

Hyperspectral Compressive Sensing Using Manifold-Structured Sparsity Prior

Lei Zhang[†], Wei Wei^{*†}, Yanning Zhang[†], Fei Li[†], Chunhua Shen[‡], Qinfeng Shi[‡]

[†]School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, China

[‡]School of Computer Science, The University of Adelaide, Australia

[†]zhanglei211@mail.nwpu.edu.cn, [†]{weiweinwpu, ynzhang}@nwpu.edu.cn, [‡]{chunhua.shen, javen.shi}@adelaide.edu.au

Abstract

To reconstruct hyperspectral image (HSI) accurately from a few noisy compressive measurements, we present a novel manifold-structured sparsity prior based hyperspectral compressive sensing (HCS) method in this study. A matrix based hierarchical prior is first proposed to represent the spectral structured sparsity and spatial unknown manifold structure of HSI simultaneously. Then, a latent variable Bayes model is introduced to learn the sparsity prior and estimate the noise jointly from measurements. The learned prior can fully represent the inherent 3D structure of HSI and regulate its shape based on the estimated noise level. Thus, with this learned prior, the proposed method improves the reconstruction accuracy significantly and shows strong robustness to unknown noise in HCS. Experiments on four real hyperspectral datasets show that the proposed method outperforms several state-of-the-art methods on the reconstruction accuracy of HSI.

1. Introduction

Hyperspectral image (HSI) is a 3D data cube which contains a series of 2D spatial images over continuous spectral bands and each pixel has a spectrum [5]. Its abundant spectral information is helpful for object identification, which facilitates a variety of applications on HSI [22, 1]. However, the high cost on imaging, storage and transmission incurred by the huge volume of HSI limits its application.

Recently, compressive sensing (CS) provides a brand-new framework for image acquisition and compression. It has been proved that a sparse signal can be recovered from a few compressive measurements under some mild conditions [12]. Thus only a few measurements need to be captured during the imaging procedure, which greatly reduces the resource expense on imaging. Since HSI can be converted into a sparse signal by transformation [16] or unmixing [23] strategies, hyperspectral compressive sensing (HCS) meth-

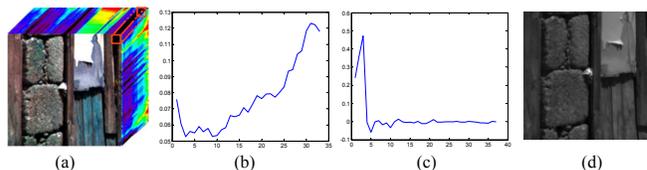


Figure 1. Illustration of spectral sparsity and spatial correlation in the wavelet transformation of HSI. (a) HSI cube with a marked pixel. (b) The spectrum of the marked pixel. (c) The wavelet transformation coefficient vector of the marked spectrum. (d) The image composed of the first transformation coefficient of spectra from all pixels according to pixels' spatial arrangement.

ods [16, 20, 18] have drawn much attention for HSI compression. However, how to reconstruct HSI accurately from a few noisy measurements is still challenging.

Regularization methods [18, 20, 16] are effective to deal with the reconstruction in HCS. One of their most important concerns is to learn a proper sparsity prior for HSI. Since high spectral resolution gives a continuous spectrum at each pixel, HSI can be sparsified pixel by pixel (see Figure 1) and a group of sparse vectors are generated. Though many regularizers (e.g., ℓ_0 norm and ℓ_1 norm etc.) can depict the sparsity in each of those sparse vectors, they usually treat coefficients in the vector independently without considering its structure, which is crucial to improving the reconstruction accuracy of standard sparse learning [17].

For those sparse vectors generated from HSI, two kinds of structures are often considered. One is the structure in each sparse vector, the other is the structure among different sparse vectors (see Figure 1, where the transformation coefficient vectors of all pixels are spatially correlated). They are termed as intra-vector structure and inter-vector structure in this study. To explore the intra-vector structure, Chen [9] proposed a tree structured sparsity prior for the wavelet transformation coefficients of image. Wipf et al. [27] represented the implicit structure in a sparse vector by an empirical Bayes framework. Zhang et al. [28] proposed a matrix-based reweighted Laplace prior to capture the structure in each column vector from a sparse matrix assuming those column vectors are independent. For inter-vector structure, Cotter et al. [10] extended the classical fo-

*The corresponding author.

cal undetermined system solver to multiple vectors to explore the inter-vector structure. Zhang et al. [29, 26] adopted a block sparse Bayes learning framework to learn the correlation among multiple vectors (state-of-the-art method on inter-vector structure based CS).

Though these two structures based methods can improve the reconstruction accuracy over those methods without considering the structure information, they can only capture either the spectral or spatial structure in HSI. Actually both the spectral and spatial structures are crucial for modeling HSI. Moreover, 3D structure of HSI (i.e. the correlation among any two coefficients in HSI) requires these two structures to be further correlated. For simplicity, we term their interrelation result as the 3D structured sparsity in this study. However, none of those existing methods can capture such kind of 3D structured sparsity of HSI.

In this study, we propose a novel manifold-structured sparsity based HCS (MSHCS) method to reconstruct HSI accurately from noisy measurements. A matrix based hierarchical prior is proposed to represent the spectral structured sparsity and spatial unknown manifold structure of HSI simultaneously, which represents the 3D structured sparsity of HSI. To the best of our knowledge, this is the first paper to capture the 3D structured sparsity of HSI. To make this prior fit the 3D structure of HSI well and be robust to the unknown noise in HCS, a latent variable Bayes model is introduced to learn the sparsity prior and estimate the noise jointly from measurements. Thus, with this learned prior, the proposed method improves the reconstruction accuracy significantly and shows strong robustness to unknown noise in HCS. Experiments on four real hyperspectral datasets demonstrate that the proposed method outperforms other 6 state-of-the-art HCS methods on the reconstruction accuracy. For example, the proposed MSHCS exceeds other methods on PSNR at least 4db on the Face dataset when the sampling rate is 0.09 and the SNR of measurements is 15db.

2. The Proposed Method

Given a HSI of n_b bands as $\mathcal{X} \in \mathbb{R}^{n_r \times n_c \times n_b}$, which contains n_r rows and n_c columns in the image of each band, we reshape the image of each band in \mathcal{X} as a row vector to form a 2D matrix $X \in \mathbb{R}^{n_b \times n_p}$, where $n_p = n_r \times n_c$. Each column of X denotes the spectrum of one pixel, while each row of X denotes a vectorized image from one band. For convenience, we term the column and row of X as the spectral dimension and spatial dimension in this study. In HCS, X is sampled by a random sampling matrix $A \in \mathbb{R}^{m_b \times n_b}$ ($m_b < n_b$) along the spectral dimension as

$$F = AX + N, \quad (1)$$

where $F \in \mathbb{R}^{m_b \times n_p}$ is the noisy compressive measurements of X and $N \in \mathbb{R}^{m_b \times n_p}$ denotes noise. In this study,

we mainly focus on reconstructing X from the noisy measurements F given A . Since X is not sparse, a dictionary $\Psi \in \mathbb{R}^{n_b \times n_d}$ is employed to transform X into a group of sparse coefficient vectors $Y \in \mathbb{R}^{n_d \times n_p}$ as $X = \Psi Y$. Each column of Y , denoted as $Y_{.i}$, is a sparse vector. Ψ is known in this study, which can be orthogonal basis ($n_b = n_d$), over-complete basis ($n_b < n_d$) or dictionary learned from training examples. Hence, the reconstruction can be reduced to estimating the most likely Y from F as

$$Y_{opt} = \arg \max_Y p(Y|F) \propto p(F|Y) p(Y). \quad (2)$$

To this end, noise N is assumed to obey a matrix normal distribution $\mathcal{MN}(0, \Sigma_n, I)$ and the rows of N are uncorrelated, where I denotes an identity matrix with proper size. Thus, $\Sigma_n = \text{diag}(\boldsymbol{\lambda})$ is a $\boldsymbol{\lambda}$ -dependent diagonal matrix¹, where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_{m_b}]^T$. Since $X = \Psi Y$, the likelihood of HCS can be formulated as

$$p(F|Y, \boldsymbol{\lambda}) = \frac{\exp\left\{-\frac{1}{2} \|A\Psi Y - F\|_{\Sigma_n}^2\right\}}{(2\pi)^{m_b n_p / 2} |\Sigma_n|^{n_p / 2}}, \quad (3)$$

where $\|Q\|_{\Sigma_n} = \sqrt{\text{tr}(Q^T \Sigma_n^{-1} Q)}$ represents a weighted trace norm. In the following subsections, we will introduce how to model the prior $p(Y)$ and estimate its parameters.

2.1. Manifold-Structured Sparsity Prior

As discussed before, X can be transformed into a group of coefficient vectors Y in Ψ -determined space, which has 3D structured sparsity. To model this 3D structured sparsity, we propose a novel manifold-structured sparsity prior with a hierarchical structure as Figure 2 shows.

First, we represent the sparse matrix Y with a matrix norm distribution as

$$p(Y|\Sigma_{ry}, \Sigma_{cy}) = \frac{\exp\left\{-\frac{1}{2} \text{tr}(\Sigma_{cy}^{-1} Y^T \Sigma_{ry}^{-1} Y)\right\}}{(2\pi)^{n_d n_p / 2} |\Sigma_{ry}|^{n_p / 2} |\Sigma_{cy}|^{n_d / 2}}, \quad (4)$$

where covariance matrix Σ_{ry} depicts the correlation among the coefficients within each sparse vector of Y and Σ_{cy} describes the correlation among all of these sparse vectors.

2.1.1 Spectral Structured Sparsity

Previous works [11, 23] show that Laplace distribution is appropriate to model the distribution of image sparsity. However, Laplace prior is unable to capture the structure in sparse vector and it puts undemocratic penalization on nonzero coefficients [7]. To alleviate these problems, we adopt the reweighted Laplace prior proposed in [28] to represent the spectral structured sparsity in each column vector

¹For a vector \boldsymbol{x} , $\text{diag}(\boldsymbol{x})$ denotes a diagonal matrix with elements from \boldsymbol{x} . For a matrix X , $\text{diag}(X)$ denotes extracting the diagonal elements from X to form a vector.

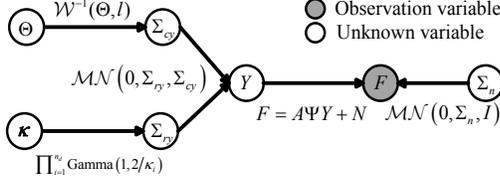


Figure 2. The hierarchical structure of the manifold-structured sparsity prior based HCS.

of Y . Specifically, we first define $\Sigma_{ry} = \text{diag}(\gamma)$, where $\gamma = [\gamma_1, \dots, \gamma_{n_d}]^T$. Then, a Gamma distribution is imposed on the unknown γ .

$$p(\gamma|\kappa) = \prod_{i=1}^{n_d} \text{Gamma}\left(1, \frac{2}{\kappa_i}\right) = \prod_{i=1}^{n_d} \frac{\kappa_i}{2} \exp\left(-\frac{\kappa_i \gamma_i}{2}\right). \quad (5)$$

When $\Sigma_{cy} = I$, the hierarchical prior in Eqs. (4) (5) equals to a reweighted Laplace prior on each Y_i as [28], which can capture the structured sparsity within each sparse vector.

2.1.2 Spatial Manifold Structure

The inter-vector structure of Y can be viewed as an unknown manifold [15, 19], which is implicitly modeled by describing the correlation among all sparse vectors in Y as [4, 15] have done. Since Σ_{cy} in Eq. (4) describes the correlation among all sparse vectors in matrix Y , a reasonable Σ_{cy} is capable to represent the desired manifold structure in Y , which will be clarified further in subsection 2.4. To learn Σ_{cy} more flexibly, we further assume Σ_{cy} obey an inverse-Wishart distribution as

$$p(\Sigma_{cy}|\Theta, l) = \mathcal{W}^{-1}(\Theta, l) = \frac{|\Theta|^{l/2} \exp\left\{-\frac{1}{2}\text{tr}(\Theta \Sigma_{cy}^{-1})\right\}}{2^{n_p l/2} \Gamma_{n_p}(l/2) |\Sigma_{cy}|^{(n_p+l+1)/2}}, \quad (6)$$

where l is a constant implying the freedom degree, Γ_{n_p} is a multivariate Gamma function and $\Theta \in \mathbb{R}^{n_p \times n_p}$ is the reference covariance matrix. In this prior, Σ_{cy} is encouraged to fit the reference covariance matrix Θ by minimizing the Bregman divergence between Σ_{cy} and Θ . Through learning Σ_{cy} , the hierarchical prior represented in Eqs. (4) (6) can capture the unknown manifold structure robustly to random noise (We will clarify this in Subsection 2.4).

The proposed manifold-structured sparsity prior unifies the spectral structured sparsity and spatial manifold structure into a matrix normal distribution as Eq. (4). It is noticeable that $p(Y|\Sigma_{ry}, \Sigma_{cy}) \neq p(Y|\Sigma_{ry}, I)p(Y|I, \Sigma_{cy})$ in general case, which implies that the proposed prior considers the correlation between spectral structured sparsity of each vector and the manifold structure among those sparse vectors. Consequently, the proposed sparsity prior can represent the 3D structured sparsity of HSI. More evidence will be provided in Subsection 2.4.

2.2. Latent Variable based Sparsity Prior Learning

To make the proposed prior fit the distribution of desired HSI and robust to unknown noise, a latent variable Bayes model is introduced to learn the noise variance λ , the prior-related parameters κ , γ , Σ_{cy} and Θ from the noisy measurements. Let $\mathbf{f} = \text{vec}(F)$, $\mathbf{y} = \text{vec}(Y)$, $\mathbf{n} = \text{vec}(N)$ and $\Phi = I \otimes (A\Psi)$, the model in Eq. (1) amounts to $\mathbf{f} = \Phi \mathbf{y} + \mathbf{n}$, where \otimes denotes the Kronecker product. Then, the likelihood in Eq. (3) can be represented as

$$p(\mathbf{f}|\mathbf{y}, \lambda) = \frac{\exp\left\{-\frac{1}{2}\|\mathbf{f} - \Phi \mathbf{y}\|_{I \otimes \Sigma_n}^2\right\}}{(2\pi)^{m_b n_p/2} |I \otimes \Sigma_n|^{1/2}}. \quad (7)$$

Similarly, the sparsity prior in Eq. (4) equals to

$$p(\mathbf{y}|\gamma, \Sigma_{cy}) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^T \Sigma_y^{-1} \mathbf{y}\right)}{(2\pi)^{n_d n_p/2} |\Sigma_y|^{1/2}}, \quad \Sigma_y = \Sigma_{cy} \otimes \Sigma_{ry}. \quad (8)$$

Thus, all unknown variables can be inferred by

$$\begin{aligned} & \max_{\lambda \geq 0, \gamma \geq 0, \kappa, \Sigma_{cy}, \Theta} p(\lambda, \gamma, \kappa, \Sigma_{cy}, \Theta | \mathbf{f}) \propto \\ & = \int p(\mathbf{f}|\mathbf{y}, \lambda) p(\mathbf{y}|\gamma, \Sigma_{cy}) p(\gamma|\kappa) p(\Sigma_{cy}|\Theta, l) d\mathbf{y}, \end{aligned} \quad (9)$$

where the latent variable \mathbf{y} is integrated out and flat priors are implicitly adopted for $p(\lambda)$, $p(\kappa)$ and $p(\Theta)$ for computational convenience. It can be proved that the optimization in Eq. (9) equals to minimizing the following cost function

$$\begin{aligned} \mathcal{L}(\lambda, \gamma, \kappa, \Sigma_{cy}, \Theta) &= -2 \log p(\lambda, \gamma, \kappa, \Sigma_{cy}, \Theta | \mathbf{f}) \\ &\equiv \mathbf{f}^T \Sigma_{by}^{-1} \mathbf{f} + \log |\Sigma_{by}| + \sum_{i=1}^{n_d} (\kappa_i \gamma_i - 2 \log \kappa_i) \\ &\quad + \text{tr}(\Theta \Sigma_{cy}^{-1}) + (n_p + l + 1) \log |\Sigma_{cy}| - l \log |\Theta|, \end{aligned} \quad (10)$$

where $\Sigma_{by} = I \otimes \Sigma_n + \Phi \Sigma_y \Phi^T$ and the constant term is removed. With the learned variables by minimizing Eq. (10), we can directly recover the sparse signal Y by the following maximum a posterior(MAP) estimation.

$$\mathbf{y}_{opt} = \arg \max_{\mathbf{y}} p(\mathbf{y}|\mathbf{f}) \propto p(\mathbf{f}|\mathbf{y}, \lambda) p(\mathbf{y}|\gamma, \Sigma_{cy}). \quad (11)$$

However, it is complicated to solve the nonconvex optimization in Eq. (10) directly. Instead, we transform the cost function into an intuitive regularized regression formula as [20, 18]. Similar as [27], $\mathbf{f}^T \Sigma_{by}^{-1} \mathbf{f}$ can be reformulated as

$$\mathbf{f}^T \Sigma_{by}^{-1} \mathbf{f} = \min_{\mathbf{y}} \|\Phi \mathbf{y} - \mathbf{f}\|_{I \otimes \Sigma_n}^2 + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y}. \quad (12)$$

Substituting Eq. (12) into Eq. (10), we can integrate sparse signal recovery, sparsity prior learning and noise estimation into a unified optimization framework as:

$$\min_{\mathbf{y}, \lambda \geq 0, \gamma \geq 0, \kappa, \Sigma_{cy}, \Theta} \mathcal{L}(\mathbf{y}, \lambda, \gamma, \kappa, \Sigma_{cy}, \Theta), \quad (13)$$

where

$$\begin{aligned} \mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \Sigma_{cy}, \Theta) &= \|\Phi \mathbf{y} - \mathbf{f}\|_{I \otimes \Sigma_n}^2 + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y} \\ &+ \log |\Sigma_{by}| + \sum_{i=1}^{n_d} (\kappa_i \gamma_i - 2 \log \kappa_i) + \text{tr}(\Theta \Sigma_{cy}^{-1}) \\ &+ (n_p + l + 1) \log |\Sigma_{cy}| - l \log |\Theta|. \end{aligned} \quad (14)$$

It can be proved as [27] that minimizing $\mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \Sigma_{cy}, \Theta)$ and $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \Sigma_{cy}, \Theta)$ result in the same solution over $\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \Sigma_{cy}, \Theta$. Therefore, this unified framework equals to minimizing Eq. (10) assisted with a MAP estimation in Eq. (11). *In this framework, on one hand, the estimated noise will result in the shape of sparsity prior to be regulated for better fitting the image distribution. On the other hand, the better learned sparsity prior will promote the accuracy of noise estimation. Thus, this framework guarantees the accurate reconstruction of HSI from noisy measurements.*

2.3. Optimization Procedure

In this section, we will first estimate the reference covariance matrix Θ from measurements, then optimize the remaining unknown variables from Eq. (13).

2.3.1 Reference Covariance Matrix Estimation

Since the sparse matrix Y preserves the manifold structure of X in spatial domain [15], Θ can be defined based on the manifold structure of X to provide a prior guess of Σ_{cy} . Assuming the manifold structure of X is defined on a fully connected graph [25], Θ can be denoted as the inverse of the Laplacian matrix of the graph [15] as $\Theta^{opt} = (D - M)^{-1}$. M is the similarity weight matrix of X and D is a diagonal matrix with $D_{ii} = \sum_j M_{ij}$, where M_{ij} denotes the element of M at position (i, j) . Generally, M is defined by the Gaussian similarity function [25] over two pixels of X as

$$M_{ij} = \exp(-\|X_{.i} - X_{.j}\|_2^2 / \sigma), \quad (15)$$

where the similar pixel pair $(X_{.i}, X_{.j})$ is given a larger weight M_{ij} and vice versa. σ is a predetermined scalar. However, the unknown X makes computing M intractable.

In CS, it has been proved that compressive measurements preserve the geometry structure of the high-dimensional signal well (i.e., $\|X_{.i} - X_{.j}\|_2^2 \approx \|F_{.i} - F_{.j}\|_2^2$ with a small distortion) when the sampling matrix satisfies the restricted isometry property (RIP) [3]. In this study, Gaussian random matrix is employed as the sampling matrix, which satisfies RIP with high probability [2]. Therefore, the measurements F preserves the geometrical structure of X well and M can be defined directly on measurements F . To improve the robustness to noise, local feature in spatial domain of F is utilized to define M as

$$M_{ij} = \exp(-\|\mathcal{N}_i^F - \mathcal{N}_j^F\|_F^2 / \sigma), \quad (16)$$

where $\mathcal{N}_i^F \in \mathbb{R}^{m_b \times k^2}$ contains all measurements within a square neighborhood of fixed size $k \times k$ centered at point $F_{.i}$ in the spatial domain of F . $F_{.i}$ denotes the measurements of pixel $X_{.i}$. $\|\cdot\|_F$ denotes the Frobenius norm. Thus, \mathcal{N}_i^F records the local feature centered at point $F_{.i}$ and $\|\mathcal{N}_i^F - \mathcal{N}_j^F\|_F^2$ denotes the dissimilarity between local features, which is more robust to the noise than that defined between two individual points as Eq. (15). Then, $\Theta = (D - M)^{-1}$, which gives a good approximation to Θ^{opt} and avoids the over-fitting resulted from estimating Θ by the unified framework in Eq. (13).

2.3.2 Parameter Estimation

Since learning parameters in the vector space as Eq. (13) results in the cost of computation and memory to be prohibitive, we back-map the algorithm into original matrix space as Eq. (1) to reduce the cost by an approximation as

$$(I \otimes \Sigma_n + \Phi \Sigma_y \Phi^T)^{-1} \approx \Sigma_{cy}^{-1} \otimes (\Sigma_n + A \Psi \Sigma_{ry} \Psi^T A^T)^{-1}. \quad (17)$$

This approximation is inspired by [26], but the proposed method has more sophisticated form than that in [26]. This approximation performs well over a wide range of conditions (see Section 3). When $\boldsymbol{\lambda} = \mathbf{0}$ or $\Sigma_{cy} = I$, equality in Eq. (17) holds. With this approximation, we perform an alternative minimization scheme [28] to solve the optimization problem in Eq. (13), which reduces the original problem into several simpler subproblems and optimizes the variables in each subproblem alternatively.

(1) Sparse Signal Recovery: Solving for Y . Given $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$, the subproblem over \mathbf{y} can be formulated as

$$\min_{\mathbf{y}} \|\Phi \mathbf{y} - \mathbf{f}\|_{I \otimes \Sigma_n}^2 + \mathbf{y}^T \Sigma_y^{-1} \mathbf{y}. \quad (18)$$

The solution² is

$$Y^{new} = \Sigma_{ry} \Psi^T A^T (\Sigma_n + A \Psi \Sigma_{ry} \Psi^T A^T)^{-1} F. \quad (19)$$

(2) Sparsity Learning: Solving for $\boldsymbol{\gamma}$. With the fixed Y , $\boldsymbol{\lambda}$, Σ_{cy} and $\boldsymbol{\kappa}$, the subproblem over $\boldsymbol{\gamma}$ simplifies to

$$\min_{\boldsymbol{\gamma} \geq 0} \sum_{i=1}^{n_d} \frac{Y_{i.} \Sigma_{cy}^{-1} Y_{i.}^T}{\gamma_i} + n_p \log |\Sigma_n + A \Psi \Sigma_{ry} \Psi^T A^T| + \sum_{i=1}^{n_d} \kappa_i \gamma_i, \quad (20)$$

where $Y_{i.}$ denotes the i th row of Y . A concave conjugate function [6] is used to convert this nonconvex optimization into a convex one. Consequently, the solution² over $\boldsymbol{\gamma}$ is

$$\begin{aligned} \boldsymbol{\alpha} &= \text{diag}[\Sigma_{ry} - \Sigma_{ry} \Psi^T A^T (\Sigma_n + A \Psi \Sigma_{ry} \Psi^T A^T)^{-1} A \Psi \Sigma_{ry}], \\ \gamma_i^{new} &= (\sqrt{4\kappa_i (Y_{i.} \Sigma_{cy}^{-1} Y_{i.}^T + n_p \alpha_i)} + n_p^2 - n_p) / (2\kappa_i), \end{aligned} \quad (21)$$

where $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{n_d}]^T$ is an intermediate variable.

²The details of derivation can be found in the supplementary materials.

(3) Manifold Learning: Solving for Σ_{cy} . Given Y , γ and Θ , the subproblem over Σ_{cy} simplifies to

$$\min_{\Sigma_{cy}} \sum_{i=1}^{n_d} \frac{Y_i \Sigma_{cy}^{-1} Y_i^T}{\gamma_i} + \mu \log |\Sigma_{cy}| + \text{tr}(\Theta \Sigma_{cy}^{-1}), \quad (22)$$

where $\mu = m_b + n_p + l + 1$. This convex optimization gives a closed-form solution² as

$$\Sigma_{cy}^{new} = (Y^T \Sigma_{ry}^{-1} Y + \Theta + \epsilon I) / \mu, \quad (23)$$

where ϵ is a constant scalar and ϵI is introduced to make Σ_{cy} invertible. To improve the robustness to noise, $\mu = \|Y^T \Sigma_{ry}^{-1} Y + \Theta + \epsilon I\|_F$ is suggested.

(4) Noise Estimation: Solving for λ . Similarly, the subproblem over λ can be formulated as

$$\min_{\lambda \geq 0} \|A\Psi Y - F\|_{\Sigma_n}^2 + n_p \log |\Sigma_n + A\Psi \Sigma_{ry} \Psi^T A^T|, \quad (24)$$

and a concave conjugate function is adopted to convert this nonconvex optimization into a convex one. The solution² of λ is

$$\begin{aligned} \mathbf{v} &= \text{diag}[(\Sigma_n + A\Psi \Sigma_{ry} \Psi^T A^T)^{-1}], \\ \lambda_i^{new} &= \sqrt{(Q_{i,i}^T Q_{i,i}) / (n_p v_i)}, \end{aligned} \quad (25)$$

where $\mathbf{v} = [v_1, \dots, v_{m_b}]^T$ is an intermediate variable and $Q = A\Psi Y - F$.

(5) Hyperparameter Estimation: Solving for κ . Given γ , we have the subproblem over κ as

$$\min_{\kappa} \sum_{i=1}^{n_d} (\kappa_i \gamma_i - 2 \log \kappa_i). \quad (26)$$

The solution for κ is

$$\kappa_i^{new} = 2/\gamma_i. \quad (27)$$

The whole optimization procedure is summarized in Algorithm 1. The maximum iteration number $ItNum = 200$ and minimum update difference $\eta = \|Y^{new} - Y\|_2 / \|Y\| = 10^{-4}$ are employed as the stopping criteria in this study. Because the alternative minimization scheme decreases the objective function at each iteration and the objective function is proved to be bounded from below, the optimization will converge to a local minima.

2.4. Analysis of Structure and Robustness

In this subsection, we clarify the 3D structure and robustness to noise of the proposed prior.

Spatial Manifold Structure. Locally linear embedding (LLE) [4] and Laplacian eigenmap (LE) [15] are two popular manifold learning paradigms, where the manifold structure is implicitly represented by the correlation among nodes. Applied to HCS, the correlation defined in these two

Algorithm 1: Manifold-Structured Sparsity Prior based Hyperspectral Compressive Sensing (MSHCS)

Input: Random sampling matrix A , dictionary Ψ , noisy compressive measurements F .

Initialize: Noise level $\lambda = \mathbf{1}_{m_b}$, signal variance $\gamma = \mathbf{1}_{n_d}$, hyperparameter $\kappa = \mathbf{1}_{n_d}$;

preprocessing: Learn reference covariance matrix Θ ;

while *Stopping criteria is not satisfied* **do**
 1. Recover sparse signal Y by Eq. (19);
 2. Learn sparsity parameter γ by Eq. (21);
 3. Learn manifold structure Σ_{cy} by Eq. (23);
 4. Estimate noise level λ by Eq. (25);
 5. Update the hyperparameter κ by Eq. (27);

Output: Reconstruct HSI by $X_{rec} = \Psi Y_{rec}$.

methods can be viewed as the following priors on Y .

$$\begin{aligned} p_{LLE}(Y|W) &\propto \exp \left\{ -\frac{1}{2} \sum_i \|Y_{\cdot i} - \sum_j W_{ji} Y_{\cdot j}\|_2^2 \right\}, \\ p_{LE}(Y|K) &\propto \exp \left\{ -\frac{1}{2} \sum_{i,j} K_{ij} \|Y_{\cdot i} - Y_{\cdot j}\|_2^2 \right\}, \end{aligned} \quad (28)$$

where W and K are the similarity weight matrices defined on Y in LLE and LE, respectively. Let $L = (I - W)(I - W)^T$ in LLE and $L = D - K$ in LE where D is a diagonal matrix with $D_{ii} = \sum_j K_{ij}$, $p_{LLE}(Y|W)$ and $p_{LE}(Y|K)$ can be integrated into a unified distribution.

$$p(Y|L) = \frac{\exp \left\{ -\frac{1}{2} \text{tr}(YLY^T) \right\}}{(2\pi)^{-ndn_p/2} |L^{-1}|^{nd/2}}. \quad (29)$$

It can be seen that this manifold structure prior amounts to the proposed prior in Eq. (4) when $\Sigma_{cy} = L^{-1}$ and $\Sigma_{ry} = I$, i.e., the proposed prior can model the spatial manifold structure of Y by learning a reasonable Σ_{cy} .

3D Structure. The optimization in Eq. (13) equals to a standard regularized regression model as

$$\begin{aligned} \min_{\mathbf{y}} \|\Phi \mathbf{y} - \mathbf{f}\|_{I \otimes \Sigma_n}^2 + g(\mathbf{y}), \\ g(\mathbf{y}) = \min_{\gamma \geq 0, \kappa, \Sigma_{cy}} \|\mathbf{y}\|_{\Sigma_y}^2 + \log |\Sigma_{by}| + \sum_{i=1}^{n_d} (\kappa_i \gamma_i - 2 \log \kappa_i) \\ + \text{tr}(\Theta \Sigma_{cy}^{-1}) + (n_p + l + 1) \log |\Sigma_{cy}| - l \log |\Theta|, \end{aligned} \quad (30)$$

where $g(\mathbf{y})$ is the penalty function on \mathbf{y} . This standard regularized regression model equals to a Bayes MAP estimation with an implicit prior $\bar{p}(\mathbf{y}) \propto \exp \left\{ -\frac{1}{2} g(\mathbf{y}) \right\}$ and likelihood in Eq. (7) [27]. In this study, since $\log |\Sigma_{by}|$ makes Σ_n , Σ_{ry} , Σ_{cy} and Φ coupled together, the corresponding prior $\bar{p}(\mathbf{y})$ is non-factorial, i.e., there is no such a prior $\mathbf{g}(y_i)$ for each coefficient y_i that satisfies $\bar{p}(\mathbf{y}) = \prod_i \mathbf{g}(y_i)$, implying that any two coefficients of \mathbf{y} are dependent in this prior. Consequently, the proposed prior can represent the 3D structure

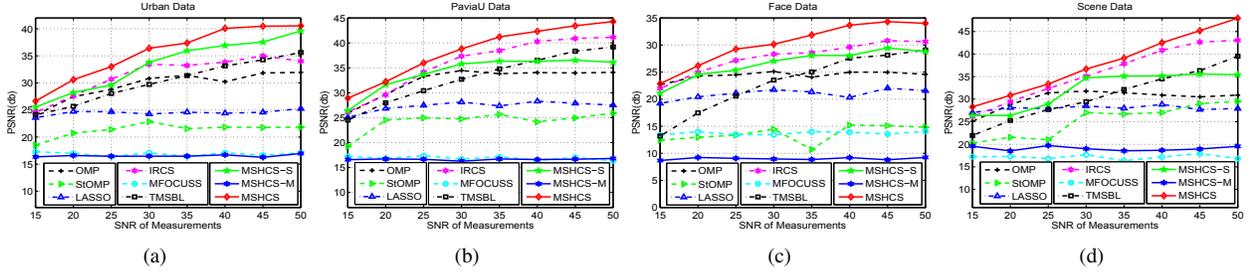


Figure 3. The PSNR curves on four datasets with sampling rate $\rho = 0.3$ and the measurements of SNR ranging from 15db to 50db.

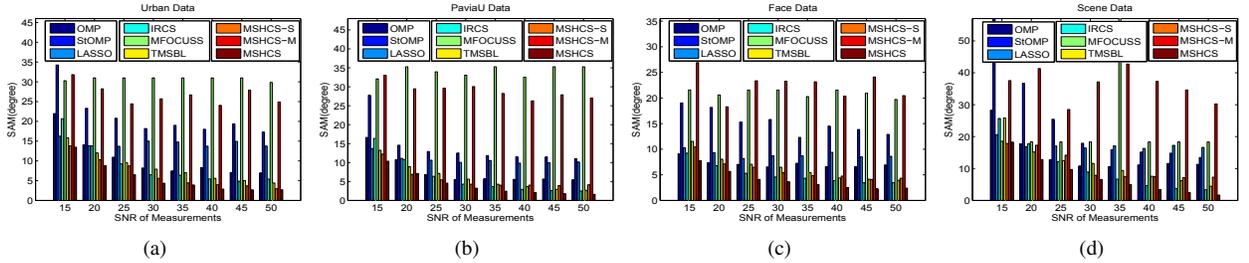


Figure 4. The SAM bar charts on four datasets with sampling rate $\rho = 0.3$ and the measurements of SNR ranging from 15db to 50db.

of HSI. Although a similar analysis is provided in [28], they only capture the structure within each sparse vector.

Robustness to Noise. According to Eqs. (21) (23) (25), the parameters γ and Σ_{cy} in spectral structured sparsity and spatial manifold structure are updated depending on the estimated noise level, and the noise level is estimated based on the learned sparsity prior in turn, which makes the proposed manifold-structured sparsity robust to noise.

3. Experimental Results and Analysis

Datasets: Four datasets are used to evaluate the performance of the proposed method, named as Urban³, PaviaU⁴, Face⁵ and Scene⁶. Urban and PaviaU are remote sensing HSI, Face comes from the CMU hyperspectral face database and Scene is the HSI of nature scene. For Urban, we select 128 continuous bands and crop 128×128 pixels in each band as the experimental data. For other datasets, we crop 128×128 pixels in each band as the experimental data.

Comparison methods: We compare the proposed method with 6 state-of-the-art methods, including Orthogonal Matching Pursuit (OMP) [24], Stagewise Orthogonal Matching Pursuit (StOMP) [13], LASSO [14], Iteratively Reweighted Compressive Sensing (IRCS) [8], the regularized Multi-variable Focal Underdetermined System Solver (MFOCUSS) [10], and the Temporally Multi-variable Sparse Bayes Learning (TMSBL) [29]. Therein, OMP, StOMP are the classical greedy algorithm based CS methods, LASSO and IRCS are the ℓ_1 and ℓ_p ($p = 0.5$) norm

based regularization methods, MFOCUSS and TMSBL are the inter-vector structure based CS methods. To further illustrate the superiority of the proposed method, we utilize two special cases of the proposed method as another two comparison methods, namely MSHCS-S and MSHCS-M. MSHCS-S only considers the spectral structured sparsity by setting $\Sigma_{cy} = I$ whereas MSHCS-M only considers the spatial manifold structure with $\Sigma_{ry} = I$.

Parameter setup: For all experiments, the 3D HSI is converted into a matrix X by vectorizing the image of each band as a row of X . Haar wavelet is chosen as the dictionary Ψ to transform X into the sparse matrix Y . A column normalized Gaussian random sampling matrix is adopted as A for all methods. The MSHCS, MSHCS-S and MSHCS-M share the same stopping criterion in Subsection 2.3.2. For other comparison methods, all parameters involved are optimally assigned as described in the reference papers.

Evaluation measures: Peak signal-to-noise ratio (PSNR) [21] and spectral angle mapper (SAM) [21] are adopted as the evaluation measures. PSNR measures the average similarity between the reconstructed image and reference image, while SAM calculates the average angles between spectrum vectors from the reconstructed image and the reference image at each pixel. Larger PSNR, or small SAM denotes higher reconstruction accuracy.

3.1. Performance on Robustness to Noise

In this subsection, X is compressed by the sampling matrix A with sampling rate $\rho = 0.3$, which is defined as the volume proportion of the measurements to the original HSI. Different levels of additive Gaussian white noise are added into the measurements F to simulate the noise in HCS, which results in the signal-noise-ratio (SNR) of F

³<http://www.tec.army.mil/Hypercube>

⁴http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

⁵<http://www.consortium.ri.cmu.edu/hsagree/index.cgi>

⁶http://personalpages.manchester.ac.uk/staff/david.foster/Hyperspectral_images_of_natural_scenes_04.html

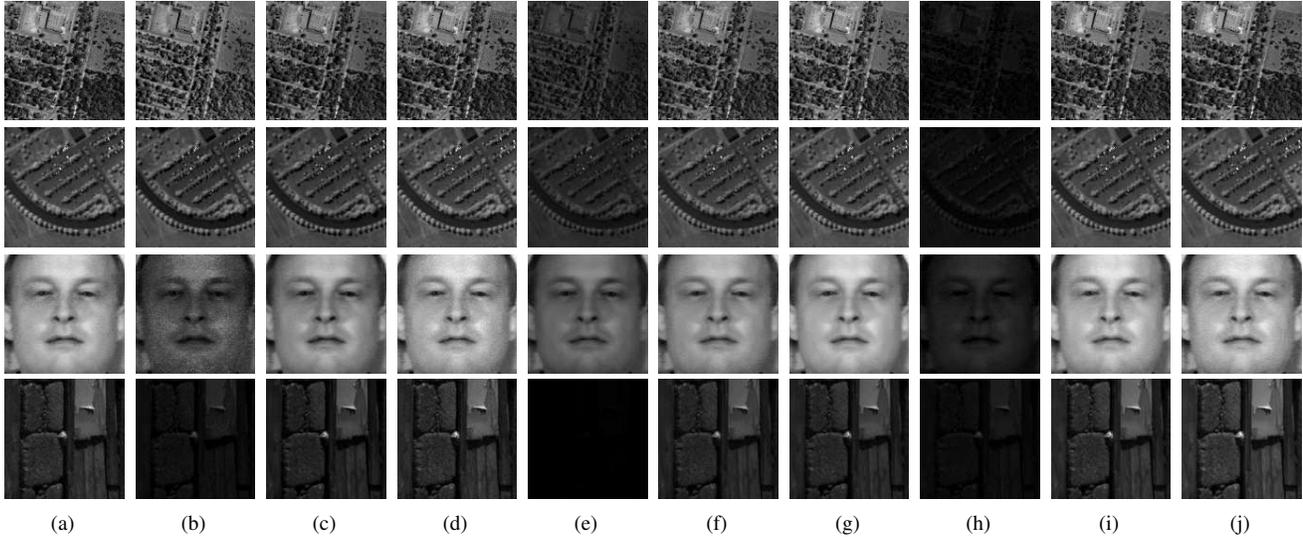


Figure 5. Visual reconstruction results of the 120th band from Urban, the 90th band from PaviaU, the 44th band from Face and the 30th band from Scene with sampling rate $\rho = 0.3$ when the SNR of measurements is $25db$. (a) OMP. (b) StOMP. (c) LASSO. (d) IRCS. (e) MFOCUSS. (f) TMSBL. (g) MSHCS-S. (h) MSHCS-M. (i) MSHCS. (j) Original bands.

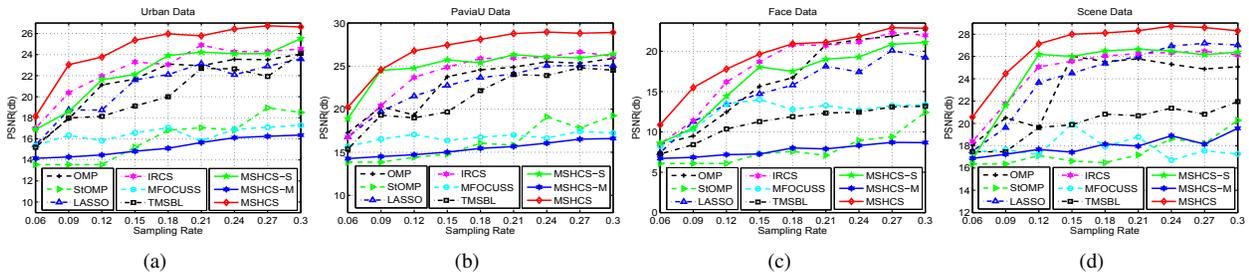


Figure 6. The PSNR curves on four datasets with the measurements of SNR = $15db$ and the sampling rate ρ ranging from 0.06 to 0.3.

ranging from $15db$ to $50db$. Given the noisy F , sampling matrix A and dictionary Ψ , we recover the sparse signal Y for all comparison methods. Then HSI is reconstructed as $\hat{X} = \Psi\hat{Y}$. After 10 Monte-Carlo runs of reconstruction, we obtain the average evaluation measures.

Under different levels of noise, the PSNR curves versus the SNR of measurements on four datasets are shown in Figure 3. It is clear that the proposed MSHCS outperforms other HCS methods on all datasets significantly. For example, when SNR of measurements is $30db$, the PSNR of MSHCS-S exceeds other competing methods at least $2.5db$ on Urban dataset, $1.4db$ on PaviaU dataset, $1.8db$ on Face dataset and $1.6db$ on Scene dataset. The bar charts of SAM values under different levels of noise on four datasets are given in Figure 4. Under different levels of noise, the SAM values of MSHCS are smaller than other methods in most of the cases. For example, when SNR of measurements is larger than $20db$, SAM values of MSHCS are smaller than 10 degree on the PaviaU dataset. All the SAM values of MSHCS on Face dataset are smaller than 8 degree under different levels of noise. Since MSHCS-M neglects the spectral sparsity of Y but it is crucial for reconstruction in HCS, larger SAM

values are given by MSHCS-M compared with MSHCS-S and MSHCS. Additionally, the visual comparison of part reconstruction results is shown in Figure 5, where MSHCS obtains the most approximate results to the original bands.

The evaluation results above demonstrate that the proposed MSHCS gives the best reconstruction result of HSI among all methods under different levels of noise.

3.2. Performance on Reconstruction Accuracy

In this subsection, we compress X by the sampling matrix A with sampling rate ρ ranging from 0.06 to 0.3. Additive Gaussian white noise is added into the measurements F and the SNR of F is $15db$. Given the noisy F , sampling matrix A and dictionary Ψ , we reconstruct HSI as Subsection 3.1 does by all methods. The average evaluation measures are given with 10 Monte-Carlo runs of reconstruction.

Under fixed noise level, the PSNR curves versus sampling rate on four datasets are shown in Figure 6. We can find that the proposed MSHCS obtains the highest PSNR values on all datasets. Specifically, when sampling rate $\rho = 0.09$, the PSNR of MSHCS exceeds other methods at least $2.6db$ on Urban dataset, $4db$ on Face dataset and $2.6db$

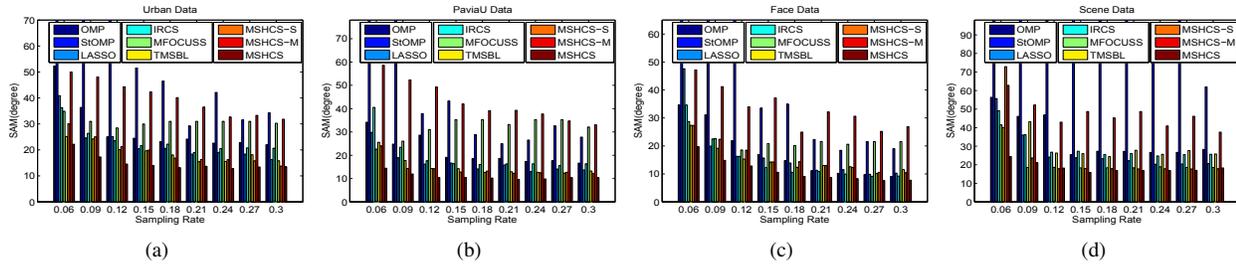


Figure 7. The SAM bar charts on four datasets with the measurements of SNR = 15db and the sampling rate ρ ranging from 0.06 to 0.3.

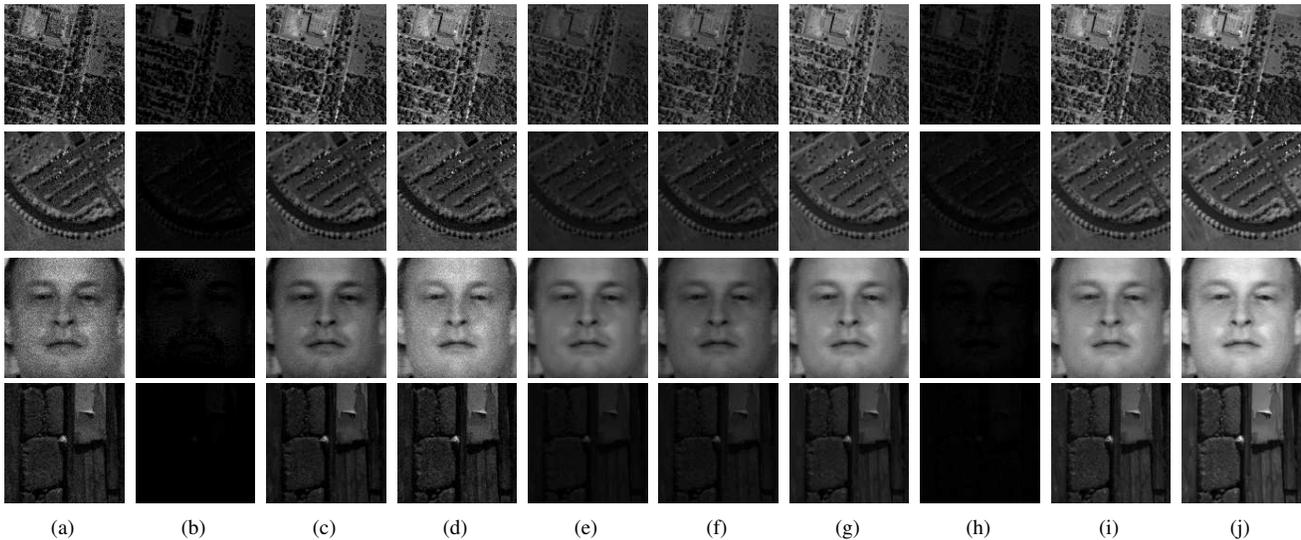


Figure 8. Visual reconstruction results of the 120th band from Urban, the 90th band from PaviaU, the 44th band from Face and the 30th band from Scene with sampling rate $\rho = 0.15$ when the SNR of measurements is 15db. (a) OMP. (b) StOMP. (c) LASSO. (d) IRCS. (e) MFOCUSS. (f) TMSBL. (g) MSHCS-S. (h) MSHCS-M. (i) MSHCS. (j) Original bands.

on Scene dataset. The bar charts of SAM values versus sampling rate on four datasets are given in Figure 7. With different sampling rates, the SAM values of MSHCS on all datasets are smaller than other competing methods in most of the cases. For example, when sampling rate $\rho > 0.06$, the SAM values of MSHCS are smaller than 12 degree on the PaviaU dataset. Similarly, without considering the spectral sparsity of Y leads to larger SAM values of MSHCS-M than that of MSHCS-S and MSHCS. The visual comparison of part reconstruction results is given in Figure 8, where the proposed MSHCS yields the most approximate results to the original bands.

These comparison results demonstrate that the proposed MSHCS outperforms other methods on the reconstruction accuracy of HSI with different sampling rates.

4. Conclusion

We have proposed a novel manifold-structured sparsity prior based HCS method to accurately reconstruct HSI from a few noisy measurements. The proposed prior represents the 3D structured sparsity of HSI for the first time through integrating the spectral structured sparsity and spa-

tial unknown manifold structure into a unified distribution with hierarchical structure. To make the prior fit the image distribution well and robustly to the random noise in HCS, the sparsity prior and unknown noise are jointly optimized from measurements by a latent Bayes model. Thus, with this learned prior, the proposed method improves the reconstruction accuracy under unknown noise corruption significantly. Extensive experimental results on four real hyperspectral datasets demonstrate the superiority of the proposed method to other 6 state-of-the-art HCS methods on the reconstruction accuracy of HSI.

5. Acknowledgement

This work is supported by the National Natural Science Foundation of China (No.61231016, No.61301192, No.61201291), Natural Science Basis Research Plan in Shaanxi Province of China (No.2013JQ8032), the NPU Foundation for Fundamental Research (No.3102015JSJ0006) and Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University (No.CX201521).

References

- [1] P. Bajorski. Target detection under misspecified models in hyperspectral images. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):470–477, 2012.
- [2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.
- [3] R. G. Baraniuk, V. Cevher, and M. B. Wakin. Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective. *Proceedings of the IEEE*, 98(6):959–971, 2010.
- [4] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. *Advances in neural information processing systems*, 16:177–184, 2004.
- [5] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379, 2012.
- [6] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [7] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.
- [8] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on*, pages 3869–3872. IEEE, 2008.
- [9] C. Chen and J. Huang. Compressive sensing mri with wavelet tree sparsity. In *Advances in neural information processing systems*, pages 1115–1123, 2012.
- [10] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *Signal Processing, IEEE Transactions on*, 53(7):2477–2488, 2005.
- [11] W. Dong, D. Zhang, and G. Shi. Centralized sparse representation for image restoration. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1259–1266. IEEE, 2011.
- [12] D. L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
- [13] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 58(2):1094–1121, 2012.
- [14] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.
- [15] S. Gao, I. W. Tsang, L.-T. Chia, and P. Zhao. Local features are not lonely—laplacian sparse coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3555–3561. IEEE, 2010.
- [16] M. Golbabaee and P. Vandergheynst. Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 2741–2744. Ieee, 2012.
- [17] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. *The Journal of Machine Learning Research*, 12:3371–3412, 2011.
- [18] C. Li, T. Sun, K. F. Kelly, and Y. Zhang. A compressive sensing and unmixing scheme for hyperspectral data processing. *Image Processing, IEEE Transactions on*, 21(3):1200–1210, 2012.
- [19] X. Lu, H. Yuan, P. Yan, Y. Yuan, and X. Li. Geometry constrained sparse coding for single image super-resolution. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1648–1655. IEEE, 2012.
- [20] G. Martin, J. M. Bioucas-Dias, and A. Plaza. Hyperspectral coded aperture (hyca): A new technique for hyperspectral compressive sensing. In *Signal Processing Conference (EU-SIPCO), 2013 Proceedings of the 21st European*, pages 1–5. IEEE, 2013.
- [21] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2949–2956. IEEE, 2014.
- [22] S. Prasad, W. Li, J. E. Fowler, and L. M. Bruce. Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(9):3474–3486, 2012.
- [23] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas. A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing. *Signal Processing, IEEE Transactions on*, 60(2):585–599, 2012.
- [24] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007.
- [25] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [26] J. Wan, Z. Zhang, J. Yan, T. Li, B. D. Rao, S. Fang, S. Kim, S. L. Risacher, A. J. Saykin, and L. Shen. Sparse bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in alzheimer’s disease. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 940–947. IEEE, 2012.
- [27] D. P. Wipf, B. D. Rao, and S. Nagarajan. Latent variable bayesian models for promoting sparsity. *Information Theory, IEEE Transactions on*, 57(9):6236–6255, 2011.
- [28] L. Zhang, W. Wei, Y. Zhang, C. Tian, and F. Li. Reweighted laplace prior based hyperspectral compressive sensing for unknown sparsity. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015.
- [29] Z. Zhang and B. D. Rao. Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning. *Selected Topics in Signal Processing, IEEE Journal of*, 5(5):912–926, 2011.