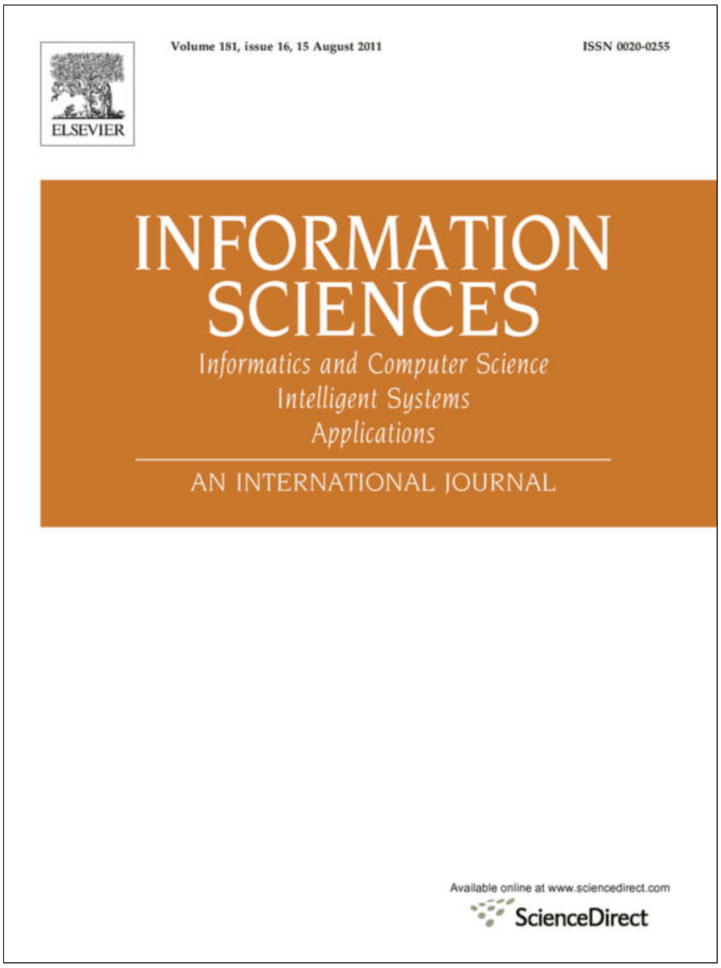

Volume 181, issue 16, 15 August 2011
ISSN 0020-0255
ELSEVIER
INFORMATION SCIENCES
Informatics and Computer Science
Intelligent Systems
Applications
AN INTERNATIONAL JOURNAL
Available online at www.sciencedirect.com
ScienceDirect

Information Sciences 181 (2011) 3336–3355

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Convergence of multi-objective evolutionary algorithms to a uniformly distributed representation of the Pareto front

Yu Chen, Xiufen Zou *, Weicheng Xie

School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 8 January 2010
Received in revised form 31 January 2011
Accepted 3 April 2011
Available online 9 April 2011

Keywords:
Multi-objective optimization
Evolutionary algorithm
Convergence
Diversity
Pareto set

ABSTRACT

In evolutionary multi-objective optimization (EMO), the convergence to the Pareto set of a multi-objective optimization problem (MOP) and the diversity of the final approximation of the Pareto front are two important issues. In the existing definitions and analyses of convergence in multi-objective evolutionary algorithms (MOEAs), convergence with probability is easily obtained because diversity is not considered. However, diversity cannot be guaranteed. By combining the convergence with diversity, this paper presents a new definition for the finite representation of a Pareto set, the *B*-Pareto set, and a convergence metric for MOEAs. Based on a new archive-updating strategy, the convergence of one such MOEA to the *B*-Pareto sets of MOPs is proved. Numerical results show that the obtained *B*-Pareto front is uniformly distributed along the Pareto front when, according to the new definition of convergence, the algorithm is convergent.

1. Introduction

Evolutionary algorithms (EAs), a type of stochastic algorithms inspired by natural evolution, can perform well in different kinds of optimization problems because they do not require the extra properties that are often associated with optimization problems, such as differentiability and continuity of the objective functions. In recent years, evolutionary multi-objective optimization (EMO) has become one of the most active research areas in the field of evolutionary computation [9]. A wide variety of multi-objective optimization algorithms (MOEAs), such as SPEA [52], SPEA2 [54], PAES [30], NSGA-II [13], IBEA [55], MSOPS [25], SMS-EMOA [16,2], MOEA/D [49,50] and MDMOEA [56], have been proposed and successfully applied to a number of real-world multi-objective optimization problems (MOPs). In addition, independent strategies have also been proposed to improve the efficiencies of MOEAs [17,45,48,6,37,46]. However, the theoretical foundations of this area are still very weak. Rudolph and Agapie [40] applied Markov chain theory to analyze some basic MOEAs, but the results are only based on certain special MOEAs with finite state space. Villalobos-Arias et al. [47] showed that Markov chains associated with the meta-heuristics for MOPs can converge geometrically to their stationary distribution, but convergence to the optimal solution set is ensured only if the best solutions in the current population are retained and copied to the next generation. Hanne [21] extended these convergence results to a type of MOEA that had an "efficiency-preserving" property, however, convergence quality (i.e., the diversity of the final results) was not considered.

Diversity has recently been included in the convergence analysis of MOEAs to obtain a finite approximation of the Pareto front. Knowles and Corne [31] analyzed the adaptive grid archive strategy proposed in [30] and proved that after a finite number of iterations, the archive always represents an ϵ-approximation of the Pareto front. However, the approximation quality depends on the granularity of the adaptive grid and on the number of allowed solutions. Moreover, these results rely

* Corresponding author.
E-mail address: xfzou@whu.edu.cn (X. Zou).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.04.004

on the additional assumption that grid boundaries remain unchanged after a finite number of iterations, which can only be fulfilled in certain special cases. Zhou and He [51] proposed a self-adaptive (μ + 1) MOEA based on adaptive "box domination" and proved that under certain general conditions, the algorithm converges in probability or almost surely to the Pareto front. However, the diversity of the final approximate Pareto front was not precisely given.

The hypervolume indicator, combining convergence and diversity in a metric, is the only known indicator compliant with the concept of Pareto-dominance [8]. Auger et al. [1] analyzed how the *optimal μ-distributions*-finite sets of μ solutions maximizing the hypervolume indicator are spread over the Pareto front, and answered the question of how to choose the reference point to obtain extremes of the front. Additionally, Friedrich et al. [18] pointed out that in some special cases the hypervolume indicator gives the best achievable multiplicative approximation ratio. However, these theoretical results only hold for bi-objective problems with continuous Pareto fronts.

Using the definition of t-dominance, Laumanns et al. [32] proposed the notion of the t-Pareto set and presented two archive-updating strategies to obtain a t-Pareto set. Additionally, Hanne [22] derived an ϵ-efficient set by proposing a new framework of MOEAs in contrast to those in [32], although the archive size is claimed to be the same as in [32]. Schütze et al. [43] presented an improved notion of the finite representation of the Pareto set called the ϵ-Pareto set. Using the updating strategies, they proposed stochastic algorithms that converge to the ϵ-Pareto set in a probabilistic way. To eliminate gaps occurring in regions in which the Pareto front is "flat", several compound updating strategies [44] have been proposed to generate a uniformly distributed ϵ-Pareto set.

Although a bounded t-Pareto set or ϵ-Pareto set can be obtained in [32,22,43,44], when MOPs with unknown properties are encountered, it is difficult to determine the value of t (or ϵ) to obtain an adequate tradeoff between the complexity of the algorithm and the distribution density of the approximate Pareto solutions. Moreover, the size of a t-Pareto set (or ϵ-Pareto set) is not determined uniquely even if the value of t (or ϵ) has been specified [32,22,43,44]. Furthermore, the gap-free representation obtained in [44] is not formally described. In this paper, a new definition of the B-Pareto set is proposed to obtain a uniformly distributed representation of a Pareto front. We also present a new convergence metric that gives rise to a novel updating strategy that maintains convergence and guarantees diversity. Accordingly, the MOEA converges almost surely to a B-Pareto front for a certain type of MOPs, namely, the regular multi-objective optimization problems (RMOPs).

This paper is structured as follows. By taking into account both convergence and diversity, Section 2 presents the definition of convergence to a B-Pareto set, which is a new concept that provides a finite representation of a Pareto set. In Section 3, convergence to a B-Pareto set of an RMOP is proved based on the definition of a convergence metric and the newly proposed updating strategy. In Section 4, numerical experiments are conducted to verify the convergence results and to show the rationality of the B-Pareto set definition. Some discussion points are presented in Section 5. Finally, some conclusions are addressed in Section 6.

2. Combining diversity with convergence to the Pareto front

MOEAs are currently widely utilized to solve MOPs

$$\min F(x) = (f_1(x), f_2(x), \ldots, f_m(x)), \tag{1}$$

where $x = (x_1, \ldots, x_n) \in S_x \subseteq \mathbb{R}^n$ and $y = (y_1, \ldots, y_m) = (f_1(x), \ldots, f_m(x)) \in S_y \subseteq \mathbb{R}^m$. S_x, the set of all feasible solutions, is called the *feasible space*, and $S_y = F(S_x)$ is called the *objective space*.

To compare the feasible solutions of an MOP, the definition of Pareto dominance is often used [12]. In this paper, the definition of dominance is provided for objective space S_y.

Definition 1. Let S_x and S_y be the feasible space and the objective space of MOP (1) respectively. $y^{(1)} = \left(y_1^{(1)}, \ldots, y_m^{(1)}\right)$ and $y^{(2)} = \left(y_1^{(2)}, \ldots, y_m^{(2)}\right)$ are two vectors in S_y.

1. (**Pareto dominance**) $y^{(1)}$ is said to (**Pareto**) **dominate** $y^{(2)}$(denoted as $y^{(1)} \prec y^{(2)}$) i.f.f.
   (a) $y^{(1)}$ **weakly dominates** $y^{(2)}$ (denoted as $y^{(1)} \preceq y^{(1)}$), i.e., $\forall i \in \{1, \ldots, m\} : y_i^{(1)} \leqslant y_i^{(2)}$;
   (b) $\exists j \in \{1, \ldots, m\} : y_j^{(1)} < y_j^{(2)}$.
2. (**Pareto set** & **Pareto front**)
   (a) $x \in S_x$ is called a **Pareto solution** of MOP (1) if there does not exist a feasible solution x' satisfying $F(x') \prec F(x)$;
   (b) The set of all Pareto solutions of MOP (1) is called the **Pareto set** of MOP (1), denoted as P_{S_x};
   (c) $PF = F(P_{S_x}) \subseteq S_y$ is called the **Pareto front** of MOP (1).

The entire Pareto front is usually not a single point but a set of points. Because of the limitations of our computational abilities, only a finite subset of the Pareto front can be generated by MOEAs. Therefore, the purpose of global optimization is to obtain a finite approximation of the Pareto front that is not only close to the true Pareto front but that is also well-distributed. To obtain a diverse and precise approximation of the Pareto front, MOPs are solved by MOEAs with a fixed population size and a fixed archive size. A universal framework for MOEAs is illustrated with Algorithm 1, where $\xi^{(t)}$ is the population at generation t, and $\eta^{(t)}$ is the intermediate population obtained using the "Generate ()" function. During the process of evolution, the archive $\mathcal{A}^{(t)}$ and population $\xi^{(t)}$ are updated using the "ArchiveUpdate ()" function and "PopUpdate ()" functions, respectively.

Algorithm 1. Multi-objective evolutionary algorithm (MOEA)

```
Set generation counter t := 1;
Initialize ξ^(t) randomly;
A^(t) := ξ^(t);
for t = 1,2,... do
    η^(t) := Generate(ξ^(t));
    A^(t+1) := ArchiveUpdate(A^(t), η^(t));
    ξ^(t+1) := PopUpdate (ξ^(t), η^(t));
end for
Output the results
```

2.1. A uniform representation of the Pareto front

As a finite approximation of the true Pareto front, the final results of an MOEA should be distributed uniformly along it. Many metrics have been defined to measure the diversity of a finite representation [12], and several updating strategies have been proposed to maintain an appropriate level of diversity and an approximately high quality with respect to the final results [13,32,16,2,22,41–43]. Some of the most favorable strategies are based on ϵ-dominance, which leads to the convergence of MOEAs to the ϵ-Pareto set [32,43].

The ϵ-Pareto set controls the distance between two adjacent points of the approximate Pareto front via ϵ, and accordingly, a finite representation of a Pareto set is defined. However, with respect to its size and members, the ϵ-Pareto set cannot be determined by solely the value of ϵ. Moreover, gaps may exist in the PF approximation [44], and determining ϵ during engineering computations is difficult because of the unknown properties of MOPs. Furthermore, the boundary points of the Pareto front are also usually difficult to obtain. Therefore, to overcome these difficulties we propose the concept of a B-Pareto set, which maintains the diversity in the finite representation of the true Pareto front with a fixed archive size.

Definition 2. Let $y^{(1)}$ and $y^{(2)}$ be two vectors in S_y. $\forall\delta \in \mathbb{R}^+$, $y^{(1)}$ is said to **weakly δ-ball dominate** $y^{(2)}$ (in short $y^{(1)} \preceq_\delta y^{(2)}$) with respect to MOP (1) if there exists at least one point $y \in U(y^{(1)},\delta)$ such that $y \preceq y^{(2)}$, where $U(y^{(1)},\delta) = \{y \in \mathbb{R}^m; \|y - y^{(1)}\|_2 \leqslant \delta\}$.

A weaker version of dominance, which is denoted weak δ-ball dominance, is defined in Definition 2, and its comparison with Pareto dominance is illustrated in Fig. 1. All vectors in the neighborhood $U(y^{(1)},\delta)$ are weakly δ-ball dominated by $y^{(1)}$, and vectors dominated by $y \in U(y^{(1)},\delta)$ are also weakly δ-ball dominated by $y^{(1)}$. Then, if $y^{(2)} \in \mathbb{R}^m$ is not weakly δ-ball dominated by $y^{(1)}$, the distance between $y^{(1)}$ and $y^{(2)}$ is necessarily greater than δ. We provide an example to explain how to determine the weak δ-ball dominance relation.

Example 1 (*Determination of the weak δ-ball dominance relation*). Let $y = (y_1,\ldots,y_m)$ and $y' = (y'_1,\ldots,y'_m)$ be two vectors in S_y. For a given $\delta > 0$,

1. if $y \preceq y'$, then $y \preceq_\delta y'$;
2. if y does not weakly dominate y', $\exists j_1,\ldots,j_k \in \{1,\ldots,m\}$ such that
$$y_{j_l} > y'_{j_l}, \quad l = 1,2,\ldots,k.$$

Denote $y'' = (y''_1,\ldots,y''_m)$, where $y''_i = \min\{y_i, y'_i\}, i = 1,\ldots,m$. Then $y \preceq_\delta y'$ if and only if $\|y - y''\|_2 \leqslant \delta$. □

Then, the so-called weak δ-Pareto front, a finite representation of the Pareto front, is obtained by Definition 3.

Definition 3. Let $\delta \in \mathbb{R}^+$.

1. A set $P_\delta \subseteq \mathbb{R}^m$ is called a **weak δ-approximate Pareto front** of MOP (1) if every point $y \in S_y$ is weakly δ-ball dominated by at least one $y' \in P_\delta$.
2. A subset P^*_δ of PF is called a **weak δ-Pareto front** if P^*_δ is a weak δ-approximate Pareto front of MOP (1).

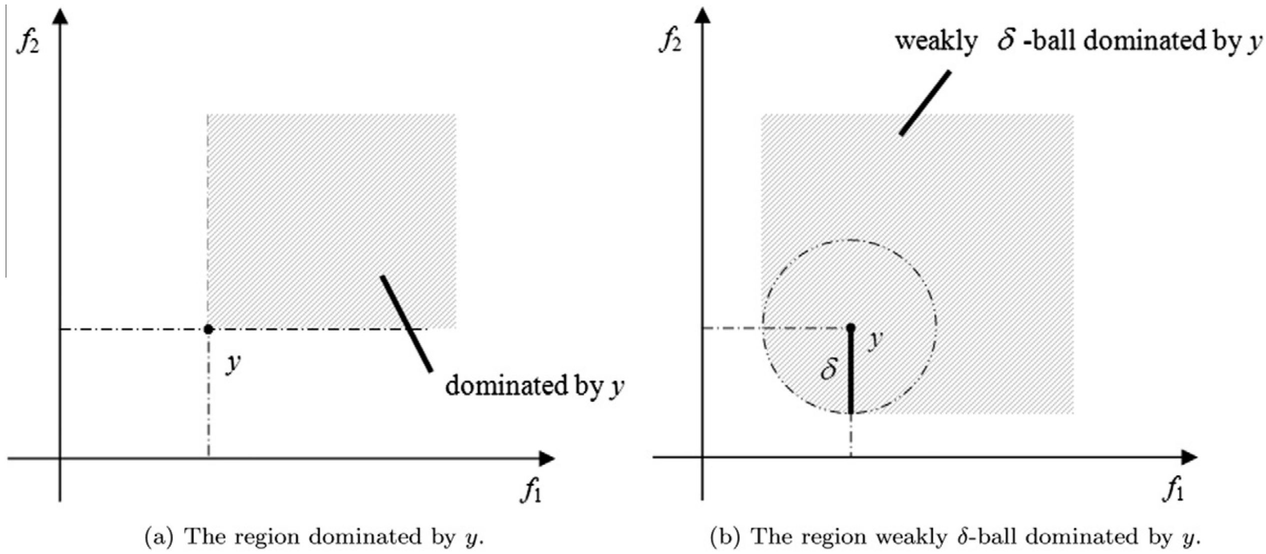


Fig. 1. The comparisons between dominance and weak δ-ball dominance. The weak δ-ball dominance is a weak version of dominance, and the distance between two points which are not weakly δ-ball dominated by each other is greater than δ.

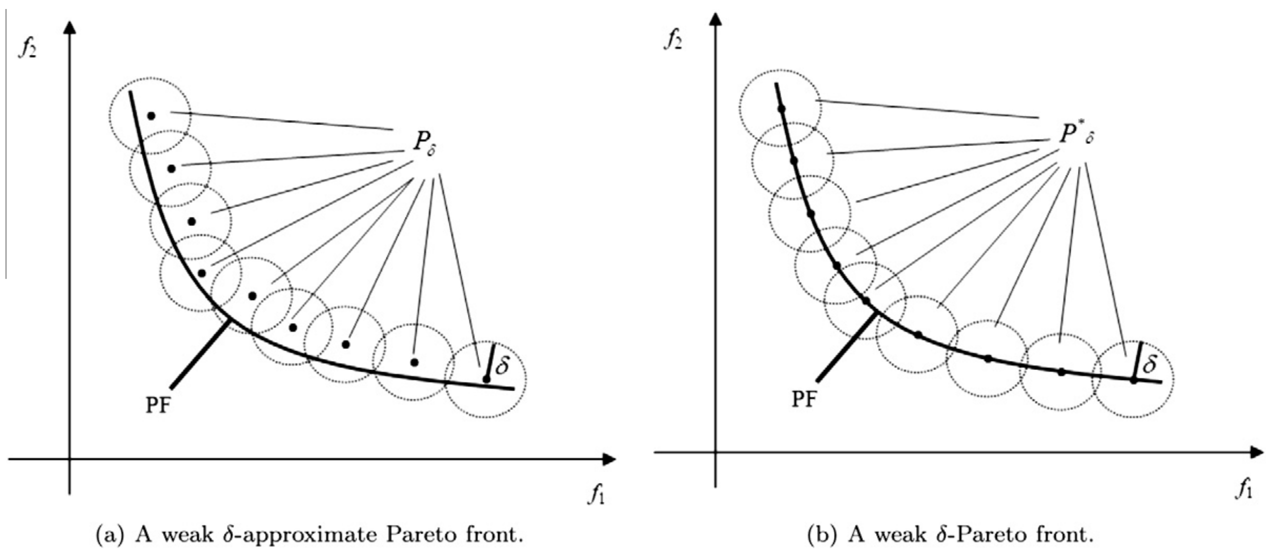


Fig. 2. The illustrations of a weak δ-approximate Pareto front (a) and a weak δ-Pareto front (b). The neighborhoods of all points in a weak δ-approximate Pareto front constitute a cover of the Pareto front. When the points are all located on the Pareto front, it comes to a weak δ-Pareto front.

Fig. 2 illustrates the difference between a weak δ-approximate Pareto front and a weak δ-Pareto front. Note that the δ-neighborhoods of all points in P_δ constitute a cover of the Pareto front, i.e., $\forall y \in PF$, there exists a $y' \in P_\delta$ satisfying $y \in U(y', \delta)$. When all points in P_δ are located on the Pareto front, the weak δ-approximate Pareto front becomes a weak δ-Pareto front. Thus, the value of δ also affects the size of P^*_δ, which influences the complexity of updating processes in MOEAs. A large value of δ will lead to a sparsely distributed P^*_δ, whereas a small value of δ will cause a large weak δ-Pareto set of high complexity. To eliminate the difficulty of choosing an appropriate value of δ, this value is set at the minimum distance between two vectors in a weak δ-Pareto front.

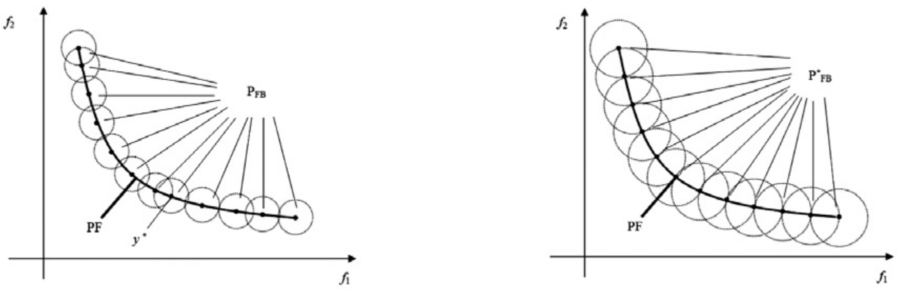
Definition 4. Let P_{FB} be the set of N vectors on the Pareto front of MOP (1), and let

$$\delta_{y^*} = dist(y^*, P_{FB} \setminus \{y^*\}) = \min_{y \in P_{FB} \setminus \{y^*\}} \|y - y^*\|_2, \quad y^* \in P_{FB}.$$

P_{FB} is said to be an **approximate** B**-Pareto front of MOP (1) of size** N, if $P_{FB} \setminus \{y^*\}$ is a weak δ_{y^*}-Pareto front of MOP (1) for a $y^* \in P_{FB}$ with

$$dist(y^*, P_{FB} \setminus \{y^*\}) = \min_{\substack{y \neq y' \\ y, y' \in P_{FB}}} \|y - y'\|_2.$$

The inverse image P_{SB} of P_{FB} is called an **approximate** B**-Pareto set of MOP (1) of size** N.



(a) An approximate B-Pareto front of MOP (1) of size N.

(b) A B-Pareto front of MOP (1) of size N.

Fig. 3. Visualizations of an approximate B-Pareto front (a) and a B-Pareto front (b) of MOP (1) of size N. The set of N neighborhoods $\{U(y,\delta^*); y \in P_{FB}\}$ constitute a finite cover of the Pareto front, where δ^* is the minimum distance between any two points in P_{FB}. When δ^* is maximized subject to the size N, it becomes to a B-Pareto front of MOP (1) of size N.

Fig. 3(a) shows a two-dimensional case of the approximate B-Pareto front. Here, the radius of neighborhoods is determined by N automatically, differing from the definition of the weak δ-approximate Pareto front. The value of $\delta_{y^*} = dist(y^*, P_{FB} \setminus \{y^*\})$ is not arbitrarily small for a given MOP with a bounded Pareto front because the definition of P_{FB} requires that all vectors on PF are weakly δ_{y^*}-ball dominated by at least one point in $P_{FB} \setminus \{y^*\}$, i.e., $\{U(y, \delta_{y^*}); y \in P_{FB} \setminus \{y^*\}\}$ is a finite cover of the Pareto front. When

$$\delta^* = \min_{y^* \in P_{FB}} \delta_{y^*} = \min_{\substack{y \neq y' \\ y, y' \in P_{FB}}} \|y - y'\|_2$$

is maximized subject to the given size N, it approaches the definitions of both the B-Pareto front and the B-Pareto set.

Definition 5. Denote $\mathcal{P_F}$ as the set of all approximate B-Pareto fronts of MOP (1) of size N. $P^*_{FB} \in \mathcal{P_F}$ is said to be a B-**Pareto front of MOP (1)** of size N if

$$div(P^*_{FB}) = \sup_{P_{FB} \in \mathcal{P_F}} div(P_{FB}),$$

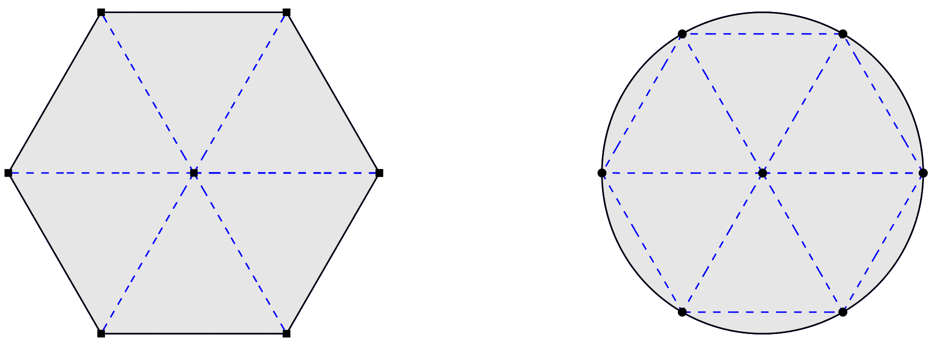
where $div(X) = \min_{\substack{x \neq x' \\ x, x' \in X}} \|x - x'\|_2$ if X is a set of N vectors in $\mathbb{R}^m$. The inverse image P^*_B of P^*_{FB} is called a B-**Pareto set of MOP(1) of size** N.

Definition 5 provides a new definition for a finite representation of a Pareto set, namely, the B-Pareto set. Under some mild conditions, the existence of the supremum of $div(P_{FB})$ can be obtained. For example, when the objective space S_y of MOP (1) is compact, the maximum value of $div(P_{FB})$ can be achieved from the continuity of $div(P_{FB})$. Furthermore, it can make members of F^*_{FB} distributed as uniformly as possible by maximizing $div(P_{FB})$ on the condition that its cardinality remains unchanged. Specifically, the B-Pareto front of size N, as illustrated in Fig. 3(b), contains N points distributed uniformly along the true Pareto front when it is simply connected. In addition, the automatic determination of the diversity of the B-Pareto set avoids the problem of choosing an appropriate value for the given precision, which greatly influences the efficiency of MOEAs.

In fact, the B-Pareto front of an MOP of size N is not necessarily determined uniquely. Fig. 4(a) and (b) describes two special cases. When the Pareto front is a regular hexagon, the B-Pareto front of size seven is the set of points illustrated by real squares (Fig. 4(a)). However, if the Pareto front is a disc, then the B-Pareto front of size seven is not uniquely determined, as illustrated by real circles in Fig. 4(b). In this case, one point on the B-Pareto front is the center of the disk, and the other six points are the vertices of any inscribed regular hexagon of the circle.

2.2. Definition of convergence

Most importantly, MOEAs should converge to the global Pareto set. Two popular ways of implementing convergence analysis of EAs are as follows. The first approach involves modeling the EAs as Markov chains such that the convergence of EAs can be investigated by studying the properties of Markov chains [35,10,39,38]. This method is widely applied to analyze EAs with finite state space and has been extended to the convergence analysis of MOEAs [40,47]. In the second approach, based on the objective values of individuals, an additional metric function is defined to study the convergence of EAs. In this case, the stochastic convergence of EAs can be studied by analyzing the convergence of a random variable sequence (*r.v.s.*). The theoretical analyses of real-coded EAs are mainly studied in this way, as in [39,3,15,4,5,26,27]. Additionally, this method has been applied by He and Yao [28,29] to analyses of some discrete-coded EAs, and by Laummans et al. [33] to successfully analyze the convergence properties of a discrete-coded MOEA.



(a) The illustration of the uniquely determined B-Pareto front of size seven.

(b) The illustration of one of the B-Pareto fronts of size seven.

Fig. 4. Two cases of the B-Pareto front. In case (a), the B-Pareto front of size seven is uniquely determined, and in case (b), the seven points in B-Pareto front of size seven is not uniquely determined, where the six points on the circle can be located on the vertexes of any inscribed regular hexagon of the circle.

In this paper, we will also define a scalar function to compare feasible solutions, and consequently the convergence of MOEAs is investigated by defining a convergence metric and a diversity metric. Convergence to a B-Pareto set is defined by the convergence of two different metrics. First, we introduce the stochastic convergence of an *r.v.s.*.

Definition 6. Let $\{X^{(t)}, t = 1,2,\ldots\}$ be an *r.v.s.* $\{X^{(t)}; t = 1,2,\ldots\}$ is said to **converge almost surely to** X^*, denoted by $X^{(t)} \xrightarrow{a.s.} X^*$, if

$$P\left(\lim_{t\to\infty} X^{(t)} = X^*\right) = 1.$$

To quantify the closeness a feasible solution to the Pareto set, a non-negative function $d(x)$ should be defined to satisfy (1) $d(x) \geqslant 0$ and (2) $d(x) = 0 \iff x \in P_{S_x}, \forall x \in S_x$. Then, the convergence metric and the diversity metric of archives are defined as follows:

Definition 7. Let $\mathcal{A}$ be the archive of size N in an MOEA. The convergence metric $D(\mathcal{A})$ and diversity metric $DIV(\mathcal{A})$ are respectively defined as

$$D(\mathcal{A}) = \frac{1}{N}\sum_{x\in\mathcal{A}} d(x) \tag{2}$$

and

$$DIV(\mathcal{A}) = \min_{\substack{x\neq x'\\ x,x'\in\mathcal{A}}} \|F(x) - F(x')\|_2. \tag{3}$$

By (2), the convergence metric of archives is defined such that (1) $D(\mathcal{A}) \geqslant 0$ and (2) $D(\mathcal{A}) = 0 \iff \mathcal{A} \subseteq P_{S_x}, \forall \mathcal{A} \subseteq S_x$. As such, it quantifies the distance from $\mathcal{A}$ to the true Pareto set. The metric $DIV(\mathcal{A})$, which is defined as the minimum distance between any two solutions, reaches the maximum value when $\mathcal{A}$ becomes the B-Pareto set. The following example shows the feasibility of constructing these two metrics.

Example 2 (*The convergence metric and diversity metric for discrete-coded MOEAs).* Let $x \in \mathcal{A}$, where $\mathcal{A}$ is the archive of a discrete-coded MOEA of size N. Then non-dominated sorting [13] can be performed on the solutions space to evaluate the feasible solutions. Define the function

$$d(x) = rank(x) - 1,$$

where $rank(x)$ is the rank of feasible solution x in the feasible space S_x. The value of $rank(x)$ is 1 when $F(x)$ belongs to the first non-dominated front, 2 when $F(x)$ belongs to the second non-dominated front, and so on. It follows that (1) $d(x) \geqslant 0$ and (2) $d(x) = 0 \iff x \in P_{S_x}, \forall x \in S_x$. Thus, the convergence metric of the MOEA defined by

$$D(\mathcal{A}) = \frac{1}{N}\sum_{x\in\mathcal{A}} d(x)$$

satisfies (1) $D(\mathcal{A}) \geqslant 0$ and (2) $D(\mathcal{A}) = 0$ if and only if $\mathcal{A} \subseteq P_{S_x}, \forall \mathcal{A} \subseteq S_x$. Meanwhile, $DIV(\mathcal{A})$ can also be defined according to (3). $\square$

According to Definition 6, the convergence of an MOEA can be defined by the convergence of $\{D(\mathcal{A}^{(t)}); t = 1,2,\ldots\}$.

Definition 8. Let P_{S_x} be the Pareto set of MOP (1), and let $\{\mathcal{A}^{(t)}; t = 1, 2, \ldots\}$ be the archive sequence of an MOEA. The MOEA is said to **converge almost surely** to P_{S_x} if $D(\mathcal{A}^{(t)}) \xrightarrow{a.s.} 0$, i.e.,

$$P\Big(\lim_{t\to\infty} D(\mathcal{A}^{(t)}) = 0\Big) = 1.$$

Here $D(\cdot)$ is the convergence metric defined by (2).

Definition 8 describes the convergence of MOEAs by that of *r.v.s.* without respect to the diversity of the final approximation of the Pareto set. If the diversity sequence $\{DIV(\mathcal{A}^{(t)}); t = 1, 2, \ldots\}$ also converges, we say that the MOEA converges almost surely to a B-Pareto set of MOP (1) of size N.

Definition 9. Let $\{\mathcal{A}^{(t)}, t = 1, 2, \ldots\}$ be the archive sequence of an MOEA of size N. The MOEA is said to **converge almost surely to the B-Pareto set of MOP (1) of size** N if

1. it converges almost surely to the P_{S_x} of MOP (1);
2. $\{DIV(\mathcal{A}^{(t)}), t = 1, 2, \ldots\}$ converges almost surely to $\sup_{P_{SB}\in\mathcal{P}_S} DIV(P_{SB})$, where $\mathcal{P}_S$ is the set of all approximate B-Pareto sets of MOP (1) of size N.

Definition 5 presents the notion of a B-Pareto set using the uniform distribution of the B-Pareto front, and Example 2 shows that the convergence metric and diversity metric of the archive can also be defined according to the distribution of $F(\mathcal{A}^{(t)})$. Thus, the convergence to the B-Pareto set is defined in Definition 9. If the image $F(\mathcal{A}^{(t)})$ of the archive approximates the B-Pareto front in objective space S_y, we say that the archive sequence $\mathcal{A}^{(t)}$ converges to the B-Pareto set in feasible space S_x. In fact, a similar definition can be performed on the population sequence of MOEAs without an archive if the population sequence $\{\xi^{(t)}; t = 1, 2, \ldots\}$ of the MOEA is used to define the convergence metric $D(\cdot)$ and diversity metric $DIV(\cdot)$. In this paper, $D(\mathcal{A})$ and $DIV(\mathcal{A})$ are both defined using the archive of the MOEA described in Algorithm 1.

3. Convergence to a *B*-Pareto front for regular multi-objective optimization problems (RMOPs)

As an initial analysis, we consider MOPs with bounded and simply connected Pareto fronts, known as regular multi-objective optimization problems (RMOPs). An example of an RMOP is as follows:

$$\begin{cases} \min F(x) = (f_1(x), f_2(x), \ldots, f_m(x)), \\ x \in [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n. \end{cases} \tag{4}$$

Assume that $F = (f_1, f_2, \ldots, f_m)$ is convex and continuous. The following investigates the convergence of an MOEA to its B-Pareto set.

3.1. Construction of the convergence metric

As presented above, the convergence metric of an MOEA is based on the definition of $d(x)$. First, we discuss some properties of the Pareto front of MOP (1).

Lemma 1. $\forall i \in \{1, 2, \ldots, m\}$, *there exists an implicit function*

$$y_i = h_i(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)$$

defined by the Pareto front of MOP (1).

Proof. Suppose that $y^{(1)} = (y_1, \ldots, y_{i-1}, y_i^{(1)}, y_{i+1}, \ldots, y_m)$ and $y^{(2)} = (y_1, \ldots, y_{i-1}, y_i^{(2)}, y_{i+1}, \ldots, y_m)$ are two different points in the objective space S_y of MOP (1), where $i \in \{1, \ldots, m\}$. Then either $y^{(1)} \prec y^{(2)}$ or $y^{(2)} \prec y^{(1)}$ holds. Thus $y^{(1)}$ and $y^{(2)}$ cannot be simultaneously located on the Pareto front and there exists an implicit function $y_i = h_i(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)$ defined by the Pareto front of MOP (1). □

Using Lemma 1, we can now denote the Pareto front of RMOP (4) as

$$G = \{y = (y_1, \cdots, y_m) \in S_y; y_i = h_i(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)\}, \quad \forall i = 1, 2, \ldots, m.$$

Then

$$G_{h_i}(C_i) = \{y = (y_1, \ldots, y_m) \in S_y; y_i = h_i(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m) + C_i\} \tag{5}$$

represents a bounded hypersurface parallel to PF when C_i is a non-negative real number. So, $\forall i \in \{1, 2 \ldots, m\}$, the objective space S_y of RMOP (4) can be divided into two parts:

1. $\mathcal{S}_i$, in which all points are located on $G_{h_i}(C_i)$ for some given non-negative number C_i;

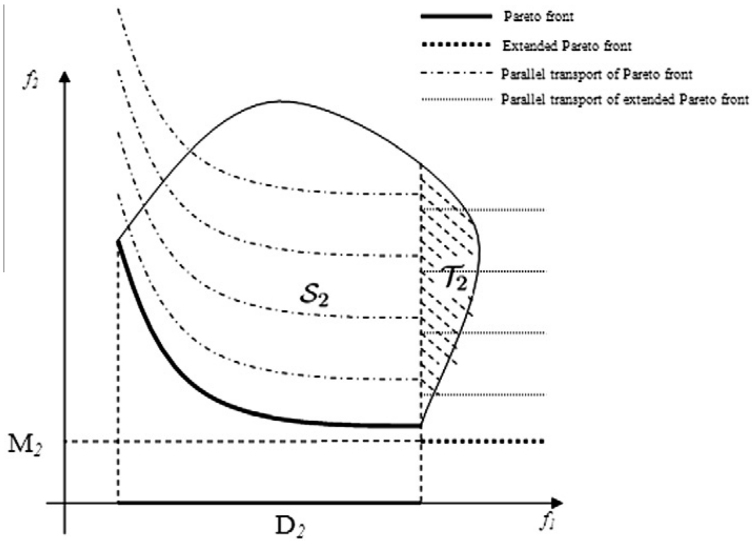


Fig. 5. The objective space S_y of RMOP (4) for a two-dimensional case. Here M_2 is a number equal to or less than the infimum of y_2. The Pareto front can be denoted as $G = \{y = (y_1,y_2) \in S_y; y_2 = h_2(y_1)\}$, where D_2, the definition domain of h_2, is the projection of $\mathcal{S}_2$. Any point in $\mathcal{S}_2$ can be included in $G_{h_2}(C_2) = \{y = (y_1, y_2) \in S_y; y_2 = h_2(y_1) + C_2\}$ for some given non-negative number C_2, but all points in $\mathcal{T}_2$ cannot be included in $G_{h_2}(C_2) = \{y = (y_1, y_2) \in S_y; y_2 = h_2(y_1) + C_2\}$ for any given non-negative number C_2. If we extend h_2 to H_2 according to (6), then any point in S_y can be included in $G_{H_2}(C_2) = \{y = (y_1, y_2) \in S_y; y_2 = H_2(y_1) + C_2\}$ for one and only one given non-negative number C_2.

2. $\mathcal{T}_i$, in which all points cannot be located on $G_{h_i}(C_i)$ for any non-negative number C_i.

See Fig. 5 for an illustration of a two-dimensional case. It is obvious that when S_y is convex, the points in $\mathcal{T}_2$ are not located on the curve $y_2 = h_2(y_1) + C_2$ for any non-negative real number C_2, and thus, the Pareto front PF should be extended to cover all points in S_y. Define

$$H_i(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m) = \begin{cases} h_i(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m), & \text{if } (y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m) \in D_i; \\ M_i, & \text{if } (y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m) \notin D_i; \end{cases} \tag{6}$$

Note that D_i, the definition domain of $h_i(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m)$, is the projection of PF on the subspace $\{y = (y_1,\ldots,y_m) \in \mathbb{R}^m; y_i = 0\}$, and M_i is a given real number equal to or less than the infimum of y_i. Then, any point $y = (y_1,\ldots,y_m)$ in S_y can be given a non-negative value C_i described as $d_i(y_1,\ldots,y_m)$, because it is on

$$G_{H_i}(C_i) = \{y = (y_1,\ldots,y_m) \in S_y; y_i = H_i(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_m) + C_i\}$$

for one and only one $C_i \geqslant 0$. Taking $\sum_{i=1}^m C_i$ as the measure of an individual x, we have the following definition.

Definition 10. $\forall x \in S_x$, the non-negative function

$$d(x) = \sum_{i=1}^m d_i(f_1(x),\ldots,f_m(x)) \tag{7}$$

is called the **fitness function of individual** x.

The fitness function $d(x)$ identifies the difference between x and the Pareto set and has the following properties.

Lemma 2. *Let $y^{(1)} = F(x^{(1)})$ and $y^{(2)} = F(x^{(2)})$. If $y^{(1)} \prec y^{(2)}$, then $d(x^{(1)}) < d(x^{(2)})$.*

Proof. See Appendix A for the proof. □

Lemma 3. *When the MOEA is applied to solve RMOP* (4), *it holds that $d(x) \geqslant 0$, $\forall x \in S_x$. Furthermore, $d(x) = 0$ if and only if $x \in P_{S_x}$.*

Proof. From the definition of $d(x)$, $d(x) \geqslant 0$, $\forall x \in S_x$. Furthermore, if $x \in P_{S_x}$ (i.e., $F(x) \in PF$), $d(x)$ is equal to zero. In contrast, if $x \notin P_{S_x}$, there exists $x^* \in P_{S_x}$ such that $x^* \prec x$, i.e., $0 = d(x^*) < d(x)$. Thus, this results in $d(x) = 0$ if and only if $x \in P_{S_x}$. □

Then the convergence metric of MOEAs can be achieved by (2). In the following section, we investigate the MOEA that guarantees convergence to a B-Pareto set of the RMOP (4).

3.2. Description of the MOEA

We apply the MOEA described in Algorithm 1 to solve RMOP (4), where the "*Generate ()*" and "*ArchiveUpdate ()*" functions are illustrated with Algorithms 2 and 3, respectively.

Algorithm 2. "Generate ()" Function

Input: ξ;
Set $\eta = \emptyset$;
for all $x \in \xi$
  $x' = \text{mutate}(x)$;
  $\eta = \eta \cup \{x'\}$;
end for
Output η

Algorithm 3. "ArchiveUpdate ()" function

Input: $\mathcal{A}, \eta$;
for all non-dominated solutions x' in η **do**
  for all $x \in \mathcal{A}$ **do**
    if $x' \prec x$ **then**
      $\mathcal{A} = \mathcal{A} \cup \{x'\} \setminus \{x\}$;
      BREAK; {'BREAK' means it will end the inner loop}
    end if
  end for
  if $\nexists x \in \mathcal{A} : x \prec x'$ **then**
    if $\exists x \in \mathcal{A} : x' \lhd x$ **then**
      $\mathcal{A} = \mathcal{A} \cup \{x'\} \setminus \{x\}$;
    end if
  end if
end for
if $rand() < p$ **then**
  Set $\mathcal{A}' = \emptyset$;
  for all $x \in \mathcal{A}$ **do**
    $\mathcal{A}' = \mathcal{A}' \cup \text{mutate}(x)$;
  end for
  if $\mathcal{A}' \lhd \mathcal{A}$ **then**
    $\mathcal{A} = \mathcal{A}'$;
  end if
end if
Output $\mathcal{A}$

1. In Algorithm 2, η is the intermediate population consisting of candidates generated by function "mutate ()" with the mutation operator

$$x' = x + z, \tag{8}$$

where the continuous joint density function of z is denoted as $\varphi(z,x)$. Although the "Generate ()" functions of most state-of-the-art MOEAs consist of both crossover and mutation operators, it is appropriate to consider one merely consisting of a mutation operator because some crossover operators (particularly for real-coded MOEAs such as linear crossover [12] and simulated binary crossover (SBX) [11]) can be thought of as mutation operators performed on any one of the individuals included in the crossover operations.

2. Algorithm 3 describes the archive-updating strategies utilized in this study by which the archive $\mathcal{A}^{(t+1)}$ is obtained from $\mathcal{A}^{(t)}$ and $\eta^{(t)}$. Example 1 shows that the determination of weak δ-ball dominance is more complicated than Pareto dominance and ϵ-dominance, and thus, in the algorithm, the B-Pareto set is generated using a composed updating strategy without the computation of weak δ-ball dominance. Here, $x' \lhd x$ means that for two feasible solutions x and x', it holds that

$$d(x') < d(x)$$

and

$$\min_{y \in F(\mathcal{A} \setminus \{x\})} \|F(x') - y\|_2 > \min_{y \in F(\mathcal{A} \setminus \{x\})} \|F(x) - y\|_2.$$

$\mathcal{A}' \lhd \mathcal{A}$ indicates that

$$D(\mathcal{A}') < D(\mathcal{A})$$

and

$$DIV(\mathcal{A}') > DIV(\mathcal{A})$$

hold at the same time. $Size(\mathcal{A})$ returns the cardinality of $\mathcal{A}$, and $rand()$ returns a random number uniformly distributed in $(0,1)$. Three different updating strategies of archives are as follows:

S1 If there exist in $\mathcal{A}$ some individuals dominated by the newly generated candidate x', one of the dominated individuals is deleted from the archive and x' is added to the archive.

S2 If the new candidate x' is non-dominated with all individuals in $\mathcal{A}$, an individual $x \in \mathcal{A}$ will be updated by x' if $x' \lhd x$.

S3 With a given probability p, an intermediate archive $\mathcal{A}'$ is generated by the mutations performed on all the individuals in $\mathcal{A}$. $\mathcal{A}'$ will replace $\mathcal{A}$ if $\mathcal{A}' \lhd \mathcal{A}$.

3.3. Convergence to the B-Pareto set of RMOP (4)

Based on the above preliminary work, the convergence theorems of MOEAs can be obtained. The following results are introduced and subsequently proven.

Lemma 4. *Let $\phi(z,x) > 0$ be a continuous function defined in $\mathbb{R}^n \times S_x$. $\forall \delta > 0, x, x^* \in S_x$,*

$$L_\delta(x^*, x) = \int_{x+z \in U(x^*,\delta) \cap S_x} \phi(z,x) dz$$

is continuous in $S_x \times S_x$, where $U(x^*, \delta) = \{x \in \mathbb{R}^n; \|x - x^*\| \leqslant \delta\}$.

Proof. See Appendix B for the proof. □

Lemma 5. *$\forall \varepsilon > 0, x \in S_x$, the new feasible solution x' generated by (8) is located in $U(x^*, \varepsilon)$ with a positive probability, i.e.,*

$$P_\varepsilon(x^*, x) \triangleq \mathbf{P}\{x' \in U(x^*, \varepsilon) \cap S_x\} > 0,$$

where $U(x^*, \varepsilon) = \{x' \in \mathbb{R}^n; \|x' - x^*\|_2 \leqslant \varepsilon\}$ is the neighborhood of any feasible solution x^*.

Proof. Because the feasible space S_x of RMOP (4) is a hypercube in $\mathbb{R}^n$, $S_x \cap U(x^*, \varepsilon)$ is a compact set with a positive Lebesgue measure. Because $\varphi(z,x) > 0$ is continuous in S_x, we see that its minimum value exists in $S_x \cap U(x^*, \varepsilon)$ and is greater than zero. Thus, we conclude that

$$P_\varepsilon(x^*, x) = \mathbf{P}\{x' \in S_x \cap U(x^*, \varepsilon)\} = \int_{z \in S_x \cap U(x^*, \varepsilon) - x} \varphi(z,x) dz > 0$$

for all $x, x^* \in S_x$, where $S_x \cap U(x^*, \varepsilon) - x = \{y; y + x \in S_x \cap U(x^*, \varepsilon)\}$. □

Remark. $\forall x \in S_x$, $x' = x + z \in S_x$, $\varphi(z,x) > 0$ means that the probability density function of the newly generated candidate x', which is a random vector in $\mathbb{R}^n$, is always greater than zero when $x' \in S_x$. The joint density function $\phi(z,x)$ satisfying this condition is usually utilized in real-coded EAs. For example, in evolutionary programmings (EPs), z is usually set as a random vector with a Gaussian distribution, the joint density function of which is continuous and positive for all $z \in \mathbb{R}^n$. In addition, some other distributions, such as the Leѵ́y distribution, also have joint density functions satisfying the condition of Lemmas 4 and 5.

Proposition 1. *Suppose that the size of the archive is N. $\forall \varepsilon > 0$, $t > 0$, there exists a real number $M(\varepsilon)$ located in $(0,1]$ such that*

$$\mathbf{P}\{D(\mathcal{A}^{(t+1)}) < \varepsilon | D(\mathcal{A}^{(t)}) > \varepsilon\} \geqslant M(\varepsilon)$$

if the continuous joint density function $\varphi(z,x)$ *in* (8) *is greater than zero for all* $x \in S_x$, $z \in S_x - x$.

Proof. Of the three strategies **S1**, **S2** and **S3**, only **S1** and **S2** could possibly influence convergence. Therefore, only strategy **S1** is considered in this proof because strategies **S2** and **S3** both ensure that the convergence metric value $D(\mathcal{A}^{(t)})$ is non-increasing.

If $D(\mathcal{A}^{(t)}) > \varepsilon$, there exists at least one individual $x \in \mathcal{A}^{(t)}$ with $d(x) > \varepsilon$. By Lemma 3, we know that x is not a Pareto solution. Thus, there exists a Pareto solution $x^* \in P_{S_x}$ such that $F(x^*) \prec F(x)$. From the regularity of the RMOPs, there exists a $\delta > 0$ such that $d(x') < \varepsilon$, $\forall x' \in U(x^*, \delta) \cap S_x$. Moreover, Lemma 5 shows that the new individual x' generated from x by (8) satisfies

$$P_\delta(x^*, x) \triangleq \mathbf{P}\{x' \in U(x^*, \delta) \cap S_x\} > 0.$$

That is, according to **S1**, an individual $x \in \mathcal{A}^{(t)}$ with $d(x) > \varepsilon$ is updated by a candidate solution $x' \in \eta^{(t)}$ with probability

$$P_\delta(x^*, x) \triangleq \mathbf{P}\{x' \in U(x^*, \delta) \cap S_x\} > 0,$$

where $d(x') < \varepsilon$. Moreover, $\forall x^* \in P_B^*$, Lemma 4 shows that $P_\delta(x^*,x)$ is continuous, and as a result, its positive minimum value P_δ exists. Let μ be the number of individuals in $\mathcal{A}^{(t)}$ with $d(x) > \varepsilon$. Then, updating strategy **S1** ensures that with a probability greater than $(P_\varepsilon)^\mu$, all the individuals x in $A^{(t+1)}$ satisfy $d(x) < \varepsilon$. Therefore, if all three updating strategies **S1**, **S2** and **S3** are included in the updating process,

$$\mathbf{P}\{D(\mathcal{A}^{(t+1)}) < \varepsilon | D(\mathcal{A}^{(t)}) > \varepsilon\} \geqslant P_\varepsilon^\mu \triangleq M(\varepsilon) > 0$$

always holds. Because P_ε is the positive minimum value of $\mathbf{P}\{x' \in U(x^*, \varepsilon)\}$, we have $0 < P_\varepsilon \leqslant 1$. Accordingly, $M(\varepsilon) \triangleq P_\varepsilon^\mu$ is located in $(0,1]$. □

Proposition 1 shows that if the convergence metric value of the archive is greater than a given positive value ε, then the next archive will have a convergence metric of ε with a positive probability. Then, by the non-increasing $D(\mathcal{A}^{(t)})$, we come to the conclusion that $D(\mathcal{A}^{(t)})$ converges almost surely to zero, i.e., the MOEA consisting of Algorithms 1, 2 and 3 converges almost surely to the Pareto set. To show this result, an elementary probability lemma is introduced [34].

Lemma 6. *Let* $\{X(t), t = 1,2,\ldots\}$ *be an r.v.s. If*

$$\sum_{t=1}^{+\infty} \mathbf{P}\left\{|X(t)| \geqslant \frac{1}{l}\right\} < +\infty$$

for every positive integer l, then $P\{X(t) \nrightarrow 0\} = 0$, *where* $X(t) \nrightarrow 0$ *indicates that X(t) does not converge to zero when t tends to infinity.*

Theorem 1. *The MOEA consisting of Algorithms* 1, 2 and 3 *converges almost surely to the Pareto set of RMOP* (4) *if the continuous joint density function* $\varphi(z,x)$ *is greater than zero for all* $x \in S_x$, $z \in S_x - x$.

Proof. Let $\{\mathcal{A}^{(t)}, t = 1,2,\ldots\}$ be the archive sequence of the MOEA consisting of Algorithms 1, 2 and 3. Proposition 1 shows that $\forall \varepsilon > 0, t > 0$, there exists a real number $0 < M(\varepsilon) \leqslant 1$ such that

$$\mathbf{P}\{D(\mathcal{A}^{(t+1)}) < \varepsilon | D(\mathcal{A}^{(t)}) > \varepsilon\} \geqslant M(\varepsilon).$$

Because the convergence metric value $D(\mathcal{A}^{(t)})$ does not increase in the evolving process, we have

$$\begin{aligned}
\mathbf{P}\left\{D(\mathcal{A}^{(t+1)}) \geqslant \frac{1}{l}\right\} &= \mathbf{P}\left\{D(\mathcal{A}^{(t)}) \geqslant \frac{1}{l}\right\} \cdot \mathbf{P}\left\{D(\mathcal{A}^{(t+1)}) \geqslant \frac{1}{l} | D(\mathcal{A}^{(t)}) \geqslant \frac{1}{l}\right\} \\
&= \mathbf{P}\left\{D(A^{(t-1)}) \geqslant \frac{1}{l}\right\} \cdot \mathbf{P}\left\{D(\mathcal{A}^{(t)}) \geqslant \frac{1}{l} | D(\mathcal{A}^{(t-1)}) \geqslant \frac{1}{l}\right\} \cdot \mathbf{P}\left\{D(\mathcal{A}^{(t+1)}) \geqslant \frac{1}{l} | D(\mathcal{A}^{(t)}) \geqslant \frac{1}{l}\right\} = \cdots \\
&\leqslant \mathbf{P}\left\{D(\mathcal{A}^{(1)}) \geqslant \frac{1}{l}\right\} \cdot \left(1 - M\left(\frac{1}{l}\right)\right)^t
\end{aligned}$$

for every positive integer l. Thus

$$\sum_{t=1}^{+\infty} \mathbf{P}\left\{D(\mathcal{A}^{(t+1)}) \geqslant \frac{1}{l}\right\} = \mathbf{P}\left\{D(\mathcal{A}^{(1)}) \geqslant \frac{1}{l}\right\} \cdot \sum_{t=1}^{+\infty} \left(1 - M\left(\frac{1}{l}\right)\right)^t < +\infty$$

holds for every positive integer l. From Lemma 6, $\mathbf{P}\{D(\mathcal{A}^{(t+1)})\nrightarrow 0\} = 0$, which is equivalent to $\mathbf{P}\{D(\mathcal{A}^{(t+1)}) \to 0\} = 1$, and so, $\mathbf{P}\{\lim_{t\to\infty} D(\mathcal{A}^{(t+1)}) = 0\} = 1$. Thus, the MOEA consisting of Algorithms 1, 2 and 3 converges almost surely to the Pareto set of RMOP (4). □

Theorem 1 shows the convergence of the MOEA consisting of Algorithms 1, 2 and 3 without respect to diversity. However, the diversity of the final results can be well maintained and even converges almost surely to the B-Pareto set of the RMOP (4).

Theorem 2. *If the continuous joint density function $\varphi(z,t)$ in (8) is greater than zero for all $x \in S_x$, $z \in S_x - x$, the MOEA consisting of Algorithms 1, 2 and 3 converges almost surely to the B-Pareto set of RMOP (4) of size N, where N is the size of the archive equal to that of the population.*

Proof. Let $\{\mathcal{A}^{(t)}, t = 1, 2, \ldots\}$ be the archive sequence of the MOEA consisting of Algorithms 1, 2 and 3. We distinguish three different update strategies.

S1 If the new intermediate individual x' generated by Algorithm 2 dominates $x \in \mathcal{A}^{(t)}$, it will replace x in $\mathcal{A}^{(t+1)}$. However, this update may lead to a decrease in $DIV(\mathcal{A}^{(t)})$ because $dist(x', \mathcal{A}^{(t)} \setminus \{x\})$ may be less than $dist(x, \mathcal{A}^{(t)} \setminus \{x\})$.

S2 When x' is generated by Algorithm 2 such that

$$d(x') < d(x)$$

and

$$\min_{y\in F(\mathcal{A}\setminus\{x\})} \|F(x') - y\|_2 > \min_{y\in F(\mathcal{A}\setminus\{x\})} \|F(x) - y\|_2,$$

$x \in \mathcal{A}^{(t)}$ will also be updated by x'. Because this substitution does not reduce the minimum distance between two individuals in the archive,

$$DIV(\mathcal{A}^{(t+1)}) \geqslant DIV(\mathcal{A}^{(t)})$$

always holds after this update.

S3 The third replacement strategy updates the entire population with a given probability p. If the intermediate archive $\mathcal{A}'^{(t)}$ generated by performing mutations on all individuals in $\mathcal{A}^{(t)}$ satisfies

$$D(\mathcal{A}'^{(t)}) < D(\mathcal{A}^{(t)})$$

and

$$DIV(\mathcal{A}'^{(t)}) > DIV(\mathcal{A}^{(t)}),$$

then $\mathcal{A}^{(t+1)}$ is set to $\mathcal{A}'^{(t)}$. Similar to the second strategy, the value of $DIV(\mathcal{A}^{(t)})$ does not decrease after this update.

Now archive $\mathcal{A}^{(t)}$ can be represented by the sum of two different parts, i.e.,

$$\mathcal{A}^{(t)} = \mathcal{X}^{(t)} + \Delta\mathcal{X}^{(t)},$$

where $\mathcal{X}^{(t)}$ denotes the non-increasing part of the archive sequence, and $\Delta\mathcal{X}^{(t)}$ is the disturbance generated by **S1**. Then, if the $\mathcal{A}^{(t)}$ is updated by **S2** and **S3**, we have

$$\mathcal{A}^{(t)} = \mathcal{X}^{(t)} \quad \text{and} \quad \Delta\mathcal{X}^{(t)} = \mathbf{0},$$

where $\mathbf{0}$ is the set of N zero vectors in $\mathbb{R}^n$. If $\mathcal{A}^{(t)}$ is updated by **S1**, then $\mathcal{X}^{(t)} = \mathcal{A}^{(t-1)}$ and $\Delta\mathcal{X}^{(t)} \neq \mathbf{0}$. When the population does not change at generation t, both $\mathcal{X}^{(t)} = \mathcal{A}^{(t-1)}$ and $\Delta\mathcal{X}^{(t)} = \mathbf{0}$ hold. Now let us analyze the respective properties of $\mathcal{X}^{(t)}$ and $\Delta\mathcal{X}^{(t)}$.

1. $\{DIV(\mathcal{X}^{(t)}); t = 1, 2, \ldots\}$ is a bounded and monotonically increasing sequence that converges to a finite value, denoted $DIV(\mathcal{X}^{(\infty)})$. Suppose that

$$DIV(\mathcal{X}^{(\infty)}) < DIV(P_B^*).$$

Then from the continuity of $\|\cdot\|_2$, there exists a neighborhood of the B-Pareto set P_B^*, denoted

$$U(P_B^*, \varepsilon) \triangleq \prod_{x^*\in P_B^*} U(x^*, \varepsilon) \cap S_x = \prod_{x^*\in P_B^*} \{x \in S_x; \|x - x^*\|_2 < \varepsilon\},$$

such that $\forall \mathcal{X} \in U(P_B^*, \varepsilon)$,

$$D(\mathcal{X}'^{(t)}) > D(\mathcal{X}) > D(P_B^*) = 0, \quad \forall 1 < t < \infty$$

and

$$DIV(\mathcal{X}'^{(t)}) < DIV(\mathcal{X}) < DIV(P_B^*), \quad \forall 1 < t < \infty,$$

where $\prod_{x^* \in P_B^*} U(x^*, \varepsilon)$ is the Cartesian product of $U(x^*, \varepsilon)$. Moreover, Lemma 5 implies that with a positive probability $\mathbf{P}\{x' \in U(x^*, \varepsilon)\}$, the intermediate individual x' generated by mutation (8) is located in the neighborhood $U(x^*, \varepsilon)$, and so $\mathcal{X}^{(t+1)}$ is in $U(P_B^*, \varepsilon)$ with a positive probability $\prod_{x^* \in P_B^*} \mathbf{P}\{x' \in U(x^*, \varepsilon)\}$. Similar to the proof of Theorem 1, we conclude that

$$DIV(P_B^*) - DIV(\mathcal{X}'^{(t)}) \xrightarrow{a.s.} 0,$$

i.e.,

$$DIV(\mathcal{X}'^{(t)}) \xrightarrow{a.s.} DIV(P_B^*).$$

2. If the archive is updated by **S1**, then

$$\Delta\mathcal{X}^{(t)} = \{0, \ldots, x'^{(t)} - x^{(t)}, \ldots, 0\} \neq \mathbf{0},$$

where $x^{(t)} \in \mathcal{A}^{(t)}$, and $x'^{(t)}$ is the intermediate individual generated satisfying $x'^{(t)} \prec x^{(t)}$. Moreover, Theorem 1 shows that

$$D(\mathcal{A}^{(t)}) \xrightarrow{a.s.} 0$$

and so,

$$x'^{(t)} - x^{(t)} \xrightarrow{a.s.} 0$$

also holds.
In conclusion,

$$DIV(\mathcal{A}^{(t)}) = DIV(\mathcal{X}'^{(t)} + \Delta\mathcal{X}^{(t)})$$

converges almost surely to $DIV(P_B^*)$. Combined with the result proven in Theorem 1 that $D(\mathcal{A}^{(t)}) \xrightarrow{a.s.} 0$, we prove that the MOEA consisting of Algorithms 1, 2 and 3 converges almost surely to the B-Pareto set of RMOP (4) of size N. $\square$

3.4. Comparisons with other fixed population size methods

There are also other efficient EMO methods that work with a fixed population (or archive) size. To demonstrate the efficiency of the proposed method, we compare it with two excellent representatives of these methods, namely, the hypervolume-based approaches [55,16,2] and SPEA2 [54]. The test problem is the ZDT1 problem [53], and population size and archive size are both set to three. The Pareto front of the ZDT1 problem is $\{y = (y_1, y_2); y_2 = 1 - \sqrt{y_1}, y_1 \in [0, 1]\}$.

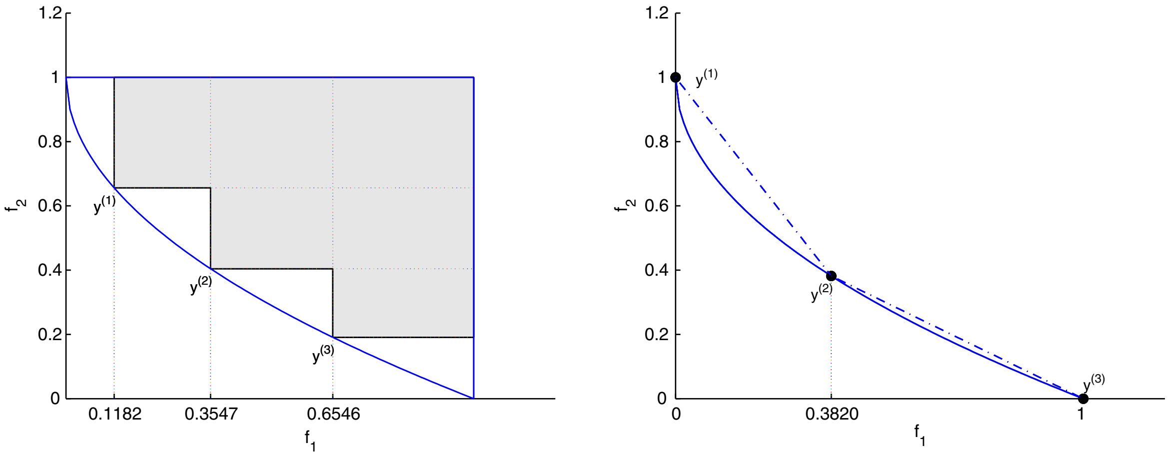
3.4.1. Comparison with hypervolume-based approaches

Hypervolume-based approaches [55,16,2] aim to obtain a population (or archive) with the greatest dominated hypervolume. Let the reference point be (1,1), and denote the three points on the Pareto front as $y^{(1)}$, $y^{(2)}$ and $y^{(3)}$, where $y^{(i)} = (y_1^{(i)}, y_2^{(i)}), i = 1, 2, 3$. Then $(y^{(1)}, y^{(2)}, y^{(3)})$ is the optimal solution of the maximum problem

$$\begin{cases} \max f(y^{(1)}, y^{(2)}, y^{(3)}) = (1 - y_1^{(1)})(1 - y_2^{(1)}) + (y_2^{(1)} - y_2^{(3)})(1 - y_1^{(3)}) + (y_2^{(1)} - y_2^{(2)})(y_1^{(3)} - y_1^{(2)}) \\ s.t. \;\; y_2^{(i)} = 1 - \sqrt{y_1^{(i)}}; \\ \qquad y_1^{(i)} \in [0, 1], \quad i = 1, 2, 3. \end{cases}$$

From the differentiability of $f(y^{(1)}, y^{(2)}, y^{(3)})$, the optimal solution is located on one of the stationary points. We obtain approximate values of $y^{(1)}$, $y^{(2)}$ and $y^{(3)}$, which are (0.1182,0.6562), (0.3547,0.4044) and (0.6546,0.1909), respectively. The method proposed in this paper forces the convergence of the archive to the B-Pareto front of size three, which contains three points $y^{(1)} = (0,1)$, $y^{(2)} \approx (0.3820, 0.3819)$ and $y^{(3)} = (1,0)$.

Fig. 6 illustrates the results of the two compared methods. Fig. 6(a) shows that the distance between $y^{(1)}$ and $y^{(2)}$ are not equal to the distance between $y^{(2)}$ and $y^{(3)}$, and this is because the curvature of the Pareto front of ZDT1 differs at distinct positions. Thus, we conclude that the results of the hypervolume-based methods are sensitive to the shape of the Pareto front. Moreover, the reference point also influences the results of hypervolume-based methods. With the reference point (1,1), the results of the hypervolume-based methods cannot reach the boundary points of the Pareto front. Although the left boundary point of ZDT1 can be included if the reference point is selected appropriately, the right extreme point of the Pareto



(a) The best results obtained by the hypervolume-based methods.

(b) The best results obtained by the method proposed in this paper.

Fig. 6. Comparison between the hypervolume-based methods and the method proposed in this paper. When the archive size and the population size are both three, the best representation of the Pareto front obtained by the hypervolume-based methods is illustrated by (a), and the best representation of the Pareto front obtained by the method proposed in this paper is illustrated by (b). With a given reference point (1, 1), results of the hypervolume-based methods cannot obtain the boundary point of the Pareto front, whereas the results obtained in this paper is a *B*-Pareto front of the ZDT1 problem, in which all points are distributed uniformly on the Pareto front. (a) The best results obtained by the hypervolume-based methods. (b) The best results obtained by the method proposed in this paper.

front cannot be obtained regardless of the reference point [1]. Even if the extreme point could be obtained by setting the reference point in a certain manner, the reference point is hard to obtain when the properties of the investigated MOP are not clear. However, Fig. 6(b) shows that the method proposed in this paper obtains the *B*-Pareto front of the MOP, which always contains the boundary points of the Pareto front and is uniformly distributed on the Pareto front.

3.4.2. Comparison with SPEA2

SPEA2 [54], also a high performance MOEA with a fixed archive size, is able to obtain an approximation of the Pareto front with both high precision and good diversity. SPEA2 compares individuals according to their fitness values, and the runtime of the fitness assignment procedure is dominated by a density estimator with complexity $\mathcal{O}(M^2 \log M)$, where M is the sum of the population size and archive size [54]. When the population and the archive are of the same size N, the complexity of the fitness assignment procedure is $\mathcal{O}(N^2 \log N)$.

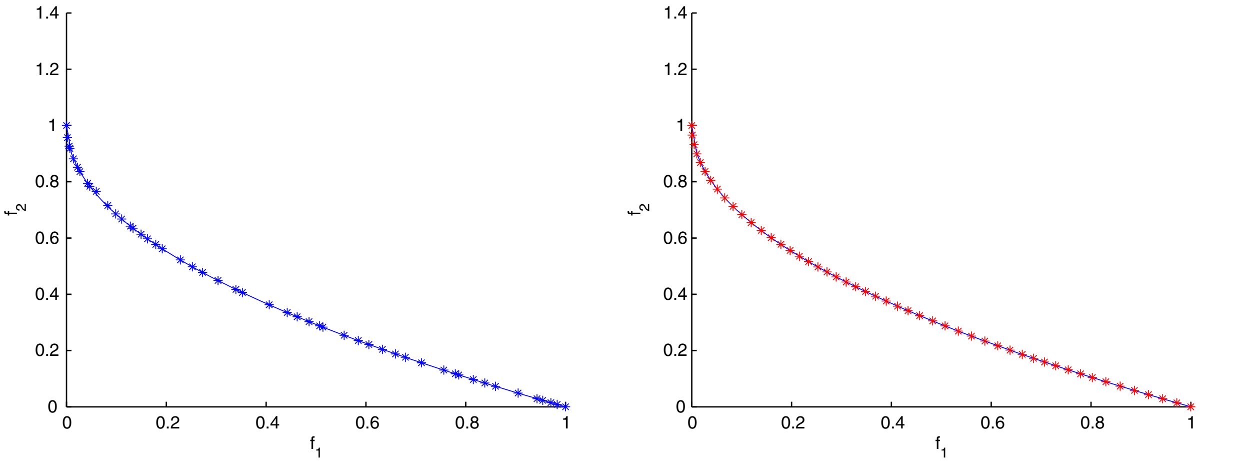
The method proposed in this paper contains both a convergence metric computation procedure and a diversity computation procedure. At each iteration, the convergence metric computation is of the order $\mathcal{O}(N)$. Because the determination of relation '$\lhd$' includes computations of the distance between the candidate and all individuals in the archive, the complexity of the diversity computation is $\mathcal{O}(N^2)$. Thus, the total complexity of the updating procedure is $\mathcal{O}(N^2)$, which is less than the complexity of SPEA2.

4. Numerical results

To demonstrate the appropriateness of the new definition of *B*-Pareto set and the effect of the new archive-updating strategy, numerical experiments are presented in this paper. The comparison is based on a well-known MOEA, namely, NSGAII [13], which has been successfully applied in specific scientific and engineering fields. Using the new updating strategy proposed in this paper, the intermediate population of newly generated individuals is introduced to update the extra archive attached to the NSGAII procedure at each generation. Because the elements in the archive are not included in the process of recombination, the same candidate individuals were generated for different updating strategies, which are the selection strategy of NSGAII and the new proposed archive-updating strategy. Other influencing factors are excluded in the comparison. The numerical results have been presented to confirm the theoretical results by performing the experiments on the test problems ZDT1, ZDT2 [53] and DTLZ2 [14].

4.1. Two-dimensional cases

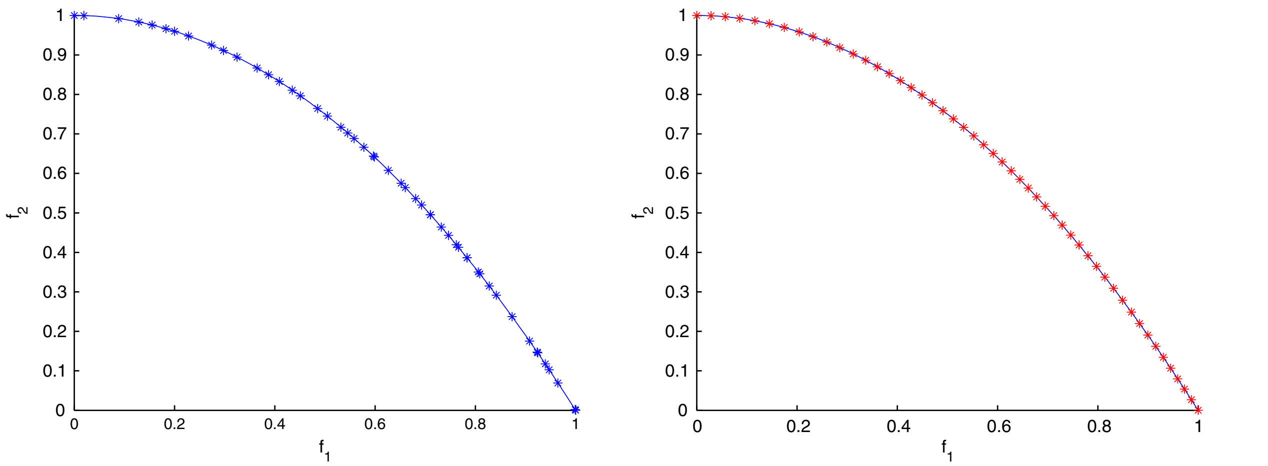
At first, the comparison was performed using the 30-dimensional ZDT1 and ZDT2 problems. To obtain a sparsely distributed finite representation of the Pareto front to observe its uniformity, the population size and the archive size were set to



(a) The results of NSGAII.

(b) The results of the new updating strategy proposed in this paper.

Fig. 7. Illustrations of the final results of problem ZDT1. (a) The final population of NSGAII, and (b) the final archive obtained by the proposed updating strategy.



(a) The results of NSGAII.

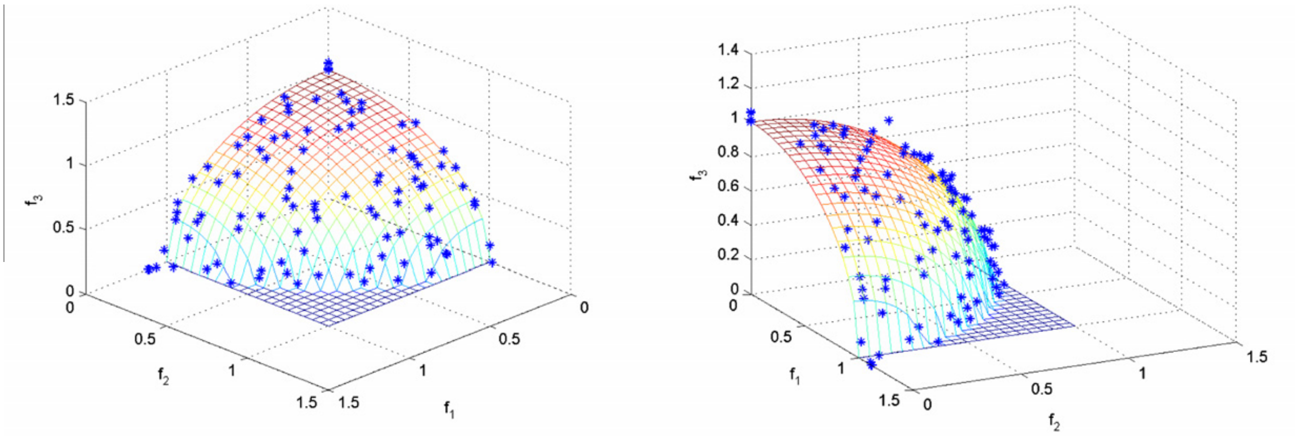
(b) The results of new updating strategy proposed in this paper.

Fig. 8. Illustrations of the final results of problem ZDT2. (a) The final population of NSGAII, and (b) the final archive obtained by the proposed updating strategy.

50, and the two parameters in NSGAII, namely, the distribution indices η_c and η_m, were set to 15 and 20, respectively. After 300 iterations of the algorithms, the final results are illustrated in Figs. 7 and 8.

Because the diversity of the population can be well-preserved in the procedure of NSGAII, the SBX crossover and Polynomial mutation [13] operators can be looked upon as a mutation operator that can be described in terms of (8) such that the joint density function $\varphi(z,x)$ is continuous and greater than zero. ZDT1 and ZDT2 have bounded and simply connected Pareto fronts; thus, the conditions of Theorem 2 are satisfied.

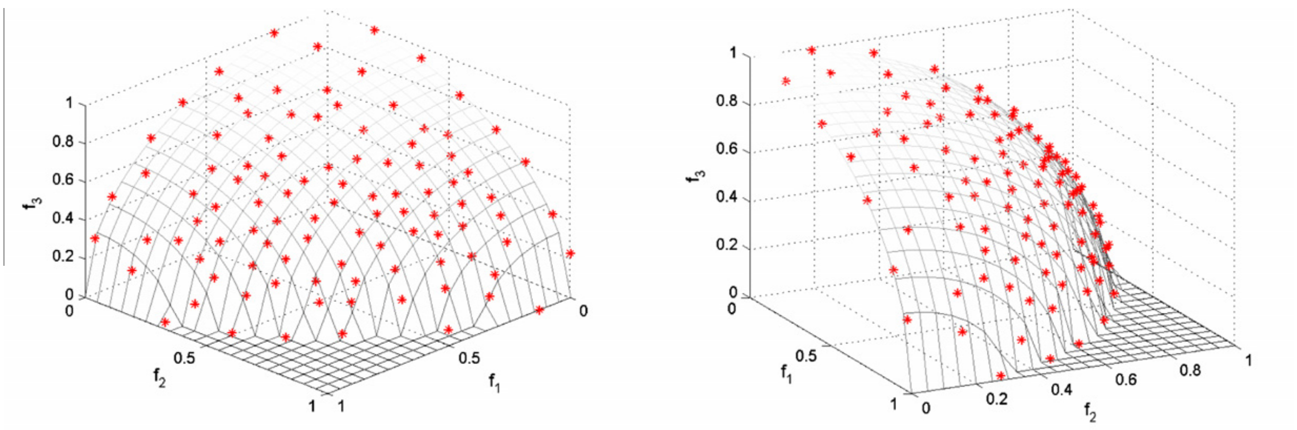
In this case, the vectors in the *B*-Pareto front were distributed uniformly along the Pareto front, and the numerical results verified the theoretical results obtained in this paper, i.e., the archive sequence of the MOEA converges almost surely to a *B*-Pareto set of RMOP (4) with size 50. Numerical results also show that the gaps did not appear in regions where the Pareto front is "flat". The Pareto fronts of ZDT1 and ZDT2 are convex and concave, respectively, and the curvatures of the Pareto fronts of ZDT1 and ZDT2 were not equal at different parts, which can influence the distribution of final results for some existing updating strategies, such as the strategy proposed in [43], according to which some gaps may occur in regions where the Pareto front is "flat" due to the nature of ϵ-dominance [44]. However, the new archive-updating strategy proposed in this paper ignores the differences in different parts of the Pareto front, and the *B*-Pareto front, which is a uniformly distributed representation of the Pareto front of RMOP (4), was obtained in the end.



(a) The observation of the 3-D plot from the view point (1,1,1).

(b) The observation of the 3-D plot from the view point (1,-0.5,0.5).

Fig. 9. The final results of NSGAII for problem DTLZ2.



(a) The observation of the 3-D plot from the view point (1,1,1).

(b) The observation of the 3-D plot from the view point (1,-0.5,0.5)

Fig. 10. The final results of the new proposed updating strategy for problem DTLZ2.

4.2. A three-dimensional case

The theoretical results can also be extended to the m-dimensional case ($m \geqslant 3$). A three-dimensional case, the 3-D DTLZ2 problem, was investigated to evaluate the efficiency of the new updating strategy. In this case, the population size and archive size were both set to 100, and the values of η_c and η_m were set to 15 and 20, respectively. Numerical results obtained from 300 iterations of the algorithm were illustrated in Figs. 9 and 10.

Obviously, the results obtained using the proposed archive-updating strategy were closer to the Pareto front, and were distributed more uniformly along the Pareto front than the final results of NSGAII. Accordingly, we can draw the conclusion that the new updating strategy greatly improves the final result. In contrast, the final results of NSGAII were not satisfactory, and the *B*-Pareto front was not generated. This is because all vectors generated by NSGAII were not distributed regularly in the objective space, and consequently, after a finite number of iterations, the selected individuals could not be scattered uniformly along the Pareto front. Therefore, new strategies for the candidate generation should be designed to obtain a good approximation of the *B*-Pareto front after more iterations of the MOEAs.

5. Discussions

In this study, we focused on the theoretical research, which is difficult in multi-objective evolutionary algorithms. We proposed a new definition for the finite representation of the Pareto front, namely, the *B*-Pareto front (and correspondingly, the *B*-Pareto set in the feasible space) and a convergence metric of MOEAs by combining the convergence with diversity. In contrast to the concept of ϵ-Pareto set, the *B*-Pareto set can always maintain a fixed size *N*, and gaps do not appear in regions where the Pareto front is "flat", which makes it more appropriate for use in the theoretical analysis of an MOEA.

There are also other efficient EMO methods that work with a fixed population size, e.g., SPEA2 [54] and the hypervolume-based approaches [55,16,2]. The goal of our approach and these methods are the same: to consider both convergence and diversity. To the best of our knowledge, however, no report has theoretically proved the convergence of the hypervolume-based methods and SPEA2. Moreover, comparisons of them through a simple example indicate that the hypervolume-based methods sometimes cannot reach the boundary points of the Pareto front, which can be reached by our proposed method. In addition, the time complexity of SPEA2 is higher than that of our proposed method.

In this study, the non-negative function $d(\cdot)$ has only been defined for the case that the Pareto front is simply connected, which should be extended in further works. Moreover, runtime analysis of real-coded MOEAs is also one of our future goals because other than runtime analysis of discrete multi-objective optimization problems, a rigorous analysis of runtime of real-coded MOEAs has not been explored [20,33,7,36,19,23,24]. Numerical results show that our new updating strategy can improve the efficiency of NSGAII, and thus, designing an efficient MOEA based on the theoretical results presented here would be an important future work. Although the results presented in this study can be improved, they can serve as a significant foundation for new algorithm design and for theoretical analysis.

6. Conclusions

In this paper, we present a new concept of the *B*-Pareto front by combining convergence with diversity, and theoretically show how to obtain the *B*-Pareto front of an RMOP. The convergence of the MOEA was guaranteed by the compound-updating strategy, which is based on a new fitness function of individuals and a method for preserving diversity. Numerical results confirm the appropriateness of theoretical results. Moreover, comparisons with NSGAII, the hypervolume-based approaches and SPEA2 show that the new archive-updating strategy is quite efficient. In conclusion, the theoretical analysis, numerical experiments and comparisons with other methods show that our method is competitive.

Acknowledgements

The authors thank the anonymous reviewers for their helpful comments and suggestions, and also thank Prof. Yuren Zhou at the South China University of Technology, China for his useful discussions. This work was supported by the Chinese National Natural Science Foundation under Grant No. 61070007 and the Key Project of National Natural Science Foundation of China under Grant No. 51039005.

Appendix A. Proof of Lemma 2

Proof. Denote

$$y^{(1)} = F(x^{(1)}) = (f_1(x^{(1)}), \ldots, f_m(x^{(1)})) = (y_1^{(1)}, \ldots, y_m^{(1)})$$

and

$$y^{(2)} = F(x^{(2)}) = (f_1(x^{(2)}), \ldots, f_m(x^{(2)})) = (y_1^{(2)}, \ldots, y_m^{(2)}).$$

If $y^{(1)} \prec y^{(2)}$, then $\exists j_1, \ldots, j_l \in \{1, 2, \ldots, m\}$, where

$$y_{j_t}^{(1)} < y_{j_t}^{(2)}, \quad t = 1, \ldots, l$$

and it holds that $y_k^{(1)} = y_k^{(2)}$ when $\forall k \neq j_t$, $t = 1, \ldots, l$.

By Lemma 1, $\forall i \in \{1, 2, \ldots, m\}$, there exists an implicit function

$$y_i = h_i(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)$$

specified by the Pareto front *PF*. Defining H_i according to (6), a class of hypersurfaces

$$G_{H_i}(C_i) = \{y = (y_1, \ldots, y_m) \in S_y; y_i = H_i(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m) + C_i\}$$

can be derived. Divide the objective space S_y of RMOP (4) into two parts:

1. $\mathcal{S}_i$, all points in which are located on $G_{h_i}(C_i)$ for some given non-negative number C_i;
2. $\mathcal{T}_i$, all points in which cannot be located on $G_{h_i}(C_i)$ for any non-negative number C_i,where G_{h_i} is defined by (5). Then, two different cases can be distinguished.

1. If there exists one and only one $j \in \{1, 2, \ldots, m\}$ such that $y_j^{(1)} < y_j^{(2)}$, then for all $i \in \{1, 2, \ldots, m\}$, one of the followings is true.

(a) $y^{(1)} \in \mathcal{S}_i, y^{(2)} \in \mathcal{S}_i$.

i. If $i = j$, we have

$$H_i\left(y_1^{(1)},\ldots,y_{i-1}^{(1)},y_{i+1}^{(1)},\ldots,y_m^{(1)}\right) = H_i\left(y_1^{(2)},\ldots,y_{i-1}^{(2)},y_{i+1}^{(2)},\ldots,y_m^{(2)}\right),$$

because $y_k^{(1)} = y_k^{(2)}$ always holds when $k \neq j$. So, from

$$d_i\left(y_1^{(l)},\ldots,y_m^{(l)}\right) = y_i^{(l)} - H_i\left(y_1^{(l)},\ldots,y_{i-1}^{(l)},y_{i+1}^{(l)},\ldots,y_m^{(l)}\right), \quad l = 1,2,$$

then

$$d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) < d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right) \tag{A.1}$$

when $i = j$.

ii. When $i \neq j$, hypothesize that

$$d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) > d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right). \tag{A.2}$$

From

$$y_i^{(l)} - d_i\left(y_1^{(l)},\ldots,y_m^{(l)}\right) = H_i\left(y_1^{(l)},\ldots,y_{i-1}^{(l)},y_{i+1}^{(l)},\ldots,y_m^{(l)}\right) = h_i\left(y_1^{(l)},\ldots,y_{i-1}^{(l)},y_{i+1}^{(l)},\ldots,y_m^{(l)}\right), \quad l = 1,2, \tag{A.3}$$

it follows that

$$\left(y_1^{(l)},\ldots,y_{i-1}^{(l)},y_i^{(l)} - d_i(y_1^{(l)},\ldots,y_m^{(l)}),y_{i+1}^{(l)},\ldots,y_m^{(l)}\right) \in PF, \quad l = 1,2. \tag{A.4}$$

Note that (A.2) implies $y_i^{(1)} - d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) < y_i^{(2)} - d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right)$, which leads to the fact that

$$\left(y_1^{(1)},\ldots,y_{i-1}^{(1)},y_i^{(1)} - d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right),y_{i+1}^{(1)},\ldots,y_m^{(1)}\right) \prec \left(y_1^{(2)},\ldots,y_{i-1}^{(2)},y_i^{(2)} - d_i(y_1^{(2)},\ldots,y_m^{(2)}),y_{i+1}^{(2)},\ldots,y_m^{(2)}\right).$$

This contradicts (A.4). Thus, when $i \neq j$,

$$d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) \leqslant d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right). \tag{A.5}$$

By (A.1) and (A.5),

$$d_i(y_1^{(1)},\ldots,y_m^{(1)}) \leqslant d_i(y_1^{(2)},\ldots,y_m^{(2)})$$

always holds when $y^{(1)} \in \mathcal{S}_i, y^{(2)} \in \mathcal{S}_i$.

(b) $y^{(1)} \in \mathcal{S}_i, y^{(2)} \in \mathcal{T}_i$.If $i = j$, then $y_i^{(1)} < y_i^{(2)}$. Then both $y^{(1)}$ and $y^{(2)}$ are located in $\mathcal{S}_i$. In this case, $i \neq j$ by necessity. From (6), it follows that

$$d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) < d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right);$$

(c) $y^{(1)} \in \mathcal{T}_i, y^{(2)} \in \mathcal{T}_i$. Because $y^{(1)} \prec y^{(2)}$, then $d_i(y_1^{(1)},\ldots,y_m^{(1)}) \leqslant d_i(y_1^{(2)},\ldots,y_m^{(2)})$, and $d_i(y_1^{(1)},\ldots,y_m^{(1)}) < d_i(y_1^{(2)},\ldots,y_m^{(2)})$ holds if and only if $i = j$.Based on (a), (b) and (c), we have

$$d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) \leqslant d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right),$$

where the strict inequality

$$d_i\left(y_1^{(1)},\ldots,y_m^{(1)}\right) < d_i\left(y_1^{(2)},\ldots,y_m^{(2)}\right)$$

holds if $i = j$. Thus,

$$y^{(1)} \prec y^{(2)} \;\Rightarrow\; d(x^{(1)}) < d(x^{(2)})$$

when there exists one and only one $j \in \{1,2,\ldots,m\}$ such that $y_j^{(1)} < y_j^{(2)}$.

2. When there exist l indices $j_1, \ldots, j_l \in \{1,2,\ldots,m\}$ such that

$$y_{j_t}^{(1)} < y_{j_t}^{(2)}, \quad t = 1,2\ldots,l,$$

without loss of generality, it is assumed that $1 \leqslant j_1 < j_2 < \cdots < j_{l-1} < j_l \leqslant m$. Denoting that

$$y^{(j_1)} = y^{(1)} + \left(0,\ldots,0,y_{j_1}^{(2)} - y_{j_1}^{(1)},0,\ldots,0\right),$$
$$\ldots$$
$$y^{(j_{l-1})} = y^{(j_{l-2})} + \left(0,\ldots,0,y_{j_{l-1}}^{(2)} - y_{j_{l-1}}^{(1)},0,\ldots,0\right),$$
$$y^{(2)} = y^{(j_{l-1})} + \left(0,\ldots,0,y_{j_l}^{(2)} - y_{j_l}^{(1)},0,\ldots,0\right),$$

then we have

$$d(x^{(1)}) = \sum_{i=1}^{m} d_i(y^{(1)}) < \sum_{i=1}^{m} d_i(y^{(j_1)}) < \cdots < \sum_{i=1}^{m} d_i(y^{(j_{k-1})}) < \sum_{i=1}^{m} d_i(y^{(2)}) = d(x^{(2)}).$$

In conclusion, if $y^{(1)} \prec y^{(2)}$, then $d(x^{(1)}) < d(x^{(2)})$. □

Appendix B. Proof of Lemma 4

Proof. $\forall x \in S_x$, suppose that $x + z$ and $x + \Delta x + z$ are both in $U(x^*,\delta) \cap S_x$. Then

$$L_\delta(x^*, x + \Delta x) = \int_{(x+\Delta x)+z \in U(x^*,\delta)\cap S_x} \phi(z,x)dz.$$

Thus

$$\begin{aligned}
\|L_\delta(x^*, x+\Delta x) - L_\delta(x^*, x)\| &= \left\| \int_{(x+\Delta x)+z\in U(x^*,\delta)\cap S_x} \phi(z,x)dz - \int_{x+z\in U(x^*,\delta)\cap S_x} \phi(z,x)dz \right\| \\
&= \left\| \int_{z\in U(x^*,\delta)\cap S_x-(x+\Delta x)} \phi(z,x)dz - \int_{z\in U(x^*,\delta)\cap S_x - x} \phi(z,x)dz \right\| \\
&= \left\| \int_{z\in U(x^*,\delta)\cap S_x-(x+\Delta x)\setminus(U(x^*,\delta)\cap S_x-x)} \phi(z,x)dz - \int_{z\in U(x^*,\delta)\cap S_x-x\setminus(U(x^*,\delta)\cap S_x-(x+\Delta x))} \phi(z,x)dz \right\| \\
&\leqslant \left\| \int_{z\in U(x^*,\delta)\cap S_x-(x+\Delta x)\setminus U(x^*,\delta)\cap S_x-x} \phi(z,x)dz \right\| + \left\| \int_{z\in U(x^*,\delta)\cap S_x-x\setminus U(x^*,\delta)\cap S_x-(x+\Delta x)} \phi(z,x)dz \right\|.
\end{aligned}$$

Because the Lebesgue measures of $U(x^*,\delta) \cap S_x - (x + \Delta x)\setminus(U(x^*,\delta) \cap S_x - x)$ and $U(x^*,\delta) \cap S_x - x\setminus(U(x^*,\delta) \cap S_x - (x + \Delta x))$ converge to zero when $\|\Delta x\|$ tends to 0, we have

$$\|L_\delta(x^*, x+\Delta x) - L_\delta(x^*, x)\| \to 0,$$

when $\|\Delta x\|$ tends to 0. Thus, $L_\delta(x^*,x)$ is continuous for all $x \in S_x$. Similarly, we can prove that $L_\delta(x^*,x)$ is continuous about x^* as well. □

References

[1] A. Auger, J. Bader, D. Brockhoff, E. Zitzler, Theory of the hypervolume indicator: optimal μ-distributions and the choice of the reference point, in: Proceedings of the International Workshop on Foundations of Genetic Algorithms (FOGA 2009), ACM, New York, 2009, pp. 87–102.
[2] N. Beume, B. Naujoks, M. Emmerich, SMS-MOEA: multiobjective selection based on dominated hypervolume, European Journal of Operational Research 181 (2007) 1653–1669.
[3] H.G. Beyer, The Theory of Evolution Strategies, Springer, 2001.
[4] H.G. Beyer, H.P. Schwefel, I. Wegener, How to analyse evolutionary algorithms, Theoretical Computer Science 287 (2002) 101–130.
[5] A. Bienvenüe, O. Francois, Global convergence for evolution strategies in spherical problems: some simple proofs and difficulties, Theoretical Computer Science 306 (2003) 269–289.
[6] K. Bringmann, T. Friedrich, An efficient algorithm for computing hypervolume contributions, Evolutionary Computation 18 (2010) 383–402.
[7] D. Brockhoff, T. Friedrich, N. Hebbinghaus, Do additional objectives make a problem harder?, in: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO '07), ACM, New York, 2007, pp 765–772.
[8] D. Brockhoff, T. Friedrich, F. Neumann, Analyzing hypervolume indicator based algorithms, in: Proceedings of 10th International Conference on Parallel Problem Solving from Nature (PPSN X), LNCS, vol. 5199, Springer, Berlin, 2008, pp. 651–660.
[9] C.A.C. Coello, G.B. Lamont, Applications of Multi-Objective Evolutionary Algorithms, World Scientific, 2004.
[10] H. Dawid, A markov chain analysis of genetic algorithms with a state dependent fitness function, Complex System 8 (1994) 407–417.
[11] K. Deb, R. Agrawal, Simulated binary crossover for continuous search space, Complex System 9 (1995) 115–148.
[12] K. Deb, Multi-objective Optimization Using Evolutionary Algorithms, Wiley, Chichester, UK, 2001.
[13] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2002) 182–197.

[14] K. Deb, L. Thiele, M. Laumanns, E. Zitzler, Scalable multi-objective optimization test problems, in: Proceedings of IEEE Congress on Evolutionary Computation (CEC'02), IEEE Press, Piscataway, New Jersey, 2002, pp. 825–830.
[15] S. Droste, T. Jansen, I. Wegener, On the analysis of the (1 + 1) evolutionary algorithm, Theoretical Computer Science 276 (2002) 51–82.
[16] M. Emmerich, N. Beume, B. Naujoks, An EMO algorithm using the hypervolume measure as selection criterion, in: Proceedings of the 3rd International Conference on Evolutionary Multi-Criterion Optimization (EMO'2005), Springer, Berlin, 2005, pp. 62–76.
[17] M. Farina, P. Amato, A fuzzy definition of optimality'' for many-criteria optimization problems, IEEE Transactions on Systems, Man, and Cybernetics, Part A 34 (2004) 315–326.
[18] T. Friedrich, C. Horoba, F. Neumann, Multiplicative approximations and the hypervolume indicator, in: Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation (GECCO 2009), ACM, Montral, Canada, 2009, pp. 571–578.
[19] T. Friedrich, N. Hebbinghaus, F. Neumann, Plateaus can be harder in multi-objective optimization, Theoretical Computer Science 411 (2010) 854–864.
[20] O. Giel, Expected runtimes of a simple multi-objective evolutionary algorithm, in: Proceedings of IEEE Congress on Evolutionary Computation(CEC'03), IEEE Press, Piscataway, New Jersey, 2003, pp. 1918–1925.
[21] T. Hanne, On the convergence of multiobjective evolutionary algorithms, European Journal of Operational Research 117 (1999) 553–564.
[22] T. Hanne, A multiobjective evolutionary algorithm for approximating the efficient set, European Journal of Operational Research 176 (2007) 1723–1734.
[23] C. Horoba, F. Neumann, Benefits and drawbacks for the use of ϵ-dominance in evolutionary multi-objective optimization, in: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation, GECCO'08, ACM, New York, 2008, pp. 641–648.
[24] C. Horoba, F. Neumann, Additive approximations of Pareto-optimal sets by evolutionary multi-objective algorithms, in: Proceedings of the 10th ACM SIGEVO workshop on Foundations of genetic algorithms, ACM, New York, 2009, pp. 79–86.
[25] E.J. Hughes, Evolutionary many-objective optimisation: many once or one many?, in: Proceedings of IEEE Congress on Evolutionary Computation(CEC'05), IEEE Press, Piscataway, New Jersey, 2005, pp 222–227.
[26] J. Jägersküpper, How the (1+1) ES using isotropic mutations minimizes positive definite quadratic forms, Theoretical Computer Science 361 (2006) 38–56.
[27] J. Jägersküpper, Algorithm analysis of a basic evolutionary algorithm for continuous optimization, Theoretical Computer Science 379 (2007) 329–347.
[28] J. He, X. Yao, From an individual to a population: an analysis of first hitting time of population-based evolutionary algorithms, IEEE Transactions on Evolutionary Computation 6 (2002) 495–511.
[29] J. He, X. Yao, A study of drift analysis for estimating computation time of evolutionary algorithms, Natural Computing 3 (2004) 21–35.
[30] J.D. Knowles, D.W. Corne, Approximating the nondominated front using the pareto archived evolution strategy, Evolutionary Computation 8 (2000) 149–172.
[31] J.D. Knowles, D.W. Corne, Properties of an adaptive archiving algorithm for storing nondominated vectors, IEEE Transactions on Evolutionary Computation 7 (2003) 100–116.
[32] M. Laumanns, L. Thiele, K. Deb, E. Zitzler, Combining convergence and diversity in evolutionary multiobjective optimization, Evolutionary Computation 10 (2002) 263–282.
[33] M. Laumanns, L. Thiele, E. Zitzler, Running time analysis of multiobjective evolutionary algorithms on pseudo-boolean functions, IEEE Transactions on Evolutionary Computation 8 (2004) 170–182.
[34] M. Loève, Probability Theory I, 4th ed., Springer-Verlag, New York, USA, 1977.
[35] S.W. Mahfound, Finite markov chain models of an alternative selection strategy for the genetic algorithm, Complex System 7 (1993) 155–170.
[36] F. Neumann, Expected runtimes of a simple evolutionary algorithm for the multi-objective minimum spannning tree problem, European Journal of Operational Research 181 (2007) 1620–1629.
[37] B.Y. Qu, P.N. Suganthan, Multi-objective evolutionary algorithms based on the summation of normalized objectives and diversified selection, Information Sciences 180 (2010) 3170–3181.
[38] D. Reynolds, J. Gomatam, Stochastic modelling of genetic algorithms, Artificial Intelligence 82 (1996) 303–330.
[39] G. Rudolph, Convergence analysis of canonical genetic algorithms, IEEE Transactions on Neural Networks 5 (1994) 96–101.
[40] G. Rudolph, A. Agapie, Convergence properties of some multi-objective evolutionary algorithms, in: Proceedings of on IEEE Conference Evolutionary Computation, IEEE Press, Piscataway, New Jersey, 2000, pp. 1010–1016.
[41] O. Schütze, C.A.C. Coello, S. Mostaghim, E.G. Talbi, M. Dellnitz, Hybridizing evolutionary strategies with continuation methods for solving multi-objective problems, Engineering Optimization 40 (2008) 383–402.
[42] O. Schütze, C.A.C. Coello, E. Tantar, E.G. Talbi, Computing finite size representations of the set of approximate solutions of an MOP with stochastic search algorithms, in: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation (GECCO '08), ACM, New York, 2008, pp. 713–720.
[43] O. Schütze, M. Laumanns, C.A.C. Coello, Convergence of stochastic search algorithms to finite size Pareto set approximations, Journal of Global Optimization 41 (2008) 559–577.
[44] O. Schütze, M. Laumanns, E. Tantar, C.A.C. Coello, E.G. Talbi, Convergence of stochastic search algorithms to gap-free Pareto front approximations, in: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO'07), ACM, New York, 2007, pp. 892–901.
[45] C. Shi, Z. Yan, K. Lü, Z. Shi, B. Wang, A dominance tree and its application in evolutionary multi-objective optimization, Information Sciences 179 (2009) 3540–3560.
[46] H.K. Singh, T. Ray, W. Smith, C-PSA: Constrained Pareto simulated annealing for constrained multi-objective optimization, Information Sciences 180 (2010) 2499–2513.
[47] M. Villalobos-Arias, C.A.C. Coello, O. Hernández-Lerma, Asymptotic convergence of metaheuristics for multiobjective optimization problems, Soft Computing 10 (2006) 1001–1005.
[48] Y. Wang, Y. Yang, Particel swarm optimization with preference order ranking for multi-objective optimization, Information Sciences 179 (2009) 1944–1959.
[49] Q. Zhang, H. Li, MOEA/D: a multiobjective evolutionary algorith based on decomposition, IEEE Transactions on Evolutionary Computation 11 (2007) 712–731.
[50] Q. Zhang, A. Zhou, Y. Jin, RM-MEDA: a regularity model based multiobjective estimation of distribution algorithm, IEEE Transactions on Evolutionary Computation 12 (2008) 41–63.
[51] Y. Zhou, J. He, Convergence analysis of a self-adaptive multi-objective evolutionary algorithm based on grids, Information Processing Letters 104 (2007) 117–122.
[52] E. Zitzler, L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach, IEEE Transactions on Evolutionary Computation 3 (1999) 257–271.
[53] E. Zitzler, K. Deb, L. Thiele, Comparison of multiobjective evolutionary algorithms: empirical results, Evolutionary Computation. 8 (2000) 173–195.
[54] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: improving the strength Pareto evolutionary algorithm. TIK-Report No. 103, Computer Engineering and Networks Laboratory (TIK), Zurich, Switzerland, Swiss Federal Institute of Technology (ETH) Zurich, 2001.
[55] E. Zitzler, S. Künzli, Indicator-based selection in multiobjective search, in: Proceedings of the International Conference on Parallel Problem Solving form Nature (PPSNIII), LNCS, vol. 866, Springer, Berlin, 2004, pp. 832–842.
[56] X. Zou, Y. Chen, M. Liu, L. Kang, A new evolutionary algorithm for solving many-objective optimization problems, IEEE Transactions on System, Man and Cybernetics, Part B 38 (2008) 1402–1412.