Multiplicative Approximations, Optimal Hypervolume Distributions, and the Choice of the Reference Point

Tobias Friedrich   friedrich@uni-jena.de
Lehrstuhl Theoretische Informatik I, Fakultät für Mathematik und Informatik,
Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

Frank Neumann   frank.neumann@adelaide.edu.au
Optimisation and Logistics, School of Computer Science,
The University of Adelaide, Adelaide, SA 5005, Australia

Christian Thyssen   christian.thyssen@ls2.cs.tu-dortmund.de
Lehrstuhl 2, Fakultät für Informatik,
Technische Universität Dortmund, 44221 Dortmund, Germany

Abstract

Many optimization problems arising in applications have to consider several objective functions at the same time. Evolutionary algorithms seem to be a very natural choice for dealing with multi-objective problems as the population of such an algorithm can be used to represent the trade-offs with respect to the given objective functions. In this paper, we contribute to the theoretical understanding of evolutionary algorithms for multi-objective problems. We consider indicator-based algorithms whose goal is to maximize the hypervolume for a given problem by distributing μ points on the Pareto front. To gain new theoretical insights into the behavior of hypervolume-based algorithms we compare their optimization goal to the goal of achieving an optimal multiplicative approximation ratio. Our studies are carried out for different Pareto front shapes of bi-objective problems. For the class of linear fronts and a class of convex fronts, we prove that maximizing the hypervolume gives the best possible approximation ratio when assuming that the extreme points have to be included in both distributions of the points on the Pareto front. Furthermore, we investigate the choice of the reference point on the approximation behavior of hypervolume-based approaches and examine Pareto fronts of different shapes by numerical calculations.

Keywords

Approximations, Hypervolume Indicator, Multi-objective optimization, Theory.

1 Introduction

Multi-objective optimization [15] deals with the task of optimizing several objective functions at the same time. Here, several attributes of a given problem are employed as objective functions and are used to define a partial order, called preference order, on the solutions, for which the set of minimal (maximal) elements is sought. Usually,

the objective 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. Due to this, such problems usually do not have a single optimal function value. Instead, there is a set of optimal objective vectors which represents the different trade-offs of the different objective functions. Solutions that cannot 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*.

In contrast to single-objective optimization, in multi-objective optimization the task is not to compute a single optimal solution but a set of solutions representing the different trade-offs with respect to the given objective functions. Most of the best-known single-objective polynomially solvable problems like shortest path or minimum spanning tree become NP-hard when at least two weight functions have to be optimized at the same time. In this sense, multi-objective optimization is generally considered as more difficult than single-objective optimization.

Another, more promising, approach to deal with multi-objective optimization problems is to apply general stochastic search algorithms that evolve a set of possible solutions into a set of solutions that represent the trade-offs with respect to the objective functions. Well-known approaches in this field are evolutionary algorithms [2] and ant colony optimization [14]. Especially, multi-objective evolutionary algorithms (MOEAs) have been shown to be very successful when dealing with multi-objective problems [11, 12]. Evolutionary algorithms work with a set of solutions called population which is evolved over time by applying crossover and mutation operators to produce new possible solutions for the underlying multi-objective problem. Due to this population-based approach, they are in a natural way well-suited for dealing with multi-objective optimization problems.

A major problem when dealing with multi-objective optimization problems is that the number of different trade-offs may be too large. This implies that not *all* trade-offs can be computed efficiently, i. e., in polynomial time. In the discrete case the Pareto front may grow exponentially with respect to the problem size and may be even infinite in the continuous case. In such a case, it is not possible to compute the whole Pareto front efficiently and the goal is to compute a good approximation consisting of a not too large set of Pareto-optimal solutions. It has been observed empirically that MOEAs are able to obtain good approximations for a wide range of multi-objective optimization problems.

The aim of this paper is to contribute to the theoretical understanding of MOEAs in particular with respect to their approximation behavior. Many researchers have worked on how to use evolutionary algorithms for multi-objective optimization problems and how to find solutions being close to the Pareto front and covering all parts of 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 popular approach to achieve the mentioned objectives is to use the hypervolume indicator [33] for measuring the quality of a population. This approach has gained increasing interest in recent years (see e. g. [4, 22, 24, 35]). 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. Auger, Bader, Brockhoff, and Zitzler [1] have shown that the slope of the front determines which objective vectors maximize the value of the hypervolume when dealing with continuous Pareto fronts. Comparing hypervolume-optimal sets and best-approximation sets in the worst-case, Bringmann and Friedrich [9] observed that maximizing the hypervolume aligns well with an additive approximation while a good multiplicative approximation is achieved when the hypervolume of logarithmized axes is maximized [21]. Whether additive or multiplicative approximation is appropriate, depends on the meaning of an objective: Additive approximation is invariant to shifting the objective function; multiplicative approximation is invariant to scaling the objective function. While Bringmann and Friedrich [9] studied the worst-case behavior, we examine the properties of common specific Pareto fronts. The aim of this paper is to further increase the theoretical understanding of the hypervolume indicator and examine its multiplicative 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 a 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 [28]. We follow this approach and examine the approximation ratio of a population with respect to all objective vectors of the Pareto front.

The advantage of the approximation ratio is that it gives a meaningful scalar value which allows us to compare the quality of solutions between different functions, different population sizes, and even different dimensions. This is not the case for the hypervolume indicator. A specific dominated volume does not give a priori any information how well a front is approximated. Also, the hypervolume measures the space relative to an arbitrary reference point (cf. Section 2.1). This (often unwanted) freedom of choice not only changes the distribution of the points, but also makes the hypervolumes of different solutions measured relative to a (typically dynamically changing) reference point very hard to compare.

Our aim is to examine whether a given solution set of μ search points maximizing the hypervolume (called the optimal hypervolume distribution) gives a good approximation measured with respect to the approximation ratio. We do this by investigating two classes of objective functions having two objectives each and analyze the optimal distribution for the hypervolume indicator and the one achieving the optimal approximation ratio.

In a first step, we assume that both sets of μ points have to include both optimal points regarding the given two single objective functions. We point out situations where maximizing the hypervolume provably leads to the best approximation ratio achievable by choosing μ Pareto-optimal solutions. After these theoretical investigations, we carry out numerical investigations to see how the shape of the Pareto 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. These initial theoretical and experimental results investigating the correlation between the hypervolume indicator and multiplicative approximations have been published as a conference version in [20].

This paper extends its conference version in Section 4 to the case where the optimal hypervolume distribution is dependent on the chosen reference point. The reference point is a crucial parameter when applying hypervolume-based algorithms. It determines the area in the objective space where the algorithm focuses its search. As the hypervolume indicator itself, it is hard to understand the impact of the choice of the reference point. Different studies have been carried out on this topic and initial results on the optimal hypervolume distribution in the dependence of the reference point have been obtained in [1] and [10]. We provide new insights into how the choice of the reference point may affect the approximation behavior of hypervolume-based algorithms. In our studies, we relate the optimal hypervolume distribution with respect to a given reference to the optimal approximation ratio obtainable when having the freedom to choose the μ points arbitrarily.

The rest of the paper is structured 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 under the assumption that both extreme points have to be included in the two distributions and reports on our numerical investigations Pareto fronts having different shapes. In Section 4, we generalize our results and study the impact of the reference point on the optimal hypervolume distribution and relate this choice to the best possible overall approximation ratio when choosing μ points. Finally, we finish with some concluding remarks.

2 The Hypervolume Indicator and Multiplicative Approximations

In this paper, we consider bi-objective maximization problems $P\colon \mathcal{S} \to \mathbb{R}^2$ 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 $f\colon [x_{\min}, x_{\max}] \to \mathbb{R}$ is a continuous, differentiable, and strictly monotonically decreasing function. This allows us to denote with f not only the actual function $f\colon [x_{\min}, x_{\max}] \to \mathbb{R}$, 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 $X^* = \{x_1^*, x_2^*, \dots, x_\mu^*\}$ of μ Pareto-optimal search points $(x_i^*, f(x_i^*))$ that constitutes a good approximation of the front f.

2.1 Hypervolume indicator

The hypervolume (HYP) measures the volume of the dominated portion of the objective space. It was first introduced for performance assessment in multi-objective optimization by Zitzler and Thiele [33]. Later on it was used to guide the search in various hypervolume-based evolutionary optimizers [4, 16, 22, 25, 31, 35].

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* $r = (r_1, r_2, \dots, r_d)$. The *hypervolume* $\mathrm{HYP}_r(Y)$ *of a solution set* Y *in dependence of a given reference point* $r = (r_1, r_2, \dots, r_d)$ is then defined as

$$\mathrm{HYP}_r(Y) := \mathrm{VOL}\left(\bigcup_{(y_1, \dots, y_d) \in Y} [r_1, y_1] \times \cdots \times [r_d, y_d] \right)$$

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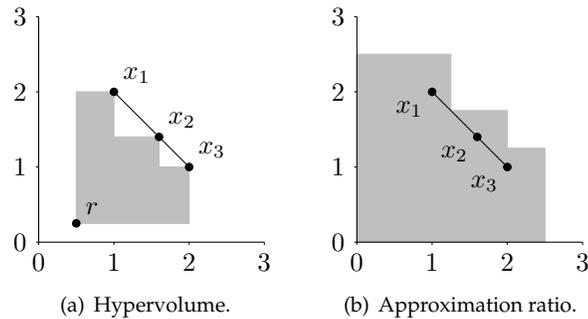
Figure 1: Point distribution $X = \{1, 1.6, 2\}$ for the linear front $f \colon [1,2] \to [1,2]$ with $f(x) = 3 - x$, which achieves a hypervolume of $\mathrm{HYP}(X) = 1.865$ with respect to the reference point $r = (0.5, 0.25)$ and an approximation ratio of $\mathrm{APP}(X) = 1.25$. The shaded areas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

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

The hypervolume indicator is a popular second-level sorting criterion in many recent multi-objective evolutionary algorithms for several reasons. Besides having a very intuitive interpretation, it is also the only common indicator that is strictly Pareto-compliant [34]. 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. It has further been shown by Bringmann and Friedrich [9] that the worst-case approximation factor of all possible Pareto fronts obtained by any hypervolume-optimal set of fixed size μ is asymptotically equal to the best worst-case approximation factor achievable by any set of size μ.

In the last years, the hypervolume has become very popular 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 [30] and Knowles [23]. For $d \leq 3$ it can be solved in (asymptotically optimal) time $\mathcal{O}(n \log n)$ [19]. The currently best asymptotic runtime for $d \in \{4,5,6\}$ is $\mathcal{O}(n^{(d-1)/2} \log n)$ [29]. The best known bound for large dimensions $d \geq 7$ is $\mathcal{O}(n^{(d+2)/3})$ [5].

On the other hand, Bringmann and Friedrich [6] proved that all hypervolume algorithms must have a superpolynomial runtime in the number of objectives (unless $\mathbf{P} = \mathbf{NP}$). Assuming the widely accepted exponential time hypothesis, the runtime must even be at least $n^{\Omega(d)}$ [8]. As this dashes the hope for fast and exact hypervolume algorithms, there are several estimation algorithms [3, 6, 7] for approximating the hypervolume based on Monte Carlo sampling.

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. We call a Pareto front convex if the function defining the Pareto front is a convex function. Otherwise, we call the Pareto front concave. Note that this differs from the notation

used in [20].

The *approximation ratio* APP(X) *of a solution set* X with respect to f is defined according to [28] as follows.

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

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

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

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

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 [26]. In this paper, an algorithmic framework for discrete multi-objective optimization is proposed which converges to a $(1 + \varepsilon)$-approximation of the Pareto front.

3   Results independent of the reference point

The goal of this paper is to relate the above definition of approximation to the optimization goal implicitly defined by the hypervolume indicator. Using the hypervolume, the choice of the reference point decides which parts of the front are covered. In this section we avoid the additional influence of the reference point by considering only solutions where both extreme points have to be included. The influence of the reference point is studied in Section 4.

All the functions that we consider in this paper have positive and bounded domains and codomains. Furthermore, the functions that are under consideration don't have infinite or zero derivative at the extremes. 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]. Assuming that $x_{\min}$ and $x_{\max}$ have to be included in the optimal hypervolume distribution, the value of the volume is in this section independent of the choice of the reference point. Therefore, we write HYP(X) instead of HYP$_r(X)$.

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

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

of cardinality μ which contain $(x_{\min}, f(x_{\min}))$ and $(x_{\max}, f(x_{\max}))$. 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*

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

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

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

consists of μ points that minimize the approximation ratio with respect to f. For brevity, we will also use $X_{\mathrm{opt}}(\mu, f)$ in Figures 5–7 as a short form to refer to both sets $X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f)$ and $X_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)$.

Note that "optimal hypervolume distributions" are also called "optimal μ-distributions" [1, 10] or "maximum hypervolume set" [9] in the literature.

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 $X_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)$. Later on, we consider two classes of functions f on which the optimal hypervolume distribution $X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f)$ is equivalent to the optimal approximation distribution $X_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)$ and therefore provably leads to the best achievable approximation ratio.
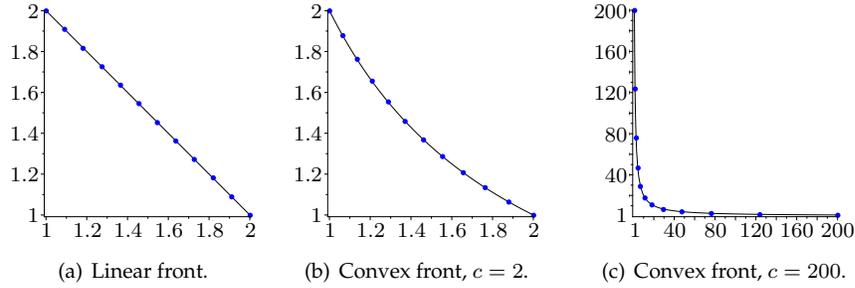
3.1 Optimal approximations

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 $f : [x_{\min}, x_{\max}] \to \mathbb{R}$ 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 $x_i \le x_{i+1}$ for all $1 \le i < \mu$. If there is a constant $\delta > 1$ and a set $= Z = \{z_1, \ldots, z_{\mu-1}\}$ with $x_i \le z_i \le x_{i+1}$ and $\delta = \frac{z_i}{x_i} = \frac{f(z_i)}{f(x_{i+1})}$ for all $1 \le i < \mu$, then $X = X_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)$ is the optimal approximation distribution with approximation ratio δ.*

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 $x'_i \le x'_{i+1}$, $1 \le i < \mu$, and show a contradiction.

The points z_i, $1 \le i \le \mu - 1$, are the points that are worst approximated by the set X. Each z_i is approximated by a factor of δ. Hence, in order to obtain a better approximation than the one achieved by the set X, the points z_i have to be approximated within a ratio of less than δ. We now assume that there is a point z_i for which a better approximation is achieved by the set X'. Getting a better approximation of z_i than δ means that there is at least one point $x'_j \in X'$ with $x_i < x'_j < x_{i+1}$ as otherwise z_i is approximated within a ratio of at least $\frac{z_i}{x_i} = \frac{f(z_i)}{f(x_{i+1})} = \delta$.

We assume w. l. o. g. that $j \le i + 1$ and show that there is at least one point z with $z \le z_i$ that is not approximated by a factor of δ or that $x'_1 > x_{\min}$ holds. To approximate all points z with $x_{i-1} \le z \le x'_j$ by a factor of δ, the inequality $x_{i-1} < x'_{j-1} < x_i$

Figure 2: Optimal point distribution $X_{\mathrm{opt}}^{\mathrm{HYP}}(12, f) = X_{\mathrm{opt}}^{\mathrm{APP}}(12, f)$ for (a) the linear front $f\colon [1,2] \to [1,2]$ with $f(x) = 3 - x$ and (b,c) the convex fronts $f\colon [1,c] \to [1,c]$ with $f(x) = c/x$. The respective optimal hypervolume distributions and optimal approximation distributions are equivalent in all three cases.

has to hold as otherwise z_{i-1} is approximated within a ratio of more than δ by X'. We iterate the arguments. In order to approximate all points in $x_{i-s} \leq y \leq x_{i-s+1}$, $x_{i-s} < x'_{j-s} < x_{i-s+1}$ has to hold as otherwise z_{i-s} is not approximated within a ratio of δ by X'. Considering $s = j - 1$ either one of the points z, $x_{i-j+1} \leq y \leq x_{i-j+2}$ is not approximated within a ratio of δ by X' or $x_{\min} = x_1 \leq x_{i-j+1} < x'_1$ holds, which contradicts the assumption that X' includes $x_{\min}$ and constitutes an approximation better than δ.

The case $j > i + 1$ can be handled symmetrically, by showing that either $x'_n < x_{\max}$ or there is a point $z \geq z_{i+1}$ that is not approximated within a ratio of δ by X'. This completes the proof. □

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 Analytic results for linear fronts

The distribution of points maximizing the hypervolume for linear fronts has already been investigated in [1, 17]. 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

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

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

Auger et al. [1] and Emmerich et al. [17] 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

$$x_i = x_{\min} + \frac{i-1}{\mu-1} \cdot (x_{\max} - x_{\min})$$

$$= 1 + \frac{i-1}{\mu-1} \cdot \left(\frac{1-d}{c} - 1\right) \tag{1}$$

due to Theorem 6 in [1].

The following theorem shows that the optimal approximation distribution coincides with the optimal hypervolume distribution.

Theorem 2 *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*

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

Proof. We determine the approximation ratio that the optimal hypervolume distribution $X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f) = \{x_1, \ldots, x_\mu\}$ using μ points achieves. Let $\tilde{x}$, $x_i < \tilde{x} < x_{i+1}$, 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.

Furthermore, 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.

Using

$$x_i = 1 + \frac{i-1}{\mu-1} \cdot \left(\frac{1-d}{c} - 1\right)$$

and resolving the Equation 2 with respect to $\tilde{x}$, we get

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

For the approximation ratio we get

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

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. $\square$

Figure 2 (a) shows the optimal distribution for $f(x) = 3 - x$ and $\mu = 12$.

3.3 Analytic results for a class of convex fronts

We now consider the distribution of μ points on a convex 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 distribution includes $x_{\min}$ and $x_{\max}$. For the class of convex 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. $x_1 \leq x_2 \leq \cdots \leq x_\mu$, is then given by

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

We consider a Pareto front given by the function

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

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

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

Hence, to maximize the hypervolume we have to find μ points minimizing

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

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. Furthermore, we have $1 = x_1 < x_2 < \ldots < x_\mu = c$ as the equality of two points implies that one of them can be exchanged for another unchosen point on the Pareto front and thereby increases the hypervolume.

We work under these assumptions and aim to find a set of points X that minimizes the function h. To do this, we consider the gradient vector given by the partial derivatives

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

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

$$
\begin{aligned}
x_2 &= x_\mu^{1/(\mu-1)} &= c^{1/(\mu-1)}, \\
x_3 &= x_2^2 &= c^{2/(\mu-1)}, \\
&\vdots & \vdots \\
x_{\mu-1} &= x_2^{\mu-2} &= c^{(\mu-2)/(\mu-1)}.
\end{aligned}
\tag{3}
$$

The following theorem shows that the optimal approximation distribution coincides with the optimal hypervolume distribution.

Theorem 3 *Let* $f\colon [1,c] \to [1,c]$ *be a convex front with* $f(x) = c/x$ *where* $c > 1$ *is an arbitrary constant. Then*

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

Proof. We determine the approximation ratio that the optimal hypervolume distribution $X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f) = \{x_1, \ldots, x_\mu\}$ using μ points achieves. 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

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

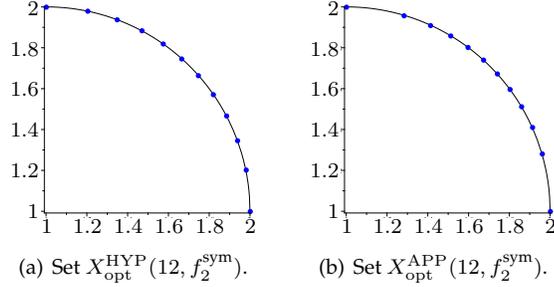Hence, the set of search points maximizing the hypervolume achieves an approximation ratio of

$$
\frac{c^{(2i-1)/(2\mu-2)}}{c^{(i-1)/(\mu-1)}} = c^{1/(2\mu-2)}.
\tag{4}
$$

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. $\square$

Figure 2 shows the optimal distribution for $\mu = 12$ and $c = 2$ as well as $c = 200$.

3.4 Numerical evaluation for fronts of different shapes

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.

(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 3: 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 $\text{APP}(X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{sym}})) \approx 1.021$.

Our goal is to study the optimal hypervolume distribution for different shapes of Pareto fronts and investigate how the shape of such a front influences the approximation behavior of the hypervolume indicator. We examine a family of fronts of the shape x^p where $p > 0$ is a parameter that determines the degree of the polynomial describing the Pareto front. Furthermore, we allow scaling in both dimensions.

The Pareto fronts that we consider can be defined by a function of the form $f_p \colon [x_1, x_\mu] \to [y_\mu, y_1]$ 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. 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 $f_p^{\text{asy}} \colon [1, 201] \to [1, 2]$.

Note, that choosing $p = 1$ corresponds to the well-known test function DTLZ1 [13]. For $p = 2$ the shape of the front corresponds to functions DTLZ2, DTLZ3, and DTLZ4.

Our goal is to study the optimal hypervolume distribution for our parametrized family of Pareto fronts and relate it to an optimal multiplicative approximation. Therefore, 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.

As in Section 3, we assume that both extreme have to be included in both distributions. For the optimal hypervolume distribution, it suffices to find the $x_2, x_3, \ldots, x_{\mu-1}$

(a) Set $X_{\text{opt}}^{\text{HYP}}(12, f_2^{\text{asy}})$.   (b) Set $X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{asy}})$.
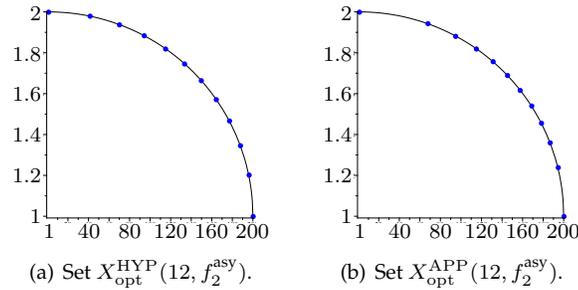
Figure 4: Optimal point distributions for asymmetric front f_2^{asy}. Note that the optimal hypervolume distribution and the optimal approximation distribution differ in this case. The set of points maximizing the hypervolume yields an approximation ratio of $\text{APP}(X_{\text{opt}}^{\text{HYP}}(12, f_2^{\text{asy}})) \approx 1.038$, which is 0.839% larger than the optimal approximation ratio $\text{APP}(X_{\text{opt}}^{\text{APP}}(12, f_2^{\text{asy}})) \approx 1.030$.

that maximize the dominated hypervolume, that is, the solutions 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 [18].

In the optimal multiplicative approximation, 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. We used the Optimization package of Maple 15.

In the following, we present the results that have been obtained by our numerical investigations. We first examine the case of f_2. Figures 3 and 4 show different point distributions for f_2. It can be observed that the hypervolume distribution differs from the optimal distribution. Figures 3(a) and 3(b) show the distributions for the symmetric front
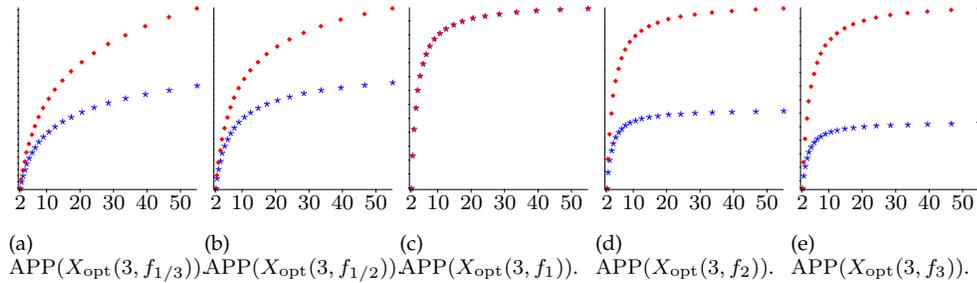
$$f_2(x) = 1 + \sqrt{1 - (x - 1)^2}$$

with $(x_1, y_1) = (1, 2)$ and $(x_\mu, y_\mu) = (2, 1)$. Figures 4(a) and 4(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 3(a) and 4(a) while the relative positions achieving an optimal approximation change with scaling (cf. Figures 3(b) and 4(b)). Hence, the relative position of

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

Figure 5: 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). We omit the values of the y-axis as we are only interested in the relative comparison (◆ vs. ⋆) for each front f_p. Note that as analytically predicted in Theorem 2, both curves coincide in (c) for the linear function f_1 independent of the scaling.

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 3 and 4 the optimal multiplicative approximation factor for the symmetric and asymmetric case is 1.021 (Figure 3(b)) and 1.030 (Figure 4(b)), respectively, while the hypervolume only achieves an approximation of 1.025 (Figure 3(a)) and 1.038 (Figure 4(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 5 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 concave 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 convex 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 6 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
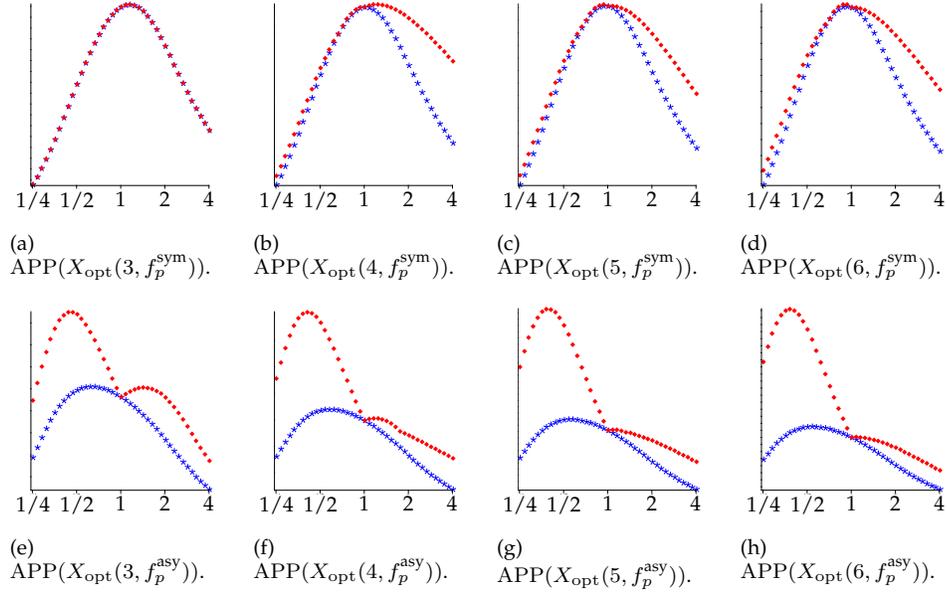
14
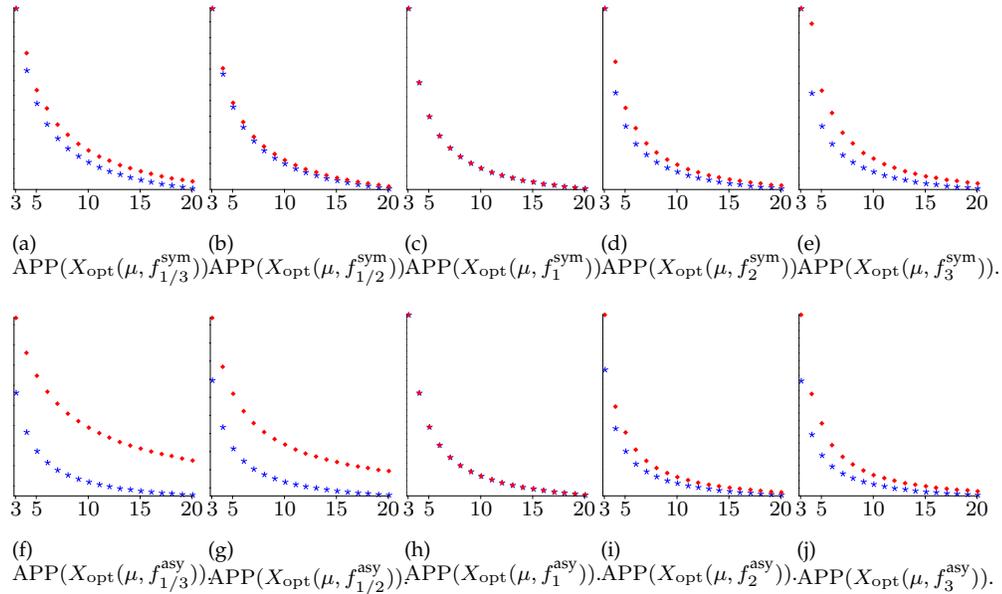
Figure 6: Approximation ratio of the optimal hypervolume distribution ($\blacklozenge$) and the optimal approximation distribution ($\star$) *depending on the number of points* μ for symmetric and asymmetric fronts f_p and different parameters p (cf. Definition 2). We omit the values of the y-axis as we are only interested in the relative comparison ($\blacklozenge$ vs. $\star$) for each front f_p. Note that (c) and (h) show that the approximation ratio of the optimal hypervolume distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f_1^{\mathrm{sym}}))$ and the optimal approximation distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, f_1^{\mathrm{sym}}))$ 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).

measure is especially poor even for larger μ for convex asymmetric fronts, that is, f_p^{asy} with $p < 1$ (e.g. Figures 6(f) and 6(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 [27]. The hypervolume was thought to prefer convex regions to concave regions [32] 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 7 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 convex ($p < 1$) fronts are much better approximated by the hypervolume than the concave ($p > 1$) fronts (cf. Figure 7(a)–(d)). For f_p^{asy} this is surprisingly the other way around (cf. Figure 7(e)–(h)).

4 Influence of the reference point

In all previous investigations, we have not considered the impact of the reference point. To allow a fair comparison we assumed that the optimal approximation distribution and the optimal hypervolume distribution have to include both extreme points. This is clearly not optimal when considering the optimal approximation distribution. There-

(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 7: 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. We omit the values of the y-axis as we are only interested in the relative comparison (♦ vs. ⋆) for each front f_p and population size μ. Note that (a) shows that the approximation ratio of the optimal hypervolume distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{HYP}}(3, f_p^{\mathrm{sym}}))$ and the optimal approximation distribution $\mathrm{APP}(X_{\mathrm{opt}}^{\mathrm{APP}}(3, f_p^{\mathrm{sym}}))$ are equivalent for all examined p.

fore, we relax our assumption and allow any set consisting of μ points and raise the question how the optimal approximation distribution looks in this case. Considering the hypervolume indicator, the question arises whether this optimal approximation distribution can be achieved by choosing a certain reference point. Therefore, the goal of this section is to examine the impact of the reference point for determining optimal approximation distributions.

For this we have to redefine parts of the notation. We mark all variables with a hat (like ^) to make clear that we do not require the extreme points to be included anymore.

Consider again a Pareto front f. We now let $\mathcal{X}_r(\mu, f)$ be the set of all subsets of

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

of cardinality μ, where we do *not* assume that $(x_{\min}, f(x_{\min}))$ and $(x_{\max}, f(x_{\max}))$ have to be necessarily contained. We also have to redefine the notion of optimal hypervolume distributions and optimal approximation distribution.

Definition 3 *The optimal hypervolume distribution*

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f) := \underset{X \in \mathcal{X}_r(\mu, f)}{\mathrm{argmax}} \ \mathrm{HYP}_r(X)$$

consists of μ points that maximize the hypervolume with respect to f. The optimal approxima-

tion distribution

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f) := \operatorname*{argmin}_{X \in \mathcal{X}_r(\mu, f)} \mathrm{APP}(X)$$

consists of μ points that minimize the approximation ratio with respect to f.

4.1 Optimal approximations

Similar to Lemma 1, the following lemma states conditions for an optimal approximation distribution which does not have to contain the extreme points.

Lemma 4 *Let $f \colon [x_{\min}, x_{\max}] \to \mathbb{R}$ be a Pareto front and $X = \{x_1, \ldots, x_\mu\}$ a solution set with $x_i < x_{i+1}$ for all $1 \leq i < \mu$. If there is a ratio $\delta > 1$ and a set $Z = \{z_1, \ldots, z_{\mu-1}\}$ with $x_i < z_i < x_{i+1}$ for all $1 \leq i < \mu$ such that*

- *$z_i = \delta \cdot x_i$ for all $1 \leq i \leq \mu$ (where $z_\mu = x_{\max}$) and*
- *$f(z_i) = \delta \cdot f(x_{i+1})$ for all $0 \leq i < \mu$ (where $z_0 = x_{\min}$)*

then $X = \widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)$ is the optimal approximation distribution with approximation ratio δ.

Proof. For each i, $1 \leq i \leq \mu - 1$, z_i is the worst approximated point in the interval $[x_i, x_{i+1}]$. Furthermore, $z_0 = x_{\min}$ is the worst approximated point in the interval $[x_{\min}, x_1]$ and $z_\mu = x_{\max}$ is the worst approximated point in the interval $[x_\mu, x_{\max}]$. This implies that the approximation ratio of X is

$$\delta = \max \left\{ \frac{f(z_0)}{f(x_1)}, \frac{z_\mu}{x_\mu}, \frac{z_i}{x_i}, \frac{f(z_i)}{f(x_{i+1})} \;\middle|\; 1 \leq i \leq \mu - 1 \right\}.$$

Assume there is a different solution set $X' = \{x'_1, \ldots, x'_\mu\}$ with $x'_i < x'_{i+1}$ for all $1 \leq i < \mu$ and approximation ratio at most δ.

Since $X' \neq X$ there is an index i with $x'_i \neq x_i$. Consider the smallest such index. We distinguish the two cases $x'_i < x_i$ and $x'_i > x_i$.

Assume $x'_i < x_i$. Consider the point $z'_i = \delta \cdot x'_i$. Since $z'_i = \delta \cdot x'_i < \delta \cdot x_i = z_i$, we derive $x'_{i+1} < x_{i+1}$ as otherwise $\frac{f(z'_i)}{f(x'_{i+1})} \geq \frac{f(z'_i)}{f(x_{i+1})} > \delta$ would contradict our assumption that X' achieves an approximation ratio of at most δ. Repeating the argument $(\mu - i)$-times leads to $x'_\mu < x_\mu$, which gives $\delta \cdot x'_\mu < \delta \cdot x_\mu = x_{\max}$. This implies that the approximation of $x_{\max}$ by X' is $\frac{x_{\max}}{x'_\mu} > \delta$ which contradicts the assumption that X' achieves an approximation ratio of at most δ.

Assume $x'_i > x_i$. Then all points within $(z_{i-1}, f^{-1}(\delta \cdot f(x'_i)))$ are not δ-approximated. The interval is not empty since $f^{-1}(\delta \cdot f(x'_i)) > x'_i > x_i > z_{i-1}$ due to $\delta > 1$ and f strictly monotonically decreasing. We have another contradiction.

Altogether, we get that $X = \widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)$ is the unique set achieving an approximation ratio of at most δ and therefore an optimal approximation distribution. $\qquad\square$

The previous lemma can be used to compute the overall optimal approximation distribution of μ for a given function describing the Pareto front. In the following, we will use this to compare it to the optimal hypervolume distribution depending on the

chosen reference point. Again we consider the class of linear fronts and the class of convex fronts given in Section 3.

4.2 Analytic results for linear fronts

We first consider linear fronts. The optimal multiplicative approximation factor can be easily determined with Lemma 4 as shown in the following theorem.

Theorem 5 *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*

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f) = \{x_1, \ldots, x_\mu\},$$

where $x_i = \frac{d\,(\mu c - i\,(c+d-1))}{c\,(c+(\mu+1)\,d-1)}$, $1 \le i \le \mu$, and

$$\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)) = \frac{c + (\mu + 1)\,d - 1}{\mu d}.$$

Proof. Using Lemma 4, we get the following system of $2\,(\mu + 1)$ linear equations

$$
\begin{aligned}
z_0 &= 1, \\
z_i &= \delta\, x_i && \text{for } i = 1 \ldots \mu, \\
z_\mu &= (1 - d)/c, \\
c\, y_i + d &= \delta\,(c\, x_{i+1} + d) && \text{for } i = 0 \ldots \mu - 1.
\end{aligned}
$$

The unique solution of this system of linear equations is

$$
\begin{aligned}
\delta &= \frac{c + (\mu + 1)\,d - 1}{\mu d}, \\
x_i &= \frac{d\,(\mu\,c - i\,(c + d - 1))}{c\,(c + (\mu + 1)\,d - 1)} && \text{for } i = 1 \ldots \mu, \\
z_i &= 1 - i\,\frac{c + d - 1}{\mu\,c} && \text{for } i = 0 \ldots \mu,
\end{aligned}
$$

which proves the claim. $\qquad\square$

It remains to analyze the approximation factor achieved by an optimal hypervolume distribution. The impact of the reference point for the class of linear functions has been investigated by Brockhoff in [10]. Using his results, we can conclude the following theorem.

Theorem 6 *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. Let $\mu \ge 2$ and*

$$
\begin{aligned}
M_1 &:= \min\left\{ c + d - r_1,\ \frac{\mu}{\mu - 1} \cdot (c + d - 1), \right. \\
&\qquad\qquad \left. (d - 1) + \frac{d + (\mu + 1)\,c}{\mu} + \frac{r_2 + d - 1}{\mu c} + \frac{d - 1}{\mu c^2} \right\}, \\
M_2 &:= \min\left\{ \frac{1 - d}{c} - r_2,\ \frac{\mu}{\mu - 1} \cdot \frac{1 - c - d}{c},\ \frac{1 - c - d}{c} + \frac{r_1 - c - d}{\mu c} \right\}.
\end{aligned}
$$

Then the optimal hypervolume distribution with respect to the reference point r is

$$\widehat{X}_{\text{opt}}^{\text{HYP}}(\mu, r, f) = \{x_1, \ldots, x_\mu\}$$

where

$$x_i = \frac{M_1 - d + 1}{c} + i \cdot \frac{d - M_1 + (M_2 + 1)\,c - 1}{c\,(\mu + 1)}.$$

Theorem 2 follows immediately from Theorem 3 of Brockhoff [10] by translating their minimization setting into our maximization setting. Knowing the set of points which maximize the hypervolume, we can now determine the achieved approximation depending on the chosen reference point.

Theorem 7 *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. Let $\mu \geq 2$ and M_1 and M_2 defined as in Theorem 6, then*

$$\text{APP}(\widehat{X}_{\text{opt}}^{\text{HYP}}(\mu, r, f)) = \max\{A_\ell, A_c, A_r\}$$

where

$$A_\ell := \frac{(c + d)\,(\mu + 1)}{\mu + c + d + M_1\mu + M_2 c},$$

$$A_c := \frac{d\,(\mu + 1)}{d\mu + c + 2d - 1 - M_1 + M_2 c},$$

$$A_r := \frac{(1 - d)\,(\mu + 1)}{c\mu - d + 1 + M_1 + M_2 c\mu}.$$

Proof. We want to determine the approximation ratio of the optimal hypervolume distribution $\widehat{X}_{\text{opt}}^{\text{HYP}}(\mu, r, f) = \{x_1, \ldots, x_\mu\}$ as defined in Theorem 6. For this, we distinguish between three cases. The approximation ratio of the inner points $\tilde{x}$ with $x_1 \leq \tilde{x} \leq x_\mu$ can be determined as in the proof of Theorem 2. It suffices to plug the definition of x_i and x_{i+1} from Theorem 6 into equation 2. Let $\tilde{x}$ be the solution of this linear equation. Then the inner approximation factor is

$$A_c = \frac{\tilde{x}}{x_i} = \frac{d\,(\mu + 1)}{d\mu + c + 2d - 1 - M_1 + M_2 c},$$

which is independent of i.

It remains to determine the outer approximation factors. The approximation factor of the points $\tilde{x}$ with $1 \leq \tilde{x} \leq x_1$ is maximized for $\tilde{x} = 1$. The left approximation factor is therefore

$$A_\ell = \frac{c + d}{f(x_1)} = \frac{(c + d)\,(\mu + 1)}{\mu + c + d + M_1\mu + M_2 c}.$$

The approximation factor of the points $\tilde{x}$ with $x_\mu \leq \tilde{x} \leq (1 - d)/c$ is maximized for $\tilde{x} = x_\mu$. The right approximation factor is therefore

$$A_r = \frac{1 - d}{c x_\mu} = \frac{(1 - d)\,(\mu + 1)}{c\mu - d + 1 + M_1 + M_2 c\mu}.$$

The overall approximation factor is then the largest approximation factor of the three parts, that is, $\max\{A_\ell, A_c, A_r\}$. $\qquad\square$

4.3 Analytic results for a class of convex fronts

We now consider convex fronts and investigate the overall optimal multiplicative approximation first which does not have to include the extreme points. The following theorem shows how such an optimal approximations looks like and will serve later for the comparison to an optimal hypervolume distribution in dependence of the chosen reference point.

Theorem 8 *Let $f\colon [1,c] \to [1,c]$ be a convex front with $f(x) = c/x$ where $c > 1$ is an arbitrary constant. Then*

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f) = \{x_1, \ldots, x_\mu\},$$

where $x_i = c^{\frac{2i-1}{2\mu}}$, $1 \le i \le \mu$, and $\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)) = c^{\frac{1}{2\mu}}$.

Proof. Using Lemma 4, we have $z_0 = x_{\min} = 1$ and $z_\mu = x_{\max} = c$. Furthermore,

$$f(z_0) = \delta f(x_1)$$
$$\Leftrightarrow \quad c = \delta \cdot \frac{c}{x_1}$$
$$\Rightarrow \quad x_1 = \delta.$$

We have $z_i = \delta x_i$, $1 \le i \le \mu$, and

$$f(z_i) = \delta \cdot f(x_{i+1})$$
$$\Leftrightarrow \quad \frac{c}{z_i} = \delta \frac{c}{x_{i+1}}$$
$$= \quad \frac{c}{\delta x_i} = \delta \frac{c}{x_{i+1}}$$
$$\Rightarrow \quad \frac{x_{i+1}}{x_i} = \delta^2$$

for $1 \le i < \mu$. This implies $x_i = \delta^{2i-1}$, $1 \le i \le \mu$. Furthermore,

$$y_\mu = \delta \, x_\mu$$
$$\Leftrightarrow \quad c = \delta^{2\mu}$$
$$\Rightarrow \quad \delta = c^{\frac{1}{2\mu}}$$

This implies $x_i = c^{\frac{2i-1}{2\mu}}$, $1 \le i \le \mu$ and $\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)) = c^{\frac{1}{2\mu}}$ which completes the proof. $\qquad\square$

Now, we consider the optimal hypervolume distribution depending on the choice of the reference point and compare it to the optimal multiplicative approximation.
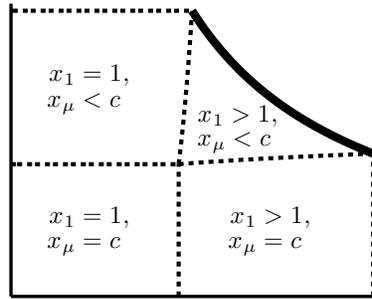
Theorem 9 *Let $f\colon [1,c] \to [1,c]$ be a convex front with $f(x) = c/x$ where $c > 1$ is an arbitrary constant. Then*

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f) = \{x_1, \ldots, x_\mu\},$$

where x_i, $1 \le i \le \mu$, depend on the choice of the reference point r as follows:

1. *If $r_1 \le c^{-1/(\mu-1)}$ and $r_2 \le c^{-1/(\mu-1)}$, then $x_1 = 1$, $x_\mu = c$, $x_i = c^{(i-1)/(\mu-1)}$ for $1 \le i \le \mu$, and*

$$\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) = c^{\frac{1}{2\mu-2}}.$$

Figure 8: Position of the extremal points of optimal hypervolume distributions depending on the reference point for concave functions $f: [1, c] \to [1, c]$ with $f(x) = c/x$ and $c > 1$. Depending on the position of the reference point, the leftmost point of a optimal hypervolume distribution x_1 is either at the border ($x_1 = x_{\min} = 1$) or inside the domain ($x_1 > x_{\min} = 1$). Similarly, the rightmost point x_μ is either at the border ($x_\mu = x_{\max} = c$) or inside the domain ($x_\mu < x_{\max} = c$). (Note that the figure looks very similar for linear functions.)

2. If $r_1 \leq c\, r_2^\mu$ and $r_2 \leq c\, r_1^\mu$, then $x_1 > 1$, $x_\mu < c$, $x_i = (c^i\, r_1^{\mu-i+1} / r_2^i)^{1/(\mu+1)}$ for $1 \leq i \leq \mu$, and

$$\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) = \max\left\{ \left(\frac{c \cdot r_1^\mu}{r_2}\right)^{\frac{1}{\mu+1}}, \left(\frac{c}{r_1 r_2}\right)^{\frac{1}{2(\mu+1)}}, c\left(\frac{r_2^\mu}{c^\mu \cdot r_1}\right)^{\frac{1}{\mu+1}} \right\}.$$

3. If $r_2 \geq c^{-1/(\mu-1)}$, $r_2 \leq c$, and $r_2 \geq cr_1^\mu$, then $x_1 = 1$, $x_i = (c/r_2)^{(i-1)/\mu}$ for $1 \leq i \leq \mu$, and

$$\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) = \max\left\{ \left(\frac{c}{r_2}\right)^{\frac{1}{2\mu}}, c\left(\frac{c}{r_2}\right)^{\frac{\mu}{\mu-1}} \right\}.$$

4. If $r_1 \geq c^{-1/(\mu-1)}$, $r_1 \leq c$ and $r_1 \geq c\, r_2^\mu$, then $x_\mu = c$, $x_i = r_1(c/r_1)^{i/\mu}$ for $1 \leq i \leq \mu$, and

$$\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) = \max\left\{ \left(\frac{c}{r_1}\right)^{\frac{1}{2\mu}}, \left(cr_1^{\mu-1}\right)^{\frac{1}{\mu}} \right\}.$$

Proof. In order to proof Theorem 9, we distinguish four cases, namely whether $x_1 = 1$ or $x_1 > 1$ and whether $x_\mu = c$ or $x_\mu < c$. Figure 8 gives an illustration of the four cases.

The first case $x_1 = 1$ and $x_\mu = c$ corresponds to the previous situation where we required that both extreme points are included. The statement of Theorem 9 for this case follows immediately from equations 3 and 4 in Section 3.3.

The second case $x_1 > 1$ and $x_\mu < c$ is more involved. First note that we consider only points that have a positive contribution with respect to the given reference point. Therefore, we assume that $r_1 < x_1$ and $r_2 < f(x_\mu)$ holds.

The hypervolume of a set of points $X = \{x_1, \ldots, x_\mu\}$, where w.l.o.g. $x_1 \leq x_2 \leq \cdots \leq x_\mu$, with respect to a reference point $r = (r_1, r_2)$ with $r_1 < x_1$ and $r_2 < f(x_\mu)$ is

then given by

$$
\begin{aligned}
\text{HYP}(X, r) ={}& (x_1 - r_1) \cdot (f(x_1) - r_2) \\
&+ (x_2 - x_1) \cdot (f(x_2) - r_2) \\
&\cdots \\
&+ (x_\mu - x_{\mu-1}) \cdot (f(x_\mu) - r_2) \\
={}& c \cdot \mu + r_1 r_2 \\
&- c(r_1/x_1 + x_1/x_2 + x_2/x_3 + \cdots + x_{\mu-1}/x_\mu) \\
&- x_\mu \cdot r_2.
\end{aligned}
$$

In order to maximize the hypervolume, we consider the function

$$
h(x_1, \ldots, x_\mu) = c \left(\frac{r_1}{x_1} + \frac{x_1}{x_2} + \frac{x_2}{x_3} + \cdots + \frac{x_{\mu-1}}{x_\mu} \right) + x_\mu \cdot r_2
$$

and compute its partial derivatives.

We have $1 < x_1 < x_2 < \ldots < x_\mu < c$ as the equality of two points implies that one of them can be exchanged for another and thereby increases the hypervolume. We work under these assumptions and aim to find a set of points X that minimizes the function h. To do this, we consider the gradient vector given by the partial derivatives

$$
h'(x_1, \ldots, x_\mu) = \left(\frac{c}{x_2} - \frac{c r_1}{x_1^2}, \frac{c}{x_3} - \frac{c x_1}{x_2^2}, \ldots, \frac{c}{x_\mu} - \frac{c x_{\mu-2}}{x_{\mu-1}^2}, r_2 - \frac{c x_{\mu-1}}{x_\mu^2} \right).
$$

This implies that h can be minimized by setting

$$
\begin{aligned}
x_2 &= x_1^2/r_1 \\
x_3 &= x_2^2/x_1 &&= x_1^3/r_1^2 \\
x_4 &= x_3^2/x_2 &&= x_1^4/r_1^3 \\
&\vdots \\
x_\mu &= x_{\mu-1}^2/x_{\mu-2} &&= x_1^\mu/r_1^{\mu-1} \\
x_\mu^2 &= c x_{\mu-1}/r_2.
\end{aligned}
$$

Hence with

$$
\frac{c x_{\mu-1}}{r_2} = \frac{c x_1^{\mu-1}}{r_1^{\mu-2} r_2} = \frac{x_1^{2\mu}}{r_1^{2(\mu-1)}}
$$

we get

$$
x_1 = \min \left\{ \max \left\{ \left(\frac{c r_1^\mu}{r_2} \right)^{\frac{1}{\mu+1}}, 1 \right\}, c \right\}.
$$

As we can assume $x_1 \geq 1$ and $x_\mu \leq c$, we get for $r_2 \leq c r_1^\mu$ and $r_1 \leq c r_2^\mu$ that

$$
x_i = \left(\frac{c^i \cdot r_1^{\mu-i+1}}{r_2^i} \right)^{\frac{1}{\mu+1}}
$$

for $1 \leq i \leq \mu$. It now remains to determine the achieved approximation factor. For this, we proceed as in Theorem 3 and use that

$$
\frac{x}{x_i} = \frac{f(x)}{f(x_{i+1})} \implies x = \sqrt{x_i \cdot x_{i+1}}.
$$

This gives an approximation factor of the inner points of

$$\frac{x_{i+1}}{x} = \left(\frac{c}{r_1 r_2}\right)^{\frac{1}{2(\mu+1)}}.$$

For the upper end points the approximation is

$$\frac{c}{f(x_1)} = x_1 = \left(\frac{c \cdot r_1^\mu}{r_2}\right)^{\frac{1}{\mu+1}}.$$

For the lower end points the approximation is

$$\frac{c}{x_\mu} = c \cdot \left(\frac{r_2^\mu}{c^\mu \cdot r_1}\right)^{\frac{1}{\mu+1}}.$$

Hence the overall approximation factor in the second case is

$$\max\left\{\left(\frac{c \cdot r_1^\mu}{r_2}\right)^{\frac{1}{\mu+1}}, \left(\frac{c}{r_1 r_2}\right)^{\frac{1}{2(\mu+1)}}, c \cdot \left(\frac{r_2^\mu}{c^\mu \cdot r_1}\right)^{\frac{1}{\mu+1}}\right\}.$$

The third case $x_1 = 1$ and $x_\mu < c$ fixes only the left end of the front. Here, we consider the function

$$h(x_2, \ldots, x_\mu) = c\left(\frac{1}{x_2} + \frac{x_2}{x_3} + \frac{x_3}{x_4} + \cdots + \frac{x_{\mu-1}}{x_\mu}\right) + x_\mu \cdot r_2.$$

Not that in contrast to the second case, $h(\cdot)$ does not depend on r_1. We can assume without loss of generality that $1 = x_1 < x_2 < \ldots < x_\mu < c$. The partial derivatives are therefore

$$h'(x_2, \ldots, x_\mu) = \left(\frac{c}{x_3} - \frac{c}{x_2^2}, \frac{c}{x_4} - \frac{cx_2}{x_3^2}, \ldots, \frac{c}{x_\mu} - \frac{cx_{\mu-2}}{x_{\mu-1}^2}, r_2 - \frac{cx_{\mu-1}}{x_\mu^2}\right).$$

This implies that h can be minimized by setting

$$\begin{aligned}
x_3 &= x_2^2 \\
x_4 &= x_3^2/x_2 &= x_2^3 \\
x_5 &= x_4^2/x_3 &= x_2^4 \\
&\vdots \\
x_\mu &= x_{\mu-1}^2/x_{\mu-2} &= x_2^{\mu-1} \\
x_\mu^2 &= \frac{cx_{\mu-1}}{r_2}.
\end{aligned}$$

Starting with $x_\mu^2 = x_\mu^2$ we get,

$$\frac{cx_{\mu-1}}{r_2} = \frac{cx_2^{\mu-2}}{r_2} = x_2^{2(\mu-1)}$$

and

$$x_2 = \min\left\{\max\left\{\left(\frac{c}{r_2}\right)^{\frac{1}{\mu}}, 1\right\}, c\right\}.$$

Again using that $x_2 \geq 1$ and $x_\mu \leq c$ and assuming that $r_2 \leq c$ and $r_2 \geq c^{-\frac{1}{\mu-1}}$, we get

$$x_i = \left(\frac{c}{r_2}\right)^{\frac{i-1}{\mu}}.$$

This results in an approximation factor for the inner points of

$$\frac{x_{i+1}}{\sqrt{x_i \cdot x_{i+1}}} = \left(\frac{c}{r_2}\right)^{\frac{1}{2\mu}}.$$

For the upper end points the approximation is

$$\frac{c}{f(x_1)} = x_1 = 1.$$

For the lower end points the approximation is

$$\frac{c}{x_\mu} = c\left(\frac{r_2}{c}\right)^{\frac{\mu-1}{\mu}}.$$

Hence the overall approximation factor in the third case is

$$\max\left\{\left(\frac{c}{r_2}\right)^{\frac{1}{2\mu}}, c\left(\frac{c}{r_2}\right)^{\frac{\mu}{\mu-1}}\right\}.$$

The fourth case $x_1 > 1$ and $x_\mu = c$ fixes the right end of the front. We consider the function

$$h(x_1, \ldots, x_{\mu-1}) = c\left(\frac{r_1}{x_1} + \frac{x_1}{x_2} + \frac{x_2}{x_3} + \cdots + \frac{x_{\mu-2}}{x_{\mu-1}} + \frac{x_{\mu-1}}{c}\right)$$

and compute its partial derivatives

$$h'(x_1, \ldots, x_{\mu-1}) = \left(\frac{c}{x_2} - \frac{cr_1}{x_1^2}, \frac{c}{x_3} - \frac{cx_1}{x_2^2}, \ldots, \frac{c}{x_{\mu-1}} - \frac{cx_{\mu-3}}{x_{\mu-2}^2}, 1 - \frac{cx_{\mu-2}}{x_{\mu-1}^2}\right).$$

This implies that h can be minimized by setting

$$
\begin{array}{rclcl}
x_2 & = & x_1^2/r_1 & & \\
x_3 & = & x_2^2/x_1 & = & x_1^3/r_1^2 \\
x_4 & = & x_3^2/x_2 & = & x_1^4/r_1^3 \\
& \vdots & & & \\
x_{\mu-1} & = & x_{\mu-2}^2/x_{\mu-3} & = & x_1^{\mu-1}/r_1^{\mu-2} \\
x_{\mu-1}^2 & = & cx_{\mu-2}. & &
\end{array}
$$

Setting

$$cx_{\mu-2} = cx_1^{\mu-2}/r_1^{\mu-3} = x_1^{2\mu-2}/r_1^{2\mu-4}$$

we get

$$x_1 = \min\left\{ \max\left\{ \left(cr_1^{\mu-1} \right)^{\frac{1}{\mu}}, 1 \right\}, c \right\}.$$

Using that $x_1 \geq 1$ and $x_{\mu-1} \leq c$ and assuming $r_2 \geq c^{-1/(\mu-1)}$ and $r_2 \leq c$ gives

$$x_i = r_1 \left(\frac{c}{r_1} \right)^{\frac{i}{\mu}}.$$

This results in an approximation factor for the inner points of

$$\frac{x_{i+1}}{\sqrt{x_i \cdot x_{i+1}}} = \left(\frac{c}{r_1} \right)^{\frac{1}{2\mu}}$$

For the upper end points the approximation is

$$\frac{c}{f(x_1)} = x_1 = \left(cr_1^{\mu-1} \right)^{\frac{1}{\mu}}.$$

For the lower end points the approximation is

$$\frac{c}{x_\mu} = 1.$$

Hence the overall approximation factor for the fourth case is

$$\max\left\{ \left(\frac{c}{r_1} \right)^{\frac{1}{2\mu}}, \left(cr_1^{\mu-1} \right)^{\frac{1}{\mu}} \right\},$$

which finishes the proof. □

### 4.4   Numerical evaluation for two specific fronts

We now use the theoretical results of this Section 4 on the approximation factor depending on the reference point and study two specific fronts as an example.

First, we consider the linear front $f\colon [1, 2] \to [1, 2]$ with $f(x) = 3 - x$. A plot of this front is shown in Figure 2 (a). For $\mu = 10$, Theorem 5 gives that the optimal distribution

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f) = \left\{ \frac{33}{31}, \frac{36}{31}, \frac{39}{31}, \frac{42}{31}, \frac{45}{31}, \frac{48}{31}, \frac{51}{31}, \frac{54}{31}, \frac{57}{31}, \frac{60}{31} \right\}$$

achieves the (optimal) approximation of $\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)) = 31/30$. With Theorem 7 we can now also determine the approximation factor of optimal hypervolume distributions depending on the reference point r. For some specific reference points we get

$$
\begin{aligned}
\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) &= & 2 & \quad \text{for } r = (2, 1), \\
\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) &= & 4/3 & \quad \text{for } r = (3/2), \\
\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) &= & 22/21 & \quad \text{for } r = (1, 1), \\
\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) &= & 31/30 & \quad \text{for } r = (30/31, 30/31), \\
\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{HYP}}(\mu, r, f)) &= & 27/26 & \quad \text{for } r \leq (8/9, 8/9).
\end{aligned}
$$
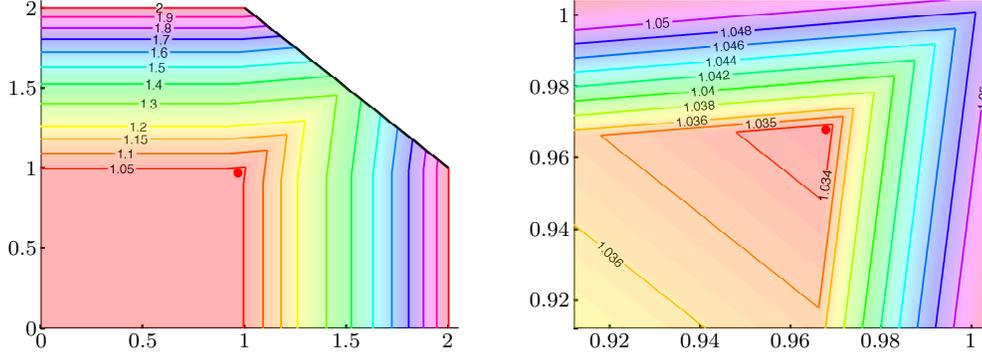
Figure 9: Approximation factor of optimal hypervolume distribution depending on the reference point for the linear function $f\colon [1,2] \to [1,2]$ with $f(x) = 3-x$ for a population size of $\mu = 10$. The right figure shows a closeup view of the area around the reference reference point with the best approximation ratio, which is marked with a red dot.

Figure 9 shows a plot of the approximation factor depending on the reference point. We observe that if $r_2 > 32\, r_1 - 30$ or $r_1 > 32\, r_2 - 30$, the approximation factor is only determined by the inner approximation factor A_c (cf. Theorem 7). Moreover, for $r_1 > 10\, r_2 - 8$ and $r_2 > 10\, r_1 - 8$ the approximation factor only depends on r_1 and r_2, respectively. For $r \le (8/9, 8/9)$ it is constant. The optimal approximation factor is achieved for the reference point

$$\left(\frac{c^2 + d\,(c+1+\mu) - 1}{c + d\,(\mu+1) - 1}, \frac{c^2 + d\,(c+1+\mu) - 1}{c + d\,(\mu+1) - 1} \right) = (30/31, 30/31).$$

Let us now consider a specific convex function $f\colon [1,c] \to [1,c]$ with $f(x) = c/x$ and $c = 2$ for a population size of $\mu = 10$. The function is shown in Figure 2(b). Theorem 8 gives that the optimal distribution

$$\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f) = \left\{ 2^{\frac{1}{20}}, 2^{\frac{3}{20}}, 2^{\frac{5}{20}}, \cdots, 2^{\frac{15}{20}}, 2^{\frac{17}{20}}, 2^{\frac{19}{20}} \right\}$$

achieves the (optimal) approximation factor of

$$\mathrm{APP}(\widehat{X}_{\mathrm{opt}}^{\mathrm{APP}}(\mu, f)) = 2^{1/20} \approx 1.0353.$$

With Theorem 9 we can determine the approximation factor of optimal hypervolume distributions depending on the reference point r. Figure 10 shows the behavior of the approximation factor depending on the choice of the reference point r. We observe that for $r_2 > c\, r_1^{\mu}$ and $r_1 > c\, r_2^{\mu}$, the approximation factor only depends on r_1 and r_2, respectively. For
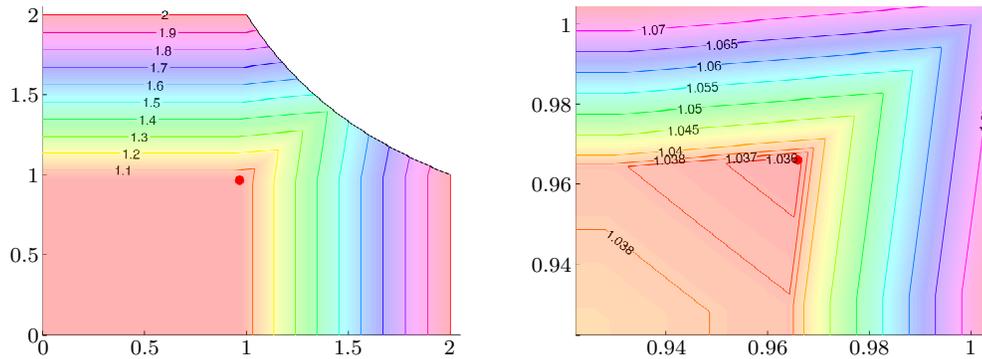
$$r \le (c^{1/(1-\mu)}, c^{1/(1-\mu)}) = (2^{-1/9}, 2^{-1/9}) \approx (0.926, 0.926)$$

the approximation factor is invariably

$$c^{1/(2\mu - 2)} = 2^{1/18} \approx 1.0393.$$

The optimal approximation factor is achieved for the reference point

$$(c^{-1/2\mu}, c^{-1/2\mu}) = (2^{-1/20}, 2^{-1/20}) \approx (0.966, 0.966).$$

Figure 10: Approximation factor of optimal hypervolume distribution depending on the reference point for the convex function $f \colon [1,2] \to [1,2]$ with $f(x) = 2/x$ for a population size of $\mu = 10$. The right figure shows a closeup view of the area around the reference point with the best approximation ratio, which is marked with a red dot.

5   Conclusions

Evolutionary algorithms have been shown to be very successful for dealing with multi-objective optimization problems. This is mainly due to the fact that such problems are hard to solve by traditional optimization methods. The use of the population of an evolutionary algorithm to approximate the Pareto front seems to be a natural choice for dealing with these problems. The use of 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. We started by considering the case where we assumed that the extreme points with respect to the given objective functions have to be included in both distributions. Considering linear fronts and a class of convex fronts we have pointed out that the hypervolume indicator gives provably the best multiplicative approximation ratio that is achievable. To gain further insights into the optimal hypervolume distribution and its relation to multiplicative approximations, we carried out numerical investigations. These investigations point out that the shape as well the scaling of the objectives heavily influences the approximation behavior of the hypervolume indicator. Examining fronts with different 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.

After having obtained these results, we analyzed the impact of the reference points on the hypervolume distribution and compared the multiplicative approximation ratio obtained by this indicator to the overall optimal approximation that does not have to contain the extreme points. In general, the choice of the reference point determines

the approximation ratio that a hypervolume-based algorithm can achieve. However, it is hard to determine the reference point that optimizes the approximation ratio as it depends on the multi-objective problem under consideration. Our investigations show that also in this case the hypervolume distribution can lead to an overall optimal approximation when the reference point is chosen in the right way for the class of linear and convex functions under investigation. Furthermore, our results point out the impact of the choice of the reference point with respect to the approximation ratio that is achieved as shown in Figures 9 and 10.

Our results provide insights into the connection of the optimal hypervolume distribution and approximation ratio for special classes of functions describing the Pareto fronts of multi-objective problems having two objectives. For future work, it would be interesting to obtain results for broader classes of functions as well as problems having more than 2 objectives.

Acknowledgement

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 *10th International Workshop on Foundations of Genetic Algorithms (FOGA '09)*, pages 87–102. ACM Press, 2009.

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

[3] J. Bader and E. Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19:45–76, 2011.

[4] 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.

[5] K. Bringmann. An improved algorithm for Klee's measure problem on fat boxes. *Computational Geometry: Theory and Applications*, 45(5-6):225–233, 2012.

[6] 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.

[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. Parameterized average-case complexity of the hypervolume indicator. In *15th Annual Conference on Genetic and Evolutionary Computation Conference (GECCO '13)*, pages 575–582. ACM Press, 2013.

[9] K. Bringmann and T. Friedrich. Approximation quality of the hypervolume indicator. *Artificial Intelligence*, 195:265–290, 2013.

[10] D. Brockhoff. Optimal μ-distributions for the hypervolume indicator for problems with linear bi-objective fronts: Exact and exhaustive results. In *8th International Conference on Simulated Evolution And Learning (SEAL '10)*, volume 6457 of *LNCS*, pages 24–34, 2010.

[11] C. A. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, 2002.

[12] K. Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, Chichester, UK, 2001.

[13] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *IEEE Congress on Evolutionary Computation (CEC '02)*, pages 825–830, 2002.

[14] M. Dorigo and T. Stützle. *Ant Colony Optimization*. MIT Press, 2004.

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

[16] M. T. M. Emmerich, N. Beume, and B. Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. In *3rd International Conference on Evolutionary Multi-Criterion Optimization (EMO '05)*, pages 62–76. Springer, 2005.

[17] M. T. M. Emmerich, A. H. Deutz, and N. Beume. Gradient-based/evolutionary relay hybrid for computing Pareto front approximations maximizing the S-metric. In *4th International Workshop on Hybrid Metaheuristics (HM '07)*, volume 4771 of *LNCS*, pages 140–156. Springer, 2007.

[18] R. Fletcher. *Practical methods of optimization*. Wiley-Interscience, New York, NY, USA, 1987.

[19] C. M. Fonseca, L. Paquete, and M. López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. In *IEEE Congress on Evolutionary Computation (CEC '06)*, pages 1157–1163, 2006.

[20] T. Friedrich, C. Horoba, and F. Neumann. Multiplicative approximations and the hypervolume indicator. In *11th Annual Conference on Genetic and Evolutionary Computation (GECCO '09)*, pages 571–578. ACM Press, 2009.

[21] T. Friedrich, K. Bringmann, T. Voß, and C. Igel. The logarithmic hypervolume indicator. In *11th International Workshop on Foundations of Genetic Algorithms (FOGA '11)*, pages 81–92. ACM Press, 2011.

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

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

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

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

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

[27] 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 *10th Annual Conference on Genetic and Evolutionary Computation (GECCO '08)*, pages 665–672. ACM Press, 2008.

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

[29] H. Yıldız and S. Suri. On Klee's measure problem for grounded boxes. In *ACM Symposium on Computational Geometry (SoCG '12)*, pages 111–120, 2012.

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

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

[32] E. Zitzler and L. Thiele. An evolutionary approach for multiobjective optimization: The strength Pareto approach. Tik report, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, May 1998.

[33] 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.

[34] 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.

[35] 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)*, volume 4403 of *LNCS*, pages 862–876. Springer, 2007.