arXiv:1907.01149v2 [eess.IV] 8 Mar 2020

Hyperspectral Super-Resolution via Global-Local Low-Rank Matrix Estimation

Ruiyuan Wu[†], Wing-Kin Ma[†], Xiao Fu[‡], and Qiang Li[⋆]

[†]Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR of China

[‡]School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, USA

[⋆]School of Information and Communications Engineering, University of Electronic Science and Technology of China, China

E-mails: rywu@ee.cuhk.edu.hk, wkma@ee.cuhk.edu.hk, xiao.fu@oregonstate.edu, lq@uestc.edu.cn

March 10, 2020

Abstract

Hyperspectral super-resolution (HSR) is a problem that aims to estimate an image of high spectral and spatial resolutions from a pair of co-registered multispectral (MS) and hyperspectral (HS) images, which have coarser spectral and spatial resolutions, respectively. In this paper we pursue a low-rank matrix estimation approach for HSR. We assume that the spectral-spatial matrices associated with the whole image and the local areas of the image have low-rank structures. The local low-rank assumption, in particular, has the aim of providing a more flexible model for accounting for local variation effects due to endmember variability. We formulate the HSR problem as a global-local rank-regularized least-squares problem. By leveraging on the recent advances in non-convex large-scale optimization, namely, the smooth Schatten-p approximation and the accelerated majorization-minimization method, we develop an efficient algorithm for the global-local low-rank problem. Numerical experiments on synthetic, semi-real and real data show that the proposed algorithm outperforms a number of benchmark algorithms in terms of recovery performance.

1 Introduction

Hyperspectral (HS) sensors have limited spatial resolution as a tradeoff for achieving high spectral resolution. Such a tradeoff is due to hardware limitations and the measurement mechanism. In a nutshell, a certain amount of light energy reflected by the scene is required for each spectral band of an HS pixel to achieve sufficiently high SNRs. For a sensor with coarse spectral resolution, enough energy can be acquired from a small area by accumulating energy of a wide range of spectral bands.

When the spectral resolution increases, the area sensed by a pixel needs to be enlarged to acquire the same amount of energy for each spectral band, which leads to a lower spatial resolution. How to enhance the spatial resolution of HS images has been a subject of great interest. Recently, the idea of using an additional multispectral (MS) image—which has finer spatial resolution than the HS but possesses only several coarse spectral bands—for HS spatial resolution enhancement has shed new light on the subject. This MS-aided enhancement problem is called *hyperspectral super-resolution* (HSR) or HS-MS data fusion. One approach for HSR is to adapt pansharpening techniques in fusion of panchromatic and HS images [1]. Another approach, which is currently more popular, is based on low-dimensional data models. The low-dimensional model approach assumes that the spectral pixels of the target high-spatial-resolution HS image, or the super-resolution (SR) image for short, lie in a low-dimensional subspace. This assumption is particularly reasonable in the linear spectral mixture scenario: Since the aforesaid scenario has every spectral pixel described as a linear combination of the spectral signatures of the underlying endmembers, the spectral pixels lie in a subspace spanned by the endmember spectral signatures. Also, since the number of endmembers is often much smaller than the number of HS spectral bands, the subspace dimension is low. The low-dimensional model may also be applicable to some classes of nonlinear spectral mixtures such as the bilinear mixture model [2, 3]. The low-dimensional model approach has strong connections to hyperspectral unmixing (HU). To be specific, insights and methods in HU are quite often used in the low-dimensional model approach. A comparative review has shown that the low-dimensional model approach can lead to better recovery than those from the pansharpening approach, assuming no or negligible HS-MS co-registration error [1].

To facilitate our discussion later, in this paper we taxonomize the existing low-dimensional model-based HSR methods into two types.

1. *Matrix factorization:* This type of methods models the spectral-spatial matrix of the SR image as a product of two matrix factors—one being the spectral dictionary, and another the coefficients for low-dimensional representation, and it seeks to jointly estimate the two matrix factors from the observed HS-MS pair. Coupled non-negative matrix factorization (CNMF) [4], a pioneering HSR method, falls into this type. As its name suggests, CNMF exploits the non-negativity of the matrix factors. Subsequent research follows the same route and exploits other problem structures—sparsity [5], the sum-to-one abundance condition from the linear spectral mixture model [6, 7], non-local pixel similarity [8], and many more—to attempt to improve recovery quality.

2. *Dictionary-based regression:* This type of methods also assumes that the spectral-spatial matrix of the SR image is the product of a spectral dictionary and the associated coefficient matrix. The difference is that it does not seek to jointly estimate the spectral dictionary and the coefficients. It first determines the spectral dictionary via some easy way, and then uses that spectral dictionary to perform regression to recover the coefficients. A typical example is HySure [9], which extracts the spectral dictionary by applying vertex component analysis (VCA) [10] to the HS image, and then recovers the coefficients by applying spatial total variation-regularized linear regression to the HS-MS image pair. Other methods include [11, 12], which exploit the local low-rank structure; this will be further discussed later. The dictionary-based regression methods are easy to implement compared to the matrix factorization methods.

Research on these two types of methods is mostly focused on the practical aspects, and it is worthwhile to note that some specific methods have recently been shown to exhibit theoretical

recovery guarantees as well [13, 14]—which supports the soundness of the low-dimensional model approach via a theoretical lens.

Matrix factorization and dictionary-based regression are considered most representative in HSR methods, although there are others. For example, tensor factorization has recently been studied for HSR [15–20]. The tensor model is also a low-dimensional model, and it exploits not only the spectral-spatial structure but also the two-dimensional spatial structure. Tensor factorization is shown to exhibit favorable sufficient conditions on exact recovery guarantees [16, 17]. In addition, deep learning for HSR has most recently gained growing interest [21–23].

Under the low-dimensional model, HSR can be seen as a problem of recovering a low-rank matrix from incomplete observations; this will be elaborated upon in Sections 3 and 4. From this perspective, the problem is nearly the same as the matrix completion problem which has drawn widespread interest in fields such as recommender systems, machine learning and mathematical optimization [24–26]. The problem in matrix completion is that we have a matrix with many missing entries, and we aim to recover the missing entries from the available entries. The main assumption in matrix completion is that the matrix to be recovered has low rank structure. This assumption is the same as the low-dimensional model assumption in HSR. In matrix completion we see two main types of methods. One is matrix factorization, which shares the same rationale as matrix factorization for HSR. Another is low-rank matrix estimation. This approach does not predetermine the target dimension of the low-dimensional subspace, or the target rank of the matrix to be recovered, which is the case in matrix factorization. Instead of using the matrix factorization model, it seeks the minimum rank matrix for accomplishing the task. A well-known method in low-rank matrix estimation is the nuclear norm minimization method [24]. It is a convex solution, and the idea is to approximate the hard-to-handle rank function by the nuclear norm which is convex. Non-convex rank approximation, such as the Schatten-p approximation, was also considered for approximating the low-rank problem better [27]. Back to HSR, while we have seen numerous studies on matrix factorization and dictionary-based regression, we see far less on low-rank matrix estimation.

In addition, and as a different issue, the existing low-dimensional model-based HSR methods are usually not designed to account for the endmember variability (EV) effects due, for instance, to illumination conditions and intrinsic spectral variability of the materials [28]. In low-dimensional models, EV means that the spectral dictionary can vary in space. Taking a step back to HU, we have seen studies that use the matrix factorization method to deal with endmember variability [29–31]. In that regard, a possibility one can consider is to adapt such matrix factorization methods to the HSR application. We are however unaware of such development as of the writing of this paper.

In this paper, our objective is to explore the potential of low-rank matrix estimation in HSR. Our study is not a direct application of the existing low-rank matrix estimation methods, such as the nuclear norm minimization method. Our formulation takes the possibility of EV into consideration. We posit a global-local low-rank structure with the SR image, in which not only the spectral-spatial matrix of the whole SR image has low-rank structure, but that of each local area also has another low-rank structure. This assumption means that each local area can have its low-dimensional representation. The local low-rank assumption provides the model with the flexibility to account for EV. Moreover, since the low-dimensional subspaces of the local areas should be related, particularly the neighboring ones, we also assume that the whole spectral-spatial matrix has low rank and utilize it to tie the local subspaces together. As we will see, our global-local low-rank matrix estimation leads to a fairly clean formulation. In comparison, if one applies the

EV-present matrix factorization methods in HU to HSR, the resulting formulation would be more complicated.

The arising challenge and our proposed solution should be described. The global-local low-rank matrix estimation problem is a non-convex large-scale problem. For example, to recover an SR image of 100 spectral bands and of size 200×200, our problem requires us to handle $100 \times 200^2 = 4,000,000$ optimization variables. An efficient optimization strategy is clearly needed. We attack the problem by leveraging on the recent advances in non-convex large-scale optimization, namely, the smooth Schatten-p approximation [27] and an accelerated version of the majorization-minimization (MM) method [32]. As mentioned, the smooth Schatten-p approximation, albeit non-convex, approximates rank better in comparison with the convex nuclear norm. Also, its smooth nature enables us to access powerful machinery in smooth optimization. The accelerated MM method is a combination of inexact MM and the accelerated projected gradient (PG) method. Our recent research in another context [32]—which also deals with large problem sizes—has suggested that this type of accelerated methods runs very fast in practice. Using the above two techniques, we develop a fast algorithm called Global-Local lOw-Rank promotIng Algorithm (GLORIA). As will be shown by numerical results, GLORIA gives competitive recovery performance compared to the state of the arts and related methods. We conducted semi-real experiments on five different datasets, and GLORIA consistently ranks first or second in performance indicators such as peak SNR and spectral angle mappers. We also provide results on synthetic and real data experiments, in which GLORIA also exhibits promising performance.

Before we proceed to the description of our method, we should mention related works. First, the dictionary-based regression method in [11], HSR-LDL-EIA, also utilizes some kind of local low-rank structures. HSR-LDL-EIA considers the linear spectral mixture model and assumes that the number of endmembers in each local area is very small and no greater than the number of MS bands. With that assumption, the HSR problem can be easily solved in a local-area-by-local-area fashion—which is what HSR-LDL-EIA does. The local low-rank assumption used in HSR-LDL-EIA is not the same as the one used by us. As discussed earlier, our local low-rank assumption is to cater for EV. Second, the recent work in [12] takes insight from the local low-rank assumption in HSR-LDL-EIA and proposes a dictionary-based regression method using local nuclear norm-regularized linear regression. Again, the local low-rank assumption in [12] is founded on the argument that the number of endmembers in each local area is small. We should also mention the works [33, 34] which follow similar rationales as those in [11, 12].

Let us summarize our contributions.

1. We consider low-rank matrix estimation for HSR, which has not been studied in prior works. We propose a global-local low-rank approach which aims to account for EV effects.

2. The global-local low-rank approach requires us to tackle a large-scale non-convex optimization problem. We custom-develop an efficient algorithm for such purpose, using recent advances in large-scale non-convex optimization. As will be shown in numerical experiments, our algorithm has competitive recovery performance.

Readers who are interested in trying our algorithm can find the source codes at `https://github.com/REIYANG/GLORIA`.

2 Background

2.1 The Measurement Model

Let us begin by providing a concise review of the background. Fig. 1 depicts the scenario. We have a scene observed by an HS sensor and an MS sensor. The MS sensor has a lower spectral resolution than the HS sensor, while the HS sensor has a lower spatial resolution than the MS sensor. The goal of HSR is to use the observed MS and HS images to construct a higher resolution image whose spectral resolution is identical to that of the HS sensor, and spatial resolution the MS. For convenience, the image we seek to construct will be called the super-resolution (SR) image. As a common assumption (see, e.g., [4]), the HS image is modeled as a spatially degraded version of the SR image by means of spatial blurring and down-sampling. Also, the MS image is modeled as spectrally degraded version of the SR image by means of spectral bandpass averaging.

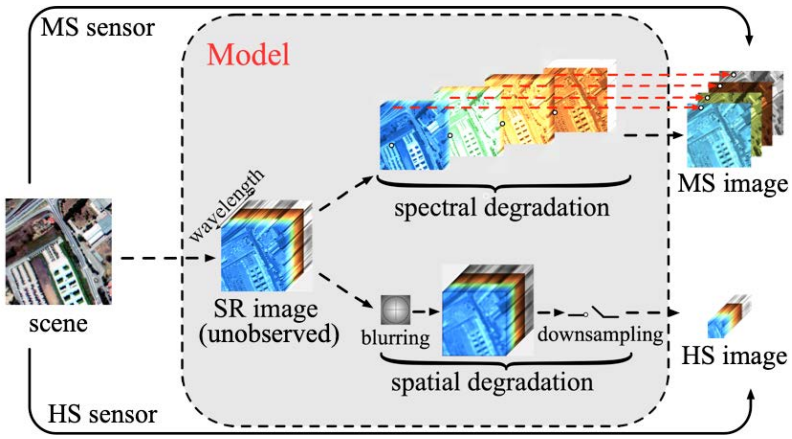


Figure 1: The HSR scenario.

The HSR data model is as follows. Assuming that the HS and MS images are co-registered, we model the HS and MS images as

$$\boldsymbol{Y}_{\mathrm{M}} = \boldsymbol{F}\boldsymbol{X} + \boldsymbol{V}_{\mathrm{M}}, \tag{1a}$$

$$\boldsymbol{Y}_{\mathrm{H}} = \boldsymbol{X}\boldsymbol{G} + \boldsymbol{V}_{\mathrm{H}}, \tag{1b}$$

where $\boldsymbol{Y}_{\mathrm{M}} \in \mathbb{R}^{M_{\mathrm{M}} \times L}$ and $\boldsymbol{Y}_{\mathrm{H}} \in \mathbb{R}^{M \times L_{\mathrm{H}}}$ are the spectral-spatial matrices of the observed MS and HS images, respectively (resp.); M_{M} and M are the numbers of spectral bands of the MS and HS images, resp., with $M_{\mathrm{M}} < M$; L and L_{H} are the numbers of pixels of the MS and HS images, resp., with $L_{\mathrm{H}} < L$; $\boldsymbol{X} \in \mathbb{R}^{M \times L}$ is the spectral-spatial matrix of the SR image; $\boldsymbol{F} \in \mathbb{R}^{M_{\mathrm{M}} \times M}$ and $\boldsymbol{G} \in \mathbb{R}^{L \times L_{\mathrm{H}}}$ describe the measurement responses that lead to the MS and HS images, resp.; $\boldsymbol{V}_{\mathrm{M}}$ and $\boldsymbol{V}_{\mathrm{H}}$ are noise. Note that $\boldsymbol{F}$ designates the relative spectral bandpass responses from the SR image to the MS image, while $\boldsymbol{G}$ specifies the spatial blurring and down-sampling responses that result in the HS image. The measurement response matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ are assumed to be known, and in practice $\boldsymbol{F}$ and $\boldsymbol{G}$ can be acquired either by calibration [4] or by estimation from the HS and MS images [9, 35]. Furthermore, the MS and HS images are measured by means of reflectance, with values lying between 0 and 1. As such, we may assume that $x_{ij} \in [0, 1]$, for all i, j.

2.2 The Matrix Factorization Model

Next, we describe the matrix factorization model which is the core assumption for matrix factorization and dictionary-based regression methods in HSR. In the matrix factorization model we assume that the SR image $\boldsymbol{X}$ can be factored as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{2}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is the spectral dictionary; $\boldsymbol{S} \in \mathbb{R}^{N\times L}$ is the coefficient matrix; N is the target rank or model order, which is pre-fixed and is often chosen to be much less than M and L. In many existing studies, the model (2) is seen as the linear spectral mixture model in which the columns $\boldsymbol{a}_i$'s of $\boldsymbol{A}$ are interpreted as spectral signatures of the endmembers of the scene, and the columns $\boldsymbol{s}_i$'s of $\boldsymbol{S}$ the associated abundances of the pixels. Under the model (2), the matrix factorization methods seek to find $(\boldsymbol{A}, \boldsymbol{S})$ from $(\boldsymbol{Y}_{\mathrm{M}}, \boldsymbol{Y}_{\mathrm{H}})$ by minimizing a data-fitting loss function, and thereby reconstruct $\boldsymbol{X}$ by $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$. For example, in CNMF [4], the idea is to solve

$$\min_{\boldsymbol{A}\geq\boldsymbol{0},\boldsymbol{S}\geq\boldsymbol{0}} \ \|\boldsymbol{Y}_{\mathrm{M}} - \boldsymbol{F}\boldsymbol{A}\boldsymbol{S}\|_F^2 + \|\boldsymbol{Y}_{\mathrm{H}} - \boldsymbol{A}\boldsymbol{S}\boldsymbol{G}\|_F^2, \tag{3}$$

where $\boldsymbol{X} \geq \boldsymbol{0}$ means that $\boldsymbol{X}$ is element-wise non-negative; $\|\cdot\|_F$ denotes the Frobenius norm. CNMF, as well as other matrix factorization formulations, considers structured factors to better exploit the underlying problem structure. CNMF utilizes the fact that, under the linear spectral mixture model, $\boldsymbol{A}$ and $\boldsymbol{S}$ are non-negative. Moreover, in the dictionary-based regression methods, one first determines $\boldsymbol{A}$ from $\boldsymbol{Y}_{\mathrm{H}}$ by methods such as principal component analysis (PCA) or VCA. Then, $\boldsymbol{S}$ is estimated by solving the data-fitting problem like the one in (3), but with $\boldsymbol{A}$ fixed. In estimating $\boldsymbol{S}$, regularization such as total variation may be added for problem structure exploitation [9]. A common, and key, concept behind the various matrix factorization and dictionary-based regression methods is that although $\boldsymbol{X}$ is a high-dimensional matrix, we hold the belief that every spectral pixel $\boldsymbol{x}_i$ of $\boldsymbol{X}$ should lie in a low-dimensional subspace spanned by $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$.

As an alternative view, matrix factorization and dictionary-based regression may be regarded as methods for estimating a low-rank matrix $\boldsymbol{X}$ from the incomplete HS-MS observations. Specifically, the low-dimensional subspace assumption with $\boldsymbol{X}$ implies $\mathrm{rank}(\boldsymbol{X}) = \mathrm{rank}(\boldsymbol{A}\boldsymbol{S}) \leq N$.

3 Global-Local Low-Rank Formulation

This section describes the main development of this paper, global-local low-rank matrix estimation.

3.1 A Brief Review of Low-Rank Matrix Estimation

The low-rank matrix estimation methods have recently become popular in the context of matrix completion [24,26]. Let us first describe how the *de facto* standard in low-rank matrix estimation, namely, the nuclear norm approximation, can be applied to HSR. In low-rank matrix estimation, we assume $\boldsymbol{X}$ to be a low-rank matrix. This assumption can be interpreted as requiring the columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ to lie in a low-dimensional subspace. The matrix factorization model (2) can also be seen as constraining $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ to lie in a low-dimensional subspace, with the subspace dimension no greater than N. Hence, both the low-rank matrix estimation and matrix factorization methods exploit low-dimensional data structures. The idea with low-rank matrix estimation is to find a

low-rank $\boldsymbol{X}$ whose data fitting loss is small. A common low-rank matrix estimation formulation is as follows

$$\min_{\boldsymbol{X}\in\mathbb{R}^{M\times L}} \ \ell(\boldsymbol{X}) + \gamma\,\mathrm{rank}(\boldsymbol{X}), \tag{4}$$

where $\gamma > 0$ is given and is called a regularization parameter; $\ell : \mathbb{R}^{M\times L} \to \mathbb{R}$ denotes the data-fitting loss function and is given by

$$\ell(\boldsymbol{X}) = \tfrac{1}{2}\|\boldsymbol{Y}_{\mathrm{M}} - \boldsymbol{F}\boldsymbol{X}\|_F^2 + \tfrac{1}{2}\|\boldsymbol{Y}_{\mathrm{H}} - \boldsymbol{X}\boldsymbol{G}\|_F^2. \tag{5}$$

Let us give a brief comparison of low-rank matrix estimation and matrix factorization. Matrix factorization methods, such as the one in (3), require one to pre-determine the target rank N. There is no such rank constraint in low-rank matrix estimation. The rank of $\boldsymbol{X}$ itself serves as the regularization for the low-rank matrix recovery endeavor, and the parameter γ controls the balance between low rankness and goodness of data fitting.

The challenge with solving problem (4) is that the rank function in (4) is hard to handle; it is non-convex and non-differentiable. The state of the art handles this issue by applying the nuclear norm approximation. The reader is referred to the literature [26] for detailed description of the concept, and here we concisely explain the idea. The rank of $\boldsymbol{X}$ is identical to the number of nonzero singular values of $\boldsymbol{X}$, and hence we can express $\mathrm{rank}(\boldsymbol{X})$ as

$$\mathrm{rank}(\boldsymbol{X}) = \sum_{i=1}^{\min\{M,L\}} u(\sigma_i(\boldsymbol{X})), \tag{6}$$

where $\sigma_i(\boldsymbol{X}) \geq 0$ denotes the ith largest singular value of $\boldsymbol{X}$; $u(y) = 1$ if $y > 0$, and $u(y) = 0$ if $y = 0$. The idea with nuclear norm approximation is to approximate $\mathrm{rank}(\boldsymbol{X})$ by removing u from (6), which leads to the following approximate function:

$$\|\boldsymbol{X}\|_* = \sum_{i=1}^{\min\{M,L\}} \sigma_i(\boldsymbol{X}).$$

The above function is called the nuclear norm and is known to be convex [24, 26]. Applying this nuclear norm approximation of rank to problem (4) gives rise to the following problem:

$$\min_{\boldsymbol{X}\in\mathbb{R}^{M\times L}} \ \ell(\boldsymbol{X}) + \gamma\|\boldsymbol{X}\|_*. \tag{7}$$

The advantage of the above approximation is that it is a convex problem. Moreover, problem (7) can be efficiently solved by methods such as the accelerated proximal gradient method [36] and ADMM [37].

3.2 The Global-Local Low-Rank Model

We consider a *global-local* low-rank assumption for the HSR problem. To explain the idea, consider the illustration in Fig. 2. We segment the SR image into a number of local patches. Our belief is that each local patch exhibits its own local rank structure. Or, the low-dimensional subspace of one patch does not need to be the same as that of another. This assumption appears to make sense since real images may have local variation effects due to EV. Moreover, we still keep the old low-rank

assumption with $\boldsymbol{X}$. This is because the low-dimensional subspace of one patch should be related to those of its neighboring patches, and such correlations may result in a low-rank $\boldsymbol{X}$ in the global sense (with a higher global rank than the local ranks). Alternatively speaking, the low-dimensional subspace of the whole $\boldsymbol{X}$ plays the role of tying together the low-dimensional subspaces of the local patches.

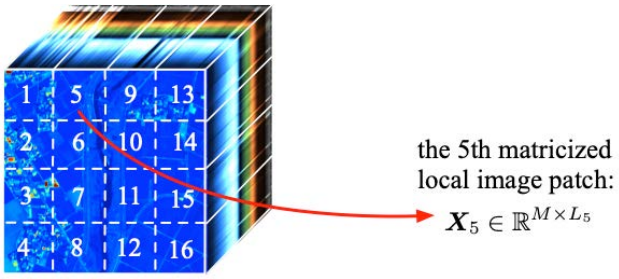


Figure 2: Segmentation of an image into patches.

To write down the global-local low-rank assumption, we assume that the pixel indices of the image are arranged such that $\boldsymbol{X}$ can be conveniently expressed as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_P \end{bmatrix},$$

where each $\boldsymbol{X}_i \in \mathbb{R}^{M \times L_i}$ is the spectral-spatial matrix of a local area, or local patch, of the image; P is the number of patches; L_i is the number of pixels of patch i. For example, as illustrated in Fig. 2, we can divide the image into equal-space rectangular blocks as our local patches. Other ways to form the local patches, e.g., via segmentation [11], may also be considered. We assume that every $\boldsymbol{X}_i$ is a low-rank matrix, and $\boldsymbol{X}$ is also a low-rank matrix.

The global-local low-rank assumption stated above is fairly general and does not restrict itself to specific mixture models such as the linear spectral mixture model. On the other hand, we can better understand the assumption by a more concrete example, in which the linear spectral mixture model is used, as follows. Suppose we model each $\boldsymbol{X}_i$ to follow the linear spectral mixture model

$$\boldsymbol{X}_i = \boldsymbol{A}_i \boldsymbol{S}_i, \tag{8}$$

where $\boldsymbol{A}_i \in \mathbb{R}^{M \times N}$ and $\boldsymbol{S}_i \in \mathbb{R}^{N \times L_i}$ are the endmember and abundance matrices of patch i; N is the total number of endmembers in the whole SR image $\boldsymbol{X}$. In this model, we assume the presence of EV by allowing the endmember matrix $\boldsymbol{A}_i$ to be different for each patch. Note that our model assumes that EV appears at the patch level, not at the pixel level, and this can be justified if the local region of each patch is small enough such that the endmember spectral signatures experience little or no variation within the patch. We also want to impose an assumption that $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_P$ are correlated, since, in reality, they should be variations of one another. Such correlation would mean that $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_P$ can be linearly represented by a "global" basis $\boldsymbol{B}$, whose dimension is no less than N, but not significantly greater than N owing to the correlations. This further means that $\boldsymbol{X}$ lives in a low-dimensional subspace with $\boldsymbol{B}$ as its basis. Our global low-rank assumption is to exploit the global low-dimensional structure.

Additionally, in the above motivating model example, at first sight one would be tempted to say that the rank of $\boldsymbol{X}_i$ is universally given by $\operatorname{rank}(\boldsymbol{X}_i) = N$ (under the slightly technical premise

that $\boldsymbol{A}_i$ and $\boldsymbol{S}_i$ have full column rank and full row rank, resp.). In fact, it is reasonable to assume non-identical rank($\boldsymbol{X}_i$). In practice, it is likely that each local region is composed of a small number of endmembers, rather than all of the endmembers. Thus we can assume that among all the rows $\boldsymbol{s}_i^1, \ldots, \boldsymbol{s}_i^N$ of $\boldsymbol{S}_i$, only N_i of them are nonzero (or active). Consequently we have rank($\boldsymbol{X}_i$) = N_i (again, under the technical premise that $\boldsymbol{A}_i$ has full column rank and that the nonzero $\boldsymbol{s}_i^j$'s are linearly independent), which is our local low-rank assumption.

(a) Chikusei (b) Indian Pines (c)Washington DC Mall (d) University of Pavia

(e) Moffett Field (f) Cuprite

Figure 3: Color composite images of the tested HS images.

3.3 An Experiment

To support our argument that the global-local low-rank structure would be a reasonable assumption, we perform the following numerical experiment. We take real HS images and numerically evaluate their global and low rank values. The images come from six different datasets, namely, Chikusei, Cuprite, Indian Pines, University of Pavia, Washington DC Mall and Moffett Field. They are shown in Fig. 3 in color composite forms. For each image, we obtain the local patches $\boldsymbol{X}_i$'s by the equal-space rectangular segmentation in Fig. 2. For each local patch $\boldsymbol{X}_i$ we evaluate its rank in an approximate manner, specifically, by finding the smallest integer r such that

$$\frac{\sum_{j=1}^{r} \sigma_j(\boldsymbol{X}_i)^2}{\sum_{j=1}^{\min\{M,L_i\}} \sigma_j(\boldsymbol{X}_i)^2} \geq 0.9999.$$

Table 1 shows the approximate ranks of the tested images under different patch sizes. One can clearly see that all the tested images exhibit global-local low-rank characteristics. For example, for the Chikusei image, the global rank is 10 while the average local rank for $P = 16^2$ is around 6.5.

Table 1: The local approximate ranks of the real HS datasets. The patch size refers to the spectral-spatial dimension, $M \times L_i$, of the local patches.

P	Chikusei (480 × 480 pixels, 128 bands)		Cuprite (480 × 480 pixels, 187 bands)	
	patch size	approx. rank	patch size	approx. rank
1^2	$128 \times 230,400$	10	$187 \times 230,400$	9
2^2	$128 \times 57,600$	8.75±0.96	$187 \times 57,600$	8.25±1.26
3^2	$128 \times 25,600$	8.33±1.73	$187 \times 25,600$	7.44±1.33
4^2	$128 \times 14,400$	8.19±1.83	$187 \times 14,400$	7.13±1.20
5^2	$128 \times 9,216$	7.92±1.89	$187 \times 9,216$	6.76±1.13
6^2	$128 \times 6,400$	7.81±1.98	$187 \times 6,400$	6.64±1.13
8^2	$128 \times 3,600$	7.38±2.13	$187 \times 3,600$	6.38±1.06
10^2	$128 \times 2,304$	7.08±2.29	$187 \times 2,304$	6.12±1.09
12^2	$128 \times 1,600$	6.89±2.34	$187 \times 1,600$	6.03±1.02
15^2	$128 \times 1,024$	6.52±2.35	$187 \times 1,024$	5.84±1.00
16^2	128×900	6.47±2.36	187×900	5.79±1.01
P	Indian Pine (120 × 120 pixels, 178 bands)		University of Pavia (240 × 240 pixels, 103 bands)	
	patch size	approx. rank	patch size	approx. rank
1^2	$178 \times 14,400$	22	$103 \times 57,600$	22
2^2	$178 \times 3,600$	19.75±1.71	$103 \times 14,400$	20.75±1.89
3^2	$178 \times 1,600$	17.56±4.10	$103 \times 6,400$	20.44±2.07
4^2	178×900	16.63±3.77	$103 \times 3,600$	20.44±2.66
5^2	178×576	15.20±3.94	$103 \times 2,304$	20.28±2.84
6^2	178×400	14.69±4.44	$103 \times 1,600$	20.03±3.08
8^2	178×255	13.36±4.47	103×900	19.80±3.43
P	Washington DC Mall (240 × 240 pixels, 191 bands)		Moffett Field (240 × 240 pixels, 187 bands)	
	patch size	approx. rank	patch size	approx. rank
1^2	$191 \times 57,600$	6	$187 \times 57,600$	13
2^2	$191 \times 14,400$	5.50±0.58	$187 \times 14,400$	12.00±2.71
3^2	$191 \times 6,400$	5.22±0.67	$187 \times 6,400$	11.56±2.65
4^2	$191 \times 3,600$	4.50±1.03	$187 \times 3,600$	11.00±2.94
5^2	$191 \times 2,304$	4.52±1.26	$187 \times 2,304$	10.68±2.87
6^2	$191 \times 1,600$	4.39±1.23	$187 \times 1,600$	10.56±2.92
8^2	191×900	4.20±1.17	187×900	10.08±2.89

3.4 The Global-Local Low-Rank Matrix Estimation Formulation

Under the global-local low-rank assumption in the preceding subsection, it is natural to formulate the HSR problem as the following global-local low-rank matrix estimation problem:

$$\min_{\boldsymbol{X} \in \mathcal{X}} \ \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \gamma_i \operatorname{rank}(\boldsymbol{X}_i), \tag{9}$$

where $\gamma_0, \gamma_1, \cdots, \gamma_P > 0$ are given regularization parameters; we denote $\boldsymbol{X}_0 = \boldsymbol{X}$ for notational convenience; ℓ has been defined in (5); $\mathcal{X} \subseteq \mathbb{R}^{M \times L}$ is given by $\mathcal{X} = [0,1]^{M \times L}$.

As reviewed previously, the standard approach to handle problem (4) is to approximate each $\text{rank}(\boldsymbol{X}_i)$ by the nuclear norm $\|\boldsymbol{X}_i\|_*$. Here we pursue a different option, namely, the the smooth Schatten-p approximation [27]. To put it into context, let us first define a notation. Given a symmetric $n \times n$ positive definite matrix $\boldsymbol{A}$ and a number p, we define

$$\boldsymbol{A}^p = \boldsymbol{U}\boldsymbol{\Lambda}^p\boldsymbol{U}^T, \tag{10}$$

where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ constitute the eigen-decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$; note that $\boldsymbol{U}$ is orthogonal, $\boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \ldots \geq \lambda_n > 0$, and $\boldsymbol{\Lambda}^p = \text{Diag}(\lambda_1^p, \ldots, \lambda_n^p)$. The smooth Schatten-$p$ function of an $M \times L$ matrix $\boldsymbol{X}$, with $M \leq L$, is defined as

$$\phi_{p,\tau}(\boldsymbol{X}) = \sum_{i=1}^{M} \left(\sigma_i(\boldsymbol{X})^2 + \tau\right)^{p/2} = \text{tr}\left((\boldsymbol{X}\boldsymbol{X}^T + \tau\boldsymbol{I})^{p/2}\right),$$

where $p > 0, \tau > 0$ are given. This function has the following properties:

(i) $\phi_{p,\tau}$ is smooth (or has derivatives of all orders);

(ii) $\phi_{p,\tau}$ is convex for $p \geq 1$, and non-convex for $p < 1$;

(iii) as $\tau \to 0$, $\phi_{1,\tau}(\boldsymbol{X}) \to \|\boldsymbol{X}\|_*$;

(iv) as $p \to 0, \tau \to 0$, $\phi_{p,\tau}(\boldsymbol{X}) \to \text{rank}(\boldsymbol{X})$.

As can be seen in the above properties, the smooth Schatten-p function is a smooth approximation of rank. Ideally we would like to choose very small p and τ so that $\phi_{p,\tau}(\boldsymbol{X})$ closely approximates $\text{rank}(\boldsymbol{X})$, but using very small p and τ will also make $\phi_{p,\tau}$ poorly behaved (e.g., very large Lipschitz constant of the gradient of $\phi_{p,\tau}$).

By replacing $\text{rank}(\boldsymbol{X}_i)$ in problem (9) with the smooth Schatten-p function, we obtain the Schatten-p approximation of the global-local low-rank matrix estimation formulation (9) as follows:

$$\min_{\boldsymbol{X} \in \mathcal{X}} \ \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \gamma_i \phi_{p,\tau}(\boldsymbol{X}_i), \tag{11}$$

where p and τ are given. We will be interested in the case of $p < 1$. The corresponding problem (11) is non-convex, but we found that, empirically, using $p < 1$ results in better recovery performance than using $p = 1$ (the smooth nuclear norm case).

4 Global-Local Low-Rank Algorithm

In this section, we develop an algorithm for the global-local low-rank matrix estimation formulation (9). Problem (9) is a large-scale optimization problem with ML optimization variables, and computational efficiency is a main concern in algorithm design. The algorithm to be presented is custom-designed for problem (9), where we exploit the problem structure for computational efficiency. The main optimization concepts used in our algorithm design are majorization-minimization (MM) and the accelerated projected gradient method.

4.1 Majorization-Minimization

Firstly, we describe the MM method. For notational convenience, let

$$f(\boldsymbol{X}) = \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \gamma_i \phi_{p,\tau}(\boldsymbol{X}_i). \tag{12}$$

In MM we seek a function $g(\boldsymbol{X}; \bar{\boldsymbol{X}})$, called a majorant of f, that satisfies

$$g(\boldsymbol{X}; \bar{\boldsymbol{X}}) \geq f(\boldsymbol{X}), \quad g(\bar{\boldsymbol{X}}; \bar{\boldsymbol{X}}) = f(\bar{\boldsymbol{X}}), \quad \forall \boldsymbol{X}, \bar{\boldsymbol{X}} \in \mathcal{X}.$$

We also require that, for any given $\bar{\boldsymbol{X}} \in \mathcal{X}$, $g(\cdot; \bar{\boldsymbol{X}})$ is convex and continuously differentiable. Given a starting point $\boldsymbol{X}^0 \in \mathcal{X}$, the MM method handles problem (9) by iteratively solving

$$\boldsymbol{X}^{k+1} = \arg\min_{\boldsymbol{X} \in \mathcal{X}} \; g(\boldsymbol{X}; \boldsymbol{X}^k), \quad k = 0, 1, \ldots \tag{13}$$

where the problem at each iteration in (13) is a convex problem. The MM iteration (13) is known to guarantee convergence to a stationary solution to problem (9) [38]. We identify a majorant for problem (9) by resorting to the following result.

Fact 1 *[27, Sec. 3.1 and Appendix A] For $0 < p \leq 1$, the smooth Schatten-p function admits an alternative characterization*

$$\phi_{p,\tau}(\boldsymbol{X}) = \arg\min_{\boldsymbol{W} \in \mathbb{S}^M_{++}} \; \psi_{p,\tau}(\boldsymbol{X}, \boldsymbol{W}), \tag{14}$$

where $\mathbb{S}^M_{++}$ denotes the set of all $M \times M$ symmetric and positive definite matrices;

$$\psi_{p,\tau}(\boldsymbol{X}, \boldsymbol{W}) = \tfrac{p}{2}\mathrm{tr}(\boldsymbol{W}(\boldsymbol{X}\boldsymbol{X}^T + \tau \boldsymbol{I})) + \tfrac{2-p}{p}\mathrm{tr}(\boldsymbol{W}^{\frac{p}{p-2}}).$$

Also, the minimum in (14) is uniquely attained at

$$\boldsymbol{W}^\star = (\boldsymbol{X}\boldsymbol{X}^T + \tau \boldsymbol{I})^{\frac{p}{2}-1}.$$

By applying Fact 1 to (12), and noting $\psi_{p,\tau}(\boldsymbol{X}, \boldsymbol{W}) \geq \psi_{p,\tau}(\boldsymbol{X}, \boldsymbol{W}^\star) = \phi_{p,\tau}(\boldsymbol{X})$ for any $\boldsymbol{W} \in \mathbb{S}^M_{++}$, we obtain the following majorant

$$g(\boldsymbol{X}; \boldsymbol{X}^k) = \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \gamma_i \psi_{p,\tau}(\boldsymbol{X}_i, \boldsymbol{W}_i^k),$$

where

$$\boldsymbol{W}_i^k = (\boldsymbol{X}_i^k (\boldsymbol{X}_i^k)^T + \tau \boldsymbol{I})^{\frac{p}{2}-1}, \quad i = 0, \ldots, P. \tag{15}$$

Note that computing $\boldsymbol{W}_i^k$ requires computing the eigendecomposition of $\boldsymbol{X}_i^k (\boldsymbol{X}_i^k)^T + \tau \boldsymbol{I}$, which takes $\mathcal{O}(M^3)$ floating point operations; cf. (10). It is easy to show that this g is convex and continuously differentiable. Also, solving the MM iteration $\min_{\boldsymbol{X} \in \mathcal{X}} g(\boldsymbol{X}; \boldsymbol{X}^k)$ in (13) is the same as solving

$$\min_{\boldsymbol{X} \in \mathcal{X}} \; \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \tfrac{p\gamma_i}{2}\mathrm{tr}(\boldsymbol{W}_i^k \boldsymbol{X}_i \boldsymbol{X}_i^T). \tag{16}$$

The above problem is a quadratically regularized least-squares, with an iteratively reweighted quadratic regularizer. Hence, the MM method developed above can be interpreted as an iteratively regularizer-reweighted least-squares method. In fact, if we remove the bound constraints $\boldsymbol{X} \in \mathcal{X}$ we can solve problem (16) in closed form. However, we would like to keep the bound constraints, and this leads to the development in the next subsection.

4.2 Accelerated Projected Gradient for Solving the MM Iteration

Secondly, we describe an iterative method for solving the MM iteration in (16). We employ the accelerated projected gradient (APG) method [39–41], which is a fast first-order algorithm. The APG method for solving problem (16) is as follows. Let

$$g_k(\boldsymbol{X}) = \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \tfrac{p\gamma_i}{2} \mathrm{tr}(\boldsymbol{W}_i^k \boldsymbol{X}_i \boldsymbol{X}_i^T) \tag{17}$$

for convenience. Given a starting point $\boldsymbol{X}^0 \in \mathcal{X}$, we perform the recursion

$$\boldsymbol{X}^{j+1} = \Pi_{\mathcal{X}} \left(\boldsymbol{Z}^j - \tfrac{1}{\beta_j} \nabla g_k(\boldsymbol{Z}^j) \right), \ j = 0, 1, \ldots \tag{18}$$

Here, $1/\beta_j > 0$ is the step size; $\nabla g_k(\boldsymbol{X})$ denotes the gradient of g_k at $\boldsymbol{X}$; $\Pi_{\mathcal{X}}(\boldsymbol{Y}) = \arg\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{Y} - \boldsymbol{X}\|_F^2$ denotes the projection onto $\mathcal{X}$; $\boldsymbol{Z}^j$ is the extrapolated point of the $\boldsymbol{X}^j$'s and is given by

$$\boldsymbol{Z}^j = \boldsymbol{X}^j + \alpha_j (\boldsymbol{X}^j - \boldsymbol{X}^{j-1});$$

where $\boldsymbol{X}^{-1} = \boldsymbol{X}^0$; the sequence $\{\alpha_j\}$, called the extrapolation sequence, is given by

$$\alpha_j = \frac{\xi_{j-1} - 1}{\xi_j}, \quad \xi_j = \frac{1 + \sqrt{1 + 4\xi_{j-1}^2}}{2}, \tag{19}$$

with $\xi_{-1} = 0$. The APG iteration (18) is efficient to implement. The gradient $\nabla g_k(\boldsymbol{X})$ is given by

$$\begin{aligned} &\nabla g_k(\boldsymbol{X}) \\ =&\nabla \ell(\boldsymbol{X}) + \sum_{i=0}^{P} \tfrac{p\gamma_i}{2} \nabla \left(\mathrm{tr}(\boldsymbol{W}_i^k \boldsymbol{X}_i \boldsymbol{X}_i^T) \right) \\ =&\nabla \ell(\boldsymbol{X}) + p(\gamma_0 \boldsymbol{W}_0^k \boldsymbol{X} + [\ \gamma_1 \boldsymbol{W}_1^k \boldsymbol{X}_1, \ldots, \gamma_P \boldsymbol{W}_P^k \boldsymbol{X}_P\ ]). \end{aligned} \tag{20}$$

Also, we have

$$\nabla \ell(\boldsymbol{X}) = \boldsymbol{F}^T \boldsymbol{F} \boldsymbol{X} - \boldsymbol{F}^T \boldsymbol{Y}_{\mathrm{M}} + \boldsymbol{X} \boldsymbol{G} \boldsymbol{G}^T - \boldsymbol{Y}_{\mathrm{H}} \boldsymbol{G}^T. \tag{21}$$

It can be verified that, given $\boldsymbol{W}_0^k, \ldots, \boldsymbol{W}_P^k$, computing (20)–(21) takes $\mathcal{O}(M(ML + \mathrm{nnz}(\boldsymbol{G})))$ floating-point operations, where $\mathrm{nnz}(\boldsymbol{G})$ denotes the number of nonzero elements of $\boldsymbol{G}$. We should note that a large number of elements of $\boldsymbol{G}$ are zero. Recall that $\boldsymbol{G}$ describes the spatial degradation process of local spatial blurring and downsampling. One can show that the number of nonzero elements of each column of $\boldsymbol{G}$ depends on the size of the blurring kernel, and in practice the blurring kernel size is often small. Also, the projection operation $\Pi_{\mathcal{X}}$ for the bound set $\mathcal{X} = [0, 1]^{M \times L}$ is merely a clipping function, i.e.,

$$\Pi_{\mathcal{X}}(\boldsymbol{Y}) = \max\{\boldsymbol{0}, \min\{\boldsymbol{1}, \boldsymbol{Y}\}\}, \tag{22}$$

where $\boldsymbol{0}$ and $\boldsymbol{1}$ denote all-zero and all-one matrices, resp., and min and max are taken in the element-wise manner.

We complete the APG development by specifying our step-size rule. The APG method guarantees convergence to the optimal solution if β_k is chosen to be a Lipschitz constant of ∇g_k [40,41]. We have the following result.

Fact 2 *A Lipschitz constant of* ∇g_k *in* (17) *is*

$$L_{g_k} = \lambda_{\max}(\boldsymbol{F}^T\boldsymbol{F} + p\gamma_0\boldsymbol{W}_0^k) + \lambda_{\max}(\boldsymbol{G}\boldsymbol{G}^T) + p\max_{i=1,\ldots,P}\gamma_i\lambda_{\max}(\boldsymbol{W}_i^k), \tag{23}$$

where $\lambda_{\max}(\cdot)$ *denotes the largest eigenvalue of the argument.*

The proof of Fact 2 is shown in Appendix A. The computational cost of (23) is mainly with computing the largest eigenvalues of $M\times M$ symmetric matrices, and there are $P+2$ such eigenvalues to compute. The eigenvalue $\lambda_{\max}(\boldsymbol{G}\boldsymbol{G}^T)$ can be computed offline, and the eigenvalues $\lambda_{\max}(\boldsymbol{W}_i^k)$ are byproducts of computing $\boldsymbol{W}_i^k$ in (15) (again, cf. (10)). It suffices to calculate $\lambda_{\max}(\boldsymbol{F}^T\boldsymbol{F}+p\gamma_0\boldsymbol{W}_0^k)$ at every MM iteration. Hence, computing (23) takes $\mathcal{O}(M^3)$ floating-point operations.

4.3 Algorithm Speedup via Inexact MM

Let us summarize the MM method developed in the last two subsections: We perform the MM iteration (13). Every iteration requires solving a regularized least-squares (with bound constraints) exactly, and we do that by applying the APG method in (18)–(22). The setback with this exact MM approach is that while the APG method is considered fast, it still takes time to solve each MM problem exactly.

The algorithm we finally adopt is an inexact MM method. At each MM iteration, we apply the APG method with one iteration only. Specifically, we replace the exact MM iteration (13) by

$$\boldsymbol{X}^{k+1} = \Pi_{\mathcal{X}}\left(\boldsymbol{Z}^k - \tfrac{1}{\beta_k}\nabla g_k(\boldsymbol{Z}^k)\right), \quad k = 0, 1, \ldots \tag{24a}$$

$$\boldsymbol{Z}^k = \boldsymbol{X}^k + \alpha_k(\boldsymbol{X}^k - \boldsymbol{X}^{k-1}), \tag{24b}$$

where α_k is given by (19); β_k is chosen as $\beta_k = 1/L_{g_k}$, with L_{g_k} given by (23). By using this inexact MM update we hope that the total number of iterations (i.e., the sum of APG iterations incurred by all the MM iterations) may be reduced, and the runtime improved. Based on our numerical experience, the inexact MM method runs much faster than the exact MM. Also, it is shown in [32] that, under some technical assumptions, the above inexact MM method guarantees convergence to a stationary solution.

We summarize the inexact MM algorithm in pseudo-form in Algorithm 1. We call this algorithm Global-Local lOw-Rank promotIng Algorithm (GLORIA). It can be verified that the complexity of GLORIA is $\mathcal{O}(M(ML + PM^2 + \text{nnz}(\boldsymbol{G})))$ per iteration.

Algorithm 1 GLORIA

input a starting point $\boldsymbol{X}^0$
$\boldsymbol{X}^{-1} = \boldsymbol{X}^0, \xi_0 = 0$;
$k = 0$;
repeat
  $\xi_{k+1} = \left(1 + \sqrt{1 + 4\xi_k}\right)/2$;
  $\alpha_k = (\xi_k - 1)/\xi_{k+1}$;
  $\boldsymbol{Z}^k = \boldsymbol{X}^k + \alpha_k(\boldsymbol{X}^k - \boldsymbol{X}^{k-1})$;
  $\nabla \ell(\boldsymbol{Z}^k) = \boldsymbol{F}^T\boldsymbol{F}\boldsymbol{Z}^k - \boldsymbol{F}^T\boldsymbol{Y}_{\mathrm{M}} + \boldsymbol{Z}^k\boldsymbol{G}\boldsymbol{G}^T - \boldsymbol{Y}_{\mathrm{H}}\boldsymbol{G}^T$;
  compute the eigendecomposition $\boldsymbol{U}_i\boldsymbol{\Lambda}_i\boldsymbol{U}_i^T := \boldsymbol{Z}_i^k(\boldsymbol{Z}_i^k)^T + \tau\boldsymbol{I}, \ \ i = 0, 1, \ldots, P$;
  $\boldsymbol{W}_i^k = \boldsymbol{U}_i\boldsymbol{\Lambda}_i^{\frac{p}{2}-1}\boldsymbol{U}_i^T, \ \ i = 0, 1, \ldots, P$;
  $\nabla g_k(\boldsymbol{Z}^k) =$
    $\nabla \ell(\boldsymbol{Z}^k) + p\gamma_0\boldsymbol{W}_0^k\boldsymbol{Z}_0^k + p[\ \gamma_1\boldsymbol{W}_1^k\boldsymbol{Z}_1^k, \ldots, \gamma_P\boldsymbol{W}_P^k\boldsymbol{Z}_P^k\ ]$;
  $L_{g_k} =$
  $\lambda_{\max}(\boldsymbol{F}^T\boldsymbol{F} + p\gamma_0\boldsymbol{W}_0^k) + \lambda_{\max}(\boldsymbol{G}\boldsymbol{G}^T) + p\max_{i=1,\ldots,P}\gamma_i\lambda_{M,i}^{\frac{p}{2}-1}$;
  $\tilde{\boldsymbol{X}}^{k+1} = \boldsymbol{Z}^k - (1/L_{g_k})\nabla g_k(\boldsymbol{Z}^k)$;
  $\boldsymbol{X}^{k+1} = \max\{\boldsymbol{0}, \min\{\boldsymbol{1}, \tilde{\boldsymbol{X}}^{k+1}\}\}$;
  $k := k + 1$;
until some stopping criterion is satisfied.
output $\boldsymbol{X}_{\mathrm{est}} = \boldsymbol{X}^k$.

5 Numerical Results

We performed extensive numerical experiments to benchmark GLORIA against a number of existing algorithms. The benchmarked algorithms we choose are considered most representative in the context or are related to our method. Namely, they are GSA [42], GLP [43], CNMF [4], FUMI [6], HySure [9], the nuclear norm minimization (NNM) method in (7), and LRSR [12]. GSA and GLP are pansharpening-based methods; CNMF and FUMI are representative methods in matrix factorization-based HSR; HySure is a representative method in dictionary-based regression, and it employs spatial total variation regularization in its regression; LRSR is another dictionary-based regression method which uses local nuclear-norm regularization; NNM is regarded as the baseline method for low-rank matrix estimation in the context of matrix completion, and thus we include it in our experiments. All the algorithms are implemented on a desktop computer with Intel Core i7-5760X@3GHz CPU and 32GB memory. Codes are written in MATLAB R2015a.

The parameter settings of GLORIA are as follows, unless specified otherwise. The parameters of the smooth Schatten-p function are $p = 1/2$, $\tau = 1$. The local patches are obtained by dividing the image into equal-spaced rectangular blocks, as in Fig. 2. The regularization parameters γ_i's are chosen to be identical; i.e., $\gamma_0 = \gamma_1 = \cdots = \gamma_P = \gamma$. We initialize the algorithm by using a $[0, 1]$-uniform i.i.d. generated $\boldsymbol{X}$. We stop the algorithm when the relative change of the objective function is below 10^{-5} or when the number of iterations exceeds 100.

The parameter settings of the benchmarked algorithms basically follow the recommended settings in [12, 44]. In addition, for matrix factorization and dictionary-based regression methods, we fix the target rank as $N = 30$. We use VCA to initialize the matrix factorization algorithms, and we stop the algorithms by the same stopping rule as that for GLORIA. Also, the NNM method is implemented by applying the APG method in [45] to problem (7).

The performance measures employed for evaluating the recovery performance are the peak SNR (PSNR), spectral angle mapper (SAM), Erreur Relative Globale Adimensionnelle de Sythèse

(ERGAS) and universal image quality index (UIQI). They have been extensively used in the HSR literature, and we refer the reader to [1] for their definitions.

5.1 Semi-Real Data Experiments

First, we consider semi-real data experiments. The experiments were based on the widely-used Wald's protocol [46], where we take a real HS image as the ground-truth SR image $\boldsymbol{X}$ and use it to generate the observed MS and HS images, $\boldsymbol{Y}_{\rm H}$ and $\boldsymbol{Y}_{\rm M}$, through the model (1). We consider the following real HS datasets.

1. *Hyperspec-VNIR-C Chikusei:* This dataset was acquired by the Headwall Hyperspec-VNIR-C imaging sensor [47]. It covers 128 spectral bands whose wavelength range is from 363nm to 1,018nm. We take a 480×480 subimage from this dataset as our SR image.

2. *AVIRIS Indian Pines:* This dataset was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) HS sensor [48]. The wavelength range is from 400nm to 2,500nm. It has 178 spectral bands after dropping bands that are corrupted by water absorption. In the experiment, a 120×120 subimage is used.

3. *HYDICE Washington DC Mall:* This dataset was taken by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) HS sensor [49]. We take a subimage of this dataset, which has 240×240 pixels and 191 clean spectral bands. The wavelength range is from 400nm to 2,500nm.

4. *ROSIS University of Pavia:* This dataset was measured by the Reflective Optics System Imaging Spectrometer (ROSIS) HS sensor. This dataset has 103 spectral bands whose wavelength range is from 430nm to 850nm. We take a 240×240 subimage from this dataset as the SR image.

5. *AVIRIS Moffett Field:* This dataset, recorded by AVIRIS HS sensor, has 187 uncorrupted spectral bands. The wavelength range is from 400nm to 2,500nm. We take a 240×240 subimage from this dataset.

These five images have been displayed in Fig. 3. The settings with the spectral and spatial measurement response matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ should also be described. The matrix $\boldsymbol{F}$ is chosen such that it is equivalent to the spectral response of either the Landsat 4 TM sensor [50] (6 bands, with spectral coverage from 400nm to 2,500nm) or the IKONOS sensor [51] (4 bands, with spectral coverage from 450nm to 900nm). We choose the Landsat 4 TM sensor response for Indian Pines, Washington DC Mall and Moffett Field, and the IKONOS sensor for Chikusei and University of Pavia; such choosing is to match the spectral coverage of the HS images. As discussed previously, $\boldsymbol{G}$ corresponds to the process of spatial blurring and downsampling. The blurring function is a 11×11 Gaussian point spread function, with variance 1.7^2. The downsampling is done every 4 pixels, both horizontally and vertically. Furthermore, the noise terms $\boldsymbol{V}_{\rm H}$ and $\boldsymbol{V}_{\rm M}$ are randomly generated following an i.i.d. mean-zero Gaussian distribution. We fix the SNRs at $\mathsf{SNR}_{\rm M} = \mathsf{SNR}_{\rm H} = 25$dB.

Table 2 summarizes some of the settings with the experiments. There, we also show the settings with the patch number P of GLORIA. The regularization parameter of GLORIA is fixed as $\gamma = 20/(\mathsf{SNR}_{\rm M} + \mathsf{SNR}_{\rm H})$. We ran 50 independent trials in each image. The obtained performance is shown in Table 3, where, for each performance measure, we use blue, brown and red boldfaced letters to mark the best, second best and third best algorithms. To give the reader an additional

reference on the performance, we show the SAM maps of the various algorithms in Figs. 4–8; note that the SAM maps shown are from one realization.

From Table 3 we see that, except for runtimes, GLORIA ranks best or second best in all of the performance measures. From Figs. 4–8, we also see that GLORIA yields good results compared to the other algorithms—and this is particularly so for the Indian Pine image in Fig. 5. In fact, we observe that even the NNM method, which is the baseline low-rank matrix estimation method and can be regarded as the precursor of our global-local low-rank pursuit, works reasonably. The above reported results suggest that the exploitation of low-rank spectral-spatial data structures in HSR is a working idea. We speculate that the good performance of GLORIA compared to the other algorithms is because GLORIA exploits the local low-rank data structure, which may provide better robustness to the EV effects. We will use synthetic experiments to examine the EV effects in the next subsection.

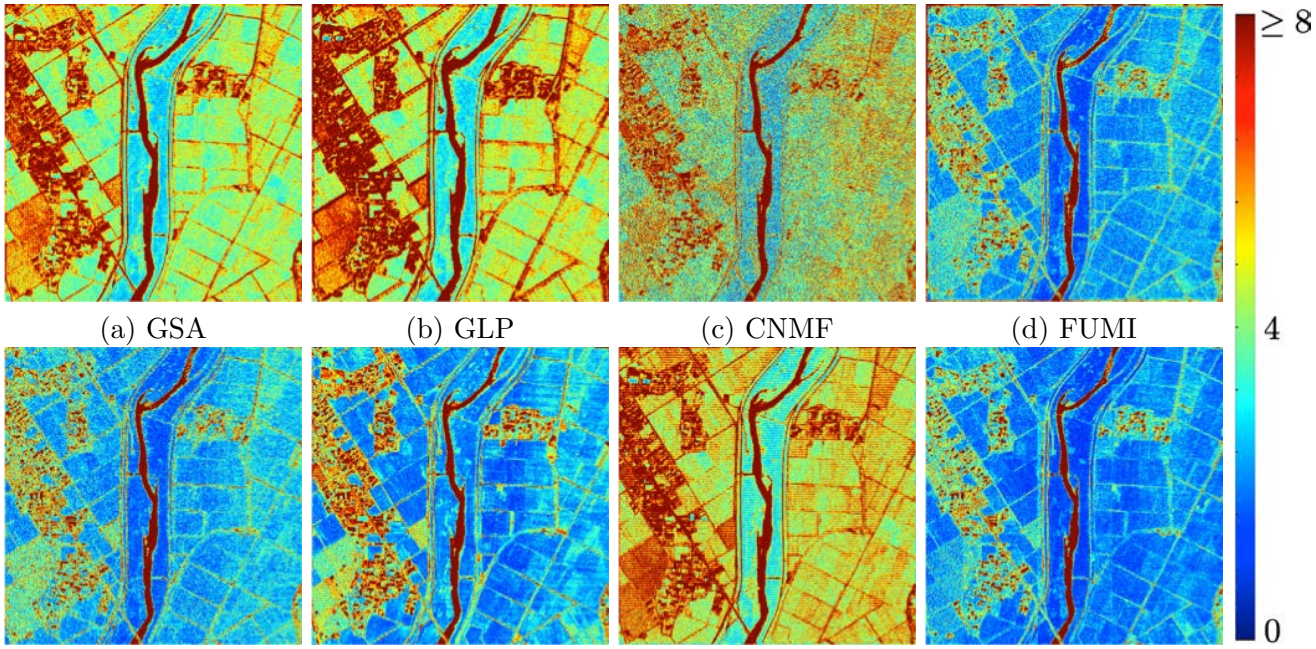
We should also discuss the runtime performance. The best algorithms are GSA and GLP, which do not perform very well in recovery performance. Let us compare GLORIA to the representative CNMF and HySure methods. GLORIA runs faster than CNMF and HySure for the cases of Chikusei and University of Pavia, and slower for the cases of Washington DC Mall and Moffett Field. On this issue, we should note that GLORIA deals with ML optimization variables, e.g., $29,491,200$ variables in Chikusei. In comparison, CNMF and HySure require $(M+L)N$ and NL optimization variables, resp., which amount to $6,915,840$ and $6,912,000$ variables, resp., in Chikusei. In terms of runtime per variable, GLORIA is considered efficient.

Table 2: Some settings with the semi-real data experiments.

	Chikusei	Indian Pines	Washinton DC Mall	University of Pavia	Moffett Field
MS response	IKONOS	Landsat 4 TM	Landsat 4 TM	IKONOS	Landsat 4 TM
Image size ($M \times L$)	128×230,400	178×14,400	191×57,600	103×57,600	187×57,600
Patch size ($M \times L_i$)	128×3,600	178×900	190×3,600	103×3,600	187×3,600
Patch number P	64	16	16	16	16

Table 3: Average performance of the algorithms on semi-real datasets.

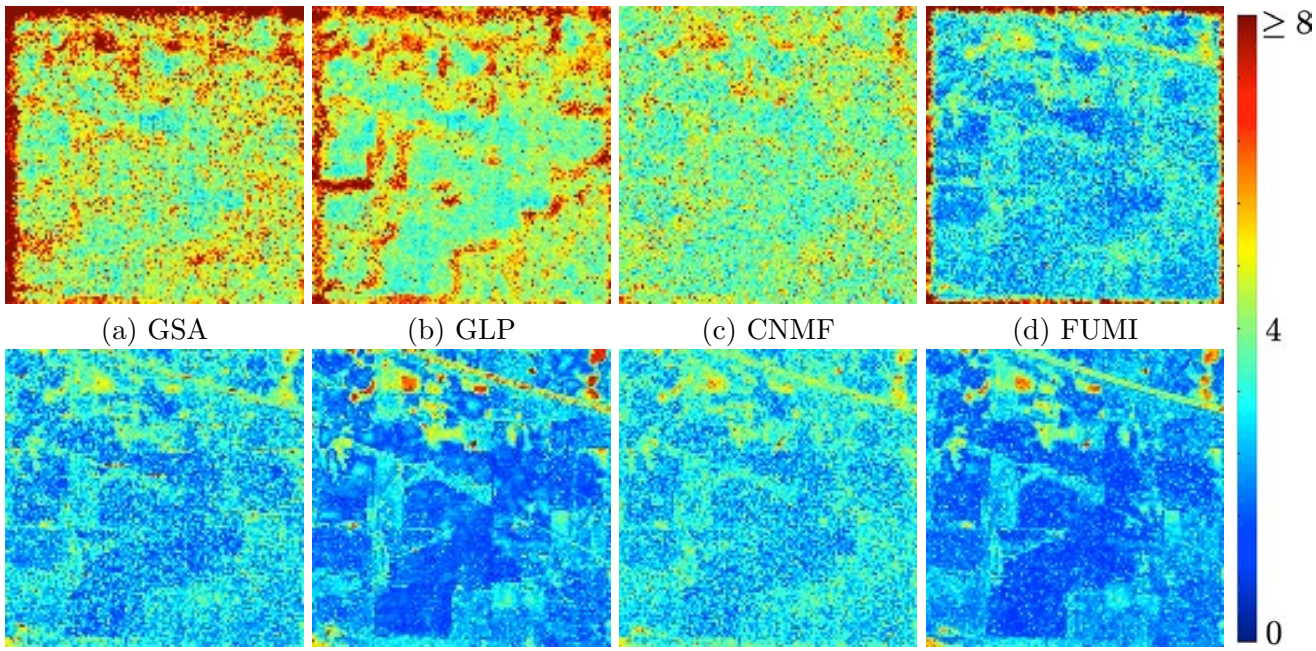
Method	Time (sec.)	PSNR	SAM	ERGAS	UIQI
Ideal value	0	∞	0	0	1
Dataset - Chikusei					
GSA	**1.44±0.15**	33.11±0.01	4.85±0.01	6.59±0.01	0.787±0.000
GLP	**43.44±8.84**	29.13±0.00	5.36±0.01	7.92±0.01	0.733±0.000
CNMF	147.54±64.93	34.13±0.11	4.21±0.08	5.24±0.11	0.811±0.004
FUMI	293.31±24.18	35.18±0.04	**2.55±0.03**	**4.03±0.03**	**0.882±0.001**
HySure	130.88±11.80	**36.15±0.15**	**2.84±0.09**	**4.31±0.08**	0.863±0.003
LRSR	279.26±43.04	**35.84±0.09**	2.88±0.03	4.41±0.06	**0.880±0.002**
NNM	**91.95±12.48**	32.63±0.01	4.96±0.00	7.49±0.01	0.764±0.000
GLORIA	119.99±21.78	**37.73±0.01**	**2.31±0.00**	**3.59±0.00**	**0.894±0.000**
Dataset - Indian Pines					
GSA	**0.44±0.08**	19.07±2.24	7.19±9.30	4.09±2.78	0.512±0.067
GLP	**2.38±0.51**	18.06±0.48	4.84±0.09	3.66±0.20	0.397±0.034
CNMF	16.57±4.04	22.64±0.22	4.29±0.10	2.43±0.07	0.523±0.009
FUMI	11.40±0.30	24.85±0.15	**2.69±0.05**	1.76±0.04	**0.762±0.007**
HySure	10.96±0.37	**26.70±0.16**	2.74±0.04	**1.41±0.03**	0.698±0.008
LRSR	18.80±0.28	**27.98±0.14**	**2.53±0.03**	**1.18±0.02**	**0.796±0.004**
NNM	**6.81±0.89**	26.43±0.04	2.93±0.01	1.71±0.01	0.679±0.002
GLORIA	14.56±0.92	**29.09±0.03**	**2.28±0.00**	**1.04±0.00**	**0.804±0.001**
Dataset - Washington DC Mall					
GSA	**0.97±0.09**	20.09±0.10	6.51±0.03	20.35±0.38	0.654±0.004
GLP	**7.68±1.83**	18.20±0.09	7.00±0.09	14.85±0.06	0.582±0.005
CNMF	31.74±16.95	25.71±0.19	3.84±0.08	6.94±0.22	0.772±0.007
FUMI	44.33±0.69	**29.17±0.10**	**2.49±0.02**	**2.91±0.06**	**0.915±0.004**
HySure	31.28±8.81	**28.82±0.33**	**3.00±0.07**	3.74±0.25	**0.872±0.009**
LRSR	71.92±1.55	27.91±0.16	3.85±0.10	**3.59±0.15**	0.869±0.007
NNM	**27.63±1.71**	27.57±0.12	3.28±0.02	8.05±0.03	0.813±0.005
GLORIA	57.06±8.14	**29.36±0.23**	**2.85±0.02**	**3.18±0.20**	**0.888±0.002**
Dataset - University of Pavia					
GSA	**0.49±0.05**	30.50±0.02	6.79±0.03	3.05±0.01	0.900±0.001
GLP	**4.45±1.14**	25.71±0.07	7.88±0.11	5.54±0.06	0.764±0.005
CNMF	27.08±21.05	35.86±0.18	4.09±0.09	1.84±0.04	0.946±0.002
FUMI	36.72±0.81	35.29±0.04	**3.12±0.01**	1.76±0.01	**0.962±0.000**
HySure	27.46±9.73	**36.71±0.14**	**3.34±0.02**	**1.55±0.01**	**0.961±0.001**
LRSR	58.29±1.20	36.43±0.09	3.45±0.05	1.65±0.03	0.959±0.001
NNM	**10.55±0.41**	**36.81±0.01**	3.48±0.00	**1.55±0.00**	0.957±0.000
GLORIA	16.68±0.50	**37.51±0.01**	**3.16±0.00**	**1.46±0.00**	**0.964±0.000**
Dataset - Moffett Field					
GSA	**0.87±0.06**	30.56±0.02	5.84±0.01	2.79±0.01	0.843±0.000
GLP	**6.50±0.70**	28.18±0.19	5.30±0.08	3.53±0.07	0.775±0.013
CNMF	**17.38±9.60**	36.08±0.12	3.19±0.06	1.57±0.03	0.919±0.002
FUMI	35.07±1.31	34.65±0.02	**2.48±0.02**	1.68±0.01	**0.946±0.000**
HySure	26.72±7.39	36.90±0.19	2.73±0.05	1.41±0.02	0.939±0.002
LRSR	55.78±0.69	**36.97±0.12**	2.58±0.04	**1.36±0.02**	**0.942±0.001**
NNM	22.79±3.52	**37.49±0.10**	**2.51±0.03**	**1.34±0.01**	0.940±0.001
GLORIA	44.38±5.46	**38.23±0.01**	**2.30±0.00**	**1.22±0.00**	**0.949±0.000**



(a) GSA (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

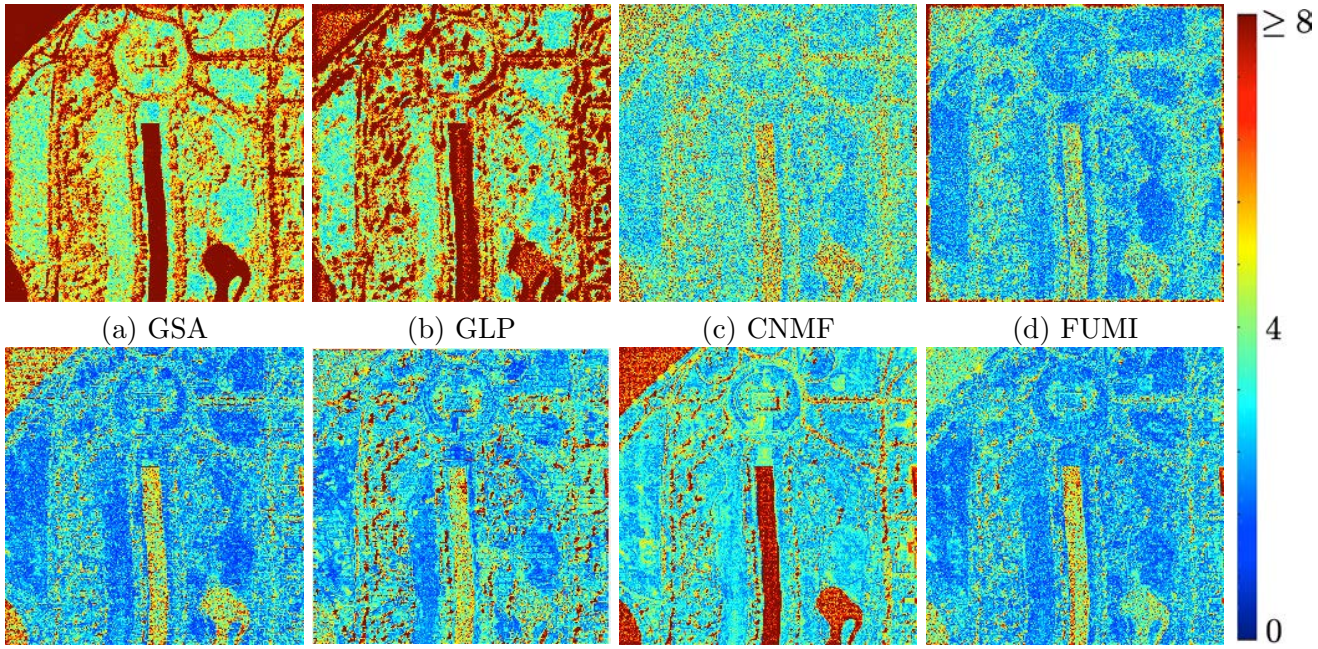
Figure 4: SAM maps of the Chikusei dataset.



(a) GSA (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

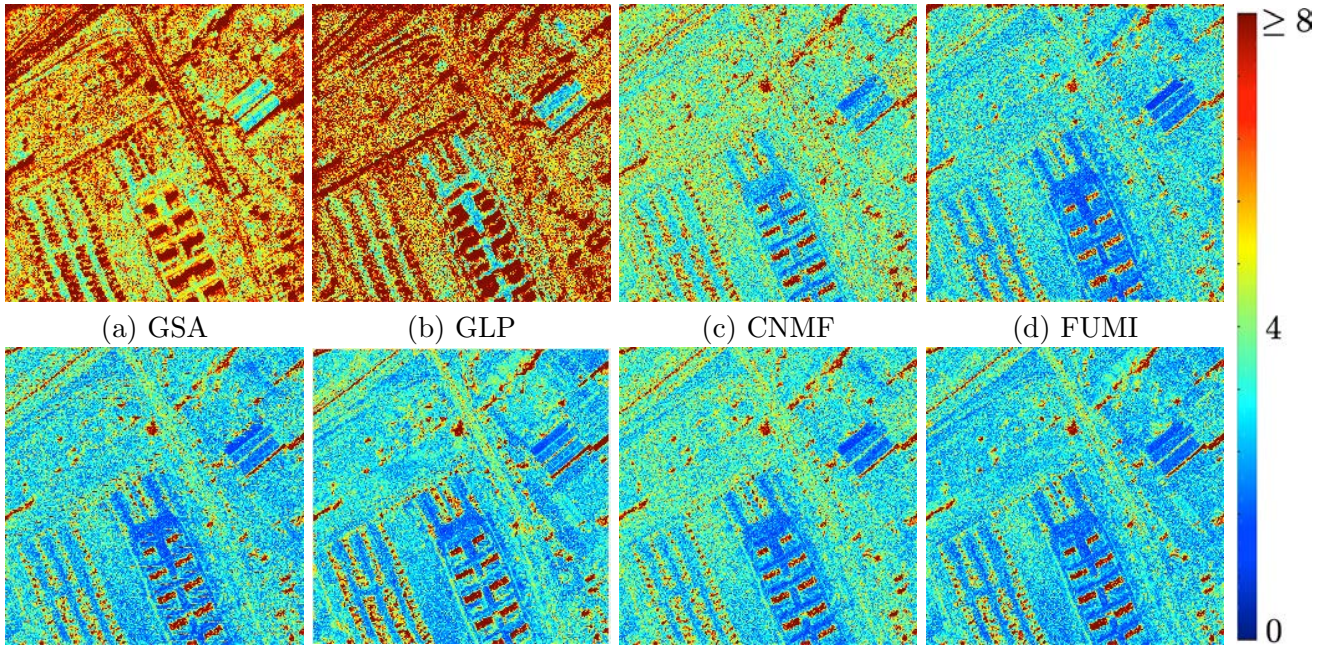
Figure 5: SAM maps of the Indian Pines dataset.



(a) GSA (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

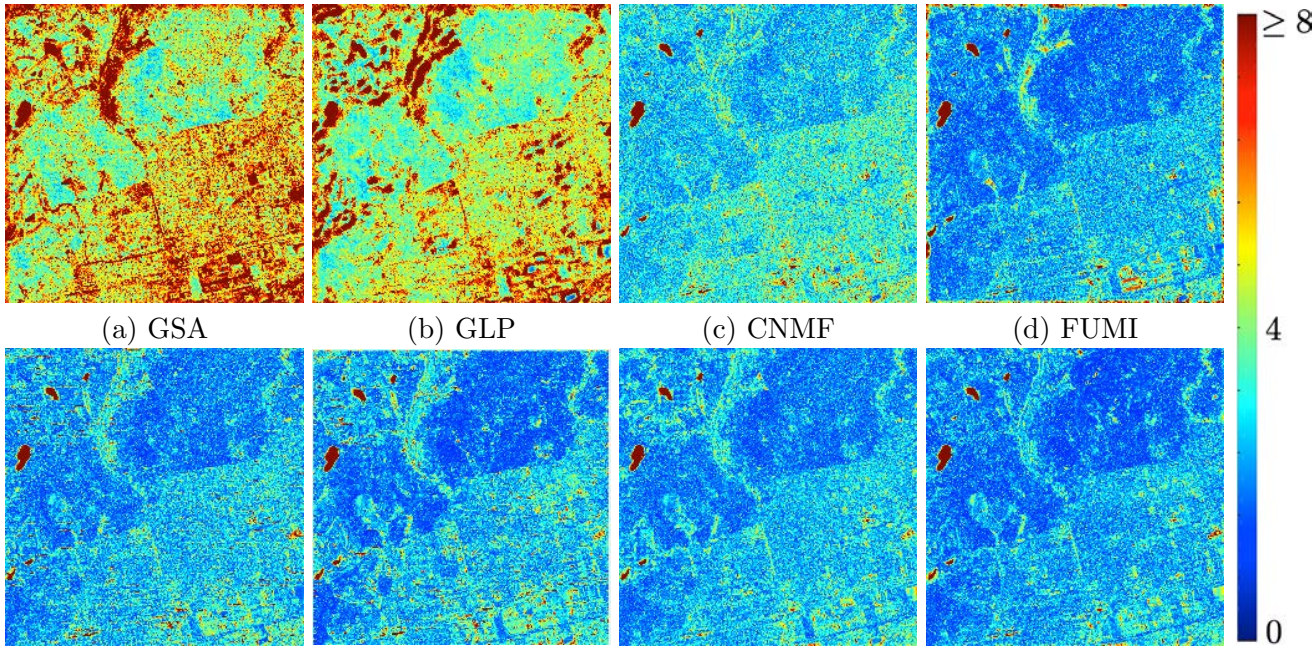
Figure 6: SAM maps of the algorithms, the Washington DC Mall dataset.



(a) GSA (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

Figure 7: SAM maps of the University of Pavia dataset.



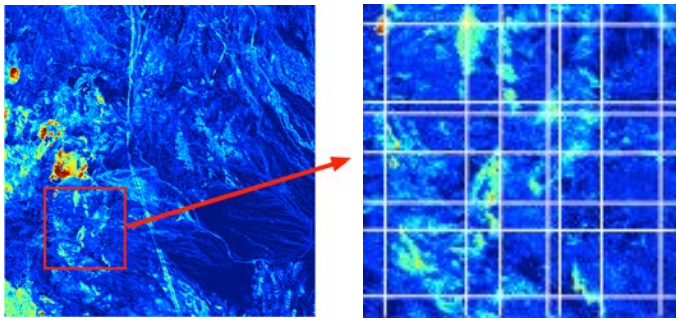
(a) GSA (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

Figure 8: SAM maps of the Moffett Field dataset.

5.2 Synthetic Data Experiments

Second, we consider synthetic data experiments. The way we prepare the experiment is similar to the one in [9,11], with the difference that EV is also involved. The procedure is described as follows. We use the local-patch-wise, and EV-present, linear spectral mixture model (8) to generate the SR image. The number of endmembers is $N = 5$. The generations of the local endmember matrix $\boldsymbol{A}_i$ and abundance matrix $\boldsymbol{S}_i$ will be described shortly. Each local patch is rectangular, but its horizontal and vertical lengths are, in each trial, random. Fig. 9(b) shows one such arrangement. We obtain $P = 64$ such patches, and it is important to note that none of the algorithms, including GLORIA, has knowledge of such patch arrangement. In GLORIA we will apply the same equi-spaced rectangular segmentation as before, and there will be mismatches between the actual patches and the patches presumed by GLORIA. Doing so is to provide a more realistic simulation as, in reality, it is impossible to exactly know how EV changes in space.



(a) (b)

Figure 9: Synthetic SR image generation: (a) abundance map of one material (estimated from the Cuprite dataset); (b) segmentation of the abundance map.

The endmember matrix $\boldsymbol{A}_i$ is chosen as a collection of the spectral signatures of five materials, specifically, Actinolite, Albite, Muscovite, Olivine and Topaz. To simulate the EV effect, for each patch and for each material we randomly pick one variation of that material from the U.S. geological survey (USGS) spectral library [52]. The abundance matrix $\boldsymbol{S}$ is chosen as the abundance maps extracted from a real HS image, namely, the AVIRIS Cuprite dataset; the extraction is done by applying an HU method called SVMAX [53]. In each trial, we randomly cropped a 120×120 submap from AVIRIS Cuprite; see the illustration in Fig. 9(a) where the submap is marked as a red rectangle. The abundance maps are then extracted from that submap.

Some other simulation settings are as follows. The spectral measurement response matrix corresponds to the spectral response of the Landsat 4 TM sensor. The regularization parameter of GLORIA is $\gamma = 40/(\mathsf{SNR}_{\mathrm{M}} + \mathsf{SNR}_{\mathrm{H}})$. We ran 100 independent trials. Table 4 shows the results, where again the best three algorithms are marked in blue, brown and red. As can be seen, GLORIA generally gives the best HSR recovery performance. This suggests that GLORIA has the flexibility to accommodate the EV effect. Also, GLORIA works better when the patch size is smaller, or when the number of patches P increases. Another observation is that at the lower SNR, i.e., 15dB, GLORIA works considerably better than the state-of-art methods.

Previously we mentioned that GLORIA, an inexact MM scheme using accelerated projected gradient, runs very fast. To give the reader some idea, we conduct the following runtime test. We benchmark GLORIA against the exact MM scheme and the inexact MM scheme via the nominal projected gradient (PG). The exact MM scheme solves each MM iteration exactly via the accelerated PG method (see Section 4.2). The inexact MM scheme via the nominal PG removes the extrapolation by setting $\alpha_k = 0$ in the inexact MM iteration (24). The previously described synthetic experiment is used to test the three MM schemes, and we consider SNR= 25dB and $P = 16$. The runtime results, shown in Table 5, indicate significant runtime advantages of GLORIA over the other two candidates.

Table 4: Average performance of the algorithms on the synthetic data.

$\mathsf{SNR}_\mathrm{M}/\mathsf{SNR}_\mathrm{H}$ - 15dB					
Method		PSNR	SAM	ERGAS	UIQI
Ideal value		∞	0	0	1
GSA		10.98±1.27	19.29±4.71	9.49±1.83	0.216±0.055
GLP		14.76±0.32	11.09±0.18	5.90±0.17	0.273±0.048
CNMF		15.73±0.41	10.52±0.52	5.43±0.22	0.299±0.050
FUMI		17.63±0.16	7.69±0.42	4.32±0.17	**0.436±0.055**
HySure		17.02±0.51	9.21±0.32	4.77±0.17	0.363±0.064
LRSR		**18.82±0.72**	**6.98±0.54**	**3.92±0.26**	0.434±0.074
NNM		17.29±0.50	8.88±0.20	4.43±0.10	0.372±0.065
GLORIA	$P=1$	18.46±0.23	7.90±0.10	4.00±0.06	0.432±0.060
	$P=9$	**21.46±0.12**	**5.03±0.13**	**3.00±0.07**	**0.583±0.051**
	$P=16$	**21.76±0.14**	**4.61±0.10**	**2.83±0.06**	**0.598±0.053**
$\mathsf{SNR}_\mathrm{M}/\mathsf{SNR}_\mathrm{H}$ - 25dB					
GSA		18.64±2.05	9.43±9.14	4.91±3.69	0.608±0.102
GLP		19.41±0.47	4.01±0.14	3.40±0.21	0.604±0.068
CNMF		24.71±0.35	3.89±0.20	1.97±0.10	0.729±0.048
FUMI		22.54±0.34	2.47±0.14	2.51±0.08	0.853±0.034
HySure		**31.33±0.26**	**1.26±0.09**	**0.90±0.04**	**0.923±0.021**
LRSR		30.63±0.22	**1.51±0.12**	1.01±0.05	**0.914±0.021**
NNM		27.30±0.44	2.78±0.05	1.42±0.03	0.820±0.045
GLORIA	$P=1$	28.99±0.21	2.21±0.08	1.25±0.03	0.870±0.029
	$P=9$	**30.70±0.22**	1.62±0.08	**1.00±0.03**	0.909±0.023
	$P=16$	**31.32±0.21**	**1.42±0.10**	**0.90±0.04**	**0.924±0.020**
$\mathsf{SNR}_\mathrm{M}/\mathsf{SNR}_\mathrm{H}$ - 35dB					
GSA		21.31±2.55	7.85±11.01	4.38±4.35	0.796±0.109
GLP		21.07±0.53	1.68±0.09	2.79±0.19	0.782±0.070
CNMF		32.25±0.35	1.76±0.12	0.89±0.06	0.925±0.020
FUMI		23.45±0.30	1.47±0.11	2.32±0.09	0.949±0.010
HySure		**35.93±0.42**	**1.02±0.10**	**0.55±0.04**	**0.977±0.006**
LRSR		34.89±0.40	1.03±0.11	0.63±0.04	**0.971±0.007**
NNM		34.87±0.32	1.19±0.02	0.59±0.01	0.959±0.014
GLORIA	$P=1$	35.48±0.51	1.09±0.08	0.58±0.05	0.966±0.012
	$P=9$	**36.18±0.22**	**1.01±0.03**	**0.52±0.01**	0.970±0.011
	$P=16$	**36.73±0.49**	**0.96±0.09**	**0.53±0.04**	**0.974±0.008**

Table 5: Runtime performance of various MM schemes.

	Exact MM	Inexact MM via nominal PG	Inexact MM via APG (GLORIA)
Time (sec.)	214.16±114.83	102.48±2.44	24.25±0.35

5.3 Real Data Experiments

Finally, we test the algorithms on real data. The dataset is the one used in [54]. The HS image was acquired by the Hyperion HS sensor. It covers a spectral range from 400nm to 2,500nm, and has 89 spectral bands after removing 131 noisy bands. The MS image was captured by the MS sensor mounted on the Sentinel-2A satellite. It has 13 bands, and we adopt 4 bands whose central wavelengths are 490nm, 560nm, 665nm and 842nm, resp. Readers are referred to [54] for further details. After pre-processing such as co-registration and cropping, we obtain the HS-MS image pair. The image pair is illustrated in Fig. 10. The setting is $(M, M_{\rm M}, L, L_{\rm H}, N) = (89, 4, 360^2, 120^2, 30)$.

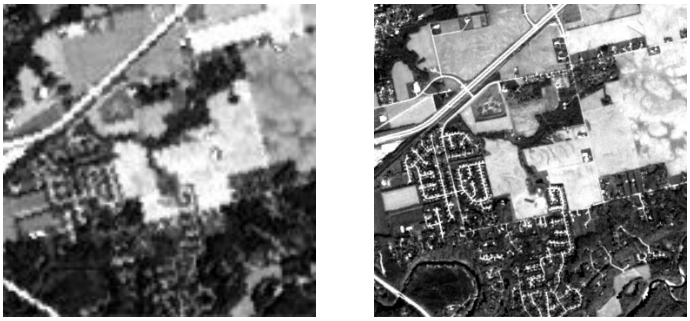


Figure 10: The tested HS-MS image pair: left) the 5th band of the HS image, right) the 1st band of the MS image

We employ the algorithm in [9] to estimate the spectral and spatial measurement responses $\boldsymbol{F}$ and $\boldsymbol{G}$. Empirically we found that, for the tested HS-MS pair, this algorithm happens to yield a poor estimate of $\boldsymbol{G}$. To mend this issue, we consider a second-stage estimation: we fix the blurring kernel to be a Gaussian kernel and estimate its standard deviation σ^2 by the nonlinear least-squares $\min_{\sigma^2>0} \|\boldsymbol{y}_{\rm M}^1 \circ g(\sigma^2) - \boldsymbol{f}^1 \boldsymbol{Y}_{\rm H}\|^2$, where $g(\sigma^2)$ denotes the Gaussian kernel with variance σ^2, $\circ$ corresponds to the 2-D convolution operation, and $\boldsymbol{y}_{\rm M}^1$ and $\boldsymbol{f}^1$ are the first row of $\boldsymbol{Y}_{\rm M}$ and the estimated $\boldsymbol{F}$, resp.

We follow the same simulation settings as those in semi-real experiments, except that we use $\gamma = 10$ for GLORIA. Figs. 11-12 illustrate the 5th and 20th bands of the original MS image and recovered images. From the figures, we note that GLP does not work well compared to the other algorithms. Moreover, if we zoom in the recovered images, we can see that the images recovered by HySure and LRSR have strip noise, while the images recovered by CNMF and NNM have pepper noise. In comparison, the images recovered by FUMI and GLORIA appear smoother.

(a) HS image (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

Figure 11: The 5th bands of the HS image and recovered images

(a) HS image (b) GLP (c) CNMF (d) FUMI

(e) HySure (f) LRSR (g) NNM (h) GLORIA

Figure 12: The 20th bands of the HS image and recovered images

6 Conclusion

In this paper we explored the route of low-rank matrix estimation for HSR. By positing a low-rank spectral-spatial data structure, both globally and locally, we built an algorithmic solution, called GLORIA, that exploits such structure for HSR. Our extensive numerical studies, which include semi-real data experiments on five different datasets, one synthetic experiment and one real experiment, show that exploiting global-local low-rank structure not only is a working idea but also provides satisfactory reconstruction results. Our global-local low-rank exploitation is made possible by customizing an efficient first-order strategy in large-scale structured optimization. We close this paper by naming a future direction. It would be interesting to study how low-rank matrix estimation would be useful in other hyperspectral problems such as multisource and multitemporal data fusion.

Appendix A

By definition, a constant L is said to be a Lipschitz constant of ∇g_k on $\mathcal{X}$ if $\|\nabla g_k(\boldsymbol{X})-\nabla g_k(\boldsymbol{Y})\|_F \leq L\|\boldsymbol{X}-\boldsymbol{Y}\|_F$ for any $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{X}$. From (20) we have

$$\begin{aligned}
&\|\nabla g_k(\boldsymbol{X})-\nabla g_k(\boldsymbol{Y})\|_F && \text{(25a)}\\
\leq&\|(\boldsymbol{F}^T\boldsymbol{F}+p\gamma_0\boldsymbol{W}_0^k)(\boldsymbol{X}-\boldsymbol{Y})\|_F+\|(\boldsymbol{X}-\boldsymbol{Y})\boldsymbol{G}\boldsymbol{G}^T\|_F\\
&\qquad+p\sum_{i=1}^{P}\|\gamma_i\boldsymbol{W}_i^k(\boldsymbol{X}_i-\boldsymbol{Y}_i)\|_F && \text{(25b)}\\
\leq&\Big(\lambda_{\max}(\boldsymbol{F}^T\boldsymbol{F}+p\gamma_0\boldsymbol{W}_0^k)+\lambda_{\max}(\boldsymbol{G}\boldsymbol{G}^T)\\
&\qquad+p\max_{i=1,\ldots,P}\gamma_i\lambda_{\max}(\boldsymbol{W}_i^k)\Big)\|\boldsymbol{X}-\boldsymbol{Y}\|_F. && \text{(25c)}
\end{aligned}$$

for any $\boldsymbol{X},\boldsymbol{Y}\in\mathbb{R}^{M\times L}$, where (25b) is due to the triangle inequality; (25c) is due to i) the inequality $\|\boldsymbol{A}\boldsymbol{B}\|_F\leq\|\boldsymbol{A}\|_2\|\boldsymbol{B}\|_F$ in which $\|\boldsymbol{A}\|_2$ denotes the spectral norm of $\boldsymbol{A}$, and ii) the identity $\|\boldsymbol{A}\|_2=\lambda_{\max}(\boldsymbol{A})$ for positive semidefinite $\boldsymbol{A}$. The proof of Fact 2 is complete.

References

[1] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, 2015.

[2] 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, 2013.

[3] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, 2014.

[4] 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, 2012.

[5] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2013, pp. 1409–1413.

[6] Q. Wei, J. 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, 2016.

[7] R. Wu, H.-T. Wai, and W.-K. Ma, "Hybrid inexact BCD for coupled structured matrix factorization in hyperspectral super-resolution," to appear in *IEEE Trans. Signal Process.*, 2020.

[8] R. Dian, S. Li, L. Fang, and Q. Wei, "Multispectral and hyperspectral image fusion with spatial-spectral sparse representation," *Inf. Fusion*, vol. 49, pp. 262–270, 2019.

[9] M. Simões, J. 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, 2015.

[10] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.

[11] M. A. Veganzones, M. Simões, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 274–288, 2016.

[12] R. Dian, S. Li, L. Fang, and J. Bioucas-Dias, "Hyperspectral image super-resolution via local low-rank and sparse representations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2018, pp. 4003–4006.

[13] Q. Li, W.-K. Ma, and Q. Wu, "Hyperspectral super-resolution: Exact recovery in polynomial time," in *IEEE Workshop Stat. Signal Process.*, 2018, pp. 378–382.

[14] H. Liu, R. Wu, and W.-K. Ma, "Is there any recovery guarantee with coupled structured matrix factorization for hyperspectral super-resolution?" in *Proc. IEEE Int. Workshop Comput. Adv. Mult.-Sens. Adapt. Process. (CAMSAP)*, 2019, available online at https://arxiv.org/pdf/1907.12728.pdf.

[15] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, 2018.

[16] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, 2018.

[17] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Coupled tensor low-rank multilinear approximation for hyperspectral super-resolution," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2019, pp. 5536–5540.

[18] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 30, no. 9, pp. 2672–2683, 2019.

[19] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multirank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, 2019.

[20] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, pp. 1–12, 2019.

[21] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, 2017.

[22] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 305–319, 2018.

[23] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn Syst.*, no. 99, pp. 1–11, 2018.

[24] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[25] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6535–6579, 2016.

[26] Y. Chen and Y. Chi, "Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 14–31, 2018.

[27] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *J. Mach. Learn. Res.*, vol. 13, no. Nov, pp. 3441–3473, 2012.

[28] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, 2013.

[29] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4904–4917, 2015.

[30] 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, 2015.

[31] L. Drumetz, J. Chanussot, C. Jutten, W.-K. Ma, and A. Iwasaki, "Spectral variability aware blind hyperspectral image unmixing based on convex geometry," *IEEE Trans. Image Process.*, vol. 29, pp. 4568–4582, 2020.

[32] M. Shao, Q. Li, W.-K. Ma, and A. M.-C. So, "A framework for one-bit and constant-envelope precoding over multiuser massive MISO channels," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5309–5324, 2019.

[33] Y. Zhou, L. Feng, C. Hou, and S.-Y. Kung, "Hyperspectral and multispectral image fusion based on local low rank and coupled spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5997–6009, 2017.

[34] Y. Qu, W. Wang, R. Guo, B. Ayhan, C. Kwan, S. Vance, and H. Qi, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, 2018.

[35] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 419–426, 2013.

[36] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pac. J. Optim.*, vol. 6, no. 615-640, p. 15, 2010.

[37] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

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

[39] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $\mathcal{O}(1/k^2)$," in *Doklady AN USSR*, vol. 269, 1983, pp. 543–547.

[40] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[41] A. Beck, *First-Order Methods in Optimization*. MOS-SIAM Series on Optimization, SIAM, 2017.

[42] 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, 2007.

[43] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sensing*, vol. 72, no. 5, pp. 591–596, 2006.

[44] 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, 2017.

[45] R. Wu, Q. Li, X. Fu, and W.-K. Ma, "A convex low-rank regularization method for hyperspectral super-resolution," in *IEEE Workshop Stat. Signal Process.*, 2018, pp. 383–387.

[46] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Jan. 2000, pp. 99–105.

[47] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Space Application Laboratory, University of Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016.

[48] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 44, no. 2-3, pp. 127–143, 1993.

[49] R. W. Basedow, D. C. Carmer, and M. E. Anderson, "HYDICE system: Implementation and performance," in *Proc. SPIEs 1995 Symp. OE/Aerospace Sens. Dual Use Photonics*, vol. 2480, 1995, pp. 258–268.

[50] G. Chander, B. L. Markham, and D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors," *Remote Sens. Environ.*, vol. 113, no. 5, pp. 893–903, 2009.

[51] G. Dial, H. Bowen, F. Gerlach, J. Grodecki, and R. Oleszczuk, "IKONOS satellite, imagery, and products," *Remote Sens. Environ.*, vol. 88, no. 1-2, pp. 23–36, 2003.

[52] R. F. Kokaly, R. N. Clark, G. A. Swayze, K. E. Livo, T. M. Hoefen, N. C. Pearson, R. A. Wise, W. M. Benzel, H. A. Lowers, R. L. Driscoll *et al.*, "USGS spectral library version 7," US Geological Survey, Tech. Rep., 2017.

[53] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, 2011.

[54] J. Yang, Y.-Q. Zhao, and J. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, p. 800, 2018.