Convergence of Hypervolume-Based Archiving Algorithms I: Effectiveness

Karl Bringmann
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Tobias Friedrich
Max-Planck-Institut für Informatik
Saarbrücken, Germany

ABSTRACT

The core of hypervolume-based multi-objective evolutionary algorithms is an archiving algorithm which performs the environmental selection. A $(\mu + \lambda)$-archiving algorithm defines how to choose μ children from μ parents and λ offspring together. We study theoretically $(\mu + \lambda)$-archiving algorithms which never decrease the hypervolume from one generation to the next.

Zitzler, Thiele, and Bader (IEEE Trans. Evolutionary Computation, 14:58–79, 2010) proved that all $(\mu + 1)$-archiving algorithms are *ineffective*, which means there is an initial population such that independent of the used reproduction rule, a set with maximum hypervolume cannot be reached. We extend this and prove that for $\lambda < \mu$ all archiving algorithms are ineffective. On the other hand, *locally optimal algorithms*, which maximize the hypervolume in each step, are effective for $\lambda = \mu$ and can always find a population with hypervolume at least half the optimum for $\lambda < \mu$.

We also prove that there is *no* hypervolume-based archiving algorithm which can always find a population with hypervolume greater than $1/(1 + 0.1338\,(1/\lambda - 1/\mu))$ times the optimum.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]:
Analysis of Algorithms and Problem Complexity

General Terms

Measurement, Archiving Algorithms,
Performance, Hypervolume Indicator

Keywords

Multiobjective Optimization, Theory,
Performance Measures, Selection

1. INTRODUCTION

A typical evolutionary algorithm requires (i) a reproduction rule to generate new individuals and (ii) a selection rule to choose a subset of individuals from a larger population. In contrast to the single-objective case, the greater challenge of evolutionary multi-objective optimization (EMO) is the selection rule. In the single-objective case, alternatives are usually judged by a single real-valued objective function which defines a linear order and therefore a complete ranking of all alternatives. On the other hand, the Pareto order of multi-objective optimization only defines a partial order. For ranking incomparable candidate solutions, we have to use a so-called second level sorting criterion. Unfortunately, every known criterion has their own advantages and drawbacks.

Many state-of-the-art EMO algorithms use the hypervolume [17] as a second level sorting criterion. Examples for hypervolume-based algorithms are the multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES, [7, 15]), the SMS-EMOA [1], and the indicator-based evolutionary algorithm (IBEA, [16]).

Most multi-objective evolutionary algorithms (MOEAs) are elitist, that is, they keep an external archive in order to capture the output of the search process. As the set of visited solutions can become very large, most algorithms assume that the size of the archive is upper bounded by some fixed value μ. The two typical elitist selection strategies are (i) choosing the μ best of the λ offspring (comma strategy) or (ii) choosing the μ best of the λ offspring and μ parents together (plus strategy). We consider the latter and describe the selection step by a $(\mu+\lambda)$-*archiving algorithm* which defines how to choose a new population of μ children from the union of μ parents and λ offspring (cf. Algorithm 2).

1.1 Convergence

We want to study the convergence behavior of multi-objective evolutionary algorithms (MOEAs). The first rigorous study of the convergence properties of general MOEAs appeared around the turn of the millennium. The papers by Hanne [6] and Rudolph and Agapie [13] show that certain types of elitism lead to convergence to a subset of the Pareto front. If the archive is unbounded, asymptotic convergence to the whole Pareto front is also relatively simple to show if the mutation operator connects all solutions with nonzero probability and no nondominated solution is discarded.

In contrast to these results, we are interested in convergence results for bounded archives. Several authors [2, 5, 8–12, 18] have considered $(\mu + 1)$-MOEAs, that is, generation processes which produce a sequence of solutions for which an archive of fixed maximum size μ is maintained. Most of this work is empirical. We refer the reader to the recent paper by López-Ibáñez, Knowles, and Laumanns [12] which reviews the properties of six different $(\mu + 1)$-archiving algorithms in greater detail. They have also implemented several archiving algorithms within a common framework. Their experiments show that hypervolume-based archiving is to be preferred over other methods for example for clustered points.

On the theoretical side, Laumanns, Thiele, Deb, and Zitzler [11],

GECCO'11, July 12–16, 2011, Dublin, Ireland.

Knowles and Corne [8], and Schütze, Laumanns, Coello, Dellnitz, and Talbi [14] studied ε-based archiving algorithms. Moreover, hypervolume-based archiving algorithms were studied in Zitzler et al. [18] in general and in Beume et al. [2] for concrete fronts. We extend the results of the last two works in two directions: We consider offspring sizes $\lambda > 1$ and consider not only the question of converging to a maximum, but whether we converge at least to a good approximation of the maximal hypervolume.

1.2 Archiving Algorithms

We want to study the convergence behavior of hypervolume-based $(\mu + \lambda)$-MOEAs. Their shared feature is that their selection rule uses the hypervolume. We want to abstract from the reproduction step and study convergence only depending on the used hypervolume-based archiving algorithm.

Our results hold for different classes of archiving algorithms. We start with defining the two most important ones. Precise definitions and results are stated in the main part of the paper.

- An archiving algorithm is *non-decreasing* if it chooses the population of children such that the dominated hypervolume does not decrease compared to the parent generation (cf. Definition 2.4). This is certainly a desirable property for any hypervolume-based archiving algorithm. It is also a necessary assumption in order to prove our lower bounds (cf. Theorem 5.1).

- An archiving algorithm is *locally optimal* if it chooses the population of children such that the dominated hypervolume is maximized (cf. Definition 2.5). For example the archiving algorithm of SMS-EMOA [1] falls in this class. However, locally optimal archiving algorithms are not efficiently computable in general [4].

1.3 Choice of the Offspring

In order to rigorously study the impact of archiving algorithms on the convergence, we have to decide how an offspring is generated. As the offspring generation is done on some arbitrary search space, it seems impossible to come up with realistic probabilistic assumptions. We assume a best-case offspring generation and say an archiving algorithm is *effective* if there exists a sequence of offsprings such that the algorithm reaches an optimum (cf. Definition 3.1). This corresponds to the notion of λ-*greedy* used by Zitzler et al. [18].

Another approach is a worst-case perspective on the offspring generation. The problem is that an adversary who selects the offspring is very strong and can limit the search to a very small part of the search space. Therefore it is impossible to reach the optimum in this case. We do not consider worst-case offspring selection in this paper, but discuss it briefly in Section 8.

1.4 Results

Zitzler et al. [18] proved that all non-decreasing $(\mu + 1)$-archiving strategies are ineffective (cf. Theorem 3.2). On the other hand, it is easy to see that a $(\mu + \mu)$-archiving strategy can directly jump to the optimal set (cf. Theorem 3.4). We extend this and prove that indeed all locally optimal $(\mu + \mu)$-archiving strategies are effective. The status for example of $(\mu + 2)$-archiving strategies was open so far. In fact, Zitzler et al. [18, p. 71] state it as an open research issue "whether other values for λ with $1 < \lambda < \mu$ are sufficient to guarantee convergence". We answer this question in the negative and prove that for $\lambda < \mu$ all non-decreasing $(\mu + \lambda)$-archiving strategies are ineffective (cf. Theorem 3.5). This shows that only for $\lambda = \mu$ non-decreasing hypervolume-based archiving strategies can be effective. Note that the restriction to non-decreasing strategies is necessary (cf. Theorem 5.1).

These results on the effectiveness raise the following question: How close to an optimal set are the best reachable sets for $\lambda < \mu$? To measure this, we call an archiving strategy α-*approximate* if it can always reach a set with a hypervolume at least $1/\alpha$ times the largest possible hypervolume (cf. Definition 4.1). We prove that no non-decreasing $(\mu + \lambda)$-archiving algorithm can be better than $\left(1+0.1338\left(\frac{1}{\lambda}-\frac{1}{\mu}\right)\right)$-approximate (cf. Theorem 4.2). On the other hand, every $(\mu + \lambda)$-archiving algorithm which chooses a children population with larger hypervolume than the parent population (if there is one), reaches a 2-approximation (cf. Theorem 4.3). Note that in these results we omitted all summands and factors of arbitrarily small $\varepsilon > 0$. For details see the respective theorems.

The outline is as follows. In Section 2 we introduce the basic concepts and notations. Sections 3 and 4 shows our results on exact and approximate effectiveness. In Section 5 we discuss why we only study non-decreasing algorithms. The proofs of the two main results are given in Sections 6 and 7. We finish with some concluding remarks and plans for future work in Section 8.

2. PRELIMINARIES

We consider the maximization of vector-valued objective functions $f\colon \mathcal{X} \to \mathbb{R}^d$. Here, $\mathcal{X}$ denotes an arbitrary search space consisting of all alternatives of the decision problem. The feasible points $\mathcal{Y} := f(\mathcal{X})$ are called the objective space. We consider the following abstract framework of a MOEA.

Algorithm 1: General $(\mu + \lambda)$-MOEA

1 $P^0 \leftarrow$ initialize with μ individuals
2 **for** $i \leftarrow 1$ **to** N **do**
3 $\quad Q^i \leftarrow$ generate λ offspring
4 $\quad P^i \leftarrow$ select μ individuals from $P^{i-1} \cup Q^i$

In this paper we make no assumptions on the specific search space $\mathcal{X}$, nor an assumption on how the points are initialized (cf. line 1 of Algorithm 1), nor an assumption how offspring is generated (cf. line 3 of Algorithm 1). Therefore, we assume that the initial population is chosen worst-case and the offspring generation is best-case. Our main concern is the way how the population of children is chosen (cf. line 4 of Algorithm 1). We will formally define and discuss different *archiving algorithms* in Sections 2.2 and 2.3.

We use the terms archive and population synonymously for the set of current solutions P^i of Algorithm 1. In concrete MOEAs, populations are subsets of the search space. As we do not want to assume any structural properties of the search space, we abstract from the search space and will *only work on the objective space* $\mathcal{Y} \subseteq \mathbb{R}^d$ in the remainder. We therefore also identify individuals with points in the d-dimensional Euclidean space. Only considering the objective space also means that the archiving algorithm called in line 4 of Algorithm 1 only has access to the points in the objective space and does not know the respective preimages in the search space. As several points in the search space can map to the same point in the objective space, we can not assume that our population is free of duplicate or multiple points. Hence a population is a multiset of points from the objective space $\mathcal{Y}$. As multiplicities are insignificant for all of our results, we avoid the notion of multisets and instead use the following definition.

DEFINITION 2.1. *A population P is a finite subset of $\mathbb{R}^d$. If an objective space $\mathcal{Y} \subseteq \mathbb{R}^d$ is fixed, we require $P \subseteq \mathcal{Y}$. We call P a μ-population if $|P| \leqslant \mu$.*

In the remainder of this Section 2 we first define the hypervolume quality indicator in Section 2.1 and general archiving algorithms in Section 2.2, and then use both to classify hypervolume-based archiving algorithms in Section 2.3.

2.1 Hypervolume Indicator

The hypervolume indicator $\text{HYP}(P)$ [17] of a population P is the volume of the union of regions of the objective space which are dominated by P and bounded by a reference point R. Here domination refers to the following dominance relation for points in the objective space $\mathcal{Y} \subseteq \mathbb{R}^d$:

$$(x_1, x_2, \ldots, x_d) \preceq (y_1, y_2, \ldots, y_d)$$
$$\text{iff } x_1 \leqslant y_1, x_2 \leqslant y_2, \ldots, \text{and } x_d \leqslant y_d.$$

Formally, the hypervolume $\text{HYP}(P)$ of a population P is defined as

$$\text{HYP}(P) := \int_{\mathbb{R}^d} A_P(x)\, dx$$

where the attainment function $A_P \colon \mathbb{R}^d \to \{0,1\}$ is an indicator function on the objective space which describes the space above the reference point R which is dominated by P, that is, $A_P(x) = 1$ if $R \preceq x$ and there is a $p \in P$ such that $x \preceq p$, and $A_P(x) = 0$ otherwise.

In this work we will fix the reference point w.l.o.g. to $R = 0^d$, since translations do not change any of our results. This means that the reference point is globally fixed and known to the archiving algorithm.

The aim of a hypervolume-based MOEA is finding a set P^* of size μ which maximizes the hypervolume, that is,

$$\text{HYP}(P^*) = \text{maxHYP}_\mu(\mathcal{Y})$$

where we define for all $Y \subset \mathbb{R}^d$,

$$\text{maxHYP}_\mu(Y) := \sup_{\substack{P \subseteq Y \\ |P| \leqslant \mu}} \text{HYP}(P).$$

In the remainder of the paper, the set Y will often be finite. In these cases, the supremum in the definition of $\text{maxHYP}_\mu(Y)$ becomes a maximum. However, for infinite sets the supremum is necessary in general.

Most hypervolume-based algorithms like the steady-state MO-CMA-ES [7, 15] and the SMS-EMOA [1], remove the individual contributing the least hypervolume to the population. The *contribution* of a point p to a population P is

$$\text{CON}_P(p) := \text{HYP}(P) - \text{HYP}(P - p),$$

where we use the notation $P - p$ for $P \setminus \{p\}$. We also use $P + p$ to shurtcut $P \cup \{p\}$ throughout the paper.

Note that according to the definition of $\text{CON}_P(p)$, the contributing hypervolume of a dominated individual is zero. Unfortunately, $\text{CON}_P(p)$ is NP-hard to approximate [4], as opposed to $\text{HYP}(P)$ which can be approximated efficiently [3]. This implies that all contribution-based archiving algorithms are computationally very expensive.

2.2 Archiving Algorithms

We now specify more formally how to choose the μ offsprings in line 4 of Algorithm 1. For this, we consider the following general framework of an archiving algorithm.

Algorithm 2: General $(\mu + \lambda)$-archiving algorithm

input : μ-population P, λ-population Q
output: μ-population P' with $P' \subseteq P \cup Q$

Note that any $(\mu + \lambda)$-archiving algorithm is also a $(\mu + \lambda')$-archiving algorithm for any $\lambda' < \lambda$, as we then allow only a subset of the inputs, namely with smaller offspring population Q. We do not make any assumptions on the runtime of an archiving algorithm. In fact, as hypervolume computation is #P-hard [3], most hypervolume-based archiving algorithms are not computable in polynomial time in the number of objectives d. We will use the following notation to describe an archiving algorithm.

DEFINITION 2.2. *A $(\mu + \lambda)$-archiving algorithm $\mathcal{A}$ is a partial mapping $\mathcal{A} \colon 2^{\mathbb{R}^d} \times 2^{\mathbb{R}^d} \mapsto 2^{\mathbb{R}^d}$ such that for a μ-population P and a λ-population Q, $\mathcal{A}(P,Q)$ is a μ-population and $\mathcal{A}(P,Q) \subseteq P \cup Q$.*

For convenience, we sometimes drop the prefix $(\mu + \lambda)$ and just refer to an archiving algorithm without specifying μ and λ. Note that (in contrast to e.g. Knowles and Corne [8, Def. 4]), we require $\mathcal{A}(P,Q) \subseteq P \cup Q$ in Definition 2.2. This assumption is crucial for most of our results. With this, we can now formally describe the generation process of Algorithm 1 as follows.

DEFINITION 2.3. *Let P^0 be a μ-population and $Q^1, \ldots, Q^N$ a sequence of λ-populations. Then*

$$P^i := \mathcal{A}(P^{i-1}, Q^i) \qquad \text{for all } i = 1, \ldots, N.$$

We also set

$$\begin{aligned}\mathcal{A}(P^0, Q^1, \ldots, Q^i) &:= \mathcal{A}(\mathcal{A}(P^0, Q^1, \ldots, Q^{i-1}), Q^i) \\ &= \mathcal{A}(\ldots \mathcal{A}(\mathcal{A}(P^0, Q^1), Q^2), \ldots, Q^i) \\ &= P^i \qquad \text{for all } i = 1, \ldots, N.\end{aligned}$$

2.3 Hypervolume-based Archiving Algorithms

We now specify two types of hypervolume-based archiving algorithms. The first one only requires the archiving algorithms to never return a solution with a smaller hypervolume:

DEFINITION 2.4. *An archiving algorithm $\mathcal{A}$ is* non-decreasing, *if for all inputs P and Q we have*

$$\text{HYP}(\mathcal{A}(P,Q)) \geqslant \text{HYP}(P).$$

All reasonable hypervolume-based archiving algorithms are non-decreasing. However, the class also contains ineffective algorithms like the algorithm which always returns P. In fact, in the following Sections 3 and 4 we do not consider algorithms which are not non-decreasing. We will justify this imposed minimum requirement in Theorem 5.1 in Section 5.

While our negative results hold for all non-decreasing archiving algorithms (cf. Theorems 3.2 and 3.5), we apparently need a stronger requirement to show positive results. The second type of archiving algorithms we are looking at is the one usually considered to be the best. It always returns a population which maximizes the hypervolume among all given points. As there may be more than one population maximizing the hypervolume, there is more than one archiving algorithm covered by this definition.

DEFINITION 2.5. *An archiving algorithm $\mathcal{A}$ is* locally optimal*, if for all inputs P and Q we have*

$$\mathrm{HYP}(\mathcal{A}(P,Q)) = \mathrm{maxHYP}_\mu(P \cup Q).$$

These are the two most important classes of hypervolume-based archiving algorithms. Most of our positive results below (cf. Theorems 3.4 and 4.3) hold for slightly larger classes of archiving algorithms. In these cases, we define the specific class in the respective theorem.

We are now prepared to study effectiveness, an approach to analyzing the quality of archiving algorithms with respect to the goal of maximizing the hypervolume, in the following Section 3.

3. EFFECTIVENESS

Without any additional assumptions on the specific MOEA and problem at hand, we can only assume the initial population to be worst-case. A best-case view makes no sense, as then the initial population already maximizes the hypervolume. On the other hand, there are two possible ways to choose offspring: worst-case and best-case. We consider the case of a best-case choice of the offspring and analyzes which archiving algorithms are *effective*, that is, are able to reach the optimum. We plan to complement this with a worst-case perspective on the choice of the offspring in future work. See Section 8 for details.

More formally, this section elaborates whether for a given archiving algorithm $\mathcal{A}$ and all finite objective spaces $\mathcal{Y}$ and initial populations $P^0 \subseteq \mathcal{Y}$, there is a sequence of offsprings such that the archiving algorithm run on P^0 and the sequence of offsprings generates a population maximizing the hypervolume on $\mathcal{Y}$. As discussed above, this corresponds to a worst-case view on the problem (i.e., objective space $\mathcal{Y}$ and initial population P^0), but a best-case view on the drawn offspring. This is summarized in the following discussion.

DEFINITION 3.1. *A $(\mu+\lambda)$-archiving algorithm $\mathcal{A}$ is* effective*, if for all finite sets $\mathcal{Y} \subset \mathbb{R}^d$ and μ-populations $P^0 \subseteq \mathcal{Y}$ there exists an $N \in \mathbb{N}$ and a sequence of λ-populations $Q^1, \dots, Q^N \subseteq \mathcal{Y}$ such that*

$$\mathrm{HYP}(\mathcal{A}(P^0, Q^1, \dots, Q^N)) = \mathrm{maxHYP}_\mu(\mathcal{Y}).$$

Here, we require the objective spaces $\mathcal{Y}$ to be finite, as infinite objective spaces do not necessarily have a hypervolume maximizing μ-population. This is no real restriction as for infinite objective spaces the following negative results remain valid. We first state a result of Zitzler et al. [18].

THEOREM 3.2. *There is no effective non-decreasing $(\mu+1)$-archiving algorithm.*

Note that we have reformulated the statement of [18, Cor. 4.6] in our notation defined above. The notation used in [18] is very different as they examine general set-based MOEAs and not specifically archiving algorithms. We do not give a separate proof for Theorem 3.2 as it directly follows from Theorem 3.5 below. Theorem 3.2 assumes $\lambda = 1$. The corresponding result for $\lambda = \mu$ follows from [18, Thm. 4.4]:

THEOREM 3.3. *There is an effective non-decreasing $(\mu+\mu)$-archiving algorithm.*

We also do not give a proof for Theorem 3.3 as it follows from Theorem 3.4 below. Since Theorem 3.3 is only an existential statement, it is natural to ask how effective non-decreasing $(\mu+\mu)$-archiving algorithms look like. The following theorem shows that all locally optimal $(\mu+\mu)$-archiving algorithms are effective.

THEOREM 3.4. *Let $\mathcal{A}$ be a $(\mu+\mu)$-archiving algorithm such that for all μ-populations P and Q,*

$$\mathrm{HYP}(\mathcal{A}(P,Q)) \geqslant \mathrm{HYP}(Q).$$

Then $\mathcal{A}$ is effective.

In particular: All locally optimal $(\mu+\mu)$-archiving algorithms are effective.

Proof. Let $\mathcal{Y}$ be any finite objective space and $P^0 \subset \mathcal{Y}$ of size μ. Moreover, let P^* maximize the hypervolume on $\mathcal{Y}$, i.e., $\mathrm{HYP}(P^*) = \mathrm{maxHYP}_\mu(\mathcal{Y})$. Then setting $Q^1 := P^*$ we have $\mathrm{HYP}(\mathcal{A}(P^0, Q^1)) \geqslant \mathrm{HYP}(Q^1) = \mathrm{maxHYP}_\mu(\mathcal{Y})$, so P^1 maximizes the hypervolume. □

Note that Theorem 3.4 for finite objective spaces also holds for infinite objective spaces that have a hypervolume maximizing μ-population. In general, however, there is no μ-population maximizing the hypervolume on an infinite objective space, so that no statement as above holds.

Zitzler et al. [18, p. 71] pointed out that it is open, whether there are effective non-decreasing $(\mu+\lambda)$-archiving algorithms for $1 < \lambda < \mu$. We answer this question in the negative and prove the following theorem.

THEOREM 3.5. *There is no effective non-decreasing $(\mu+\lambda)$-archiving algorithm for $\lambda < \mu$.*

Again, we do not give a separate proof for Theorem 3.5 as it follows from its stronger counterpart Theorem 4.2 below. In order to prove Theorem 3.5 directly, one would show that there is an objective space and a suboptimal initial population P^0 such that any change of less than μ points of P^0 decreases the hypervolume indicator. However, the populations constructed that way have a hypervolume which is very close to the optimal one. Hence, the question arises of whether we at least arrive at a good approximation of the maximum hypervolume. We study this question in the following Section 4.

4. APPROXIMATE EFFECTIVENESS

Above negative results on the effectiveness raise the question of approximate effectiveness. To study this, we apply the following definition.

DEFINITION 4.1. *Let $\alpha \geqslant 1$. A $(\mu+\lambda)$-archiving algorithm $\mathcal{A}$ is* α-approximate *if for all sets $\mathcal{Y} \subset \mathbb{R}^d$ with finite* $\mathrm{maxHYP}_\mu(\mathcal{Y})$ *and μ-populations $P^0 \subseteq \mathcal{Y}$ there is an $N \in \mathbb{N}$ and a sequence of λ-populations $Q^1, \dots, Q^N \subseteq \mathcal{Y}$ such that*

$$\mathrm{HYP}(\mathcal{A}(P^0, Q^1, \dots, Q^N)) \geqslant \frac{1}{\alpha}\,\mathrm{maxHYP}_\mu(\mathcal{Y}).$$

We first examine what is the best approximation ratio we can hope for and prove a lower bound for the approximation ratio of all non-decreasing algorithms. To do so, we explicitly construct an objective space with two unconnected local maxima and show the following theorem. The full proof is given in Section 6.

THEOREM 4.2. *There is no $\left(1 + 0.1338\left(\frac{1}{\lambda} - \frac{1}{\mu}\right) - \varepsilon\right)$-approximate non-decreasing $(\mu+\lambda)$-archiving algorithm for any $\varepsilon > 0$.*

A bound of the form $1 + c(1/\lambda - 1/\mu)$ is very natural, as for $\lambda = \mu$ we get 1, and there is indeed an effective archiving algorithm in this case by Theorem 3.4. The proven constant, however, may be far from being tight. Maybe surprisingly, Theorem 4.2 indeed does not hold without the restriction to non-decreasing archiving algorithms. Theorem 5.1 in Section 5 shows that for all μ and λ there are archiving algorithms which are not non-decreasing, but effective.

Complementing the lower bound of Theorem 4.2, we can also prove the following upper bound on α. The proof of the following Theorem 4.3 can be found in Section 7.

THEOREM 4.3. *Let $\mathcal{A}$ be a non-decreasing $(\mu+\lambda)$-archiving algorithm. Assume that $\mathcal{A}$ is "improving if possible", i.e., for populations P and Q with*

$$\mathrm{maxHYP}_\mu(P \cup Q) > \mathrm{HYP}(P)$$

we have

$$\mathrm{HYP}(\mathcal{A}(P,Q)) > \mathrm{HYP}(P).$$

Then $\mathcal{A}$ is $(2+\varepsilon)$-approximate for any $\varepsilon > 0$.

In particular: All locally optimal $(\mu+\lambda)$-archiving algorithms are $(2+\varepsilon)$-approximate for any $\varepsilon > 0$.

The proof of Theorem 4.3 only considers offspring sets of size 1. Because of the respective lower bound of $(1.1338 - 0.1338/\mu)$ for $\lambda = 1$ from Theorem 4.2, this proof method can not show a better approximation ratio than a constant. We conjecture that a proof which handles offspring sets of size $\lambda > 1$ can show the following result.

CONJECTURE 4.4. *All locally optimal $(\mu+\lambda)$-archiving algorithms are $1 + \mathcal{O}(1/\lambda)$-approximate.*

5. WHY ONLY NON-DECREASING?

We finally want to discuss why we require all archiving algorithms to be non-decreasing. The reason is that otherwise none of the negative results and lower bounds from above hold as there is an effective $(\mu+\lambda)$-archiving algorithm for all $\mu, \lambda \in \mathbb{N}$. Such an algorithm is very simple: Given an ancestral population P and an offspring population Q, it returns the symmetric difference of both sets if this is not larger than μ and otherwise returns P directly. The algorithm is described in more detail in Algorithm 3. This algorithm is not non-decreasing, very unnatural, and does not guide in a sensible direction. However, for technical reasons one can prove the following statement.

THEOREM 5.1. *For any $\mu, \lambda \in \mathbb{N}$ there is an effective (not necessarily non-decreasing) $(\mu+\lambda)$-archiving algorithm.*

Proof. We study the following $(\mu+\lambda)$-archiving algorithm and prove that it is indeed effective.

Algorithm 3: An effective $(\mu+\lambda)$-archiving algorithm

```
P' := (P \ Q) ∪ (Q \ P)
if |P'| ⩽ μ then
    return P'
else
    return P
```

To show that Algorithm 3 is effective, let $\mathcal{Y} \subset \mathbb{R}^d$ be finite and $P^0 \subseteq \mathcal{Y}$ a μ-population. Moreover, let $P^* \subseteq \mathcal{Y}$ be a μ-population with $\mathrm{HYP}(P^*) = \mathrm{maxHYP}_\mu(\mathcal{Y})$. Write $P^0 \setminus P^* = \{p_1^0, \ldots, p_\mu^0\}$ (with possibly some of the p_i^0 being equal) and $P^* \setminus P^0 = \{p_1^*, \ldots, p_\mu^*\}$. Let $Q^{2i-1} = \{p_i^0\}$ and $Q^{2i} = \{p_i^*\}$ for $i = 1, \ldots, \mu$. On this offspring Algorithm 3 works as desired: After every second offspring generation one point of P^0 is replaced by a point from P^* so that after 2μ generations we arrive at P^*.

On the way there we even always have populations of size μ or $\mu - 1$ (as long as $|P| = |P^*| = \mu$). If we have $\lambda \geqslant 2$, we can even stick to populations of size μ by inserting every two offspring generations at once, i.e., $Q^1 \cup Q^2$, then $Q^3 \cup Q^4$, and so on. This shows that the possibility of the proven theorem does not stem from a faulty definition of μ-populations, as one might guess. ☐

This justifies why we always assume the archiving algorithms to be at least non-decreasing. Theorem 5.1 shows that Theorems 3.2, 3.5 and 4.2 do not hold for general archiving algorithms, which are not required to be non-decreasing.

6. PROOF OF THE LOWER BOUND

In this section we prove the lower bound of $\left(1 + 0.1338\left(\frac{1}{\lambda} - \frac{1}{\mu}\right) - \varepsilon\right)$ on the approximation factor of all non-decreasing $(\mu+\lambda)$-archiving algorithm as stated in Theorem 4.2. Note that this also implies that there is no effective non-decreasing $(\mu+\lambda)$-archiving algorithm for $\lambda < \mu$ as stated in Theorem 3.5.

Proof of Theorem 4.2. Let $\mu, \lambda \in \mathbb{N}$, $\lambda < \mu$. We construct an objective space $\mathcal{Y}$ and initial population P^0 as follows. Set $\mathcal{Y} = \{p_1, \ldots, p_{2\mu+1}\}$ with $p_i = (x_i, y_i)$ and

$$\begin{aligned} x_i &= \alpha^i - 1, && \text{for } i \text{ even},\\ y_i &= \alpha^{2\mu+2-i} - 1, && \text{for } i \text{ even},\\ x_i &= \gamma\alpha^i - 1, && \text{for } i \text{ odd},\\ y_i &= \gamma\alpha^{2\mu+2-i} - 1, && \text{for } i \text{ odd}, \end{aligned}$$

where $1 < \gamma < \alpha$. Figure 1 on the next page shows an illustration of the points for $\mu = 3$. Additionally, set $P^0 = \{p_2, p_4, \ldots, p_{2\mu}\}$. It is easy to see (but not needed for the proof) that $P^* = \{p_1, p_3, \ldots, p_{2\mu-1}\}$ maximizes the hypervolume on $\mathcal{Y}$. Alternatively, one could look at $P^* - p_1 + p_{2\mu+1}$.

We want to choose γ and α in such a way that P^0 is a local maximum from which one cannot escape exchanging only λ points. Thus, no non-decreasing selection policy with offspring size λ finds a better population than P^0. We then continue with proving that $\mathrm{HYP}(P^*)$ is sufficiently larger than the hypervolume of P^0.

For showing this, define $A := \mathrm{CON}_{\mathcal{Y}}(p_{2i})$ and $B := \mathrm{CON}_{\mathcal{Y}}(p_{2i+1})$. It can be easily seen that this is independent of the choice of i and that $A < B$. Moreover, we consider the area dominated by both, p_{2i} and p_{2i+1}, namely $C := \mathrm{HYP}(\mathcal{Y}) - \mathrm{HYP}(\mathcal{Y} - p_{2i} - p_{2i+1}) - A - B$. Those areas are depicted in Figure 1. Observe that this is again independent of i and one gets the same area considering p_{2i} and p_{2i-1}.

Now, let $Q^1 \subseteq \mathcal{Y}$ be a λ-population and consider any μ-population $P^1 \subseteq P^0 \cup Q^1$ with $P^0 \neq P^1$. We want to choose α and γ in such a way that $\Delta\mathrm{HYP} := \mathrm{HYP}(P^1) - \mathrm{HYP}(P^0) < 0$, so that we have to stick to P^0. For this, let $H := \mathrm{HYP}(\mathcal{Y})$, so that we have $\mathrm{HYP}(P^0) = H - (\mu+1)B$. For P^1, observe that there is an index i with $p_i, p_{i+1} \notin P^1$ (as otherwise $P^1 = P^0$). These two points dominate together an area of C that is not dominated by P^1. Moreover, every point $p_i \in \mathcal{Y}$, $p_i \notin P^1$ adds another A or B to $H - \mathrm{HYP}(P^1)$, depending on i being even or odd. Letting k be the number of points of odd index in P^1 we thus have

$$\mathrm{HYP}(P^1) \leqslant H - C - (\mu + 1 - k)\,B - k\,A.$$

Thus, we have $\Delta\mathrm{HYP} \leqslant k(B - A) - C$. As the offspring size

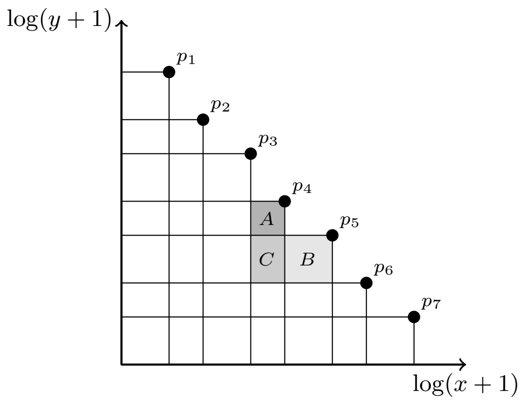

Figure 1: A schematic log-log **plot of the example used in the proof of Theorem 4.2. The considered areas** A, B, C **are indicated.**

$|Q^1| \leqslant \lambda$ we have $k \leqslant \lambda$ and thus

$$\Delta\text{HYP} \leqslant \lambda(B - A) - C.$$

We want to choose α and γ such that the right hand side from above is less than 0. We compute

$$\begin{aligned} A &= (x_{2i} - x_{2i-1})(y_{2i} - y_{2i+1}) \\ &= (\alpha^{2i} - \gamma\alpha^{2i-1})(\alpha^{2\mu+2-2i} - \gamma\alpha^{2\mu+2-2i-1}) \\ &= \alpha^{2\mu+2}(1 - \gamma/\alpha)^2. \end{aligned}$$

Similarly, we see that

$$\begin{aligned} B &= \alpha^{2\mu+2}(\gamma - 1/\alpha)^2, \\ B - A &= \alpha^{2\mu+2}(\gamma^2 - 1)(1 - 1/\alpha^2), \\ C &= \alpha^{2\mu+1}(1 - \gamma/\alpha)(\gamma - 1/\alpha). \end{aligned}$$

Now, $\lambda(B - A) - C < 0$ turns into a quadratic inequality in γ. We solve it and get

$$\gamma < \frac{\alpha^2 + 1 + (\alpha^2 - 1)\sqrt{4\alpha^2\lambda^2 + 1}}{2\alpha(\lambda(\alpha^2 - 1) + 1)}. \tag{1}$$

Simple calculations show that this bound is always greater than 1 and less equal α (at least for $\alpha \geqslant 2$, $\lambda \geqslant 1$ this is easy to show). Hence, there is no contradiction to $\gamma > 1$ and we can choose γ arbitrarily close to the right hand side from above. Thus, for $\alpha \geqslant 2$ and $\gamma > 1$ satisfying equation (1) no $(\mu + \lambda)$-archiving algorithm can escape from P^0.

All that is left to show is that $\text{HYP}(P^*)$ is sufficiently greater than $\text{HYP}(P^0)$. Above we saw that $\text{HYP}(P^0) = H - (\mu + 1)B$, where $H = \text{HYP}(\mathcal{Y})$. Now, observe that $\text{HYP}(P^*) = H - \mu A - B - C$, where the B stems from $p_{2\mu+1}$ not being in P^* and the C from $p_{2\mu+1}$ and $p_{2\mu}$ not being in P^*. We, thus, have

$$\Delta\text{HYP} := \text{HYP}(P^*) - \text{HYP}(P^0) = \mu(B - A) - C.$$

Let $\varepsilon > 0$. By choosing γ (dependent on α) sufficiently near to the right hand side of equation (1) we have $0 > \lambda(B - A) - C \geqslant -\varepsilon$ and, hence,

$$\begin{aligned} \Delta\text{HYP} &\geqslant (\mu - \lambda)(B - A) - \varepsilon \\ &= (\mu - \lambda)\alpha^{2\mu+2}(\gamma^2 - 1)(1 - 1/\alpha^2) - \varepsilon. \end{aligned}$$

We compute $\text{HYP}(P^0)$ as follows, where we set $x_0 := 0$:

$$\begin{aligned} \text{HYP}(P^0) &= \sum_{i=1}^{\mu} (x_{2i} - x_{2(i-1)})\, y_{2i} \\ &= \sum_{i=1}^{\mu} (\alpha^{2i} - \alpha^{2(i-1)})\, \alpha^{2\mu+2-2i} \\ &= \mu\alpha^{2\mu+2}(1 - 1/\alpha^2). \end{aligned}$$

Now, the approximation ratio of any $(\mu+\lambda)$-archiving algorithm on $\mathcal{Y}$ with initial population P^0 is, as it cannot escape P^0,

$$\begin{aligned} \frac{\max\text{HYP}_\mu(\mathcal{Y})}{\text{HYP}(P^0)} &\geqslant \frac{\text{HYP}(P^*)}{\text{HYP}(P^0)} \\ &= 1 + \frac{\Delta\text{HYP}}{\text{HYP}(P^0)} \\ &\geqslant 1 + (1 - \frac{\lambda}{\mu})(\gamma^2 - 1) - \varepsilon, \end{aligned}$$

for $\alpha \geqslant \sqrt{2}$ and, thus, $\text{HYP}(P^0) \geqslant 1$, so that we can bound $\varepsilon/\text{HYP}(P^0) \leqslant \varepsilon$. For maximizing the right hand side we will plug in $\alpha = 1 + \sqrt{6}$, so that γ is bounded from above and below by constants. This way, choosing γ sufficiently near to the right hand side of equation (1), we get

$$\begin{aligned} &\frac{\max\text{HYP}_\mu(\mathcal{Y})}{\text{HYP}(P^0)} \geqslant 1 - 2\varepsilon \\ &+ \left(1 - \frac{\lambda}{\mu}\right)\left(\left(\frac{\alpha^2 + 1 + (\alpha^2 - 1)\sqrt{4\alpha^2\lambda^2 + 1}}{2\alpha(\lambda(\alpha^2 - 1) + 1)}\right)^2 - 1\right) \end{aligned}$$

We consider the bracket on the right hand side separately. This is

$$\begin{aligned} &\left(\frac{\alpha^2 + 1 + (\alpha^2 - 1)\sqrt{4\alpha^2\lambda^2 + 1}}{2\alpha(\lambda(\alpha^2 - 1) + 1)}\right)^2 - 1 \\ &= \frac{(\alpha^2 + 1 + (\alpha^2 - 1)\sqrt{4\alpha^2\lambda^2 + 1})^2 - 4\alpha^2(\lambda(\alpha^2 - 1) + 1)^2}{4\alpha^2(\lambda(\alpha^2 - 1) + 1)^2} \\ &\geqslant \frac{(\alpha^2 + 1 + (\alpha^2 - 1)\sqrt{4\alpha^2\lambda^2})^2 - 4\alpha^2(\lambda(\alpha^2 - 1) + 1)^2}{4\alpha^2(\lambda(\alpha^2 - 1) + \lambda)^2} \\ &= \frac{(\alpha^2 - 1)^2 + 4(\alpha - 1)^3\alpha(\alpha + 1)\lambda}{4\alpha^6\lambda^2}. \end{aligned}$$

Plugging this in and simplifying, we get

$$\begin{aligned} &\frac{\max\text{HYP}_\mu(\mathcal{Y})}{\text{HYP}(P^0)} \\ &\geqslant 1 - 2\varepsilon + \frac{(\mu - \lambda)(\alpha - 1)^2(\alpha + 1)(\alpha + 1 + 4(\alpha - 1)\alpha\lambda)}{4\alpha^6\lambda^2\mu} \\ &\geqslant 1 - 2\varepsilon + \frac{(\mu - \lambda)(\alpha - 1)^2(\alpha + 1)(4(\alpha - 1)\alpha\lambda)}{4\alpha^6\lambda^2\mu}. \end{aligned}$$

Now, the right hand side gets maximal for $\alpha = 1 + \sqrt{6}$. Plugging this in we get

$$\begin{aligned} \frac{\max\text{HYP}_\mu(\mathcal{Y})}{\text{HYP}(P^0)} &\geqslant 1 + \frac{12(3 + \sqrt{6})}{(1 + \sqrt{6})^5}\left(\frac{1}{\lambda} - \frac{1}{\mu}\right) - 2\varepsilon \\ &\geqslant 1 + 0.1338 \cdot \left(\frac{1}{\lambda} - \frac{1}{\mu}\right) - 2\varepsilon. \end{aligned}$$

This finishes the proof. □

7. PROOF OF THE UPPER BOUND

In this section we prove the upper bound of $2 + \varepsilon$ on the approximation factor of all $(\mu + \lambda)$-archiving algorithms that choose a population of children with greater hypervolume than the parent population (if there is one), as stated in Theorem 4.3. We conjecture that the upper bound of $2 + \varepsilon$ on the approximation factor is not tight and it might be possible to improve it asymptotically to 1.

Proof of Theorem 4.3. Let $\varepsilon > 0$, $\mathcal{Y} \subset \mathbb{R}^d$ with $\mathrm{maxHYP}_\mu(\mathcal{Y}) < \infty$ and $P^0 \subseteq \mathcal{Y}$ be a μ-population. By definition of maxHYP there exists a μ-population $P^* \subseteq \mathcal{Y}$ with $\mathrm{HYP}(P^*) \geqslant (1 + \varepsilon/2)^{-1}\, \mathrm{maxHYP}_\mu(\mathcal{Y})$.

We try to add the points of P^* one-by-one to the current population, starting with P^0. This means we set $Q^1 \cup \ldots \cup Q^\mu = P^*$ with $|Q^i| \leqslant 1$. After one run through all these Q^i the population P^μ might have changed from P^0. If this is the case we run again through these Q^i, until no change happens. This requires at most μ runs through the Q^i, if the archiving algorithm fulfills the premise, so by setting $N = \mu^2$ and $Q^{i+k\mu} = Q^i$ for $i, k \in \mathbb{N}$ we get to a population P^N which is stable under insertions of single points from P^*. We show that in this case $\mathrm{HYP}(P^N) \geqslant \frac{1}{2}\mathrm{HYP}(P^*)$.

We have $|P^N| \leqslant \mu$, but consider any point $p \in P^*$ not dominated by any point in P^N. If $|P^N| < \mu$ then $P^N + p$ has greater hypervolume, which contradicts P^N being stable under insertions of single points from P^*. Thus, either $|P^N| = \mu$ or there is no such point p, in which case $\mathrm{HYP}(P^N) \geqslant \mathrm{HYP}(P^*) \geqslant (1 + \varepsilon/2)^{-1}\, \mathrm{maxHYP}_\mu(\mathcal{Y})$, which proves the claim in this case. Hence, we can assume $|P^N| = \mu$.

Consider a point $\tilde{p} \in P^N$ with minimal contribution $\mathrm{CON}_{P^N}(p)$. The contribution of a point p measures the volume of the space that is dominated by p alone, so that we have

$$\sum_{p \in P^N} \mathrm{CON}_{P^N}(p) \leqslant \mathrm{HYP}(P^N).$$

Now, as $\tilde{p}$ has minimal contribution and $|P^N| = \mu$, we get

$$\mathrm{CON}_{P^N}(\tilde{p}) \leqslant \frac{1}{\mu} \sum_{p \in P^N} \mathrm{CON}_{P^N}(p).$$

Let $q \in P^*$ and consider $P' := P^N - \tilde{p} + q$. This population has $\mathrm{HYP}(P') \leqslant \mathrm{HYP}(P^N)$, as P^N is stable under insertions of single points from P^*. Thus, we have

$$\begin{aligned}\mathrm{HYP}(P') &= \mathrm{CON}_{P'}(q) - \mathrm{CON}_{P^N}(\tilde{p}) + \mathrm{HYP}(P^N)\\ &\leqslant \mathrm{HYP}(P^N),\end{aligned}$$

which gives

$$\mathrm{CON}_{P'}(q) \leqslant \mathrm{CON}_{P^N}(\tilde{p}).$$

Additionally, as $P' \subset P^N + q$ we have

$$\mathrm{CON}_{P^N+q}(q) \leqslant \mathrm{CON}_{P'}(q).$$

Together, these inequalities imply

$$\mathrm{CON}_{P^N+q}(q) \leqslant \frac{1}{\mu}\mathrm{HYP}(P^N).$$

Now, consider the population $P^N \cup P^*$ (of size in $[\mu, 2\mu]$). We have $\mathrm{HYP}(P^N \cup P^*) \geqslant \mathrm{HYP}(P^*)$. Moreover, we can write the former as

$$\begin{aligned}\mathrm{HYP}(P^N \cup P^*) &= \mathrm{HYP}(P^N) + \sum_{i=1}^{\mu} \mathrm{CON}_{P^N+q_1+\ldots+q_i}(q_i)\\ &\leqslant \mathrm{HYP}(P^N) + \sum_{i=1}^{\mu} \mathrm{CON}_{P^N+q_i}(q_i)\\ &\leqslant \mathrm{HYP}(P^N) + \sum_{i=1}^{\mu} \frac{1}{\mu}\,\mathrm{HYP}(P^N)\\ &\leqslant 2\mathrm{HYP}(P^N).\end{aligned}$$

This yields the desired inequality

$$\mathrm{HYP}(P^N) \geqslant \frac{1}{2}\,\mathrm{HYP}(P^*) \geqslant (2 + \varepsilon)^{-1}\,\mathrm{maxHYP}_\mu(\mathcal{Y}).$$

and finishes the proof. □

8. DISCUSSION AND FUTURE WORK

We have studied theoretically the convergence of archiving algorithms used in hypervolume-based multi-objective evolutionary algorithms. We proved that non-decreasing $(\mu + \lambda)$-archiving algorithms can only be effective for $\lambda \geqslant \mu$ and that they cannot achieve an approximation of the maximum hypervolume by a factor of more than $1/(1 + 0.1338\,(1/\lambda - 1/\mu))$. On the positive side, we proved that the popular (but computationally very expensive) locally optimal algorithms are effective for $\lambda = \mu$ and can always find a population with hypervolume at least half the optimum for $\lambda < \mu$. We conjecture that the lower bound of one half can be improved to a value which is asymptotically 1, which is a natural follow-up question.

Let us further remark that it makes no sense to take a best-case view on the initial population (then the initial population already maximizes the hypervolume) nor does it make sense to take a best-case view on the objective space (then it contains only the population maximizing the hypervolume). The only two deterministic approaches for the offspring generation are best-case and worst-case. Following the work of Zitzler et al. [18], we have studied the best-case scenario in detail. However, we believe it is equally interesting to study the worst-case. Thus, in the future we plan to examine archiving algorithms in a setting where an adversary chooses the offspring.

It might also be interesting to study average-case models. The disadvantage of this is that it introduces the necessity of modelling other parts of the (until now) abstract EMO algorithm, namely the generation of the initial population and offspring as well as the structure of the search space.

Note that our approach for analyzing $(\mu + \lambda)$-archiving algorithms does not apply to comma strategies, i.e., (μ, λ)-MOEAs, where the archiving algorithm can only choose from the offspring points. The reason is that in this case (in best- as well as worst-case offspring generation) the final population $P^N = \mathcal{A}(P^0, Q^1, \ldots, Q^N)$ only depends on the last offspring Q^N and hence every locally optimal archiving algorithm is trivially as effective as it can be, which is not an interesting result.

Acknowledgements

We thank Lothar Thiele for pointing us to Theorem 3.2 and posing the question whether Theorem 3.5 or the opposite holds.

References

[1] N. Beume, B. Naujoks, and M. T. M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181: 1653–1669, 2007.

[2] N. Beume, B. Naujoks, M. Preuss, G. Rudolph, and T. Wagner. Effects of 1-greedy $\mathcal{S}$-metric-selection on innumerably large Pareto fronts. In *Proc. 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO '09)*, Vol. 5467 of *LNCS*, pp. 21–35, 2009.

[3] K. Bringmann and T. Friedrich. Approximating the volume of unions and intersections of high-dimensional geometric objects. *Computational Geometry: Theory and Applications*, 43:601–610, 2010.

[4] K. Bringmann and T. Friedrich. Approximating the least hypervolume contributor: NP-hard in general, but fast in practice. In *Proc. 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO '09)*, pp. 6–20. Springer, 2009.

[5] D. W. Corne and J. D. Knowles. Some multiobjective optimizers are better than others. In *Proc. Congress on Evolutionary Computation (CEC '03)*, pp. 2506–2512. IEEE Press, 2003.

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

[7] C. Igel, N. Hansen, and S. Roth. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15:1–28, 2007.

[8] J. Knowles and D. Corne. *Metaheuristics for Multiobjective Optimisation*, Vol. 535 of *Lecture Notes in Economics and Mathematical Systems*, chapter Bounded Pareto Archiving: Theory and Practice, pp. 39–64. Springer, 2004.

[9] J. D. Knowles and D. Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Trans. Evolutionary Computation*, 7:100–116, 2003.

[10] J. D. Knowles, D. W. Corne, and M. Fleischer. Bounded archiving using the Lebesgue measure. In *Proc. IEEE Congress on Evolutionary Computation (CEC '03)*, Vol. 4, pp. 2490–2497. IEEE Press, 2003.

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

[12] M. López-Ibáñez, J. D. Knowles, and M. Laumanns. On sequential online archiving of objective vectors. In *Proc. 6th International Conference on Evolutionary Multi-Criterion Optimization (EMO '11)*, Vol. 6576 of *LNCS*, pp. 46–60. Springer, 2011.

[13] G. Rudolph and A. Agapie. Convergence properties of some multi-objective evolutionary algorithms. In *Proc. Congress on Evolutionary Computation (CEC '00)*, pp. 1010–1016. IEEE Press, 2002.

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

[15] T. Suttorp, N. Hansen, and C. Igel. Efficient covariance matrix update for variable metric evolution strategies. *Machine Learning*, 75:167–197, 2009.

[16] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Proc. 8th International Conference Parallel Problem Solving from Nature (PPSN VIII)*, Vol. 3242 of *LNCS*, pp. 832–842. Springer, 2004.

[17] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans. Evolutionary Computation*, 3:257–271, 1999.

[18] E. Zitzler, L. Thiele, and J. Bader. On set-based multiobjective optimization. *IEEE Trans. Evolutionary Computation*, 14:58–79, 2010.