

Research Article

Active Set Type Algorithms for Nonnegative Matrix Factorization in Hyperspectral Unmixing

Li Sun ¹, Congying Han ², and Ziwen Liu ²

¹College of Information Sciences and Engineering, Shandong Agricultural University, 271018, Taian, China

²School of Mathematical Science, University of Chinese Academy of Sciences, 100049 Beijing, China

Correspondence should be addressed to Li Sun; sunlishi@hotmail.com

Received 30 June 2019; Accepted 30 July 2019; Published 21 November 2019

Guest Editor: Marco Perez-Cisneros

Copyright © 2019 Li Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyperspectral unmixing is a powerful method of the remote sensing image mining that identifies the constituent materials and estimates the corresponding fractions from the mixture. We consider the application of nonnegative matrix factorization (NMF) for the mining and analysis of spectral data. In this paper, we develop two effective active set type NMF algorithms for hyperspectral unmixing. Because the factor matrices used in unmixing have sparse features, the active set strategy helps reduce the computational cost. These active set type algorithms for NMF is based on an alternating nonnegative constrained least squares (ANLS) and achieve a quadratic convergence rate under the reasonable assumptions. Finally, numerical tests demonstrate that these algorithms work well and that the function values decrease faster than those obtained with other algorithms.

1. Introduction

Hyperspectral data are acquired by high-resolution imaging sensors that record hundreds of continuous narrow spectral band images. The goals of hyperspectral imaging are to obtain the spectrum for each pixel in the image of the scene and identify materials or process objects. Because the spatial pixel size obtained by hyperspectral sensors is often large, more than one material can contribute to the spectrum measured from a single pixel, leading to the existence of mixed pixels. To improve the accuracy of the material classification and recognition process, hyperspectral unmixing (HU) is necessary. Certain objects leave unique fingerprints in the electromagnetic spectrum, and these spectral signatures enable the identification of the materials that compose a scanned pixel. HU aims at identifying materials (endmembers) present in a captured image and their corresponding fractional abundances using high spectral resolution of hyperspectral images.

In the past few decades, researchers have proposed many algorithms for HU based on the linear mixing model. When the endmember matrix is known a priori,

one can effectively recover the fractional abundances. Recently, nonnegative matrix factorization (NMF), which is consistent with the linear mixing model, has been proposed as a powerful tool for identifying the components of a certain dataset. In fact, the nonnegative constraints of NMF support the physical meaning of HU, and from a mathematical perspective, NMF divides the original matrix into two matrices. The first matrix contains the basis vectors, which approximate each column of the original matrix with the combination coefficient of a certain column of the second matrix. Based on the above description, NMF provides a new approach to HU [1].

In recent years, many methods have been proposed to solve the NMF problem. Berry et al. [2] divided NMF algorithms into three general classes: multiplicative updating algorithms, gradient descent algorithms, and alternating least squares algorithms. The prototypical multiplicative updating (MU) rules proposed by Lee and Seung [3] have been widely used as the baseline in relevant studies. Another widely used method of solving NMF problems is the projected gradient method [4]. With additional projections, the nonnegativity constraint is satisfied. Most projected gradient methods provide strong optimization properties [5–8].

Third, the alternating least squares (ALS) algorithm [2, 9], which alternatively fixes one matrix and improves the other with a simple least squares computation, is commonly applied. All the negative elements are set to 0 to meet the nonnegativity constraint.

To produce useful decompositions, some constraints on the scales of U or V are needed. Hoyer proposed the nonnegative sparse coding in [10, 11], and NMF with sparse constraint had been applied to unmix hyperspectral data [12–16]. Cai et al. presented graph-regularized nonnegative matrix factorization (GNMF) [17, 18]. GNMF combined with sparseness constraint could preserve the geometrical structure of the data and represent it in low dimensional space when applying in HU [19]. Also, orthogonal matrix factorization [20] and minimum volume-constrained NMF [21] have been put forward to enhance the decomposition results. Recently, multilayer NMF [22, 23] and deep NMF [24] complete the matrix decomposition in different layers, and they reduce the reconstruction error efficiently. In [25], an algorithm based on a clustered multitask network was proposed, which considered sparsity, clustering, and neighborhood information. Extension of different regularized NMF models has received significant attention in data mining [23, 26–34]. Most of these regularized NMF models are solved by the MU method because it is easy to implement and often yields good results. However, the MU method converges slowly due to a first-order convergence rate. To accelerate the convergence rate, some (quasi-)Newton type methods have been applied to solve NMF [35–37].

In this paper, we propose an active set type Newton method to solve nonnegative least squares problems. Because the combination coefficients in the abundance matrix are generally sparse, active set type algorithms are a good choice for solving NMF problems [37–40]. In [41–43], active set identification functions are employed to predict the constraints that are active at the solution. After the estimated active set is determined, the variables in the active set become active variables, and the rest of the variables are inactive variables. Finally, the direction in the subspace spanned by inactive variables is determined. These active set type algorithms are efficient in solving large-scale optimization problems.

This paper is organized as follows. In Section 2, we provide the basic definitions of important terms and the two active set identification techniques. In Section 3, we present the proposed active set type Newton method, which is based on the alternative nonnegative least squares approach. The convergence properties are also discussed. The proposed algorithms are compared to other commonly used NMF algorithms in Section 4.

We use the following notation in this paper. A superscript k is used to indicate the iteration number. If H is a matrix with elements h_{ij} , the double subscript indicates the element is in the i -th row and j -th column of H . Let I be an index set, such that $I = \{i_1, \dots, i_r\}$. We denote H_I as the $|I| \times |I|$ submatrix of H , where $i, j \in I$. If v is an n -element vector, we denote v_I as the subvector with components v_i , where $i \in I$. Some basic definitions and assumptions are stated in the next section.

2. Problem Formulation and Active Set Strategy

2.1. Nonnegative Matrix Factorization Problem Formulation.

In general, NMF can be mathematically formulated as follows. Given an input matrix $W \in R_+^{m \times n}$, in which each element of W is nonnegative, and a integer $r \ll \min\{m, n\}$, NMF aims to find the reduced rank nonnegative matrices $U \in R_+^{m \times r}$ and $V \in R_+^{r \times n}$, such that

$$W \approx UV. \quad (1)$$

We note that the spectral signature of each pixel of the hyperspectral image can be obtained from the combination of the nonnegative spectral signatures of the constitutive endmembers. In such a case, U is often considered the endmember matrix and V is the fractional matrix [44]. These results imply that endmember extraction and abundance maps creation can be simultaneously accomplished through NMF.

There are many methods measuring the accuracy of approximation given by (1). In this paper, we focus on the most common choice, which is the Frobenius norm. We solve the following minimization problem to find U and V :

$$\begin{aligned} \min \quad & f(U, V) = \frac{1}{2} \|W - UV\|_F^2, \\ \text{s.t.} \quad & U, V \geq 0, \end{aligned} \quad (2)$$

where the inequalities $U \geq 0$ and $V \geq 0$ indicate that all the elements of U and V are nonnegative.

By alternatively fixing one matrix and improving the other, the two-variable minimization problem (2) can be converted into a alternative nonnegative least squares (ANLS) minimization subproblems as follows:

$$\begin{aligned} \text{(ANLS1)} \quad & \min_{\text{s.t. } V \geq 0} f(U^k, V) = \frac{1}{2} \|W - U^k V\|_F^2, \\ \text{(ANLS2)} \quad & \min_{\text{s.t. } U \geq 0} f(U, V^{k+1}) = \frac{1}{2} \|W - UV^{k+1}\|_F^2. \end{aligned} \quad (3)$$

Because ANLS2 easily transformed to ANLS1 with the transposition of matrices W, U , and V and, when solving these equations, we are more accustomed to the form of $Ax = b$ rather than $x^T A^T = b^T$, we focus on solving ANLS1. We use ANLS uniformly, to correspond to ANLS1.

In HU, U represents the endmember matrix, the columns of which contain the spectra of the endmembers. After U is given, each row of the solution V provides the distribution of each endmember. An efficient solution for ANLS is the focus of the present work. Methods that rigorously solve for ANLS might add to the computational burden, neglect the nonnegative constraints, and solve the corresponding unconstrained least square problems; then, any negative values are simply set to zeros, which may yield poor estimates that are not the truly least squares [45, 46].

2.2. Basic Definition and Assumption. In fact, the ANLS can be treated as a large-scale bound-constrained optimization

problem, and we present three active set type optimization algorithms for solving ANLS.

The gradient of the objective function of the ANLS [3] is

$$\begin{aligned}\nabla_U f(U, V) &= (UV - W)V^T, \\ \nabla_V f(U, V) &= U^T(UV - W).\end{aligned}\quad (4)$$

From the KKT optimality conditions in [3], (U, V) is a stationary point of problem (2), if the multipliers λ_U and λ_V exist such that

$$\begin{cases} \nabla_U f(U, V) - \lambda_U = 0, \\ \nabla_V f(U, V) - \lambda_V = 0, \\ U \geq 0, \\ V \geq 0, \\ U \circ \lambda_U = 0, \\ V \circ \lambda_V = 0, \end{cases}\quad (5)$$

where \circ denotes the Hadamard product, which takes two matrices of the same dimensions and produces another matrix in which each element i, j is the product of the elements i, j of the original two matrices, i.e., $(V \circ \lambda_V)_{ij} = v_{ij}(\lambda_V)_{ij}$, $i = 1, \dots, r$, $j = 1, \dots, n$.

2.3. Two Active Set Identification Techniques. A constraint of the form $v_{ij} \geq 0$, $i = 1, \dots, r$, $j = 1, \dots, n$ is said to be active at the optimal solution V^* , if strict equality holds at v_{ij}^* , i.e., $v_{ij}^* = 0$, and inactive at v_{ij}^* , if strict inequality holds. For any feasible solution of the ANLS, the current iteration V is classified into two parts: active variables and inactive variables.

To facilitate the discussion, we concentrate on a vector of the columns of one matrix and use $\text{vec}(V)$ instead of V . Let $I(\text{vec}(V))$ be the index set that contains the estimated indices of the active variables. Then, the active set $I(\text{vec}(V)) = \{i : \text{vec}(V)_i = 0\}$ and the remaining index of $\text{vec}(V)$ is denoted by $F(\text{vec}(V)) = \{1, 2, \dots, nr\} \setminus I(\text{vec}(V))$.

We present two active set identification techniques that help estimate the active set of the optimal solution of the ANLS problem. The active set strategy will improve the accuracy of the ANLS solution, and the computational cost is reduced.

(i) Identify the active set with a constant

With a small positive constant ζ , we obtain an estimate of the active set:

$$(AS1)I(\text{vec}(V)) = \{i : \text{vec}(V)_i \leq \zeta\}.\quad (6)$$

(ii) Identify the active set with the multiplier

From the first equality of the KKT condition, we obtain the approximate multiplier:

$$\lambda_i^k = \begin{cases} \text{vec}(\nabla_V f(U^k, V))_i, & \text{if } \text{vec}(V)_i = 0, \\ 0, & \text{otherwise.} \end{cases}\quad (7)$$

Then, we have the second active set identification function, that is,

$$(AS2)I(\text{vec}(V)) = \{i : \text{vec}(V)_i \leq \varepsilon \lambda_i^k\},\quad (8)$$

where ε is a small positive constant.

We use $I^k = I(\text{vec}(V^k))$ and $F^k = F(\text{vec}(V^k))$ in the following equations. At iteration k , the variables of $\text{vec}(V)$ with indices in I^k are called active variables, and the remaining variables are called inactive variables.

A nonnegative matrix V^* is said to be a stationary point of the ANLS, if for every $i = 1, 2, \dots, nr$,

$$\begin{cases} \text{vec}(\nabla_V f(U, V))_i \geq 0, & \forall i \in I^*, \\ \text{vec}(\nabla_V f(U, V))_i = 0, & \forall i \in F^*, \end{cases}\quad (9)$$

where $I^* = \{i : \text{vec}(V^*)_i = 0\}$ and $F^* = \{1, 2, \dots, nr\} \setminus I^*$. Strict complementary condition holds for V^* , if the strict inequalities hold in the first constraint of (9).

3. Active Set Type Newton Algorithm for ANLS

The columns of V are used to generate a vector $\text{vec}(V)$, and the ANLS can be written as follows:

$$\begin{aligned}f(U, V) &= \frac{1}{2} \text{vec}(V)^T \begin{pmatrix} U^T U & & \\ & \ddots & \\ & & U^T U \end{pmatrix} \text{vec}(V) \\ &\quad - \text{tr}\left((W^T U)^T V\right) + \frac{1}{2} \text{tr}(W^T W),\end{aligned}\quad (10)$$

where $\text{tr}(\cdot)$ is the trace of a certain matrix [3, 4].

We assume that following standard assumption holds.

Assumption. U is a full-column rank matrix.

This assumption can be easily satisfied because r is a small number. Then, the Hessian matrix B^k of the ANLS is

$$B^k = \begin{pmatrix} U^T U & & \\ & \ddots & \\ & & U^T U \end{pmatrix} \in \mathbb{R}^{nr \times nr}.\quad (11)$$

Under the assumption, $U^T U$ is a positive definite matrix. It follows that ANLS is a strictly convex optimization problem and after the Newton method is applied, a second order local convergent rate is obtained. The empirical test of the full-column rank assumption indicated that this assumption is satisfied in the random case (see more details in [37]).

3.1. Active Set Type Algorithm for Solving the ANLS Problem.

In this section, we introduce an active set type algorithm to solve the ANLS problem. The Newton method is employed to update the inactive variables, and the active variables are updated in a simple manner, that is, by directly proceeding to the boundary or inside the feasible region.

3.1.1. Search Direction of the Inactive Variables. By extracting the information related to the inactive variables from B^k , we obtain

$$B_{F^k}^k = P_{F^k}^{kT} B^k P_{F^k}^k, \quad (12)$$

where $P_{F^k}^k \in R^{rn \times m_k}$, m_k is the number of elements in F^k , and the column of $P_{F^k}^k$ is the unit vector e_i . If we assume that $F^k = \{i_1, i_2, \dots, i_{m_k}\}$, the columns of $P_{F^k}^k$ are e_{i_l} , where $l = 1, 2, \dots, m_k$. In this case, e_{i_l} is the i_l -th column of the $rn \times rn$ identity matrix.

Lemma 1. *Under the above assumption, $B_{F^k}^k$ is a positive definite matrix.*

Proof. Because U is full-column rank matrix, $U^T U$ is a positive definite matrix and B^k is positive definite.

Then, $\forall x \neq 0$, and it follows that

$$x^T B_{F^k}^k x = x^T (P_{F^k}^{kT} B^k P_{F^k}^k) x = (P_{F^k}^k x)^T B^k (P_{F^k}^k x) > 0. \quad (13)$$

Consequently, $B_{F^k}^k$ is a positive definite matrix.

Because $B_{F^k}^k$ is a positive definite matrix, it follows that $\forall z \in R^{m_k}$, where $z \neq 0$, and

$$\lambda_1 \|z\|^2 \leq z^T B_{F^k}^k z \leq \lambda_n \|z\|^2, \quad (14)$$

where λ_1 and λ_n are the smallest and largest eigenvalues of $B_{F^k}^k$, respectively.

The search direction of the inactive variables $\text{vec}(V)_{J^k} \in R^{m_k}$ is finally defined in (14). The hyperspectral data are sparse, which indicates that there may be many estimated active variables. As a result, the computational cost is significantly reduced.

$$d_{F^k}^k = (B_{F^k}^k)^{-1} [\text{vec}(\nabla_V f(U^k, V))]_{F^k}. \quad (15) \quad \square$$

3.1.2. Search Direction of the Active Variables. The active variables are updated in a simple way:

$$d_i^k = \begin{cases} -\text{vec}(\nabla_V f(U^k, V))_i, & \text{if } \text{vec}(V^k)_i - \text{vec}(\nabla_V f(U^k, V))_i \geq 0, \\ -\text{vec}(V^k)_i, & \text{if } \text{vec}(V^k)_i - \text{vec}(\nabla_V f(U^k, V))_i \leq 0. \end{cases} \quad (16)$$

The above definition of the search direction will help d^k satisfy the descent condition (for further details, see [41–43]), and with d^k , the next iteration is guaranteed to be a feasible point. After the search direction $d^k = (d_{I^k}^k, d_{F^k}^k)$ is computed, a step size is chosen with the Armijo rule. At the k -th iteration, we propose Algorithm 1 to solve the ANLS problem.

Theorem 1. *Suppose that the assumption and strict complementary slackness assumption are satisfied, the sequence $\{V^k\}$ generated by Algorithm 1 converges to the optimal solution V^* and the convergence rate of $\{V^k - V_F^*\}$ is at least quadratic.*

Proof. Because the Hessian matrix in (10) is positive definite, $f(U, V)$ is strictly convex and the optimization problem in (10) has a unique solution V^* . \square

With (14) and strict complementary slackness assumption, we obtain from Proposition 4 in [47] that sub-algorithm 3.1 converges to the optimal solution V^* and the rate of $\{V^k - V_F^*\}$ is at least quadratic.

Now, we can present the new algorithm for NMF (Algorithm 2).

Theorem 2. *Any limit point in the sequence $\{U^k, V^k\}$ generated by the AS is a stationary point of (2).*

The above convergence result is directly from the corollary in Reference [48].

4. Numerical Tests

In this section, the proposed algorithm with active set identification technique AS1 or AS2 is marked as AS1 or AS2. We compare AS1 and AS2 with the main stream approaches to solve (2), that is, the multiplicative updating (MU) method and projected gradient (PG) method. We add the experimental results of the projected Newton (PN) method. The PN method uses the projected Newton direction to update the factor matrices in ANLS, while AS1 and AS2 use projected Newton direction to update the inactive variables, and active variables are updated with simple rule based on the information of the gradient. PG and PN and AS1 and AS2 have the similar rules in the stopping criterion and step size updating. Because algorithms are sensitive to the initial point when solving NMF problems, five algorithms start with the same initial random generated matrices U^0 and V^0 . All these five algorithms use the same parameters related to the stopping criterion, that is,

(i) tol: tolerance for a relative stopping criterion, $\text{tol} = 1e - 16$

(ii) timelimit: limit of time, $\text{timelimit} = 5$

(iii) maxiter: limit of the iterations, $\text{maxiter} = 1e6$

All algorithms are implemented in Matlab; experimental results on synthetic and real-world data are, respectively, presented in Sections 4.1 and 4.2.

The stopping criterion of the proposed algorithms is similar to that adopted by Lin:

$$\begin{aligned} \|\nabla_U^P f(U^{k+1}, V^k)\|_F &\leq \text{tol}, \\ \|\nabla_V^P f(U^{k+1}, V^{k+1})\|_F &\leq \text{tol}, \end{aligned} \quad (17)$$

where the projected gradient $\nabla^P f(x)$ is defined as follows:

$$\nabla^P f(x)_i = \begin{cases} \nabla f(x)_i, & \text{if } x_i > 0, \\ \min(0, \nabla f(x)_i), & \text{if } x_i = 0. \end{cases} \quad (18)$$

The stopping tolerance tol is updated in a similar way as shown in Reference [4].

Input: $W \in R_+^{m \times n}$, r and $U^k \in R_+^{m \times r}$.

Output: V^{k+1} such that $V^{k+1} = \arg \min_{V \geq 0} (1/2) \|W - U^k V\|_F^2$.

Step 1 (computation of the search direction)

Compute the estimated active set I^k with one of the definitions (AS1) or (AS2).

Compute the search direction of the active and inactive variables with (15) and (16).

Step 2 (update the factor matrix)

Let $\text{vec}(V^{k+1}) = P_+ [\text{vec}(V^k) + \alpha^k d^k]$, where the step size is $\alpha^k = (1/2)^{t_k}$ and t_k is the first nonnegative integer that satisfies the following inequality: $f(U^k, V^{k+1}) - f(U^k, V^k) \leq \sigma \alpha^k \sum \nabla_V f(U^k, V)_i \text{vec}(d^k)_i$.

ALGORITHM 1: Newton method with an active set strategy (NMAS).

Input: $W \in R_+^{n \times m}$, r , $U^0 \in R_+^{n \times r}$, $k := 0$, $\sigma = \varepsilon = 10^{-4}$.

Output: The approximation of the two factor matrices U and V .

Step 1 (update the factor matrices alternatively)

If the convergence criterion is satisfied, then stop; else,

update $V^{k+1} = \arg \min_{V \geq 0} f(U^k, V) = (1/2) \|W - U^k V\|_F^2$ with Algorithm 1 and,

update $U^{k+1} = \arg \min_{U \geq 0} f(U, V^{k+1}) = (1/2) \|W - UV^{k+1}\|_F^2$ with Algorithm 1.

Step 2

$k := k + 1$, go to Step 1.

ALGORITHM 2: Active set type algorithm for NMF (AS).

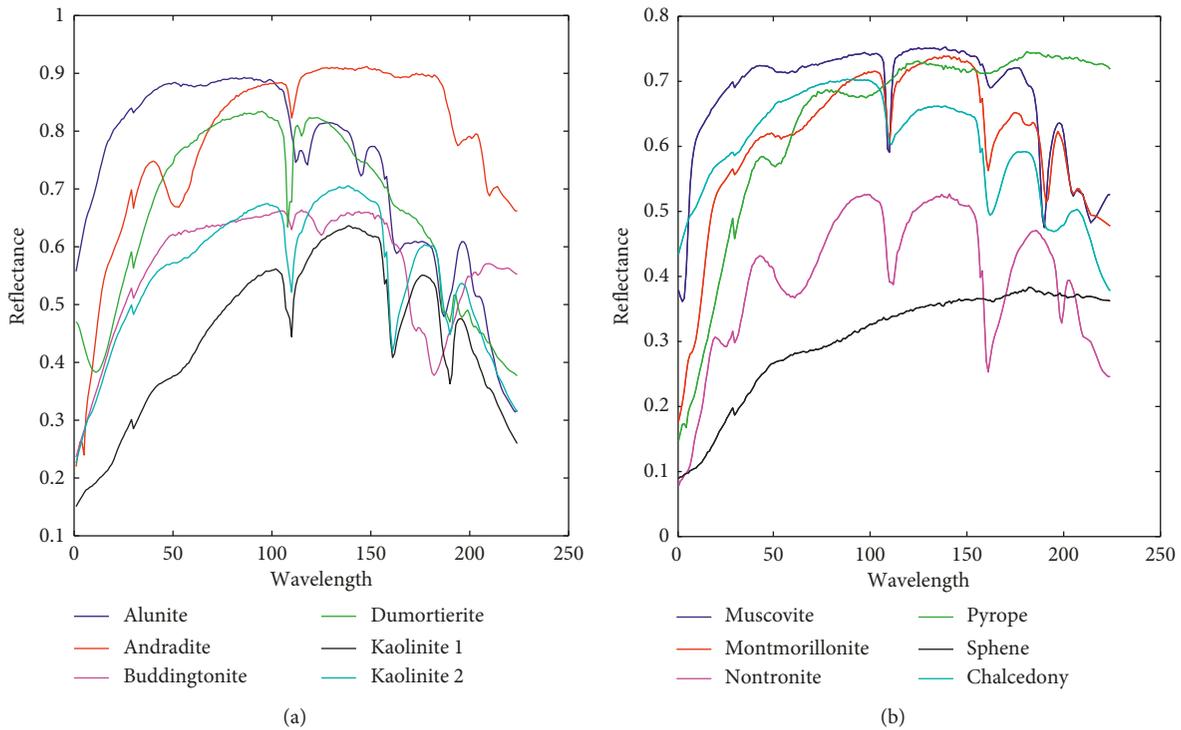


FIGURE 1: Selected spectral signatures from the USGS.

4.1. *Simulation Test 1.* We start with generating the original matrix W . The spectral signatures are chosen from the United States Geological Survey (USGS) digital spectral library. Figure 1 shows twelve of these endmember signatures.

The spectral signatures contain 224 spectral reflectance bands in the wavelength range of $0.38 - 2.5 \mu\text{m}$. The

endmember matrix U has twelve columns, representing the twelve spectral signatures, and the abundance matrix V is obtained from Dirichlet distribution of parameter $(0.1, \dots, 0.1)$, which satisfies the nonnegative constraints. The sum of coordinates is 1. Let the original matrix $W = UV$. Figure 2 displays the square root of the final objective function

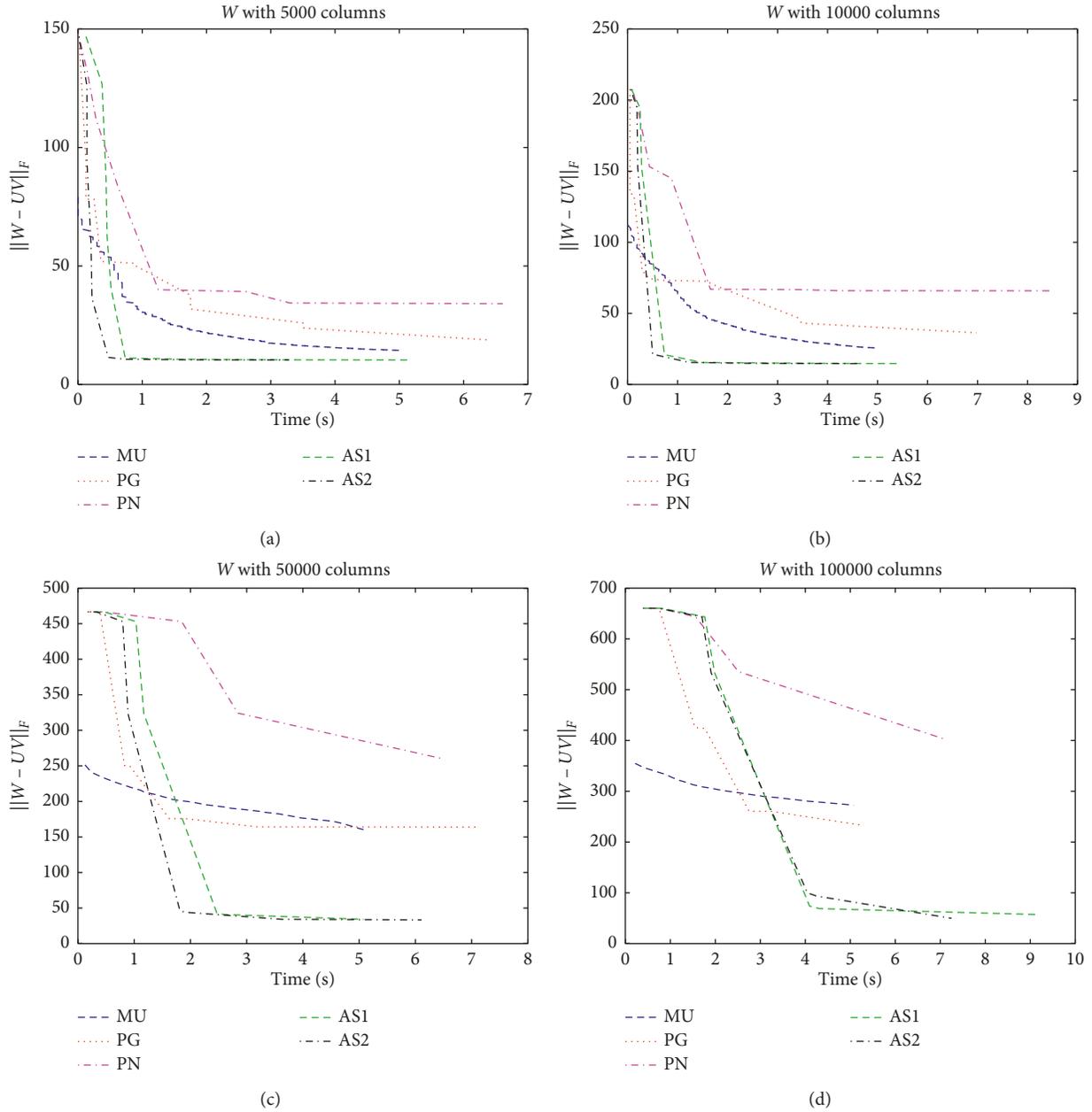


FIGURE 2: Synthetic data results.

value $\|W - UV\|_F$ depending on the time. We set the number of columns of W to 5000, 10000, 50000, and 100000, and for each W , we plot the objective value vs. time curve.

We perform a quantitative analysis with added noise, that is, $W = \max(0, UV + N)$, where $N = \xi \cdot \text{randn}(m, n)$ in Matlab notation (Gaussian distribution of mean zero and standard deviation). Figure 3 shows the square root of the final objective value vs. standard deviation ξ .

Figures 2 and 3 show that the active set type methods perform well. Specifically, the function value curves slowly decrease, and the results of the AS method rapidly decrease, which indicates that if the dataset is well structured, AS should be able to find the solution much more easily than the other methods. Generally speaking, PN should perform better than PG because the second-order information is

employed with the Newton step. As far as we are concerned, there are two reasons for the failure of PN: the computation of the inverse matrix is time consuming and the sparsity causes the singularity of the Hessian matrix, especially in the end of the iteration. After the active set strategies are applied in the projected Newton method, the algorithms become more stable and time saving. Because the searching of the step size α_k is a time-consuming operation, the curves of PG, PN, AS1, and AS2 are polygonal lines.

4.2. Tests of Hyperspectral Unmixing. Finally, we evaluate the effectiveness of the active set type method with the real-world data. Urban is one of the most widely used

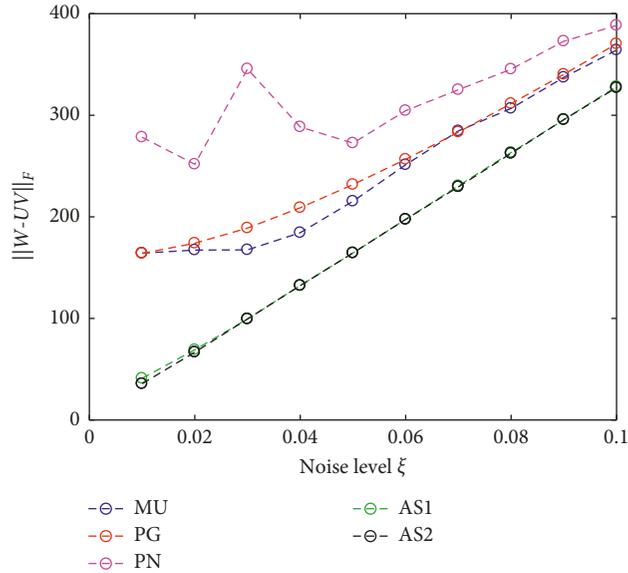


FIGURE 3: $\|W - UV\|$ depending on the noise $N = \xi \cdot \text{randn}(m, n)$.

TABLE 1: SAD comparison for the ground truth and estimated endmember signature.

Endmember	MU	PG	PN	AS1	AS2
Asphalt road	0.4102	0.4330	0.1205	0.2383	0.2099
Grass	0.3093	0.4083	0.2068	0.2719	0.2557
Tree	0.1481	0.3621	0.2849	0.2159	0.2195
Roof	0.2832	0.3063	0.5307	0.3137	0.3412

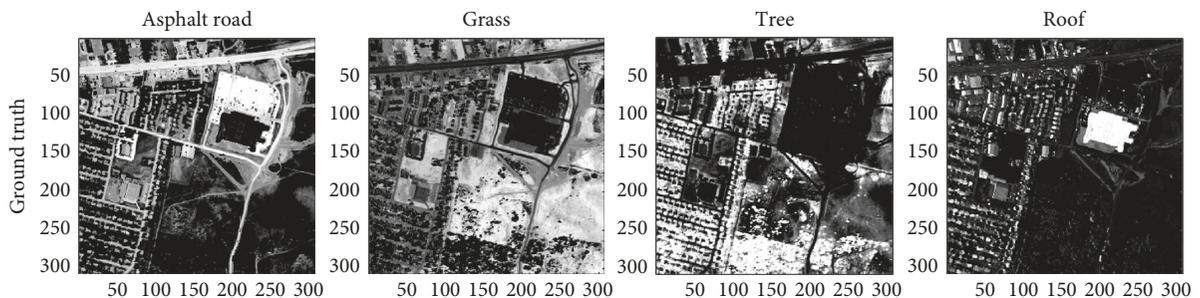


FIGURE 4: Ground truth of Urban with 4 endmembers.

hyperspectral data used in the hyperspectral unmixing study. There are 307×307 pixels, each of which corresponds to a $2 \times 2 \text{ m}^2$ area. In this image, there are 210 wavelengths ranging from $0.4 \sim 2.5 \mu\text{m}$. Due to dense water vapor and atmospheric effects, we remove several channels and keep 162 channels [49–51]. There are four endmembers in these data: asphalt road, grass, tree, and roof. Set the number of endmembers $r = 4$; we factorize the original matrix with MU, PG, PN, AS1, and AS2. In the numerical test, we set $\text{timelimit} = 15$. All of these five algorithms successfully separate the endmember matrix and the abundance matrix. We compare five methods in the spectral angle distance (SAD) and the abundance maps of the 4 endmembers. The SAD is used to compare the similarity of the endmembers. Let V_a, V_b be the spectral signature of endmember a and b . Then,

$$\text{SAD}(V_a, V_b) = \arccos \frac{(V_a, V_b)}{\|V_a\| \cdot \|V_b\|}. \quad (19)$$

The smaller the SAD, the more similarities in endmembers a and b . We compute the SAD between our unmixed endmember signatures and the ground truth of these four endmembers, whose spectral signatures are extracted from the USGS library.

Table 1 shows that our proposed active set type algorithms are consistently better or comparable to all the other methods. Figure 4 shows the ground truth for the abundance fractions of these endmembers, that is, roof, asphalt, grass, and tree in order from left to right.

In order to give an intuitive comparison of the HU results, we illustrate the abundance maps on Urban data in Figure 5, which is shown in gray scale.

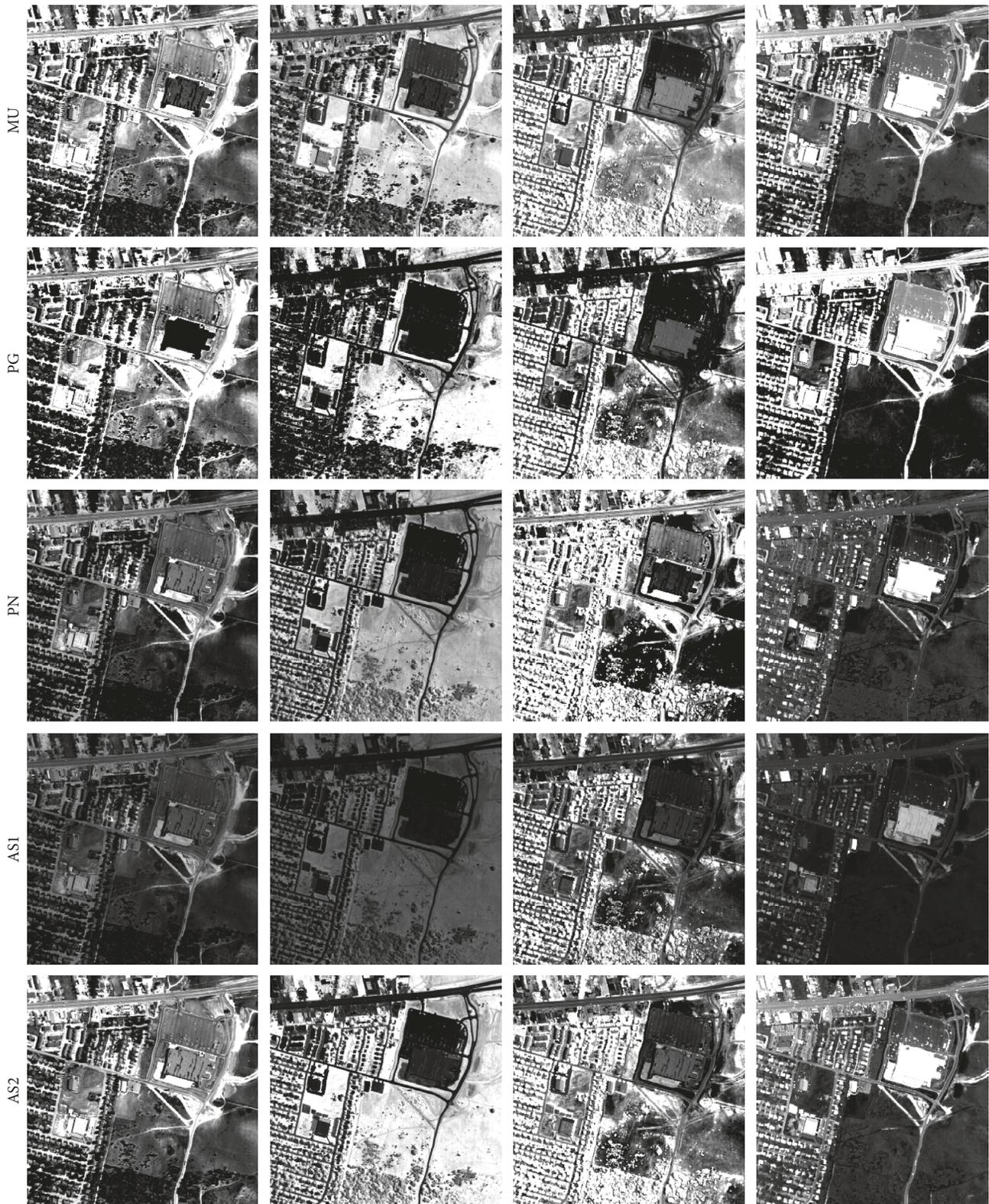


FIGURE 5: Unmixing with five algorithms.

Because the second-order information of the objective function is used in the proposed active set type Newton algorithms, the run time is reduced. Based on the help of the active set strategy, AS1 and AS2 are more stable than PG and

PN. When dealing with hyperspectral data, we found that PG and PN are sensitive to the initial matrices U^0 and V^0 and the values of the parameters in the algorithm. Additionally, PG might stop because the stopping condition of the ANLS

subproblem is satisfied, which indicates that the descent condition does not meet under the maximum inner iteration constraint of the inner loop.

4.3. Summary and Discussion. We present two active set type Newton algorithms for solving NMF problems based on an alternating nonnegative constrained least squares approach. Based on an active strategy, the value of the objective function decreases at the beginning of the iterations, which makes the algorithms more robust in various applications. Finally, we make some remarks about the future work. First, it is said that a good initialization matrix can improve the speed and accuracy of the solutions as it produces faster convergence to the improved local minima. A good clustering method which helps the estimated endmember signatures or given the endmember matrix based on the prior ground truth data is worthy of studying. Second, efficient constraints should be added to the NMF model; these help to improve the accuracy of HU. The endmember matrix and abundance matrix both have their own characteristics, such as sparsity of V and column linear independent of U . Third, an optimal solution for the ANLS subproblem corresponding to one factor is not required before the update of the other factor is performed. Generally speaking, the algebra-relevant algorithms, such as MU and ALS, perform simple, while it is at higher computational cost with the algorithms based on the optimization scheme to find an accurate solution of ANLS subproblem. This indicates that the inexact Newton or limited memory quasi-Newton method with certain active set strategy might work.

Data Availability

The data used to support this study are available from <http://lesun.weebly.com/hyperspectral-data-set.html>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Science Foundation of China (11701337, 41271235, and 11301307), Shandong Provincial (China) Natural Science Foundation (ZR2016DM03), and the Project of Shandong Province Higher Educational Science and Technology Program (J16LI16).

References

- [1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon et al., "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [2] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.
- [3] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, pp. 535–541, Cambridge, MA, USA, 2000.
- [4] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007.
- [5] R. Zdunek and A. Cichocki, "Fast nonnegative matrix factorization algorithms using projected gradient approaches for large-scale problems," *Computational Intelligence and Neuroscience*, vol. 2008, Article ID 939567, 13 pages, 2008.
- [6] N. Guan, D. Tao, Z. Luo, and B. Yuan, "NeNMF: an optimal gradient method for nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, 2012.
- [7] X. Li, W. Zhang, and X. Dong, "A class of modified FR conjugate gradient method and applications to non-negative matrix factorization," *Computers & Mathematics with Applications*, vol. 73, no. 2, pp. 270–276, 2017.
- [8] Y. Huang, H. Liu, and S. Zhou, "An efficient monotone projected Barzilai-Borwein method for nonnegative matrix factorization," *Applied Mathematics Letters*, vol. 45, pp. 12–17, 2015.
- [9] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization," *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012.
- [10] P. O. Hoyer, "Non-negative sparse coding," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 557–565, Martigny, Switzerland, September 2002.
- [11] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [12] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 161–173, 2009.
- [13] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4267–4279, 2016.
- [14] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3909–3921, 2017.
- [15] S. Khoshshokhan, R. Rajabi, and H. Zayyani, "Sparsity-constrained distributed unmixing of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1279–1288, 2019.
- [16] S. Khoshshokhan, R. Rajabi, and H. Zayyani, "Distributed unmixing of hyperspectral data with sparsity constraint," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017.
- [17] D. Cai, X. He, X. Wu, and J. Han, "Non-negative Matrix Factorization on Manifold," in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pp. 63–72, Pisa, Italy, December 2008.
- [18] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.

- [19] R. Rajabi and H. Ghassemian, "Sparsity constrained graph regularized NMF for spectral unmixing of hyperspectral data," *Journal of the Indian Society of Remote Sensing*, vol. 43, no. 2, pp. 269–278, 2015.
- [20] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal non-negative matrix T-factorizations for clustering," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 126–135, Philadelphia, PA, USA, August 2006.
- [21] R. Schachtner, G. Pöppel, A. M. Tomé, and E. W. Lang, "Minimum determinant constraint for non-negative matrix factorization," in *Independent Component Analysis and Signal Separation*, pp. 106–113, Springer, Berlin, Germany, 2009.
- [22] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 38–42, 2015.
- [23] X. Li, X. Shen, Z. Shu, Q. Ye, and C. Zhao, "Graph regularized multilayer concept factorization for data representation," *Neurocomputing*, vol. 238, pp. 139–151, 2017.
- [24] X.-R. Feng, H.-C. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 6245–6257, 2018.
- [25] S. Khoshsokhan, R. Rajabi, and H. Zayyani, "Clustered multitask non-negative matrix factorization for spectral unmixing of hyperspectral data," *Journal of Applied Remote Sensing*, vol. 13, no. 2, Article ID 026509, 2019.
- [26] C. Lin and M. Pang, "Graph regularized nonnegative matrix factorization with sparse coding," *Mathematical Problems in Engineering*, vol. 2015, Article ID 239589, 11 pages, 2015.
- [27] X. G. Dai, C. D. Li, and B. Q. Xiang, "Graph sparse non-negative matrix factorization algorithm based on the inertial projection neural network," *Complexity*, vol. 2018, Article ID 2743678, 12 pages, 2018.
- [28] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: the no-pure-pixel case," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 10, pp. 5530–5546, 2015.
- [29] M. Arngren, M. N. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using bayesian non-negative matrix factorization with volume prior," *Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 479–496, 2011.
- [30] Z. Z. Yu, Y. H. Liu, B. Li, S. C. Pang, and C. C. Jia, "Incremental graph regulated nonnegative matrix factorization for face recognition," *Journal of Applied Mathematics*, vol. 2014, Article ID 928051, 10 pages, 2014.
- [31] 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*, vol. 45, no. 3, pp. 765–777, 2007.
- [32] N. Akhtar and A. Mian, "RCMF: robust constrained matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3354–3366, 2017.
- [33] F. Pompili, N. Gillis, P.-A. Absil, and F. Glineur, "Two algorithms for orthogonal nonnegative matrix factorization with application to clustering," *Neurocomputing*, vol. 141, pp. 15–25, 2014.
- [34] Z. Li, X. Wu, and H. Peng, "Nonnegative matrix factorization on orthogonal subspace," *Pattern Recognition Letters*, vol. 31, no. 9, pp. 905–911, 2010.
- [35] R. Zdunek and A. Cichocki, "Nonnegative matrix factorization with quadratic programming," *Neurocomputing*, vol. 71, no. 10–12, pp. 2309–2320, 2008.
- [36] S. Z. Lai, H. B. Li, and Z. T. Zhang, "A symmetric rank-one quasi-Newton method for nonnegative matrix factorization," *ISRN Applied Mathematics*, vol. 2014, Article ID 846483, 11 pages, 2014.
- [37] P. Gong and C. Zhang, "Efficient nonnegative matrix factorization via projected Newton method," *Pattern Recognition*, vol. 45, no. 9, pp. 3557–3565, 2012.
- [38] R. Zdunek, "Regularized nonnegative matrix factorization: geometrical interpretation and application to spectral unmixing," *International Journal of Applied Mathematics and Computer Science*, vol. 24, no. 2, pp. 233–247, 2014.
- [39] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 2, pp. 713–730, 2008.
- [40] C. Zhang, L. Jing, and N. Xiu, "A new active set method for nonnegative matrix factorization," *SIAM Journal on Scientific Computing*, vol. 36, no. 6, pp. A2633–A2653, 2014.
- [41] L. Sun, G. P. He, Y. L. Wang, and C. Y. Zhou, "An accurate active set newton method for large scale bound constrained optimization," *Applications of Mathematics*, vol. 56, no. 3, pp. 297–314, 2011.
- [42] L. Sun, G. He, Y. Wang, and L. Fang, "An active set quasi-Newton method with projected search for bound constrained minimization," *Computers & Mathematics with Applications*, vol. 58, no. 1, pp. 161–170, 2009.
- [43] L. Sun, L. Fang, and G. P. He, "An active set strategy based on the multiplier function or the gradient," *Applications of Mathematics*, vol. 50, no. 4, pp. 291–304, 2010.
- [44] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 29–47, 2006.
- [45] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics*, vol. 11, no. 5, pp. 393–401, 1997.
- [46] M. H. V. Benthem and M. R. Keenan, "Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems," *Journal of Chemometrics*, vol. 18, no. 10, pp. 441–450, 2004.
- [47] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM Journal on Control and Optimization*, vol. 20, no. 2, pp. 221–246, 1982.
- [48] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Operations Research Letters*, vol. 26, no. 3, pp. 127–136, 2000.
- [49] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 101–118, Feb. 2014.
- [50] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5412–5427, 2014.
- [51] F. Zhu, Y. Wang, B. Fan, G. Meng, and C. Pan, "Effective spectral unmixing via robust representation and learning-based sparsity," 2014, <http://arxiv.org/abs/1409.0685v3>.

