The Maximum Hypervolume Set Yields Near-optimal Approximation

Karl Bringmann
Universität des Saarlandes
Saarbrücken, Germany

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

ABSTRACT

In order to allow a comparison of (otherwise incomparable) sets, many evolutionary multiobjective optimizers use indicator functions to guide the search and to evaluate the performance of search algorithms. The most widely used indicator is the hypervolume indicator. It measures the volume of the dominated portion of the objective space.

Though the hypervolume indicator is very popular, it has not been shown that maximizing the hypervolume indicator is indeed equivalent to the overall objective of finding a good approximation of the Pareto front. To address this question, we compare the optimal approximation factor with the approximation factor achieved by sets maximizing the hypervolume indicator. We bound the optimal approximation factor of n points by $1 + \Theta(1/n)$ for arbitrary Pareto fronts. Furthermore, we prove that the same asymptotic approximation ratio is achieved by sets of n points that maximize the hypervolume indicator. This shows that the speed of convergence of the approximation ratio achieved by maximizing the hypervolume indicator is asymptotically optimal.

This implies that for large values of n, sets maximizing the hypervolume indicator quickly approach the optimal approximation ratio. Moreover, our bounds show that also for relatively small values of n, sets maximizing the hypervolume indicator achieve a near-optimal approximation ratio.

Categories and Subject Descriptors

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

General Terms

Theory, Algorithms, Measurement, Performance

Keywords

Multiobjective Optimization, Performance Measures, Hypervolume Indicator, Approximation

1. INTRODUCTION

How to compare Pareto sets lies at the heart of research in multiobjective optimization. Most popular are methods that assign to each Pareto set a real number that somehow reflects its quality. Such functions are called unary indicator functions. A very desirable property of such indicator functions is that they should be strictly Pareto compliant [31]. That is, the indicator should value a Pareto set A higher than a Pareto set B if A dominates B. The only known unary indicator that has this feature is the *hypervolume indicator* (HYP). It measures the volume of the dominated portion of the objective space (for a formal definition see equation (2.1)). The hypervolume indicator was first introduced for performance assessment in multiobjective optimization by Zitzler and Thiele [30]. Later on it was used to guide the search in various hypervolume-based evolutionary optimizers (e.g. [6, 13, 15, 17, 29, 32]). Hypervolume-based optimizers have become very popular in recent years. They all aim (in different ways) to find a solution or Pareto set that maximizes HYP. However, it is not obvious that this aim is the same as the original objective of finding a good approximation set of the Pareto front.

The distribution of the points of sets maximizing the hypervolume has been examined in several previous papers. Some empirical studies observed that a solution set that maximizes HYP is somehow "well distributed" [13, 16, 17]. Others observed that "convex regions may be preferred to concave regions" [20, 30] while further authors argued that HYP is "biased towards the boundary solutions" [11]. For the number of points $n \to \infty$, Auger et al. [2] proved that the density of points only depends on the gradient, but not on their position on the front. On last years GECCO, Friedrich et al. [14] examined the approximation ratio of fronts maximizing the hypervolume. For linear and reciprocal functions they could prove that maximizing HYP achieves an optimal approximation while on other functions they showed empirically that both might differ. In contrast, in this paper we provide a rigorous analysis of the approximation quality of hypervolume maximizing sets. This issue was wide open so far though it is crucial for understanding the implicit optimization goal when using the hypervolume indicator as a quality measure for populations.

The quality of single-objective optimization problems is typically measured by its multiplicative approximation factor. For maximization problems this is the ratio between the optimal value and the best found value. This notion generalizes gently to our multiobjective setting. We say a Pareto set is an α-approximation if it approximately dominates the

Pareto curve, that is, if for every point on the Pareto curve, the Pareto set contains a point that is at least as good approximately (within a factor α) in all objectives. For a selection of papers using this approach, see e.g. [9, 12, 22, 23, 25] or most recently [10] and references therein.

To define this properly, let us look at a maximization problem with a front that can be described by a monotonically decreasing function $f\colon [a, A] \to [b, B]$ with $0 < a < A$, $0 < b < B$ as shown in Figure 1. Then the approximation ratio (cf. Definition 2.1) of a tupel $X := (x_1, \ldots, x_n)$ (called solution set) is the least α such that for each $x \in [a, A]$ there is an $x_i \in X$ with

$$x \leqslant \alpha\, x_i \quad \text{and} \quad f(x) \leqslant \alpha\, f(x_i).$$

The approximation ratio does not depend on the scaling of $[a, A]$ and $[b, B]$. This can be seen by observing that for fixed constants $\mu, \nu > 0$, the function $f'\colon [\mu\, a, \mu\, A] \to [\nu\, b, \nu\, B]$ with $f'(x) = \nu\, f(x/\mu)$ achieves the same approximation ratio α with the solution set $X' := (\mu x_1, \ldots, \mu x_n)$. However, the approximation ratio significantly depends on the proportions A/a and B/b. To see this, let us look at a function $f_\varepsilon\colon [1, A] \to [A^{-1}, 1]$ with

$$f_\varepsilon(x) := \begin{cases} 1 & \text{for } x \leqslant 1 + \varepsilon, \\ 1/x & \text{otherwise.} \end{cases} \tag{1.1}$$

It is easy to see that for $0 < \varepsilon < A$ there is exactly one point on the front, namely $(1 + \varepsilon, 1) \in [a, A] \times [b, B]$, which maximizes the dominated space[1]. The approximation achieved by this point is $A/(1+\varepsilon)$ while $(\sqrt{A}, 1/\sqrt{A}) \in [a, A] \times [b, B]$ achieves the optimal approximation ratio $\sqrt{A}$. Hence for $\varepsilon \to 0$, the approximation ratio of the solution set maximizing the hypervolume is off by a factor of $\sqrt{A}$ from the optimal ratio. This shows that the approximation ratio of sets maximizing the hypervolume can be very large for small numbers of points. However, this paper proves that this is *not* the case for sufficiently large solution sets.

We are not interested in bounds for the approximation ratio of specific functions. Instead, we take a worst-case perspective and look at all[2] functions $f\colon [a, A] \to [b, B]$ with $0 < a < A$, $0 < b < B$ and $f(a) = B$, $f(A) = b$. We prove in Corollary 3.2 that for this class of functions there are solution sets of size n with an (optimal) approximation ratio of

$$1 + \frac{\log(\min\{A/a, B/b\})}{n} = 1 + \Theta\left(\frac{1}{n}\right). \tag{1.2}$$

The asymptotic here is in the size of the solution set n as the class of functions is fixed and so is $[a, A]$ and $[b, B]$. Above bound gives a lower bound for the approximation ratio which can be achieved by any solution set. For sets which maximize the hypervolume we are able to prove an upper bound of

$$1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n - 4} = 1 + \Theta\left(\frac{1}{n}\right) \tag{1.3}$$

for the approximation ratio (see Corollary 4.7 for details).

[1]We are assuming here that the size of the dominated space is measured relative to a common reference point $R = (0, 0)$. For the formal definition of HYP, see equation (2.1) in Section 2.1.

[2]We restrict our attention to functions where there exists a set maximizing the hypervolume indicator. For technical details, see the definition of $\mathcal{F}$ at the beginning of Section 2.

The asymptotic in n shows that the speed of convergence to an approximation ratio of 1 for sets maximizing the hypervolume is of the same order as for sets with optimal approximation ratio. Note that this is surprising in several respects. First, it is good news (compare with the negative example of equation (1.1)!) that the set maximizing the hypervolume gives a multiplicative approximation at all. Especially as the set of fronts we look at is rather large, it was unexpected that we find an approximation guarantee for HYP that holds for all fronts uniformly. Second, it is even better news that with growing n the convergence to a perfect approximation ratio of 1 is of best order possible.

Of course, the dependence on $[a, A]$ and $[b, B]$ in the approximation ratio of the optimal solution set in equation (1.2) is less than in our upper bound given in equation (1.3) for the approximation factor for sets maximizing the hypervolume. However, our bound for the approximation ratio of sets maximizing the hypervolume is only an upper bound. Therefore we can only conclude that their approximation ratio is close to the optimal one, but not how far it is off. To give an intuition how tight both bounds still are, Figure 4 at the end of the paper gives an example for a certain class of functions.

The outline of the paper is as follows. In Section 2, we give precise definitions of the concepts introduced informally above. Section 3 then gives a tight bound for the optimal approximation factor. The approximation factor of the hypervolume indicator is analyzed in Section 4. We finish with a discussion of the implications and some open problems.

2. PRELIMINARIES

Throughout this study we consider bi-criterion maximization problems mapping from an arbitrary search space $\mathcal{S}$ to $\mathbb{R}^2$. The restriction to maximization problems is technically inconsequential, but burdens the exposition and notation. Similar results can be achieved for minimization problems.

2.1 Modeling a Front

We restrict ourselves to Pareto fronts that can be written as $\{(x, f(x)) \mid x \in [a, A]\}$ where $f\colon [a, A] \to [b, B]$ is a (not necessarily strictly) monotonically decreasing, upper semi-continuous[3] function with $f(a) = B$, $f(A) = b$ for some $0 < a < A$, $0 < b < B$. We write $\mathcal{F} = \mathcal{F}_{[a,A]\to[b,B]}$ for the set of all such functions f, and fix $[a, A]$, $[b, B]$ for the rest of the paper. We use the term *front* for both, the set of points $\{(x, f(x)) \mid x \in [a, A]\}$, and the function f.

The condition of f being upper semi-continuous cannot be relaxed further as without it there does not have to exist a set of points maximizing the hypervolume indicator (see equation (4.1) for a non-semi-continuous function which has no solution set maximizing HYP). Moreover, it allows us to define a meaningful inverse function $f^{-1}\colon [b, B] \to [a, A]$ by setting $f^{-1}(y) := \max\{x \in [a, A] \mid f(x) \geqslant y\}$. If f is not upper semi-continuous, this maximum does not necessarily exist.

[3]Semi-continuity is a weaker property than normal continuity. A function f is said to be upper semi-continuous if for all points x of its domain, $\limsup_{y\to x} f(y) \leqslant f(x)$. Intuitively speaking this means that for all points x the function values for arguments near x are either close to $f(x)$ or less than $f(x)$. For more details see e.g. Rudin [24].

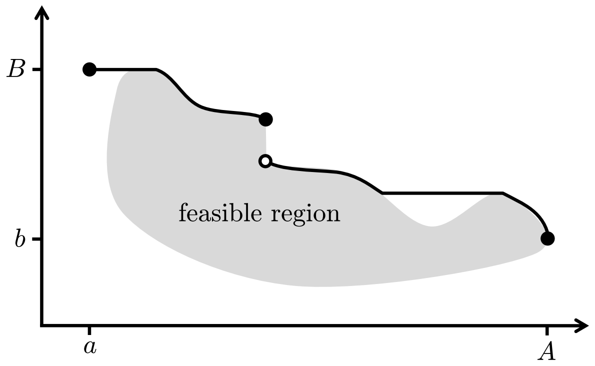

Figure 1: Illustration how an arbitrary feasible region of a maximization problem implies a monotonically decreasing, upper semi-continuous front as defined in Section 2.1.

See Figure 1 for an example of a function that is a front which satisfies above definition. Note that there are some points $(x, f(x))$ on the front (the black curve), that are obviously not part of the feasible region (the gray area). However, all such points are dominated by a point $(x', f(x'))$ that is contained in the feasible region. Hence, these added points change neither the set optimally approximating the front nor the set maximizing HYP. Thus, our modeling of the feasible region does not change the values we are interested in. Also note that at the x-coordinate where $f(x)$ makes a discrete "jump", we have defined $f(x)$ to be the greater of the two values. Defining $f(x)$ as the lesser value (the limit from right) would destroy upper semi-continuity.

A further example of disconnected Pareto fronts that fit in our framework can be found in Figure 2, which can be seen as the modeling of a discrete feasible region.

Note that the set $\mathcal{F}$ of fronts we consider is a very general one. Most papers that theoretically examine the hypervolume indicator assume that the front is continuous and differentiable (e.g. [1, 2, 14]), and are thus not able to give results about discrete fronts, which we can.

2.2 Approximation of a Front

Fix an $n \in \mathbb{N}$. For fixed $[a, A]$ we call a tupel $X = (x_1, \ldots, x_n), a \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant A$ a *solution set* and write $\mathcal{X}$ for the set of all such solution sets. We intend to find a solution set $X \in \mathcal{X}$ that constitutes a good approximation of a front f. According to the custom of approximation algorithms, we measure the quality of a front f by its multiplicative approximation factor. For this we use the following definition of Papadimitriou and Yannakakis [22].

Definition 2.1. *Let $f \in \mathcal{F}$ and $X = (x_1, x_2, \ldots, x_n) \in \mathcal{X}$. The solution set X is an* α-approximation *of f if for each $x \in [a, A]$ there is an $x_i \in X$ with*

$$x \leqslant \alpha\, x_i \quad \text{and} \quad f(x) \leqslant \alpha\, f(x_i)$$

where $\alpha \in \mathbb{R}$, $\alpha \geqslant 1$. The approximation ratio *of X with respect to f is defined as*

$$\alpha(f, X) := \inf\{\alpha \in \mathbb{R} \mid X \text{ is an } \alpha\text{-approximation of } f\}.$$

This definition of approximation has also been used to examine the approximation of the hypervolume indicator empirically in [14] and is similar to the definition of ε-dominance given in [18, 19].

The quality of an algorithm which calculates a solution set of size n for the Pareto curves in $\mathcal{F}$ has to be compared with the respective optimal approximation factor defined as follows.

Definition 2.2. *For fixed $[a, A]$, $[b, B]$, and n, let*

$$\alpha_{\mathrm{OPT}} := \sup_{f \in \mathcal{F}} \inf_{X \in \mathcal{X}} \alpha(f, X).$$

The value α_{OPT} is chosen such that every front in $\mathcal{F}$ can be approximated by n points to a factor of α_{OPT}, and there is a front which cannot be approximated better. In Section 3 we compute this value exactly.

2.3 Hypervolume indicator

In geometric terms, the hypervolume indicator of a solution set $X \in \mathcal{X}$ measures the volume of the space dominated by the points $(x_i, f(x_i))$ for $x_i \in X$. This space is truncated at a fixed foot point called the *reference point* $R = (R_x, R_y)$. The *hypervolume* $\mathrm{HYP}(X)$ of a solution set $X = (x_1, \ldots, x_n)$ is then defined as

$$\mathrm{HYP}(X) := \textsc{vol}\left(\bigcup_{i=1}^{n} [R_x, x_i] \times [R_y, f(x_i)]\right) \tag{2.1}$$

with $\textsc{vol}(\,\cdot\,)$ being the usual Lebesgue measure.

There are many algorithms for exact calculation of the hypervolume (e.g. [4, 5, 13, 21, 26–28]). It is known that the hypervolume cannot be calculated exactly in time polynomial in the number of dimensions unless P=NP [7]. However, there are several polynomial-time approximation algorithms [3, 7, 8] based on Monte Carlo sampling.

2.4 Approximation Factor of HYP

Recall that $n \in \mathbb{N}$ is fixed for the rest of the paper. We want to know how good the solution set of size n maximizing the hypervolume indicator approximates a Pareto front. Two problems arise with this formulation: First, there might be no solution set which maximizes the hypervolume indicator. More precisely, there might be a sequence of solution sets of size n with greater and greater hypervolume indicator value, but the limit of the values is not taken by any solution set of size n. We show in Lemma 4.1 that this cannot happen as we require the fronts to be upper semi-continuous. Second, there might be several solution sets maximizing HYP. We show that this can indeed occur and that the approximation factors of these solution sets can be very different. We focus on the worst case and hence give upper bounds for the approximation factor of all sets maximizing the hypervolume. This approach is elaborated in the following definition.

Definition 2.3. *For fixed $[a, A], [b, B], n$, and $f \in \mathcal{F}$ let*

$$\mathcal{X}^f_{\mathrm{HYP}} := \left\{X \in \mathcal{X} \;\middle|\; \mathrm{HYP}(X) = \max_{Y \in \mathcal{X}} \mathrm{HYP}(Y)\right\} \text{ and}$$

$$\alpha_{\mathrm{HYP}} := \sup_{f \in \mathcal{F}} \sup_{X \in \mathcal{X}^f_{\mathrm{HYP}}} \alpha(f, X).$$

The set $\mathcal{X}^f_{\mathrm{HYP}}$ is the set of all solution sets that maximize HYP. The value α_{HYP} is chosen such that for every front f in $\mathcal{F}$ every solution set maximizing HYP approximates f by a factor of at most α_{HYP}.

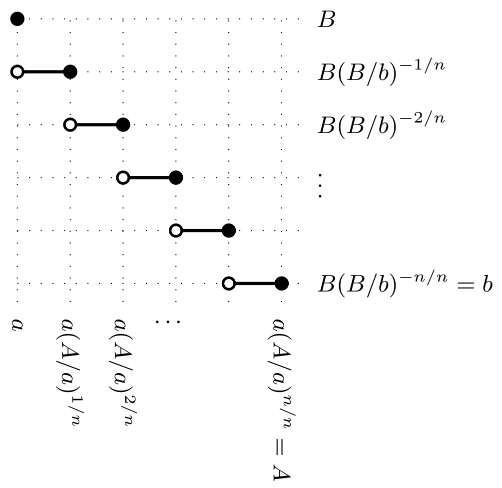

Figure 2: Front f used for the lower bound construction in the proof of Theorem 3.1.

3. BOUNDING THE BEST APPROXIMATION FACTOR α_{OPT}

In this section we examine the optimal approximation factor α_{OPT}. Recall that no set of n points can achieve a better approximation factor than α_{OPT}. This is the reason why bounds for α_{OPT} are important for comparison before examining α_{HYP} in Section 4. We can prove the following tight bound for α_{OPT}.

THEOREM 3.1. $\alpha_{\text{OPT}} = \min\{A/a, B/b\}^{1/n}$.

Proof. We first show $\alpha_{\text{OPT}} \leqslant (A/a)^{1/n}$. For this, let $\alpha := (A/a)^{1/n}$ and $x_i := a\,\alpha^{i-1}$ for $i \in \{1,\dots,n\}$. The x_i are an α-approximation of f as the point $(x_i, f(x_i))$ α-dominates all points $(x, f(x))$ for $x_i \leqslant x \leqslant x_{i+1}$. Hence, $\alpha_{\text{OPT}} \leqslant \alpha = (A/a)^{1/n}$.

To show that analogously $\alpha_{\text{OPT}} \leqslant (B/b)^{1/n}$, let $\alpha := (B/b)^{1/n}$ and $x_i := f^{-1}(B\alpha^{-i})$ for $i \in \{1,\dots,n\}$. Then $f(x_i) \geqslant B\alpha^{-i}$ and no point $(x, f(x))$ has $f(x_i) > f(x) > B\alpha^{-i}$. Hence, the point $(x_i, f(x_i))$ α-dominates all points $(x, f(x))$ with $B\alpha^{-i} \leqslant f(x) \leqslant B\alpha^{-i+1}$ and we get $\alpha_{\text{OPT}} \leqslant \alpha = (B/b)^{1/n}$.

It remains to prove the lower bound $\alpha_{\text{OPT}} \geqslant \min\{A/a, B/b\}^{1/n}$. For this, we set $f(x) := B(B/b)^{-i/n}$ for $a(A/a)^{(i-1)/n} < x \leqslant a(A/a)^{i/n}$ and $i \in \{0,\dots,n\}$. Then f is a front which consists of $(n+1)$ levels. It is illustrated in Figure 2. Let us now consider a solution set $(x_1,\dots,x_n)$ consisting of n points. As f has $n+1$ levels, the pigeonhole principle gives that there is at least one level where there is none of the n points. This implies that the rightmost point in this level is only approximated by a factor of $\min\{(A/a)^{1/n}, (B/b)^{1/n}\}$. $\square$

COROLLARY 3.2. *For all $n \geqslant \log(\min\{A/a, B/b\})/\varepsilon$ and $\varepsilon \in (0,1)$,*

$$\alpha_{\text{OPT}} \geqslant 1 + \frac{\log(\min\{A/a, B/b\})}{n},$$
$$\alpha_{\text{OPT}} \leqslant 1 + (1+\varepsilon)\,\frac{\log(\min\{A/a, B/b\})}{n}.$$

Proof. Both inequalities follow directly from Theorem 3.1. For the first inequality note that $e^x \geqslant 1 + x$ for all $x \in \mathbb{R}$. For the second we upper bound e^x with $0 \leqslant x \leqslant \varepsilon$ by

$$\begin{aligned} e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!} \leqslant 1 + \sum_{k=1}^{\infty} \frac{x^k}{2^{k-1}} \leqslant 1 + x \sum_{k=0}^{\infty} \frac{\varepsilon^k}{2^k} \\ &= 1 + x \frac{1}{1-\varepsilon/2} \leqslant 1 + (1+\varepsilon)x, \end{aligned}$$

as $(1+\varepsilon)\,(1-\varepsilon/2) \geqslant 1$. $\square$

Observe that for a fixed function class $\mathcal{F}$ (and hence fixed $[a, A]$, $[b, B]$) Corollary 3.2 implies

$$\alpha_{\text{OPT}} = 1 + \Theta(1/n)$$

as claimed in equation (1.2).

4. BOUNDING THE APPROXIMATION RATIO α_{HYP} OF THE MAXIMUM HYPERVOLUME SOLUTION SET

In this section we examine α_{HYP}, the approximation ratio of a set maximizing the hypervolume indicator. We start by showing that without upper semi-continuity there does not necessarily exist a solution set maximizing HYP. To see this, consider the front $f\colon [1,2] \to [1,2]$ with

$$f(x) := \begin{cases} 1 & \text{for } x = 2, \\ 2 & \text{for } 1 \leqslant x < 2. \end{cases} \tag{4.1}$$

and reference point $R = (0,0)$. The one element solution set $X = (2-\varepsilon)$ achieves $\text{HYP}(X) = 4 - 2\,\varepsilon$ for each $\varepsilon > 0$. However, no solution set X can have $\text{HYP}(X) = 4$, as $f(2) = 1 < 2$. Thus, there exists no solution set maximizing HYP as there is an infinite series of solution sets with greater and greater hypervolume indicator, but the limit (which is 4) is not taken by any solution set.

Next we prove that conditioning on our fronts being upper semi-continuous implies that there are sets maximizing HYP. In more detail, there is a solution set X of size n which maximizes HYP among all solution sets of size n.

LEMMA 4.1. *Let $f \in \mathcal{F}$, $n \in \mathbb{N}$. Then there exists a (not necessarily unique) solution set $X \in \mathcal{X}$ maximizing the hypervolume indicator $\text{HYP} = \text{HYP}_f$ on $\mathcal{X}$.*

Proof. If we compute HYP in each of the intervals $[x_i, x_{i+1}]$ and sum up, we can easily see that we have for a solution set $X = (x_1,\dots,x_n)$

$$\text{HYP}(X) = \sum_{i=1}^{n} (x_i - x_{i-1})\,(f(x_i) - R_y),$$

where $x_0 := R_x$ and $R = (R_x, R_y)$ is the reference point. Now, for a series of solution sets $(X^{(j)})_{j\in\mathbb{N}}$ with $X^{(j)} \to X$

for $j \to \infty$ in $[a, A]^n$ we have

$$\begin{aligned}\lim_{j\to\infty} \mathrm{HYP}(X^{(j)}) &= \lim_{j\to\infty} \sum_{i=1}^{n} (x_i^{(j)} - x_{i-1}^{(j)})\,(f(x_i^{(j)}) - R_y) \\ &= \sum_{i=1}^{n} (x_i - x_{i-1}) \left(\lim_{j\to\infty} f(x_i^{(j)}) - R_y \right) \\ &\leqslant \sum_{i=1}^{n} (x_i - x_{i-1})\,(f(x_i) - R_y) \\ &= \mathrm{HYP}(X),\end{aligned}$$

where we used $x_i - x_{i-1} \geqslant 0$ and the upper semi-continuity of f. Thus, we showed that HYP is upper semi-continuous, too.

As an upper semi-continuous function, HYP has a maximum on the compact set $\mathcal{X}$ of all solution sets. □

To confirm that this set is indeed not unique in general, let us look again at the introductory negative example function f_ε from equation (1.1).By choosing $\varepsilon = 0$ and $A = 4$ we get a function $f_0\colon [1,4] \to [1/4,1]$ with $f_0(x) = 1/x$. With reference point $R = (0,0)$ we get $\mathrm{HYP}((x)) = x\,(1/x) = 1$ for all $x \in [1,4]$. Therefore the set of solution sets of size $n = 1$ which maximize HYP is far from unique as $\mathcal{X}^f_{\mathrm{HYP}} = \{(x) \mid x \in [1,4]\}$. Moreover, this example shows that the approximation factors of two solution sets maximizing HYP can differ significantly as the solution set (1) achieves an approximation factor of 4, while the solution set (2) achieves an approximation factor of 2.

We now want to give bounds on the approximation factor achieved by the sets maximizing the hypervolume indicator. For this recall that the *contribution* of a point $x \in X$ to the hypervolume of a solution set $X \in \mathcal{X}$ is the volume dominated by x and no other element of X. More formally, the contribution of a point x is $\mathrm{HYP}(X) - \mathrm{HYP}(X \setminus x)$. In the following we mainly deal with the minimal contribution defined as follows.

Definition 4.2. *Let $f \in \mathcal{F}$, $n \geqslant 3$ and $X = (x_1, \ldots, x_n) \in \mathcal{X}$. Then the minimal contribution of this solution set X is*

$$\textsc{MinCon}(X) := \min_{2\leqslant i\leqslant n-1} \big(x_i - x_{i-1}\big)\big(f(x_i) - f(x_{i+1})\big).$$

Figure 3 gives an illustration of Definition 4.2. Note that above definition of $\textsc{MinCon}(X)$ is independent of the reference point R, as it only considers the minimal contribution of any of the points $x_2, \ldots, x_{n-1}$. Restricted to these $(n-2)$ inner points, it corresponds to the definition of $\textsc{MinCon}(X)$ in [8].

We first show the following upper bound for $\textsc{MinCon}(X)$.

Lemma 4.3. *Let $f \in \mathcal{F}$, $n \geqslant 3$, and $X = (x_1 \ldots, x_n) \in \mathcal{X}$. Then,*

$$\textsc{MinCon}(X) \leqslant \frac{(x_n - x_1)\,(f(x_1) - f(x_n))}{(n-2)^2}.$$

Proof. Let $a_i := x_i - x_{i-1}$ for $2 \leqslant i \leqslant n$ and $b_i := f(x_i) - f(x_{i+1})$ for $1 \leqslant i \leqslant n-1$. Plugging this notation in Definition 4.2 gives $\textsc{MinCon}(X) = \min_{2\leqslant i\leqslant n-1} a_i b_i$ and

$$a_i \geqslant \textsc{MinCon}(X)/b_i \qquad \text{for all } 2 \leqslant i \leqslant n-1.$$

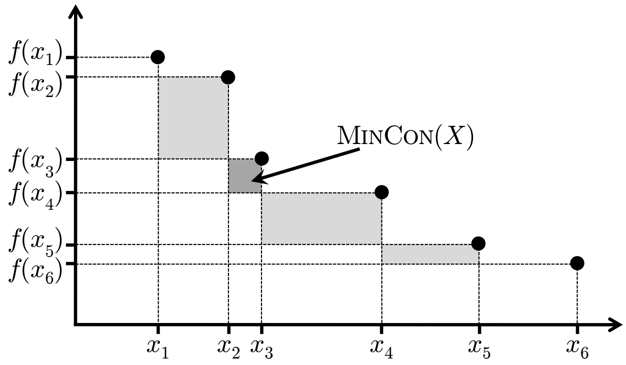

Figure 3: The minimal contribution $\textsc{MinCon}(X)$ of a solution set $X = (x_1, x_2, \ldots, x_6)$ is in Definition 4.2 defined to be the least hypervolume contribution $\mathrm{HYP}(X) - \mathrm{HYP}(X \setminus x)$ for $x \in \{x_2, x_3, x_4, x_5\}$.

This implies

$$\begin{aligned}\sum_{i=2}^{n-1} \textsc{MinCon}(X)/b_i &\leqslant \sum_{i=2}^{n-1} a_i \leqslant \sum_{i=2}^{n} a_i \\ &= \sum_{i=2}^{n} x_i - \sum_{i=1}^{n-1} x_i = x_n - x_1,\end{aligned}$$

and therefore

$$\textsc{MinCon}(X) \leqslant \frac{x_n - x_1}{\sum_{i=2}^{n-1} 1/b_i}.$$

We can now use that the harmonic mean is less than the arithmetic mean, that is,

$$\frac{n-2}{\sum_{i=2}^{n-1} 1/b_i} \leqslant \frac{\sum_{i=2}^{n-1} b_i}{n-2}$$

to obtain

$$\begin{aligned}\textsc{MinCon}(X) &\leqslant \frac{(x_n - x_1)\,\sum_{i=2}^{n-1} b_i}{(n-2)^2} \\ &\leqslant \frac{(x_n - x_1)\,(f(x_1) - f(x_n))}{(n-2)^2}.\end{aligned}$$ □

First, we calculate the approximation ratio of the "inner points", i.e., points $x \in [x_1, x_n]$. In a second step we determine how well the "outer points" x with $x < x_1$ or $x > x_n$ are approximated.

Theorem 4.4. *Let $f \in \mathcal{F}$ and $n > 4$. Every solution set $(x_1, \ldots, x_n) \in \mathcal{X}^f_{\mathrm{HYP}}$ achieves a*

$$1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}$$

multiplicative approximation of all points $(x, f(x))$ with $x \in [x_1, x_n]$.

Proof. Assume there is a point x, which is not approximated by a factor of

$$\alpha := 1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}. \tag{4.2}$$

Let i be such that $x_i < x < x_{i+1}$ and therefore

$$\begin{aligned} x &> \alpha\, x_i, \\ f(x) &> \alpha\, f(x_{i+1}). \end{aligned} \tag{4.3}$$

We know

$$\textsc{MinCon}(X) \geqslant (x - x_i)\,(f(x) - f(x_{i+1})), \tag{4.4}$$

as otherwise one could replace the point contributing $\textsc{MinCon}(X)$ by x and increase the hypervolume, which is a contradiction.

Plugging equations (4.3) in equation (4.4), we get

$$\textsc{MinCon}(X) > (\alpha - 1)^2\, x_i\, f(x_{i+1}). \tag{4.5}$$

Observe that there are i points in $\{x_1, \dots, x_n\} \cap [a, x_i]$. Hence for $3 \leqslant i \leqslant n-1$ we can upper bound the minimal contribution using Lemma 4.3 on the points $x_1, \dots, x_i$ by

$$\begin{aligned} \textsc{MinCon}(X) &\leqslant (x_i - x_1)\,(f(x_1) - f(x_i))/(i-2)^2 \\ &\leqslant x_i\, B/(i-2)^2. \end{aligned} \tag{4.6}$$

Analogously, for $1 \leqslant i \leqslant n-3$ we can upper bound the minimal contribution using Lemma 4.3 on the points $x_{i+1}, \dots, x_n$ by

$$\begin{aligned} \textsc{MinCon}(X) &\leqslant (x_n - x_{i+1})\,(f(x_{i+1}) - f(x_n))/(n-i-2)^2 \\ &\leqslant A\, f(x_{i+1})/(n-i-2)^2. \end{aligned} \tag{4.7}$$

Combining equation (4.5) with equations (4.6) and (4.7), we get for $3 \leqslant i \leqslant n-3$ that

$$(\alpha - 1)^2\, x_i\, f(x_{i+1}) < \min\left\{ \frac{x_i\, B}{(i-2)^2}, \frac{A\, f(x_{i+1})}{(n-i-2)^2} \right\}$$

or, equivalently,

$$\alpha < 1 + \min\left\{ \frac{\sqrt{B/f(x_{i+1})}}{i-2}, \frac{\sqrt{A/x_i}}{n-i-2} \right\}$$

which yields with $x_i \geqslant a$ and $f(x_{i+1}) \geqslant b$ that

$$\alpha < 1 + \min\left\{ \frac{\sqrt{B/b}}{i-2}, \frac{\sqrt{A/a}}{n-i-2} \right\} \tag{4.8}$$

for $3 \leqslant i \leqslant n-3$.

Now, the right hand side of equation (4.8) gets maximal if the two terms are equal since one of them is monotonically increasing in i and the other one is monotonically decreasing in i. As this happens exactly for $i = 2 + \frac{(n-4)\sqrt{B/b}}{\sqrt{A/a}+\sqrt{B/b}}$, we get the upper bound

$$\alpha < 1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}$$

for $3 \leqslant i \leqslant n-3$. This contradicts with equation (4.2) and proves that every point $(x, f(x))$ with $x \in [x_3, x_{n-2}]$ is multiplicatively approximated by a factor of α.

It remains to show a contradiction to equation (4.3) for $i = 1, 2$ and $i = n-2, n-1$. For $i = 1, 2$ we get from equations (4.5) and (4.7) that

$$\alpha < 1 + \frac{\sqrt{A/a}}{n-i-2} \leqslant 1 + \frac{\sqrt{A/a}}{n-4}.$$

which contradicts with equation (4.2).

Finally, for $i = n-2, n-1$ we get from equations (4.5) and (4.6) that

$$\alpha < 1 + \frac{\sqrt{B/b}}{i-2} \leqslant 1 + \frac{\sqrt{B/b}}{n-4}.$$

which also contradicts with equation (4.2) and finishes the proof. □

It remains to examine the approximation factor of the "outer points" x with $x < x_1$ or $x > x_n$.

Theorem 4.5. *Let $f \in \mathcal{F}$, $n > 3$, and $R = (R_x, R_y) \leqslant (0,0)$ be the reference point. Every solution set $(x_1, \dots, x_n) \in \mathcal{X}^f_{\text{HYP}}$ achieves a*

$$1 + \frac{A}{(a - R_x)\,(n-2)^2}$$

multiplicative approximation of all points $(x, f(x))$ with $x < x_1$, and a

$$1 + \frac{B}{(b - R_y)\,(n-2)^2}$$

multiplicative approximation of all points $(x, f(x))$ with $x > x_n$.

Proof. We show the theorem only for $x \leqslant x_1$. The case $x \geqslant x_n$ follows by symmetry in the two objectives.

The approximation factor of any $x \leqslant x_1$ is exactly $f(x)/f(x_1)$. This is maximized for $x = a$, so that the approximation factor of any $x \leqslant x_1$ is at most $B/f(x_1)$. We show that $B/f(x_1)$ is less than $1 + \frac{A}{(a-R_x)\,(n-2)^2}$.

Using Lemma 4.3 on the points $x_1, \dots, x_n$ we get that

$$\begin{aligned} \textsc{MinCon}(X) &\leqslant (x_n - x_1)\,(f(x_1) - f(x_n))/(n-2)^2 \\ &\leqslant A\, f(x_1)/(n-2)^2. \end{aligned} \tag{4.9}$$

Let x_i be a point with contribution $\textsc{MinCon}(X)$. We define another solution set $X' := X \setminus \{x_i\} \cup \{a\}$ which contains the point a instead of the point x_i. By definition,

$$\text{HYP}(X') = \text{HYP}(X) - \textsc{MinCon}(X) + (a - R_x)\,(B - f(x_1)).$$

Together with $\text{HYP}(X) \geqslant \text{HYP}(X')$ this yields

$$\textsc{MinCon}(X) \geqslant (a - R_x)\,(B - f(x_1)). \tag{4.10}$$

Combining equations (4.9) and (4.10), we finally get the desired

$$\frac{B}{f(x_1)} \leqslant 1 + \frac{A}{(a - R_x)\,(n-2)^2}.$$

□

Together Theorems 4.4 and 4.5 imply the following corollary.

Corollary 4.6. *Let $f \in \mathcal{F}$, $n > 4$, and let $R = (R_x, R_y) \leqslant (0,0)$ be the reference point. Then*

$$\alpha_{\text{HYP}} \leqslant 1 + \max\left\{ \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}, \frac{A}{(a - R_x)\,(n-2)^2}, \frac{B}{(b - R_y)\,(n-2)^2} \right\}.$$

For sufficiently large n *or* sufficiently small coordinates of the reference point, the two last terms in Corollary 4.6 are less than the first one. More precisely, it is easy to see the following slightly simplified result.

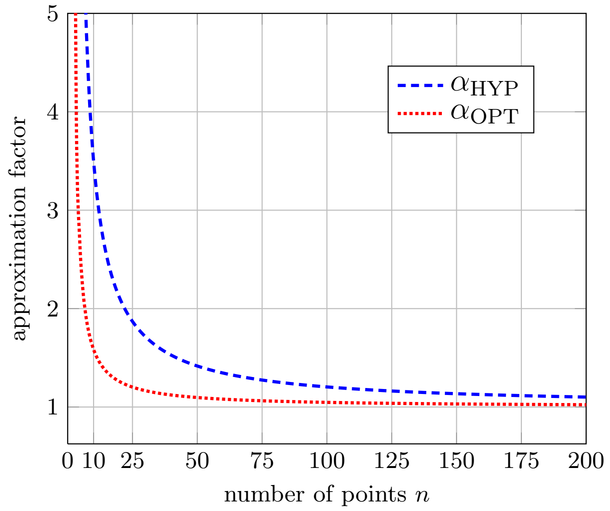

Figure 4: Comparison of the proven bounds for the approximation factors for all functions $f \in \mathcal{F}_{[1,100]\to[1,100]}$. α_{OPT} (⋯⋯) gives the tight bound of Theorem 3.1 for the best possible approximation factor for sets of size n. This is compared with our upper bound for α_{HYP} (- - - -) of Corollary 4.7 which measures the approximation factor achieved by sets of size n maximizing the hypervolume assuming a reference point $R \leqslant (-10, -10)$.

COROLLARY 4.7. *Let $f \in \mathcal{F}, n > 4$, and let $R = (R_x, R_y) \leqslant (0, 0)$ be the reference point. If* either

- $n \geqslant 2 + \max\{\sqrt{A/a}, \sqrt{B/b}\}$ or
- $R_x \leqslant -\sqrt{A\,a}/n$, $R_y \leqslant -\sqrt{B\,b}/n$,

we have

$$\alpha_{\mathrm{HYP}} \leqslant 1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}.$$

Hence, for a fixed function class $\mathcal{F}$ (and, thus, fixed $[a, A]$, $[b, B]$), we get the same asymptotic bound for α_{HYP} as we got for α_{OPT}, that is,

$$\alpha_{\mathrm{HYP}} = 1 + \Theta(1/n).$$

as claimed in equation (1.3).

5. DISCUSSION

The hypervolume indicator is used more and more to measure the quality of a population in evolutionary multiobjective algorithms. This indirectly changes the optimization goal and it was so far not known whether this new goal also gives a good approximation of the Parto front. We have examined the approximation factor α_{HYP} of the sets maximizing the hypervolume indicator and the optimal approximation factor α_{OPT}. We could prove that for all monotonically decreasing, upper semi-continuous Pareto fronts, the asymptotic behavior of α_{HYP} and α_{OPT} in the number of points n is the same, namely $1 + \Theta(1/n)$. However, the constant factor hidden by the Θ might be larger for α_{HYP}.

To give a simple illustration of this, let us look at the function class $\mathcal{F} = \mathcal{F}_{[1,100]\to[1,100]}$. As in [2], we choose the reference point R such that it does not disturb the approximation. According to Corollary 4.7, we can choose $R = (-10, -10)$. Theorem 3.1 and Corollary 3.2 give that over all such functions $f \in \mathcal{F}$ the best approximation factor for n points is $\alpha_{\mathrm{OPT}} = 100^{1/n} \approx 1 + 4.6/n$. On the other hand, Corollary 4.7 shows that the set maximizing the hypervolume indicator gives an approximation ratio of $\alpha_{\mathrm{HYP}} \leqslant 1 + 20/(n-4)$. These two bounds are shown in Figure 4. It can be seen that for small numbers of points ($n \leqslant 10$), also the optimal approximation factor can be large while for large solution sets (say, $n \geqslant 100$), the multiplicative approximation achieved by the set maximizing the hypervolume is small, namely $\alpha_{\mathrm{HYP}} \leqslant 1.21$. This is only slightly greater than $\alpha_{\mathrm{OPT}} \approx 1.05$ for $n = 100$.

Open Problems

While we have proven the first result showing that maximizing the hypervolume indicator indeed yields an approximation of the Pareto front, this also raises a couple of pressing new questions. In this paper we looked only at maximization problems. Similar results for minimization and mixed problems seem possible with the same techniques and would be desirable.

As evolutionary multiobjective optimization is used not only with two objectives, it would further be of great interest to examine how higher dimensional analogs of α_{OPT} and α_{HYP} behave and prove bounds for both.

Moreover, it also seems interesting to determine the approximation factors of other indicators as this value provides a rigorous way to compare indicators. Ultimately we are looking for an efficiently computable indicator which has similar favorable properties as the hypervolume indicator, but maybe yields an even better approximation factor.

References

[1] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Investigating and exploiting the bias of the weighted hypervolume to articulate user preferences. In *Proc. 11th annual Conference on Genetic and Evolutionary Computation (GECCO '09)*, pp. 563–570, 2009.

[2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: optimal μ-distributions and the choice of the reference point. In *Proc. 10th International Workshop on Foundations of Genetic Algorithms (FOGA '09)*, pp. 87–102, 2009.

[3] J. Bader and E. Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 2010+. To appear.

[4] N. Beume. S-Metric calculation by considering dominated hypervolume as Klee's measure problem. *Evolutionary Computation*, 17:477–492, 2009.

[5] N. Beume and G. Rudolph. Faster S-metric calculation by considering dominated hypervolume as Klee's measure problem. In *Proc. Second International Conference on Computational Intelligence (IASTED '06)*, pp. 233–238, 2006.

[6] 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.

[7] K. Bringmann and T. Friedrich. Approximating the volume of unions and intersections of high-dimensional geometric objects. In *Proc. 19th International Symposium on Algorithms and Computation (ISAAC '08)*, Vol. 5369 of *LNCS*, pp. 436–447, 2008.

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

[9] T. C. E. Cheng, A. Janiak, and M. Y. Kovalyov. Bicriterion single machine scheduling with resource dependent processing times. *SIAM J. on Optimization*, 8:617–630, 1998.

[10] C. Daskalakis, I. Diakonikolas, and M. Yannakakis. How good is the Chord algorithm? In *Proc. 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '10)*, pp. 978–991, 2010.

[11] K. Deb, M. Mohan, and S. Mishra. Evaluating the ε-domination based multi-objective evolutionary algorithm for a quick computation of pareto-optimal solutions. *Evolutionary Computation*, 13:501–525, 2005.

[12] I. Diakonikolas and M. Yannakakis. Small approximate pareto sets for biobjective shortest paths and other problems. *SIAM Journal on Computing*, 39:1340–1371, 2009.

[13] M. T. M. Emmerich, N. Beume, and B. Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. In *Proc. Third International Conference on Evolutionary Multi-Criterion Optimization (EMO '05)*, pp. 62–76, 2005.

[14] T. Friedrich, C. Horoba, and F. Neumann. Multiplicative approximations and the hypervolume indicator. In *Proc. 11th annual Conference on Genetic and Evolutionary Computation (GECCO '09)*, pp. 571–578, 2009.

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

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

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

[18] R. Kumar and N. Banerjee. Running time analysis of a multiobjective evolutionary algorithm on simple and hard problems. In *Proc. 8th International Workshop on Foundations of Genetic Algorithms (FOGA '05)*, Vol. 3469 of *Lecture Notes in Computer Science*, pp. 112–131. Springer, 2005.

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

[20] G. Lizarraga-Lizarraga, A. Hernandez-Aguirre, and S. Botello-Rionda. G-metric: an m-ary quality indicator for the evaluation of non-dominated sets. In *Proc. 10th annual Conference on Genetic and Evolutionary Computation (GECCO '08)*, pp. 665–672, 2008.

[21] B. Naujoks, N. Beume, and M. T. M. Emmerich. Multi-objective optimisation using S-metric selection: application to three-dimensional solution spaces. In *Proc. IEEE Congress on Evolutionary Computation (CEC '05)*, pp. 1282–1289, 2005.

[22] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Proc. 41st annual Symposium on Foundations of Computer Science (FOCS '00)*, pp. 86–92, 2000.

[23] C. H. Papadimitriou and M. Yannakakis. Multiobjective query optimization. In *Proc. 20th ACM Symposium on Principles of Database Systems (PODS '01)*, pp. 52–59, 2001.

[24] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, 3rd edition, 1987.

[25] S. Vassilvitskii and M. Yannakakis. Efficiently computing succinct trade-off curves. *Theor. Comput. Sci.*, 348: 334–356, 2005.

[26] R. L. While, L. Bradstreet, L. Barone, and P. Hingston. Heuristics for optimizing the calculation of hypervolume for multi-objective optimization problems. In *Proc. IEEE Congress on Evolutionary Computation (CEC '05)*, pp. 2225–2232, 2005.

[27] X. Zhou, N. Mao, W. Li, and C. Sun. A fast algorithm for computing the contribution of a point to the hypervolume. In *Proc. Third International Conference on Natural Computation (ICNC '07)*, Vol. 4, pp. 415–420, 2007.

[28] E. Zitzler. Hypervolume metric calculation, 2001. Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, See ftp://ftp.tik.ee.ethz.ch/pub/people/zitzler/hypervol.c.

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

[30] 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.

[31] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans. Evolutionary Computation*, 7:117–132, 2003.

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