Reweighted Laplace Prior Based Hyperspectral Compressive Sensing for Unknown Sparsity

Lei Zhang*†, Wei Wei*†, Yanning Zhang†, Chunna Tian‡, Fei Li†
†School of Computer Science, Northwestern Polytechnical University, Xi'an, 710072, China
‡School of Electronic Engineering, Xidian University, Xi'an, 710071, China
†zhanglei211@mail.nwpu.edu.cn, †{weiweinwpu, ynzhang}@nwpu.edu.cn, ‡chnatian@xidian.edu.cn

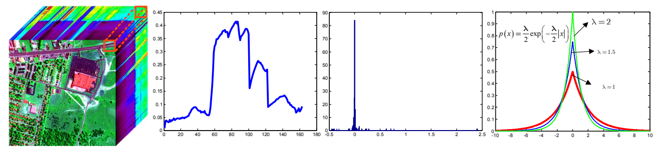
Abstract

Compressive sensing(CS) has been exploited for hypespectral image(HSI) compression in recent years. Though it can greatly reduce the costs of computation and storage, the reconstruction of HSI from a few linear measurements is challenging. The underlying sparsity of HSI is crucial to improve the reconstruction accuracy. However, the sparsity of HSI is unknown in reality and varied with different noise, which makes the sparsity estimation difficult. To address this problem, a novel reweighted Laplace prior based hyperspectral compressive sensing method is proposed in this study. First, the reweighted Laplace prior is proposed to model the distribution of sparsity in HSI. Second, the latent variable Bayes model is employed to learn the optimal configuration of the reweighted Laplace prior from the measurements. The model unifies signal recovery, prior learning and noise estimation into a variational framework to infer the parameters automatically. The learned sparsity prior can represent the underlying structure of the sparse signal very well and is adaptive to the unknown noise, which improves the reconstruction accuracy of HSI. The experimental results on three hyperspectral datasets demonstrate the proposed method outperforms serveral state-of-the-art hyperspectral CS methods on the reconstruction accuracy.

1. Introduction

Hyperspectral image(HSI) is a 3D data cube containing both spectral and spatial information. It consists of a series of 2D spatial images over many continuous spectral bands [3]. The spectral information makes it possible to identify and quantify distinct material substance, which facilitates a variety of applications of HSI, such as landform classification [18], mineral exploration [2], etc. However, the 3D structure of HSI increases the costs of storage, computation and transmission, which limits the usage of HSI.

*indicates equal contributions.



(a) (b) (c) (d)

Figure 1. (a) A hyperspectral image with a marked pixel(red cube). (b) The spectrum curve of the marked pixel. (c) The histogram of the transformed coefficients in wavelet domain. (d) The curves of the Laplace prior with different parameters, which is suitable to depict the sparsity of spectrum.

Compressive sensing(CS) is a newly proposed method for image compression. Its theory shows that a sparse signal can be recovered with high probability from a few linear measurements [6], which makes it possible to compress image during the image acquisition. It can dramatically reduce the consumption of imaging resource compared with traditional compression methods such as DPCM, JPEG, JPEG2000, etc. Since sparsification methods such as transformation based methods [9, 12](e.g. wavelet transform) or unmixing based methods with a redundant endmember dictionary [10, 11, 13, 21] can transform HSI into a sparse signal(see Figure 1(a)(b)(c)), a series of hyperspectral compressive sensing(HCS) methods have been proposed [9, 12, 15, 16, 19]. For HCS, how to recover the original HSI from a few measurements is a challenging problem. ℓ_0 norm, ℓ_1 norm [20] and Laplace prior are prevalently adopted to constrain the sparsity of HSI for reconstruction. Recently, to reduce the undemocratic penalization of ℓ_1 norm on the nonzero coefficients of a sparse signal, Candes et al. [4] proposed a reweighted ℓ_1 norm which has resulted in a state-of-the-art CS method. Chartrand et al. [5] further extended this idea to a reweighted ℓ_p $(0 \leq p < 2)$ norm based method. However, two important issues have been neglected in those norm-based or prior-based methods when depicting the sparsity of signal. 1) These sparse regularization terms(e.g. ℓ_0, ℓ_1 norm) assume the sparse coefficients are independent, which neglect the structure information of sparse coefficients. 2) The adaptability of these regularization terms to the unknown noise is unsolved.

To model the distribution of sparsity in HSI and make the method adaptive to the unknown noise, we propose a novel matrix-based reweighted Laplace prior by utilizing the characteristics of HSI. Then, a latent variable Bayes model [25] is employed to learn the hyperparameters of the reweighted Laplace prior from the measurements. This Bayes model unifies the signal recovery, prior learning and noise estimation into a variational framework, where all unknown variables are inferred automatically. The learned sparsity prior can capture the underlying structure of the sparse signal and is adaptive to the unknown noise, which improves the reconstruction accuracy of HSI. The experimental results on three hyperspectral datasets demonstrate that the proposed method outperforms several state-of-the-art HCS methods on the reconstruction accuracy of HSI.

2. The Proposed Method

We denote a 3D hyperspectral image as $\mathcal{X} \in \mathbb{R}^{n_r \times n_c \times n_b}$ where n_r, n_c and n_b represent the dimension of image height, width and bands, respectively. For convenience, we rearrange $\mathcal{X}$ into a 2D matrix $X \in \mathbb{R}^{n_b \times n_p}$ by reshaping the image of each band as a row of X, $n_p = n_r \times n_c$. In HCS, X is linearly sampled by a Gaussian or Bernoulli random sampling matrix $A \in \mathbb{R}^{m_b \times n_b} (m_b < n_b)$ as

$$G = AX + N \tag{1}$$

where $G \in \mathbb{R}^{m_b \times n_p}$ is the measurements of X and $N \in \mathbb{R}^{m_b \times n_p}$ denotes noise. The reconstruction of X is an inverse problem, which tries to recover X from the measurements G given A. Since X is not sparse, we transform X into a sparse signal Y by introducing a known linear basis matrix D as $X = DY$. It is noticeable that the choice of linear basis matrix D can be diverse. When D is an orthogonal basis, Y contains the sparse transformation coefficients of X. Similarly, when D is a collection of endmembers or a redundant dictionary, Y is a sparse abundance matrix or composed of the sparse representation coefficients, respectively.

Given D, the reconstruction task turns to recover the sparse Y from G as

$$Y_{opt} = \arg\max_{Y} p\left(Y|G\right). \tag{2}$$

In this study, noise N in Eq. (1) is assumed to obey matrix normal distribution $\mathcal{MN}\left(0, \Sigma_n, I\right)$, where Σ_n controls the noise level and I is an identity matrix. The likelihood $p(G|X, \Sigma_n) = \mathcal{MN}\left(AX, \Sigma_n, I\right)$. We assume the noise in each row of G is uncorrelated in this study, thus the covariance matrix $\Sigma_n = diag\left(\boldsymbol{\lambda}\right)$[1] is a $\boldsymbol{\lambda}$-dependent diagonal matrix, where $\boldsymbol{\lambda} = \left[\lambda_1, ..., \lambda_{m_b}\right]^T$. Since $X = DY$, we have the following likelihood of HCS

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

where $\left\|Q\right\|_{\Sigma_n} = \sqrt{\mathbf{tr}\left(Q^T \Sigma_n^{-1} Q\right)}$ represents a weighted trace norm.

2.1. Reweighted Laplace Prior

Though Laplace prior [1, 14] is suitable to depict the sparsity of HSI as Figure 1(d) shows, it is unable to capture the underlying structure of the sparse signal, and it puts undemocratic penalization on nonzero coefficients of a sparse signal. In addition, traditional Laplace prior is not conjugate to the Gaussian likelihood in Eq. (3). To solve these problems, a data-driven hierarchical reweighted Laplace sparsity prior is proposed in this study as Figure 2 shows.

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

$$p\left(Y|\boldsymbol{\gamma}\right) = \frac{\exp\left\{-\frac{1}{2}\|Y\|_{\Sigma_y}^2\right\}}{\left(2\pi\right)^{n_b n_p/2}\left|\Sigma_y\right|^{n_p/2}}, \quad \Sigma_y = diag\left(\boldsymbol{\gamma}\right), \tag{4}$$

where $diag(\boldsymbol{\gamma})$ denotes a diagonal matrix with the diagonal elements from $\boldsymbol{\gamma}$. $\boldsymbol{\gamma} = \left[\gamma_1, ..., \gamma_{n_b}\right]^T$ controls the variation of each row in Y. The smaller $\boldsymbol{\gamma}$ denotes the stronger sparsity of Y. $Y = [\boldsymbol{y}_1, ..., \boldsymbol{y}_{n_p}]$ and $\boldsymbol{y}_i \in \mathbb{R}^{n_b}$ is the ith column of Y. This prior implies each column of Y obeys the zero mean Gaussian distribution under covariance matrix Σ_y.

Then, a Gamma distribution is imposed on the unknown $\boldsymbol{\gamma}$ as follows

$$p\left(\boldsymbol{\gamma}|\boldsymbol{\kappa}\right) = \prod_{i=1}^{n_b} \mathrm{Gamma}(1, \frac{2}{\kappa_i}) = \prod_{i=1}^{n_b} \frac{\kappa_i}{2}\exp\left(-\frac{\kappa_i \gamma_i}{2}\right) \tag{5}$$

where $\boldsymbol{\kappa} = [\kappa_1, ..., \kappa_{n_b}]^T$ controls the shape of the joint Gamma distribution. It can be proved that the proposed hierarchical prior in Figure 2 equals to a reweighted Laplace prior[2] for each column of Y through the following maximum a posteriori(MAP) estimation

$$p\left(\boldsymbol{y}_i\right) \propto \int p\left(\boldsymbol{y}_i|\boldsymbol{\gamma}\right) p\left(\boldsymbol{\gamma}|\boldsymbol{\kappa}\right) d\boldsymbol{\gamma} = \frac{\exp\left(-\left\|K\boldsymbol{y}_i\right\|_1\right)}{2^{n_b}|K|^{-1}} \tag{6}$$

where we adopt a flat prior on $p(\boldsymbol{\kappa})$ and omit it. In Eq. (6), $K = diag([\sqrt{\kappa_1}, ..., \sqrt{\kappa_{n_b}}]^T)$.

The proposed hierarchical prior is a general form of Laplace prior. When all elements of $\boldsymbol{\kappa}$ are identical to a fixed scalar, the proposed prior degenerates to the traditional Laplace prior. By adjusting K, the proposed prior

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

[2]The detailed derivation can be found in the supplementary material.

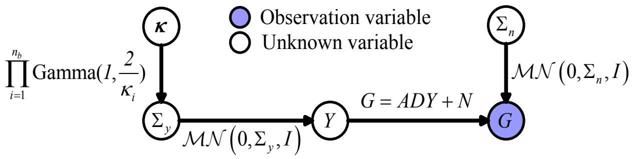


Figure 2. The hierarchical structure of the reweighted Laplace prior based HCS.

can reduce the undemocratic penalization of the traditional Laplace prior, which is similar as the reweighted ℓ_1 norm in [4] does. But the hierarchical structure guarantees the proposed prior is conjugate to the likelihood in Eq. (3) which is different with Ref. [4]. More importantly, the sparsity prior obtained by the proposed method is more flexible to noise by learning $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$, and can capture the underlying structure of the sparse signal. We will analyze these merits in detail in Subsection 2.4.

2.2. Latent Variable Based Sparsity Learning

If the noise variable $\boldsymbol{\lambda}$, the variables of sparsity prior $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ are given in advance, Y can be inferred directly by the following MAP estimation

$$\max_Y p(Y|G) = \max_Y p(G|Y, \boldsymbol{\lambda}) p(Y|\boldsymbol{\gamma}). \tag{7}$$

However, the appropriate sparsity prior of Y and noise covariance are both unknown in reality. In this study, a latent variable Bayes model is adopted to learn the reweighted Laplace prior and the noise covariance simultaneously from the measurements. Those unknown variables can be learned by the following MAP estimation

$$\begin{aligned}&\max_{\boldsymbol{\lambda}\geq 0, \boldsymbol{\gamma}\geq 0, \boldsymbol{\kappa}} p(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}|G) \\ =&\max_{\boldsymbol{\lambda}\geq 0, \boldsymbol{\gamma}\geq 0, \boldsymbol{\kappa}} \int p(G|Y, \boldsymbol{\lambda}) p(Y|\boldsymbol{\gamma}) p(\boldsymbol{\gamma}|\boldsymbol{\kappa})\, dY\end{aligned} \tag{8}$$

where $p(\boldsymbol{\kappa})$ and $p(\boldsymbol{\lambda})$ adopt flat priors and the latent variable Y is integrated out. It can be proved that Eq. (8) is equivalent to minimizing the following cost function

$$\begin{aligned}&\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}) \triangleq -2\log p(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}|G) \\ &= \mathbf{tr}\left(n_{n_p}^{-1} G^T \Sigma_{by}^{-1} G\right) + \log|\Sigma_{by}| + \sum_{i=1}^{n_b} \frac{\kappa_i\gamma_i - 2\log\kappa_i}{n_p}\end{aligned} \tag{9}$$

where $\Sigma_{by} = \Sigma_n + AD\Sigma_y D^T A^T$ is termed as model covariance. This cost function consists of three parts from left to right, which are data-fitting term, volume-based [24] and hyperprior-based regularization terms. The data-fitting term encourages the model covariance to fit the empirical covariance of measurements. The volume-based regularization term $\log|\Sigma_{by}|$ attempts to degenerate the volume of high dimensional space determined by the model covariance Σ_{by}. The hyperprior-based regularization term implicitly prevents undemocratic penalization on different sparse coefficients of the signal which will be discussed in Subsection 2.3. Eq. (9) provides a tradeoff among these three terms. Ideally, we can obtain the optimal estimation of $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ using Eq. (9), then recover the unknown Y with $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ using Eq. (7).

However, it is difficult to estimate $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ from the nonconvex optimization in Eq. (9). Instead of optimizing nonconvex $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa})$ with the assist of MAP estimation to reconstruct the sparse signal Y as Eq. (7), we transform the cost function into an equivalent regularized regression formula, which is more intuitive as Ref. [16, 15] do in HCS. According to Ref. [25, 26], the data-fitting term equals to a minimization problem over Y as

$$\mathbf{tr}\left(n_{n_p}^{-1} G^T \Sigma_{by}^{-1} G\right) = \min_Y \left\| ADY - \frac{G}{\sqrt{n_p}} \right\|_{\Sigma_n}^2 + \|Y\|_{\Sigma_y}^2. \tag{10}$$

Then, the transformed cost function $\mathcal{L}(Y, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa})$ is given as

$$\begin{aligned}\mathcal{L}(Y, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}) &= \left\| ADY - \frac{G}{\sqrt{n_p}} \right\|_{\Sigma_n}^2 + \|Y\|_{\Sigma_y}^2 \\ &+ \log|\Sigma_{by}| + \frac{1}{n_p}\sum_{i=1}^{n_b}(\kappa_i\gamma_i - 2\log\kappa_i)\end{aligned} \tag{11}$$

which is the upper bound of $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa})$. It can be proved that the solution over $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa})$ of minimizing $\mathcal{L}(Y, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa})$ as Eq. (12) is equivalent to the solution of minimizing $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa})$ as [24]. In Eq. (12), we unify signal recovery, sparsity learning and noise estimation into one framework.

$$\min_{Y, \boldsymbol{\lambda}\geq 0, \boldsymbol{\gamma}\geq 0, \boldsymbol{\kappa}} \mathcal{L}(Y, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\kappa}) \tag{12}$$

2.3. Optimization Procedure

The proposed model involves multiple variables and is hard to be minimized directly. The alternative minimization scheme, which reduces the original problem into several simpler subproblems, is adopted in this study. We address the optimization of each variable alternately.

(1) **Sparse Signal Recovery: Optimizing for** Y. In this subproblem, we fix $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$, optimizing Y by

$$\min_Y \left\| ADY - \frac{G}{\sqrt{n_p}} \right\|_{\Sigma_n}^2 + \|Y\|_{\Sigma_y}^2. \tag{13}$$

The closed-form solution for Y is given as

$$Y^{new} = \Sigma_y D^T A^T \Sigma_{by}^{-1} \frac{G}{\sqrt{n_p}}. \tag{14}$$

(2) **Sparsity Learning : Optimizing for** $\boldsymbol{\gamma}$. Given Y, $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$, the sub-problem over $\boldsymbol{\gamma}$ is derived as

$$\min_{\boldsymbol{\gamma}\geq 0} \|Y\|_{\Sigma_y}^2 + \log|\Sigma_{by}| + \sum_{i=1}^{n_b} \frac{\kappa_i\gamma_i - 2\log\kappa_i}{n_p}. \tag{15}$$

To solve this nonconvex optimization problem, a concave conjugate function is introduced to find an upper bound for $\log|\Sigma_{by}|$ to relax Eq. (15) to a convex optimization. $\boldsymbol{\gamma}$ can be estimated[2] as

$$\begin{aligned} \boldsymbol{z} &= diag\left(\Sigma_y - \Sigma_y D^T A^T \Sigma_{by}^{-1} A D \Sigma_y\right) \\ \gamma_i^{new} &= \frac{\sqrt{4\kappa_i n_p\left(\bar{y}_i^2 + z_i\right) + n_p^2} - n_p}{2\kappa_i} \end{aligned} \tag{16}$$

where $\bar{\boldsymbol{y}} = [\bar{y}_1, ..., \bar{y}_{n_b}]^T = diag\left(YY^T\right)$ and $\boldsymbol{z} = [z_1, ..., z_{n_b}]^T$ is an intermediate variable.

(3) **Noise Estimation: Optimizing for $\boldsymbol{\lambda}$**. With the recovered sparse signal Y and prior parameter $\boldsymbol{\gamma}$, we have the following sub-problem over $\boldsymbol{\lambda}$

$$\min_{\boldsymbol{\lambda}\geq 0} \left\| ADY - \frac{G}{\sqrt{n_p}} \right\|_{\Sigma_n}^2 + \log|\Sigma_{by}|. \tag{17}$$

Similar to the optimization of $\boldsymbol{\gamma}$, a concave conjugate function is utilized to yield a closed-form solution[2] as

$$\boldsymbol{\alpha} = diag\left(\Sigma_{by}^{-1}\right), \ \lambda_i^{new} = \sqrt{\frac{\bar{q}_i^2}{\alpha_i}} \tag{18}$$

where $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_{m_b}]$, $\boldsymbol{q} = [\bar{q}_1, ..., \bar{q}_{m_b}]^T = diag\left(QQ^T\right)$ and $Q = ADY - G/\sqrt{n_p}$.

(4) **Hyperparameter Estimation: Optimizing for $\boldsymbol{\kappa}$**. Given the learned $\boldsymbol{\gamma}$, we derive the sub-problem over $\boldsymbol{\kappa}$ as

$$\min_{\boldsymbol{\kappa}} \sum_{i=1}^{n_b} (\kappa_i\gamma_i - 2\log\kappa_i) \tag{19}$$

which results in a closed-form solution for $\boldsymbol{\kappa}$ as

$$\kappa_i^{new} = \frac{2}{\gamma_i}. \tag{20}$$

The whole optimization procedure is summarized in Algorithm 1. Since the alternative minimization scheme decreases the objective function in each iteration and the objective function is proved to be bounded from below, the optimization converges as Ref. [25, 26].

2.4. Benefits of Sparsity Learning

In this section, we will discuss the benefits of the learned reweighted Laplace prior. It is clear that the unified optimization framework in Eq. (12) equals to a standard regularized regression model over sparse Y as

$$\begin{aligned} &\min_{Y} \ (ADY - \frac{G}{\sqrt{n_p}})^T \Sigma_n^{-1} (ADY - \frac{G}{\sqrt{n_p}}) + \mathbf{g}(Y), \\ &\mathbf{g}(Y) = \min_{\boldsymbol{\gamma}\geq 0, \boldsymbol{\kappa}} \|Y\|_{\Sigma_y}^2 + \log\left|\Sigma_n + AD\Sigma_y D^T A^T\right| \\ &\qquad + \sum_{i=1}^{n_b} \frac{\kappa_i\gamma_i - 2\log\kappa_i}{n_p} \end{aligned} \tag{21}$$

Algorithm 1: Reweighted Laplace Prior based Hyperspectral Compressive Sensing(RLPHCS)

Input: Random measurement matrix A, linear basis matrix D, linear measurement G.
Initialize: Noise level $\boldsymbol{\lambda} = \mathbf{1}_{m_b}$, signal variance $\boldsymbol{\gamma} = \mathbf{1}_{n_b}$, hyperparameter $\boldsymbol{\kappa} = \mathbf{1}_{n_b}$.
while *Stopping criteria is not satisfied* **do**
1. Update the intermediate variable $\Sigma_{by} = \Sigma_n + AD\Sigma_y D^T A^T$;
2. Update the weighted sparse signal Y by Eq. (14);
3. Update the signal variance $\boldsymbol{\gamma}$ by Eq. (16);
4. Update the noise level $\boldsymbol{\lambda}$ by Eq. (18);
5. Update the hyperparameter $\boldsymbol{\kappa}$ by Eq. (20);

Postprocessing: Compensate the sparse signal by $Y_{rec} = \sqrt{n_p}Y$;
Output: Reconstructed HSI $X_{rec} = DY_{rec}$.

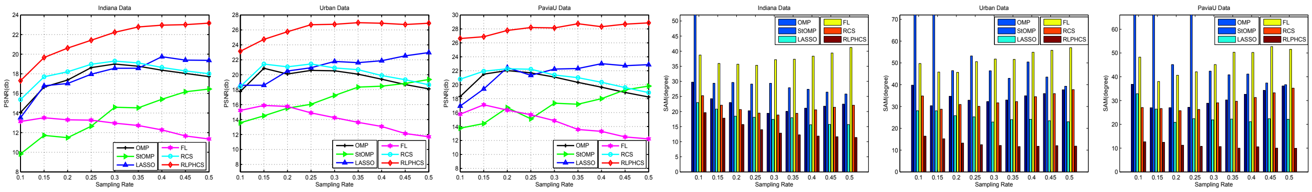
where $\mathbf{g}(Y)$ is a penalty function. It can be proved as Ref. [25] that this standard regularized regression model amounts to a Bayes MAP estimation with an implicit prior $\bar{p}(Y) = \exp\{-\frac{1}{2}\mathbf{g}(Y)\}$ and likelihood in Eq. (3). Since $\log\left|\Sigma_n + AD\Sigma_y D^T A^T\right|$ makes $\boldsymbol{\gamma}$ coupled with the noise variance $\boldsymbol{\lambda}$, sampling matrix A and linear basis matrix D, $\mathbf{g}(Y)$ is non-separable, viz., there is no such function $\mathfrak{g}$ which satisfies $\mathbf{g}(Y) = \sum_i \mathfrak{g}(Y_{i.})$, where $Y_{i.}$ is the ith row of Y. This non-separable penalty function corresponds to a nonfactorial implicit prior $\bar{p}(Y)$. It implies all rows of Y are dependent under an implicit structure. Therefore, the sparsity learning guarantees the reweighted Laplace prior can capture the underlying structure of the sparse Y. On the contrary, for each column of Y, the commonly used ℓ_0 or ℓ_l norm can be formulated as the summation of the constraint on each element Y_{ij}, which implies the structure of Y(i.e. the relation among different rows) is ignored for HCS.

In addition, in terms of Eq. (16), we can intuitively find that $\boldsymbol{\gamma}$ is adaptively regulated with the estimated noise level $\boldsymbol{\lambda}$, which is included in Σ_{by}. It implies the learned sparsity can adaptively regulate its configuration based on the unknown noise. Since the element with a large magnitude in Y corresponds to a large γ_i, these elements are divided by a large γ_i in the reweighted Laplace prior according to Eq. (6) and $\boldsymbol{\kappa}$ updated in Eq. (20) to reduce the undemocratic penalization effect on elements of Y.

3. Experimental Results and Analysis

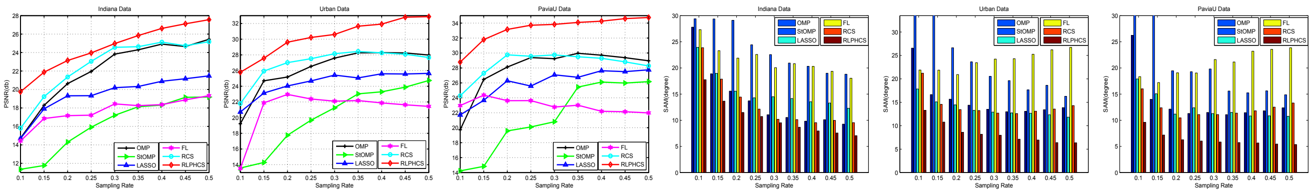
3.1. Experimental Setup

Data sets: In this study, three datasets are utilized to evaluate the performance of the proposed method, named as



(a) (b) (c) (d) (e) (f)

Figure 3. The PSNR curves and SAM bar charts of different methods on three datasets under SNR=$10db$. (a)(b)(c) PNSR curves. (d)(e)(f) SAM bar charts.



(a) (b) (c) (d) (e) (f)

Figure 4. The PSNR curves and SAM bar charts of different methods on three datasets under SNR=$20db$. (a)(b)(c) PNSR curves. (d)(e)(f) SAM bar charts.

INDIANA[3], URBAN[4] and PAVIAU[5]. After removing the noisy and water absorption bands from each data, the resulted INDIANA data is of 145×145 pixels and 169 bands, the URBAN data is of 307×307 pixels and 163 bands, and the resulted PAVIAU data is of 610×340 pixels and 103 bands. In the following experiments, we crop a subimage of 200×200 pixels from each band of URBAN and PAVIAU as the test data. The INDIANA data is employed directly.

Comparison methods: We compare the proposed method with 5 state-of-the-art methods. They are Orthogonal Matching Pursuit(OMP) [22], Stagewise Orthogonal Matching Pursuit(StOMP) [7], LASSO [8], Fast Laplace(FL) [1] and Reweighted ℓ_1 norm based Compressive Sensing(RCS) [4]. These traditional CS methods are extended to HCS by reconstructing each spectrum of HSI. In this paper, orthogonal transformation based method is used to transform HSI into a sparse signal. Due to the limitation on page length, comparison results with unmixing based HCS methods can be found in the supplementary material.

Parameter setup: For all experiments, the 3D HSI is converted into a 2D matrix X by vectorizing the image of each band as a row of X. A Gaussian random sampling matrix is adopted as A for all methods. For the proposed method, the maximum iteration number N_{It} and the minimum update difference $\eta = \|Y^{new} - Y\|_2 / \|Y\|_2$ are the only two parameters which need to be manually tuned for terminating the algorithm. We set $N_{It} = 400$ and $\eta = 10^{-4}$. For the comparison methods, all parameters involved are optimally assigned as described in the reference papers.

[3]http://cobweb.ecn.purdue.edu/ biehl/MultiSpec/
[4]http://www.tec.army.mil/Hypercube
[5]http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

Evaluation measures: Peak signal-to-noise ratio(PSNR) [17], structure similarity(SSIM) [23] and spectral angle mapper(SAM) [17] are adopted as the evaluation measures. PSNR and SSIM measure the average similarity and structure consistency between the reconstructed image and the reference image, respectively, while SAM calculates the average angles between spectrum vectors from the reconstructed image and the reference image at each pixel. Therefore, larger PSNR, SSIM and smaller SAM denote higher reconstruction accuracy.

3.2. Performance Evaluation on Orthogonal Transformation Based HCS

In this section, Haar wavelet is chosen to form the linear basis matrix D , with which X can be transformed into a sparse signal Y. We compress all data by the random sampling matrix A with sampling rate ρ ranging from 0.1 to 0.5. The sampling rate ρ denotes the volume proportion between measurements G and original data X. To simulate the random noise in CS, we put three levels of additive Gaussian white noise into the measurements, which results in the signal noise ratio(SNR) of measurements to be $10db$, $20db$ and $30db$, respectively. OMP, StOMP, LASSO, FL and RCS are implemented as the comparison methods. Given the measurements G and random sampling matrix A, we recover the sparse signal $\hat{Y}$ by all competing methods. Then HSI is reconstructed as $\hat{X} = D\hat{Y}$ with the Haar wavelet basis matrix D. After 10 Monte-Carlo runs of reconstruction, we obtain the following average evaluation measures.

Under three levels of noise, the average PSNR curves versus sampling rate of different methods on three datasets are shown in Figure 3(a)(b)(c), Figure 4(a)(b)(c) and Figure 5(a)(b)(c). It is clear that the proposed RLPHCS outperforms all other HCS methods on each dataset dramatically. For example, when SNR of measurements is $10db$,

Table 1. The SSIM scores of different methods on three datasets under SNR = $10db$.

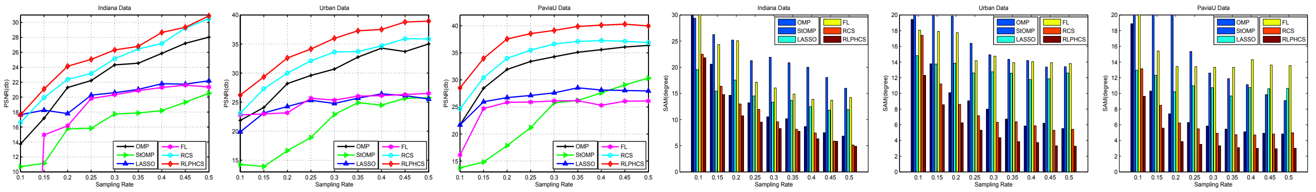
Results on INDIANA dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.1695	0.2127	0.2068	0.2218	0.2055	0.1854	0.1596	0.1436	0.1279
StOMP [7]	0.0329	0.158	0.1410	0.1995	0.2371	0.2893	0.3661	0.4162	0.4251
LASSO [8]	0.3267	0.4583	0.4311	0.4365	0.4394	0.4331	0.4348	0.4234	0.4064
FL [1]	0.0605	0.0652	0.0609	0.0616	0.0476	0.0440	0.0374	0.0313	0.0288
RCS [4]	0.2351	0.2805	0.2616	0.2545	0.2337	0.2063	0.1779	0.1567	0.1386
RLPHCS	**0.5752**	**0.6216**	**0.6415**	**0.7016**	**0.7195**	**0.7243**	**0.7272**	**0.7338**	**0.7368**
Results on URBAN dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.2866	0.4806	0.4208	0.4281	0.4206	0.3969	0.3629	0.3275	0.3028
StOMP [7]	0.0117	0.0787	0.1742	0.2055	0.3184	0.3589	0.3627	0.4113	0.4206
LASSO [8]	0.4589	0.4798	0.5528	0.5802	0.6318	0.6076	0.5993	0.6074	0.6080
FL [1]	0.2253	0.2249	0.2363	0.1969	0.1772	0.1618	0.1403	0.1112	0.1014
RCS [4]	0.3851	0.5147	0.4901	0.4817	0.4488	0.4247	0.3813	0.3523	0.3219
RLPHCS	**0.7478**	**0.8042**	**0.8449**	**0.8578**	**0.8627**	**0.873**	**0.8657**	**0.8722**	**0.8728**
Results on PAVIAU dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.3336	0.4711	0.4721	0.4361	0.4097	0.3690	0.3340	0.3010	0.2764
StOMP [7]	0.006	0.0462	0.2007	0.0965	0.2667	0.2483	0.3120	0.367	0.4165
LASSO [8]	0.2722	0.4524	0.6098	0.5647	0.5824	0.5779	0.6037	0.5692	0.5716
FL [1]	0.1699	0.2783	0.2369	0.2108	0.1781	0.1355	0.1280	0.1062	0.1001
RCS [4]	0.4711	0.4905	0.4880	0.4732	0.4265	0.3971	0.3650	0.3278	0.2971
RLPHCS	**0.8196**	**0.8282**	**0.8594**	**0.8651**	**0.8709**	**0.8733**	**0.8721**	**0.8762**	**0.8787**

Table 2. The SSIM scores of different methods on three datasets under SNR = $20db$.

Results on INDIANA dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.2413	0.4355	0.5296	0.5424	0.5627	0.5944	0.5923	0.5588	0.5576
StOMP [7]	0.1777	0.2026	0.3412	0.5553	0.5888	0.6546	0.682	0.7397	0.7213
LASSO [8]	0.4617	0.6908	0.763	0.7693	0.8025	0.8043	0.8234	0.8289	0.8313
FL [1]	0.196	0.2762	0.3137	0.2122	0.2459	0.2146	0.2314	0.233	0.2482
RCS [4]	0.3509	0.5209	0.5948	0.5996	0.5993	0.5927	0.5711	0.5377	0.5139
RLPHCS	**0.8109**	**0.827**	**0.8503**	**0.8589**	**0.8644**	**0.8774**	**0.883**	**0.8898**	**0.8934**
Results on URBAN dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.4984	0.7759	0.7614	0.7842	0.7986	0.7919	0.7795	0.7632	0.745
StOMP [7]	0.0177	0.0662	0.3587	0.5053	0.6652	0.7434	0.7700	0.7743	0.8483
LASSO [8]	0.7081	0.8320	0.8309	0.871	0.8859	0.8794	0.8938	0.8862	0.9011
FL [1]	0.5818	0.5841	0.6186	0.5762	0.5629	0.5474	0.5210	0.5091	0.5017
RCS [4]	0.6639	0.8285	0.8203	0.8244	0.821	0.8057	0.7847	0.7626	0.7418
RLPHCS	**0.8822**	**0.9064**	**0.9362**	**0.9452**	**0.9444**	**0.9528**	**0.9567**	**0.9603**	**0.9605**
Results on PAVIAU dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.4539	0.7546	0.8091	0.8286	0.8202	0.8116	0.7935	0.7779	0.7538
StOMP [7]	0.0354	0.0769	0.5133	0.5605	0.6077	0.8043	0.8214	0.8158	0.8166
LASSO [8]	0.6334	0.7694	0.8686	0.8444	0.8748	0.8728	0.8962	0.8969	0.8940
FL [1]	0.6371	0.6929	0.6375	0.6396	0.5744	0.5907	0.5250	0.5190	0.5148
RCS [4]	0.7117	0.8118	0.8580	0.8373	0.8259	0.8065	0.7866	0.7595	0.731
RLPHCS	**0.9179**	**0.9363**	**0.9481**	**0.9515**	**0.9519**	**0.9542**	**0.9548**	**0.9564**	**0.9568**

Table 3. The SSIM scores of different methods on three datasets under SNR = $30db$.

Results on INDIANA dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.2721	0.4199	0.6582	0.7201	0.7593	0.7579	0.8201	0.8209	0.8474
StOMP [7]	0.1237	0.2116	0.5299	0.6159	0.6698	0.6869	0.7293	0.7475	0.7989
LASSO [8]	0.7045	0.7705	0.7473	0.8478	0.8676	0.8771	0.8881	0.8995	0.8984
FL [1]	0.1571	0.2698	0.3248	0.4799	0.4793	0.4976	0.5726	0.5520	0.4794
RCS [4]	0.4766	0.6129	0.7306	0.7761	0.7896	0.7897	0.8222	0.8252	0.8271
RLPHCS	**0.6805**	**0.8800**	**0.9123**	**0.9298**	**0.9356**	**0.9307**	**0.9445**	**0.9494**	**0.9551**
Results on URBAN dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.6827	0.7992	0.8998	0.9106	0.9239	0.9516	0.9565	0.9511	0.9617
StOMP [7]	0.0660	0.0385	0.4117	0.5409	0.8068	0.8664	0.8292	0.8929	0.8726
LASSO [8]	0.7010	0.8382	0.8587	0.8912	0.8943	0.9030	0.9185	0.9113	0.9062
FL [1]	0.7500	0.726	0.7169	0.8299	0.8093	0.8312	0.8436	0.8331	0.8490
RCS [4]	0.7328	0.8945	0.9258	0.9396	0.9499	0.9481	0.9503	0.9526	0.9469
RLPHCS	**0.9063**	**0.9501**	**0.9701**	**0.9765**	**0.9825**	**0.9854**	**0.9869**	**0.9883**	**0.9885**
Results on PAVIAU dataset									
	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$	$\rho = 0.30$	$\rho = 0.35$	$\rho = 0.40$	$\rho = 0.45$	$\rho = 0.50$
OMP [22]	0.6897	0.8627	0.9328	0.9467	0.956	0.9597	0.9632	0.9657	0.9653
StOMP [7]	0.0056	0.0780	0.3543	0.6096	0.8212	0.8467	0.8872	0.9089	0.9270
LASSO [8]	0.7630	0.8606	0.8822	0.8959	0.8961	0.9223	0.9037	0.9115	0.9063
FL [1]	0.4177	0.7570	0.8383	0.8500	0.8788	0.8791	0.8586	0.8702	0.8794
RCS [4]	0.7870	0.9105	0.9433	0.9549	0.9592	0.9583	0.9562	0.9521	0.9480
RLPHCS	**0.9302**	**0.9652**	**0.9781**	**0.9809**	**0.9825**	**0.9839**	**0.9847**	**0.9848**	**0.9848**



(a) (b) (c) (d) (e) (f)

Figure 5. The PSNR curves and SAM bar charts of different methods on three datasets under SNR=$30db$. (a)(b)(c) PNSR curves. (d)(e)(f) SAM bar charts.

RLPHCS exceeds other competing methods on PSNR at least $1.9db$ on INDANA data, $3.3db$ on URBAN data and $4.9db$ on PAVIAU data, respectively.

The bar charts of SAM values under three different levels of noise on three datasets are given in Figure 3(d)(e)(f), Figure 4(d)(e)(f) and Figure 5(d)(e)(f), respectively. Under each noise level, the SAM values of RLPHCS are almost smaller than all other methods on each dataset. For example, when SNR of measurements is $20db$ and $30db$, SAM values of RLPHCS are smaller than 10 degree on the PAVIAU dataset when the sampling rate is larger than 0.1.

The comparison results of SSIM under three levels of noise on three datasets are shown in Table 1, Table 2 and Table 3, respectively. It can be seen that RLPHCS has the highest SSIM scores among all methods. Specifically, when SNR of measurements is $20db$, the SSIM scores of the proposed RLPHCS are larger than 0.9 on PAVIAU data when the sampling rate $\rho \geq 0.1$. The visual reconstruction results with the sampling rate $\rho = 0.2$ on the measurements of SNR =$20db$ are shown in Figure 6, where the proposed RLPHCS also yields the most approximate results to the original bands visually.

The results evaluated by three measures indicate that the proposed RLPHCS outperforms other methods on the reconstruction accuracy of HSI under different levels of noise. When SNR is lower, the advantage of the proposed method is more obvious. This is because RLPHCS can capture the underlying structure of the sparse signal and regulate the configuration of the sparsity prior adaptively to cope with the unknown noise.

4. Conclusions

In this study, we propose a novel reweighted Laplace prior based HCS method, where the configuration of the reweighted Laplace prior is learned by a latent variable Bayes model. The learned prior depicts the underlying structured sparsity of HSI very well and is adaptive to the unknown noise. This helps the algorithm reconstruct the HSI precisely. Moreover, the signal recovery, prior learning and noise estimation are integrated into a unified variational framework where all unknown variables are inferred automatically. In the end, plenty experimental results illustrate the superiority of the proposed method over several state-of-the-art HCS methods on the reconstruction accuracy.

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), and Basic Science Research Fund in Xidian University (No.JB140222).

References

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Bayesian compressive sensing using laplace priors. *Image Processing, IEEE Transactions on*, 19(1):53–63, 2010.

[2] 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.

[3] J. Bioucas-Dias and A. Plaza. Hyperspectral unmixing: Geometrical, statistical, and sparse regression-based approaches. In *Proc. SPIE*, volume 7830, page 78300A, 2010.

[4] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

[5] 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.

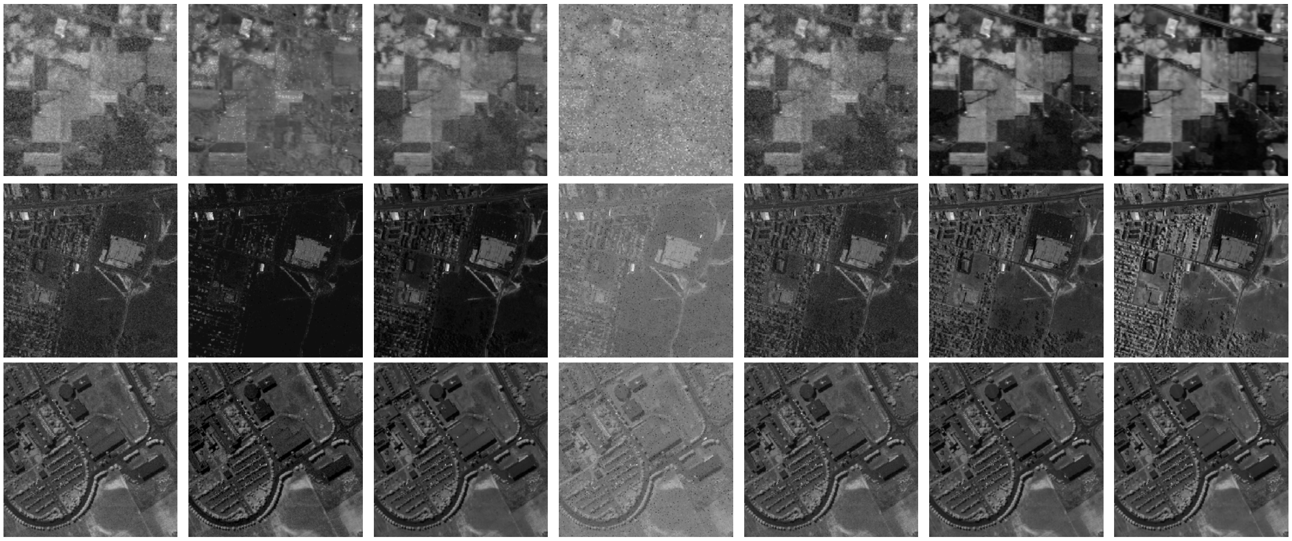
[6] D. L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[7] 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.

[8] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[9] 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.

[10] J. B. Greer. Sparse demixing of hyperspectral images. *Image Processing, IEEE Transactions on*, 21(1):219–228, 2012.



(a) (b) (c) (d) (e) (f) (g)

Figure 6. Reconstruction results of the 20th band from INDIANA, the 60th band from URBAN and the 90th band from PAVIAU with sampling rate $\rho = 0.2$ when the SNR of measurements is $20db$. All comparison methods recover the HSI in a wavelet transform based HCS framework. (a) OMP. (b) StOMP. (c) LASSO. (d) FL. (e) RCS. (f) RLPHCS. (g) Original bands.

[11] Z. Guo, T. Wittman, and S. Osher. L1 unmixing and its application to hyperspectral image enhancement. In *SPIE Defense, Security, and Sensing*, pages 73341M–73341M. International Society for Optics and Photonics, 2009.

[12] J. Hahn, C. Debes, M. Leigsnering, and A. M. Zoubir. Compressive sensing and adaptive direct sampling in hyperspectral imaging. *Digital Signal Processing*, 26:113–126, 2014.

[13] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza. Sparse unmixing of hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(6):2014–2039, 2011.

[14] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, 56(6):2346–2356, 2008.

[15] 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.

[16] G. Martin, J. M. Bioucas-Dias, and A. Plaza. Hyperspectral coded aperture (hyca): A new technique for hyperspectral compressive sensing. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5. IEEE, 2013.

[17] 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.

[18] 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.

[19] A. Rajwade, D. Kittle, T.-H. Tsai, D. Brady, and L. Carin. Coded hyperspectral imaging and blind compressive sensing. *SIAM Journal on Imaging Sciences*, 6(2):782–812, 2013.

[20] M. Tan, I. W. Tsang, and L. Wang. Matching pursuit lasso part i: Sparse recovery over big dictionary. *Signal Processing, IEEE Transactions on*, 63(3):727–741, 2015.

[21] 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.

[22] 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.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[24] D. Wipf and S. Nagarajan. A unified bayesian framework for meg/eeg source imaging. *Neuroimage*, 44(3):947–966, 2009.

[25] 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.

[26] H. Zhang and D. Wipf. Non-uniform camera shake removal using a spatially-adaptive sparse penalty. In *Advances in Neural Information Processing Systems*, pages 1556–1564, 2013.