Adaptive Diversity Maintenance and Convergence Guarantee in Multiobjective Evolutionary Algorithms

Huidong Jin
Department of Information Systems
Lingnan University
Tuen Mun, N.T., Hong Kong
hdjin@ieee.org

Man-Leung Wong
Department of Information Systems
Lingnan University
Tuen Mun, N.T., Hong Kong
mlwong@ln.edu.hk

Abstract- The issue of obtaining a well-converged and well-distributed set of Pareto optimal solutions efficiently and automatically is crucial in multi-objective evolutionary algorithms (MOEAs). Many studies have proposed different evolutionary algorithms that can progress towards Pareto optimal sets with a wide-spread distribution of solutions. However, most mathematically convergent MOEAs desire certain prior knowledge about the objective space in order to efficiently maintain widespread solutions. In this paper, we propose, based on our novel 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*. ARA complements the existing archiving techniques, and is useful to both researchers and practitioners.

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. Multi-Objective Evolutionary Algorithms (MOEAs) have been gaining increasing attention among researchers and practitioners mainly because they can be suitably applied to find multiple Pareto optimal solutions in a single run [3]. This fact enables a user to have a less-subjective search in the first phase of finding a set of well-distributed solutions. Because of inherent cooperation in an evolutionary search procedure, MOEAs are computationally promising for simultaneous discovery of multiple trade-off solutions. The features have attracted numerous researchers to develop different MOEAs [1] — from MOGA [8], NPGA [10], and NSGA [16] with skillful fitness assignment and nondominated sorting; to SPEA [20], PESA–II [2], NASA–II [5], SPEA2 [19], IMOEA [17], and DMOEA [18] with elitism, diversity estimation and maintenance; to PAES [12] (based on AGA [11]) and ϵ-MOEA [4] (based on ϵ-dominance [13]) 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 [2, 5, 10, 18, 19, 20], but few of them have convergence guarantee. Some early theoretical work has pointed out some approaches to enable MOEAs to converge to Pareto front [9, 15], but with little consideration of the distribution of the Pareto optimal solutions obtained [13]. Several recent studies have made a big pace to generate diversified and Pareto optimal solutions [4, 11, 13]. The archiving techniques in [13] and [4] desire the distribution knowledge about the Pareto front beforehand. If the parameters are not set appropriately, in some extreme cases, only a solution is archived because it ϵ-dominates all the others [11]. The Adaptive Grid Archiving (AGA) strategy has been proved to converge to a Pareto optimal set of bounded size under certain condition [11]. Unfortunately, this condition is not easily satisfied, and the solution oscillation problem has happened in practical applications [11] or been demonstrated empirically [7].

One basic idea of these efficient and successful diversity preserving mechanisms is to partition the whole objective space into mutually excluded regions, and then consider the Pareto optimality and diversity locally in these regions [2, 4, 12, 11, 13, 17]. Each region is of limited volume while the objective space is unknown in advance. This conflict makes it difficult to explore the whole solution space, and results in the unexpected difficulty in the recent work [11, 13]. In this work, we introduce the concept of open (hyper-)rectangles and apply the rectangles in the space partitioning, such that even infinite search space can be enveloped by a bounded number of rectangles. We introduce an extended Pareto dominance (E-dominance) to achieve this idea. In addition, our search space partitioning is adjusted adaptively according to the solutions found so far. The archive retains more Pareto optimal solutions in the crucial region, and some accidental solutions within the open rectangles. Therefore, our proposed Adaptive Rectangle Archiving (ARA) technique can explore the whole objective space and maintain some representative Pareto optimal solutions automatically without any prior knowledge.

In the rest of the paper, we first give a template of the MOEA 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

We focus on, without loss of generality, minimization multiobjective problems [3] in this work. For a multiobjective function Γ from $X(\subseteq \Re^d)$ to a finite set $Y(\subset \Re^m, m \geq 2)$, a decision vector $\mathbf{x}^{(1)}$ *dominates* another one $\mathbf{x}^{(2)}$, if and only if their objective vectors $\mathbf{y}^{(1)} = \Gamma(\mathbf{x}^{(1)}) = [y_1^{(1)}, y_2^{(1)}, \cdots, y_m^{(1)}]^T$ and $\mathbf{y}^{(1)} = \Gamma(\mathbf{x}^{(2)})$ satisfy

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

It is denoted as $\mathbf{x}^{(1)} \prec \mathbf{x}^{(2)}$. For convenience, we also denote it as $\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}$. Furthermore, we denote $\neg\left(\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}\right)$ as $\mathbf{y}^{(1)} \not\prec \mathbf{y}^{(2)}$. $\mathbf{y}^{(1)}$ is said 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 $\mathbf{y}^{(1)} \sim \mathbf{y}^{(2)}$. Therefore, $\mathbf{y}^{(1)} \not\prec \mathbf{y}^{(2)}$ 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 *dominates* and *nondominated* relations can be defined between an objective vector $\mathbf{y}$ and a set $A(\subseteq Y)$:

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

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

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

$$\mathbf{y} \not\prec A \iff \forall \mathbf{a} \in A, \quad \mathbf{y} \not\prec \mathbf{a}. \tag{5}$$

Given the set of vectors Y, its *Pareto front* Y^* contains all vectors $\mathbf{y}^* \in Y$ that are not dominated by any 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 in the Pareto front Y^*, which holds for almost all multiobjective problems.

For many multiobjective optimization problems, the unique Pareto front Y^* is of substantial size. Thus, the determination of Y^* is computationally prohibitive. The whole Pareto front Y^* is usually difficult to maintain. Furthermore, it is questionable to be regarded as an optimization solution [7, 13]. The value of presenting such a large set of solutions to a decision maker is doubtful in the context of decision support, instead one should provide him with a set of representative Pareto optimal solutions. 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 [11].

In order to facilitate our analysis on archiving strategies, we separate the evolutionary procedure and the archiving procedure as done in [11, 13]. Procedure 1 gives an abstract description of a MOEA with archiving. The integer

Procedure 1 ***MOEA with Archiving***

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

t denotes the iteration count, the m-dimensional objective vector $\mathbf{y}$ 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. The function EVOLUTION represents an evolutionary algorithm, where the evolutionary operator is associated with variation (recombination, mutation, and selection). It can generate a population of points, possibly using the contents of the old archive $A^{(t-1)}$. However, for convenience, it only outputs a new solution in each iteration t. ARCHIVE gets the new solution $\mathbf{y}^{(t)}$ and the old archive $A^{(t-1)}$ and determines the updated archive $A^{(t)}$. The archive is usually 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 function ARCHIVE, i.e., how to appropriately update the archive. For each $\mathbf{y}$, its additional information about the corresponding decision values could be associated to the archive, but will be of no concern in this paper. According to the requirements of MOEAs, an ideal archiving strategy should maintain solutions having the following properties:

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

Well distributed: Solutions are uniformly distributed on the whole Pareto front;

Computationally efficient: The time and memory complexity should be low;

Little prior knowledge: Little knowledge about the multiobjective problem is required beforehand.

This last property may facilitate users greatly, since most of time, we have to make decisions on some conflicting problems with little prior knowledge. As mentioned before, some information of the objective space has been stored in the archive during running. Thus, we can adjust our archive adaptively. We shall give such a technique in Section 4, after the discussion of the existing approaches in the next section.

3 MOEAs and their Limitations

We briefly discuss a number of archiving or elitism strategies in the literature of MOEAs.

Early theoretical work of MOEAs mainly concentrates on convergence. Hanne [9] 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. There is no assumption on the distribution of solutions, and arbitrary regions may become unreachable with the (negative) efficiency preservation [13]. Rudolph and Agapie [15], using stochastic process techniques, developed several sophisticated selection operators to preclude 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 [10]), clustering (SPEA [20], SPEA2 [19]), nearest neighbor distance (NAGA-II [5]), crowding count (PAES [12], PESA-II [2], DMOEA [18]), or some preselection operators [3]. Most of them are more or less successful, but they cannot ensure convergence to Pareto optimal sets.

Recently, Laumanns *et al.* [13] proposed several archiving strategies that guarantee to progress towards the Pareto front and covers the whole range of nondominated solutions. The algorithms maintain a bounded archive of nondominated 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, must either be set manually or be determined adaptively. 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. 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 solutions [11].

More recently, Knowles and Corne [11] analyzed a metric-based archiving and an adaptive grid archiving one. The metric-based strategy 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 more than a few objectives. The adaptive grid archiving strategy implemented in PAES [11] provably 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 condition is not true for many optimization problems with more than two objectives. Thus, the oscillation problem of the archive has happened in practical applications [11] or been demonstrated empirically [7].

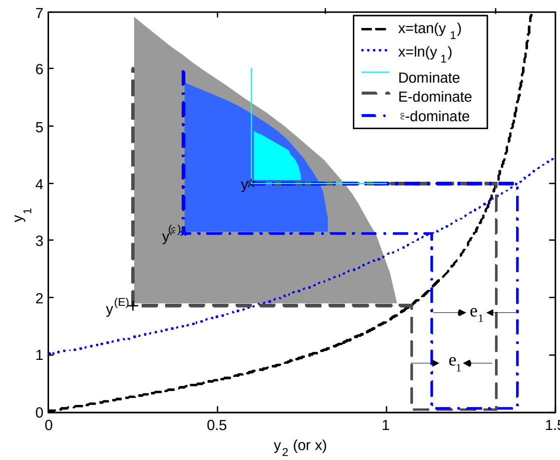

Figure 1: Illustration of E-dominance, ϵ-dominance, and Pareto dominance. The regions dominated by $\mathbf{y}$ under three different dominance relations are illustrated by three shadows respectively. The calculation of vectors $\mathbf{y}^{(E)}$ and $\mathbf{y}^{(\epsilon)}$ is illustrated in the bottom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

In order to diversify the solutions, the density estimation or diversity preservation has been locally made in some boxes for computational efficiency. However, the objective space is unknown in advance, even infinite. Thus, it is sometimes impractical to use boxes to envelop the space appropriately. This issue results in the oscillation of AGA [11] and probably poor representation of the Pareto front in [13], though they may generate widespread solutions.

4 Adaptive Rectangle Archiving Strategy

In this section, we present an Adaptive Rectangle Archiving (ARA) algorithm that address the problem of previously unknown, even infinite, Pareto fronts. In this archiving strategy, we use a self-adaption mechanism to preserve diversity according to the archived information about the objective space. In the *crucial* region, a solution is allowed to preserve in a narrow (hyper-)rectangle, and thus more Pareto solutions are archived. In the unknown, even infinite, regions, some *open* rectangles are used to envelop them. These open rectangles even may envelop infinite objective values. Within these open rectangles, some Pareto 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.

4.1 Extended Pareto Dominance

Since we need to use an archive of points to approximately dominate the whole objective space, one intuitive solution is to permit some tolerance on dominance. To achieve it, we extend the Pareto dominance as follows.

Definition 1 (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 trans-*

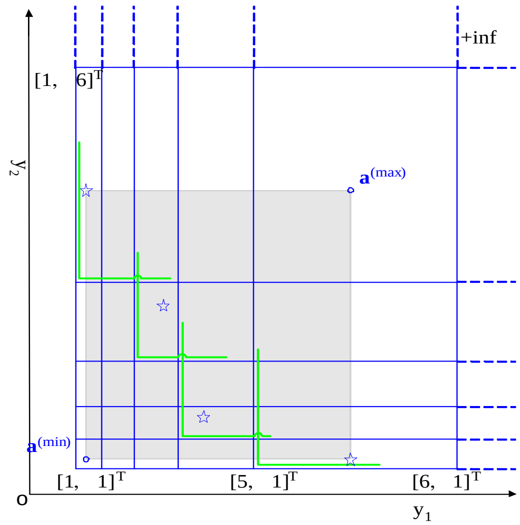

Figure 2: 2-D adaptive rectangle partitioning. The dashed line segments indicate open rectangles. The gray rectangle indicates the crucial region indicated by $\mathbf{a}^{(min)}$ and $\mathbf{a}^{(max)}$. The gray line segments indicate the region E-dominated by a solution, denoted by a pentagram.

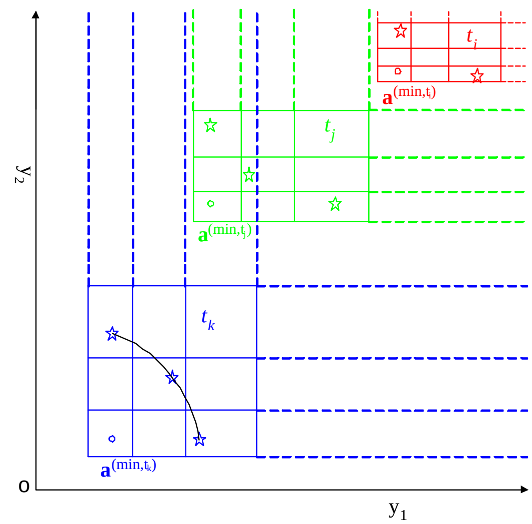

Figure 3: How the adaptive rectangles change their location and shapes in the objective space as the vectors in the archive $A^{(t)}$ change through iterations $t_i < t_j < t_k$. The bold curve indicates the Pareto front, pentagrams are archived solutions, and small circles indicate $\mathbf{a}^{(min,t)}$.

ferring function, FUN*, and a constant vector* $\mathbf{e}(>0)$*, if and only if*

$$\text{FUN}(y_i^{(1)}) - e_i \leq \text{FUN}(y_i^{(2)}), \quad \forall\, i \in \{1, \ldots, m\}. \qquad (6)$$

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 $\mathbf{y}^{(1)} \prec y^{(2)}$, then $\mathbf{y}^{(1)} \preceq_E y^{(2)}$. Furthermore, it is obvious that the E-dominance relation is transitive.

E-dominance generalizes several dominance relations. For example, it becomes ϵ-dominance [13] as $\text{FUN}(y_i) = \ln(y_i)$ and $e_i = \ln(1+\epsilon)$, the additive ϵ-dominance [13, 14] as $\text{FUN}(y_i) = y_i$ and $e_i = \epsilon$, and the Pareto dominance as $\text{FUN}(y_i) = y_i$ and $e_i = 0$.

In order to envelop unknown, possible infinite, objective values, we may employ a nonlinear transferring function, e.g., $\text{FUN}(y_i) = \tan(y_i * scale_i)$. Thus, the infinite points are transferred to $\frac{\pi}{2}$, and may be E-dominated by a bounded value, say, $\frac{\arctan(\frac{\pi}{2} - e_i)}{scale_i}$. The $scale_i$ is specified adaptively according to the solutions found so far. E-dominance is with a tangent function hereafter unless otherwise specified. 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 2 (E-approximate Pareto Set) *Let* $Y \subset \Re^m$ *be a set of vectors,* FUN *a monotonically increasing function, and* $\mathbf{e}$ *a positive 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.,*

$$\forall \mathbf{y} \in Y : \exists \mathbf{a} \in Y_E \text{ such that } \mathbf{a} \preceq_E \mathbf{y}. \qquad (7)$$

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

Definition 3 (E-Pareto Set) *Let* $Y \subseteq \Re^m$ *be a set of vectors, and* $\mathbf{e}$ *a positive. 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.,* $Y_E^* \in P(Y_E)$*, and*

2. Y_E^* *only contains Pareto optimal points of* Y*, i.e.,* $Y_E^* \subseteq Y^*$

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

Obviously, the Pareto front is the biggest E-Pareto set. 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 but also is of smaller size. Of course, an E-Pareto set is more attractive as it consists of Pareto optimal solutions only.

4.2 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 convenience, we partition the archive in ARA into two parts: $A = \{A^{(min)}, A^{(arc)}\}$. 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: $A^{(min)} = [\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \cdots, \mathbf{a}^{(m)}]$. Each element, $\mathbf{a}^{(i)}$, is initialized to be infinite, and stores the solution found so far that has the minimal value at the i^{th} objective. We have $a_i^{(i)} =$

Procedure 2 $\mathrm{ARA}(\mathbf{y}, A)$

1. ***if*** $(\mathbf{y} \prec A^{(min)} \vee \mathbf{a}^{(min)} \not\prec \mathbf{y})$ ***then***
2.   ***for all*** $i \in \{1, \cdots, m\}$ ***do***
3.   ***if*** $(a_i^{(i)} > y_i)$ ***then***
4. $\mathbf{a}^{(i)} := \mathbf{y}$;   */* **Recedes** */*
5.   ***else if*** $(\mathbf{y} \prec \mathbf{a}^{(i)})$
6. $\mathbf{a}^{(i)} := \mathbf{y}$;   */* **Dominates** */*
7.   ***end if***
8.   ***end do***
9. $A^{(arc_2)} := \emptyset$;   */* **Re-forms** $A^{(arc)}$ */*
10.   ***for all*** $\mathbf{a} \in A^{(arc)}$ *such that* $A^{(min)} \not\prec \mathbf{a}$ ***do***
11. $\textsc{InsertInRect}(\mathbf{a}, A^{(arc_2)}, A^{(min)})$;
12.   ***end do***
13. $A^{(arc)} := A^{(arc_2)}$;
14. ***else if*** $(A^{(min)} \not\prec \mathbf{y})$ */* **Updates** $A^{(arc)}$ */*
15. $\textsc{InsertInRect}(\mathbf{y}, A^{(arc)}, A^{(min)})$;
16. ***end if***
17. $A := \{A^{(min)}, A^{(arc)}\}$;

$\min\limits_{\mathbf{a}\in A^{(min)}} \{a_i\}$. Furthermore, we introduce two vectors associated with $A^{(min)}$ to describe the crucial region: $\mathbf{a}^{(min)}$ with $a_i^{(min)} = \min\limits_{\mathbf{a}\in A^{(min)}} \{a_i\}$ and $\mathbf{a}^{(max)}$ with $\mathbf{a}_i^{(max)} = \max\limits_{\mathbf{a}\in A^{(min)}} \{a_i\}$. 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 in the crucial region in 2-D case. The gray rectangle in Figure 2 indicates the crucial region, and envelops all four Pareto solutions, indicated by pentagrams.

The pseudo code of our archiving strategy, ARA, is given in Procedure 2, which is illustrated in Figure 3. At each iteration, the algorithm first checks whether the crucial region should be updated. If a new objective value is smaller than the archived one, **Recedes** replaces the old vector with the new one; 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 $\mathbf{a}^{(max)}$. If the crucial region is updated, the solutions in $A^{(arc)}$ have to be archived again (**Re-forms** $A^{(arc)}$). Thus, the minimal objective value is certainly archived, and the solutions in $A^{(arc)}$ are chosen based on the current $A^{(min)}$.

It is possible that the condition in Line 1 of Procedure

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***

    $A^{(arc)} := A^{(arc)} \bigcup \mathbf{y} \backslash D$;   */* **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***

    $A^{(arc)} := A^{(arc)} \bigcup \{\mathbf{y}\} \backslash \{\mathbf{a}\}$ */* **IntraRectDom** */*

else if $\forall \mathbf{a} \in A^{(arc)} : \textsc{Rect}(\mathbf{a}, A^{(min)}) \sim \textsc{Rect}(\mathbf{y}, A^{(min)})$

    $A^{(arc)} := A^{(arc)} \bigcup \{\mathbf{y}\}$ */* **Occupies** a rectangle */*

else

    $A^{(arc)} := A^{(arc)}$;   */* **SteadyState** */*

end if

2 holds, but neither **Recedes** nor **Dominates** is executed. It occurs only when $(\mathbf{y} = \mathbf{a}^{(min)}) \wedge (\forall i, \mathbf{a}^{(i)} = \mathbf{y})$. This rarely happens since multiobjective problems normally have more than one solution.

If the new vector $\mathbf{y}$ neither has smaller objective value nor dominates any vector in $A^{(min)}$, then it is processed by INSERTINRECT, as described 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 (hyper-)rectangles (see Function 4), where each vector uniquely belongs to one rectangle. Using the proposed E-dominance relation on these rectangles, the algorithm always maintains a set of nondominated rectangles (**InterRectDom** and **Occupies**), thus guaranteeing the E-approximate Pareto property. On the fine level, at most one solution is kept in each rectangle. Within a rectangle, each representative vector can only be replaced by a dominating one (**IntraRectDom**), which ensures convergence.

Now let us see how the function RECT in Function 4 partitions the crucial region finely while envelops the unknown regions with open rectangles based on E-dominance. Since it is difficult to automatically detect the maximal objective values in the Pareto front [11], we simply view it as infinite. As shown in Lines 2∼4 in RECT, $a_i^{(min)}$ and $+\infty$ are mapped into 1 and $\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor$, respectively. So, the rectangles are open if its coordinates contain $\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor$, e.g., [1,6] and [6,1] in Figure 2. The scale calculated in Line 2 reflects the distance between $a_i^{(max)}$ and $a_i^{(min)}$. The farther away $a_i^{(max)}$ is from $a_i^{(min)}$, the larger the scale value is. Furthermore, this scale, together with the value 1.5 in Line 4, enables $a_i^{(max)}$ to be mapped to $\left(\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor - 1\right)$, which is next to the coordinate corresponding to $+\infty$.

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

1. ***for all*** $i \in \{1, \cdots, m\}$ ***do***
2. $scale_i = \frac{\arctan\left(\frac{\pi}{2} - e_i\right)}{a_i^{(max)} - a_i^{(min)}}$;
3. $\alpha_i := \tan\left(\left(y_i - a_i^{(min)}\right) * scale_i\right)$;
4. $r_i := \left\lfloor \frac{\alpha_i}{e_i} + 1.5 \right\rfloor$;
5. ***end do***
6. ***output****: return* $\mathbf{r} = [r_1, \cdots, r_m]^T$.

Therefore, when $e_i < \frac{\pi}{4}$, $a_i^{(min)}$, $a_i^{(max)}$, and $+\infty$ are mapped onto different rectangles. Furthermore, there are $\left(\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor - 2\right)$ rectangles between $a_i^{(min)}$ and $a_i^{(max)}$. Less e_i is, more finely the crucial region is divided. An example with $e_i = \frac{\pi}{10}$ is illustrated in Figure 2. The unknown region is enveloped by some open rectangles, as indicated by dashed line segments. Clearly, the crucial region is finely divided, and a well-distributed Pareto optimal solutions are archived in $A^{(arc)}$, as indicated by pentagrams in Figure 2.

5 Convergence analysis

We now give some theorems to show that our archive converges to Pareto optimal sets, and preserves diversity of solution at the same time. We first give theoretical analysis 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.

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

Proof: We need to prove two cases: *Case 1.* the minimal objective values generated-so-far will enter the archive; *Case 2.* the objective vectors with the minimal objective values do not lose.

Case 1. To prove this point, we only need to prove $a_i^{(min,t)} = y_i^{(t)}$ when a smaller objective value is generated for some $i \in \{1, \ldots, m\}$ and $t < \tau$, i.e, when $y_i^{(t)} < a_i^{(min,t-1)}$. At 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 $\left(\mathbf{y}^{(t)} \sim \mathbf{a}^{(i,t-1)}\right)$. For the former, $\mathbf{y}^{(t)} \prec A^{(min,t-1)}$ and the rule, **Dominates**, executes. For the latter, if $\mathbf{y}^{(t)} \prec \mathbf{a}^{(min,t-1)}$, then $\mathbf{y}^{(t)} \prec \mathbf{a}^{(min,t-1)} \prec \mathbf{a}^{(i,t-1)}$. It contradicts $\mathbf{y}^{(t)} \sim \mathbf{a}^{(i,t-1)}$. So, $\mathbf{y}^{(t)} \sim \mathbf{a}^{(min,t-1)}$, and **Recedes** executes. For both situations, $\mathbf{a}^{(i,t)} = \mathbf{y}^{(t)}$. Thus, $a_i^{(min,t)} = y_i^{(t)}$.

Case 2. We only have to prove $a_i^{(min,t)} = a_i^{(min,t-1)}$ if $a_i^{(min,t-1)} < y_i^{(t)}$. Since $a_i^{(min,t-1)} < y_i^{(t)}$, we know $a_i^{(i,t-1)} < y_i^{(t)}$ and $\mathbf{y}^{(t)} \not\prec \mathbf{a}^{(i,t)}$. So, both **Dominates** and **Recedes** do not execute, and $\mathbf{a}^{(i,t)} = \mathbf{a}^{(i,t-1)}$. ■

As described in Procedures 2 and 3, one solution dominated by the archived solutions is impossible to enter $A^{(min,t)}$ or $A^{(arc,t)}$, respectively. Furthermore, as required in Lines 10 and 14, if a solution dominates $A^{(min,t)}$, it cannot enter $A^{(arc,t)}$. On the other hand, since solutions in $A^{(min)}$ must have one minimal objective value generated so far, the solution in $A^{(arc)}$ also cannot dominate $A^{(min)}$. Thus, we have the following nondominated relations among the solutions in the archive.

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 1, $A^{(arc,t)}$ collects the Pareto optimal solutions iteratively, as stated in the following theorem.

Theorem 2 *The archive* $A^{(arc,\tau)} (\neq \emptyset)$ *is an E-Pareto set of* $Y^{(\tau_0,\tau)} = A^{(arc,\tau_0)} \bigcup \left\{ \bigcup_{t=\tau_0}^{\tau} \{\mathbf{y}^{(t)}\} \right\}$ *if* $A^{(min,\tau_0)} = A^{(min,\tau)}$.

Due to the space limitation, we only sketch the proof. When $A^{(min,t)}$ does not change, the generated solutions are all inputted to INSERTINRECT. Any solution $\mathbf{y}$ must be E-dominated by $A^{(arc,t)}$, or enters the archive. Once being archived, it will not be deleted until it is replaced by a new one that E-dominated it. The solutions E-dominated by $\mathbf{y}$ are transitively E-dominated by the new one. So, $A^{(arc,t)}$ still E-dominates these solutions. Similarly, if an archived solution $\mathbf{a}$ is not Pareto optimal among the solutions generated so far, it is replaced by a Pareto optimal one by executing **InterRectDom** or **IntraRectDom**. So, $A^{(arc,t)}$ must be an E-Pareto set of the solutions inputted into INSERTINRECT when the crucial region is unchanged.

Theorems 1 and 2 state that, in ARA, the archive retains the minimal objective values and the E-Pareto solutions of the objective vectors inputted so far. The archive retains the best-so-far solutions, and this feature allows a MOEA using ARA to stop anytime. Using these features of ARA, we give the convergence results, based on an assumption of EVOLUTION, that the archiving algorithm may reach the crucial region of the Pareto front, and then E-dominates it.

Theorem 3 *If the function* EVOLUTION *gives every possible solution in the search space with a positive minimum probability, then*

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

$$\{\mathbf{a} \in Y \,|\, \exists i, a_i = \min_{\mathbf{y} \in Y}\{y_i\} \wedge (\mathbf{a} \prec \mathbf{y}, \forall \mathbf{y} \in Y \wedge y_i = a_i)\} \tag{8}$$

with probability one as $t \to \infty$.

Proof: 1. Since EVOLUTION can generate every possible solution with a positive minimum probability, according to the Borel-Cantelli Lemma (see e.g., [6, p. 201]), it is guaranteed that arbitrary 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, there exists $t_i < +\infty$ such that $y_i^{(t_i)} = \min_{\mathbf{y}\in Y}\{y_i\}$ for any i. According to Theorem 1, $a_i^{(min,t)} = \min_{\mathbf{y}\in Y}\{y_i\}$ for all $t > t_i$. Therefore, when $t > \tau_{c_1} \triangleq \max_{i=1,\dots,m}\{t_i\}$, each element of $\mathbf{a}^{(min,t)}$ reaches the minimal objective value and will not change.

2. If $A^{(min,t)}(t > \tau_{c_1})$ is not Pareto optimal in Y, there must exist $\mathbf{y}^*(\in Y^*)$ such that $\mathbf{y}^* \prec \mathbf{a}^{(i,t)} \wedge y_i^* = \min_{\mathbf{y}\in Y}\{y_i\}$ for some i. There exists $t_{i_2}(> t)$ such that $\mathbf{y}^{(t_{i_2})} = \mathbf{y}^*$. $\mathbf{y}^{(t_{i_2})} \prec \mathbf{a}^{(i,t_{i_2}-1)}$, then $\mathbf{y}^{(t_{i_2})} \prec A^{(t_{i_2}-1)}$. Thus, **Dominates** executes, and $\mathbf{a}^{(i,t_{i_2}-1)}$ is replaced by $\mathbf{y}^{(t_{i_2})}$. Once $t > \tau_{c_2} \triangleq \max_{i=1,\dots,m}\{t_{i_2}\}$, $A^{(min,t)}$ reaches a Pareto optimal set as described in Eq.(8).

Once $A^{(min,t)}$, as in Eq.(8), is Pareto optimal in Y and each member at least has a minimal objective value, 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. ■

The direct result of Theorem 3 is that $\mathbf{a}^{(max,t)}$ converges, so does the crucial region. The assumption about EVOLUTION is quite common in theoretical analysis of evolutionary algorithms [11, 15]. 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 point in a random mutation hillclimber [11]. Based on this weak assumption, we give the main convergence result of our archiving strategy below.

Theorem 4 *If* EVOLUTION *gives every possible solution in the search space with a positive minimum probability, the archive sequence* $\{A^{(arc,t)}\}$ *of ARA converges to a well-distributed E-Pareto set of the whole objective space with bounded size with probability one as* $t \longrightarrow +\infty$*, i.e.,*

- $A^{(arc,t)} \in P(Y_E^*)$;
- $2 \leq |A^{(arc,t)}| \leq \dfrac{\prod_{i=1}^{m}\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor}{\max_{i\in\{1,\dots,m\}}\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor}$ *for any given* $\mathbf{e}$ *with* $0 < e_i < \frac{\pi}{4}$.

Proof: According to Theorem 3, $A^{(min,t)}$ converges to a Pareto optimal set when $t > \tau_{c_2}$. Then, according to Theorem 2, $A^{(arc,\tau)}$ 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 $A^{(arc,t_{\mathbf{y}}+1)}$ 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 $t > \tau_{c_3}$.

Let us consider $t > \tau_{c_2}$ (Theorem 3). The rectangle envelops an archived vector in $A^{(min,t)}$ must envelop a member of $A^{(arc,t)}$ as t increases (Otherwise, the member of $A^{(min,t)}$ enters $A^{(arc,t)}$ and occupy the rectangle when t increases.). For each objective i, the coordinates of these rectangles must have two different values: 1 and $\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor - 1$, because they corresponds $a_i^{(min,t)}$ and $a_i^{(max,t)}$, respectively. So, $|A^{(arc,t)}| \geq 2$ as $t \longrightarrow \infty$.

As we can observe in RECT (Function 4), the i^{th} dimension (objective) is divided into $\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor$ segments. The objective space is divided into $\prod_{i=1}^{m}\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor$ hyper-rectangles. From each hyper-rectangle, at most one solution can be in $A^{(arc,t)}$ at the same time. Now consider the equivalence classes of hyper-rectangles where, without loss of generality, the hyper-rectangles in each class have the same coordinates in all but one dimension. There are at most $\max_{i=1,\dots,m}\left\lfloor \frac{\frac{\pi}{2}}{e_i} + 1.5 \right\rfloor$ different hyper-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 ϵ-Pareto set, which sometimes retains only one solution [13].

6 Conclusion and discussion

In this paper, we have introduced the E-(approximate) Pareto set as a novel solution 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 Pareto dominance concept in the MOEAs literature. Moreover, it is practically important as it works with Pareto fronts of bounded size without prior knowledge about multiobjective problems.

We have constructed the ARA archiving strategy that can be used in evolutionary algorithms. It can maintain the minimal objective values and well-distributed Pareto optimal solutions among the solutions generated so far (Theorems 1 and 2).

Our archiving strategy, with 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 (Theorems 3 and 4).

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

In future, we will apply ARA to real life applications. Instead of tangent, we may apply other transferring functions in order to treat different solution regions more fairly.

Our theoretical analysis is based on the assumption of finite search space, however, the E-Pareto set concept is applicable to more complicated situations. We also leave these for future work.

Acknowledgments

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. LU 3009/02E). We would like to express our sincere thanks to anonymous referees for their comments and suggestions.

Bibliography

[1] C. A. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, May 2002. ISBN 0-3064-6762-3.

[2] D. W. Corne, N. R. Jerram, J. D. Knowles, and M. J. Oates. PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization. In L. Spector and *et al.*, editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2001)*, pages 283–290, 2001.

[3] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, 2001.

[4] K. Deb, M. Mohan, and S. Mishra. Towards a quick computation of well-spread pareto-optimal solutions. In C. M. Fonseca and *et al.*, editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 222–236, April 2003.

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

[6] W. Feller. *An Introduction to Probability Theory and Practice*. Wiley, Singapore, 3rd edition, 1976.

[7] J. Fieldsend, R. Everson, and S. Singh. Using unconstrained elite archives for multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 7(3):305–323, June 2003.

[8] C. M. Fonseca and P. J. Fleming. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. In S. Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 416–423, 1993.

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

[10] J. Horn, N. Nafpliotis, and D. E. Goldberg. A Niched Pareto Genetic Algorithm for Multiobjective Optimization. In *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, volume 1, pages 82–87, June 1994.

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

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

[13] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary Computation*, 10(3):263–282, Fall 2002.

[14] H. Reuter. An approximation method for the efficiency set of multiobjective programming problems. *Optimization*, 21:905–911, 1990.

[15] G. Rudolph and A. Agapie. Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, July 2000.

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

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

[18] G. Yen and H. Lu. Dynamic multiobjective evolutionary algorithm: adaptive cell-based rank and density estimation. *IEEE Transactions on Evolutionary Computation*, 7(3):253–274, June 2003.

[19] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. In K. Giannakoglou and *et al.*, editors, *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, September 2001.

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