Abstract—Hyperspectral super-resolution (HSR) is a problem that aims to estimate an image of high spectral and spatial resolutions from a pair of coregistered multispectral (MS) and hyperspectral (HS) images, which have coarser spectral and spatial resolutions, respectively. In this article, 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 nonconvex 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.

Index Terms—Endmember variability (EV), hyperspectral (HS), hyperspectral super-resolution (HSR), low-rank matrix estimation.

I. 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 a 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 adopt pansharpening techniques in a 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. In addition, 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 article, we taxonomize the existing low-dimensional model-based HSR methods into two types.

1) Matrix Factorization: This type of method 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 nonnegativity 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], nonlocal pixel similarity [8], and many more—to attempt to improve recovery quality.

2) Dictionary-Based Regression: This type of method 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 in 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 2-D 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 III and IV. 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 the 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 (NNM) 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. Nonconvex 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 EV [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 developments as of the writing of this article.

In this article, our objective is to explore the potential of low-rank matrix estimation in HSR. This article is not a direct application of the existing low-rank matrix estimation methods, such as the NNM 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 a 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 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 nonconvex large-scale problem. For example, to recover an SR image of 100 spectral bands and of size \(200 \times 200\), 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 nonconvex 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 nonconvex, 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 10w-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 data sets, and GLORIA consistently ranks first or second in performance indicators such as peak SNR (PSNR) and spectral angle mappers (SAMs). 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 not 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] and [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 nonconvex optimization problem. We custom-develop an efficient algorithm for such purposes, using recent advances in large-scale non-convex optimization. As will be shown in numerical experiments, our algorithm has a competitive recovery performance.

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

II. BACKGROUND

A. 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 of the MS. For convenience, the image we seek to construct will be called the SR image. As a common assumption (see [4]), the HS image is modeled as a spatially degraded version of the SR image by means of spatial blurring and downsampling. Also, the MS image is modeled as a spectrally degraded version of the SR image by means of spectral bandpass averaging.

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

\[ Y_M = FX + V_M \]

\[ Y_H = XG + V_H \]

where \( Y_M \in \mathbb{R}^{M_M \times L} \) and \( Y_H \in \mathbb{R}^{M_L \times L} \) are the spectral–spatial matrices of the observed MS and HS images, respectively (resp.); \( M_M \) and \( M_L \) are the numbers of spectral bands of the MS and HS images, resp., with \( M_M < M_L \); \( L \) and \( L_H \) are the numbers of pixels of the MS and HS images, resp., with \( L_H < L \); \( X \in \mathbb{R}^{M_L \times L} \) is the spectral–spatial matrix of the SR image; \( F \in \mathbb{R}^{M_M \times M_L} \) and \( G \in \mathbb{R}^{L \times L_H} \) describe the measurement responses that lead to the MS and HS images, resp.; \( V_M \) and \( V_H \) are noise. Note that \( F \) designates the relative spectral bandpass responses from the SR image to the MS image, while \( G \) specifies the spatial blurring and downsampling responses that result in the HS image. The measurement response matrices \( F \) and \( G \) are assumed to be known, and in practice \( F \) and \( G \) 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 \( i, j \).

B. 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 \( X \) can be factored as

\[ X = AS \]

where \( A \in \mathbb{R}^{M \times N} \) is the spectral dictionary; \( S \in \mathbb{R}^{N \times L} \) is the coefficient matrix; \( N \) is the target rank or model order, which is prefixed 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 \( a_i \)'s of \( A \) are interpreted as spectral signatures of the endmembers of the
scene, and the columns $s_i$’s of $S$ the associated abundances of the pixels. Under the model (2), the matrix factorization methods seek to find $(A, S)$ from $(Y_M, Y_H)$ by minimizing a data-fitting loss function, and thereby reconstruct $X$ by $X = AS$. For example, in CNMF [4], the idea is to solve

$$\min_{A \geq 0, S \geq 0} \|Y_M - FAS\|_F^2 + \|Y_H - ASG\|_F^2$$

(3)

where $X \geq 0$ means that $X$ is elementwise nonnegative; $\| \cdot \|_F$ 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, methods such as total variation may be added for problem structure. Moreover, in the dictionary-based regression methods, one first determines $A$ from $Y_H$ by methods such as principal component analysis (PCA) or VCA. Then, $S$ is estimated by solving the data-fitting problem like the one in (3), but with $A$ fixed. In estimating $S$, 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 $X$ is a high-dimensional matrix, we hold the belief that every spectral pixel $x_i$ of $X$ should lie in a low-dimensional subspace spanned by $a_1, \ldots, a_N$.

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

III. GLOBAL–LOCAL LOW-RANK FORMULATION

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

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 $X$ to be a low-rank matrix. This assumption can be interpreted as requiring the columns $x_1, \ldots, x_L$ to lie in a low-dimensional subspace. The matrix factorization model (2) can also be seen as constraining $x_1, \ldots, x_L$ 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 $X$ whose data fitting loss is small. A common low-rank matrix estimation formulation is as follows:

$$\min_{X \in \mathbb{R}^{M \times L}} \ell(X) + \gamma \text{rank}(X)$$

(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

$$\ell(X) = \frac{1}{2}\|Y_M - FX\|_F^2 + \frac{1}{2}\|Y_H - XG\|_F^2.$$ 

(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 predetermine the target rank $N$. There is no such rank constraint in low-rank matrix estimation. The rank of $X$ itself serves as the regularization for the low-rank matrix recovery endeavor, and the parameter $\gamma$ 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 nonconvex and nondifferentiable. The state of the art handles this issue by applying the nuclear norm approximation. The reader is referred to the literature [26] for a detailed description of the concept, and here we concisely explain the idea. The rank of $X$ is identical to the number of nonzero singular values of $X$, and hence we can express $\text{rank}(X)$ as

$$\text{rank}(X) = \min_{\ell(M, L)} \sum_{i=1}^{\min\{M, L\}} u(\sigma_i(X))$$

(6)

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

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

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:

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

(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].

B. 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 $X$. 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 $X$ in the global sense (with a higher global rank than the local ranks). Alternatively speaking, the low-dimensional subspace of the whole $X$ plays the role of tying together the low-dimensional subspaces of the local patches.

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

$$X = [X_1 \ X_2 \ \cdots \ X_P]$$

where each $X_i \in \mathbb{R}^{M \times L_i}$ 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, for example, via segmentation [11], may also be considered. We assume that every $X_i$ is a low-rank matrix, and $X$ 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 $X_i$ to follow the linear spectral mixture model:

$$X_i = A_i S_i$$  \hspace{1cm} (8)

where $A_i \in \mathbb{R}^{M \times N}$ and $S_i \in \mathbb{R}^{N \times L_i}$ are the endmember and abundance matrices of patch $i$; $N$ is the total number of endmembers in the whole SR image $X$. In this model, we assume the presence of EV by allowing the endmember matrix $A_i$ 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 $A_1, \ldots, A_P$ are correlated, since, in reality, they should be variations of one another. Such a correlation would mean that $A_1, \ldots, A_P$ can be linearly represented by a “global” basis $B$, whose dimension is not less than $N$, but not significantly greater than $N$ owing to the correlations. This further means that $X$ lives in a low-dimensional subspace with $B$ as its basis. Our global low-rank assumption is to exploit the global low-dimensional structure.

Additionally, in the abovementioned motivating model example, at first sight, one would be tempted to say that the rank of $X_i$ is universally given by rank($X_i$) = $N$ (under the slightly technical premise that $A_i$ and $S_i$ have full column rank and full row rank, resp.). In fact, it is reasonable to assume nonidentical rank($X_i$). 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 $s_{i1}^{j}, \ldots, s_{iN}^{j}$ of $S_i$, only $N_i$ of them are nonzero (or active). Consequently, we have rank($X_i$) = $N_i$ (again, under the technical premise that $A_i$ has a full column rank and that the nonzero $s_i^{j}$’s are linearly independent), which is our local low-rank assumption.

### C. 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 data sets, 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 $X_i$’s by the equal-space rectangular segmentation in Fig. 2. For each local patch $X_i$, we evaluate its rank in an approximate manner, specifically, by finding the smallest integer $r$ such that

$$\frac{\sum_{j=1}^{r} \sigma_j(X_i)^2}{\sum_{j=1}^{\min(M,L_i)} \sigma_j(X_i)^2} \geq 0.9999.$$
where $U$ and $\Lambda$ constitute the eigen-decomposition $A = U \Lambda U^T$; note that $U$ is orthogonal, $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \ldots \geq \lambda_n > 0$, and $A^p = \text{Diag}(\lambda_1^p, \ldots, \lambda_n^p)$. The smooth Schatten-$p$ function of an $M \times L$ matrix $X$, with $M \leq L$, is defined as
\[
\phi_{p,\tau}(X) = \sum_{i=1}^{M} (\sigma_i(X)^2 + \tau)^{p/2} = \text{tr}((XX^T + \tau I)^{p/2})
\]
where $p > 0$, $\tau > 0$ are given. This function has the following properties.

1) $\phi_{p,\tau}$ is smooth (or has derivatives of all orders).
2) $\phi_{p,\tau}$ is convex for $p \geq 1$, and nonconvex for $p < 1$.
3) As $\tau \to 0$, $\phi_{1,\tau}(X) \to \|X\|_*$. 
4) As $p \to 0$, $\tau \to 0$, $\phi_{p,\tau}(X) \to \text{rank}(X)$.

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 $\tau$ so that $\phi_{p,\tau}(X)$ closely approximates rank($X$), but using very small $p$ and $\tau$ will also make $\phi_{p,\tau}$ poorly behaved (e.g., very large Lipschitz constant of the gradient of $\phi_{p,\tau}$). By replacing rank($X$) 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:
\[
\min_{X \in \mathcal{X}} \ell(X) + \sum_{i=0}^{p} \gamma_i \phi_{p,\tau}(X_i) \quad (11)
\]
where $p$ and $\tau$ are given. We will be interested in the case of $p < 1$. The corresponding problem (11) is nonconvex, but we found that, empirically, using $p < 1$ results in better recovery performance than using $p = 1$ (the smooth nuclear norm case).

IV. 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 MM and the accelerated PG method.

A. MM

First, we describe the MM method. For notational convenience, let
\[
f(X) = \ell(X) + \sum_{i=0}^{p} \gamma_i \phi_{p,\tau}(X_i) \quad (12)
\]
In MM, we seek a function $g(X; \tilde{X})$, called a majorant of $f$, that satisfies
\[
g(X; \tilde{X}) \geq f(X), \quad g(\tilde{X}; \tilde{X}) = f(\tilde{X}), \quad \forall X, \tilde{X} \in \mathcal{X}.
\]

| $P$ | Chikusei | Cuprite |
|-----|---------|---------|
| patch size | (480 x 480 pixels, 128 bands) | (480 x 480 pixels, 187 bands) |
| $1^2$ | 128 x 230, 400 | 10 | 187 x 230, 400 | 9 |
| $2^2$ | 128 x 57, 600 | 8.75 ± 0.96 | 187 x 57, 600 | 8.25 ± 1.26 |
| $3^2$ | 128 x 25, 600 | 8.33 ± 1.73 | 187 x 25, 600 | 7.44 ± 3.33 |
| $4^2$ | 128 x 14, 400 | 8.19 ± 1.83 | 187 x 14, 400 | 7.13 ± 1.20 |
| $5^2$ | 128 x 9, 216 | 7.92 ± 1.89 | 187 x 9, 216 | 6.76 ± 1.13 |
| $6^2$ | 128 x 6, 400 | 7.81 ± 1.98 | 187 x 6, 400 | 6.64 ± 1.13 |
| $8^2$ | 128 x 3, 600 | 7.38 ± 2.13 | 187 x 3, 600 | 6.38 ± 1.06 |
| $10^2$ | 128 x 2, 304 | 7.08 ± 2.29 | 187 x 2, 304 | 6.12 ± 1.09 |
| $12^2$ | 128 x 1, 600 | 6.89 ± 2.34 | 187 x 1, 600 | 6.03 ± 1.02 |
| $15^2$ | 128 x 1, 024 | 6.52 ± 2.35 | 187 x 1, 024 | 5.84 ± 1.00 |
| $16^2$ | 128 x 900 | 6.47 ± 2.36 | 187 x 900 | 5.79 ± 1.01 |
| $P$ | Indian Pine | University of Pavia |
| patch size | (120 x 120 pixels, 178 bands) | (240 x 240 pixels, 103 bands) |
| $1^2$ | 178 x 14, 400 | 22 | 103 x 57, 600 | 22 |
| $2^2$ | 178 x 3, 600 | 19.75 ± 1.71 | 103 x 14, 400 | 20.75 ± 1.89 |
| $3^2$ | 178 x 1, 600 | 17.36 ± 1.40 | 103 x 6, 400 | 20.44 ± 2.07 |
| $4^2$ | 178 x 900 | 16.63 ± 3.77 | 103 x 3, 600 | 20.46 ± 2.66 |
| $5^2$ | 178 x 576 | 15.20 ± 3.94 | 103 x 3, 204 | 20.28 ± 2.84 |
| $6^2$ | 178 x 400 | 14.69 ± 4.44 | 103 x 1, 600 | 20.03 ± 3.08 |
| $8^2$ | 178 x 255 | 13.36 ± 4.47 | 103 x 900 | 19.80 ± 3.43 |
| $P$ | Washington DC Mall | Moffett Field |
| patch size | (240 x 240 pixels, 191 bands) | (240 x 240 pixels, 187 bands) |
| $1^2$ | 191 x 57, 600 | 6 | 187 x 57, 600 | 13 |
| $2^2$ | 191 x 14, 400 | 5.50 ± 0.58 | 187 x 14, 400 | 12.00 ± 2.71 |
| $3^2$ | 191 x 6, 400 | 5.22 ± 0.67 | 187 x 6, 400 | 11.66 ± 2.65 |
| $4^2$ | 191 x 3, 600 | 4.50 ± 1.03 | 187 x 3, 600 | 11.00 ± 2.94 |
| $5^2$ | 191 x 2, 304 | 4.52 ± 1.26 | 187 x 2, 304 | 10.68 ± 2.87 |
| $6^2$ | 191 x 1, 600 | 4.39 ± 1.23 | 187 x 1, 600 | 10.56 ± 2.92 |
| $8^2$ | 191 x 900 | 4.20 ± 1.17 | 187 x 900 | 10.08 ± 2.89 |

Table I 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.

D. 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:
\[
\min_{X \in \mathcal{X}} \ell(X) + \sum_{i=0}^{p} \gamma_i \text{rank}(X_i) \quad (9)
\]
where $\gamma_0, \gamma_1, \ldots, \gamma_p > 0$ are given regularization parameters; we denote $X_0 = X$ for notational convenience; $\ell$ 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 rank($X_i$) by the nuclear norm $\|X_i\|_*$. Here, we pursue a different option, namely, 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 $A$ and a number $p$, we define
\[
A^p = U \Lambda^p U^T \quad (10)
\]
We also require that, for any given \( \tilde{X} \in \mathcal{X} \), \( g(\cdot; \tilde{X}) \) is convex and continuously differentiable. Given a starting point \( X^0 \in \mathcal{X} \), the MM method handles problem (9) by iteratively solving

\[
X^{k+1} = \arg \min_{X \in \mathcal{X}} g(X; X^k), \quad k = 0, 1, \ldots
\]

(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, Section III.1 and Appendix]: For \( 0 < p \leq 1 \), the smooth Schatten-\( p \) function admits an alternative characterization

\[
\phi_{p,r}(X) = \arg \min_{W \in \mathbb{S}^M_{++}} \psi_{p,r}(X, W)
\]

(14)

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

\[
\psi_{p,r}(X, W) = \frac{2}{p} \text{tr}(W(XX^T + r I)) + \frac{2-r}{p} \text{tr}(W^{-1}).
\]

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

\[
W^* = (XX^T + r I)^{-\frac{r-1}{2}}.
\]

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

\[
g(X; X^k) = \ell(X) + \sum_{i=0}^p \gamma_i \psi_{p,r}(X_i, W_i^k)
\]

where

\[
W_i^k = \left( X_i^k (X_i^k)^T + r I \right)^{-\frac{r-1}{2}}, \quad i = 0, \ldots, P.
\]

(15)

Note that computing \( W_i^k \) requires computing the eigendecomposition of \( X_i^k (X_i^k)^T + r I \), which takes \( O(M^3) \) floating point operations [see (10)]. It is easy to show that \( g \) is convex and continuously differentiable. Also, solving the MM iteration \( \min_{X \in \mathcal{X}} g(X; X^k) \) in (13) is the same as solving

\[
\min_{X \in \mathcal{X}} \ell(X) + \sum_{i=0}^p \frac{r-i}{2} \text{tr}(W_i^k X X_i^T).
\]

(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 \( X \in \mathcal{X} \), we can solve problem (16) in a closed form. However, we would like to keep the bound constraints, and this leads to the developments mentioned in Section IV-B.

B. Accelerated PG (APG) for Solving the MM Iteration

Second, we describe an iterative method for solving the MM iteration in (16). We employ the APG method [39]–[41], which is a fast first-order algorithm. The APG method for solving problem (16) is given below. Let

\[
g_k(X) = \ell(X) + \sum_{i=0}^p \frac{r-i}{2} \text{tr}(W_i^k X X_i^T)
\]

(17)

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

\[
X^{j+1} = \Pi_\mathcal{X} \left( Z^{j} - \frac{1}{\beta_j} \nabla g_k(Z^j) \right), \quad j = 0, 1, \ldots
\]

(18)

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

\[
Z^j = X^j + \alpha_j (X^j - X^{j-1})
\]

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

\[
\alpha_j = \frac{\xi_{j-1} - 1}{\xi_j}, \quad \xi_j = 1 + \frac{\sqrt{1 + 4 \xi_{j-1}^2} - 1}{2}
\]

(19)

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

\[
\nabla g_k(X) = \nabla \ell(X) + \sum_{i=0}^p \frac{r-i}{2} \text{tr}(W_i^k X X_i^T)
\]

\[
= \nabla \ell(X) + p (\gamma_0 W_0^k X + \left[ \gamma_1 W_1^k X_1, \ldots, \gamma_P W_P^k X_P \right]).
\]

(20)

Also, we have

\[
\nabla \ell(X) = F^T X F - F^T Y M + X G G^T - Y G^T.
\]

(21)

It can be verified that, given \( W_0^k, \ldots, W_p^k \), computing (20)–(21) takes \( O(M (ML + \text{nnz}(G))) \) floating-point operations, where \( \text{nnz}(G) \) denotes the number of nonzero elements of \( G \). We should note that a large number of elements of \( G \) are zero. Recall that \( G \) describes the spatial degradation process of local spatial blurring and down-sampling. One can show that the number of nonzero elements of each column of \( G \) 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, that is,

\[
\Pi_\mathcal{X}(Y) = \max\{0, \min\{1, Y\}\}
\]

(22)

where \( \text{0} \) and \( \text{1} \) denote all-zero and all-one matrices, respectively, 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 \( \beta_k \) is chosen to be a Lipschitz constant of \( \nabla g_k \) [40], [41]. We have the following result.

Fact 2: A Lipschitz constant of \( \nabla g_k \) in (17) is

\[
L_{g_k} = \lambda_{\max}(F^T F + p \gamma_0 W_0^k) + \lambda_{\max}(G G^T)
\]

\[
+ p \max_{i=1, \ldots, P} \gamma_i \lambda_{\max}(W_i^k)
\]

(23)

where \( \lambda_{\max}(\cdot) \) denotes the largest eigenvalue of the argument. The proof of Fact 2 is shown in the Appendix. 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 \( \lambda_{\max}(G G^T) \) can be computed offline, and the eigenvalues \( \lambda_{\max}(W_i^k) \) are
byproducts of computing $W_k^i$ in (15) [again, see (10)]. It suffices to calculate $\lambda_{\min}(F^TF + p\gamma_0 W_0^i)$ at every MM iteration. Hence, computing (23) takes $O(M^2)$ floating-point operations.

C. Algorithm Speedup via Inexact MM

Let us summarize the MM method developed in Sections IV-B and IV-C. 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

$$X^{k+1} = \Pi_X\left(Z^k - \frac{1}{\beta_k} \nabla g_k(Z^k)\right), \quad k = 0, 1, \ldots$$

$$Z^{k+1} = X^k + \alpha_k(X^k - X^{k-1})$$

where $\alpha_k$ is given by (19); $\beta_k$ 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 pseudoform in Algorithm 1. We call this algorithm GLORIA. It can be verified that the complexity of GLORIA is $O(M (ML + PM^2 + \text{nnz}(G)))$ per iteration.

V. 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 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 at 3-GHz CPU and 32-GB 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 $p = 1/2$ and $\tau = 1$. The local patches are obtained by dividing the image into equal-spaced rectangular blocks, as in Fig. 2. The regularization parameters $\gamma_i$’s are chosen to be identical; that is, $\gamma_0 = \gamma_1 = \cdots = \gamma_p = \gamma$. We initialize the algorithm by using a $[0,1]$-uniform i.i.d. generated $X$. 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] and [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 PSNR, SAM, Erreur Relative Globale Adimensionnelle de Synthese (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.

A. 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 $X$ and use it to generate the observed MS and HS images, $Y_H$ and $Y_M$, through the model (1). We consider the following real HS data sets.

1) Hyperspec-VNIR-C Chikusei: This data set was acquired by the Headwall Hyperspec-VNIR-C imaging sensor [47]. It covers 128 spectral bands whose wavelength range is from 363 to 1018 nm. We take a $480 \times 480$ subimage from this data set as our SR image.
2) **Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines:** This data set was captured by the AVIRIS HS sensor [48]. The wavelength range is from 400 to 2500 nm. It has 178 spectral bands after dropping bands that are corrupted by water absorption. In the experiment, a 120 $\times$ 120 subimage is used.

3) **Hyperspectral Digital Imagery Collection Experiment (HYDICE) Washington DC Mall:** This data set was taken by the HYDICE HS sensor [49]. We take a subimage of this data set, which has 240 $\times$ 240 pixels and 191 clean spectral bands. The wavelength range is from 400 to 2500 nm.

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

5) **AVIRIS Moffett Field:** This data set, recorded by AVIRIS HS sensor, has 187 uncorrupted spectral bands. The wavelength range is from 400 to 2500 nm. We take a 240 $\times$ 240 subimage from this data set.

These five images have been displayed in Fig. 3. The settings with the spectral and spatial measurement response matrices $F$ and $G$ should also be described. The matrix $F$ is chosen such that it is equivalent to the spectral response of either the Landsat 4 TM sensor [50] (six bands, with spectral coverage from 400 to 2500 nm) or the IKONOS sensor [51] (four bands, with spectral coverage from 450 to 900 nm). 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 a choice is to match the spectral coverage of the HS images. As discussed previously, $G$ corresponds to the process of spatial blurring and downsampling. The blurring function is a $11 \times 11$ Gaussian point spread function, with variance 1.7. The downsampling is done every four pixels, both horizontally and vertically. Furthermore, the noise terms $V_H$ and $V_M$ are randomly generated following an i.i.d. mean-zero Gaussian distribution.

We fix the SNRs at $\text{SNR}_M = \text{SNR}_H = 25 \text{ dB}$. The obtained performance is shown in Table III, 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 III, 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 a 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 Section V-B.

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, for example, $29491200$ variables in Chikusei. In comparison, CNMF and HySure require $(M + L)N$ and $NL$ optimization variables, respectively, which amount to $6915840$ and $6912000$ variables in Chikusei. In terms of runtime per variable, GLORIA is considered efficient.
B. Synthetic Data Experiments

Second, we consider synthetic data experiments. The way we prepare the experiment is similar to the one mentioned in [9] and [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 $A_i$ and abundance matrix $S_i$ 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.

**TABLE IV**

| Method | PSNR | SAM | BRGAS | UQI |
|--------|------|------|-------|-----|
| GSA    | 10.99±2.27 | 19.29±4.31 | 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   | 15.71±0.41 | 10.52±0.52 | 5.43±0.22 | 0.299±0.050 |
| FUMI   | 17.63±0.16 | 7.69±0.62 | 4.32±0.17 | 0.456±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±0.33 | 7.90±3.10 | 2.02±0.06 | 0.432±0.060 |
| $P = 9$ | 21.46±0.12 | 5.03±0.13 | 2.00±0.07 | 0.583±0.051 |
| $P = 16$ | 21.76±0.14 | 4.61±0.10 | 2.43±0.06 | 0.594±0.053 |

**TABLE V**

| Method | $5N_{IM}, 5N_{IM} - 152B$ |
|--------|--------------------------|
| GSA    | 21.31±2.35 | 7.85±11.01 | 4.38±4.35 | 0.796±0.109 |
| GLP    | 21.07±0.53 | 1.68±0.91 | 2.79±0.19 | 0.785±0.070 |
| CNMF   | 32.29±0.35 | 1.76±0.12 | 0.49±0.06 | 0.925±0.020 |
| FUMI   | 23.45±0.30 | 1.47±0.11 | 2.32±0.09 | 0.849±0.010 |
| HySure | 35.93±0.42 | 2.02±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    | 36.87±0.32 | 1.19±0.02 | 0.39±0.01 | 0.999±0.014 |

| GLORIA | 
|--------|-------|-------|-------|-----|
| $P = 1$ | 35.48±0.51 | 1.09±0.08 | 0.35±0.05 | 0.966±0.012 |
| $P = 9$ | 36.18±0.23 | 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 |

**Fig. 8.** SAM maps of the Moffett Field data set. (a) GSA. (b) GLP. (c) CNMF. (d) FUMI. (e) HySure. (f) LRSR. (g) NNM. (h) GLORIA.

**Fig. 9.** Synthetic SR image generation. (a) Abundance map of one material (estimated from the Cuprite data set). (b) Segmentation of the abundance map.

**Fig. 10.** Tested HS-MS image pair. (Left) Fifth band of the HS image. (Right) First band of the MS image.

The endmember matrix $A_i$ is chosen as a collection of the spectral signatures of five materials, specifically, Actinolite, Albite, Muscovite, Olivine, and Topaz. To simulate the EV changes in space.
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 $S$ is chosen as the abundance maps extracted from a real HS image, namely, the AVIRIS Cuprite data set; the extraction is done by applying an HU method called SVMAX [53]. In each trial, we randomly cropped a $120 \times 120$ 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 $\gamma = 40/(\text{SNR}_M + \text{SNR}_H)$. We ran 100 independent trials. Table IV 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.

C. Real Data Experiments

Finally, we test the algorithms on real data. The data set is the one used in [54]. The HS image was acquired by the Hyperion HS sensor. It covers a spectral range from 400 to 2500 nm 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 four bands whose central wavelengths are 490, 560, 665, and 842 nm. Readers are referred to [54] for further details. After preprocessing such as co-registration and cropping, we obtain the HS–MS image pair. The image pair is illustrated in Fig. 10. The setting is $(M, M_M, L, L_H, N) = (89, 4, 360^2, 120^2, 30)$. We employ the algorithm in [9] to estimate the spectral and spatial measurement responses $F$ and $G$. Empirically, we found that, for the tested HS-MS pair, this algorithm happens to yield a poor estimate of $G$. To rectify this issue, we consider a second-stage estimation: we fix the blurring kernel to be a Gaussian kernel and estimate its standard deviation $\sigma^2$ by

$$min_{\sigma^2 > 0} \|Y_M \circ g(\sigma^2) - f^1 Y_H\|^2,$$

where $g(\sigma^2)$ denotes the Gaussian kernel with variance $\sigma^2$, $\circ$ corresponds to the 2-D convolution operation, and $y_M^1$ and $f^1$ are the first row of $Y_M$ and the estimated $F$, respectively.

We follow the same simulation settings as those in semi-real experiments, except that we use $\gamma = 10$ for GLORIA.
Figs. 11 and 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 zoomed-in view 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.

VI. CONCLUSION

In this article, 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 data sets, 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 article by naming a future direction. It would be interesting to study how low-rank matrix estimation would be useful in other HS problems such as multisource and multitemporal data fusion.

APPENDIX

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

$$\|\nabla g_k(\mathbf{X}) - \nabla g_k(\mathbf{Y})\|_F \leq \left\| (\mathbf{F}^T \mathbf{F} + p \gamma_0 \mathbf{W}_0^k) (\mathbf{X} - \mathbf{Y}) \right\|_F + \left\| (\mathbf{X} - \mathbf{Y}) \mathbf{G} \mathbf{G}^T \right\|_F + p \sum_{i=1}^p \left\| \gamma_i \mathbf{W}_i^k (\mathbf{X}_i - \mathbf{Y}_i) \right\|_F \leq \left( \lambda_{\text{max}}(\mathbf{F}^T \mathbf{F} + p \gamma_0 \mathbf{W}_0^k) + \lambda_{\text{max}}(\mathbf{G} \mathbf{G}^T) \right) \|\mathbf{X} - \mathbf{Y}\|_F$$

for any $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times L}$, where (25b) is due to the triangle inequality; (25c) is due to: 1) the inequality $\|\mathbf{A} \mathbf{B}\|_F \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_F$ in which $\|\mathbf{A}\|_2$ denotes the spectral norm of $\mathbf{A}$ and 2) the identity $\|\mathbf{A}\|_2 = \lambda_{\text{max}}(\mathbf{A})$ for positive-semidefinite $\mathbf{A}$. The proof of Fact 2 is complete.

REFERENCES

[1] L. Loncan et al., “Hyperspectral pansharpening: A review,” IEEE Geosci. Remote Sens. Mag., vol. 3, no. 3, pp. 27–46, Sep. 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, Jan. 2014.
[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, Jun. 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, Feb. 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. Acoust, Speech Signal Process., May 2013, pp. 1409–1413.
[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, May 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.

Ruiyuan Wu (Student Member, IEEE) received the B.Eng. degree in information engineering from the South China University of Technology, Guangzhou, China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong. He is currently a Visiting Scholar with Arizona State University, Tempe, AZ, USA. His research interests include structured matrix factorization, non-convex optimization, and distributed learning techniques for data science and remote sensing.

Wing-Kin Ma (Fellow, IEEE) is currently a Professor with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong. His most recent research focuses on two distinct topics, namely, structured matrix factorization for data science and remote sensing, and MIMO transceiver design and optimization. His research interests lie in signal processing, optimization, and communications.

Dr. Ma is currently a member of the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC) and was previously a member of the Signal Processing Theory and Methods Technical Committee (SPTM-TC). He received the Research Excellence Award 2013–2014 from CUHK, the 2015 IEEE Signal Processing Magazine Best Paper Award, the 2016 IEEE Signal Processing Letters Best Paper Award, and the 2018 IEEE Signal Processing Society (SPS) Best Paper Award. He has rich experience in editorial service, for example, an Associate Editor and then later a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Lead Guest Editor of the IEEE Signal Processing Magazine Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing in 2014, and many others. He was an IEEE Signal Processing Society Distinguished Lecturer from 2018 to 2019 and a Tutorial Speaker in EUSIPCO 2011 and ICASSP 2014. He currently serves as an IEEE Signal Processing Society Regional Director-at-Large (Region 10).

Xiao Fu (Member, IEEE) received the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2014. He was a Post-Doctoral Associate with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, from 2014 to 2017. He is currently an Assistant Professor with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA. His research interests include the broad area of signal processing and machine learning.

Dr. Fu received the Best Student Paper Award at ICASSP 2014. Two of his coauthored papers received the Best Student Awards at the IEEE Computational Advances in Multi-Sensor Adaptive Processing 2015 and the IEEE Machine Learning for Signal Processing 2019. He serves as the Treasurer of the IEEE Signal Processing Society Oregon Chapter. He also serves as a member of the EURASIP Technical Area Committee in Signal Processing for Multisensor Systems for the term 2020–2023.

Qiang Li (Member, IEEE) received the B.Eng. and M.Phil. degrees in communication and information engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2005 and 2008, respectively, and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2012. From August 2011 to January 2012, he was a Visiting Scholar with the University of Minnesota, Minneapolis, MN, USA. From February 2012 to October 2013, he was a Research Associate with the Department of Electronic Engineering and the Department of Systems Engineering and Engineering Management, CUHK. Since November 2013, he has been with the School of Information and Communication Engineering, UESTC, where he is currently an Associate Professor. His research interests include efficient optimization algorithm design for wireless communications and machine learning.

Dr. Li received the First Prize Paper Award in the IEEE Signal Processing Society Postgraduate Forum Hong Kong Chapter in 2010, the Best Paper Award of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2016, and the Best Paper Award of the IEEE Signal Processing Letters 2016.