Multiplicative Approximations and the Hypervolume Indicator

Tobias Friedrich
Algorithms Group
International Computer Science Institute
Berkeley, CA, USA

Christian Horoba
Fakultät für Informatik, LS 2
Technische Universität Dortmund
Dortmund, Germany

Frank Neumann
Algorithms and Complexity Group
Max-Planck-Institut für Informatik
Saarbrücken, Germany

ABSTRACT

Indicator-based algorithms have become a very popular approach to solve multi-objective optimization problems. In this paper, we contribute to the theoretical understanding of algorithms maximizing the hypervolume for a given problem by distributing μ points on the Pareto front. We examine this common approach with respect to the achieved multiplicative approximation ratio for a given multi-objective problem and relate it to a set of μ points on the Pareto front that achieves the best possible approximation ratio. For the class of linear fronts and a class of concave fronts, we prove that the hypervolume gives the best possible approximation ratio. In addition, we examine Pareto fronts of different shapes by numerical calculations and show that the approximation computed by the hypervolume may differ from the optimal approximation ratio.

Categories and Subject Descriptors: F.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity

General Terms: Theory, Algorithms, Performance

Keywords: Approximation, evolutionary algorithms, hypervolume indicator, indicator-based algorithms, multi-objective optimization

1. INTRODUCTION

Multi-objective optimization [11] deals with the task of optimizing several objective functions at the same time. Usually, these functions are conflicting, which means that improvements with respect to one function can only be achieved when impairing the solution quality with respect to another objective function. Solutions that can not be improved with respect to any function without impairing another one are called *Pareto-optimal solutions*. The objective vectors associated with these solutions are called *Pareto-optimal objective vectors* and the set of all these objective vectors constitutes the *Pareto front*. Often the Pareto front grows exponentially with respect to the problem size or is even infinite in the continuous case. In this case, it is not possible to compute the whole Pareto front efficiently and the goal is to compute a good approximation.

Evolutionary algorithms [2] are in a natural way well-suited for dealing with multi-objective optimization problems as the population, which consists of a number of search points, can be evolved into an approximation of the Pareto front [8, 9].

Many researchers have worked on how to use evolutionary algorithms for multi-objective problems and how to achieve a good spread over the Pareto front. However, often the optimization goal remains rather unclear as it is not stated explicitly how to measure the quality of an approximation that a proposed algorithm should achieve.

One approach to achieve a good spread over the Pareto front is to use the hypervolume indicator [26] for measuring the quality of a population. This approach has gained increasing interest in recent years (see e. g. [4, 14, 15, 23]). The hypervolume indicator implicitly defines an optimization goal for the population of an evolutionary algorithm. Unfortunately, this optimization goal is rarely understood from a theoretical point of view. Recently, it has been shown in [1] that the slope of the front determines which objective vectors maximize the value of the hypervolume when dealing with continuous Pareto fronts. The aim of this paper is to further increase the theoretical understanding of the hypervolume indicator and examine its approximation behavior.

As multi-objective optimization problems often involve a vast number of Pareto-optimal objective vectors, multi-objective evolutionary algorithms use a population of fixed size and try to evolve the population into good approximation of the Pareto front. However, often it is not stated explicitly what a good approximation for a given problem is. One approach that allows a rigorous evaluation of the approximation quality is to measure the quality of a solution set with respect to its approximation ratio [20]. We follow this approach and examine the approximation ratio of a population with respect to all objective vectors of the Pareto front.

Our aim is to examine whether a given solution set of μ search points maximizing the hypervolume gives a good approximation measured with respect to the approximation ratio. We point out situations where the hypervolume provably leads to the best approximation ratio achievable by choosing μ Pareto-optimal solutions. Later on, we carry out numerical investigations to see how the shape of the Pareto

GECCO'09, July 8–12, 2009, Montréal Québec, Canada.

front influences the approximation behavior of the hypervolume indicator and point out where the approximation given by the hypervolume differs from the best one achievable by a solution set of μ points.

The outline of the paper is as follows. In Section 2, we introduce the hypervolume indicator and our notation of approximations. Section 3 gives analytic results for the approximation achievable by the hypervolume indicator and Section 4 provides further numerical investigations. Finally, we finish with some concluding remarks.

2. PRELIMINARIES

In this paper, we consider bi-objective maximization problems $P: \mathcal{S} \to \mathbb{R}^2$ for an arbitrary decision space $\mathcal{S}$. We are interested in the so-called Pareto front of P, which consists of all maximal elements of $P(\mathcal{S})$ with respect to the weak Pareto dominance relation. We restrict ourselves to problems with a Pareto front that can be written as $\{(x, f(x)) \mid x \in [x_{\min}, x_{\max}]\}$ where $f: [x_{\min}, x_{\max}] \to \mathbb{R}$ is a continuous, differentiable, and strictly monotonically decreasing function. This allows us to denote with f not only the actual function $f: [x_{\min}, x_{\max}] \to \mathbb{R}$, but also the front $\{(x, f(x)) \mid x \in [x_{\min}, x_{\max}]\}$ itself. We assume further that $x_{\min} > 0$ and $f(x_{\max}) > 0$ hold.

We intend to find a solution set $X = \{x_1, x_2, \ldots, x_\mu\}$ of μ Pareto-optimal search points $(x_i, f(x_i))$ that constitutes a good approximation of the front f.

2.1 Hypervolume indicator

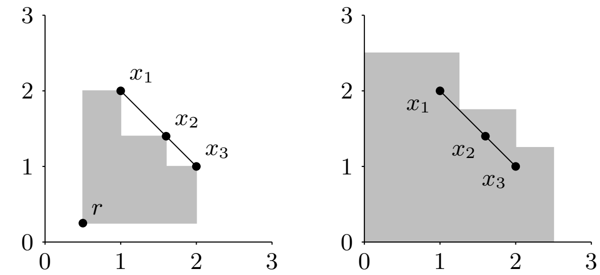
There are various indicators to measure the quality of a solution set, but there is only one widely used indicator that is strictly Pareto-compliant [27], namely the hypervolume indicator. Strictly Pareto-compliant means that given two solution sets A and B the indicator values A higher than B if the solution set A dominates the solution set B. The hypervolume (HYP) measures the volume of the dominated portion of the objective space. The hypervolume was first introduced for performance assessment in multi-objective optimization by Zitzler and Thiele [26]. Later on it was used to guide the search in various hypervolume-based evolutionary optimizers [4, 12, 14, 16, 23, 24].

Geometrically speaking, the hypervolume indicator measures the volume of the dominated space of all solutions contained in a solution set $X \subseteq \mathbb{R}^d$. This space is truncated at a fixed footpoint called the *reference point* $r = (r_1, r_2, \ldots, r_d)$. The *hypervolume* $\mathrm{HYP}(X)$ *of a solution set* X is then defined as

$$\mathrm{HYP}(X) := \textsc{vol}\left(\bigcup_{(x_1,\ldots,x_d)\in X} [r_1, x_1] \times \cdots \times [r_d, x_d]\right)$$

with $\textsc{vol}(\cdot)$ being the usual Lebesgue measure (see Figure 1(a) for an illustration).

It has become very popular recently and several algorithms have been developed to calculate it. The first one was the Hypervolume by Slicing Objectives (HSO) algorithm, which was suggested independently by Zitzler [22] and Knowles [17]. The currently best asymptotic runtime of $O(n \log n + n^{d/2})$ is obtained by Beume and Rudolph [5]. Bringmann and Friedrich [6] proved recently that it is **#P**-hard[1] in the number of dimensions to calculate HYP precisely. Therefore, all hypervolume algorithms must have a superpolynomial runtime in the number of objectives (unless **P** = **NP**). As the **#P**-hardness of HYP dashes the hope for an exact polynomial algorithm, there are polynomial estimation algorithms [3, 6, 7] for approximating the hypervolume based on Monte Carlo sampling.



(a) Hypervolume. (b) Approximation ratio.

Figure 1. Point distribution $X = \{1, 1.6, 2\}$ for the linear front $f: [1, 2] \to [1, 2]$ with $f(x) = 3 - x$, which achieves a hypervolume of $\mathrm{HYP}(X) = 1.865$ with respect to the reference point $r = (0.5, 0.25)$ and an approximation ratio of $\mathrm{APP}(X) = 1.25$. The shaded areas show the dominated portion of the objective space and the approximated portion of the objective space, respectively.

2.2 Approximations

In the following, we define our notion of approximation in a formal way.

Let $X = \{x_1, \ldots, x_\mu\}$ be a solution set and f a function that describes the Pareto front. The *approximation ratio* $\mathrm{APP}(X)$ *of a solution set* X with respect to f is defined according to [20] as follows

Definition 1. *Let* $f: [x_{\min}, x_{\max}] \to \mathbb{R}$ *and* $X = \{x_1, x_2, \ldots, x_\mu\}$. *The solution set* X *is an* δ-approximation of f *iff for each* $x \in [x_{\min}, x_{\max}]$ *there is a* $x_i \in X$ *with*

$$x \le \delta \cdot x_i \quad \text{and} \quad f(x) \le \delta \cdot f(x_i)$$

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

$$\mathrm{APP}(X) := \min\{\delta \in \mathbb{R} \mid X \text{ is a } \delta\text{-approximation of } f\}.$$

Figure 1(b) shows the area of the objective space that a certain solution set X δ-approximates for $\delta = 1.25$. Note that this area covers the entire Pareto front f. Since the objective vector $(1.25, 1.75)$ is not δ-approximated for all $\delta < 1.25$, the approximation ratio of X is 1.25.

Our definition of approximation is similar to the definition of multiplicative ε-dominance given in [18]. In this paper, an algorithmic framework for discrete multi-objective optimization is proposed which converges to a $(1 + \varepsilon)$-approximation of the Pareto front.

2.3 Our goal

The goal of this paper is to relate the above definition of approximation to the optimization goal implicitly defined by the hypervolume indicator. We want to find a distribution

[1] **#P** is the analog of **NP** for counting problems [21].

of the points such that the whole front is captured. Using the hypervolume the choice of the reference point decides which parts of the front are covered. All the functions that we consider in this paper have positive and bounded domains and codomains. Hence, choosing the reference point $r = (r_1, r_2)$ for appropriate $r_1, r_2 \leq 0$ ensures that the points $x_{\min}$ and $x_{\max}$ are contained in an optimal hypervolume distribution. A detailed calculation on how to choose the reference point such that $x_{\min}$ and $x_{\max}$ are contained in an optimal hypervolume distribution is given in [1]. To allow a fair comparison, we also require that the set minimizing the approximation ratio contains $x_{\min}$ and $x_{\max}$.

Consider a Pareto front f. There is an infinite number of possible solution sets of fixed size μ. To make this more formal, let $\mathcal{X}(\mu, f)$ be the set of all subsets of

$$\{(x, f(x)) \mid x \in [x_{\min}, x_{\max}]\} \cup \{(x_{\min}, f(x_{\min})), (x_{\max}, f(x_{\max}))\}$$

of cardinality μ. We want to compare two specific solution sets from $\mathcal{X}$ called *optimal hypervolume distribution* and *optimal approximation distribution* defined as follows.

Definition 2. *The optimal hypervolume distribution*

$$X_{\text{opt}}^{\text{HYP}}(\mu, f) := \underset{X \in \mathcal{X}(\mu, f)}{\operatorname{argmax}} \text{HYP}(X)$$

consists of μ points that maximize the hypervolume with respect to f. The optimal approximation distribution

$$X_{\text{opt}}^{\text{APP}}(\mu, f) := \underset{X \in \mathcal{X}(\mu, f)}{\operatorname{argmin}} \text{APP}(X)$$

consists of μ points that minimize the approximation ratio with respect to f. For brevity, we will in Figures 6–8 also use $X_{\text{opt}}(\mu, f)$ as a short form to refer to both sets $X_{\text{opt}}^{\text{HYP}}(\mu, f)$ and $X_{\text{opt}}^{\text{APP}}(\mu, f)$.

3. ANALYTIC RESULTS

We want to investigate the approximation ratio obtained by a solution set maximizing the hypervolume indicator in comparison to an optimal one. For this, we first examine conditions for an optimal approximation distribution $X_{\text{opt}}^{\text{APP}}(\mu, f)$. Later on, we consider two classes of functions f on which the optimal hypervolume distribution $X_{\text{opt}}^{\text{HYP}}(\mu, f)$ is equivalent to the optimal approximation distribution $X_{\text{opt}}^{\text{APP}}(\mu, f)$ and therefore provably leads to the best achievable approximation ratio.

3.1 Optimal approximations

We now consider the optimal approximation ratio that can be achieved placing μ points on the Pareto front given by the function f. The following lemma states a condition which allows to check whether a given set consisting of μ points achieves an optimal approximation ratio for a given function f.

Lemma 1. *Let $f\colon [x_{\min}, x_{\max}] \to \mathbb{R}$ be a Pareto front and $X = \{x_1, \ldots, x_\mu\}$ be an arbitrary solution set with $x_1 = x_{\min}$, $x_\mu = x_{\max}$, and $x_i \leq x_{i+1}$ for all $1 \leq i < \mu$. If there is a constant $\delta > 1$ and a set $Y = \{y_1, \ldots, y_{\mu-1}\}$ with $x_i \leq y_i \leq x_{i+1}$ and $\delta = \frac{y_i}{x_i} = \frac{f(y_i)}{f(x_{i+1})}$ for all $1 \leq i < \mu$, then $X = X_{\text{opt}}^{\text{APP}}(\mu, f)$ is the optimal approximation distribution with approximation ratio δ.*

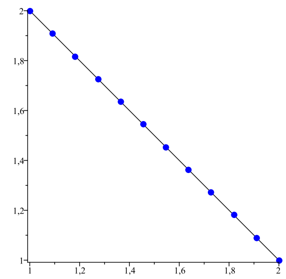


Figure 2. Optimal point distribution $X_{\text{opt}}^{\text{HYP}}(12, f) = X_{\text{opt}}^{\text{APP}}(12, f)$ for the linear front $f\colon [1, 2] \to [1, 2]$ with $f(x) = 3 - x$. The optimal hypervolume distribution and optimal approximation distribution are equivalent in this case.

Proof. We assume that a better approximation ratio than δ can be achieved by choosing a different set of solutions $X' = \{x'_1, \ldots, x'_\mu\}$ with $x'_1 = x_{\min}$, $x'_\mu = x_{\max}$, and $x'_i \leq x'_{i+1}$, $1 \leq i < \mu$, and show a contradiction.

Let y_i be the point for which a better approximation can be achieved. Getting a better approximation of y_i means that there is at least one point $x'_j \in X'$ with $x_i < x'_j < x_{i+1}$. We assume w. l. o. g. that $j \leq i$ and show that there is at least one point y with $y \leq y'_i$ that is not approximated by a factor of δ or that $x'_1 > x_{\min}$ holds. The case $j > i$ can be handled symmetrically, by showing that either $x'_n < x_{\max}$ or there is a point $y \geq y_{i+1}$ that is not approximated by a factor of δ.

To approximate all points y with $x_{i-1} \leq y \leq x_j$ by a factor of δ, the inequality $x_{i-1} < x'_{j-1}$ has to hold. Iterating the arguments in order to approximate all points in $x_{i-s} \leq y \leq x_{i-s+1}$, $x_{i-s} < x'_{j-s}$ has to hold. Considering $s = j - 1$ either one of the points y, $x_{i-j+1} \leq y \leq x_{i-j+2}$ is not approximated by a factor of δ or $x_{\min} = x_1 \leq x_{i-j+1} < x'_1$ holds, which contradicts one of our assumptions. □

We will use this lemma in the rest of the paper to check whether an approximation obtained by the hypervolume indicator is optimal as well as use these ideas to identify sets of points that achieve an optimal approximation ratio.

3.2 Linear fronts

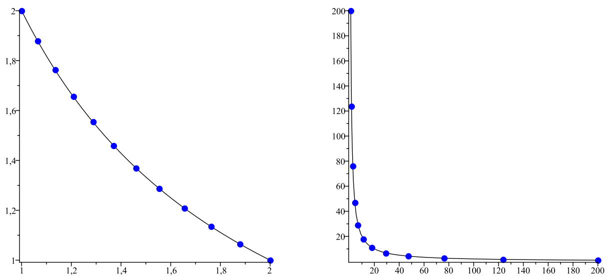
The distribution of points maximizing the hypervolume for linear fronts has already been investigated in [1]. Therefore, we start by considering the hypervolume indicator with respect to the approximation it achieves when the Pareto front is given by a linear function

$$f\colon [1, (1-d)/c] \to [1, c+d] \quad \text{with} \quad f(x) = c \cdot x + d$$

where $c < 0$ and $d > 1 - c$ are arbitrary constants.

Auger et al. [1] have shown that the maximum hypervolume of μ points on a linear front is reached when the points are distributed in an equally spaced manner. We assume that the reference point is chosen such that the extreme points of the Pareto front are included in the optimal distribution of the μ points on the Pareto front, that is, $x_1 = x_{\min} = 1$ and $x_\mu = x_{\max} = (1-d)/c$ hold. The maximal hypervolume is achieved by choosing

$$\begin{aligned} x_i &= x_{\min} + \frac{i-1}{\mu-1} \cdot (x_{\max} - x_{\min}) \\ &= 1 + \frac{i-1}{\mu-1} \cdot \left(\frac{1-d}{c} - 1\right) \end{aligned} \tag{1}$$



(a) $c = 2$. (b) $c = 200$.

Figure 3. Optimal point distribution $X^{\text{HYP}}_{\text{opt}}(12, f) = X^{\text{APP}}_{\text{opt}}(12, f)$ for two concave fronts $f\colon [1, c] \to [1, c]$ with $f(x) = c/x$. The optimal hypervolume distribution and optimal approximation distribution are equivalent in this case.

due to Theorem 6 in [1].

In the following, we consider the approximation ratio that the hypervolume using μ points achieves. Let $x_i < \tilde{x} < x_{i+1}$ be a Pareto-optimal x-coordinate. The approximation given by x_i and x_{i+1} is

$$\min\left\{\frac{\tilde{x}}{x_i}, \frac{f(\tilde{x})}{f(x_{i+1})}\right\}.$$

As f is monotonically decreasing, the worst-case approximation is attained for a point $\tilde{x}$, $x_i < \tilde{x} < x_{i+1}$, if

$$\frac{\tilde{x}}{x_i} = \frac{f(\tilde{x})}{f(x_{i+1})} \tag{2}$$

holds. Resolving the linear equations (1) and (2), we get

$$\tilde{x} = \frac{d\left((d + c - 1)i - c\mu - d + 1\right)}{c\left((\mu - 2)d - c + 1\right)}.$$

The approximation ratio resolves to

$$\frac{\tilde{x}}{x_i} = \frac{f(\tilde{x})}{f(x_{i+1})} = \frac{d(\mu - 1)}{d(\mu - 2) - c + 1}.$$

Hence, the worst-case approximation is independent of the choice of i and the same for all intervals $[x_i, x_{i+1}]$ of the Pareto front.

Lemma 1 implies that the hypervolume achieves the best possible approximation ratio on the class of linear fronts. Figure 2 shows the optimal distribution for $f(x) = 3 - x$ and $\mu = 12$. In summary, we have shown the following theorem.

Theorem 1. *Let $f\colon [1, (1 - d)/c] \to [1, c + d]$ be a linear function $f(x) = c \cdot x + d$ where $c < 0$ and $d > 1 - c$ are arbitrary constants. Then*

$$X^{\text{HYP}}_{\text{opt}}(\mu, f) = X^{\text{APP}}_{\text{opt}}(\mu, f).$$

3.3 A class of concave fronts

We now consider the distribution of μ points on a concave front maximizing the hypervolume. In contrast to the class of linear functions where an optimal approximation can be achieved by distributing the μ points in an equally spaced manner along the front, the class of functions considered in this section requires that the points are distributed exponentially to obtain an optimal approximation.

As already argued we want to make sure that optimal hypervolume distribute includes $x_{\min}$ and $x_{\max}$. For the class of concave fronts that we consider, this can be achieved by choosing the reference point $r = (0, 0)$.

The hypervolume of a set of points $X = \{x_1, \ldots, x_\mu\}$, where w. l. o. g. $x_1 \le x_2 \le \cdots \le x_\mu$, is then given by

$$\begin{aligned}
\text{HYP}(X) &= x_1 \cdot f(x_1) + x_2 \cdot f(x_2) - x_1 \cdot f(x_2) \\
&\quad + \cdots + x_\mu \cdot f(x_\mu) - x_{\mu-1} \cdot f(x_\mu) \\
&= x_1 \cdot f(x_1) + x_2 \cdot f(x_2) + \cdots + x_\mu \cdot f(x_\mu) \\
&\quad - (x_1 \cdot f(x_2) + \cdots + x_{\mu-1} \cdot f(x_\mu)).
\end{aligned}$$

We consider a Pareto front given by the function

$$f\colon [1, c] \to [1, c] \quad \text{and} \quad f(x) = c/x$$

where $c > 1$ is an arbitrary constant. Then we get

$$\text{HYP}(X) = c \cdot \mu - c \cdot \left(\frac{x_1}{x_2} + \frac{x_2}{x_3} + \cdots + \frac{x_{\mu-2}}{x_{\mu-1}} + \frac{x_{\mu-1}}{x_\mu}\right).$$

Hence, finding μ points that minimize

$$h(x_1, \ldots, x_\mu) := \left(\frac{x_1}{x_2} + \cdots + \frac{x_{\mu-1}}{x_\mu}\right)$$

maximizes the hypervolume. Setting $x_1 = 1$ and $x_\mu = c$ minimizes h, since x_1 and x_μ occur just in the first and last term of h, respectively. We consider the gradient vector given by the partial derivatives

$$\begin{aligned}
&h'(x_1, \ldots, x_\mu) \\
&\quad = \left(\frac{1}{x_2}, -\frac{x_1}{x_2^2} + \frac{1}{x_3}, \ldots, -\frac{x_{\mu-2}}{x_{\mu-1}^2} + \frac{1}{x_\mu}, -\frac{x_{\mu-1}}{x_\mu^2}\right).
\end{aligned}$$

This implies that h can be minimized by setting

$$\begin{aligned}
x_3 &= x_2^2/x_1 &&= x_2^2, \\
x_4 &= x_3^2/x_2 &&= x_2^3, \\
&\vdots && \vdots \\
x_\mu &= x_{\mu-1}^2/x_{\mu-2} &&= x_2^{\mu-1}.
\end{aligned}$$

From the last equation we get

$$\begin{aligned}
x_2 &= x_\mu^{1/(\mu-1)} &&= c^{1/(\mu-1)}, \\
x_3 &= x_2^2 &&= c^{2/(\mu-1)}, \\
&\vdots && \vdots \\
x_{\mu-1} &= x_2^{\mu-2} &&= c^{(\mu-2)/(\mu-1)}.
\end{aligned}$$

As f is monotonically decreasing, the worst-case approximation is attained for a point x, $x_i < x < x_{i+1}$, if

$$\frac{x}{x_i} = \frac{f(x)}{f(x_{i+1})}$$

holds. Substituting the coordinates and function values, we get

$$\frac{x}{x_i} = \frac{x}{c^{(i-1)/(\mu-1)}} \quad \text{and} \quad \frac{f(x)}{f(x_{i+1})} = \frac{c/x}{c/c^{i/(\mu-1)}} = \frac{c^{i/(\mu-1)}}{x}.$$

Therefore,

$$x^2 = c^{i/(\mu-1)} \cdot c^{(i-1)/(\mu-1)} = c^{(2i-1)/(\mu-1)},$$

which implies

$$x = c^{(2i-1)/(2\mu-2)}.$$

Hence, the set of search points maximizing the hypervolume achieves an approximation ratio of

$$\frac{c^{(2i-1)/(2\mu-2)}}{c^{(i-1)/(\mu-1)}} = c^{1/(2\mu-2)}.$$

We have seen that the requirements of Lemma 1 are fulfilled. Hence, an application of Lemma 1 shows that the hypervolume indicator achieves an optimal approximation ratio when the Pareto front is given by $f\colon [1,c] \to [1,c]$ with $f(x) = c/x$ where $c \in \mathbb{R}_{>1}$ is any constant. Figure 3 shows the optimal distribution for $\mu = 12$ and $c = 2$ as well as $c = 200$. In summary, we have shown the following theorem.

THEOREM 2. *Let $f\colon [1,c] \to [1,c]$ be a concave front with $f(x) = c/x$ where $c > 1$ is an arbitrary constant. Then*

$$X_{\text{opt}}^{\text{HYP}}(\mu, f) = X_{\text{opt}}^{\text{APP}}(\mu, f).$$

4. EMPIRICAL EVALUATION

The analysis of the distribution of an optimal set of search points tends to be hard or is impossible for more complex functions. Hence, resorting to numerical analysis methods constitutes a possible escape from this dilemma. This section is dedicated to the numerical analysis of a larger class of functions.

4.1 Fronts under investigation

For this, we examine a family of fronts of the shape x^p for $p > 0$. To allow for a proper scaling in both dimension we consider $f_p\colon [x_1, x_\mu] \to [y_\mu, y_1]$ with

$$f_p(x) := y_\mu - (y_\mu - y_1) \cdot \left(1 - \left(\frac{x - x_1}{x_\mu - x_1}\right)^p\right)^{1/p}.$$

We use the notation $y_i = f(x_i)$ for the function value $f(x_i)$ of a point x_i. As we assume the reference point to be sufficiently negative, the leftmost point (x_1, y_1) and the rightmost point (x_μ, y_μ) are always contained in the optimal hypervolume distribution as well as in the optimal approximation. For $p = 1$ the considered function corresponds to the test function DTLZ1 [10]. For $p = 2$ the shape of the front corresponds to DTLZ2, DTLZ3, and DTLZ4. We will mainly concentrate on two parameter sets of f_p, that is,

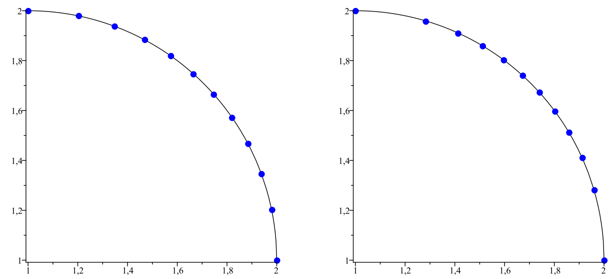
- the *symmetric* front $f_p^{\text{sym}}\colon [1,2] \to [1,2]$ and
- the *asymmetric* front $f_p^{\text{asy}}\colon [1,201] \to [1,2]$.

4.2 Our approach

We calculate for different functions f_p and $\mu \geq 3$

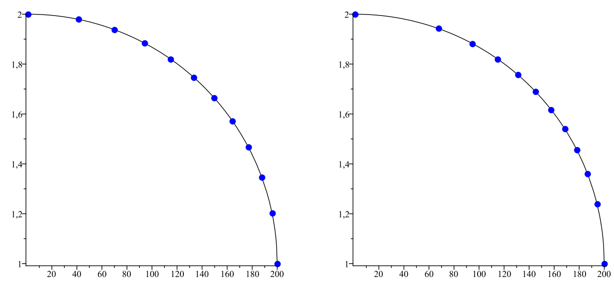
- the set of μ points $X_{\text{opt}}^{\text{HYP}}(\mu, f_p)$ which maximizes the dominated hypervolume, and
- the set of μ points $X_{\text{opt}}^{\text{APP}}(\mu, f_p)$ which minimizes the multiplicative approximation ratio.

In the first case, it suffices to find the $x_2, x_3, \ldots, x_{\mu-1}$ that maximize the dominated hypervolume, that is, the solutions



(a) Set $X_{\text{opt}}^{\text{HYP}}(12, f_2^{\text{sym}})$. (b) Set $X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{sym}})$.

Figure 4. Optimal point distributions for symmetric front f_2^{sym}. Note that the optimal hypervolume distribution and the optimal approximation distribution differ in this case. The set of points maximizing the hypervolume yields an approximation ratio of $\text{APP}(X_{\text{opt}}^{\text{HYP}}(12, f_2^{\text{sym}})) \approx 1.025$, which is 0.457% larger than the optimal approximation ratio $\text{APP}(X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{sym}})) \approx 1.021$.



(a) Set $X_{\text{opt}}^{\text{HYP}}(12, f_2^{\text{asy}})$. (b) Set $X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{asy}})$.

Figure 5. Optimal point distributions for asymmetric front f_2^{asy}. Note that the optimal hypervolume distribution and the optimal approximation distribution differ in this case. The set of points maximizing the hypervolume yields an approximation ratio of $\text{APP}(X_{\text{opt}}^{\text{HYP}}(12, f_2^{\text{asy}})) \approx 1.038$, which is 0.839% larger than the optimal approximation ratio $\text{APP}(X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{asy}})) \approx 1.030$.

of

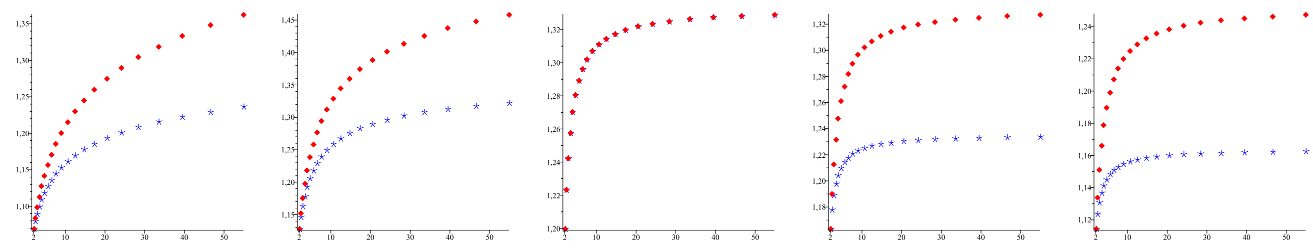
$$\operatorname*{argmax}_{x_2,\ldots,x_{\mu-1}} \left((x_2 - x_1) \cdot (f(x_2) - f(x_\mu)) + \sum_{i=3}^{\mu-1} (x_i - x_{i-1}) \cdot (f(x_i) - f(x_\mu)) \right)$$

We solve the arising nonlinear continuous optimization problem numerically by means of sequential quadratic programming [13].

In the second case, we have to solve the following system of nonlinear equations

$$\frac{z_1}{x_1} = \frac{z_2}{x_2} = \cdots = \frac{z_{\mu-1}}{x_{\mu-1}} = = \frac{f(z_1)}{f(x_2)} = \frac{f(z_2)}{f(x_3)} = \cdots = \frac{f(z_{\mu-1})}{f(x_\mu)}$$

with auxiliary variables $z_1, \ldots, z_{\mu-1}$ due to Lemma 1. The numerical solution of this system of equations can be determined easily by any standard computer algebra system.



(a) $\text{APP}(X_{\text{opt}}(3, f_{1/3}))$. (b) $\text{APP}(X_{\text{opt}}(3, f_{1/2}))$. (c) $\text{APP}(X_{\text{opt}}(3, f_1))$. (d) $\text{APP}(X_{\text{opt}}(3, f_2))$. (e) $\text{APP}(X_{\text{opt}}(3, f_3))$.

Figure 6. Approximation ratio of the optimal hypervolume distribution (◆) and the optimal approximation distribution (⋆) *depending on the scaling* x_μ of the fronts f_p (cf. Definition 2). Note that as analytically predicted in Theorem 1, both curves coincide in (c) for the linear function f_1 independent of the scaling.

4.3 Discussing the results

In the following, we present the results that have been obtained by our numerical investigations. We first examine the case of f_2. Figures 4 and 5 show different point distributions for f_2. It can be observed that the hypervolume distribution differs from the optimal distribution. Figures 4(a) and 4(b) show the distributions for the symmetric front

$$f_2(x) = 1 + \sqrt{1 - (x-1)^2}$$

with $(x_1, y_1) = (1, 2)$ and $(x_\mu, y_\mu) = (2, 1)$. Figures 5(a) and 5(b) show the asymmetric front

$$f_2(x) = 1 + \sqrt{1 - (x/200 - 1/200)^2}$$

with $(x_1, y_1) = (1, 2)$ and $(x_\mu, y_\mu) = (201, 1)$.

It can be observed that the relative positions of the hypervolume points stay the same in Figures 4(a) and 5(a) while the relative positions achieving an optimal approximation change with scaling (cf. Figures 4(b) and 5(b)). Hence, the relative position of the points maximizing the hypervolume is robust with respect to scaling. But as the optimal point distribution for a multiplicative approximation is dependent on the scaling, the hypervolume cannot achieve the best possible approximation quality.

In the example of Figures 4 and 5 the optimal multiplicative approximation factor for the symmetric and asymmetric case is 1.021 (Figure 4(b)) and 1.030 (Figure 5(b)), respectively, while the hypervolume only achieves an approximation of 1.025 (Figure 4(a)) and 1.038 (Figure 5(a)), respectively. Therefore in the symmetric and asymmetric case of f_2 the hypervolume is not calculating the set of points with the optimal multiplicative approximation.

We have already seen that scaling the function has a high impact on the optimal approximation distribution but not on the optimal hypervolume distribution. We want to investigate this effect in greater detail. The influence of scaling the parameter $x_\mu \geq 2$ of different functions $f_p \colon [1, x_\mu] \to [1, 2]$ is depicted in Figure 6 for $p = 1/3, 1/2, 1, 2, 3$. For fixed $\mu = 3$ it shows the achieved approximation ratio. As expected, the larger the asymmetry (x_μ) the larger the approximation ratios. For convex fronts ($p > 1$) the approximation ratios seem to converge quickly for large enough x_μ. The approximation of f_2 tends towards the golden ratio $\sqrt{5} - 1 \approx 1.236$ for the optimal approximation and $4/3 \approx 1.333$ for the optimal hypervolume. For f_3 they tend towards 1.164 and 1.253, respectively. Hence, for f_2 and f_3 the hypervolume is never more than 8% worse than the optimal approximation. This is different for the concave fronts ($p < 1$). There, the ratio between the hypervolume and the optimal approximation appears divergent.

Another important question is how the choice of the population size influences the relation between an optimal approximation and the approximation achieved by an optimal hypervolume distribution. We investigate the influence of the choice of μ on the approximation behavior in greater detail. Figure 7 shows the achieved approximation ratios depending on the number of points μ. For symmetric f_p's with $(x_1, y_1) = (y_\mu, x_\mu)$ and $\mu = 3$ the hypervolume achieves an optimal approximation distribution for all $p > 0$. The same holds for the linear function f_1 independent of the scaling implied by (x_1, y_1) and (y_μ, x_μ).

For larger populations, the approximation ratio of the hypervolume distribution and the optimal distribution decreases. However, the performance of the hypervolume measure is especially poor even for larger μ for concave asymmetric fronts, that is, f_p^{asy} with $p < 1$ (e.g. Figures 7(f) and 7(g)). Our investigations show that the approximation of an optimal hypervolume distribution may differ significantly from an optimal one depending on the choice of p. An important issue is whether the front is convex or concave [19]. The hypervolume was thought to prefer convex regions to concave regions [25] while [1] showed that the density of points only depends on the slope of the front and not on convexity or concavity. To illuminate the impact of convex vs. concave further, Figure 8 shows the approximation ratios depending on p. As expected, for $p = 1$ the hypervolume calculates the optimal approximation. However, the influence of the p is very different for the symmetric and the asymmetric test function. For f_p^{sym} the concave ($p < 1$) fronts are much better approximated by the hypervolume than the convex ($p > 1$) fronts (cf. Figure 8 (a)–(d)). For f_p^{asy} this is surprisingly the other way around (cf. Figure 8 (e)–(h)).

5. CONCLUSIONS

Using the hypervolume indicator to measure the quality of a population in an evolutionary multi-objective algorithm has become very popular in recent years. Understanding the optimal distribution of a population consisting of μ individuals is a hard task and the optimization goal when using the hypervolume indicator is rather unclear. Therefore, it is a challenging task to understand the optimization goal by using the hypervolume indicator as a quality measure for a population. We have examined how the hypervolume indicator approximates Pareto fronts of different shapes and related it to the best possible approximation ratio. Considering linear fronts and a class of concave fronts we have pointed out that the hypervolume indicator gives provably the best multiplicative approximation ratio that is achievable. Further, numerical investigation points out that the shape as well the scaling of the objectives heavily influences the approximation behavior of the hypervolume indicator. Examining fronts with differrent shapes we have shown that the approximation achieved by an optimal set of points with respect to the hypervolume may differ from the set of μ points achieving the best approximation ratio.

References

[1] 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 *Foundations of Genetic Algorithms Workshop (FOGA)*. ACM Press, 2009. To appear, available from `http://www.tik.ee.ethz.ch/sop/publications/`.

[2] T. Bäck, D. B. Fogel, and Z. Michalewicz, editors. *Handbook of Evolutionary Computation*. Oxford University Press, 1997.

[3] J. Bader and E. Zitzler. Hype: An algorithm for fast hypervolume-based many-objective optimization. TIK Report 286, Computer Engineering and Networks Laboratory (TIK), ETH Zürich, 2008.

[4] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

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

[6] K. Bringmann and T. Friedrich. Approximating the volume of unions and intersections of high-dimensional geometric objects. In *International Symposium on Algorithms and Computation (ISAAC)*, volume 5369 of *Lecture Notes in Computer Science*, pages 436–447. Springer, 2008.

[7] K. Bringmann and T. Friedrich. Approximating the least hypervolume contributor: NP-hard in general, but fast in practice. In *International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, volume 5467 of *Lecture Notes in Computer Science*, pages 6–20. Springer, 2009.

[8] C. A. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer, 2nd edition, 2007.

[9] K. Deb. *Multi-objective Optimization Using Evolutionary Algorithms*. Wiley, 2001.

[10] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *IEEE Congress on Evolutionary Computation (CEC)*, pages 825–830. IEEE Press, 2002.

[11] M. Ehrgott. *Multicriteria Optimization*. Springer, 2nd edition, 2005.

[12] M. Emmerich, N. Beume, and B. Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. In *International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, pages 62–76. Springer, 2005.

[13] R. Fletcher. *Practical Methods of Optimization*. Wiley, 2nd edition, 2000.

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

[15] J. Knowles and D. Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Transactions on Evolutionary Computation*, 7(2):100–116, 2003.

[16] J. Knowles, D. Corne, and M. Fleischer. Bounded archiving using the Lebesgue measure. In *IEEE Congress on Evolutionary Computation (CEC)*, pages 2490–2497. IEEE Press, 2003.

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

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

[19] 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 *Genetic and Evolutionary Computation Conference (GECCO)*, pages 665–672. ACM Press, 2008.

[20] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Symposium on Foundations of Computer Science (FOCS)*, pages 86–92. IEEE Press, 2000.

[21] C. M. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

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

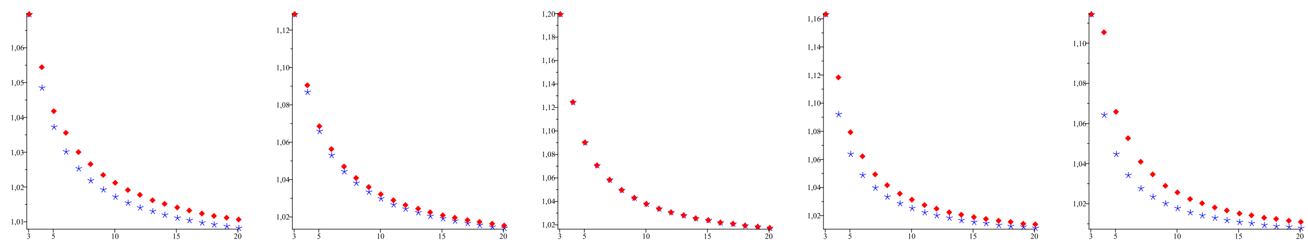
[23] E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In *International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, pages 862–876. Springer, 2007.

[24] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *International Conference on Parallel Problem Solving from Nature (PPSN)*, pages 832–842. Springer, 2004.

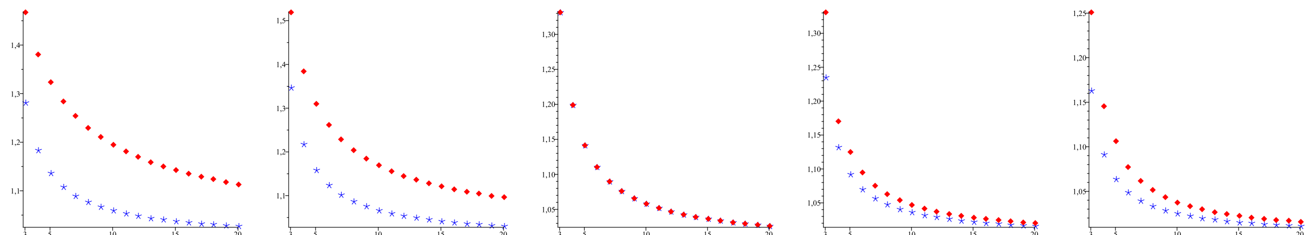
[25] E. Zitzler and L. Thiele. An evolutionary approach for multiobjective optimization: The strength Pareto approach. TIK Report 43, Computer Engineering and Networks Laboratory (TIK), ETH Zürich, 1998.

[26] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.

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

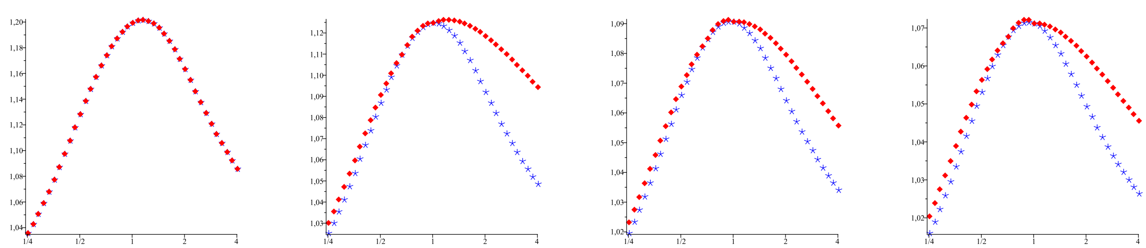


(a) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{1/3}^{\mathrm{sym}}))$. (b) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{1/2}^{\mathrm{sym}}))$. (c) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{1}^{\mathrm{sym}}))$. (d) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{2}^{\mathrm{sym}}))$. (e) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{3}^{\mathrm{sym}}))$.

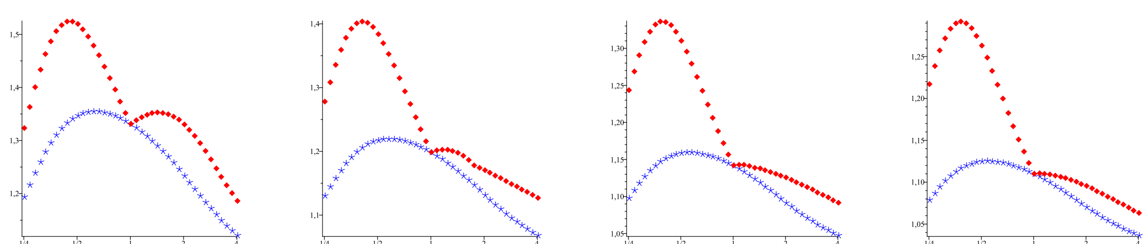


(f) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{1/3}^{\mathrm{asy}}))$. (g) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{1/2}^{\mathrm{asy}}))$. (h) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{1}^{\mathrm{asy}}))$. (i) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{2}^{\mathrm{asy}}))$. (j) $\mathrm{APP}(X_{\mathrm{opt}}(\mu, f_{3}^{\mathrm{asy}}))$.

Figure 7. Approximation ratio of the optimal hypervolume distribution (◆) and the optimal approximation distribution (⋆) *depending on the number of points* μ for symmetric and asymmetric fronts f_p and different parameters p (cf. Definition 2). Note that (c) and (h) show that the approximation ratio of the optimal hypervolume distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f_1^{\mathsf{sym}}))$ and the optimal approximation distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f_1^{\mathsf{sym}}))$ are equivalent for all examined μ. That maximizing the hypervolume yields the optimal approximation ratio can also be observed for all symmetric f_p^{sym} with $\mu = 3$ in (a)–(e).



(a) $\mathrm{APP}(X_{\mathrm{opt}}(3, f_p^{\mathrm{sym}}))$. (b) $\mathrm{APP}(X_{\mathrm{opt}}(4, f_p^{\mathrm{sym}}))$. (c) $\mathrm{APP}(X_{\mathrm{opt}}(5, f_p^{\mathrm{sym}}))$. (d) $\mathrm{APP}(X_{\mathrm{opt}}(6, f_p^{\mathrm{sym}}))$.



(e) $\mathrm{APP}(X_{\mathrm{opt}}(3, f_p^{\mathrm{asy}}))$. (f) $\mathrm{APP}(X_{\mathrm{opt}}(4, f_p^{\mathrm{asy}}))$. (g) $\mathrm{APP}(X_{\mathrm{opt}}(5, f_p^{\mathrm{asy}}))$. (h) $\mathrm{APP}(X_{\mathrm{opt}}(6, f_p^{\mathrm{asy}}))$.

Figure 8. Approximation ratio of the optimal hypervolume distribution (◆) and the optimal approximation distribution (⋆) *depending on the convexity/concavity parameter* p for symmetric and asymmetric fronts f_p and different population sizes μ (cf. Definition 2). The x-axis is scaled logarithmically. Note that (a) shows that the approximation ratio of the optimal hypervolume distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{HYP}}(3, f_p^{\mathsf{sym}}))$ and the optimal approximation distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{APP}}(3, f_p^{\mathsf{sym}}))$ are equivalent for all examined p.