Analysis for the Weakly Pareto Optimum in Multi-Objective based Hyperspectral Band Selection

Bin Pan, Zhenwei Shi and Xia Xu

Abstract

Band selection refers to finding the most representative channels from hyperspectral images. Usually, certain objective functions are designed and combined via regularization terms. A possible drawback of these methods is that they can only generate one solution in a single run with a given band number. To overcome this problem, multi-objective based methods, which were able to simultaneously obtain a series of subsets with different band numbers, were investigated for band selection. However, because the range of band selection problem is discrete, recently proposed weighted Tchebycheff based multi-objective methods may suffer weakly Pareto optimal problem. In this case, the solutions for each band number will be non-unique and no optimal solution exists. Decision makers have to manually select a unique solution for each band number. In this paper, we provide a theoretical analysis about the weakly Pareto optimal problem in band selection, and quantitatively give the boundary conditions. Moreover, we further summarize the suggestions which will help users avoid the weakly Pareto optimal problem. According to these criterions, we develop a new adaptive-penalty based boundary intersection (APBI) framework to improve the multi-objective algorithm in hyperspectral band selection. APBI mainly includes two advantages: 1) avoiding weakly Pareto optimum and 2) reducing the sensibility of the penalty factor. The theoretical analysis is further validated by contrast experiments. The results demonstrate that the weakly Pareto optimal solutions really exist in weighted Tchebycheff methods, while APBI can overcome this problem.

Index Terms

Multi-objective optimization, band selection, hyperspectral imagery, weakly Pareto optimum

The work was supported by the National Key R&D Program of China under the Grant 2017YFC1405605, and the National Natural Science Foundation of China under the Grants 61671037. *(Corresponding author: Zhenwei Shi.)*

Bin Pan and Xia Xu are with Image Processing Center, School of Astronautics, Beihang University, Beijing, China, and College of Geomatics, Shandong University of Science and Technology, Qingdao, China(e-mail: panbin@buaa.edu.cn; xuxia@buaa.edu.cn)

Zhenwei Shi (Corresponding Author) is with Image Processing Center, School of Astronautics, Beihang University, Beijing, China, and Beijing Key Laboratory of Digital Media, Beihang University, Beijing, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing, China, (e-mail: shizhenwei@buaa.edu.cn).

I. INTRODUCTION

Hyperspectral imagery (HSI) includes abundant spectral information which contributes to many remote sensing applications. However, in some certain circumstances only part of the information is required [1], [2], *i.e.*, bands are redundancy. Thus band selection methods are developed. There are two manners for HSI band selection: supervised and unsupervised. The former usually targets at some particular applications such as classification [3]–[8]. By comparison, the unsupervised band selection aims more general situations and has attracted broader research. In this paper we focus on unsupervised band selection.

Researchers usually defined certain objective functions to constrain the bands correlation [1], [9]–[18]. In [1], Wang *et al.* gave a contextual definition about the salient bands in HSI data, and a new objective is proposed. This algorithm was improved in [19] and [20] by clustering based approaches. In [9], a sequential backward based band selection method was proposed, which assumed that redundant bands presented shorter distance to the hyperplane spanned by the other bands. Some other geometry based methods could be observed in [10], [21]. In [13] overall accuracy and redundancy were combined and considered as the objectives. In [16] nonnegative matrix factorization was integrated into the band selection problem. Information evaluation was also popular for band selection, where larger information or lower similarity was usually considered as selection criterions [22], [23].

Although the above methods presented good performance, there are two difficulties that they have to address. First and foremost, the required band number must be set in advance. If users hope to obtain a series of results with different band numbers, the band selection algorithms have to be conducted repeatedly. Besides, it is difficult to determine which objective is the most effective. Although objectives can be combined via regularization terms, the regularization coefficients usually require manual setting.

To overcome these problems, multi-objective (MO) [24] based methods were used for HSI band selection. MO-based methods can directly generate a series of subsets corresponding to different numbers of bands in a single run, so as to avoid the setting of band numbers. Moreover, MO-based methods could optimize several objectives simultaneously without the setting for regularization coefficients. Another advantage of MO-based methods is that they can directly handle L0 problem without any relaxation, which is completely in line with the characteristics of HSI band selection problem. Therefore MO-based methods are promising for band selection.

Recent works about MO-based HSI band selection (MOBS for short) are not many [25]–[27]. Literature [25] was the first attempt to use an MO-based method for band selection, whereas it adopted a supervised manner. Literature [26] (GMOBS for short) and [27] (RMOBS for short) were two recently proposed unsupervised algorithms. RMOBS [27] mainly tried to solve the decision making problem in band selection. The goals of RMOBS were different from traditional MO-based methods, where users still have to set the band number manually. Among these methods GMOBS [26] was a more representative one. It provided a framework for unsupervised band selection, and promising performance was observed. Both GMOBS and RMOBS were developed under the framework of multi-objective evolutionary algorithm based on decomposition (MOEA/D) [28]. MOEA/D is a very popular framework, and it has been widely adopted by many MO-based works [26], [27], [29]–[32]. MOEA/D tries to approximate the Pareto

Front via decomposition based approaches. Usually, weighted Tchebycheff (WT) is used for decomposition [26], [27]. After sufficient iterations the Pareto Front where the final solution set lies can be determined. Overall, the most significant advantage of MOBS methods is that they can generate a series of subsets in a single running.

Unfortunately, this advantage is compromised due to the existence of weakly Pareto optimal solutions. Because the range of band selection problem is discrete, MOEA/D is hard to reach Pareto optimal. A more common situation is that several solutions will reach optimal at the same time. This situation is called weakly Pareto optimum [24]. Theoretically, these solutions are equally optimal. For some real applications such as band selection, we hope that the solution must be unique. Thus weak Pareto solutions should be avoided.

In this paper, we give a detailed discussion about the reasons and conditions of weakly Pareto optimum in band selection problem. In the following sections, we will draw the conclusion that weighted Tchebycheff based decomposition is bound to bring about weakly Pareto optimum. Moreover, we also prove that even improving the decomposition process by boundary intersection (BI) or penalty-based boundary intersection (PBI), weakly Pareto optimum would still appear in certain conditions. Finally, we give some suggestions and guidance for how to design an available approach to avoid weakly Pareto optimal problem. Based on these analysis, we propose a new adaptive-penalty based boundary intersection (APBI) framework for decomposition. APBI is a general framework which can be easily extended to other preferred forms. This paper can be considered as an extension of our conference version [33]. However, the main text and the novelties are completely reorganized. This paper pays more attention to the theoretical analysis, and provides a general solution. While literature [33] directly proposed an available algorithm without sufficient discussion. The algorithm developed in [33] can be considered as a particular case of APBI. Besides, more convincing experiments are conducted to validate the correctness of the theoretical analysis. Overall, the contributions of this paper can be summarized as follows:

- We give a theoretical discussion for the weakly Pareto optimal problem in HSI band selection, and quantitatively explain the corresponding boundary conditions. We also theoretically prove that WT based decomposition cannot reach Pareto optimal front.
- Some strategies about how to avoid weakly Pareto optimum are suggested, and a new adaptive-penalty based boundary intersection approach is developed for multi-objective based band selection.

In Section II, we will introduce the background for multi-objective optimization based band selection. The main work will be shown in Section III. Verification and comparison experiments are conducted in Section IV. We draw the conclusion in Section V. During our discussion, we will try to design various illustrations to provide more visual and intuitional explanation.

II. MOBS

In this section, we first give a brief introduction to the MOEA/D framework which is the basis of the investigated methods, and then describe the critical process of MOBS.

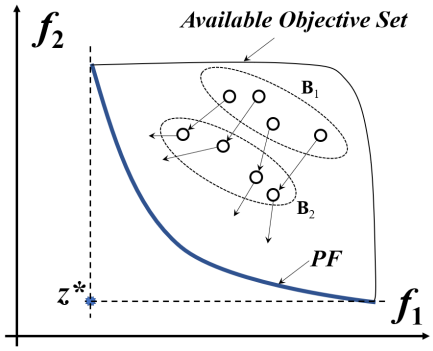


Fig. 1: An illustration about MOEA/D. $\mathbf{B}_1$ and $\mathbf{B}_2$ are populations in contiguous iterations. The curve $\widehat{\mathbf{PF}}$ is the Pareto (optimal) front.

A. Background of MOEA/D

An MO problem can be described by the following expression:

$$\min_{\boldsymbol{x}} \mathcal{F}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_{on}(\boldsymbol{x})], \tag{1}$$

where $f_i(\boldsymbol{x})$ is one of the objectives, and on is the number of objectives. Usually, the objectives in Eq. (1) are mutually conflicting, which means there is not a single solution that can minimize all these objectives simultaneously [34], [35]. This situation results in that the solution(s) for each MO problem is a set known as Pareto optimal. Compared with single objective based methods (such as using regularization terms), a significant property of MO methods is that they will generate solution sets.

MOEA/D [28] was widely employed to optimize MO problems [35] since it was proposed in 2007. Recent MO based band selection works were also developed based on MOEA/D [26], [27]. Therefore this paper focuses on the discussion about MOEA/D.

The basic idea of MOEA/D is the decomposition strategy, *i.e.*, decomposing an MO problem into a number of scalar optimization subproblems and optimizes them in a collaborative manner [35]. Fig. 1 is an illustration about MOEA/D. MOEA/D begins with an initialized solution set called *population*, and each solution in the population is called an *individual*. Then the population is iteratively updated using an evolutionary algorithm. The population tends to $\boldsymbol{z}^*$ which is a virtual ideal point. Note that during each iteration $\boldsymbol{z}^*$ will change according to the current population. After sufficient iterations the population will generally approximate the curve $\widehat{\mathbf{PF}}$ which is called *Pareto front* (PF). If $\widehat{\mathbf{PF}}$ contains the right solutions, it is called Pareto optimal front.

Based on MOEA/D, various of HSI band selection methods could be developed. Generally, these methods include two relevant parts: objectives and the optimization process.

B. Objectives

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_L]^T \in \mathbb{R}^{L \times M}$ denote the HSI data with L bands and M pixels. Define $\boldsymbol{b} = [b_1, b_2, \cdots, b_L]$ as a binary vector representation for $\boldsymbol{X}$, where $b_k = 1$ if the k-th band is selected and $b_k = 0$ otherwise. One of the advantages of MOBS methods is that they provide a feasible manner to combine several preferred objectives.

However, the selection of objectives will not affect the occurrence of weakly Pareto optimum. Thus for a fair comparison, in this paper we use the same objectives as [26]:

$$\begin{aligned} &\min_{\boldsymbol{b}} \mathcal{F}(\boldsymbol{b}) = [f_1(\boldsymbol{b}), f_2(\boldsymbol{b})] \\ &\begin{cases} f_1(\boldsymbol{b}) = \|\boldsymbol{b}\|_0 \\ f_2(\boldsymbol{b}) = \dfrac{1}{|\boldsymbol{b}|} \displaystyle\sum_{k=1}^{|\boldsymbol{b}|} \dfrac{1}{H(\boldsymbol{x}_{\boldsymbol{b}}^k)} \end{cases} \end{aligned} \tag{2}$$

where $H(\boldsymbol{x}_{\boldsymbol{b}}^k)$ is the information entropy of the k-th band. MOBS methods try to optimize the two objectives $f_1(\cdot)$ and $f_2(\cdot)$ simultaneously. It is worth noting that the problem described by $f_1(\cdot)$ is non-convex and NP-hard. Usually, convex relaxation or greedy algorithms are used for L0 problem. However, these methods cannot guarantee that the obtained solution is optimal. By comparison, MOBS methods can directly handle L0 problem without any relaxation [26], [27], [32]. The objectives used here are available choices, not the only. Users can simply replace or add new objectives for different tasks. The choice of objectives has no influence on our discussion about optimization process.

C. Optimization Process

The optimization process of MOBS methods can be described by 3 steps: (1) Initialize a solution set; (2) Calculate an ideal point according to current solutions; (3) Update the solutions based on a decomposition approach and an evolutionary strategy. During each iteration the solution set is updated. The iterations stop when the solutions keep stable. The final result is the Pareto (optimal) front which is a solution set.

1) Initialization: MOBS methods begin with an initial solution set $\boldsymbol{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_p\}$, *i.e.*, population, and p is the population size. As a subset selection problem, the individuals are usually set as zero vectors so as to get to the real solution as close as possible. After several times update, individuals in $\boldsymbol{B}$ will be diverse.

2) The Ideal Point: MOBS methods try to make the population become "better" during the iteration process, thus there should be an ideal point to guide the movement of the population. An ideal point $\boldsymbol{z}^*$ is often defined by

$$\begin{aligned} &\boldsymbol{z}^* = (z_1^*, z_2^*) \\ &\begin{cases} z_1^* = \min f_1(\boldsymbol{B}) \\ z_2^* = \min f_2(\boldsymbol{B}) \end{cases} \end{aligned} \tag{3}$$

where z_1^* is the minimum $f_1(\cdot)$ value of all the individuals, and so as z_2^*. Obviously, $\boldsymbol{z}^*$ is usually a virtual point. The target of MOBS is forcing all the individuals to get closer to the ideal point. It is worth noting that because the population keeps updating during iterations, the ideal point changes in each iteration.

3) Update: Based on the current ideal point and population, there must be a criterion which could guide the individuals approach the ideal point. In MOEA/D, the distance between the population and the ideal point is used as the criterion, where closer distance corresponds to better solution. To remove the influence by the different measurement units of the objectives, decomposition based approaches are utilized to determine the distance. In GMOBS [26] and RMOBS [27], weighted Tchebycheff decomposition was selected. The decomposition process is the key factor for the weakly Pareto optimal problem, which will be further discussed in the next section.

Based on the distance measurement by the weighted Tchebycheff decomposition, the current population will move towards the ideal point. Usually, evolutionary methods such as random flapping [26], [27] are adopted to update the population. After sufficient iterations the population will stabilize in the marginal part of the range, which is the approximate Pareto (optimal) front.

III. Analysis and Discussion

Recent papers about MOBS tended to use weighted Tchebycheff (WT) decomposition for distance measurement. In this section, we first discuss the weakly Pareto optimal problem in this approach. Besides, literature [28] suggested that boundary intersection (BI) approach can also be used for decomposition and may present better performance. In this paper we improve the MOEA/D framework by BI and further analyse its limitations. Finally, we propose a new decomposition method, adaptive-penalty based boundary intersection (APBI), to avoid the weakly Pareto optimal problem.

A. Weighted Tchebycheff Approach

Decomposition is used to convert the problem of approximating the PF into a number of scalar optimization problems [28]. For each scalar optimization problem, the solution is a point in the PF. WT decomposition measures the distance between ideal point $\boldsymbol{z}^*$ and the i-th individual $\boldsymbol{b}_i$ by

$$g_i^{\mathrm{wt}}(\boldsymbol{b}_i|\boldsymbol{\lambda}_i, \boldsymbol{z}^*) = \max_{1\leq j\leq 2}\{\lambda_i^j|f_j(\boldsymbol{b}_i) - z_j^*|\}, \tag{4}$$

where $\boldsymbol{\lambda}_i = [\lambda_i^2, \lambda_i^1]^{\mathrm{T}}$ is the weights for the i-th individual. $\boldsymbol{\lambda}_i$ is a uniformly distributed direction vector with $(\lambda_i^2 + \lambda_i^1) = 1$ is assigned. Fig. 2 is an illustration for the setting of $\boldsymbol{\lambda}$. In the initialization process, the i-th individual is assigned the $\boldsymbol{\lambda}_i$. Fig. 2 uses 11 individuals and two objectives for example. The target of using different $\boldsymbol{\lambda}$ is to guide the individuals approach different optimal directions, and each direction corresponds to a certain scalar optimization problem.

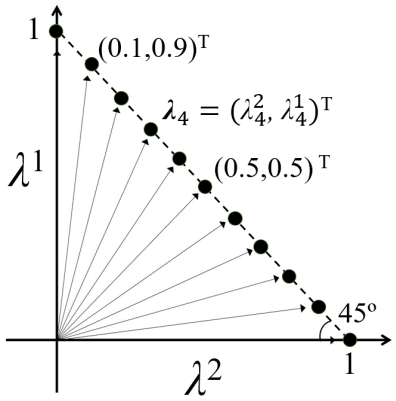


Fig. 2: An illustration about how to determine $\boldsymbol{\lambda}$. Here we set the population size p as 11 for example.

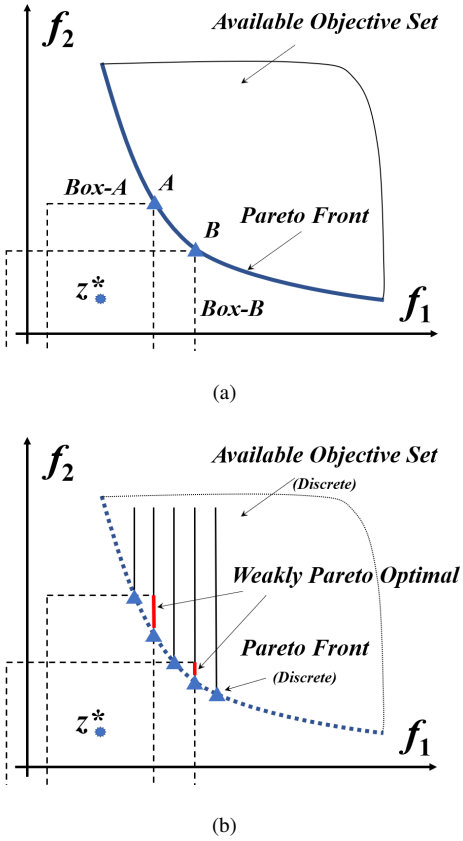
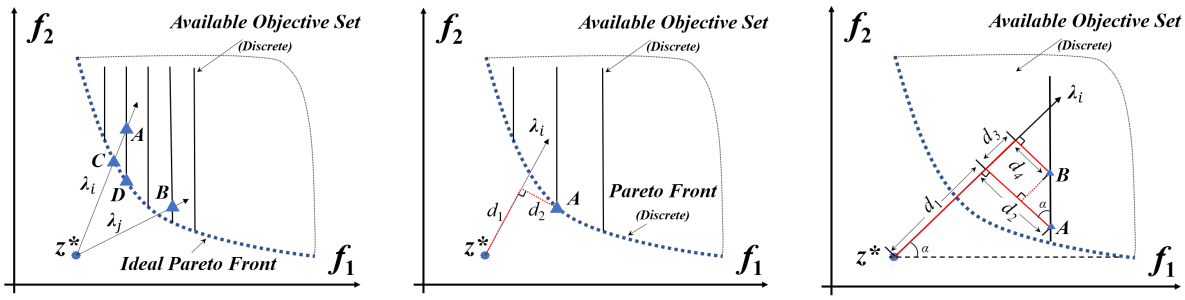


Fig. 3: Illustrations about the Pareto front by WT approach on (a) continuous range and (b) discrete range.

The weights $\boldsymbol{\lambda}$ for different individuals are the key factor that gives rise to the obtained solution set, *i.e.*, PF. However, according to Eq. (4), we find that there are many contour lines surrounding $\boldsymbol{z}^*$. Actually, each $\boldsymbol{\lambda}_i$ corresponds to a certain rectangular contour line, as shown in Fig. 3(a). In Fig. 3(a) the range (available objective set) is continuous and the intersection points A and B are two solutions in the PF. Points in Box-A (contour line) present the same distance value, according to Eq. (4). Obviously, for each $\boldsymbol{\lambda}_i$, smaller box corresponds to more optimal solutions. Assume that current PF is the Pareto optimal front. If we slightly shrink Box-A, the intersection will disappear. Although points in the shrunk box look more optimal, they have been already beyond the range. On the other hand, enlarging Box-A will generate bigger distance value. Thus current Box-A is optimal for weight $\boldsymbol{\lambda}_A$. The same is true for Box-B.

However, the situation will change in band selection problem. According to Eq. 2, the range of MOBS is discrete, or rather, many parallel lines, as shown in Fig. 3(b). Due to the manner of setting $\boldsymbol{\lambda}$, a contour line and the PF are almost impossible to exactly intersect in a single point. A more common situation is that the contour line has some overlap with the range, as displayed by the red lines in Fig. 3(b). According to Eq. (4), the points in the same red line have the same distance value. Therefore, they are equally optimal. This is just the so-called weakly Pareto optimal phenomenon. Due to the discrete characteristics of the MOBS' range, WT decomposition is bound



(a) (b) (c)

Fig. 4: BI is not suitable for discrete range problems. Although PBI can overcome this difficulty, it will also suffer weakly Pareto optimum in some conditions. (a) Illustrations for BI and (b) PBI approaches. (c) An example of weakly Pareto optimum in PBI.

to generate weakly Pareto optimal.

The reason can be described as follows. Assume that the population has reached Pareto optimal front, $(f_1(\boldsymbol{b}_i),f_2(\boldsymbol{b}_i))$ is the top-right point in the i-th box with $\boldsymbol{\lambda}_i = [\lambda_i^2, \lambda_i^1]^T$, (z_1, z_2) is the ideal point, the weakly Pareto optimal solutions will not appear only if the following constraint is met:

$$\begin{aligned} & \lambda_i^1(f_1(\boldsymbol{b}_i) - z_1) = \lambda_i^2(f_2(\boldsymbol{b}_i) - z_2), \\ \Rightarrow \quad & \frac{f_2(\boldsymbol{b}_i) - z_2}{f_1(\boldsymbol{b}_i) - z_1} = \frac{\lambda_i^1}{\lambda_i^2} \\ & = (1 - \frac{i-1}{p-1})/(\frac{i-1}{p-1}) \\ & = \frac{p-i}{i-1}, \quad i \in \{2, 3, \cdots, p\}. \end{aligned} \tag{5}$$

This is an equality constraint. In a continuous range problem such as Fig. 3(a), Eq. (5) is easy to meet. However, when the range is discrete such as Fig. 3(b), it is hardly to guarantee that the point $(f_1(\boldsymbol{b}_i),f_2(\boldsymbol{b}_i))$ is exactly located in the Pareto optimal front. Actually this is an *impossible event* according to geometric probability. Therefore, we conclude that weakly Pareto optimal solutions are bound to appear in weighted Tchebycheff based MOBS methods. Note that Eq. (5) does not require to assume that the solutions in Pareto optimal front are unique. To some extent, the points in Fig. 3 have three elements: $f_1(\boldsymbol{b}_i)$, $f_2(\boldsymbol{b}_i)$ and $\boldsymbol{\lambda}_i$. Even though two individuals were overlapped in $f_1(\cdot)$ and $f_2(\cdot)$, they cannot share the same λ value.

We further quantitatively discuss the degree of weakly Pareto optimum. Let $f_2^{\text{int}}(\boldsymbol{b}_i)$ denote the vertical coordinate value of the i-th intersection point. The length of weakly Pareto optimal solutions for the i-th individual can be obtained by

$$\begin{aligned} L_{wp}(i) &= f_2(\boldsymbol{b}_i) - f_2^{int}(\boldsymbol{b}_i) \\ &= \frac{\lambda_i^1}{\lambda_i^2}(f_1(\boldsymbol{b}_i) - z_1) + z_2 - f_2^{int}(\boldsymbol{b}_i) \\ &= \frac{p-i}{i-1}(f_1(\boldsymbol{b}_i) - z_1) + z_2 - f_2^{int}(\boldsymbol{b}_i) \end{aligned} \tag{6}$$

The monotonicity of Eq. (6) is an important factor for our discussion. Determining its monotonicity directly is

relatively difficult. Fortunately, due to the characteristics of MOBS problem, Eq. (6) can be simplified. $f_1(\cdot)$ is the sparsity with range $\mathbb{N}$. When i is not a large value, we may consider $f_1(\boldsymbol{b}_i) \approx i$. z_1 is not a variable, according to Eq. (3), z_1 could be approximated as 1. Therefore Eq. (6) could be approximated by

$$L_{\mathrm{wp}}(i) = (p-i)M + z_2 - f_2^{\mathrm{int}}(\boldsymbol{b}_i) \tag{7}$$

where $M \approx 1$ is considered as a constant. Obviously, the revised $L_{\mathrm{wp}}(i)$ is monotonous decreasing. When $i \to p$, $f_2^{\mathrm{int}}(\boldsymbol{b}_i)$ will tend to z_2 which results in $\lim_{i=p} L_{\mathrm{wp}}(i) = 0$.

Strictly speaking, the final solutions by WT is not the Pareto optimal front. Pareto optimal front contains solutions that cannot be dominated by any others [36]. Obviously, there are no dominant solutions in many situations. Based on the above analysis we can safely draw the following conclusions:

- As long as there are solutions in the overlapped regions (red lines in Fig. 3(b) with length $L_{\mathrm{wp}}(i)$), weighted Tchebycheff decomposition based MOBS methods will certainly generate weakly Pareto optimal solutions.
- The weakly Pareto optimal solutions mainly distribute in the low-sparsity regions. With the increasing of sparsity, the influence weakens.

Because WT decomposition is the most widely adopted strategy in recent MOBS methods [26], [27], there should be more studies on its properties and disadvantages.

There may be some doubt that why not directly use the non-dominated sorting method to find the dominant solutions from the weakly Pareto set. We do not select this approach because the ideas between non-dominated sorting and decomposition-based multi-objective optimization are conflicting. Non-dominated sorting can generate optimal solution set in some Pareto-optimization based multi-objective frameworks such as NSGA-II. Because of the computational cost of this kind of methods, in this paper we study a decomposition-based framework, MOEA/D. In MOEA/D the solution set is generated according to a distance measure, and theoretically non-dominated sorting and distance may not appear in a single method. Combining non-dominated sorting to MOEA/D may harm its preciseness.

B. Boundary Intersection Approach

In [28], Zhang *et al.* suggested that boundary intersection (BI) and its variation penalty-based boundary intersection (PBI) are another kinds of available composition methods, and they present several advantages over WT approach. However, in this section we will prove that even improving MOBS by BI and PBI the weakly Pareto optimal problem still exists.

Firstly, we air the conclusion that BI cannot be directly used for MOBS. BI decomposition is also a strategy to measure the distance between ideal point and individuals, which is defined by

$$\begin{aligned} & g_i^{\mathrm{bi}}(\boldsymbol{b}_i|\boldsymbol{\lambda}_i, \boldsymbol{z}^*) = d, \\ s.t. \quad & \boldsymbol{z}^* - \mathcal{F}(\boldsymbol{b}_i) = d\boldsymbol{\lambda}_i^T \end{aligned} \tag{8}$$

$g_i^{bi}(\cdot)$ is the distance which should be minimized. However, Eq. (8) is also an equality constraint, which means it only works in continuous MO problems. Fig. 4(a) is a visual explanation. In the direction $\boldsymbol{\lambda}_i$, point C is the ideal

solution which is located in the Pareto optimal front. However, because the range is discrete, C does not exist. The real intersection point is A, *i.e.*, A is located in the obtained Pareto front. Obviously, this PF is not optimal. More seriously, the point A may even not exist: the lines in the range are also discrete. Still, according to geometric probability, the existence of intersection points is an impossible event. Overall, BI approach has no solution for MOBS because of the equality constraint and the discrete range.

An improvement for BI is using a penalty method to deal with the equality constraint, which leads to penalty-based boundary intersection approach. The distance by PBI is defined as follows

$$g_i^{\text{pbi}}(\boldsymbol{b}_i|\boldsymbol{\lambda}_i, \boldsymbol{z}^*) = d_1 + \theta d_2, \tag{9}$$

where

$$\begin{aligned} d_1 &= \frac{\|(\mathcal{F}(\boldsymbol{b}_i) - \boldsymbol{z}^*)\boldsymbol{\lambda}_i\|}{\|\boldsymbol{\lambda}_i\|}, \\ d_2 &= \|\mathcal{F}(\boldsymbol{b}_i) - (\boldsymbol{z}^* + d_1\boldsymbol{\lambda}_i^T)\|. \end{aligned} \tag{10}$$

$\theta > 0$ is a preset penalty parameter. Minimizing Eq. (9) is actually the sub-problem for individual $\mathbf{b}_i$. The current population get close to $\mathbf{z}^*$ based on an evolution strategy according to the evaluation results by Eq. (9) and Eq. (10). Fig. 4(b) displays an illustration for PBI. d_1 can be considered as the projection length of an individual on the direction $\boldsymbol{\lambda}_i$, and d_2 is the vertical distance.

PBI seems able to address the discrete range problems well, however, we think that there are at least three disadvantages that PBI will suffer:

1) Weakly Pareto optimum still exists in some conditions: We take Fig. 4 for example. The distance between points A and $\boldsymbol{z}^*$ can be calculated by

$$d_A = d_1 + \theta d_2, \tag{11}$$

while that between points B and $\boldsymbol{z}^*$ is

$$\begin{aligned} d_B &= d_1 + d_3 + \theta d_4 \\ &= d_1 + d_3 + \theta(d_2 - \frac{d_3}{\tan\alpha}) \\ &= d_1 + \theta d_2 + d_3(1 - \frac{\lambda_i^2\theta}{\lambda_i^1}). \end{aligned} \tag{12}$$

d_A will be equal to d_B as long as

$$\begin{aligned} & d_3(1 - \frac{\lambda_i^2\theta}{\lambda_i^1}) = 0, \\ \Rightarrow \quad & \frac{\lambda_i^1}{\lambda_i^2} = \theta. \end{aligned} \tag{13}$$

Eq. (13) demonstrates that the penalty parameter is correlative with $\boldsymbol{\lambda}$. Because $\boldsymbol{\lambda}$ is set by uniform scale (see Fig. 2, usually hundreds of individuals), it is very likely to meet Eq. (13) and thus generate weakly Pareto optimum. Overall, PBI can avoid weakly Pareto optimum only if Eq. (13) has been found.

2) The value of penalty factor has to be manually set: The penalty factor θ could be considered as a weight for d_1 and d_2. Obviously, too large or too small penalty factor will worsen the performance of a penalty method, and this disadvantage is also presented in the original MOEA/D work by Zhang *et al.* [28]. Please note that in MOBS problem setting $\theta = 0$ does not mean PBI is equal to BI. Under this condition PBI only calculates a projection, and the shorter ones correspond to better solutions. However, for Fig. 4(a) there is no theoretical support to indicate that whether point A or D is better.

3) Fixed θ value will result in different influences on various $\boldsymbol{\lambda}$: This is because d_1 and d_2 can be considered as functions for $\boldsymbol{\lambda}$. We explain this problem via a much simplified situation, as shown in Fig. 5. The PF is simplified to a line segment with $\pi/4$ angle, and assume $d_2^A = d_2^B$. Then the distances in $\boldsymbol{\lambda}_i$ and $\boldsymbol{\lambda}_j$ can be calculated by

$$\begin{aligned} g_i^{\mathrm{pbi}}(\boldsymbol{b}_A|\boldsymbol{\lambda}_i, \boldsymbol{z}^*) =& d_1^A + \theta d_2^A, \\ g_j^{\mathrm{pbi}}(\boldsymbol{b}_B|\boldsymbol{\lambda}_i, \boldsymbol{z}^*) =& d_1^B + \theta d_2^B \\ =& d_1^B + \theta d_2^A. \end{aligned} \tag{14}$$

With the increase of α (which is determined by $\boldsymbol{\lambda}$), the value of d_1 will change. According to Fig. 5, $d_1(\boldsymbol{\lambda})$ is monotone decreasing when $\alpha \in [0, \pi/4]$ and increasing when $\alpha \in (\pi/4, \pi/2]$. This situation will result in a problem: If θ is fixed, the weights between d_1 and d_2 will be changed by $\boldsymbol{\lambda}$. To be specific, d_2 will have larger weight when $\alpha = \pi/4$ while smaller weight when $\alpha = 0$. This is not what we expect. We hope that no matter how $\boldsymbol{\lambda}$ changes the weights between d_1 and d_2 should be stable. Absolutely, fixed θ cannot meet this requirement. Here we assume the PF is a line segment with $\pi/4$ angle, so d_1 is symmetrical on $\pi/4$. In practice, selecting the minimum Euclidean distance point (point of contact) is a better choice.

Please note the real Pareto Front is usually not strictly symmetrical and the quantitative expression is unknown, and thus Fig. 5 is just a simplification for the real Pareto Front. Although this simplification may not completely accord with the real situation, their variation trends are similar: the monotonicity of d_1 changes on a certain point.

We also conclude the drawbacks of BI/PBI for the problem of MOBS:

- BI cannot directly be utilized for MOBS because there is no available solution.
- PBI will still suffer weakly Pareto optimal without reasonable setting of penalty factor.
- Manually setting the penalty factor as a fixed value is not appropriate, and such setting would introduce weights deviation.

C. Suggestions and APBI

As is discussed above, weakly Pareto optimum will occur in both WT and PBI decompositions. However, since we have concluded the theoretical reasons for weakly Pareto optimum, now we are able to develop new methods to avoid this problem. Furthermore, the new methods should also overcome the other drawbacks of PBI. Here, we give some suggestions about how to develop new decomposition approaches for MOBS:

- PBI presents more potential than WT to handle the weakly Pareto optimum problem. It is better to select PBI as the basis.

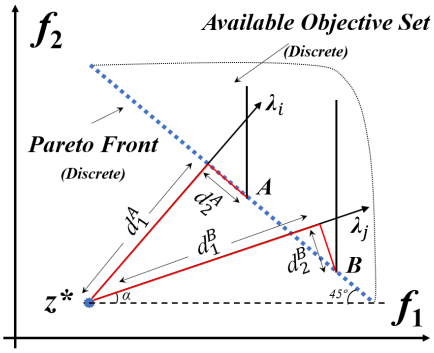


Fig. 5: A simplified illustration about the influence of θ on different $\boldsymbol{\lambda}$.

- The setting of θ and $\boldsymbol{\lambda}$ should avoid the relationship in Eq. (13).
- θ should be adaptive rather than fixed. And it should guarantee that d_1 and d_2 have nearly stable weights.

Obviously, there are many strategies to satisfy the above three criterions. In this paper, we propose an available approach, adaptive-penalty based boundary intersection, to overcome the difficulties in MOBS problem.

The key point of APBI is that the penalty factor is represented by a function of $\boldsymbol{\lambda}$, denoted by $\theta(\alpha)$ where $\alpha = \arctan(\lambda^1/\lambda^2)$. According to Section III(B)-3, $\theta(\alpha)$ should present different monotony on the two sides of a certain point. Here we use parabola function to model $\theta(\alpha)$. To simplify the problem, we still assume that the PF is a line segment and symmetrical on $\lambda_1 = \lambda_2$. Because

$$d_1^{\lambda^1=0} = d_1^{\lambda^2=0} = \sqrt{2} d_1^{\lambda^1=\lambda^2}, \tag{15}$$

we can consider that $\theta = 1$ when $\lambda^1 = 0$ or $\lambda^2 = 0$, and $\theta = 1/\sqrt{2}$ when $\lambda^1 = \lambda^2$. In this case, the parabola will pass point (0,1) and centered on $(\pi/4, 1/\sqrt{2})$. Then the parabola $\theta(\alpha)$ can be expressed by

$$\begin{aligned} \theta(\arctan(\frac{\lambda^1}{\lambda^2})) =& \frac{8(2-\sqrt{2})}{\pi^2}(\arctan(\frac{\lambda^1}{\lambda^2}) - \frac{\pi}{4})^2 + \frac{\sqrt{2}}{2} \\ \approx& 0.47(\arctan(\frac{\lambda^1}{\lambda^2} - 0.78)^2 + 0.71, \end{aligned} \tag{16}$$

and the integrated APBI based decomposition equation is obtained as follows:

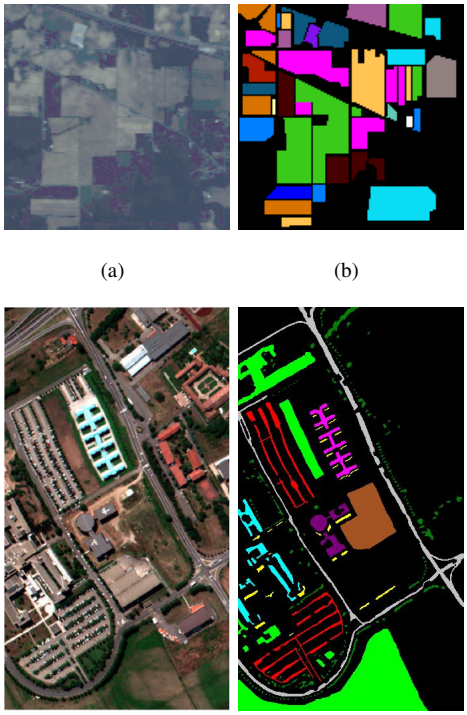
$$g_i^{\text{apbi}}(\boldsymbol{b}_i|\boldsymbol{\lambda}_i, \boldsymbol{z}^*) = d_1 + \theta(\arctan(\frac{\lambda_i^1}{\lambda_i^2}))d_2. \tag{17}$$

Eq. (17) can be substituted by any functions with similar shapes. If using a fixed θ instead, APBI degrades into PBI.

IV. Experiments

A. Setups

Because the weakly Pareto optimal problem mainly appears in WT decomposition, and recently proposed MOBS methods are all based on WT [26], [27], in the experiments we show the comparison results by WT. A typical WT based method, GMOBS [26], is used for comparison analysis. Our improved method is termed by AMOBS



(a) (b)

(c) (d)

Fig. 6: validation data sets and the corresponding ground truths. (a) IndianP and (b) its ground truth. (c) PaviaU and (d) its ground truth. Different colors represent different classes.

for short. Because the code for GMOBS is not available, we reproduce this method. Parameters in GMOBS and AMOBS are set equal as much as possible. The experiments consist of three parts:

- Validate that weakly Pareto optimal solutions really exist in GMOBS, and disappear after our improvements.
- Validate that the solutions obtained by GMOBS and AMOBS are highly similar, which means they have found very close PFs.
- Use the classification experiment as an example to validate that both GMOBS and AMOBS can improve the accuracies. Several recently proposed band selection methods are also used for comparison [14], [27].

Please note that the target of these experiments is to validate our above theoretical analysis, rather than verify that the selected bands by AMOBS is more beneficial for classification. We assume that GMOBS has already found the representative bands, but not unique. AMOBS tries to find the solution with the best theoretical interpretability. Therefore AMOBS is not necessary to outperform GMOBS in the classification accuracies. Actually because most selected bands by AMOBS and GMOBS are the same, their accuracies are also very close. In a word, this paper focuses on the discussion about weakly Pareto optimum, and it has little relationship with classification task.

B. Data sets

As validation experiments, as long as the weakly Pareto optimal phenomenon can be observed in a single data set, we can conclude that this phenomenon really exists. Therefore, it is not necessary to report the results on too many data sets. Here we only use two popular data sets which were captured by different sensors, Indian Pines(IndianP) and Pavia University (PaviaU)[1], as examples. They have been widely studied in recent hyperspectral papers [37]–[39].

- IndianP was gathered by airborne visible/infrared imaging spectrometer (AVIRIS) in Northwestern Indiana. It consists of 200 available bands after removing water absorption regions. Totally 145×145 pixels are included with 20m spatial resolution, and the wavelengths range from 0.4 to 2.5 μm. 10249 pixels are labeled and classified into 16 different classes. A false color composite image (R-G-B=bands 36-17-11) and the corresponding ground truth are shown in Fig. 6(a)-(b).
- PaviaU was acquired by reflective optics system imaging spectrometer (ROSIS-3) sensor over Pavia, northern Italy. There are 103 spectral bands and 42776 labeled pixels. It covers 610×340 pixels region with 1.3m spatial resolution. A false color composite image (R-G-B=bands 10-27-46) and the corresponding ground truth image are shown in Fig. 6(c)-(d).

TABLE I: Some selected bands on IndianP data set.

Bands Number	4	6	8	10
GMOBS	① 29 30 31 115 ② 19 35 87 118	① 28 29 30 31 44 115 ② 20 26 31 41 51 116 ③15 33 41 52 113 114 ④19 30 35 46 87 118	①20 27 28 29 30 31 34 116 ②20 28 30 31 34 41 44 51 ③15 22 33 41 52 61 113 114	①20 27 28 29 30 31 33 34 49 116 ②20 28 29 30 31 33 34 44 51 116 ③20 28 30 31 34 35 41 44 51 116 ④19 22 35 37 42 46 50 116 118 125
AMOBS	28 29 34 118	28 29 30 31 32 34	27 28 29 30 31 32 42 117	27 28 29 30 31 32 33 42 116 117

C. Weakly Pareto Optimal

This section is used to validate that weakly Pareto optimum really exists. Results on the two data sets are shown in Fig. 7 and Table I-II. Since $f_1(\cdot)$ and $f_2(\cdot)$ are error values of Eq. (2), there are no units in this figure. The normalization of objectives will not significantly affect the penalty factor. Different normalization would make the final PF steeper ore smoother, but its general shape will not change. In our experiments we do not conduct normalization. The circles in Fig. 7 denote current solutions, and ①-④ are different weak solutions. Because the solutions tend stable after sufficient iterations, current solution sets could be considered as Pareto optimal front.

[1]Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

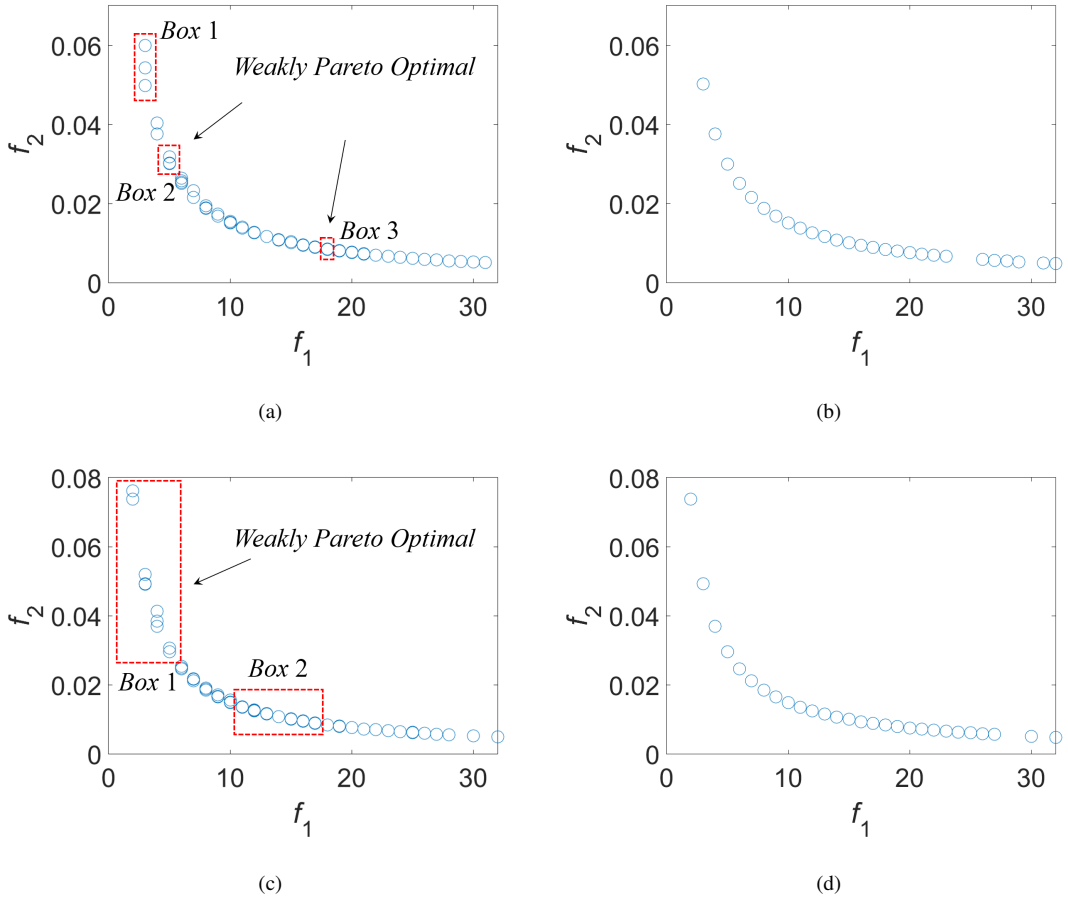


Fig. 7: Band selection results by GMOBS on (a) IndianP, (c) PaviaU, and AMOBS on (b) IndianP, (d) PaviaU. The circles denote the solutions, and the red boxes are some weakly Pareto optimal solutions.

TABLE II: Some selected bands on PaviaU data set.

Bands Number	4	6	8	10
GMOBS	①90 91 94 96 ②90 91 96 97 ③62 73 90 96 ④10 15 33 64	①88 90 91 92 96 97 ②77 79 82 90 96 100 ③55 72 88 96 98 100	①89 90 91 92 96 97 99 100 ②77 79 82 84 90 92 96 100 ③42 51 59 62 82 90 94 96	①77 79 83 90 92 93 95 96 100 101 ②53 77 79 82 83 87 88 90 96 100 ③28 30 37 46 71 75 80 91 95 96
AMOBS	88 89 93 101	88 89 93 96 101 103	86 89 90 92 93 95 101 103	86 87 89 90 92 93 96 98 101 103

The initial population size p is set as 101 for both GMOBS and AMOBS. We conduct 200 iterations for IndianP, and 40 for PaviaU. PaviaU requires less iterations because it only consists of 103 bands.

1) Results on IndianP: As we predicted in Section III(A), weakly Pareto optimal solutions appear in WT based decomposition methods. Fig. 7(a) is the result of GMOBS. It can be observed that the weakly Pareto optimal phenomenon is obvious especially when the selected bands are not many. Take Box 1 for example. Although the bottom one looks better, according to the theory of decomposition based multi-objective optimization [24], [28], the solutions in Box 1 are equally optimal. In this case, it is not appropriate to simply regard the bottom solution as the best one. With the increasing of bands number, the weakly Pareto optimum weakens. This result also confirms our conjecture. However, it is worth noting that although some solutions seem overlapped (such as Box 3), the overlapped solutions correspond to different band selection results, *i.e.*, Box 3 is still weakly Pareto optimal. By comparison, AMOBS can avoid the weakly Pareto optimum, as shown in Fig. 7(b). There is a unique solution for each bands number.

Table I displays part of the quantitative results. GMOBS presents weakly Pareto optimal solutions with 4-10 selected bands. A noticeable tendency is that the weakly Pareto solutions become more similar with the increasing of bands number. This is consistent with Fig. 7(a). The weakly Pareto solutions obtained by GMOBS look similar and present strong relevance. This is normal because they are all close to the real optimal one.

2) Results on PaviaU: Fig. 7(c) shows the results of GMOBS on PaviaU data. The variation tendency in this figure is very similar to that in Fig. 7(a). Because of the overlap between contour lines and the Pareto Front (theoretical explanation is illustrated in Fig. 3), weakly Pareto solutions can be observed. The length of weakly Pareto optimal region keeps shrinking due to the monotonicity of λ_1. AMOBS achieves better results as shown in Fig. 7(d). We note that the solution set is not integrated in Fig. 7(b) and Fig. 7(d). Actually this is inevitable both in GMOBS and AMOBS. The reason is that $\boldsymbol{\lambda}$ is equal-space sampled and the maximum number of solutions is fixed. According to the analysis in Section III, there will be at most p solutions in PF. Therefore when the PF become smooth, missed solutions appear. However, the weakly Pareto solutions are usually observed in the steep part of PF where missing solutions hardly happens. So in this paper we do not pay special attention to this problem.

Table II is the quantitative results for bands number 4-10. We find that although the selected bands by GMOBS and AMOBS are not completely the same, they present very similar results. Since WT and APBI are two different decomposition approaches, their results are not necessarily to be equal. However, we still hope that they should be similar, because the "ideal" situation is that they both find the global optimal solution. A drawback for GMOBS is that although all the weak solutions are considered as equally optimal, sometimes they present very significant difference. For example, when bands number is 4, the ④ solution is completely different from others. It is difficult to judge which one is optimal. This problem does not exist in AMOBS, and this is the major advantage of AMOBS and also our initial target.

3) Computational cost: All the experiments are conducted on a desktop with i7-6700K, 4.2Ghz and 16G RAM, Matlab 2015a. Table III reports the computational time of GMOBS and AMOBS. AMOBS is a little faster than GMOBS because the APBI decomposition is slightly more efficient than WT. However, since the two methods

TABLE III: Computational time on the two data sets.

Data/iterations	GMOBS	AMOBS
IndianP/200	106.47s	92.16s
PaviaU/40	102.44s	84.63s

are both built under the framework of MOEA/D, the gaps are not significant. PaviaU costs more time in a single iteration, while the total time is close to that of IndianP. This is because PaviaU has fewer bands but larger size, when compared with IndianP. We have to emphasize that although $\sim$100s seem time-consuming, the real time cost is not much. Both GMOBS and AMOBS can generate a series of solutions in a single running, *i.e.*, after $\sim$100s running we can obtain all the results for different bands numbers. Suppose we need the results on bands number 5-40, AMOBS only requires $<$3s for each band number. Overall, the efficiency is not the disadvantage of AMOBS.

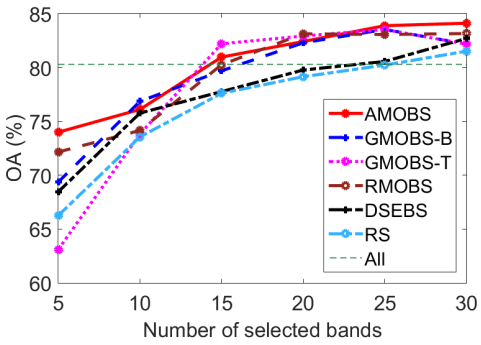
D. Classification Validation

We use classification experiments to verify that the selected bands are representative. However, AMOBS is not specially designed for classification task, and its objectives are very general. Moreover, there are many other factors besides the selected bands that may affect the classification results, which are not related to our research. Therefore, we do not expect that AMOBS can outperform other methods in accuracies. Instead, we hope to validate the following assumptions proposed in our above discussion:

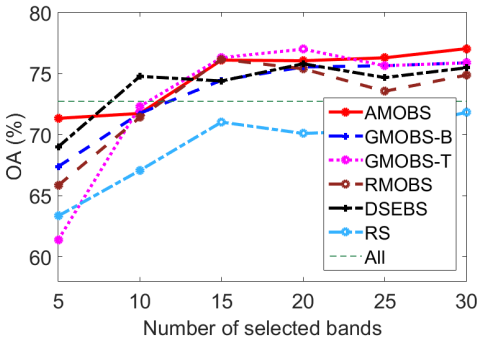
- First and foremost, we should verify that all the weakly Pareto solutions (*e.g.*, solutions in Fig. 7(a), Box 1) are effective, in which case determining a unique solution from the weakly Pareto optimal set deserves studying.
- At least, the band selection methods should outperform random selection.
- AMOBS can achieve comparative performance as some state-of-the-art band selection methods.

Based on the above three criterions, we design the classification experiments as shown in Fig. 8. Three recently proposed methods, GMOBS [26], RMOBS [27] and DSEBS [14], are used for comparison. The results by random selection (RS) and using all bands (All) are compared as well. Specially, GMOBS is divided into two manners: GMOBS-B and GMOBS-T. GMOBS-B means using the bottom solution in weakly Pareto optimal set, and GMOBS-T means using the top one. According to the WT theory, they are equally optimal. For IndianP data, the training and testing samples numbers are set as 1:4. For PaviaU data, we use 1% of all labeled samples for training and the rest for testing.

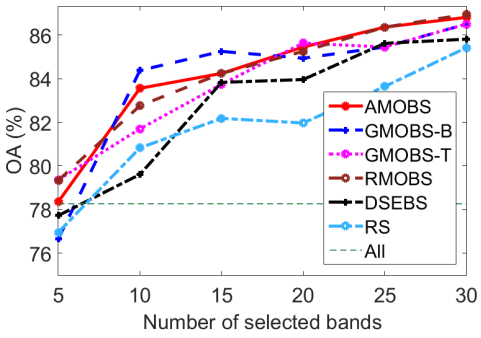
Fig. 8(a)(c) are the results by support vector machine (SVM) [40], [41] on IndianP and PaviaU data sets, and Fig. 8(b)(d) are these by k-Nearest Neighbor (KNN) on the two data sets. It can be observed that all the methods outperform RS in most conditions, which indicate that band selection is effective. The four MO based methods (AMOBS, GMOBS-B, GMOBS-T and RMOBS) present very close performance. This is not strange because they have utilized the same framework. The obtained Pareto optimal fronts do not differ a lot. We observe that GMOBS-B and GMOBS-T present comparative results. When more bands are selected, GMOBS-B and GMOBS-T present



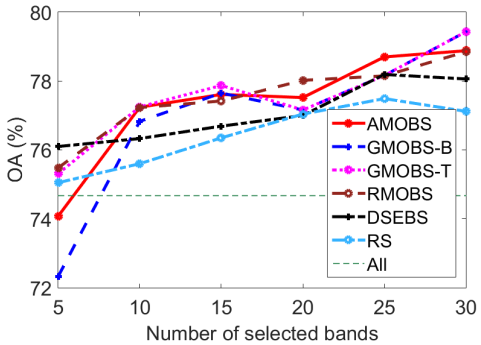
(a)



(b)



(c)



(d)

Fig. 8: Classification experiments. From top to the bottom are the results by (a) SVM on IndianP, (b) KNN on IndianP, (c) SVM on PaviaU and (d) KNN on PaviaU.

the same results. This is because weakly Pareto optimal solutions disappear, according to the characteristics of WT approach. Although RMOBS is also based on multi-objective framework, it can only obtain one solution in a single running. Therefore RMOBS does not make full use of the advantages of MO-based framework. By comparison, AMOBS not only acts as an effective MO-based band selection method, but provides theoretical explanation about the weakly Pareto optimum.

V. Conclusion

In this paper, we discussed the weakly Pareto optimal problem in multi-objective based HSI band selection. The characteristics of two popular decomposition approaches, weighted Tchebycheff and boundary intersection, were analysed. In summary, we have drawn the following conclusion: 1) WT approach is bound to generate weakly Pareto optimum, and this situation aggravates with the decreasing of selected bands; 2) BI approach is not suitable for band selection; 3) Improving BI to PBI could partly avoid weakly Pareto optimum (under certain circumstances it still happens), however, this improvement will harm the effectiveness of MO methods in other aspects. Besides these overall conclusions, we also drew several detailed sub-conclusions, as shown in Section III. The boundary conditions are quantitatively described.

Based on the analysis and discussion, we provided some suggestions about how to develop new decomposition approaches which could avoid weakly Pareto optimum. As an example, adaptive-penalty based boundary intersection was proposed. We further used popular HSI data to verify the existence of weakly Pareto optimal solutions, and demonstrated the effectiveness of our improvements. In real scenarios, our suggestions can be directly used to improve multi-objective based band selection methods.

References

[1] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, p. 1279, 2016.

[2] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical feature selection for random projection," *IEEE Transactions on Neural Networks and Learning Systems*, 2018.

[3] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer Science and Business Media, 2003.

[4] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 138–142, Jan 2011.

[5] X. Cao, T. Xiong, and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 329–333, March 2016.

[6] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by gabor filtering based deep network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1166–1178, April 2018.

[7] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "CoinNet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[8] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 6, no. 56, pp. 3173–3184, Jun 2018.

[9] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7111–7119, Nov 2014.

[10] M. G. Asl, M. R. Mobasheri, and B. Mojaradi, "Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 3774–3787, July 2014.

[11] X. Jin, Y. Liang, D. Tian, and F. Zhuang, "Particle swarm optimization using dimension selection methods," *Applied Mathematics and Computation*, vol. 219, no. 10, pp. 5185–5197, 2013.

[12] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 998–1002, May 2015.

[13] C. Sui, Y. Tian, Y. Xu, and Y. Xie, "Unsupervised band selection by integrating the overall accuracy and redundancy," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 185–189, Jan 2015.

[14] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 227–239, Jan 2016.

[15] M. Bevilacqua and Y. Berthoumieu, "Unsupervised hyperspectral band selection via multi-feature information-maximization clustering," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 540–544.

[16] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 5087–5098, Nov 2017.

[17] S. Pang, T. Li, F. Dai, and M. Yu, "Particle swarm optimization algorithm for multisalesman problem with time and capacity constraints," *Applied Mathematics and Information Sciences*, vol. 7, no. 6, pp. 2439–2444, 2013.

[18] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[19] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1431–1445, March 2016.

[20] F. Z. Q. Wang and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[21] V. S. K. Ganesan and S. Vasuki, "Maximin distance based band selection for endmember extraction in hyperspectral images using simplex growing algorithm," *Multimedia Tools and Applications*, no. 3, pp. 1–17, 2017.

[22] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, June 2006.

[23] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564–568, Oct 2008.

[24] K. Miettinen, *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Dordrecht, 1999, vol. 12.

[25] A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 12, pp. 4175–4188, 2009.

[26] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 544–557, 2016.

[27] X. Xu, Z. Shi, and B. Pan, "A new unsupervised hyperspectral band selection method based on multiobjective optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112–2116, 2017.

[28] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on evolutionary computation*, vol. 11, no. 6, pp. 712–731, 2007.

[29] J. Yu, M. Li, Y. Wang, and G. He, "A decomposition method for large-scale box constrained optimization," *Applied Mathematics and Computation*, vol. 231, no. 12, pp. 9–15, 2014.

[30] M. Zhang, L. Jiao, W. Ma, J. Ma, and M. Gong, "Multi-objective evolutionary fuzzy clustering for image segmentation with moea/d," *Applied Soft Computing*, vol. 48, pp. 621 – 637, 2016.

[31] A. Ma, Y. Zhong, D. He, and L. Zhang, "Multiobjective subpixel land-cover mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 422–435, Jan 2018.

[32] X. Xu, Z. Shi, and B. Pan, "ℓ0-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.

[33] B. Pan, L. Wang, X. Xu, and Z. Shi, "Towards weakly pareto optimal: Animprovedmulti-objective based band selection method for hyperspectral imagery," in *2018 International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018.

[34] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York, NY, USA: Wiley, 2001.

[35] A. Trivedi, D. Srinivasan, K. Sanyal, and A. Ghosh, "A survey of multiobjective evolutionary algorithms based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 3, pp. 440–462, June 2017.

[36] Y. L. Li, Y. R. Zhou, Z. H. Zhan, and J. Zhang, "A primary theoretical study on decomposition-based multiobjective evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 563–576, Aug 2016.

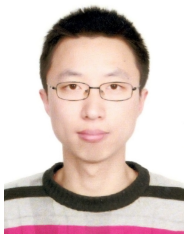
[37] B. Pan, Z. Shi, and X. Xu, "MugNet: Deep learning for hyperspectral image classification using limited samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 108–119, 2018, Deep Learning RS Data.

[38] B. Du, M. Zhang, L. Zhang, R. Hu, and D. Tao, "Pltd: Patch-based low-rank tensor decomposition for hyperspectral images," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 67–79, Jan 2017.

[39] X. Kang, P. Duan, X. Xiang, S. Li, and J. A. Benediktsson, "Detection and correction of mislabeled training samples for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

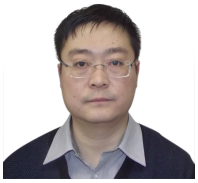
[40] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, Aug 2004.

[41] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247 – 259, 2011.



Bin Pan Bin Pan (S'18) received the B.S. degree from the School of Astronautics, Beihang University, Beijing, China, in 2013. He is currently working toward the Ph.D. degree in the Image Processing Center, School of Astronautics, Beihang University, Beijing China. He also serves as a reviewer of several journals, including *ISPRS Journal of Photogrammetry and Remote Sensing*, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* and *IEEE Geoscience and Remote Sensing Letters*.

His research interests include deep learning and hyperspectral image processing.



Zhenwei Shi Zhenwei Shi (M'13) received his Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was Visiting Scholar in the Department of Electrical Engineering and Computer Science, Northwestern University, U.S.A., from 2013 to 2014. He is currently a professor and the dean of the Image Processing Center, School of Astronautics, Beihang University. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Prof. Shi received the Best Reviewer Awards for his service to *IEEE Transactions on Geoscience and Remote Sensing* (TGRS) and *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (JSTARS) in 2017. He has authored or co-authored over 100 scientific papers in related journals and proceedings, including *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Neural Networks*, *IEEE Transactions on Geoscience and Remote Sensing*, *IEEE Transactions on Image Processing* and *IEEE Conference on Computer Vision and Pattern Recognition*. He also serves as an Associate Editor of *Infrared Physics & Technology* since 2016.

Xia Xu Xia Xu (S'18) received the B.S. And M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015. She is currently working toward the Ph.D. degree in the Image Processing Center, School of Astronautics, Beihang University, Beijing China. She also serves as a reviewer of several journals, including *IEEE Transactions on Geoscience and Remote Sensing*, *IEEE Transactions on Cybernetics* and *Applied Soft Computing*.

Her research interests include hyperspectral unmixing and multi-objective optimization.