

Elitist Archiving for Multi-Objective Evolutionary Algorithms: To Adapt or Not to Adapt

Hoang N. Luong and Peter A.N. Bosman

Centrum Wiskunde & Informatica (CWI)
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
{Hoang.Luong,Peter.Bosman}@cwi.nl
<http://www.cwi.nl>

Abstract. Objective-space discretization is a popular method to control the elitist archive size for evolutionary multi-objective optimization and avoid problems with convergence. By setting the level of discretization, the proximity and diversity of the Pareto approximation set can be controlled. This paper proposes an adaptive archiving strategy which is developed from a rigid-grid discretization mechanism. The main advantage of this strategy is that the practitioner just decides the desirable target size for the elitist archive while all the maintenance details are automatically handled. We compare the adaptive and rigid archiving strategies on the basis of a performance indicator that measures front quality, success rate, and running time. Experimental results confirm the competitiveness of the adaptive method while showing its advantages in terms of transparency and ease of use.

Keywords: Multiobjective optimization, estimation of distribution algorithms, elitist archive.

1 Introduction

Optimization problems in practice may involve more than a single objective, and often conflicting ones. A *utopian* solution, that optimizes all objectives at the same time, is unachievable. A solution \mathbf{x} can be better than another solution \mathbf{y} in some objectives, but worse in others. The optimum for such multi-objective optimization (MO) problems is thus a set of equally preferable trade-off solutions rather than a single optimal point. We formalize the terminologies and notations for MO used in this paper as follows:

1. **Multi-objective optimization.** m objective functions $f_i(\mathbf{x}), i \in \mathcal{M} = \{0, 1, \dots, m-1\}$, without loss of generality, must all be minimized. A solution vector $\mathbf{x} = (x_0, x_1, \dots, x_{k-1})$ in the decision space has an corresponding image vector $\mathbf{f}(\mathbf{x}) = (f_0(\mathbf{x}), f_1(\mathbf{x}), \dots, f_{m-1}(\mathbf{x}))$ in the objective space.
2. **Pareto dominance.** A solution \mathbf{x}^0 *dominates* a solution \mathbf{x}^1 (denoted $\mathbf{x}^0 \succ \mathbf{x}^1$) if and only if $(\forall i \in \mathcal{M} : f_i(\mathbf{x}^0) \leq f_i(\mathbf{x}^1)) \wedge (\mathbf{f}(\mathbf{x}^0) \neq \mathbf{f}(\mathbf{x}^1))$.

3. **Pareto set.** A Pareto set \mathcal{P} of size n is called a Pareto set if and only if $\neg\exists x^0, x^1 \in \mathcal{P} : x^0 \succ x^1$.
4. **Pareto optimality.** A solution x^0 is said to be *Pareto optimal* if and only if $\neg\exists x^1 : x^1 \succ x^0$.
5. **Pareto-optimal set.** The set \mathcal{P}_S of all Pareto-optimal solutions: $\mathcal{P}_S = \{x^0 | \neg\exists x^1 : x^1 \succ x^0\}$.
6. **Pareto-optimal front.** The set \mathcal{P}_F in the objective space of all image vectors corresponding to the solutions in \mathcal{P}_S in the decision space: $\mathcal{P}_F = \{f(x) = (f_0(x), f_1(x), \dots, f_{m-1}(x)) | x \in \mathcal{P}_S\}$.

The optimal solution for a multi-objective optimization problem is the Pareto-optimal set \mathcal{P}_S and its corresponding image \mathcal{P}_F . The actual size of \mathcal{P}_S and \mathcal{P}_F may be infinite or too numerous to be obtained by finite computational resources. In practice the desired result is often a representative solutions subset \mathcal{S} of \mathcal{P}_S having a reasonable size, from which decision makers are able to consider and make their final choice. This subset \mathcal{S} should have its image $f(\mathcal{S})$ well-spread along the Pareto-optimal front \mathcal{P}_F , which means diversity in the quality of trade-off solutions regarding all related objectives.

Different works have shown that elitism is crucial for the convergence of multi-objective optimization evolutionary algorithms (MOEAs) [1,2]. While elitist preservation for single objective optimization is a trivial task, in which the only best solution needs to be kept and updated along the run, multi-objective optimization requires more complicated elitism strategies. A separate data structure, called the elitist archive, is often used to keep track of the best Pareto set, in which every solution is not dominated by any other solution in the whole population nor by any other elitist solution in earlier generations. When the number of solutions on the Pareto front is large, the archive may grow to an extreme size. Large archives are furthermore computationally expensive to maintain. Because computational resources are always limited, an upper bound for the archive size is definitely compulsory. Problems occur when this upper bound is reached and new non-dominated solutions continue to be found. One way to differentiate MOEAs is how they treat this elitist archiving problem.

Laumanns et al. [3] propose ϵ -dominance and ϵ -Pareto set to address the problem of convergence and diversity of the approximate set. The ϵ -Pareto set is proved to have bounded size. However, Hernández-Díaz et al. [4] point out that the box-domination scheme for maintaining an ϵ -Pareto set prevents the archive from achieving its intended upper bound. The authors then present Pareto-adaptive ϵ -dominance (*pa* ϵ -dominance) with a curve-fitting scheme to determine several parameters in order to generate a more suitable grid depending on the problem being solved at hand. *pa* ϵ -dominance limits the types of Pareto fronts it can handle to the curves of the family $\{x_1^p + x_2^p + \dots + x_n^p = 1 : 0 \leq x_1, x_2, \dots, x_n \leq 1, 0 < p < \infty\}$, and the objective space should thus be continuous. A notable adaptive grid archiving (AGA) strategy is presented by Knowles and Corne [5]. AGA uses a grid, which can adapt its position and size, to estimate the density of the archived solutions in the objective space. When the archive is full, and a new non-dominated solution is generated in a less crowded region, a solution in

a more crowded region will be removed. If the new solution lies in an already crowded region, it will be ignored. AGA thus can control the size of the elitist archive, but its convergence cannot be guaranteed [5]. Furthermore, without *a priori* knowledge of true Pareto front ranges, determining how many regions the objective space should be divided into (to generate the grid) before the run is started is an uninformed decision, and thus could easily make the grid become too coarse-grained or too fine-grained.

In this paper, we present a new adaptive elitist archiving strategy for MOEAs. The work is based on a straightforward rigid objective-space discretization approach that was already used in earlier research [6]. With the proposed adaptive elitist archiving strategy, an optimization practitioner can straightforwardly decide her desirable archive size, and let the algorithm automatically adapt its structure. Our paper is organized as follows. In Section 2 we describe two archiving strategies: the rigid-grid discretization, and the adaptive grid discretization. Section 3 shows experimental results comparing the performance of the two strategies under different parameter settings. Section 4 concludes our paper.

2 Elitists Archiving Strategies

2.1 Rigid Grid Discretization

MOEAs with competent operators (e.g., selection, modelling, and variation operators) can generate good solutions which are distributed along the true Pareto-optimal front. Because the number of non-dominated solutions may exceed the capacity of the elitist archive, archiving strategies are needed to decide which solutions should be stored and which solutions can be discarded. To limit the elitist archive to reasonable sizes while ensuring that non-dominated solutions are potentially well distributed across their ranges, the objective space is discretized into equal hypercubes, and each hypercube is allowed to contain only one solution at a time (see, e.g. [6]). The discretization is performed by dividing each objective dimension f_i into equal segments of unit length λ_i ; for the sake of simplicity, here λ_i are set to the same λ for all objectives. Because the edge-lengths λ_i of hypercubes are determined before an MOEA run, and are fixed during the run, we refer to this method as rigid-grid discretization (RGD).

When the MOEA generates a non-dominated solution, it will go through an acceptance test to enter the elitist archive. If the new solution is (Pareto) dominated by any archive solutions, it is discarded. A new non-dominated solution can enter the elitist archive if and only if it occupies an empty hypercube or it dominates the solution that currently resides in the same hypercube. If the new non-dominated solution does not dominate the occupant, that new solution is considered as a dominated solution and is discarded as well. When a new solution is accepted into the archive, all solutions dominated by it are removed from the archive to ensure that the archive is always a Pareto set. The pseudo-code for adding a non-dominated solution into the archive is described in Fig. 1.

While keeping non-dominated solutions well-spread, RGD also prevents the elitist archive from degeneration. Degeneration happens if an MOEA prunes a

<pre> UPDATEELITISTARCHIVE($\mathcal{A}, \mathbf{x}^0, \lambda$) $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_{m-1})$ 1 if $\exists \mathbf{x}^1 \in \mathcal{A} : \mathbf{x}^1 \succ \mathbf{x}^0$ then 1.1 $\mathcal{A}' \leftarrow \mathcal{A}$ 2 else if $\exists \mathbf{x}^1 \in \mathcal{A} : \text{ISAMEBOX}(\mathbf{f}(\mathbf{x}^0), \mathbf{f}(\mathbf{x}^1), \lambda) \wedge \mathbf{x}^0 \not\succeq \mathbf{x}^1$ 2.1 $\mathcal{A}' \leftarrow \mathcal{A}$ 3 else 3.1 $\mathcal{D} \leftarrow \{\mathbf{x}^1 \in \mathcal{A} \mid \mathbf{x}^0 \succ \mathbf{x}^1\}$ 3.2 $\mathcal{A}' \leftarrow \mathcal{A} \cup \{\mathbf{x}^0\} \setminus \mathcal{D}$ 4 return \mathcal{A}' </pre>	<pre> ISAMEBOX($\mathbf{f}^0, \mathbf{f}^1, \lambda$) $\mathbf{f}^0 = (f_0^0, f_1^0, \dots, f_{m-1}^0)$, $\mathbf{f}^1 = (f_0^1, f_1^1, \dots, f_{m-1}^1)$, $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_{m-1})$ 1 for $i \leftarrow 0$ to $m-1$ do 1.1 if $\lfloor f_i^0 / \lambda_i \rfloor \neq \lfloor f_i^1 / \lambda_i \rfloor$ then 1.1.1 return false 2 return true </pre>
---	---

Fig. 1. Pseudo-code for adding a non-dominated solution \mathbf{x}^0 into the elitist archive \mathcal{A}

non-dominated solution \mathbf{x}^g from its elitist archive at iteration g , and at a later generation g' , the archive accepts a solution $\mathbf{x}^{g'}$ which would be dominated by \mathbf{x}^g if \mathbf{x}^g were still in the archive. This for instance is the case for the archiving strategies adopted in the well-known MOEAs NSGA-II and SPEA2 [7]. With RGD, there is no need to additionally prune the elitist archive; RGD just decides whether or not a solution is qualified to enter the archive. An occupant of a hypercube is removed if and only if it is dominated by a better solution which is newly accepted into the elitist archive. Because the ranges of the Pareto-optimal front are limited, the maximal number m_λ of non-dominated solutions which can be put into the grid corresponding to a discretization level λ is bounded. The elitist archive size is thus always less than or equal to m_λ . If the MOEA does not generate any better solutions, the elitist archive will stay the same. The MOEA thus converges in this sense.

Experimental results showed the effectiveness of this rigid grid discretization technique on various benchmark problems [6]. However, it requires practitioners to set the discretization levels (i.e. the hypercube sizes) before the run. If information about the ranges of feasible solutions in the objective space is not prior knowledge, then setting fixed values is problematic and raises problems such as making the archive too coarse-grained or too fine-grained. A too coarse-grained discretized objective space misses many valuable solutions, and a too fine-grained archive requires a considerable amount of computational resources to maintain. Furthermore, manually setting the hypercube sizes is not a transparent manner to control the ultimate archive size from the perspective of decision makers.

2.2 Adaptive Grid Discretization

In real-life scenarios, a practitioner may not have prior knowledge about the ranges of the Pareto-optimal front, and she still needs to control the elitist archive size around an allowable size t due to limitations in computational resources. We resolve this problem by proposing an adaptive grid discretization mechanism (AGD). Regarding available resources, the practitioner can decide her *budget* for the elitist archive before an MOEA run, and the objective space will be adaptively discretized to maintain the archive around the target size t .

The archive functions much like in the rigid case: non-dominated solutions enter the archive, and dominated solutions are removed. When the archive size deviates

```

MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM()
1  $\mathbf{P} \leftarrow \text{INITIALIZE}()$ 
2  $\mathbf{F} \leftarrow \{f(\mathbf{x}) = (f_0(\mathbf{x}), f_1(\mathbf{x}), \dots, f_{m-1}(\mathbf{x})) \mid \mathbf{x} \in \mathbf{P}\}$ 
3  $\mathcal{A} \leftarrow \emptyset$ 
4  $\boldsymbol{\lambda} \leftarrow (0, 0, \dots, 0)$ 
5 while  $\neg \text{TERMINATION CONDITIONS SATISFIED}()$  do
  5.1  $(\mathbf{S}, \mathbf{F}^S) \leftarrow \text{MAKE SELECTION}(\mathbf{P}, \mathbf{F})$ 
  5.2  $\mathbf{O} \leftarrow \text{GENERATE NEW SOLUTIONS}(\mathbf{S}, \mathcal{A})$ 
  5.3  $\mathbf{F}^O \leftarrow \{f(\mathbf{x}) = (f_0(\mathbf{x}), f_1(\mathbf{x}), \dots, f_{m-1}(\mathbf{x})) \mid \mathbf{x} \in \mathbf{O}\}$ 
  5.4 for  $i \leftarrow 0$  to  $|\mathbf{O}| - 1$  do
    5.4.1  $\mathcal{A} \leftarrow \text{UPDATE ELITIST ARCHIVE}(\mathcal{A}, \mathbf{O}[i], \boldsymbol{\lambda})$ 
  5.5 if  $|\mathcal{A}| > t^{up}$  then
    5.5.1  $(\mathcal{A}, \boldsymbol{\lambda}) \leftarrow \text{ADAPT GRID DISCRETIZATION}(\mathcal{A})$ 
  5.6  $(\mathbf{P}, \mathbf{F}) \leftarrow \text{MAKE REPLACEMENT}((\mathbf{S}, \mathbf{F}^S), (\mathbf{O}, \mathbf{F}^O))$ 
ADAPT GRID DISCRETIZATION( $\mathcal{A}$ )
1  $low \leftarrow 1$ 
2  $high \leftarrow MAX$ 
3  $\mathcal{A}' \leftarrow \mathcal{A}$ 
4  $max_i \leftarrow \max\{f_i^0, f_i^1, \dots, f_i^{|\mathcal{A}'|-1}\}, i \in \{0, 1, \dots, m-1\}$ 
5  $min_i \leftarrow \min\{f_i^0, f_i^1, \dots, f_i^{|\mathcal{A}'|-1}\}, i \in \{0, 1, \dots, m-1\}$ 
6 for  $count \leftarrow 0$  to  $N - 1$  do
  6.1  $mid = \frac{low + high}{2}$ 
  6.2  $\lambda_i = \frac{max_i - min_i}{mid}, i \in \{0, 1, \dots, m-1\}$ 
  6.3  $\boldsymbol{\lambda} \leftarrow (\lambda_0, \lambda_1, \dots, \lambda_{m-1})$ 
  6.4  $\mathcal{A} \leftarrow \emptyset$ 
  6.5 for  $j \leftarrow 0$  to  $|\mathcal{A}'| - 1$  do
    6.5.1  $\mathcal{A} \leftarrow \text{UPDATE ELITIST ARCHIVE}(\mathcal{A}, \mathcal{A}'[j], \boldsymbol{\lambda})$ 
  6.6 if  $|\mathcal{A}| < t^{low}$  then
    6.6.1  $low \leftarrow mid$ 
  6.7 else
    6.7.1  $high \leftarrow mid$ 
7 return  $\mathcal{A}, \boldsymbol{\lambda}$ 

```

Fig. 2. Pseudo-code for adaptive grid discretization. Initially, $\boldsymbol{\lambda}$ is assigned a zero vector $(0, 0, \dots, 0)$, which means that no objective-space discretization is used. MAX is the maximal number of segments which a dimension can be divided into. N is the maximum number of steps in the binary search for objective-space discretization. In this paper, we set $MAX = 2^{25}$ and $N = 25$.

too much from the target size t , the edge-lengths need to be re-determined. Because we do not want to perform the objective-space discretization every time a single non-dominated solution is generated, we allow the elitist archive to grow to an upper bound t^{up} before pruning it. To prevent the adaptation process from deleting too many solutions, we set an lower bound of t^{low} for the elitist archive size. As soon as the archive size reaches the upper bound, the objective space adaptation process is triggered. AGD first determines the ranges of all current archived solutions in the objective space, and then performs a binary search, targeted at t^{low} , for how many segments each range should be divided into. The final discretization must satisfy the condition that the archive size is greater than the lower bound and less than the upper bound (i.e. $t^{low} < t < t^{up}$). The details of AGD are described in Fig. 2. In this paper, we set t^{low} and t^{up} as $0.75*t$ and $1.25*t$, respectively. We calibrated these values by hand taking into account that if t^{up} is too large or t^{low} is too small then the actual archive size may deviate too much from the target, while setting the bounds closer to t increases the computational overhead for re-discretization too much.

AGD can be seen as a sequence of RGDs with different discretization levels λ . When changing to a new discretization, degeneration of the elitist Pareto front can happen because some non-dominated solutions are removed, but during an RGD, degeneration does not happen. Ultimately, there is an iteration g when solutions in the elitist archive already cover the ranges of the Pareto-optimal front, i.e. the MOEA is nearing the Pareto-optimal front, and the current discretization level λ^g ensures that the maximal number of non-dominated solutions which can be put into the grid is close to the target size t of the elitist archive (i.e., $m_{\lambda^g} \approx t$, and $m_{\lambda^g} \leq t^{up}$). From that iteration g , there is no need to re-discretize the objective space any more. If the MOEA does not generate any better solutions, the elitist archive will stay the same. The MOEA thus again converges in this sense.

3 Experiments

3.1 Benchmark Problems

In this paper, we select the MAMaLGaM (Multi-objective Adapted Maximum-Likelihood Gaussian Model [6]) as the MOEA to be combined with the two elitist archiving strategies above. It should be noted however that because these archiving mechanisms work independently from the generation of new solutions, they can be readily implemented in other MOEAs, including those aimed at discrete parameter spaces. We carry out a performance assessment of the two archiving strategies over 8 benchmark problems described in Table 1. ZDT $_i$, $i \in \{1, 2, 3, 6\}$ are well-known test problems proposed by Zitzler et al. [2]. GM $_i$, $i \in \{1, 2\}$ are generalizations of the MED (Multiple Euclidean Distances) problems [8]. Developed from the well-known Rosenbrock function, BD $_i$, $i \in \{1, 2\}$ were recently introduced in the MOEA literature [9]. Details about construction and difficulty of these benchmarks can be found in the referenced research.

We refer to an approximation set \mathcal{S} as the combination of the elitist archive and all the non-dominated solutions in the current population. We consider the outcome of an MOEA to be the final approximation set upon termination. To compare the quality of approximation sets, we use a performance indicator, denoted $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$.

$$D_{\mathcal{P}_F \rightarrow \mathcal{S}}(\mathcal{S}) = \frac{1}{|\mathcal{P}_F|} \sum_{f^0 \in \mathcal{P}_F} \min_{\mathbf{x} \in \mathcal{S}} \{d(\mathbf{f}(\mathbf{x}), f^0)\} \quad (1)$$

where \mathcal{P}_F is the Pareto-optimal front, $\mathbf{f}(\mathbf{x})$ is a point in objective space, which is the objective value vector of a solution $\mathbf{x} \in \mathcal{S}$, and $d(\cdot, \cdot)$ computes Euclidean distance. $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ is also referred to in literature as inverse generational distance. It can be inferred from Equation 1 that the proximity and diversity of \mathcal{S} with respect to the Pareto-optimal set \mathcal{P}_S is measured in the objective space with regard to the Pareto-optimal front \mathcal{P}_F . Because the Pareto-optimal fronts of all test problems here are continuous and thus are infinitely large, for the sake of computability, we approximated the true \mathcal{P}_F by uniformly sampling along it a subset of 5000 points.

The smaller $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ value an approximation set \mathcal{S} has, the better its quality is. In practice the Pareto-optimal front may not be known, and thus the performance indicator $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ cannot be used. However, for benchmarking purposes, where \mathcal{P}_F is available, this indicator has a two-fold advantage: it can measure both the proximity and diversity of \mathcal{S} with respect to \mathcal{P}_F . In our experiments, an MOEA run with its final \mathcal{S} having $D_{\mathcal{P}_F \rightarrow \mathcal{S}} \leq 0.01$ is considered as a successful run because such approximation set are quite close to the true Pareto-optimal fronts. Fig. 3 shows the default Pareto optimal fronts and the adaptive elitist archives with different desirable target sizes t for all the problems.

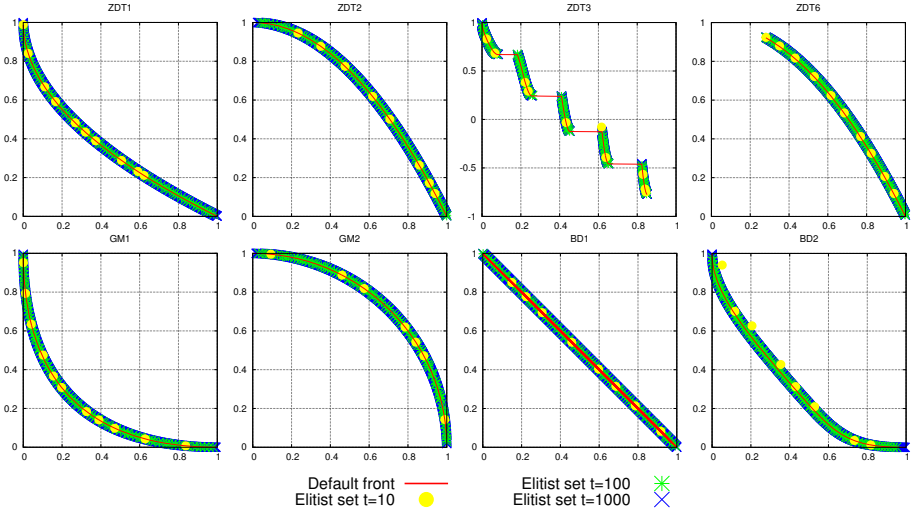


Fig. 3. For all problems: the default front and the elitist archive of 3 different archive size settings. Horizontal axis is f_0 objective value. Vertical axis is f_1 objective value.

Table 1. The MO problem test suite

Name	Objectives	IR
GM ₁	$f_0 = \ \frac{1}{2}(x - c^0)\ ^d, \quad f_1 = \ \frac{1}{2}(x - c^1)\ ^d$ $c^0 = (1, 0, 0, \dots), \quad c^1 = (0, 1, 0, 0, \dots), \quad d = 2$	$[-1; 1]^{10}$ ($l = 10$)
GM ₂	$f_0 = \ \frac{1}{2}(x - c^0)\ ^d, \quad f_1 = \ \frac{1}{2}(x - c^1)\ ^d$ $c^0 = (1, 0, 0, \dots), \quad c^1 = (0, 1, 0, 0, \dots), \quad d = \frac{1}{2}$	$[-1; 1]^{10}$ ($l = 10$)
ZDT ₁	$f_0 = x_0, \quad f_1 = \gamma(1 - \sqrt{f_0/\gamma})$ $\gamma = 1 + 9(\sum_{i=1}^{l-1} x_i/(l-1))$	$[0; 1]^{30}$ ($l = 30$)
ZDT ₂	$f_0 = x_0, \quad f_1 = \gamma(1 - (f_0/\gamma)^2)$ $\gamma = 1 + 9(\sum_{i=1}^{l-1} x_i/(l-1))$	$[0; 1]^{30}$ ($l = 30$)
ZDT ₃	$f_0 = x_0, \quad f_1 = \gamma(1 - \sqrt{f_0/\gamma} - (f_0/\gamma)\sin(10\pi f_0))$ $\gamma = 1 + 9(\sum_{i=1}^{l-1} x_i/(l-1))$	$[0; 1]^{30}$ ($l = 30$)
ZDT ₆	$f_0 = 1 - e^{-4x_0} \sin^6(6\pi x_0), \quad f_1 = \gamma(1 - (f_0/\gamma)^2)$ $\gamma = 1 + 9(\sum_{i=1}^{l-1} x_i/(l-1))^{0.25}$	$[0; 1]^{10}$ ($l = 10$)
BD ₁	$f_0 = x_0, \quad f_1 = 1 - x_0 + \gamma$ $\gamma = \sum_{i=1}^{l-2} (100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2)$	$[0; 1] \times$ $[-5.12; 5.12]^9$ ($l = 10$)
BD ₂	$f_0 = \frac{1}{l} \sum_{i=0}^{l-1} x_i^2$ $f_1 = \frac{1}{l-1} \sum_{i=0}^{l-2} (100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2)$	$[-5.12; 5.12]^{10}$ ($l = 10$)

3.2 Results

All the presented results here are averaged over 30 runs. Each run has a budget of 10^6 fitness evaluations. MAMaLGaM terminates when it uses all the allowable evaluations, or when all the distribution multipliers ≤ 0.5 . Details about the operations of MAMaLGaM and its components can be found in [6].

Fig. 4 shows convergence graphs of the $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator values from the beginning until termination for MAMaLGaM with the two elitist archiving strategies on all benchmark problems. When the elitist archive has limited volume (i.e., the target size is too small, $t = 10$, or the grid is too coarse-grained, $\lambda = 0.1$), it is less likely to achieve the desirable convergence ($D_{\mathcal{P}_F \rightarrow \mathcal{S}} \leq 0.01$). Otherwise, when having archives of adequate capacity, the MOEA achieves good convergence behavior for both variants of archiving mechanisms. Fig. 4 also shows that the greater the elitist archive is, the better $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator values it can obtain. MAMaLGaM with rigid grid of $\lambda = 0.001$ shows its superiority in most of problems because it maintains the largest number of non-dominated solutions. Table 2 shows the average numbers of solutions in the archive for each benchmark problems. Because of allowing more solutions in the elitist archive, and thus in the approximation sets, the rigid grid MOEAs do perform slightly better than the their relatively corresponding adaptive versions (i.e., $\lambda = 0.1$ vs $t = 10$, $\lambda = 0.01$ vs $t = 100$, $\lambda = 0.001$ vs $t = 1000$). The $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator values of RGD are thus slightly better than those of AGD. Note however, that this is a consequence of our choice for setting λ and t , and not because of inferiority of AGD. Doubling t would give similar $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator values. Also, if we

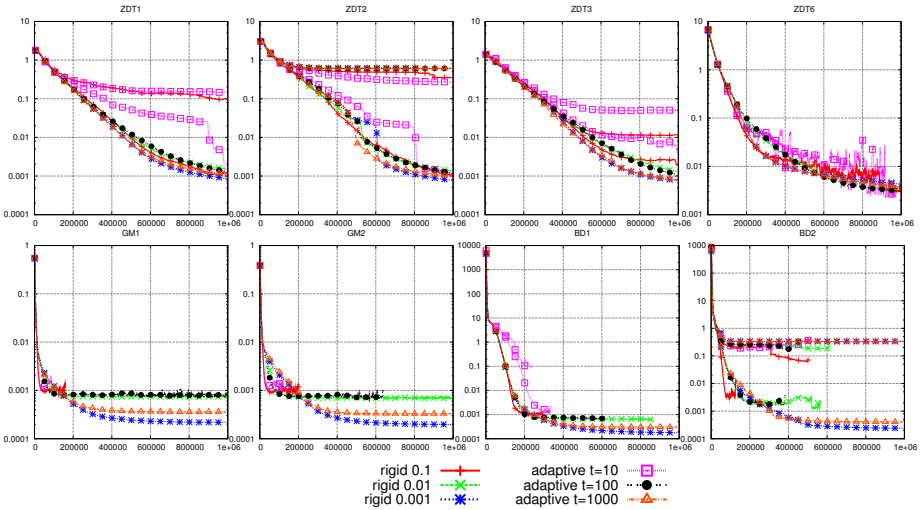


Fig. 4. Average performance of MAMaLGaM with two archiving strategies on all problems. Horizontal axis: number of evaluations (both objectives per evaluation). Vertical axis: $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$. For each algorithm averages are shown both for successful runs and unsuccessful runs, giving double occurrences of lines if some runs were unsuccessful.

Table 2. Elitist archive sizes, success rates (i.e. the percentage of times MAMaLGaM obtained $D_{\mathcal{P}_F \rightarrow S}$ indicator ≤ 0.01), and average running times (measured in seconds) of MAMaLGaM with 2 variants of elitist archiving strategies on all problems.

	BD ₁	BD ₂ ^S	GM ₁	GM ₂	ZDT ₁	ZDT ₂	ZDT ₃	ZDT ₆
ELITIST ARCHIVE SIZES								
$\lambda = 0.1$	22	16	21	21	21	16	24	17
$\lambda = 0.01$	200	166	201	197	200	194	210	165
$\lambda = 0.001$	2000	1751	1975	1954	1996	1852	2020	1576
$t = 10$	8	8	8	8	8	8	8	9
$t = 100$	98	97	100	108	90	102	103	105
$t = 1000$	1017	974	1035	1052	957	1003	1045	964
SUCCESS RATES								
$\lambda = 0.1$	100	3	100	100	93	80	90	100
$\lambda = 0.01$	100	43	100	100	100	96	100	100
$\lambda = 0.001$	100	66	100	100	100	93	100	100
$t = 10$	90	0	100	100	26	80	50	93
$t = 100$	100	66	100	100	100	96	100	100
$t = 1000$	100	66	100	100	100	93	100	100
AVERAGE RUNNING TIMES								
$\lambda = 0.1$	542	358	241	286	1101	1070	1073	1774
$\lambda = 0.01$	2015	1069	2589	2417	1161	1127	1126	2677
$\lambda = 0.001$	4892	4492	5211	5161	1903	1714	1778	4196
$t = 10$	536	552	160	160	1110	1105	1087	1224
$t = 100$	1180	719	1045	1139	1116	1106	1094	2530
$t = 1000$	4231	3834	4502	4498	1479	1424	1458	3925

terminate an MOEA run when it reaches the successful threshold ($D_{\mathcal{P}_F \rightarrow S} \leq 0.01$), it can be seen that the adaptive and rigid archives have similar convergence behavior.

Table 2 shows the percentage of runs that an MOEA finds a final approximation set with performance indicator value $D_{\mathcal{P}_F \rightarrow S} \leq 0.01$, which is considered as successful. It can be seen that MOEAs with tiny archives have lower success rate, which also means poorer reliability, in convergence. When the elitist archive has adequate capacity, regardless of being a rigid or adaptive, the optimization process will, in most of the times, converge successfully to fronts that are quite close to the true Pareto-optimal front. This is however more dependent on the capabilities of the other operators in MAMaLGaM rather than AGD.

Table 2 also demonstrates that the rigid and the adaptive archiving strategies have similar running times, which can partly reflect their computational costs. It is apparent that the greater elitist archive an MOEA has, the more *expensive* it is to maintain. While the match is tied for performance (indicator values), reliability (success rate), and efficiency (computational cost), the adaptive strategy wins over the rigid grid in terms of *transparency* with respect to desired archive size. For the rigid-grid discretization, the practitioner can indirectly and relatively influence the archive size by adjusting the λ value, but she hardly controls its actual growth without prior knowledge about the ranges of the objectives for the Pareto-optimal fronts. If our adaptive archive is employed, the practitioner simply decides the desirable target size of the elitist archive, thus its capacity, and then let all the details be handled behind the scenes.

4 Conclusions

In this paper, we have presented and compared two elitist archiving techniques for evolutionary multi-objective optimization: a rigid objective space discretization and an adaptive version. The two variants are showed to have similar convergence behavior, success rate, and running time on various benchmark problems. The advantage of the adaptive archiving strategy resides in its straightforwardness and transparency through which the practitioners can decide their desirable archive size and all the archiving processes are then automatically handled. According to the ranges of different dimensions in the objective space, our technique is able to select appropriate discretizations such that the final approximation set is well-spread with good proximity concerning the Pareto-optimal front provided that the MOEA is capable of generating such good solutions. Experimental results on benchmark problems support our above claims.

Although we only tested our adaptive archiving technique with the MAMaL-GaM, it can be implemented into other state-of-the-art MOEAs straightforwardly because it works independently from how new solutions are generated. Our technique is not limited to continuous search spaces as its design is not based on any assumptions about the continuity of functions. Our technique has potential to be applied successfully to a broad spectrum of optimization problems.

References

1. Bosman, P.A.N., Thierens, D.: The Balance between Proximity and Diversity in Multi-Objective Evolutionary Algorithms. *IEEE Trans. Evol. Comput.* 7(2), 174–188 (2003)
2. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation* 8(2), 173–195 (2000)
3. Laumanns, M., Thiele, L., Deb, K., Zitzler, E.: Combining Convergence and Diversity in Evolutionary Multiobjective Optimization. *Evolutionary Computation* 10(3), 263–282 (2002)
4. Hernandez-Diaz, A.G., Santana-Quintero, L.V., Coello Coello, C.A.: Pareto-adaptive -dominance. *Evolutionary Computation* 15(4), 493–517 (2007)
5. Knowles, J., Corne, D.: Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Trans. Evol. Comput.* 7(2), 100–116 (2003)
6. Bosman, P.A.N.: The Anticipated Mean Shift and Cluster Registration in Mixture-based EDAs for Multi-Objective Optimization. In: *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, GECCO 2010*, pp. 351–358. ACM, New York (2010)
7. López-Ibáñez, M., Knowles, J., Laumanns, M.: On Sequential Online Archiving of Objective Vectors. In: Takahashi, R.H.C., Deb, K., Wanner, E.F., Greco, S. (eds.) *EMO 2011*. LNCS, vol. 6576, pp. 46–60. Springer, Heidelberg (2011)
8. Harada, K., Sakuma, J., Kobayashi, S.: Local Search for Multiobjective Function Optimization: Pareto Descent Method. In: *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, GECCO 2006*, pp. 659–666 (2006)
9. Bosman, P.A.N., de Jong, E.D.: Combining Gradient Techniques for Numerical Multi-Objective Evolutionary Optimization. *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, GECCO 2006*, 627–634 (2006)