

Convergence of Stochastic Search Algorithms to Gap-Free Pareto Front Approximations

Oliver Schuetze
INRIA Futurs, LIFL, CNRS,
Cite Scientifique
59655 Villeneuve d'Ascq,
FRANCE

schuetze@lifl.fr

Marco Laumanns
Institute for Operations
Research
ETH Zurich, 8092 Zurich,
SWITZERLAND

laumanns@ifor.math.ethz.ch

Emilia Tantar
INRIA Futurs, LIFL, CNRS,
Cite Scientifique
59655 Villeneuve d'Ascq,
FRANCE

Emilia.Tantar@lifl.fr

Carlos A. Coello Coello
CINVESTAV-IPN, Computer
Science Department
Mexico D.F. 07300,
MEXICO

ccoello@cs.cinvestav.mx

El-ghazali Talbi
INRIA Futurs, LIFL, CNRS,
Cite Scientifique
59655 Villeneuve d'Ascq,
FRANCE

talbi@lifl.fr

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 Pareto front approximation. Since such 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.

Categories and Subject Descriptors

G.1.6 [Numerical Analysis]: Optimization; F.2.1 [Analysis of Algorithms and Problem Complexity]: Numerical Algorithms and Problems

General Terms

Algorithms, Convergence

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO '07, July 7–11, 2007, London, England, United Kingdom.
Copyright 2007 ACM 978-1-59593-697-4/07/0007 ...\$5.00.

Keywords

multi-objective optimization, convergence, ϵ -dominance, stochastic search algorithms

1. INTRODUCTION

One interesting application of multi-objective optimization and its related tools is the online-optimization of mechanical 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 efficient set, the so-called *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 [8, 14] for an operating point assignment strategy of a linear drive, and [4] 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

Here we extend the work of [13] and present new archiving

¹In 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.

strategies for the storage of the ‘essential’ solutions found by a stochastic search algorithm. The strategy used in [13] 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 multiobjective continuation methods (see [11] 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 [13] 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.

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 [3, 9], consider only the convergence to the entire Pareto set, or to a certain subset without considering the approximation quality. Most notably, the archiving/selection strategies of the state-of-the-art multiobjective evolutionary algorithms, such as NSGA-II [2], SPEA2 [15], or the adaptive grid archiving used in PAES and PESA-II [6], which focus on obtaining well-spread Pareto front approximations, do not guarantee convergence.

In [7] the issue of convergence towards a finite-size Pareto set approximation was finally addressed for a general class of iterative search algorithms. Two archiving algorithms were proposed that provably maintain a finite-size approximation of all points ever generated during the search process. This led to the claim that these archiving strategies will 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

[13].

The remainder of this article is organized as follows: Section 2 states the required background for the algorithms, which are presented in Section 3. In Section 4 we present some numerical results. Then, we demonstrate the potential for a possible hybridization with continuation methods in Section 5. Finally, we present our conclusions in Section 6.

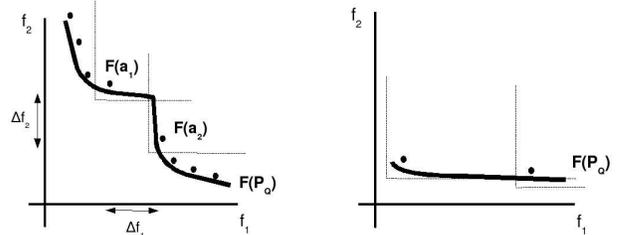


Figure 1: Examples of two ϵ -approximate Pareto sets. Dents or ‘flat’ regions in the Pareto front can lead to (possibly unwanted) gaps in the approximation. In the left figure there is a gap in the approximation between $F(a_1)$ and $F(a_2)$ both in the f_1 - as well as in the f_2 -direction.

2. BACKGROUND

In the following we consider continuous unconstrained multiobjective optimization problems

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

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

$$F: \mathbb{R}^n \rightarrow \mathbb{R}^k, \quad F(x) = (f_1(x), \dots, f_k(x)),$$

and where each $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous.

Definition 2.1 (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 $i \in \{1, \dots, k\}$. The relation \leq_p is defined analogously.

(b) A vector $x \in \mathbb{R}^n$ dominates a vector $y \in \mathbb{R}^n$ (in short: $x < y$) with respect to (MOP) if $F(x) \leq_p F(y)$ and $F(x) \neq F(y)$ (i.e., there exists a $j \in \{1, \dots, k\}$ 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 .

In the following we will define a weaker concept of dominance, so-called (absolute) ϵ -dominance, as well as two approximation concepts which will be used in the sequel.

Definition 2.2 Let $\epsilon = (\epsilon_1, \dots, \epsilon_k) \in \mathbb{R}_+^k$ 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 $F(x) - \epsilon \leq_p F(y)$ and $F(x) - \epsilon \neq F(y)$.

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 2.3 [7]

- (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.,

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

- (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 2.4 [10]

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

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

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

δ is called the uniformity level.

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

Definition 2.5 Let $u, v \in \mathbb{R}^n$ and $A, B \subset \mathbb{R}^n$. The maximum norm distance d_∞ , the semi-distance $\text{dist}(\cdot, \cdot)$ and the Hausdorff distance $d_H(\cdot, \cdot)$ are defined as follows:

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

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

(c) $\text{dist}(B, A) := \sup_{u \in B} \text{dist}(u, A)$

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

Motivated by the need for gap-free Pareto front approximations and inspired by Definitions 2.3 and 2.4 we introduce the following objects:

Definition 2.6 Let $\epsilon \in \mathbb{R}_+^k$ and $\Delta_M, \Delta_m > 0$ and denote by \mathcal{P} the Pareto set of (MOP).

- (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 $\text{dist}(F(\mathcal{P}), F(A_\epsilon)) \leq \Delta_M$.

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

- (c) 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

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

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

Note that $A_\epsilon \subset \mathcal{P}_Q$ is not postulated for ϵ -approximate Pareto sets.

Algorithm 1 gives a framework of a generic stochastic multi-objective optimization algorithm, which will be considered in this work. Theorems 2.7 and 2.8 state convergence results which are closely related to the present work, but which cannot guarantee that the limit archives do not reveal gaps in their Pareto front approximation.

Algorithm 1 Generic Stochastic Search Algorithm

```

1:  $P_0 \subset Q$  drawn at random
2:  $A_0 = \text{ArchiveUpdate}(P_0, \emptyset)$ 
3: for  $j = 0, 1, 2, \dots$  do
4:    $P_{j+1} = \text{Generate}(P_j)$ 
5:    $A_{j+1} = \text{ArchiveUpdate}(P_{j+1}, A_j)$ 
6: end for

```

Algorithm 2 $A := \text{ArchiveUpdateEps1}_\epsilon(P, A_0)$

```

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

```

Theorem 2.7 [13] Let an MOP $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$ be given, where F is continuous, let $Q \subset \mathbb{R}^n$ be a compact set and $\epsilon \in \mathbb{R}_+^k$. Further let

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

Then, an application of Algorithm 1, where $\text{ArchiveUpdateEps1}_\epsilon$ 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 w.r.t. $F|_Q$ for all $l \geq l_0$.

Algorithm 3 $A := \text{ArchiveUpdateEps2}_\epsilon(P, A_0)$

```

1:  $A := A_0$ 
2: for all  $p \in P$  do
3:   if  $\nexists a \in A : a \prec_{\epsilon/3} p$  then
4:      $A := A \cup \{p\}$ 
5:   end if
6:   for all  $a \in A$  do
7:     if  $p \prec a$  then
8:        $A := A \cup \{p\} \setminus \{a\}$ 
9:     end if
10:  end for
11: end for

```

Theorem 2.8 [13] Let (MOP) be given and $Q \subset \mathbb{R}^n$ be compact, and let there be no weak Pareto points in $Q \setminus \mathcal{P}_Q$ (where \mathcal{P}_Q denotes the set of Pareto points of $F|_Q$). Further, let F be injective and

$$\forall x \in Q \text{ and } \forall \delta > 0 : P(\exists l \in \mathbb{N} : P_l \cap B_\delta(x) \cap Q \neq \emptyset) = 1 \quad (2)$$

Then, an application of Algorithm 1, where *ArchiveUpdateEps2 ϵ* 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 an ϵ -approximate Pareto set w.r.t. $F|_Q$ for all $l \geq l_1$.

(b) There exists with probability one a $l_2 \in \mathbb{N}$ such that

$$|A_{l+1}| = |A_l|, \quad \forall l \geq l_2.$$

(c) The limit archive

$$A_\infty := \lim_{l \rightarrow \infty} A_l$$

is an ϵ -Pareto set w.r.t. $F|_Q$ with probability one.

3. THE ALGORITHMS

In the following we investigate two different strategies for the archiving of the solutions found by the algorithm leading to different limit behaviors of the sequence of archives (under certain additional conditions).

Algorithm 4 $A := \text{ArchiveUpdateTight1}_{\epsilon, \tilde{\Delta}}(P, A_0)$

```

1:  $A := A_0$ 
2: for all  $p \in P$  do
3:   if  $(\exists a \in A : a \prec p)$  or  $(\exists a_1 \in A : a_1 \prec_{\epsilon/3} p)$ 
   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:        $A := A \setminus \{a\}$ 
9:     end if
10:  end for
11:   $A := A \cup \{p\}$ 
12: end for

```

Theorem 3.1 Let an MOP $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$ be given, where F is continuous, let $Q \subset \mathbb{R}^n$ be a compact set and $\epsilon \in \mathbb{R}_+^k$. Let $\epsilon_m := \min_{i=1, \dots, k} \epsilon_i$, $\epsilon_M := \max_{i=1, \dots, k} \epsilon_i$, further let $\Delta, \tilde{\Delta} \in \mathbb{R}_+$ be given such that $\epsilon_M < \tilde{\Delta} < \Delta$, and

$$\forall x \in Q \text{ and } \forall \delta > 0 : P(\exists l \in \mathbb{N} : P_l \cap B_\delta(x) \cap Q \neq \emptyset) = 1 \quad (3)$$

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_1} is a $(\Delta, \frac{\epsilon_m}{3})$ -tight ϵ -approximate Pareto set and $A_l = A_{l+1} \forall l \geq l_1$.

PROOF. First, we turn our attention to the question of which elements are added to the archive. The crucial ex-

pression \mathcal{E} (line 3 of Algorithm 4) reads as follows:

$$\underbrace{(\exists a \in A : a \prec p)}_A \text{ or } \underbrace{(\exists a_1 \in A : a_1 \prec_{\epsilon/3} p)}_{\mathcal{B}_1} \text{ and } \underbrace{\exists a_2 \in A : d_\infty(F(a_2), F(p)) \leq \tilde{\Delta}}_{\mathcal{B}_2}$$

Since $\neg \mathcal{E} = (\neg \mathcal{A} \text{ and } \neg \mathcal{B}_1)$ 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

$$(\mathcal{E}_1) \quad \bar{A}a \in A : a \prec_{\epsilon/3} p, \text{ or}$$

$$(\mathcal{E}_2) \quad \bar{A}a \in A : a \prec p \text{ and } \forall a \in A : d_\infty(F(a), F(p)) > \tilde{\Delta}. \quad (4)$$

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 2.7 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. $F|_Q$ 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). Further, since $F(\mathcal{P}_Q)$ is compact and $A_l, l \in \mathbb{N}$, is finite it follows that

$$\text{dist}(F(\mathcal{P}_Q), F(A_l)) = \max_{y \in F(\mathcal{P}_Q)} \min_{a \in A_l} d_\infty(y, F(a))$$

That is, the claim is right for an archive A_l if for every $y \in F(\mathcal{P}_Q)$ there exists an element $a \in A_l$ such that $d_\infty(y, F(a)) \leq \Delta$. That is, $F(\mathcal{P}_Q)$ must be contained in $C_{A_l, \Delta}$, where

$$C_{A, \Delta} := \bigcup_{a \in A} B_\Delta^\infty(F(a)),$$

and $B_\Delta^\infty(x) := \{y \in \mathbb{R}^k : d_\infty(x, y) < \Delta\}$. 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) 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 with $|A_0| = n_0$ 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)) that

$$\exists i \in 1, \dots, k : f_i(p) < f_i(a) - \frac{\epsilon_i}{3}. \quad (5)$$

Note that ϵ_i in (5) can be replaced by $\tilde{\Delta}$ in case A is an ϵ -approximate Pareto set. 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 (5) holds for a_i and a_{i-1} . Since $F(Q)$ is bounded there exist $l_i, u_i \in \mathbb{R}^n, i = 1, \dots, k$, such that $F(Q) \subset [l_1, u_1] \times \dots \times [l_k, u_k]$. 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 $h \geq \lceil r/n_0 \rceil$. Denote by $a_0 \in A_0$ the root of the history. For a, a_0 it follows that

$$\exists i \in 1, \dots, k : f_i(a) < f_i(a_0) - s \frac{\epsilon_i}{3},$$

where $s \geq \lceil h/k \rceil$. For $\tilde{s} > d_{max} := 3 \max_{i=1, \dots, k} \frac{u_i - l_i}{\epsilon_i}$ (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 $F(\mathcal{P}_Q) \not\subset C_{A_1, \Delta}$ for all $l \in \mathbb{N}$. Hence there exists a sequence of image points

$$y_i \in F(\mathcal{P}_Q) \setminus C_{A_1, \Delta} \quad \forall i \in \mathbb{N}. \quad (6)$$

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

$$y_{i_j} \rightarrow y^* \text{ for } j \rightarrow \infty. \quad (7)$$

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

$$\forall (y, \tilde{y}) \in F(Q) \times U_1 : y \leq_p \tilde{y} \Rightarrow d_\infty(y, \tilde{y}) \leq \tilde{\Delta} \quad (8)$$

Let $\tilde{U}_1 := U_1 \cap B_{(\Delta - \tilde{\Delta})/2}^\infty(y^*)$. By (3) 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) 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 (8) – 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

$$\begin{aligned} d_\infty(F(a_1), \tilde{y}) &\leq d_\infty(F(a_1), \tilde{y}_1) + d_\infty(\tilde{y}_1, \tilde{y}) \\ &< \tilde{\Delta} + 2 \frac{\Delta - \tilde{\Delta}}{2} = \Delta \quad \forall \tilde{y} \in \tilde{U}_1. \end{aligned} \quad (9)$$

By (6) and (7) there exist $j_1, \tilde{l}_1 \in \mathbb{N}$ with

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

Since by (9) 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. \square

Remarks 3.2 (a) *We have chosen an archiving strategy to obtain a uniform spread of the Pareto front as motivated in Section 1. There exist on the other hand certainly reasons to aim for a uniform approximation of the Pareto set. The following little example shows, however, that in that case in addition the injectivity of F has to be postulated, at least when keeping only non-dominated solutions in the archive. For a given $\Delta > 0$*

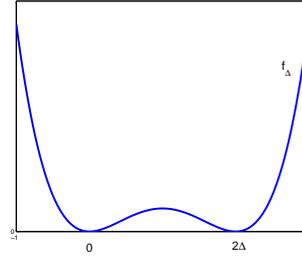


Figure 2: Example of a function f_Δ with two isolated global minima $m_1 = 0$ and $m_2 = 2\Delta$.

let f_Δ be as shown in Figure 2. That is, let f_Δ have two isolated global minima m_1 and m_2 with $m_1 < m_2$ and with $d_\infty(m_1, m_2) > \Delta$. Define $F := (f_\Delta, f_\Delta + C)$, where $C \in \mathbb{R}$ is a constant. If the domain is e.g. chosen as $A := [m_1 - \Delta, m_2 + \Delta]$, the Pareto set of the resulting MOP is given by $\mathcal{P} = \{m_1, m_2\}$. Since the probability to find a point $p_2 \in A$ which has the same objective values $F(p_2) = F(p_1)$ of a given point $p_1 \in A$ is zero in the underlying setting, it follows that the set of nondominated points of a given population consists, with probability one, of one single point. Thus, an approximation A with $\text{dist}(\mathcal{P}_Q, A) \leq \Delta$ can in general not be obtained when only nondominated points are kept in the archive.

(b) *The archive size can be bounded in analogy to the magnitudes of the archives which are maintained by using *ArchiveUpdateEps1* by (see [13])*

$$|A_l| \leq \sum_{\substack{i_1, \dots, i_{k-1}=1 \\ i_1 > \dots > i_{k-1}}}^k \prod_{j=1}^{k-1} \left\lceil 3 \frac{M_{i_j} - m_{i_j}}{\epsilon_{i_j}} \right\rceil, \quad (10)$$

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), 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 $\text{dist}(F(A), F(\mathcal{P}_Q))$ 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 $F(p) \in F(A) \setminus C_{A, \Delta}$ which dominates $a_i, i \in \{1, 2, 3\}$. In such a manner an example can be constructed with $\text{dist}(F(A), F(\mathcal{P}_Q)) = \max_{i=1, \dots, k} (M_i - m_i)$. However, we have never observed this (bad) theoretical value in our computational experiments.

Algorithm 5 shows an archiving strategy which overcomes the problem of Algorithm 4 described above but which in turn lacks the uniformity.

Theorem 3.3 *Let (MOP) be given and $Q \subset \mathbb{R}^n$ be compact, and let there be no weak Pareto points in $Q \setminus \mathcal{P}_Q$. Fur-*

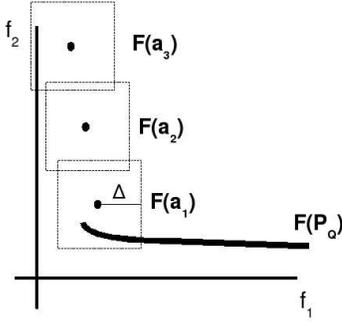


Figure 3: Possible example of a set A which could be generated by *ArchiveUpdateTight1* with $\text{dist}(F(A), F(\mathcal{P}_Q)) > \Delta$.

Algorithm 5 $A := \text{ArchiveUpdateTight2}_{\epsilon, \Delta}(P, A_0)$

```

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:      $A := A \cup \{p\}$ 
5:   end if
6:   for all  $a \in A$  do
7:     if  $p < a$  then
8:        $A := A \cup \{p\} \setminus \{a\}$ 
9:     end if
10:  end for
11: end for

```

then, let F be injective and

$$\forall x \in Q \text{ and } \forall \delta > 0 : P(\exists l \in \mathbb{N} : P_l \cap B_\delta(x) \cap Q \neq \emptyset) = 1 \quad (11)$$

Then an application of Algorithm 1, where *ArchiveUpdateTight2* $_{\epsilon, \Delta}$ 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. $F|_Q$ for all $l \geq l_1$.

(b) There exists with probability one a $l_2 \in \mathbb{N}$ such that

$$|A_{l+1}| = |A_l|, \quad \forall l \geq l_2.$$

(c) The limit archive

$$A_\infty := \lim_{l \rightarrow \infty} A_l$$

is a Δ -tight ϵ -Pareto set w.r.t. $F|_Q$ with probability one.

PROOF. All parts of the proof are analogue to parts in proofs of Theorem 2.8 and Theorem 3.1. \square

Remark 3.4 The archive size obtained by *ArchiveUpdateTight2* can be bounded analogously to the one obtained by *ArchiveUpdateEps2* [13] by

$$|A_l| \leq \prod_{i=1}^k \left\lceil 3 \frac{M_i - m_i}{\epsilon_i} \right\rceil.$$

The lower bound of $|A_\infty|$ for both new archiving strategies is obviously given by 1. For this, consider e.g. $f_1 = f_2 = \dots = f_k$ to be a convex function which takes its (unique) minimum inside Q . 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 $\epsilon \ll \Delta$, and (b) the Pareto front can fall into different connected components. However, if this is not the case, one can obtain the following result, which we state without the (obvious) proof.

Lemma 3.5 Let $\tilde{m}_i = \min_{x \in \mathcal{P}_Q} f_i(x)$ and $\tilde{M}_i = \max_{x \in \mathcal{P}_Q} f_i(x)$, $1 \leq i \leq k$. Then, when using *ArchiveUpdateTight1* $_{\epsilon, \Delta}$ or *ArchiveUpdateTight2* $_{\epsilon, \Delta}$, the archive size maintained in Algorithm 1 for the limit archive $A_\infty = \lim_{l \rightarrow \infty} A_l$ is bounded by

$$|A_\infty| \geq \max_{i=1, \dots, k} \left\lceil \frac{\tilde{M}_i - \tilde{m}_i}{2\Delta} \right\rceil \quad (12)$$

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.

4. NUMERICAL RESULTS

In this section we compare three different archiving strategies on the following example:

$$f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$$

$$f_i(x) = \sum_{\substack{j=1 \\ j \neq i}}^n (x_j - a_j^i)^2 + (x_i - a_i^i)^4, \quad (13)$$

where

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

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). Figure 4 shows examples for resulting limit sets. Hereby we have taken $N = 200,000$ randomly chosen points in $Q = [-1.5, 1.5]^3$ using the following archiving strategies:

- (ND) all nondominated points are kept,
- (Eps₁) *ArchiveUpdateEps1*, and
- (Tight₁) *ArchiveUpdateTight1*,

and using the values $\epsilon = (1, \dots, 1)$ and $\Delta = 2$. The set obtained by *ArchiveUpdateEps1* forms probably (or is near to) an ϵ -approximate Pareto set, but reveals huge gaps, which is not the case in Figure 4 (c), where *ArchiveUpdateTight1* has been used. The ‘tightest’ approximation in this case study is certainly obtained when all nondominated points are kept in the archive (see Figure 4 (b)). However, in that case the time which had to be spent to update the archive² was huge compared to both other strategies (see Table 4).

An implementation of all the archiving strategies discussed in this work including this example can be found in [1].

²The elements of the all archives were stored using a linear list.

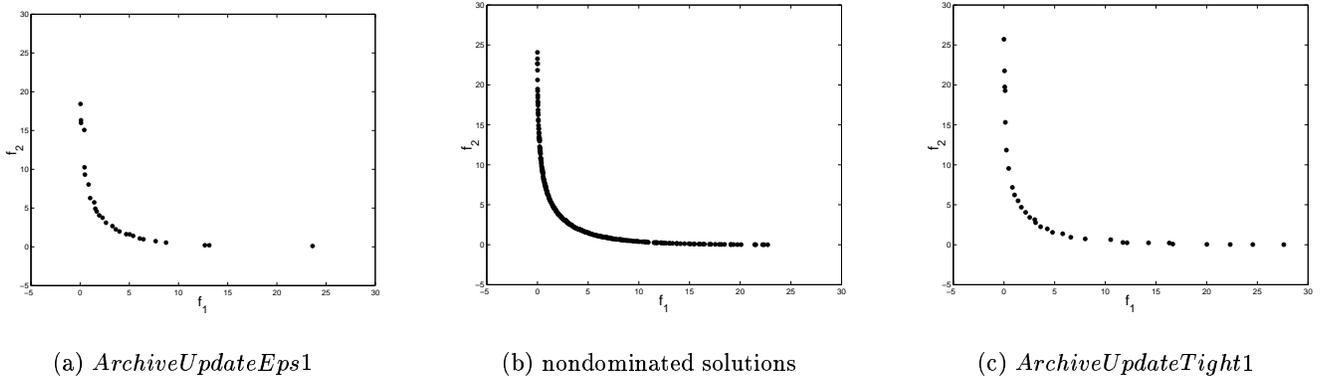


Figure 4: Three limit archives obtained by different archiving strategies.

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 (see text). We have taken the average result of 100 test runs.

	ND	Eps1	Tight1
$ A_N $	267	27	34
T	36.46	0.29	0.36

5. OUTLOOK: HYBRIDIZING WITH MULTI-OBJECTIVE CONTINUATION METHODS

In order to increase the overall computational performance, it is often desired to combine the (global) stochastic search algorithm with a local search strategy. Here we want to show that in the underlying context a hybridization with multi-objective continuation methods (e.g., [5, 12]) could be advantageous since these methods are very effective locally and can be used to compute uniformly spread solutions. In the following we construct a step size strategy for our purpose and show numerical results on one (easy) example, indicating the possible benefit of such a hybridization. For more details and examples see [13].

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 point $x_0 \in \mathcal{P}_Q$, $\Delta \in \mathbb{R}_+$, and a *search direction* $v \in \mathbb{R}^n$ with $\|v\| = 1$, the task is to find a step size $h \in \mathbb{R}_+$ such that for the next guess $y_0 = x_0 + hv$ it holds

$$\|F(x_0) - F(y_0)\|_\infty = \Theta\Delta, \quad (14)$$

where $\Theta \in (0, 1)$ is a safety factor. In case F is Lipschitz continuous there exists an $L \geq 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in Q. \quad (15)$$

The Lipschitz constant around x_0 (i.e., locally) can be estimated by

$$L_{x_0} := \|DF(x_0)\|_\infty = \max_{i=1, \dots, k} \|\nabla f_i(x_0)\|_1.$$

Combining (14) and (15), using $\|x_0 - y_0\| = h$, and assuming that h is sufficiently small, we obtain the following estimation

$$h \approx \frac{\Theta\Delta}{L_{x_0}}. \quad (16)$$

Note that this estimation only suits for small values of Δ since in the other case h will be too large, and thus L_{x_0} can not serve as a suitable Lipschitz estimation.

In Figure 5 a numerical result for MOP (13) is presented, where we have used *one* run of the continuation algorithm proposed in [12], starting with *one* single solution. 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. The resulting set consists of 23 solutions which clearly represent a uniform Pareto front approximation. To obtain the result, less than 100 Newton steps on an auxiliary function (which, however, incorporates the derivatives of the objectives) had to be performed. This result motivates for the further development of such combinations with evolutionary multiobjective optimization techniques in the future.

6. CONCLUSIONS

We have proposed two archiving strategies for obtaining gap-free ('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 $(\Delta, \epsilon_m/3)$ -tight ϵ -approximate Pareto set, that is a gap-free Pareto front approximation which provides the guaranteed uniformity level $\epsilon_m/3$. The limit set of the second strategy forms a Δ -tight ϵ -Pareto set, which offers a better approximation quality 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.

Acknowledgements

The first author would like to thank Ulrich Meier-Noe for fruitful discussions on the content of this paper. The fourth author gratefully acknowledges support from the NSF-CO-NACyT project no. 42435-Y.

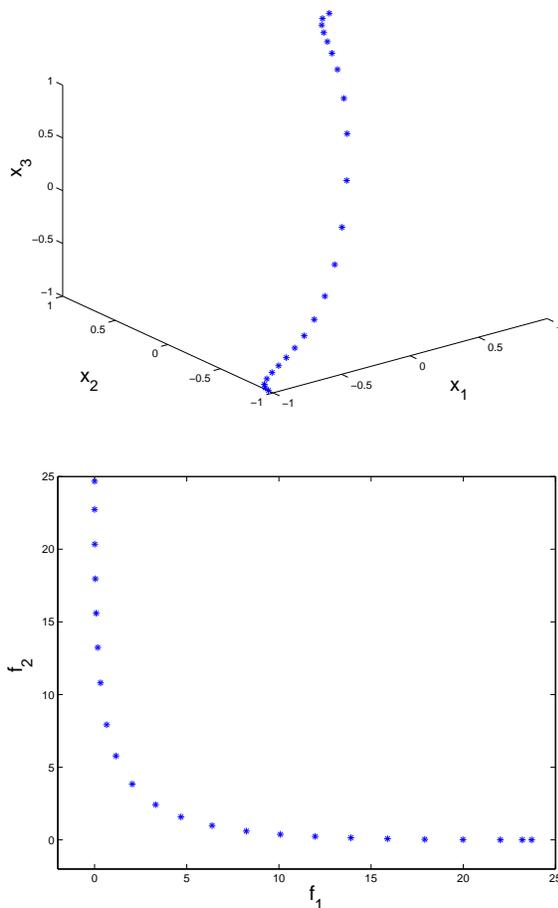


Figure 5: Result of the continuation method with step size control on MOP (13) for $n = 3$ in parameter space (above) and image space (below).

7. REFERENCES

- [1] <http://paradiseo.gforge.inria.fr>.
- [2] K. Deb, Samir Agrawal, Amrit Pratap, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In Marc Schoenauer et al., editors, *Parallel Problem Solving from Nature (PPSN VI)*, Lecture Notes in Computer Science Vol. 1917, pages 849–858. Springer, 2000.
- [3] T. Hanne. On the convergence of multiobjective evolutionary algorithms. *European Journal Of Operational Research*, 117(3):553–564, 1999.
- [4] T. Hestermeyer and O. Oberschelp. Selbstoptimierende Fahrzeugregelung - Verhaltensbasierte Adaption. In *Intelligente mechatronische Systeme*, volume 122 of *HNI-Verlagsschriftenreihe*. Heinz Nixdorf Institut, 2003.
- [5] C. Hillermeier. *Nonlinear Multiobjective Optimization - A Generalized Homotopy Approach*. Birkhäuser, 2001.
- [6] J. Knowles and D. Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Transactions on Evolutionary Computation*, 7(2):100–116, 2003.
- [7] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.
- [8] A. Pottharst, K. Baptist, O. Schütze, J. Böcker, N. Fröhlecke, and M. Dellnitz. Operating point assignment of a linear motor driven vehicle using multiobjective optimization methods. Proceedings of the 11th International Conference EPE-PEMC 2004, Riga, Latvia., 2004.
- [9] G. Rudolph and A. Agapie. On a multi-objective evolutionary algorithm and its convergence to the Pareto set. In *Congress on Evolutionary Computation (CEC2000)*, pages 1010–1016, 2000.
- [10] S. Sayin. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, 87:543–560, 2000.
- [11] O. Schütze. *Set Oriented Methods for Global Optimization*. PhD thesis, University of Paderborn, 2004. <<http://ubdata.uni-paderborn.de/ediss/17/2004/schuetze/>>.
- [12] O. Schütze, A. Dell'Aere, and M. Dellnitz. On continuation methods for the numerical treatment of multi-objective optimization problems. In Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Ralph E. Steuer, editors, *Practical Approaches to Multi-Objective Optimization*, number 04461 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum (IBFI), Schloss Dagstuhl, Germany, 2005. <<http://drops.dagstuhl.de/opus/volltexte/2005/349>>.
- [13] Oliver Schütze, Marco Laumanns, Carlos A. Coello Coello, Michael Dellnitz, and El-Ghazali Talbi. Convergence of stochastic search algorithms to finite size Pareto set approximations. Research Report 6063, INRIA, 12 2006.
- [14] K. Witting, B. Schulz, A. Pottharst, M. Dellnitz, J. Böcker, and N. Fröhlecke. A new approach for online multiobjective optimization of mechatronical systems. Submitted to the International Journal on Software Tools for Technology Transfer STTT, 2005.
- [15] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization. In K.C. Giannakoglou et al., editors, *Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001)*, pages 95–100. International Center for Numerical Methods in Engineering (CIMNE), 2002.