Convergence of Hypervolume-Based Archiving Algorithms

Karl Bringmann and Tobias Friedrich

***Abstract*— Multi-objective evolutionary algorithms typically maintain a set of solutions. A crucial part of these algorithms is the archiving, which decides what solutions to keep. A $(\mu+\lambda)$-archiving algorithm defines how to choose in each generation μ children from μ parents and λ offspring together. We study mathematically the convergence behavior of hypervolume-based archiving algorithms. We distinguish two cases for the offspring generation.**

A best-case view leads to a study of the effectiveness of archiving algorithms. It was known that all $(\mu+1)$-archiving algorithms are ineffective, which means that a set with maximum hypervolume is not necessarily reached. We prove that for $\lambda < \mu$ all archiving algorithms are ineffective. We also present upper and lower bounds for the achievable hypervolume for different classes of archiving algorithms.

On the other hand, a worst-case view on the offspring generation leads to a study of the competitive ratio of archiving algorithms. This measures how much smaller hypervolumes are achieved due to not knowing the future offspring in advance. We present upper and lower bounds on the competitive ratio of different archiving algorithms and present an archiving algorithm, which is the first known computationally efficient archiving algorithm with constant competitive ratio.

***Index Terms*—Optimization methods, Multiobjective optimization, Performance measures, Selection, Hypervolume indicator.**

I. Introduction

Many real-world optimization problems have multiple objectives like time vs. cost. This implies that in general there is no unique optimum, but an often very large (or even infinite) set of incomparable solutions which form the Pareto front. Multi-objective optimizers deal with this by trying to find a small set of trade-off solutions which approximate the Pareto front. They typically keep a bounded archive of μ points (population) in order to capture the output of the search process. In each round they generate λ new points (offspring) by mutation and crossover. The key question is then how to select μ individuals from a larger population. We consider the so-called plus selection strategy, where the next population is chosen out of the λ offspring and μ parents together. We call a specific replacement strategy 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.

The goal for *hypervolume-based* multi-objective evolutionary algorithms (MOEAs) is to maximize the hypervolume indicator of the output population, which is the volume of the dominated portion of the objective space (see Section II for a formal definition). For this type of MOEA, two archiving algorithms are known in the literature:

- A *locally optimal* archiving algorithm returns a subset of μ points from the given $\mu+\lambda$ points such that the hypervolume indicator is maximized.
- A *greedy* archiving algorithm deletes a point such that the hypervolume of the remaining points is maximal. This is repeated until only μ points are left.

Many hypervolume based algorithms like SIBEA [17], SMS-EMOA [1], or the generational MO-CMA-ES [10, 11] use greedy archiving algorithms. As locally optimal algorithms have to choose the best out of a large number, $\binom{\mu+\lambda}{\mu}$, of subsets of the given points, they are generally considered to be computationally infeasible. Note that a locally optimal archiving algorithm in general does not maximize the hypervolume over multiple generations. However, it still seems to have superior theoretical properties: It has long been known that the resulting point sets of both algorithms differ [3], and that the deleted hypervolume (the contribution of the deleted points) can even be arbitrarily larger for greedy archiving algorithms compared to locally optimal algorithms [6]. We prove in this paper that all locally optimal and all greedy archiving algorithms have to solve NP-hard problems (cf. Theorem IV.1 and Observation II.8). Hence, such algorithms are not computationally efficient unless P = NP.

We want to study the intrinsic limitations of and the potential provided by hypervolume-based archiving algorithms. Beyond the smaller classes of locally optimal and greedy archiving algorithms we thus also consider the following two natural classes of archiving algorithms:

- A *non-decreasing* archiving algorithm chooses the population of children such that the dominated hypervolume does not decrease compared to the parent generation.
- An *increasing* archiving algorithm chooses the population of children such that the dominated hypervolume increases compared to the parent generation, unless there is no subset of population and offspring with a larger dominated hypervolume.

Both are intuitively desirable properties for hypervolume-based archiving algorithms. We will see that there are algorithms which are non-decreasing, but not increasing (cf. Algorithms 4 and 5). Moreover, we prove that both classes significantly differ. There are non-decreasing archiving algorithms which are better and faster than all increasing archiving algorithms (see Sections I-B and I-C for more detailed statements).

Karl Bringmann is with Max Planck Institute for Informatics, 66123 Saarbrücken, Germany. E-mail: `karl.lastname@mpi-inf.mpg.de`

Tobias Friedrich is with Friedrich-Schiller-Universität Jena, 07743 Jena, Germany. E-mail: `lastname@uni-jena.de`

Manuscript received xxx; revised xxx.

To rigorously study the impact of archiving algorithms on convergence, we cannot concentrate only on single iterations, but have to consider multiple generations of populations. We model this long run behavior with the initial population being worst-case input to the archiving algorithm, followed by some kind of offspring generation, and we then ask whether we arrive at a population with a large hypervolume. Note that it makes no sense to take a best-case view on the initial population as then the initial population already maximizes the hypervolume. It is also not meaningful to take a best-case view on the objective space as this implies that it contains only the population maximizing the hypervolume. There are two natural assumptions on the offspring generation: best-case and worst-case. A best-case offspring generation is always 'lucky', that is, we ask whether there exists a sequence of offspring sets such that the archiving algorithm ends up in a population maximizing the hypervolume. On the other hand, a worst-case offspring generation is always 'unlucky', that is, we assume an adversary selects the offspring and ask how close the achieved hypervolume of an archiving algorithm gets compared to the achievable hypervolume if we had known which offspring would come in the future.

Assuming a best-case or worst-case view allows us studying archiving algorithms independent of specific variation operators. Both assumptions give rise to interesting results (see Sections I-A and I-B). Negative results for the best-case are very general as they show the limitations of *all* archiving algorithms. On the other hand, an algorithm with proven worst-case performance works for every offspring generation and therefore has a *guarantee* for all possible scenarios.

We summarize our results in Sections I-A, I-B and I-C. In Section II we introduce the basic concepts and notation. Section III gives some technical basics regarding the hypervolume. Section IV studies the computational complexity of increasing archiving algorithms. The main results are afterwards presented in Sections V and VI. In Section V we consider a best-case choice of the offspring and analyze which archiving algorithms are *effective*. In Section VI we consider a worst-case view on offspring generation and study the *competitiveness* of archiving algorithms.

This paper extends previous results of two conference papers of the authors [8, 9] in several directions. We present in Section III several new basic properties of the hypervolume indicator. We also introduce in Section VI-D a new technique for transferring approximation lower bounds to lower bounds for competitiveness, and present a number of other new results (e.g. Theorem V.4). Moreover, after the publication of [8], Ulrich and Thiele [15] presented an improved upper bound on the approximation achieved by increasing archiving algorithms. We now prove in Theorem V.8 an upper bound which is again stronger than the one of Ulrich and Thiele [15].

A. Results on effectiveness

Most previous work in this setting [15, 18] assumes a best-case perspective on the offspring generation. This means that we ask whether, for each population, there exists a sequence of offspring sets such that the archiving algorithm ends up in a population maximizing the hypervolume. This can be formalized with the notion of effectiveness: An archiving algorithm is *effective* if there is a sequence of offspring such that the algorithm reaches an optimum. Zitzler et al. [18] proved that all non-decreasing $(\mu+1)$-archiving strategies are ineffective (cf. Theorem V.2) while there are effective non-decreasing $(\mu+\mu)$-archiving algorithms (cf. Theorem V.3). We additionally prove in Theorem V.4 that all increasing $(\mu+\mu)$-archiving strategies are effective. Zitzler et al. [18] left open what happens for general $(\mu+\lambda)$-archiving algorithms. We answer this with Theorem V.5 and prove that all non-decreasing $(\mu+\lambda)$-archiving strategies are ineffective for $\lambda<\mu$.

In order to measure how close to an optimal set the best reachable sets for $\lambda<\mu$ are, we call an archiving algorithm α-*approximate* if it can always reach a set with a hypervolume at least $1/\alpha$ times the largest possible hypervolume. We prove in Theorem V.7 that no non-decreasing $(\mu+\lambda)$-archiving algorithm can be better than $\big(1+0.1338\,\big(\frac{1}{\lambda}-\frac{1}{\mu}\big)-\varepsilon\big)$-approximate for any $\varepsilon>0$. This bound can be tightened for a relaxed variant of the hypervolume, which is defined relative to a reference set instead of a single reference point. For this less restrictive setting, Ulrich and Thiele [15] showed a lower bound of $1+\frac{1}{2\lambda}$ for $\lambda<\mu$.

On the other hand, the authors [8, Thm. 4.3] showed that every increasing $(\mu+\lambda)$-archiving algorithm reaches a $(2+\varepsilon)$-approximation for any $\varepsilon>0$. Using that the hypervolume indicator is non-decreasing submodular, this upper bound was improved by Ulrich and Thiele [15]. We now again improve their results and show in Theorem V.8 that every increasing $(\mu+\lambda)$-archiving algorithm reaches a $\big(2-\frac{\lambda}{\mu}+\varepsilon\big)$-approximation for any $\varepsilon>0$.

B. Results on competitiveness

We can also assume a worst-case perspective on both the initial population and offspring generation. This corresponds to the well-known concept of *competitive analysis*. It has already been observed that archiving algorithms fit nicely in this classical theory developed for online algorithms [2]. López-Ibáñez, Knowles, and Laumanns [12, p. 59] suggested it as an open problem "to use competitive analysis techniques from the field of online algorithms to obtain worst-case bounds, in terms of a measure of 'regret' for archivers."

We consider the initial population and offspring as worst-case input and ask again how large a hypervolume we can get. In this case, however, the adversary, who selects the offspring, can limit the search to a very small part of the search space, and it is therefore impossible in general to reach the optimum hypervolume. This motivates the following definition. We say an archiving algorithm is α-*competitive* if for all initial populations and offspring it reaches a hypervolume which is only a factor $1/\alpha$ smaller than the hypervolume of the *best* μ *points seen* (cf. Definition VI.1).

On the negative side, we prove that all increasing archiving algorithms are at best μ-competitive (cf. Theorem VI.3). This means that there is a sequence of offspring such that the hypervolume of the μ individuals chosen iteratively by an algorithm which maximizes the hypervolume in each step is μ times larger than the maximum hypervolume achievable by another choice of μ individuals. This lower bound of μ on the competitive ratio is in fact tight for all locally optimal algorithms and all increasing $(\mu + 1)$-archiving algorithms (cf. Theorem VI.2). This implies that the notion of competitiveness measures no difference between all archiving algorithms of these two classes as they meet exactly the same bound.

However, on the positive side, we are able to design an archiving algorithm that is $4 + 2/\mu$-competitive (cf. Theorem VI.4), which implies a constant competitive ratio compared to the unbounded ratio of μ from above. It is a non-decreasing archiving algorithm which is not increasing, i.e., there are populations and offspring where we stay with the current population, although the offspring allows an increase in hypervolume. This proves that significantly better competitive ratios can be achieved for archiving algorithms which are not increasing compared to the typically used increasing archiving algorithms. The algorithm works as follows (for details see Algorithm 4): It adds offspring one by one to the current population. Considering the population and an offspring, we compute the hypervolume for exchanging the offspring with any other point in the population. We take the best exchange only if it increases the population's hypervolume by at least a certain minimal factor.

C. Results on computational efficiency

We prove that all increasing archiving algorithms solve an NP-hard problem (cf. Theorem IV.1), assuming that the number of dimensions is part of the input. This implies that all common greedy archiving algorithms are not computationally efficient for unbounded dimension unless P = NP (cf. Observation II.8). This still allows archiving algorithms which are not increasing to be computationally efficient. Indeed, we also prove that a randomized variant of our aforementioned $4+2/\mu$-competitive archiving algorithm can be made to run efficiently (cf. Theorem VI.6). Note that this is in sharp contrast to the large set of increasing archiving algorithms, which all have a worse competitive ratio (cf. Theorem VI.3) and a worse computational complexity (cf. Theorem IV.1) compared to the proposed new archiving algorithm. The underlying reason why the new algorithm can beat all increasing archiving algorithms is that approximating the hypervolume is tractable even in high dimensions [5] and for the new algorithm it is sufficient only to approximate the hypervolume, as it checks only for constant factor increases. Although the new archiving algorithm might not be used as-is in any practical MOEAs, it is a proof of concept that there are computationally efficient archiving algorithms which can beat the competitive ratio of the thus far typically used locally optimal and greedy archiving algorithms.

II. Preliminaries

This section formally introduces all necessary notation. The two most fundamental concepts are the hypervolume indicator (Section II-A) and archiving algorithms (Section II-B). The combination of both, i.e. hypervolume-based archiving algorithms, are introduced in Section II-C.

We consider maximization problems with vector-valued objective functions

$$f\colon \mathcal{X} \to \mathbb{R}^d,$$

where $\mathcal{X}$ denotes an arbitrary search space. The feasible points $\mathcal{Y} := f(\mathcal{X})$ are called the objective space. Consider the following abstract framework of a MOEA:

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

```
1 P^0 ← initialize with μ individuals
2 for i ← 1 to N do
3     Q^i ← generate λ offspring
4     P^i ← select μ individuals from P^{i−1} ∪ Q^i
```

We want to make no assumptions about 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 both the search space and the initialization are worst-case, and we assume that offspring generation is either best-case (see Section V) or worst-case (see Section VI). Our main concern is how the population of children is chosen (cf. line 4 of Algorithm 1). We will formally define and discuss different *archiving algorithms* in Sections II-B and II-C.

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.

Definition II.1. *A population P is a finite multiset and a 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 $|P| \leqslant \mu$.*

A. Hypervolume indicator

The hypervolume indicator $\text{HYP}(P)$ [16] of a finite set $P \subset \mathbb{R}^d$ is the volume of the union of regions of the objective space which are dominated by P and bounded by a reference point $r = (r_1, \ldots, r_d)$. More precisely, for $p = (p_1, \ldots, p_d) \in P$ define $\text{box}(p) := [r_1, p_1] \times \ldots \times [r_d, p_d]$ (which is only defined if $p_i \geqslant r_i$ for all i). Then

$$\text{HYP}(P) := \text{VOL}\Big(\bigcup_{p \in P} \text{box}(p)\Big),$$

where VOL is the usual Lebesgue measure in $\mathbb{R}^d$. Computing $\text{HYP}(P)$ requires time $n^{\Omega(d)}$ [4] (unless the exponential time hypothesis fails), but can be approximated very efficiently in polynomial time [5].

We 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. Additionally, HYP is now defined for any finite point set $P \subset \mathbb{R}^d_+$. Here and throughout the paper, we denote by $\mathbb{R}_+$ the positive real numbers, and we will assume that the objective space $\mathcal{Y}$ is a subset of $\mathbb{R}^d_+$ from now on.

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

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

where we define for all $Y \subseteq \mathbb{R}^d_+$,

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

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

The *contribution* of a point p to a population P is

$$\mathrm{CON}_P(p) := \mathrm{HYP}(P + p) - \mathrm{HYP}(P - p).$$

Here and throughout the paper, we use the notation $P + p$ for $P \cup \{p\}$ and $P - p$ for $P \setminus \{p\}$. Note that this definition of the contribution makes sense for $p \in P$ (in which case it is the hypervolume we lose by deleting p from P) as well as for $p \notin P$ (in which case it is the hypervolume we gain by adding p to P). Also note that according to the definition of $\mathrm{CON}_P(p)$, the contributing hypervolume of a dominated individual is zero. We further generalize CON for any finite $P, Q \subset \mathbb{R}^d_+$ by setting

$$\mathrm{CON}_P(Q) := \mathrm{HYP}(P \cup Q) - \mathrm{HYP}(P \setminus Q).$$

B. Archiving algorithms

We now specify more formally how to choose the μ individuals of the succeeding population 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 [5], 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 II.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 (or even shorter: *algorithm*) without specifying μ and λ. With this notation, we can now formally describe the generation process of Algorithm 1 as follows.

Definition II.3. *Let P^0 be a μ-population and $Q^1, \ldots, Q^N$ a sequence of λ-populations. Then we set*

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

With slight abuse of notation we also set

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

C. Hypervolume-based archiving algorithms

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

Definition II.4. *A $(\mu+\lambda)$-archiving algorithm $\mathcal{A}$ is* non-decreasing*, if for all inputs P and Q we have*

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

Most hypervolume-based archiving algorithms are non-decreasing. However, the class also contains ineffective algorithms like the algorithm which always returns P.

The second, slightly smaller class of hypervolume-based archiving algorithms is defined as follows.

Definition II.5. *A $(\mu+\lambda)$-archiving algorithm $\mathcal{A}$ is* increasing*, if it is non-decreasing and for all inputs P and Q with*

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

we have

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

Moreover, we define locally optimal and greedy archiving algorithms. Note that for both classes there is more than one archiving algorithm fulfilling the respective definition, as ties may be broken arbitrarily.

Definition II.6. *A $(\mu+\lambda)$-archiving algorithm $\mathcal{A}$ is* locally optimal*, if for all inputs P and Q we have*

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

Definition II.7. *A $(\mu+\lambda)$-archiving algorithm $\mathcal{A}$ is* greedy*, if there are functions $\mathcal{A}', a'$, $\mathcal{A}'(P) = P - a'(P)$ with $a'(P) \in \mathrm{argmin}_{p \in P} \mathrm{CON}_P(p)$ such that for all inputs P and Q we have*

$$\mathcal{A}(P,Q) = \underbrace{\mathcal{A}' \circ \ldots \circ \mathcal{A}'}_{\lambda \text{ times}}(P \cup Q).$$

The rest of the paper focuses on increasing and non-decreasing archiving algorithms. Their relation to locally optimal and greedy archiving algorithms is as follows.

Observation II.8. *Greedy $(\mu+1)$-archiving algorithms and locally-optimal $(\mu+\lambda)$-archiving algorithms are increasing archiving algorithms. Greedy $(\mu+\lambda)$-archiving algorithms are not necessarily non-decreasing archiving algorithms for $\lambda > 1$.*

This observation allows us to translate all forthcoming bounds for increasing (or non-decreasing) archiving algorithms to locally-optimal archiving algorithms and greedy $(\mu+1)$-archiving algorithms. Moreover, since the computational hardness (cf. Theorem IV.1) and the lower bound for the competitive ratio (cf. Theorem VI.3) apply to the restriction of a greedy algorithm to $\lambda = 1$, they also apply to greedy algorithms in general. However, some of our results do not hold for greedy algorithms when $\lambda > 1$.

Consider the following variant of greedy algorithms. A *non-decreasing greedy* archiving algorithm takes the output $P' = \mathcal{A}(P, Q)$ of a *greedy* archiving algorithm $\mathcal{A}$ and returns either P' or P, whichever set has higher hypervolume (where ties may be broken arbitrarily). This postprocessing makes much sense, as it prohibits decreasing the hypervolume of our population. In fact, *all of our results* for increasing or non-decreasing archiving algorithms also hold for such non-decreasing greedy archiving algorithms, except for the upper bounds in Theorems V.3, V.4 and V.8. Since we are more interested in *lower* bounds for (non-decreasing) greedy algorithms, we did not try to reprove these upper bounds.

III. Technical Basics

In this section we show basic properties of HYP and CON that will be used in later proofs. We start with some very basic facts.

Lemma III.1. *For any finite $P \subseteq P' \subset \mathbb{R}^d_+$ we have*

1) *(non-negativity)* $\mathrm{HYP}(P) \geqslant 0$,
2) *(monotonicity)* $\mathrm{HYP}(P) \leqslant \mathrm{HYP}(P')$,
3) *(empty set)* $\mathrm{HYP}(\emptyset) = 0$.

Proof. Recall that $\mathrm{HYP}(P) = \mathrm{VOL}(\bigcup_{p\in P} \mathrm{box}(p))$. Since volume in $\mathbb{R}^d$ is non-negative, so is HYP. Moreover, for $P \subseteq P'$ we have $\bigcup_{p\in P} \mathrm{box}(p) \subseteq \bigcup_{p\in P'} \mathrm{box}(p)$, so that $\mathrm{HYP}(P) \leqslant \mathrm{HYP}(P')$. Lastly, $\mathrm{HYP}(\emptyset) = \mathrm{VOL}(\emptyset) = 0$. ☐

From the above facts and the definition of CON we directly obtain similar facts about CON.

Lemma III.2. *For any finite $Q, P \subset \mathbb{R}^d_+$ we have*

1) $\mathrm{CON}_P(Q) \geqslant 0$,
2) $\mathrm{CON}_P(\emptyset) = 0$ *and* $\mathrm{CON}_\emptyset(Q) = \mathrm{HYP}(Q)$,
3) $\mathrm{CON}_P(Q) = \mathrm{CON}_{P\cup Q}(Q) = \mathrm{CON}_{P\setminus Q}(Q)$.

Proof. For $\mathrm{CON}_P(Q) = \mathrm{HYP}(P \cup Q) - \mathrm{HYP}(P \setminus Q)$ non-negativity follows from $P \setminus Q \subseteq P \cup Q$ and monotonicity of HYP.

Note that $\mathrm{CON}_P(\emptyset) = \mathrm{HYP}(P \cup \emptyset) - \mathrm{HYP}(P \setminus \emptyset) = 0$. Furthermore, $\mathrm{CON}_\emptyset(P) = \mathrm{HYP}(\emptyset \cup P) - \mathrm{HYP}(\emptyset \setminus P) = \mathrm{HYP}(P)$.

Lastly, since $P\cup Q = (P\cup Q)\cup Q = (P\setminus Q)\cup Q$ and $P\setminus Q = (P \cup Q) \setminus Q = (P \setminus Q) \setminus Q$ we have $\mathrm{CON}_P(Q) = \mathrm{HYP}(P \cup Q) - \mathrm{HYP}(P \setminus Q) = \mathrm{CON}_{P\cup Q}(Q) = \mathrm{CON}_{P\setminus Q}(Q)$. ☐

We will often make use of telescoping sums such as the following.

Lemma III.3. *Let $P = \{p_1, \ldots, p_\mu\} \subset \mathbb{R}^d_+$ and set $P_i := \{p_1, \ldots, p_i\}$. Then we have for any $0 \leqslant i \leqslant \mu$*

$$\mathrm{HYP}(P) = \mathrm{HYP}(P_i) + \sum_{j=i+1}^{\mu} \mathrm{CON}_{P_j}(p_j).$$

Proof. Follows from $\mathrm{CON}_{P_j}(p_j) = \mathrm{HYP}(P_j + p_j) - \mathrm{HYP}(P_j - p_j) = \mathrm{HYP}(P_j) - \mathrm{HYP}(P_{j-1})$. ☐

One of the most fundamental facts about HYP is that it is *submodular* [13], as has been observed by Ulrich and Thiele [15, Thm. 1]. The following Lemma III.4 presents a short proof of this property.

Lemma III.4. *(Submodularity) For any finite $P \subseteq P' \subset \mathbb{R}^d_+$ and $Q \subset \mathbb{R}^d_+$ we have*

$$\mathrm{CON}_P(Q) \geqslant \mathrm{CON}_{P'}(Q).$$

Proof. Recall that $\mathrm{HYP}(T) = \mathrm{VOL}(\bigcup_{p\in T} \mathrm{box}(p))$ for any finite $T \subset \mathbb{R}^d_+$. Let $B_T := \bigcup_{p\in T} \mathrm{box}(p)$ so that $\mathrm{HYP}(T) = \mathrm{VOL}(B_T)$. Using the definitions of CON and B_T and the facts $B_{P\setminus Q} \subseteq B_{P\cup Q}$ and $(P \cup Q) \setminus (P \setminus Q) = Q$ we obtain

$$\begin{aligned}\mathrm{CON}_P(Q) &= \mathrm{HYP}(P \cup Q) - \mathrm{HYP}(P \setminus Q)\\ &= \mathrm{VOL}(B_{P\cup Q}) - \mathrm{VOL}(B_{P\setminus Q})\\ &= \mathrm{VOL}(B_{P\cup Q} \setminus B_{P\setminus Q})\\ &= \mathrm{VOL}(B_Q \setminus B_{P\setminus Q}),\end{aligned}$$

and similarly for P'. Now, since $P \subseteq P'$ we have $B_{P\setminus Q} \subseteq B_{P'\setminus Q}$, so that $B_Q \setminus B_{P\setminus Q} \supseteq B_Q \setminus B_{P'\setminus Q}$. Hence, $\mathrm{CON}_P(Q) \geqslant \mathrm{CON}_{P'}(Q)$. ☐

We need a simple lower bound for HYP in terms of CON.

Lemma III.5. *For any finite $P \subset \mathbb{R}^d_+$ we have*

$$\mathrm{HYP}(P) \geqslant \sum_{p\in P} \mathrm{CON}_P(p).$$

Proof. Follows from Lemma III.6 below by setting $\lambda = 1$. ☐

More generally, we have the following lower bound.

Lemma III.6. *For any $P \subset \mathbb{R}^d_+$ of size μ and $\lambda \leqslant \mu$ we have*

$$\mathrm{HYP}(P) \geqslant \frac{1}{\binom{\mu-1}{\lambda-1}} \sum_{\substack{T\subseteq P\\ |T|=\lambda}} \mathrm{CON}_P(T).$$

Proof. For $\varepsilon > 0$ and $\bar{x} = (x_1, \ldots, x_d) \in \mathbb{N}^d_0$ let $A^\varepsilon_{\bar{x}} := [x_1\varepsilon, (x_1+1)\varepsilon] \times \ldots \times [x_d\varepsilon, (x_d+1)\varepsilon]$. We call $A^\varepsilon_{\bar{x}}$ (or $\bar{x}$) an *atom*. For $p \in P$ let $I^\varepsilon(p) := \{\bar{x} \in \mathbb{N}^d_0 \mid A^\varepsilon_{\bar{x}} \subseteq \mathrm{box}(p)\}$ and for $T \subseteq P$ let $I^\varepsilon(T) := \bigcup_{p\in T} I^\varepsilon(p)$. Since (the indicator function of) a box is Riemann-integrable and since $\mathrm{VOL}(A^\varepsilon_{\bar{x}}) = \varepsilon^d$ we have

$$\mathrm{VOL}(\mathrm{box}(p)) = \lim_{\varepsilon\to 0} \mathrm{VOL}\Bigg(\bigcup_{\bar{x}\in I^\varepsilon(p)} A^\varepsilon_{\bar{x}}\Bigg) = \lim_{\varepsilon\to 0} \varepsilon^d |I^\varepsilon(p)|.$$

For the same reasons, we have

$$\mathrm{HYP}(P) = \lim_{\varepsilon\to 0} \varepsilon^d |I^\varepsilon(P)|,$$

which implies for any $T \subseteq P$

$$\mathrm{CON}_P(T) = \lim_{\varepsilon\to 0} \varepsilon^d |I^\varepsilon(P) \setminus I^\varepsilon(P \setminus T)|.$$

Hence, it suffices to show

$$|I^\varepsilon(P)| \geqslant \frac{1}{\binom{\mu-1}{\lambda-1}} \sum_{\substack{T\subseteq P\\ |T|=\lambda}} |I^\varepsilon(P) \setminus I^\varepsilon(P \setminus T)|, \qquad (1)$$

then the claim follows by letting $\varepsilon \to 0$. To this end, consider any atom $\bar{x} \in I^\varepsilon(P)$. Note that all atoms that are counted in any term of inequality (1) are in $I^\varepsilon(P)$. Let $Z = \{p \in P \mid \bar{x} \in I^\varepsilon(p)\}$. Then $\bar{x} \in I^\varepsilon(P) \setminus I^\varepsilon(P \setminus T)$ if and only if $T \supseteq Z$ (since $\bar{x} \in I^\varepsilon(p) \subseteq I^\varepsilon(P \setminus T)$ for any $p \in Z \setminus T$). Hence, $A^\varepsilon_{\bar{x}}$ appears in $\binom{\mu-|Z|}{\lambda-|Z|} \leqslant \binom{\mu-1}{\lambda-1}$ summands on the right hand side of inequality (1), while it appears exactly once on the left hand side. This proves the claim.

We remark that instead of using *atoms* one could also phrase this proof using *volume elements* known from mathematical analysis. $\square$

Alternatively, we can bound HYP from above in terms of the hypervolumes of the single points.

Lemma III.7. *For any finite $P \subset \mathbb{R}^d_+$ we have*

$$\mathrm{HYP}(P) \leqslant \sum_{p \in P} \mathrm{HYP}(\{p\}).$$

Proof. Follows from Lemma III.8 below if we set $A = \emptyset$. $\square$

We can slightly generalize the above bound as follows.

Lemma III.8. *Let $A, B \subset \mathbb{R}^d_+$ be finite and set $P := A \cup B$. Then we have*

$$\mathrm{HYP}(P) \leqslant \mathrm{HYP}(A) + \sum_{b \in B} \mathrm{CON}_A(b).$$

Proof. Follows from Lemma III.9 below by setting $\lambda = 1$. $\square$

Even more general, we can prove an upper bound as follows.

Lemma III.9. *Let $A, B \subset \mathbb{R}^d_+$ be finite and set $P := A \cup B$. Let $|B| = \mu$ and $\lambda \leqslant \mu$. Then we have*

$$\mathrm{HYP}(P) \leqslant \mathrm{HYP}(A) + \frac{1}{\binom{\mu-1}{\lambda-1}} \sum_{\substack{T \subseteq B \\ |T| = \lambda}} \mathrm{CON}_A(T).$$

Proof. Since $\mathrm{HYP}(P) - \mathrm{HYP}(A) = \mathrm{HYP}(A \cup (B \setminus A)) - \mathrm{HYP}(A \setminus (B \setminus A)) = \mathrm{CON}_A(B \setminus A)$, we may prove instead the equivalent statement

$$\mathrm{CON}_A(B \setminus A) \leqslant \frac{1}{\binom{\mu-1}{\lambda-1}} \sum_{\substack{T \subseteq B \\ |T| = \lambda}} \mathrm{CON}_A(T). \tag{2}$$

As in the proof of Lemma III.6 we use *atoms* as follows. For $\varepsilon > 0$ and $\bar{x} = (x_1, \ldots, x_d) \in \mathbb{N}_0^d$ we let $A^\varepsilon_{\bar{x}} := [x_1\varepsilon, (x_1+1)\varepsilon] \times \ldots \times [x_d\varepsilon, (x_d+1)\varepsilon]$ be an atom. For $p \in P$ we let $I^\varepsilon(p) := \{\bar{x} \in \mathbb{N}_0^d \mid A^\varepsilon_{\bar{x}} \subseteq \mathrm{box}(p)\}$ and for $T \subseteq P$ we let $I^\varepsilon(T) := \bigcup_{p \in T} I^\varepsilon(p)$. As in the proof of Lemma III.6, we obtain for any $T \subseteq B$

$$\mathrm{CON}_A(T) = \lim_{\varepsilon \to 0} \varepsilon^d |I^\varepsilon(A \cup T) \setminus I^\varepsilon(A \setminus T)|.$$

Since $(A \cup T) \setminus (A \setminus T) = T$, this can be simplifed to

$$\mathrm{CON}_A(T) = \lim_{\varepsilon \to 0} \varepsilon^d |I^\varepsilon(T) \setminus I^\varepsilon(A \setminus T)|.$$

Hence, it suffices to show

$$|I^\varepsilon(B \setminus A) \setminus I^\varepsilon(A \setminus (B \setminus A))| =$$
$$|I^\varepsilon(B \setminus A) \setminus I^\varepsilon(A)| \leqslant \frac{1}{\binom{\mu-1}{\lambda-1}} \sum_{\substack{T \subseteq B \\ |T| = \lambda}} |I^\varepsilon(T) \setminus I^\varepsilon(A \setminus T)|, \tag{3}$$

then statement (2) follows by letting $\varepsilon \to 0$. Consider any atom $\bar{x} \in I^\varepsilon(B \setminus A) \setminus I^\varepsilon(A)$ (i.e., that appears on the left hand side of inequality (3)) and let $b \in B \setminus A$ with $\bar{x} \in I^\varepsilon(b)$. Then for any T with $b \in T \subseteq B$ and $|T| = \lambda$ we have $\bar{x} \in I^\varepsilon(T) \setminus I^\varepsilon(A) \subseteq I^\varepsilon(T) \setminus I^\varepsilon(A \setminus T)$. As there are $\binom{\mu-1}{\lambda-1}$ such sets T, the atom $A^\varepsilon_{\bar{x}}$ is counted at least $\binom{\mu-1}{\lambda-1}$ times on the right hand side of inequality (3). This proves inequality (3) and, thus, the claim. $\square$

IV. Computational Complexity

We first study the computational complexity of the large class of increasing archiving algorithms. This includes locally optimal and greedy archiving algorithms (cf. Observation II.8). We prove that all increasing archiving algorithms solve an NP-hard problem and are thus not computationally efficient unless P = NP. By reduction from the known hardness of computing a least contributor of a set of points, we show the following theorem.

Theorem IV.1. *All increasing archiving algorithms solve an NP-hard problem (if d is part of the input).*

Proof. We reduce from the problem of computing a least contributor of a set of points: Given $P \subseteq \mathbb{R}^d_+$ of size n, compute a point $p \in P$ with $\mathrm{CON}_P(p)$ minimal (see Section II-A for the definition). This problem is NP-hard according to [7].

Let P be an instance to the least contributor problem, and let $\mathcal{A}$ be an increasing archiving algorithm. We compute $A(p) := P \setminus \mathcal{A}(P - p, \{p\})$ for each $p \in P$. This is the point with which the archiving algorithm $\mathcal{A}$ exchanges p given population $P - p$ and offspring $\{p\}$.

Consider the graph with vertex set P and directed edges $(p, A(p))$ for each $p \in P$. This graph may have self-loops. It includes a directed cycle as a subgraph: Starting at any point and always following the unique out-edge we will at some point see an already visited point again; this means we traversed a cycle (after some initial path).

Let $(p_0, \ldots, p_{k-1})$ be such a cycle. It can have length $k = 1$, if the cycle is a self-loop. Since $A(p_i) = p_{i+1}$ (with indices modulo k) and the archiving algorithm is increasing — thus also non-decreasing — we have $\mathrm{HYP}(P - p_{i+1}) \geqslant \mathrm{HYP}(P - p_i)$ for all $i \in \{0, \ldots, k-1\}$. Hence, all $\mathrm{HYP}(P - p_i)$ are equal; in particular $\mathrm{HYP}(P - p_0) = \mathrm{HYP}(P - A(p_0))$. Since the archiving algorithm is increasing, this means that no increase was possible given population $P - p_0$ and offspring $\{p_0\}$, and, hence, that $\mathrm{HYP}(P - p_0) = \mathrm{HYP}(P) - \mathrm{CON}_P(p_0)$ is maximal among all $\mathrm{HYP}(P - p)$. In other words, $\mathrm{CON}_P(p_0)$ is minimal and p_0 is a least contributor. The same holds for all other points $p \in P$ that lie on a directed cycle in the constructed graph. Thus, we can compute a least contributor using any increasing archiving algorithm. This reduction proves that all increasing archiving algorithms solve an NP-hard problem. □

Theorem IV.1 above shows that only archiving algorithms which are not increasing in the meaning of Definition II.5 might be computationally efficient (unless P = NP). In Section VI-C we indeed present such a non-decreasing archiving algorithm, which is not increasing, but has a polynomial runtime.

V. 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. In this section we consider the best-case choice of the offspring and analyze which archiving algorithms are *effective*, that is, are able to reach the optimum. This is complemented by a worst-case perspective on the choice of the offspring in Section VI.

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 offspring such that the archiving algorithm runs on P^0 and the sequence of offspring 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 definition.

Definition V.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 result of Zitzler et al. [18] remains valid.

Theorem V.2. *There is no effective non-decreasing $(\mu+1)$-archiving algorithm (for $\mu > 1$).*

Note that we have reformulated the statement of [18, Cor. 4.6] in our notation defined above. We do not give a separate proof for Theorem V.2 as it directly follows from Theorem V.5 below. Theorem V.2 assumes $\lambda = 1$. The corresponding result for $\lambda = \mu$ follows from [18, Thm. 4.4]:

Theorem V.3. *There is an effective non-decreasing $(\mu+\mu)$-archiving algorithm.*

We do not give a direct proof for Theorem V.3 as it follows from Theorem V.4 below. In order to show Theorem V.3, observe that there is an increasing $(\mu+\mu)$-archiving algorithm. Theorem V.4 below shows that this increasing $(\mu+\mu)$-archiving algorithm is also effective. As every increasing archiving algorithm is also non-decreasing, this proves Theorem V.3.

Since Theorem V.3 is only an existential statement, it is natural to ask what effective non-decreasing $(\mu+\mu)$-archiving algorithms look like. The authors showed in [8, Thm. 3.4] that all $(\mu+\mu)$-archiving algorithms $\mathcal{A}$ with $\mathrm{HYP}(\mathcal{A}(P,Q)) \geqslant \mathrm{HYP}(Q)$ for all P, Q are effective. This implies that all locally optimal $(\mu+\mu)$-archiving algorithms are effective. The following Theorem V.4 shows another generalization of Theorem V.3, which also implies that all locally optimal $(\mu+\mu)$-archiving algorithms are effective.

Theorem V.4. *All increasing $(\mu+\mu)$-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., $\mathrm{HYP}(P^*) = \mathrm{maxHYP}_\mu(\mathcal{Y})$. We set $Q^i := P^*$ for $i = 1, \dots, N$ and N sufficiently large. Then all populations satisfy $P^i \subseteq P^0 \cup P^*$ for $i = 0, \dots, N$. Note that as long as $\mathrm{HYP}(P^i) < \mathrm{HYP}(P^*)$ we have $\mathrm{maxHYP}_\mu(P^i \cup Q^{i+1}) = \mathrm{HYP}(Q^{i+1}) = \mathrm{HYP}(P^*) > \mathrm{HYP}(P^i)$, so an improvement is possible. Hence, any increasing archiving algorithm will choose a subset $P^{i+1} \subseteq P^i \cup Q^{i+1}$ with $\mathrm{HYP}(P^{i+1}) > \mathrm{HYP}(P^i)$. Since there are at most $\binom{2\mu}{\mu}$ different subsets of $P^0 \cup P^*$, at the latest after $N = \binom{2\mu}{\mu}$ iterations we reach $P^N = P^*$. □

Note that Theorem V.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. Hence Theorem V.4 does not hold for all infinite objective spaces.

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 V.5. *There is no effective non-decreasing $(\mu+\lambda)$-archiving algorithm for $\lambda < \mu$.*

Again, we do not give a separate proof for Theorem V.5 as it follows from its stronger counterpart Theorem V.7 below. In order to prove Theorem V.5 directly, one would construct 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 V-A.

A. Approximate Effectiveness

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

Definition V.6. *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* 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* $Q^1, \ldots, Q^N \subseteq \mathcal{Y}$ *such that*

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

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. Note that this also implies that there is no effective non-decreasing $(\mu+\lambda)$-archiving algorithm for $\lambda < \mu$ as stated in Theorem V.5. In the proof we explicitly construct an objective space with two unconnected local maxima that have sufficiently different hypervolume.

Theorem V.7. *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* $\varepsilon > 0$.

Proof. 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 HYP(P^*) is sufficiently larger than the hypervolume of P^0.

For showing this, define $A := \mathrm{CON}_{\mathcal{Y}}(p_{2i})$ and $B := \mathrm{CON}_{\mathcal{Y}}(p_{2i+1})$. Observe 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 $\Delta\mathrm{HYP}_1 := \mathrm{HYP}(P^1) - \mathrm{HYP}(P^0) < 0$, 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 $\Delta\mathrm{HYP}_1 \leqslant k(B - A) - C$. As the offspring size $|Q^1| \leqslant \lambda$ we have $k \leqslant \lambda$ and thus

$$\Delta\mathrm{HYP}_1 \leqslant \lambda(B - A) - C.$$

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

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

Simple calculations show that this bound is always greater than 1 and less than or 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 (4) no $(\mu+\lambda)$-archiving algorithm can escape from P^0.

All that is left to show is that HYP(P^*) is sufficiently greater than HYP(P^0). Above we saw that $\mathrm{HYP}(P^0) = H - (\mu+1)B$, where $H = \mathrm{HYP}(\mathcal{Y})$. Now, observe that $\mathrm{HYP}(P^*) = H - \mu A - B - C$, 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

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

Let $\varepsilon > 0$. By choosing γ (dependent on α) sufficiently near to the right hand side of equation (4) we have $0 > \lambda(B - A) - C \geqslant -\varepsilon$ and, hence,

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

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

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

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,

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

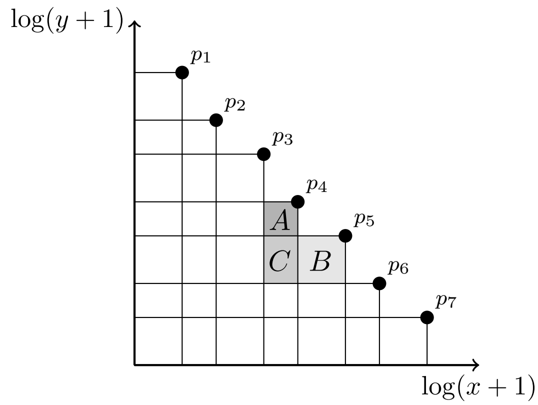


Figure 1: A schematic log-log plot of the example used in the proof of Theorem V.7. The considered areas A, B, C are indicated.

for $\alpha \geqslant \sqrt{2}$ and, thus, $\mathrm{HYP}(P^0) \geqslant 1$, so that we can bound $\varepsilon/\mathrm{HYP}(P^0) \leqslant \varepsilon$. 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 (4), we get

$$\frac{\mathrm{maxHYP}_\mu(\mathcal{Y})}{\mathrm{HYP}(P^0)} \geqslant 1 - 2\varepsilon + \left(1 - \frac{\lambda}{\mu}\right)\left(\left(\frac{\alpha^2 + 1 + (\alpha^2 - 1)\sqrt{4\alpha^2\lambda^2 + 1}}{2\alpha(\lambda(\alpha^2 - 1) + 1)}\right)^2 - 1\right).$$

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

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

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

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

This finishes the proof. □

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 V.4. The proven constant, however, may be far from being tight. Maybe surprisingly, Theorem V.7 indeed does not hold without the restriction to non-decreasing archiving algorithms. Theorem V.9 in Section V-B shows that for all μ and λ there are archiving algorithms which are not non-decreasing, but effective.

Complementing the lower bound of Theorem V.7, the authors [8, Thm. 4.3] proved for all $\varepsilon > 0$ an upper bound of $2+\varepsilon$ on the approximation ratio achieved by all increasing $(\mu+\lambda)$-archiving algorithms. This was improved by Ulrich and Thiele [15], who showed that every increasing $(\mu + 1)$-archiving algorithm reaches a $\left(2 - \frac{1}{\mu}\right)$-approximation. More precisely, they showed that every increasing $(\mu+\lambda)$-archiving algorithm reaches a $\left(2 - \frac{\lambda - p}{\mu}\right)$-approximation, where $\mu = q\,\lambda - p$ with $p < \lambda$ and $p, q \in \mathbb{N}_{\geqslant 0}$. We can now improve both results and prove the following theorem.

Theorem V.8. *Let $\mathcal{A}$ be an increasing $(\mu + \lambda)$-archiving algorithm, $\lambda \leqslant \mu$. Then $\mathcal{A}$ is $\left(2 - \frac{\lambda}{\mu} + \varepsilon\right)$-approximate for any $\varepsilon > 0$. If the objective space is finite we can even set $\varepsilon = 0$.*

This is better than the bound of [15] if λ does not divide μ. For $\lambda \to \mu$ it approaches 1 and for $\lambda = \mu$ it attains 1.

Proof of Theorem V.8. 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 as a supremum, there exists a μ-population $P^* \subseteq \mathcal{Y}$ with $\mathrm{HYP}(P^*) \geqslant (1 + \varepsilon/2)^{-1}\,\mathrm{maxHYP}_\mu(\mathcal{Y})$. If $\mathcal{Y}$ is finite we even have $\varepsilon = 0$.

As (best-case) offspring we choose the size-λ subsets of P^*. This we do until the current population $P = P^N$ is stable under insertions of such sets, i.e., until

$$\mathrm{maxHYP}_\mu(P \cup Q) = \mathrm{HYP}(P) \text{ for all } Q \subseteq P^*,\ |Q| = \lambda.$$

We show that for any such stable solution P we have $(2 - \frac{\lambda}{\mu})\mathrm{HYP}(P) \geqslant \mathrm{HYP}(P^*)$, so that $(2 - \frac{\lambda}{\mu} + \varepsilon)\mathrm{HYP}(P) \geqslant \mathrm{maxHYP}_\mu(\mathcal{Y})$, proving the result. To this end, let $S \subset P$, $|S| = \lambda$ with $\mathrm{CON}_P(S)$ minimal among all such sets, and set $P' := P \setminus S$. By monotonicity of HYP and Lemma III.9 we have

$$\begin{aligned}
\mathrm{HYP}(P^*) &\leqslant \mathrm{HYP}(P^* \cup P') \\
&\leqslant \mathrm{HYP}(P') + \frac{1}{\binom{\mu-1}{\lambda-1}} \sum_{\substack{T \subseteq P^* \\ |T| = \lambda}} \mathrm{CON}_{P'}(T). \qquad (5)
\end{aligned}$$

Observe that $P' \cup T$ is reachable from P by exchanging S with T, where $T \subseteq P^*$, $|T| = \lambda$. Thus, by stability of P

$$\mathrm{HYP}(P' \cup T) \leqslant \mathrm{HYP}(P) = \mathrm{HYP}(P' \cup S),$$

implying

$$\mathrm{CON}_{P'}(T) \leqslant \mathrm{CON}_{P'}(S).$$

Together with inequality (5) we get

$$\begin{aligned}\mathrm{HYP}(P^*) &\leqslant \mathrm{HYP}(P') + \frac{\binom{\mu}{\lambda}}{\binom{\mu-1}{\lambda-1}}\,\mathrm{CON}_{P'}(S)\\ &= \mathrm{HYP}(P') + \tfrac{\mu}{\lambda}\,\mathrm{CON}_{P'}(S)\\ &= \mathrm{HYP}(P) + \left(\tfrac{\mu}{\lambda}-1\right)\,\mathrm{CON}_{P'}(S). \qquad (6)\end{aligned}$$

Note that $\mathrm{CON}_{P'}(S) = \mathrm{CON}_P(S)$ (by Lemma III.2.(3) and $P' = P \setminus S$). As S minimizes the contribution, we have

$$\mathrm{CON}_P(S) \leqslant \frac{1}{\binom{\mu}{\lambda}} \sum_{\substack{T\subseteq P\\ |T|=\lambda}} \mathrm{CON}_P(T).$$

Moreover, by Lemma III.6 we have

$$\sum_{\substack{T\subseteq P\\ |T|=\lambda}} \mathrm{CON}_P(T) \leqslant \binom{\mu-1}{\lambda-1}\,\mathrm{HYP}(P).$$

Together, we obtain

$$\mathrm{CON}_{P'}(S) \leqslant \frac{\binom{\mu-1}{\lambda-1}}{\binom{\mu}{\lambda}}\mathrm{HYP}(P) = \tfrac{\lambda}{\mu}\mathrm{HYP}(P).$$

Plugging this into inequality (6) yields the desired bound. □

B. Effectiveness and Decreasing Algorithms

We now briefly discuss why the previous results required 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 V.9. *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

$P' := (P \setminus Q) \cup (Q \setminus P)$
if $|P'| \leqslant \mu$ **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 $P^0 \setminus P^* = \{p^0_1, \ldots, p^0_\mu\}$ (with possibly some of the p^0_i being equal) and $P^* \setminus P^0 = \{p^*_1, \ldots, p^*_\mu\}$. Let $Q^{2i-1} = \{p^0_i\}$ and $Q^{2i} = \{p^*_i\}$ for $i = 1, \ldots, \mu$. Algorithm 3 works as desired on this offspring: 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 justifies why we assumed the archiving algorithms to be non-decreasing in this Section V. Theorem V.9 shows that Theorems V.2, V.5 and V.7 do not hold for general archiving algorithms, which are not required to be non-decreasing.

VI. Competitiveness

In contrast to the last section, we now take a worst-case view on offspring generation (as well as the initial population), so we want to bound $\mathcal{A}(P^0, Q^1, \ldots, Q^N)$ for any initial population P^0 and any offspring generations $Q^1, \ldots, Q^N$. Observe that in this setting all results have to be independent of the specific objective space $\mathcal{Y}$ and we cannot hope to reach $\mathrm{maxHYP}_\mu(\mathcal{Y})$ in general. The only aim can be achieving a hypervolume as good as the maximum hypervolume among all μ-populations which are subsets of the points we have seen so far, that is,

$$\mathrm{maxHYP}_\mu\left(P^0 \cup \bigcup_{i=1}^{N} Q^i\right),$$

which can be arbitrarily smaller than $\mathrm{maxHYP}_\mu(\mathcal{Y})$.

This allows us to view archiving algorithms as an online problem where the algorithm is fed with new offspring in a serial fashion and has to decide which individual it should keep in the population without knowing the entire input. To measure the ‘regret’ of an archiving algorithm we define its competitive ratio α as follows.

Definition VI.1. *Let P^0 be a μ-population and Q^i be λ-populations for $1 \leqslant i \leqslant N$. Then $I := (P^0, Q^1, \ldots, Q^N)$ is an* instance. *We also set*

$$\mathrm{Obs}(I) := P^0 \cup \bigcup_{i=1}^{N} Q^i.$$

An archiving algorithm $\mathcal{A}$ is α-competitive (for some $\alpha \geqslant 1$) if for all instances $I = (P^0, Q^1, \ldots, Q^N)$ we have

$$\mathcal{A}(P^0, Q^1, \ldots, Q^N) \geqslant \frac{1}{\alpha}\,\mathrm{maxHYP}_\mu(\mathrm{Obs}(I)).$$

A. Increasing algorithms are at best μ-competitive

It is easy to show an upper bound on the competitive ratio of μ for a very large class of archiving algorithms. It applies to all non-decreasing archiving algorithms with the following property: If a single offspring point $q \in Q$ alone dominates a larger hypervolume than all points in the current population together, then the algorithm should take this point q (or do something even better). Note that all locally optimal archiving algorithms satisfy this condition.

Theorem VI.2. *Let $\mathcal{A}$ be a non-decreasing archiving algorithm such that for all inputs P and Q and points $q \in Q$ we have*

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

Then $\mathcal{A}$ is μ-competitive.

In particular: All locally optimal $(\mu+\lambda)$-archiving algorithms and all increasing $(\mu+1)$-archiving algorithms are μ-competitive.

Proof. Let $I := (P^0, Q^1, \ldots, Q^N)$ be an instance and $P^* \subseteq \mathrm{Obs}(I)$ be a μ-population with $\mathrm{HYP}(P^*) = \max\mathrm{HYP}_\mu(\mathrm{Obs}(I))$. By Lemma III.7 we have

$$\mathrm{HYP}(P^*) \leqslant \sum_{p\in P^*} \mathrm{HYP}(\{p\}),$$

so there is a point $p^* \in P^*$ with

$$\mathrm{HYP}(\{p^*\}) \geqslant \frac{1}{|P^*|}\mathrm{HYP}(P^*) \geqslant \frac{1}{\mu}\mathrm{HYP}(P^*).$$

Either $p^* \in P^0$ (in which case we set $r := 0$) or $p^* \in Q^r$ for some r. Consider $P^i = \mathcal{A}(P^0, Q^1, \ldots, Q^i)$. We have $\mathrm{HYP}(P^r) \geqslant \mathrm{HYP}(\{p^*\})$ by the extra assumption and $\mathrm{HYP}(P^r) \leqslant \mathrm{HYP}(P^{r+1}) \leqslant \ldots \leqslant \mathrm{HYP}(P^N)$ by $\mathcal{A}$ being non-decreasing. Taken together these prove the claim. □

Observe that even a very simple algorithm fulfills the premises of the above theorem: It considers the offspring one-by-one and replaces its current population P^i with $\{q\}$, if the offspring point q has greater hypervolume than P^i. This requires only one costly computation of $\mathrm{HYP}(P^0)$; all other populations consist of only a single point in the objective space.

Perhaps surprisingly, no increasing archiving algorithm is better than this simple algorithm in the worst case: For them, the bound of Theorem VI.2 is tight.

Theorem VI.3. *No increasing $(\mu+\lambda)$-archiving algorithm is $(\mu-\varepsilon)$-competitive for any $\varepsilon > 0$.*

In the proof of this theorem, we explicitly construct a bad 2-dimensional instance; see Figure 2 for an example with $\mu = 5$. The initial population consists of the points $p_1, \ldots, p_4$ and the rightmost point of the q_i's. Then every offspring consists of (λ copies of) a q_i slightly to the left and above the old one, so that any increasing algorithm has to exchange the two points. This way, the population will always consist of the points $p_1, \ldots, p_4$ and one of the q_i's, with the latter point being dragged to the left. The optimal population, however, consists of five nicely spaced q_i's, which has (by choosing the free parameters correctly) a hypervolume that is nearly a factor μ larger than the hypervolume of the population of the increasing algorithm.

Proof. We construct an instance $I = (P^0, Q^1, \ldots, Q^N)$ as follows. For reals $a, A > 0$ to be chosen later and $j \in \{1, \ldots, \mu-1\}$, we set

$$p_j = (A + j\,a, (\mu-j)a),$$

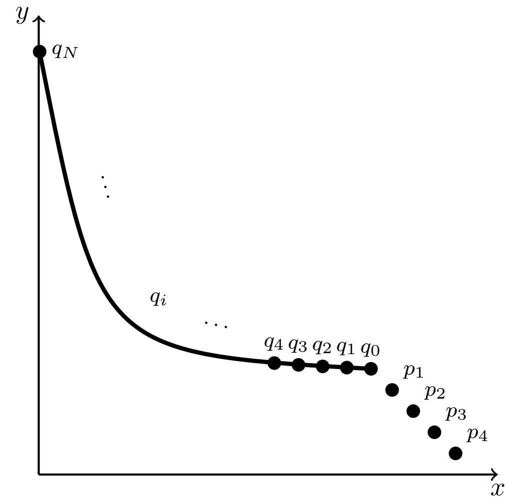


Figure 2: Illustration of the example used in the proof of Theorem VI.3.

and $B := (\mu-1)a$. Moreover, for $\delta > 0$, $0 < \rho < 1$ to be chosen later and $i \in \{0, \ldots, N\}$, we set $q_i = (x_i, y_i)$ with

$$x_i := A\rho^i,$$
$$y_i := B + \frac{1+\delta\, i}{x_i}.$$

These points are depicted in Figure 2. Setting $P^0 := \{q_0, p_1, \ldots, p_{\mu-1}\}$ and $Q^i := \{q_i\}$ (or λ copies of q_i), we get an instance I.

We show that $\mathcal{A}(P^{i-1}, Q^i) = P^i$ with $P^i = \{q_i, p_1, \ldots, p_{\mu-1}\}$ for $\mathcal{A}$ being an increasing archiving algorithm. To do this, we have to show that the exchange of q_{i-1} with q_i increases the hypervolume and is the only increasing exchange. This means we have to show $\mathrm{HYP}(P^i) > \mathrm{HYP}(P^{i-1})$ and $\mathrm{HYP}(P^{i-1}) \geqslant \mathrm{HYP}(P^{i-1} + q_i - p_j)$ for any $1 \leqslant j \leqslant \mu-1$, as those are the only possible exchanges. We have

$$\mathrm{HYP}(P^i) = \mathrm{HYP}(\{p_1, \ldots, p_{\mu-1}\}) + \mathrm{CON}_{P^i}(q_i),$$

where $\mathrm{CON}_{P^i}(q_i) = x_i(y_i - B) = 1 + \delta i$, and $p_1, \ldots, p_{\mu-1}$ are collinear points, so that their hypervolume can easily be calculated, yielding

$$\mathrm{HYP}(P^i) = A\,B + \tbinom{\mu}{2}a^2 + 1 + \delta\, i. \tag{7}$$

This gives $\mathrm{HYP}(P^i) > \mathrm{HYP}(P^{i-1})$ right away. Moreover, by inspection of the constructed instance we have

$$\begin{aligned}&\mathrm{HYP}(P^{i-1} + q_i - p_j)\\ &= \mathrm{HYP}(P^{i-1}) - \mathrm{CON}_{P^{i-1}}(p_j) + \mathrm{CON}_{P^{i-1}}(q_i),\end{aligned}$$

with $\mathrm{CON}_{P^{i-1}}(p_j) \geqslant a^2$ and

$$\begin{aligned}\mathrm{CON}_{P^{i-1}}(q_i) &= x_i(y_i - y_{i-1})\\ &= 1 + \delta\, i - \rho(1 + \delta(i-1))\\ &= (1-\rho)(1 + \delta(i-1)) + \delta.\end{aligned}$$

Hence, for

$$a^2 \geqslant (1-\rho)(1 + \delta\, N) + \delta, \tag{8}$$

we have $\mathrm{HYP}(P^{i-1}) \geqslant \mathrm{HYP}(P^{i-1} + q_i - p_j)$ for any $1 \leqslant j \leqslant \mu - 1$, and $P^i = \mathcal{A}(P^{i-1}, Q^i)$ indeed holds.

Lastly, we need a lower bound on $\mathrm{maxHYP}_\mu(\mathrm{Obs}(I))$. For this we require that μ divides N and consider $P = \{q_{iN/\mu} \mid 0 \leqslant i \leqslant \mu - 1\}$. If we consider instead $P' = \{q'_{iN/\mu} \mid 0 \leqslant i \leqslant \mu - 1\}$, with $q'_i = (x_i, y'_i)$, $y'_i = B + 1/x_i$, the hypervolume decreases only, as we decrease the y-coordinates. Hence, we have $\mathrm{maxHYP}_\mu(\mathrm{Obs}(I)) \geqslant \mathrm{HYP}(P) \geqslant \mathrm{HYP}(P')$. All that is left to show is

$$\mathrm{HYP}(P') > (\mu - \varepsilon)\mathrm{HYP}(P^N). \tag{9}$$

We compute $\mathrm{HYP}(P')$ to be

$$\begin{aligned}\mathrm{HYP}(P') &= \mathrm{HYP}(\{q'_0\}) + \sum_{i=1}^{\mu-1} x_{iN/\mu}(y'_{iN/\mu} - y'_{(i-1)N/\mu}) \\ &= A(B + 1/A) + \sum_{i=1}^{\mu-1}(1 - \rho^{N/\mu}) \\ &= A\,B + \mu - (\mu - 1)\rho^{N/\mu}. \end{aligned} \tag{10}$$

Then equation (9) is fulfilled (using equations (7) and (10)) if

$$\begin{aligned}&A\,B + \mu - (\mu - 1)\rho^{N/\mu} > \\ &(\mu - \varepsilon)\left(A\,B + \tbinom{\mu}{2}a^2 + 1 + \delta\,N\right).\end{aligned}$$

Rearranging this, we get

$$\begin{aligned}\varepsilon >&(\mu - 1 - \varepsilon)A\,B + (\mu - \varepsilon)\tbinom{\mu}{2}a^2 + \\ &(\mu - 1)\rho^{N/\mu} + (\mu - \varepsilon)\delta\,N.\end{aligned}$$

This inequality is fulfilled by setting

$$\begin{aligned}A &:= \varepsilon/(4\mu B), \\ a^2 &:= \varepsilon/(4\mu^3), \\ \delta &:= \min\{\varepsilon/(4\mu N), a^2/2\}, \text{ and} \\ N &:= \mu\left\lceil \log_\rho(\varepsilon/(4\mu))\right\rceil.\end{aligned}$$

As we can assume $\varepsilon \leqslant 1$, we have $\delta \leqslant \min\{1/N, a^2/2\}$, so that requirement (8) can be simplified to $a^2 \geqslant 2(1-\rho) + a^2/2$. We set $\rho := 1 - a^2/4$ to satisfy it. Noting that there is no cyclic dependence in these definitions, we conclude the proof. □

B. A competitive non-decreasing algorithm

We showed in the preceeding Section VI-A that increasing archiving algorithms only achieve an unbounded competitive ratio of $\geqslant \mu$. This result implies that the notion of competitiveness is not suited for comparing different increasing archiving algorithms: All of them have an unbounded competitiveness of at least μ and this is tight even for very simple algorithms. This leaves open whether there are non-decreasing, but not increasing archiving algorithms with better, e.g., constant competitive ratio. This would imply that some archiving algorithm that is not locally optimal achieves a better competitive ratio than all locally optimal archiving algorithms.

In this section (Section VI-B) we present a non-decreasing (and not increasing) archiving algorithm which indeed achieves constant competitiveness. In the following section (Section VI-C) we will present a randomized variant of the algorithm which is also computationally efficient.

Theorem VI.4. *There is a $(4 + 2/\mu)$-competitive non-decreasing $(\mu + 1)$-archiving algorithm.*

Note that we can easily build a $(\mu + \lambda)$-archiving algorithm with the same competitiveness from the $(\mu + 1)$-archiving algorithm guaranteed by Theorem VI.4 by feeding the λ offspring one by one to the $(\mu + 1)$-archiving algorithm.

We do not prove Theorem VI.4 directly, as it follows from the proof of Theorem VI.6 below. However, one such archiving algorithm $\mathcal{A}_{\mathrm{comp}}$ is given in Algorithm 4. This non-locally-optimal algorithm improves on the locally optimal algorithms with respect to the competitive ratio and hence is better suited for hypervolume-based selection in the worst case. Note that this does not imply that this algorithm should be used in practice, as worst-case optimality is usually not needed.

Algorithm 4: Competitive $(\mu + 1)$-archiving algorithm $\mathcal{A}_{\mathrm{comp}}$

input : μ-population P, offspring $\{q\}$
output: μ-population P'

```
1 foreach p ∈ P + q do
2     H_p ← HYP(P + q − p)
3 p' ← argmax{H_p | p ∈ P}
4 if H_p' > (1 + 1/μ) H_q then
5     return P + q − p'
6 else
7     return P
```

Unfortunately, the runtime of $\mathcal{A}_{\mathrm{comp}}$ cannot be polynomial in $\mu + d$ (unless P = NP) as the exact hypervolume calculation in line 2 of Algorithm 4 is #P-hard [5]. However, this also holds for all increasing archiving algorithms as shown in Theorem IV.1.

C. A computationally efficient randomized competitive non-decreasing archiving algorithm

We now propose a randomized variant of $\mathcal{A}_{\mathrm{comp}}$ which improves on all increasing algorithms not only with respect to the competitive ratio, but also the runtime. It is a randomized algorithm which meets the competitive ratio bound only with a certain high probability. Hence, we need to redefine competitiveness (and non-decreasing) to include randomized algorithms.

Definition VI.5. *Let $\alpha \geqslant 1$ and $p\colon \mathbb{N} \to [0, 1]$. A randomized archiving algorithm $\mathcal{A}$ is α-*competitive with probability *p if for all instances $I = (P^0, Q^1, \ldots, Q^N)$ we have*

$$\mathcal{A}(P^0, Q^1, \ldots, Q^N) \geqslant \frac{1}{\alpha}\,\mathrm{maxHYP}_\mu(\mathrm{Obs}(I))$$

with probability at least $p(N)$.

We call a randomized archiving algorithm $\mathcal{A}$ non-decreasing with probability *p if for all instances $I = (P^0, Q^1, \ldots, Q^N)$ we have $\mathrm{HYP}(\mathcal{A}(P^0, Q^1, \ldots, Q^i)) \geqslant \mathrm{HYP}(\mathcal{A}(P^0, Q^1, \ldots, Q^{i-1}))$ for all $i \in \{1, \ldots, N\}$ with probability at least $p(N)$.*

Our proposed algorithm $\mathcal{A}_{\text{eff}}$ is given in Algorithm 5. It takes additional parameters $\varepsilon, \delta > 0$ and is $(4 + 2/\mu + \varepsilon)$-competitive with probability $p(N) = 1 - N\delta$.

Algorithm 5: Randomized Efficient Competitive $(\mu+1)$-archiving algorithm $\mathcal{A}_{\text{eff}}$

input : μ-population P, offspring $\{q\}$, error bound ε, error probability δ
output: μ-population P'

```
1 ε' ← ε/104
2 c ← 1 + 2ε'
3 foreach p ∈ P + q do
4     H_p ← compute (1 + ε'/μ)-approximation of
      HYP(P + q − p) with error probability δ/(μ + 1)
5 p' ← argmax{H_p | p ∈ P}
6 if H_{p'} > (1 + c/μ) H_q then
7     return P + q − p'
8 else
9     return P
```

The critical feature of $\mathcal{A}_{\text{eff}}$ is line 4. It makes use of the hypervolume approximation scheme of Bringmann and Friedrich [5] which computes with probability at least $1 - \delta$ a $(1 + \varepsilon)$-approximation of the hypervolume of a given set of μ points in $\mathbb{R}^d_+$ in time $\mathcal{O}(\log(1/\delta)\mu d/\varepsilon^2)$. It is clear from the hypervolume approximation algorithm used here that $\mathcal{A}_{\text{eff}}$ is efficiently computable, namely with a run-time of at most $\mathcal{O}(\log(\mu/\delta)\mu^4 d/\varepsilon^2)$. Note that we aimed only for a polynomial runtime and did not try to optimize the algorithm for a better runtime bound.

We can prove that $\mathcal{A}_{\text{eff}}$ is competitive. The following theorem states our result.

Theorem VI.6. *Let* $0 < \varepsilon \leqslant 1$. *Algorithm* $\mathcal{A}_{\text{eff}}$ *is a randomized* $(\mu + 1)$*-archiving algorithm which is non-decreasing and* $(4 + 2/\mu + \varepsilon)$*-competitive with probability* $p(N) = 1 - N\delta$. *Moreover, it has a deterministic runtime polynomial in* μ, λ, d, $\log(1/\delta)$, *and* $1/\varepsilon$.

Observe that by setting $\varepsilon = 0$, so that all hypervolume approximations are in fact exact computations, Algorithm 5 becomes the same as Algorithm 4. This implies that the proof of Theorem VI.6 is a proof of Theorem VI.4, too.

The perhaps surprising probability bound is caused by the (necessary) assumption that every call to the hypervolume approximation algorithm indeed returns a $(1 + \varepsilon'/\mu)$-approximation. The factor N in the probability can easily be cancelled out by a sufficiently small δ, as the runtime depends only logarithmically on $1/\delta$.

Proof. Let $I = (P^0, Q^1, \ldots, Q^N)$ be an instance. Consider the probability that every hypervolume approximation of $\mathcal{A}_{\text{eff}}$ on I indeed lies in the specified bounds, i.e., we have

$$(1-\varepsilon'/\mu)\text{HYP}(P+q-p) \leqslant H_p \leqslant (1+\varepsilon'/\mu)\text{HYP}(P+q-p)$$

for every computation of H_p in $\mathcal{A}_{\text{eff}}$. For a single call, this happens with probability at least $1-\delta/(\mu+1)$. Furthermore, we have at most $N(\mu + 1)$ hypervolume approximations running $\mathcal{A}_{\text{eff}}$ on I, as in every invocation of $\mathcal{A}_{\text{eff}}$, at most $\mu + 1$ hypervolume approximations are computed. With the union bound, we arrive at a probability of at least $p(N) = 1-N\delta$ that all the hypervolume approximations lie within the specified bounds. Thus, in the remainder we can assume that all hypervolume approximations of $\mathcal{A}_{\text{eff}}$ indeed lie in the specified bounds.

To show that $\mathcal{A}_{\text{eff}}$ is non-decreasing, note that in every iteration either we stay with the current population or the hypervolume increases by a constant factor, in which case

$$\begin{aligned}\Big(1 + \frac{\varepsilon'}{\mu}\Big)\,\text{HYP}(P^i) &\geqslant H_{p'} \\ &\geqslant \Big(1 + \frac{c}{\mu}\Big)\,H_q \geqslant \Big(1 - \frac{\varepsilon'}{\mu}\Big)\Big(1 + \frac{c}{\mu}\Big)\,\text{HYP}(P^{i-1}).\end{aligned}$$

For $c = 1 + 2\varepsilon' \geqslant 1$, $\mu \geqslant 1$ and $\varepsilon' \leqslant 1/3$, which is true by assumption, we have

$$\Big(1 - \frac{\varepsilon'}{\mu}\Big)\Big(1 + \frac{c}{\mu}\Big) \geqslant \Big(1 + \frac{\varepsilon'}{\mu}\Big),$$

implying that the algorithm is non-decreasing.

It remains to prove that $\mathcal{A}_{\text{eff}}$ is $(4+2/\mu+\varepsilon)$-competitive. To this end, let $P^i = \mathcal{A}_{\text{eff}}(P^{i-1}, Q^i)$ for $i = 1, \ldots, N$. Consider $\hat{P} := \bigcup_{i=0}^N P^i$, the set of all points in $\text{Obs}(I)$ that were taken by $\mathcal{A}_{\text{eff}}$ at some point. Let $P^* \subseteq \text{Obs}(I)$ be a μ-population with $\text{HYP}(P^*) = \text{maxHYP}_\mu(\text{Obs}(I))$. By monotonicity of HYP and Lemma III.8 we have

$$\begin{aligned}\text{maxHYP}_\mu(\text{Obs}(I)) = \text{HYP}(P^*) &\leqslant \text{HYP}(\hat{P} \cup P^*) \\ &\leqslant \text{HYP}(\hat{P}) + \sum_{q \in P^* \setminus \hat{P}} \text{CON}_{\hat{P}}(q). \quad (11)\end{aligned}$$

To continue, we bound $\text{HYP}(\hat{P})$ as well as the contribution of any point $q \in P^* \setminus \hat{P}$ to $\hat{P}$ in terms of $\text{HYP}(P^N)$. This will yield the desired bound for $\text{HYP}(P^N)$. We start with the latter.

Consider a point not chosen by the algorithm, $q \in \text{Obs}(I) \setminus \hat{P}$. We have $Q^i = \{q\}$ for some $1 \leqslant i \leqslant N$. Let $\tilde{p} \in P^{i-1} + q$ with $\text{CON}_{P^{i-1}+q}(\tilde{p})$ minimal among all $p \in P^{i-1} + q$. Note that we can have $\tilde{p} = q$, if q has the least contribution, but choosing q does not decrease the hypervolume enough, i.e., by a factor $1 + \frac{c}{\mu}$. Using Lemma III.5 we have

$$\begin{aligned}\text{CON}_{P^{i-1}+q}(\tilde{p}) &\leqslant \frac{1}{\mu+1} \sum_{p \in P^{i-1}+q} \text{CON}_{P^{i-1}+q}(p) \\ &\leqslant \frac{1}{\mu+1}\,\text{HYP}(P^{i-1} + q) \\ &= \frac{1}{\mu+1}\big(\text{HYP}(P^{i-1}) + \text{CON}_{P^{i-1}}(q)\big)\,. \quad (12)\end{aligned}$$

Let $p' = \text{argmax}\{H_p \mid p \in P^{i-1}\}$, see Algorithm 5. The point q was not taken by the algorithm, so we have

$$\begin{aligned}&\Big(1 - \frac{\varepsilon'}{\mu}\Big)\,\text{HYP}(P^{i-1} + q - \tilde{p}) \\ &\leqslant H_{\tilde{p}} \leqslant H_{p'} \leqslant \Big(1 + \frac{c}{\mu}\Big)\,H_q \\ &\leqslant \Big(1 + \frac{c}{\mu}\Big)\Big(1 + \frac{\varepsilon'}{\mu}\Big)\,\text{HYP}(P^{i-1}), \quad (13)\end{aligned}$$

where the factors $(1 \pm \varepsilon'/\mu)$ stem from the H_p being approximations. The equality $\mathrm{HYP}(P^{i-1} + q - \tilde{p}) = \mathrm{HYP}(P^{i-1}) + \mathrm{CON}_{P^{i-1}}(q) - \mathrm{CON}_{P^{i-1}+q}(\tilde{p})$ can be seen by replacing CON by its definition. Using this, inequality (12), and $\frac{1}{1-\varepsilon'/\mu} \leqslant 1 + \frac{2\varepsilon'}{\mu}$ for $\varepsilon' \leqslant 1/2$, we can simplify inequality (13) to

$$\mathrm{CON}_{P^{i-1}}(q) \leqslant \beta \mathrm{HYP}(P^{i-1}),$$

where

$$\beta = \left(1 + \frac{1}{\mu}\right)\left(1 + \frac{c}{\mu}\right)\left(1 + \frac{\varepsilon'}{\mu}\right)\left(1 + \frac{2\varepsilon'}{\mu}\right) - 1.$$

Together with submodularity (Lemma III.4) and using $\hat{P} \supseteq P^{i-1}$ and $\mathrm{HYP}(P^N) \geqslant \mathrm{HYP}(P^{i-1})$ we obtain

$$\mathrm{CON}_{\hat{P}}(q) \leqslant \mathrm{CON}_{P^{i-1}}(q) \leqslant \beta \mathrm{HYP}(P^{i-1}) \leqslant \beta \mathrm{HYP}(P^N). \tag{14}$$

Plugging in the definition of $c = 1 + 2\varepsilon'$, simplifying and roughly bounding the number of terms involving ε'/μ together with their coefficients, and using $\mu \geqslant 1$, yields

$$\beta \leqslant \frac{2}{\mu} + \frac{1}{\mu^2} + 64\frac{\varepsilon'}{\mu}. \tag{15}$$

Now we bound $\mathrm{HYP}(\hat{P})$. Consider $I = \{i \mid P^i \neq P^{i-1}, 1 \leqslant i \leqslant N\}$, the indices where P^i changed, and let p^i be the unique point in $P^{i-1} \setminus P^i$, i.e., the point we deleted in round $i \in I$. Then we have for $i \in I$ and every $p \in P^i + p^i$

$$\begin{aligned} \left(1 + \frac{\varepsilon'}{\mu}\right) \mathrm{HYP}(P^i) &\geqslant H_{p_i} \geqslant H_p \\ &\geqslant \left(1 - \frac{\varepsilon'}{\mu}\right) \mathrm{HYP}(P^i + p^i - p). \end{aligned}$$

Since $\mathrm{HYP}(P^i + p^i - p) = \mathrm{HYP}(P^i) + \mathrm{CON}_{P^i}(p^i) - \mathrm{CON}_{P^i+p^i}(p)$, this is equivalent to

$$\begin{aligned} &\frac{2\varepsilon'}{\mu}\left(1 - \frac{\varepsilon'}{\mu}\right)^{-1} \mathrm{HYP}(P^i) + \mathrm{CON}_{P^i+p^i}(p) \\ &\qquad \geqslant \mathrm{CON}_{P^i}(p^i). \end{aligned}$$

Summing over all $p \in P^i + p^i$ and using $\mathrm{HYP}(P) \geqslant \sum_{p \in P} \mathrm{CON}_P(p)$ (Lemma III.5) we get

$$\begin{aligned} &2\varepsilon'\left(1 + \frac{1}{\mu}\right)\left(1 - \frac{\varepsilon'}{\mu}\right)^{-1} \mathrm{HYP}(P^i) + \mathrm{HYP}(P^i + p^i) \\ &\qquad \geqslant (\mu + 1)\mathrm{CON}_{P^i}(p^i). \end{aligned}$$

This yields, after substituting $\mathrm{HYP}(P^i + p^i) = \mathrm{HYP}(P^i) + \mathrm{CON}_{P^i}(p^i)$ again,

$$\mathrm{CON}_{P^i}(p^i) \leqslant \gamma \mathrm{HYP}(P^i),$$

with $\gamma = \frac{1}{\mu}\big(1 + 2\varepsilon'\big(1 + \frac{1}{\mu}\big)\big(1 - \frac{\varepsilon'}{\mu}\big)^{-1}\big)$. Now, by a telescoping sum and submodularity (Lemma III.4) we have

$$\begin{aligned} \mathrm{HYP}(\hat{P}) &= \mathrm{HYP}(P^N) + \sum_{i \in I} \mathrm{CON}_{P^N \cup \ldots \cup P^i}(p^i) \\ &\leqslant \mathrm{HYP}(P^N) + \sum_{i \in I} \mathrm{CON}_{P^i}(p^i) \\ &\leqslant \mathrm{HYP}(P^N) + \gamma \sum_{i \in I} \mathrm{HYP}(P^i). \end{aligned}$$

As we go to a new population only if we have an improvement of a factor of at least $(1 + c/\mu)$, but we deal with approximations, we get $\delta \cdot \mathrm{HYP}(P^i) \geqslant \mathrm{HYP}(P^{i-1})$ for $i \in I$ and $\delta := (1 + c/\mu)^{-1}(1 - \varepsilon'/\mu)^{-1}(1 + \varepsilon'/\mu)$. Plugging this into the above inequality yields

$$\begin{aligned} \mathrm{HYP}(\hat{P}) &\leqslant \mathrm{HYP}(P^N)\left(1 + \gamma \sum_{i=0}^{N-1} \delta^i\right) \\ &\leqslant \mathrm{HYP}(P^N)\left(1 + \frac{\gamma}{1 - \delta}\right). \end{aligned} \tag{16}$$

Plugging in the definitions of $c = 1 + 2\varepsilon'$, δ and γ, simple calculations and rough estimations using $\mu \geqslant 1$ and $\varepsilon' \leqslant 1/6$ show that

$$\frac{\gamma}{1 - \delta} \leqslant 1 + \frac{1}{\mu} + 40\varepsilon'. \tag{17}$$

Now we can take equation (11), plug in equations (14) and (16), and simplify using equations (15) and (17) to get

$$\begin{aligned} \mathrm{HYP}(P^*) &\leqslant \mathrm{HYP}(\hat{P}) + \sum_{q \in P^* \setminus \hat{P}} \mathrm{CON}_{\hat{P}}(q) \\ &\leqslant \left(1 + \frac{\gamma}{1 - \delta} + \mu\beta\right) \mathrm{HYP}(P^N) \\ &\leqslant \left(4 + \frac{2}{\mu} + 104\varepsilon'\right) \mathrm{HYP}(P^N) \\ &= \left(4 + \frac{2}{\mu} + \varepsilon\right) \mathrm{HYP}(P^N). \end{aligned}$$

□

D. Lower bound for Competitiveness

It is useful to relate the notion of approximation from Section V with the notion of competitiveness introduced in Section VI-A. The following lemma shows that competitiveness implies approximation.

Lemma VI.7. *If $\mathcal{A}$ is an α-competitive $(\mu + \lambda)$-archiving algorithm, then $\mathcal{A}$ is also an $(\alpha + \varepsilon)$-approximate $(\mu + 1)$-archiving algorithm for all $\varepsilon > 0$.*

To ease the application of Lemma VI.7, we first state a direct implication of its contraposition in the following Lemma VI.8. It allows us to transfer lower bounds for approximation to lower bounds for competitiveness.

Lemma VI.8. *If there is no α-approximate non-decreasing $(\mu + 1)$-archiving algorithm, then there is no $(\alpha - \varepsilon)$-competitive non-decreasing $(\mu + \lambda)$-archiving algorithm for any $\varepsilon > 0$ and $\lambda \geqslant 1$.*

We can now easily combine Theorem V.7 and Lemma VI.8 and get the following corollary.

Theorem VI.9. *There is no $\big(1.1338 - 0.1338/\mu - \varepsilon\big)$-competitive non-decreasing $(\mu + \lambda)$-archiving algorithm.*

Observe that the structure of the bound shown in Theorem VI.9 is natural as for $\mu = 1$ it proves that there is no $(1-\varepsilon)$-competitive non-decreasing $(1+1)$-archiving algorithm while a greedy $(1 + 1)$-archiving algorithm is obviously 1-competitive. For $\mu \geqslant 2$, Theorem VI.9 implies that there is no $(\mu+\lambda)$-archiving algorithm with a competitive ratio of 1.06 or less.

Proof of Lemma VI.7. Let $\mathcal{A}$ be an α-competitive $(\mu+\lambda)$-archiving algorithm. Let $\mathcal{Y} \subset \mathbb{R}_+^d$ with finite $\mathrm{maxHYP}_\mu(\mathcal{Y})$ and $P^0 \subseteq \mathcal{Y}$ a μ-population. Since $\mathcal{Y}$ may be infinite there may not be a hypervolume maximizing set in it. However, the supremum in the definition of maxHYP guarantees the existence of a μ-population $P^* \subseteq \mathcal{Y}$ with $\mathrm{HYP}(P^*) \geqslant \frac{\alpha}{\alpha+\varepsilon} \mathrm{maxHYP}_\mu(\mathcal{Y})$. For $P^* = \{p_1^*, \ldots, p_\mu^*\}$ we set $Q^i := \{p_i^*\}$ for $i = 1, \ldots, \mu$. Consider the instance $I = (P^0, Q^1, \ldots, Q^\mu)$. $\mathcal{A}$ is α-competitive, so we have

$$\mathcal{A}(P^0, Q^1, \ldots, Q^\mu) \geqslant \frac{1}{\alpha} \mathrm{maxHYP}_\mu(\mathrm{Obs}(I)),$$

but $\mathrm{Obs}(I)$ contains P^*, so $\mathrm{maxHYP}_\mu(\mathrm{Obs}(I)) \geqslant \mathrm{HYP}(P^*)$. Putting everything together we see that there exists a sequence of offspring such that $\mathcal{A}(P^0, Q^1, \ldots, Q^\mu) \geqslant \frac{1}{\alpha+\varepsilon} \mathrm{maxHYP}_\mu(\mathcal{Y})$, so $\mathcal{A}$ is an $(\alpha+\varepsilon)$-approximate $(\mu+1)$-archiving algorithm. □

VII. Conclusion

The first question when theoretically analyzing evolutionary algorithms is typically convergence. We considered an abstract hypervolume-based multi-objective evolutionary algorithm (MOEA) without problem-specific assumptions on the structure of the search space or algorithm-specific assumptions on the generation of the initial and offspring population.

Assuming the offspring generation to be best-case, 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 (for any $\lambda \to \infty$, even if $\mu \gg \lambda$), but leave this as an open question. For practically used archiving algorithms our results suggest using $\lambda \geqslant \mu$ as this is necessary for being able to end up in a population maximizing the hypervolume even when assuming an optimal offspring generation.

We also studied the behavior of archiving algorithms assuming the offspring generation to be worst-case. For this setting, we have proven that increasing archiving algorithms are computationally inefficient and achieve only an unbounded competitive ratio of μ. The same holds for greedy archiving algorithms used in hypervolume-based MOEAs like SIBEA [17], SMS-EMOA [1], and the generational MO-CMA-ES [10, 11]. In sharp contrast to this, we presented a non-decreasing archiving algorithm which not only achieves a constant competitive ratio of $4 + 2/\mu$, but is also efficiently computable. This new archiving algorithm can be implemented efficiently, but it remains open to find a practical and computationally efficient archiving algorithm which is at the same time effective and competitive, that is, performing well for best-case and worst-case offspring generation.

We focused on best-case and worst-case offspring generation. Another interesting option would be an average-case offspring generation. It might be the case that in this setting locally optimal archiving algorithms show a better performance, as results such as Theorem VI.3 might do not hold in the average-case. The disadvantage of such a model would be the necessity of modeling other parts of the (until now) abstract MOEA, namely the generation of the initial population and offspring as well as the structure of the search space. Such average-case results would therefore be necessarily less general. It might be possible to compensate this by considering a smoothed model [14] instead.

While this paper solely focused on hypervolume-based MOEAs, it is worth studying the convergence behavior of other MOEAs which are, e.g., based on the ε-indicator. This might yield interesting results for comparing different indicators.

Acknowledgements

This research endavor was triggered by Lothar Thiele, who pointed us to Theorem V.2 and posed the question whether Theorem V.5 holds or the opposite.

Karl Bringmann is a recipient of the *Google Europe Fellowship in Randomized Algorithms*, and this research is supported in part by this Google Fellowship. Tobias Friedrich is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 618091 (SAGE).

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] A. Borodin and R. El-Yaniv. *Online computation and competitive analysis*, Vol. 2. Cambridge University Press Cambridge, 1998.

[3] L. Bradstreet, L. Barone, and L. While. Maximising hypervolume for selection in multi-objective evolutionary algorithms. In *IEEE Congress on Evolutionary Computation (CEC '06)*, pp. 6208–6215, 2006.

[4] K. Bringmann and T. Friedrich. Parameterized average-case complexity of the hypervolume indicator. In *15th Annual Conference on Genetic and Evolutionary Computation (GECCO '13)*, pp. 575–582. ACM Press, 2013.

[5] 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.

[6] K. Bringmann and T. Friedrich. An efficient algorithm for computing hypervolume contributions. *Evolutionary Computation*, 18:383–402, 2010.

[7] K. Bringmann and T. Friedrich. Approximating the least hypervolume contributor: NP-hard in general, but fast in practice. *Theoretical Computer Science*, 425:104–116, 2012.

[8] K. Bringmann and T. Friedrich. Convergence of hypervolume-based archiving algorithms I: Effectiveness.

In *13th Annual Conference on Genetic and Evolutionary Computation (GECCO '11)*, pp. 745–752. ACM Press, 2011.

[9] K. Bringmann and T. Friedrich. Convergence of hypervolume-based archiving algorithms II: Competitiveness. In *14th Annual Conference on Genetic and Evolutionary Computation (GECCO '12)*, pp. 457–464. ACM Press, 2012.

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

[11] C. Igel, T. Suttorp, and N. Hansen. Steady-state selection and efficient covariance matrix update in the multi-objective CMA-ES. In *4th International Conference on Evolutionary Multi-Criterion Optimization (EMO '07)*, Vol. 4403 of *LNCS*, pp. 171–185. Springer, 2007.

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

[13] G. Nemhauser, L. Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions – I. *Mathematical Programming*, 14:265–294, 1978.

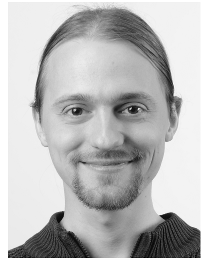
[14] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51:385–463, 2004.

[15] T. Ulrich and L. Thiele. Bounding the effectiveness of hypervolume-based $(\mu + \lambda)$-archiving algorithms. In *6th International Conference on Learning and Intelligent Optimization (LION '12)*, pp. 235–249, 2012.

[16] 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.

[17] E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In *4th International Conference on Evolutionary Multi-Criterion Optimization (EMO '07)*, Vol. 4403 of *LNCS*, pp. 862–876. Springer, 2007.

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



Karl Bringmann received his MSc degrees in computer science and mathematics from Saarland University in Saarbrücken, Germany, in 2011. Since 2011 he is a PhD student in the Algorithms and Complexity Group at the Max Planck Institute for Informatics in Saarbrücken, Germany. In 2012 he received the Google Europe Fellowship in Randomized Algorithms. His research interests lie in randomized algorithms (including sampling algorithms and evolutionary algorithms), computational geometry, and lower bounds.

Tobias Friedrich received his MSc in computer science from the University of Sheffield, UK, in 2003 and his diploma in mathematics from the University of Jena, Germany, in 2005. In 2007 he obtained a PhD in computer science from Saarland University, Germany. Afterwards, he was a postdoc in the Algorithms Group of the International Computer Science Institute, Berkeley, USA. From 2011 till 2012 he was senior researcher at the Max Planck Institute for Informatics and independent research group leader at the Cluster of Excellence on "Multimodal Computing and Interaction" in Saarbrücken. Since August 2012 he is full professor and chair of theoretical computer science at the University of Jena, Germany. The central topics of his work are randomized methods in mathematics and computer science and randomized algorithms (both classical and evolutionary).