

Signal Subspace Identification for Incomplete Hyperspectral Image With Applications to Various Inverse Problems

Chia-Hsiang Lin^{ID}, Member, IEEE, and Si-Sheng Young, Student Member, IEEE

Abstract—In hyperspectral remote sensing (HRS), signal subspace identification (SSID) is a critical step in many widely renowned HRS algorithms, while the accuracy of the subspace identification relies on the complete information of the data pixels. However, as the sensor arrays would be partially damaged after the satellites are launched, hyperspectral pixels are quite often incompletely acquired. Even for those renowned algorithms, they simply remove those incomplete pixels when computing the hyperspectral signal subspace. Nevertheless, even if some spectral bands of a given incomplete pixel are missing, we intuitively believe that the remaining bands of that pixel should still contribute to the accuracy of the subspace identification. We design a computationally efficient algorithm, termed as subspace identification for incomplete signals of hyperspectral image (SISHY), to utilize the information embedded in those incomplete pixels. To this end, we prove a lemma that allows us to reformulate the algebraic identification problem into an affine geometry problem, thereby allowing us to flexibly add suitable regularizer for better identification result as needed. SISHY judiciously associates the regularized subspace identification problem with a denoising operator, thereby allowing an efficient algorithm implementation and yielding a physically interpretable data matrix completion result as a byproduct. Experiments demonstrate that the SISHY algorithm does improve the efficacy of the subsequent tasks of unmixing, inpainting, and classification.

Index Terms—Classification, hyperspectral unmixing (HU), image inpainting, inverse problems, signal subspace identification (SSID).

I. INTRODUCTION

A. Overview

HYPERSPECTRAL remote sensing (HRS) has become one of the key techniques in space exploration due to the strong and provable material identifiability of hyperspectral images [1], [2]. In numerous HRS tasks, signal subspace identification (SSID) is an essential algorithmic step in improving the performance and in reducing the computational complexity. For example, in the hyperspectral unmixing (HU) task,

Manuscript received 20 September 2023; revised 1 February 2024; accepted 10 March 2024. Date of publication 18 March 2024; date of current version 1 April 2024. This work was supported in part by the EINSTEIN Program of National Science and Technology Council (NSTC), Taiwan, under Grant MOST 111-2636-E-006-028; and in part by the Emerging Young Scholar Program (i.e., 2030 Program) of NSTC, Taiwan, under Grant NSTC 112-2628-E-006-017. (Corresponding author: Chia-Hsiang Lin.)

Chia-Hsiang Lin is with the Department of Electrical Engineering and the Min Wu School of Computing, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: chiahsiang.steven.lin@gmail.com).

Si-Sheng Young is with the Department of Electrical Engineering, Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: q38121509@gs.ncku.edu.tw).

Digital Object Identifier 10.1109/TGRS.2024.3378705

which aims at unambiguously separating the mixed spectra into pure signature vectors corresponding to the underlying materials, the widely renowned HU method called vertex component analysis (VCA) [3] also adopts the SSID as its first algorithmic step. Similarly, another very fast HU algorithm known as hyperplane-based Craig simplex identification (HyperCSI) [4] employs the SSID to perform dimension reduction (DR), allowing us to use the signal sparsity in the DR space for real-time signature estimation. For another example, considering that the hyperspectral image is often incompletely acquired (due to broken sensors) or damaged during the transmission from satellites to the ground stations, hyperspectral image inpainting is another critical task in HRS. The well-known fast hyperspectral inpainting (FastHyIn) method also adopts SSID technique to transfer the image domain into the eigenimage domain, where the self-similarity regularization can be effectively performed thereby yielding a computationally efficient algorithm [5]. The last example in HRS is the hyperspectral classification task, for which a popular method known as superpixelwise principal component analysis (SuperPCA) incorporates the SSID stage into the learning framework thereby leading to superior unsupervised feature extraction [6]. In fact, numerous representative classification methods also rely on SSID to extract the discriminative features across various source domains [7], [8], [9].

In fact, SSID also plays a critical role in bridging two core inverse imaging frameworks, i.e., convex optimization (CO) and deep learning (DE), thereby leading to the so-called CODE small-data machine learning theory in recent HRS literature [10]. Simply speaking, CO often involves more complicated mathematics, but does not rely on big data collection that is time consuming and expensive. On the contrary, DE often relies on big data but avoids sophisticated mathematics. By CODE, one can employ the SSID to tailor the so-called Q -quadratic norm [11] regularization function, which elegantly bridges CO and DE to avoid big data and math-heavy optimization, and such a regularization scheme has been proven to yield superior capability in challenging HRS missions like change detection [12]. Other applications of SSID include the quantum DE, as well as miniaturized satellite hyperspectral compressed sensing, as investigated in very recent hyperspectral satellite literature [13], [14]. The accuracy of the subspace identification greatly relies on the information of completely observed pixels in the entire hyperspectral image. Nevertheless, the onboard sensor arrays would

be partially damaged after the satellites are launched, resulting in incompletely acquired hyperspectral pixels, or even if some hyperspectral pixel is completely acquired, it may be corrupted during the transmission from the satellites to the ground stations. Both effects lead to missing information in the hyperspectral image. Even for the aforementioned renowned HRS algorithms, they simply remove those incomplete pixels (i.e., pixels with missing spectral bands) when computing the hyperspectral signal subspace; obviously, this will degrade the SSID performance as well as the efficacy of the subsequent HRS missions.

B. Related Work

Intuitively, even if some spectral bands of a given incomplete pixel are missing, we believe that the remaining bands of that pixel should still contribute to the accuracy of the subspace identification. We aim to design a computationally efficient algorithm to exploit the information embedded in those incomplete pixels. Below we recall some related works. Though there are numerous SSID methods, including maximum noise fraction (MNF) [15], projection pursuit (PP) [16], and hyperspectral signal identification by minimum error (HySime) [17], they are not able to address the scenarios wherein the hyperspectral images are incompletely observed. Some variations of the principal component analysis (PCA) methods were proposed to tackle the missing data, including extended PCA (EPCA) [18], least-squares PCA (LSPCA) [19], probabilistic PCA (PPCA) [20], and variational Bayesian PCA (VBPCA) [21]. EPCA tackles the incomplete observations based on the best low-rank approximation. Specifically, EPCA interpolates the missing data by low-rank approximation, which can be obtained by singular value decomposition (SVD). LSPCA works well if the missing rate is low, but suffers from overfitting issue and rank deficiency [19]. As proper prior information or regularization can prevent overfitting, the probabilistic learning method of PPCA incorporates the latent variable and noise, both following normal distributions, where the incomplete case can be element-wise formulated and solved by the expectation–maximization (EM) algorithm [22]. To refrain the overfitting, VBPCA [21] considers the parameters in the priors of PPCA as more complicated distributions, and uses the approximated probability density function in minimizing the cost function by the EM algorithm. Typically, the VBPCA is more robust to overfitting thanks to its sensitivity to posterior probability mass function [19], while the computation requirement increases more rapidly than PPCA when the missing rate gets large. Although probabilistic-based algorithms achieve feasible identification of signal subspace, the high-computational complexity is less practical. So, Grassmannian rank-one update subspace estimation (GROUSE) [23] presents an efficient method with linear computational complexity for highly sparse data. GROUSE is an incremental gradient algorithm implemented on the Grassmannian manifold of subspace. Grassmannian robust adaptive subspace tracking algorithm (GRASTA) [24] is another GROUSE-based robust subspace tracking method to address SSID for incomplete data.

The performance of GROUSE-based methods is often trapped at local minima limiting the efficacy of Grassmannian

manifold [25]. Parallel estimation and tracking by recursive least squares (PETRELS) [25], a modification of the projection approximation subspace tracking (PAST) method [26] can converge to the true subspace under specific prior assumptions not restricted by the Grassmannian manifold. PAST is proposed for subspace identification with complete data, while PETRELS is a second-order stochastic gradient descent method that can rebuild the incomplete elements via least-squares estimation. Numerous algorithms similar to PETRELS were also explored using different priors [27], [28], [29], [30], [31]. Robust online subspace estimation and tracking algorithm (ROSETA) [29] combines the advantages of PETRELS and GRASTA, executing the adaptive step on the robust sparsity-promoting cost function for improving the convergence rate during the subspace identification phase. Furthermore, PETRELS adopts a constant false alarm rate (CFAR) detector that can mitigate the outlier effect, yielding the method termed PETRELS-CFAR [30]. Advanced PETRELS for robust SSID was also proposed under the CO framework known as alternating direction method of multipliers (ADMM) [11], leading to the SSID method called P-ADMM, which is more efficient for large-scale missing scenario compared to the naive PETRELS [31]. On the other hand, improved PETRELS (IPETRELS) is designed in [32] by incorporating a subspace bounded regularization term. We should also mention the biological research package in [33], originally developed for microarray and metabolite data often containing many missing values, can also be applied to solve the SSID problem for incompletely observed data matrix. Another mainstream of robust SSID is recursive projected compressive sensing (ReProCS) algorithm [34], [35]. The original ReProCS proposed in [34] provides an efficient solution with upper-bounded reconstruction error under mild assumptions. The nearly optimal robust subspace tracking (NORST) [36] relies on the ReProCS framework with improved SSID performance, and it is further modified to address the issue of missing data, resulting in a memory-efficient SSID method called NORST-miss [37]. Another much simpler approach than ReProCS is federated over-air robust subspace tracking with missing entries (Fed-OA-RSTMiss) [38]. Fed-OA-RSTMiss develops an incomplete SSID algorithm holding piecewise constant subspace change assumption without any changing upper-bound constraints, thereby leading to a more efficient SSID algorithm.

C. Contributions

Nevertheless, the aforementioned SSID methods either suffer from heavy computational burdens or yield limited SSID performances. More seriously, though some methods appear to have good quantitative SSID performances, they work poorly when their SSID results are used for the subsequent core HRS tasks. In this article, we aim to propose a computationally efficient SSID algorithm that really assists solving crucial inverse problems encountered in HRS, including HU, inpainting, and classification. Our algorithm, termed subspace identification for incomplete signals of hyperspectral image (SISHY), elegantly introduces an auxiliary variable to avoid

the approximated data covariance matrix being deteriorated. The auxiliary variable can be viewed as the data matrix completion result, and this physical meaning allows SISHY to incorporate flexible regularization functions to promote desired properties of a valid hyperspectral image. We also show that SISHY has very simple algorithmic form, because our derivation proves that the physical meaning of incorporating the additional regularizers is nothing but the fundamental denoising problem. Experimental comparisons among numerous SSID algorithms, as well as three HRS applications, will demonstrate the superiority of SISHY. The main contributions are itemized as follows.

- 1) We first propose a heuristic SSID algorithm (i.e., Algorithm 1) to address the SSID problems for the incompletely observed hyperspectral images, which works well when the missing rate is not too high. Then, to tackle higher missing rates, we need to let all the available entry-level information be fully exerted, as done in the optimization criterion (10). This criterion, together with Lemma 2 that transforms the algebra-based SSID formulation into a geometry-based SSID formulation, allows us to derive the SISHY algorithm (i.e., Algorithm 2). Algorithm 2 is highly efficient (thanks to the affine geometry) and avoids repeatedly computing some matrix inversion terms (see the discussion below Lemma 2).
- 2) Considering that SSID may find particular interests in other areas, we leave a degree of freedom to allow incorporating a regularization term ϕ (depending on the desired prior) into the SSID formulation. We also demonstrate how to exploit such a flexibility by investigating the self-similarity prior-based ϕ , and solve the corresponding regularized SSID formulation using plug-and-play strategy, thereby leading to the regSISHY algorithm (i.e., Algorithm 3).
- 3) The proposed SSID technique is outstanding in terms of both the computational speed and performance efficiency. More importantly, our SSID technique does help improve the performance of numerous renowned HRS algorithms, including HyperCSI (HU), FastHyIn (hyperspectral inpainting), and SuperPCA (hyperspectral classification), double confirming the superiority of SISHY. Remarkably, our technique has elegantly introduced an auxiliary variable, which not only allows us to incorporate the regularization term but also avoids the approximated data covariance matrix being deteriorated, and this auxiliary variable can be judiciously viewed as the data matrix completion result.

The remaining parts of this article will be organized as follows. In Section II, we design the SSID criteria that are then implemented by customized algorithms. In Section III, we experimentally demonstrate the efficacy of SISHY, and show that SISHY-driven benchmark unmixing/inpainting/classification algorithms do yield better HRS performances. Conclusions are drawn in Section IV, while mathematical proofs are presented in Appendix to keep the main article clean.

Some standard notations are collectively presented hereinafter. For a given matrix \mathbf{M} , $[\mathbf{M}]_{m,:}$ is a row vector defined by the m th row of \mathbf{M} and $[\mathbf{M}]_{ij}$ is the (i, j) th entry of \mathbf{M} . Also, \mathbf{M}^T and \mathbf{M}^{-1} denote the matrix transpose and inversion, respectively. The matrix \mathbf{I}_n is the $n \times n$ identity matrix and $\|\cdot\|_F$ is the Frobenius norm. The vector $\mathbf{1}_p$ is the p -dimensional all-one vector, and $[\mathbf{v}]_m$ is the m th entry of the vector \mathbf{v} . For a given set \mathcal{S} , the operators $\text{conv } \mathcal{S}$ and $\text{aff } \mathcal{S}$, respectively, represent convex hull and affine hull of \mathcal{S} . For any positive integer Z , we define $\mathcal{I}_Z \triangleq \{1, \dots, Z\}$.

II. PROPOSED SSID FORMULATION AND SISHY ALGORITHM

SSID is critical in numerous science and engineering applications. As a critical but yet to be well-resolved problem, we aim to derive SSID from the incompletely observed data \mathbf{X}_{Ω} . One specific aim is to design a computationally efficient algorithm for subspace identification in HRS. Moreover, considering that SSID often serves as just the very first step in various HRS tasks, another aim is to ensure that the developed SISHY algorithm does help upgrade the performances in the subsequent HRS missions, including unmixing, inpainting, and classification. These aims will be achieved in Section II, and experimentally demonstrated in Section III.

The SSID problem can be interpreted probabilistically using a latent variable model [20], [39], [40], but the resulting SSID methods could take several hours on normal desktop computers [41], violating our aim. Thus, we turn to pursue optimization-based methods; it is known that SSID is an optimal data analysis method for data compression in the mean-squared-error (MSE) sense. Specifically, when the M-band, L -pixel data $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L] \in \mathbb{R}^{M \times L}$ is completely observed, the naive version of the SSID problem can be formulated as

$$(\mathbf{E}^*, \mathbf{S}^*) := \arg \min_{\mathbf{E} \in \mathbb{R}^{M \times N}, \mathbf{S} \in \mathbb{R}^{N \times L}} \frac{1}{2} \|\mathbf{X} - \mathbf{ES}\|_F^2 \quad (1)$$

where the optimal solution \mathbf{E}^* will then define the identified subspace that best represent the data \mathbf{X} in the MSE sense, and \mathbf{S} is just the corresponding coefficient matrix. Here, the model-order N can be estimated using some information theory-driven algorithms [42], and will be assumed known a priori in this work. The formulation (1) can be further customized for hyperspectral data, as illustrated next.

It is known that the hyperspectral pixels $\mathbf{x}_1, \dots, \mathbf{x}_L$ actually belong to a low-dimensional convex hull $\text{conv}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ [11], in which $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the hyperspectral signatures (a.k.a. endmembers) of the $n \ll M$ materials underlying the image \mathbf{X} [43]. Therefore, for hyperspectral data applications, instead of identifying the data subspace directly, we may more economically identify the data affine hull because we have the following geometry relation, i.e.,

$$\mathbf{x}_i \in \text{conv}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \subseteq \text{aff}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}.$$

To be rigorous, as an affine hull \mathcal{A} can be written as

$$\mathcal{A} = \{\mathbf{x} \mid \mathbf{x} = \mathbf{Es} + \mathbf{d}, \forall s\}$$

with $\mathbf{d} \in \mathbb{R}^M$ being the shifting vector, the SSID formulation can be more economically customized for the hyperspectral data, as the following MSE minimization problem, i.e.,

$$(\mathbf{E}^*, \mathbf{S}^*, \mathbf{d}^*) := \arg \min_{\mathbf{E}, \mathbf{S}, \mathbf{d}} \frac{1}{2} \|\mathbf{X} - (\mathbf{E}\mathbf{S} + \mathbf{d}\mathbf{1}_L^T)\|_F^2. \quad (2)$$

The closed-form solution to (2) is given in Lemma 1, whose proof can be easily derived and is thus omitted for conciseness.

Lemma 1: Assume $N \leq M \leq L$. A closed-form solution to (2) is given by

$$\mathbf{d}^* = \frac{1}{L} \mathbf{X} \mathbf{1}_L \quad (3)$$

$$\mathbf{E}^* = [\mathbf{q}_1, \dots, \mathbf{q}_N]$$

$$\mathbf{S}^* = \mathbf{E}^{*T} (\mathbf{X} - \mathbf{d}^* \mathbf{1}_L^T) \quad (4)$$

where $\mathbf{q}_i \in \mathbb{R}^M$ denotes the i th principal eigenvector of the mean-removed data covariance matrix $\mathbf{C} \triangleq (\mathbf{X} - \mathbf{d}^* \mathbf{1}_L^T)(\mathbf{X} - \mathbf{d}^* \mathbf{1}_L^T)^T \in \mathbb{R}^{M \times M}$. \square

Note that $N \leq M \leq L$ is a natural assumption for real hyperspectral images.

By Lemma 1, we see that (2) can be equivalently reformulated as

$$(\mathbf{E}^*, \mathbf{S}^*, \mathbf{d}^*) := \arg \min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}} \frac{1}{2} \|\mathbf{X} - (\mathbf{E}\mathbf{S} + \mathbf{d}\mathbf{1}_L^T)\|_F^2 \quad (5)$$

where $\mathbb{U} \subset \mathbb{R}^{M \times N}$ collects all the semiunitary matrices \mathbf{E} (i.e., matrix satisfying $\mathbf{E}^T \mathbf{E} = \mathbf{I}_N$). The reason is that the global optimal solution \mathbf{E}^* presented in Lemma 1 already satisfies the semiunitary property; thus, adding the additional constraint “ $\mathbf{E} \in \mathbb{U}$ ” in (5) does not alter the physical meaning of the SSID problem. However, the additional constraint “ $\mathbf{E} \in \mathbb{U}$ ” is practically useful to reduce the computational complexity of the subsequent signal-processing tasks; for example, if \mathbf{E}^* is not semiunitary, the computation of \mathbf{S}^* would not be as simple as that presented in Lemma 1. As we will see, the proposed SISHY algorithm will need to compute \mathbf{S}^* iteratively, implying that the constraint “ $\mathbf{E} \in \mathbb{U}$ ” is practically useful. It is interesting to remark that traditional probabilistic PCA may be algebraically defined using the requirements that \mathbf{E} is semiunitary, and that for each $k = 1, \dots, N$, the first k columns of \mathbf{E} span the k -dimensional “principal subspace” [19]. Another remark is that the above two specific requirements make the solution to (5) practically unique (except for changing the sign, and in the special case of having equal eigenvalues) [19]. Next, we should consider the case of incompletely observed \mathbf{X} .

Inspired by Lemma 1, we first propose a heuristic SSID algorithm (i.e., Algorithm 1), and discuss its limitations. Roughly speaking, the heuristic algorithm is to compute the \mathbf{d} merely using the observable data \mathbf{X}_{Ω} , and to compute \mathbf{E} by estimating the covariance matrix also merely using \mathbf{X}_{Ω} . Recall the notation of the data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L] \in \mathbb{R}^{M \times L}$ with each data point $\mathbf{x}_i \in \mathbb{R}^M$. As the data may not be completely observed, we define $\Omega \subseteq \{(m, n) \mid m \in \mathcal{I}_M, n \in \mathcal{I}_L\}$ as the support set of indices of those observable entries in \mathbf{X} , and denote \mathbf{X}_{Ω} as the observable data. For the complement entry $(m, n) \in \bar{\Omega} \triangleq (\mathcal{I}_M \times \mathcal{I}_L) \setminus \Omega$, the entry $[\mathbf{X}_{\Omega}]_{mn}$ is simply set as 0 for convenience, meaning that the m th band in the n th pixel is missing.

Accordingly, given $\Omega \subseteq \{(m, n) \mid m \in \mathcal{I}_M, n \in \mathcal{I}_L\}$ defined previously, we further define the notations.

- 1) Let $\Omega^m \subseteq \{1, \dots, L\}$ be the set collecting column indices (i.e., pixel indices) of observable entries in the m th row (i.e., the m th band) of \mathbf{X} .
- 2) Let $\Omega_n \subseteq \{1, \dots, M\}$ be the set collecting row indices (i.e., band indices) of observable entries in the n th column (i.e., the n th pixel) of \mathbf{X} .

Then, we are ready to describe a heuristic SSID algorithm inspired by Lemma 1. Specifically, when \mathbf{X} is incompletely observed, it is natural to approximate $\mathbf{d}^* \in \mathbb{R}^M$ by

$$[\widehat{\mathbf{d}}]_m = \frac{1}{|\Omega^m|} [\mathbf{X}_{\Omega}]_{m,:} \mathbf{1}_L \quad \forall m = 1, \dots, M \quad (6)$$

where $|\mathcal{S}|$ denotes the cardinality of the set \mathcal{S} . As missing entries in \mathbf{X}_{Ω} are set as 0, the meaning of (6) is averaging over those observable entries for each spectral band [see (3)]. To understand (6), we first notice that (3) approximates the m th entry of \mathbf{d}^* by averaging over the L pixels in the m th band of \mathbf{X} . Thus, when some of the L pixels are missing, it is natural to approximate the m th entry by the average of the remaining $|\Omega^m|$ observable pixels in the m th band, and this is exactly what (6) has done. When the missing rate is not too high, the approximations from (3) to (6) are expected to be similar, but we still call this intuitive approach a heuristic algorithm (see Algorithm 1). Note that the approximation in (6) reduces back to the true data mean \mathbf{d}^* when \mathbf{X} is completely observed (i.e., $\Omega^m = \mathcal{I}_L$). Therefore, (6) can be viewed as a generalized version of (3). On the other hand, inspired by Lemma 1 as well, it is natural to approximate \mathbf{E}^* by

$$\widehat{\mathbf{E}} = [\widehat{\mathbf{q}}_1, \dots, \widehat{\mathbf{q}}_N] \quad (7)$$

in which $\widehat{\mathbf{q}}_i \in \mathbb{R}^M$ denotes the i th principal eigenvector of the approximated mean-removed data covariance matrix $\widehat{\mathbf{C}}$, whose (i, j) th entry is defined as

$$[\widehat{\mathbf{C}}]_{ij} := \frac{1}{|\Omega^i \cap \Omega^j|} [\widehat{\mathbf{U}}]_{i,:} [\widehat{\mathbf{U}}]_{j,:}^T \quad (8)$$

where $\widehat{\mathbf{U}} \equiv \mathbf{X}_{\Omega} - \widehat{\mathbf{d}} \mathbf{1}_L^T$ on Ω (and 0 on $\bar{\Omega}$). To understand (7), note that \mathbf{E}^* is completely determined by the mean-removed data covariance matrix \mathbf{C} (see Lemma 1), whose (i, j) th entry (i.e., $[\mathbf{C}]_{ij}$) is then completely determined by the i th and j th rows/bands of the data \mathbf{X} (see the definition of \mathbf{C} in Lemma 1). Naturally, when the i th and j th bands are not completely observed, we simply turn to use those $|\Omega^i \cap \Omega^j|$ pixels observable in “both” bands (i.e., the intersection of both bands) to estimate the (i, j) th entry $[\mathbf{C}]_{ij}$, and this is exactly what (8) has done. Therefore, $\widehat{\mathbf{E}}$ will also reduce back to \mathbf{E}^* when the given data are completely observable, and (7) shows a more generalized version than Lemma 1. Finally, to approximate \mathbf{S} , we check (4) to see that $s_n = \mathbf{E}^{*T} (\mathbf{x}_n - \mathbf{d}^*)$ when \mathbf{x}_n is completely observed (i.e., $\Omega_n = \mathcal{I}_M$). If \mathbf{x}_n is incompletely observed, we can turn to approximate it using $s_n = \mathbf{E}^{*T} \mathbf{v}$, where $\mathbf{v} = \mathbf{x}_n - \mathbf{d}^*$ on Ω_n (and 0 on $\bar{\Omega}_n$). The rational to set 0 on $\bar{\Omega}_n$ is that \mathbf{v} is a mean-removed vector, motivating us to assume that $[\mathbf{v}]_i$ has an expectation value of 0 when $i \in \bar{\Omega}_n$. All in all, we can approximate the n th

column of \mathbf{S}^* by

$$\hat{s}_n = \begin{cases} \hat{\mathbf{E}}^T(\mathbf{x}_n - \hat{\mathbf{d}}), & \text{if } \Omega_n = \mathcal{I}_M \\ \hat{\mathbf{E}}^T \mathbf{v}, & \text{if } \Omega_n \neq \mathcal{I}_M. \end{cases} \quad (9)$$

The heuristic SSID Algorithm for incompletely observed data \mathbf{X}_{Ω} is summarized in Algorithm 1, which is expected to work nicely when the missing rate is not large.

Algorithm 1 Heuristic SSID Algorithm With Missing Data

- 1: Given incompletely observed hyperspectral image \mathbf{X}_{Ω} .
- 2: Approximate \mathbf{d}^* by $\hat{\mathbf{d}}$ using (6).
- 3: Compute $\hat{\mathbf{C}}$ using (8).
- 4: Approximate \mathbf{E}^* by $\hat{\mathbf{E}}$ using $\hat{\mathbf{C}}$ and (7).
- 5: Approximate \mathbf{S}^* by $\hat{\mathbf{S}} := [\hat{s}_1, \dots, \hat{s}_L]$ using (9).
- 6: Output the SSID result $(\mathbf{E}^*, \mathbf{S}^*, \mathbf{d}^*) \approx (\hat{\mathbf{E}}, \hat{\mathbf{S}}, \hat{\mathbf{d}})$.

However, when there are too many missing entries, implying that there are very few intersections of observable pixel entries between different spectral bands, Algorithm 1 may not work well, because the approximated mean-removed data covariance matrix $\hat{\mathbf{C}}$ may even not be a positive semidefinite (PSD) matrix. To get a sense, we design a simple toy example, for which the incomplete data are

$$\mathbf{X} = \begin{bmatrix} 2.1 & 1.2 & - & 1 & 0.1 \\ 2.1 & - & 1.1 & - & 0.1 \\ - & 2.1 & 1.1 & 0.1 & - \end{bmatrix}$$

where “–” denotes missing entry. By (6) and (8), one can verify that Algorithm 1 yields the following non-PSD mean-removed data covariance matrix (with one negative eigenvalue), i.e.,

$$\hat{\mathbf{C}} = \begin{bmatrix} 0.505 & 1 & 0.1 \\ 1 & 2/3 & 0 \\ 0.1 & 0 & 2/3 \end{bmatrix}$$

because there are too few intersections of observable pixel entries among the bands (i.e., $|\Omega^i \cap \Omega^j|$ is too small). We also use two real hyperspectral data acquired by National Aeronautics and Space Administration (NASA) over two representative landscapes (i.e., city and mountain), as displayed in Fig. 1, where the commonly seen missing patterns are exactly the same as those used in [10]. The two patterns correspond to slight missing and serious missing, respectively (see Fig. 1). As it turns out, both data lead to non-PSD matrices; the city data has 4 negative eigenvalues, while the mountain data has six negative eigenvalues. To resolve the dilemma, we will trickily incorporate an auxiliary variable into the SSID criterion (5), as will be detailed below.

To this end, we first rewrite the SSID criterion (5) particularly for the incomplete data \mathbf{X}_{Ω} as

$$\begin{aligned} & \min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}} \frac{1}{2} \|\mathbf{X} - (\mathbf{E}\mathbf{S} + \mathbf{d}\mathbf{1}_L^T)\|_F^2 \\ & \equiv \min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}} \sum_{\ell=1}^L \frac{1}{2} \|\mathbf{x}_\ell - (\mathbf{E}\mathbf{s}_\ell + \mathbf{d})\|_2^2 \\ & \equiv \min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}} \sum_{\ell=1}^L \sum_{m=1}^M \frac{1}{2} [x_{m\ell} - (\mathbf{E}_{m,:}\mathbf{s}_\ell + d_m)]^2 \end{aligned} \quad (10)$$

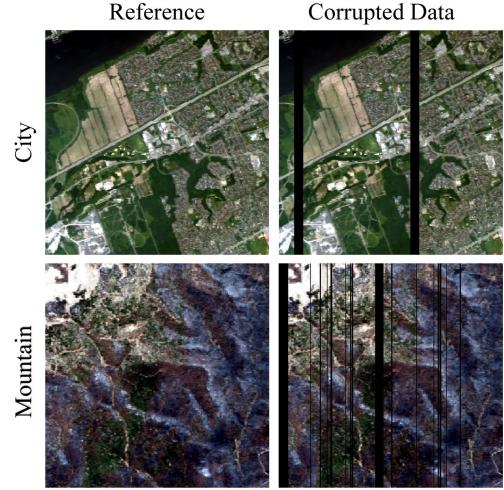


Fig. 1. NASA’s hyperspectral images lead to non-PSD mean-removed data covariance matrices.

where $\mathbf{S} = [s_1, \dots, s_L]$, $x_{m\ell} \triangleq [\mathbf{X}]_{m,\ell}$, $\mathbf{E}_{m,:}$ is the m th row of \mathbf{E} , and d_m is the m th entry of \mathbf{d} . From the above reformulation (10), one can naturally extend the SSID criterion to the case where the data are incompletely observed, i.e.,

$$\min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}} \sum_{(m,\ell) \in \Omega} \frac{1}{2} [x_{m\ell} - (\mathbf{E}_{m,:}\mathbf{s}_\ell + d_m)]^2. \quad (11)$$

Though (11) may be directly solved by alternating optimization, this approach would be quite time-consuming. The reason is that for each s_n , the corresponding subproblem requires us to compute one specific projection matrix $(\mathbf{E}_{\Omega_n}^T \mathbf{E}_{\Omega_n})^{-1} \mathbf{E}_{\Omega_n}^T$, where $\Omega_n \subseteq \{1, \dots, M\}$ collects band indices of observable entries in the n th pixel \mathbf{x}_n , and \mathbf{E}_{Ω_n} is a submatrix of \mathbf{E} collecting rows with indices specified by Ω_n . It is not feasible to precompute all the possible projection matrix partly because there are in total 2^M possible index sets Ω_n , and partly because the variable \mathbf{E} will be updated alternatively. So, judicious optimization trick for solving (11) is necessary.

We will introduce an auxiliary variable, which not only allows us to solve (11) efficiently, but also avoids the aforementioned non-PSD dilemma caused by high missing rate. Specifically, we propose to incorporate the auxiliary variable $\mathbf{Z} \in \mathbb{R}^{M \times L}$, together with an additional constraint, into the SSID criterion (11), leading to

$$\min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}, \mathbf{Z} \in \Omega_Z} \frac{1}{2} \|\mathbf{Z} - (\mathbf{E}\mathbf{S} + \mathbf{d}\mathbf{1}_L^T)\|_F^2 \quad (12)$$

where $\Omega_Z \subseteq \mathbb{R}^{M \times L}$ collects those $M \times L$ matrices whose entries on Ω are the same as \mathbf{X}_{Ω} . Note that with the additional constraint of “ $\mathbf{Z} \in \Omega_Z$,” the SSID criteria (11) and (12) are equivalent, as stated and proved in the following lemma.

Lemma 2: With the constraint “ $\mathbf{Z} \in \Omega_Z$,” the SSID criteria (11) and (12) are equivalent. \square

The proof of Lemma 2 is relegated to Appendix A.

Though equivalent, the formulation (12) has some advantages compared to (11). For example, we no longer need to compute the L projection matrices “ $(\mathbf{E}_{\Omega_n}^T \mathbf{E}_{\Omega_n})^{-1} \mathbf{E}_{\Omega_n}^T$ ” (involving matrix inversion) for each pixel n at each iteration. Moreover, the newly introduced quantity \mathbf{Z} just serves as a

matrix/tensor completion result, and can be used to avoid the aforementioned non-PSD dilemma during the optimization procedure. So, the remaining question is how to implement the SSID criterion (12). At first glance, there are four block variables, but after careful observation, block variables $(\mathbf{E}, \mathbf{S}, \mathbf{d})$ can be merged to form a subproblem that is exactly the SSID problem [with the virtual data \mathbf{Z} ; see (5)]. Note that there is no missing entry in the virtual data \mathbf{Z} , and hence no non-PSD dilemma. Therefore, we alternatively solve the following two subproblems.

1) To update the block variable \mathbf{Z} , we need to solve

$$\mathbf{Z}_i := \arg \min_{\mathbf{Z} \in \Omega_Z} \frac{1}{2} \|\mathbf{Z} - (\mathbf{E}_{i-1} \mathbf{S}_{i-1} + \mathbf{d}_{i-1} \mathbf{1}_L^T)\|_F^2 \quad (13)$$

whose closed-form solution is given as $[\mathbf{Z}_i]_{mn} := [\mathbf{X}_\Omega]_{mn}$ for $(m, n) \in \Omega$ (computed once for all) and $[\mathbf{Z}_i]_{mn} := [\mathbf{E}_{i-1} \mathbf{S}_{i-1} + \mathbf{d}_{i-1} \mathbf{1}_L^T]_{mn}$ for $(m, n) \notin \Omega$.

2) To update the block variable $(\mathbf{E}, \mathbf{S}, \mathbf{d})$, we need to solve

$$(\mathbf{E}_i, \mathbf{S}_i, \mathbf{d}_i) := \arg \min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}} \frac{1}{2} \|\mathbf{Z}_i - (\mathbf{E}\mathbf{S} + \mathbf{d}\mathbf{1}_L^T)\|_F^2$$

which is exactly the SSID problem with the virtual data \mathbf{Z}_i [see (5)] and can be solved by Lemma 1.

The proposed SISHY algorithm has been completed and summarized in Algorithm 2.

Algorithm 2 SISHY Algorithm for SSID Criterion (12)

- 1: **Given** incompletely observed hyperspectral image \mathbf{X}_Ω .
 - 2: Set $i := 1$. Initialize $(\mathbf{E}_0, \mathbf{S}_0, \mathbf{d}_0)$ by Algorithm 1.
 - 3: **repeat**
 - 4: Update \mathbf{Z}_i by the closed-form solution below (13).
 - 5: Update $(\mathbf{E}_i, \mathbf{S}_i, \mathbf{d}_i)$ by applying Lemma 1 on \mathbf{Z}_i .
 - 6: Update $i := i + 1$.
 - 7: **until** the prespecified stopping criterion is met.
 - 8: **Output** the matrix completion result $\mathbf{Z}^* := \mathbf{Z}_{i-1}$ and the SSID result $(\mathbf{E}^*, \mathbf{S}^*, \mathbf{d}^*) := (\mathbf{E}_{i-1}, \mathbf{S}_{i-1}, \mathbf{d}_{i-1})$.
-

Through the above discussion, the auxiliary variable \mathbf{Z} has been associated with a clear physical meaning as the matrix completion result (i.e., image inpainting results) [44]. Thus, in case an image inpainting result is desired, a practically meaningful method to obtain a better matrix completion output \mathbf{Z} is to regularize the solution of the SSID criterion (12), by forcing some desired property ϕ (e.g., self-similarity, sparsity, or continuity) on \mathbf{Z} , leading to

$$\min_{\mathbf{E} \in \mathbb{U}, \mathbf{S}, \mathbf{d}, \mathbf{Z} \in \Omega_Z} \frac{1}{2} \|\mathbf{Z} - (\mathbf{E}\mathbf{S} + \mathbf{d}\mathbf{1}_L^T)\|_F^2 + \beta\phi(\mathbf{Z}) \quad (14)$$

where ϕ is a regularization function tailored to describe the desired structural property of \mathbf{Z} , and β is the regularization parameter. In this work, we test one of the most powerful regularization schemes by designing ϕ to force the image self-similarity structure [45].

Finally, we implement the regularized SSID criterion (14). Again, the four block variables can be merged into two parts, i.e., $(\mathbf{E}, \mathbf{S}, \mathbf{d})$ and \mathbf{Z} . Solving the first part $(\mathbf{E}, \mathbf{S}, \mathbf{d})$ is exactly the same as the unregularized version, and can be addressed by applying Lemma 1 to the virtual data \mathbf{Z}_i [see (5)]. Solving

the second part \mathbf{Z} is trickier due to the implicit form of the self-similarity function ϕ , but this can be efficiently handled by the so-called plug-and-play (PnP) strategy [46]. Specifically, to update \mathbf{Z} in (14), we need to solve

$$\begin{aligned} \mathbf{Z}_i &:= \arg \min_{\mathbf{Z} \in \Omega_Z} \frac{1}{2} \|\mathbf{Z} - (\mathbf{E}_{i-1} \mathbf{S}_{i-1} + \mathbf{d}_{i-1} \mathbf{1}_L^T)\|_F^2 + \beta\phi(\mathbf{Z}) \\ &= \arg \min_{\mathbf{Z} \in \Omega_Z} \frac{1}{2} \|\mathbf{Z} - \mathcal{IM}\|_F^2 + \beta\phi(\mathbf{Z}) \end{aligned} \quad (15)$$

which is exactly the denoising problem. To see it, considering the noisy image (with the noise level $\sigma^2 := \beta$ [47, Eq. (20)]) $\mathcal{IM} \triangleq \mathbf{E}_{i-1} \mathbf{S}_{i-1} + \mathbf{d}_{i-1} \mathbf{1}_L^T$, (15) can be regarded as finding a clean image \mathbf{Z} (close to \mathcal{IM} and on Ω_Z) that has the self-similarity structure ϕ ; in other words, the solution of (15), \mathbf{Z}_i , can be obtained by denoising \mathcal{IM} using a self-similarity denoiser, followed by projecting the denoising result onto Ω_Z (let Π be the projector here). The most well-known self-similarity denoiser would be the block-matching 3-D (BM3D) denoiser [48] (denoted as f) that can be applied to implement (15), where the BM3D denoiser can be applied in the N_p -dimensional eigenimage space to improve the computational efficiency [5]. In this article, we set $N_p := 3$ while the noise-level σ^2 is automatically determined by the noise estimation (NE) method [17]. More detailed analysis of the PnP setting are presented in Section III-C. All in all, the PnP strategy implements the problem type of (15) by $\mathbf{Z}_i := \Pi[f(\mathcal{IM})]$, and the overall algorithm for the regularized SSID criterion (14) is summarized in Algorithm 3.

Algorithm 3 SISHY Algorithm for SSID Criterion (14)

- 1: **Given** incompletely observed hyperspectral image \mathbf{X}_Ω .
 - 2: Set $i := 1$. Initialize $(\mathbf{E}_0, \mathbf{S}_0, \mathbf{d}_0)$ by Algorithm 1.
 - 3: **repeat**
 - 4: Update \mathbf{Z}_i by the $\mathbf{Z}_i := \Pi[f(\mathbf{E}_{i-1} \mathbf{S}_{i-1} + \mathbf{d}_{i-1} \mathbf{1}_L^T)]$.
 - 5: Update $(\mathbf{E}_i, \mathbf{S}_i, \mathbf{d}_i)$ by applying Lemma 1 on \mathbf{Z}_i .
 - 6: Update $i := i + 1$.
 - 7: **until** the prespecified stopping criterion is met.
 - 8: **Output** the matrix completion result $\mathbf{Z}^* := \mathbf{Z}_{i-1}$ and the SSID result $(\mathbf{E}^*, \mathbf{S}^*, \mathbf{d}^*) := (\mathbf{E}_{i-1}, \mathbf{S}_{i-1}, \mathbf{d}_{i-1})$.
-

As will be experimentally demonstrated, in the tough HRS data inpainting mission, Algorithm 3 outperforms Algorithm 2 at the cost of higher computational time. So, Algorithm 2 is also good when the computational time is a key concern, as will be seen in our experimental investigations (see Section III-C). In Section III, we will apply the proposed SISHY algorithm to various critical inverse problems in HRS, and will demonstrate the superiority of SISHY over existing benchmark SSID algorithms.

III. EXPERIMENTAL RESULTS

In this section, we demonstrate the superiority of the proposed SISHY algorithm, which yields strong SSID results from incomplete hyperspectral data and does well improve the performances of several HRS tasks. In Section III-A, we compare SISHY with well-known benchmark SSID algorithms, including P-ADMM [31], IPETRELS [32], NORST-miss [37],

GRASTA [24], ROSETA [29], LSPCA [19], and VBPCA [21]. There are many HRS algorithms that need to perform SSID as their very first step, so considering the incomplete data scenarios, we replace their SSID methods by our SISHY. As it turns out, the replacement of SISHY has improved the HRS performances in several applications, including HU (see Section III-B), hyperspectral inpainting (see Section III-C), and hyperspectral classification (see Section III-D). In this section, for tables that investigating quantitative comparisons, we use boldfaced number to indicate the best performance in a specific scenario. Besides, all experimental comparisons between baselines and SISHY are averaged over ten independent runs to avoid experimental bias.

As for the experimental setting, the penalty parameter, regularization parameter, step-size threshold, and forgetting factor in P-ADMM and IPETRELS, are set as 1.5, 1.5, 0, and 0.999, respectively. The parameters in NORST-miss are set as $K := 32$, $\alpha := 60$, and $\omega_{\text{eval}} := 6.65 \times 10^{-4}$, where the notation usage here exactly follows the article [37] for convenience. GRASTA has a penalty parameter $\rho := 2$, convergence tolerance 1×10^{-8} , and the maximum iteration of 200. The initial step size in ROSETA can be denoted as $\mu_0 = (C2^{-l})/(1 + \eta_0)$, where $C := 10$, $\eta_0 := 80$, and $l := 0$ are the default settings suggested by the authors. The parameters in Fed-OA-RSTMiss (Fed-RST) are set as $\xi := 25$ and $\alpha := 60$, with threshold 7.5961E-4, where the notation usages simply follow from the original paper, and these are default settings in their demo [38]. Furthermore, the subspace basis estimated by the above algorithms are all standardized into an orthogonal one for better SSID performance. All experiments are conducted on the computational equipment with Core-i7-11700 CPU (2.50-GHz speed and 16-GB RAM) under the computing environment MATLAB R2022a.

A. Quantitative SSID Performance Evaluations

To evaluate the SSID performance quantitatively, we need to know the ground-truth hyperspectral subspace. To this end, we synthesize the 224-band hyperspectral image $X = AY + N$ according to a standard procedure [4], where columns of the signature matrix $A \in \mathbb{R}^{224 \times 6}$ are formed by six hyperspectral material signatures randomly selected from the U.S. Geological Survey (USGS) library [49] (i.e., Jarosite, Pyrope, Dumortierite, Buddingtonite, Muscovite, and Goethite), and $N \in \mathbb{R}^{224 \times 16900}$ models the random Gaussian noise corruption of 25 dB signal-to-noise ratio (SNR). As for the abundance matrix $Y \in \mathbb{R}^{6 \times 16900}$, its six rows are formed by the vectorized version of the abundance maps shown in Fig. 2, which well captures the continuity and sparsity nature of hyperspectral abundances and is thus widely used in synthesizing hyperspectral data [50]. As each column of Y forms a set of convex combination coefficients [50], the noise-free version $X = AY$ then defines a 5-D ground-truth subspace \mathcal{S} (i.e., affine hull formed by the above six hyperspectral signature vectors), and the aim of SSID is to exactly identify such a subspace.

To evaluate the similarity between \mathcal{S} and the identified subspace $\widehat{\mathcal{S}}$, we use the frequently adopted measures, including the subspace estimation performance (SEP) [30] and the chordal distance [17], as defined below. Inspired by the Grassmannian

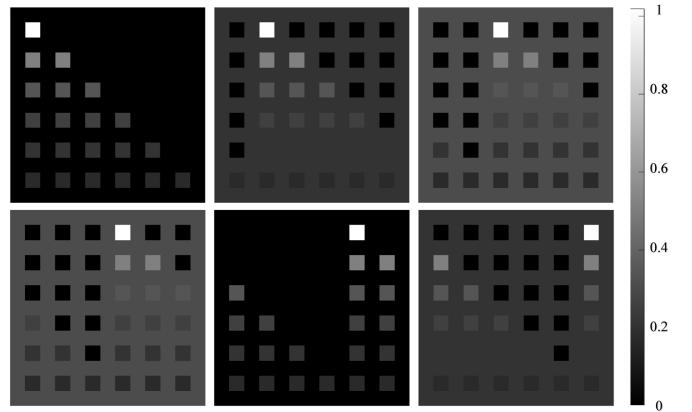


Fig. 2. Ground-truth abundance maps sized of 130×130 commonly used for synthesizing hyperspectral data, where each subblock contains 10×10 pixels.

space theory, the chordal distance (CD) d_{ch} between the two subspaces is defined as

$$d_{\text{ch}}(\mathcal{S}, \widehat{\mathcal{S}}) = \frac{1}{\sqrt{2}} \|\mathbf{U} - \widehat{\mathbf{U}}\|_F$$

where $\mathbf{U} \triangleq \mathbf{E}\mathbf{E}^T$ with \mathbf{E} being an orthonormal basis of \mathcal{S} , and $\widehat{\mathbf{U}} \triangleq \widehat{\mathbf{E}}\widehat{\mathbf{E}}^T$ with $\widehat{\mathbf{E}}$ being an orthonormal basis of $\widehat{\mathcal{S}}$. Another metric for measuring the distance between two subspaces is given by

$$\text{SEP} = \frac{\text{Tr}(\mathbf{E}^T(\mathbf{I} - \widehat{\mathbf{E}}\widehat{\mathbf{E}}^T)\mathbf{E})}{\text{Tr}(\mathbf{E}^T(\widehat{\mathbf{E}}\widehat{\mathbf{E}}^T)\mathbf{E})} \quad (16)$$

where ‘‘Tr’’ denotes the matrix trace operator. For both indices d_{ch} and SEP, a lower index value implies higher similarity between \mathcal{S} and $\widehat{\mathcal{S}}$.

The results are summarized in Figs. 3 and 4, where a wide range of missing rates are tested. Given a missing rate, the missing entries in the hyperspectral data X are randomly selected. Not surprisingly, as the missing rate increases, the performances of most SSID methods tend to become weaker in terms of both chordal distance d_{ch} and SEP. In particular, P-ADMM and IPETRELS are more sensitive to the missing rates. Remarkably, only VBPCA and the proposed SISHY algorithm are robust against large missing rates, as can be seen from their nearly zero values of both chordal distance and SEP. However, the computational time of SISHY (i.e., Algorithm 2) is about $50 \times$ faster than VBPCA, as can be seen from Table I wherein the computational time of each investigated SSID method is displayed. The reason may be that Bayesian learning is often more computationally heavy. In the subsequent experiments, we will further evaluate these SSID methods by applying them to real-world HRS missions.

Beyond the above quantitative evaluation, we perform a convergence analysis for the proposed SISHY based on the same synthesized hyperspectral data with different missing rates (i.e., 40%, 50%, and 60%) and two representative indexes (i.e., SEP and the CD d_{ch}). Specifically, we evaluate the SEP/CD curves of SISHY with various iteration numbers (ranging from 1 to 30) across 40%, 50%, and 60% missing rates, as displayed in Fig. 5. From the convergence curves

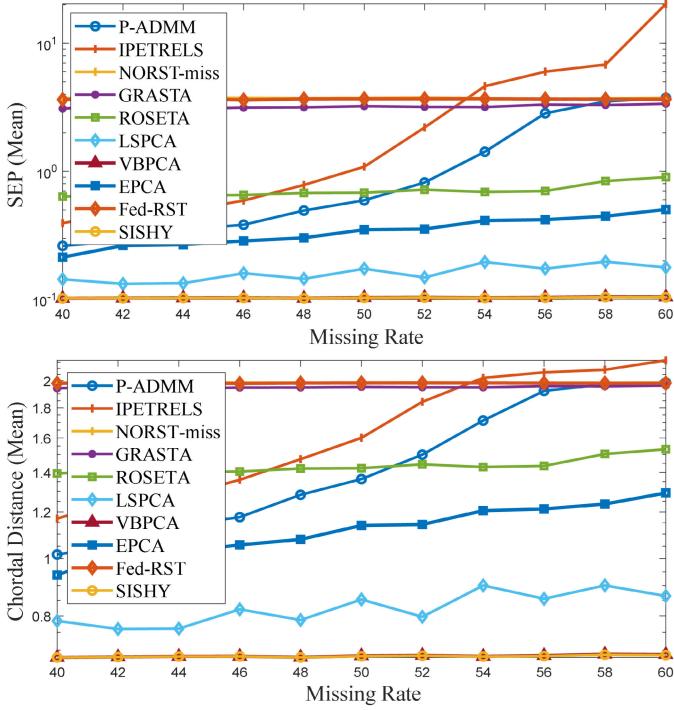


Fig. 3. Chordal distance and SEP mean curves of SSID algorithms over different missing rates.

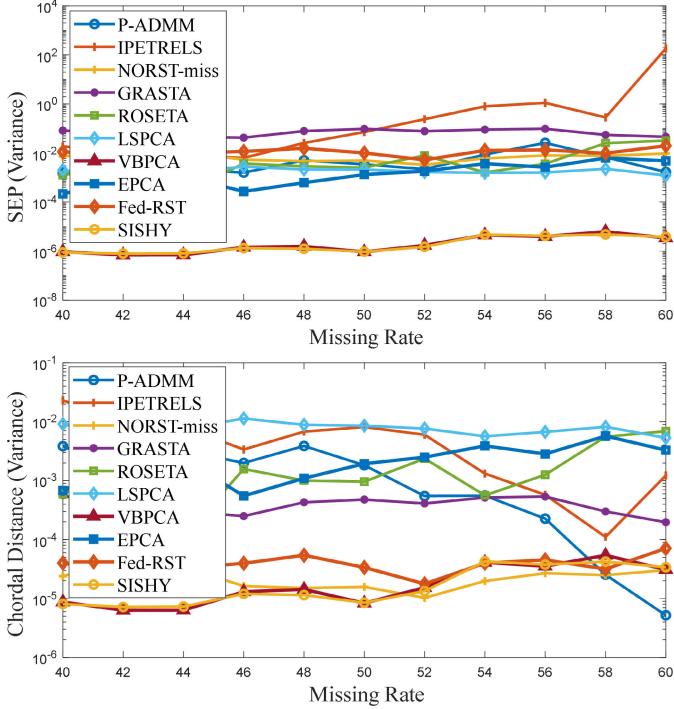


Fig. 4. Chordal distance and SEP variance curves of SSID algorithms over different missing rates.

illustrated in Fig. 5, one can observe that both CD and SEP curves decrease radically (within ten iterations) for all the missing rates, indicating a promising convergence property of SISHY. Even with a large missing rate (e.g., 60%), the proposed SISHY eventually achieves small SEP and CD values. On the other hand, one can also observe that the more challenging case (i.e., larger missing rate) naturally correspond

TABLE I
RUNNING TIME T OF DIFFERENT SSID ALGORITHMS

Methods	T (seconds)
P-ADMM [31]	17.782 ± 0.280
IPETRELS [32]	16.245 ± 0.323
NORST-miss [37]	13.507 ± 0.133
GRASTA [24]	12.179 ± 0.120
ROSETA [29]	18.842 ± 0.255
LSPCA [19]	5.977 ± 0.009
VPCA [21]	65.100 ± 0.468
EPCA [18]	6.949 ± 0.018
Fed-RST [38]	8.114 ± 0.153
SISHY	1.420 ± 0.001

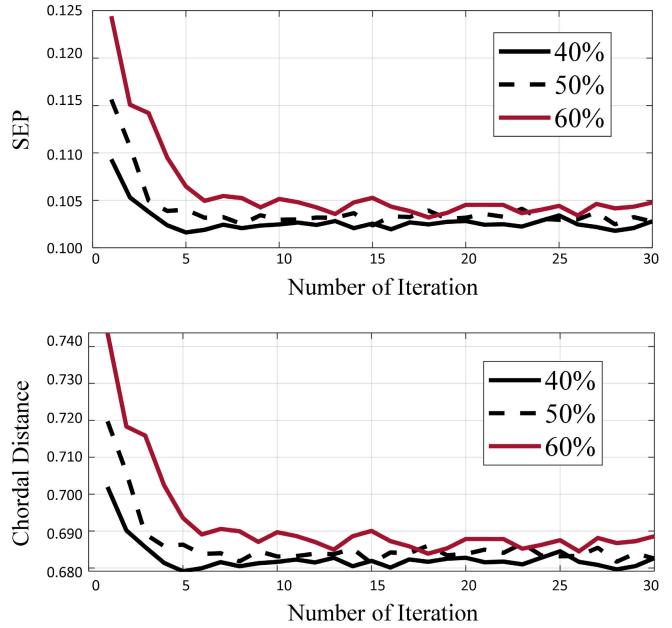


Fig. 5. Convergence curves of SISHY over different number of iterations for different missing rates, and the results are averaged over ten independent runs.

to a slower convergence rate, as expected. In summary, the proposed SISHY has a fast convergence rate, and hence we empirically suggest ten to 15 iterations that are sufficient for effective identification.

B. Evaluation Using Blind Source Separation

In this section, we evaluate the proposed SISHY algorithm by applying it to the critical signal processing task known as blind source separation, which can be employed to conduct the HU task in HRS [51]. The HU mission aims at blindly recovering both the endmember matrix \mathbf{A} (containing the pure material signatures) and the abundance matrix \mathbf{Y} , directly from the observable data $\mathbf{X} \approx \mathbf{AY}$. Each column of the data matrix \mathbf{X} can be viewed as a mixture of the hyperspectral signature vectors (i.e., columns in \mathbf{A}), as introduced in Section III-A. More specifically, we use exactly the same data \mathbf{X} used in Section III-A, and the aim is to recover the hyperspectral signatures of the six minerals (Jarosite, Pyrope, Dumortierite, Buddingtonite, Muscovite, and Goethite) in \mathbf{A} , as well as the abundance maps in \mathbf{Y} . This HU mission can be well done by calling the benchmark HyperCSI algorithm [4] when the hyperspectral image \mathbf{X} is completely observed, but the ratio of

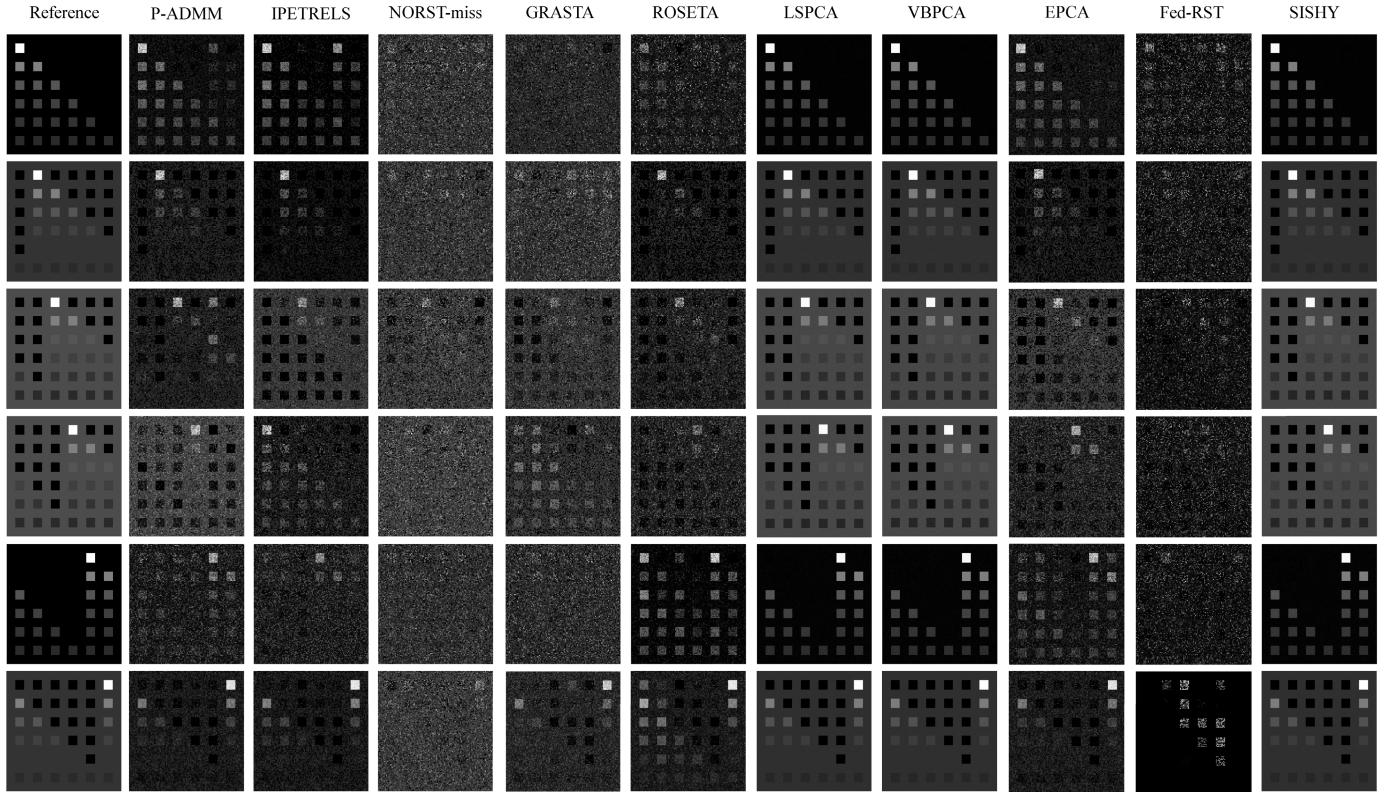


Fig. 6. Visual evaluation of different SSID methods on the source separation task HU using the separated abundance maps, where the first column shows the set of ground-truth abundance maps frequently used in the HRS study [50].

the incompletely observed hyperspectral pixels is up to 60% in this experiment. Many of those missing pixels are located on the boundary of the so-called signature endmember simplex, destroying the data topology that is critical for HU [52].

When there are missing entries in X , HyperCSI does not work well as its very first step is to perform SSID. To solve the dilemma, we perform the SSID step of HyperCSI by those SSID algorithms that can handle the missing data (e.g., SISHY Algorithm 2). As will be seen, when plugging different SSID methods into HyperCSI, the resulting HU performances are quite deviated. To quantitatively evaluate the accuracy of the recovered hyperspectral signatures A in the HU mission, we adopt the frequently used root-mean-square spectral angle error ϕ_{en} for measuring the disparity between true endmembers and estimated endmembers, and the definition of ϕ_{en} can be found in [4, Eq. (32)]. The quantitative results are summarized in Table II, where lower values of ϕ_{en} imply better HU results. The computational time is also displayed in Table II. We also investigate the visual performance by displaying the separated abundance maps Y for the 40 dB SNR scenario in Fig. 6. The HU algorithm, HyperCSI, leverages various SSID algorithms to recover the simplex structure from the incomplete hyperspectral data X .

From Table II, one can see that for some methods (e.g., LSPCA), the SSID performance is quite sensitive to the noise corruption, while the proposed SISHY algorithm exhibits more robust and stable performances. Among the existing peer methods, the performance of VBPCA is still outstanding. P-ADMM also yields pretty good quantitative HU results,

TABLE II
QUANTITATIVE EVALUATION OF DIFFERENT SSID METHODS ON THE BLIND SOURCE SEPARATION TASK (I.E., HU), IN TERMS OF ϕ_{en} (DEGREES) AND AVERAGE RUNNING TIME T (SECONDS)

Methods	ϕ_{en} (degrees)					T (sec.)
	20	25	30	35	40	
P-ADMM	3.652	3.699	3.418	5.547	3.340	28.175
IPETRELS	5.081	9.904	3.574	6.594	4.778	28.562
NORST-miss	9.832	9.805	10.409	10.545	11.168	14.881
GRASTA	9.935	12.198	9.675	10.287	10.898	14.814
ROSETA	9.227	8.086	7.476	7.945	5.431	22.736
LSPCA	7.692	1.652	1.205	1.150	1.144	6.323
VBPCA	1.879	1.187	0.750	0.475	0.330	109.883
EPCA	4.021	3.690	3.540	3.461	3.380	1.630
Fed-RST	11.287	11.252	10.064	9.393	10.805	5.988
SISHY	1.579	1.055	0.705	0.454	0.319	1.228

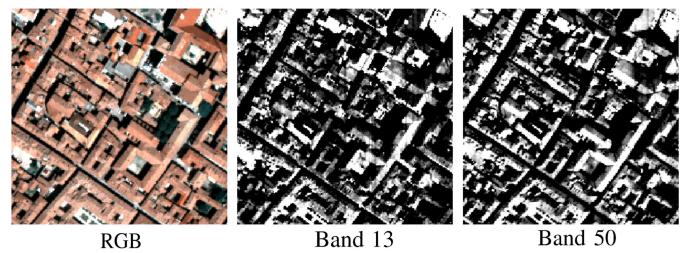


Fig. 7. Ground truth of some investigated bands of the Pavia Center hyperspectral data acquired over northern Italy.

showing the value of the convex analysis approach. On the other hand, we evaluate the HU performance by examining the visual quality of the separated abundance maps, as shown in

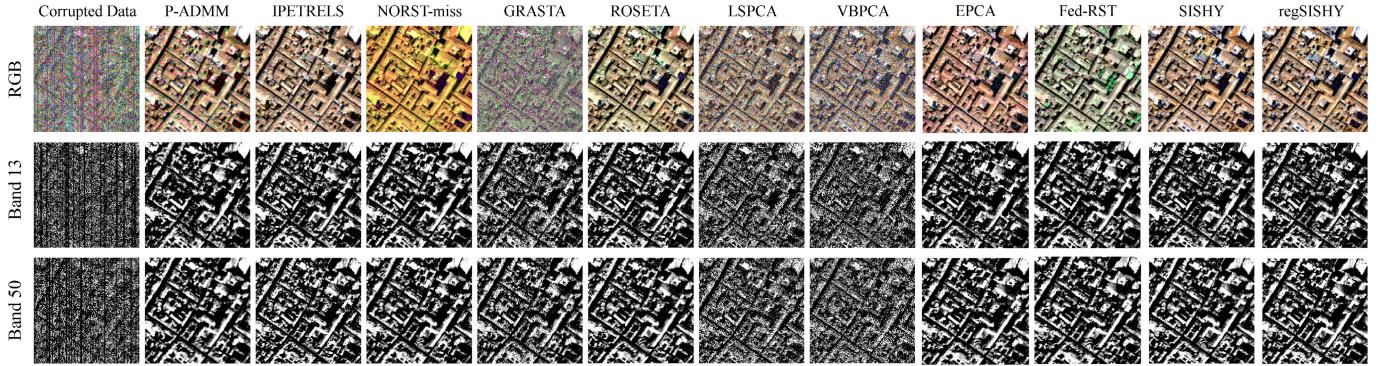


Fig. 8. Inpainting results from highly damaged Pavia Center (northern Italy) hyperspectral image.

Fig. 6. The incompletely observed hyperspectral data do cause negative impacts on the final HU outcome, as can be clearly seen from columns 2 to 6 in Fig. 6. One can also observe that only LSPCA, VBPCA, as well as the proposed SISHY algorithm, still yield visually pleasant HU results, as can be seen from the high-quality abundance map recovery results in Fig. 6. Finally, one should notice that while SISHY achieves the state-of-the-art performances, SISHY is more than one order of magnitude faster than P-ADMM, and about two orders of magnitude faster than VBPCA (see Table II). This is truly remarkable.

C. Evaluation Using Image Inpainting

In HRS, another critical task is image inpainting because remotely sensed images are quite often incompletely acquired or missing during transmission, and reconstruction of hyperspectral image often relies on the SSID technology. The hyperspectral inpainting (HI) task can be well done by calling the FastHyIn algorithm [5], but FastHyIn performs the SSID stage by simply removing those incomplete pixels. Nevertheless, as discussed previously, those incomplete pixels should still contain useful information for better SSID. So, in this experiment, we study the impact on the final HI performance of FastHyIn, using different SSID methods that can handle the missing data scenarios (i.e., SSID methods that can use the information of those incomplete pixels). Besides, we further analyze different aspects of the proposed SISHY (e.g., parameter sensitivity and limitations) with respect to the final HI performance of FastHyIn. Hence the remaining parts of the HI experiments are summarized as follows. In Section III-C1, we organize the data and quantitative metrics for the HI experiment. Then, a comprehensive quantitative and qualitative evaluation is presented in Section III-C2. Next, we conduct a concise analysis of parameter selection for regSISHY in Section III-C3. Finally, we discuss the limitation of our method in Section III-C4.

1) *Data Description:* This study is conducted on the widely investigated Pavia Center data [5], captured by the reflective optics system imaging spectrometer (ROSI) sensor over northern Italy. The clean version of some representative bands in the studied 201×201 region of interest (ROI) is shown in Fig. 7. There are in total 80 bands used in this HI experiment,

where each band is first corrupted by Gaussian noise of standard deviation 0.26, followed by further corrupting the data by the commonly seen stripe pattern with 15% missing stripes [10], leading to the corrupted data shown in the first column of Fig. 8.

2) *Qualitative and Quantitative Evaluation:* To evaluate the HI results achieved with different SSID methods, we adopt the structural similarity (SSIM), peak SNR (PSNR), and spectral angle mapper (SAM), as the performance indices [10]. Higher value of SSIM indicates better global reconstruction; higher value of PSNR implies better spatial reconstruction; lower value of SAM means better spectral reconstruction. The quantitative evaluation results are summarized in Table IV, where one can see that the unregularized SISHY algorithm (i.e., Algorithm 2) has already yielded superior HI results than existing SSID methods in terms of all the three indices. Not surprisingly, the regularized version of SISHY, denoted by regSISHY (i.e., Algorithm 3), yields even stronger results than SISHY in terms of all the three indices, at the cost of longer computational time. Fig. 8 shows the qualitative evaluation of the HI results, where one can see that only IPETRELS, SISHY, and regSISHY yield visually pleasant HI results. The visual result obtained from P-ADMM has slight luminance distortion, and other peer methods either fail to reconstruct the image or suffer from some color distortion.

In this HI experiment, ROSETA is the strongest peer method (see Tables III and IV), and its performance is nearly the same as the regSISHY algorithm in terms of computational time and SSIM/PSNR/SAM. Though the self-similarity structure promoted by regSISHY has improved the HI performance, the unregularized SISHY is also outstanding. Remarkably, even for this strongest peer method, the unregularized SISHY algorithm is already sufficient to outperform ROSETA, but SISHY is now nearly $20\times$ faster than ROSETA after removing the regularization mechanism. So, the unregularized version of SISHY is sufficiently good especially when the computational time is a key concern.

Mathematically, we did not make any specific assumption for the types of remote sensing images; instead, our technique is developed merely based on those very general and widely accepted properties of hyperspectral images (e.g., self-similarity). Thus, our method should be very general, and this has been experimentally found to be true. Due to space

TABLE III

QUANTITATIVE EVALUATION OF DIFFERENT SSID METHODS ON THE INPAINTING MISSION (WITHOUT OUTLIER CORRUPTION) FOR THE PAVIA CENTER HYPERSPECTRAL DATA. “NORST” DENOTES “NORST-MISS” TO CONSTRAINT THE TABLE WIDTH. FOR EACH INDEX, WE USE RED COLOR TO MARK THE BEST METHOD, AND USE BLUE COLOR TO MARK THE SECOND-BEST METHOD

Methods	P-ADMM	IPETRELS	NORST	GRASTA	ROSETA	LSPCA	VPCA	EPCA	Fed-RST	SISHY	regSISHY
Mean SSIM(\uparrow)	0.668	0.694	0.681	0.278	0.681	0.381	0.379	0.798	0.690	0.802	0.805
Mean PSNR(\uparrow)	27.489	27.786	25.235	25.450	29.170	20.357	20.374	30.403	25.255	30.367	30.424
Mean SAM(\downarrow)	13.200	12.844	15.956	21.367	11.445	31.425	31.656	9.693	15.959	9.764	9.726
Var SSIM	0.000	0.001	0.000	0.001	0.017	0.000	0.000	0.000	0.000	0.000	0.000
Var PSNR	0.129	0.454	0.329	0.485	0.178	0.050	0.037	0.054	0.334	0.056	0.054
Var SAM	0.212	0.292	0.195	4.678	0.360	0.006	0.001	0.014	0.255	0.015	0.012
Running Time	24.530	23.785	10.392	8.252	16.434	6.001	95.264	0.940	10.348	0.802	14.330

TABLE IV

QUANTITATIVE EVALUATION OF DIFFERENT SSID METHODS ON THE INPAINTING MISSION (WITH OUTLIER CORRUPTION) FOR THE PAVIA CENTER HYPERSPECTRAL DATA. “NORST” DENOTES “NORST-MISS” TO CONSTRAINT THE TABLE WIDTH. FOR EACH INDEX, WE USE RED COLOR TO MARK THE BEST METHOD, AND USE BLUE COLOR TO MARK THE SECOND-BEST METHOD

Methods	P-ADMM	IPETRELS	NORST	GRASTA	ROSETA	LSPCA	VPCA	EPCA	Fed-RST	SISHY	regSISHY
Mean SSIM(\uparrow)	0.623	0.626	0.627	0.237	0.523	0.370	0.361	0.688	0.600	0.670	0.677
Mean PSNR(\uparrow)	20.949	20.898	21.098	18.894	19.425	19.232	19.440	21.602	20.145	21.577	21.545
Mean SAM(\downarrow)	10.717	10.753	11.321	26.535	18.758	20.292	20.469	8.326	15.659	8.937	8.631
Var SSIM	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
Var PSNR	0.067	0.081	0.044	0.171	0.130	0.045	0.030	0.067	0.178	0.065	0.068
Var SAM	0.024	0.048	0.075	4.067	1.081	0.007	0.004	0.004	1.640	0.002	0.001
Running Time	26.988	25.498	10.147	8.292	17.302	6.197	93.828	1.764	10.907	1.052	15.158

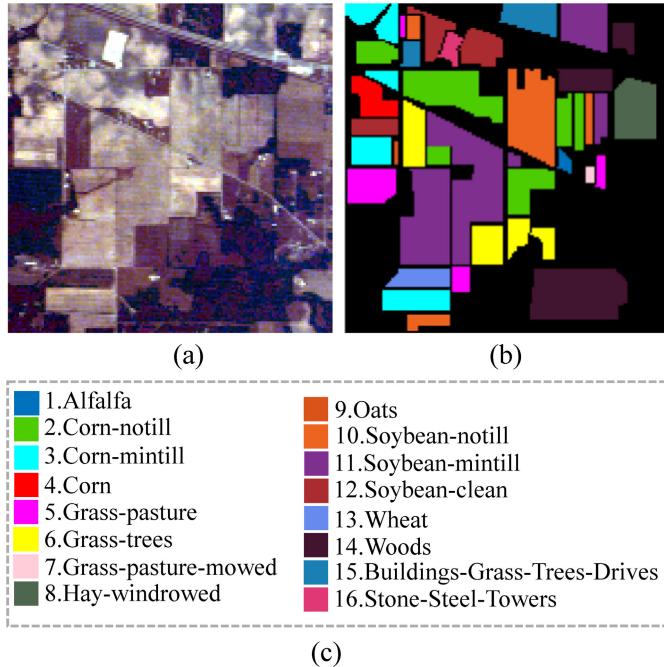


Fig. 9. Indian Pine hyperspectral data. (a) RGB composition. (b) Ground-truth classification map. (c) Class number and color index.

limitation, the results are omitted here, but we remark that by testing other image types like the ShellKey data used in [10], we found that our regSISHY algorithm still shows outstanding performance than other peer methods.

3) *Parameter Sensitivity*: Next, we investigate the parameter sensitivity of regSISHY based on the final HI result for better comprehension. In this evaluation, we jointly optimize the regularization parameter β (corresponding to the

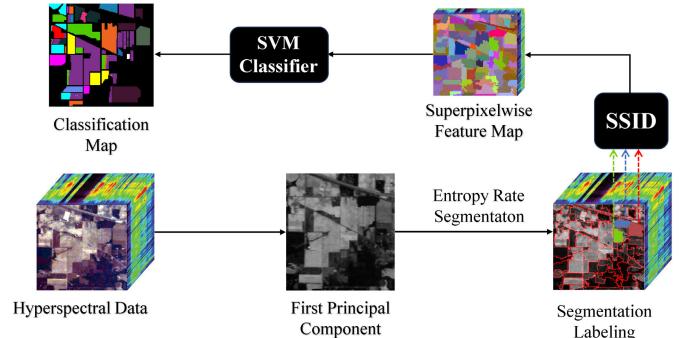


Fig. 10. Schematic of SuperPCA classification method.

noise level σ^2 [47]) and the eigenspace dimension $N_p \in \{1, \dots, 9\}$. Specifically, when N_p is fixed, the optimal β can be automatically computed using the NE method [17], as discussed in Section II. Thus, in the joint optimization, the outer loop is for N_p with the automatic inner loop for β , leading to the optimal setting of $N_p^* = 3$ with respect to PSNR. At the optimal point, we have $\sigma^* \equiv \sqrt{\beta^*} = 4E-5$, and we found that $N_p^* = 3$ well matches our default setting (see Section II).

4) *Discussion*: After demonstrating the superiority of our methods, we further discuss their limitations. Hence, we incorporate a widely occurring scenario in HRS, i.e., outlier interference [53], to investigate the limitations of the proposed SISHY and regSISHY. More specifically, we leverage the same stripe pattern and noise type mentioned in Section III-C2, followed by adding the outlier interference and setting signal-to-outlier-ratio (SOR) as SOR: = 10 dB, where the approach of generating outlier and the precise definition of SOR can be found in [54]. The quantitative results are summarized in

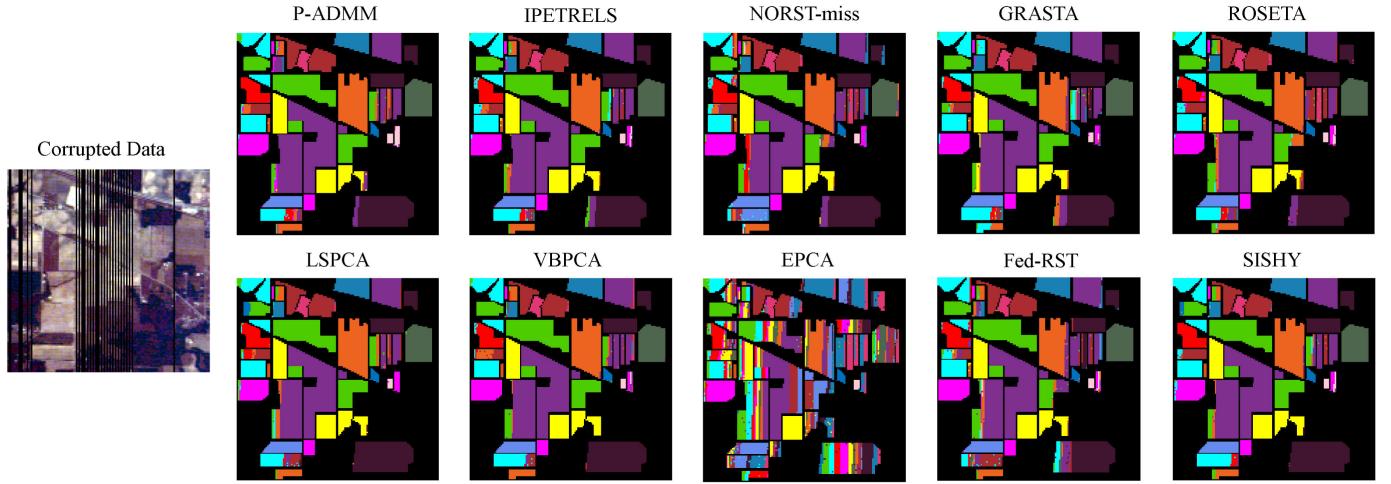


Fig. 11. Classification maps using different SSID algorithms tested on the Indian Pines data with serious stripe missing pattern.

TABLE V
CLASSIFICATION PERFORMANCE OF SUPERPCA USING DIFFERENT SSID METHODS ON THE INDIAN PINES DATA

Class No.	P-ADMM	IPETRELS	NORST-miss	GRASTA	ROSETA	LSPCA	VBPCA	EPCA	Fed-RST	SISHY
1	97.83±0.05	97.83±0.05	97.93±2.12	100±0.00	98.70±0.04	98.70±0.04	97.83±0.05	92.61±0.91	96.09±0.05	98.70±0.04
2	86.60±0.27	89.48±0.24	78.12±0.30	82.12±0.24	79.12±0.15	85.90±0.21	88.73±0.16	23.97±0.12	79.83±0.30	88.56±0.17
3	86.11±0.07	86.64±0.07	66.61±0.14	83.19±0.13	84.26±0.08	86.61±0.45	90.90±0.10	30.84±0.40	71.09±0.14	90.92±0.16
4	94.59±0.22	93.33±0.20	84.30±0.29	94.30±0.14	92.17±0.32	92.95±0.26	93.19±0.17	49.76±0.34	88.55±0.21	95.27±0.16
5	93.66±0.04	94.15±0.03	93.73±0.02	91.74±0.04	94.13±0.03	95.45±0.07	94.61±0.08	65.65±0.32	91.77±0.06	96.27±0.08
6	96.77±0.01	97.29±0.04	96.01±0.06	95.11±0.12	95.56±0.36	97.20±0.09	98.24±0.02	44.94±0.39	92.49±0.17	98.19±0.02
7	95.71±0.23	96.43±0.23	93.57±0.35	96.43±0.23	95.71±0.31	97.14±0.12	96.43±0.49	90.71±1.02	96.43±0.23	97.14±0.12
8	100±0.00	100±0.00	94.35±0.20	100±0.00	100±0.00	99.64±0.00	99.64±0.00	36.29±0.42	95.67±0.14	99.64±0.00
9	88.00±0.89	89.00±8.85	85.00±1.76	88.00±1.29	82.00±2.24	96.00±0.24	96.00±0.25	88.00±1.56	88.00±1.56	100±0.00
10	86.88±0.29	87.17±0.31	82.61±0.64	85.05±0.21	83.46±0.44	90.76±0.12	90.23±0.14	39.96±0.18	77.92±0.59	90.31±0.13
11	94.91±0.14	93.36±0.17	85.44±0.34	92.90±0.05	87.86±0.09	92.35±0.12	90.54±0.04	26.01±0.43	83.98±0.24	93.68±0.07
12	92.77±0.11	91.72±0.14	87.67±0.19	89.43±0.11	86.89±0.19	89.43±0.12	89.50±0.19	57.42±0.26	83.06±0.06	92.58±0.08
13	99.43±0.00	99.43±0.00	95.77±0.00	99.43±0.00	99.43±0.00	99.43±0.00	99.43±0.00	65.66±0.43	99.43±0.00	99.43±0.00
14	92.93±0.35	93.18±0.31	90.53±0.28	88.36±0.47	92.87±0.34	95.91±0.18	98.16±0.21	24.84±0.16	80.33±0.28	95.34±0.45
15	95.81±0.39	95.84±0.41	91.26±0.29	94.30±0.26	93.43±0.47	97.87±0.01	95.65±0.08	66.18±0.16	81.57±0.12	98.71±0.00
16	96.67±0.04	96.83±0.03	96.83±0.09	98.73±0.03	96.03±0.13	97.30±0.07	97.30±0.04	67.46±0.60	96.83±0.16	98.10±0.05
OA(%)	92.29±0.03	92.40±0.03	85.39±0.03	89.73±0.01	88.45±0.02	92.31±0.03	92.81±0.02	36.01±0.02	83.16±0.03	93.63±0.01
AA(%)	93.67±0.01	93.87±0.05	88.71±0.02	92.44±0.01	91.35±0.03	94.54±0.01	94.78±0.02	54.40±0.02	87.69±0.02	95.80±0.00
$\kappa \times 100$	91.16±0.04	91.29±0.04	83.36±0.03	88.23±0.01	86.79±0.02	91.20±0.04	91.78±0.03	30.18±0.02	80.72±0.04	92.70±0.02

Table IV. With the outlier interference, not only the proposed SISHY and regSISHY expose their limitations, but the other peer methods are also supremely influenced by the outlier corruption. Even though the proposed SISHY and regSISHY are not very robust against the outlier interference, they still show their robustness with a performance superior than most peer methods. Furthermore, SISHY has the fastest inference time for the outlier-corrupted data. We may take the outlier effect into account for the future research.

D. Evaluation Using HRS Classification

Finally, the SuperPCA classification is conducted in this article based on two renowned HRS data, i.e., Indian Pines, and Houston 2018 data [8]. In Section III-D1, we perform a comprehensive experiment using the Indian Pines data. On the other hand, the experiment using the Houston 2018 data is presented in Section III-D2.

1) *Indian Pines Dataset*: The Indian Pine dataset, whose ROI is shown in Fig. 9(a) with the ground-truth labeling given in (b), is captured by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor [55], and the tested hyperspectral cube is of size $145 \times 145 \times 200$ after removing water-vapor

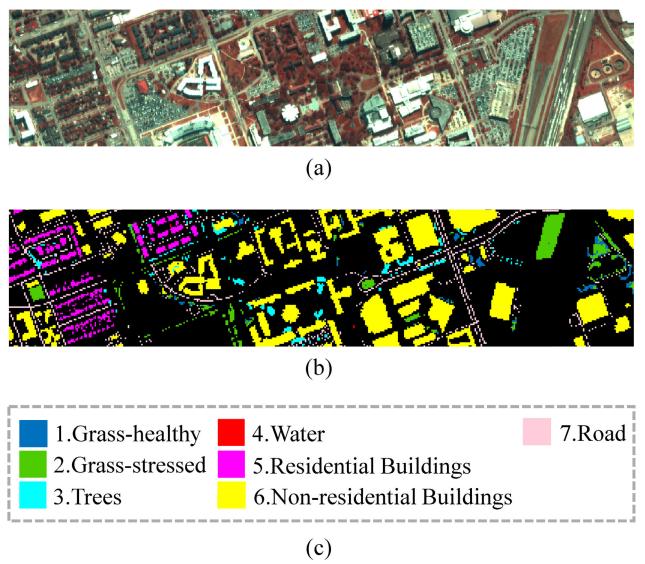


Fig. 12. Houston 2018 hyperspectral data. (a) RGB composition. (b) Ground-truth classification map. (c) Class number and color index.

corruption bands. In the reference label, 16 different classes [see Fig. 9(c)] with 10 249 labeled pixels and corresponding

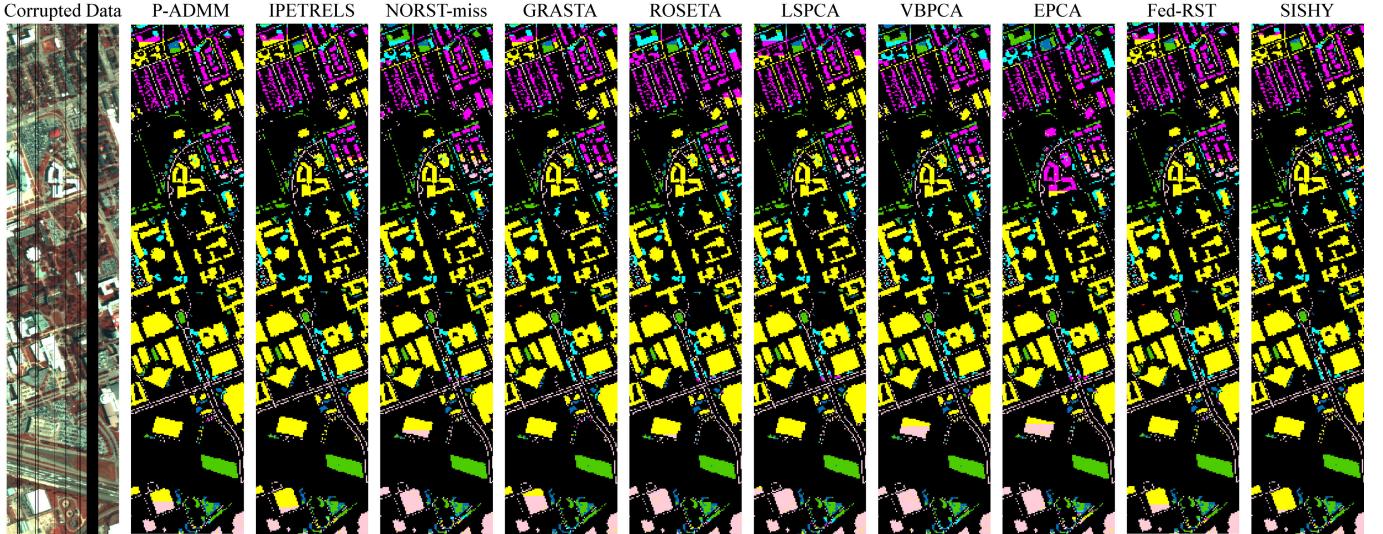


Fig. 13. Classification maps using different SSID algorithms tested on the high-resolution Houston 2018 data with serious stripe missing pattern.

classification map are shown in Fig. 9(b), and readers may refer to [6] for more details. To see how SuperPCA relies on effective SSID, its flowchart is given in Fig. 10. The dimension of the subspace (i.e., number of features) and the number of superpixels are set to 30, and 100, respectively. For the SVM classifier (see Fig. 10), the training samples are randomly selected 30 samples from each class; for those classes with pixels fewer than 30 (e.g., Oats class), just half of the pixels are selected as the training samples.

In the scenarios with missing data, the resulting insufficient training samples often lead to weaker classification results because SuperPCA relies on the SSID technique, as can be seen from Fig. 10. We adopt the stripe missing corruption pattern widely encountered in the remote sensing area, as displayed in Fig. 11, where about 23% regions are damaged with each damaged pixel (i.e., pixel covered by the stripe missing pattern) containing 11 missing spectral bands. For this challenging scenario, the SuperPCA classification results achieved by different SSID methods are summarized in Fig. 11 and Table V. From the classification maps in Fig. 11, one can see that P-ADMM, IPETRELS, GRASTA, VBPCA, and the proposed SISHY have relatively good visual quality, especially in the lower right corner (woods category). To fairly evaluate the classification performance, the quantitative classification accuracy for each category has been summarized in the upper half of Table V, which well echoes the visual results in Fig. 11. Specifically, among the peer methods, P-ADMM, IPETRELS, GRASTA, and VBPCA have relatively good quantitative performance, while the proposed SISHY algorithm still yields the best classification performance; this can be seen from the averaged accuracy (AA) over the 16 categories, for which SISHY achieves the state-of-the-art result of 95.8% (see Table V). Since kappa coefficient (κ) and overall accuracy (OA) are also commonly used quantitative performance measures [6], they are also displayed in Table V. Again, the proposed SISHY algorithm yields the state-of-the-art performance in terms of both κ and OA.

2) *Houston 2018 Dataset*: The Houston 2018 dataset, whose ROI is shown in Fig. 12(a) with the ground-truth labeling given in 12(b), is obtained by the national center for airborne laser mapping (NCALM) over the University of Houston [56], and the tested hyperspectral cube is of size $210 \times 954 \times 48$ with a high spatial resolution (i.e., GSD 1 m). In the reference label, 7 different classes [see Fig. 12(c)] with 53 200 labeled pixels and corresponding classification map are shown in Fig. 12(b), and more details can be found in [8]. The dimension of the subspace and the number of superpixels are set to 30 and 50, respectively. For the SVM classifier, the training samples are randomly selected from each class, resulting in a total of 5% of labeled samples for training.

In the Houston scenarios, a serious stripe missing corruption pattern is displayed in Fig. 13, where about 15% regions are damaged with each damaged pixel containing 25% missing spectral bands. For this scenario, we summarize the SuperPCA classification results achieved by different SSID methods in Fig. 13 and Table VI. From Fig. 13, P-ADMM, IPETRELS, GRASTA, Fed-RST, and the proposed SISHY yield promising qualitative results. However, we remark that the proposed SISHY still outperforms these benchmarks. More specifically, SISHY achieves more precise classification results in the yellow class (i.e., nonresidential buildings), which can be observed in the bottom-left and upper-left corners of the classification map. To further evaluate the classification performance, we also conduct a quantitative comparison of each class and the results are summarized in Table VI. From Table VI, one can observe that SISHY achieves high accuracy in numerous classes (e.g., grass-healthy and grass-stressed). Besides, SISHY has superior accuracy in the sixth class (i.e., nonresidential buildings), which well echoes the qualitative comparison. Though the performance of GRASTA is truly outstanding with the best AA index, the proposed SISHY algorithm again achieves the best performance in terms of both the OA and κ on the Houston 2018 dataset.

TABLE VI
CLASSIFICATION PERFORMANCE OF SUPERPCA USING DIFFERENT SSID METHODS ON THE HOUSTON 2018 DATA

Class No.	P-ADMM	IPETRELS	NORST-miss	GRASTA	ROSETA	LSPCA	VBCPA	EPCA	Fed-RST	SISHY
1	95.50±0.03	95.06±0.02	94.42±0.02	95.57±0.03	94.23±0.03	90.74±0.11	95.98±0.01	93.48±0.02	94.67±0.02	95.39±0.01
2	92.40±0.03	92.50±0.02	90.29±0.02	93.01±0.03	91.11±0.02	93.34±0.03	90.46±0.02	83.98±0.02	92.86±0.02	93.32±0.01
3	92.51±0.06	92.21±0.05	86.54±0.02	93.50±0.05	87.60±0.03	84.92±0.04	88.17±0.01	85.29±0.02	92.69±0.03	91.18±0.01
4	92.73±0.63	92.73±0.63	84.55±1.33	97.27±0.17	85.45±1.19	22.73±7.65	08.18±1.23	88.18±0.84	90.91±0.83	83.64±1.45
5	87.81±0.03	87.25±0.04	84.60±0.04	88.14±0.02	83.23±0.03	81.90±0.03	84.85±0.01	81.86±0.00	88.06±0.03	87.47±0.02
6	90.88±0.01	90.89±0.01	85.42±0.05	90.02 ±0.00	84.73±0.01	81.98±0.04	84.14±0.02	72.36±0.02	90.40±0.01	92.73±0.00
7	85.41±0.12	85.10±0.13	78.79±0.01	86.95±0.08	79.89±0.02	70.23±0.19	70.28±0.09	69.69±0.01	85.77±0.07	82.36±0.01
OA(%)	90.23±0.00	90.13±0.00	85.20±0.02	90.00±0.00	84.87±0.00	81.86±0.01	83.51±0.00	74.93±0.01	90.02±0.00	91.04±0.01
AA(%)	91.03±0.04	90.82±0.04	86.37±0.03	92.07±0.02	86.61±0.03	75.12±0.17	74.58±0.04	82.12±0.02	90.76±0.04	89.44±0.04
$\kappa \times 100$	83.26±0.01	83.07±0.01	75.28±0.03	83.00±0.01	74.89±0.01	70.37±0.02	72.54±0.01	61.05±0.01	82.96±0.01	84.35±0.00

IV. CONCLUSION AND FUTURE WORK

We have developed the SISHY technique for computationally efficient subspace identification of incomplete hyperspectral data, and also for physically interpretable data matrix completion as a byproduct. We have clearly presented the design philosophy of the proposed SSID criteria, leading to three SSID algorithms. Algorithm 1 is naive but suitable for cases of low missing rates, while Algorithm 2 goes a step further to avoid the non-PSD covariance matrix approximation and to bypass tremendous amount of inverse matrix computations. Algorithm 3 allows future users to flexibly plug suitable regularizers of interest to get physically meaningful data matrix completion results, which also alludes better SSID results, and such a flexibility actually introduces a new degree of freedom for future users to explore other potential applications of SISHY. We have demonstrated the use of self-similarity regularization, and shown that the induced SSID problem can be judiciously transformed into a denoising problem, thereby allowing us to solve it elegantly and efficiently using a benchmark self-similarity denoiser. Comprehensive experiments have proven the superiority of SISHY not only in accurate SSID but, more importantly, in really improving the efficacy of the subsequent HRS missions (e.g., HU, inpainting, and classification). Based on the concise form of the developed SSID algorithms, further application to implement the quantum low-rank module of HyperQUEEN (i.e., hyperspectral quantum deep network) [13] for quantum image compression is an interesting future research line.

APPENDIX

A. Proof of Lemma 2

Since $\Omega_Z \subseteq \mathbb{R}^{M \times L}$ is defined as the set collecting those $M \times L$ matrices whose entries on Ω are the same as X_Ω , the constraint “ $Z \in \Omega_Z$ ” implies that

$$\begin{aligned} \|Z - (ES + d\mathbf{1}_L^T)\|_F^2 &= \sum_{\Omega} [z_{m\ell} - (E_{m,:}s_\ell + d_m)]^2 + \sum_{\bar{\Omega}} [z_{m\ell} - (E_{m,:}s_\ell + d_m)]^2 \\ &= \sum_{\Omega} [x_{m\ell} - (E_{m,:}s_\ell + d_m)]^2 + \sum_{\bar{\Omega}} [z_{m\ell} - (E_{m,:}s_\ell + d_m)]^2 \end{aligned}$$

for which we make two observations. First, the first term

$$f_1(E, S, d) \triangleq \sum_{(m,\ell) \in \Omega} [x_{m\ell} - (E_{m,:}s_\ell + d_m)]^2$$

becomes a constant term with respect to (w.r.t.) Z . Second, as $Z \in \Omega_Z$ does not pose any constraint on the domain $\bar{\Omega}$, we observe that minimization of the second term

$$f_2(E, S, d, Z) \triangleq \sum_{(m,\ell) \in \bar{\Omega}} [z_{m\ell} - (E_{m,:}s_\ell + d_m)]^2$$

w.r.t. Z yields zero, i.e.,

$$\min_{Z \in \Omega_Z} f_2(E, S, d, Z) = \min_{Z \in \Omega_Z} \sum_{\bar{\Omega}} [z_{m\ell} - (E_{m,:}s_\ell + d_m)]^2 = 0.$$

Based on the above observations that the first term becomes a constant w.r.t. Z , and that the second term vanishes when optimizing w.r.t. Z , we can reformulate (12) as follows:

$$\begin{aligned} &\min_{E \in \mathbb{U}, S, d, Z \in \Omega_Z} \frac{1}{2} \|Z - (ES + d\mathbf{1}_L^T)\|_F^2 \\ &= \min_{E \in \mathbb{U}, S, d} \min_{Z \in \Omega_Z} \|Z - (ES + d\mathbf{1}_L^T)\|_F^2 \\ &= \min_{E \in \mathbb{U}, S, d} \min_{Z \in \Omega_Z} f_1(E, S, d) + f_2(E, S, d, Z) \\ &= \min_{E \in \mathbb{U}, S, d} \left[f_1(E, S, d) + \min_{Z \in \Omega_Z} f_2(E, S, d, Z) \right] \\ &= \min_{E \in \mathbb{U}, S, d} \left[f_1(E, S, d) + 0 \right] \\ &= \min_{E \in \mathbb{U}, S, d} \sum_{(m,\ell) \in \Omega} [x_{m\ell} - (E_{m,:}s_\ell + d_m)]^2 \end{aligned}$$

which is exact the SSID criterion (11). Therefore, the SSID criteria (11) and (12) are indeed equivalent, and the proof has been completed. ■

ACKNOWLEDGMENT

The authors thank the Taiwan Space Agency (TASA) and the National Center for High-performance Computing (NCHC) for providing the computing resources.

REFERENCES

- C. Lin, W. Ma, W. Li, C. Chi, and A. Ambikapathi, “Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, Oct. 2015.
- C.-H. Lin, R. Wu, W.-K. Ma, C.-Y. Chi, and Y. Wang, “Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization,” *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 1651–1679, Jan. 2018.
- 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, Apr. 2005.

- [4] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1946–1961, Apr. 2016.
- [5] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
- [6] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
- [7] Y. Zhang, W. Li, W. Sun, R. Tao, and Q. Du, "Single-source domain expansion network for cross-scene hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 32, pp. 1498–1512, 2023.
- [8] Y. Zhang, W. Li, M. Zhang, Y. Qu, R. Tao, and H. Qi, "Topological structure and semantic information transfer network for cross-scene hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 6, pp. 2817–2830, Jun. 2023.
- [9] M. Zhang, X. Zhao, W. Li, Y. Zhang, R. Tao, and Q. Du, "Cross-scene joint classification of multisource data with multilevel domain adaption network," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–13, Apr. 2023, doi: [10.1109/TNNLS.2023.3262599](https://doi.org/10.1109/TNNLS.2023.3262599).
- [10] C.-H. Lin, Y.-C. Lin, and P.-W. Tang, "ADMM-ADAM: A new inverse imaging framework blending the advantages of convex optimization and deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5514616.
- [11] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. Boca Raton, FL, USA: CRC Press, 2017.
- [12] T.-H. Lin and C.-H. Lin, "Hyperspectral change detection using semi-supervised graph neural network and convex deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5515818.
- [13] C.-H. Lin and Y.-Y. Chen, "HyperQUEEN: Hyperspectral quantum deep network for image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5511720.
- [14] C.-H. Lin and T.-H. Lin, "All-addition hyperspectral compressed sensing for metasurface-driven miniaturized satellite," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5503215.
- [15] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 1, pp. 65–74, Feb. 1988.
- [16] C.-I. Chang and A. Ifarraguerri, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2529–2538, 2000.
- [17] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [18] K. Ji, Y. Shen, Q. Chen, and T. Feng, "Extended principal component analysis for spatiotemporal filtering of incomplete heterogeneous GNSS position time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5801419.
- [19] A. Ilin and T. Raiko, "Practical approaches to principal component analysis in the presence of missing values," *J. Mach. Learn. Res.*, vol. 11, no. 66, pp. 1957–2000, 2010.
- [20] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Statist. Soc. B, Stat. Methodol.*, vol. 61, no. 3, pp. 611–622, Jan. 1999.
- [21] Q. Jiang, X. Yan, and B. Huang, "Neighborhood variational Bayesian multivariate analysis for distributed process monitoring with missing data," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2330–2339, Nov. 2019.
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.
- [23] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Sep. 2010, pp. 704–711.
- [24] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1568–1575.
- [25] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5947–5959, Dec. 2013.
- [26] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.
- [27] J. Feng, H. Xu, and S. Yan, "Online robust PCA via stochastic optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 404–412.
- [28] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2663–2677, May 2015.
- [29] H. Mansour and X. Jiang, "A robust online subspace estimation and tracking algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4065–4069.
- [30] N. Linh-Trung, V.-D. Nguyen, M. Thameri, T. Minh-Chinh, and K. Abed-Meraim, "Low-complexity adaptive algorithms for robust subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1197–1212, Dec. 2018.
- [31] L. T. Thanh, N. V. Dung, N. L. Trung, and K. Abed-Meraim, "Robust subspace tracking with missing data and outliers via ADMM," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, A Coruna, Spain, Sep. 2019, pp. 1–5.
- [32] L. T. Thanh, N. V. Dung, N. L. Trung, and K. Abed-Meraim, "Robust subspace tracking with missing data and outliers: Novel algorithm with convergence guarantee," *IEEE Trans. Signal Process.*, vol. 69, pp. 2070–2085, 2021.
- [33] W. Stacklies, H. Redestig, M. Scholz, D. Walther, and J. Selbig, "PcaMethods—A bioconductor package providing PCA methods for incomplete data," *Bioinformatics*, vol. 23, no. 9, pp. 1164–1167, May 2007.
- [34] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, "Recursive robust PCA or recursive sparse recovery in large but structured noise," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 5007–5039, Aug. 2014.
- [35] P. Narayananmurthy and N. Vaswani, "Provable dynamic robust PCA or robust subspace tracking," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1547–1577, Mar. 2019.
- [36] P. Narayananmurthy and N. Vaswani, "Nearly optimal robust subspace tracking," in *Proc. Int. Conf. Mach. Learn.*, Stockholm, Sweden, 2018, pp. 3701–3709.
- [37] P. Narayananmurthy, V. Daneshpajoh, and N. Vaswani, "Provable subspace tracking from missing data and matrix completion," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4245–4260, Aug. 2019.
- [38] P. Narayananmurthy, N. Vaswani, and A. Ramamoorthy, "Federated over-air subspace tracking from incomplete and corrupted data," *IEEE Trans. Signal Process.*, vol. 70, pp. 3906–3920, 2022.
- [39] C. M. Bishop, "Variational principal components," in *Proc. 9th Int. Conf. Artif. Neural Networks*, Edinburgh, U.K., Sep. 1999, pp. 509–514.
- [40] S. Roweis, "EM algorithms for PCA and SPCA," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 10, 1997, pp. 626–632.
- [41] J. Lutinnen, A. Ilin, and J. Karhunen, "Bayesian robust PCA of incomplete data," *Neural Process. Lett.*, vol. 36, no. 2, pp. 189–202, Oct. 2012.
- [42] C. Lin, C. Chi, L. Chen, D. J. Miller, and Y. Wang, "Detection of sources in non-negative blind source separation by minimum description length criterion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4022–4037, Sep. 2018.
- [43] C. Lin and J. M. Bioucas-Dias, "Nonnegative blind source separation for ill-conditioned mixtures via John ellipsoid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2209–2223, May 2021.
- [44] C.-H. Lin, Y. Liu, C.-Y. Chi, C.-C. Hsu, H. Ren, and T. Q. S. Quek, "Hyperspectral tensor completion using low-rank modeling and convex functional analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, Feb. 2023, doi: [10.1109/TNNLS.2023.3243808](https://doi.org/10.1109/TNNLS.2023.3243808).
- [45] C.-H. Lin and J. M. Bioucas-Dias, "An explicit and scene-adapted definition of convex self-similarity prior with application to unsupervised Sentinel-2 super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3352–3365, May 2020.
- [46] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, 2020.
- [47] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, Nov. 2021.

- [48] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [49] R. Clark et al. (2007). *USGS Digital Spectral Library Splib06a: U.S. Geological Survey, Digital Data Series 231*. [Online]. Available: <https://www.usgs.gov/labs/spectroscopy-lab>
- [50] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [51] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [52] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.
- [53] T.-H. Chan, A. Ambikapathi, W.-K. Ma, and C.-Y. Chi, "Robust affine set fitting and fast simplex volume max-min for hyperspectral end-member extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3982–3997, Jul. 2013.
- [54] Y.-R. Syu, C.-H. Lin, and C.-Y. Chi, "An outlier-insensitive unmixing algorithm with spatially varying hyperspectral signatures," *IEEE Access*, vol. 7, pp. 15086–15101, 2019.
- [55] (Feb. 6, 2013). *AVIRIS USA Hyperspectral Data Cube*. [Online]. Available: <https://aviris.jpl.nasa.gov/>
- [56] B. Le Saux, N. Yokoya, R. Hansch, and S. Prasad, "2018 IEEE GRSS data fusion contest: Multimodal land use classification [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 1, pp. 52–54, Mar. 2018.



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

From 2015 to 2016, he was a Visiting Student of Virginia Tech, Arlington, VA, USA. He is currently an Associate Professor with the Department of Electrical Engineering and the Miin Wu School of Computing, National Cheng Kung University (NCKU), Tainan, Taiwan. Before joining NCKU, he held research positions at The Chinese University of Hong Kong, Hong Kong, in 2014 and 2017, NTHU from 2016 to 2017, and the University of Lisbon (ULisboa), Lisbon, Portugal, from 2017 to 2018. He was an Assistant Professor with the Center for Space and Remote Sensing Research, National Central University, Taoyuan City, Taiwan, in 2018, and a Visiting Professor with ULisboa, in 2019. His research interests include network science, quantum computing, convex geometry and optimization, blind signal processing, and imaging science.

Dr. Lin received the Emerging Young Scholar Award from National Science and Technology Council (NSTC) in 2023, the Future Technology Award from NSTC in 2022, the Outstanding Youth Electrical Engineer Award from The Chinese Institute of Electrical Engineering (CIEE) in 2022, the Best Young Professional Member Award from IEEE Tainan Section in 2021, the Prize Paper Award from IEEE Geoscience and Remote Sensing Society (GRS-S) in 2020, the Top Performance Award from Social Media Prediction Challenge at ACM Multimedia in 2020, and The 3rd Place from AIM Real World Super-Resolution Challenge at IEEE International Conference on Computer Vision (ICCV) in 2019. He received the Ministry of Science and Technology (MOST) Young Scholar Fellowship, together with the EINSTEIN Grant Award, from 2018 to 2023. In 2016, he was a recipient of the Outstanding Doctoral Dissertation Award from the Chinese Image Processing and Pattern Recognition Society and the Best Doctoral Dissertation Award from the IEEE GRS-S.



Si-Sheng Young (Student Member, IEEE) is currently pursuing the Ph.D. degree with the Intelligent Hyperspectral Computing Laboratory (IHCL), Department of Electrical Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan.

His research interests include convex optimization, deep learning, anomaly detection, data fusion, and imaging inverse problems.

Mr. Young received the Merit Award from The Grand Challenge "Computing for the Future," Miin Wu School of Computing, NCKU, in 2023, as well as the highly competitive "Pan Wen Yuan Scholarship" from the Industrial Technology Research Institute, Hsinchu, Taiwan.