A Convex Optimization-Based Coupled Nonnegative Matrix Factorization Algorithm for Hyperspectral and Multispectral Data Fusion

Chia-Hsiang Lin[ID], Fei Ma, Chong-Yung Chi, *Senior Member, IEEE*, and Chih-Hsiang Hsieh

Abstract—Fusing a low-spatial-resolution hyperspectral data with a high-spatial-resolution (HSR) multispectral data has been recognized as an economical approach for obtaining HSR hyperspectral data, which is important to accurate identification and classification of the underlying materials. A natural and promising fusion criterion, called coupled nonnegative matrix factorization (CNMF), has been reported that can yield high-quality fused data. However, the CNMF criterion amounts to an ill-posed inverse problem, and hence, advisable regularization can be considered for further upgrading its fusion performance. Besides the commonly used sparsity-promoting regularization, we also incorporate the well-known sum-of-squared-distances regularizer, which serves as a convex surrogate of the volume of the simplex of materials' spectral signature vectors (i.e., endmembers), into the CNMF criterion, thereby leading to a convex formulation of the fusion problem. Then, thanks to the biconvexity of the problem nature, we decouple it into two convex subproblems, which are then, respectively, solved by two carefully designed alternating direction method of multipliers (ADMM) algorithms. Closed-form expressions for all the ADMM iterates are derived via convex optimization theories (e.g., Karush–Kuhn–Tucker conditions), and furthermore, some matrix structures are employed to obtain alternative expressions with much lower computational complexities, thus suitable for practical applications. Some experimental results are provided to demonstrate the superior fusion performance of the proposed algorithm over state-of-the-art methods.

Index Terms—Alternating direction method of multipliers (ADMM), convex optimization, coupled nonnegative matrix factorization (CNMF), data fusion, hyperspectral data.

I. INTRODUCTION

HIGH-spatial-resolution (HSR) hyperspectral images are of paramount importance to accurate identification and classification of the underlying materials [1]–[3]. However, the spatial resolution of hyperspectral sensors is

generally more limited than that of multispectral sensors (or panchromatic sensors) [4]. While direct acquisition of HSR hyperspectral images is expensive, fusion of multisensor images serves as a more economical alternative for obtaining such images, and has received enormous attention in the past decade in the Geoscience and Remote Sensing Society (GRSS) [1]–[3], [5], [6], leading to the annual IEEE GRSS Data Fusion Contest. An effective approach for obtaining an HSR hyperspectral image is to fuse a low-spatial-resolution (LSR) hyperspectral image with an HSR multispectral image (e.g., RGB data [7]), and both are acquired over the same scene and possibly under similar atmospheric/illumination conditions [8]–[10]. Acquisition of such complementary hyperspectral/multispectral data becomes increasingly available due to the high-resolution optical satellites and the ever improving revisit cycles (see [10] and the references therein).

One category of hyperspectral/multispectral data fusion methods is rooted in the pansharpening technique [11]–[16], which has been used to enhance the spatial resolution of multispectral imagery (by fusing it with an HSR panchromatic image) [5]. For example, the multiresolution analysis-based pansharpening methods [17] can be adapted for hyperspectral and multispectral data fusion, by linearly combining multispectral band images (via linear regression) to synthesize HSR image for each hyperspectral band—the so-called hypersharpening multiresolution analysis (HSMRA) technique [17], [18]. Another representative pansharpening method, called component substitution (CS) [13], can also be adapted for hyperspectral and multispectral data fusion, by applying it to each multispectral band and the corresponding grouped hyperspectral bands (constructed by correlation-based clustering) [10].

Another category of fusion methods models the multispectral and hyperspectral data as spectrally and spatially degraded versions of the high spectral/spatial resolution data (denoted by **Z**), respectively, and considers the fusion problem as an inverse problem of recovering **Z** from the observable hyperspectral/multispectral data [see (1) and (2)]. Since such an inverse problem is ill-posed in general [19, Sec. III.A], incorporating a generic prior (e.g., Gaussian prior assigned to the quantity **Z** [20]) has been considered to design various fusion methods in the original image domain [21] or in the transformed wavelet domain [22], and fusion methods in this category are collectively referred to as Bayesian inference

method (BIM) [19]. In [23], by introducing a noise-corrupted version of $\mathbf{Z}$ (denoted by $\mathbf{Z}'$), the LSR hyperspectral data are modeled as a blurred version of $\mathbf{Z}'$. Naturally, $\mathbf{Z}'$ is considered as missing data, and the expectation–maximization algorithm can be used to estimate $\mathbf{Z}$ [23]. As the E-step and M-step can be elegantly interpreted as a deblurring stage and a denoising stage, respectively, the method proposed in [23] actually performs both image fusion and restoration in the meantime. In [20], the Gaussian prior is assigned to (dimension-reduced and vectorized) $\mathbf{Z}$. As the method [20] proposed to estimate the associated hyperparameters from the data (instead of fixing them *a priori*), the induced joint posterior distribution is too complicated to be directly maximized. Hence, Gibbs sampler [24] (implemented by the Hamilton Monte Carlo technique [25]) is adopted to generate a collection of samples that asymptotically follow the posterior distribution, and then, these samples are averaged to estimate $\mathbf{Z}$. As Bayesian fusion framework usually involves complicated, nonconvex likelihood or posterior, the corresponding algorithms are often not computationally cheap [20]. In view of this, reconsidering the fusion problem under the convex optimization framework is quite appealing. A representative method in this line, termed hyperspectral superresolution (HySure) [26], proposes a convex formulation based on the so-called vector total variation regularizer [27] to simultaneously incorporate the spatial and spectral data characteristics, and solves the problem via the split augmented Lagrangian shrinkage algorithm [28]. In the pioneering work [26], the superresolution image $\mathbf{Z}$ adheres to a low-rank model $\mathbf{Z} = \mathbf{AS}$, where $\mathbf{A}$ is assumed known and $\mathbf{S}$ is considered a dimension-reduced version of $\mathbf{Z}$, leading to a concise convex formulation of the fusion problem with variable $\mathbf{S}$.

A natural way for hyperspectral/multispectral data fusion is based on spectral unmixing [29], for which both $\mathbf{A}$ and $\mathbf{S}$ are considered unknowns. Fusion methods in this category extract the spectral information by learning the material signature (also known as endmember) matrix $\mathbf{A}$ from hyperspectral data, and the extracted spectral information and the multispectral data are used to reconstruct the HSR material abundance matrix $\mathbf{S}$ (usually via sparse regression). Then, these spectral information and spatial information are combined to form the desired HSR hyperspectral fused data under the linear mixing model [29]. A representative method in this line is based on the promising coupled nonnegative matrix factorization (CNMF) criterion [30]. As aforementioned, $\mathbf{Z}$ may not be uniquely inversed from its spectrally and spatially degraded versions (i.e., the observable multispectral and hyperspectral data), making the fusion an ill-posed problem in general [19]. Just like prior information is needed in Bayesian fusion criteria, the CNMF criterion also requires regularization. The most commonly used regularizers are to promote the sparsity of abundance maps based on ℓ_0-norm, or its convex surrogate ℓ_1-norm [31]. Another commonly used regularizer in spectral unmixing is based on the volume of the endmember simplex as considered in the minimum-volume-constrained nonnegative matrix factorization [32]. Although such a simplex volume minimization approach has been theoretically

proven to be effective in yielding high-fidelity material signatures [33], [34], the simplex volume regularizer is nonconvex [35] and hence may degrade the fusion efficiency. In view of this, the sum-of-squared distances (SSD) between all the simplex vertices, termed SSD regularizer [36], can be used as an effective convex surrogate of the simplex volume.

In this paper, following the philosophy in [26], but considering both $\mathbf{A}$ and $\mathbf{S}$ as unknowns under the CNMF framework, we formulate the fusion as a biconvex problem that can be elegantly solved by the convex optimization theory. Specifically, we simultaneously incorporate both the convex sparsity-promoting ℓ_1-norm regularizer and the convex volume-demoting SSD regularizer into the CNMF criterion. This regularization strategy has been adopted for CNMF previously [37], termed volume and sparsity constrained CNMF (VSC-CNMF). Nevertheless, the formulation in [37] only considers the data fitting error of hyperspectral data [37, eq. (13)]; although this allows adopting the multiplicative update rules to solve the induced problem [37], ignoring the data fitting error of multispectral data actually violates the design of the original (unregularized) CNMF framework [30], leading to not much fusion performance difference compared with the unregularized one (see Section IV). Instead, we propose a convex optimization-based CNMF (CO-CNMF) algorithm to solve the fusion problem with both data fitting terms taken into consideration. By merit of the biconvexity nature, we employ the block coordinate descent method [38] to decouple the regularized CNMF problem into two convex subproblems, and solve each subproblem by a carefully designed alternating direction method of multipliers (ADMM) [39]. We not only derive closed-form solutions for the ADMM iterates, but also exploit some intrinsic matrix structures to refine the closed-form expressions for two-core ADMM iterates. The refined expressions are proved with much lower computational complexity than their original expressions. This is very important in practice as ADMM may lead to computationally complex algorithm (even if each iterate is equipped with a closed-form expression) when the dimension of data is large, such as in hyperspectral imagery analysis [40]. Moreover, we prove that CO-CNMF also holds the stationarity convergence property, serving as the counterpart of the NMF convergence theory [41]. The superior fusion performance of the proposed CO-CNMF algorithm, over some benchmark fusion methods, will be demonstrated by experiments based on Wald's protocol [42].

The remaining part of this paper is organized as follows. In Section II, we present the signal model that describes the spectral/spatial relationships among the hyperspectral data, the multispectral data, and the fused data, followed by our formulation of the regularized CNMF problem. Then, we develop a convex optimization algorithm to solve the fusion problem in Section III, and present some experimental results and performance comparison in Section IV. Finally, we draw some conclusions in Section V. Some detailed derivations and theoretical proofs are given in the appendices.

The following notations will be used in the ensuing presentation. $\|\cdot\|_p$ denotes the p-norm. $\|\cdot\|_F$ denotes the Frobenius norm. Boldface $\mathbf{0}$, $\mathbf{1}$, and $\mathbf{I}$, respectively, denote the all-zero
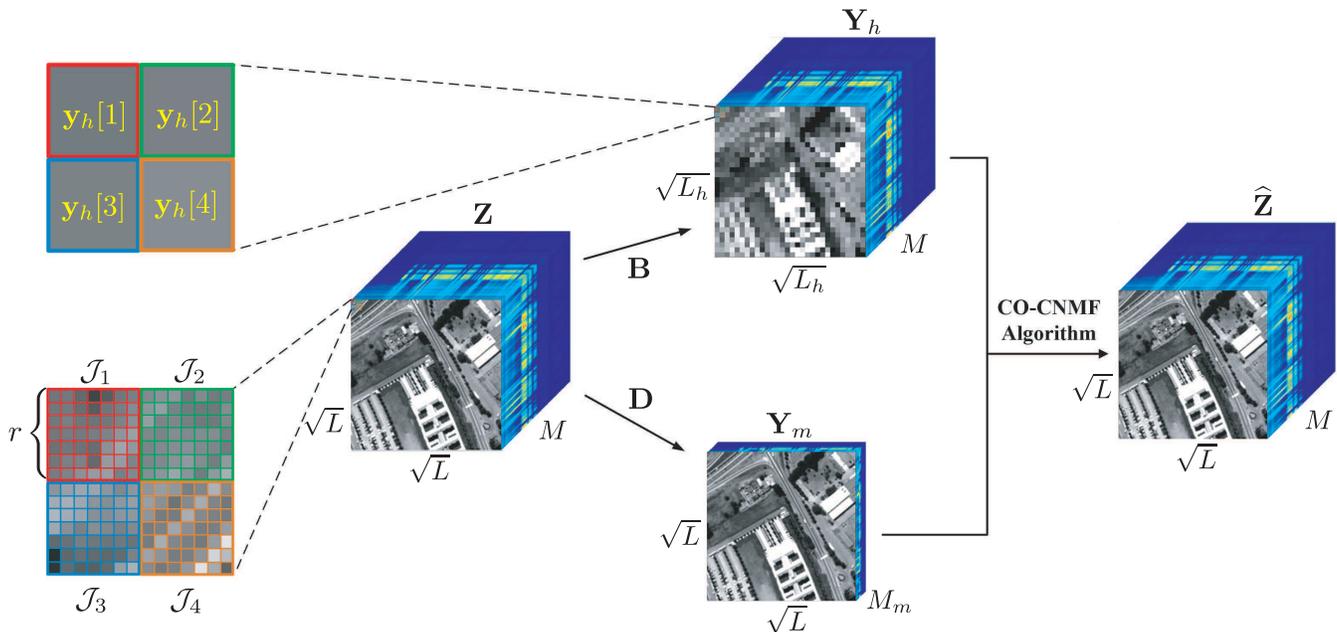
Fig. 1. Observations, $\mathbf{Y}_h$ and $\mathbf{Y}_m$, can be modeled as spatially blurred (i.e., $L_h < L$) and spectrally downsampled (i.e., $M_m < M$) versions of the HSR hyperspectral data $\mathbf{Z}$ [see (1) and (2)]. Each pixel $\mathbf{y}_h[n]$ (in the image $\mathbf{Y}_h$) corresponds to the spatial region $\mathcal{J}_n$, which covers r^2 HSR pixels in $\mathbf{Z}$ [see (19)]. Therefore, assuming perfect spatial coregistration [30], the spectral mixture $\mathbf{y}_h[n]$ can be obtained by intermingling the r^2 spectral mixtures that, respectively, correspond to the r^2 subregions of $\mathcal{J}_n$ [see (20)]. The aim is to fuse the LSR hyperspectral data $\mathbf{Y}_h$ and the HSR multispectral data $\mathbf{Y}_m$, thereby yielding an HSR hyperspectral data $\widehat{\mathbf{Z}}$, via convex formulation and optimization.

vector, all-one vector, and identity matrix of proper dimension (indicated in their subscripts). $\mathbf{e}_i^{(m)}$ is the ith m-dimensional unit vector. $\mathcal{I}_L \triangleq \{1, \ldots, L\}$ for any positive integer L. conv $\mathcal{S}$ denotes the convex hull of the set $\mathcal{S}$ [38]. $\mathbf{vec}(\mathbf{X})$ is the vector formed by sequentially stacking the columns of the matrix $\mathbf{X}$. $\otimes$ stands for the Kronecker product. $[\cdot]_+$ denotes the orthogonal projection onto nonnegative orthant of the Euclidean space. $\succeq$ is the componentwise inequality.

II. SIGNAL MODEL AND PROBLEM FORMULATION

To obtain the desired HSR hyperspectral data $\mathbf{Z} \in \mathbb{R}^{M \times L}$, an economical approach is to fuse the HSR multispectral data $\mathbf{Y}_m \in \mathbb{R}^{M_m \times L}$ and the LSR hyperspectral data $\mathbf{Y}_h \in \mathbb{R}^{M \times L_h}$, where M_m and L (respectively, M and L_h) denote the number of spectral bands and the number of pixels in $\mathbf{Y}_m$ (respectively, $\mathbf{Y}_h$), respectively. The observations, $\mathbf{Y}_h$ and $\mathbf{Y}_m$, can be modeled as spatially degraded (i.e., $L_h < L$) and spectrally degraded (i.e., $M_m < M$) versions of the fused data $\mathbf{Z}$; to be precise, we have (see Fig. 1)

$$\mathbf{Y}_m = \mathbf{DZ} + \mathbf{E}_m \tag{1}$$
$$\mathbf{Y}_h = \mathbf{ZB} + \mathbf{E}_h \tag{2}$$

where the spectral response transform matrix $\mathbf{D} \in \mathbb{R}^{M_m \times M}$ downsamples the hyperspectral bands of $\mathbf{Z}$ leading to the multispectral data $\mathbf{Y}_m$, the spatial spread transform matrix $\mathbf{B} \in \mathbb{R}^{L \times (L/r^2)}$ blurs the HSR data $\mathbf{Z}$ leading to the LSR data $\mathbf{Y}_h$, in which the blurring factor $r \triangleq (L/L_h)^{1/2}$ is assumed to be an integer [19], and $\mathbf{E}_h$ and $\mathbf{E}_m$ are the residuals [30]. $\mathbf{B}$ and $\mathbf{D}$ can be obtained by image registration and radiometric calibration, and so they are assumed to be known

a priori [30]. Hence, as suggested by (1) and (2), the fusion problem is to recover the high spectral/spatial resolution data $\mathbf{Z}$ from the observable $\mathbf{Y}_m$ and $\mathbf{Y}_h$—a linear inverse problem with $\mathbf{Z}$ being the causal factor [28]. Note that there are other approaches to formulate the fusion as an inverse problem by considering different causal factors (e.g., materials' abundances) under the Bayesian framework [19]. Such an inverse problem is ill-posed in general (e.g., in the pansharpening case, i.e., $M_m = 1$) [19, Sec. III.A], and hence requires regularization (in the CNMF framework) or prior information (in the Bayesian framework); for instance, the Gaussian prior has been assigned to the quantity $\mathbf{Z}$ [20, Sec. 3.2].

Furthermore, in (1) and (2), each hyperspectral pixel of $\mathbf{Z} = [\mathbf{z}[1], \ldots, \mathbf{z}[L]]$ can be modeled as a mixture of the underlying materials' spectral signature vectors (also known as endmembers), namely [2], [43]

$$\mathbf{z}[n] = \sum_{i=1}^{N} s_i[n]\, \mathbf{a}_i = \mathbf{As}[n], \quad n \in \mathcal{I}_L \tag{3}$$

where the ith column of $\mathbf{A} \triangleq [\mathbf{a}_1, \ldots, \mathbf{a}_N] \succeq \mathbf{0}_{M \times N}$ is the ith endmember signature with N denoting the number of endmembers, and $\mathbf{s}[n] \triangleq [s_1[n], \ldots, s_N[n]]^T \succeq \mathbf{0}_N$ is the nth abundance vector. Note that (3) can be written in the matrix form as $\mathbf{Z} = \mathbf{AS}$, where both $\mathbf{A}$ and $\mathbf{S} \triangleq [\mathbf{s}[1], \ldots, \mathbf{s}[L]]$ are assumed to be nonnegative [2], [43]. Regarding the sum-to-one assumption of $\mathbf{S}$ (i.e., $\mathbf{S}^T \mathbf{1}_N = \mathbf{1}_L$), it may not perfectly hold in real data. Indeed, some benchmark fusion methods consider the sum-to-one assumption [30], while some do not [19], [26]. Our empirical experiences suggest that the sum-to-one constraint could have a negative

impact on the fusion performance, and hence, we opt not to enforce it in our method.

A natural data fusion criterion arising from (1)–(3) has been proposed in [30] to minimize data fitting residuals[1]

$$\text{CNMF}(\mathbf{A}, \mathbf{S}) \triangleq \|\mathbf{Y}_h - \mathbf{ASB}\|_F^2 + \|\mathbf{Y}_m - \mathbf{DAS}\|_F^2 \quad (4)$$

while, as such criterion itself amounts to an ill-posed problem, we further incorporate some regularization mechanisms based on the physical considerations on $\mathbf{A}$ and $\mathbf{S}$. Specifically, the (regularized) CNMF criterion reconstructs the fused data $\mathbf{Z} = \mathbf{AS}$ by solving the following problem:

$$\min_{\mathbf{A}, \mathbf{S}} \ \frac{1}{2} \text{CNMF}(\mathbf{A}, \mathbf{S}) + \lambda_1 \, \phi_1(\mathbf{A}) + \lambda_2 \, \phi_2(\mathbf{S})$$
$$\text{s.t. } \mathbf{A} \succeq \mathbf{0}_{M \times N}, \quad \mathbf{S} \succeq \mathbf{0}_{N \times L} \quad (5)$$

where $\text{CNMF}(\mathbf{A}, \mathbf{S})$ is the original (unregularized) CNMF criterion [30], $\lambda_1 > 0$ and $\lambda_2 > 0$ are the regularization parameters, and the two regularizers are selected as

$$\phi_1(\mathbf{A}) \triangleq \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2, \quad \phi_2(\mathbf{S}) \triangleq \|\mathbf{S}\|_1. \quad (6)$$

$\phi_1(\mathbf{A})$ is the well-known SSD regularizer in hyperspectral imagery, serving as the convex surrogate[2] of the volume of the simplex $\text{conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ [36]. Note that the simplex volume minimization approach has been theoretically proven to be capable of yielding high-fidelity endmember estimates [33], suggesting that $\phi_1(\mathbf{A})$ be an ideal and natural regularizer for the CNMF problem (5) [37]. $\phi_2(\mathbf{S})$, which is the convex envelope of $\|\mathbf{S}\|_0$ (i.e., the number of nonzero elements in $\mathbf{S}$) [38], is assigned to reflect the abundance sparsity in practical scenarios [44].

Although the regularization strategy of (5) has been considered in [37] for judiciously incorporating the nature of $\mathbf{A}$ and $\mathbf{S}$, the (revised) multiplicative update rules adopted in [37] may not effectively solve the regularized CNMF problem (5), leading to marginal fusion performance improvement compared with the unregularized one [30]. In Section III, we solve (5) with a more delicately designed algorithm

---

[1]It is possible to consider different weights in the CNMF criterion, i.e., $\text{CNMF}(\mathbf{A}, \mathbf{S}) \triangleq \|\mathbf{Y}_h - \mathbf{ASB}\|_F^2 + \rho \|\mathbf{Y}_m - \mathbf{DAS}\|_F^2$ for some $\rho \neq 1$. For example, VSC-CNMF considers $\rho = 0$ [37]. In case that a prior knowledge is available for selecting a suitable $\rho \geq 0$ (but $\rho \neq 1$), one can redefine $(\mathbf{C}_1^{(\mathbf{A})}, \mathbf{C}_2^{(\mathbf{S})}, \mathbf{y})$ in (7) and (8) as

$$\mathbf{C}_1^{(\mathbf{A})} \triangleq [(\mathbf{B}^T \otimes \mathbf{A})^T, \ \sqrt{\rho}(\mathbf{I}_L \otimes \mathbf{DA})^T]^T$$
$$\mathbf{C}_2^{(\mathbf{S})} \triangleq [((\mathbf{SB})^T \otimes \mathbf{I}_M)^T, \ \sqrt{\rho}(\mathbf{S}^T \otimes \mathbf{D})^T]^T$$
$$\mathbf{y} \triangleq [\text{vec}(\mathbf{Y}_h)^T, \ \text{vec}(\sqrt{\rho}\mathbf{Y}_m)^T]^T$$

when simplifying the criterion as a single convex quadratic term, while other developments still remain valid. However, when such a prior knowledge is unavailable, assigning equal weights in fitting the entries of $(\mathbf{Y}_h, \mathbf{Y}_m)$ (i.e., $\rho = 1$) is more reasonable [30] and will be considered in this paper.
[2]From some basic matrix analysis, one can verify that $\log(\det(\bar{\mathbf{A}}^T \bar{\mathbf{A}})) \leq \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2$, where $\bar{\mathbf{A}} \triangleq [\mathbf{a}_1 - \mathbf{a}_N, \ldots, \mathbf{a}_{N-1} - \mathbf{a}_N]$ $(N \geq 2)$, and $\log(\cdot)$ [respectively, $\det(\cdot)$] denotes natural logarithm (respectively, determinant). By plugging the simplex volume expression $\text{vol}(\text{conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}) = (1/((N-1)!))(\det(\bar{\mathbf{A}}^T \bar{\mathbf{A}}))^{1/2}$ [33] into the above inequality, we see that $\log(\text{vol}(\text{conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\})) \leq \phi_1(\mathbf{A}) \ \forall N \geq 2$. Therefore, the SSD regularizer can be viewed as a convex upper bound approximation of the (log-transformed) simplex volume.

using convex optimization tools [38]. The proposed algorithm, as will be seen, shows that the regularization strategy in the formulation (5) indeed yields much better fusion performance than the unregularized one.

III. CONVEX OPTIMIZATION-BASED DATA FUSION ALGORITHM

In this section, we solve (5) with an algorithm devised based on convex optimization theories, such as ADMM and Karush–Kuhn–Tucker (KKT) conditions [38]. We not only derive closed-form expressions for each algorithmic step, but also exploit the structure/sparsity of the matrices (involved in these expressions) to greatly reduce the computational complexity.

We begin by noticing that although (5) is a nonconvex problem, it is a biconvex one. To be precise, when $\mathbf{A}$ (respectively, $\mathbf{S}$) is fixed, (5) is a convex problem with variable $\mathbf{S}$ (respectively, $\mathbf{A}$) (see [45, Definition 1.3] for the definition of a biconvex problem). To see it, we define $\mathbf{s} \triangleq \text{vec}(\mathbf{S})$, $\mathbf{a} \triangleq \text{vec}(\mathbf{A})$, and $\mathbf{y} \triangleq [\text{vec}(\mathbf{Y}_h)^T, \ \text{vec}(\mathbf{Y}_m)^T]^T$. Then, the biconvexity of (5) can be easily seen by expressing $\text{CNMF}(\mathbf{A}, \mathbf{S})$ as a (convex) quadratic form in terms of $\mathbf{s}$ (for fixed $\mathbf{A}$ [see (7)]), and as a (convex) quadratic form in terms of $\mathbf{a}$ (for fixed $\mathbf{S}$ [see (8)]). Precisely, with some matrix manipulations, we have

$$\text{CNMF}(\mathbf{A}, \mathbf{S}) = \left\| \mathbf{C}_1^{(\mathbf{A})} \mathbf{s} - \mathbf{y} \right\|_2^2 \quad (7)$$
$$= \left\| \mathbf{C}_2^{(\mathbf{S})} \mathbf{a} - \mathbf{y} \right\|_2^2 \quad (8)$$

where

$$\mathbf{C}_1^{(\mathbf{A})} \triangleq [(\mathbf{B}^T \otimes \mathbf{A})^T, \ (\mathbf{I}_L \otimes \mathbf{DA})^T]^T$$
$$\in \mathbb{R}^{(ML_h + M_m L) \times NL} \quad (9)$$
$$\mathbf{C}_2^{(\mathbf{S})} \triangleq [((\mathbf{SB})^T \otimes \mathbf{I}_M)^T, \ (\mathbf{S}^T \otimes \mathbf{D})^T]^T$$
$$\in \mathbb{R}^{(ML_h + M_m L) \times NM}. \quad (10)$$

Since (5) is a biconvex problem, alternating optimization is suitable for solving it.

Specifically, the proposed CO-CNMF algorithm is initialized with $\mathbf{A}^0$, which is obtained by the HyperCSI algorithm [43] (that is fast and reproducible for estimating the endmember signatures [43]), followed by alternatively solving the following two convex subproblems until convergence:

$$\mathbf{S}^{k+1} \in \arg \min_{\mathbf{S} \succeq \mathbf{0}_{N \times L}} \ \frac{1}{2} \text{CNMF}(\mathbf{A}^k, \mathbf{S}) + \lambda_2 \, \phi_2(\mathbf{S}) \quad (11)$$

$$\mathbf{A}^{k+1} \in \arg \min_{\mathbf{A} \succeq \mathbf{0}_{M \times N}} \ \frac{1}{2} \text{CNMF}(\mathbf{A}, \mathbf{S}^{k+1}) + \lambda_1 \, \phi_1(\mathbf{A}) \quad (12)$$

where the superscript "k" in $\mathbf{S}^k$ and $\mathbf{A}^k$ denotes the iteration number rather than their power. The proposed CO-CNMF algorithm is summarized in Algorithm 1, whose convergence property is detailed in Proposition 1 with proof relegated to Appendix A.

Proposition 1: Assuming that Algorithms 2 and 3 exactly solve (11) and (12), respectively, the sequence $\{(\mathbf{A}^k, \mathbf{S}^k)\}$ generated by Algorithm 1 converges to a stationary point [38] of (5).

Algorithm 1 CO-CNMF Algorithm for Solving (5)

1: **Given** $\mathbf{Y}_h$, $\mathbf{Y}_m$, $\mathbf{B}$ and $\mathbf{D}$.
2: Set $k := 0$, and initial $\mathbf{A}^0$ (obtained by HyperCSI [43]).
3: **repeat**
4:    Solve (11) by Algorithm 2, and update $\mathbf{S}^{k+1}$;
5:    Solve (12) by Algorithm 3, and update $\mathbf{A}^{k+1}$;
6:    $k := k + 1$;
7: **until** the predefined stopping criterion is met.
8: **Output** HSR hyperspectral fused data $\mathbf{Z} = \mathbf{A}^k \mathbf{S}^k$.

Remark 1: Note that both (11) and (12) are convex problems. Therefore, exact minimization for them (assumed in Proposition 1) is achievable through some off-the-shelf convex solvers (e.g., CVX [46]). However, these general-purpose solvers can be quite slow especially when the problem size is large. We are hence motivated to develop customized convex solvers for efficiently handling (11) and (12) based on ADMM, that is known to be effective for big data (convex) optimization [39]. Note that under the very mild condition that the linear equality constraint $\boldsymbol{v}_1 = \boldsymbol{M}\boldsymbol{v}_2$ has $\boldsymbol{M}^T\boldsymbol{M}$ to be invertible (here, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the two primal variables in ADMM formulation), a limit point of the sequence $\{(\boldsymbol{v}_1^j, \boldsymbol{v}_2^j)\}$ generated by ADMM (j is the ADMM iteration number) must be a global minimizer [38, Lemma 9.2]. Obviously, the above condition holds true for both ADMM Algorithms 2 and 3 [see (14) and (23), for both of which $\boldsymbol{M}^T\boldsymbol{M}$ in the above condition reduces to the identity matrix and hence invertible] that will be devised in Sections III-A and III-B, respectively. Note that the global minimizer of (11) [respectively, (12)] can be easily recovered from that of (14) [respectively, (23)] by devectorization.

Algorithm 2 ADMM Algorithm for Solving (11)

1: **Given** $\mathbf{Y}_h$, $\mathbf{Y}_m$, $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{A}^k$.
2: Set $j := 0$, $\mathbf{x}^0 := \mathbf{0}_{NL}$ and $\boldsymbol{v}^0 := \mathbf{0}_{NL}$.
3: **repeat**
4:    Update $\mathbf{s}^{j+1}$ by (21);
5:    Update $\mathbf{x}^{j+1}$ by (22);
6:    Update $\boldsymbol{\mu}^{j+1} := \boldsymbol{\mu}^j + \mathbf{s}^{j+1} - \mathbf{x}^{j+1}$;
7:    $j := j + 1$;
8: **until** the predefined stopping criterion is met.
9: **Output** $\mathbf{S}^{k+1}$.

A. ADMM for Solving (11)

Besides the large problem size, another bottleneck for efficiently solving (11) consists in the coupled ℓ_2- and ℓ_1-norm terms. Since ADMM blends the decomposability of the dual ascent method with the superior convergence property of the method of multipliers [39], this motivates us to solve (11) via ADMM, which allows handling ℓ_2- and ℓ_1-norm terms separately and leads to closed-form expressions for each algorithmic step.

For the given convex problem (11), we have more than one way to reformulate it into the ADMM form [39]. For

---

Algorithm 3 ADMM Algorithm for Solving (12)

1: **Given** $\mathbf{Y}_h$, $\mathbf{Y}_m$, $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{S}^{k+1}$.
2: Set $j := 0$, $\mathbf{z}^0 := \mathbf{0}_{MN}$ and $\widetilde{\boldsymbol{v}}^0 := \mathbf{0}_{MN}$.
3: **repeat**
4:    Update $\mathbf{a}^{j+1}$ by (27);
5:    Update $\mathbf{z}^{j+1}$ by (29);
6:    Update $\widetilde{\boldsymbol{v}}^{j+1} := \widetilde{\boldsymbol{v}}^j + \widetilde{\eta}(\mathbf{a}^{j+1} - \mathbf{z}^{j+1})$;
7:    $j := j + 1$;
8: **until** the predefined stopping criterion is met.
9: **Output** $\mathbf{A}^{k+1}$.

---

instance, (11) can be reformulated as

$$\min_{\mathbf{s},\mathbf{x},\mathbf{z}\in\mathbb{R}^{NL}} \frac{1}{2}\left\|\mathbf{C}_1^{(\mathbf{A}^k)}\mathbf{s} - \mathbf{y}\right\|_2^2 + \lambda_2\|\mathbf{x}\|_1 + I_+(\mathbf{z})$$
$$\text{s.t. } \mathbf{s} = \mathbf{x}, \quad \mathbf{x} = \mathbf{z}$$

as considered in [31] [see (6) and (9)], where

$$I_+(\mathbf{z}) \triangleq \begin{cases} 0, & \text{if } \mathbf{z} \succeq \mathbf{0}_{MN} \\ \infty, & \text{otherwise.} \end{cases} \tag{13}$$

The above reformulation can separately handle the ℓ_2-norm term, the ℓ_1-norm term, and the nonnegative constraint, in (11), by the primal variables $\mathbf{s}$, $\mathbf{x}$, and $\mathbf{z}$, respectively. However, it induces totally five ADMM iterates (three for primal updates and two for dual updates) [31], making the algorithm not computationally efficient.

We will show that by the KKT theory [38], we actually only need a total of three ADMM iterates (two for primal updates and one for dual update) [see (15)]. The key for this is to deal with the ℓ_1-norm term and the nonnegative constraint [in (11)] by one single primal variable [i.e., $\mathbf{x}$ in (14)]; this makes that the nondifferentiable ℓ_1-norm term becomes differentiable on the nonnegative-constrained domain (i.e., $\mathbb{R}_+^{NL}$), thus allowing the adoption of KKT theorem for simple closed-form ADMM iterate (see Appendix C). Specifically, we reformulate (11) as

$$\min_{\mathbf{s},\mathbf{x}\in\mathbb{R}^{NL}} \frac{1}{2}\|\mathbf{C}_1\mathbf{s} - \mathbf{y}\|_2^2 + \lambda_2\|\mathbf{x}\|_1$$
$$\text{s.t. } \mathbf{s} = \mathbf{x}, \quad \mathbf{x} \succeq \mathbf{0}_{NL} \tag{14}$$

where we use $\mathbf{C}_1$ to denote $\mathbf{C}_1^{(\mathbf{A}^k)}$ for notational simplicity [see (9)]. Note that the primal variable $\mathbf{x}$ simultaneously handles the ℓ_1-norm term and the nonnegative constraint, which greatly simplifies the expression of the augmented Lagrangian (involved in ADMM), thereby improving the computational efficiency.

The augmented Lagrangian of (14) is given by

$$\mathcal{L}(\mathbf{s}, \mathbf{x}, \boldsymbol{v}) = \frac{1}{2}\|\mathbf{C}_1\mathbf{s} - \mathbf{y}\|_2^2 + \lambda_2\|\mathbf{x}\|_1 + \boldsymbol{v}^T(\mathbf{s} - \mathbf{x}) + \frac{\eta}{2}\|\mathbf{s} - \mathbf{x}\|_2^2$$

where $\boldsymbol{v} \in \mathbb{R}^{NL}$ is the only dual variable associated with the equality constraint in (14), and $\eta > 0$ is the penalty parameter. Then, ADMM solves (14) by iteratively updating the two

primal variables and the dual variable as follows [47]:

$$\mathbf{s}^{j+1} \in \arg\min_{\mathbf{s}\in\mathbb{R}^{NL}} \mathcal{L}(\mathbf{s}, \mathbf{x}^j, \boldsymbol{v}^j) \tag{15a}$$

$$\mathbf{x}^{j+1} \in \arg\min_{\mathbf{x}\in\mathbb{R}_+^{NL}} \mathcal{L}(\mathbf{s}^{j+1}, \mathbf{x}, \boldsymbol{v}^j) \tag{15b}$$

$$\boldsymbol{v}^{j+1} = \boldsymbol{v}^j + \eta(\mathbf{s}^{j+1} - \mathbf{x}^{j+1}) \tag{15c}$$

where $\mathbf{x}^0$ and $\boldsymbol{v}^0$ are initialized by $\mathbf{0}_{NL}$ (or by warm start [39]). Note that (15) only involves three ADMM iterates.

By defining the scaled dual variable $\boldsymbol{\mu} \triangleq \boldsymbol{v}/\eta$ [39], (15a) and (15b) can be conveniently expressed in the following scaled form:

$$\mathbf{s}^{j+1} \in \arg\min_{\mathbf{s}\in\mathbb{R}^{NL}} \frac{1}{2}\|\mathbf{C}_1\mathbf{s} - \mathbf{y}\|_2^2 + \frac{\eta}{2}\|\mathbf{s} - \mathbf{x}^j + \boldsymbol{\mu}^j\|_2^2 \tag{16a}$$

$$\mathbf{x}^{j+1} \in \arg\min_{\mathbf{x}\in\mathbb{R}_+^{NL}} \lambda_2\|\mathbf{x}\|_1 + \frac{\eta}{2}\|\mathbf{s}^{j+1} - \mathbf{x} + \boldsymbol{\mu}^j\|_2^2 \tag{16b}$$

with (15c) being simplified as $\boldsymbol{\mu}^{j+1} = \boldsymbol{\mu}^j + \mathbf{s}^{j+1} - \mathbf{x}^{j+1}$. Note that (16b) is generally referred to as nonnegative-constrained proximal mapping, and its solution can be obtained using proximity theory [48], [49].

Next, we derive closed-form solutions for (16a) and (16b), respectively. Since (16a) is an unconstrained quadratic problem, its closed-form solution can be derived as

$$\mathbf{s}^{j+1} = (\mathbf{C}_1^T\mathbf{C}_1 + \eta\,\mathbf{I}_{NL})^{-1}(\mathbf{C}_1^T\mathbf{y} + \eta\,\mathbf{x}^j - \boldsymbol{v}^j). \tag{17}$$

However, the computational complexity of (17) can be easily verified to be $\mathcal{O}((NL)^2(ML_h + M_mL + NL)) = \mathcal{O}(L_h(NL)^2\xi)$, where $\xi \triangleq \max\{M, M_mr^2, Nr^2\}$. Since L is usually several tens of thousands, such complexity may be too high for practical application. To reduce the complexity, we practically consider an assumption that each hyperspectral pixel $\mathbf{y}_h[n]$ (i.e., the nth column of $\mathbf{Y}_h$) can be obtained by blurring (or linear combining) the r^2 HSR hyperspectral pixels (in $\mathbf{Z}$) of the same spatial region covered by the pixel $\mathbf{y}_h[n]$ (see Fig. 1). Mathematically, we assume that there exists a partition $\{\mathcal{J}_1, \ldots, \mathcal{J}_{L_h}\}$ of the index set $\mathcal{I}_L$ such that

$$\mathcal{I}_L = \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{L_h} \tag{18}$$

$$|\mathcal{J}_n| = r^2 \quad \forall n \in \mathcal{I}_{L_h} \tag{19}$$

$$\mathbf{y}_h[n] = \mathbf{Z}_{\mathcal{J}_n}\mathbf{g} + \mathbf{e}_h[n] \quad \forall n \in \mathcal{I}_{L_h} \tag{20}$$

where $\mathcal{J}_n$ collects the indices of pixels (in $\mathbf{Z}$) corresponding to the spatial region covered by $\mathbf{y}_h[n]$, $\mathbf{e}_h[n]$ denotes the nth column of $\mathbf{E}_h$ [see (2)], $\mathbf{Z}_{\mathcal{J}_n} \in \mathbb{R}^{M\times r^2}$ is the matrix collecting the r^2 columns of $\mathbf{Z}$ with indices specified by $\mathcal{J}_n$, and $\mathbf{g}$ is the blurring (or combining) coefficient vector [19], [30], [31], [50]. Some works model the blurring effect by a uniformly downsampling procedure [31, Sec. 3.2], for which $\mathbf{g}$ is a normalized all-one vector; some works model the blurring effect via a symmetric Gaussian convolution kernel [19, Sec. II.A], for which $\mathbf{g}$ is a vectorized convolution kernel (also known as Gaussian point spread function [30], [50]). In the following lemma, we derive another closed-form solution of (16a), which has much lower computational complexity than (17), for arbitrary $\mathbf{g}$:

Lemma 1: Let $\mathbf{y}_h \triangleq \mathbf{vec}(\mathbf{Y}_h)$, $\mathbf{y}_m \triangleq \mathbf{vec}(\mathbf{Y}_m)$, and $\bar{\mathbf{C}} \triangleq [(\mathbf{g}^T \otimes \mathbf{A}^k)^T, (\mathbf{I}_{r^2} \otimes \mathbf{DA}^k)^T]^T \in \mathbb{R}^{(M+r^2M_m)\times(r^2N)}$.

Then, (16a) has the following solution:

$$\begin{aligned}\mathbf{s}^{j+1} :=\ & \eta[\mathbf{I}_{L_h} \otimes (\bar{\mathbf{C}}^T\bar{\mathbf{C}} + \eta\,\mathbf{I}_{r^2N})^{-1}](\mathbf{x}^j - \boldsymbol{\mu}^j)\\ & + \{\mathbf{I}_{L_h} \otimes [(\bar{\mathbf{C}}^T\bar{\mathbf{C}} + \eta\,\mathbf{I}_{r^2N})^{-1}(\mathbf{g}^T \otimes \mathbf{A}^k)^T]\}\mathbf{y}_h\\ & + \{\mathbf{I}_{L_h} \otimes [(\bar{\mathbf{C}}^T\bar{\mathbf{C}} + \eta\,\mathbf{I}_{r^2N})^{-1}(\mathbf{I}_{r^2} \otimes \mathbf{DA}^k)^T]\}\mathbf{y}_m\end{aligned} \tag{21}$$

whose computational complexity is given by

$$\mathcal{O}((NL + (Nr^2)^2)\xi)$$

where $\xi \triangleq \max\{M, M_mr^2, Nr^2\}$.

The proof of Lemma 1 is relegated to Appendix B. On the other hand, the closed-form solution for (16b) is given by

$$\mathbf{x}^{j+1} = [\mathbf{s}^{j+1} + \boldsymbol{\mu}^j - (\lambda_2/\eta)\mathbf{1}_{NL}]_+ \tag{22}$$

which can be derived from the KKT conditions [38], detailed in Appendix C. The resulting ADMM algorithm is summarized in Algorithm 2.

B. ADMM for Solving (12)

The bottleneck for solving (12) stems from its nonnegativity constraint. Without this constraint, (12) can be solved with a closed-form solution. Inspired by this observation, we employ the decomposability of ADMM to separately handle the nonnegativity constraint in (12). To this end, we reformulate (12) in a form wherein the primal variables can be split into several blocks, with the associated objective function separable across this splitting [39], that is

$$\begin{aligned}\min_{\mathbf{a},\mathbf{z}\in\mathbb{R}^{MN}} \quad & \frac{1}{2}\|\mathbf{C}_2\,\mathbf{a} - \mathbf{y}\|_2^2 + \lambda_1\phi_1(\mathbf{A}) + I_+(\mathbf{z})\\ \text{s.t.} \quad & \mathbf{a} = \mathbf{z}\end{aligned} \tag{23}$$

where we use $\mathbf{C}_2$ to denote $\mathbf{C}_2^{(\mathbf{S}^{k+1})}$ for notational simplicity [see (10)], and $I_+(\mathbf{z})$ is the indicator function defined by (13).

Then, with $\mathbf{z}^0$ and $\widetilde{\boldsymbol{v}}^0$ (the dual variable) initialized by $\mathbf{0}_{MN}$, ADMM solves (23) via the following iterative procedure:

$$\mathbf{a}^{j+1} \in \arg\min_{\mathbf{a}\in\mathbb{R}^{MN}} \mathcal{L}(\mathbf{a}, \mathbf{z}^j, \widetilde{\boldsymbol{v}}^j) \tag{24a}$$

$$\mathbf{z}^{j+1} \in \arg\min_{\mathbf{z}\in\mathbb{R}^{MN}} \mathcal{L}(\mathbf{a}^{j+1}, \mathbf{z}, \widetilde{\boldsymbol{v}}^j) \tag{24b}$$

$$\widetilde{\boldsymbol{v}}^{j+1} = \widetilde{\boldsymbol{v}}^j + \widetilde{\eta}(\mathbf{a}^{j+1} - \mathbf{z}^{j+1}) \tag{24c}$$

where $\mathcal{L}(\mathbf{a}, \mathbf{z}, \widetilde{\boldsymbol{v}})$ is the augmented Lagrangian of problem (23) defined as

$$\begin{aligned}\mathcal{L}(\mathbf{a}, \mathbf{z}, \widetilde{\boldsymbol{v}}) = & \frac{1}{2}\|\mathbf{C}_2\mathbf{a} - \mathbf{y}\|_2^2 + \lambda_1\phi_1(\mathbf{A}) + I_+(\mathbf{z})\\ & + \widetilde{\boldsymbol{v}}^T(\mathbf{a} - \mathbf{z}) + \frac{\widetilde{\eta}}{2}\|\mathbf{a} - \mathbf{z}\|_2^2\end{aligned}$$

in which $\widetilde{\eta} > 0$ is the penalty parameter. Next, we solve (24a) and (24b).

To derive a closed-form solution for (24a), we define $\mathbf{P}_{ij} \triangleq (\mathbf{e}_i^{(N)} - \mathbf{e}_j^{(N)})^T \otimes \mathbf{I}_M$ and rewrite $\phi_1(\mathbf{A})$ as

$$\phi_1(\mathbf{A}) = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\mathbf{P}_{ij}\,\mathbf{a}\|_2^2. \tag{25}$$

By (25), one can see that (24a) is actually an unconstrained (convex) quadratic problem, whose closed-form solution can be derived as

$$\mathbf{a}^{j+1} = \left(\mathbf{C}_2^T \mathbf{C}_2 + \lambda_1 \mathbf{P}^T \mathbf{P} + \widetilde{\eta} \, \mathbf{I}_{MN}\right)^{-1} \left(\mathbf{C}_2^T \mathbf{y} + \widetilde{\eta} \, \mathbf{z}^j - \widetilde{\boldsymbol{v}}^j\right) \tag{26}$$

where $\mathbf{P} \in \mathbb{R}^{(0.5MN(N-1)) \times (MN)}$ is the matrix formed by stacking all the $\mathbf{P}_{ij} \in \mathbb{R}^{M \times (MN)}$. The computational complexity of (26) can be verified to be $\mathcal{O}(N^4 M^3 + (NM)^2 \xi')$, where $\xi' \triangleq \max\{M L_h, M_m L\}$, which is much lower than that of (17) (note that $M \ll L$ for typical scenarios). However, we still derive another closed-form solution, which is more computationally efficient than (26), in the following lemma.

Lemma 2: Assume that $N \leq M M_m$. Then, (24a) has the following solution:

$$\mathbf{a}^{j+1} = (\lambda_1 \mathbf{P}^T \mathbf{P} + \widetilde{\eta} \, \mathbf{I}_{MN} + ((\mathbf{S}^{k+1}\mathbf{B})(\mathbf{S}^{k+1}\mathbf{B})^T) \otimes \mathbf{I}_M$$
$$+ (\mathbf{S}^{k+1}(\mathbf{S}^{k+1})^T) \otimes (\mathbf{D}^T \mathbf{D}))^{-1} (\mathbf{C}_2^T \mathbf{y} + \widetilde{\eta} \, \mathbf{z}^j - \widetilde{\boldsymbol{v}}^j) \tag{27}$$

whose computational complexity is given by

$$\mathcal{O}(N^4 M^3 + NM\xi')$$

where $\xi' \triangleq \max\{M L_h, M_m L\}$.

The proof of Lemma 2 is relegated to Appendix D. Usually, N is less than ten, while M is several hundreds. Therefore, the premise of $N \leq M M_m$ is true in general. To solve (24b), we notice that it can be conveniently expressed in the following scaled form:

$$\mathbf{z}^{j+1} \in \arg\min_{\mathbf{z} \in \mathbb{R}^{MN}} I_+(\mathbf{z}) + \frac{\widetilde{\eta}}{2} \|\mathbf{a}^{j+1} - \mathbf{z} + \widetilde{\boldsymbol{\mu}}^j\|_2^2 \tag{28}$$

where $\widetilde{\boldsymbol{\mu}} \triangleq \widetilde{\boldsymbol{v}}/\widetilde{\eta}$ is the scaled dual variable [39]. Equation (28) is generally referred to as the proximity operator for the indicator function $I_+(\mathbf{z})$ [39], with a closed-form solution given by

$$\mathbf{z}^{j+1} = [\mathbf{a}^{j+1} + \widetilde{\boldsymbol{\mu}}^j]_+ = [\mathbf{a}^{j+1} + \widetilde{\boldsymbol{v}}^j/\widetilde{\eta}]_+. \tag{29}$$

The resulting algorithm is summarized in Algorithm 3.

Remark 2: Algorithm 1 will stop when the relative change of the objective values of (5) is smaller than 10^{-3}, which is a standard stopping criterion commonly used in biconvex optimization [45, Sec. 4.2.1]. On the other hand, Algorithms 2 and 3 will stop when their associated primal/dual residuals are smaller than 10^{-3}, which is also a standard stopping criterion in ADMM [39]. For Algorithm 2, the primal residual is $\|\mathbf{s}^j - \mathbf{x}^j\|_2$, and the dual residual is $\|-\eta(\mathbf{x}^j - \mathbf{x}^{j-1})\|_2$ [39]. For Algorithm 3, the primal residual is $\|\mathbf{a}^j - \mathbf{z}^j\|_2$, and the dual residual is $\|-\widetilde{\eta}(\mathbf{z}^j - \mathbf{z}^{j-1})\|_2$ [39].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Design

To evaluate the performance of the proposed CO-CNMF algorithm, we adopt the widely used Wald's protocol to design our experiments [42]. Specifically, the HSR multispectral data $\mathbf{Y}_m$ and LSR hyperspectral data $\mathbf{Y}_h$ are, respectively, obtained by downsampling and blurring an observed HSR hyperspectral
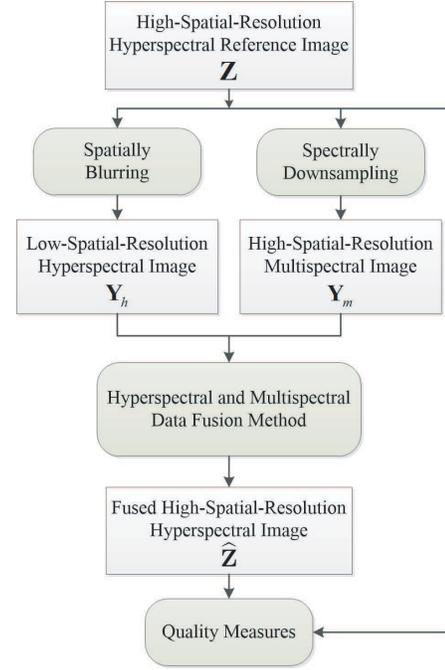
Fig. 2. Flowchart for the Wald's protocol.

reference image $\mathbf{Z}$ (see Fig. 2). Next, the proposed CO-CNMF algorithm, as well as six state-of-the-art fusion methods, including the original (unregularized) CNMF [30], BIM [19], HySure [26], VSC-CNMF [37], HSMRA [17], and CS [13], are used to fuse $\mathbf{Y}_m$ and $\mathbf{Y}_h$. The quality of the obtained fused HSR hyperspectral image (denoted as $\widehat{\mathbf{Z}}$) is then evaluated by the similarity between $\widehat{\mathbf{Z}}$ and the reference image $\mathbf{Z}$.

Some popular metrics have been proposed in the literature [5], [19], [30], [51] to quantitatively measure the similarity between $\widehat{\mathbf{Z}}$ and $\mathbf{Z}$, including peak signal-to-noise ratio (PSNR) for spatial quality measure, spectral angle mapper (SAM) for spectral quality measure, and root mean squared error (RMSE) and *erreur relative globale adimensionnelle de synthèse* (ERGAS) for global quality measures. Let $\widehat{\mathbf{Z}}^{(m)}$ and $\mathbf{Z}^{(m)}$ denote the mth rows of $\widehat{\mathbf{Z}}$ and $\mathbf{Z}$, respectively. Then, the definitions of the above quantities are given as follows.

1) The PSNR is defined as

$$\text{PSNR} = \frac{1}{M} \sum_{m=1}^{M} \text{PSNR}_m$$

where PSNR_m measures the spatial quality in the mth spectral band, $\forall m \in \mathcal{I}_M$, defined as

$$\text{PSNR}_m = 10 \log_{10} \left(\frac{\max\{\hat{z}_{mn}^2 \mid n \in \mathcal{I}_L\}}{\frac{1}{L}\|\widehat{\mathbf{Z}}^{(m)} - \mathbf{Z}^{(m)}\|_2^2} \right) \tag{30}$$

where $\hat{z}_{mn}$ denotes the nth entry in the vector $\widehat{\mathbf{Z}}^{(m)}$. The larger the value of PSNR, the higher the spatial quality of the fused image $\widehat{\mathbf{Z}}$.

2) The SAM measures the spectral distortion as

$$\text{SAM} = \frac{1}{L}\sum_{n=1}^{L}\arccos\left(\frac{(\hat{\mathbf{z}}[n])^T\mathbf{z}[n]}{\|\hat{\mathbf{z}}[n]\|_2\cdot\|\mathbf{z}[n]\|_2}\right)$$

where $\hat{\mathbf{z}}[n]$ denotes the nth column of $\widehat{\mathbf{Z}}$. The smaller the value of SAM, the better the spectral quality of $\widehat{\mathbf{Z}}$.

3) The RMSE measures the global quality by

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\text{RMSE}_m^2}$$

where

$$\text{RMSE}_m = \frac{1}{\sqrt{L}}\|\widehat{\mathbf{Z}}^{(m)} - \mathbf{Z}^{(m)}\|_2. \tag{31}$$

The smaller the value of RMSE, the higher the global quality of the fused image $\widehat{\mathbf{Z}}$.

4) Another global quality measure, ERGAS, is to quantize the relative dimensionless global error, defined as

$$\text{ERGAS} = \frac{100}{r}\sqrt{\frac{1}{M}\sum_{m=1}^{M}\frac{\text{RMSE}_m^2}{\mu_{\mathbf{Z}^{(m)}}^2}}$$

where r is the blurring factor, $\mu_{\mathbf{Z}^{(m)}} \triangleq (1/L)\mathbf{Z}^{(m)}\mathbf{1}_L$ is the mean of the row vector $\mathbf{Z}^{(m)}$, and RMSE_m has been defined in (31). The smaller the value of ERGAS, the higher the global quality of $\widehat{\mathbf{Z}}$.

Furthermore, the computational time T (in seconds) is used as the measure of computational efficiency. All the algorithms under test are implemented using Mathworks MATLAB R2013a, and the computer facility is equipped with Core-i7-4790K CPU with 4-GHz speed and 16-GB random access memory.

B. Data Generation

For the reference image $\mathbf{Z}$ to be used in Wald's protocol (see Section IV-A), we consider three benchmark data sets, acquired by three different hyperspectral sensors, respectively, as listed in the following.

1) The first data set (with a spatial resolution of 1.3 m) was acquired by the reflective optics system imaging spectrometer (ROSIS) sensor over Pavia University, Northern Italy [52]. The ROSIS sensor is composed of 115 bands, while, after removing those corrupted by water-vapor absorption, only 103 bands (between 430 and 860 nm) are used in our experiments. Then, a uniform spectral response transform matrix $\mathbf{D} \in \mathbb{R}^{4\times103}$, approximately corresponding to the Landsat TM bands 1–4 (covering 445–516, 506–595, 632–698, and 757–853 nm regions, respectively), downsamples the reference data to generate the multispectral data $\mathbf{Y}_m$ [26], [53], [54].

2) The second data set was taken over the Washington DC mall by the hyperspectral digital imagery collection experiment (HYDICE) sensor in 1995 [55], and a total of 191 spectral bands (ranging from 400 to 2500 nm) are used in our experiments. Then, a downsampling matrix $\mathbf{D} \in \mathbb{R}^{6\times191}$, corresponding to the Landsat
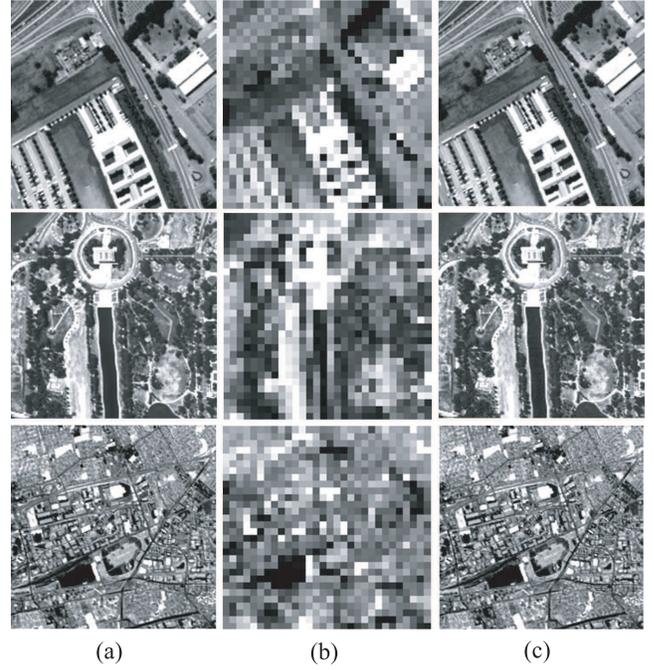
Fig. 3. Fiftieth band of (a) reference image $\mathbf{Z}$, (b) blurred LSR image $\mathbf{Y}_h$, and (c) reconstructed HSR image obtained by the proposed CO-CNMF algorithm, for the three data sets, i.e., (Top row) Pavia University, (Middle row) Washington DC, and (Bottom row) Moffett Field.

TM bands 1–5 and 7 (covering 450–520, 520–600, 630–690, 760–900, 1550–1750, and 2080–2350 nm regions, respectively), uniformly filters the HYDICE data to generate the multispectral data $\mathbf{Y}_m$ [26], [30], [54].

3) The third data set, provided by Jet Propulsion Laboratory, National Aeronautics and Space Administration, was acquired over Moffett Field, Santa Clara, CA, USA, by Airborne Visible/Infrared Imaging Spectrometer sensor in 1997 [56], with a total of 183 bands (ranging from 400 to 2500 nm) used in our experiments. Then, a spectral response transform matrix $\mathbf{D} \in \mathbb{R}^{6\times183}$, corresponding to the Landsat TM bands 1–5 and 7, uniformly downsamples the reference data to generate the multispectral data $\mathbf{Y}_m$ [26], [30], [54].

For each of the three reference images, we select an $L = 210 \times 210$ commonly studied subscene [19], [26], [30], and the associated 50th band images are displayed in the left column of Fig. 3. Then, a spatial spread transform matrix $\mathbf{B} \in \mathbb{R}^{44100\times900}$, corresponding to the Gaussian point spread function, with variance 2 and blurring factor $r = 7$ [30], [50], [57], blurs the reference data $\mathbf{Z}$ to generate the LSR hyperspectral data $\mathbf{Y}_h$ [50], [58], [59]. Finally, $\mathbf{Y}_m$ and $\mathbf{Y}_h$ are contaminated by zero-mean additive Gaussian noise with SNRs being 30 and 35 dB, respectively [19]. Each $\mathbf{Y}_h$ contains $L_h = 30 \times 30$ pixels, and the associated 50th band LSR images are displayed in the middle column of Fig. 3.

C. Model-Order Selection and Parameter Setting

When the linear mixing model (3) is violated, it can even be helpful to cover nonlinear effects with a larger

Fig. 4. Fusion performance of the CNMF algorithm (black dashed line) and the proposed CO-CNMF algorithm (blue solid line), in terms of (a) SAM and (b) RMSE, with respect to different model orders N, for the three data sets, i.e., Pavia University (○), Washington DC (□), and Moffett Field (△).

TABLE I

FUSION PERFORMANCE OF THE PROPOSED CO-CNMF ALGORITHM, IN TERMS OF SAM AND RMSE, WITH RESPECT TO DIFFERENT REGULARIZATION PARAMETERS λ_1 AND λ_2, FOR THE THREE DATA SETS

Dataset	Performance Measures	Regularizer Parameters λ_1 ($\lambda_2 = 0$)				
		10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
Pavia University	SAM	2.960	2.956	**2.954**	2.955	3.034
	RMSE	89.488	89.479	**89.403**	89.486	93.499
Washington DC	SAM	2.015	2.012	2.010	**2.002**	2.104
	RMSE	110.249	110.236	110.205	**110.033**	114.414
Moffett Field	SAM	2.209	2.204	**2.201**	2.216	2.346
	RMSE	91.438	91.426	**91.385**	92.920	105.749

Dataset	Performance Measures	Regularizer Parameters λ_2 ($\lambda_1 = 0$)				
		10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
Pavia University	SAM	2.979	2.966	**2.956**	2.960	2.965
	RMSE	89.746	89.530	**89.489**	89.495	89.510
Washington DC	SAM	2.021	2.016	**2.012**	2.013	2.014
	RMSE	110.745	110.440	**110.238**	110.245	110.252
Moffett Field	SAM	2.221	2.209	**2.205**	2.206	2.208
	RMSE	91.729	91.609	**91.429**	91.440	91.464

TABLE II

PERFORMANCE COMPARISON OF VARIOUS FUSION ALGORITHMS WITH THE DATA SET OF PAVIA UNIVERSITY

Methods	Performance Measures				Running Time
	PSNR	SAM	RMSE	ERGAS	T (seconds)
CNMF [30]	35.920	3.470	133.068	1.424	9.943
BIM [19]	36.081	3.310	130.202	1.391	151.882
HySure [26]	35.613	4.193	138.138	1.481	28.246
VSC-CNMF [37]	35.651	3.269	138.121	1.429	12.170
HSMRA [17]	32.453	3.943	191.905	2.090	**0.285**
CS [13]	37.670	3.948	115.586	1.205	0.359
CO-CNMF	**40.122**	**2.952**	**88.075**	**0.975**	15.966

TABLE III

PERFORMANCE COMPARISON OF VARIOUS FUSION ALGORITHMS WITH THE DATA SET OF WASHINGTON DC

Methods	Performance Measures				Running Time
	PSNR	SAM	RMSE	ERGAS	T (seconds)
CNMF [30]	32.163	2.408	147.209	1.524	13.097
BIM [19]	32.365	2.408	155.343	1.281	151.839
HySure [26]	29.925	3.637	199.745	1.724	29.276
VSC-CNMF [37]	31.933	2.425	160.795	1.458	17.860
HSMRA [17]	29.257	3.197	241.029	1.775	**0.527**
CS [13]	33.348	2.895	154.212	**1.155**	0.678
CO-CNMF	**33.634**	**1.979**	**108.760**	1.328	24.335

TABLE IV

PERFORMANCE COMPARISON OF VARIOUS FUSION ALGORITHMS WITH THE DATA SET OF MOFFETT FIELD

Methods	Performance Measures				Running Time
	PSNR	SAM	RMSE	ERGAS	T (seconds)
CNMF [30]	35.327	2.428	158.697	1.224	13.506
BIM [19]	37.347	2.179	124.893	0.971	151.703
HySure [26]	37.009	3.132	144.507	1.075	29.682
VSC-CNMF [37]	34.480	2.366	171.795	1.373	18.743
HSMRA [17]	33.920	2.748	183.352	1.453	**0.498**
CS [13]	36.022	3.422	157.959	1.166	0.650
CO-CNMF	**41.343**	**2.159**	**87.398**	**0.678**	26.965

model-order N, as reported in [60]. In the CNMF framework [30], the additional endmembers in $\mathbf{A}$ can then be explained as containing both shading and material endmembers. Note that the nonlinear effect can also be interpreted by the variability of endmembers [40, Sec. V.B]. Because setting a larger N allows capturing the effect by associating a single material with multiple endmembers, this strategy can hence deal with spectrally varying scenes, as reported in the original CNMF work [30]. Another possible strategy is to adopt nonlinear modeling for $\mathbf{Z}$ [61], but it may induce a cumbersome fusion algorithm as considering multiple reflections of solar radiation can lead to far more complicated (nonconvex) formulations. To investigate a proper setting of N, we test the fusion performance of the proposed CO-CNMF algorithm (with $\lambda_1 = \lambda_2 = 0.001$ and $\eta = \widetilde{\eta} = 1$) with respect to different model orders. The results are demonstrated in Fig. 4, where one can see that setting a model-order of $N > 10$ does not help in improving the data fusion performances (in terms of both SAM and RMSE) for all the three data sets. Therefore, we set $N = 10$ for the CO-CNMF algorithm in all the subsequent experiments, though the orders estimated by the benchmark model selection algorithms [62], [63] are all no larger than 8 for all the three data sets. As assumed in many state-of-the-art fusion methods, we will once again confirm that the linear mixing model is sufficient for yielding promising fusion performance (see Section IV-D).

On the other hand, from Fig. 4, we see that for all the three data sets, the proposed CO-CNMF algorithm outperforms the original (unregularized) CNMF algorithm (in terms of both

SAM and RMSE) for all the tested model orders. This shows that the ill-posed CNMF problem indeed requires regularization, and that the adopted SSD regularizer $\phi_1(\mathbf{A})$ and the sparsity-promoting regularizer $\phi_2(\mathbf{S})$ are suitable for regularizing the CNMF solution. To understand which regularizer is

TABLE V
SUMMARY OF COMPLEXITIES (AND THEIR OMs ON REAL DATA) AND PRT, IN SECONDS, OF THE NAIVE UPDATES (17) [RESPECTIVELY, (26)] AND ITS IMPROVEMENT (21) [RESPECTIVELY, (27)], WHERE $\xi \triangleq \max\{M, M_m r^2, N r^2\}$ AND $\xi' \triangleq \max\{ML_h, M_m L\}$

Variable	Equation	Complexity	Pavia University		Washington DC		Moffett Field	
			OM	PRT	OM	PRT	OM	PRT
$\mathbf{s}^{j+1}$	(17)	$\mathcal{O}\left(L_h(NL)^2\xi\right)$	$\mathcal{O}(8.577 \times 10^{16})$	32.59	$\mathcal{O}(8.577 \times 10^{16})$	54.51	$\mathcal{O}(8.577 \times 10^{16})$	61.94
	(21)	$\mathcal{O}\left((NL + (Nr^2)^2)\xi\right)$	$\mathcal{O}(3.337 \times 10^{8})$	0.089	$\mathcal{O}(3.337 \times 10^{8})$	0.130	$\mathcal{O}(3.337 \times 10^{8})$	0.116
$\mathbf{a}^{j+1}$	(26)	$\mathcal{O}\left(N^4 M^3 + (NM)^2\xi'\right)$	$\mathcal{O}(1.981 \times 10^{11})$	12.10	$\mathcal{O}(1.035 \times 10^{12})$	22.09	$\mathcal{O}(9.474 \times 10^{11})$	21.14
	(27)	$\mathcal{O}\left(N^4 M^3 + NM\xi'\right)$	$\mathcal{O}(1.111 \times 10^{10})$	0.049	$\mathcal{O}(7.018 \times 10^{10})$	0.302	$\mathcal{O}(6.177 \times 10^{10})$	0.229

more effective, we disable ϕ_1 (respectively, ϕ_2) and identify an optimal setting for λ_2 (respectively, λ_1). The results are demonstrated in Table I, where the setting of 0.001 is optimal (or nearly optimal) for both λ_1 or λ_2 (when one is disabled). Furthermore, the SSD regularizer seems slightly more effective than the ℓ_1-norm regularizer for all the three data sets, because the setting $(\lambda_1, \lambda_2) = (0.001, 0)$ always yields slightly smaller SAM/RMSE than the setting $(\lambda_1, \lambda_2) = (0, 0.001)$. Note that enabling both regularizers, i.e., $\lambda_1 = \lambda_2 = 0.001$, further upgrades the fusion performance (see Tables II–IV). In the ensuing experiments, the regularization parameters will be set as $\lambda_1 = \lambda_2 = 0.001$; the parameters can be tuned if users have further prior knowledge on their data sets; for instance, if the materials in a scene of interest are known to be highly sparse, the users may increase λ_2 for the sparsity-promoting ℓ_1-norm regularizer ϕ_2. Moreover, the setting of penalty parameters can affect the convergence speed [39]; we empirically found that the standard setting $\eta = \tilde{\eta} = 1$ [39, Ch. 11] has good convergence speed, so it will be adopted in the ensuing experiments. Although the above setting may not be optimal, it generally yields good fusion performance (see Fig. 4). The reconstructed HSR images, obtained by CO-CNMF under the above parameter setting, for all the three selected subscenes, are also displayed in the right column of Fig. 3, where one can see that the reconstructed images hold high resemblance to their respective reference images.

D. Performance Comparison and Discussion

To quantitatively compare the performances of the proposed CO-CNMF algorithm with the state-of-the-art algorithms, including CNMF [30], BIM [19], HySure [26], VSC-CNMF [37], HSMRA [17], and CS [13], we evaluate their fusion performances in terms of PSNR, SAM, RMSE, and ERGAS, and the results are displayed in Tables II (Pavia University), III (Washington DC), and IV (Moffett Field). The boldface numbers in these tables indicate the best performance (i.e., the largest PSNR, or the smallest SAM/RMSE/ERGAS) of all the image fusion algorithms under test.

For all the three data sets, one can observe that the proposed CO-CNMF algorithm has the highest PSNR value (the best spatial quality), and the smallest SAM value (the best spectral shape preservation capability). Moreover, CO-CNMF also yields the best global fusion quality for all the three data sets, as indicated by the smallest RMSE/ERGAS values (except for the ERGAS value on the Washington DC data set, in terms of which CS performs the best). Note that VSC-CNMF does not show better performance than the (unregularized) CNMF;

this indicates that without considering both data fitting errors (for $\mathbf{Y}_h$ and $\mathbf{Y}_m$), regularization itself may not be very effective to yield better fusion results [37].

As for computational efficiency, HSMRA performs the best, while BIM takes the longest running time. For all the three data sets, under the stopping criteria specified in Remark 2, Algorithm 1 terminates within 80 outer iterations, while, for each outer iteration k, Algorithm 2 (respectively, Algorithm 3) terminates within 50 (respectively, 20) inner iterations. Let us emphasize that the amounts of computational time of CO-CNMF for all the three data sets are more than 2000 s (around two orders of magnitude (OMs) longer than those shown in Tables II–IV), if we simply use the naive closed-form expressions (17) (for updating $\mathbf{s}^{j+1}$) and (26) (for updating $\mathbf{a}^{j+1}$) in the ADMM iterates. To understand the effectiveness of (21) (derived in Lemma 1 for updating $\mathbf{s}^{j+1}$) and (27) (derived in Lemma 2 for updating $\mathbf{a}^{j+1}$), we summarize the associated computational complexities (and their OMs for the three data sets), as well as the per-iteration running time (PRT), in Table V. One can see that (21) [respectively, (27)] improves the computational efficiency of (17) [respectively, (26)] by several orders when updating $\mathbf{s}^{j+1}$ (respectively, $\mathbf{a}^{j+1}$). In fact, the *averaged* PRT (averaged over inner iterations j) of (21) and (27) are much less than that shown in Table V, because some quantities (independent of j) only need to be computed once. For example, for updating $\mathbf{s}^{j+1}$ using (21), both $\mathbf{V}_1 \triangleq \eta[\mathbf{I}_{L_h} \otimes (\bar{\mathbf{C}}^T\bar{\mathbf{C}} + \eta\ \mathbf{I}_{r^2 N})^{-1}]$ and $\mathbf{v}_2 \triangleq \{\mathbf{I}_{L_h} \otimes [(\bar{\mathbf{C}}^T\bar{\mathbf{C}} + \eta\mathbf{I}_{r^2 N})^{-1}(\mathbf{g}^T \otimes \mathbf{A}^k)^T]\}\mathbf{y}_h + \{\mathbf{I}_{L_h} \otimes [(\bar{\mathbf{C}}^T\bar{\mathbf{C}} + \eta\mathbf{I}_{r^2 N})^{-1}(\mathbf{I}_{r^2} \otimes \mathbf{D}\mathbf{A}^k)^T]\}\mathbf{y}_m$ only need to be computed once (during the first iteration $j = 1$), because they are independent of j; in the subsequent iterations $j > 1$, we only need to compute the multiplication $\mathbf{v}_3^j \triangleq \mathbf{V}_1(\mathbf{x}^j - \boldsymbol{\mu}^j)$ and the addition $\mathbf{s}^{j+1} := \mathbf{v}_3^j + \mathbf{v}_2$. This implementation trick greatly reduces the computational time of CO-CNMF; using Pavia University data set for example, after the second iteration, (21) only takes 0.040 (instead of 0.089 s), while (27) only takes 0.005 s (instead of 0.049 s), greatly showing the effectiveness of Lemmas 1 and 2.

Furthermore, to compare the fusion performances with respect to different spectral bands, we also display the PSNR$_m$ [see (30)] and RMSE$_m$ [see (31)] curves of these fusion algorithms, for all the three data sets, in Fig. 5. For the Washington DC data set, VSC-CNMF (respectively, CS) outperforms the other algorithms in the Landsat TM band 5 (respectively, 7), while CO-CNMF has the best fusion performance in the other band regions. For Pavia University and Moffett Field data sets, the proposed CO-CNMF algorithm significantly outperforms
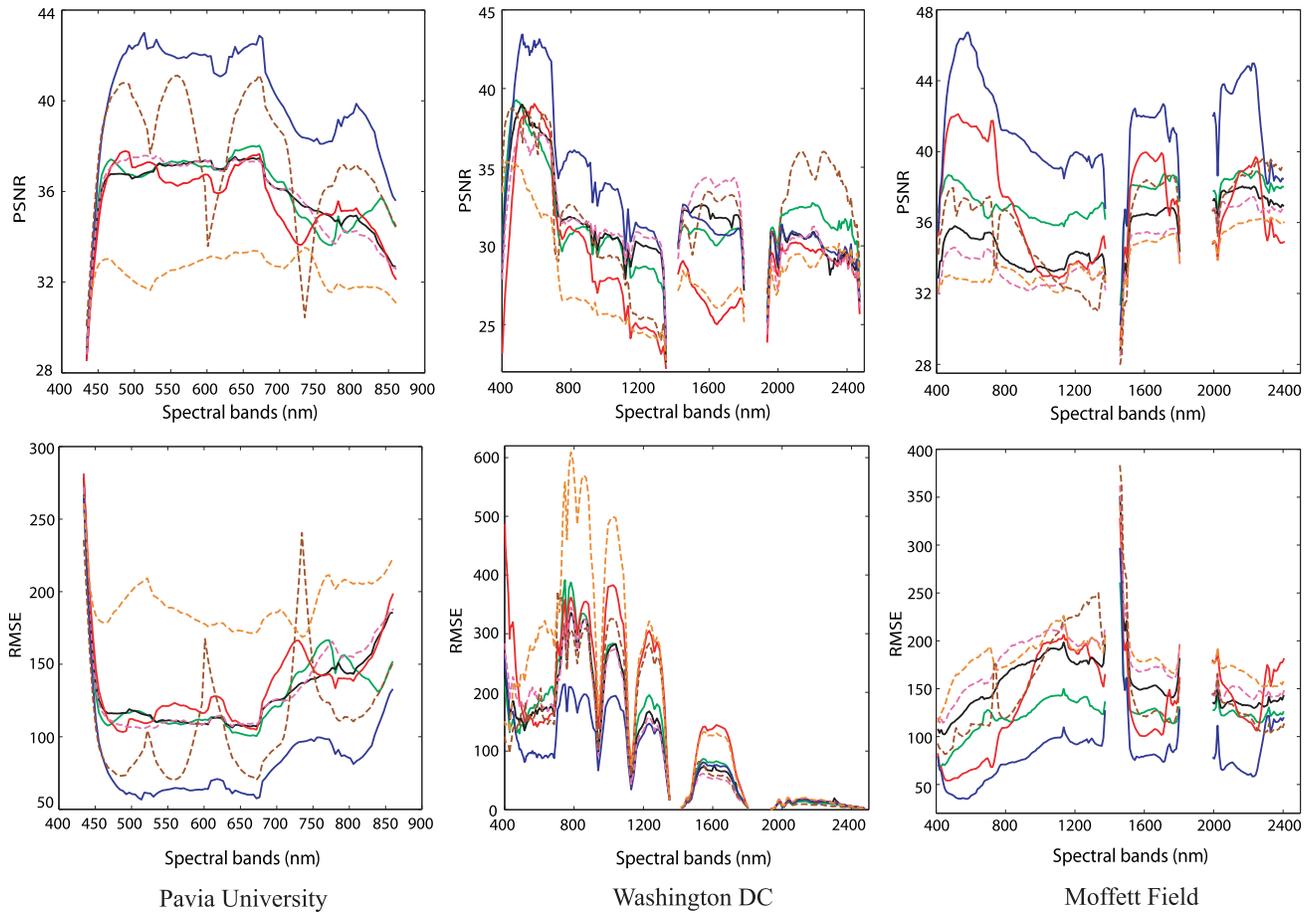
Fig. 5. PSNR_m and RMSE_m curves (versus spectral band m) of the proposed CO-CNMF algorithm (blue line) and six state-of-the-art fusion algorithms, including CNMF (black line), BIM (green line), HySure (red line), VSC-CNMF (pink dashed line), HSMRA (orange dashed line), and CS (brown dashed line), for the three data sets, i.e., (left column) Pavia University, (middle column) Washington DC, and (right column) Moffett Field.

the other six state-of-the-art algorithms for all the band regions. These experimental results well support the validity of the regularized CNMF criterion (5), as well as the efficacy of the carefully designed convex optimization method.

E. Performance Evaluation for Imperfect Coregistration Case

Image coregistration is to transform different data sets (acquired from different sensors) into the same coordinate system [64]. Although perfect coregistration (for $\mathbf{Y}_m$ and $\mathbf{Y}_h$) is assumed in many state-of-the-art fusion methods, the impact of this hypothesis violation is seldom studied in the literature. In this section, we study a typical misregistration case that the spatial regions covered by $\mathbf{Y}_m$ and $\mathbf{Y}_h$ are not perfectly aligned [64]. Specifically, the experimental design is still based on Wald's protocol, but we introduced another reference image $\mathbf{Z}_\delta$ (for generating $\mathbf{Y}_m$), where the coordinate system of $\mathbf{Z}_\delta$ is deviated from the original reference image $\mathbf{Z}$ by δ pixels (both horizontally and vertically); δ is an integer parameter and $\delta = 0$ corresponds to the perfect coregistration case. A graphical illustration is given in Fig. 6. Then, $\mathbf{Y}_m$ (respectively, $\mathbf{Y}_h$) is obtained as a spectrally downsampled version of $\mathbf{Z}_\delta$ (respectively, a spatially blurred version of $\mathbf{Z}$), and $\widehat{\mathbf{Z}}$ is obtained by fusing the imperfectly coregistered data (see Fig. 2).
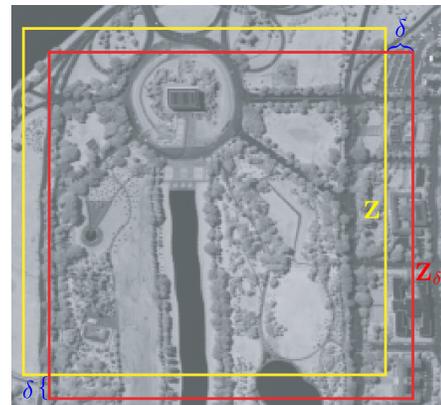
Fig. 6.   Spatial regions covered by the two reference images $\mathbf{Z}$ (for generating $\mathbf{Y}_h$) and $\mathbf{Z}_\delta$ (for generating $\mathbf{Y}_m$) are deviated by δ pixels (both horizontally and vertically), so the coordinate systems of the resultant $\mathbf{Y}_h$ and $\mathbf{Y}_m$ are not perfectly coregistered.

The experimental results for Washington DC data set are given in Fig. 7, where BIM is less susceptible to misregistration, because the overcomplete dictionary is learned from overlapping patches of a rough estimate of (dimension-reduced) $\mathbf{Z}$ [19]; note that the patch-overlapping strategy can improve the robustness against misregistration (besides reducing block
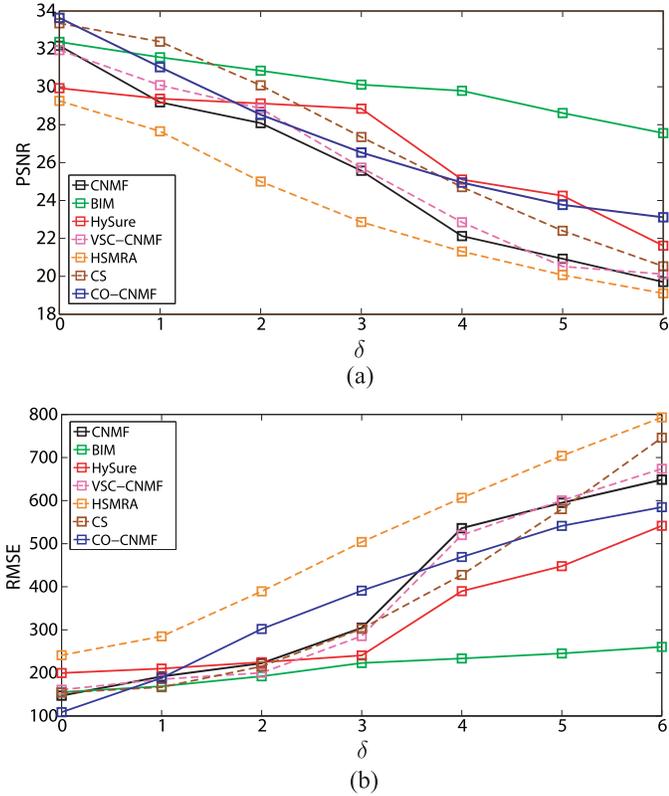
Fig. 7. (a) PSNR and (b) RMSE curves (versus δ) of various fusion methods for the Washington DC data set.

artifacts) [65]. However, the fusion performances of all the methods under test, including BIM, degrade when the two coordinate systems (with respect to $\mathbf{Y}_m$ and $\mathbf{Y}_h$) get more deviated (or when δ becomes larger). The above results indicate that the coregistration issue (though out of the scope of this paper) is important for yielding better fusion performance.

V. CONCLUSION

Adopting the CNMF criterion with the ℓ_1-norm and SSD regularizers incorporated, we have presented a convex formulation for the hyperspectral and multispectral data fusion problem, which was then solved by the proposed ADMM-based CO-CNMF algorithm. In CO-CNMF, all the ADMM iterates are updated using closed-form expressions. These expressions were derived from convex optimization theories and have been further refined by carefully observing some matrix structures embedded in their naive counterparts, thereby yielding significant computational complexity reduction (see Lemmas 1 and 2) for practical applications. For performance evaluation, we have also provided extensive experimental results based on Wald's protocol, using three benchmark data sets acquired by three different hyperspectral sensors. These results have demonstrated the superior fusion performance of the proposed CO-CNMF algorithm over state-of-the-art fusion methods.

APPENDIX

A. Proof of Proposition 1

We begin with observing that in (5), the constraint sets for both $\mathbf{A}$ and $\mathbf{S}$ are closed sets, because they are

intersections of finite closed half-spaces [38]. Moreover, as it can be easily verified that CNMF($\mathbf{A}, \mathbf{S}$), $\phi_1(\mathbf{A})$, and $\phi_2(\mathbf{S})$ are all continuous functions [66], their linear combination [i.e., the objective function of (5)] is also continuous [66]. The above observations, together with [45, Th. 4.7], imply that the alternating convex minimization adopted in Algorithm 1 (see [45, Algorithm 4.1]) must generate a sequence $\{(\mathbf{A}^k, \mathbf{S}^k)\}$ that converges to a partial optimum (see [45, Definition 4.1] for the definition of partial optimum). On the other hand, because the ℓ_1-norm regularizer can be simplified as a differentiable function (i.e., $\phi_2(\mathbf{S}) \triangleq \|\mathbf{S}\|_1 = \mathbf{1}_N^T \mathbf{S} \mathbf{1}_L$) when $\mathbf{S} \succeq \mathbf{0}_{N \times L}$, the objective function of (5) is hence differentiable on the problem domain [38]. Finally, by [45, Corollary 4.3], a partial optimum must also be a stationary point, if the objective function of (5) is differentiable and biconvex. Therefore, the proof of Proposition 1 has been completed. ∎

B. Proof of Lemma 1

By (18) and (19) and $r^2 = L/L_h$, one can see that

$$\mathcal{J}_{n_1} \cap \mathcal{J}_{n_2} = \emptyset \text{ for any } n_1 \neq n_2. \tag{32}$$

From (20) and (32), there exists a permutation matrix $\Pi \in \mathbb{R}^{L \times L}$ such that $\mathbf{Z}\Pi = [\mathbf{Z}_{\mathcal{J}_1}, \ldots, \mathbf{Z}_{\mathcal{J}_{L_h}}]$ and

$$\mathbf{Y}_h = (\mathbf{Z}\Pi)(\mathbf{I}_{L_h} \otimes \mathbf{g}) + \mathbf{E}_h. \tag{33}$$

Comparing (2) and (33), we can take the spatial spread transform matrix as

$$\mathbf{B} = \Pi (\mathbf{I}_{L_h} \otimes \mathbf{g}) \in \mathbb{R}^{L \times L_h}. \tag{34}$$

Moreover, to simplify the proof, we can assume without loss of generality that the columns of $\mathbf{Z}$ have already been partitioned (i.e., $\mathbf{Z} = [\mathbf{Z}_{\mathcal{J}_1}, \ldots, \mathbf{Z}_{\mathcal{J}_{L_h}}]$), implying that the permutation matrix $\Pi = \mathbf{I}_L$, which, together with (34) and the fact of $L = r^2 L_h$, yields

$$\mathbf{B}^T \otimes \mathbf{A}^k = \mathbf{I}_{L_h} \otimes (\mathbf{g}^T \otimes \mathbf{A}^k) \tag{35}$$

$$\mathbf{I}_L \otimes (\mathbf{D}\mathbf{A}^k) = \mathbf{I}_{L_h} \otimes (\mathbf{I}_{r^2} \otimes \mathbf{D}\mathbf{A}^k). \tag{36}$$

From (9), (35), and (36), we observe that

$$\mathbf{C}_1^T \mathbf{C}_1 = \mathbf{I}_{L_h} \otimes (\bar{\mathbf{C}}^T \bar{\mathbf{C}}) \tag{37}$$

where $\mathbf{C}_1$ denotes $\mathbf{C}_1^{(\mathbf{A}^k)}$ for notational simplicity [see (9)]. Then, by (17) and (37), the solution of (16a) is given by

$$\mathbf{s}^{j+1} = (\mathbf{C}_1^T \mathbf{C}_1 + \eta \mathbf{I}_{NL})^{-1} \mathbf{v} \tag{38}$$

$$= [\mathbf{I}_{L_h} \otimes (\bar{\mathbf{C}}^T \bar{\mathbf{C}} + \eta \mathbf{I}_{r^2 N})^{-1}] \mathbf{v} \tag{39}$$

where the vector $\mathbf{v} \triangleq \mathbf{C}_1^T \mathbf{y} + \eta \mathbf{x}^j - \eta \boldsymbol{\mu}^j$. Note that the inversion formula in (39) is much computationally economical than (38), because we only need to compute the inverse of a $(Nr^2) \times (Nr^2)$ matrix in (39) [instead of an $(NL) \times (NL)$ matrix in (38)].

However, the computation of $\mathbf{v} = \mathbf{C}_1^T \mathbf{y} + \eta \mathbf{x}^j - \eta \boldsymbol{\mu}^j$ is still time-consuming, due to multiplications involved in $\mathbf{C}_1^T \mathbf{y}$. Directly simplifying the expression of $\mathbf{C}_1^T \mathbf{y}$ (so as to reduce the complexity) does not seem feasible. To solve this bottleneck, we come up with directly dealing with the multiplication

chain, i.e., $(\mathbf{C}_1^T \mathbf{C}_1 + \eta \mathbf{I}_{NL})^{-1} \mathbf{C}_1^T \mathbf{y}$ (instead of solely dealing with $\mathbf{C}_1^T \mathbf{y}$). Specifically, we have

$$
\begin{aligned}
&\left(\mathbf{C}_1^T \mathbf{C}_1 + \eta \mathbf{I}_{NL}\right)^{-1} \mathbf{C}_1^T \mathbf{y} \\
&= \left(\mathbf{C}_1^T \mathbf{C}_1 + \eta \mathbf{I}_{NL}\right)^{-1} \left[\mathbf{I}_{L_h} \otimes (\mathbf{g}^T \otimes \mathbf{A}^k)^T\right] \mathbf{y}_h \\
&\quad + \left(\mathbf{C}_1^T \mathbf{C}_1 + \eta \mathbf{I}_{NL}\right)^{-1} [\mathbf{I}_{L_h} \otimes (\mathbf{I}_{r^2} \otimes \mathbf{DA}^k)^T] \mathbf{y}_m \\
&= \{\mathbf{I}_{L_h} \otimes [(\bar{\mathbf{C}}^T \bar{\mathbf{C}} + \eta \mathbf{I}_{r^2 N})^{-1} (\mathbf{g}^T \otimes \mathbf{A}^k)^T]\} \mathbf{y}_h \\
&\quad + \{\mathbf{I}_{L_h} \otimes [(\bar{\mathbf{C}}^T \bar{\mathbf{C}} + \eta \mathbf{I}_{r^2 N})^{-1} (\mathbf{I}_{r^2} \otimes \mathbf{DA}^k)^T]\} \mathbf{y}_m \quad (40)
\end{aligned}
$$

where the first equality is due to (35) and (36); the second equality is due to (37) and some simple matrix manipulations. Note that, in (40), we only need to compute $[(\bar{\mathbf{C}}^T \bar{\mathbf{C}} + \eta \mathbf{I}_{r^2 N})^{-1} (\mathbf{g}^T \otimes \mathbf{A}^k)^T]$ and $[(\bar{\mathbf{C}}^T \bar{\mathbf{C}} + \eta \mathbf{I}_{r^2 N})^{-1} (\mathbf{I}_{r^2} \otimes \mathbf{DA}^k)^T]$ once, and then blockwisely copy them L_h times, followed by the matrix-vector multiplications. Combining (39) and (40) yields (21).

The computational complexity of the first term in (21) is dominated by the matrix–matrix multiplication for $\bar{\mathbf{C}}^T \bar{\mathbf{C}}$ [requiring $\mathcal{O}((Nr^2)^2(M + r^2 M_m))$], the inverse of an $(Nr^2) \times (Nr^2)$ matrix [requiring $\mathcal{O}((Nr^2)^3)$], and L_h multiplications of an $(Nr^2) \times (Nr^2)$ matrix and an (Nr^2)-dimensional vector [requiring $\mathcal{O}(L_h(Nr^2)^2)$].

The computational complexity of the second term in (21) is dominated by the matrix–matrix multiplication for $\bar{\mathbf{C}}^T \bar{\mathbf{C}}$ [requiring $\mathcal{O}((Nr^2)^2(M + r^2 M_m))$], the inverse of an $(Nr^2) \times (Nr^2)$ matrix [requiring $\mathcal{O}((Nr^2)^3)$], the multiplication of an $(Nr^2) \times (Nr^2)$ matrix and an $(Nr^2) \times M$ matrix [requiring $\mathcal{O}((Nr^2)^2 M)$], and L_h multiplications of an $(Nr^2) \times M$ matrix and an M-dimensional vector [requiring $\mathcal{O}(L_h Nr^2 M)$].

The computational complexity of the third term in (21) is dominated by the matrix–matrix multiplication for $\bar{\mathbf{C}}^T \bar{\mathbf{C}}$ [requiring $\mathcal{O}((Nr^2)^2(M + r^2 M_m))$], the inverse of an $(Nr^2) \times (Nr^2)$ matrix [requiring $\mathcal{O}((Nr^2)^3)$], the multiplication of an $(Nr^2) \times (Nr^2)$ matrix and an $(Nr^2) \times (M_m r^2)$ matrix [requiring $\mathcal{O}((Nr^2)^2 M_m r^2)$], and L_h multiplications of an $(Nr^2) \times (M_m r^2)$ matrix and an $(M_m r^2)$-dimensional vector [requiring $\mathcal{O}(L_h Nr^2 M_m r^2)$].

All in all, the computational complexity of (21) is $\mathcal{O}(Nr^2(Nr^2 + L_h)\xi) = \mathcal{O}((NL + (Nr^2)^2)\xi)$, where $\xi \triangleq \max\{M, M_m r^2, Nr^2\}$. ∎

C. Derivation of (22)

KKT conditions are not directly applicable in solving (16b), because the ℓ_1-norm term is not differentiable [38]. This issue can be solved by simultaneously handling the ℓ_1-norm term and the nonnegative constraint, using one single primal variable $\mathbf{x}$, in the ADMM formulation (14).

To be precise, by the fact of $\|\mathbf{x}\|_1 = \mathbf{1}_{NL}^T \mathbf{x}$, $\forall \mathbf{x} \in \mathbb{R}_+^{NL}$, we reformulate (16b) into a differentiable form

$$
\mathbf{x}^{j+1} \in \arg \min_{\mathbf{x} \in \mathbb{R}_+^{NL}} \lambda_2 \mathbf{1}_{NL}^T \mathbf{x} + \frac{\eta}{2} \|\mathbf{s}^{j+1} - \mathbf{x} + \mu^j\|_2^2 \quad (41)
$$

whose closed-form solution can then be obtained by solving the associated KKT conditions [38], that is

$$
\begin{aligned}
\mathbf{x}^\star &\succeq \mathbf{0}_{NL} \quad &(42a) \\
\widetilde{\boldsymbol{\lambda}}^\star &\succeq \mathbf{0}_{NL} \quad &(42b) \\
\widetilde{\lambda}_i^\star x_i^\star &= 0 \quad \forall i \quad &(42c) \\
\lambda_2 \mathbf{1}_{NL} + \frac{\eta}{2}(2\mathbf{x}^\star - 2(\mathbf{s}^{j+1} + \mu^j)) - \widetilde{\boldsymbol{\lambda}}^\star &= \mathbf{0}_{NL} \quad &(42d)
\end{aligned}
$$

where $\widetilde{\boldsymbol{\lambda}}$ is the vector consisting of all the Lagrange multipliers for the (componentwise) inequality constraints in (41); x_i and $\widetilde{\lambda}_i$, respectively, denote the ith entry of $\mathbf{x}$ and $\widetilde{\boldsymbol{\lambda}}$; the superscript "$\star$" denotes primal and dual optimal solutions of (41). From (42b) and (42d), it can be seen that the primal optimal solution must satisfy

$$
\mathbf{x}^\star \succeq \mathbf{s}^{j+1} + \mu^j - (\lambda_2/\eta)\mathbf{1}_{NL}
$$

which, together with (42a) and the complementary slackness (42c), yields the closed-form solution

$$
\mathbf{x}^{j+1} = \mathbf{x}^\star = [\mathbf{s}^{j+1} + \mu^j - (\lambda_2/\eta)\mathbf{1}_{NL}]_+
$$

where the operator $[\cdot]_+$ is applied componentwise. ∎

D. Proof of Lemma 2

We recall that $\mathbf{C}_2$ denotes $\mathbf{C}_2^{(\mathbf{S}^{k+1})}$ for notational simplicity [see (10)]. The computational complexity of (26) mainly comes from the computation of $\mathbf{C}_2^T \mathbf{C}_2$; note that the computation of $\mathbf{P}^T \mathbf{P}$ is very fast due to the high sparsity of $\mathbf{P}$. Hence, our primary objective is to reduce the complexity in computing $\mathbf{C}_2^T \mathbf{C}_2$. Note that $\mathbf{C}_2^T \mathbf{C}_2$ does not have the elegant diagonal structure as $\mathbf{C}_1^T \mathbf{C}_1$ does [see (37)]. Therefore, we need other strategies to reduce the complexity to be presented next.

We begin by noticing from (10) that

$$
\begin{aligned}
\mathbf{C}_2^T \mathbf{C}_2 = &((\mathbf{S}^{k+1}\mathbf{B})^T \otimes \mathbf{I}_M)^T ((\mathbf{S}^{k+1}\mathbf{B})^T \otimes \mathbf{I}_M) \\
&+ ((\mathbf{S}^{k+1})^T \otimes \mathbf{D})^T ((\mathbf{S}^{k+1})^T \otimes \mathbf{D}). \quad (43)
\end{aligned}
$$

We handle the first term and the second term in (43) by their respective structures in the following. For ease of the ensuing presentation, for a given matrix $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_c]$, we use $[\mathbf{M}]_{a:b}$ ($1 \le a \le b \le c$) to denote the matrix $[\mathbf{m}_a, \ldots, \mathbf{m}_b]$.

First, since $\mathbf{S}^{k+1} \in \mathbb{R}^{N \times L}$ and $\mathbf{B} \in \mathbb{R}^{L \times L_h}$, if we directly compute $(\mathbf{S}^{k+1}\mathbf{B})^T$ in the first term in (43), the complexity is $\mathcal{O}(NL_h L)$. However, by (34), we actually only need to compute

$$
\begin{aligned}
\mathbf{v}' = \mathbf{g}^T &\left[[\mathbf{S}^{k+1}]_{1:r^2}^T, [\mathbf{S}^{k+1}]_{r^2+1:2r^2}^T, \ldots, [\mathbf{S}^{k+1}]_{(L_h-1)r^2+1:L}^T\right] \\
&\in \mathbb{R}^{1 \times NL_h} \quad (44)
\end{aligned}
$$

and then reshape this vector to obtain the matrix $(\mathbf{S}^{k+1}\mathbf{B})^T \in \mathbb{R}^{L_h \times N}$ (whose first row is given by $[\mathbf{v}']_{1:N}$, and the second row is formed by $[\mathbf{v}']_{N+1:2N}$, and so on), resulting in a total complexity of $\mathcal{O}(NL_h r^2) = \mathcal{O}(NL) \ll \mathcal{O}(NL_h L)$. Then, the above strategy for obtaining $(\mathbf{S}^{k+1}\mathbf{B})^T$, together with the observation that

$$
\begin{aligned}
((\mathbf{S}^{k+1}\mathbf{B})^T \otimes \mathbf{I}_M)^T ((\mathbf{S}^{k+1}\mathbf{B})^T \otimes \mathbf{I}_M) \\
= ((\mathbf{S}^{k+1}\mathbf{B})(\mathbf{S}^{k+1}\mathbf{B})^T) \otimes \mathbf{I}_M \quad (45)
\end{aligned}
$$

makes the complexity of the first term in (43) only

$$\mathcal{O}(NL + N^2 L_h). \tag{46}$$

To handle the second term in (43), we employ two properties of Kronecker product, i.e., $(\mathbf{M}_1 \otimes \mathbf{M}_2)^T = \mathbf{M}_1^T \otimes \mathbf{M}_2^T$ and $(\mathbf{M}_1 \otimes \mathbf{M}_2)(\mathbf{M}_3 \otimes \mathbf{M}_4) = \mathbf{M}_1\mathbf{M}_3 \otimes \mathbf{M}_2\mathbf{M}_4$ (for arbitrary matrices $\mathbf{M}_i$ of proper dimension), which lead to

$$((\mathbf{S}^{k+1})^T \otimes \mathbf{D})^T((\mathbf{S}^{k+1})^T \otimes \mathbf{D}) = [\mathbf{S}^{k+1}(\mathbf{S}^{k+1})^T] \otimes (\mathbf{D}^T\mathbf{D}). \tag{47}$$

Note that the computation of the left-hand side of (47) costs $\mathcal{O}(N^2 M^2 L M_m)$, while the computation of the right-hand side only costs

$$\mathcal{O}(N^2 L + M^2 M_m + N^2 M^2). \tag{48}$$

Combining (26), (43), (45), and (47) yields (27). Besides (46) and (48), the remaining complexity for computing (27) is dominated by the matrix–matrix multiplication for $\mathbf{P}^T\mathbf{P}$ [requiring $\mathcal{O}((MN)^2 MN(N-1))$], the inverse of an $(MN) \times (MN)$ matrix [requiring $\mathcal{O}((MN)^3)$], the multiplication of an $(MN) \times (ML_h + M_m L)$ matrix and an $(ML_h + M_m L)$-dimensional vector [requiring $\mathcal{O}((MN)(ML_h + M_m L))$], and the multiplication of an $(MN) \times (MN)$ matrix and an (MN)-dimensional vector [requiring $\mathcal{O}((MN)^2)$]. All in all, the complexity of (27) is given by (see $M_m < M$)

$$\mathcal{O}(N^2 L + M^3 N^4 + M^2 N L_h + MNM_m L)$$

which reduces to $\mathcal{O}(N^4 M^3 + (NM)\xi')$ under the premise of $N \le MM_m$. Hence, the proof of Lemma 2 is completed. ∎

ACKNOWLEDGMENT

REFERENCES

[1] J. M. Bioucas Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[3] S. Chaudhuri and K. Kotwal, *Hyperspectral Image Fusion*. New York, NY, USA: Springer, 2013.

[4] C.-I. Chang, Ed., *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ, USA: Wiley, 2007.

[5] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[6] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[7] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 2329–2336.

[8] N. Yokoya and A. Iwasaki, "Hyperspectral and multispectral data fusion mission on hyperspectral imager suite (HISUI)," in *Proc. IEEE IGARSS*, Melbourne, VIC Australia, Jul. 2013, pp. 4086–4089.

[9] C. Lanaras, E. Baltsavias, and K. Schindler, "Advances in hyperspectral and multispectral image fusion and spectral unmixing," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XL-3/W3, pp. 451–458, Aug. 2015.

[10] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[11] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014.

[12] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.

[13] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[14] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[15] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "25 years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2012.

[16] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.

[17] J. G. Liu, "Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.

[18] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, Jun. 2015.

[19] Q. Wei, J. M. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[20] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 3176–3180.

[21] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[22] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009.

[23] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3453–3462, Sep. 2012.

[24] G. Casella and E. I. George, "Explaining the Gibbs sampler," *Amer. Statist.*, vol. 46, no. 3, pp. 167–174, 1992.

[25] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of multiband images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.

[26] M. Simões, J. M. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

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

[28] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[29] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 963–978, Sep. 2011.

[30] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[31] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A nonnegative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE ICASSP*, Vancouver, BC, Canada, May 2013, pp. 1409–1413.

[32] 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.

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

[34] C.-H. Lin, A. Ambikapathi, W.-C. Li, and C.-Y. Chi, "On the endmember identifiability of Craig's criterion for hyperspectral unmixing: A statistical analysis for three-source case," in *Proc. IEEE ICASSP*, Vancouver, BC, Canada, May 2013, pp. 2139–2143.

[35] C.-H. Lin, R. Wu, W.-K. Ma, C.-Y. Chi, and Y. Wang. "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization." Unpublished paper, 2017. [Online]. Available: https://arxiv.org/abs/1708.02883

[36] 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.

[37] Y. Zhang, Y. Wang, Y. Liu, C. Zhang, M. He, and S. Mei, "Hyperspectral and multispectral image fusion using CNMF with minimum endmember simplex volume and abundance sparsity constraints," in *Proc. IEEE IGARSS*, Milan, Italy, Jul. 2015, pp. 1929–1932.

[38] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. Boca Raton, FL, USA: CRC Press, 2017.

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

[40] P.-A. Thouvenin, N. Dobigeon, and 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.

[41] C. J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1589–1596, Nov. 2007.

[42] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[43] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1946–1961, Apr. 2016.

[44] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, May 2014.

[45] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," *Math. Methods Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007.

[46] M. Grant and S. Boyd. (Apr. 2011). *CVX: MATLAB Software for Disciplined Convex Programming, Version 1.21*. [Online]. Available: http://cvxr.com/cvx/

[47] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[48] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "Blind and fully constrained unmixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5510–5518, Dec. 2014.

[49] É. Thiébaut, F. Soulez, and L. Denis, "Exploiting spatial sparsity for multiwavelength imaging in optical interferometry," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 30, no. 2, pp. 160–170, 2013.

[50] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[51] Z. H. Nezhad, A. Karami, R. Heylen, and P. Scheunders, "Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2377–2389, Jun. 2016.

[52] *ROSIS Free Pavia University Data*. Accessed: Oct. 25, 2016. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[53] *IKONOS Satellite Imagery*. Accessed: Nov. 1, 2016. [Online]. Available: https://apollomapping.com/imagery/high-resolution-imagery/ikonos

[54] M. A. Bendoumi, M. He, and S. Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6574–6583, Oct. 2014.

[55] R. W. Basedow, D. C. Carmer, and M. E. Anderson, "HYDICE system: Implementation and performance," *Proc. SPIE*, vol. 2480, pp. 258–267, Jun. 1995.

[56] *AVIRIS Free Standard Data Products*. Accessed: Nov. 4, 2016. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[57] Q. Wei, J. M. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.

[58] T. Stathaki, *Image Fusion*. New York, NY, USA: Academic, 2008.

[59] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.

[60] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3586–3594.

[61] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[62] C.-H. Lin, C.-Y. Chi, L. Chen, D. J. Miller, and Y. Wang, "Detection of sources in non-negative blind source separation by minimum description length criterion," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2017.2749279.

[63] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[64] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[65] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao. "Sparse representation based multi-sensor image fusion: A review." Unpublished paper, 2017. [Online]. Available: https://arxiv.org/abs/1702.03515

[66] R. L. Wheeden and A. Zygmund, *Measure and Integral: An Introduction to Real Analysis*. New York, NY, USA: Marcel Dekker, 1977.

Chia-Hsiang Lin received the B.S. degree in electrical engineering and the Ph.D. degree in communications engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2010 and 2016, respectively.

From 2015 to 2016, he was a Visiting Student with Virginia Tech, Blacksburg, VA, USA. He was also a Post-Doctoral Researcher with NTHU from 2016 to 2017, and a Visiting Scholar with The Chinese University of Hong Kong, Hong Kong, in 2014 and 2017. He is currently a Post-Doctoral Researcher with the Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal. His research interests include network science, game theory, convex geometry and optimization, blind source separation, and imaging science.

Dr. Lin received the Outstanding Doctoral Dissertation Award from the Chinese Image Processing and Pattern Recognition Society in 2016 and the Best Doctoral Dissertation Award from the IEEE Geoscience and Remote Sensing Society in 2016.

Fei Ma received the B.S. degree in communication engineering and the M.S. degree in computer science from Liaoning Technical University, Fuxin, China, in 2001 and 2004, respectively, and the Ph.D. degree in circuit and system from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010.

He held a post-doctoral position at the University of Electronic Science and Technology of China, Chengdu, China, in 2014. Then, he conducted research as a Visiting Scholar with National Tsing Hua University, Hsinchu, Taiwan, from 2015 to 2016. He is currently a Faculty Member with the School of Electronic and Information Engineering, Liaoning Technical University, Huludao, China. His research interests include hyperspectral image processing, radar signal processing, pattern recognition, and convex optimization.

Chong-Yung Chi (S'83–M'83–SM'89) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1983.

From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA, USA. He has been a Professor with the Department of Electrical Engineering since 1989 and the Institute of Communications Engineering (ICE) since 1999 (also the Chairman of ICE during 2002–2005), National Tsing Hua University, Hsinchu, Taiwan. He has authored over 220 technical papers, including over 80 journal papers (mostly in the IEEE TRANSACTIONS ON SIGNAL PROCESSING), over 130 peer-reviewed conference papers, four book chapters, and two books, including a new textbook *Convex Optimization for Signal Processing and Communications from Fundamentals to Applications* (CRC Press, 2017) (popularly used in major universities in China). His research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, and biomedical and hyperspectral image analysis.

Dr. Chi has been a member of the Sensor Array and Multichannel Technical Committee since 2013, and is a member of the IEEE Signal Processing Society. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2001 to 2005 and from 2012 to 2015, and the IEEE SIGNAL PROCESSING LETTERS from 2006 to 2010.

Chih-Hsiang Hsieh received the B.S. degree from the Department of Communications Engineering, National Central University, Taoyuan City, Taiwan, in 2015, and the M.S. degree from the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2017.

He is currently a Graduate Student with the Institute of Communications Engineering, National Tsing Hua University. His research interests include convex optimization, hyperspectral imaging, and data fusion.