12-1-2010

Adaptive, convergent, and diversified archiving strategy for multiobjective evolutionary algorithms

Huidong JIN
CSIRO Mathematics, Informatics and Statistics, Australia

Man Leung WONG
Lingnan University, Hong Kong

Recommended Citation

Jin, H., & Wong, M. L. (2010). Adaptive, convergent, and diversified archiving strategy for multiobjective evolutionary algorithms. Expert Systems with Applications, 37(12), 8462-8470. doi: 10.1016/j.eswa.2010.05.032

Adaptive, Convergent, and Diversified Archiving Strategy for Multiobjective Evolutionary Algorithms

Huidong Jin[*], Man-Leung Wong[†]

June 1, 2010

Abstract

It is crucial to obtain automatically and efficiently a well-distributed set of Pareto optimal solutions in multi-objective evolutionary algorithms (MOEAs). Many studies have proposed different evolutionary algorithms that can progress towards the Pareto front with a widely spread distribution of solutions. However, most theoretically convergent MOEAs necessitate certain prior knowledge about the Pareto front in order to efficiently maintain widespread solutions. In this paper, we propose, based on the new E-dominance concept, an Adaptive Rectangle Archiving (ARA) strategy that overcomes this important problem. The MOEA with this archiving technique provably converges to well-distributed Pareto optimal solutions without prior knowledge about the Pareto front. ARA complements the existing archiving techniques, and is useful to both researchers and practitioners.

Keywords: Evolutionary computation, multi-objective evolutionary algorithms, diversified archiving, convergence, ϵ-Pareto set

[*]Dr. Huidong Jin is with CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra ACT 2601, Australia. Email: hdjin@ieee.org. Tel: +61 2 62167258; Fax: +61 2 62167111.

[†]Dr. Man-Leung Wong is with the Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, Hong Kong. Email: mlwong@ln.edu.hk. Please direct correspondence to Dr. Man-Leung Wong at email address mlwong@ln.edu.hk or the corresponding address above.

1 Introduction

Most real-life optimization problems or decision-making problems are multi-objective in nature, since they normally have several (possibly conflicting) objectives that must be satisfied at the same time [Coello Coello et al. (2002); Deb (2001); Jin et al. (2004); Jin and Sendhoff (2003)]. Multi-Objective Evolutionary Algorithms (MOEAs) have been gaining increasing attention among researchers and practitioners because they are suitably applied to find multiple Pareto optimal solutions in a single run [Deb (2001)]. This fact enables a decision-maker to have a less-subjective search in the first phase of finding a set of well-distributed solutions. In addition, because of inherent cooperation in an evolutionary search procedure, MOEAs are computationally promising for simultaneous discovery of multiple Pareto optimal solutions. These features have attracted numerous researchers to develop different MOEAs [Coello Coello et al. (2002)] — from MOGA [Fonseca and Fleming (1993)], NPGA [Horn et al. (1994)], and NSGA [Srinivas and Deb (1994)] with skillful fitness assignment and nondominated sorting; to SPEA [Zitzler and Thiele (1999)], PESA–II [Corne et al. (2001)], NSGA–II [Deb et al. (2002)], SPEA2 [Zitzler et al. (2002)], IMOEA [Tan et al. (2001)], and DMOEA [Yen and Lu (2003)] with elitism, diversity estimation and maintenance; to PAES [Knowles and Corne (2000)] (based on AGA [Knowles and Corne (2003)]) and ϵ-MOEA [Deb et al. (2003)] (based on ϵ-dominance [Laumanns et al. (2002)]) with sound diversity and convergence guarantee.

Despite the great success of these MOEAs, there have been few successful attempts of developing *convergence-guaranteed* and *computationally efficient* procedures that maintain a *well-distributed Pareto optimal set* with *little prior knowledge about the objective space.* Most MOEAs may get widespread solutions using different diversity exploitation mechanisms [Corne et al. (2001); Deb et al. (2002); Horn et al. (1994); Yen and Lu (2003); Zitzler et al. (2002); Zitzler and Thiele (1999)], but few of them have convergence guarantee. Some early theoretical work has pointed out some approaches to enable MOEAs to converge to Pareto front [Hanne (1999); Rudolph and Agapie (2000)], but with little consideration of the distribution of the Pareto optimal solutions obtained [Laumanns et al. (2002)]. Several recent studies have made substantial progress to maintain diversified and Pareto optimal solutions using archiving [Knowles and Corne (2003); Laumanns et al. (2002)]. These archiving techniques have been successfully and widely used in handling multiobjective problems [Coello Coello et al. (2004); Deb et al. (2003); Knowles and Corne (2004, 2000)]. The recent theoretical research, based on no-free-lunch theorems, substantiates that MOEAs which use different archiving schemes can differ in overall performance, i.e., the performance averaged over all possible problems [Corne and Knowles (2003)]. Thus, archiving has been becoming an essential part of MOEAs. However, the archiving techniques in [Laumanns et al. (2002)] desire the distribution knowledge about the Pareto front beforehand. If the parameters are not set appropriately, in some extreme cases, only one solution is archived because it ϵ-dominates all the others [Knowles and Corne (2003, 2004)]. In addition, the Adaptive Grid Archiving (AGA) strategy has been proved to converge to a Pareto optimal set of bounded size under certain conditions [Knowles and Corne (2003)]. Unfortunately, this condition is not easily satisfied, and the solution oscillation problem has happened in practical applications [Knowles and Corne (2003)] or has been demonstrated empirically [Fieldsend et al. (2003); Knowles and Corne (2004)].

One basic idea of these effective and efficient diversity preserving mechanisms is to partition the whole objective space into mutually excluding regions, and then consider the Pareto optimality and diversity locally in these regions [Corne et al. (2001); Deb et al. (2003); Knowles and Corne (2003, 2000); Laumanns et al. (2002); Tan et al. (2001); Yen and Lu (2003)]. Each region is of limited volume while the range of the Pareto front is not known *a priori*. This conflict makes it difficult to explore the whole objective space, and results in the unexpected difficulty in the recent work [Knowles and Corne (2003); Laumanns et al. (2002)]. In this work, we introduce the new concept of *open rectangles* and apply the (open) rectangles[1] in the space partitioning, such that even an infinite objective space can be enveloped by a limited number of rectangles. We introduce an Extended Pareto dominance (E-dominance) to achieve this idea. In addition, our objective space partitioning is adjusted adaptively according to the solutions found so far. The archive retains well-distributed Pareto optimal solutions in the crucial region enveloped by the normal rectangles, and some accidental solutions in the region enveloped by the open rectangles. Therefore, our proposed Adaptive Rectangle Archiving (ARA) strategy can explore the whole objective space, and maintain representative Pareto optimal solutions efficiently and automatically. Moreover, these archived solutions approximately dominate the whole Pareto front without any prior knowledge.

In the rest of the paper, we first give a template of MOEAs with archiving. Then, in Section 3, we review the existing MOEAs and discuss why they do not have sound convergence and diversity guarantee when no prior knowledge is available. In section 4, the E-dominance concept and the E-Pareto set are introduced, and ARA is proposed to retain an E-Pareto set, which approximately dominates the whole Pareto front. This is supported by the theoretical results, both based on iterations and infinite treads, given in Section 5. In Section 6, conclusive comments and possible future research are discussed.

2 Preliminaries

Without loss of generality, we focus on minimization multiobjective problems in this paper. However, either by using the *duality* principle [Deb (2001)] or by simple modifications to the domination definitions, these definitions and algorithms can be used to handle maximization or

[1]Strictly speaking, they are *rectangular (hyper)parallelepipeds* and *open rectangular (hyper)parallelepipeds* in high dimensional objective space, respectively. For the sake of convenience, we use rectangles and open rectangles, terms in 2-dimensional space, to indicate them throughout the paper.

combined minimization and maximization problems.

For a multiobjective function Γ from $X(\subseteq \Re^d)$ to a finite set $Y(\subset \Re^m, m \geq 2)$, a decision vector $\mathbf{x}^* = [x_1^*, x_2^*, \cdots, x_d^*]^T$ is *Pareto optimal* if and only if for any other decision vector $\mathbf{x} \in X$, their objective vectors $\mathbf{y}^* = \Gamma(\mathbf{x}^*) = [y_1^*, y_2^*, \cdots, y_m^*]^T$ and $\mathbf{y} = \Gamma(\mathbf{x})$ holds either

$$y_i^* \leq y_i \text{ for any objective } i(1 \leq i \leq m),$$

or there exist two different objectives i, j such that

$$(y_i^* < y_i) \wedge \left(y_j^* > y_j\right).$$

Thus, for a Pareto optimal decision vector $\mathbf{x}^*$, there exists no decision vector $\mathbf{x}$ which would decrease some objective values without causing a simultaneous increase in at least one other objective. These Pareto optimal decision vectors are good trade-offs for the multiobjective optimization problem. For finding these vectors, dominance in the objective space plays an important role. An objective vector $\mathbf{y}^{(1)} = \Gamma(\mathbf{x}^{(1)}) = \left[y_1^{(1)}, y_2^{(1)}, \cdots, y_m^{(1)}\right]^T$ *dominates* another objective vector $\mathbf{y}^{(2)} = \Gamma(\mathbf{x}^{(2)})$ if and only if the former is partially less than the latter in each objective, i.e.,

$$\begin{cases} y_i^{(1)} \leq y_i^{(2)}, & \forall i \in \{1, \cdots, m\} \\ y_j^{(1)} < y_j^{(2)}, & \exists j \in \{1, \cdots, m\}. \end{cases} \tag{1}$$

It is denoted as $\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}$. For notational convenience, $\mathbf{y}^{(1)}$ is defined to be *incomparable* with $\mathbf{y}^{(2)}$ if $\neg(\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)} \vee \mathbf{y}^{(2)} \prec \mathbf{y}^{(1)} \vee \mathbf{y}^{(1)} = \mathbf{y}^{(2)})$. It is denoted as $\mathbf{y}^{(1)} \sim \mathbf{y}^{(2)}$. We also denote $\neg\left(\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}\right)$ as $\mathbf{y}^{(1)} \not\prec \mathbf{y}^{(2)}$. That means $(\mathbf{y}^{(1)} = \mathbf{y}^{(2)} \vee \mathbf{y}^{(2)} \prec \mathbf{y}^{(1)} \vee \mathbf{y}^{(1)} \sim \mathbf{y}^{(2)})$.

Likewise, the *dominance* and *incomparability* relations can be defined between an objective vector $\mathbf{y}$ and a set $A(\subseteq Y)$:

$$\mathbf{y} \prec A \iff \exists \mathbf{a} \in A, \quad \mathbf{y} \prec \mathbf{a} \tag{2}$$

$$A \prec \mathbf{y} \iff \exists \mathbf{a} \in A, \quad \mathbf{a} \prec \mathbf{y} \tag{3}$$

$$\mathbf{y} \sim A \iff \forall \mathbf{a} \in A, \quad \mathbf{y} \sim \mathbf{a} \tag{4}$$

$$\mathbf{y} \nprec A \iff \forall \mathbf{a} \in A, \quad \mathbf{y} \nprec \mathbf{a} \tag{5}$$

$$A \nprec \mathbf{y} \iff \forall \mathbf{a} \in A, \quad \mathbf{a} \nprec \mathbf{y}. \tag{6}$$

Given the set of objective vectors Y, its *Pareto front* Y^* contains all vectors $\mathbf{y}^* \in Y$ that are not dominated by any other vector $\mathbf{y} \in Y$. That is, $Y^* = \{\mathbf{y}^* \in Y | \nexists \mathbf{y} \in Y, \mathbf{y} \prec \mathbf{y}^*\}$. We call its subset a *Pareto optimal set.* Each $\mathbf{y}^* \in Y^*$ is *Pareto optimal* or *nondominated.* A Pareto optimal solution reaches a good tradeoff among these conflicting objectives: one objective cannot be improved without worsening any other objective. In this paper, we assume that there are at least two different values for each objective on the Pareto front Y^*, which holds for almost all multiobjective problems.

In the general case, it is impossible to find an analytic expression of the Pareto front. The normal procedure to find the Pareto front is to compute the objective values of decision vectors sufficiently enough, and then determine the Pareto optimal vectors to form the Pareto front [Coello Coello et al. (2004)]. However, for many multiobjective optimization problems, the Pareto front Y^* is of substantial size, and the determination of Y^* is computationally prohibitive. Thus, the whole Pareto front Y^* is usually difficult to get and maintain. Furthermore, it is questionable to regard the whole Pareto front as an ideal answer [Fieldsend et al. (2003); Laumanns et al. (2002)]. The value of presenting such a large set of solutions to a decision maker is also doubtful in the context of decision support. Usually, a small set of representative Pareto optimal solutions are expected. Finally, in a solution set of bounded size, preference information could be used to steer the process to certain parts of the search space. Therefore, all practical implementations of MOEAs have maintained (off-line) a bounded archive of best (nondominated) solutions found so far [Knowles and Corne (2003)].

Procedure 1 ***A MOEA with Archiving***

1. $t := 0$, $A^{(0)} := \emptyset$;
2. ***Repeat:***
3. $\quad t := t + 1$;
4. $\quad \mathbf{y}^{(t)} := \textsc{Evolution}(A^{(t-1)})$; /* ***Generates a solution*** */
5. $\quad A^{(t)} := \textsc{Archive}(A^{(t-1)}, \mathbf{y}^{(t)})$; /* ***Update Archive*** */
6. ***Termination:*** *Until stopping criterion fulfilled;*
7. ***Output:*** $A^{(t)}$, t.

In order to facilitate our analysis on archiving strategies, we separate the evolutionary procedure and the archiving procedure as done in [Coello Coello et al. (2004); Knowles and Corne (2003); Laumanns et al. (2002)]. Procedure 1 gives an abstract description of a MOEA with archiving. The integer t denotes the iteration count, the m-dimensional objective vector $\mathbf{y}^{(t)}$ is the solution generated at iteration t, and the set $A^{(t)}$ is the archive at iteration t and should contain a representative subset of the objective space Y. Evolution represents an evolutionary algorithm, where the evolutionary operator is associated with variation (recombination, mutation, and selection). It can generate a population of solutions, possibly using the contents of the old archive $A^{(t-1)}$. However, it only outputs a new solution at each iteration t to facilitate our theoretical analyses. Archive gets the new solution $\mathbf{y}^{(t)}$ and the old archive $A^{(t-1)}$, and then determines the updated archive $A^{(t)}$. The archive may be used in two ways: On one hand, it is used to store the best representative solutions found so far; On the other hand, the evolutionary operator exploits this archive to steer the search to promising regions.

This paper mainly deals with the procedure Archive, i.e., how to appropriately update the archive. For each objective vector $\mathbf{y}$, additional information about the corresponding decision values could be associated to the archive, but is of no concern in this paper. It is also out of our concern of keeping different decision vectors that map to the same point in the objective space. According to the requirements of MOEAs, an ideal archiving strategy should have the following four properties.

Pareto optimality: They converge to the Pareto front in each run;

Good distribution: Archived solution are distributed on the whole Pareto front in a well-defined sense;

Computational efficiency: The time and memory complexity should be low;

Little prior knowledge requirement: Little knowledge about the multiobjective problem, especially the Pareto front, is necessitated beforehand.

This last property may facilitate us greatly, since most of time, we have to make decisions on some conflicting problems with little prior knowledge. However, most existing effective archiving strategies require the knowledge about the Pareto front *a priori*. As mentioned above, useful information of the objective space may be stored in the archive during running. Archiving can be adjusted adaptively. We present such a archiving strategy in Section 4, after the discussion of the existing techniques in the next section.

3 Related work

We briefly discuss a number of archiving or elitist strategies in the literature of MOEAs. Early theoretical work of MOEAs mainly concentrates on convergence, or the Pareto optimality. Hanne (1999) gave a convergence proof for a $(\mu+\lambda)$-MOEA with Gaussian mutation distributions over a compact real search space by the application of a (negative) efficiency preservation selection scheme, which only accepts new solutions dominating at least one of the archived solutions.

The algorithm is efficient, but puts no attention on the distribution of solutions, and arbitrary regions may become unreachable [Knowles and Corne (2003); Laumanns et al. (2002)]. Based on [Rudolph (1998)], Rudolph and Agapie (2000) applied stochastic process techniques to show that MOEAs with a fixed-size archive and a sophisticated selection operator can avoid the problem of deterioration. Their algorithms with evolutionary operators having a positive transition probability matrix provably converge to the Pareto optimal ones, but they do not guarantee a good distribution of the solutions archived.

A number of elitist MOEAs have been developed to address diversity of the archived solutions. The diversity exploitation mechanisms include mating restriction, fitness sharing (NPGA [Horn et al. (1994)]), clustering (SPEA [Zitzler and Thiele (1999)], SPEA2 [Zitzler et al. (2002)]), nearest neighbor distance (NSGA-II [Deb et al. (2002)]), crowding count (PAES [Knowles and Corne (2000)], PESA-II [Corne et al. (2001)], DMOEA [Yen and Lu (2003)]), or some preselection operators [Deb (2001)]. Most of them are quite successful, but they cannot ensure convergence to Pareto optimal sets.

Recently, Laumanns et al. (2002) proposed several archiving strategies that guarantee to progress towards the Pareto front and covers roughly the whole range of Pareto optimal solutions. The algorithms maintain a bounded archive of Pareto optimal solutions that is iteratively updated in the presence of a new solution based on the concept of ϵ-dominance. However, the ϵ value, which determines solution resolution, would either be set manually or be determined adaptively [Laumanns et al. (2002)]. In the former case, the size of the archive is bounded only by a function of the objective space ranges, which is usually unknown in advance and not easy to set [Knowles and Corne (2004)]. Whereas in the latter case, ϵ may become arbitrarily large, and thus only poor representatives of the sequence of solutions presented to the archive are retained. In some extreme cases, only one solution is archived since it ϵ-dominates all other Pareto optimal solutions [Knowles and Corne (2003, 2004)].

More recently, Knowles and Corne (2003) analyzed a metric-based archiving and an Adaptive Grid Archiving (AGA) strategies. The metric-based one requires S-metric which assigns a scalar value to each possible approximation set reflecting its quality and fulfilling certain monotonicity conditions. Convergence is then defined as the achievement of a local optimum of the quality function. However, its computational overhead is prohibitively high for optimization problems with several objectives. Their AGA strategy, implemented in PAES [Knowles and Corne (2000)], maintains solutions in some critical hyperboxes of the Pareto front once they have been found. The strategy is provably convergent when the Pareto front spans the feasible objective space in all objectives. This prerequisite is not satisfied for most optimization problems with more than two objectives. Thus, the oscillation problem of the archive has happened in practical applications [Knowles and Corne (2003)] or been demonstrated empirically [Fieldsend et al. (2003); Knowles and Corne (2004)]. The archiving strategy in [Fieldsend et al. (2003)] aims at archiving all the Pareto optimal solutions and is normally not computationally efficient.

In order to diversify the archived solutions, the density estimation or diversity preservation has been locally made in some boxes for computational efficiency. However, the objective space is usually unknown in advance and sometimes infinite. Thus, it is often impractical to use boxes to envelop the space appropriately. This issue results in the oscillation of AGA [Knowles and Corne (2003)] and probably poor representation of the Pareto front in [Laumanns et al. (2002)], though they may generate widespread Pareto optimal solutions.

4 Adaptive Rectangle Archiving Strategy

4.1 Sequential Archiving and Range of Pareto Fronts

We first introduce two definitions for the range of the Pareto front Y^*.

Definition 1 (Nadir) *The* nadir *of the Pareto front Y^* consists of the minimal objective values* *of the Pareto front, i.e., its* i^{th} *element* $\text{Nadir}(Y^*)_i = \min_{\mathbf{y}^* \in Y^*} y_i^*$.

Definition 2 (Zenith) *The* zenith Y^* *consists of the maximal objective values of the Pareto front, i.e.,* $\text{Zenith}(Y^*)_i = \max_{\mathbf{y}^* \in Y^*} y_i^*$.

$\text{Nadir}(Y^*)$ and $\text{Zenith}(Y^*)$ indicate the range of the whole Pareto front. A *deterministic archiving* algorithm gives a deterministic output (an archive) for an input sequence of solutions. At each iteration, its archiving decision is solely based on the current archive and the new input solution. That means the archiving algorithm is explicitly not allowed to access to previous solutions from the input sequence, except the ones in the archive. It is easy for a deterministic archiving algorithm to maintain the nadir of the Pareto front as in Section 4.3 and [Knowles and Corne (2003)]. Unfortunately, if the archive size is smaller than the Pareto front, a deterministic archiving algorithm seems impossible to maintain the zenith of the Pareto front, even every possible solution is inputted infinite times.

Theorem 1 *If the archive size is smaller than the Pareto front, i.e.,* $|A| < |Y^*|$, *no deterministic archiving strategy can guarantee to maintain the zenith of the Pareto front in a sequential manner.*

Proof: We use a counterexample to show it. Let $Y^* = \{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(n)}\}$, $n = |Y^*| > 2$, every two vectors are incomparable, and $y_i^{(k)} < y_i^{(j)}$ for $1 \le k < j \le n$ and certain $i \in \{1, \cdots, m\}$. We have $\text{Zenith}(Y^*)_i = y_i^{(n)}$. We further assume there exist n mutual incomparable objective vectors $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \cdots, \mathbf{v}^{(n)}\}$ such that $v_i^{(k)} = 1 + y_i^{(n)}$, $\mathbf{y}^{(k)} \prec \mathbf{v}^{(k)}$ for $k = 1, \cdots, n$, and $\mathbf{v}^{(k)} \sim \mathbf{y}^{(j)}$ for $k \ne j$. It is worth pointing out that such an example exists once $m > 2$.

For any archiving algorithm, the final archive A should be a subset of Y^*. Therefore, if there exists a non-Pareto optimal solution, say, $\mathbf{v}^{(k)}$, it should be substituted later by the new input $\mathbf{y}^{(k)}$ which dominates $\mathbf{v}^{(k)}$. Otherwise, if it could not be substituted, the maximal value in the

i^{th} objective of the archive will be $1 + y_i^{(n)}$, which is larger than Zenith$(Y^*)_i$. That means, the dominated solution in the archive should be substituted by one which dominates it.

On the other hand, since $|A| < n$, there must exist a Pareto optimal solution, say $\mathbf{y}^{(n)}$, not included in the archive, denoted by $A^{(n)}$. If the new input is $\mathbf{v}^{(n)}$, there are two possibilities of the archiving decision: reject it or accept it. If it is rejected, the archiving algorithm will not accept the input $\mathbf{y}^{(n)}$ too, because $\mathbf{y}^{(n)}$ and $\mathbf{v}^{(n)}$ have the same relationship with the current archive $A^{(n)}$ and should be treated in the same way. Then the maximal value in the i^{th} objective of the archive will always be smaller than $y_i^{(n)}$. If it is accepted, one Pareto optimal solution, say, $\mathbf{y}^{(k)}(k < n)$ should be removed. Later on, $\mathbf{v}^{(n)}$ in the archive will be replaced by $\mathbf{y}^{(n)}$ because of dominance. Now, the archive, denoted by $A^{(k)}$, cannot reject $\mathbf{v}^{(k)}$ because the comparison relation between $\mathbf{v}^{(k)}$ and $A^{(k)}$ is the same as the relation between $\mathbf{v}^{(n)}$ and $A^{(n)}$. Therefore, the maximal value of the archive in the i^{th} objective is varying; it is larger or smaller than Zenith$(Y^*)_i$ sometimes. So, no deterministic archiving can determine the maximal objective values of the Pareto front. ■

As we can see, there are two goals of archiving in order to retain the zenith of the Pareto front, trying to find the Pareto optimal solutions and trying to identify the maximal objective values on the Pareto front. These two goals may drive the archive to move in two conflicting directions when the archive size is smaller than the Pareto optimal set.

4.2 Extended Pareto Dominance

In this section, we present an Adaptive Rectangle Archiving (ARA) algorithm that addresses the problem of unknown Pareto fronts. In this archiving strategy, we use a self-adaption mechanism to preserve diversity according to the archived information about the objective space. To overcome the problem of unavailability of the maximal objective values on the Pareto front, we partition the whole objective space into two non-overlapping regions, *the crucial region* and *the* *open region.* In the crucial region, a solution is allowed to be preserved in a bounded rectangle[1], and thus many representative Pareto optimal solutions are archived. In the unknown, even infinite, open region, some *open rectangles* are used to envelop a solution. These open rectangles may even envelop infinite objective values. Within these open rectangles, some Pareto optimal solutions are selected to be archived. The rectangles are specified according to our extended Pareto dominance concept, which is defined below, followed by the description of ARA.

Since we need to use an bounded archive of objective vectors to approximately dominate the whole Pareto front, a method is to permit some tolerance on dominance. To achieve it, we first extend the Pareto dominance as follows.

Definition 3 (E-dominance) *Let* $\mathbf{y}^{(1)}$ *and* $\mathbf{y}^{(2)}$ *be two objective vectors.* $\mathbf{y}^{(1)}$ *is said to* E-dominate $\mathbf{y}^{(2)}$ *for a transferring function,* $\textsc{Fun}$*, and a constant vector* $\mathbf{e}(\geq \mathbf{0})$*, if and only if*

$$\textsc{Fun}(y_i^{(1)}) - e_i \leq \textsc{Fun}(y_i^{(2)}), \quad \forall\, i \in \{1, \cdots, m\}. \tag{7}$$

It is denoted as $\mathbf{y}^{(1)} \preceq_E \mathbf{y}^{(2)}$*.*

The transferring function should be continuous and monotonously increasing. This ensures that E-dominance may be implied by the traditional dominance, i.e., if $\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}$, then $\mathbf{y}^{(1)} \preceq_E \mathbf{y}^{(2)}$. When $\mathbf{y}^{(1)} \preceq_E \mathbf{y}^{(2)}$ and $\mathbf{y}^{(2)} \preceq_E \mathbf{y}^{(3)}$, $\mathbf{y}^{(1)} \preceq_E \mathbf{y}^{(3)}$ does not always hold. Thus, the E-dominance relation is not transitive.

E-dominance generalizes several dominance relations. For example, it becomes ϵ-dominance [Laumanns et al. (2002)] as $\textsc{Fun}(y_i) = \ln(y_i)$ and $e_i = \ln(1+\epsilon)$, additive ϵ-dominance [Hanne (1999); Reuter (1990)] as $\textsc{Fun}(y_i) = y_i$ and $e_i = \epsilon$, and the Pareto dominance as $\textsc{Fun}(y_i) = y_i$ and $e_i = 0$.

In order to envelop unknown, possible infinite, objective values, we may employ a nonlinear transferring function, e.g.[2], $\textsc{Fun}(y_i) = \arctan(y_i * \text{scale}_i)$. Thus, even the infinite values are

[2]Another example is $\textsc{Fun}(y_i) = \frac{y_i * \text{scale}_i}{1+y_i * \text{scale}_i}$. Here y_i is supposed to be not less than 0, which may be easily got

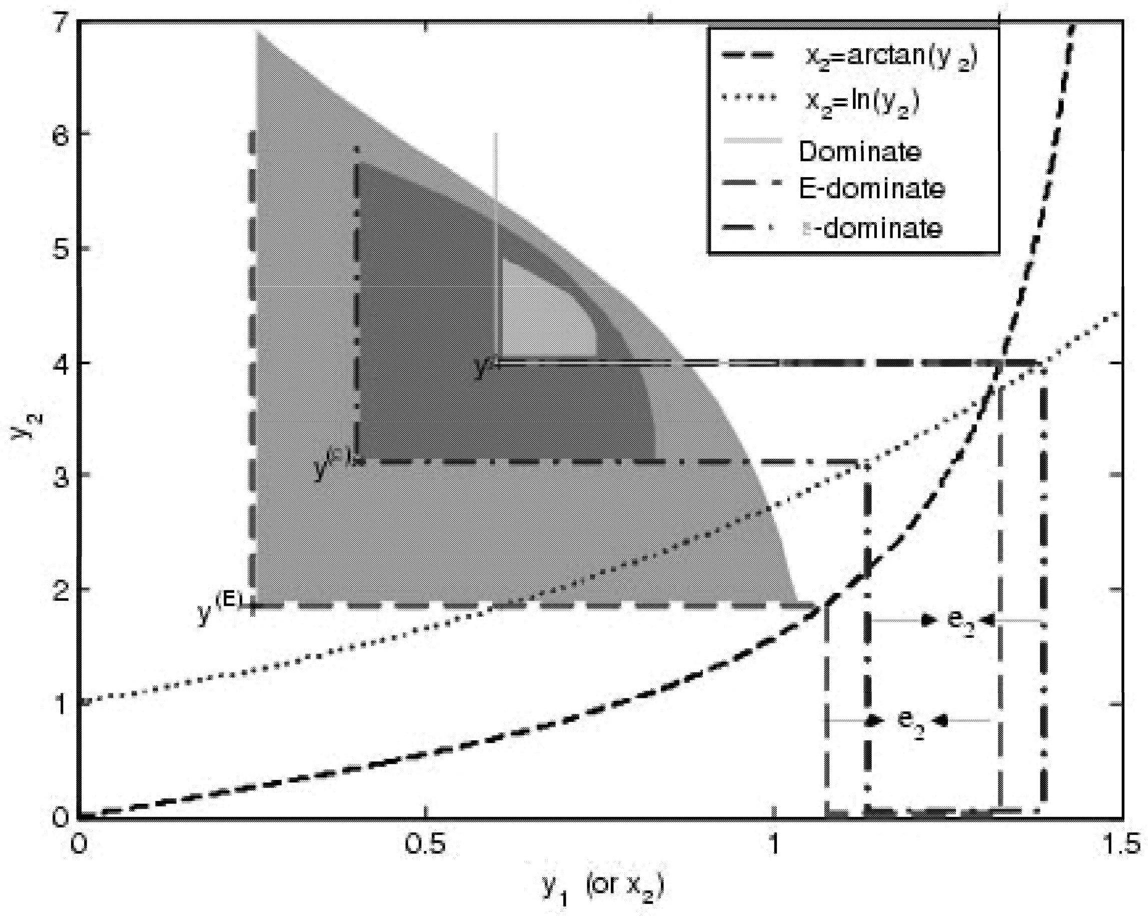

Figure 1: Illustration of Pareto dominance, E-dominance, and ϵ-dominance. The regions dominated by $\mathbf{y}$ under three different dominance relations are illustrated by three shadows respectively. The calculation of vectors $\mathbf{y}^{(E)}$ and $\mathbf{y}^{(\epsilon)}$ is illustrated in the bottom right corner. Two transferring functions are indicated by two curves.

transferred to a limited value, say $\frac{\pi}{2}$, and may be E-dominated by a bounded value, say, $\frac{\tan(\frac{\pi}{2}-e_i)}{\text{scale}_i}$. The scale_i can be specified adaptively according to the solutions found so far. In this paper, we exemplify E-dominance using the arctangent function though the definitions and algorithms are applicable to other transferring functions. The comparison among E-dominance, ϵ-dominance, and Pareto dominance is illustrated in Figure 1. Based on the E-dominance relation, we have the following definitions.

Definition 4 (E-approximate Pareto Set) *Let $Y \subset \Re^m$ be a set of objective vectors,* FUN *a monotonically increasing continuous function, and* $\mathbf{e}$ *a nonnegative vector. Then a set Y_E is called an* E-approximate Pareto set *of Y, if any vector $\mathbf{y} \in Y$ is E-dominated by at least one vector $\mathbf{a} \in Y_E$, i.e.,*

$$\forall \mathbf{y} \in Y : \exists \mathbf{a} \in Y_E \textit{ such that } \mathbf{a} \preceq_E \mathbf{y}. \tag{8}$$

The set of all E-approximate Pareto sets of Y is denoted as $P_E(Y)$.

As mentioned above, since finding the whole Pareto front of an arbitrary set Y is usually not practical because of its usually large size, one needs to be less ambitious in general. An E-approximate Pareto set is a practical solution concept as it not only represents all vectors Y approximately but also may be of smaller size. The E-approximate Pareto set is of theoretical interest. A further refinement of the concept leads to the following definition.

Definition 5 (E-Pareto Set) *Let $Y \subseteq \Re^m$ be a set of vectors,* FUN *a monotonically increasing continuous function, and* $\mathbf{e}$ *a nonnegative vector. Then a set $Y_E^* \subseteq Y$ is called an* E-Pareto set *of Y, if*

1. *Y_E^* is an E-approximate Pareto set of Y, i.e., $Y_E^* \in P_E(Y)$, and*
2. *Y_E^* only contains Pareto optimal vectors of Y, i.e., $Y_E^* \subseteq Y^*$*

after a linear transformation as shown in Section 4 or in [Knowles and Corne (2004)].

The set of all E-Pareto sets of Y is denoted as $P_E^*(Y)$.

Compared with an E-approximate Pareto set, an E-Pareto set seems more attractive as it consists of Pareto optimal solutions only. E-Pareto sets are not unique. We can archive a bounded E-Pareto set for any Pareto front even with infinite objective value as shown in Section 4.3. There are many different concepts for Pareto set approximation in the literature [Laumanns et al. (2002)]. Most of them deals with infinite sets, which are of theoretical interest [Hanne (1999)] but are not practical for our purpose of producing and maintaining a representative subset. Some of them, say ϵ-dominance [Laumanns et al. (2002)] and additive ϵ-dominance [Hanne (1999); Reuter (1990)] may produce bounded archives to represent the whole Pareto front if its range is given. But they are difficult to generate bounded representative subset when the range of Pareto front is previously unknown, and even infinite.

4.3 Archiving Procedure

Our adaptive archiving strategy basically has two features. One is to determine the crucial region adaptively. The other one is to find an E-Pareto set based on the E-dominance concept. For descriptive convenience, we partition the archive in ARA into two parts, i.e., $A = \{A^{(min)}, A^{(arc)}\}$. Thus, $A^{(t)} = \{A^{(min,t)}, A^{(arc,t)}\}$. The purpose of $A^{(arc)}$ is to maintain an E-Pareto set according to the solution space information collected in $A^{(min)}$. $A^{(min)}$ is an array: $A^{(min)} = \left[\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \cdots, \mathbf{a}^{(m)}\right]$. Each element, $\mathbf{a}^{(i)}$, is initialized with infinite, and stores the solution that has the minimal value found so far in the i^{th} objective. We have $a_i^{(i)} = \min\limits_{\mathbf{a}^{(k)} \in A^{(min)}} \{a_i^{(k)}\}$. Furthermore, we introduce two vectors associated with $A^{(min)}$ to indicate the crucial region, $\mathbf{a}^{(min)}$ with $a_i^{(min)} = a_i^{(i)} = \min\limits_{\mathbf{a}^{(k)} \in A^{(min)}} \{a_i^{(k)}\}$ and $\mathbf{a}^{(max)}$ with $a_i^{(max)} = \max\limits_{\mathbf{a}^{(k)} \in A^{(min)}} \{a_i^{(k)}\}$. The crucial region, whose member dominates $\mathbf{a}^{(max)}$ but is dominated by $\mathbf{a}^{(min)}$, contains most Pareto optimal solutions generated so far, and so it is decisive for archiving. For example, all solutions dominated by $\mathbf{a}^{(max)}$ are not Pareto optimal. Especially, all Pareto optimal solutions are located

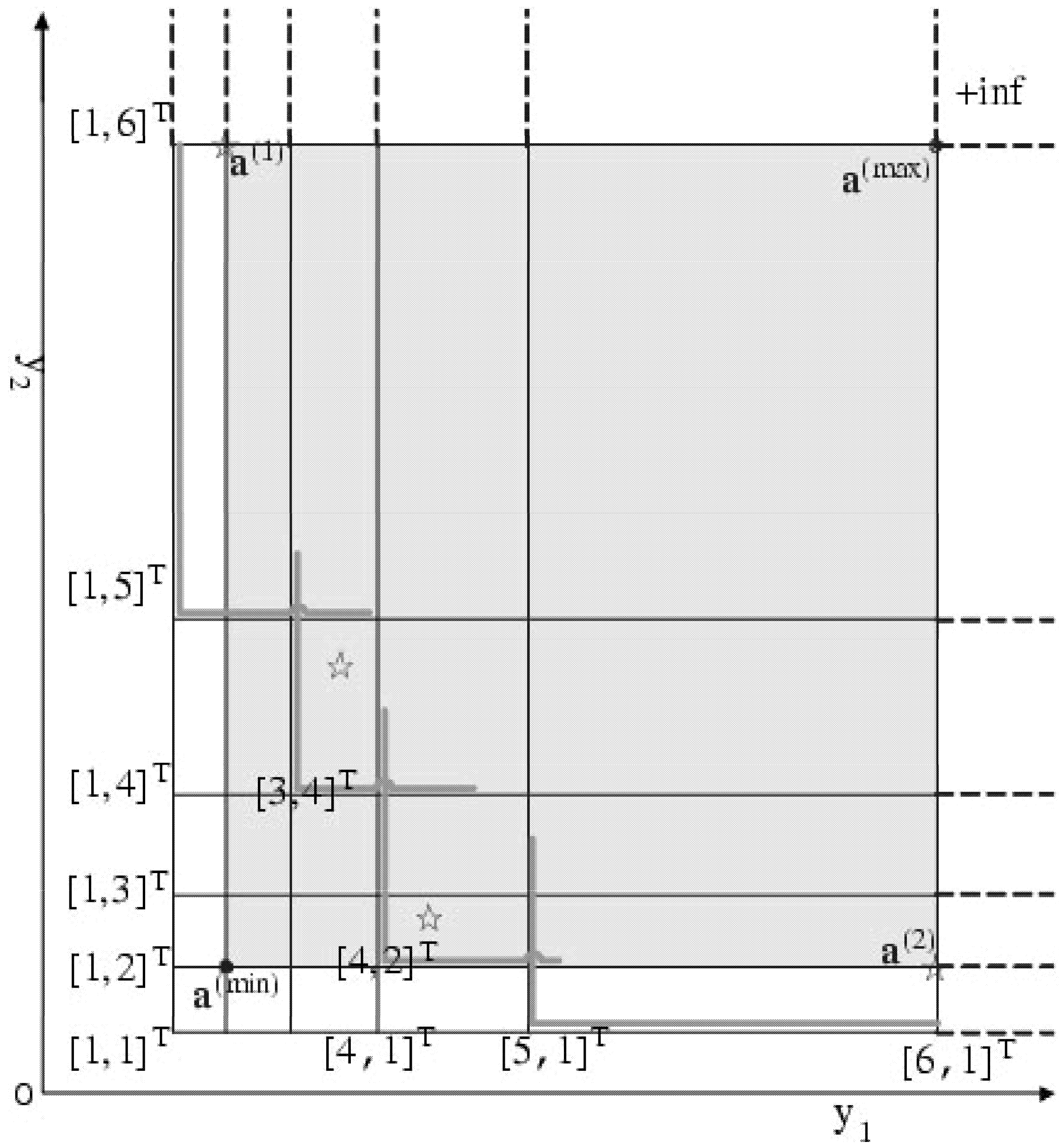

Figure 2: Adaptive rectangle partitioning of a 2-D objective space. The dashed line segments indicate open rectangles. The gray rectangle indicates the crucial region indicated by $\mathbf{a}^{(min)}$ and $\mathbf{a}^{(max)}$. The gray line segments indicate the region E-dominated by a solution, denoted by a pentagram.

in the crucial region in 2-D case. The gray rectangle in Figure 2 indicates the crucial region, and envelops all four Pareto optimal solutions, indicated by pentagrams.

Procedure 2 ARA($\mathbf{y}, A$) */* Adaptive Rectangle Archiving */*

if $\left(\left(\mathbf{y} \prec A^{(min)}\right) \vee \left(\mathbf{y} \sim \mathbf{a}^{(min)}\right)\right)$ ***then***
 for all $i \in \{1, \cdots, m\}$ ***do***
  if $\left(y_i < a_i^{(i)}\right)$ ***then***
   $\mathbf{a}^{(i)} := \mathbf{y}$;   /* RECEDES */
  else if $(\mathbf{y} \prec \mathbf{a}^{(i)})$
   $\mathbf{a}^{(i)} := \mathbf{y}$;   /* DOMINATES */
  end if
 end do
 $A^{(arc_2)} := \emptyset$;   /* RE-FORMS $A^{(arc)}$ */
 for all $\mathbf{a} \in A$ *such that* $A^{(min)} \not\prec \mathbf{a}$ ***do***
  INSERTINRECT$\left(\mathbf{a}, A^{(arc_2)}, A^{(min)}\right)$;
 end do
 $A^{(arc)} := A^{(arc_2)}$;
else if $(A^{(min)} \not\prec \mathbf{y})$ /* UPDATES $A^{(arc)}$ */
 INSERTINRECT$\left(\mathbf{y}, A^{(arc)}, A^{(min)}\right)$;
end if
$A := \{A^{(min)}, A^{(arc)}\}$;

The pseudo code of our archiving strategy, ARA, is given in Procedure 2, which is illustrated in Figure 3. In each iteration, the algorithm first checks whether the crucial region determined by $A^{(min)}$ should be updated. If a new objective value is smaller than the minimal one of the archive, RECEDES replaces the archived vector with the new one, and the minimal objective value is archived; If the new vector dominates a vector in $A^{(min)}$, DOMINATES will also replace the old vector with the new one. This operation ensures the convergence of $A^{(min)}$ to the Pareto front, and then the convergence of $\mathbf{a}^{(min)}$ and $\mathbf{a}^{(max)}$. If the crucial region is updated, the solutions in $A^{(arc)}$ have to be archived again (RE-FORMS $A^{(arc)}$), such that the solutions in $A^{(arc)}$ are always chosen based on the current $A^{(min)}$.

If the input vector $\mathbf{y}$ neither has a smaller objective value nor dominates any vector in $A^{(min)}$, then it is processed by InsertInRect, whose pseudo code is given in Procedure 3. The procedure mainly chooses representative Pareto optimal solutions based on the crucial region specified by $A^{(min)}$. It can be described at two levels. On the coarse level, the objective space is discretized by dividing it into rectangles (see Function 4), where each vector uniquely belongs to one (open) rectangle. Using the proposed E-dominance relation on these rectangles, the algorithm always maintains a set of nondominated rectangles through InterRectDom and Occupies. They guarantee the E-approximate Pareto property. On the fine level, at most one solution is maintained for each rectangle, which assures well-distribution of solutions. Within a rectangle, each representative vector can only be replaced by a dominating one (IntraRectDom), which ensures the convergence to the Pareto front.

Procedure 3 is similar to Algorithm 3 in [Laumanns et al. (2002)], which generates an ϵ-Pareto set. However, they are based on different dominance relations and they partition the objective space in different ways. Only Procedure 3 can handle an unknown objective space in a reasonable way. This idea is embodied in the function Rect outlined in Function 4.

The function Rect gives a possible implementation which partitions the crucial region finely while envelops the unknown regions with open rectangles based on E-dominance. Since it is arduous to detect automatically the maximal objective values on the Pareto front [Knowles and Corne (2003)], we simply treat it as infinite. As shown in Lines 2-4 in Rect, only $a_i^{(min)}$ is mapped into 1, while $+\infty$ is mapped into $\left\lceil \frac{\frac{\pi}{2}}{e_i} + 1 \right\rceil$. So, the rectangles are open if its "coordinates" contain $\left\lceil \frac{\frac{\pi}{2}}{e_i} + 1 \right\rceil$, e.g., $[1,6]^T$ and $[6,1]^T$ in Figure 2. In Line 2, scale_i is calculated according to the difference between $a_i^{(max)}$ and $a_i^{(min)}$. The farther away $a_i^{(max)}$ is from $a_i^{(min)}$, the larger the scale_i value is. Furthermore, this scale, together with the constant of 1 in Line 4,

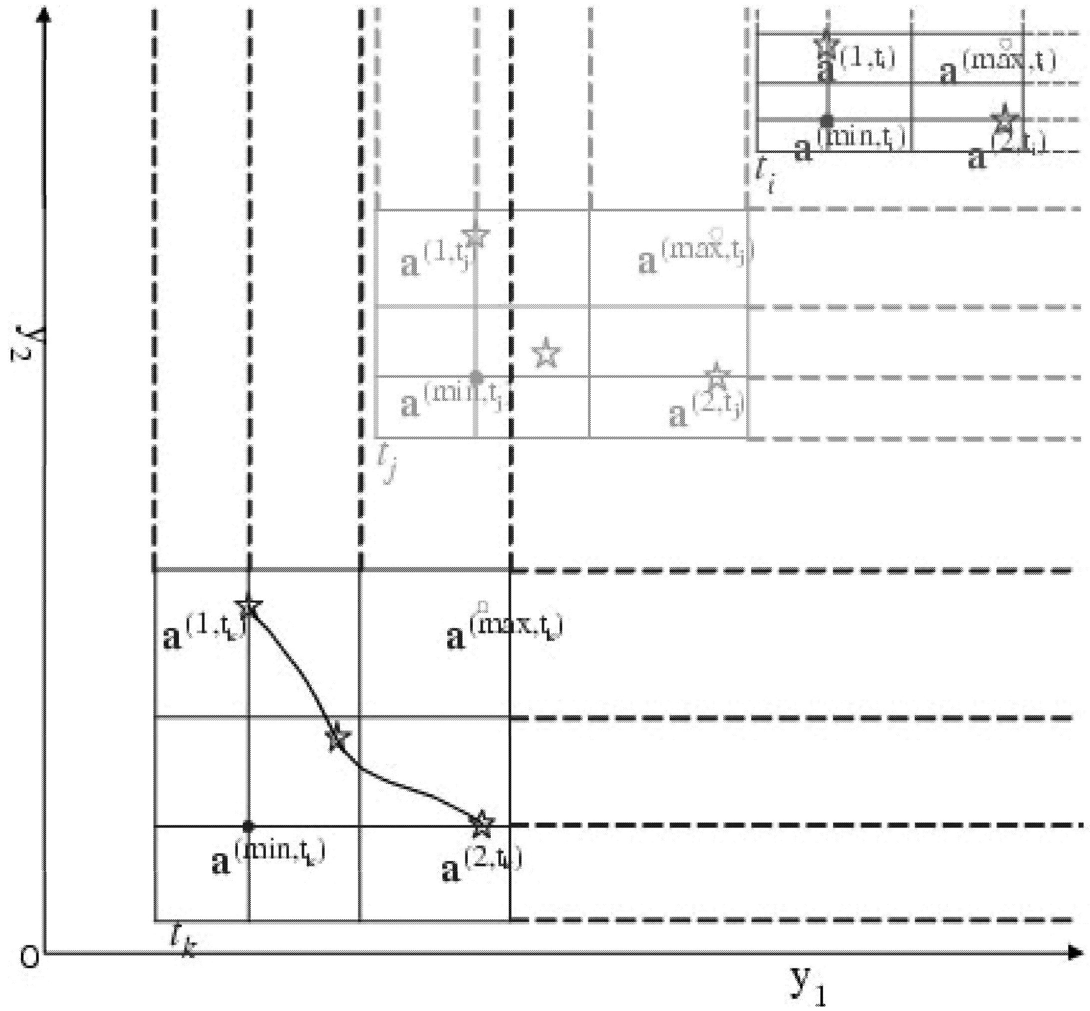

Figure 3: Illustration of adaptive rectangles who change their location and shapes in the objective space as the objective vectors in the archive $A^{(t)}$ change through iterations $t_i < t_j < t_k$. The curve indicates the Pareto front, pentagrams indicate archived solutions, small discs indicate $\mathbf{a}^{(min,t)}$, and small circles indicate $\mathbf{a}^{(max,t)}$.

Procedure 3 $\textsc{InsertInRect}(\mathbf{y}, A^{(arc)}, A^{(min)})$

$D := \{\mathbf{a} \in A^{(arc)} | \textsc{Rect}(\mathbf{y}, A^{(min)}) \prec \textsc{Rect}(\mathbf{a}, A^{(min)})\}$;
if $D \neq \emptyset$ ***then***
    $A^{(arc)} := (A^{(arc)} \bigcup \mathbf{y}) \backslash D$; /* $\textsc{InterRectDom}$ */
else if $\exists \mathbf{a} \in A^{(arc)} : (\textsc{Rect}(\mathbf{a}, A^{(min)}) = \textsc{Rect}(\mathbf{y}, A^{(min)})) \wedge (\mathbf{y} \prec \mathbf{a})$ ***then***
    $A^{(arc)} := A^{(arc)} \bigcup \{\mathbf{y}\} \backslash \{\mathbf{a}\}$ /* $\textsc{IntraRectDom}$ */
else if $\forall \mathbf{a} \in A^{(arc)} : \textsc{Rect}(\mathbf{a}, A^{(min)}) \sim \textsc{Rect}(\mathbf{y}, A^{(min)})$
    $A^{(arc)} := A^{(arc)} \bigcup \{\mathbf{y}\}$ /* $\textsc{Occupies}$ *a rectangle* */
else
    $A^{(arc)} := A^{(arc)}$; /* $\textsc{SteadyState}$ */
end if

enables $a_i^{(max)}$ to be mapped to $\left\lceil \frac{\frac{\pi}{2}}{e_i} \right\rceil$, e.g., 5 in Figure 2. It is next to the "coordinate" corresponding to $+\infty$. Therefore, when $e_i < \frac{\pi}{4}$, $a_i^{(min)}$, $a_i^{(max)}$, and $+\infty$ are mapped into different rectangles. Furthermore, there are $\left(\left\lceil \frac{\frac{\pi}{2}}{e_i} \right\rceil - 1 \right)$ rectangles between $a_i^{(min)}$ and $a_i^{(max)}$. The less e_i is, the more finely the crucial region is partitioned. An example with $e_i = \frac{\pi}{10}$ is illustrated in Figure 2. The open region is enveloped by some open rectangles, which are indicated by dashed line segments. Finally, the crucial region is finely partitioned, and a well-distributed Pareto optimal solutions are archived in $A^{(arc)}$, which are indicated by pentagrams in Figure 2.

5 Theoretical analyses

We now give some theorems to show that the archive of ARA converges to Pareto optimal sets under appropriate conditions, and at the same time preserves diversity of solution. We first give theoretical analyses on each iteration of Procedures 2 and 3.

The following theorem shows that the lower boundaries of archive $A^{(t)}$, i.e., $\mathbf{a}^{(min,t)}$, retain the minimal objective values generated so far.

Function 4 RECT($\mathbf{y}$, $A^{(min)}$)

for all $i \in \{1, \cdots, m\}$ ***do***

$\quad \text{scale}_i = \frac{\tan\left(\frac{\pi}{2} - e_i\right)}{a_i^{(max)} - a_i^{(min)}}$;

$\quad \alpha_i := \arctan\left(\left(y_i - a_i^{(min)}\right) * \text{scale}_i\right)$; *\/* Arctangent transferring function *\/*

$\quad r_i := 1 + \left\lceil \frac{\alpha_i}{e_i} \right\rceil$; */* Rectangle's i^{th} "Coordinate" */*

end do

output: *return* $\mathbf{r} = [r_1, \cdots, r_m]^T$.

Theorem 2 *Let* $Y^{(\tau)} = \bigcup_{t=1}^{\tau} \{\mathbf{y}^{(t)}\}$ *be the set of objective vectors generated by* EVOLUTION. *Then the archive* $A^{(\tau)}$ *contains the minimal objective values of* $Y^{(\tau)}$. *That is,* $a_i^{(min,\tau)} = \min_{t=1,\cdots,\tau} \{y_i^{(t)}\}$.

Proof: We need to prove two different cases:

Case 1. the minimal objective values generated-so-far will enter the archive;

Case 2. the objective vectors with the minimal objective values in the archive will not lose.

Since $A^{(min,t)}$ is a part of $A^{(t)}$, we only need to prove these two cases based on $A^{(min,t)}$.

For Case 1, we only need to prove $a_i^{(min,t)} = y_i^{(t)}$ when a smaller objective value is generated for any $i \in \{1, \cdots, m\}$ and $t \leq \tau$, i.e, when $y_i^{(t)} < a_i^{(min,t-1)}$. On this iteration, we have $\left(\mathbf{y}^{(t)} \prec \mathbf{a}^{(min,t-1)}\right)$ or $\left(\mathbf{y}^{(t)} \sim \mathbf{a}^{(min,t-1)}\right)$. Since $a_i^{(min,t-1)} = a_i^{(i,t-1)}$, we have either $\left(\mathbf{y}^{(t)} \prec \mathbf{a}^{(i,t-1)}\right)$ or $\left(\mathbf{y}^{(t)} \sim \mathbf{a}^{(i,t-1)}\right)$. If $\mathbf{y}^{(t)} \prec \mathbf{a}^{(i,t-1)}$, then $\mathbf{y}^{(t)} \prec A^{(min,t-1)}$, and the operation DOMINATES, executes. When $\mathbf{y}^{(t)} \sim \mathbf{a}^{(i,t-1)}$, if $\mathbf{y}^{(t)} \prec \mathbf{a}^{(min,t-1)}$, then $\mathbf{y}^{(t)} \prec \mathbf{a}^{(min,t-1)} \prec \mathbf{a}^{(i,t-1)}$ (or $= \mathbf{a}^{(i,t-1)}$). It contradicts $\mathbf{y}^{(t)} \sim \mathbf{a}^{(i,t-1)}$. Thus, $\mathbf{y}^{(t)} \sim \mathbf{a}^{(min,t-1)}$, and the operation RECEDES executes. For both situations, $\mathbf{a}^{(i,t)} = \mathbf{y}^{(t)}$. Thus, $a_i^{(min,t)} = y_i^{(t)}$.

For Case 2, we have to prove $a_i^{(min,t)} = a_i^{(min,t-1)}$ if $a_i^{(min,t-1)} \leq y_i^{(t)}$ for any $i \in \{1, \cdots, m\}$. If $a_i^{(min,t-1)} < y_i^{(t)}$, $a_i^{(i,t-1)} < y_i^{(t)}$ and then $\mathbf{y}^{(t)} \not\prec \mathbf{a}^{(i,t)}$. Thus, both DOMINATES and RECEDES will not execute, and $\mathbf{a}^{(i,t)} = \mathbf{a}^{(i,t-1)}$. If $a_i^{(min,t-1)} = y_i^{(t)}$, only the operation DOMINATES may execute, which still leads to $a_i^{(min,t)} = a_i^{(min,t-1)}$. The proof is completed. ■

According to Procedures 2 and 3, a new solution dominated by the archived solutions in $A^{(min)}$ (or $A^{(arc)}$) is impossible to enter $A^{(min)}$ (or $A^{(arc)}$). Thus, any solution in $A^{(min)}$ (or $A^{(arc)}$) cannot dominate other solutions in $A^{(min)}$ (or $A^{(arc)}$). As required by the conditions in Lines 10 and 14 of Procedure 2, if a new solution is dominated by $A^{(min)}$, it cannot enter $A^{(arc)}$. Furthermore, according to the condition in Line 1 of Procedure 2, any solution dominating $A^{(min)}$ will enter $A^{(min)}$ first before inputting into $A^{(arc)}$. It means that a solution in $A^{(arc)}$ also cannot dominate $A^{(min)}$. Thus, we have the following nondominated relations among the solutions in the archive $A = \{A^{(min)}, A^{(arc)}\}$.

Lemma 1 *Members of $A^{(t)}$ are either nondominated or equal to one another, i.e., $\forall \mathbf{a}^0, \mathbf{a}^1 \in A^{(t)}$, $(\mathbf{a}^0 \sim \mathbf{a}^1) \vee (\mathbf{a}^0 = \mathbf{a}^1)$.*

Similar to that $A^{(min,t)}$ retains the minimal objective values inputted so far in Theorem 2, $A^{(arc,t)}$ collects the Pareto optimal solutions iteratively, as stated in the following theorem.

Theorem 3 *The archive $A^{(arc,t)}$ is an E-Pareto set of $Y^{(t_0,t)} \triangleq \left\{ \bigcup_{\tau=t_0}^{t} \{\mathbf{y}^{(\tau)}\} \right\} \bigcup A^{(arc,t_0)}$ if $A^{(min,t)} = A^{(min,t_0)}$ and $t > t_0 > 0$; $A^{(t)}$ is an E-Pareto set of $Y^{(t_0,t)} \bigcup A^{(min,t_0)}$ if $A^{(min,t)} = A^{(min,t_0)}$.*

Proof: The second statement is a direct consequence of the first statement and Lemma 1. Thus we only need to prove the first one.

We consider an extreme scenario first: $\mathbf{a}^{(1,t)} = \cdots = \mathbf{a}^{(m,t)} = \mathbf{a}^{(min,t)}$. At this iteration, $A^{(arc,t)}$ can contain only $\mathbf{y} = \mathbf{a}^{(min,t)}$. In fact, if $\mathbf{y}(\in A^{(arc,t)}) \neq \mathbf{a}^{(min,t)}$, then $\mathbf{y} \sim \mathbf{a}^{(min,t)}$ according to Lemma 1. However, this contradicts Theorem 2 that $\mathbf{a}^{(min,t)}$ dominates all solutions generated so far. So, $A^{(arc,t)} = \left\{\mathbf{a}^{(min,t)}\right\}$ dominates, and then E-dominates, the solutions generated so far. The theorem is correct.

Note that each update on $A^{(min,t)}$ (either DOMINATES or RECEDES), $A^{(arc,t)}$ will be updated by using Procedure 3 with the new $A^{(min,t)}$ (as required by RE-FORMS $A^{(arc)}$ in Procedure 2).

So we can consider Procedure 3 by simply assuming $A^{(arc,t_0)} = \emptyset$.

If the statement $A^{(arc,t_1)} \in P_E^*(Y^{(t_0,t_1)})$ is not true, for some $t_1 > t_0$. According to Definition 5, there are two possible cases, which will be proved impossible below,

Case 1. Some $\mathbf{y}^{(\tau_0)}$ ($\tau_0 \leq t_1$) is not E-dominated by any member of $Y^{(t_0,t_1)}$ and not in $A^{(arc,t_1)}$;

Case 2. Some $\mathbf{y}^{(\tau_0)}$ ($\tau_0 \leq t_1$) is in $A^{(arc,t_1)}$ but E-dominated by one member of $Y^{(t_0,t_1)}$.

For the first case, for $\mathbf{y}^{(\tau_0)}$ not being in $A^{(arc,t_1)}$, it can either have been rejected at $t = \tau_0$ or accepted at $t = \tau_0$ and removed later on. However, removal only takes place when some new $\mathbf{y}^{(\tau_1)}(t_1 \geq \tau_1 > \tau_0)$ enters $A^{(arc,\tau_1)}$, which dominates $\mathbf{y}^{(\tau_0)}$ (IntraRectDom) or $\textsc{Rect}(\mathbf{y}^{(\tau_1)}, A^{(min,\tau_1)}) \prec \textsc{Rect}(\mathbf{y}^{(\tau_0)}, A^{(min,\tau_1)})$ (InterRectDom). When $A^{(min,t)}$ doesn't change during $[t_0, t_1]$ as assumed by the theorem, both dominance relations are transitive. In addition, they both imply E-dominance. Thus, there will always be a member of $A^{(arc,t_1)}$ that E-dominates $\mathbf{y}^{(\tau_0)}$, which contradicts the assumption. On the other hand, $\mathbf{y}^{(\tau_0)}$ will be rejected only if there is another $\mathbf{a}^{(0)} \in A^{(arc,\tau_0)}$ with $\mathbf{a}^{(0)} \prec \mathbf{y}^{(\tau_0)}$ or $\textsc{Rect}(\mathbf{a}^{(0)}, A^{(min,\tau_0)}) \prec \textsc{Rect}(\mathbf{y}^{(\tau_0)}, A^{(min,\tau_0)})$. When $A^{(min,t)}$ doesn't change during $[t_0, t_1]$, both dominance relations are transitive. With the same argument as above, there must exist a solution $\mathbf{a} \in A^{(arc,t_1)}$ such that $\mathbf{a} \prec \mathbf{y}^{(\tau_0)}$ or $\textsc{Rect}(\mathbf{a}, A^{(min,t_1)}) \prec \textsc{Rect}(\mathbf{y}^{(\tau_0)}, A^{(min,t_1)})$. Both imply E-dominance, which contradicts the assumption that $\mathbf{y}^{(\tau_0)}$ is not E-dominated by any member of $Y^{(t_0,t_1)}$.

For the second case, since $\mathbf{y}^{(\tau_0)}(\in A^{(arc,\tau_0)})$ is not in the Pareto front of $Y^{(t_0,t_1)}$, there exists $\mathbf{y}^{(\tau_1)}$ with $t_0 \leq \tau_1(\neq \tau_0) \leq t_1$ on the Pareto front of $Y^{(t_0,t_1)}$ such that $\mathbf{y}^{(\tau_1)} \prec \mathbf{y}^{(\tau_0)}$. This implies $\textsc{Rect}(\mathbf{y}^{(\tau_1)}, A^{(min,\tau_0)}) \prec \textsc{Rect}(\mathbf{y}^{(\tau_0)}, A^{(min,\tau_0)})$ or $(\textsc{Rect}(\mathbf{y}^{(\tau_1)}, A^{(min,\tau_0)}) = \textsc{Rect}(\mathbf{y}^{(\tau_0)}, A^{(min,\tau_0)}) \wedge \mathbf{y}^{(\tau_1)} \prec \mathbf{y}^{(\tau_0)})$. Hence, if $\tau_1 < \tau_0$, $\mathbf{y}^{(\tau_0)}$ would not have entered the archive $A^{(arc,\tau_0)}$ (SteadyState). If $\tau_1 > \tau_0$, $\mathbf{y}^{(\tau_0)}$ would have been removed from $A^{(arc,\tau_1)}$ when $t = \tau_1$ (InterRectDom or IntraRectDom). Thus, $\mathbf{y}^{(\tau_0)}$ is not in $A^{(arc,t_1)}$, which contradicts the assumption. This completes the proof. ∎

Theorems 2 and 3 state that, the archiving strategy ARA is quite greedy and it retains the minimal objective values and the E-Pareto optimal solutions of the objective vectors inputted so far. The archive retains the "best-so-far" solutions, and this feature enables the MOEA using ARA to be stopped anytime. Using these features of ARA, we give the convergence results that ARA reaches a good crucial region, and then E-dominate the whole Pareto front based on a weak assumption on EVOLUTION.

Theorem 4 *If the function* EVOLUTION *gives every possible solution in the decision space with a positive minimum probability, then, with probability one as* $t \to +\infty$,

1. *the lower boundaries of archive* $A^{(t)}$ *of ARA,* $\mathbf{a}^{(min,t)}$*, converge to the minimal objective values of the whole Pareto front, and*
2. $\{A^{(min,t)}\}$ *converges to a Pareto optimal set*

$$\left\{\mathbf{a}^{(i)}, i \in \{1, \cdots, m\} \,\middle|\, \left(a_i^{(i)} = \min_{\mathbf{y}\in Y}\{y_i\}\right) \wedge (\mathbf{a}\prec\mathbf{y} \vee \mathbf{a}\sim\mathbf{y}, \forall \mathbf{y} \in Y \wedge y_i = a_i)\right\}. \tag{9}$$

Proof: *1.* Since EVOLUTION can generate every possible solution with a positive minimum probability, according to the Borel-Cantelli Lemma (see e.g., [(Feller, 1976, p. 201)]), it is guaranteed that any solution is generated infinitely often and that the waiting time for the first occurrence as well as for the second, and so forth is finite with probability 1. Thus, for any $i \in \{1, \cdots, m\}$, there exists $t_i(< +\infty)$ such that $\mathbf{y}_i^{(t_i)} = \min_{\mathbf{y}\in Y}\{y_i\}$. According to Theorem 2, $a_i^{(min,t)} = \min_{\mathbf{y}\in Y}\{y_i\}$ for all $t > t_i$. Therefore, when $t > \tau_{c_1} \triangleq \max_{i\in\{1,\cdots,m\}}\{t_i\}$, each element of $\mathbf{a}^{(min,t)}$ reaches the minimal objective value of the objective space and will not change again.

2. If $A^{(min,t)}(t > \tau_{c_1})$ is not Pareto optimal in Y, there must exist $\mathbf{y}^*(\in Y^*)$ such that $\left(\mathbf{y}^* \prec \mathbf{a}^{(i,t)}\right) \wedge \left(y_i^* = a_i^{(i,t)} = \min_{\mathbf{y}\in Y}\{y_i\}\right)$ for some i. There must exist $t_{o_i}(> t)$ such that $\mathbf{y}^{(t_{o_i})} = \mathbf{y}^*$. Thus, $\mathbf{y}^{(t_{o_i})} \prec \mathbf{a}^{(i,t_{o_i}-1)}$, and then $\mathbf{y}^{(t_{o_i})} \prec A^{(t_{o_i}-1)}$. Thus, DOMINATES of Procedure 2 executes, and $\mathbf{a}^{(i,t_{o_i}-1)}$ is replaced by $\mathbf{y}^{(t_{o_i})}$. Once $t > \tau_{c_2} \triangleq \max_{i=1,\cdots,m}\{t_{o_i}\}$, $A^{(min,t)}$ reaches a Pareto optimal set as described in Eq.(9).

Once $A^{(min,t)}$ is Pareto optimal in Y and each member at least has a minimal value on one objective, there is not a vector $\mathbf{y}$ that either dominates $A^{(min,t)}$ or $y_i < a_i^{(i,t)}$. The condition in Line 1 of Procedure 2 cannot be satisfied. Neither DOMINATES nor RECEDES executes. Therefore, $A^{(min,t)}$ becomes stable. This completes the proof. ■

A direct consequence of Theorem 4 is that $\mathbf{a}^{(max,t)}$ converges, so does the crucial region. It is worth pointing out that $a_i^{(max,t)} > a_i^{(min,t)}$ for $1 \leq i \leq m$ because at least two different values in each objective are assumed to be on the Pareto front. The assumption about EVOLUTION is quite common in theoretical analyses of evolutionary algorithms [Knowles and Corne (2003); Rudolph and Agapie (2000)]. It is true whenever, for example, a mutation is applied to every bit in a binary string with some small probability, the standard method of generating a new solution in a random mutation hillclimber [Knowles and Corne (2003)]. Based on this weak assumption again, we give the main convergence result of our archiving strategy below.

Theorem 5 *If* EVOLUTION *gives every possible solution in the search space with a positive minimum probability, the archive sequences* $\left\{A^{(arc,t)}\right\}$ *and* $\left\{A^{(t)}\right\}$ *of ARA converges to bounded-sized E-Pareto sets of the whole objective space with probability one as* $t \longrightarrow +\infty$*, i.e.,*

- $A^{(arc,t)}, A^{(t)} \in P_E^*(Y)$;
- $2 \leq \left|A^{(arc,t)}\right| \leq \dfrac{\prod_{i=1}^{m}\left\lceil \frac{\frac{\pi}{2}}{e_i}+1\right\rceil}{\max_{i\in\{1,\cdots,m\}}\left\lceil \frac{\frac{\pi}{2}}{e_i}+1\right\rceil}$ *for any given* $\mathbf{e}$ *with* $0 < e_i < \frac{\pi}{4}$.
- $2 \leq \left|A^{(t)}\right| \leq m + \dfrac{\prod_{i=1}^{m}\left\lceil \frac{\frac{\pi}{2}}{e_i}+1\right\rceil}{\max_{i\in\{1,\cdots,m\}}\left\lceil \frac{\frac{\pi}{2}}{e_i}+1\right\rceil}$ *for any given* $\mathbf{e}$ *with* $0 < e_i < \frac{\pi}{4}$.

Proof: According to Theorem 4, $A^{(min,t)}$ converges to a Pareto optimal set when $t > \tau_{c_2}$. In addition, according to Lemma 1, the members of $A^{(t)} = \{A^{(min,t)}, A^{(arc,t)}\}$ don't dominate each other. Thus, we can prove directly the statement about $A^{(t)}$ based on the one about $A^{(arc,t)}$. We only prove the statement about $A^{(arc,t)}$ below.

According to Theorems 3 and 4, $A^{(arc,\tau)}$ is an E-Pareto set of $\left\{\bigcup_{t=\tau_{c_2}}^{\tau}\{\mathbf{y}^{(t)}\}\right\}\bigcup A^{(arc,\tau_{c_2})}$. Evolution generates any solution infinitely often and that the waiting time for the first occurrence as well as for the second, and so forth is finite with probability 1, so, for each solution $\mathbf{y}\in Y$, there exists $t_{\mathbf{y}}(\tau_{c_2}<t_{\mathbf{y}}<+\infty)$ such that $\mathbf{y}^{(t_{\mathbf{y}})}=\mathbf{y}$. Then $A^{(arc,t_{\mathbf{y}})}$ must E-dominate $\mathbf{y}$. Since $|Y|$ is finite, $\tau_{c_3}\triangleq\max_{\mathbf{y}\in Y}\{t_{\mathbf{y}}\}<+\infty$. Thus, $A^{(arc,t)}$ is an E-Pareto set of Y as $t>\tau_{c3}$.

According to the proof of Theorem 3, for each solution $\mathbf{y}$ E-dominated by $A^{(arc)}$, there must exist $\mathbf{a}\in A^{(arc)}$ such that $\textsc{Rect}(\mathbf{a},A^{(min)})\prec\textsc{Rect}(\mathbf{y},A^{(min)})$ or $\textsc{Rect}(\mathbf{a},A^{(min)})=\textsc{Rect}(\mathbf{y},A^{(min)})$. When $t>\tau_{c_2}$, $A^{(min,t)}$ becomes stable and $a_i^{(min,t)}<a_i^{(max,t)}$ for $i=1,\cdots,m$. According to Function 4, the "coordinates" of these rectangles occupied by the members of $A^{(min,t)}$ must have two different values: 1 and $\left\lceil\frac{\frac{\pi}{2}}{e_i}\right\rceil$ for each objective. When $t>\tau_{c3}$, $A^{(arc,t)}$ E-dominates $A^{(min,t)}$. That means the "coordinate" of the rectangle occupied by every member of $A^{(min,t)}$ must be dominated or equal to the "coordinate" of the rectangle occupied by one member of $A^{(arc,t)}$. Then $A^{(arc,t)}$ has at least two members. Otherwise, if $A^{(arc,t)}$ has only one member $\mathbf{a}$, $\textsc{Rect}(\mathbf{a},A^{(min)})=[1,1,\cdots,1]^T$. This leads to $\mathbf{a}=\mathbf{a}^{(min,t)}$. Then $\mathbf{a}\prec A^{(min,t)}$, which contradicts Theorem 2. So, $\left|A^{(arc,t)}\right|\geq 2$ as $t\longrightarrow+\infty$.

As we can observe in Function 4, the i^{th} dimension (objective) is divided into $\left\lceil\frac{\frac{\pi}{2}}{e_i}+1\right\rceil$ segments. The objective space is divided into $\prod_{i=1}^{m}\left\lceil\frac{\frac{\pi}{2}}{e_i}+1\right\rceil$ rectangles in total. From each rectangle, at most one solution can be in $A^{(arc,t)}$ at the same time. Now consider the equivalence classes of rectangles where, without loss of generality, the rectangles in each class have the same "coordinates" in all but one dimension. There are at most $\max_{i=1,\cdots,m}\left\lceil\frac{\frac{\pi}{2}}{e_i}+1\right\rceil$ different rectangles in each class constituting a chain of dominating rectangles. Hence, only one solution from each of these classes can be a member of $A^{(arc,t)}$ at the same time. This completes the proof. ■

This theorem states that the archive of ARA can finally E-dominate the whole Pareto front. It also states that the archive size is bounded, given an appropriate vector $\mathbf{e}$. In addition, there are at least two different Pareto optimal solutions in the archive. This point is different from the ϵ-Pareto set, which sometimes retains only one solution [Laumanns et al. (2002)].

6 Conclusion and discussion

In this paper, we have introduced the E-Pareto set as a novel concept for evolutionary multiobjective optimization. It is theoretically attractive as it helps to construct algorithms with the desired convergence and distribution properties, and it generalizes the dominance concepts in the Multi-Objective Evolutionary Algorithms (MOEAs) literature. Moreover, it is practically important as it works with Pareto fronts of bounded size without prior knowledge about optimization problems.

We have constructed the Adaptive Rectangle Archiving (ARA) archiving strategy that can be used in MOEAs. It can maintain the minimal objective values and well-distributed Pareto optimal solutions among the solutions generated so far in the sense of E-dominance (Theorems 2 and 3).

Our archiving strategy, with an appropriate assumption on the solution generation procedure, can retain the minimal objective values and a well-distributed approximation of the whole Pareto front with probability 1 in the sense of E-dominance (Theorems 4 and 5).

When the knowledge about the distribution of the Pareto front is not available, an end user can easily set an appropriate vector $\mathbf{e}$, and then ARA can provide a representative, well-distributed Pareto optimal set. So, our archiving strategy complements the existing ones.

Our archiving technique is based on the arctangent transferring function, but the underlying principle is easily applicable to other transferring functions. For example, we may use a new transferring function and design a different Rect function to treat different solution regions more uniformly. We will study this topic and its applications on benchmark and real-life problems.

Our theoretical analyses have assumed a bounded-sized objective space, but the E-Pareto set and the E-approximation Pareto set concepts are also applicable to more complicated real-life problems, such as multiobjective data mining problems [Jin et al. (2003)]. Another interesting research direction is, following the adaptive grid archiving algorithm, to design a new archiving algorithm which always maintains a fixed-sized, well-distributed archive. These topics are subject to our future research interest.

References

Coello Coello, C. A., Toscano Pulido, G., and Salazar Lechuga, M. (2004). "Handling Multiple Objectives With Particle Swarm Optimization." *IEEE Transactions on Evolutionary Computation*, 8(3), 256–279.

Coello Coello, C. A., Van Veldhuizen, D. A., and Lamont, G. B. (2002). *Evolutionary Algorithms for Solving Multi-Objective Problems.* Kluwer Academic Publishers.

Corne, D. and Knowles, J. (2003). "Some Multiobjective Optimizers are Better than Others." *Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003)*, Vol. 4. 2506–2512.

Corne, D. W., Jerram, N. R., Knowles, J. D., and Oates, M. J. (2001). "PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization." *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2001)*, L. Spector, E. D. Goodman, A. Wu, W. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon, and E. Burke, eds. 283–290.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons.

Deb, K., Mohan, M., and Mishra, S. (2003). "Towards a Quick Computation of Well-Spread Pareto-Optimal Solutions." *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, eds. 222–236.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA–II." *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.

Feller, W. (1976). *An Introduction to Probability Theory and Practice*. Wiley, 3rd edition.

Fieldsend, J. E., Everson, R. M., and Singh, S. (2003). "Using Unconstrained Elite Archives for Multiobjective Optimization." *IEEE Transactions on Evolutionary Computation*, 7(3), 305–323.

Fonseca, C. M. and Fleming, P. J. (1993). "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization." *Proceedings of the Fifth International Conference on Genetic Algorithms*, S. Forrest, ed., University of Illinois at Urbana-Champaign. 416–423.

Hanne, T. (1999). "On the convergence of multiobjective evolutionary algorithms." *European Journal of Operational Research*, 117(3), 553–564.

Horn, J., Nafpliotis, N., and Goldberg, D. E. (1994). "A Niched Pareto Genetic Algorithm for Multiobjective Optimization." *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, Vol. 1. 82–87.

Jin, H.-D., Shum, W., Leung, K.-S., and Wong, M.-L. (2004). "Expanding self-organizing map for data visualization and cluster analysis." *Information Sciences*, 163, 157–173.

Jin, H.-D., Wong, M.-L., and Leung, K.-S. (2003). "Scalable model-based clustering by working on data summaries." *Proceedings of Third IEEE International Conference on Data Mining (ICDM 2003)*, Melbourne, Florida, USA. 91–98.

Jin, Y. and Sendhoff, B. (2003). "Trade-Off between Performance and Robustness: An Evolutionary Multiobjective Approach." *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, eds. 237–251.

Knowles, J. and Corne, D. (2003). "Properties of an Adaptive Archiving Algorithm for Storing Nondominated Vectors." *IEEE Transactions on Evolutionary Computation*, 7(2), 100–116.

Knowles, J. and Corne, D. (2004). "Bounded Pareto Archiving: Theory and Practice." *Metaheuristics for Multiobjective Optimisation*, X. Gandibleux, M. Sevaux, K. Sörensen, and V. T'kindt, eds. 39–64.

Knowles, J. D. and Corne, D. W. (2000). "Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy." *Evolutionary Computation*, 8(2), 149–172.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). "Combining Convergence and Diversity in Evolutionary Multi-objective Optimization." *Evolutionary Computation*, 10(3), 263–282.

Reuter, H. (1990). "An approximation method for the efficiency set of multiobjective programming problems." *Optimization*, 21, 905–911.

Rudolph, G. (1998). "Evolutionary Search for Minimal Elements in Partially Ordered Finite Sets." *Evolutionary Programming VII, Proceedings of the 7th Annual Conference on Evolutionary Programming*, V. Porto, N. Saravanan, D. Waagen, and A. Eiben, eds. 345–353.

Rudolph, G. and Agapie, A. (2000). "Convergence Properties of Some Multi-Objective Evolutionary Algorithms." *Proceedings of the 2000 Conference on Evolutionary Computation*, Vol. 2. 1010–1016.

Srinivas, N. and Deb, K. (1994). "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms." *Evolutionary Computation*, 2(3), 221–248.

Tan, K., Lee, T., and Khor, E. (2001). "Evolutionary Algorithms with Dynamic Population Size and Local Exploration for Multiobjective Optimization." *IEEE Transactions on Evolutionary Computation*, 5(6), 565–588.

Yen, G. G. and Lu, H. (2003). "Dynamic Multiobjective Evolutionary Algorithm: Adaptive Cell-Based Rank and Density Estimation." *IEEE Transactions on Evolutionary Computation*, 7(3), 253–274.

Zitzler, E., Laumanns, M., and Thiele, L. (2002). "SPEA2: Improving the Strength Pareto Evolutionary Algorithm." *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, K. Giannakoglou, D. Tsahalis, J. Periaux, P. Papailou, and T. Fogarty, eds. 95–100.

Zitzler, E. and Thiele, L. (1999). "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach." *IEEE Transactions on Evolutionary Computation*, 3(4), 257–271.