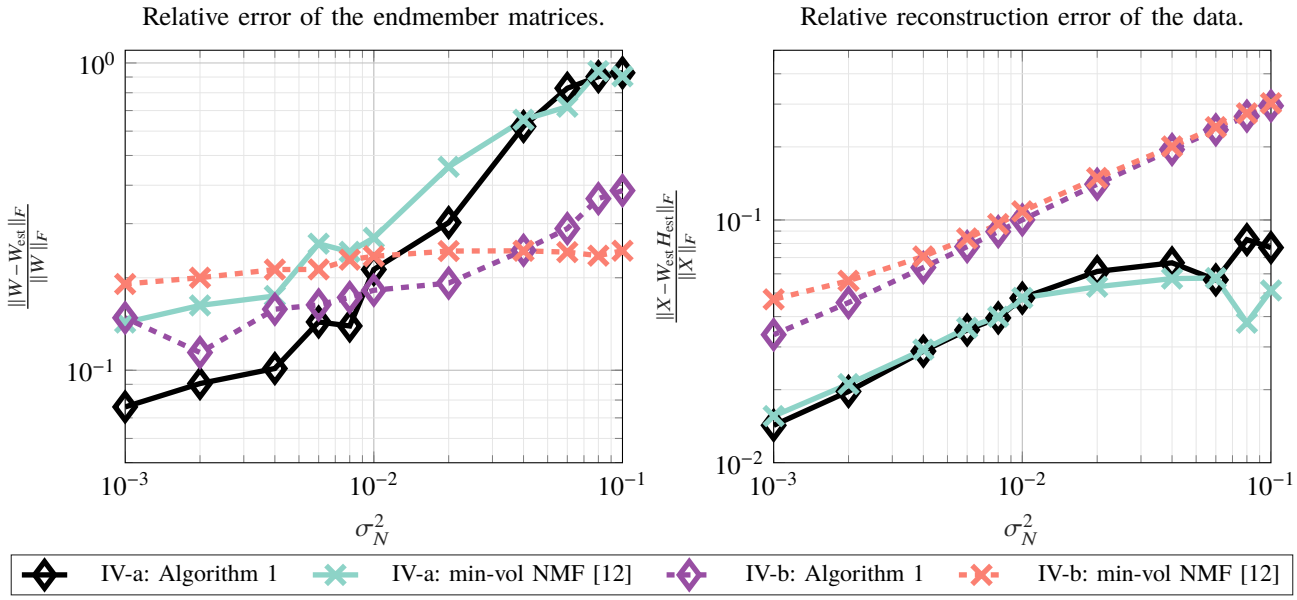

Fig. 2: Relative error comparisons for each the two experiments averaged over 20 runs. Experiment IV-a (solid lines) has 1 rare endmember (the last column of W in Equation (18)) present in 1% of samples, and Experiment IV-b (dashed lines) has 2 rare endmembers (the dashed spectra in Figure 3) present in 2% of samples.

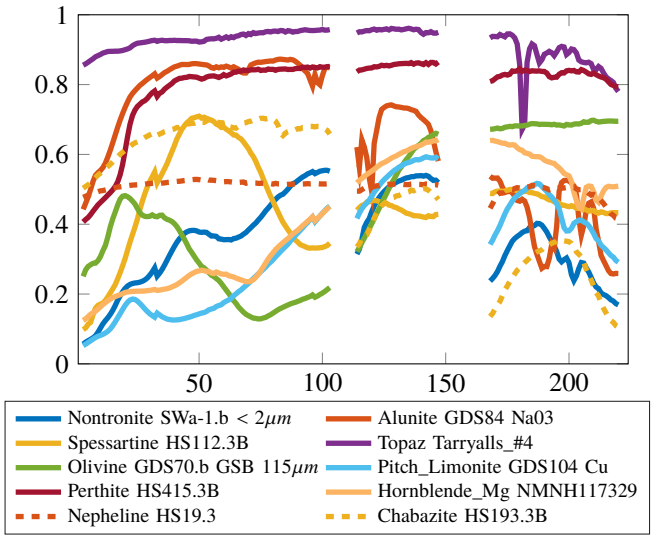

Fig. 3: Endmember spectra used for Experiment IV-b. Spectral signatures are taken from the USGS Spectral Library [17] with water absorption bands removed. The two dashed spectra are the rare endmembers.

[2] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, 2015.

[3] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of nonnegative matrix factorization," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 328–332, 2018.

[4] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications." *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 59–80, 2019.

[5] C.-I. Chang, W. Xiong, H.-M. Chen, and J.-W. Chai, "Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 504–520, 2011.

[6] S. Ravel, S. Bourennane, and C. Fossati, "Hyperspectral images unmixing with rare signals," in *Proc. 6th Eur. Workshop is. Inf. Process. (EUVIP)*. IEEE, 2016.

[7] C. Fossati, S. Bourennane, and A. Cailly, "Unmixing improvement based on bootstrap for hyperspectral imagery," in *Proc. 6th Eur. Workshop Vis. Inf. Process. (EUVIP)*. IEEE, 2016.

[8] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1364–1377, 2010.

[9] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, 1994.

[10] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, 2014.

[11] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6254–6268, 2016.

[12] V. Leplat, A. M. Ang, and N. Gillis, "Minimum-volume rank-deficient nonnegative matrix factorizations," in *Proc. IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP)*. IEEE, 2019.

[13] A. M. S. Ang and N. Gillis, "Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4843–4853, 2019.

[14] N. Z. Shor, *Minimization methods for non-differentiable functions*. Springer Science & Business Media, 2012, vol. 3.

[15] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[16] N. Gillis, "Successive nonnegative projection algorithm for robust nonnegative blind source separation," *SIAM J. Imaging Sci.*, vol. 7, no. 2, pp. 1420–1450, 2014.

[17] R. N. Clark, G. A. Swayze, A. Gallagher, T. King, and W. Calvin, "The U.S. Geological Survey digital spectral library," *Open File Rep.*, pp. 93–592, 1993.

[18] K. Sim and R. Hartley, "Removing outliers using the l_∞ norm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2006.