Ke Li[1], Álvaro Fialho[2], and Sam Kwong[1]

[1] Department of Computer Science, City University of Hong Kong, Hong
jerryli3@student.cityu.edu.hk, cssamk@cityu.edu.hk
[2] LIX, École Polytechnique, Palaiseau, France
fialho@lix.polytechnique.fr

Abstract. Differential Evolution (DE) is a simple yet powerful evo
tionary algorithm, whose performance highly depends on the setting
some parameters. In this paper, we propose an adaptive DE algorith
for multi-objective optimization problems. Firstly, a novel tree neighb
hood density estimator is proposed to enforce a higher spread betwe
the non-dominated solutions, while the Pareto dominance strength
used to promote a higher convergence to the Pareto front. These t
metrics are then used by an original replacement mechanism based
a three-step comparison procedure; and also to port two existing ad
tive mechanisms to the multi-objective domain, one being used for
autonomous selection of the operators, and the other for the adapt
control of DE parameters CR and F. Experimental results confirm
superior performance of the proposed algorithm, referred to as Ad
MODE, when compared to two state-of-the-art baseline approaches, a
to its static and partially-adaptive variants.

Keywords: Multi-Objective Optimization, Differential Evolution, T
Neighborhood Density, Parameter Control, Adaptive Operator Selecti

1 Introduction

Differential Evolution, proposed by Storn and Price [15], is a popular
cient population-based, direct heuristic for solving global optimization
in continuous search spaces. The main benefits brought by DE are i
structure, ease of use, fast convergence speed and robustness, which
to be widely applied to many real-world applications. For the generatic
solutions (trial vectors), each individual (target vector) is combined wi
by means of different forms of weighted sums (mutation strategies). Orig
case the newly generated solution has a better fitness value than its co
ing parent, it replaces its parent in the population for the next genera
aim of these iterations is basically to find a proper direction for the se
cess towards the optimum, by following the quality distribution of the
in the current population.

Evolutionary Multiobjective Optimization With Hybrid Selection Principles

Ke Li, Kalyanmoy Deb
Department of Electrical and Computer Engineering
Michigan State University
East Lansing, MI 48864, USA
Email: keli.genius@gmail.com, kdeb@egr.msu.edu

Qingfu Zhang
Department of Computer Science
City University of Hong Kong
Tat Chee Avenue, Hong Kong SAR
Email: qingfu.zhang@cityu.edu.hk

Abstract—**Achieving balance between convergence and diversity is a basic issue in evolutionary multiobjective optimization (EMO). In this paper, we propose a hybrid EMO algorithm that assigns different selection principles to two separate and co-evolving archives. Particularly, one archive maintains a repository with a competitive selection pressure towards the Pareto-optimal front (PF), the other preserves a population with a satisfied distribution in the objective space. Furthermore, to exploit guidance information towards the Pareto-optimal set (PS), we develop a restricted mating selection mechanism to select mating parents from each archive for offspring generation. Empirical studies are conducted on a set of benchmark problems with complicated PSs. Experimental results demonstrate the effectiveness and competitiveness of our proposed algorithm in balancing convergence and diversity.**

I. INTRODUCTION

A multiobjective optimization problem (MOP) can be stated as follows:

$$\begin{array}{ll} \text{minimize} & \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_m(\mathbf{x}))^T \\ \text{subject to} & \mathbf{x} \in \mathbf{\Omega} \end{array} \qquad (1)$$

where $\mathbf{\Omega} = \Pi_{i=1}^{n}[a_i, b_i] \subseteq \mathbb{R}^n$ is the decision (variable) space, and $\mathbf{x} = (x_1, \cdots, x_n)^T \in \mathbf{\Omega}$ is a candidate solution. $\mathbf{F} : \mathbf{\Omega} \to \mathbb{R}^m$ constitutes m real-valued objective functions, and $\mathbb{R}^m$ is called the objective space. The attainable objective set is defined as $\Theta = \{\mathbf{F}(\mathbf{x})|\mathbf{x} \in \mathbf{\Omega}\}$. $\mathbf{x}^1$ is said to Pareto dominate $\mathbf{x}^2$ (denoted as $\mathbf{x}^1 \preceq \mathbf{x}^2$) if and only if $f_i(\mathbf{x}^1) \leq f_i(\mathbf{x}^2)$ for every $i \in \{1, \cdots, m\}$ and $f_j(\mathbf{x}^1) < f_j(\mathbf{x}^2)$ for at least one $j \in \{1, \cdots, m\}$. A solution $\mathbf{x}^*$ is Pareto-optimal to (1) if there is no other solution $\mathbf{x} \in \mathbf{\Omega}$ such that $\mathbf{x} \preceq \mathbf{x}^*$. $\mathbf{F}(\mathbf{x}^*)$ is then called a Pareto-optimal (objective) vector. The set of all Pareto-optimal solutions is called the Pareto-optimal set (PS). Accordingly, the set of all Pareto-optimal vectors, PF= $\{\mathbf{F}(\mathbf{x}) \in \mathbb{R}^m|\mathbf{x} \in \mathrm{PS}\}$, is called the Pareto-optimal front (PF). Evolutionary algorithms have been recognized as especially desirable for solving MOPs mainly because they can obtain a PF approximation in a single simulation run. Over the last two decades, various evolutionary multiobjective optimization (EMO) algorithms have been proposed to solve MOPs (e.g., [1]–[9]).

In multiobjective optimization, there are two basic issues: 1) convergence towards the PF and 2) diversity of the obtained points in representing the PF. Though sounds quite intuitive, it is far from trivial to achieve these two goals simultaneously. According to the ways of balancing convergence and diversity, current EMO algorithms can be categorized into three groups. The first one is Pareto-based methods (e.g., NSGA-II [10]), which uses Pareto dominance to emphasize convergence and use density estimation to maintain population diversity. Due to the direct use of dominance concept, this kind of method is good at controlling the selection pressure (e.g., ϵ-dominance [11]) of convergence towards the PF. The second category is decomposition-based methods (e.g., MOEA/D [12]), which decomposes a MOP in question into several subproblems and uses a population-based algorithm to solve them simultaneously. Due to the use of a set of evenly spread weight vectors for specifying the search directions, decomposition-based method eases the tasks of maintaining population diversity. The last group is indicator-based methods (e.g., IBEA [13]), which assigns fitness values to solutions according to a performance indicator, such as hypervolume [14]. Since this indicator aggregates the information of both convergence and diversity, it naturally balances the convergence and diversity during the evolution process. However, the selection pressure for neither convergence nor diversity is quantitatively controllable by simply using a single indicator.

To address the balance between convergence and diversity, this paper designs a hybrid EMO algorithm by taking advantages of different selection principles. It is a extension along our recent work on dual-population paradigm [15]. Our major idea is to maintain two separate and co-evolving archives during evolution process. In particular, one archive uses a modified dominance concept, termed as grid-dominance, to strengthen the selection pressure towards the PF, the other uses a decomposition-based technique to maintain a set of well-distributed solutions. Furthermore, to exploit guidance information towards the PS, we develop a restricted mating selection mechanism to select mating parents for offspring generation. It is worth noting that using two or multiple selection principles (or archives) is not a brand new technique in the EMO community (e.g., [16]–[18]). Differently, this paper aims at maintaining two separate and co-evolving populations, which respectively handles convergence and diversity of the

evolution process. It paves a new avenue to integrate different selection principles.

The rest of this paper is organized as follows. Section II presents the technical details of our proposed algorithm step by step. In Section III, several EMO algorithms are included in empirical studies and comparisons. Section IV concludes this paper and provides some future directions.

II. Proposed Algorithm

The flow chart of our proposed hybrid EMO algorithm, HEMOA/GD for short, is given in Fig. 1. It maintains two co-evolving populations, which are with distinct objectives and selection principles. In general, one population, named as convergence archive $P^C = \{\dot{\mathbf{x}}^1, \cdots, \dot{\mathbf{x}}^M\}$, helps promote the population convergence towards PF; the other one, termed as diversity archive $P^D = \{\ddot{\mathbf{x}}^1, \cdots, \ddot{\mathbf{x}}^N\}$, helps preserve population distribution in the objective space. During the reproduction process, a restricted mating selection mechanism is applied to select mating parents from each of P^C and P^D, respectively, for offspring generation. Afterwards, the generated offspring solution, inheriting promising properties from both archives, is used to update each of of P^C and P^D, respectively, based on its corresponding selection principle. In the following paragraphs, we will illustrate each module of HEMOA/GD in detail step by step.
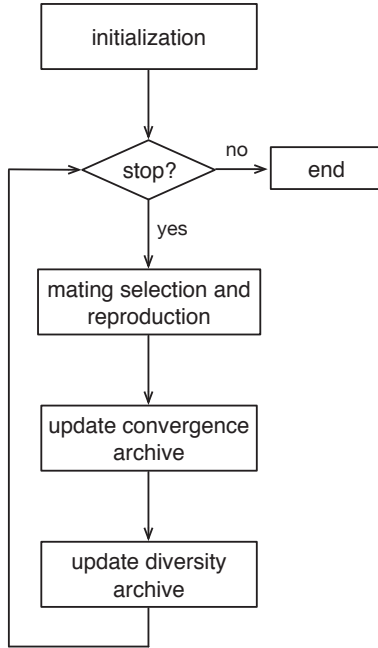

Fig. 1: Flow chart of HEMOA/GD.

A. Design of Convergence Archive

The major purpose of the P^C is to maintain a repository that provides competitive selection pressure towards PF. To facilitate the selection process, the objective space is divided into several grid cells. Inspired by the adaptive genetic algorithm proposed by Knowles [19], the size and location of this grid structure is progressively updated during the evolution process [2]. Here we only illustrate the grid setup in one dimension, without loss of generality. First of all, the lower and upper boundaries of the grid structure on the k-th dimension (denoted as lb_k and ub_k, respectively) are determined as:

$$lb_k = z_k^* - (z_k^{nad} - z_k^*)/(2 \times div) \tag{2}$$

$$ub_k = z_k^{nad} + (z_k^{nad} - z_k^*)/(2 \times div) \tag{3}$$

where $\mathbf{z}^* = (z_1^*, \cdots, z_m^*)^T$ and $\mathbf{z}^{nad} = (z_1^{nad}, \cdots, z_m^{nad})^T$ are the ideal and nadir objective vectors, and div is a user-defined parameter that determines the number of divisions on each dimension. A parameter sensitivity study on div is performed in Section III-D. Since the ideal and nadir objective vectors are usually unknown a priori, we use their approximations instead, where $z_i^* = \min\{f_i(\mathbf{x})|\mathbf{x} \in P^C \cup P^D\}$ and $z_i^{nad} = \max\{f_i(\mathbf{x})|\mathbf{x} \in P^C \cup P^D\}$, where $i \in \{1, \cdots, m\}$. In this case, the grid structure is with div^m grid cells and the width of each grid cell on the k-th dimension (denoted as $width_k$) is evaluated as:

$$width_k = (ub_k - lb_k)/div \tag{4}$$

For each solution $\dot{\mathbf{x}}$ in P^C, the coordinate of its resided grid cell $\mathbf{C}(\dot{\mathbf{x}})$ on the k-th dimension is calculated as:

$$C_k(\dot{\mathbf{x}}) = \lfloor (f_k(\dot{\mathbf{x}}) - lb_k)/width_k \rfloor \tag{5}$$

where $k \in \{1, \cdots, m\}$. Fig. 2 gives a simple example of the grid setup on the k-th dimension, where $div = 6$. The coordinates of grid cells of all solutions (from left to right) on the k-th dimension are 0, 1, 2, 3, 4, 4 and 5, respectively.
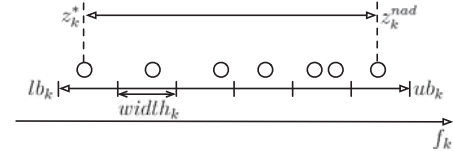

Fig. 2: Grid setup on the k-th dimension.

For further discussion of the selection principle in P^C, we introduce several concepts of grid-dominance relationship.

Definition 1: Let $\mathbf{x}^i$ and $\mathbf{x}^j$ reside in different grid cells, say grid cell $\mathbf{C}(\mathbf{x}^i)$ and grid cell $\mathbf{C}(\mathbf{x}^j)$, respectively. $\mathbf{x}^i$ is said to grid-dominate $\mathbf{x}^j$, denoted as $\mathbf{x}^i \preceq_G \mathbf{x}^j$, if and only if

1) $\forall k \in \{1, \cdots, m\}$: $C_k(\mathbf{x}^i) \leq C_k(\mathbf{x}^j)$.
2) $\exists l \in \{1, \cdots, m\}$: $C_l(\mathbf{x}^i) < C_l(\mathbf{x}^j)$.

This grid-dominance relationship establishes a strict partial order between solutions. To prove this, we study its four properties as follows.

Property 1: The grid-dominance is irreflexive.

Proof: We would like to show that $\forall \mathbf{x} \in \Omega$, $\mathbf{x} \npreceq_G \mathbf{x}$. $\forall k \in \{1, \cdots, m\}$, we have $C_k(\mathbf{x}) = C_k(\mathbf{x})$, thus the second condition in Definition 1 will never be met. ∎

Property 2: The grid-dominance is asymmetric.

Proof: We would like to show that if $\mathbf{x} \preceq_G \mathbf{y}$, $\mathbf{y} \npreceq_G \mathbf{x}$. Assume that $\mathbf{x} \preceq_G \mathbf{y}$ and $\mathbf{y} \preceq_G \mathbf{x}$, we have $\forall k \in \{1, \cdots, m\}$,

$C_k(\mathbf{y}) \leq C_k(\mathbf{x})$ and $\exists l \in \{1, \cdots, m\}, C_l(\mathbf{y}) < C_l(\mathbf{x})$. This should contradict the assumption that $\mathbf{x} \preceq_G \mathbf{y}$, where $\exists i \in \{1, \cdots, m\}, C_k(\mathbf{x}) < C_k(\mathbf{y})$. ∎

Property 3: The grid-dominance is transitive.

Proof: We would like to show that if $\mathbf{x} \preceq_G \mathbf{y}$ and $\mathbf{y} \preceq_G \mathbf{z}$, then $\mathbf{x} \preceq_G \mathbf{z}$.

Given that $\mathbf{x} \preceq_G \mathbf{y}$ and $\mathbf{y} \preceq_G \mathbf{z}$, we have $\forall k \in \{1, \cdots, m\}, C_k(\mathbf{x}) \leq C_k(\mathbf{y})$ and $C_k(\mathbf{y}) \leq C_k(\mathbf{z})$. Thus, $\forall k \in \{1, \cdots, m\}, C_k(\mathbf{x}) \leq C_k(\mathbf{z})$. Moreover, $\exists l \in \{1, \cdots, m\}, C_l(\mathbf{y}) < C_l(\mathbf{z})$, thus for such l, $C_l(\mathbf{x}) \leq C_l(\mathbf{y}) < C_l(\mathbf{z})$. Consequently, we can have $\mathbf{x} \preceq_G \mathbf{z}$. ∎

Property 4: The grid-dominance defines a strict partial order relationship.

Proof: Since the grid-dominance is irreflecxive, asymmetric and transitive, then it defines a strict partial order relationship. ∎

One of the major considerations of the grid-dominance relationship is the preservation of the Pareto dominance. In the following, we will further study its compatibility and completeness [20] with the Pareto dominance relationship as done in [21].

Definition 2: The relationship $\mathbf{x} \triangleright \mathbf{y}$ is said to be compatible with the Pareto dominance if and only if

$$\mathbf{x} \triangleright \mathbf{y} \Rightarrow \mathbf{x} \preceq \mathbf{y} \tag{6}$$

Definition 3: The relationship $\mathbf{x} \triangleright \mathbf{y}$ is said to be complete with the Pareto dominance if and only if

$$\mathbf{x} \preceq \mathbf{y} \Rightarrow \mathbf{x} \triangleright \mathbf{y} \tag{7}$$

Now, we can have the following theorem.

Theorem 1: Given a population of solutions, the grid-dominance is:

1) complete with the Pareto dominance relationship.
2) compatible with the non-Pareto dominance relationship.

Proof: Given that $\mathbf{x} \preceq \mathbf{y}$, we have $\forall k \in \{1, \cdots, m\}, f_k(\mathbf{x}) \leq f_k(\mathbf{y})$, and $\exists l \in \{1, \cdots, m\}, f_k(\mathbf{x}) < f_k(\mathbf{y})$. If the grid structure is determined, lb_k and $width_k$ are determined for all $k \in \{1, \cdots, m\}$. Then, we can have $\forall k \in \{1, \cdots, m\}, \lfloor (f_k(\mathbf{x}) - lb_k)/width_k \rfloor \leq \lfloor (f_k(\mathbf{y}) - lb_k)/width_k \rfloor$, and $\exists l \in \{1, \cdots, m\}, \lfloor (f_k(\mathbf{x}) - lb_k)/width_k \rfloor < \lfloor (f_k(\mathbf{y}) - lb_k)/width_k \rfloor$. Consequently, the grid-dominance is said to be complete with the Pareto dominance. On the other hand, due the rounding down operation, $\exists k \in \{1, \cdots, m\}$, we might have $\lfloor (f_k(\mathbf{x}) - lb_k)/width_k \rfloor = \lfloor (f_k(\mathbf{y}) - lb_k)/width_k \rfloor$ with $f_k(\mathbf{x}) > f_k(\mathbf{y})$. In this case, $\mathbf{x} \preceq_G \mathbf{y}$, but $\mathbf{y} \not\preceq \mathbf{x}$. Consequently, the grid-dominance is said to be compatible with the non-Pareto dominance relationship. ∎

In summary, the grid-dominance does not contradict the order induced by the Pareto dominance. More importantly, it has the ability to differentiate between the non-dominated solutions in a partial manner, depending on the width of the grid cell. This fact provides the grid-dominance a stronger selection pressure than the Pareto dominance, which is especially beneficial for many-objective optimization. Fig. 3

gives an illustration of the grid structure in a two-dimensional case. As for $\mathbf{x}^1$, it grid-dominates the entire region ABCDA, while according to the definition of Pareto dominance, $\mathbf{x}^1$ only dominate the region EFCGE. Specifically, under the Pareto dominance relationship, $\mathbf{x}^1 \preceq \mathbf{x}^2$, but $\mathbf{x}^1 \not\preceq \mathbf{x}^3$; whereas under the grid-dominance relationship, $\mathbf{x}^1 \preceq_G \mathbf{x}^2$ and $\mathbf{x}^1 \preceq_G \mathbf{x}^3$.
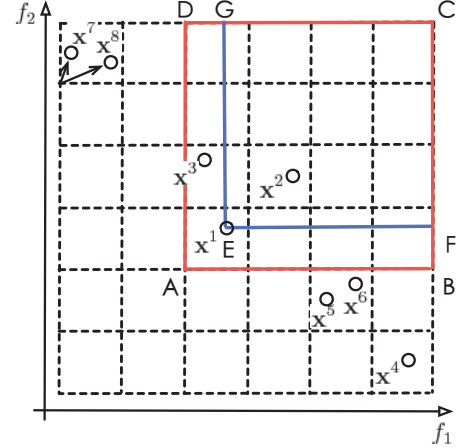

Fig. 3: An illustration of the grid structure and selection principle of P^C.

For an offspring $\overline{\mathbf{x}}$, we at first find its grid cell location. Then, it is compared with solutions in P^C according to the grid-dominance relationship. In this case, we might have the following four scenarios:

1) If $\exists \dot{\mathbf{x}} \in P^C$, satisfying $\overline{\mathbf{x}} \preceq_G \dot{\mathbf{x}}$, $\overline{\mathbf{x}}$ will be accepted as a member in P^C while $\dot{\mathbf{x}}$ is excluded from P^C; otherwise, $\overline{\mathbf{x}}$ is not accepted. As shown in Fig. 3, $\mathbf{x}^1 \preceq_G \mathbf{x}^2$ and $\mathbf{x}^1 \preceq_G \mathbf{x}^3$, so $\mathbf{x}^2$ and $\mathbf{x}^3$ will be excluded from P^C.
2) If $\nexists \dot{\mathbf{x}} \in P^C$, satisfying $\overline{\mathbf{x}} \preceq_G \dot{\mathbf{x}}$ (and vice versa), and no other solution share the same grid cell with $\overline{\mathbf{x}}$, it will be directly accepted as a member in P^C. For example, $\mathbf{x}^4$ will stay in P^C, since no other solution grid-dominate $\mathbf{x}^4$.
3) If $\overline{\mathbf{x}}$ shares the same grid cell with a solution $\dot{\mathbf{x}}$ in P^C and $\overline{\mathbf{x}} \preceq \dot{\mathbf{x}}$, $\overline{\mathbf{x}}$ will be accepted as a member in P^C and $\dot{\mathbf{x}}$ is excluded from P^C; otherwise, $\overline{\mathbf{x}}$ is not accepted. For example, in Fig. 3, $\mathbf{x}^5$ and $\mathbf{x}^6$ share the same grid cell, and $\mathbf{x}^6$ will be excluded from P^C since $\mathbf{x}^5 \preceq \mathbf{x}^6$.
4) Inspired from the idea of ϵ-dominance [16], if $\overline{\mathbf{x}}$ shares the same grid cell with a solution $\dot{\mathbf{x}}$ in P^C and they are non-dominated with each other, the one with the smaller distance to the left corner of their shared grid cell will be remained in P^C, while the other one is excluded. Specifically, the coordinate of the left corner $\mathbf{c}(\overline{\mathbf{x}})$ of the grid cell resided by $\overline{\mathbf{x}}$ is determined as:

$$c_k(\overline{\mathbf{x}}) = lb_k + width_k \times C_k(\overline{\mathbf{x}}) \tag{8}$$

where $k \in \{1, \cdots, m\}$. For example, in Fig. 3, $\mathbf{x}^7$ and $\mathbf{x}^8$ share the same grid cell, and $\mathbf{x}^8$ is excluded from P^C since it has a larger distance to $\mathbf{c}(\mathbf{x}^8)$.

It is obvious that at most one solution is allowed to be resided in a grid cell according to the selection principle of P^C. In other words, each solution in P^C resides in a different grid cell. Therefore, this selection principle not only strengthens the selection pressure towards the PF, but also helps to maintain the population distribution in P^C. Moreover, since the total number of grid cells is div^m, the size of P^C is automatically bounded.

B. Design of the Diversity Archive

The primary objective of P^D is to maintain a population with a well distribution in the objective space. To this end, we define a set of weight vectors $W = \{\mathbf{w}^1, \cdots, \mathbf{w}^N\}$, according to the method introduced in [22]. For each $\mathbf{w}^i = (w_1^i, \cdots, w_m^i)^T$, where $i \in \{1, \cdots, N\}$, we have $\sum_{k=1}^m w_k^i = 1$. Each weight vector not only defines a subproblem, which is a decomposition of the original MOP, but also specifies a search direction in the objective space. Here we use the method suggested in [7] as the aggregation function of a subproblem, since its search direction is the corresponding weight vector. As weight vectors are evenly distributed, their optimal solutions are expected to be with a satisfied distribution in the objective space. Furthermore, these weight vectors divide the objective space into N subregions Ω^1 to Ω^N:

$$\Omega^i = \{\mathbf{F}(\mathbf{x}) \in \mathbb{R}_+^m | \langle \mathbf{F}(\mathbf{x}), \mathbf{w}^i \rangle \leq \langle \mathbf{F}(\mathbf{x}), \mathbf{w}^j \rangle\} \quad (9)$$

where $i \in \{1, \cdots, N\}$ and $\langle \mathbf{F}(\mathbf{x}), \mathbf{w} \rangle$ is the Euclidian distance between $\mathbf{F}(\mathbf{x})$ and $\mathbf{w}$. Fig. 4 gives an example of six evenly distributed weight vectors $\mathbf{w}^1$ to $\mathbf{w}^6$, and their corresponding subregions are Ω^1 to Ω^6, respectively. For example, in Fig. 4, $\mathbf{x}^1$ is associated with Ω^2. At last, each weight vector is assigned a neighborhood, consisting of its T $(1 \leq T \leq N)$ nearest weight vectors.
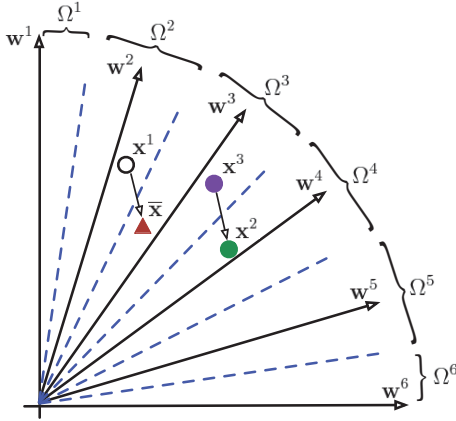

Fig. 4: Illustration of weight vectors and subregions.

After the setup of subproblems and subregions, the selection principle in P^D is relatively simple. We at first determine the corresponding subregion Ω^k associated with $\overline{\mathbf{x}}$ as:

$$k = \underset{i \in \{1, \cdots, N\}}{\operatorname{argmin}} \langle \overline{\mathbf{F}}(\overline{\mathbf{x}}), \mathbf{w}^i \rangle \quad (10)$$

where $\langle \overline{\mathbf{F}}(\overline{\mathbf{x}}), \mathbf{w}^i \rangle$ is the acute angle between $\overline{\mathbf{F}}(\overline{\mathbf{x}})$ and $\mathbf{w}^i$, $\overline{\mathbf{F}}(\overline{\mathbf{x}})$ is the normalized objective vector of $\overline{\mathbf{x}}$, and its k-th individual objective function is normalized as:

$$\overline{f}_k(\overline{\mathbf{x}}) = \frac{f_k(\overline{\mathbf{x}})}{\sum_{i=1}^m f_i(\overline{\mathbf{x}})} \quad (11)$$

where $k \in \{1, \cdots, m\}$. Note that the subregion association relationship of solutions already in P^D has been done during the initialization process. Afterwards, $\overline{\mathbf{x}}$ replaces the solution $\ddot{\mathbf{x}} \in P^D$ associated with Ω^k in case it has a better aggregation function value; otherwise, it is not accepted. For the sake of diversity preservation, only one solution is allowed to be associated with a subregion in this way.

C. Reproduction with Restricted Mating Selection Mechanism

Due to the vector-based property and reportedly good performance on continuous optimization problems [23], we use the differential evolution (DE) operator for offspring generation. In particular, a trial vector $\mathbf{u} = \{u_1, \cdots, u_n\}$ is generated as follows:

$$u_j = \begin{cases} x_j^i + F \times (x_j^{r_1} - x_j^{r_2}) & \text{if } rand < CR \text{ or } j = j_{rand} \\ x_j & \text{otherwise} \end{cases} \quad (12)$$

where $\mathbf{x}^i$ is the solution associated with subregion Ω^i, $j \in \{1, \cdots, n\}$, $rand \in [0, 1]$, CR and F are two control parameters and j_{rand} is a random integer uniformly chosen from 1 to n. To further increase the population diversity, a polynomial mutation [24] is conducted upon $\mathbf{u}$ to obtain an offspring solution $\overline{\mathbf{x}}$. Note that violated decision variables are simply set to their nearest boundaries.

As for the traditional DE operator [23], $\mathbf{x}^{r_1}$ and $\mathbf{x}^{r_2}$ in equation (12) are usually selected in a random manner. This random mating selection mechanism has a strong exploration ability. However, since no guidance information towards the optima has been considered, it might cause a degeneration problem. For example, in Fig. 5, if we set $\mathbf{x}^i \leftarrow \mathbf{x}^1$, $\mathbf{x}^{r_1} \leftarrow \mathbf{x}^2$ and $\mathbf{x}^{r_2} \leftarrow \mathbf{x}^3$, based on equation (12), the offspring might be the green star $\mathbf{u}^1$. In this case, although $\mathbf{x}^1$ and $\mathbf{x}^3$ are close to the PS (the blue dotted curve), $\mathbf{u}^1$ is guided away from the PS. On the other hand, if we set the mating parents as $\mathbf{x}^i \leftarrow \mathbf{x}^4$, $\mathbf{x}^{r_1} \leftarrow \mathbf{x}^6$ and $\mathbf{x}^{r_2} \leftarrow \mathbf{x}^5$, according to equation (12), we might obtain the offspring solution $\mathbf{u}^2$, which is closer to the PS this time.

Based on the above discussions, we propose a restricted mating selection mechanism to select mating parents for offspring generation. This mechanism aims at exploiting the guidance information towards the PS, so that the offspring is expected to approach the PS as much as possible. As discussed in Section II-A, P^C provides a population of solutions with good convergence property, thus $\mathbf{x}^{r_1}$ is selected from P^C. On the other hand, as discussed in Section II-B, P^D maintains a population of well-distributed solutions, thus $\mathbf{x}^{r_2}$ is selected from P^D. In this way, we might have a large chance to obtain the direction information towards the PS. To be specific, as for the current subregion Ω^i, we randomly choose two
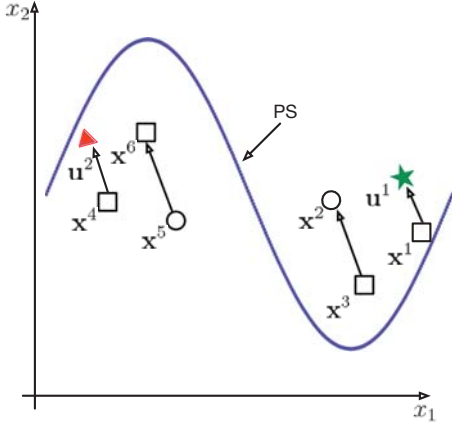
903

Fig. 5: A simple illustration on the restricted mating selection mechanism.

neighboring subregions Ω^j and Ω^k from its neighborhood. Afterwards, we randomly choose a solution $\dot{\mathbf{x}} \in P^C$ associated with Ω^j, and set $\mathbf{x}^{r_1} \leftarrow \dot{\mathbf{x}}$. Note that if there does not exist such solution in P^C, we just choose a solution from P^D in a similar way. At last, we choose another solution $\ddot{\mathbf{x}} \in P^D$ associated with Ω^k, and set $\mathbf{x}^{r_2} \leftarrow \ddot{\mathbf{x}}$. Let us consider a simple example shown in Fig. 4, let $\mathbf{x}^i \leftarrow \mathbf{x}^1$. Assume that $\mathbf{x}^2$, associated with Ω^4, is selected from the P^C, and set $\mathbf{x}^{r_1} \leftarrow \mathbf{x}^2$; the solution $\mathbf{x}^3$, associated with Ω^3, is selected from the P^D, and set $\mathbf{x}^{r_2} \leftarrow \mathbf{x}^3$. After the reproduction operation, we might obtain the offspring $\overline{\mathbf{x}}$ (denoted as the red triangular).

III. EMPIRICAL STUDIES

A. Performance Metrics and Benchmark Problems

This paper considers two widely used metrics (i.e., inverted generational distance (IGD) [25] and hypervolume (HV) [14]) to evaluate the performance of an EMO algorithm. Both of them are able to give a comprehensive assessment on convergence and diversity. Generally speaking, IGD calculates the average distance of each sample point along the PF away from the obtained solution set, and a lower value is better. HV calculates the volume of the objective space dominated by the obtained solution set and bounded by a reference points. A larger HV value is preferable. Interested readers are recommended to their original papers for more details. Furthermore, the reference points for HV calculation are set as $(2.0, 2.0)^T$ for two-objective cases and $(2.0, 2.0, 2.0)^T$ for three-objective ones. Seven recently proposed MOP instances [26], i.e., MOP1 to MOP7, are used as benchmark problems in this paper.

B. EMO Algorithms for Comparisons and Parameter Settings

In this paper, we choose four well-known EMO algorithms for comparisons. They are: NSGA-II [10] using a Pareto-based method, IBEA [13] using an indicator-based method, MOEA/D [27] and MOEA/D-M2M [26] using a decomposition-based method. All algorithms are implemented in JAVA, except MOEA/D-M2M in MATLAB[1]. According to the standard settings in [26], population size N is set to 100 for MOP1 to MOP5, and 300 for MOP 6 to MOP7. According to the recommendations of the original papers, Table I summarizes the general parameter settings of different EMO algorithms.

C. Comparisons with Other EMO Algorithms

MOP instances, modified from ZDT [28] and DTLZ [29] test suites, are recently proposed benchmark problems. For MOP1 to MOP7, some Pareto-optimal solutions are much easier to obtain than others at the early stage of evolution. Population diversity is very important for solving these problems. As reported in [30], some state-of-the-art EMO algorithms, e.g., MOEA/D and NSGA-II, have significant difficulties in tackling these problems. Comparison results on the IGD and HV values are given in Table II. Fig. 6 to Fig. 8 plot the final solutions obtained in the run with the median IGD value of each EMO algorithm for each instance. Comparing to the other algorithms, HEMOA/GD is very competitive as it wins on all MOP instances. According to Wilcoxon's rank sum test, all better metric values obtained by HEMOA/GD are with statistical significance.

To be specific, MOP1 is a modification of ZDT1 instance [28]. From the experimental results shown in Fig. 6, we find that only the solutions obtained by HEMOA/GD and MOEA/D-M2M are able to approximate the entire PF, while the other algorithms only find some extreme points along each individual objective. MOP2 and MOP3 have the same concave PF shape in the objective space. It is clear that the solutions obtained by HEMOA/GD not only well converge to the PF, but also well distribute along the whole PF. By contrast, although MOEA/D-M2M can approximate the whole PF, the distribution of solutions is not as smooth as that of HEMOA/GD. The other algorithms even cannot converge to the PF. MOP4, extended from ZDT3 instance [28], has three disconnected PF segments. From the results shown in Fig. 6, we notice that both HEMOA/GD and MOEA/D-M2M find some dominated solutions between the rightmost and middle segments of the PF. Nevertheless, the IGD and HV values obtained by HEMOA/GD are still better than MOEA/D-M2M. Again, solutions obtained by other algorithms only approximate very limited regions of the PF. MOP5 is an extension of MOP1. Solutions obtained by HEMOA/GD are clearly better than other algorithms in terms of convergence and diversity. Although MOEA/D-M2M can approximate the whole PF, some of its solutions, distributed on the leftmost and rightmost corners, cannot converge to the PF as shown in Fig. 6. MOP6 is a three-objective instance, which is developed from DTLZ1 instance [29]. As shown in Fig. 7, HEMOA/GD not only converge to the PF, but also well spread among the whole PF. Although the solutions obtained by MOEA/D-M2M can approximate most regions of the PF, many of them are

---

[1]NSGA-II, IBEA, MOEA/D and our proposed HEMOA/GD is developed upon jMetal framework, downloaded from http://jmetal.sourceforge.net. The source codes can be obtained from the authors upon requests.

TABLE I: Parameter settings of different EMO algorithms

EMO Algorithms	Parameter Settings
HEMOA/GD	$p_m = 1/nreal, \mu_m = 20, CR = 1.0, F = 0.5, \delta = 0.9, T = 20, n_r = 2, div = \lfloor 0.3 \times N \rfloor$
NSGA-II	$p_c = 0.9, p_m = 1/nreal, \mu_c = 20, \mu_m = 20$
IBEA	$p_c = 0.9, p_m = 1/nreal, \mu_c = 20, \mu_m = 20, \kappa = 0.05$
MOEA/D	$p_m = 1/nreal, \mu_m = 20, CR = 1.0, F = 0.5, \delta = 0.9, T = 20, n_r = 2$
MOEA/D-M2M	$p_m = 1/nreal$, $Max\_gen=[(300,000 - 6 \times N)/N]$, $K = S = 10$ for MOP1 to MOP5, $K = S = 17$ for MOP6 to MOP7

not converged. By contrast, other algorithm only find some of the boundary solutions. MOP7 is modified from DTLZ2 instance [29]. Solutions obtained by HEMOA/GD are better than those obtained by MOEA/D-M2M, which have many gaps along the PF. Similar to MOP6, solutions obtained by other algorithms crowd on the boundary of $f_1 - f_2$ plane.

D. Sensitivity Studies of div

The number of divisions on each dimension div is a major parameter in the convergence archive. This subsection investigates the performance of HEMOA/GD with different div settings, where $div = \lfloor \alpha \times N \rfloor$, $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 1.5, 2.0, 3.0\}$. Fig. 9 presents the errorbars of IGD values on MOP1 to MOP7 instances with different div settings. From this figure, it is clear that the performance HEMOA/GD is not sensitive to the setting of div in most cases. Generally speaking, $\alpha = 0.1$ is always the worst choice, while $\alpha = 0.3$ is the best one. In fact, div controls the granularity of grid structure in the convergence archive, and thus the selection pressure of grid-dominance relationship. Specifically, the smaller the div, the larger the width of a grid cell. This also indicates a stronger selection pressure. On the other hand, the selection pressure will be reduced with the increase of div. In an extreme case, grid-dominance degenerates to Pareto-dominance when $div \to \infty$. It is interesting to note that the performance of HEMOA/GD does not decrease too much when α becomes big. It implies that Pareto-dominance might also be applicable to the design of selection principle in the convergence archive. This observation confirms the rationality of a more general dual-population paradigm design developed in our recent work [15].

IV. CONCLUSIONS

This paper presents a novel way for balancing convergence and diversity in EMO. Two separate and co-evolving populations (i.e., convergence archive and diversity archive) are maintained during the evolution process. By using a grid-dominance relationship, the convergence archive maintains a repository having a competitive selection pressure towards the PF; while the diversity archive, using a decomposition-based selection principle, preserves a population of well-distributed solutions. To further exploit guidance information towards the PS, a restricted mating selection mechanism is proposed to select mating parents for offspring generation. The performance of our proposed HEMOA/GD is investigated on a set of benchmark problems which pose significant challenges

in balancing convergence and diversity. Empirical results fully demonstrate the competitiveness of HEMOA/GD, considering the fact that it performs better than the other algorithms on all benchmark problems.

Future work includes the investigation of HEMOA/GD on a wider range test problems, e.g,. MOPs with a large number of objectives, problems in dynamic, noisy and uncertain environments, and some real-world application problems.

REFERENCES

[1] K. Li, Á. Fialho, and S. Kwong, "Multi-objective differential evolution with adaptive control of parameters and operators," in *LION'11: Proc. of the 5th International Conference Learning and Intelligent OptimizatioN*, C. A. C. Coello, Ed. Springer LNCS, Jan. 2011, pp. 473–487.

[2] K. Li, S. Kwong, R. Wang, J. Cao, and I. J. Rudas, "Multi-objective differential evolution with self-navigation," in *SMC'12: Proc. of the 2012 IEEE International Conference on Systems, Man, and Cybernetics*. Seoul, Korea(South): IEEE, 2012, pp. 508–513.

[3] K. Li, S. Kwong, J. Cao, M. Li, J. Zheng, and R. Shen, "Achieving balance between proximity and diversity in multi-objective evolutionary algorithm," *Information Sciences*, vol. 182, no. 1, pp. 220–242, 2012.

[4] K. Li, S. Kwong, R. Wang, K.-S. Tang, and K.-F. Man, "Learning paradigm based on jumping genes: A general framework for enhancing exploration in evolutionary multiobjective optimization," *Information Sciences*, vol. 226, pp. 1–22, 2013.

[5] K. Li and S. Kwong, "A general framework for evolutionary multiobjective optimization via manifold learning," *Neurocomputing*, vol. 146, pp. 65–74, 2014.

[6] K. Li, Á. Fialho, S. Kwong, and Q. Zhang, "Adaptive operator selection with bandits for a multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 1, pp. 114–130, 2014.

[7] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable matching based selection in evolutionary multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 6, pp. 909–923, 2014.

[8] K. Li, K. Deb, Q. Zhang, and S. Kwong, "An evolutionary many-objective optimization algorithm based on dominance and decomposition," *IEEE Transactions on Evolutionary Computation*, 2014, accepted for publication.

[9] K. Li, S. Kwong, Q. Zhang, and K. Deb, "Interrelationship-based selection for decomposition multiobjective optimization," *IEEE Transactions on Cybernetics*, 2014, accepted for publication.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[11] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evolutionary Computation*, vol. 10, no. 3, pp. 263–282, 2002.

[12] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, pp. 712–731, Dec. 2007.

[13] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *PPSN VIII: Proc. of the 8th International Conference on Parallel Problem Solving from Nature*. Springer, 2004, pp. 832–842.

TABLE II: Performance comparisons on IGD and HV values

		HEMOA/GD	NSGA-II	IBEA	MOEA/D	MOEA/D-M2M
MOP1	IGD	**1.7111E-2(2.51E-3)**	3.645E-1(4.19E-3)[†]	3.630E-1(6.02E-3)[†]	3.453E-1(4.72E-2)[†]	1.750E-2(7.29E-4)[†]
	HV	**3.6404(3.29E-3)**	3.0772(1.11E-2)[†]	3.0803(1.55E-2)[†]	3.1102(8.83E-2)[†]	3.6368(1.24E-3)[†]
MOP2	IGD	**7.086E-3(5.39E-3)**	3.543E-1(5.55E-7)[†]	3.480E-1(1.60E-2)[†]	2.326E-1(5.45E-2)[†]	1.565E-2(4.15E-3)[†]
	HV	**3.3208(1.02E-2)**	3.0000(2.02E-6)[†]	3.0037(9.59E-3)[†]	3.0664(4.72E-2)[†]	3.3066(7.57E-3)[†]
MOP3	IGD	**4.478E-3(1.04E-4)**	1.119E-1(1.15E-2)[†]	1.740E-1(1.86E-2)[†]	1.295E-1(5.12E-2)[†]	1.860E-2(7.79E-3)[†]
	HV	**3.2082(3.43E-4)**	3.1148(1.44E-2)[†]	3.0547(1.88E-2)[†]	3.0784(5.11E-2)[†]	3.1866(1.41E-2)[†]
MOP4	IGD	**1.013E-2(4.43E-3)**	3.107E-1(2.04E-2)[†]	3.035E-1(1.64E-2)[†]	2.885E-1(3.31E-2)[†]	1.111E-2(2.57E-3)[†]
	HV	**3.5050(4.34E-3)**	3.1413(7.67E-3)[†]	3.1431(7.09E-3)[†]	3.1557(2.26E-2)[†]	3.5033(4.30E-3)[†]
MOP5	IGD	**1.692E-2(1.94E-3)**	2.796E-1(3.26E-2)[†]	3.031E-1(1.63E-2)[†]	3.165E-1(5.47E-3)[†]	2.166E-2(1.17E-3)[†]
	HV	**3.6397(2.91E-3)**	3.0753(1.42E-1)[†]	2.9384(1.40E-1)[†]	2.7598(1.14E-1)[†]	3.6266(3.01E-3)[†]
MOP6	IGD	**4.325E-2(1.85E-3)**	3.044E-1(5.75E-6)[†]	3.012E-1(1.44E-2)[†]	2.952E-1(1.88E-2)[†]	6.681E-2(1.12E-3)[†]
	HV	**7.7766(2.17E-3)**	7.4978(6.46E-5)[†]	7.5018(2.01E-2)[†]	7.5096(2.83E-2)[†]	7.7346(3.21E-3)[†]
MOP7	IGD	**7.362E-2(3.04E-3)**	3.509E-1(7.99E-6)[†]	3.552E-1(1.44E-3)[†]	3.477E-1(1.85E-2)[†]	8.786E-2(2.08E-3)[†]
	HV	**7.3669(3.61E-3)**	7.2130(3.80E-5)[†]	7.2067(2.29E-3)[†]	7.2150(2.13E-2)[†]	7.2771(4.09E-3)[†]

Wilcoxon's rank sum test at a 0.05 significance level is performed between HEMOA/GD and each of the other algorithms. [†] denotes the performance of the corresponding algorithm is significantly worse than HEMOA/GD. The best mean is highlighted in boldface with gray background.
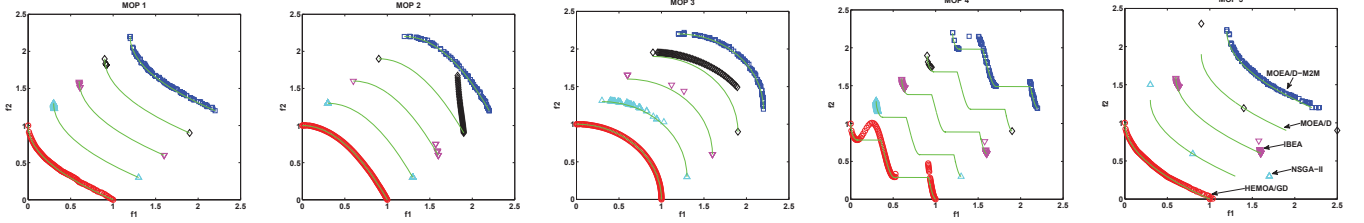

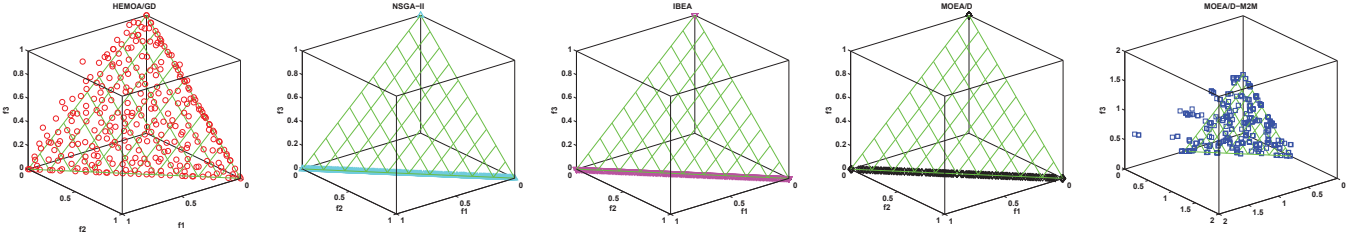Fig. 6: Final solutions obtained by different algorithms on MOP1 to MOP5.

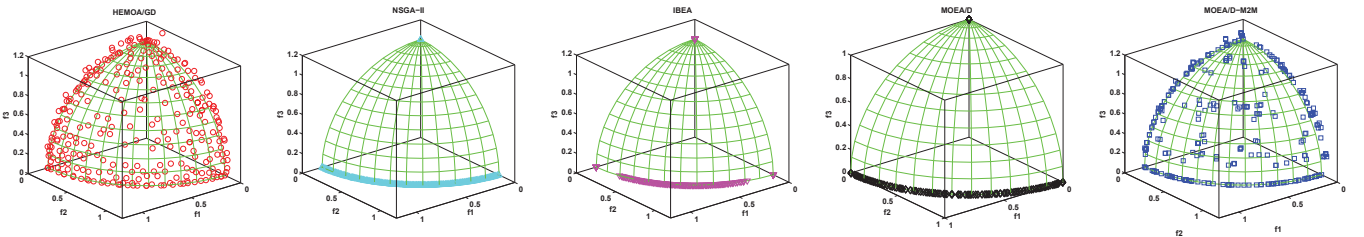Fig. 7: Final solutions obtained by different algorithms on MOP6.

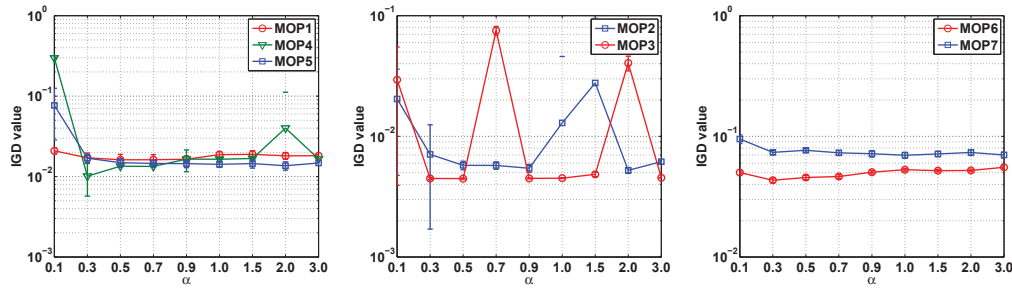Fig. 8: Final solutions obtained by different algorithms on MOP7.

Fig. 9: Parameter sensitivity study on div.

[14] E. Zitzler and L. Thiele, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

[15] K. Li, S. Kwong, and K. Deb, "A dual-population paradigm for evolutionary multiobjective optimization," *Information Sciences*, 2015, accepted for publication.

[16] K. Deb, M. Mohan, and S. Mishra, "Evaluating the ϵ-domination based multi-objective evolutionary algorithm for a quick computation of pareto-optimal solutions," *Evolutionary Computation*, vol. 13, no. 4, pp. 501–525, 2005.

[17] B. Li, J. Li, K. Tang, and X. Yao, "An improved two archive algorithm for many-objective optimization," in *CEC'14: Proc. of the 2014 IEEE Congress on Evolutionary Computation*, 2014, pp. 2869–2876.

[18] K. Praditwong and X. Yao, "A new multi-objective evolutionary optimisation algorithm: The two-archive algorithm," in *CIS'06: Proc. of the 2006 International Conference on Computational Intelligence and Security*, 2006, pp. 95–104.

[19] J. Knowles and D. Corne, "Properties of an adaptive archiving algorithm for storing nondominated vectors," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 100–116, Apr. 2003.

[20] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, and V. da Fonseca, "Performance assessment of multiobjective optimizers: an analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, Apr. 2003.

[21] L. Ben Said, S. Bechikh, and K. Ghedira, "The r-dominance: A new dominance relation for interactive evolutionary multicriteria decision making," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 801–818, Oct. 2010.

[22] I. Das and J. E. Dennis, "Normal-Boundary Intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems," *SIAM Journal on Optimization*, vol. 8, pp. 631–657, 1998.

[23] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.

[24] K. Deb and M. Goyal, "A combined genetic adaptive search (geneas) for engineering design," *Computer Science and Informatics*, vol. 26, pp. 30–45, 1996.

[25] P. Bosman and D. Thierens, "The balance between proximity and diversity in multiobjective evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 174–188, Apr. 2003.

[26] H. Liu, F. Gu, and Q. Zhang, "Decomposition of a multiobjective optimization problem into a number of simple multiobjective subproblems," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 450–455, 2014.

[27] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances," in *CEC'09: Proc. of the 2009 IEEE Congress on Evolutionary Computation*, 2009, pp. 203–208.

[28] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.

[29] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization*, ser. Advanced Information and Knowledge Processing, A. Abraham, L. Jain, and R. Goldberg, Eds. Springer London, 2005, pp. 105–145.

[30] M. Baïou and M. Balinski, "Many-to-many matching: stable polyandrous polygamy (or polygamous polyandry)," *Discrete Applied Mathematics*, vol. 101, no. 1-3, pp. 1–12, Apr. 2000.