Towards Convergence: A Gradient-Based Multiobjective Method With Greedy Hash for Hyperspectral Unmixing

Ruiying Li, Bin Pan, Xia Xu, Tao Li and Zhenwei Shi

Abstract

Multiobjective optimization aims at addressing the conflicting objectives, which has been introduced to improve the performance of sparse hyperspectral unmixing. Recently proposed multiobjective unmixing methods usually employ evolutionary algorithms to improve the unmixing accuracy. However, evolutionary algorithms may suffer the challenge of convergence, in which case the reasonability of the solutions is hard to guarantee. To solve the problem of convergence, in this paper, we present a new gradient-based multiobjective unmixing method, which explores the optimization direction in a theoretically reliable manner. Furthermore, considering the mathematical model of hyperspectral sparse unmixing where sparsity error objective of selected endmembers is discrete, we develop a greedy hash based coding approach which is able to well describe the discrete constraints imposed on endmembers. The major components of the proposed method are a search approach and an update approach. In the search approach, we construct the pareto descent direction via a gradient-based strategy, which contributes to converging to an optimal continuous solution by searching along this direction. In the update approach, we update discrete binary endmember via hash coding under the guidance of greedy principle, which allows our method to handle the problem of discrete objective. The major contribution of the proposed method is designing a new framework that can get the optimal discrete endmembers in a convergent way. Moreover, we provide the theoretical analysis and proof for the convergence. Synthetic and real-world experiments have indicated the advantages of our algorithm when compared with evolutionary multiobjective unmixing methods.

Index Terms

Hyperspectral unmixing, multi-objective optimization, convergence, greedy hash

The work was supported by the National Key Research and Development Program of China under Grant 2022ZD0160401, the National Natural Science Foundation of China under the Grant 62001251, 62001252 and 62272248, and the Beijing-Tianjin-Hebei Basic Research Cooperation Project under the Grant F2021203109. (Corresponding author: Bin Pan)

Ruiying Li and Bin Pan (corresponding author) are with the School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University, Tianjin 300071, China. (e-mail: liruiying@mail.nankai.edu.cn; panbin@nankai.edu.cn).

Xia Xu and Tao Li are with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: xuxia@nankai.edu.cn; litao@nankai.edu.cn).

Zhenwei Shi is with the image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

I. INTRODUCTION

Hyperspectral remote sensing image carries consecutive spectral information of the land-covers and can provide abundant spectral information in many applications [1], [2]. However, the problem of low resolution in hyperspectral images (HSI) is inevitably as these images are generally acquired by satellites from high altitude, which results in mixing pixels [3], [4]. Mixing pixels can be separated to explore the pure materials spectra by hyperspectral unmixing, which is currently related to the different type of model assumptions [5]. The mixing model assumptions can be described as a linear mixing model (LMM) or a nonlinear mixing model [6]–[8]. LMM is easily implemented as it supposes that the collected spectral is consist of several pure spectra (endmembers) weighted by their corresponding fractions (abundances) in the form of linear way.

Linear mixing model, because of its simplicity and flexibility, has promoted a large amount of researches. These researches can be roughly divided into five kinds of unmixing methods [5], [9]. Geometrical approaches usually extract pure spectra by maximizing the simplex of the data [10]–[14]. The statistical unmixing method relies on a probability distribution [15]–[17]. Nonnegative matrix decomposition describes observed image data to endmembers and abundance by factorizing HSI into two low-rank nonnegative matrices [18]–[22]. A large number of deep learning unmixing methods were based on autoencoder [23]–[27]. Sparse regression-based unmixing utilizes the set of prior spectral signatures, which named as spectral library A. This known spectral library is assumed to contain all materials in observed HSI whose number of endmembers is usually very small compared to the dimensionality of A [28], [29].

Due to the independence to pure pixels, sparse unmixing has obtained great interest among researchers since proposed. One of the mainly used approaches is aimed at solving l_0-norm minimization problem [30]–[33]. The spectral *a priori* information was adopted in [34] to improve unmixing accuracy. In order to get better unmixing performance, spectral variability, total variation regularization, nonlocal spatial information, and collaborative sparsity among neighboring pixels were used to exploit the information presented in HSI [**?**], [35]–[41].

However, the contradiction of objective functions makes the adjustment of regularization parameters a difficulty in sparse unmixing. To address this open issue, multiobjective optimization methods have been designed for hyperspectral unmixing [42]–[44]. We regard hyperspectral unmixing problem as a multiobjective optimization problem (MOP) because of its two objectives, namely reconstruction error and sparsity error. In this case, we can minimize conflicting objectives simultaneously without setting the weight coefficient [45]–[48]. Several multiobjective unmixing methods have been designed under the basis of traditional evolutionary algorithms (EA) to extract endmembers [49]–[53]. In spite of their advantages, EA applied in multiobjective unmixing are unable to get the convergence.

Recently, researchers have attempted to tackle MOP in a convergent way by adopting multiobjective gradient descent algorithms [54]–[57], which became the motivation of our work. Multiobjective gradient descent algorithms can be regarded as natural extension of single objective gradient descent, which construct the optimization direction to search optimal solution and get convergence [56]. Although these algorithms can handle MOP effectively, it is not feasible to directly adopt them in hyperspectral unmixing as there is a general problem: Endmembers are expected to be binary so that they can corresponding to an encoded spectral library. However, existing gradient-based

multiobjective methods for MOP can not directly address discrete constraints imposed on decision vectors.

In this paper, we propose a gradient-based multiobjective unmixing method, which simultaneously handles the problem of convergence and the discrete constraints imposed on endmembers. The proposed method that abbreviated as Gradient-based Multi-Objective method with Greedy Hash (GMOGH) mainly includes a search approach and an update approach. The gradient-based search approach contributes to converging to an optimal continuous solution by searching along a pareto descent direction, which is constructed via a gradient-based strategy with multiobjective gradient functions. In order to handle the problem of discrete objective and update discrete binary endmember, the hash-based update approach develops a hash coding strategy under the guidance of greedy principle. Moreover, we provide the theoretical analysis to prove the convergence of GMOGH. Overall, the major contributions of GMOGH can be summarized as follows:

- We develop a gradient-based search approach, which could construct a pareto descent direction to solve the problem of convergence.
- We provide a hash-based update approach to describe the discrete constraints improsed on endmembers by greedy hash coding.
- We present theoretical proof to support the convergence of GMOGH.

The remainder of this paper is structured as follows. Section II expresses the background containing LMM, sparse unmixing and mathematical description of multiobjective optimization. Section III presents the proposed algorithm and theory. Section IV describes the experimental results on both synthetic data sets and real data sets. The code is available online[1].

II. Background

In this section, we briefly introduce LMM, sparse unmixing and mathematical description of multiobjective optimization.

A. Linear Mixing Model

Let $y_i \in R^{L\times 1}$ indicates the spectrum of ith mixed pixel in observed HSI, which can be estimated by a combination of endmembers [5]. This combination is expressed in a linear way:

$$y_i = Mx_i + w_i \tag{1}$$

where L denotes the number of bands, $M = \{m_1, ..., m_N\} \in R^{L\times N}$ denotes the mixed matrix, m_j denotes the jth endmember and N denotes total number of endmembers. $x_i \in R^{N\times 1}$ denotes abundance and w_i denotes noise or error. If the observed HSI includes B mixed pixels, then it can be described as:

$$Y = MX + W \tag{2}$$

where $Y = \{y_1, y_2, ..., y_B\}$ indicates the observed spectrum, $X = \{x_1, x_2, ..., x_B\}$ indicates abundance matrix, M indicates endmember matrix, $W = \{w_1, w_2, ..., w_B\}$ indicates noise. Considering the abundance fractions

[1]Code is released at https://levir.buaa.edu.cn/Publication.htm

geometrically, abundance nonnegativity constraint (ANC) and abundance sum-to-one constraint (ASC) should be applied.

$$ANC : X \geq 0 \tag{3a}$$

$$ASC : 1_N^T A = 1_B^T \tag{3b}$$

It should be noted that [31] claims ASC is a strong constraint, and an automatic imposition to a generalized ASC is given by ANC. Then we only consider ANC in this paper.

B. Sparse Unmixing

Sparse unmixing is an unmixing method based on LMM, which assumes that a known spectral library is obtained from prior information and many spectral samples are contained in it. Let $A \in R^{N \times m}$ denote this library, which means that there are m spectral samples and N spectral bands in A. Under this assumption, a portion of these known spectral samples can be combined into observed HSI in a linear manner. The selected spectral samples is a very small fraction of endmembers present in A, which is sparse. Sparse representation can be factored into the unmixing problem, whose goal then goes from selecting real endmembers to finding the optimal known samples in spectral library A. Let $X \in R^m$ indicates fractional abundance, then hyperspectral unmixing is rewritten as

$$\begin{aligned} &\min_X \|X\|_0 \\ &s.t. \|Y - AX\|_F \leq \delta \ and \ X \geq 0 \end{aligned} \tag{4}$$

Since problem (4) is NP-hard, it is often transformed into a l_1-norm relaxation problem with regularization

$$\begin{aligned} &\min_X \frac{1}{2} \|Y - AX\|_F + \lambda \|X\|_1 \\ &s.t. X \geq 0 \end{aligned} \tag{5}$$

In this case, the results of sparse unmixing largely depend on the regular term coefficient λ, which needs to be set manually. Therefore, the selection of weight coefficient becomes a major difficulty in sparse unmixing.

C. Mathematical Description of Multiobjective Optimization

In MOP, we are interested in minimizing many conflicting ovjectives simultaneously. Assuming there exists N objectives need to be optimized, MOP can be expressed as:

$$\begin{aligned} &\min F(x) = [f_1(x), f_2(x), ..., f_N(x)]^T \\ &X \in \Omega \end{aligned} \tag{6}$$

where x is the decision variable and Ω is the decision space [51]. It is hard to improve all the objective functions simultaneously in MOP since objectives are conflicting, which means that a decrement in one function must be accompanied by an increment of other functions. Thus MOP is aimed at finding a Pareto optimal which brings all objectives a balance. The relevant definitions are as follows.

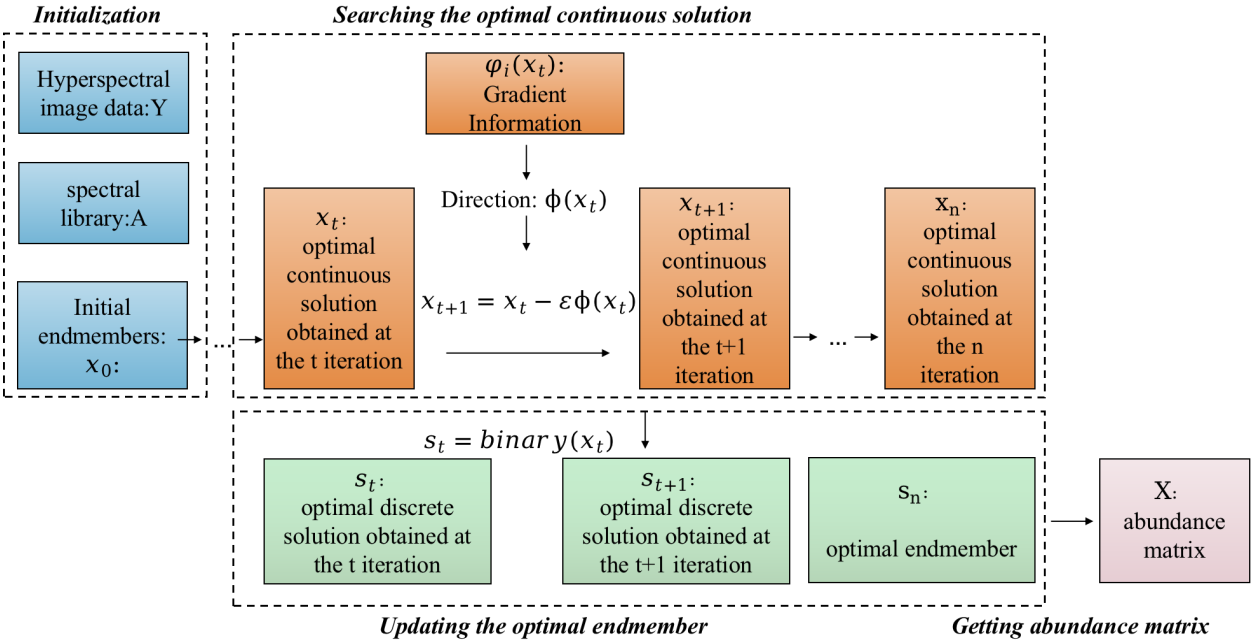


Fig. 1. The flowchart of GMOGH: a simplified illustration for the optimization process works on elements selection.

Definition 1 (Pareto dominate): α and β are termed as decision vectors in a minimum optimization problem, α is said to dominate β, noted as $\alpha \prec \beta$ if and only if

$$\begin{aligned} &\forall i \in \{1, 2, ..., N\} : f_i(\alpha) \leq f_i(\beta) \\ &\wedge \exists j \in \{1, 2, ..., N\} : f_j(\alpha) < f_j(\beta) \end{aligned} \tag{7}$$

Definition 2 (Pareto optimal): Decision vector α is called Pareto optimal if and only if:

$$\begin{aligned} &\forall \alpha^{'} \in \Omega : \alpha \prec \alpha^{'} \\ &\text{regarding the whole parameter space } \Omega \end{aligned} \tag{8}$$

Definition 3 (Pareto set): The collection of Pareto optimal points are called the Pareto set:

$$\left\{ \alpha^* | \forall \alpha^{'} \in \Omega : \alpha^* \prec \alpha^{'} \right\} \tag{9}$$

Definition 4 (Pareto front): The collection of function values $F(\alpha^*)$ of all the Pareto points α^* is called the Pareto front:

$$\{F(\alpha^*) | \alpha^* \in Pareto\ Set\} \tag{10}$$

Definition 5 (Descent Direction): $d = (d_1, d_2, ..., d_m)$ is called a Descent Direction in MOP if d satisfies:

$$(- \bigtriangledown f_i(x))^T d \geq 0, i = (1, ..., N) \tag{11}$$

III. METHODOLOGY

In this section, a gradient-based multiobjective model is proposed for hyperspectral unmixing problem (5). However, we first regard sparse unmixing problem (5) as two conflicting objectives: $\|Y - AX\|_F$ measures the error between real image data and reconstructed image data and $\lambda \|X\|_1$ is the endmember sparsity. Then the sparse unmixing problem (5) with two objectives can be expressed as a multiobjective optimization problem:

$$\begin{aligned} &\min_s F(s) = [f_1(s), f_2(s)]^T \\ &f_1(s) = \|Y - A_s X_s\|_F \\ &f_2(s) = |\,\|s\|_1 - k| \\ &s.t.\ s \in \{0,1\}^m \end{aligned} \tag{12}$$

It should be noted that problem (12) is not a simple transformation for problem (5). We consider unmixing problem as a binary vector optimization problem by encoding the whole spectral library [52]. $s = [s_1, s_2, ..., s_m] \in \{0,1\}^m$ is a binary representation of known spectral library A, "$s_i = 1$" indicates that the corresponding spectrum in A is regarded as one of endmembers exist in image data Y, and "$s_i = 0$" otherwise. $A_s = A \times diag(s)$ is the subset of spectral library corresponding to s, X_s is the abundance matrix, $\|s\|_1$ is the number of binary vectors with the value $s_i = 1$, and can also be considered as the number of endmembers estimated by this algorithm. When k is unknown, it can be estimated by the HySime [58] endmember estimation algorithm. Eq.(12) is an endmember selection problem which takes discrete binary s as decision vector and optimizes s to get optimal endmember. In order to address unmixing problem completely, we calculate the corresponding abundances via nonnegative least squares.

A. Framework of GMOGH

The flowchart of GMOGH is shown in Fig.1. We solve problem (12) by using GMOGH, which mainly plays a role in endmember selection. The framework of GMOGH includes four parts: initialization, searching the optimal continuous solution, updating the optimal endmembers and getting abundance matrix.

1) Initialization: GMOGH begins with an initial continuous solution set $X = (x^1, ..., x^p)$, its size is p. Individual $x^i = (x_1^i, ..., x_m^i)$ corresponds to a continuous solution during the optimization process. In GMOGH, we set x_i^j by randomly generating a number between 0 and 1 for initialization.

2) Searching the optimal continuous solution: The gradient-based search approach in GMOGH is developed to search optimal continuous solution without considering about discrete binary constraint. In the searching process, gradient information is utilized to construct Pareto descent direction (PDD). Along PDD, by setting a small step size, we can search optimal continuous solution x until convergence.

3) Updating the optimal endmember: The hash-based update approach in GMOGH is developed to update endmembers. When the optimal continuous solution x is acquired in search approach, we hope to get the optimal discrete option s in each iteration under the guidance of greedy principle. Moreover, s is approximated as the closest discrete point to x, which can be calculated by a hash coding strategy in this approach.

4) Getting abundance matrix: GMOGH mainly works on endmember selection to get s. When s is determined, the corresponding unique certainty of the endmember is considered, and the abundance matrix can be solved by using nonnegative least squares method.

B. The Gradient-based Search Approach in GMOGH

Motivation of the gradient-based search approach in GMOGH is to iteratively search the optimal continuous solution along a PDD which ensures that all the objectives can be improved simultaneously. In this section, without considering discrete constraints, we search continuous solutions with equation

$$x' \leftarrow x + \varepsilon\phi(x) \tag{13}$$

where $\phi(x)$ is the search direction and ε is a small step size. Assume that ε is a constant, in this case, the core operation of search approach becomes the determining of $\phi(x)$ which is constructed by using gradient information.

We now introduce the details of ensuring $\phi(x)$. In multiobjective gradient descent algorithms, gradient of the ith objective function can be denoted as $\bigtriangledown f_i(x) \doteq \varphi_i(x)$. We consider a vector set $\{\mu_i^*(x)\}_{i=1}^m$, which is the solution of

$$\min_{\mu_i} \left\| \sum_{i=1}^{m} \mu_i \varphi_i(x) \right\| \ \ s.t.\ \forall i, \sum_{i=1}^{m} \mu_i = 1, \mu_i \geq 0 \tag{14}$$

Suppose there is a direction $\varphi^*(x)$ satisfies

$$\varphi^*(x) \propto -\sum_{i=1}^{m} \mu_i^*(x)\varphi_i(x) \tag{15}$$

Under this assumption, $\varphi^*(x)$ is a PDD, which is proved in *Theorem 1*. Multiobjective gradient descent algorithms can monotonically decrease all the objectives simultaneously with $\varphi^*(x)$ and a fixed small constant ε. However, concerning the problem about local optima and the diversity between particles, $\phi(x)$ is computed by an enhanced PDD based on $\varphi^*(x)$ in this paper.

We take $\mathcal{F}$ to be the unit ball of a reproducing kernel Hilbert space with a positive definite kernel $k(x, x')$, for example, we can use truncated (σ, m_d) Gauss Kernel as

$$k(x, y) = (det(2\pi\sigma^2 I))^{-1/2} exp\left\{ -\frac{\|x-y\|^2}{2\sigma^2} \right\} 1_{\|x-y\| \leq m_d\sigma} \tag{16}$$

On the basis of $\varphi^*(x_j)$ and $k(x, y)$, we can get $\phi(x)$ as the final descent direction, which makes endmember iteration along PDD so that the endmembers are pushed towards the Pareto set [57], $\phi(x)$ is

$$\phi(x_i) = \frac{1}{n}\sum_{j=1}^{n} \varphi^*(x_j) - \alpha \bigtriangledown_{x_j} k(x_i, x_j) \tag{17}$$

where α is a positive coefficient that controls the importance of the second term and we set α as -0.5.

C. The Hash-based Update Approach in GMOGH

From section B, without considering discrete constraint, we construct the search approach Eq.(13) and the search PDD $\phi(x)$ to get optimal continuous solution x. In this section, we take discrete constraint $s \in \{0, 1\}^m$ into consideration. However, the optimization problem will become NP-hard after considering the discrete constraint. In

order to solve this NP-hard problem, we utilize greedy algorithm which selects the optimal option in each iteration and ultimately reaches an optimal discrete point, that is the optimal endmember [59].

Under the guidance of greedy principle, the hash-based update approach updates endmembers towards the optimal discrete solution in each iteration greedily and gets the global optimal endmember finally. Suppose x calculated by Eq.(13) is the optimal continuous solution. We regard optimal discrete solution s as the closest discrete point to x, which can be calculated by a hash coding approach in each iteration, that is $binary(x)$:

$$binary(x) = \begin{cases} 0, & \text{if } x < \beta, \\ 1, & \text{if } x \ \geq \ \beta \end{cases} \tag{18}$$

where β is a fixed constant. In order to make a proper choice on hyper-parameter β, we regard it as a critical point. By adopting $binary(x)$, we can not only get optimal discrete solution but also ensure that the selected endmembers are truth based on sparsity error. Therefore, β is expected to satisfy the following properties.

Property 1: β can be employed as a distinct boundary to accurately distinguish whether the corresponding spectrum is selected as real endmember. On the basis of property (1), we constrain β with properties (2) and (3).

Property 2: When the location is set as 1, it is expected that the corresponding selected endmembers are real optimal solution, that is, β is required not to be too small. Otherwise, the number of endmembers is overestimated, causing some mistake endmembers to be selected. So β needs to satisfy: $\beta \geq \ 0.3$.

Property 3: When the location is set as 0, it is expected that the corresponding selected endmembers are not real optimal solutions, that is, β is required not to be too large. Otherwise, the number of endmembers is underestimated, causing some real endmembers not to be selected. So β needs to satisfy: $\beta \leq \ 0.99$.

Property 4: The specific values of β will be considered within the range of $[0.3, \ 0.99]$ and obtained through parameter sensitivity experiments, which are displayed in Section IV.

In order to get the proper value of β, we conduct parameter sensitive experiments on both synthetic data set 1 and synthetic data set 2. Finally, parameter β is fixed at $\beta = 0.9$ to get the optimal result.

Concretely we use the following equation as our update approach:

$$s' = bianry(x + \varepsilon\phi(x)) \tag{19}$$

We update endmembers by Eq.(19), which can not only remove the discrete binary constraints but also ultimately reaches optimal endmembers. Moreover, getting optimal discrete solution in each iteration can help us observe the variation of endmembers directly.

As mentioned above, in this section, we propose a gradient-based multiobjective unmixing framwork works on endmember selection and the whole process of GMOGH is shown in Algorithm 1. The algorithm time complexity of GMOGH is calculated as $O(n)$.

D. Theoretical Analysis of GMOGH

In this section, we prove the theoretical analysis of GMOGH.

Theorem 1: $\varphi^*(x)$ is a PDD.

Proof: Under our assumption we know that:

$$\varphi^*(x) \propto -\sum_{i=1}^{m} \mu_i^*(x)\varphi_i(x) \tag{20}$$

Then we suppose:

$$\varphi^*(x) = -k\sum_{i=1}^{m} \mu_i^*(x)\varphi_i(x) \tag{21}$$

where k is a positive constant, $k \geq 0$.

Then we verify the condition to be a PDD, that is, check Eq.(11) is true or not:

$$\begin{aligned}(-\bigtriangledown f_i(x))^T d &= (-\bigtriangledown f_i(x))^T \varphi^*(x) \\ &= -k\sum_{i=1}^{m} \mu_i^*(x)\varphi_i^T(x)\varphi_i(x) \\ &\leq 0\end{aligned} \tag{22}$$

as $\mu_i^*(x) \geq 0,\ k \geq 0,\ \varphi_i^T(x)\varphi_i(x) \geq 0$.

That is to say, condition Eq.(11) has been satisfied and $\varphi^*(x)$ is a PDD.

Theorem 2: $binary(x)$ has no impact on convergence.

Proof: We update discrete solution s by using Eq.(19), at each iteration, we can find that:

$$\begin{aligned}&s_1 = binary(x_1) \\ &s_2 = binary(x_2) = binary(x_1 + \varepsilon\phi(x_1)) \\ &..... \\ &s_n = binary(x_n) = binary(x_{n-1} + \varepsilon\phi(x_{n-1})) \\ &= binary(x_1 + \varepsilon\sum_{i=1}^{n-1} \phi(x_i))\end{aligned} \tag{23}$$

It is obvious to know that $\{s_n\}$ is bounded as $s_n^i = \{0, 1\}$ and $\{s_n\}$ is only related to $\{x_n\}$. That is to say, the convergence of $\{s_n\}$ equals to the convergence of $\{x_n\}$. So $\{s_n\}$ has no impact on the whole convergence of GMOGH.

Theorem 3: Objective functions with their gradient functions satisfy the following conditions

1) Each objective function f_i is C^3, i.e., third-order continuously differentiable.

2) $c_1 I \preceq \bigtriangledown^2 f_i \preceq c_2 I, \left|\partial^3_{i,j,l} f_i\right| \leq c_3$ where c_1, c_2, c_3 are constants and $\bigtriangledown^2 f_i$ is the hessian matrix of f_i. $A \preceq B$ indicates that $B - A$ is a positive semi-definite matrix.

3) $\|g^*\|^2$ *is* C^1 , g^* is L-Lipschitz.

Proof: Consider about our objective functions Eq.(12) with their gradient information Eq.(15). Note that the kth row of library A is A_k, the kth column of abundance X is $W_{(k)}$, the element in A is a_{ij}. With these notations in place, the objective functions can be written in the following equivalent forms:

$$\begin{aligned}f_1(s) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_{ij} - \sum_{k=1}^{n} A_{ik} \cdot diag(s)_{kk} \cdot W_{kj}\right| \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_{ij} - \sum_{k=1}^{n} a_{ik} \cdot s_k \cdot w_{kj}\right|\end{aligned} \tag{24}$$

Algorithm 1: Pseudocode of GMOGH

Input: hyperspectral image **Y**, known spectral library **A**

Output: optimal endmember s^*, abundance **X**

Initialization:

an initial population $x = (x^1, ..., x^p)$, population size p,

individual $x^i = (x_1^i, ..., x_m^i)$, iteration number T

Endmember selection:

while $t < T$ **do:**

for i=1,2,...,p do:

calcute PDD of every individual:

$\phi(x_t) = \frac{1}{n} \sum_{j=1}^{n} \varphi^*(x_j) - \alpha \bigtriangledown_{x_j} k(x_t, x_j)$

search the optimal continuous solution:

$x_{t+1} = x_t + \varepsilon\phi(x_t)$

update the optimal discrete solution:

$s_{t+1} = binary(x_{t+1})$

update the optimal spectrum:

$A_{s_{t+1}} = A \times diag(s_{t+1})$

t=t+1

Return the final endmember s^* and cmopute the

corresponding spectral signatures A_{s^*}

Abundance Estimation:

get abundance matrix by using nonnegative least

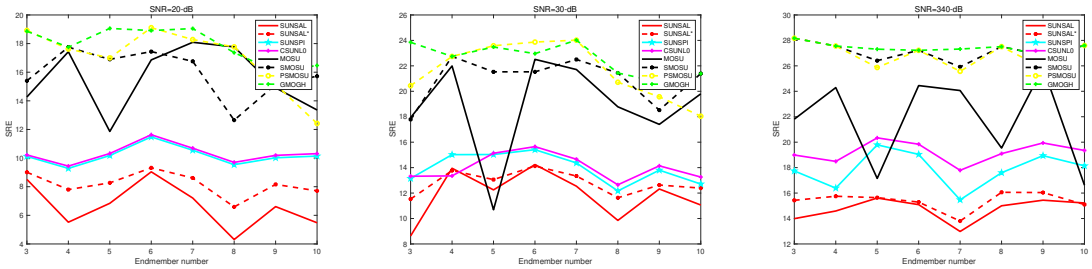
squares algorithm:

$X = arg\min_{x \geq 0} \|Y - A_{s^*} X\|_2$

$$f_2(s) = \sum_{i=1}^{n} |s_i| - k \tag{25}$$

Then we take the derivative of the functions:

$$\begin{aligned} \frac{\partial f_1(s)}{\partial s} &= (\frac{\partial f_1(s)}{\partial s_1}, \frac{\partial f_1(s)}{\partial s_2}, ..., \frac{\partial f_1(s)}{\partial s_n}) \\ &= (\sum_{i=1}^{n}\sum_{j=1}^{n} |-a_{i1}w_{1j}|, ..., \sum_{i=1}^{n}\sum_{j=1}^{n} |-a_{in}w_{nj}|) \\ &= (A_1 W_{(1)}, A_2 W_{(2)}, ..., A_n W_{(n)}) \end{aligned} \tag{26}$$



(a) (b) (c)

Fig. 2. Comparison of SRE results obtained by SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU and GMOGH on synthetic data 1. Subplots (a)–(c) respectively present the results for 20-/30-/40-dB data. For each subplot, endmember number setted as the abscissa label is varying from 3 to 10.

$$\begin{aligned}\frac{\partial f_2(s)}{\partial s} &= (\frac{\partial f_2(s)}{\partial s_1}, \frac{\partial f_2(s)}{\partial s_2}, ..., \frac{\partial f_2(s)}{\partial s_n}) \\ &= (\frac{\partial(|s_1 + ... + s_n| - k)}{\partial s_1}, ..., \frac{\partial(|s_1 + ... + s_n| - k)}{\partial s_n}) \\ &= (1, .., 1)\end{aligned} \tag{27}$$

the last equation holds under the condition: $s_i \geq 0$, and

$$\frac{\partial^2 f_1(s)}{\partial s^2} = \frac{\partial^3 f_1(s)}{\partial s^3} = \frac{\partial^2 f_2(s)}{\partial s^2} = \frac{\partial^3 f_2(s)}{\partial s^3} \tag{28}$$

exists and continuous as their first-order derivation is independent of s. So $\bigtriangledown^2 f_i = 0$ satisfies condition 2.

The gradient information

$$\begin{aligned}\varphi^*(s) &= \lambda_1(s)\varphi_1(s) + \lambda_2(s)\varphi_2(s) \\ &= (\lambda_1(s)A_1W_{(1)} + \lambda_2(s), ..., \lambda_1(s)A_nW_{(n)} + \lambda_2(s))\end{aligned} \tag{29}$$

is bounded as $\lambda_1(s) + \lambda_2(s) = 1$. And

$$\left|\frac{\partial\varphi^*(s)}{\partial s}\right| = \left|\frac{\partial\lambda_1(s)\varphi_1(s)}{\partial s} + \frac{\partial\lambda_2(s)\varphi_2(s)}{\partial s}\right| \leq k \tag{30}$$

as $\lambda_1(s)$ and $\lambda_2(s)$ are solutions of an optimization problem, where k is a constant.

On the basis of *Therorem 3* and the proof in [57], we can know that kernel embedding of φ^* can be bounded by σ in Eq.(16) and $\|\varphi^*(x)\|$ can be arbitrarily small.

As we all know, gradient $\|\varphi^*(x)\| = 0$ indicates that x is in the Pareto set, so φ^* is very small equals to convergence, which has been completed.

IV. Experiments

In this section, we evaluate the validity of GMOGH by conducting two synthetic experiments quantitatively and three real-world experiment qualitatively. In the first synthetic experiment, the results of endmember assessment and unmixing accuracy are considered since GMOGH mainly works on endmember selection. In the second

TABLE I

SELECTED ENDMEMBERS OF PMOSU AND GMOGH ON SYNTHETIC DATA 1 WITH 30-DB NOISE

K	Real Endmembers	PMOSU	GMOGH
3	$1-3$	$1-3,351$	$1-3$
4	$1-4$	$1-4$	$1-4$
5	$1-5$	$1-5$	$1-3,210-212$
6	$1-6$	$1-6$	$1-5$
7	$1-7$	$1-7$	$1-7$
8	$1-8$	$1-8,157$	$1-8$
9	$1-9$	$1-9,278$	$1-9$
10	$1-10$	$1-10,441$	$1-10$

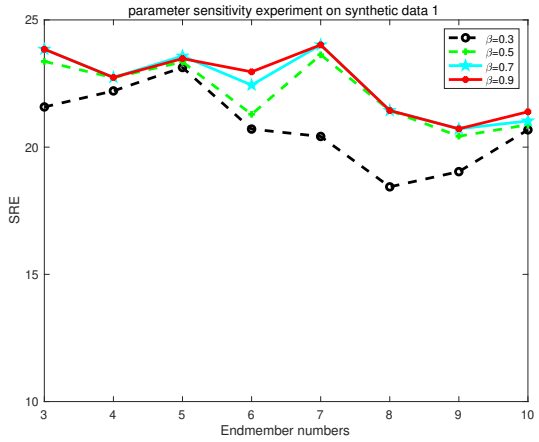


Fig. 3. The parameter sensitivity experiment on synthetic data 1 by GMOGH. The number of endmembers is varying from 3 to 10 and the value of β varies from 0.3 to 0.9 at interval of 0.2.

synthetic experiment, we consider both quantitative unmixing acuracy and qualitative abundance maps. In real-world experiments, only qualitative analysis of abundance maps are available in these cases. Some sparse unmixing methods and multiobjective unmixing methods are compared with GMOGH to evaluate the validity. The unmixing accuracy of these methods can be verified by the metric of signal-to-reconstruction error (SRE). We define SRE

TABLE II
TPR/FPR RESULT OF MOSU, SMOSU, PMOSU AND GMOGH ON SYNTHETIC DATA 1

SNR	METHOD	k	3	4	5	6	7	8	9	10
20-dB	MOSU	TPR	1	0.750	0.600	0.333	1	**1**	0.556	0.900
		FPR	0.002	0.004	0.006	0.008	0.002	**0**	0.008	0.002
	SMOSU	TPR	1	**1**	1	1	0.857	0.625	0.778	0.900
		FPR	0.004	**0**	0.008	0.006	0.010	0.012	0.014	0.006
	PMOSU	TPR	1	1	1	**1**	1	**1**	0.750	0.900
		FPR	0.002	0.002	0.002	**0**	0.002	**0**	0.004	0.006
	GMOGH	TPR	1	**1**	1	0.833	1	0.750	0.778	0.900
		FPR	0.004	**0**	0.006	0.004	0.006	0.008	0.010	0.006
30-dB	MOSU	TPR	1	1	0.400	1	1	0.750	0.889	0.900
		FPR	0.002	0.002	0.008	0.002	0.002	0.006	0.004	0.002
	SMOSU	TPR	1	**1**	1	1	1	**1**	0.778	**1**
		FPR	0.002	**0**	0.002	0.002	0.002	**0**	0.014	**0**
	PMOSU	TPR	1	**1**	**1**	**1**	**1**	1	1	1
		FPR	0.002	**0**	**0**	**0**	**0**	0.002	0.002	0.002
	GMOGH	TPR	**1**	**1**	0.750	0.833	**1**	**1**	**1**	**1**
		FPR	**0**	**0**	0.006	0	**0**	**0**	**0**	**0**
40-dB	MOSU	TPR	1	1	0.600	1	0.857	0.250	1	0.200
		FPR	0.002	0.002	0.006	0.002	0.002	0.014	0.002	0.021
	SMOSU	TPR	**1**	**1**	0.800	**1**	0.857	**1**	**1**	**1**
		FPR	**0**	**0**	0	**0**	0	**0**	**0**	**0**
	PMOSU	TPR	**1**	**1**	1	**1**	1	**1**	1	**1**
		FPR	**0**	**0**	0.002	**0**	0.002	**0**	0.002	**0**
	GMOSU	TPR	**1**	**1**	**1**	**1**	**1**	0.875	**1**	**1**
		FPR	**0**	**0**	**0**	**0**	**0**	0	**0**	**0**

for abundance matrix as:

$$SRE = 10 \times log\left(\frac{E(\|X_{real}\|_F^2)}{E(\|X_{real} - X_{estimation}\|_F^2)}\right) \tag{31}$$

where X_{real} is the abundance matrix of real image data and $X_{estimation}$ is the estimation of each algorithm. We argue that one is better than others if its SRE is larger.

Since GMOGH works on the process of endmember selection, we consider another criterion for endmember selection in [46] which are true positive rate (TPR) and false positive rate (FPR). An endmember is viewed as positive if it exists in real hyperspectral data and negative else. We also call an endmember as predicted positive if it is selected in algorithm. TPR and FPR are calcuted by True Negative (TN), True Positive (TP), False Negative (FN) and False Positive (FP):

$$TPR = \frac{TP}{TP + FN} \tag{32a}$$

$$FPR = \frac{FP}{FP + TN} \tag{32b}$$

where TN denotes negative and predicted negative case, TP denotes positive and predicted positive case, FN denotes positive but predicted negative case, FP denotes negative but predicted positive case.

In general, the values of TPR and FPR are between 0 and 1. The result of TPR is 1 means that algorithm gets all the endmembers in real data and FPR equals to 0 means that algorithm has no extra endmembers. That both of these cases happened implies an accurate result in the algorithm, as it selects the real endmembers.

GMOGH is compared with some sparse unmixing methods and multi-objective unmixing methods, namely SUNSAL [30], SUNSPI [34], CSUNL0 [32], MOSU [53], SMOSU [52] and PMOSU [47]. SUNSAL relaxes the l_0 sparsity into l_1 norm. The spectral *a priori* information was adopted in SUNSPI to improve unmixing accuracy. CSUNL0 aims at solving l0 problem directly. MOSU, SMOSU and PMOSU are multiobjective unmixing methods based on evolutionary algorithms for endmember selection. MOSU considers three objectives simultaneously and SMOSU considers two objectives. The pruning operation is realized as an objective in PMOSU, which realizes simultaneous library pruning and unmixing by a multiobjective approach.

A. Experimental Results With Synthetic Data 1

We select ground-truth spectra from the Chapter 1 of the United States Geological Survey (USGS) digital spectral library (splib06a) to generate synthetic data 1. Splib06a contains 224 spectral bands that collected under 498 spectral signatures and this library also can be used as the matrix A in Eq.(5). We set the real number of endmembers k from 3 to 10 in synthetic data 1, which contains 64×64 pixels. Dirichlet distribution is the basis on which we generate fractional abundances. Moreover, we manually choose different noises in order to simulate the different conditions in the natural environment. The signal-to-noise ratios (SNR) of our experimental noises are 20-/30-/40-dB, respectively. SNR is defined as follow:

$$SNR = 10 \times log\left(\frac{E(\|Y\|_F^2)}{E(\|N\|_F^2)}\right) \tag{33}$$

TABLE III
SRE RESULTS OF SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU AND GMOGH ON SYNTHETIC DATA 1

SNR	K	SUNSAL	SUNSAL*	SUNSPI	CSUNL0	MOSU	SMOSU	PMOSU	GMOGH
20-dB	3	8.51	9.01	10.12	10.22	14.27	15.4	**18.92**	18.86
	4	5.52	7.80	9.28	9.44	17.43	**17.75**	17.62	**17.75**
	5	6.84	8.27	10.18	10.32	11.86	16.9	17.10	**19.06**
	6	9.04	9.32	11.48	11.63	16.86	17.44	**19.11**	18.91
	7	7.21	8.61	10.55	10.69	18.09	16.76	18.28	**19.05**
	8	4.31	6.59	9.54	9.70	17.77	17.65	**17.77**	17.37
	9	6.61	8.16	10.02	10.19	14.91	15.14	15.15	**16.04**
	10	5.48	7.71	10.14	10.31	13.36	15.72	12.42	**16.47**
30-dB	3	8.63	11.53	13.15	13.32	17.93	17.79	20.44	**23.85**
	4	13.95	13.77	15.02	15.35	21.99	**22.74**	**22.74**	**22.74**
	5	12.25	13.06	15.03	15.13	10.69	21.53	**23.58**	23.48
	6	14.24	14.12	15.41	15.65	22.51	21.53	**23.87**	22.96
	7	12.54	13.34	14.37	14.66	21.71	22.50	**24.02**	**24.02**
	8	9.85	11.63	12.17	12.65	18.78	**21.44**	20.71	**21.44**
	9	12.32	12.63	13.8	14.14	17.4	18.52	19.57	**20.72**
	10	11.06	12.39	12.71	13.26	19.8	**21.39**	18.04	**21.39**
40-dB	3	13.99	15.43	17.74	18.99	21.81	**28.17**	**28.17**	**28.17**
	4	14.59	15.75	16.39	18.49	24.29	**27.56**	**27.56**	**27.56**
	5	15.61	15.64	19.79	20.34	17.15	26.41	25.87	**27.31**
	6	15.09	15.29	19.04	19.84	24.45	**27.22**	**27.22**	**27.22**
	7	12.98	13.80	15.47	17.79	24.06	25.92	25.57	**27.32**
	8	15.00	16.06	17.60	19.10	19.54	**27.51**	**27.51**	**27.51**
	9	15.44	16.04	18.94	19.94	25.83	**26.82**	25.86	**26.82**
	10	15.22	15.10	18.14	19.35	16.62	**27.60**	**27.60**	**27.60**

TABLE IV
COMPARISON OF SRE RESULTS OBTAINED BY SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU AND GMOGH ON SYNTHETIC DATA 2 FOR 20-/30-/40-DB.

SNR	SUNSAL	SUNSAL*	SUNSPI	CSUNL0	MOSU	SMOSU	PMOSU	GMOGH
20-dB	8.77	9.03	12.34	12.54	17.91	18.44	**19.40**	**19.40**
30-dB	13.33	14.62	16.87	17.21	20.59	**24.32**	**24.32**	**24.32**
40-dB	16.84	16.90	20.78	22.60	28.65	**29.32**	**29.32**	**29.32**

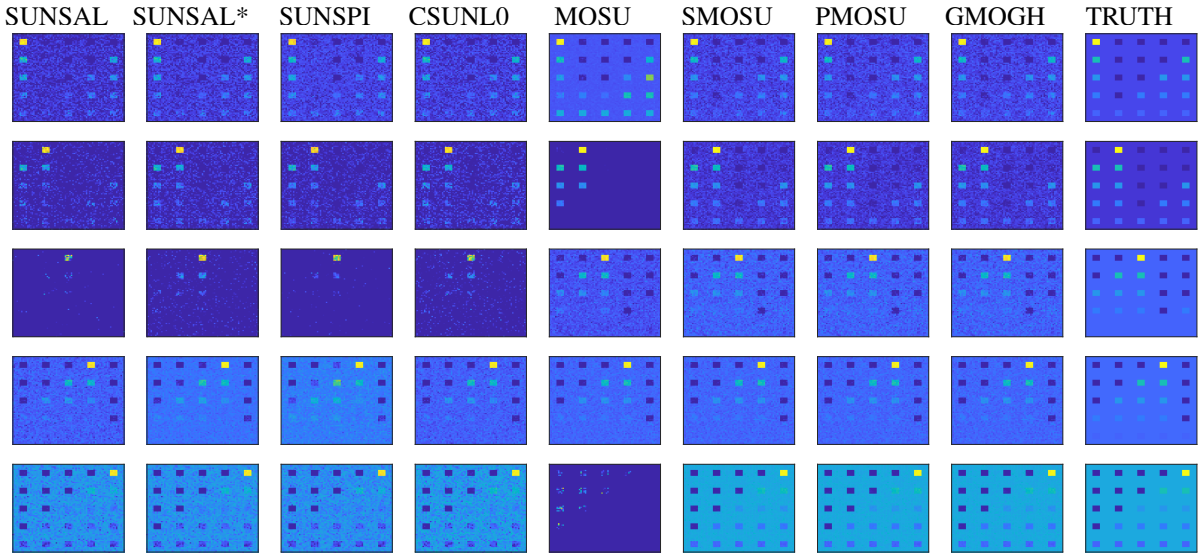


Fig. 4. Different abundance maps from top to bottom corresponding to endmembers number from 1 to 5 on synthetic data 2. These pictures from left to right are respectively obtained by SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU, GMOGH and the truth with 30-dB noise as an example.

The observed image data Y in Eq.(12) is obtained based on fractional abundances and the selected endmembers. We multiply them complying with the LMM. In fact, We select real endmembers by setting spectra at the front for simplicity. We take synthetic data 1 to dsiplay the results of endmember assessment and unmixing accuracy.

1) Endmember assessment: We compare real endmembers and selected endmembers by algorithms to evaluate the performance of GMOGH. Table I shows the selected endmembers' serial numbers by PMOSU and GMOGH under 30-dB noise as an example. There are 498 spectra in spectral library A which are numbered in order from 1st to the 498th. For example, "1-3" in the first column indicates the spectra from 1st to 3rd are selected in synthetic data, "1-3,351" in the second column indicates the spectra from 1st to 3rd and 351th are selected in algorithm. It is observed that GMOGH outperforms PMOSU in most cases and GMOGH can optimize to get the correct endmembers except when $k = 5$ and $k = 6$ under 30-dB noise.

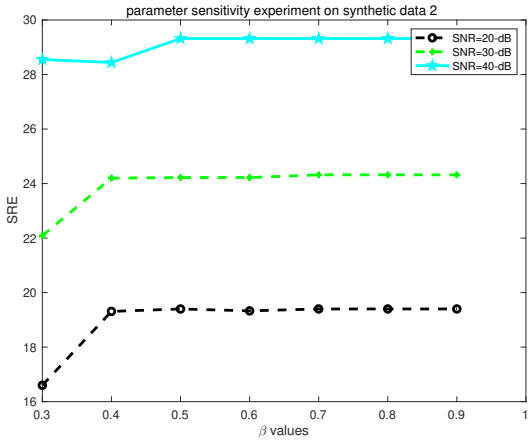


Fig. 5. The parameter sensitivity experiment on synthetic data 2 by GMOGH. Different lines respectively present the results for 20-/30-/40-dB and the value of β varies from 0.3 to 0.9 at intervals of 0.1.

Tables II presents the TPR and FPR results for MOSU, SMOSU, PMOSU and GMOGH because both of them are unmixing methods based on endmember selection. The endmember assessment results of four methods on synthetic data 1 are shown under three noise conditions, respectively. TPR=1 and FPR=0 are viewed as the best situation since it denotes that all endmembers have been found. In many cases, the reason why TPR results of these four methods are less than 1 is that they losses several active endmembers and the FPR results are larger than 0 denotes that they obtain more redundant endmembers. From Table II, the TPR and FPR results of GMOGH is better than other endmember selection algorithms because GMOGH can get TPR=1 and FPR=0 in most cases which indicates the optimal endmembers have been found.

2) Unmixing accuracy: Tables III presents the SRE results on synthetic data 1 under 20-/30-/40-dB noises cases, respectively, for SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU and GMOGH. From the table, the SRE results of GMOGH is larger than the other methods for a large proportion of $k = 3$–10 cases, so Table I indicates that GMOGH outperforms both sparse unmixing and multiobjective mixing methods, especially with high SNR. In some cases, for example, in 40-dB noise cases, SMOSU and PMOSU performs as well as GMOGH when k=3,4,6,8 and 10. It represents that as a method based on endmember selection, both of them have found the best endmembers and after reconstructing with abundance, they can get the same results.

Fig.2 shows the SRE results by various algorithms under 20-/30-/40-dB noises cases, respectively. The horizontal coordinates of the pictures are the number of endmembers. By analyzing the y-coordinate in Fig.2, it is clear that the weaker the noises the higher unmixing accuracy of algorithms and GMOGH surpasses the other seven methods in most cases. For example, in Fig.2(a), GMOGH performs better than other methods when the number of endmembers equals to 4,5,7,9 and 10. From Fig.2(b) and Fig.2(c), SMOSU, PMOSU and GMOGH are get same and high accuracy, this is because all the real endmembers are selected by these multiobjective based unmixing methods.

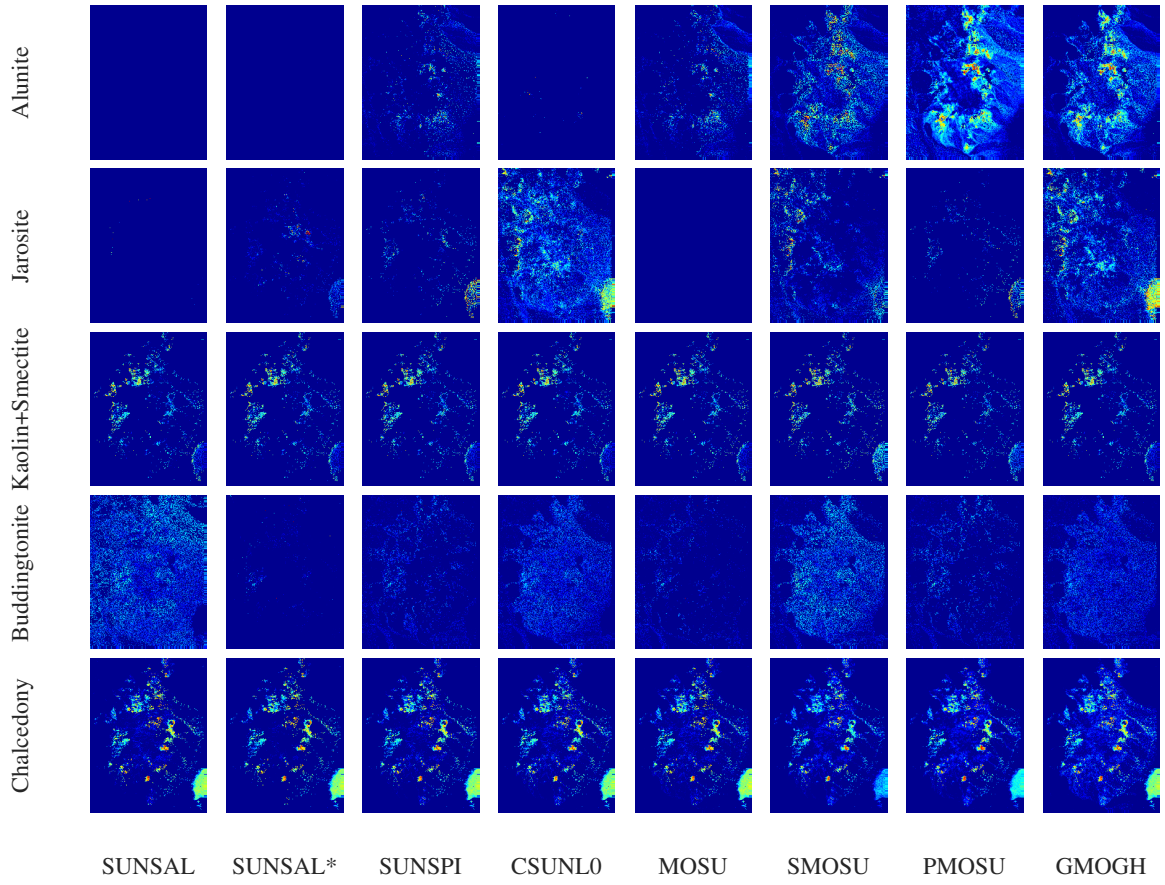


Fig. 6. Comparison of abundance maps on Cuprite data. (Left to right) Abundance maps obtained by SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU and GMOGH respectively. (Top to bottom) Maps corresponding to Alunite, Jarosite, Kaolinite+Smectite, Buddingtonite, Chalcedony.

3) Parameter sensitivity experiment: On synthetic data 1, we not only compare the endmember assessment and unmixing accuracy of different algorithms, but also conduct parameter sensitivity experiment. In order to verify the influence of parameter β in Eq.(18) and select an appropriate parameter value, we fix SNR at 30-dB and consider some possible values of β. Fig.3 shows the SRE results achieved by GMOGH with different endmember numbers and various parameter values. The abscissa of picture is set as the number of endmembers which varies from 3 to 10 and four broken lines represent four kinds of β whose value ranges from 0.3 to 0.9 at intervals of 0.2. We think that the higher location line is, the more reliable β is. It obviously to see that majority optimal SRE results are achieved when $\beta = 0.9$ except endmember number equals to 5. Moreover, we find that $\beta \geq 0.5$ results in similar SRE results of GMOGH. Because of this stability, GMOGH is supposed insensitive to parameter β. This assumption can be verified in subsection.*B*.

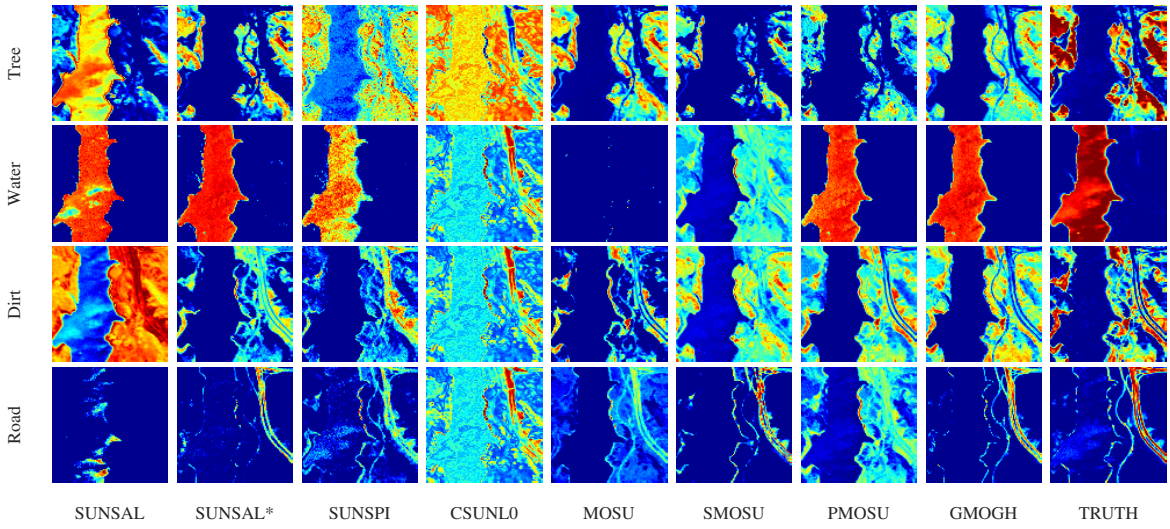


Fig. 7. Comparison of abundance maps on Jasper data set. (Left to right) Abundance maps obtained by SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU, GMOGH and TRUTH respectively. (Top to bottom) Maps corresponding to Tree, Water, Dirt, Road.

B. Experimental Results With Synthetic Data 2

1) Comparison of different unmixing methods: The method of generating data in literature [35] is also applicable to obtaining data 2. The image size of synthetic data 2 is 75×75 pixels and 5 equals to real endmembers number, which is fixed rather than varied. Synthetic data 2 constructs not only pure regions but also mixed regions, which are generated by using mixtures ranging from two to five endmembers. The spatial distribution forms of endmembers are distinct square regions. The background pixels in image data are consisted by mixtures of the same five endmembers. After completing the above operations, the image date was contaminated with correlated noise, whose SNR are 20-/30-/40-dB respectively.

The results of abundance estimation in synthetic data 2 under 20-/30-/40-dB noise cases are presented in Table IV. From Table IV, SRE results of PMOSU and GMOGH are larger than the other methods for each different SNR cases. The abundance maps obtained by different methods is shown in Fig.4 under 30-dB with five endmembers as an example. The endmembers selected by SMOSU, PMOSU and GMOGH are the same and real. So their estimated abundance maps look closer to the truth, which presents the equal result with Table II.

2) Parameter sensitivity experiment: On synthetic data 2, we also conduct parameter sensitivity experiment to verify the assumption of immunity and select proper value of parameter β in Eq.(18) under different noise conditions. In this experiment, we fix the number of endmember at k=5. Fig.5 shows the SRE results achieved by GMOGH with different SNR conditions and various parameter values. The abscissa of picture is set as the value of β which varies from 0.3 to 0.9 and three broken lines represent three kinds of noise. This phenomenon indicates that stable results can be obtained at $\beta \geq 0.5$. Moreover, the stability in three noise conditions verifies that GMOGH is insensitive to β. Combined with the results in Fig.3, we set $\beta = 0.9$ in order to pursue an optimal and accurate unmixing result.

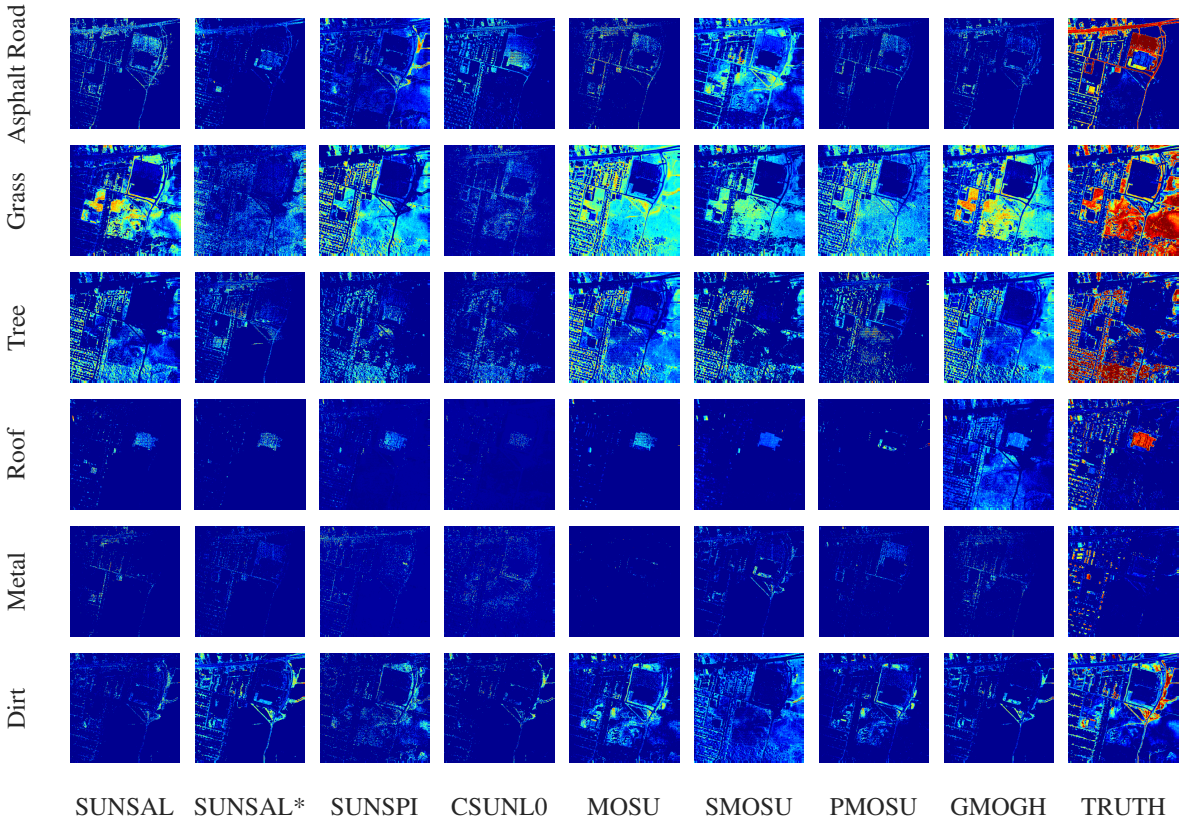


Fig. 8. Comparison of abundance maps on Urban data. (Left to right) Abundance maps obtained by SUNSAL, SUNSPI, CSUNL0, MOSU, SMOSU, PMOSU, GMOGH and TRUTH respectively. (Top to bottom) Maps corresponding to Asphalt Road, Grass, Tree, Roof, Metal, Dirt.

C. Experimental Results With Real-World Data

Cuprite data set, Jasper data set and Urban data set are adopted as real image data sets in this experiment. The spectral library A is obtained from the first chapter of splib06a with 498 endmembers and A is employed as the prior known spectral library. The reason why we do this operation is that this chapter contains signatures of all the materials contained in these real data sets. Before actual experiments, we remove the noisy bands in A accordingly. It should be noted that the performance of unmixing algorithms in real-World data are difficult to evaluate by quantitative criteria, so we provide qualitative and visualized assessments. What is more, the truth abundance maps can only be adopted as a reference, but not as a criterion to evaluate the effectiveness of algorithms.

There are 250×190 pixels in the original image data of Cuprite, which contains 224 bands but the final image data contains only 188 bands. This is because a few useless bands are discarded from the original ones. The abundance maps of five different materials are presented in Fig.6, these pictures are qualitative results obtained by different unmixing methods. Compared in Fig.6, images of GMOGH achieve a clearer outline for Alunite, Jarosite and Chalcedony.

Four endmembers are contained in Jasper: Tree, Water, Dirt, Road, which represented in Fig.7 from top to bottom.

There are 100×100 pixels in Jasper. Each pixel is recorded at 224 channels. The abundance maps of four different materials are presented in Fig.7, these pictures are qualitative results obtained by different unmixing methods and ground truth. We compare the result of an unmixing algorithm with truth maps. In Fig.7, the maps of GMOGH are closer to truth for Dirt and Road.

Urban data set has six different materials which are termed as Asphalt Road, Grass, Tree, Roof, Metal and Dirt respectively. This data set has 307×307 pixels in it. Each pixel is recorded at 210 channels. The abundance maps are presented in Fig.8, which has the same arrangement as Fig.7. Compared in Fig.8, maps of GMOGH seem closer to truth for Grass, Tree and Roof.

Based on the analysis of three data sets, we can draw the conclusion that the performance of multiobjective unmixing methods on real data sets is better than that of the sparse unmixing methods. Moreover, compared with other multiobjective methods, GMOGH can get convergence results and get good results to some extent. Therefore, the validity of GMOGH is verified on real data sets.

V. CONCLUSION

Multiobjective optimization is introduced in hyperspectral unmixing problem recently, thus reducing the adjusting of regularization coefficient. Despite the success of multiobjective unmixing by applying evolutionary algorithms, a limitation of these approaches has been the fact that they do not converge. To circumvent this limitation, in this paper, GMOGH has been specifically developed. In the search method, gradient information of objective functions have been used in generating pareto descent direction; in the update method, greedy hash coding has been taken into consideration to get binary optimal endmember. Furthermore, some theoretical analyses are presented for GMOGH.

Both simulated and real hyperspectral data sets experiments are used to test the performance of GMOGH. Sparse unmixing algorithms and multiobjective unmixing algorithms are compared with GMOGH. Experimental results indicate that the gradient-based method displays its ability in hyperspectral unmixing.

REFERENCES

[1] T. Guo, F. Luo, L. Fang, and B. Zhang, "Meta-pixel-driven embeddable discriminative target and background dictionary pair learning for hyperspectral target detection," *Remote Sensing*, vol. 14, no. 3, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/3/481

[2] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-d cnn framework for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 3–13, 2019.

[3] S. Jia, Z. Wang, Q. Li, X. Jia, and M. Xu, "Multiattention generative adversarial network for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[4] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1183–1194, 2019.

[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*, vol. 5, no. 2, pp. 354–379, 2012.

[6] Y. Su, X. Xu, J. Li, H. Qi, P. Gamba, and A. Plaza, "Deep autoencoders with multitask learning for bilinear hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8615–8629, 2021.

[7] M. Li, B. Yang, and B. Wang, "Spectral–spatial reweighted robust nonlinear unmixing for hyperspectral images based on an extended multilinear mixing model," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[8] J. Gu, B. Yang, and B. Wang, "Nonlinear unmixing for hyperspectral images via kernel-transformed bilinear mixing models," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[9] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep Autoencoder Networks for Hyperspectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4309–4321, 2019.

[10] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[11] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Imaging Spectrometry V*, M. R. Descour and S. S. Shen, Eds., vol. 3753, International Society for Optics and Photonics. SPIE, 1999, pp. 266 – 275. [Online]. Available: https://doi.org/10.1117/12.366289

[12] X. Jiang, Y. Zhao, M. Gong, T. Zhan, and M. Zhang, "A vertex-directed evolutionary algorithm for multiobjective endmember estimation," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[13] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4112–4122, 2011.

[14] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "MiSiCNet: Minimum simplex convolutional network for deep hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[15] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, 2009.

[16] H. Liu, Y. Lu, Z. Wu, Q. Du, J. Chanussot, and Z. Wei, "Bayesian unmixing of hyperspectral image sequence with composite priors for abundance and endmember variability," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[17] Y. Woodbridge, U. Okun, G. Elidan, and A. Wiesel, "Unmixing k -gaussians with application to hyperspectral imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7281–7293, 2019.

[18] L. Dong, X. Lu, G. Liu, and Y. Yuan, "A novel NMF guided for hyperspectral unmixing from incomplete and noisy data," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[19] X. Li, X. Zhang, Y. Yuan, and Y. Dong, "Adaptive relationship preserving sparse NMF for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[20] J. Peng, Y. Zhou, W. Sun, Q. Du, and L. Xia, "Self-paced nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1501–1515, 2021.

[21] 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, special Issue devoted to the Haifa 2005 conference on matrix theory. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002437950500340X

[22] 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.

[23] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2015, pp. 1–4.

[24] M. Zhao, S. Shi, J. Chen, and N. Dobigeon, "A 3-D-CNN framework for hyperspectral unmixing with spectral variability," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[25] X. Xu, X. Song, T. Li, Z. Shi, and B. Pan, "Deep autoencoder for hyperspectral unmixing via global-local smoothing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[26] S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 482–496, 2019.

[27] K. T. Shahid and I. D. Schizas, "Unsupervised hyperspectral unmixing via nonlinear autoencoders," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[28] J. Peng, W. Sun, H.-C. Li, W. Li, X. Meng, C. Ge, and Q. Du, "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 1, pp. 10–43, 2022.

[29] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.

[30] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2010, pp. 1–4.

[31] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained $l_p - l_2$ optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1142–1146, 2013.

[32] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using l_0 norm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5495–5508, 2018.

[33] X. Zhang, Y. Yuan, and X. Li, "Sparse unmixing based on adaptive loss minimization," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[34] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral *A Priori* information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 770–783, 2015.

[35] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[36] Y. Zhong, R. Feng, and L. Zhang, "Non-local sparse unmixing for hyperspectral remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1889–1909, 2014.

[37] J. Huang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Nonlocal tensor-based sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 8, pp. 6854–6868, 2021.

[38] G. Zhang, S. Mei, B. Xie, M. Ma, Y. Zhang, Y. Feng, and Q. Du, "Spectral variability augmented sparse unmixing of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[39] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, 2014.

[40] H. Li, R. Feng, L. Wang, Y. Zhong, and L. Zhang, "Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 629–647, 2021.

[41] D. He, F. Li, S. Zhang, Y. Chen, C. Deng, and S. Wang, "Dual reweighted low-rank sparse unmixing with total variation regularization," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*.

[42] X. Xu and Z. Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 54–69, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616306529

[43] M. Song, Y. Zhong, A. Ma, X. Xu, and L. Zhang, "A joint spectral unmixing and subpixel mapping framework based on multiobjective optimization," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[44] E. Liu, Z. Wu, and H. Zhang, "A multiobjective method leveraging spatial–spectral relationship for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[45] X. Jiang, M. Gong, T. Zhan, K. Sheng, and M. Xu, "Efficient two-phase multiobjective sparse unmixing approach for hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2418–2431, 2021.

[46] X. Xu, Z. Shi, B. Pan, and X. Li, "A classification-based model for multi-objective hyperspectral sparse unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9612–9625, 2019.

[47] X. Xu, B. Pan, Z. Chen, Z. Shi, and T. Li, "Simultaneously multiobjective sparse unmixing and library pruning for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3383–3395, 2021.

[48] Y. Wei, X. Xu, B. Pan, T. Li, and Z. Shi, "A multiobjective group sparse hyperspectral unmixing method with high correlation library," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7114–7127, 2022.

[49] Q. Cheng, B. Du, L. Zhang, and R. Liu, "ANSGA-III: A multiobjective endmember extraction algorithm for hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 2, pp. 700–721, 2019.

[50] Z. Wang, J. Wei, J. Li, P. Li, and F. Xie, "Evolutionary multiobjective optimization with endmember priori strategy for large-scale hyperspectral sparse unmixing," *Electronics*, vol. 10, no. 17, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/17/2079

[51] K. Deb, *Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction*. London: Springer London, 2011, pp. 3–34. [Online]. Available: https://doi.org/10.1007/978-0-85729-652-8_1

[52] X. Xu, Z. Shi, and B. Pan, "l_0-based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 141, pp. 46–58, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271618301163

[53] M. Gong, H. Li, E. Luo, J. Liu, and J. Liu, "A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 2, pp. 234–248, 2017.

[54] W. Gao, Y. Wang, L. Liu, and L. Huang, "A gradient-based search method for multi-objective optimization problems," *Information Sciences*, vol. 578, pp. 129–146, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025521007416

[55] "A stochastic multiple gradient descent algorithm," *European Journal of Operational Research*, vol. 271, no. 3, pp. 808–817, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221718304831

[56] J.-A. Désidéri, "Multiple-gradient descent algorithm (MGDA) for multiobjective optimization," *Comptes Rendus Mathematique*, vol. 350, no. 5, pp. 313–318, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1631073X12000738

[57] X. Liu, X. Tong, and Q. Liu, "Profiling pareto front with multi-objective stein variational gradient descent," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 14 721–14 733. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/7bb16972da003e87724f048d76b7e0e1-Paper.pdf

[58] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[59] S. Su, C. Zhang, K. Han, and Y. Tian, "Greedy hash: Towards fast optimization for accurate hash coding in cnn," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper/2018/file/13f3cf8c531952d72e5847c4183e6910-Paper.pdf