Computing Gap Free Pareto Front Approximations with Stochastic Search Algorithms

Oliver Schütze schuetze@cs.cinvestav.mx
CINVESTAV-IPN, Departamento de Computación, México D.F. 07360, México

Marco Laumanns laumanns@ifor.math.ethz.ch
Institute for Operations Research, ETH Zürich 8092, Zürich, Switzerland

Emilia Tantar Emilia.Tantar@lifl.fr
INRIA Lille Nord Europe, Parc Scientifique de la Haute Borne, 40 avenue Halley, 59650, Villeneuve d'Ascq, France

Carlos A. Coello Coello ccoello@cs.cinvestav.mx
CINVESTAV-IPN, Departamento de Computación, México D.F. 07360, México

El-Ghazali Talbi talbi@lifl.fr
INRIA Lille Nord Europe, Parc Scientifique de la Haute Borne, 40 avenue Halley, 59650, Villeneuve d'Ascq, France

Abstract

Recently, a convergence proof of stochastic search algorithms toward finite size Pareto set approximations of continuous multi-objective optimization problems has been given. The focus was on obtaining a finite approximation that captures the entire solution set in some suitable sense, which was defined by the concept of ϵ-dominance. Though bounds on the quality of the limit approximation—which are entirely determined by the archiving strategy and the value of ϵ—have been obtained, the strategies do not guarantee to obtain a gap free approximation of the Pareto front. That is, such approximations A can reveal gaps in the sense that points f in the Pareto front can exist such that the distance of f to any image point $f(a)$, $a \in A$, is 'large'. Since such gap free approximations are desirable in certain applications, and the related archiving strategies can be advantageous when memetic strategies are included into the search process, we are aiming in this work for such methods. We present two novel strategies that accomplish this task in the probabilistic sense and under mild assumptions on the stochastic search algorithm. In addition to the convergence proofs we give some numerical results to visualize the behavior of the different archiving strategies. Finally, we demonstrate the potential for a possible hybridization of a given stochastic search algorithm with a particular local search strategy—multi-objective continuation methods—by showing that the concept of ϵ-dominance can be integrated into this approach in a suitable way.

©200X by the Massachusetts Institute of Technology

1 Introduction

In a variety of engineering and economic problems several objectives have to be optimized concurrently. One widely accepted class of algorithms for the approximation of the solution set (*Pareto set*) of such *multi-objective optimization problems* (MOPs) is given by evolutionary strategies. A typical evolutionary multi-objective (EMO) algorithm consists, roughly speaking, of a process to generate new candidate solutions (the generator) and a strategy to store and update a 'suitable' subset of the obtained data according to the given task (the archiver). Under mild assumptions about the generator, the limit approximation set is determined almost entirely by the archiving strategy. By limit behavior we mean the behavior of the sequence of archives A_l, $l \in \mathbb{N}$, which is generated by the stochastic search algorithm in the course of the computation for iteration step $l \to \infty$. The task of this work is to develop archivers which aim for finite size and gap free approximations of the solution set, which we motivate in the following.
One interesting application of multi-objective optimization and its related tools is the online-optimization of mechatronical systems. One approach to this problem is as follows: first, all relevant (conflicting) objectives of the underlying system are collected and used to formulate a multi-objective optimization problem. This problem is then solved numerically by approximating the Pareto set (denote the approximation by $\mathcal{P}$) offline. This set serves further on as the basis for the online control by providing a repository of reference operating points: the 'optimal' point (or optimal compromise) $p(\lambda) \in \mathcal{P}$ is determined online—i.e., while running the system—according to the current situation or demand λ of the system and is used as the actual *operating point*. Since $\lambda = \lambda(t)$ varies with the time, this 'optimal' point has to be updated over and over again, according to the sensitivity of the system. See (Pottharst et al., 2004; Witting et al., 2008) for an operating point assignment strategy of a linear drive, and (Hestermeyer and Oberschelp, 2003) for an online-adjustment of an active suspension system. Crucial for the *stability* of the system is that the switch from one point or system setting $p(\lambda_1)$ to the next one $p(\lambda_2)$ can *not* be done arbitrarily, but has to be carried out as smoothly as possible. That is, large and abrupt qualitative changes—(amongst others) in terms of the changes in the influential objective values—have to be avoided. Thus, it is required in these applications—and certainly in others as well—to obtain a gap free (and preferably uniformly spread) Pareto front approximation.[1]
There exist on the other hand certainly also scenarios where a smooth changeover of the *parameter* values over time is of particular interest. The algorithms we consider here are, hovewer, not totally suited—but also not designed—for such cases, since these algorithms are entirely based on the dominance and ϵ-dominance relations, which are defined in objective space: consider a point $f \in F(\mathcal{P})$ of the image of the Pareto set (the Pareto front), where $F : \mathbb{R}^n \to \mathbb{R}^k$ is the function of given objectives and consider that f has several preimages $x_i \in \mathcal{P}$, $i = 1, \ldots, s$. Then the archives of the subsequent archiving strategies will probably contain (and retain) one approximate solution of one preimage x_i, $i \in \{1, \ldots, s\}$, after sufficiently many iterations, but (i) the index i depends on the order of the incoming solutions and is thus not controllable, and (ii) further preimages $x_j \in \{1, \ldots, s\} \backslash \{i\}$ (or points nearby) will not be accepted further on. For this we refer to archiving strategies which aim for the approximation of the entire Pareto set or a superset of it (e.g., (Rudolph et al., 2007; Schütze et al., 2007; Schütze

[1] In case the Pareto front falls into different connected components, further techniques (e.g., interpolation strategies among 'neighboring' system settings) have in addition to be considered, but such cases are not part of this work.

et al., 2008)).

Here we extend the work of (Schütze et al., 2008) and present new archiving strategies for the storage of the 'essential' solutions found by a stochastic search algorithm. The strategy used in (Schütze et al., 2008) is entirely based on the concept of ϵ-dominance, which does not consider the distances between solutions in the archive. This can lead to gaps in the approximation set, in particular when some portion of the front is flat or contains a dent (see Figure 1). Though we agree that these phenomena do not occur too often in practice, it is desirable, from a theoretical viewpoint, to have a search algorithm—including a suitable archiving strategy—which can exclude these unwanted gaps in the approximation.

Another important aspect is that the sole usage of the concept of ϵ-dominance in the archiving strategy can cause inefficiencies for the resulting search algorithm, in particular when hybridized with a local search procedure. For instance, when using multi-objective continuation methods (see (Schütze, 2004) for a combination of this technique with evolutionary strategies), where the underlying idea is to move *along* the efficient set, an unsuited archiving strategy as the one proposed in (Schütze et al., 2008) could lead to difficulties, although these methods are (in principle) very effective locally. For this, consider a Pareto front such as the one displayed in Figure 1 (left) and assume that the archive is given by $A = \{a_1\}$. If the continuation method is started with $\{a_1\}$ and merely the concept of ϵ-dominance is used for the archiving strategy this could lead to the fact that no points p on the front with $f_1(p) > f_1(a_1)$ are kept by the archiver. The reason is that there is a relatively large portion of the front near $F(a_1)$, where all points are ϵ-dominated by a_1—a 'barrier' which is hard or impossible to overcome by this (or any other) local search strategy. Figure 1 (right) shows a situation which is more extreme.

In this work we propose two different archiving strategies and prove convergence with probability one to gap free (and thus 'tight') Pareto front approximations. The limit set of the first strategy is a tight ϵ-approximate Pareto set which provides a guaranteed uniformity level, while the limit set of the second strategy forms a tight ϵ-Pareto set, which, however, lacks the uniformity. A previous study of the current work can be found in (Schuetze et al., 2007). While in (Schuetze et al., 2007) mainly the limit behavior of the archivers was studied, this work offers in addition more discussion and results which are intended for a better understanding of the effect of the novel strategies.

The remainder of this article is organized as follows: Section 2 states the background required for the understanding of the sequel. In Section 3, we propose the sets of interest, and in Section 4 two algorithms which aim for their approximation. Section 5 deals with the integration of the archivers into iterative search methods such as multi-objective evolutionary algorithms (MOEAs). In Section 6, we present some numerical results, and finally, we present our conclusions in Section 7.

2 Background and Related Work

In the following we consider continuous unconstrained multi-objective optimization problems

$$\min_{x \in \mathbb{R}^n} \{F(x)\}, \tag{MOP}$$

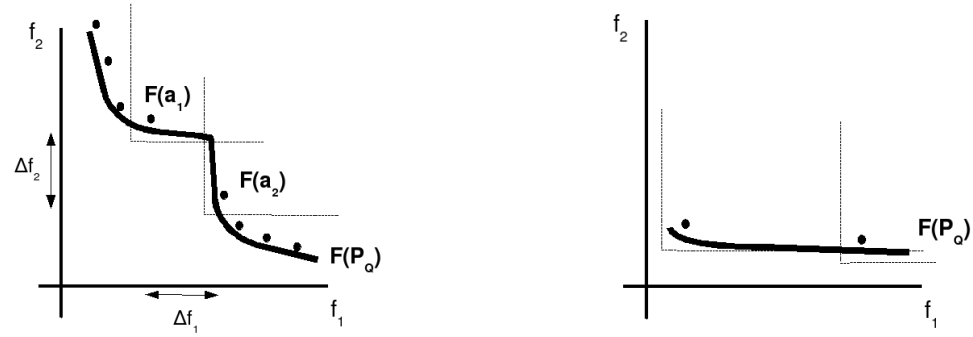


Figure 1: Examples of two ϵ approximate Pareto sets. Dents or 'flat' regions in the Pareto front can lead to (possibly unwanted) gaps in the approximation. In the left figure there is a gap in the approximation between $F(a_1)$ and $F(a_2)$ both in f_1- as in the f_2-direction.

where the function F is defined as the vector of the objective functions

$$F : \mathbb{R}^n \to \mathbb{R}^k, \qquad F(x) = (f_1(x), \ldots, f_k(x)),$$

and where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is continuous. Later we will restrict the search to a compact set $Q \subset \mathbb{R}^n$, the reader may think of as an n-dimensional box

$$Q = B_{l,u} := \{x \in \mathbb{R}^n \ : \ l_i \leq x_i \leq u_i, \ i = 1, \ldots, n\}, \tag{2.1}$$

where $l, r \in \mathbb{R}^n$ with $l_i \leq u_i, i = 1, \ldots, n$. In the next definition we state the classical concept of optimality for MOPs.

Definition 1 *(a) Let $v, w \in \mathbb{R}^k$. Then the vector v is* less than *w ($v <_p w$), if $v_i < w_i$ for all $i \in \{1, \ldots, k\}$. The relation $\leq_p$ is defined analogously.*

(b) A vector $y \in \mathbb{R}^n$ is dominated *by a vector $x \in \mathbb{R}^n$ (in short: $x \prec y$) with respect to (MOP) if $F(x) \leq_p F(y)$ and $F(x) \neq F(y)$ (i.e., there exists a $j \in \{1, \ldots, k\}$ such that $f_j(x) < f_j(y)$.*

(c) A point $x \in \mathbb{R}^n$ is called Pareto optimal *or a* Pareto point *if there is no $y \in \mathbb{R}^n$ which dominates x.*

Denote by P_Q the set of Pareto points (or Pareto set) of $F|_Q$, where $Q \subset \mathbb{R}^n$ is the domain. The image $F(P_Q)$ of the Pareto set is called the *Pareto front*. Both sets consist typically—i.e., under mild regularity assumptions on the objectives—not of finitely many points as for scalar optimization problems, but form $(k-1)$-dimensional objects. In order to guarantee convergence toward P_Q, $F(P_Q)$ or other related objects, one often has to assume that there are no *weak* Pareto points outside P_Q (see e.g., (Ehrgott, 2005; Schütze et al., 2008) for discussions). Such points are defined as follows:

Definition 2 *A point $x \in Q$ is called a* weak Pareto point *if there exists no point $y \in Q$ such that $F(y) <_p F(x)$.*

One important question for an archiving strategy is if it is capable to 'capture' the incoming data in a suitable way. This holds in particular for multi-objective optimization problems due to the dimensionality of the solution set. An 'ideal' archiver in this sense is certainly one which stores in the current archive A for every candidate solution x which has been found by the generation process an element $a \in A$ such that a is equal to x or dominates it. However, since this leads at least for continuous models to a sequence of archives A_i with $|A_i| \to \infty$ for $i \to \infty$, this demand on an archiver is not adequate for practical use. In the following we will define a weaker concept of dominance, so-called (absolute) ϵ-dominance (Loridan, 1984), as well as two approximation concepts which will be used for our further studies since they allow for finite size representations of the solution sets.

Definition 3 *Let $\epsilon = (\epsilon_1, \ldots, \epsilon_k) \in \mathbb{R}^k_+$ and $x, y \in \mathbb{R}^n$. x is said to ϵ-dominate y (in short: $x \prec_\epsilon y$) with respect to (MOP) if*

(i) $f_i(x) - \epsilon_i \leq f_i(y) \quad \forall i = 1, \ldots, k$, *and*

(ii) $f_j(x) - \epsilon_j < f_j(y)$ *for at least one* $j \in \{1, \ldots, k\}$.

Denote by $d(\cdot, \cdot)$ any distance and by $|| \cdot ||$ any norm. Further, let $B_\delta(x_0) := \{x \in \mathbb{R}^n : \|x - x_0\| < \delta\}$ be the open ball with center $x_0 \in \mathbb{R}^n$ and radius $\delta \in \mathbb{R}_+$.

Definition 4 *(Laumanns et al., 2002)*

(a) Let $\epsilon \in \mathbb{R}^k_+$. A set $A_\epsilon \subset \mathbb{R}^n$ is called an ϵ-approximate Pareto set *of (MOP) if every point $x \in \mathbb{R}^n$ is ϵ-dominated by at least one $a \in A_\epsilon$, i.e.*

$$\forall x \in \mathbb{R}^n \ : \ \exists a \in A_\epsilon : \quad a \prec_\epsilon x.$$

(b) A set $A^*_\epsilon \subset \mathbb{R}^n$ is called an ϵ-Pareto set *if A^*_ϵ is an ϵ-approximate Pareto set and if every point $a \in A^*_\epsilon$ is a Pareto point of (MOP).*

Definition 5 *(Sayin, 2000)*

(a) Let $\Delta > 0$ and let $D \subset Z$ be a discrete set. D is called a d_Δ-representation of Z if for any $z \in Z$, there exists $y \in D$ such that $d(z, y) \leq \Delta$.

(b) Let $Z \subset \mathbb{R}^n$ be any set and let D be a d_Δ-representation of Z. Then D is called a δ-uniform d_Δ-representation if

$$\min_{x,y \in D, x \neq y} d(x, y) \geq \delta.$$

δ is called the uniformity level.

Despite the existence of suitable approximation concepts, investigations on the *convergence* of particular algorithms towards such approximation sets, that is, their ability to obtain a suitable Pareto set approximation in the limit, have remained rare. Several studies, such as (Hanne, 1999; Rudolph and Agapie, 2000; Schütze et al., 2008), consider only the convergence to the entire Pareto set, or to a certain subset without considering the approximation quality. The only work which deals with

the computation of gap free representation of the Pareto front, to the authors' best knowledge, is (Fliege, 2004). However, since in that work scalarization methods on convex MOPs are investigated, the approach is not applicable for our purpose.

Finally, the issue of stochastic convergence towards finite-size Pareto set approximations was raised in the area of evolutionary multi-objective optimization, mostly under the assumption of finite search spaces. One option is to use Markov chain results assuming the underlying search processes to be Markovian (Rudolph, 1998). Another option is to define an order homomorphism of the natural dominance relation of approximation sets into a totally ordered set of quality values, thus enforcing a monotonicity of the sequence of solution sets maintained by an algorithm. As shown in (Knowles, 2002; Knowles and Corne, 2003), this entails convergence to a subset of the Pareto set as a local optimum of the quality indicator, but no approximation guarantee could given. Knowles and Corne (2003) also analyzed the adaptive grid archiving proposed in (Knowles and Corne, 2000) and proved that after finite time, even though the solution set itself might permanently oscillate, it will always represent an ϵ-approximation whose approximation quality depends on the granularity of the adaptive grid and on the number of allowed solutions. The results depend on the additional assumption that the grid boundaries converge after finite time, which is fulfilled in certain special cases.

In (Laumanns et al., 2002), two archiving algorithms were proposed that provably maintain a finite-size approximation of all points ever generated during the search process. As an immediate corollary, these archiving strategies were claimed to ensure convergence to a Pareto set approximation of given quality for any iterative search algorithm that fulfills certain mild assumptions about the process to generate new search points. While this claim holds trivially in the case of discrete (or discretized) search spaces, its extension to the continuous case is not straightforward, and was only recently given in (Schütze et al., 2008). A restriction to *discretized* models, however, can lead to problems when, e.g., memetic strategies are used (metaheuristic search algorithms mixed with local search strategies which use step size control themselves).

Next we define some distances between points as well as between different sets.

Definition 6 *Let $u, v \in \mathbb{R}^n$ and $A, B \subset \mathbb{R}^n$. The maximum norm distance d_∞, the semi-distance $dist(\cdot,\cdot)$ and the* Hausdorff distance *$d_H(\cdot,\cdot)$ are defined as follows:*

(a) $d_\infty(u,v) := \max\limits_{i=1,\ldots,n} |u_i - v_i|$

(b) $dist(u,A) := \inf\limits_{v \in A} d_\infty(u,v)$

(c) $dist(B,A) := \sup\limits_{u \in B} dist(u,A)$

(d) $d_H(A,B) := \max\{dist(A,B), dist(B,A)\}$

Algorithm 1 gives a framework of a generic stochastic multi-objective optimization algorithm, which will be considered in this work. Here, $Q \subset \mathbb{R}^n$ denotes the domain of the MOP, P_j the candidate set (or population) of the generation process at iteration step j, and A_j the corresponding archive. Algorithms 2 and 3 show two archiving

strategies which aim for the computation of ϵ-approximate Pareto sets and ϵ-Pareto sets, respectively. The difference between these two archivers is the strategy to accept candidate solutions coming from the generation process (for details we refer to (Schütze et al., 2008) or to the discussion below on the difference of Algorithms 4 and 5 which is similar). Convergence results can be found in Theorems 1 and 2, which are closely related to the corresponding results in the present work, however, Algorithms 2 and 3 can not guarantee that the limit archives do not reveal gaps in the Pareto front.

Algorithm 1 Generic Stochastic Search Algorithm

$P_0 \subset Q$ drawn at random
$A_0 = ArchiveUpdate(P_0, \emptyset)$
for $j = 0, 1, 2, \ldots$ **do**
  $P_{j+1} = Generate(P_j)$
  $A_{j+1} = ArchiveUpdate(P_{j+1}, A_j)$
end for

Algorithm 2 $A := ArchiveUpdateEps1\ (P, A_0)$

1: $A := A_0$
2: **for all** $p \in P$ **do**
3: **if** $\exists a \in A :\ a \prec_{\Theta_\epsilon} p$ **then**
4: CONTINUE ▷ *do not execute lines* 6 – 11
5: **end if**
6: **for all** $a \in A$ **do**
7: **if** $p \prec a$ **then**
8: $A := A \backslash \{a\}$
9: **end if**
10: **end for**
11: $A := A \cup \{p\}$
12: **end for**

Theorem 1 *(Schütze et al., 2008) Let an MOP $F : \mathbb{R}^n \to \mathbb{R}^k$ be given, where F is continuous, let $Q \subset \mathbb{R}^n$ be a compact set and $\epsilon \in \mathbb{R}^k_+$. Further let*

$$\forall x \in Q \text{ and } \forall \delta > 0 : \qquad P\left(\exists l \in \mathbb{N} \ :\ P_l \cap B_\delta(x) \cap Q \neq \emptyset\right) = 1 \tag{2.2}$$

Then an application of Algorithm 1, where $ArchiveUpdateEps1$ is used to update the archive, leads to a sequence of archives $A_l, l \in \mathbb{N}$, such that there exists with probability one an $l_0 \in \mathbb{N}$ such that A_l is an ϵ-approximate Pareto set for all $l \geq l_0$.

Theorem 2 *(Schütze et al., 2008) Let (MOP) be given and $Q \subset \mathbb{R}^n$ be compact, and let there be no weak Pareto points in $Q \backslash P_Q$. Further, let F be injective and*

$$\forall x \in Q \text{ and } \forall \delta > 0 : \qquad P\left(\exists l \in \mathbb{N} \ :\ P_l \cap B_\delta(x) \cap Q \neq \emptyset\right) = 1 \tag{2.3}$$

Then an application of Algorithm 1, where $ArchiveUpdateEps2$ is used to update the archive, leads to a sequence of archives $A_l, l \in \mathbb{N}$, where the following holds:

(a) There exists with probability one an $l_0 \in \mathbb{N}$ such that A_l is an ϵ-approximate Pareto set for all $l \geq l_0$.

Algorithm 3 $A := ArchiveUpdateEps2\ (P, A_0)$

```
A := A_0
for all p ∈ P do
    if ∄a ∈ A : a ≺_{Θ_ε} p then
        A := A ∪ {p}
    end if
    for all a ∈ A do
        if p ≺ a then
            A := A ∪ {p}\{a}
        end if
    end for
end for
```

(b)

$$\lim_{l\to\infty} dist(A_l, P_Q) = 0, \quad \textit{with probability one.}$$

3 The Sets of Interest

Motivated by the need for gap free Pareto front approximations and inspired by Definitions 4 and 5 we introduce here several objects.
First we define Δ_M-tight ϵ-(approximate) Pareto sets. A set $A_\epsilon \subset \mathbb{R}^n$ is such a set if (i) it is an ϵ-(approximate) Pareto set and if (ii) the maximal distance of a point in the Pareto front to the image of an archive element is not larger than a threshold Δ_M. Condition (i) refers to the approximation quality of A_ϵ in the sense of ϵ-dominance (i.e., measured in image space), and condition (ii) refers to the 'tightness' of the representation.

Definition 7 *Let* $\epsilon \in \mathbb{R}^k_+$.

(a) A set $A_\epsilon \subset \mathbb{R}^n$ *is called a* Δ_M-tight ϵ-approximate Pareto set *of (MOP) if* A_ϵ *is an* ϵ-approximate Pareto set *of (MOP) and*

$$dist(F(P_Q), F(A_\epsilon)) \leq \Delta_M. \tag{3.1}$$

(b) A set $A^*_\epsilon \subset \mathbb{R}^n$ *is called a* Δ_M-tight ϵ-Pareto set *if* F^*_ϵ *is an* ϵ-Pareto set *of (MOP) and* $d_H(F(P_Q), F(A^*_\epsilon)) \leq \Delta_M$.

A more descriptive way to express condition (3.1) is as follows: it is $dist(F(P_Q), F(A_\epsilon)) \leq \Delta_M$ if for every $y \in F(P_Q)$ there exists an element $a \in A_\epsilon$ such that $d_\infty(y, F(a)) \leq \Delta_M$. In other words, $F(P_Q)$ has to be contained in the 'box collection' C_{A_ϵ,Δ_M}, where

$$C_{A,\Delta} := \bigcup_{a\in A} B^\infty_\Delta(F(a)), \tag{3.2}$$

and $B^\infty_\Delta(x) := \{y \in \mathbb{R}^k \,:\, d_\infty(x, y) < \Delta\}$.
Since we are further interested in uniform approximations of the sets of interest, we add the uniformity level Δ_m to the objects in Definition 7.

Definition 8 *Let* $\epsilon \in \mathbb{R}^k_+$.

(a) A set $A_\epsilon \subset \mathbb{R}^n$ with $|A_\epsilon| \geq 2$ is called a (Δ_M, Δ_m)-tight ϵ-approximate Pareto set *if A_ϵ^* is an* Δ_M-tight ϵ-approximate Pareto set *of (MOP) and*

$$dist(F(a), F(A \backslash \{a\})) \geq \Delta_m, \quad \forall a \in A.$$

(b) A (Δ_M, Δ_m)-tight ϵ-Pareto set *is defined analogously.*

To get an impression about these objects, we consider the following MOP:

$$\begin{aligned} F : Q \subset \mathbb{R}^2 &\to \mathbb{R}^2 \\ F(x) &= x \end{aligned} \tag{3.3}$$

where

$$Q = [0,4]^2 \cap \{(x_1, x_2) \in \mathbb{R}^2 | x_1 \geq 4 - x_2\}. \tag{3.4}$$

The Pareto set is given by $P_Q = \{(x, 4-x) | x \in [0,4]\}$. Let

$$A_1 = \{a_1 = (0,4), a_2 = (1,3), a_3 = (2,2), a_4 = (3,1), a_5 = (4,0)\} \tag{3.5}$$

Since all elements $a_i \in P_Q$ and by distribution of the $a_i's$ the set A_1 forms a $(1/2, 1/2)$-Pareto set (see also Figure 2). Since further

$$\begin{aligned} dist(F(P_Q), F(A_1)) = \max_{p \in P_Q} \min_{a \in A_1} \|p - a\|_\infty = 1/2, \quad \text{and} \\ dist(F(a), F(A_1 \backslash \{a\})) \geq 1, \quad \forall a \in A_1, \end{aligned} \tag{3.6}$$

A_1 can be viewed as a $(1/2, 1)$-tight $(1/2, 1/2)$-Pareto set. Let $A_2 := A_1 \cup \{a_6 = (2.6, 1.6)\}$ (see Figure 2), then the uniformity level changes to $\delta = \|a_6 - a_3\|_\infty = 0.6$, and since $a_6 \notin P_Q$, A_2 forms a $(1/2, 0.6)$-tight $(1/2, 1/2)$-approximate Pareto set.
Note that not for every value of Δ_m, Δ_M, and ϵ the objects defined above exist for a given MOP, depending on the shape of the Pareto front. For instance, the domain Q of MOP (3.3) is included in $[0,4]^2$. Hence, if $\Delta_m > 4$ is chosen only one solution will be kept in the archive, and thus, no (Δ_M, Δ_m)-tight ϵ-Pareto set can be obtained for $\|\epsilon\|_\infty < 2$ since in this case at least two elements have to be kept. Similarly, no (Δ_M, Δ_m)-tight ϵ-Pareto set exists for $\Delta_M < \Delta_m/2$, independently of the value of ϵ.
Despite these potential problems, however, the values of ϵ, Δ_M and Δ_m have to be adjusted a priori. ϵ represents the (maximal) tolerable loss of a solution compared to an 'optimal' one, and is thus relatively easy to identify in a real world application (e.g., (Ruhe and Fruhwirt, 1990; White, 1998; Schütze et al., 2008)). Δ_M determines the maximal distance between two solutions and is thus depending on the search algorithm (e.g., on the step size of the local search procedure) and/or on the preference of the decision maker. In any case, $\Delta_M \geq \max_{i=1,\ldots,k} \epsilon_i$ can be chosen since two solutions which are closer together can, from a practical point of view, be regarded as equal. Finally, the value of the uniformity level Δ_m has to be chosen to be less or equal than $\min_{i=1,\ldots,k} \epsilon_i$, since otherwise it cannot be guaranteed that the limit archive set forms an ϵ-(approximate) Pareto set.

4 The Algorithms

In the following we propose two different strategies for archiving the solutions found by the algorithm and investigate some of their properties. The main focus is on the

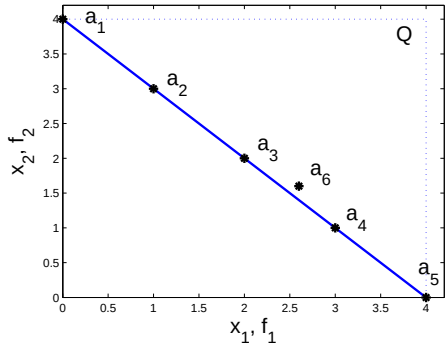


Figure 2: The set $A_1 := \{a_1, \ldots, a_5\}$ forms a $(1/2, 1)$-tight $(1/2, 1/2)$-Pareto set of MOP (3.3), and the set $A_2 := A_1 \cup \{a_6\}$ a $(1/2, 0.6)$-tight $(1/2, 1/2)$-approximate Pareto set.

limit behaviors of the sequence of archives which result by the use of these strategies under certain additional conditions on the generation process. It will be shown that the archives obtained by the first archiver $ArchiveUpdateTight1$ form with probability one after finitely many steps and under certain assumptions a $(\Delta, \Theta\epsilon_m)$-tight ϵ-approximate Pareto set of a given MOP, where Δ, ϵ and Θ are given and $\epsilon_m = \min_{i=1,\ldots,k} \epsilon_i$. In order to maintain the uniformity level $\Theta\epsilon_m$ a certain discretization in image space has to be done, which prevents the elements of the archives to converge toward the Pareto set. The second archiver, $ArchiveUpdateTight2$, allows for such a convergent behavior, but, in turn, the uniformity of the archive gets lost.

4.1 Archiver $ArchiveUpdateTight1$

Algorithm 4 $A := ArchiveUpdateTight1_{\epsilon,\tilde{\Delta}}\ (P, A_0)$

1: $A := A_0$
2: **for all** $p \in P$ **do**
3: **if** $(\exists a \in A : a \prec p)$ or $(\exists a_1 \in A :\ a_1 \prec_{\Theta\epsilon} p$ and $\exists a_2 \in A : d_\infty(F(a_2), F(p)) \leq \tilde{\Delta})$ **then**
4: CONTINUE ▷ *do not execute lines* 6 – 11
5: **end if**
6: **for all** $a \in A$ **do**
7: **if** $p \prec a$ **then**
8: $A := A \backslash \{a\}$
9: **end if**
10: **end for**
11: $A := A \cup \{p\}$
12: **end for**

First we consider the strategy $ArchiveUpdateTight1$, which is presented in Algorithm 4. Given $\epsilon \in \mathbb{R}^k_+$ and $\Delta \in \mathbb{R}_+$, the archiver accepts a new candidate solution $p \in Q$ if (i) there exists no element a of the current archive A such that a $\Theta\epsilon$-dominates p or (ii) there exists no $a \in A$ which dominates p *and* the distance of $F(a)$ to $F(p)$ for all elements $a \in A$ is larger than a threshold $\tilde{\Delta}$ (see line 3 of Algorithm 4 or (4.4)). If either

(i) or (ii) is true, p is added to the archive (line 11 of Algorithm 4) and all elements $a \in A \backslash \{p\}$ which are dominated by p are discarded from the archive (lines 7 to 10). Thus, only nondominated solutions are stored. The auxiliary variables $0 < \Theta < 1$ and $0 < \tilde{\Delta} < \Delta$ are required to guarantee convergence of the sequence of archives in the limit (see Theorem 4), but in practice they can be set to $\Theta = 1$ and $\tilde{\Delta} = \Delta$.
To prevent that the archives are unbounded (see discussion in Section 2), Algorithm 4 uses an 'exclusion strategy' which is based on ϵ-dominance: new candidate solutions x can only be added to the archive A if there does not already exist a solution $a \in A$ such that $a \prec_{\Theta\epsilon} x$ (second term in line 3 of Algorithm 4). This strategy allows on one hand to bound the number of entries of the archive (see Section 4.3) but by this it follows on the other hand that candidates x can be discarded even if they dominate elements $a \in A$.
This suboptimality in the selection mechanism of the archiver, however, is restricted to the value of ϵ which can be considered to be 'small'. The following result (Theorem 3) shows that the distance of the elements of the archive toward the observed Pareto front (i.e., the set of non-dominated solutions found by the generator so far) does not sum up. To be more precise, it will be shown that entries of the resulting archive form a $\Theta\epsilon$-approximate Pareto set of the set of points which have been discovered during the run of the algorithm. Thus, monotonicity of the sequence of archives is ensured and cycling or deterioration (Hanne, 1999) cannot occur.

Lemma 1 *Let $A_0, P \subset \mathbb{R}^n$ be finite sets, $\epsilon \in \mathbb{R}^k_+$, $0 < \Theta < 1$, $0 < \tilde{\Delta} < \Delta$, and $A := ArchiveUpdateTight1\ (P, A_0)$. Then the following holds:*

$$\forall x \in P \cup A_0 : \quad \exists a \in A : a \prec_{\Theta\epsilon} x.$$

Proof: Let $P = \{p_1, p_2, \ldots, p_l\}, l \in \mathbb{N}$. Without loss of generality we assume that all points p_i are considered in this ordering (i.e., in the for-loop in line 2 of Algorithm 4). Denote by A_i the resulting archive after p_i has been considered, and thus, $A = A_n$.
Let $x \in P \cup A_0$. There are two cases we have to distinguish.
Case A: $x \in A_0$. In that case the statement follows since points a are only discarded from the archive if in turn another point p with $p \prec a$ is inserted (see lines 7, 8 and 11 of Algorithm 4), and by the transitivity of $\prec$.
Case B: $x \in P$, i.e., there exists $j \leq l$ such that $x = p_j$, and thus, x is considered in the j-th step of the algorithm. In case x is added to A_j, the statement follows analogously to Case A since $A = ArchiveUpdateTight1\ (\{p_{j+1}, \ldots, p_l\}, A_j)$, and $x \in A_j$. If x is not added to A_j there are two possibilities (see line 3 of Algorithm 4): (i) there exists an $a \in A_j$ such that $a \prec x$, or (ii) there exists an $a_1 \in A_j$ with $a_1 \prec_{\Theta\epsilon} x$. In both cases x gets $\Theta\epsilon$-dominated by a point $\tilde{a} \in A_j$. Thus, again analogue to Case A, a point $a^* \in A$ exists which is equal to $\tilde{a}$ or dominating it, and thus $\Theta\epsilon$-dominating x. ∎

Theorem 3 *Let $l \in \mathbb{N}$, $\epsilon \in \mathbb{R}^k_+$, $0 < \Theta < 1$, and $0 < \tilde{\Delta} < \Delta$. Further, let $A_i, P_i, i = 0, \ldots, l$, be as defined in Algorithm 1, where $ArchiveUpdateTight1_{\epsilon,\tilde{\Delta}}$ is used to update the archive. Then*

$$\forall x \in \bigcup_{i=1}^{l} P_i : \quad \exists a \in A_l : a \prec_{\Theta\epsilon} x. \tag{4.1}$$

Proof: The proof is done via induction over l. For $l = 1$ we have

$$A_1 = ArchiveUpdateTight1_{\epsilon,\tilde{\Delta}}(P_0, A_0),$$

and the claim follows by Lemma 1. For the induction hypothesis, suppose the claim (4.1) is right for $l-1>1$. Let $x \in \bigcup_{i=1}^{l} P_i$. If $x \in \bigcup_{i=1}^{l-1} P_i$ there exists by induction hypothesis an element $a \in A_{l-1}$ such that $a \prec_{\Theta\epsilon} x$. Thus, a is also a member of the subsequent archive A_l or is replaced by an element $p \in P_l$ with $p \prec a$. In both cases there exists an element in A_l which $\Theta\epsilon$-dominates x. In case $x \in P_l$ the claim follows again by Lemma 1 since $A_l = ArchiveUpdateTight1_{\epsilon,\tilde{\Delta}}(P_l, A_{l-1})$, and the proof is complete. □

Next, we are interested in the limit behavior of the sequence A_i of archives within the use of a stochastic search procedure. To guarantee convergence, we have to make several (mild) assumptions on the model as well as on the process to generate new candidate solutions. Since we primarily address continuous optimization, we assume that F is continuous and that the domain Q is compact (e.g., an n-dimensional box defined by box constraints). By this it follows that the image $F(Q)$ is bounded which allows for a finite size Pareto set approximation in the sense of Definitions 4, 7 and 8 for every value of $\epsilon \in \mathbb{R}_+^k$. Note that this property is always true for discrete MOPs (i.e., $|Q| < \infty$). Further, we have to make the following assumption on the generation process (see also (Schütze, 2004; Schütze et al., 2008)):

$$\forall x \in Q \text{ and } \forall \delta > 0: \quad P\left(\exists l \in \mathbb{N} \,:\, P_l \cap B_\delta(x) \cap Q \neq \emptyset\right) = 1, \tag{4.2}$$

where $P(A)$ denotes the probability for event A. Assumption (4.2) says that every neighborhood $U \cap Q$ of every point gets 'visited' by $Generate()$ after finitely many steps with probability one. The following consideration shows that we cannot assume less: if (4.2) does not hold, there exists with probability one a point $x \in Q$ and a neighborhood $\tilde{U} = U \cap Q$ of x such that no candidate solution $p \in P_l$ lies in $\tilde{U}$ for all $l \in \mathbb{n}$. Thus, no convergence can be guaranteed since a part of the Pareto set can be contained in $\tilde{U}$ which is never 'visited'.
We point out that our results also hold for discrete models. In case the MOP is discrete, assumption (4.2) reads as

$$\forall x \in Q: \quad P\left(\exists l \in \mathbb{N} \,:\, x \in P_l\right) = 1, \tag{4.3}$$

which is e.g. fulfilled if *Generate()* is a homogeneous finite Markov chain with irreducible transition matrix ((Rudolph, 1998; Rudolph and Agapie, 2000)).
The next theorem shows that under these assumptions a generic stochastic search algorithm (Algorithm 1) coupled with the archiver in Algorithm 4 generates a sequence of archives which forms with probability one after finitely many steps a $(\Delta, \Theta\epsilon_m)$-tight ϵ-approximate Pareto set.

Theorem 4 *Let an MOP $F : \mathbb{R}^n \to \mathbb{R}^k$ be given, where F is continuous, let $Q \subset \mathbb{R}^n$ be a compact set and $\epsilon \in \mathbb{R}_+^k$. Let $\epsilon_m := \min_{i=1,\ldots,k} \epsilon_i$, $\epsilon_M := \max_{i=1,\ldots,k} \epsilon_i$, further let $\Delta, \tilde{\Delta} \in \mathbb{R}_+$ be given such that $\epsilon_M < \tilde{\Delta} < \Delta$, let $0 < \Theta < 1$, and let assumption (4.2) be fulfilled. Then an application of Algorithm 1, where $ArchiveUpdateTight1_{\epsilon,\tilde{\Delta}}$ is used to update the archive, leads to a sequence of archives, such that there exists with probability one an $l_1 \in \mathbb{N}$ such that A_l is a $(\Delta, \Theta\epsilon_m)$-tight ϵ-approximate Pareto set for all $l \geq l_1$.*

Proof: First, we turn our attention to the question of which elements are added to the archive. The crucial expression $\mathcal{E}$ (line 3 of Algorithm 4) reads as follows:

$$\underbrace{(\exists a \in A : a \prec p)}_{\mathcal{A}} \text{ or } (\underbrace{\exists a_1 \in A : \; a_1 \prec_{\Theta\epsilon} p}_{\mathcal{B}_1} \text{ and } \underbrace{\exists a_2 \in A : d_\infty(F(a_2), F(p)) \leq \tilde{\Delta}}_{\mathcal{B}_2})$$

Since $\neg\mathcal{E} = (\neg\mathcal{A}\,\text{and}\,\neg\mathcal{B}_1)\,\text{or}\,(\neg\mathcal{A}\,\text{and}\,\neg\mathcal{B}_2)$ and since $\neg\mathcal{B}_1$ implies $\neg\mathcal{A}$ it follows that points $p \in \mathbb{R}^n$ are added to a given archive A if (and only if) one of the two following expressions is true

$$\begin{aligned} &(\mathcal{E}_1) \quad \nexists a \in A : a \prec_{\Theta\epsilon} p, \text{ or} \\ &(\mathcal{E}_2) \quad \nexists a \in A : a \prec p \text{ and } \forall a \in A : d_\infty(F(a), F(p)) > \tilde{\Delta}. \end{aligned} \tag{4.4}$$

Now we are in the position to prove the theorem. By $\mathcal{E}_1$ it follows that all points, which are added by $ArchiveUpdateEps1$ to the archive are also added by $ArchiveUpdateTight1$. Thus, by Theorem 1 it follows that there exists with probability one a $l_0 \in \mathbb{N}$ such that A_l is an ϵ-approximate Pareto set w.r.t. $F\big|_Q$ for all $l \geq l_0$, since points $a \in A_l$ are only removed from the archive if in turn another point $\tilde{a}$ is added which dominates a (if $x \prec y$ and $y \prec_\epsilon z$ it follows that $x \prec_\epsilon z$).
It remains to show the 'tightness' of the limit archive. The uniformity level ϵ_m follows directly by an inductive argument and using the 'exclusion strategy' (4.4). This and the fact that $F(Q)$ is bounded is the reason that the size of the archive is bounded above for a given MOP by a number $n_0 = n_0(\epsilon, F(Q))$, which will be needed for further consideration (see Section 4.3 for more details).
As mentioned in Section 3 (see (3.2) and related discussion) the claim is right for an archive A_l if

$$F(P_Q) \subset C_{A_l,\Delta}. \tag{4.5}$$

Assume that A_l is an ϵ-approximate Pareto set for all $l \geq l_0$ and let $l \geq l_0$. By construction of $ArchiveUpdateTight1$ it follows that if $F(\mathcal{P}_Q) \subset C_{A_{l_1},\Delta}$ this inclusion holds for all $l \geq l_1$ since in this case no further point will be added to the archive (since the expressions $\mathcal{E}_1$ and $\mathcal{E}_2$ in (4.4) will be false for all further candidates). That is, it is sufficient to show the existence of such a number l_1. In the following we will do this by contradiction: first we show that by using $ArchiveUpdateTight1$ and under the assumptions made above only finitely many replacements can be done during the run of the algorithm. Then—under the assumption that there exists no number l_1 with the above property—we construct a contradiction by showing that infinitely many replacements have to be done during the run of the algorithm with the given setting.

Let a finite archive A_0 be given. If a point $p \in \mathbb{R}^n$ replaces a point $a \in A_0$ (see lines 8 and 11 of Algorithm 4) it follows by construction of $ArchiveUpdateTight1$ (see also (4.4)) that

$$\exists i \in 1, \ldots, k : \quad f_i(p) < f_i(a) - \Theta\epsilon_i. \tag{4.6}$$

Since the relation '$\prec$' is transitive, there exists for every $a \in A$ a 'history' of replaced points $a_i \in A_{l_i}$ where Equation (4.6) holds for a_i and a_{i-1}. Since $F(Q)$ is bounded there exist $l_i, u_i \in \mathbb{R}, i = 1, \ldots, k$, such that $F(Q) \subset [l_1, u_1] \times \ldots \times [l_k, u_k]$. After r replacements there exists at least one $a \in A_{l(r)}$ such that the length h of the history of a is at least $h \geq \lceil r/n_0 \rceil$, where n_0 is the maximal number of entries in the archive (see above). Denote by $a_0 \in A_0$ the root of the history. For a, a_0 it follows that

$$\exists i \in 1, \ldots, k : \quad f_i(a) < f_i(a_0) - s\Theta\epsilon_i,$$

where $s \geq \lceil h/k \rceil$. For $\tilde{s} > d_{max} := \Theta^{-1} \max_{i=1,\ldots,k} \frac{u_i - l_i}{\epsilon_i}$ (which is given for $\tilde{r} > n_0 k d_{max} + n_0 + 1$) we obtain a contradiction since in that case $f_i(a) < l_i$ and

thus $F(a) \notin F(Q)$. Hence it follows that there can be done only finitely many such replacements during the run of an algorithm.

Assume that such an l_1 as claimed above does not exist, that is, that $F(\mathcal{P}_Q) \not\subset C_{A_l,\Delta}$ for all $l \in \mathbb{N}$. Hence there exists a sequence of image points

$$y_i \in F(\mathcal{P}_Q) \backslash C_{A_i,\Delta} \quad \forall i \in \mathbb{N}. \tag{4.7}$$

Since $F(Q)$ is compact there exists an accumulation point $y^* \in F(\mathcal{P}_Q)$, that is, there exists a subsequence $\{i_j\}_{j\in\mathbb{N}}$ with

$$y_{i_j} \to y^* \text{ for } j \to \infty. \tag{4.8}$$

Since $y^* \in F(\mathcal{P}_Q)$ there exists a neighborhood U_1 of y^* such that the following holds

$$\forall (y, \tilde{y}) \in F(Q) \times U_1 : \; y \leq_p \tilde{y} \;\Rightarrow\; d_\infty(y, \tilde{y}) \leq \tilde{\Delta} \tag{4.9}$$

Let $\tilde{U}_1 := U_1 \cap B^\infty_{(\Delta-\tilde{\Delta})/2}(y^*)$. By (4.2) it follows that there exists with probability one an $l_1 \in \mathbb{N}$ and an $\tilde{x}_1 \in P_{l_0+l_1}$ with $\tilde{y}_1 = F(\tilde{x}_1) \in \tilde{U}_1$. By construction of $ArchiveUpdateTight1$ there exists an element $a_1 \in A_{l_0+l_1}$ such that $d_\infty(F(a_1), \tilde{y}_1) < \tilde{\Delta}$ (due to (4.4) there are three possibilities: $\mathcal{E}_2$ is false and thus there already exists an $a_1 \in A_{l_0+l_1}$ which (a) dominates $\tilde{x}$—in this case the claim follows with (4.9)—or (b) where $d_\infty(F(a_1), \tilde{y}_1) \leq \tilde{\Delta}$, or $\mathcal{E}_2$ is true and thus (c) $a_1 = \tilde{x}_1$ is added to the archive). Thus we have

$$d_\infty(F(a_1), \tilde{y}) \leq d_\infty(F(a_1), \tilde{y}_1) + d_\infty(\tilde{y}_1, \tilde{y}) < \tilde{\Delta} + 2\frac{\Delta - \tilde{\Delta}}{2} = \Delta \quad \forall \tilde{y} \in \tilde{U}_1. \tag{4.10}$$

By (4.7) and (4.8) there exist $j_1, \tilde{l}_1 \in \mathbb{N}$ with

$$y_{i_{j_1}} \in \tilde{U}_1 \backslash C_{l_0+l_1+\tilde{l}_1,\Delta}.$$

Since by (4.10) it holds that $d_\infty(y_{i_{j_1}}, F(a_1)) < \Delta$ it follows that $a_1 \notin A_{l_0+l_1+\tilde{l}_1}$, which is only possible via a replacement in Algorithm 4 (lines 8 and 11).
In an analogous way a sequence $\{a_i\}_{i\in\mathbb{N}}$ of elements can be constructed which have to be replaced by other elements. Since this leads to a sequence of infinitely many replacements this is a contradiction to the assumption, and the proof is complete. □

Note that the 'exclusion strategy' (4.4) prevents convergence of the elements of the archives toward the Pareto set. This is due to the fact that for points $x \in Q$ which are 'nearly' optimal, the set of points in Q which (i) dominate x and (ii) are not ϵ-dominated by x can be empty. Such nearly optimal archive entries will hence never be replaced by other better solutions, and thus, the elements in the archive stop improving when using Algorithm 4 at a certain stage, depending on the value of ϵ. This leads directly to the next archiver.

4.2 Archiver $ArchiveUpdateTight2$

The second archiving strategy we consider here, $ArchiveUpdateTight2$ which is shown in Algorithm 5, overcomes the problem stated above by changing the criterion to accept a candidate solution (which is in fact the only difference between Algorithms 4

and 5). In turn, by using Algorithm 5 the uniformity of the archive solutions can not be guaranteed any more and the (theoretical and pessimistic) upper bounds on such archives increases by one order of magnitude compared to the first archiving strategy (see Section 4.3). The difference of the two archivers $ArchiveUpdateTight1$ and $ArchiveUpdateTight2$ is the strategy to accept a candidate solution $p \in P$. Given an archive A_0 $ArchiveUpdateTight2$ accepts p if (i) either term $\mathcal{E}_1$ or $\mathcal{E}_2$ of (4.4) is true (line 3 of Algorithm 5) or (ii) if there exists an element $a \in A_0$ which is dominated by p (line 8 of Algorithm 5). In case (i) also $ArchiveUpdateTight1$ accepts the candidate solution, the difference of both archivers is case (ii), which is not considered in $ArchiveUpdateTight1$.

Under the same assumptions as made above for Algorithm 4 the following theorems show that the same monotonicity result on the approximation quality can be obtained, and that the distance $dist(A_l, P_Q)$ of the archives A_l to the Pareto set P_Q vanishes for $l \to \infty$ (the elements of the archive 'converge' to the Pareto set). Thus, if the limit archive exists (the sequence $|A_l|$ of the magnitudes of the archives is not necessarily converging), this set forms a Δ-tight ϵ-Pareto set. Though we have to assume in Theorem 6 that F has to be injective to guarantee the convergence, this property is in fact not relevant in practice (see e.g., Section 6.2).

Algorithm 5 $A := ArchiveUpdateTight2_{\epsilon,\tilde{\Delta}}\ (P, A_0)$

1: $A := A_0$
2: **for all** $p \in P$ **do**
3: **if** $\mathcal{E}_1$ is true or $\mathcal{E}_2$ is true **then** ▷ see (4.4)
4: $A := A \cup \{p\}$
5: **end if**
6: **for all** $a \in A$ **do**
7: **if** $p \prec a$ **then**
8: $A := A \cup \{p\} \backslash \{a\}$
9: **end if**
10: **end for**
11: **end for**

Theorem 5 *Let $l \in \mathbb{N}$, $\epsilon \in \mathbb{R}^k_+$, $0 < \Theta < 1$, and $0 < \tilde{\Delta} < \Delta$. Further, let $A_i, P_i, i = 0, \ldots, l$, be as defined in Algorithm 1, where $ArchiveUpdateTight2_{\epsilon,\tilde{\Delta}}$ is used to update the archive. Then*

$$\forall x \in \bigcup_{i=1}^{l} P_i : \quad \exists a \in A_l : a \prec_{\Theta\epsilon} x. \tag{4.11}$$

Proof: Analogue to proof of Theorem 3. ∎

Theorem 6 *Let (MOP) be given and $Q \subset \mathbb{R}^n$ be compact, and let there be no weak Pareto points in $Q \backslash \mathcal{P}_Q$. Further, let F be injective and let assumption (4.2) be fulfilled. Then an application of Algorithm 1, where $ArchiveUpdateTight2_{\epsilon,\tilde{\Delta}}$ is used to update the archive, leads to a sequence of archives $A_l, l \in \mathbb{N}$, where the following holds:*

(a) There exists with probability one a $l_1 \in \mathbb{N}$ such that A_l is a Δ-tight ϵ-approximate Pareto set w.r.t. $F\big|_Q$ for all $l \geq l_1$.

(b)

$$\lim_{l\to\infty} dist(A_l, P_Q) = 0, \quad \textit{with probability one.}$$

Proof: All parts of the proof are analogue to parts in proofs of Theorem 2 and Theorem 4. ∎

4.3 Bounds on the Limit Archive Sizes

Since we are aiming for finite size representations of the Pareto front, the bounds of the magnitudes of the archives obtained by the new archiving strategies are of particular interest which we address in this section.

Upper Bounds The upper bounds on the archive sizes which result by the novel archiving strategies can be derived in analogy to the bounds for the archivers presented in (Schütze et al., 2008):
The maximal archive size maintained by $ArchiveUpdateTight1$ is the same as for $ArchiveUpdateEps1$ and given by

$$|A_l| \leq \sum_{i=1}^{k} \prod_{\substack{j=1\\ j\neq i}}^{k} \left\lceil \frac{M_j - m_j}{\Theta\epsilon_j} \right\rceil, \tag{4.12}$$

where $m_i = \min_{x\in Q} f_i(x)$, $M_i = \max_{x\in Q} f_i(x), 1 \leq i \leq k$, and $|A_0| = 1$. Note that the magnitude can only be influenced by the value of ϵ. The existence of this bound is due to the 'exclusion strategy' (4.4), which makes it possible that the sequence of archives converges after finitely many steps. On the other hand, exactly this feature prevents that we can guarantee $dist(F(A), F(\mathcal{P}_Q))$ and thus $d_H(F(A), F(\mathcal{P}_Q))$ to be small (say $\leq \Delta$), as the following example shows (compare to Figure 3): assume that the elements a_3, a_2, a_1 are inserted into the archive in this order. By construction of $ArchiveUpdateTight1$, these points will not be removed in the subsequent steps since there exists no point p with $F(p) \in F(A)\backslash C_{A,\Delta}$ which dominates a_i, $i \in \{1, 2, 3\}$. In such a manner an example can be constructed with $dist(F(A), F(\mathcal{P}_Q)) = max_{i=1,..,k}(M_i - m_i)$. However, this (bad) theoretical value has never been observed in computations.

The maximal archive size obtained by $ArchiveUpdateTight2$ is equal to the one obtained by $ArchiveUpdateEps2$:

$$|A_l| \leq \prod_{i=1}^{k} \left\lceil \frac{M_i - m_i}{\Theta\epsilon_i} \right\rceil,$$

where m_i, M_i are as defined above and $|A_0| = 1$.

Lower Bounds The lower bound of $|A_\infty|$ for both new archiving strategies is obviously given by 1. For this, consider e.g. $f_1 = f_2 = \ldots = f_k$ to be a convex function which takes its (unique) minimum inside Q which leads to $|P_Q| = 1$. Though desired, it is hardly possible to provide meaningful lower bounds for general MOPs since (a) the archive size is in practice mainly determined by the value of ϵ, in particular when the

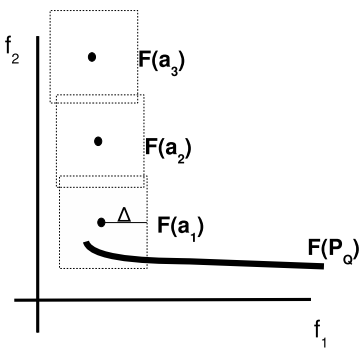


Figure 3: Possible example of a set which was generated by $ArchiveUpdateTight1$ with $dist(F(A), F(\mathcal{P}_Q)) > \Delta$.

entries of ϵ are much smaller than Δ, and since (b) the Pareto front can fall into different connected components. However, if this is not the case, one can obtain the following result in the bi-objective case, which we state without the (obvious) proof.

Proposition 1 *Let $\tilde{m}_i = \min_{x \in \mathcal{P}_Q} f_i(x)$ and $\tilde{M}_i = \max_{x \in \mathcal{P}_Q} f_i(x), i = 1, 2$, and let $F(P_Q)$ be connected. Then, when using $ArchiveUpdateTight1_{\epsilon,\tilde{\Delta}}$ or $ArchiveUpdateTight2_{\epsilon,\tilde{\Delta}}$, the archive size maintained in Algorithm 1 for the limit archive is bounded by*

$$|A_\infty| \geq \max_{i=1,\ldots,k} \left\lceil \frac{\tilde{M}_i - \tilde{m}_i}{2\Delta} \right\rceil \tag{4.13}$$

An analogous statement for $k > 2$—e.g., by estimating the Pareto front by a k-Simplex in objective space where the vertices are the minima of the objectives—, however, does not hold since the $(k-1)$-dimensional volume of the Pareto front can be arbitrarily small.

5 Integration into Iterative Search Methods

Since this work deals with the design of archiving strategies, the question that naturally arises is how these methods can be integrated *efficiently* into an iterative stochastic search process such as a MOEA, which we address here.
One obvious benefit of a MOEA which is equipped with an (external) archive compared to a 'classical' MOEA with fixed population size is certainly given by the convergence properties of the archiver (see Thms. 4 and 6 of this paper, or the results of the works discussed in Section 1). Another possible advantage is that the globality of the search will be increased: in case the external archive is considered for the mating pool, the number of 'well-converged' and 'well-distributed' parent solutions is increased leading to a potentially more thorough search around the Pareto set. In fact, in (Deb et al., 2005), where a MOEA with an external archive has been studied, it has been observed that this algorithm is

'successful in finding well-converged and well-distributed solutions with a much smaller computational effort than a number of state-of-the-art MOEAs'.

This algorithm, ϵ-MOEA, can be viewed as a possible prototype for the integration of an archiving strategy into the evolutionary search process. The procedure is as follows: the algorithm contains a population (P_l in iteration step l) and an archive (A_l in step l). A new offspring o is created by crossover of an archive solution $a_0 \in A_l$ and an individual $p \in P_l$. Both sets A_l and P_l are then updated by o (following different strategies). This process is repeated until a prescribed termination criterion is fulfilled. Apparently, *any* archiving strategy for the update of the external archive A_l can be used, including for instance the ones proposed in this paper.
However, apart from this universally applicable prototype for the integration of the archivers into a MOEA, there is a particular property of the two strategies proposed above which needs special attention: variations of existing archive entries which are too small will be discarded from the archiver for the subsequent archive. Due to the 'exclusion strategy' (4.4) used in Algorithms 4 and 5 there exists for every solution a from a given archive A a neighborhood U_a of a such that *every* point $u \in U_a$ will be rejected from the archive by a. Regarding this, it has to be noted that not every crossover (as well as mutation) strategy is suitable for a coupling with $ArchiveUpdateTight1$ or $ArchiveUpdateTight2$. For instance, there is a potential conflict in terms of efficiency when using e.g. the simulated binary crossover operator (SBX) and the polynomial mutation, probably the most commonly used operators for crossover and mutation (Deb, 2001). When $a_0 \in A_l$ and $p \in P_l$ are used e.g. for crossover[2], the probability is relatively high that the offspring o is contained in U_{a_0} (or in U_p which leads to the same problem) since the probability density has a peak at a_0 (and a second peak at p). As it has been demonstrated on numerous benchmark and real world problems, SBX and polynomial mutation are highly valuable for the evolution of P_l. Thus, the operators will also be beneficial for the evolution of A_l, but probably only in the long run. In order to obtain a greedy strategy for the evolution of A_l, which is desired for fast convergence, both SBX and polynomial mutation do not seem to be well suited since in case fast convergence is sought a minimal distance of the offspring o to a_0 is required.
In case gradient information is available (which is for instance assumed in many recent studies dealing with memetic strategies, e.g., (Brown and Smith, 2005; Bosman and de Jong, 2005; Sharma et al., 2007; Harada et al., 2007; Schütze et al., 2008)), the minimal distance of a_0 and o such that $o \notin U_{a_0}$—i.e., that o is possibly accepted by the archiver—can be estimated as follows: since $\tilde{\Delta} < \Delta$ we can assume that we are (ideally) interested in a point o in the neighborhood of a_0 such that

$$\|F(a_0) - F(o)\|_\infty = \Delta, \tag{5.1}$$

since such point will at least not be discarded due to the exclusion strategies of the archers proposed above. In case F is Lipschitz continuous there exists an $L \geq 0$ such that

$$\|F(x) - F(y)\|_\infty \leq L\|x - y\|_\infty, \quad \forall x, y \in Q. \tag{5.2}$$

Since we are heading in this paper for gap free approximations we can assume that Δ is 'small', and thus, that a_0 and o are close to each other. Hence, it is sufficient to estimate the Lipschitz constant L *locally* around a_0, which can be done by

$$L_{a_0} := \|DF(a_0)\|_\infty = \max_{i=1,\ldots,k} \|\nabla f_i(a_0)\|_1, \tag{5.3}$$

[2] An analogue statement holds for the polynomial mutation.

in case the gradients are available, where $DF(a_0) \in \mathbb{R}^{m \times n}$ denotes the Jacobian and $\nabla f_i(a_0) \in \mathbb{R}^n$, $i = 1, \ldots, m$, the objectives' gradients. An alternative way to approximate L_{a_0} without using gradient information is e.g.

$$\tilde{L}_{a_0} := \frac{\|F(a_0) - F(\tilde{a}_0)\|}{\|a_0 - \tilde{a}_0\|}, \tag{5.4}$$

where $\tilde{a}_0$ is close to a_0. Approximation (5.4) is certainly less accurate than (5.3), but we think that accuracy does not play an important role in this context.
Using such a local approximation L_{a_0} (e.g., (5.3) or (5.4)), the distance of a_0 and an offspring o which satisfies (5.1) can be estimated by

$$\|a_0 - o\|_\infty \approx \frac{\Delta}{L_{a_0}} \tag{5.5}$$

Using this, a neighborhood search (or mutation) can e.g. be realized as follows: assume that the difference of $F(a_0)$ and $F(o)$ should be in the range $[\Delta, \lambda\Delta]$, where $\lambda > 1$, then it follows analogue to (5.5) that $\|a_0 - o\| \in [\Delta/L_{a_0}, \lambda\Delta/L_{a_0}]$, and thus, the offspring can e.g. be taken uniformly at random from the pierced sphere $B^\infty_{\frac{\lambda\Delta}{L_{a_0}}}(a_0) \backslash B^\infty_{\frac{\Delta}{L_{a_0}}}(a_0)$:

(1) compute L_{a_0}

(2) choose $d \in B_1(a_0)$ and $h \in [\Delta/L_{a_0}, \lambda\Delta/L_{a_0}]$ uniformly at random

(3) set $o := a_0 + h\frac{d}{\|d\|_\infty}$

We do, however, not investigate the efficiency of this mutation strategy here since this would go beyond the scope of this paper. Instead, we give evidence that the step size control (5.5) can be used in a particular local search strategy leading to a possible efficient hybridization of this method with a MOEA which is equipped with the archivers presented in Algorithms 4 and 5 (which was one motivation for the need of gap free approximations).
To be more precise, we want to demonstrate that in the underlying context a hybridization with multi-objective continuation methods (e.g., (Hillermeier, 2001), (Schütze et al., 2005)) could be advantageous since both the concept of ϵ-dominance as well as the tightness can be directly integrated into it (see also (Schütze, 2004) for a similar study).

The basic idea of multi-objective continuation methods is, roughly speaking, to move *along* the set of (local) Pareto points. To be more precise, in the course of the algorithm one is faced with the following setting: given a (locally optimal) solution $x_0 \in Q$ and a *search direction* $v \in \mathbb{R}^n$ with $\|v\| = 1$ (obtained via linearization of the solution set at x_0), the task is to find a 'suitable' step size $h_0 \in \mathbb{R}_+$ for the next guess $y_0 = x_0 + h_0 v$. Motivated by previous considerations one can e.g. ask for a step size h_0 such that

$$\|F(x_0) - F(y_0)\|_\infty = \Delta, \tag{5.6}$$

where Δ is the tightness value taken in Algorithms 4 and 5 (alternatively, Δ in (5.6) can be replaced e.g., by $\|\epsilon\|_\infty$ for Algorithms 2 and 3). Following the discussion made

above and using a local approximation of L_{x_0} as (5.3) or (5.4), one can use the following step size control:

$$h_0 = \frac{\Delta}{L_{x_0}} \tag{5.7}$$

Note that this estimation only holds for small values of Δ since the L_{x_0} is a local approximation. If Δ is too large, L_{x_0} can not serve as a suitable Lipschitz estimation, and the value of h_0 may not be suitable. The following two examples show, however, that the control (5.7) can be beneficial for small values of Δ. Such a step size control would be interesting for hybrids of continuation methods with EMO strategies (e.g., (Harada et al., 2007; Schütze et al., 2008)) since in this case the archivers presented above could efficiently be integrated into the entire algorithm as external archives, which we want to demonstrate on the following two examples.

5.1 Example A

In order to understand the possible impact of the discussion made above on the continuation methods, we first apply the step size control on an academic example:

$$\begin{aligned} F &: \mathbb{R}^2 \to \mathbb{R}^2 \\ F(x) &= \begin{pmatrix} (x_1-1)^4 + (x_2-1)^4 \\ (x_1+1)^2 + (x_2+1)^2 \end{pmatrix} \end{aligned} \tag{5.8}$$

The Pareto set of MOP (5.8) is given by

$$\mathcal{P} = \left\{ \begin{pmatrix} x \\ x \end{pmatrix} : x \in [-1,1] \right\}.$$

Figure 4 shows two different discretizations of $\mathcal{P}$ and $F(\mathcal{P})$. In Figure 4 (a) the Pareto set is approximated by points $x_i, i = 1, \ldots, N$, which are placed equidistant in parameter space:

$$x_i = \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \frac{2i}{N} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

Next, the Pareto set was discretized using the adaptive step size control which is proposed above:

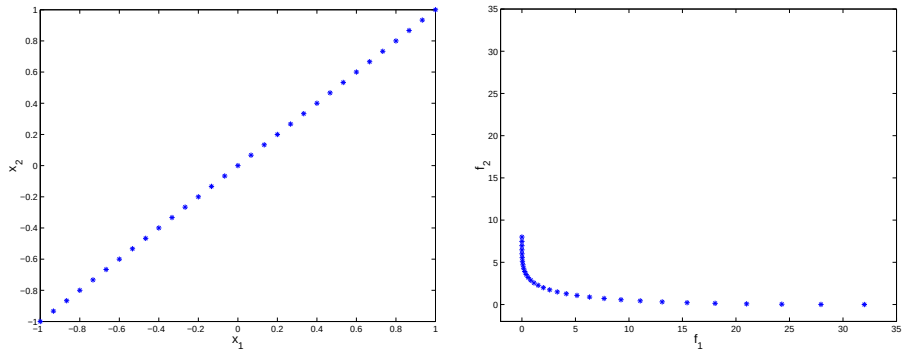
$$x_0 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad x_{i+1} = x_i + h_i \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

where h_i is taken from (5.7) and $v_i = (1/\sqrt{2}, 1/\sqrt{2})^T$ was chosen as the search direction. Figure 4 (b) shows the discretization points x_i for $\epsilon = (1,1)$, $\Delta = 1$, and $\Theta = 0.99$ yielding a satisfying distribution of the solutions on the Pareto front.

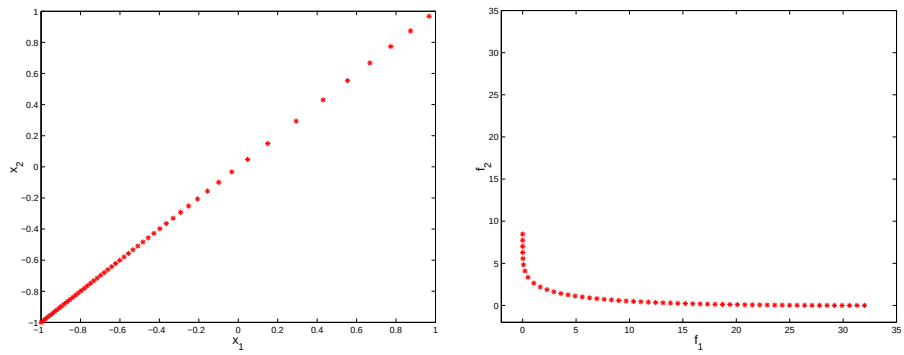
5.2 Example B

Next we consider the following MOP:

$$\begin{aligned} f_1, f_2 &: \mathbb{R}^n \to \mathbb{R} \\ f_i(x) &= \sum_{\substack{j=1 \\ j \neq i}}^{n} (x_j - a_j^i)^2 + (x_i - a_i^i)^4, \end{aligned} \tag{5.9}$$



(a) fixed step size



(b) adaptive step size

Figure 4: Discretizations of the Pareto set of MOP (5.8) with (a) fixed step size and (b) adaptive step size control.

where

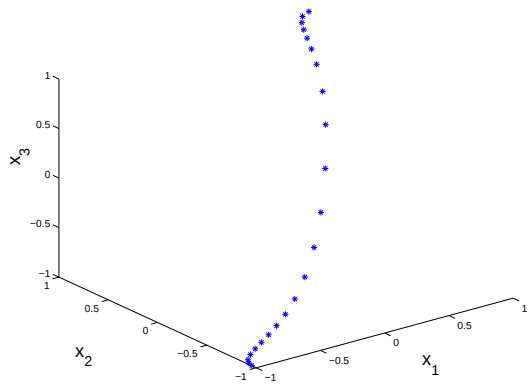
$$\begin{aligned} a^1 &= (1,1,1,1,\ldots) && \in \mathbb{R}^n \\ a^2 &= (-1,-1,-1,-1,\ldots) && \in \mathbb{R}^n, \end{aligned}$$

In Figures 5 and 6 some numerical results are presented, where we have used the continuation method proposed in (Schütze et al., 2005). To be more precise, we have applied the step size control on the distance between the current solution and the *predictor*, since this point mainly determines the distance of two solutions.
Figure 5 shows the result for $n = 3$, $\epsilon = (2,2)$, and $\Delta = 2$. In total, 23 solutions were obtained. This fits quite well with the bound which is given in Section 4.3 when choosing $\Theta = 1$ and when replacing M_i by the maximal value of f_i on the Pareto set, $\tilde{M}_i := \max_{x \in P_Q} f_i(x) = 24$, $i = 1,2$, and $m_i = 0$. In that case we obtain:

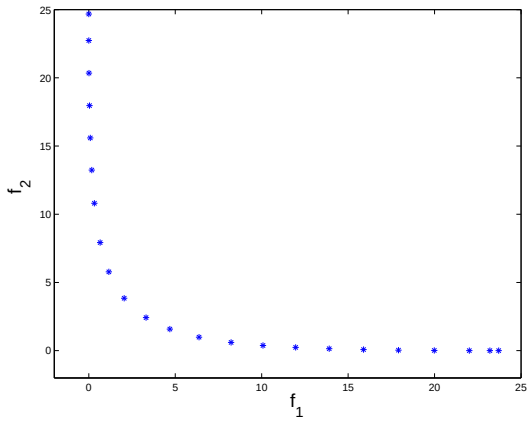
$$|A_i| \leq \left\lceil \frac{24-0}{2} \right\rceil + \left\lceil \frac{24-0}{2} \right\rceil = 24. \tag{5.10}$$

This example shows the significant difference between the archivers $ArchiveUpdateEpsi$, $i = 1,2$ which are 'merely' based on ϵ-dominance and the novel

strategies which aim for a tightness of the approximation: if points with images near to the middle of the Pareto front in Figure 5 (b) are inserted into the archive, e.g. the points m_1 with $F(m_1) = (5, 2)$ and m_2 with $F(m_1) = (2, 5)$, then no more points with images at the ends of the Pareto curve will be added further on since these are all ϵ-dominated either by m_1 or m_2. The resulting approximation would form an ϵ-approximate Pareto set, but would apparently not 'describe' the appropriate Pareto set graphically. This would change, however, for one of the novel archiving strategies, which we investigate more in detail in the next section.



(a) Parameter space

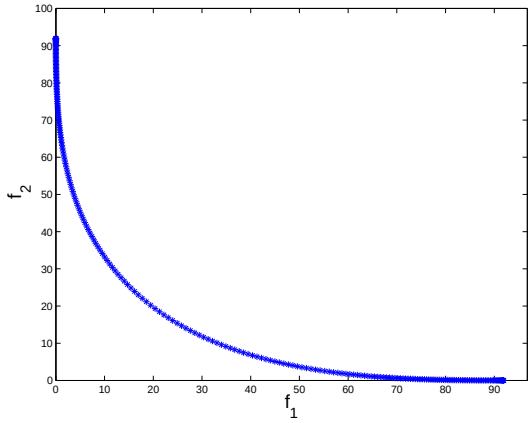


(b) Image space

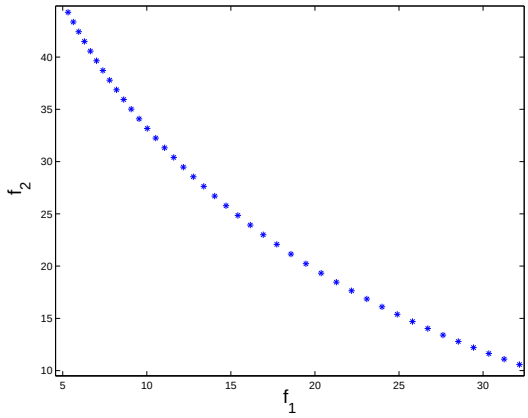
Figure 5: Result of the continuation method with step size control on MOP (5.9) for $n = 3$ in parameter space and image space.

6 Numerical Results

Here we make a comparative study on three test problems in order to illustrate the effect of the different archiving strategies. For the subsequent comparisons we have used the following archiving strategies:



(a) All solutions in objective space



(b) Zoom of (a)

Figure 6: Result of the continuation method with step size control on MOP (5.9) for $n = 20$ in image space: all solutions (a) and zoom (b).

(ND)	$ArchiveUpdateND$,
(Eps1)	$ArchiveUpdateEps1$,
(Tight1)	$ArchiveUpdateTight1$, and
(Tight2)	$ArchiveUpdateTight2$,

where $ArchiveUpdateND$ is the archiver which stores all nondominated solutions, i.e.,

$$ArchiveUpdateND(P, A_0) := \{x \in P \cup A_0 : y \not\prec x \ \forall y \in P \cup A_0\}.$$

For an investigation of the convergence properties of $ArchiveUpdateND$ we refer to (Schütze et al., 2008).
To obtain a fair comparison of the different archivers we have decided to take a random search operator for the generation process (the same sequence of points for all settings). The computations have been done on an Intel Xeon 3.2 GHz processor. An implementation of all the archiving strategies discussed in this work including these examples can be found in

Table 1: Comparison of the magnitudes of the final archive ($|A_N|$, rounded) and the corresponding update times (T, in seconds) for different archiving strategies and for MOP (5.9). We have taken the average result of 100 test runs.

	ND	Eps_1	Tight_1
$\|A_N\|$	267	27	34
T	36.46	0.29	0.36

`http://paradiseo.gforge.inria.fr/index.php?n=Paradiseo.Features`

6.1 Example 1

First, we compare the first three different archving strategies on MOP (5.9) from the previous example. We have taken $N = 200,000$ randomly chosen points in $Q = [-1.5, 1.5]^3$, and the values $\epsilon = (1, 1)$ and $\Delta = 2$. The set obtained by $ArchiveUpdateEps1$ forms probably (or is near to) an ϵ-approximate Pareto set, but reveals gaps, which is not the case in Figure 7 (c), where $ArchiveUpdateTight1$ has been used, and where the solutions are much more regularly distributed. The 'tightest' approximation in this case study is certainly obtained when all nondominated points are kept in the archive (see Figure 7 (b)). However, in that case the time which had to be spent to update the archive[3] was huge compared to the two other strategies (see Table 1).

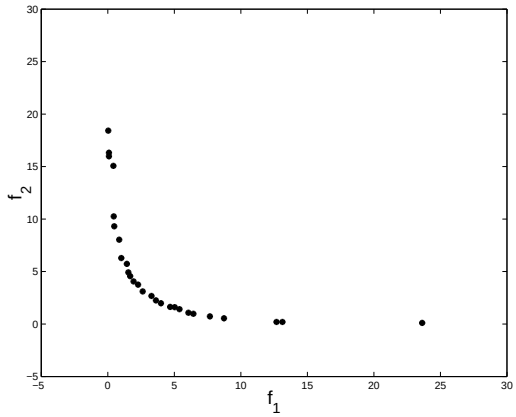
6.2 Example 2

Next we consider the following parameter dependent MOP (Witting and von Molo, 2006):

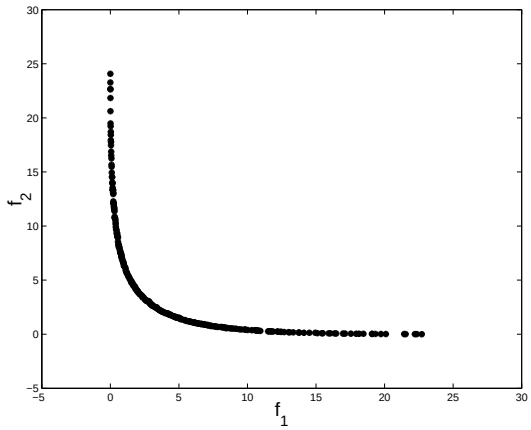
$$\begin{aligned} & f_1, f_2 : \mathbb{R}^2 \to \mathbb{R} \\ & f_1(x,y) = \frac{1}{2}(\sqrt{1+(x+y)^2} + \sqrt{1+(x-y)^2} + x - y) + \lambda \cdot e^{-(x-y)^2} \\ & f_2(x,y) = \frac{1}{2}(\sqrt{1+(x+y)^2} + \sqrt{1+(x-y)^2} - x + y) + \lambda \cdot e^{-(x-y)^2} \end{aligned} \tag{6.1}$$

Figure 8 shows examples for resulting limit sets. Hereby we have taken $N = 10,000$ randomly chosen points in $Q = [-1.5, 1.5]^2$ and $\lambda = 0.85$ for the value of the additional parameter, as well as the values $\epsilon = (0.1, 0.1)$, and $\Delta = 0.1$ for the archiving strategies (using $\tilde{\Delta} = \Delta$ and $\Theta = 1$). Also in this case the result of $ArchiveUpdateEps1$ reveals gaps in the approximation, which does not occur when using $ArchiveUpdateTight1$ or $ArchiveUpdateTight2$. Note that the differences of the latter two solutions are subtle as expected: both approximations are similar, but the final solutions obtained by $ArchiveUpdateTight2$ have converged better. The difference in the magnitudes of the archive sizes in this example (and in all other examples examined by the authors) do not differ significantly, though the theoretical upper bounds for both strategies do. Also in this example, $ArchiveUpdateND$ delivers the 'tightest' approximation containing by

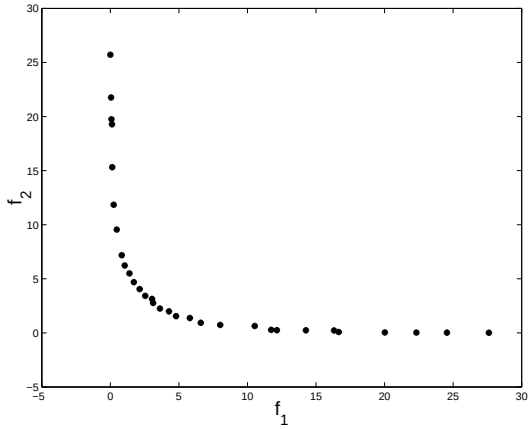
[3] The elements of the all archives were stored using a linear list.



(a) $ArchiveUpdateEps1$



(b) nondominated solutions



(c) $ArchiveUpdateTight1$

Figure 7: Three limit achives for MOP (5.9) obtained by different archiving strategies.

Table 2: Distances of the images of the final archives to the approximation $F(A_P)$ of the Pareto front of MOP (6.1) (averaged over 100 test runs). The Hausdorff distances of the solutions obtained by $ArchiveUpdateTight1$ and $ArchiveUpdateTight2$ are nearly to the optimum which is given here by $\Delta = 0.1$.

	Eps1$_1$	Tight1	Tight2	ND
$dist(F(A_{final}), F(A_{\mathcal{P}}))$	0.1029	0.1029	0.0079	0.0135
$dist(F(A_{\mathcal{P}}), F(A_{final}))$	0.9209	0.1092	0.1042	0.0290
$d_H(F(A_{final}), F(A_{\mathcal{P}}))$	0.9209	0.1092	0.1042	0.0290

far the most elements which results in a longer running time for the update process. Since the Pareto set of this MOP is given analytically by

$$\mathcal{P} = \left\{ \begin{pmatrix} & x \\ - & x \end{pmatrix} : x \in [-1.5, 1.5] \right\} \tag{6.2}$$

this allows us to have a closer look at the approximation qualities of the obtained solutions. Figure 8 shows the distances between the final archives and the Pareto front. For the latter we have used the following discretization of (6.2):

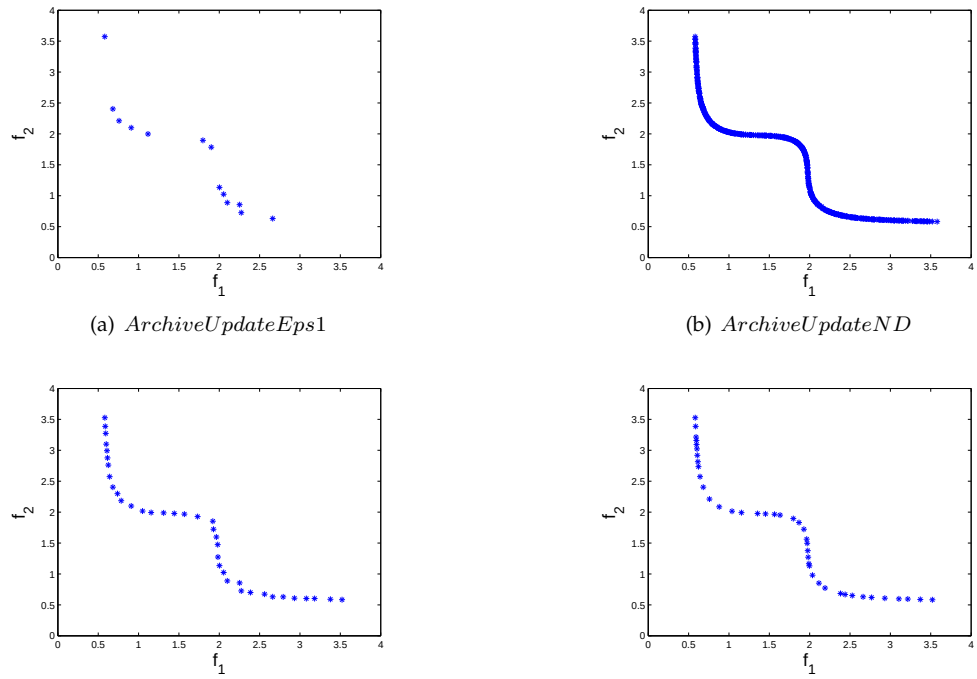
$$A_{\mathcal{P}} := \left\{ \begin{pmatrix} & x_i \\ - & x_i \end{pmatrix} : x_i = -1.5 + 3i/500,\ i = 0, \ldots, 500 \right\} \tag{6.3}$$

The Hausdorff distances of the two final archives of the novel archivers are close to 0.1 which is the optimum in this case since this is the chosen value of Δ. The main difference between the two solutions is the value of $dist(F(A_{final}), F(A_{\mathcal{P}}))$ which means that the elements obtained by $ArchiveUpdateTight2$ are nearer to the Pareto front. The gaps which can be observed when using $ArchiveUpdateEps1$ (Figure 8 (a)) are reflected by its relatively large value of $dist(F(A_{final}), F(A_{\mathcal{P}}))$ in Table 2. Best values are obtained by $ArchiveUpdateND$, but this has to be 'paid' by the much larger amount of elements in the archive. Further, it can be argued that approximations with Hausdorff distances less than 0.1 (the values of Δ and $\epsilon_i, i = 1, 2$) to the Pareto front are not needed.

6.3 Example 3

Finally we consider a 3-objective model. For this, we extend MOP (5.9) by adding a third objective which is analogue to f_1 and f_2, and choose $a_3 = (1, -1, 1, -1, \ldots) \in \mathbb{R}^n$. In order to demonstrate one possible benefit of ϵ-dominance based archivers against the classical archiver $ArchiveUpdateND$, which stores all nondominated solutions, we fix in this example the running time of the different algorithms (i.e., in our case the stochastic search algorithm coupled with the different archiver). Figure 9 shows one comparative result for $n = 10$ and where the running time was fixed to 5 minutes. Denote by A the final archive when $ArchiveUpdateND$ was used, and by B the resulting archive coming from $ArchiveUpdateTight1$. The magnitudes are $|A| = 914$ and $|B| = 529$. It can be observed that compared to A the spread of the solutions of B is much better while a larger region of the image space is 'covered' though its magnitude is less. To measure the approximation quality we use the *epsilon indicator* (Zitzler et al., 2003), where $I_\epsilon(A, B)$ gives the smallest value of $\bar{\epsilon} \in \mathbb{R}$ such that A is an ϵ-approximate Pareto set of B where $\epsilon = (\bar{\epsilon}, \ldots, \bar{\epsilon})$, i.e.,

$$I_\epsilon(A, B) := \min\{\bar{\epsilon} \in \mathbb{R} \mid \forall b \in B\ \exists a \in A :\ a \prec_\epsilon b\}. \tag{6.4}$$



(a) $ArchiveUpdateEps1$ (b) $ArchiveUpdateND$

(c) $ArchiveUpdateTight1$ (d) $ArchiveUpdateTight2$

Figure 8: Four limit achives for MOP (6.1) obtained by different archiving strategies. The magnitudes are $|A_{eps1}| = 13$, $|A_{ND}| = 511$, $|A_{tight1}| = 37$, $|A_{tight2}| = 41$. See Table 2 for their approximation qualities.

In our case, we obtain for A and B

$$I_\epsilon(A, B) = 0.9624 \quad \text{and} \quad I_\epsilon(B, A) = 0.85775, \tag{6.5}$$

indicating that B is a (slightly) better approximation of the Pareto front than A. If the *relative* ϵ-dominance

$$x \prec_\epsilon^{\text{rel}} y \Leftrightarrow F(x) \leq_p (1 + \epsilon)F(y) \tag{6.6}$$

is used for the epsilon indicator, the difference of the approximation qualities gets more significant:

$$I_\epsilon^{rel}(A, B) = 1.082 \quad \text{and} \quad I_\epsilon^{rel}(B, A) = 0.183. \tag{6.7}$$

These values can be interpreted as follows: if the objective values of A are scaled up to 20 percent, then A is entirely dominated by B. On the other hand, B has to be scaled up by more than 100 percent (i.e., the values have to be doubled) such that it gets dominated by A.

The main reason for the difference of the approximations is that an archiver based on ϵ-dominance accepts in general less solutions than all nondominated ones which

makes the update process much faster, and thus, more points can be evaluated by the generator within the given time budget. In this case the algorithm using $ArchiveUpdateND$ evaluated $6.6e5$ different test points while $3.5e7$ points where evaluated using $ArchiveUpdateTight1$ within the same time.

(a) $ArchiveUpdateND$ (b) $ArchiveUpdateTight1$

Figure 9: Two different solution sets for the MOP in Example 3. We have chosen $\epsilon = (1, 1, 1)$, $\Theta = 1$, and $\Delta = 5$.

7 Conclusions and Future Work

We have proposed two archiving strategies for obtaining finite size and gap free (or 'tight') Pareto front approximations by stochastic search algorithms and have proven the convergence of the resulting archives. The limit set using the first archiver forms with probability one a $(\Delta, \Theta\epsilon_m)$-tight ϵ-approximate Pareto set, that is, a gap free Pareto front approximation—measured by the value of Δ—which provides the guaranteed uniformity level $\Theta\epsilon_m$. The limit set of the second strategy forms a Δ-tight ϵ-Pareto set, which offers a better approximation quality measured in the Hausdorff-sense, but in turn lacks the uniformity.
For future work, the development of an archiving strategy which produces a sequence of archives leading to a (Δ, ϵ_m)-tight ϵ-Pareto set would be of particular interest. It could also be interesting to integrate the archiving strategies directly into the stochastic search process (as e.g. done in (Deb et al., 2005) for an evolutionary algorithm) in order to obtain a fast and reliable multi-objective optimization algorithm. Finally, the analysis of the archiving strategies could be advanced. The main focus in this paper was on the limit behavior of the sequence of archives, but there are also further interesting topics worth investigating. One question which naturally arises is the speed of the convergence. Related works for single-objective optimization problems (e.g., (Voigt et al., 1995; Rudolph, 1997b; Rudolph, 1997a)) show that this not straightforward.

Acknowledgements The first author would like to thank Ulrich Meier-Noe and the anonymous reviewers for their valuable comments which greatly helped them to improve the contents of this paper. The fourth author is also affiliated to the UMI-LAFMIA 3175 CNRS.

References

Bosman, P. A. and de Jong, E. D. (2005). Exploiting Gradient Information in Numerical Multi-Objective Evolutionary Optimization. In Beyer et al., H.-G., editor, *2005 Genetic and Evolutionary Computation Conference (GECCO'2005)*, volume 1, pages 755–762, New York, USA. ACM Press.

Brown, M. and Smith, R. E. (2005). Directed multi-objective optimization. *International Journal on Computers, Systems, and Signals*, 6(1):3–17.

Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley.

Deb, K., Mohan, M., and Mishra, S. (2005). Evaluating the ϵ-dominated based multi-objective evolutionary algorithm for a quick computation of Pareto-optimal solutions. *Evolutionary Computation*, 13(4):501–525.

Ehrgott, M. (2005). *Multicriteria Optimization*. Springer.

Fliege, J. (2004). Gap-free computation of Pareto-points by quadratic scalarizations. *Mathematical Methods of Operations Research*, 59:69–89.

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

Harada, K., Sakuma, J., and Kobayashi, S. (2007). Uniform Sampling of Local Pareto-Optimal Solution Curves by Pareto Path Following and its Applications in Multi-objective GA. In Thierens, D., editor, *2007 Genetic and Evolutionary Computation Conference (GECCO'2007)*, volume 1, pages 813–820, London, UK. ACM Press.

Hestermeyer, T. and Oberschelp, O. (2003). Selbstoptimierende Fahrzeugregelung - Verhaltensbasierte Adaption. In *Intelligente mechatronische Systeme*, volume 122 of *HNI-Verlagsschriftenreihe*. Heinz Nixdorf Institut.

Hillermeier, C. (2001). *Nonlinear Multiobjective Optimization - A Generalized Homotopy Approach*. Birkhäuser.

Knowles, J. D. (2002). *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. PhD thesis, The University of Reading, Department of Computer Science, Reading, UK.

Knowles, J. D. 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. D. and Corne, D. W. (2000). Approximating the non-dominated 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 multiobjective optimization. *Evolutionary Computation*, 10(3):263–282.

Loridan, P. (1984). ϵ-solutions in vector minimization problems. *Journal of Optimization, Theory and Application*, 42:265–276.

Pottharst, A., Baptist, K., Schütze, O., Böcker, J., Fröhlecke, N., and Dellnitz, M. (2004). Operating point assignment of a linear motor driven vehicle using multiobjective optimization methods. Proceedings of the 11th International Conference EPE-PEMC 2004, Riga, Latvia.

Rudolph, G. (1997a). Convergence rates of evolutionary algorithms for a class of convex objective functions. *Control and Cybernetics*, 26(3):275–390.

Rudolph, G. (1997b). Local convergence rates of simple evolutionary algorithms with cauchy mutations. *IEEE Transactions on Evolutionary Computation*, 1(4):249–258.

Rudolph, G. (1998). Finite Markov chain results in evolutionary computation: A tour d'horizon. *Fundamenta Informaticae*, 35:67–89.

Rudolph, G. and Agapie, A. (2000). Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, New Jersey. IEEE Press.

Rudolph, G., Naujoks, B., and Preuss, M. (2007). Capabilities of EMOA to Detect and Preserve Equivalent Pareto Subsets. In Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., and Murata, T., editors, *Evolutionary Multi-Criterion Optimization, 4th International Conference, EMO 2007*, pages 36–50, Matshushima, Japan. Springer. Lecture Notes in Computer Science Vol. 4403.

Ruhe, G. and Fruhwirt, B. (1990). ϵ-optimality for bicriteria programs and its application to minimum cost flows. *Computing*, 44:21–34.

Sayin, S. (2000). Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, 87:543–560.

Schuetze, O., Laumanns, M., Tantar, E., Coello, C. A. C., and ghazali Talbi, E. (2007). Convergence of Stochastic Search Algorithms to Gap-Free Pareto Front Approximations. In Thierens, D., editor, *2007 Genetic and Evolutionary Computation Conference (GECCO'2007)*, volume 1, pages 892–899, London, UK. ACM Press.

Schütze, O. (2004). *Set Oriented Methods for Global Optimization*. PhD thesis, University of Paderborn. <http://ubdata.uni-paderborn.de/ediss/17/2004/schuetze/>.

Schütze, O., Coello, C. A. C., Mostaghim, S., Talbi, E.-G., and Dellnitz, M. (2008). Hybridizing evolutionary strategies with continuation methods for solving multi-objective problems. *Engineering Optimization*, 40(5):383–402.

Schütze, O., Coello, C. A. C., and Talbi, E.-G. (2007). Approximating the ϵ-efficient set of an MOP with stochastic search algorithms. In Gelbukh, A. and Morales, A. F. K., editors, *Mexican International Conference on Artificial Intelligence (MICAI-2007)*, pages 128–138. Springer-Verlag Berlin Heidelberg.

Schütze, O., Coello, C. A. C., Tantar, E., and Talbi, E.-G. (2008). Computing Finite Size Representations of the Set of Approximate Solutions of an MOP with Stochastic Search Algorithms. In *2008 Genetic and Evolutionary Computation Conference (GECCO'2008)*, pages 713–720. ACM Press, Atlanta, USA. ISBN 978-1-60558-131-6.

Schütze, O., Dell'Aere, A., and Dellnitz, M. (2005). On continuation methods for the numerical treatment of multi-objective optimization problems. In Branke, J., Deb, K., Miettinen, K., and Steuer, R. E., editors, *Practical Approaches to Multi-Objective Optimization*, number 04461 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum (IBFI), Schloss Dagstuhl, Germany. <http://drops.dagstuhl.de/opus/volltexte/2005/349>.

Schütze, O., Laumanns, M., Coello, C. A. C., Dellnitz, M., and Talbi, E.-G. (2008). Convergence of stochastic search algorithms to finite size Pareto set approximations. *Journal of Global Optimization*, 41(4):559–577.

Schütze, O., Vasile, M., and Coello Coello, C. A. (2008). Approximate Solutions in Space Mission Design. In Rudolph, G., Jansen, T., Lucas, S., Poloni, C., and Beume, N., editors, *Parallel Problem Solving from Nature–PPSN X*, pages 805–814. Springer. Lecture Notes in Computer Science Vol. 5199, Dortmund, Germany.

Sharma, D., Kumar, A., Deb, K., and Sindhya, K. (2007). Hybridization of SBX based NSGA-II and sequential quadratic programming for solving multi-objective optimization problems. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 3003–3010, Singapore. IEEE Press.

Voigt, H.-M., Mühlenbein, H., and Cvetković, D. (1995). Fuzzy recombination for the breeder genetic algorithm. In Eshelman, L., editor, *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA'95)*, pages 104–113, San Francisco, California, USA. Morgan Kaufmann.

White, D. J. (1998). Epsilon-dominating solutions in mean-variance portfolio analysis. *European Journal of Operational Research*, 105(3):457–466.

Witting, K., Schulz, B., Pottharst, A., Dellnitz, M., Böcker, J., and Fröhleke, N. (2008). A new approach for online multiobjective optimization of mechatronical systems. *International Journal on Software Tools for Technology Transfer*, 10:223–231.

Witting, K. and von Molo, M. H. (2006). Private communication.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C., and Fonseca, V. (2003). Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132.