

# Gradient Type Methods for Linear Hyperspectral Unmixing

Fangfang Xu<sup>1,\*</sup>, Yating Wang<sup>1</sup>, Yanyan Li<sup>1</sup>, Lu Liu<sup>2</sup> and Tonghua Tian<sup>3</sup>

<sup>1</sup> College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, P.R. China.

<sup>2</sup> Institute of Software, Chinese Academy of Sciences, Beijing 100190, P.R. China.

<sup>3</sup> School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA.

Received 2 January 2021; Accepted 7 July 2021

---

**Abstract.** Hyperspectral unmixing (HU) plays an important role in terrain classification, agricultural monitoring, mineral recognition and quantification, and military surveillance. The existing model of the linear HU requires the observed vector to be a linear combination of the vertices. Due to the presence of noise, or any other perturbation source, we relax this linear constraint and penalize it to the objective function. The obtained model is solved by a sequence of gradient type steps which contain a projection onto the simplex constraint. We propose two gradient type algorithms for the linear HU, which can find vertices of the minimum volume simplex containing the observed hyper-spectral vectors. When the number of given pixels is huge, the computational time and complexity are so large that solving HU efficiently is usually challenging. A key observation is that our objective function is a summation of many similar simple functions. Then the computational time and complexity can be reduced by selecting a small portion of data points randomly. Furthermore, a stochastic variance reduction strategy is used. Preliminary numerical results showed that our new algorithms outperformed state-of-the-art algorithms on both synthetic and real data.

**AMS subject classifications:** 65K05, 68W20, 90C30

**Key words:** Hyperspectral unmixing, minimum volume simplex, linear mixture model, alternating minimization, proximal gradient method, adaptive moments method, stochastic variance reduction strategy.

---

## 1 Introduction

Hyperspectral unmixing (HU) is a source separation problem, which is widely used in terrain classification, agricultural monitoring, mineral recognition and quantification,

---

\*Corresponding author. Email address: xuff@sdu.edu.cn (F. Xu)

and military surveillance [1, 5, 18, 22]. HU aims at decomposing pixel spectra in a scene into materials, and their corresponding fractional abundances. The materials are also called endmembers, which are generally considered to represent the pure materials present in the image. The set of abundances at each pixel is considered to represent the percentage of each endmember that is occupied in the pixel. The materials in hyperspectral unmixing are statistically dependent and combine in a linear or nonlinear fashion. Because the linear model of hyperspectral unmixing is simple, whose physical meaning is clear, and the solution is satisfactory, it is very suitable for hyperspectral unmixing. We only consider linear hyperspectral unmixing [14, 23, 26, 30] in this paper.

Algorithms of linear hyperspectral unmixing mainly fall into four types: geometrical based approaches [4, 27], statistical based approaches [24, 25], sparse regression based approaches [6], and spatial/spectral joint analysis [29]. The geometrical based approaches can be categorized into two main classes of methods: pure pixel (PP) [27] based and minimum volume (MV) [14] based methods. The pure pixel based algorithms still belong to the minimum volume class but assume the presence in the data of at least one pure pixel per endmember. This kind of algorithm finds the set of the purest pixels in the data, see for instances pixel purity index (PPI) [7, 8], N-FINDR [34], iterative error analysis (IEA) [28], vertex component analysis (VCA) [27], simplex growing (SGA) [12], sequential maximum angle convex cone (SMACC) [16], alternating volume maximization (AVMAX) [11]. The minimum volume approaches seek a mixing matrix  $M$  that minimizes the volume of the simplex defined by its columns, referred to as  $\text{conv}(M)$ , subject to the constraint that  $\text{conv}(M)$  contains the observed spectral vectors. The pure pixel constraint is no longer enforced, resulting in a much harder nonconvex optimization problem. The minimum volume approaches include minimum volume simplex analysis (MVSA) [20], simplex identification via variable splitting and augmented Lagrangian (SISAL) [4], minimum volume enclosing simplex (MVES) [10], iterative constrained endmembers (ICE) [3], convex cone analysis (CCA) [17], etc. Geometrical based approaches have a light computational burden and clear conceptual meaning, but may lead to poor results in highly mixed scenarios. Statistical methods are powerful alternative in highly mixed scenarios at the cost of higher computational complexity. They are mainly based on independent component analysis, Bayesian method and non-negative matrix factorization. In sparse regression based unmixing, endmembers are assumed to exist in a huge spectral library, and each pixel of the image can be expressed by the linear combinations of a number of spectra in the spectral library. Because the endmembers are very rare compared to the spectral library, images can be sparsely expressed by the spectral library. Sparse regression based unmixing is an area with strong links to compressed sensing, least angle regression, basis and matching pursuits. Spatial/spectral joint analysis suppose that pixels are not isolated alone, but in a 3D natural scene. The endmembers of the hyperspectral image can be extracted by combining the spectral and spatial information of the surrounding pixels.

The existing model of linear HU requires the observed vector to be a linear combination of the vertices. Due to the presence of noise, or contaminations from other sources,

the spectral vectors may lie outside the true data simplex. In this paper, we construct a new model which relaxes the equality constraint and penalizes it onto the objective function. Then we propose two gradient type algorithms to solve the corresponding linear HU models. Since the number of pixels may be huge, the computational time and complexity are so large that solving HU efficiently is challenging. A key observation is that our objective function is a summation of many similar simple functions. The information involved in observed vectors may have redundancy, so we can reduce computational time and complexity by picking up a portion of data points randomly. Furthermore, many first-order randomized algorithms have been presented recently, including stochastic gradient method [32], batch gradient method, and some noise reduction methods, such as Prox-SVRG [35] and SAGA [15]. The projected mini-batch gradient descent method is a randomized algorithm of linear HU, its model is based on a nonnegative matrix factorization and the online kernel-based nonnegative matrix factorization method [21, 36]. We will apply the stochastic variance reduction strategy to reduce the computational cost in this paper.

The remainder of this article is organized as follows. In Section 2, we propose a new linear hyperspectral unmixing model. Three gradient type algorithms are given in Section 3. When the number of given pixels is large, we apply the stochastic variance reduction strategy to HU and design a proximal stochastic variance reduction gradient algorithm. Section 4 presents simulation numerical results on simulated and real data. Section 5 ends the paper by presenting a few concluding remarks.

The following notations will be used throughout this paper. Upper (lower) case letters are used for matrices (column vectors). All vectors are column vectors, and the subscript  $(\cdot)^\top$  denotes matrix and vector transposition.  $I_n$  stands for the  $n \times n$  identity matrix, and  $\mathbf{1}_p$  denotes the vector of all ones. The Frobenius norm of  $X \in \mathbb{R}^{m \times n}$  is defined as  $\|X\|_F = \sqrt{\sum_{i,j} |X_{i,j}|^2}$ . The Euclidean inner product between two vectors  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^n$  is defined as  $\langle x, y \rangle = \sum_i (x_i y_i) = x^\top y$ . The Euclidean inner product between two matrices  $X \in \mathbb{R}^{m \times n}$  and  $Z \in \mathbb{R}^{m \times n}$  is defined as  $\langle X, Z \rangle = \sum_{i,j} (X_{i,j} Z_{i,j}) = \text{trace}(X^\top Z)$ .  $\hat{1}(A)$  denotes the indicator function of the set  $A$ . The inequality  $X \geq \mathbf{0}$  is element-wise, which means  $X_{ij} \geq 0$  for all entries  $(i, j)$ . Likewise, the equality  $X = Z$  means  $X_{ij} = Z_{ij}$  for all entries  $(i, j)$ .

## 2 A model for linear hyperspectral unmixing

Assuming that there are  $p$  endmembers in a given scene with spectral signatures  $m_i \in \mathbb{R}^B$ ,  $i=1, \dots, p$ , where  $B \geq p$  denotes the number of spectral bands. In the linear mixture model, the observed spectral vectors are in the convex hull of endmember spectral signatures, i.e.,

$$Y = MS,$$

where  $Y = [y_1, \dots, y_n] \in \mathbb{R}^{B \times n}$  denotes a matrix holding the observed spectral vectors  $y_i \in \mathbb{R}^B$ ,  $S = [s_1, \dots, s_n] \in \mathbb{R}^{p \times n}$  is a matrix holding the respective fractions and  $M = [m_1, \dots, m_p] \in$

$\mathbb{R}^{B \times p}$  is the mixing matrix containing the endmembers. That is to say, for each observed vector,  $y_i = Ms_i$ ,  $i = 1, \dots, n$ . Since the components of  $s_i$  are nonnegative and their summation are equal to one, the fractional abundance vectors  $s_i$  ( $i = 1, \dots, n$ ) belong to the standard  $p$ -simplex set  $S_p = \{s \in \mathbb{R}^p : s \geq \mathbf{0}, \mathbf{1}_p^\top s = 1\}$ .

The number of endmembers  $p$  is usually much smaller than the number of bands  $B$ . Assuming that the linear model is a good approximation, spectral vectors lie in or very close to a low-dimensional linear subspace. Therefore, a signal subspace identification algorithm is required as the first preprocessing step. The signal space is denoted by  $\mathcal{U}$ , and let the columns of  $E = [e_1, \dots, e_p]$  be an orthonormal basis for  $\mathcal{U}$ , where  $e_i \in \mathbb{R}^B$ , for  $i = 1, \dots, p$ . The coordinates of the orthogonal projection of a spectral vector  $y \in \mathbb{R}^B$  onto  $\mathcal{U}$ , with respect to the basis  $E$ , are given by  $y_{\mathcal{U}} = E^\top y \in \mathbb{R}^p$ . Thus, we have

$$y_{\mathcal{U}} = E^\top Ms.$$

From now on, we assume that the observed data set has been projected onto the signal subspace, and for simplicity of notation, we still represent the projected vectors as

$$Y = MS,$$

where  $Y \in \mathbb{R}^{p \times n}$  and  $M \in \mathbb{R}^{p \times p}$ .

Since the volume defined by the column of  $M$  is proportional to  $\log|\det(M)|$ . Linear hyperspectral unmixing can be conducted by minimizing  $\log|\det(M)|$  with respect to  $s_i$  belonging to the standard  $p$ -simplex set and the linear constraint  $Y = MS$ . The corresponding model was used in [4]:

$$\begin{aligned} \min_{M, S} \quad & \log|\det(M)|, \\ \text{s.t.} \quad & Y = MS, \\ & S \geq \mathbf{0}_{p \times n}, \quad \mathbf{1}_p^\top S = \mathbf{1}_n^\top. \end{aligned} \quad (2.1)$$

Let  $Q = M^{-1}$ , the equality constraint  $Y = MS$  is equivalent to  $QY = S$ . Moreover,  $\log|\det(M)|$  is equal to  $-\log|\det(Q)|$ . Due to the presence of noise, or any other perturbation source, the spectral vectors may lie outside the true data simplex. We relax the equality constraint  $QY = S$  and penalize it onto the objective function. A new minimum volume class model can be obtained as follows:

$$\begin{aligned} \min_{Q, S} \quad & \frac{1}{2} \|QY - S\|_F^2 - \lambda \log|\det(Q)|, \\ \text{s.t.} \quad & S \geq \mathbf{0}_{p \times n}, \quad \mathbf{1}_p^\top S = \mathbf{1}_n^\top. \end{aligned} \quad (2.2)$$

where  $\lambda$  is the parameter used to balance the noise and simplex volume. Let  $u(Q, S) = \frac{1}{2} \|QY - S\|_F^2$ ,  $h(Q) = -\lambda \log|\det(Q)|$ ,  $v(S) = \hat{I}_{S_p}(S)$ , and  $\phi(Q, S) = u(Q, S) + v(S)$ . Then Model (2.2) is equivalent to

$$\min_{Q, S} \Psi(Q, S) = u(Q, S) + h(Q) + v(S). \quad (2.3)$$

$\widehat{I}_{\mathcal{S}_p}(S)$  is convex because the set  $\mathcal{S}_p$  is convex. Moreover,  $u(Q, S)$  is jointly convex in  $(Q, S)$ , and its gradient  $\nabla u(Q, S)$  is Lipschitz continuous. Model (2.3) can be solved by the alternating minimization method [9]. Alternating minimization method for model (2.3) is derived by successively minimizing  $\Psi(Q, S)$  with respect to one of the variables  $Q$  and  $S$  at a time while fixing the other at its most recent value. We firstly find  $\hat{S} := \hat{S}(QY)$  that minimizes  $\Psi(Q, S)$ . Denotes

$$f(Q) = \min_S \phi(Q, S), \quad (2.4)$$

which is equivalent to

$$\begin{aligned} \min_S \quad & \frac{1}{2} \|QY - S\|_F^2, \\ \text{s.t.} \quad & S \geq \mathbf{0}_{p \times n}, \quad \mathbf{1}_p^\top S = \mathbf{1}_n^\top. \end{aligned} \quad (2.5)$$

The above problem is essentially a projection onto the simplex constraints and its solution is available in Lemma 2.1 [33].

**Lemma 2.1.** *Let  $\alpha_j = (QY)_j$  be the  $j$ th column of  $QY$ , we sort  $\alpha_j$  into  $u$  such that  $u_1 \geq u_2 \geq \dots \geq u_p$ , and find  $\rho = \max\{1 \leq l \leq p: u_l + \frac{1}{l}(1 - \sum_{i=1}^l u_i) > 0\}$ . The closed-form solution of model (2.5) is:*

$$\hat{s}_j = \max\{\alpha_j + \eta_j, 0\}, \quad j = 1, \dots, n$$

where  $\eta_j = \frac{1}{\rho}(1 - \sum_{i=1}^{\rho} u_i)$ .

After  $\hat{S}$  is determined,  $\hat{Q}$  can be solved by

$$\min_Q \phi(Q) = f(Q) + h(Q). \quad (2.6)$$

Finally, model (2.2) is transformed into a problem with respect to the variable  $Q$  itself. Note that model (2.6) is a nonconvex problem due to the nonconvexity of  $h(Q)$ . Since  $u(Q, S)$  is jointly convex with respect to  $Q$  and  $S$  and it is strongly convex with respect to  $S$ , the function  $f(Q)$  is differentiable with respect to  $Q$  and its gradient is:

$$\nabla f(Q) = \nabla \phi(Q, \hat{S}) = (QY - \hat{S})Y^\top, \quad (2.7)$$

where  $\hat{S}$  is the unique solution of model (2.5). Since model (2.5) is actually a projection onto a convex set and the solution operator  $\hat{S}(QY)$  is non-expansive, we have

$$\|\hat{S}(QY) - \hat{S}(\bar{Q}Y)\|_F \leq \|(Q - \bar{Q})Y\|_F \leq \|Q - \bar{Q}\|_F \|Y\|_F.$$

Consequently, we obtain

$$\begin{aligned} \|\nabla f(Q) - \nabla f(\bar{Q})\|_F &\leq \|(Q - \bar{Q})Y\|_F + \|(\hat{S}(QY) - \hat{S}(\bar{Q}Y))Y^\top\|_F \\ &\leq 2\|Y\|_F^2 \|Q - \bar{Q}\|_F. \end{aligned}$$

Hence, the gradient  $\nabla f(Q)$  is Lipschitz continuous with a constant  $2\|Y\|_F^2$ . If  $\det(Q) \neq 0$ , the gradient of  $\phi(Q)$  is

$$\nabla \phi(Q) = (QY - \hat{S})Y^\top - \lambda Q^{-1}. \quad (2.8)$$

### 3 Gradient type methods

In this section, we solve model (2.6) in detail and present three gradient type methods.

#### 3.1 Proximal gradient method

The proximal gradient algorithm is a special gradient type method, it is mainly used to solve the optimization problem whose objective function is not differentiable. The idea of the proximal gradient algorithm is to apply the proximal operator to a gradient descent step with respect to the differentiable part. We compute  $Q^{k+1}$  of model (2.6) with one proximal gradient step:

$$\begin{aligned} Q^{k+1} &= \arg\min_Q \left\langle \nabla f(Q^k), Q - Q^k \right\rangle + \frac{1}{2\tau} \|Q - Q^k\|_F^2 + h(Q), \\ &= \arg\min_Q \frac{1}{2} \|Q - (Q^k - \tau \nabla f(Q^k))\|_F^2 + \tau h(Q), \end{aligned} \quad (3.1)$$

$$= \text{Prox}_{\tau h}(W^k), \quad (3.2)$$

where  $W^k = Q^k - \tau \nabla f(Q^k)$ ,  $\tau = (2 + \gamma) \|Y\|_F^2$  with  $\gamma > 0$ , and

$$\text{Prox}_{\tau h}(W^k) = \arg\min_Q \frac{1}{2} \|Q - W^k\|_F^2 + \tau h(Q).$$

Next, let us see how to compute  $Q^{k+1}$  in detail.

**Lemma 3.1.** *The solution of proximal operator in Eq. (3.2) is*

$$\text{Prox}_{\tau h}(W) = U\tilde{Q}V^\top, \quad (3.3)$$

where  $U$  and  $V$  are orthogonal matrices in the SVD decomposition of  $W$ ,  $\tilde{Q} = \text{diag}\{q_1, \dots, q_p\}$ ,  $q_i = \frac{w_i + \sqrt{w_i^2 + 4\tau\lambda}}{2}$  ( $i = 1, \dots, p$ ),  $w_i > 0$  is the  $i$ th singular value of  $W$ .

*Proof.* Let  $U\tilde{W}V^\top$  be the singular value decomposition of  $W$ , where  $\tilde{W} = \text{diag}\{w_1, \dots, w_p\}$ ,  $w_i$  is the  $i$ th singular value of  $W$ ,  $w_i > 0$ ,  $U$  and  $V$  are orthogonal matrices. Let  $\mathbf{w} = (w_1, \dots, w_p)^\top$ . Model (3.1) is equivalent to

$$\min_{\mathbf{q} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{q} - \mathbf{w}\|_F^2 - \tau \lambda \sum_{i=1}^p \log |q_i|.$$

Note that the above problem is separable,  $q_i$  ( $i = 1, 2, \dots, p$ ) is the optimal solution of

$$\min_{q_i} \frac{1}{2} (q_i - w_i)^2 - \tau \lambda \log |q_i|. \quad (3.4)$$

The solution of model (3.4) is

$$q_i = \frac{w_i + \sqrt{w_i^2 + 4\tau\lambda}}{2}, \quad w_i > 0.$$

Therefore, the solution of proximal operator in Eq. (3.2) is  $\text{Prox}_{\tau h}(W) = U\tilde{Q}V^\top$ , where  $\tilde{Q} = \text{diag}\{q_1, \dots, q_p\}$ .  $\square$

To sum up, we obtain the proximal gradient method framework for linear hyperspectral unmixing problem. The framework is illuminated in Algorithm 1.

---

**Algorithm 1:** Proximal Gradient Method (PGM)

---

```

1 Given  $Q^0, \lambda > 0, k=0$ ;
2 while stopping criterion not satisfied do
3   Given  $Q^k$ , calculate  $S$  by Lemma 2.1 and  $\nabla f(Q^k)$  by Eq. (2.7);
4   Compute  $Q^{k+1}$  by Eq. (3.3);
5    $k \leftarrow k+1$ .
```

---

Model (2.6) falls into the framework of [9]. The convergence of Algorithm 1 can be guaranteed by Proposition 3 in [9], and the generated sequence of Algorithm 1 is globally convergent to a critical point of model (1), as shown in Theorem 3.1.

**Theorem 3.1.** Let  $\{Q^k\}_{k \in \mathcal{N}}$  be a sequence generated by Algorithm 1. Assume that  $\varphi(Q)$  is bounded from below on  $\{Q \mid \varphi(Q) \leq \varphi(Q^0)\}$ .

(1) The sequence  $\{Q^k\}_{k \in \mathcal{N}}$  has finite length, that is

$$\sum_{k=1}^{\infty} \|Q^{k+1} - Q^k\|_F < \infty.$$

(2) The sequence  $\{Q^k\}_{k \in \mathcal{N}}$  converges to a critical point  $Q^*$  of model (2.6).

*Proof.* Since  $\varphi(Q^k)$  is bounded from below  $\{Q \mid \varphi(Q) \leq \varphi(Q^0)\}$  and  $f(Q)$  is nonnegative,  $h(Q^k)$  is bounded which implies that  $Q^k$  is bounded. Because  $h(Q)$  satisfies the KL property, the convergence of Algorithm 1 can be guaranteed by Proposition 3 in [9].  $\square$

In practice,  $\tau$  is usually computed by the BB (Barzilai and Borwein) [2] method. At the  $k$ -th iteration, let  $t = Q^k - Q^{k-1}$  and  $z = \nabla \varphi(Q^k) - \nabla \varphi(Q^{k-1})$ , then

$$\tau_k = \langle t, t \rangle / \langle t, z \rangle, \quad (3.5)$$

or

$$\tau_k = \langle t, z \rangle / \langle z, z \rangle. \quad (3.6)$$



### 3.2 Adaptive moments method

Since model (2.6) is non-convex, the solution is not necessarily optimal. The adaptive momentum method (ADAM) is robust and suitable for non-convex and large-scale problems. It corrects the iteration direction and adjusts the step size adaptively [19]. In this subsection, we apply Adam to solve model (2.6) and get a gradient descent algorithm. The update rules of Adam for model (2.6) are given by

$$\begin{aligned} g_k &= \nabla \varphi(Q^k), \\ H_k &= \rho_1 H_{k-1} + (1 - \rho_1) g_k, \quad \hat{H}_k = \frac{H_k}{1 - \rho_1^k}, \\ G_k &= \rho_2 G_{k-1} + (1 - \rho_2) g_k \odot g_k, \quad \hat{G}_k = \frac{G_k}{1 - \rho_2^k}, \\ Q^{k+1} &= Q^k - \tau_k \frac{\hat{H}_k}{\sqrt{\hat{G}_k + \delta}}, \end{aligned}$$

where  $\odot$  is component-wise multiplication and the division is also component-wise.  $G_k$  is a weighted mean of  $G_{k-1}$  and the accumulated squared gradients. The parameter  $\tau_k$  is the BB step size and  $\delta$  is a small constant for numerical stability. The parameters  $\rho_1$  and  $\rho_2$  are decaying rates. The second and the third equations above perform a correction on  $H_k$  and  $G_k$ .

Adaptive moments method for linear hyperspectral unmixing problem is described as Algorithm 2.

---

**Algorithm 2:** Adaptive Moments Method (ADAM)

---

- 1 Given  $Q^0$ ,  $\tau_0 > 0$ ,  $\lambda > 0$ ,  $\rho_1$ ,  $\rho_2$ ,  $g_0 = \nabla \varphi(Q^0)$ ,  $H_0 = (1 - \rho_1)g_0$ ,  $G_0 = (1 - \rho_2)g_0 \odot g_0$ ,  $\delta = 1e-7$ ,  $k = 0$ ;
  - 2 **while** *stopping criterion not satisfied* **do**
  - 3     Given  $Q^k$ , calculate  $S$  by Lemma 2.1;
  - 4     Calculate the gradient  $g_k$  by Eq. (2.8);
  - 5     Compute  $H_k = \rho_1 H_{k-1} + (1 - \rho_1)g_k$ ,  $G_k = \rho_2 G_{k-1} + (1 - \rho_2)g_k \odot g_k$ ;
  - 6     Correct bias:  $\hat{H}_k = \frac{H_k}{1 - \rho_1^k}$ ,  $\hat{G}_k = \frac{G_k}{1 - \rho_2^k}$ ;
  - 7      $Q^{k+1} = Q^k - \tau_k \frac{\hat{H}_k}{\sqrt{\hat{G}_k + \delta}}$ , where  $\tau_k$  is calculated by Eq. (3.5) or Eq. (3.6);
  - 8      $k \leftarrow k + 1$ .
- 

The convergence of ADAM depends on the suitable stepsizes and algorithm parameters. Although model (2.6) is non-convex, ADAM converge to the first-order stationary solutions under a set of mild sufficient conditions. For details of convergence of ADAM for non-convex problems, please refer to [13].



### 3.3 Proximal gradient method by variance reduction

When the number of given pixels is large, we use proximal stochastic variance reduction gradient algorithm (Prox-SVRG) to reduce computational time and complexity in this section. Prox-SVRG progressively reduces the variance of the stochastic gradient. More specifically, Prox-SVRG computes the full gradient at the beginning of each stage, followed by  $m$  stochastic gradient inner iterations.

Note that the objective function of model (2.6) can be written as

$$\frac{1}{2} \|QY - S\|_F^2 - \lambda \log |\det(Q)| = \frac{1}{2} \sum_{i=1}^n \|Qy_i - s_i\|_2^2 - \lambda \log |\det(Q)|.$$

Therefore, model (2.6) is equivalent to

$$\min_Q \frac{1}{n} \sum f_i(Q) - \frac{\lambda}{n} \log |\det(Q)|, \quad (3.7)$$

where

$$f_i(Q) = \min_{s_i} \frac{1}{2} \|Qy_i - s_i\|_2^2, \quad \text{s.t. } s_i \geq \mathbf{0}, \quad \mathbf{1}_p^\top s_i = 1. \quad (3.8)$$

The optimal solution of model (3.8) is denoted by  $\hat{s}_i := \hat{s}_i(Qy_i)$ . Then the gradient of  $f_i(Q)$  is  $\nabla f_i(Q) = (Qy_i - \hat{s}_i)y_i^\top$ , and the gradient of  $F(Q) = \frac{1}{n} \sum_{i=1}^n f_i(Q)$  is  $\nabla F(Q) = \frac{1}{n} (QY - \hat{S})Y^\top$ . Let  $R(Q) = -\frac{\lambda}{n} \log |\det(Q)|$ , and it admits a simple proximal mapping obviously. Therefore, we can apply Prox-SVRG to model (2.6). Our stochastic algorithm for linear hyperspectral unmixing is presented in Algorithm 3.

---

**Algorithm 3:** Proximal Gradient Method via Variance Reduction (PGMVR)

---

```

1 Given  $\tilde{Q}_0$  ;
2 for  $t=1,2,\dots$  do
3    $\tilde{Q} = \tilde{Q}_{t-1}$ ;
4   solve  $S$  by Lemma 2.1;
5    $\tilde{v}_t = \nabla F_P(\tilde{Q}) = \frac{1}{n} (\tilde{Q}Y - S)Y^\top$ ;
6    $\tau_t = \frac{1}{m} \|\tilde{Q}_t - \tilde{Q}_{t-1}\|_2^2 / |(\tilde{Q}_t - \tilde{Q}_{t-1})^\top (\tilde{v}_t - \tilde{v}_{t-1})|$ ;
7    $Q_0 = \tilde{Q}$ ;
8   for  $k=1,2,\dots,m$  do
9     pick  $I_k \subset \{1,2,\dots,n\}$  randomly ;
10     $v_k = \nabla f_{I_k}(Q_{k-1}) - \nabla f_{I_k}(\tilde{Q}) + \tilde{v}_t$ ;
11     $Q_k = \text{Prox}_{\tau_t R}(Q_{k-1} - \tau_k v_k)$ ;
12   $\tilde{Q}_t = Q_m$  ;
13  If stopping criterion is satisfied, break.
```

---

Because  $R(Q)$  is non-convex, problem (3.7) is the sum of convex function and non-convex regularizer. The convergence of Prox-SVRG demands convexity or strong convexity on both  $F(Q)$  and  $R(Q)$ , therefore PGMVR may diverge [35]. But we will demonstrate the efficiency of PGMVR in the numerical experiments.

## 4 Numerical results

In this section, we verify the effectiveness of our proposed algorithms for linear hyperspectral unmixing problems. A series of experiments are carried out to show that our algorithms are suitable for linear HU.

The main contribution of this paper is proposing the new model (2.2) of linear hyperspectral unmixing. PGM, ADAM, PGMVR are all based on model (2.2). We compare the proposed method with related methods. The details of these evaluated algorithms (including our algorithms) are listed as follows:

- The proximal gradient method (PGM in short), presented in Subsection 3.1.
- The adaptive moments method (ADAM in short), presented in Subsection 3.2.
- The proximal gradient method via variance reduction (PGMVR in short), presented in Subsection 3.3.
- The simplex identification via split augmented Lagrangian (SISAL in short). It is a geometrical method. In SISAL, the positivity constraints, forcing the spectral vectors to belong to the convex hull of the endmember signatures, are replaced by soft constraints. The obtained problem is solved by a sequence of augmented Lagrangian optimizations. The code for this method is downloaded from <http://www.lx.it.pt/~bioucas/code.htm>.
- The vertex component analysis (VCA in short). It is a classic geometrical method that needs the existence of pure pixels for each endmember. Different with the other algorithms that estimate the endmembers and abundances simultaneously, VCA can only estimate the endmembers. The abundances are estimated by solving a constrained least square Problem. The code for this algorithm is obtained from <http://www.lx.it.pt/~bioucas/code.htm>.
- The projected mini-batch gradient descent method (MBSGD in short) [21]. It uses a mini-batch of samples or pixels every time. This algorithm is a stochastic gradient descent approach. Since its original code is not available, we implement it ourselves.
- The online kernel-based nonnegative matrix factorization method (OKNMF in short) [36]. It is an online learning method using the mini-batch strategy. The code for this algorithm is obtained from <http://www.honeine.fr/paul/publi/17.oknmf.zip>.

All algorithms are written in MATLAB. All experiments are performed on a Lenovo PC with Intel(R) Core(TM) i7-6500 CPU at 2.50GHz and 8G of memory running Windows 7 and MATLAB 2016a.

#### 4.1 Simulated data

The data was generated according to the linear observation model. The abundance fractions are Dirichlet distributed with parameter  $\mu_i = 1, i = 1, \dots, p$ . The mixing matrix  $M$  is randomly generated by  $\text{rand}(L, p)$ . To ensure that no pure pixel is present, we discarded all pixels with any abundance fractions larger than 0.8. See Fig. 1 for simulated data. Every star dot is an endmember, while blue dots denote simulated data point. Note that there is no pure pixel in simulated data.

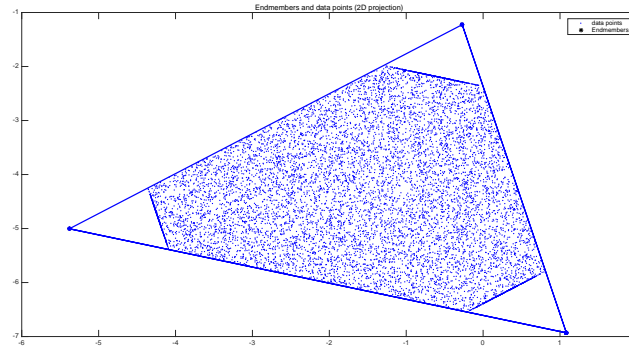


Figure 1: Diagram of simulated data.

All the parameters of SISAL are set to be their defaults values. The stopping criterion is  $\|(QY - S)Y^\top - \lambda Q^{-\top}\|_F < 10^{-4}$ , where  $Q^{-\top}$  denotes the transport matrix of  $Q^{-1}$ . The maximum number of iterations is set to be 2000 in PGM and ADAM, and 100 in PGMVR. The other parameters in PGM, ADAM, and PGMVR are set as Table 1. The other parameters of MBSGD, and OKNMF are set to be their defaults values. SNR is defined as

$$\text{SNR} = 10 \log \left( 10 \frac{\text{signal voltage}}{\text{noise voltage}} \right).$$

The meaning of  $\text{SNR} = \infty$  is that there is no noise in the data set. Because noise exists in the hyperspectral images, therefore we only consider  $\text{SNR} = 10, 20$  and  $30$  in this subsection.

In order to evaluate the proposed method, we use the spectral angle distance (SAD) [31]. SAD is used to evaluate the performance of estimated endmembers, which is an angle distance between an estimated endmember and its corresponding ground truth. It

Table 1: Parameter settings.

Methods	Parameters
PGM	$\tau_0 = 1$
ADAM	$\tau_0 = 1 \quad \rho_1 = 0.8 \quad \rho_2 = 0.9 \quad \delta = 1e-7$
PGMVR	$\tau_0 = 1 \quad m = 50$

Table 2: The average SADs vs. SNRs on simulated data.

SAD	SNR		
	10	20	30
PGM	0.0096	0.0109	0.0038
ADAM	0.0688	0.0108	0.0039
PGMVR	0.0091	0.0107	0.0038
SISAL	0.0852	0.0216	0.0043
VCA	0.0912	0.1237	0.1388
MBSGD	0.0230	0.0495	0.0545
OKNMF	0.2020	0.2035	0.2037

is defined as

$$\text{SAD}(M_i, \hat{M}_i) = \arccos \left( \frac{M_i^\top \hat{M}_i}{\|M_i\| \cdot \|\hat{M}_i\|} \right), \quad (4.1)$$

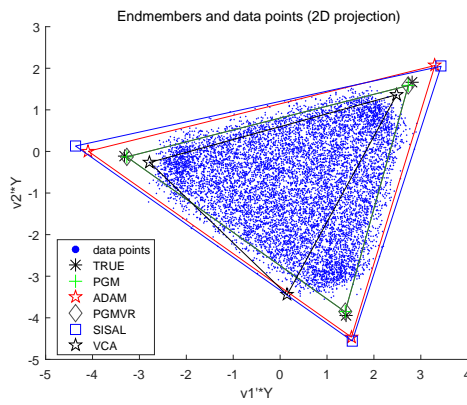
where  $M_i$  denotes the ground truth of one endmember, and  $\hat{M}_i$  is the corresponding estimated result. The smaller SAD, the better performance.

We first perform numerical experiments to study the relationship between SAD and the number of pixels for  $p = 3$  and  $N = 10000$ . When the number of pixels  $N$  is 10000, we use the whole data set to identify the endmembers. Experiment results for  $p = 3$  and  $N = 10000$  are reported in Tables 2 and 3. The endmembers found by five algorithms are compared in Fig. 2. As we can see, VCA does not perform well because there are no pure pixels in the simulated data. Because VCA is very fast, it is used to provide an initial value  $Q_0$  for PGM, ADAM, and PGMVR.

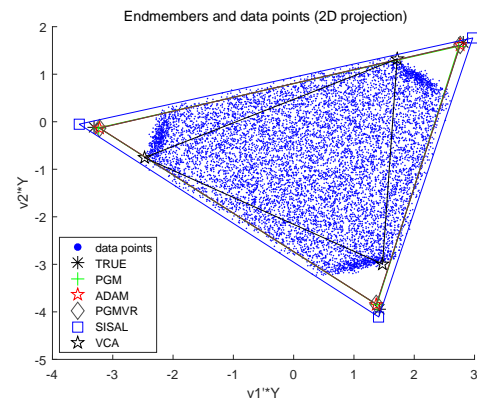
When the number of pixels  $N$  is larger than 10000, we select  $n$  columns from  $Y$  randomly to generate a submatrix  $\hat{Y}$ . Then  $\hat{Y}$  is used in all models to find endmembers. The procedure is based on the founding that most pixels are redundant to recover only 3 endmembers. To test the efficiency of the stochastic variance reduction strategy, we do some numerical experiments on large data sets.  $n$  is set to be  $0.1N$ . Let  $p = 3$  and  $\text{SNR} = 20$ . The numerical experiment results are listed in Tables 4 and 5. Model (2.2) performs very well on large data sets for noise case in comparison with results from other approaches.

Table 3: The average time vs. SNRs on simulated data.

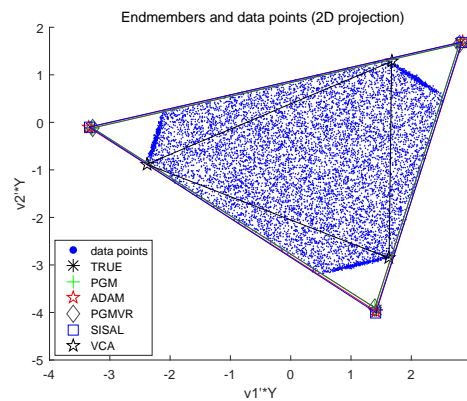
Time(s)	SNR		
	10	20	30
PGM	0.1174	0.1386	0.4399
ADAM	3.0399	2.2754	3.0773
PGMVR	0.3898	0.7711	2.7701
SISAL	6.8153	6.2807	8.7522
VCA	0.0626	0.0612	0.0567
MBSGD	1.7141	1.6148	1.6145
OKNMF	73.0677	65.5441	64.5498



(b) SNR = 10



(c) SNR = 20



(d) SNR = 30

Figure 2: Average SAD of different algorithms for different SNRs.

Table 4: The average SADs and time on big data for noise case ( $N=100000$ ).

Algorithms	PGM	ADAM	PGMVR	SISAL	VCA	MBSGD	OKNMF
SAD	0.0143	0.0124	0.0103	0.0146	0.1356	0.0391	0.2102
Time (s)	0.0220	0.1389	0.3294	1.4032	0.0143	0.1944	6.2752

Table 5: The average SADs and time on big data for noise case ( $N=1000000$ ).

Algorithms	PGM	ADAM	PGMVR	SISAL	VCA	MBSGD	OKNMF
SAD	0.0121	0.0117	0.0118	0.0237	0.1204	0.0496	0.2054
Time (s)	0.1358	2.6395	0.9019	6.2508	0.0695	1.7616	73.6146

## 4.2 Real data

We now perform some experiments on four real data examples [37–39], see Fig. 3. They are available at [http://www.escience.cn/people/feiyunZHU/Dataset\\\_GT.html](http://www.escience.cn/people/feiyunZHU/Dataset\_GT.html), and widely used in the hyperspectral unmixing study.

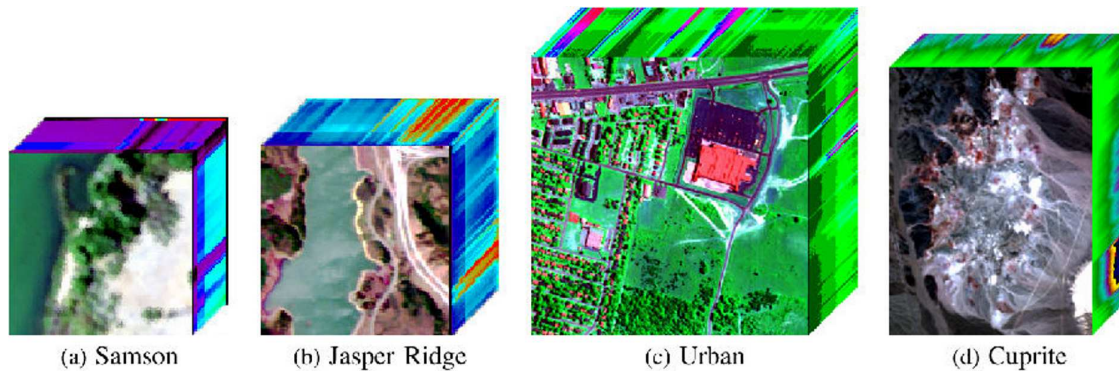


Figure 3: Four real hyperspectral images, ie, Samson, Jasper Ridge, Urban and Cuprite.

The size of Samson is  $95 \times 95 \times 156$ , in which there are three endmembers, i.e. soil, tree, and water. The size of Jasper Ridge is  $100 \times 100 \times 198$ , and its endmembers include road, soil, water, and tree. Urban has the largest size:  $307 \times 307 \times 162$ . Ground truth contains 6 endmembers: asphalt, grass, tree, roof, metal, and dirt. Cuprite's size is  $250 \times 190 \times 224$ , and it has 12 endmembers. Due to dense water vapor and atmospheric effects, we remain 188 channels after the common preprocess for Cuprite. Each data set is first transformed to a  $B \times n$  matrix  $Y$ . Then it is preprocessed according to Section 2. Note that the ground truth  $M$  of each data set is only used to measure the effectiveness of the algorithms.

Firstly, we perform numerical experiments to study the relationship between SAD in estimating the endmember signatures and the number of pixels. The results are shown in Fig. 4. The figure shows that SAD tends to be decreased as the number of pixels is

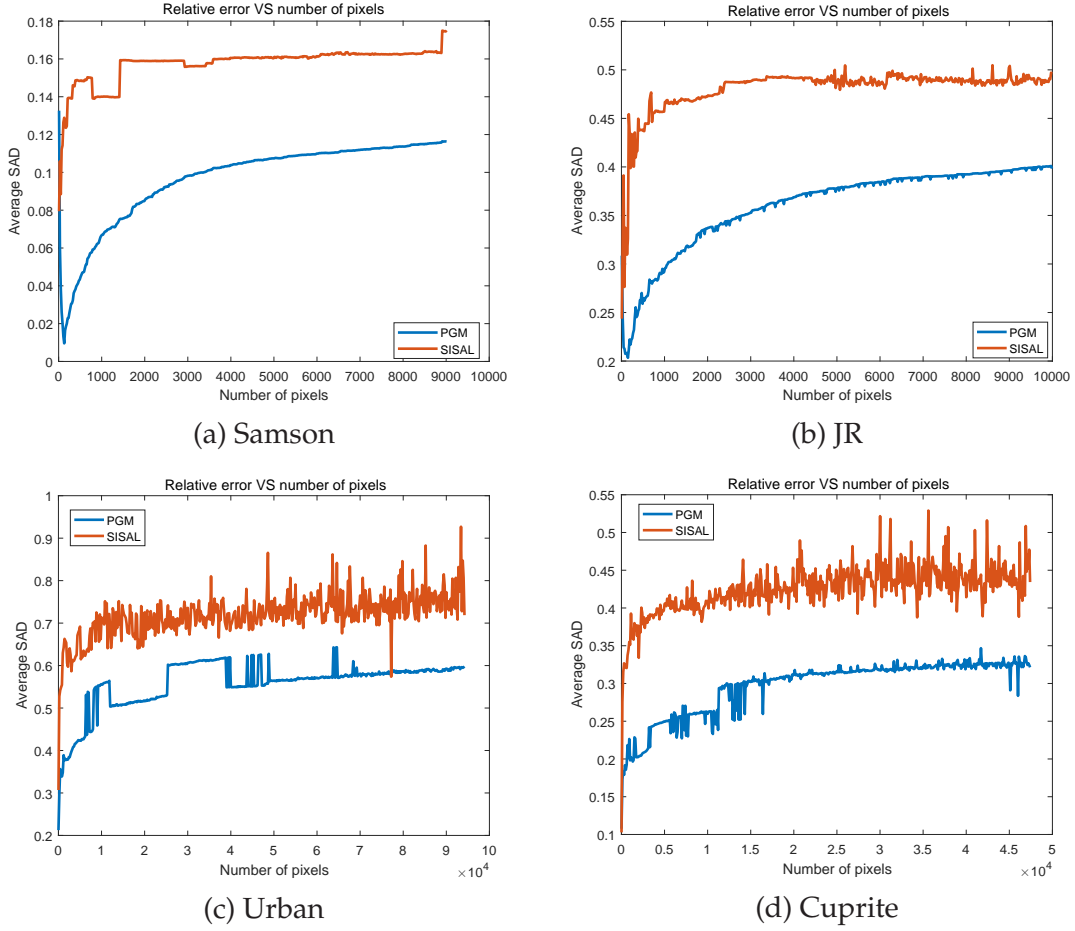


Figure 4: Average SAD of different algorithms for different data examples.

increased at the early stage, then SAD tends to be increased afterwards. Since the best number of pixels is unknown as a prior, we choose 100 pixels since it usually is enough to recover the endmembers for the four data sets. That is, we set  $n = 100$  to identify the endmember signatures  $M$ . We repeat the above procedure for 50 times and calculate the average SADs. The box plots are shown in Fig. 5.

We use PGM, ADAM, and SISAL to identify the endmembers and fractional abundances of four real hyperspectral images. All the parameters of SISAL are set to be their defaults values. The maximum number of iterations is set to be 1000 in PGM, and ADAM. The maximum number of iterations in PGMVR is set to be 10.  $\lambda$  is set to be 2. The other parameters are the same as in Table 1. Here, we ignore VCA because there are no pure pixels in this data set. The average spectral angle distances are list in Tables 6 to 9. Compared with non-stochastic algorithms and the other two stochastic algorithms, PGMVR is comparable to all other algorithms.



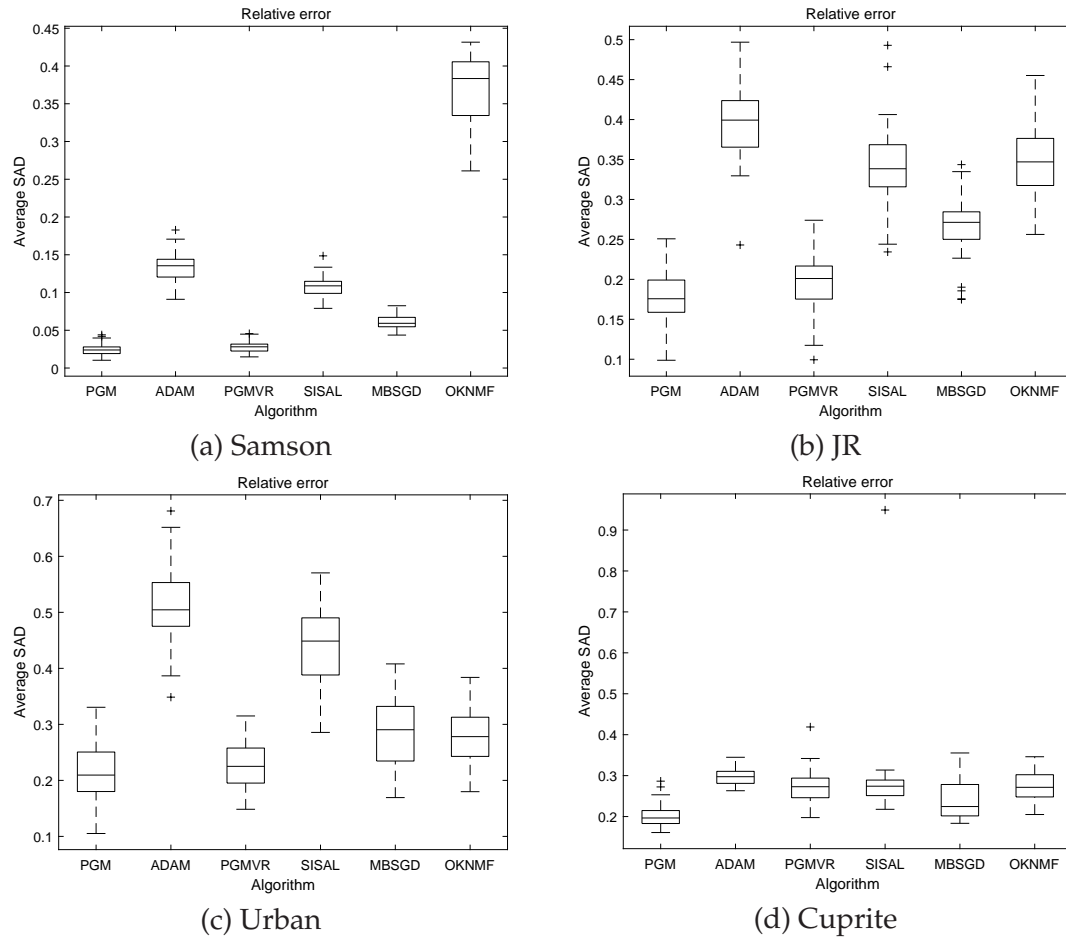
Figure 5: Box plot of SAD for  $n = 100$ .

Table 6: SAD and time of different algorithms for Samson.

Algorithms	SAD				Time(s)
	1-rock	2-tree	3-water	Mean value	
PGM	0.0157	0.0146	0.0198	0.0167	0.0101
ADAM	0.1154	0.0706	0.2477	0.1446	0.0470
PGMVR	0.0190	0.0172	0.0272	0.0211	0.0139
SISAL	0.0207	0.0613	0.2625	0.1148	0.0929

The spectral signatures of different algorithms are presented in Figs. 6 and 7. Figs. 8 and 9 shows the abundances of different algorithms. The first row is the ground truth, the following rows are respectively given by PGM, ADAM, PGMVR, and SISAL in Figs. 8

Table 7: SAD and time of different algorithms for Jasper Ridge.

Algorithms	SAD					Time(s)
	1-tree	2-water	3-dirt	4-road	Mean value	
PGM	0.0547	0.0157	0.0390	0.7097	0.2048	0.0206
ADAM	0.2120	0.2524	0.5411	0.7053	0.4277	0.0643
PGMVR	0.0545	0.0485	0.0289	0.7266	0.2146	0.0150
SISAL	0.2211	0.1710	0.4052	0.8265	0.4059	0.1158

Table 8: SAD and time of different algorithms for Urban.

Algorithms	SAD							Time(s)
	1-asphalt	2-grass	3-tree	4-roof	5-metal	6-dirt	Mean value	
PGM	0.1565	0.1103	0.0580	0.1990	0.2365	0.5654	0.2210	0.4520
ADAM	0.6198	0.4562	0.5639	0.6116	0.8695	0.9647	0.6809	0.1354
PGMVR	0.1510	0.1004	0.0632	0.1987	0.1830	0.5248	0.2035	0.0359
SISAL	0.4318	0.2653	0.1906	0.5563	0.6702	0.7462	0.4767	0.0729

Table 9: SAD and time of different algorithms for Cuprite.

SAD	Algorithms			
	PGM	ADAM	PGMVR	SISAL
1-Alunite	0.1724	0.1453	0.1850	0.1482
2-Andradite	0.0685	0.1188	0.0870	0.1046
3-Buddingtonite	0.1067	0.2023	0.1364	0.1325
4-Dumortierite	0.0802	0.1120	0.1412	0.1116
5-Kaolinite <sub>1</sub>	0.0872	0.1293	0.1037	0.1267
6-Kaolinite <sub>1</sub>	0.0599	0.2146	0.1122	0.1697
7-Muscovite	0.1518	0.2873	0.1892	0.2059
8-Montmorillonite	0.0921	0.2202	0.1336	0.2399
9-Nontronite	0.1155	0.2000	0.1322	0.2504
10-Pyrope	0.0523	0.2578	0.1551	0.2436
11-Sphene	0.1849	0.6078	0.2498	0.4795
12-Chalcedony	1.0490	1.3520	0.9327	1.3181
Mean value	0.1850	0.3206	0.2132	0.2942
Time(s)	0.4483	0.1764	0.0150	0.0987

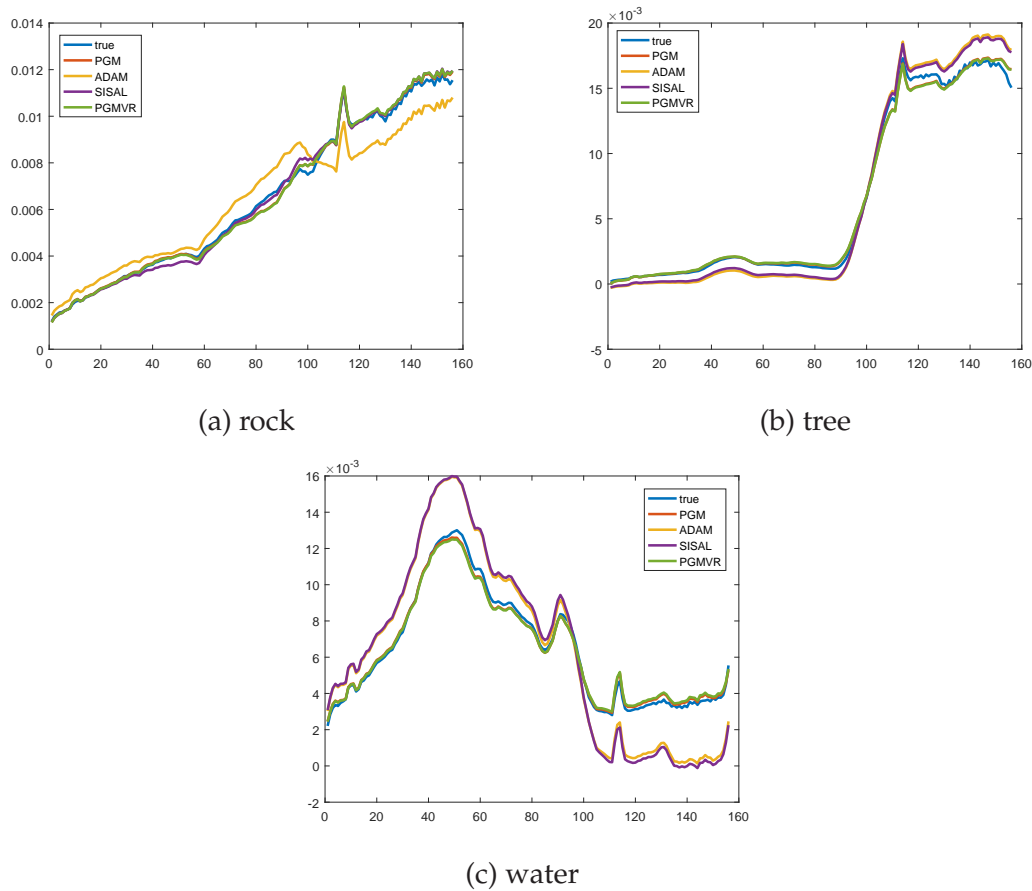


Figure 6: Spectral Signatures estimated by different algorithms compared with the ground truth of Samson.

and 9. Because the endmembers of Urban and Cuprite are too many, here we do not present their spectral signatures and abundances because of the limit of space. From Figs. 6 and 7, we observe that spectral signatures of endmembers are similar to ground truth, only some parts deviate slightly. From Figs. 8-9, we observe that abundances of SISAL are all darker than PGM and PGMVR.

Obviously, model (2.2) performs very well on real data, and the stochastic variance reduction strategy plays an important role in large scale problems.

## 5 Conclusions and future works

In this paper, we introduce a new linear unmixing model and its corresponding algorithms PGM and ADAM. The unmixing is achieved by finding the minimum volume simplex containing the hyperspectral data. This optimization problem is solved by a se-

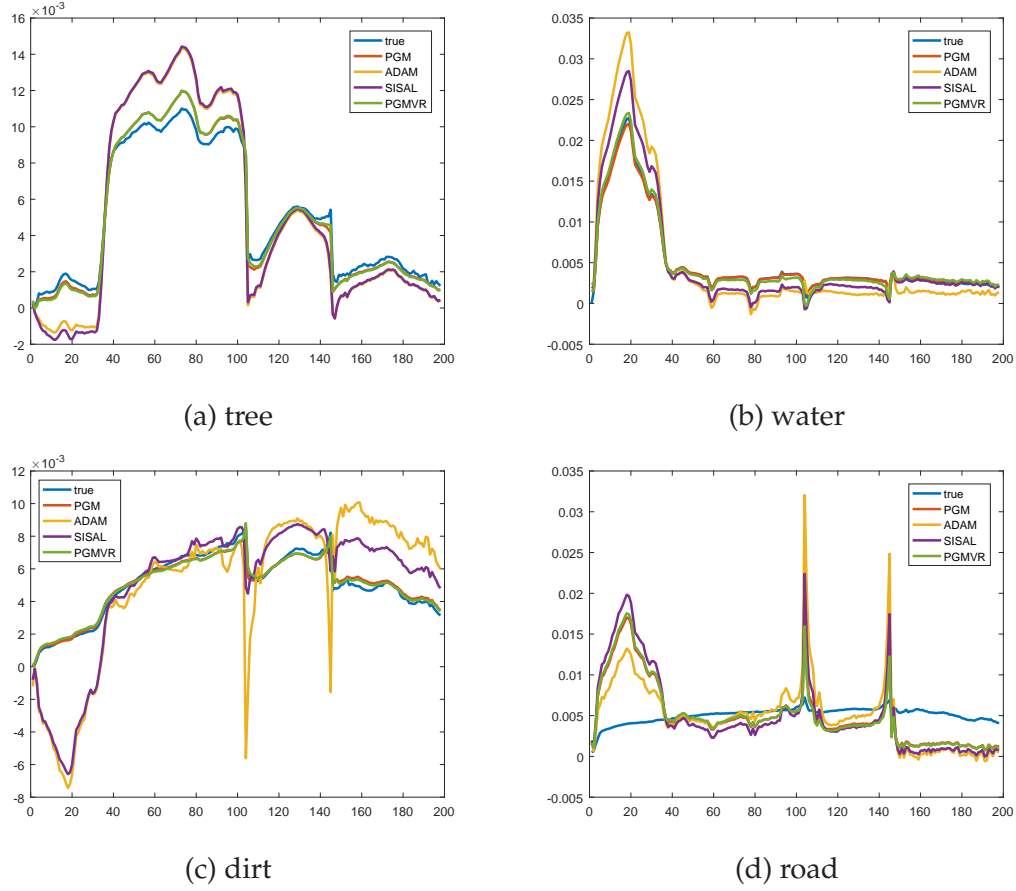


Figure 7: Spectral Signatures estimated by different algorithms compared with the ground truth of Jasper Ridge.

quence of proximal gradient steps in PGM while a special sequence of gradient steps with an adaptive learning rate in ADAM. To deal with large data sets, we present a random algorithm PGMVR to reduce the computational complexity. The numerical results demonstrate the efficiency of these algorithms.

## Acknowledgments

The work of Fangfang Xu, Yating Wang, and Yanyan Li was supported in part by NSFC grant 11901359 and Shandong Provincial Natural Science Foundation grant ZR2019QA017. The work of Lu Liu and Tonghua Tian was supported in part by NSFC grant 11831002 and 91730302 and the National Basic Research Project under the grant 2015CB856002.

The authors are grateful to Prof. Zaiwen Wen and Prof. Bin Dong for the helpful discussion and valuable suggestions.

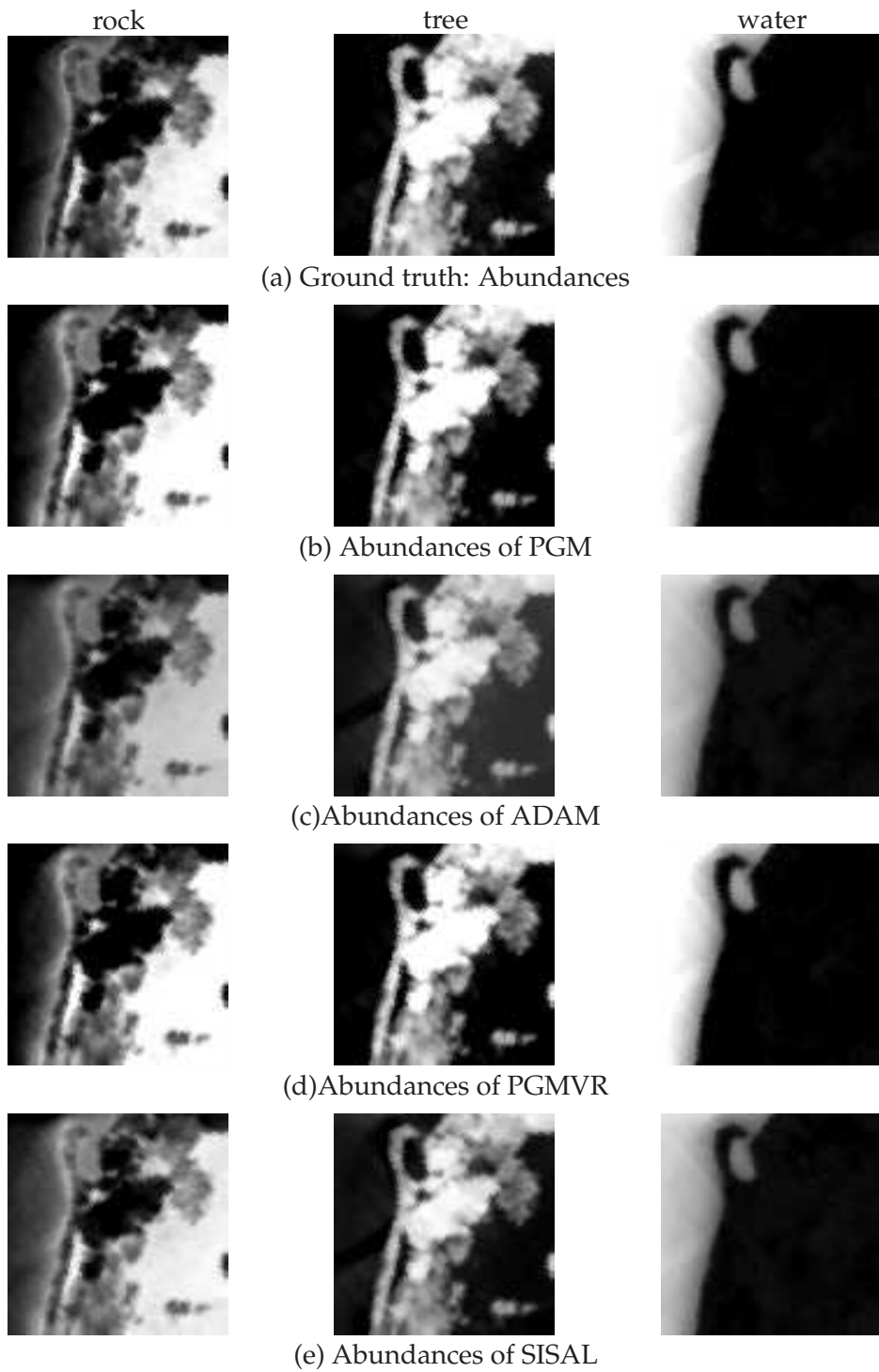


Figure 8: Abundances estimated by different algorithms compared with the ground truth of Samson.

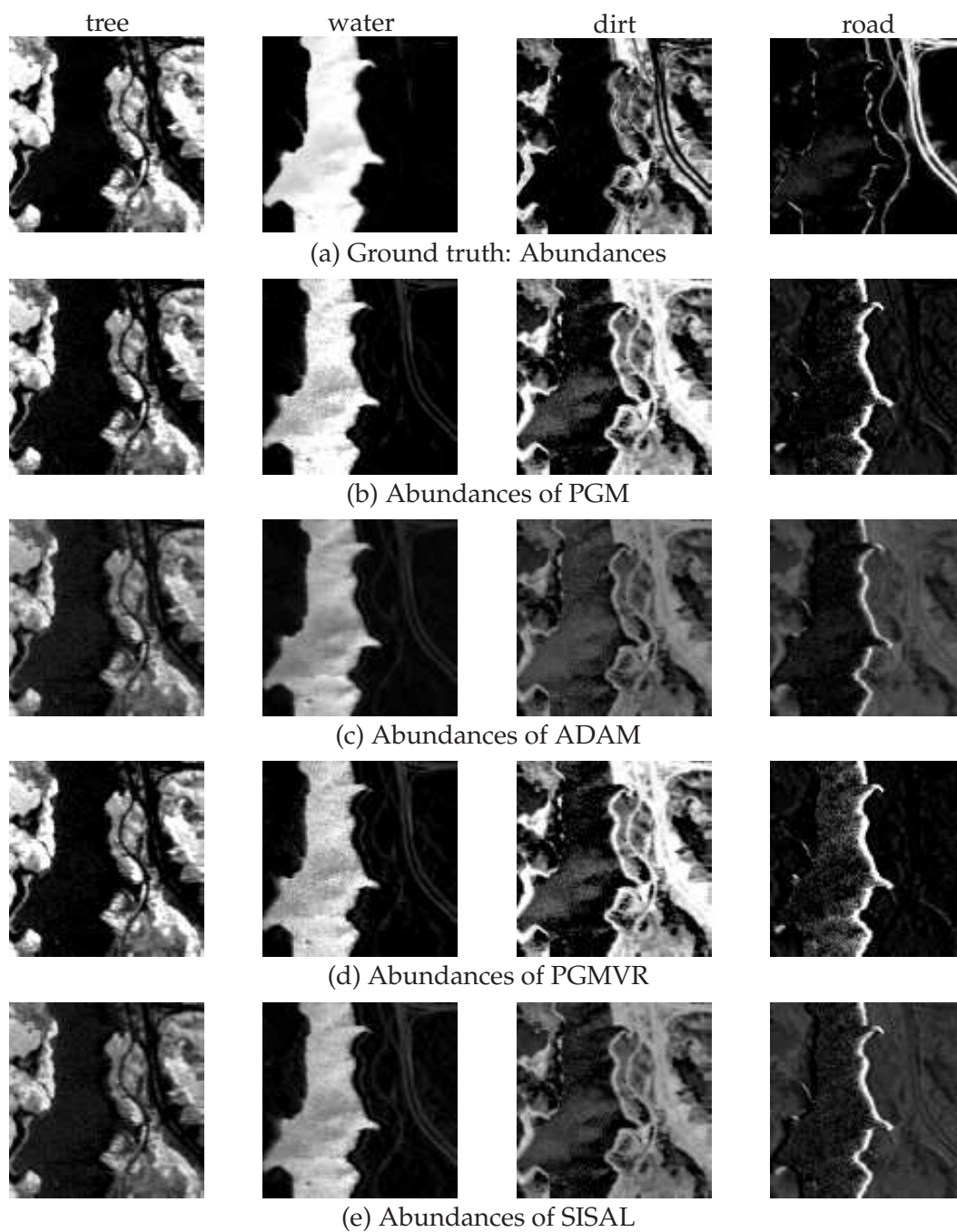


Figure 9: Abundances estimated by different algorithms compared with the ground truth of Jasper Ridge.

## References

- [1] J. B. ADAMS, M. O. SMITH, AND P. E. JOHNSON, *Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 Site*, Journal of Geophysical Research, 91 (1986), pp. 8098–8112.
- [2] J. BARZILAI AND J. M. BORWEIN, *Two-point step size gradient methods*, IMA Journal of Numerical Analysis, 8 (1988), pp. 141–148.
- [3] M. BERMAN, H. KIIVERI, R. LAGERSTROM, A. ERNST, R. DUNNE, AND J. F. HUNTINGTON, *ICE: A statistical approach to identifying endmembers in hyperspectral images*, IEEE transactions on Geoscience and Remote Sensing, 42 (2004), pp. 2085–2095.
- [4] J. M. BIOUCAS-DIAS, *A variable splitting augmented Lagrangian approach to linear spectral unmixing*, in: Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on, IEEE, 2009, pp. 1–4.
- [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*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5 (2012), pp. 354–379.
- [6] J. M. BIOUCASDIAS AND M. A. T. FIGUEIREDO, *Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing*, Optimization and Control, (2010), pp. 1–4.
- [7] J. W. BOARDMAN, *Mapping target signatures via partial unmixing of AVIRIS data*, JPL Airborne Geoscience Workshop, 1995.
- [8] ———, *Automating spectral unmixing of AVIRIS data using convex geometry concepts*, JPL Airborne Geoscience Workshop, 2013.
- [9] J. BOLTE, S. SABACH, AND M. TEBoulLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.
- [10] T. CHAN, C. CHI, Y. HUANG, AND W. MA, *Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing*, IEEE International Conference on Acoustics, Speech and Signal Processing, 57 (2009), pp. 1089–1092.
- [11] T. CHAN, W. MA, A. AMBIKAPATHI, AND C. CHI, *A simplex volume maximization framework for hyperspectral endmember extraction*, IEEE Transactions on Geoscience and Remote Sensing, 49 (2011), pp. 4177–4193.
- [12] C. CHANG, C. WU, W. LIU, AND Y. OUYANG, *A new growing method for simplex-based end-member extraction algorithm*, IEEE Transactions on Geoscience and Remote Sensing, 44 (2006), pp. 2804–2819.
- [13] X. CHEN, S. LIU, R. SUN, AND M. HONG, *On the convergence of a class of Adam-type algorithms for non-convex optimization*, in conference paper at ICLR 2019, 2019.
- [14] M. CRAIG, *Minimum-volume transforms for remotely sensed data*, IEEE Transactions on Geoscience and Remote Sensing, 32 (1994), pp. 542–552.
- [15] A. DEFAZIO, F. BACH, AND S. LACOSTE-JULIEN, *SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives*, in Advances in Neural Information Processing Systems, 2014, pp. 1646–1654.
- [16] J. H. GRUNINGER, A. J. RATKOWSKI, AND M. L. HOKE, *The sequential maximum angle convex cone (SMACC) endmember model*, in: Defense and Security, International Society for Optics and Photonics, 2004, pp. 1–14.
- [17] A. IFARRAGUERRI AND C. CHANG, *Multispectral and hyperspectral image analysis with convex cones*, IEEE Transactions on Geoscience and Remote Sensing, 37 (1999), pp. 756–770.



- [18] N. KESHAVA AND J. F. MUSTARD, *Spectral unmixing*, IEEE Signal Processing Magazine, 2002.
- [19] D. P. KINGMA AND J. L. BA, *Adam: A method for stochastic optimization*, International Conference on Learning Representations, 2015.
- [20] J. LI, A. AGATHOS, AND D. E. A. ZAHARIE, *Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing*, in: IEEE Transactions on Geoscience & Remote Sensing, vol. 53, IEEE, 2015, pp. 5067–5082.
- [21] J. LI, X. LI, AND L. ZHAO, *Hyperspectral unmixing via projected mini-batch gradient descent*, in: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, Texas, USA, 2017.7.23-2017.7.28, pp. 1133–1136.
- [22] S. LIANGROCAPART AND M. PETROU, *Mixed pixels classification*, Remote Sensing, (1998), pp. 72–83.
- [23] L. MIAO AND H. QI, *Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization*, IEEE Transactions on Geoscience and Remote Sensing, 45 (2007), pp. 765–777.
- [24] J. M. P. NASCIMENTO AND J. M. BIOUCASDIAS, *Hyperspectral unmixing algorithm via dependent component analysis*, (2007), pp. 4033–4036.
- [25] —, *Hyperspectral unmixing based on mixtures of Dirichlet components*, IEEE Transactions on Geoscience and Remote Sensing, 50 (2012), pp. 863–878.
- [26] J. M. P. NASCIMENTO AND J. M. B. DIAS, *Does independent component analysis play a role in unmixing hyperspectral data?*, IEEE Transactions on Geoscience and Remote Sensing, 43 (2005), pp. 175–187.
- [27] —, *Vertex component analysis: A fast algorithm to unmix hyperspectral data*, IEEE Transactions on Geoscience and Remote Sensing, 43 (2005), pp. 898–910.
- [28] R. NEVILLE, K. STAENZ, T. SZEREDI, J. LEFEBVRE, AND P. HAUFF, *Automatic endmember extraction from hyperspectral data for mineral exploration*, in: Proc. 21st Can. Symp. Remote Sens, 1999, pp. 21–24.
- [29] A. PLAZA, P. MARTINEZ, R. PEREZ, AND J. PLAZA, *Spatial/spectral endmember extraction by multidimensional morphological operations*, IEEE Transactions on Geoscience and Remote Sensing, 40 (2002), pp. 2025–2041.
- [30] A. PLAZA, P. MARTINEZ, R. M. PEREZ, AND J. PLAZA, *A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data*, IEEE Transactions on Geoscience and Remote Sensing, 42 (2004), pp. 650–663.
- [31] Y. QIAN, S. JIA, J. ZHOU, AND A. ROBLES-KELLY, *Hyperspectral unmixing via  $l_{1/2}$  sparsity-constrained nonnegative matrix factorization*, IEEE Transactions on Geoscience and Remote Sensing, 49 (2011), pp. 4282–4297.
- [32] H. ROBBINS AND S. MONRO, *A stochastic approximation method*, The Annals of Mathematical Statistics, (1951), pp. 400–407.
- [33] W. WANG AND M. A. CARREIRAPERPINAN, *Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application*, Mathematics, (2013).
- [34] M. E. WINTER, *N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data*, in: SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, International Society for Optics and Photonics, 1999, pp. 266–275.
- [35] L. XIAO AND T. ZHANG, *A proximal stochastic gradient method with progressive variance reduction*, SIAM Journal on Optimization, 24 (2014), pp. 2057–2075.
- [36] F. ZHU AND P. HONEINE, *Online nonnegative matrix factorization based on kernel machines*, in: Proc. 23rd European Conference on Signal Processing (EUSIPCO), Nice, France, 31 Aug.-4 Sept., 2015, pp. 2381–2385.

- [37] F. ZHU, Y. WANG, B. FAN, G. MENG, AND C. PAN, *Effective spectral unmixing via robust representation and learning-based sparsity*, Computer Science, (2014).
- [38] F. ZHU, Y. WANG, B. FAN, G. MENG, S. XIANG, AND C. PAN, *Spectral unmixing via data-guided sparsity*, Computer Science, (2014).
- [39] F. ZHU, Y. WANG, S. XIANG, B. FAN, AND C. PAN, *Structured sparse method for hyperspectral unmixing*, ISPRS Journal of Photogrammetry and Remote Sensing, 88 (2014), pp. 101–118.