Spectral Unmixing With Perturbed Endmembers

Reza Arablouei

***Abstract*—We consider the problem of supervised spectral unmixing with a fully-perturbed linear mixture model where the given endmembers, as well as the observations of the spectral image, are subject to perturbation due to noise, error, mismatch, etc. We calculate the Fisher information matrix and the Cramer-Rao lower bound associated with the estimation of the abundance matrix in the considered fully-perturbed linear spectral unmixing problem. We develop an algorithm for estimating the abundance matrix by minimizing a constrained and regularized maximum-log-likelihood objective function using the block coordinate-descend iterations and the alternating direction method of multipliers. We analyze the convergence of the proposed algorithm theoretically and perform simulations with real hyperspectral image datasets to evaluate its performance. The simulation results corroborate the efficacy of the proposed algorithm in mitigating the adverse effects of perturbation in the endmembers.**

***Keywords*—Alternating direction method of multipliers, block coordinate-descent, Cramer-Rao lower bound, hyperspectral unmixing, instrumental variable, perturbed endmembers, total least-squares, total variation.**

I. Introduction

SPECTRAL unmixing is an important way of extracting knowledge from spectral image data. Each pixel of a multiband spectral image comprises the measurements related to several parts (bands) of the electromagnetic spectrum. The imaged area corresponding to each pixel often contains multiple materials that have distinct spectral signatures. The popular linear mixture model suggests that the spectral measurements (spectrum) of each pixel is a possibly noisy observation of a linear combination of the spectra of the materials present in the scene captured by the spectral image. These materials are considered to have unique spectral signatures called endmembers. The weight in the linear combination representing the contribution of an endmember to the spectrum of a pixel is called the abundance of that endmember in the pixel [1]–[7].

Since, in practice, the mixing is realized with actual photons, the abundances are expected to be nonnegative. In addition, if one assumes that the considered linear mixture model accounts for all endmembers as well as the composition of the spectra of all pixels, it is reasonable to expect the abundances of each pixel add up to one. However, this assumption may not strictly hold when there is nonlinearity in the mixture process or variability in the endmembers [8], [9].

Linear spectral unmixing is the practice of inferring the abundance values for all pixels through adopting the linear mixture model and exploiting prior knowledge or assumptions that ensure the feasibility and enhance the accuracy of the inference. If the endmembers are not known beforehand, the process of spectral unmixing is a blind (unsupervised) source separation task and involves the inference of both endmembers and abundances. The extraction of the endmembers and the estimation of the abundances can be done simultaneously or at separate stages [10]–[20].

When an over-complete set of candidate endmembers, such as a library of spectral signatures, is available, semi-supervised spectral unmixing algorithms based on spares regression can yield both the endmembers and the abundances. Usually the number of available signatures in spectral libraries is large but the number of endmembers present in a typical spectral image is comparatively small. Therefore, linear sparse regression techniques can efficiently determine the endmembers as an optimal subset of the given spectral library as well as estimating the abundances by employing appropriate sparsity-inducing regularizers [21]–[27].

If the endmembers are already known or identified in a previous processing stage, the spectral unmixing is an instance of supervised source separation and can be cast as a constrained linear regression problem. In [28], a constrained least-squares algorithm for supervised spectral unmixing is proposed that generalizes the active-set-based nonnegative least-squares algorithm popularized by [29] via augmenting the underlying system of linear equations to incorporate the abundance sum-to-one constraint. In [30], the spectral unmixing problem is treated in the Bayesian estimation context and the prior distribution for the abundances is specified to satisfy the abundance nonnegativity and sum-to-one constraints. Then, an algorithm based on the Markov-chain Monte Carlo method is developed to estimate the abundances by drawing samples from the associated posterior distribution. In [31], the underlying constrained least-squares problem is solved using a primal-dual interior-point method claimed to be amenable to parallel computing. In [32], Dykstra's algorithm for finding projections onto the intersection of convex sets is employed to estimate the abundances, which are required to satisfy the nonnegativity and sum-to-one constraints.

The above-mentioned supervised spectral unmixing algorithms assume that the endmembers are known exactly. However, in practice, the endmembers are extracted from real spectral images that may be contaminated by measurement noise or error. In addition, the available knowledge of the endmembers might not exactly match the true endmembers of the spectral image at hand since the spectral signature of the same material may be slightly altered in different images or distinct but confusingly similar spectral signatures may be mixed up. Unaccounted nonlinear mixing phenomena or unsuspected variability in the endmembers throughout the same image can also be sources of perturbation in the endmembers.

In this paper, we consider a supervised linear spectral unmixing problem where both the spectral image data and

R. Arablouei is with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Pullenvale QLD 4069, Australia (email: reza.arablouei@csiro.au).

the available knowledge of the endmembers are subject to perturbation. We calculate the corresponding (un)constrained Fisher information matrix (FIM) and Cramer-Rao lower bound (CRLB) for the estimation of the abundance matrix based on the considered model. In order to account for the perturbations on the observations of both spectral image and endmembers, we formulate a constrained weighted maximum-log-likelihood objective function and regularize it using two penalty terms, i.e., the vector total-variation of the abundance matrix and the sum-squared-distance of the endmembers. We develop an algorithm for solving the formulated optimization problem by utilizing the block coordinate-descent (BCD) iterations [33]–[35] together with the alternating direction method of multipliers (ADMM) [36]–[40]. We examine the convergence properties of the proposed algorithm theoretically and evaluate its unmixing performance through simulation experiments with three real hyperspectral image datasets. The simulation results show that the proposed algorithm copes with the endmember perturbations effectively and yields a significantly more accurate estimation of the abundance matrix compared with the conventional constrained least-squares algorithms that are oblivious to any perturbation in the endmembers.

II. Data Model

Let us denote a multiband spectral image that is made of N pixels and L spectral bands in the matrix form as $\mathbf{X} \in \mathbb{R}^{L\times N}$. There are K linearly-independent endmembers[1] associated with this image that are arranged as the columns of the endmember matrix $\mathbf{E} \in \mathbb{R}^{L\times K}$. According to the linear mixture model, each column of $\mathbf{X}$ representing the spectral values of the corresponding pixel can be written as a linear (conical) combination of the endmembers. Therefore, $\mathbf{X}$ can be factorized as

$$\mathbf{X} = \mathbf{E}\mathbf{A} \tag{1}$$

where the columns of $\mathbf{A} \in \mathbb{R}^{K\times N}$ contain the fractional abundances of the endmembers for all pixels. In supervised spectral unmixing, we are primarily interested in estimating the abundance matrix $\mathbf{A}$ given the observations/measurements of $\mathbf{X}$ and the knowledge of $\mathbf{E}$ extracted from the available data or taken from a library of candidate endmembers based on some prior information.

In practice, multiband spectral images are collected via spectral imaging devices that are prone to measurement noise or error. The linear mixture model (1) is often not perfect either. Therefore, we presume to have access to a perturbed observation of the original spectral image $\mathbf{X}$ that is denoted by $\tilde{\mathbf{X}}$ and expressed as

$$\tilde{\mathbf{X}} = \mathbf{E}\mathbf{A} + \mathbf{P}_{\mathbf{X}} \tag{2}$$

where $\mathbf{P}_{\mathbf{X}} \in \mathbb{R}^{L\times N}$ is the perturbation matrix factoring in measurement noise/error as well as possible shortfall of the linear mixture model. In addition, we consider that instead of the exact values of the endmember matrix $\mathbf{E}$, we have access to a perturbed version of it denoted by $\tilde{\mathbf{E}}$ and related to $\mathbf{E}$ as

$$\tilde{\mathbf{E}} = \mathbf{E} + \mathbf{P}_{\mathbf{E}} \tag{3}$$

where $\mathbf{P}_{\mathbf{E}} \in \mathbb{R}^{L\times K}$ represents the endmember perturbation. The equations (2) and (3) make up our fully-perturbed linear mixture model.

We assume that the perturbation matrices $\mathbf{P}_{\mathbf{X}}$ and $\mathbf{P}_{\mathbf{E}}$ are statistically independent of each other and have matrix normal distributions such that

$$\mathbf{P}_{\mathbf{X}} \sim \mathcal{MN}_{L\times N}\left(\mathbf{0}_{L\times N}, \mathbf{\Sigma}, \mathbf{I}_N\right) \tag{4}$$

$$\mathbf{P}_{\mathbf{E}} \sim \mathcal{MN}_{L\times K}\left(\mathbf{0}_{L\times K}, \mathbf{\Xi}, \mathbf{I}_K\right) \tag{5}$$

where $\mathbf{0}_{L\times N}$ is an $L \times N$ matrix with all zero entries, $\mathbf{I}_L$ is the $L \times L$ identity matrix, and $\mathbf{\Sigma} \in \mathbb{R}^{L\times L}$ and $\mathbf{\Xi} \in \mathbb{R}^{L\times L}$ are diagonal matrices with positive diagonal entries denoting the respective row-covariance matrices for $\mathbf{P}_{\mathbf{X}}$ and $\mathbf{P}_{\mathbf{E}}$. Note that we set the column-covariance matrices to identity to reflect the realistic assumption that the perturbations pertaining to different pixels or endmembers are independent and identically distributed. However, the diagonal row-covariance matrices imply that the perturbations are independent of each other in the spectral domain but may have different variances at different bands.

It is worth mentioning that the aforesaid fully-perturbed linear mixture model is essentially different from the ones recently proposed in the literature that have been particularly designed to cope with endmember variability or nonlinear mixing effects such as those of [41]–[43]. The perturbed linear mixing model utilized in [41] explicitly accounts for the spatial and spectral variability of the endmembers by explaining the endmember variability in each pixel through perturbations added to the reference endmembers. In [42], an extended linear mixing model that allows for pixelwise spatially-coherent local variations of the endmembers is adopted. A special case of this model stipulates that each pixel is a linear combination of scaled versions of the reference endmembers where the scaling factors may differ among pixels. The model assumed in [43] is an augmentation of the linear mixture model with an additive residual term that is meant to incorporate the effects of endmember variability, nonlinear mixing, mismodeling, and outliers. Therefore, in general, the mixture models used in [41]–[43] and similar works assume that the effective endmembers can vary from pixel to pixel, hence the perturbations of the reference endmembers can be different at each pixel. However, in our model, i.e., (2) and (3), we assume that the mixing occurs linearly and the endmembers are the same for all pixels while we have access only to a single perturbed version of the endmember matrix.

III. Problem

In view of (2)-(5), we have

$$\tilde{\mathbf{X}} \sim \mathcal{MN}_{L\times N}\left(\mathbf{E}\mathbf{A}, \mathbf{\Sigma}, \mathbf{I}_N\right)$$

$$\tilde{\mathbf{E}} \sim \mathcal{MN}_{L\times K}\left(\mathbf{E}, \mathbf{\Xi}, \mathbf{I}_K\right).$$

[1] Technically, the linear mixture model implicates that the endmembers are *affinely* independent. However, since the number of endmembers K is usually much smaller that the number of spectral bands L, it is often safe to assume that the endmembers are linearly independent.

Thus, the probability density functions of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{E}}$ parameterized over the unknown matrix parameters $\mathbf{E}$ and $\mathbf{A}$ are written as

$$p_{\tilde{\mathbf{X}}}\left(\tilde{\mathbf{X}};\mathbf{E},\mathbf{A}\right) \propto \exp\left[-\frac{1}{2}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\tilde{\mathbf{X}}-\mathbf{E}\mathbf{A}\right)\right\|_{\mathsf{F}}^{2}\right]$$

$$p_{\tilde{\mathbf{E}}}\left(\tilde{\mathbf{E}};\mathbf{E}\right) \propto \exp\left[-\frac{1}{2}\left\|\boldsymbol{\Xi}^{-1/2}\left(\tilde{\mathbf{E}}-\mathbf{E}\right)\right\|_{\mathsf{F}}^{2}\right]$$

where $\|\cdot\|_{\mathsf{F}}$ denotes the Frobenius norm and $\propto$ means proportional up to a constant that does not depend on $\mathbf{E}$, $\mathbf{A}$, or their observations. Therefore, we define a weighted likelihood function [44] of the unknowns $\mathbf{E}$ and $\mathbf{A}$ given their observations as

$$\begin{aligned}\ell\left(\mathbf{E},\mathbf{A}\mid\tilde{\mathbf{X}},\tilde{\mathbf{E}}\right) &= p_{\tilde{\mathbf{X}}}\left(\tilde{\mathbf{X}};\mathbf{E},\mathbf{A}\right)\left[p_{\tilde{\mathbf{E}}}\left(\tilde{\mathbf{E}};\mathbf{E}\right)\right]^{\alpha}\\ &\propto \exp\left[-\frac{1}{2}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\tilde{\mathbf{X}}-\mathbf{E}\mathbf{A}\right)\right\|_{\mathsf{F}}^{2}-\frac{\alpha}{2}\left\|\boldsymbol{\Xi}^{-1/2}\left(\tilde{\mathbf{E}}-\mathbf{E}\right)\right\|_{\mathsf{F}}^{2}\right]\end{aligned}$$

where α is a positive weight parameter. Hence, the corresponding weighted log-likelihood function is given by

$$\begin{aligned}l\left(\mathbf{E},\mathbf{A}\right) = &-\frac{1}{2}\left\|\boldsymbol{\Sigma}^{-1/2}\left(\tilde{\mathbf{X}}-\mathbf{E}\mathbf{A}\right)\right\|_{\mathsf{F}}^{2}\\ &-\frac{\alpha}{2}\left\|\boldsymbol{\Xi}^{-1/2}\left(\tilde{\mathbf{E}}-\mathbf{E}\right)\right\|_{\mathsf{F}}^{2}.\end{aligned} \tag{6}$$

Remark 1: The weighted log-likelihood function (6) is equivalent to the canonical log-likelihood function when $\alpha = 1$. We consider this weighted form of the log-likelihood function since it has the advantage of being viewed as an approximation of the genuine log-likelihood function when there is correlation between $\mathbf{P}_{\mathbf{X}}$ and $\mathbf{P}_{\mathbf{E}}$. In such cases, an appropriate value of α can help achieve a balance between the contributions and effects of the two summands of the log-likelihood function.

A. Cramer-Rao Lower Bound

In order to gain insights into the theoretical performance limits given the above-mentioned data model and the weighted log-likelihood function (6), we derive the CRLB for the estimation of $\mathbf{A}$ while treating $\mathbf{E}$ as a nuisance matrix parameter.

For the convenience of the notation, we define

$$\begin{aligned}\mathbf{e} &= \operatorname{vec}\left\{\mathbf{E}\right\}\\ \mathbf{a} &= \operatorname{vec}\left\{\mathbf{A}^{\mathsf{T}}\right\}\end{aligned}$$

where $\operatorname{vec}\{\cdot\}$ is the vectorization operator that stacks the columns of its matrix argument on top of each other. The gradients of the function $l\left(\mathbf{E},\mathbf{A}\right)$ with respect to its arguments are calculated as

$$\begin{aligned}\mathbf{g}_{\mathbf{e}} &= \frac{\partial l\left(\mathbf{E},\mathbf{A}\right)}{\partial \mathbf{e}^{\mathsf{T}}}\\ &= \operatorname{vec}\left\{\boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{X}}-\mathbf{E}\mathbf{A}\right)\mathbf{A}^{\mathsf{T}}+\alpha\boldsymbol{\Xi}^{-1}\left(\tilde{\mathbf{E}}-\mathbf{E}\right)\right\}\\ \mathbf{g}_{\mathbf{a}} &= \frac{\partial l\left(\mathbf{E},\mathbf{A}\right)}{\partial \mathbf{a}^{\mathsf{T}}}\\ &= \operatorname{vec}\left\{\left(\tilde{\mathbf{X}}^{\mathsf{T}}-\mathbf{A}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}\right)\boldsymbol{\Sigma}^{-1}\mathbf{E}\right\}.\end{aligned}$$

It is easy to verify the following expectations

$$\begin{aligned}\mathbb{E}\left[\mathbf{g}_{\mathbf{e}}\right] &= \mathbf{0}_{KL\times 1}\\ \mathbb{E}\left[\mathbf{g}_{\mathbf{a}}\right] &= \mathbf{0}_{KN\times 1},\end{aligned}$$

which indicate the satisfaction of the relevant regularity conditions [45]. Hence, the corresponding Fisher information matrix (FIM) is expressed as

$$\begin{aligned}\mathbf{F} &= \begin{bmatrix}\mathbf{F}_{11} & \mathbf{F}_{12}\\ \mathbf{F}_{21} & \mathbf{F}_{22}\end{bmatrix}\\ &= -\mathbb{E}\begin{bmatrix}\frac{\partial \mathbf{g}_{\mathbf{e}}}{\partial \mathbf{e}^{\mathsf{T}}} & \frac{\partial \mathbf{g}_{\mathbf{e}}}{\partial \mathbf{a}^{\mathsf{T}}}\\ \frac{\partial \mathbf{g}_{\mathbf{a}}}{\partial \mathbf{e}^{\mathsf{T}}} & \frac{\partial \mathbf{g}_{\mathbf{a}}}{\partial \mathbf{a}^{\mathsf{T}}}\end{bmatrix}.\end{aligned} \tag{7}$$

In Appendix A, we calculate the partial derivatives in (7) and show that

$$\mathbf{F} = \begin{bmatrix}\mathbf{A}\mathbf{A}^{\mathsf{T}}\otimes\boldsymbol{\Sigma}^{-1}+\mathbf{I}_K\otimes\alpha\boldsymbol{\Xi}^{-1} & \mathbf{P}_{L,K}\left(\boldsymbol{\Sigma}^{-1}\mathbf{E}\otimes\mathbf{A}\right)\\ \left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\otimes\mathbf{A}^{\mathsf{T}}\right)\mathbf{P}_{K,L} & \mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\otimes\mathbf{I}_N\end{bmatrix} \tag{8}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{P}_{L,K}$ and $\mathbf{P}_{K,L}$ are the vec-permutation matrices of appropriate dimensions as explained in Appendix A.

We show in Appendix B that $\mathbf{F}$ is non-singular and invertible. Hence, according to the block matrix inversion formula [46], the $KN\times KN$ lower-right block of $\mathbf{F}^{-1}$ is the inverse of the Schur complement of $\mathbf{F}_{11}$, i.e., $\left(\mathbf{F}_{22}-\mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)^{-1}$. Therefore, the CRLB on the variance or mean square-error (MSE) of any unbiased estimator of $\mathbf{A}$, denoted by $\hat{\mathbf{A}}$, is given by [45]

$$\mathbb{E}\left[\left\|\hat{\mathbf{A}}-\mathbf{A}\right\|_{\mathsf{F}}^{2}\right] \geq \operatorname{tr}\left\{\left(\mathbf{F}_{22}-\mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)^{-1}\right\} \tag{9}$$

where $\operatorname{tr}\{\cdot\}$ is the matrix trace operator. We show in Appendix C that when the row-covariance matrix of perturbation of the endmembers is a scalar multiple of the row-covariance matrix of the perturbation of the observed spectral image, i.e., $\boldsymbol{\Xi} = \eta\boldsymbol{\Sigma}$ with $\eta > 0$, (9) becomes

$$\mathbb{E}\left[\left\|\hat{\mathbf{A}}-\mathbf{A}\right\|_{\mathsf{F}}^{2}\right] \geq \operatorname{tr}\left\{\left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1}\right\}\left(N+\frac{\eta}{\alpha}\left\|\mathbf{A}\right\|_{\mathsf{F}}^{2}\right). \tag{10}$$

Remark 2: The FIM is regular (non-singular) thanks to the additive term $\mathbf{I}_K\otimes\alpha\boldsymbol{\Xi}^{-1}$ with $\alpha > 0$ in its upper-left block (see Appendix B), which in turn is due to the second additive term of the log-likelihood function [see (6)]. This makes the abundance matrix $\mathbf{A}$ identifiable even with no constraint on $\mathbf{A}$ or $\mathbf{E}$. Indeed, the second summand on the right-hand side of (6) serves to eliminate the identifiability ambiguities inherent to the first summand, which, if considered alone, entails a weighted nonnegative matrix factorization (NMF) problem [47].

Remark 3: When the perfect knowledge of the endmembers is available, the FIM reduces to

$$\mathbf{F} = \mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\otimes\mathbf{I}_N$$

and the CRLB on the MSE of $\hat{\mathbf{A}}$ to

$$\mathbb{E}\left[\left\|\hat{\mathbf{A}}-\mathbf{A}\right\|_{\mathsf{F}}^{2}\right]\geq\operatorname{tr}\left\{\left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1}\right\}. \tag{11}$$

Therefore, one can argue that the perturbation of the endmembers incurs an information loss regarding the estimation of $\mathbf{A}$ quantifiable by $\mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}$ and a consequent increase in the CRLB. When $\boldsymbol{\Xi}=\eta\boldsymbol{\Sigma}$, as shown in Appendix C, this information loss can be expressed as

$$\mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}=\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\otimes\mathbf{A}^{\mathsf{T}}\left(\mathbf{A}\mathbf{A}^{\mathsf{T}}+\frac{\alpha}{\eta}\mathbf{I}_{K}\right)^{-1}\mathbf{A}$$

and, as seen from (10) and (11), the increase in the CRLB is by a factor of $N+\frac{\eta}{\alpha}\left\|\mathbf{A}\right\|_{\mathsf{F}}^{2}$. Notice that a carefully selected non-unity value for α can help curb the information loss and the resultant increase in the CRLB incurred by the perturbation in the endmembers, although at the cost of model inaccuracy and increased estimation bias.

B. Constrained Cramer-Rao Lower Bound

It is shown in [48]–[50] that the nonnegativity (or any other inequality) constraint on the unknown parameter (in our case, $\mathbf{A}$) does not affect the CRLB. However, equality constraints such as the sum-to-one constraint on the columns of $\mathbf{A}$ can affect the CRLB. At the presence of equality constraints, the full-rank FIM corresponding to the unconstrained case needs to be substituted by a reduced-rank version that is obtained by projecting the full-rank FIM onto the tangent hyperplane of the constraint set [48].

We can express the sum-to-one constraint on the columns of $\mathbf{A}$ as $\mathbf{1}_{K}^{\mathsf{T}}\mathbf{A}=\mathbf{1}_{N}^{\mathsf{T}}$ or equivalently

$$\begin{aligned}\mathbf{c}\left(\mathbf{A}\right)&\triangleq\mathbf{A}^{\mathsf{T}}\mathbf{1}_{K}-\mathbf{1}_{N}\\&=\mathbf{0}_{N\times1}\end{aligned} \tag{12}$$

where $\mathbf{1}_{Q}$ denotes a $Q\times1$ vector with all entries being unity. The Jacobian of the constraint function $\mathbf{c}\left(\mathbf{A}\right)$ is computed as

$$\begin{aligned}\mathbf{J}&=\frac{\partial\mathbf{c}\left(\mathbf{A}\right)}{\partial\mathbf{a}^{\mathsf{T}}}\\&=\mathbf{1}_{K}^{\mathsf{T}}\otimes\mathbf{I}_{N}.\end{aligned}$$

Considering Theorem 1 of [49], the CRLB of MSE for estimation of $\mathbf{A}$ under the constraint (12) is given by

$$\mathbb{E}\left[\left\|\hat{\mathbf{A}}-\mathbf{A}\right\|_{\mathsf{F}}^{2}\right]\geq\operatorname{tr}\left\{\mathbf{H}\left[\mathbf{H}^{\mathsf{T}}\left(\mathbf{F}_{22}-\mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\mathbf{H}\right]^{-1}\mathbf{H}^{\mathsf{T}}\right\} \tag{13}$$

where $\mathbf{H}\in\mathbb{R}^{KN\times(K-1)N}$ is a matrix whose columns form an orthonormal basis for the nullspace of the constraint Jacobian matrix $\mathbf{J}$. It is clear from (13) that the constrained CRLB is invariant to any automorphism on $\mathbf{H}$. Therefore, $\mathbf{H}$ only requires to have linearly-independent columns that span the nullspace of $\mathbf{J}$. Hence, we set $\mathbf{H}=\mathbf{Y}\otimes\mathbf{I}_{N}$ where the ith column of $\mathbf{Y}\in\mathbb{R}^{K\times(K-1)}$, denoted by $\mathbf{y}_{i}$, is constructed as

$$\mathbf{y}_{i}=\begin{bmatrix}\mathbf{1}_{i}\\-i\\\mathbf{0}_{(K-i-1)\times1}\end{bmatrix},i=1,\cdots,K-1,$$

so that $\mathbf{Y}$ spans the orthogonal complement of $\mathbf{1}_{K}^{\mathsf{T}}$.

We show in Appendix C that, when $\boldsymbol{\Xi}=\eta\boldsymbol{\Sigma}$, (13) turns into

$$\begin{aligned}&\mathbb{E}\left[\left\|\hat{\mathbf{A}}-\mathbf{A}\right\|_{\mathsf{F}}^{2}\right]\\&\geq\operatorname{tr}\left\{\mathbf{Y}\left(\mathbf{Y}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathsf{T}}\right\}\left(N+\frac{\eta}{\alpha}\left\|\mathbf{A}\right\|_{\mathsf{F}}^{2}\right).\end{aligned} \tag{14}$$

C. Constrained Maximum-Likelihood Estimation

The constrained maximum-likelihood estimates of $\mathbf{A}$ and $\mathbf{E}$ can be found by maximizing the weighted log-likelihood function (6) with respect to its arguments while constraining the entries of $\mathbf{A}$ to be nonnegative or additionally the columns of $\mathbf{A}$ to add up to unity. As the knowledge of the perturbation covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Xi}$ is often unavailable, we drop them from the weighted log-likelihood objective function by setting them to identity and rely on the parameter α to adjust for their possible disparity. Thus, the resulting optimization problem associated with the constrained maximum-likelihood estimation of $\mathbf{A}$ (and $\mathbf{E}$ as a nuisance or an auxiliary matrix variable) is

$$\begin{aligned}\min_{\mathbf{E},\mathbf{A}}\quad&\frac{1}{2}\left\|\tilde{\mathbf{X}}-\mathbf{E}\mathbf{A}\right\|_{\mathsf{F}}^{2}+\frac{\alpha}{2}\left\|\tilde{\mathbf{E}}-\mathbf{E}\right\|_{\mathsf{F}}^{2}\\\text{s.t.}\quad&\mathbf{A}\in\mathcal{S}\end{aligned} \tag{15}$$

where $\mathcal{S}$ is the set of all $K\times N$ matrices that have entries in the range $[0,1]$. If the sum-to-one constraint on the columns of $\mathbf{A}$ is also desired, $\mathcal{S}$ will be the standard $(K-1)$-simplex.

The nonnegativity constraint is required to obtain physically plausible abundance values. We include the abundance sum-to-one constraint as an option and not a necessity as, in some scenarios, it may not be practically justifiable due to issues such as endmember variability [1]. We do not impose any constraint on $\mathbf{E}$ since, in this paper, we are only interested in the estimation of $\mathbf{A}$ and treat $\mathbf{E}$ in (15) and the subsequent related optimization problems as an auxiliary matrix variable. In addition, in our numerical experiments, we did not find any noticeable benefit in constraining the entries of $\mathbf{E}$ to be nonnegative, specifically, for low to moderate levels of perturbation in the endmembers.

D. Penalties

In addition to the above-mentioned constraints on the abundance matrix $\mathbf{A}$, we add two penalty terms to the objective function in (15). The first term is an isotropic vector total-variation penalty [51]–[53], defined as

$$\left\|\nabla\mathbf{A}\right\|_{2,1}=\left\|\begin{bmatrix}\mathbf{A}\mathbf{D}_{h}\\\mathbf{A}\mathbf{D}_{v}\end{bmatrix}\right\|_{2,1}$$

where $\left\|\cdot\right\|_{2,1}$ is the $\ell_{2,1}$-norm operator that returns the sum of ℓ_{2}-norms of all the columns of its matrix argument, and $\mathbf{D}_{h}$ and $\mathbf{D}_{v}$ are discrete differential matrix operators that apply independently on all abundance bands and, respectively, yield the horizontal and vertical first-order backward differences (gradients) of the abundance bands in the spatial domain.

Natural images are known to mostly consist of piecewise homogeneous regions with few discontinuities and abrupt changes at object boundaries or edges. The vector total-variation regularization allows us to take advantage of this prior knowledge by promoting sparsity in the spatial gradient of the abundance bands, i.e., local differences between the values for adjacent pixels, while encouraging the local differences to be spatially aligned across different bands [53].

The second penalty term that we add to the objective function is the so-called sum-squared-distance of endmembers, defined as

$$\|\mathbf{EW}\|_{\mathsf{F}}^2 = \left\|\mathbf{E}\left(\mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^{\mathsf{T}}\right)\right\|_{\mathsf{F}}^2.$$

This term is the sum of squared distances between all the endmembers, which are considered to be the vertices of the simplex that circumscribes the spectral data of all pixels in the image. It was originally proposed in [11] within the context of blind hyperspectral unmixing as a computationally-efficient surrogate for the volume of the simplex formed by the endmembers and has since been reproposed through various interpretations, e.g., in [12]. Adding this penalty term to the objective function can help the estimates of the endmember matrix stay close to the true value hence aid in estimating a more accurate abundance matrix.

Consequently, in order to develop an effective algorithm for unmixing multiband spectral images based on the fully-perturbed linear mixture model, we propose to solve the following optimization problem that incorporates the above constraints and penalty terms

$$\begin{aligned}\min_{\mathbf{E},\mathbf{A}} \quad & \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{EA}\right\|_{\mathsf{F}}^2 + \frac{\alpha}{2}\left\|\tilde{\mathbf{E}} - \mathbf{E}\right\|_{\mathsf{F}}^2 + \beta\|\nabla\mathbf{A}\|_{2,1} \\ & + \frac{\gamma}{2}\|\mathbf{EW}\|_{\mathsf{F}}^2 \\ \text{s.t.} \quad & \mathbf{A} \in \mathcal{S}\end{aligned} \tag{16}$$

where $\beta > 0$ and $\gamma > 0$ are regularization parameters that together with α govern the trade-offs arising from assigning different weights to various terms of the objective function.

IV. Algorithm

A. Block Coordinate-Descent Iterations

We use the block coordinate-descent (BCD) algorithm to solve (16) with respect to $\mathbf{E}$ and $\mathbf{A}$ alternatively and in an iterative fashion. Therefore, we repeat the following alternating minimizations until the iterates converge

$$\begin{aligned}\mathbf{E}^{(n)} = \underset{\mathbf{E}}{\operatorname{argmin}} \quad & \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{EA}^{(n-1)}\right\|_{\mathsf{F}}^2 + \frac{\alpha}{2}\left\|\tilde{\mathbf{E}} - \mathbf{E}\right\|_{\mathsf{F}}^2 \\ & + \frac{\gamma}{2}\|\mathbf{EW}\|_{\mathsf{F}}^2\end{aligned} \tag{17}$$

$$\begin{aligned}\mathbf{A}^{(n)} = \underset{\mathbf{A}}{\operatorname{argmin}} \quad & \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{E}^{(n)}\mathbf{A}\right\|_{\mathsf{F}}^2 + \beta\|\nabla\mathbf{A}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{A} \in \mathcal{S}.\end{aligned} \tag{18}$$

Here, a variable with the superscript (n) denotes the estimate of its respective parameter at the nth iteration.

The minimization with respect to $\mathbf{E}$ in (17) is straightforward and results in

$$\begin{aligned}\mathbf{E}^{(n)} = & \left(\tilde{\mathbf{X}}\mathbf{A}^{(n-1)\mathsf{T}} + \alpha\tilde{\mathbf{E}}\right) \\ & \times \left(\mathbf{A}^{(n-1)}\mathbf{A}^{(n-1)\mathsf{T}} + \alpha\mathbf{I}_K + \gamma\mathbf{W}\right)^{-1}\end{aligned}$$

where we use the fact that the matrix $\mathbf{W}$ is symmetric and idempotent, i.e., $\mathbf{WW}^{\mathsf{T}} = \mathbf{W}$.

B. Alternating Direction Method of Multipliers

For the minimization with respect to $\mathbf{A}$ in (18), we define a set of auxiliary matrix variables, called $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4 \in \mathbb{R}^{K\times N}$, and apply variable splitting to break the problem into smaller pieces. Therefore, we find an estimate of $\mathbf{A}^{(n)}$ by solving the following problem

$$\begin{aligned}\min_{\mathbf{A},\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{V}_4} \quad & \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{E}^{(n)}\mathbf{A}\right\|_{\mathsf{F}}^2 + \beta\left\|\begin{bmatrix}\mathbf{V}_2\\ \mathbf{V}_3\end{bmatrix}\right\|_{2,1} + \imath_{\mathcal{S}}(\mathbf{V}_4) \\ \text{s.t.} \quad & \mathbf{V}_1 = \mathbf{A}, \mathbf{V}_2 = \mathbf{V}_1\mathbf{D}_h, \mathbf{V}_3 = \mathbf{V}_1\mathbf{D}_v, \\ & \mathbf{V}_4 = \mathbf{A}\end{aligned} \tag{19}$$

where $\imath_{\mathcal{S}}(\cdot)$ is the indicator function that takes the value of zero when its matrix argument is in the set $\mathcal{S}$ and the value of infinity otherwise.

The augmented Lagrangian function for the constrained optimization problem (19) is given by

$$\begin{aligned}&\mathcal{L}(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4) \\ &\quad = \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{E}^{(n)}\mathbf{A}\right\|_{\mathsf{F}}^2 + \beta\left\|\begin{bmatrix}\mathbf{V}_2\\ \mathbf{V}_3\end{bmatrix}\right\|_{2,1} + \imath_{\mathcal{S}}(\mathbf{V}_4) \\ &\quad\quad + \frac{\mu}{2}\|\mathbf{A} - \mathbf{V}_1 - \mathbf{G}_1\|_{\mathsf{F}}^2 + \frac{\mu}{2}\|\mathbf{V}_1\mathbf{D}_h - \mathbf{V}_2 - \mathbf{G}_2\|_{\mathsf{F}}^2 \\ &\quad\quad + \frac{\mu}{2}\|\mathbf{V}_1\mathbf{D}_v - \mathbf{V}_3 - \mathbf{G}_3\|_{\mathsf{F}}^2 + \frac{\lambda}{2}\|\mathbf{A} - \mathbf{V}_4 - \mathbf{G}_4\|_{\mathsf{F}}^2\end{aligned} \tag{20}$$

where $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4 \in \mathbb{R}^{K\times N}$ are the scaled Lagrange multipliers and $\mu \geq 0$ and $\lambda \geq 0$ are the penalty parameters. We use the alternating direction method of multipliers (ADMM) to minimize the augmented Lagrangian function (20) in an iterative fashion. At each iteration, we alternate the minimization with respect to the main unknown variable $\mathbf{A}$ and the auxiliary variables; then, we update the scaled Lagrange multipliers. Hence, we compute the iterates as

$$\begin{aligned}\mathbf{A}^{(n,m)} = \underset{\mathbf{A}}{\operatorname{argmin}}\, \mathcal{L}\Big(&\mathbf{A}, \mathbf{V}_1^{(m-1)}, \cdots, \mathbf{V}_4^{(m-1)}, \\ &\mathbf{G}_1^{(m-1)}, \cdots, \mathbf{G}_4^{(m-1)}\Big)\end{aligned} \tag{21}$$

$$\begin{aligned}\left\{\mathbf{V}_1^{(m)}, \cdots, \mathbf{V}_4^{(m)}\right\} = \underset{\mathbf{V}_1,\cdots,\mathbf{V}_4}{\operatorname{argmin}}\, \mathcal{L}\Big(&\mathbf{A}^{(n,m)}, \mathbf{V}_1, \cdots, \mathbf{V}_4, \\ &\mathbf{G}_1^{(m-1)}, \cdots, \mathbf{G}_4^{(m-1)}\Big)\end{aligned} \tag{22}$$

$$\mathbf{G}_1^{(m)} = \mathbf{G}_1^{(m-1)} - \left(\mathbf{A}^{(n,m)} - \mathbf{V}_1^{(m)}\right)$$
$$\mathbf{G}_2^{(m)} = \mathbf{G}_2^{(m-1)} - \left(\mathbf{V}_1^{(m)}\mathbf{D}_h - \mathbf{V}_2^{(m)}\right)$$
$$\mathbf{G}_3^{(m)} = \mathbf{G}_3^{(m-1)} - \left(\mathbf{V}_1^{(m)}\mathbf{D}_v - \mathbf{V}_3^{(m)}\right)$$
$$\mathbf{G}_4^{(m)} = \mathbf{G}_4^{(m-1)} - \left(\mathbf{A}^{(n,m)} - \mathbf{V}_4^{(m)}\right).$$

Note that the above ADMM iterations are sub-iterations within the BCD iterations of (17) and (18). That is why we use m as the index for the ADMM iterations to differentiate them from the BCD iterations indexed by n. At the end of the ADMM iterations within every BCD iteration, we set the value of $\mathbf{A}^{(n)}$ equal to the value of $\mathbf{A}^{(n,m)}$ for the last ADMM iteration.

The solution of (21) is

$$\mathbf{A}^{(n,m)} = \left[\mathbf{E}^{(n)\mathsf{T}}\mathbf{E}^{(n)} + (\mu+\lambda)\,\mathbf{I}_K\right]^{-1}\left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{X}} + \mu\left(\mathbf{V}_1^{(m-1)} + \mathbf{G}_1^{(m-1)}\right) + \lambda\left(\mathbf{V}_4^{(m-1)} + \mathbf{G}_4^{(m-1)}\right)\right]. \tag{23}$$

In order to improve the estimation performance, we replace $\mathbf{E}^{(n)}$ in (23) with the perturbed endmember matrix $\tilde{\mathbf{E}}$ and change (23) to

$$\mathbf{A}^{(n,m)} = \left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{E}} + (\mu+\lambda)\,\mathbf{I}_K\right]^{-1}\left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{X}} + \mu\left(\mathbf{V}_1^{(m-1)} + \mathbf{G}_1^{(m-1)}\right) + \lambda\left(\mathbf{V}_4^{(m-1)} + \mathbf{G}_4^{(m-1)}\right)\right]. \tag{24}$$

The method of instrumental variable (IV) [54]–[57] has inspired us to make this modification and take advantage of the available, albeit perturbed, knowledge of the endmember matrix, i.e., $\tilde{\mathbf{E}}$, in the estimation of the abundance matrix. We expect this to be particularly beneficial when the perturbation in the endmember matrix, i.e., $\mathbf{P}_{\mathbf{E}}$, is not too large to cause a substantial distortion. In Appendix E, we sketch the reason behind the expected benefit of using $\tilde{\mathbf{E}}$ as an IV in (24). Our simulation results presented in Section VI ahead also attest to the performance advantages of employing this IV technique.

Remark 4: We do not apply the IV method in the orthodox manner but the way we utilize it is similar to solving (21) while replacing $\mathbf{E}^{(n)}$ with $\tilde{\mathbf{E}}$ in (19), then using $\mathbf{E}^{(n)}$ as an IV and calculating $\mathbf{A}^{(n,m)}$ as an IV estimate rather than the ordinary least-squares solution. In that sense, $\mathbf{E}^{(n)}$ is a good IV since it is highly dependant on $\tilde{\mathbf{E}}$ (as $\tilde{\mathbf{E}}$ is the perturbed version of the true endmember matrix $\mathbf{E}$ and $\mathbf{E}^{(n)}$ is an estimate of $\mathbf{E}$) but is not independently correlated with $\tilde{\mathbf{X}}$, i.e., it is not correlated with $\mathbf{P}_{\mathbf{X}}$, the perturbation in $\tilde{\mathbf{X}}$ [54].

The optimization in (22) can be decoupled with respect to $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$, and $\mathbf{V}_4$, i.e., (22) can be written as

$$\left\{\mathbf{V}_1^{(m)}, \mathbf{V}_2^{(m)}, \mathbf{V}_3^{(m)}\right\} = \underset{\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3}{\operatorname{argmin}}\ \mathcal{L}\left(\mathbf{A}^{(n,m)}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4^{(m-1)}, \mathbf{G}_1^{(m-1)}, \cdots, \mathbf{G}_4^{(m-1)}\right) \tag{25}$$

$$\mathbf{V}_4^{(m)} = \underset{\mathbf{V}_4}{\operatorname{argmin}}\ \mathcal{L}\left(\mathbf{A}^{(n,m)}, \mathbf{V}_1^{(m)}, \mathbf{V}_2^{(m)}, \mathbf{V}_3^{(m)}, \mathbf{V}_4, \mathbf{G}_1^{(m-1)}, \cdots, \mathbf{G}_4^{(m-1)}\right). \tag{26}$$

The solution of (25) can be achieved through cyclic optimizations with respect to $\mathbf{V}_1$ and $\mathbf{V}_2, \mathbf{V}_3$ until the iterates converge. However, since an exact solution of (22) is not necessary at each iteration of the ADMM, we perform these optimizations only once at each ADMM iteration. Therefore, we replace (25) with

$$\mathbf{V}_1^{(m)} = \underset{\mathbf{V}_1}{\operatorname{argmin}}\ \mathcal{L}\left(\mathbf{A}^{(n,m)}, \mathbf{V}_1, \mathbf{V}_2^{(m-1)}, \mathbf{V}_3^{(m-1)}, \mathbf{V}_4^{(m-1)}, \mathbf{G}_1^{(m-1)}, \cdots, \mathbf{G}_4^{(m-1)}\right) \tag{27}$$

$$\left\{\mathbf{V}_2^{(m)}, \mathbf{V}_3^{(m)}\right\} = \underset{\mathbf{V}_2,\mathbf{V}_3}{\operatorname{argmin}}\ \mathcal{L}\left(\mathbf{A}^{(n,m)}, \mathbf{V}_1^{(m)}, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4^{(m-1)}, \mathbf{G}_1^{(m-1)}, \cdots, \mathbf{G}_4^{(m-1)}\right) \tag{28}$$

The solution of (27) is

$$\mathbf{V}_1^{(m)} = \left[\mathbf{A}^{(n,m)} - \mathbf{G}_1^{(m-1)} + \left(\mathbf{V}_2^{(m-1)} + \mathbf{G}_2^{(m-1)}\right)\mathbf{D}_h^{\mathsf{T}} + \left(\mathbf{V}_3^{(m-1)} + \mathbf{G}_3^{(m-1)}\right)\mathbf{D}_v^{\mathsf{T}}\right]\left(\mathbf{D}_h\mathbf{D}_h^{\mathsf{T}} + \mathbf{D}_v\mathbf{D}_v^{\mathsf{T}} + \mathbf{I}_N\right)^{-1}.$$

In addition, the problem (28) can be decomposed with respect to the columns of $\mathbf{V}_2^{(m)}$ and $\mathbf{V}_3^{(m)}$. Then, each subproblem can be solved by making use of the Moreau proximity operator of the $\ell_{2,1}$-norm calculated via column-wise vector-soft-thresholding [58], [59]. Therefore, by defining

$$\mathbf{Z}^{(m)} = \begin{bmatrix}\mathbf{V}_1^{(m)}\mathbf{D}_h - \mathbf{G}_2^{(m-1)} \\ \mathbf{V}_1^{(m)}\mathbf{D}_v - \mathbf{G}_3^{(m-1)}\end{bmatrix},$$

the jth columns of $\mathbf{V}_2^{(m)}$ and $\mathbf{V}_3^{(m)}$, denoted by $\mathbf{v}_{2,j}^{(m)}$ and $\mathbf{v}_{3,j}^{(m)}$, respectively, are computed in terms of the jth column of $\mathbf{Z}^{(m)}$, denoted by $\mathbf{z}_j^{(m)}$, as

$$\begin{bmatrix}\mathbf{v}_{2,j}^{(m)} \\ \mathbf{v}_{3,j}^{(m)}\end{bmatrix} = \frac{\max\left\{\left\|\mathbf{z}_j^{(m)}\right\|_2 - \frac{\beta}{\mu}, 0\right\}}{\left\|\mathbf{z}_j^{(m)}\right\|_2}\mathbf{z}_j^{(m)}.$$

Finally, the solution of (26) is the projection of the term $\mathbf{A}^{(n,m)} - \mathbf{G}_4^{(m-1)}$ onto the set $\mathcal{S}$, i.e.,

$$\mathbf{V}_4^{(m)} = \Pi_{\mathcal{S}}\left\{\mathbf{A}^{(n,m)} - \mathbf{G}_4^{(m-1)}\right\}.$$

If $\mathcal{S}$ is the set of all $K \times N$ matrices that have entries in the range $[0, 1]$ (associated with the abundance nonnegativity constraint), the projection onto it can be realized by setting the negative entries to zero and the entries greater than one to one. On the other hand, if $\mathcal{S}$ is the standard $(K-1)$-simplex (associated with the abundance nonnegativity and sum-to-one constraints), projection onto it can be realized using, e.g., the algorithm of [60].

Algorithm 1: The proposed RCTLS-IV algorithm.

initialize:

$\mathbf{A}^{(0)} = \mathbf{0}_{K\times N}$

for $n = 1, 2, \ldots,$ *until a stopping criterion is met* **do**

$\mathbf{E}^{(n)} = \left(\tilde{\mathbf{X}}\mathbf{A}^{(n-1)\mathsf{T}} + \alpha\tilde{\mathbf{E}}\right)\left(\mathbf{A}^{(n-1)}\mathbf{A}^{(n-1)\mathsf{T}} + \alpha\mathbf{I}_K + \gamma\mathbf{W}\right)^{-1}$

initialize:

$\mathbf{V}_1^{(0)} = \mathbf{V}_2^{(0)} = \mathbf{V}_3^{(0)} = \mathbf{V}_4^{(0)} = \mathbf{0}_{K\times N}$

$\mathbf{G}_1^{(0)} = \mathbf{G}_2^{(0)} = \mathbf{G}_3^{(0)} = \mathbf{G}_4^{(0)} = \mathbf{0}_{K\times N}$

for $m = 1, 2, \ldots,$ *until a stopping criterion is met* **do**

$\mathbf{A}^{(n,m)} = \left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{E}} + (\mu+\lambda)\,\mathbf{I}_K\right]^{-1}\left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{X}} + \mu\left(\mathbf{V}_1^{(m-1)} + \mathbf{G}_1^{(m-1)}\right) + \lambda\left(\mathbf{V}_4^{(m-1)} + \mathbf{G}_4^{(m-1)}\right)\right]$

$\mathbf{V}_1^{(m)} = \left[\mathbf{A}^{(n,m)} - \mathbf{G}_1^{(m-1)} + \left(\mathbf{V}_2^{(m-1)} + \mathbf{G}_2^{(m-1)}\right)\mathbf{D}_h^{\mathsf{T}} + \left(\mathbf{V}_3^{(m-1)} + \mathbf{G}_3^{(m-1)}\right)\mathbf{D}_v^{\mathsf{T}}\right]\left(\mathbf{D}_h\mathbf{D}_h^{\mathsf{T}} + \mathbf{D}_v\mathbf{D}_v^{\mathsf{T}} + \mathbf{I}_N\right)^{-1}$

$\mathbf{Z}^{(m)} = \begin{bmatrix}\mathbf{V}_1^{(m)}\mathbf{D}_h - \mathbf{G}_2^{(m-1)} \\ \mathbf{V}_1^{(m)}\mathbf{D}_v - \mathbf{G}_3^{(m-1)}\end{bmatrix}$

for $j = 1, \ldots, N$ **do**

$\begin{bmatrix}\mathbf{v}_{2,j}^{(m)} \\ \mathbf{v}_{3,j}^{(m)}\end{bmatrix} = \dfrac{\max\left\{\left\|\mathbf{z}_j^{(m)}\right\|_2 - \dfrac{\beta}{\mu}, 0\right\}}{\left\|\mathbf{z}_j^{(m)}\right\|_2}\mathbf{z}_j^{(m)}$

$\mathbf{V}_4^{(m)} = \Pi_{\mathcal{S}}\left\{\mathbf{A}^{(n,m)} - \mathbf{G}_4^{(m-1)}\right\}$

$\mathbf{G}_1^{(m)} = \mathbf{G}_1^{(m-1)} - \left(\mathbf{A}^{(n,m)} - \mathbf{V}_1^{(m)}\right)$

$\mathbf{G}_2^{(m)} = \mathbf{G}_2^{(m-1)} - \left(\mathbf{V}_1^{(m)}\mathbf{D}_h - \mathbf{V}_2^{(m)}\right)$

$\mathbf{G}_3^{(m)} = \mathbf{G}_3^{(m-1)} - \left(\mathbf{V}_1^{(m)}\mathbf{D}_v - \mathbf{V}_3^{(m)}\right)$

$\mathbf{G}_4^{(m)} = \mathbf{G}_4^{(m-1)} - \left(\mathbf{A}^{(n,m)} - \mathbf{V}_4^{(m)}\right)$

$\mathbf{A}^{(n)} \leftarrow \mathbf{A}^{(n,m)}$

We summarize the proposed algorithm, which we call regularized constrained total least-squares with instrumental variable (RCTLS-IV), in Algorithm 1. We use this name in view of the similarity of the proposed algorithm to the total least-squares methods [61]–[63] in performing estimation based on a linear errors-in-variables model. Algorithm 2 is a version of the proposed algorithm that uses only one iteration of the ADMM algorithm with warm start at each BCD iteration. In this algorithm, as a result of running a single inner ADMM iteration, the inner iteration index (m) is merged with the outer BCD iteration index (n).

C. Computational Complexity

It is clear that the most expensive steps are the calculations of $\mathbf{E}^{(n)}$, $\mathbf{A}^{(n,m)}$, and $\mathbf{V}_1^{(m)}$. The calculation of $\mathbf{E}^{(n)}$ requires two matrix-matrix multiplications with complexities of $\mathcal{O}(KLN)$ and $\mathcal{O}\left(K^2N\right)$ as well as the solution of L systems of linear equation with a common $K \times K$ coefficient matrix with the complexity of $\mathcal{O}\left(K^2L\right)$. The calculation of $\mathbf{A}^{(n,m)}$ requires two matrix-matrix multiplications with complexities of $\mathcal{O}\left(K^2L\right)$ and $\mathcal{O}(KLN)$ as well as the solution of N systems of linear equation with a common $K \times K$ coefficient matrix with the complexity of $\mathcal{O}\left(K^2N\right)$. For the calculation of $\mathbf{V}_1^{(m)}$, the matrix multiplications involving the differential matrix operators $\mathbf{D}_h$ and $\mathbf{D}_v$ and their transposes can be implemented using only subtractions. The multiplication by $\left(\mathbf{D}_h\mathbf{D}_h^{\mathsf{T}} + \mathbf{D}_v\mathbf{D}_v^{\mathsf{T}} + \mathbf{I}_N\right)^{-1}$ can also be efficiently performed using the fast Fourier transform (FFT) algorithm and the circular convolution theorem [64] with the complexity of $\mathcal{O}(KN\log N)$, when the differential operators apply with periodic boundaries. Consequently, the proposed algorithm has a computation complexity of order $\mathcal{O}(KN\max\{L, \log N\}) + \mathcal{O}\left(K^2\max\{L, N\}\right)$ per iteration or simply $\mathcal{O}(KLN)$ per iteration in the likely case of having $N \geq L \geq \log N$.

Algorithm 2: The proposed RCTLS-IV algorithm with a single inner ADMM iteration.

initialize:

$\mathbf{A}^{(0)} = \mathbf{0}_{K\times N}$

$\mathbf{V}_1^{(0)} = \mathbf{V}_2^{(0)} = \mathbf{V}_3^{(0)} = \mathbf{V}_4^{(0)} = \mathbf{0}_{K\times N}$

$\mathbf{G}_1^{(0)} = \mathbf{G}_2^{(0)} = \mathbf{G}_3^{(0)} = \mathbf{G}_4^{(0)} = \mathbf{0}_{K\times N}$

for $n = 1, 2, \ldots,$ *until a stopping criterion is met* **do**

$\mathbf{E}^{(n)} = \left(\tilde{\mathbf{X}}\mathbf{A}^{(n-1)\mathsf{T}} + \alpha\tilde{\mathbf{E}}\right)\left(\mathbf{A}^{(n-1)}\mathbf{A}^{(n-1)\mathsf{T}} + \alpha\mathbf{I}_K + \gamma\mathbf{W}\right)^{-1}$

$\mathbf{A}^{(n)} = \left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{E}} + (\mu+\lambda)\,\mathbf{I}_K\right]^{-1}\left[\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{X}} + \mu\left(\mathbf{V}_1^{(n-1)} + \mathbf{G}_1^{(n-1)}\right) + \lambda\left(\mathbf{V}_4^{(n-1)} + \mathbf{G}_4^{(n-1)}\right)\right]$

$\mathbf{V}_1^{(n)} = \left[\mathbf{A}^{(n)} - \mathbf{G}_1^{(n-1)} + \left(\mathbf{V}_2^{(n-1)} + \mathbf{G}_2^{(n-1)}\right)\mathbf{D}_h^{\mathsf{T}} + \left(\mathbf{V}_3^{(n-1)} + \mathbf{G}_3^{(n-1)}\right)\mathbf{D}_v^{\mathsf{T}}\right]\left(\mathbf{D}_h\mathbf{D}_h^{\mathsf{T}} + \mathbf{D}_v\mathbf{D}_v^{\mathsf{T}} + \mathbf{I}_N\right)^{-1}$

$\mathbf{Z}^{(n)} = \begin{bmatrix}\mathbf{V}_1^{(n)}\mathbf{D}_h - \mathbf{G}_2^{(n-1)} \\ \mathbf{V}_1^{(n)}\mathbf{D}_v - \mathbf{G}_3^{(n-1)}\end{bmatrix}$

for $j = 1, \ldots, N$ **do**

$\begin{bmatrix}\mathbf{v}_{2,j}^{(n)} \\ \mathbf{v}_{3,j}^{(n)}\end{bmatrix} = \dfrac{\max\left\{\left\|\mathbf{z}_j^{(n)}\right\|_2 - \dfrac{\beta}{\mu}, 0\right\}}{\left\|\mathbf{z}_j^{(n)}\right\|_2}\mathbf{z}_j^{(n)}$

$\mathbf{V}_4^{(n)} = \Pi_{\mathcal{S}}\left\{\mathbf{A}^{(n)} - \mathbf{G}_4^{(n-1)}\right\}$

$\mathbf{G}_1^{(n)} = \mathbf{G}_1^{(n-1)} - \left(\mathbf{A}^{(n)} - \mathbf{V}_1^{(n)}\right)$

$\mathbf{G}_2^{(n)} = \mathbf{G}_2^{(n-1)} - \left(\mathbf{V}_1^{(n)}\mathbf{D}_h - \mathbf{V}_2^{(n)}\right)$

$\mathbf{G}_3^{(n)} = \mathbf{G}_3^{(n-1)} - \left(\mathbf{V}_1^{(n)}\mathbf{D}_v - \mathbf{V}_3^{(n)}\right)$

$\mathbf{G}_4^{(n)} = \mathbf{G}_4^{(n-1)} - \left(\mathbf{A}^{(n)} - \mathbf{V}_4^{(n)}\right)$

V. Convergence

The objective function of the optimization problem (16) is nonconvex and may have multiple local minima at various coordinate points of its augmented matrix argument $[\mathbf{E}, \mathbf{A}]$. Therefore, the proposed algorithm, which is based on the BCD and ADMM algorithms, is not guaranteed to find the globally optimal solution of (16). Nonetheless, as we discuss below, the proposed algorithm converges at least to a local minimum of the objective function. We have observed experimentally that this local minimum is often very close to the global optimum when the perturbations are not too large.

First, we study the convergence properties of the BCD algorithm of (17) and (18) applied to solve (16). Incorporating the constraint $\mathbf{A} \in \mathcal{S}$ into the objective function in (16) gives the following equivalent unconstrained objective function

$$f(\mathbf{E}, \mathbf{A}) = \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{E}\mathbf{A}\right\|_{\mathsf{F}}^2 + \frac{\alpha}{2}\left\|\tilde{\mathbf{E}} - \mathbf{E}\right\|_{\mathsf{F}}^2 + \beta\left\|\nabla\mathbf{A}\right\|_{2,1} + \frac{\gamma}{2}\left\|\mathbf{E}\mathbf{W}\right\|_{\mathsf{F}}^2 + \imath_{\mathcal{S}}(\mathbf{A})$$

that can be decomposed as

$$f(\mathbf{E},\mathbf{A}) = f_0(\mathbf{E},\mathbf{A}) + f_1(\mathbf{A})$$

where

$$f_0(\mathbf{E},\mathbf{A}) = \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{E}\mathbf{A}\right\|_{\mathsf{F}}^2 + \frac{\alpha}{2}\left\|\tilde{\mathbf{E}} - \mathbf{E}\right\|_{\mathsf{F}}^2 + \frac{\gamma}{2}\|\mathbf{E}\mathbf{W}\|_{\mathsf{F}}^2$$
$$f_1(\mathbf{A}) = \beta\|\nabla\mathbf{A}\|_{2,1} + \imath_{\mathcal{S}}(\mathbf{A}).$$

The function f is continuous on its effective domain and has compact level sets. It is also biconvex in $(\mathbf{E},\mathbf{A})$, i.e., it is convex over $\mathbf{E}$ if $\mathbf{A}$ is fixed and vice versa. In addition, f_0 is Gâteaux-differentiable on its domain, which is open, and f_1 is convex. Therefore, according to Lemma 3.1 and Theorem 4.1(b) of [33], the iterates generated by the BCD algorithm of (17) and (18) are defined and every cluster point is a stationary point of f. In other words, the algorithm converges to a local minimum of the objective function f.

Next, we examine the convergence properties of the ADMM algorithm that we use to solve (18) or equivalently (19). To this end, we rewrite (19) as

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}} \quad & g_1(\mathbf{U}) + g_2(\mathbf{V}) \\ \text{s.t.} \quad & \mathbf{B}\mathbf{U} + \mathbf{C}\mathbf{V} = \mathbf{0}_{4N\times K}\end{aligned}$$

where

$$g_1(\mathbf{U}) = \frac{1}{2}\left\|\tilde{\mathbf{X}} - \mathbf{E}^{(n)}\mathbf{U}^{\mathsf{T}}\right\|_{\mathsf{F}}^2,$$
$$g_2(\mathbf{V}) = \beta\left\|\begin{bmatrix}\mathbf{V}_2\\ \mathbf{V}_3\end{bmatrix}\right\|_{2,1} + \imath_{\mathcal{S}}(\mathbf{V}_4),$$
$$\mathbf{U} = \mathbf{A}^{\mathsf{T}}, \quad \mathbf{V} = [\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{V}_4]^{\mathsf{T}},$$
$$\mathbf{B} = \begin{bmatrix}\mathbf{I}_N\\ \mathbf{0}_N\\ \mathbf{0}_N\\ \mathbf{I}_N\end{bmatrix}, \text{ and } \mathbf{C} = \begin{bmatrix}-\mathbf{I}_N & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_N\\ \mathbf{D}_h^{\mathsf{T}} & -\mathbf{I}_N & \mathbf{0}_N & \mathbf{0}_N\\ \mathbf{D}_v^{\mathsf{T}} & \mathbf{0}_N & -\mathbf{I}_N & \mathbf{0}_N\\ \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_N & \mathbf{I}_N\end{bmatrix}.$$

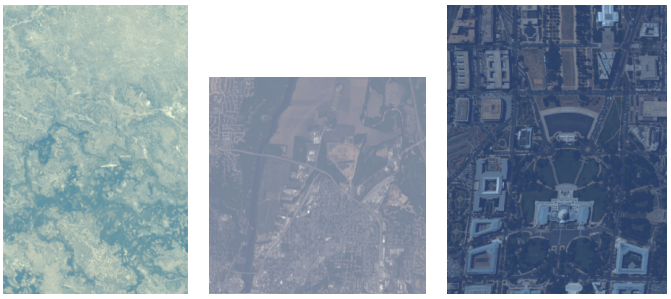
The functions g_1 and g_2 are closed, proper, and convex and the matrix $\mathbf{B}$ has full column rank. Therefore, according to [36, Theorem 1], the ADMM algorithm described in Section IV will converge to a solution of (19), if there exists any, with any pair of positive penalty parameters μ and λ regardless of the initial values of the iterates.

VI. Evaluation

To evaluate the performance of the proposed algorithm in comparison with its closest contenders, we consider the problem of linear unmixing of a hyperspectral image given its noisy observations and a perturbed version of its endmembers as explained below.

In our simulations, we use parts of three publicly available hyperspectral image datasets, namely, Botswana, Indian Pines [65], and Washington DC[2] after removing the uncalibrated, high-noise, and water-absorption spectral bands. The RGB illustrations of the considered parts are shown in Fig. 1 and their spatial and spectral resolutions are given in Table I.

[2]https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html



(a) Botswana (b) Indian Pines (c) Washington DC

Fig. 1. An RGB illustration of the hyperspectral images used in the simulations.

TABLE I
The Spatial and Spectral Resolutions of the Hyperspectral Image Datasets Used in the Simulations

	number of rows	number of columns	number of bands
Botswana	400	255	145
Indian Pines	300	300	200
Washington DC	400	305	191

The Botswana dataset has been collected by the Hyperion sensor aboard NASA's Earth Observing 1 (EO-1) satellite[3,4], the Indian Pines dataset by NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument[5], and the Washington DC dataset through the Hyperspectral Digital Imagery Collection Experiment (HYDICE) [66]. All three datasets cover the visible near-infrared (VNIR) and short-wavelength infrared (SWIR) ranges.

To create the ground truth for our experiments, We extract $K = 10$ endmembers from each dataset using the vertex component analysis (VCA) algorithm [10] and use the extracted endmembers to unmix the dataset using the SUnSAL algorithm [21]. We take the output of the VCA and SUnSAL algorithms as the ground-truth endmember and abundance matrices, respectively, and multiply them to generate the ground-truth hyperspectral images for each dataset. We then generate the perturbed hyperspectral images and the perturbed endmember matrices by contaminating them with additive zero-mean white Gaussian noise to achieve a target signal-to-noise ratio (SNR) for both.

We compare the performance of the proposed RCTLS-IV algorithm with those of four other algorithms, which we call constrained least-squares (CLS), constrained least-squares with total-variation regularization (CLS-TV), constrained total least-squares (CTLS), and constrained total least-squares with instrumental variable (CTLS-IV). The CLS algorithm solves the following optimization problem

$$\begin{aligned}\min_{\mathbf{A}} \quad & \frac{1}{2}\left\|\tilde{\mathbf{X}} - \tilde{\mathbf{E}}\mathbf{A}\right\|_{\mathsf{F}}^2, \\ \text{s.t.} \quad & \mathbf{A}\in\mathcal{S}\end{aligned}$$

[3]https://eo1.usgs.gov/sensors/hyperion

[4]http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[5]http://aviris.jpl.nasa.gov/data/free_data.html

the CLS-TV algorithm solves

$$\begin{array}{ll}\min\limits_{\mathbf{A}} & \frac{1}{2}\left\|\tilde{\mathbf{X}}-\tilde{\mathbf{E}}\mathbf{A}\right\|_{\mathsf{F}}^{2}+\beta\left\|\nabla\mathbf{A}\right\|_{2,1},\\ \text{s.t.} & \mathbf{A}\in\mathcal{S}\end{array}$$

the CTLS algorithm solves

$$\begin{array}{ll}\min\limits_{\mathbf{E},\mathbf{A}} & \frac{1}{2}\left\|\tilde{\mathbf{X}}-\mathbf{E}\mathbf{A}\right\|_{\mathsf{F}}^{2}+\frac{\alpha}{2}\left\|\tilde{\mathbf{E}}-\mathbf{E}\right\|_{\mathsf{F}}^{2},\\ \text{s.t.} & \mathbf{A}\in\mathcal{S}\end{array}\tag{29}$$

and the CTLS-IV algorithm solves the same problem as the CTLS algorithm does, i.e., (29), nonetheless using the instrumental-variable technique mentioned in Section IV. To implement these algorithms, we use the same methodology employed to develop the proposed RCTLS-IV algorithm, i.e., utilizing the BCD and ADMM algorithms, as well as identical parameter values.

Note that the CLS algorithm solves the same problem as in [28] while taking $\tilde{\mathbf{E}}$ as the endmember matrix and the CLS-TV algorithm [22], [24] extends it by incorporating a total-variation regularization. In addition, the objective function of the CTLS algorithm is very similar to that of the R-CoNMF algorithm proposed in [20]. The CTLS-IV algorithm can also be considered a version of the proposed RCTLS-IV algorithm that lacks the regularization penalties detailed in Section III-D.

We use two performance metrics, namely, the inverse normalized mean square error (INMSE), also known as the signal-to-reconstruction-error ratio (SRE) [24], defined as

$$10\log_{10}\frac{\left\|\mathbf{A}\right\|_{\mathsf{F}}^{2}}{\mathbb{E}\left[\left\|\hat{\mathbf{A}}-\mathbf{A}\right\|_{\mathsf{F}}^{2}\right]}$$

and the inverse normalized bias (INB) defined as

$$10\log_{10}\frac{\left\|\mathbf{A}\right\|_{\mathsf{F}}^{2}}{\left\|\mathbb{E}\left[\hat{\mathbf{A}}-\mathbf{A}\right]\right\|_{\mathsf{F}}^{2}}$$

where $\hat{\mathbf{A}}$ denotes the estimated abundance matrix. Higher values of these metrics indicate better performance. We approximate the expectations by averaging over 100 independent trials. At each trial, we run the algorithms for 200 iterations and initialize by $\mathbf{A}^{(0)}=\mathbf{0}_{K\times N}$. In the CTLS, CTLS-IV and RCTLS-IV algorithms, at each outer BCD iteration, we implement only one inner iteration of the ADMM algorithm with warm start as shown in Algorithm 2 for the RCTLS-IV algorithm. The data and MATLAB scripts used to produce the simulation results can be found at [67].

The proposed RCTLS-IV algorithm has five tunable parameters, the regularization parameters α, β, and γ and the ADMM penalty parameters μ and λ. The automatic tuning of the values of these parameters is an interesting and challenging subject for which there exists a few strategies such as those proposed in [58] and [71]. The values of μ and λ impact the convergence speed of the proposed algorithm and, as long as being within an appropriate range, have little influence on the accuracy of the proposed algorithm. Therefore, in our experiments with all considered datasets, we tune them so that the algorithms sufficiently converge within 200 iterations. On the other hand, the values of α, β, and γ affect the performance of the proposed algorithm in subtle ways. In Fig. 2, we plot the performance metrics of the proposed algorithm for unmixing the Botswana dataset with $\mathrm{SNR}=30$ dB versus a number of values for α, β, and γ. The results indicate that the performance of the proposed algorithm is not highly sensitive to the values of these regularization parameters. In addition, choosing the proper parameter values may involve a trade-off between the highest attainable INMSE and INB. Consequently, we tune the parameters α, β, and γ only coarsely to obtain approximately the best performance for each dataset. We give these values in Table II. Note that the difference in the order of magnitude of the parameter values is due to the scaling and dimension of the signals and perturbations and does not imply any meaningful difference in their significance.

TABLE II
THE PARAMETER VALUES USED IN THE SIMULATIONS

	α	β	γ	μ	λ
Botswana	100	6×10^{-6}	50	10^{-3}	10^{-5}
Indian Pines	250	1.2×10^{-5}	50	10^{-3}	10^{-5}
Washington DC	150	5×10^{-5}	50	10^{-2}	5×10^{-5}

Fig. 3 depicts how the INMSE of different algorithms evolve over iterations. Fig. 4 shows the values of the INMSE for each abundance band corresponding to an endmember, i.e., the INMSE for each row of $\hat{\mathbf{A}}$. Fig. 5 shows the values of the INMSE for each pixel, i.e., the INMSE of each column of $\hat{\mathbf{A}}$, sorted in an ascending order. Figs. 3–5 contain the results for all considered algorithms and datasets when the SNR is approximately 30 dB. Figs. 6–8 display the abundance values for three endmembers of each considered dataset estimated through a single run of the CLS, CTLS, and RCTLS-IV algorithms together with the corresponding ground-truth values in the form of two-dimensional abundance maps (images). We have not included the results for the CLS-TV and CTLS-IV algorithms in Figs. 6–8 since their estimated abundance maps are visually very similar to those of the CLS and RCTLS-IV algorithms, respectively. The SNR in the experiments resulting in Figs. 6–8 was approximately 30 dB. In Table III, we provide the values of the performance metrics for all considered algorithms and datasets as well as the corresponding CRLBs for three SNRs of 25, 30, and 35 dB. We also give the average algorithm run times (using MATLAB, a 2.9GHz Core-i7 CPU, and 24GB of DDR3 RAM) in Table IV.

We observe in Figs. 3–5 that, as expected, accounting for the perturbations in the endmembers (in all algorithms but the CLS and CLS-TV algorithms) significantly improves the unmixing performance evident by the reduction in both MSE and bias (increase in INMSE and INB). The IV technique employed in the CTLS-IV and RCTLS-IV algorithms and the additional regularizations used in the RCTLS-IV algorithm further enhance the performance in terms of both estimation accuracy and convergence speed, although with varying degrees depending on the considered dataset. Understandably, these benefits come at the cost of computational complexity

TABLE III
THE VALUES OF THE PERFORMANCE METRICS FOR DIFFERENT ALGORITHMS, DATASETS, AND SNRS TOGETHER WITH THE CORRESPONDING CRLBS

SNR = 25 dB						
	Botswana		Indian Pines		Washington DC	
	INMSE (dB)	INB (dB)	INMSE (dB)	INB (dB)	INMSE (dB)	INB (dB)
CLS	6.360	6.983	8.266	8.786	5.606	5.697
CLS-TV	6.409	6.974	8.290	8.781	5.607	5.696
CTLS	7.897	8.718	10.861	11.631	11.023	11.467
CTLS-IV	9.212	12.601	12.737	17.457	12.881	17.884
RCTLS-IV	9.927	13.572	13.394	17.774	13.246	17.726
CRLB	5.497		10.866		15.535	
SNR = 30 dB						
	Botswana		Indian Pines		Washington DC	
	INMSE (dB)	INB (dB)	INMSE (dB)	INB (dB)	INMSE (dB)	INB (dB)
CLS	9.037	9.807	13.101	14.044	7.545	7.636
CLS-TV	9.154	9.752	13.180	14.021	7.547	7.635
CTLS	10.274	11.243	14.091	15.003	14.447	14.987
CTLS-IV	14.302	18.849	17.931	21.736	18.489	24.759
RCTLS-IV	15.322	19.431	18.277	22.026	18.542	24.559
CRLB	10.497		15.866		20.535	
SNR = 35 dB						
	Botswana		Indian Pines		Washington DC	
	INMSE (dB)	INB (dB)	INMSE (dB)	INB (dB)	INMSE (dB)	INB (dB)
CLS	13.352	14.481	18.649	20.391	11.763	11.923
CLS-TV	13.426	14.145	18.804	20.255	11.744	11.892
CTLS	14.567	15.422	19.078	20.136	17.562	18.044
CTLS-IV	19.185	24.645	22.630	27.418	23.362	30.667
RCTLS-IV	20.076	23.616	23.314	27.804	23.457	30.032
CRLB	15.497		20.866		25.535	

TABLE IV
THE AVERAGE RUN TIMES OF THE CONSIDERED ALGORITHMS WITH EACH CONSIDERED DATASET IN SECONDS

	Botswana	Indian Pines	Washington DC
CLS	4.749	4.208	5.983
CLS-TV	30.285	27.414	39.625
CTLS	11.741	11.908	16.568
CTLS-IV	11.917	12.084	16.639
RCTLS-IV	38.318	35.332	51.111

and algorithm run time. Figs. 6–8 also testify to the merits of the proposed RCTLS-IV algorithm by revealing that its estimated abundance maps resemble the ground truth visibly better than those of its contenders.

The CRLB values presented in Table III provide a useful guide for the estimation performance that can be expected in each scenario. However, in some cases, they appear to be smaller than the experimentally-obtained INMSE values for the CTLS-IV or RCTLS-IV algorithms. We blame this on two facts. First, although the CTLS-IV and RCTLS-IV algorithms yield substantial reductions in the estimation bias compared to the CLS, CLS-TV, and CTLS algorithms, they are not strictly unbiased. This is a repercussion of the abundance nonnegativity constraint couple with the assumption of unbounded Gaussian distribution for the perturbations. Second, in the derivation of the CRLB, we dismissed the possible effects of the abundance nonnegativity constraint on the grounds that the CRLB is only a local bound and is not impacted by any strict inequality constraint.

It is argued in [48] that only equality (active) constraints on the unknown parameters affect the CRLB while inequality (inactive) constraints do not, i.e., inequality constraints do not contribute any information to the model nor impact the performance potential. This is justified by observing that as the test points approach the true parameters in the Chapman-Robbins bound, they become interior points of the constraint set. However, this only happens when the model perturbations are sufficiently small so that the estimates fluctuate around the true parameters within the feasible set, i.e., the nonnegative orthant in our case. Calculating a more accurate CRLB-like performance bound that takes into account the effects of inequality constraints, especially the nonnegativity constraint, without assuming that the estimates asymptotically fall into the feasible region is an interesting topic for future research.

VII. CONCLUSION

We studied a fully-perturbed linear mixture model where there exist perturbations on both the spectral image measurements and the available knowledge of the endmembers. This model is more realistic compared with the conventional linear mixture model that assumes perturbation only on the spectral measurements. The main reason is that, in practice, the endmembers are often extracted from imperfect and noisy multiband spectral images hence are subject to noise, error, or mismatch.

In order to gain insights into the theoretical limits of the expected performance of estimating the abundance matrix under a fully-perturbed linear mixture model, we calculated the relevant Fisher information matrix and the Cramer-Rao lower bound (CRLB) while taking into account the abundance sum-to-one constraint as well. This was done under the assumption that the perturbations on the spectral measurements and the endmembers are statistically independent and have zero-mean Gaussian distributions.

We posed the problem of inferring the abundance matrix as a constrained maximum-likelihood estimation problem and regularized the pertinent objective function by employing two penalty terms that promote smoothness in the spatial domain and geometrical containment in the spectral domain. We solved the formulated problem using a two-block coordinate-descent scheme along with an instance of the alternating direction method of multipliers. We also applied a modification to the developed algorithm by drawing inspirations from the notion of instrumental-variable estimation. Our simulation results verified the usefulness of the proposed algorithm as well as the validity of the theoretically-obtained CRLB.

For future research, it will be interesting to characterize the estimation bias of the proposed algorithm theoretically and devise a compensation mechanism to mitigate the bias. Considering different distributions for the perturbations can also be of interest, particularly to account for impulsive noise or missing data. Moreover, it will be beneficial to calculate a more accurate CRLB that effectively allows for possible drifts in the estimation performance induced by the abundance nonnegativity constraint.

Appendix A
Calculation of the FIM

Here, we derive the Jacobian matrices required for the calculation of the FIM.

The upper-left block is calculated as

$$\begin{aligned}\mathbf{F}_{11} &= -\mathbb{E}\left[\frac{\partial \mathbf{g}_{\mathbf{e}}}{\partial \mathbf{e}^{\mathsf{T}}}\right] \\ &= \frac{\partial}{\partial \mathbf{e}^{\mathsf{T}}}\operatorname{vec}\left\{\boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\} + \frac{\partial}{\partial \mathbf{e}^{\mathsf{T}}}\operatorname{vec}\left\{\alpha\boldsymbol{\Xi}^{-1}\mathbf{E}\right\} \\ &= \mathbf{A}\mathbf{A}^{\mathsf{T}} \otimes \boldsymbol{\Sigma}^{-1} + \mathbf{I}_K \otimes \alpha\boldsymbol{\Xi}^{-1}\end{aligned}$$

where $\otimes$ denotes the Kronecker product and we use the property [68]

$$\operatorname{vec}\{\mathbf{U}\mathbf{V}\mathbf{W}\} = \left(\mathbf{W}^{\mathsf{T}} \otimes \mathbf{U}\right)\operatorname{vec}\{\mathbf{V}\}. \tag{30}$$

For the upper-right block, we have

$$\begin{aligned}\mathbf{F}_{12} &= -\mathbb{E}\left[\frac{\partial \mathbf{g}_{\mathbf{e}}}{\partial \mathbf{a}^{\mathsf{T}}}\right] \\ &= -\mathbb{E}\left[\frac{\partial}{\partial \mathbf{a}^{\mathsf{T}}}\operatorname{vec}\left\{\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{X}}\mathbf{A}^{\mathsf{T}}\right\}\right] + \frac{\partial}{\partial \mathbf{a}^{\mathsf{T}}}\operatorname{vec}\left\{\boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\} \\ &= -\mathbb{E}\left[\mathbf{I}_K \otimes \boldsymbol{\Sigma}^{-1}\tilde{\mathbf{X}}\right] + \left(\mathbf{A} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{E}\right)\mathbf{P}_{K,N} \\ &\quad + \mathbf{I}_K \otimes \boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{A} \\ &= \left(\mathbf{A} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{E}\right)\mathbf{P}_{K,N} = \mathbf{P}_{L,K}\left(\boldsymbol{\Sigma}^{-1}\mathbf{E} \otimes \mathbf{A}\right)\end{aligned} \tag{31}$$

where in addition to (30) we use the property [68]

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}^{\mathsf{T}}}\operatorname{vec}\{\mathbf{U}\mathbf{V}\} &= \left(\mathbf{V}^{\mathsf{T}} \otimes \mathbf{I}_P\right)\frac{\partial}{\partial \mathbf{w}^{\mathsf{T}}}\operatorname{vec}\{\mathbf{U}\} \\ &\quad + \left(\mathbf{I}_R \otimes \mathbf{U}\right)\frac{\partial}{\partial \mathbf{w}^{\mathsf{T}}}\operatorname{vec}\{\mathbf{V}\}\end{aligned}$$

for matrix functions $\mathbf{U} : S \to \mathbb{R}^{P\times Q}$ and $\mathbf{V} : S \to \mathbb{R}^{Q\times R}$ defined and differentiable on an open set S and the argument $\mathbf{W} = \operatorname{vec}^{-1}\{\mathbf{w}\} \in S$. Moreover, $\mathbf{P}_{K,N}$ is the so-called vec-permutation matrix of order K, N that is the result of $\partial \operatorname{vec}\{\mathbf{A}\}/\partial \mathbf{a}^{\mathsf{T}}$ and satisfies $\mathbf{P}_{K,N}\operatorname{vec}\{\mathbf{A}^{\mathsf{T}}\} = \operatorname{vec}\{\mathbf{A}\}$. The vec-permutation matrix of order Q, R, $\mathbf{P}_{Q,R}$, is a $QR \times QR$ matrix partitioned into an $R \times Q$ array of $Q \times R$ submatrices where the (i, j)th submatrix has one at its (j, i)th entry and zeros elsewhere. The last line of (31) is due to the property of the vec-permutation matrices in reversing the order of the Kronecker products and the fact that $\mathbf{P}_{Q,R}^{\mathsf{T}} = \mathbf{P}_{Q,R}^{-1} = \mathbf{P}_{R,Q}$ [69].

It is easy to show that the lower-left block is the transpose of the upper-right block, i.e.,

$$\begin{aligned}\mathbf{F}_{21} &= \mathbf{F}_{12}^{\mathsf{T}} \\ &= \mathbf{P}_{N,K}\left(\mathbf{A}^{\mathsf{T}} \otimes \mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\right) = \left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{A}^{\mathsf{T}}\right)\mathbf{P}_{K,L}.\end{aligned}$$

Finally, the lower-right block is computed as

$$\begin{aligned}\mathbf{F}_{22} &= -\mathbb{E}\left[\frac{\partial \mathbf{g}_{\mathbf{a}}}{\partial \mathbf{a}^{\mathsf{T}}}\right] \\ &= \frac{\partial}{\partial \mathbf{a}^{\mathsf{T}}}\operatorname{vec}\left\{\mathbf{A}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right\} \\ &= \mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E} \otimes \mathbf{I}_N.\end{aligned}$$

Appendix B
Verifying the non-singularity (invertibility) of the FIM

Since α is positive, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Xi}$ are positive-definite, and the endmembers are linearly independent, $\mathbf{F}_{11}$ and $\mathbf{F}_{22}$ are full-rank. The Schur complement of $\mathbf{F}_{22}$ is also full-rank as it can be calculated as

$$\begin{aligned}&\mathbf{F}_{11} - \mathbf{F}_{12}\mathbf{F}_{22}^{-1}\mathbf{F}_{21} = \mathbf{A}\mathbf{A}^{\mathsf{T}} \otimes \boldsymbol{\Sigma}^{-1} + \mathbf{I}_K \otimes \alpha\boldsymbol{\Xi}^{-1} \\ &\quad - \mathbf{P}_{L,K}\left[\boldsymbol{\Sigma}^{-1}\mathbf{E}\left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1}\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{A}\mathbf{A}^{\mathsf{T}}\right]\mathbf{P}_{K,L} \\ &\overset{(a)}{=} \mathbf{A}\mathbf{A}^{\mathsf{T}} \otimes \boldsymbol{\Sigma}^{-1} + \mathbf{I}_K \otimes \alpha\boldsymbol{\Xi}^{-1} \\ &\quad - \mathbf{A}\mathbf{A}^{\mathsf{T}} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{E}\left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1}\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1} \\ &= \mathbf{I}_K \otimes \alpha\boldsymbol{\Xi}^{-1} \\ &\quad + \mathbf{A}\mathbf{A}^{\mathsf{T}} \otimes \boldsymbol{\Sigma}^{-1}\left[\mathbf{I}_L - \mathbf{E}\left(\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1}\mathbf{E}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\right]\end{aligned}$$

where (a) is due to the property of the vec-permutation matrices in reversing the order of Kronecker products. Consequently, $\mathbf{F}$ is non-singular and invertible [46].

APPENDIX C
DERIVATION OF (10)

When $\boldsymbol{\Xi} = \eta\boldsymbol{\Sigma}$, we have

$$\mathbf{A}\mathbf{A}^\mathsf{T} \otimes \boldsymbol{\Sigma}^{-1} + \mathbf{I}_K \otimes \alpha\boldsymbol{\Xi}^{-1} = \left(\mathbf{A}\mathbf{A}^\mathsf{T} + \frac{\alpha}{\eta}\mathbf{I}_K\right) \otimes \boldsymbol{\Sigma}^{-1}.$$

Therefore, we can write

$$\begin{aligned}
&\mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12} = \mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E} \otimes \mathbf{I}_N \\
&\quad - \mathbf{P}_{N,K}\left[\mathbf{A}^\mathsf{T}\left(\mathbf{A}\mathbf{A}^\mathsf{T} + \frac{\alpha}{\eta}\mathbf{I}_K\right)^{-1}\mathbf{A} \otimes \mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right]\mathbf{P}_{K,N} \\
&\overset{(b)}{=} \mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E} \otimes \mathbf{I}_N - \mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E} \otimes \mathbf{A}^\mathsf{T}\left(\mathbf{A}\mathbf{A}^\mathsf{T} + \frac{\alpha}{\eta}\mathbf{I}_K\right)^{-1}\mathbf{A} \\
&= \mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E} \otimes \left[\mathbf{I}_N - \mathbf{A}^\mathsf{T}\left(\mathbf{A}\mathbf{A}^\mathsf{T} + \frac{\alpha}{\eta}\mathbf{I}_K\right)^{-1}\mathbf{A}\right]
\end{aligned} \tag{32}$$

and consequently

$$\begin{aligned}
&\left(\mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)^{-1} \\
&= \left(\mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1} \otimes \left[\mathbf{I}_N - \mathbf{A}^\mathsf{T}\left(\mathbf{A}\mathbf{A}^\mathsf{T} + \frac{\alpha}{\eta}\mathbf{I}_K\right)^{-1}\mathbf{A}\right]^{-1} \\
&\overset{(c)}{=} \left(\mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1} \otimes \left(\mathbf{I}_N + \frac{\eta}{\alpha}\mathbf{A}^\mathsf{T}\mathbf{A}\right).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\operatorname{tr}\left\{\left(\mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)^{-1}\right\} \\
&\qquad = \operatorname{tr}\left\{\left(\mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\right)^{-1}\right\}\left(N + \frac{\eta}{\alpha}\left\|\mathbf{A}\right\|_\mathsf{F}^2\right).
\end{aligned}$$

In the above equations, (b) is due to the property of the vec-permutation matrices in reversing the order of Kronecker products and (c) is thanks to the Woodbury matrix identity, i.e.,

$$\left(\mathbf{B} + \mathbf{U}\mathbf{C}\mathbf{V}\right)^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{V}\mathbf{B}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{B}^{-1}.$$

APPENDIX D
DERIVATION OF (14)

Considering (32) and $\mathbf{H} = \mathbf{Y} \otimes \mathbf{I}_N$, we have

$$\begin{aligned}
&\mathbf{H}^\mathsf{T}\left(\mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\mathbf{H} \\
&= \mathbf{Y}^\mathsf{T}\mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{Y} \otimes \left[\mathbf{I}_N - \mathbf{A}^\mathsf{T}\left(\mathbf{A}\mathbf{A}^\mathsf{T} + \frac{\alpha}{\eta}\mathbf{I}_K\right)^{-1}\mathbf{A}\right]
\end{aligned}$$

and consequently

$$\begin{aligned}
&\mathbf{H}\left[\mathbf{H}^\mathsf{T}\left(\mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\mathbf{H}\right]^{-1}\mathbf{H}^\mathsf{T} \\
&\qquad = \mathbf{Y}\left(\mathbf{Y}^\mathsf{T}\mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{Y}\right)^{-1}\mathbf{Y}^\mathsf{T} \otimes \left(\mathbf{I}_N + \frac{\eta}{\alpha}\mathbf{A}^\mathsf{T}\mathbf{A}\right)
\end{aligned}$$

then

$$\begin{aligned}
&\operatorname{tr}\left\{\mathbf{H}\left[\mathbf{H}^\mathsf{T}\left(\mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\mathbf{H}\right]^{-1}\mathbf{H}^\mathsf{T}\right\} \\
&\quad = \operatorname{tr}\left\{\mathbf{Y}\left(\mathbf{Y}^\mathsf{T}\mathbf{E}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{E}\mathbf{Y}\right)^{-1}\mathbf{Y}^\mathsf{T}\right\}\left(N + \frac{\eta}{\alpha}\left\|\mathbf{A}\right\|_\mathsf{F}^2\right).
\end{aligned}$$

APPENDIX E
BENEFIT OF THE INSTRUMENTAL VARIABLE TRICK

If we assume that, after a sufficient number of iterations, $\mathbf{E}^{(n)}$ is uncorrelated with the perturbations $\mathbf{P}_\mathbf{E}$ and $\mathbf{P}_\mathbf{X}$ in addition to ignoring the terms multiplied by μ and λ, (23) and (24) can be approximated as

$$\mathbf{A}^{(n,m)} \approx \left(\mathbf{E}^{(n)\mathsf{T}}\mathbf{E}^{(n)}\right)^{-1}\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{X}}$$

and

$$\mathbf{A}_{\mathsf{IV}}^{(n,m)} \approx \left(\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{E}}\right)^{-1}\mathbf{E}^{(n)\mathsf{T}}\tilde{\mathbf{X}},$$

respectively, where we use the subscript IV to differentiate the two estimates. Applying the expectation operator together with using Slutsky's theorem [70] gives

$$\begin{aligned}
\mathbb{E}\left[\mathbf{A}^{(n,m)}\right] &\approx \left(\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\mathbf{E}^{(n)}\right]\right)^{-1}\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\right]\mathbb{E}\left[\tilde{\mathbf{X}}\right] \\
&\approx \left(\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\mathbf{E}^{(n)}\right]\right)^{-1}\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\right]\mathbf{E}\mathbf{A}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\mathbf{A}_{\mathsf{IV}}^{(n,m)}\right] &\approx \left(\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\right]\mathbb{E}\left[\tilde{\mathbf{E}}\right]\right)^{-1}\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\right]\mathbb{E}\left[\tilde{\mathbf{X}}\right] \\
&\approx \left(\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\right]\mathbf{E}\right)^{-1}\mathbb{E}\left[\mathbf{E}^{(n)\mathsf{T}}\right]\mathbf{E}\mathbf{A} \\
&\approx \mathbf{A}.
\end{aligned}$$

This shows that the employed IV scheme specifically drives $\mathbf{A}_{\mathsf{IV}}^{(n,m)}$ towards the true $\mathbf{A}$ in the mean sense regardless of the proximity of $\mathbf{E}^{(n)}$ to the true $\mathbf{E}$. Therefore, it can accelerate the convergence of $\mathbf{A}_{\mathsf{IV}}^{(n,m)}$ and improve its accuracy by pushing it towards $\mathbf{A}$, albeit when the perturbations are not too large so that $\mathbf{E}^{(n)}$ becomes progressively less correlated with $\mathbf{P}_\mathbf{E}$ and $\mathbf{P}_\mathbf{X}$ as the iterations evolve.

REFERENCES

[1] J. M. B. Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

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

[4] G. A. Shaw and H.-H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, 2003.

[5] J. Greer, "Sparse demixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 219–228, Jan. 2012.

[6] R. Arablouei, E. Goan, S. Gensemer, and B. Kusy, "Fast and robust push-broom hyperspectral imaging via DMD-based scanning," in *Proc. SPIE 9948, Novel Optical Systems Design and Optimization XIX*, San Diego, CA, USA, Aug./Sep. 2016, id. 99480A.

[7] R. Arablouei and F. de Hoog, "Hyperspectral image recovery via hybrid regularization," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5649–5663, Dec. 2016.

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

[9] N. Dobigeon, Y. Altmann, N. Brun, and S. Moussaoui, "Linear and nonlinear unmixing in hyperspectral imaging", in *Data Handling in Science and Technology*, vol. 30, Elsevier, 2016, ch. 6, pp. 185–224.

[10] J. Nascimento and J. M. 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.

[11] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, 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.

[12] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.

[13] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.

[14] 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.

[15] J. Li, A. Agathos, D. Zaharie, J. M. B. Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[16] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evolution Remote Sens.*, Grenoble, France, Aug. 2009, pp. 1–4.

[17] 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.

[18] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4194–4209, Nov. 2011.

[19] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *IEEE Int. Geosci. Remote Sens. Symp.*, Barcelona, Spain, Jul. 2007, pp. 4033–4036.

[20] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016.

[21] J. M. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evolution Remote Sens.*, Reykjavik, Iceland, Jun. 2010, vol. 1, pp. 1–4.

[22] M.D. Iordache, "A sparse regression approach to hyperspectral unmixing," Ph.D. dissertation, Inst. Superior Tecnico, TU Lisbon, Lisbon, Portugal, 2011.

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

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

[25] 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.

[26] M.-D. Iordache, J. M. 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.

[27] H. K. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 9, pp. 4257–4266, Sep. 2016.

[28] D. C. Heinz and C.-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[29] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-hall, 1974.

[30] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[31] E. Chouzenoux, M. Legendre, S. Moussaoui, and J. Idier, "Fast constrained least squares spectral unmixing using primal-dual interior-point optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 59–69, 2014.

[32] R. Heylen, M. A. Akhter, and P. Scheunders, "On using projection onto convex sets for solving the hyperspectral unmixing problem," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1522–1526, 2013.

[33] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, Jun. 2001.

[34] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[35] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[36] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[37] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[38] R. Glowinski and A. Marroco, "Sur lapproximation, par éléments finis dordre un, et la résolution, par pénalisation-dualité dune classe de problèmes de Dirichlet non linéaires," *Revue française dautomatique, informatique, recherche opérationnelle. Analyse numérique*, vol. 9, no. 2, pp. 41–76, 1975.

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

[40] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split-Bregman," Center of Applied Mathematics, University of California, Los Angeles, Tech. Rep. 09–31, 2009.

[41] P. A. Thouvenin, N. Dobigeon, J. Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.

[42] L. Drumetz, M. A. Veganzones, S. Henrot, , "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.

[43] A. Halimi, J. M. Bioucas-Dias, N. Dobigeon, G. S. Buller, and S. McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 146–159, Jun. 2017.

[44] C. Agostinelli and L. Greco, "A weighted strategy to handle likelihood uncertainty in Bayesian inference," *Computational Statistics*, vol. 28, no. 1, pp. 319–339, Feb. 2013.

[45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[46] D. Bernstein, *Matrix Mathematics*. Princeton University Press., 2005.

[47] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" In *Proc. Advances Neural Information Processing Systems*, Vancouver, BC, Canada, 2003, pp. 1141–1148.

[48] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[49] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[50] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramér-Rao bound with a singular Fisher information matrix," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 453–456, 2009.

[51] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, Nov. 1992.

[52] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[53] X. Bresson and T. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, vol. 2, no. 4, pp. 455–484, Nov. 2008.

[54] J. Angrist and A. Krueger, "Instrumental variables and the search for identification: From supply and demand to natural experiments," *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 69-85, 2001.

[55] R. J. Bowden and D. A. Turkington, *Instrumental Variables*. Cambridge, UK: Cambridge Univ. Press, 1984.

[56] J. D. Sargan, "The estimation of economic relationships using instrumental variables," *Econometrica*, vol. 26, pp. 393–415, Jul. 1958.

[57] A. Wald, "The fitting of straight lines if both variables are subject to error," *Annals Mathematical Statistics*, vol. 11, pp. 284-300, Sep. 1940.

[58] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[59] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, New York, NY, USA: Springer-Verlag, 2011, pp. 185–212.

[60] L. Condat, "Fast projection onto the simplex and the $\ell 1$ ball," *Mathematical Programming*, Series A, vol. 158, pp. 575–585, 2016.

[61] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, pp. 2283–2302, 2007.

[62] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.

[63] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numerical Anal.*, vol. 17, no. 6, pp. 883–893, Dec. 1980.

[64] T. G. Stockham, Jr., "High-speed convolution and correlation," in *Proc. ACM Spring Joint Computer Conf.*, New York, NY, USA, Apr. 1966, pp. 229–233.

[65] M. Baumgardner, L. Biehl, and D. Landgrebe, "220 band AVIRIS hyperspectral image data set: June 12, 1992 Indian pine test site 3," Purdue University Research Repository, 2015. doi:10.4231/R7RX991C

[66] P. A. Mitchell, "Hyperspectral digital imagery collection experiment (HYDICE)", in *Proc. SPIE 2587*, Nov. 1995, pp. 70–95.

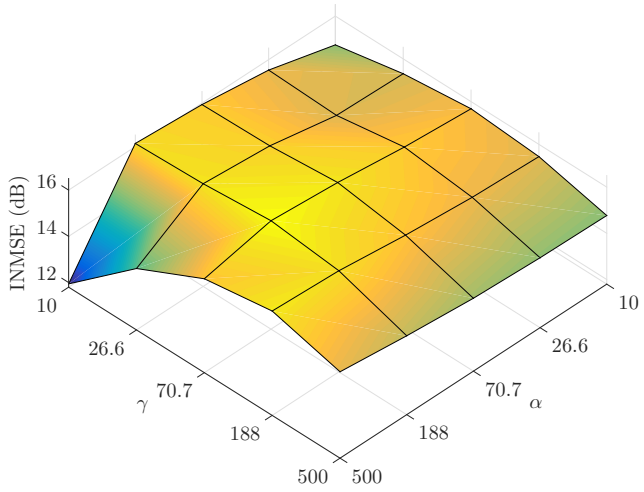
[67] https://github.com/Reza219/Spectral-unmixing-with-perturbed-endmembers

[68] J. R. Magnus, "Matrix differential calculus with applications to simple, Hadamard, and Kronecker products," *J. Mathematical Psychology*, vol. 29, pp. 474–492, 1985.

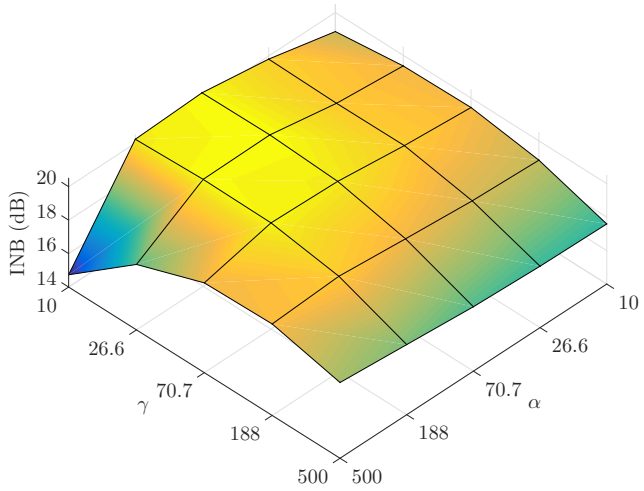
[69] H. V. Henderson and S. R. Searle, "The vec-permutation matrix, the vec operator and Kronecker products: a review," *Linear and Multilinear Algebra*, vol. 9, no. 4, pp. 271–288, 1981.

[70] G. Grimmett and D.Stirzaker, *Probability and Random Processes*. 3rd ed., Oxford, UK: Oxford Univ. Press, 2001.

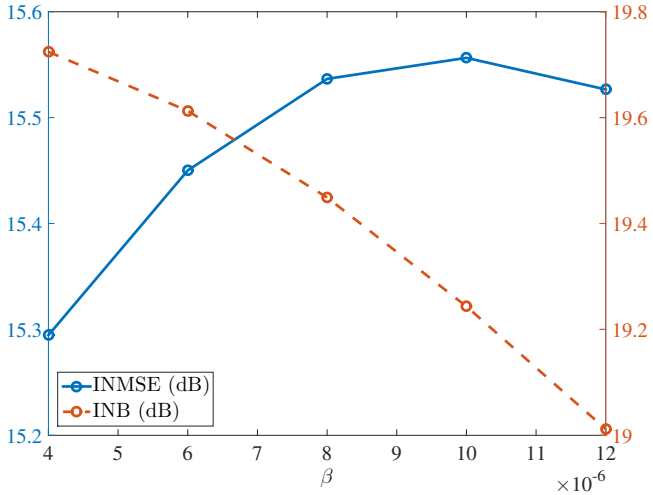
[71] G. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.



(a) INMSE for different values of α and γ while $\beta = 6 \times 10^{-6}$.

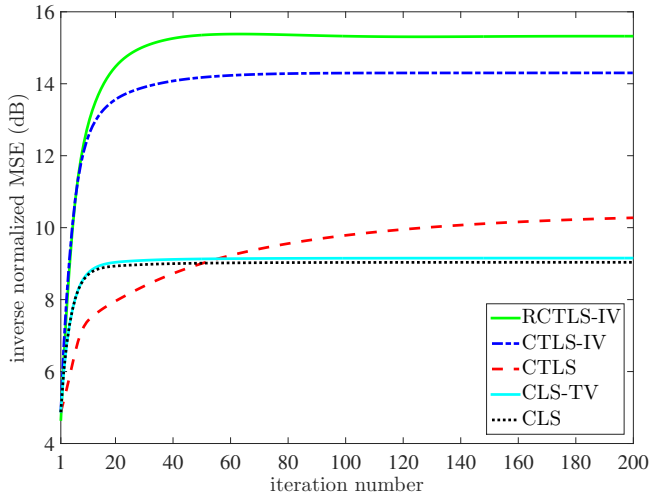


(b) INB for different values of α and γ while $\beta = 6 \times 10^{-6}$.

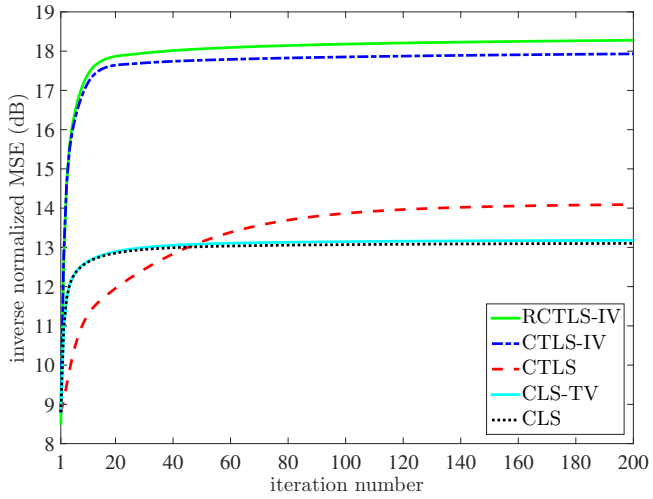


(c) INMSE and INB for different values of β while $\alpha = 100$ and $\gamma = 50$.

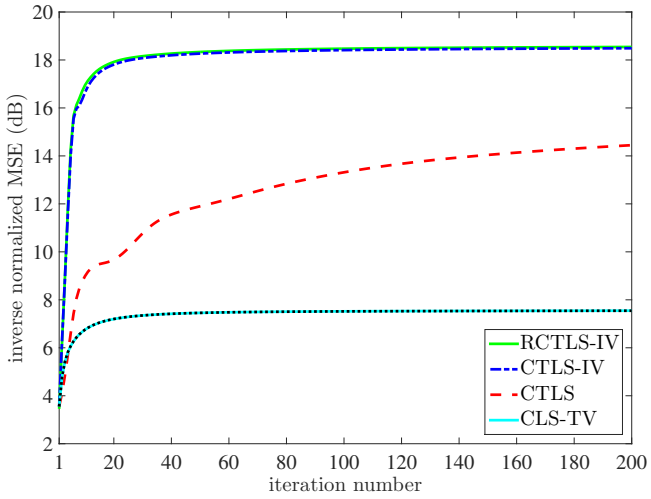
Fig. 2. The performance of the proposed algorithm with different values of the regularization parameters α, β, and γ when unmixing the Botswana dataset with SNR $= 30$ dB.



(a) Botswana

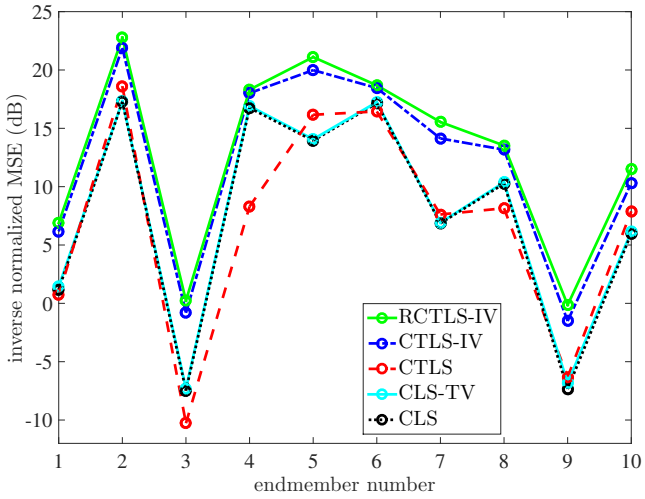


(b) Indian Pines

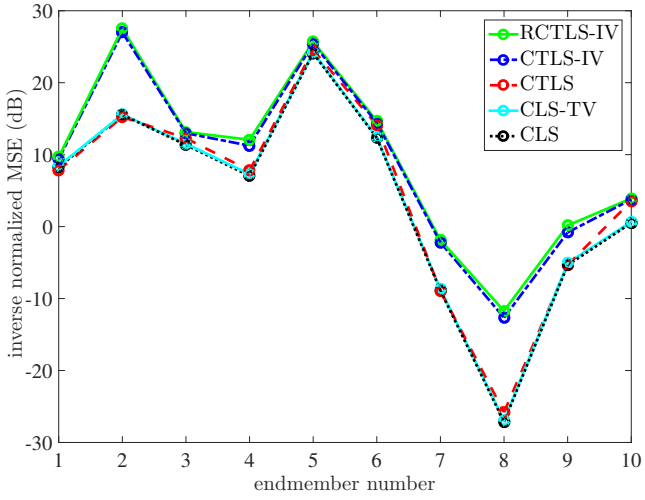


(c) Washington DC

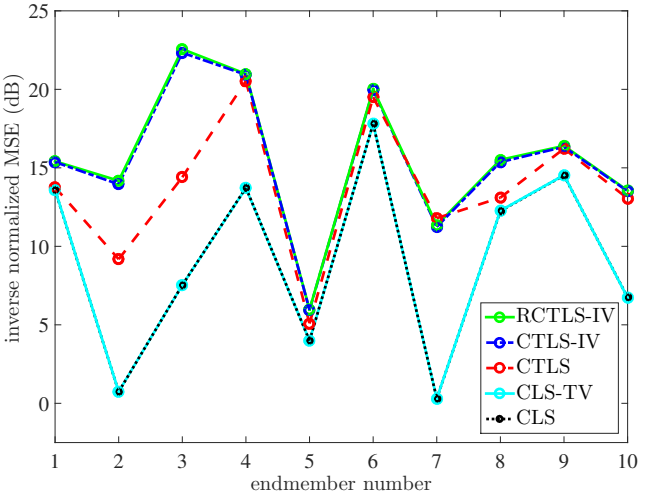
Fig. 3. The evolution of the INMSE over iterations for different algorithms and datasets with SNR $= 30$ dB.



(a) Botswana

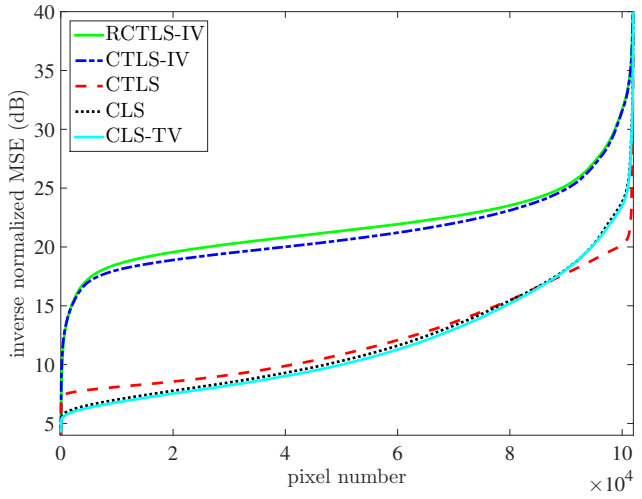


(b) Indian Pines

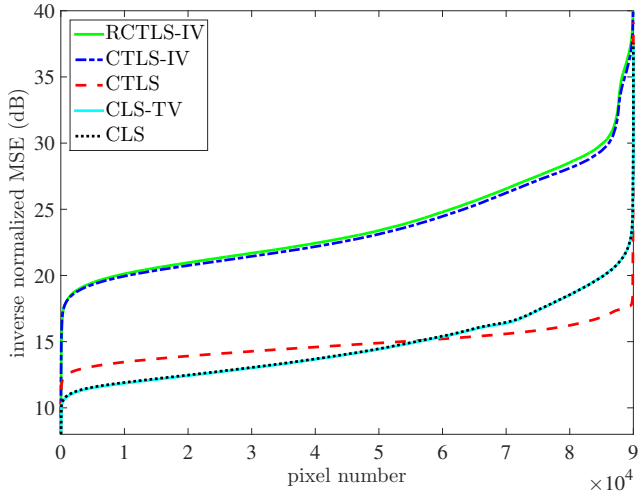


(c) Washington DC

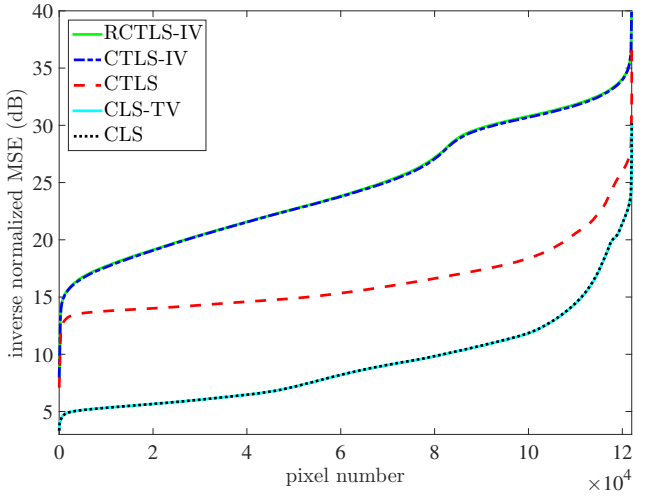
Fig. 4. The INMSE of abundance bands corresponding to each endmember for different algorithms and datasets with SNR = 30 dB.



(a) Botswana



(b) Indian Pines



(c) Washington DC

Fig. 5. The INMSE of each pixel, sorted in the ascending order, for different algorithms and datasets with SNR = 30 dB.

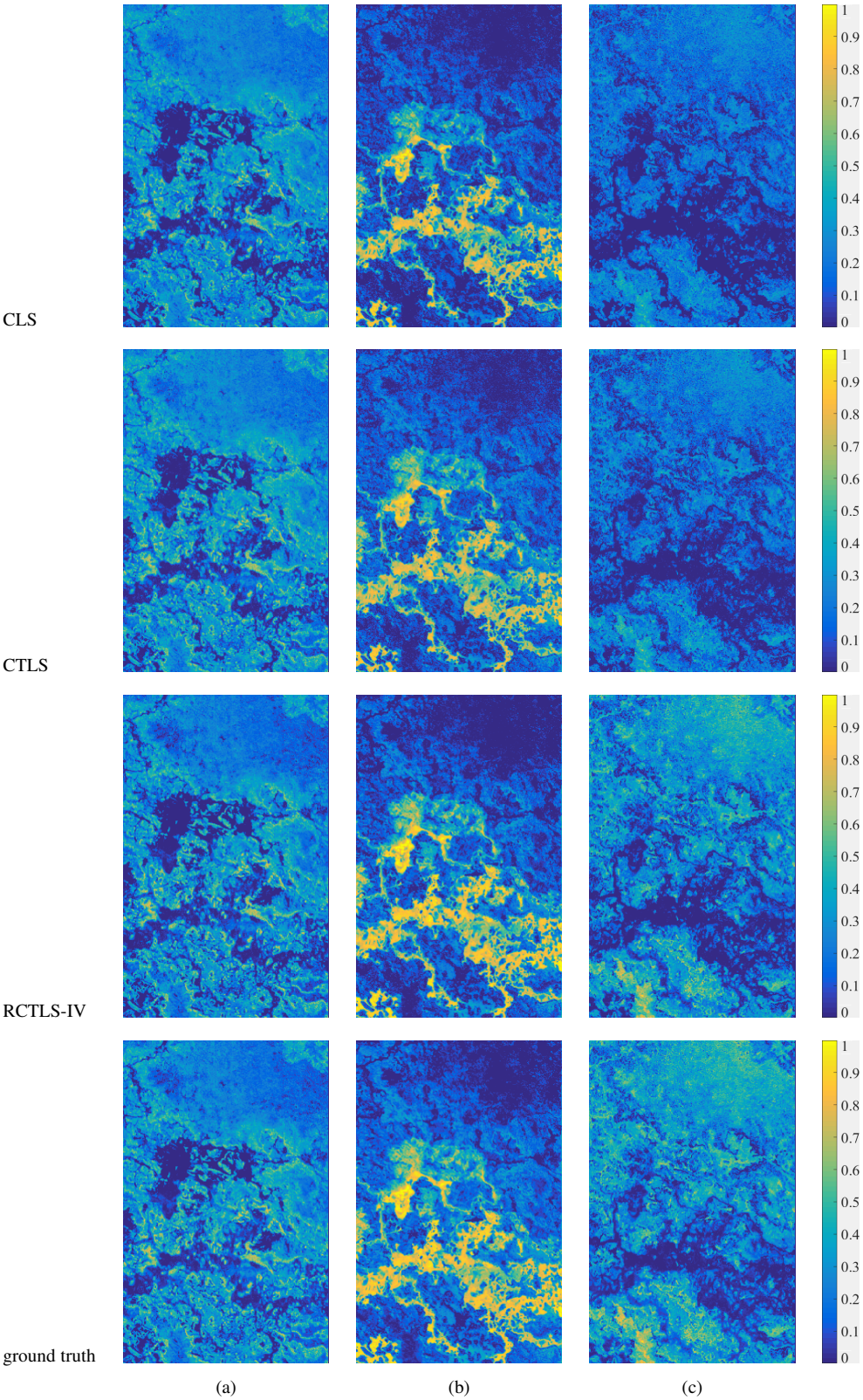


Fig. 6. The abundance maps corresponding to three endmembers, (a)–(c), estimated by the CLS, CTLS, and RCTLS-IV algorithms for Botswana dataset with $\mathrm{SNR} = 30$ dB together with the ground truth.

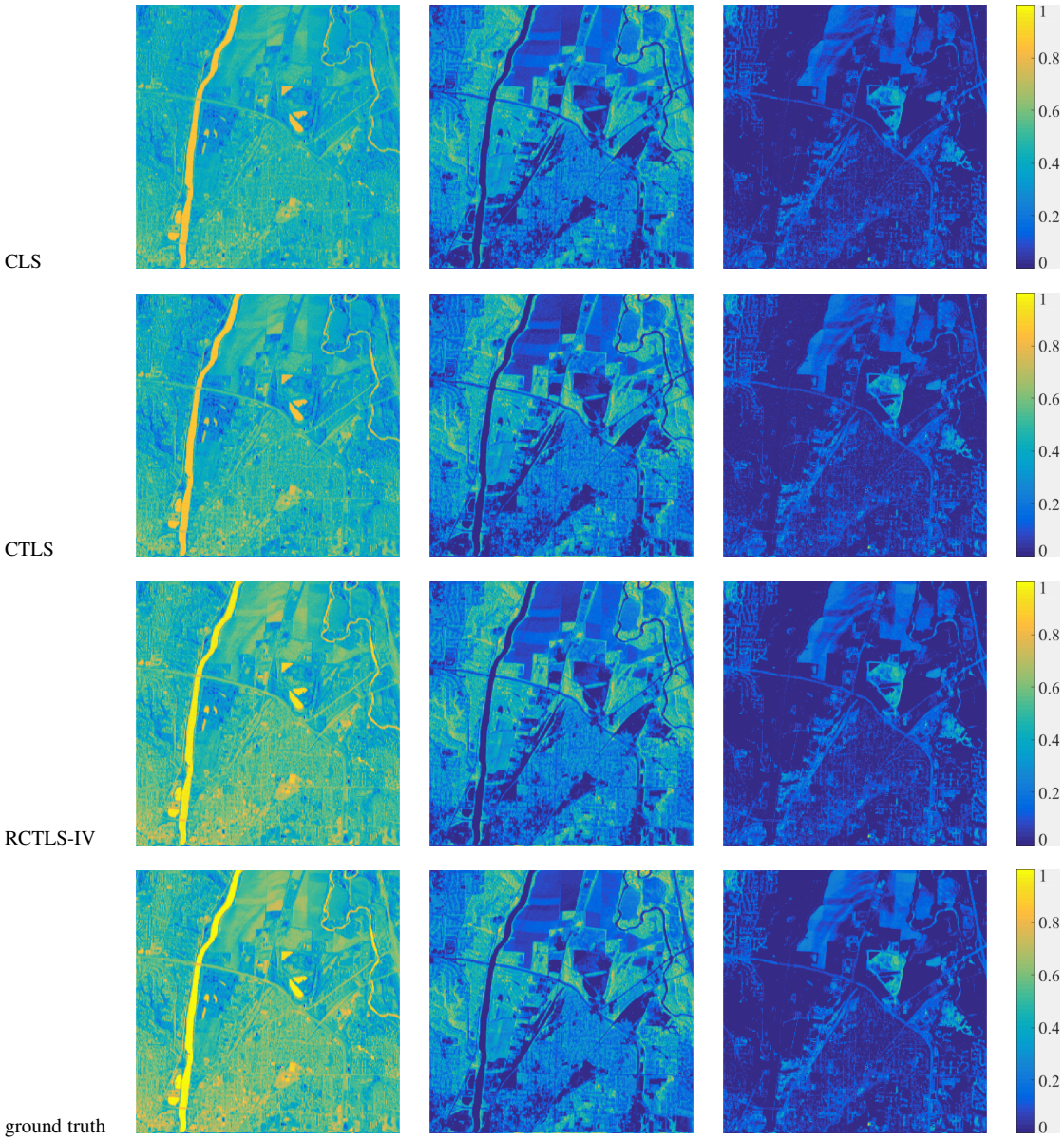


Fig. 7. The abundance maps corresponding to three endmembers, (a)–(c), estimated by the CLS, CTLS, and RCTLS-IV algorithms for Indian Pines dataset with $\mathrm{SNR} = 30$ dB together with the ground truth.

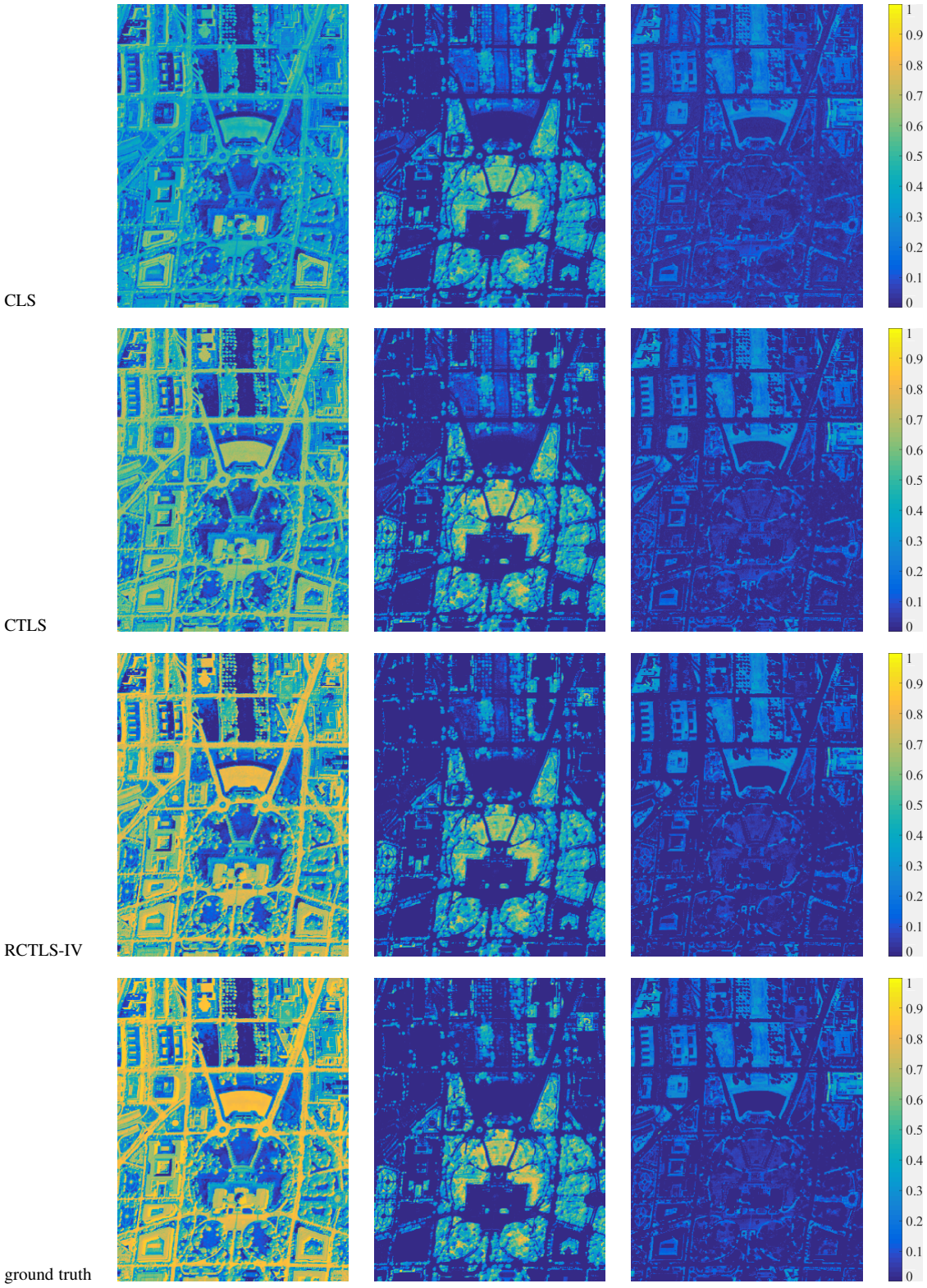


Fig. 8. The abundance maps corresponding to three endmembers, (a)–(c), estimated by the CLS, CTLS, and RCTLS-IV algorithms for Washington DC dataset with $\mathrm{SNR} = 30$ dB together with the ground truth.