Theory of the Hypervolume Indicator: Optimal μ-Distributions and the Choice of the Reference Point

Anne Auger*   Johannes Bader†

*TAO Team INRIA Saclay
LRI Paris Sud University
91405 Orsay Cedex, France
anne.auger@inria.fr

Dimo Brockhoff†   Eckart Zitzler†

†Computer Engineering and Networks Laboratory
ETH Zurich
8092 Zurich, Switzerland
firstname.lastname@tik.ee.ethz.ch

ABSTRACT

The hypervolume indicator is a set measure used in evolutionary multiobjective optimization to evaluate the performance of search algorithms and to guide the search. Multiobjective evolutionary algorithms using the hypervolume indicator transform multiobjective problems into single objective ones by searching for a finite set of solutions maximizing the corresponding hypervolume indicator. In this paper, we theoretically investigate how those *optimal μ-distributions*—finite sets of μ solutions maximizing the hypervolume indicator—are spread over the Pareto front of biobjective problems. This problem is of high importance for practical applications as these sets characterize the preferences that the hypervolume indicator encodes, i.e., which types of Pareto set approximations are favored.

In particular, we tackle the question whether the hypervolume indicator is biased towards certain regions. For linear fronts we prove that the distribution is uniform with constant distance between two consecutive points. For general fronts where it is presumably impossible to characterize exactly the distribution, we derive a limit result when the number of points grows to infinity proving that the empirical density of points converges to a density proportional to the square root of the negative of the derivative of the front. Our analyses show that it is not the shape of the Pareto front but only its slope that determines how the points that maximize the hypervolume indicator are distributed. Experimental results illustrate that the limit density is a good approximation of the empirical density for small μ. Furthermore, we analyze the issue of where to place the reference point of the indicator such that the extremes of the front can be found if the hypervolume indicator is optimized. We derive an explicit lower bound (possibly infinite) ensuring the presence of the extremes in the optimal distribution. This result contradicts the common belief that the reference point has to be chosen close to the nadir point: for certain types of fronts, we show that no finite reference point allows to have the extremes in the optimal μ-distribution.

Categories and Subject Descriptors

F.m [**Theory of Computation**]: Miscellaneous; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Theory

Keywords

Hypervolume indicator, multiobjective optimization

1. MOTIVATION

The field of evolutionary multiobjective optimization is mainly concerned with the issue of approximating the Pareto-optimal set, and various algorithms have been proposed for this purpose. In recent years, search algorithms based on the hypervolume indicator [27], a set measure reflecting the volume enclosed by a Pareto front approximation and a reference set, have become increasingly popular, e.g., [16, 9, 14]. They overcome the problems arising with density-based multiobjective evolutionary algorithms [21], since the hypervolume indicator guarantees strict monotonicity regarding Pareto dominance [11, 28]. Furthermore, recent extensions [23, 1] have broadened the applicability of this set measure with respect to preference articulation and high-dimensional objectives spaces. Hence, we expect that the scientific interest in the hypervolume indicator for search and performance assessment will grow further.

These developments motivate why several researchers have been trying to better understand the hypervolume indicator from a theoretical perspective, e.g., [11, 28, 23]. One key result is that a set of solutions achieving the maximum hypervolume for a specific problem covers the entire Pareto front [11]. However, the corresponding set may contain an infinite number of solutions, while in practice usually bounded solution sets are considered. Limiting the number of points to, let us say μ, changes the situation: in this case, only a portion of the Pareto front can be covered, and how the points are distributed over the front depends on several aspects, in particular on the front characteristics and the choice of the reference set. The resulting placement of points reflects the bias of the hypervolume indicator, and this issue has not been investigated rigorously so far.

For instance, Zitzler and Thiele [27] indicated that, when optimizing the hypervolume in maximization problems, "convex regions may be preferred to concave regions", which is

also stated in [18], whereas Deb et al. [7] argued that "[...] the hyper-volume measure is biased towards the boundary solutions". Knowles and Corne observed that a local optimum of the hypervolume indicator "seems to be 'well-distributed'" [16] which was also confirmed empirically [17, 9]. Some authors also addressed the choice of the reference set which usually contains just a single reference point. Knowles and Corne [17] demonstrated the impact of the reference point on the outcomes of selected multiobjective evolutionary algorithms based on an experimental study. Furthermore, rules of thumb exist, e.g., many authors recommend to use the corner of a space that is a little bit larger than the actual objective space as the reference point. Examples include the corner of a box 1% larger than the objective space in [15] or a box that is larger by an additive term of 1 than the extremal objective values obtained as in [3]. In various publications, the reference point is chosen as the nadir point of the investigated solution set, e.g., in [20, 19, 13], while others recommend rescaling of the objective values everytime the hypervolume indicator is computed [25].

The above statements about the bias of the hypervolume indicator seem to be contradictory and up to now no theoretical results are available that could confirm or falsify any of these hypotheses. This paper provides a theory that addresses the bias issue for the biobjective case and thereby contributes to a theoretical understanding of the principles underlying the hypervolume indicator. To this end, we will first formally state the setting considered in this paper (Section 2) and present some fundamental results on the optimal distributions of points on the Pareto front (Section 3). Section 4 investigates the influence of a reference point on the placement of a finite set of points. Afterwards, we will mathematically derive optimal placements of points for Pareto fronts that can be described by a line (Section 5). Later, we extend these results to general front shapes assuming that the number of points of the distributions converges to infinity (Section 6) and provide heuristic methods to determine the optimal distributions of μ points (Section 7).

2. PROBLEM STATEMENT

Throughout this study, we consider a bicriterion optimization problem $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}^2$ consisting of two objectives $(\mathcal{F}_1(x), \mathcal{F}_2(x)) = \mathcal{F}(x)$ which are without loss of generality to be minimized. The optimal solutions (Pareto optima) for this problem are given by the minimal elements of the ordered set $(\mathbb{R}^d, \preceq)$ where $\preceq$ stands for the weak Pareto dominance relation $\preceq := \{(x, y) \,|\, x, y \in \mathbb{R}^d \wedge \forall 1 \le i \le 2 : \mathcal{F}_i(x) \le \mathcal{F}_i(y)\}$. We assume that the overall optimization goal is to approximate the set of Pareto optima—the Pareto set—by means of a solution set and that the hypervolume indicator [27] is used as a measure to quantify the quality of a solution set. The image of the Pareto set under $\mathcal{F}$ is called Pareto front or just front for short.

The hypervolume indicator I_H gives, roughly speaking, the volume of the objective subspace that is dominated by a solution set $A \subset \mathbb{R}^d$ under consideration; it can be defined as follows on the basis of a reference set $R \subset \mathbb{R}^2$:

$$I_H(A) := \lambda(H(A, R)) \qquad (1)$$

where

- the set $H(A, R) := \{(z_1, z_2) \in \mathbb{R}^2 \,;\, \exists x \in A \,, \exists (r_1, r_2) \in R \,:\, \forall 1 \le i \le 2 \,:\, \mathcal{F}_i(x) \le z \le r_i\}$ denotes the
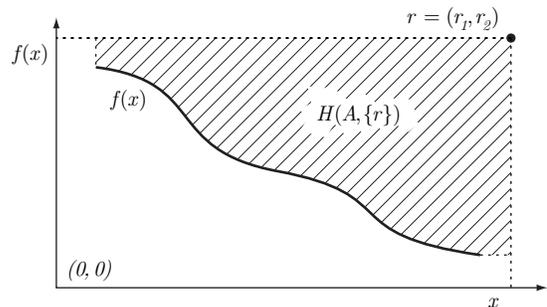

Figure 1: The hypervolume $H(A, R)$ corresponds to the set of objective vectors (hatched area) that is dominated by the solution set A, here described by a function $f(x)$, and which is enclosed by the reference set $R = r$;

set of objective vectors that are enclosed by the front $\mathcal{F}(A) := \{\mathcal{F}(x) | x \in A\}$ given by A and the reference set R, see Figure 1;

- the symbol λ stands for the Lebesgue measure with $\lambda(H(A, R)) = \int_{\mathbb{R}^2} \mathbf{1}_{H(A,R)}(z) dz$ and $\mathbf{1}_{H(A,R)}$ being the characteristic function of $H(A, R)$.

In the following, the common case of a single reference point $r = (r_1, r_2) \in \mathbb{R}^2$ is considered only, i.e., $R = \{r\}$.

It is known that the maximum hypervolume value possible is only achievable whenever the solution set A contains for each point z on the Pareto front at least one corresponding solution $x \in A$ with $\mathcal{F}(x) = z$, i.e., the image of A under $\mathcal{F}$ contains the Pareto front [11]; however, this theoretical result assumes that A can contain any number of solutions, even infinitely many. In practice, the size of A is usually restricted, e.g., by the population size when an evolutionary algorithm is employed, and therefore the question is how the indicator I_H influences the optimal selection of a *finite* number of μ solutions.

For reasons of simplicity, we will consider only the objective vectors in the following and remark that in the biobjective case the Pareto front can be described in terms of a function f mapping the image of the Pareto set under $\mathcal{F}_1$ onto the image of the Pareto set under $\mathcal{F}_2$. We assume that the image of $\mathcal{F}_1$ is a closed interval $[x_{min}, x_{max}]$ and define f as the function describing the Pareto front:

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

An example is given in Figure 1 where a front is represented in terms of this function $f(x)$. Since f represents the shape of the trade-off surface, we can conclude that, for minimization problems, f is strictly monotonically decreasing in $[x_{min}, x_{max}]$[1]. Furthermore, we only consider continuous functions f.

Now, a set of μ points on the Pareto front is entirely determined by the x-coordinates respectively the $\mathcal{F}_1$ values of these points, here denoted as $(x_1^\mu, \ldots, x_\mu^\mu)$, and f. Without loss of generality, it is assumed that $x_i^\mu \le x_{i+1}^\mu$,

[1]If f is not strictly monotonically decreasing, we can find Pareto-optimal points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ with $x_1, x_2 \in [x_{min}, x_{max}]$ such that, without loss of generality, $x_1 < x_2$ and $f(x_1) \le f(x_2)$, i.e., $(x_1, f(x_1))$ is dominating $(x_2, f(x_2))$.
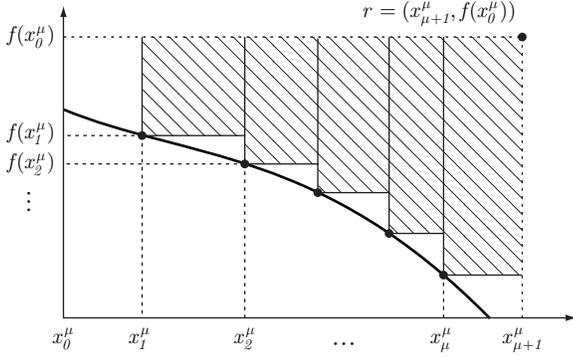
Figure 2: Hypervolume indicator computation for μ solutions $(x_1^\mu, f(x_1^\mu)), \ldots, (x_\mu^\mu, f(x_\mu^\mu))$ and the reference point $r = (r_1, r_2)$ in the biobjective case as defined in Eq. 2.

for $i = 1, \ldots, \mu - 1$ and for notation convenience, we set $x_{\mu+1}^\mu := r_1$ and $f(x_0^\mu) := r_2$ (see Figure 2). The hypervolume enclosed by these points can be, in analogy to Eq. 1, easily determined: it is the sum of rectangles of width $(x_{i+1}^\mu - x_i^\mu)$ and height $(f(x_0^\mu) - f(x_i^\mu))$ and therefore equals

$$I_H^\mu((x_1^\mu, \ldots, x_\mu^\mu)) := \sum_{i=1}^\mu (x_{i+1}^\mu - x_i^\mu)(f(x_0^\mu) - f(x_i^\mu)) \ . \quad (2)$$

Here, the hypervolume indicator I_H^μ can be seen as a μ-dimensional function of $(x_1^\mu, \ldots, x_\mu^\mu)$.

Before we can characterize a set of points maximizing the hypervolume (i.e., Eq. 2), we need to be sure that the problem is well-defined, i.e., that there exists at least one set of points maximizing Eq. 2. The existence is proven in the next theorem assuming that f is continuous.

THEOREM 1. **(Existence of optimal μ-distributions)** *If the function f describing the Pareto front is continuous, there exists (at least) one set of μ points maximizing the hypervolume.*

PROOF. Equation 2 defines a μ-dimensional function of $(x_1^\mu, \ldots, x_\mu^\mu)$. If f is moreover continuous, I_H in Eq. 2 is continuous and upper bounded by the hypervolume contribution of the entire front, i.e., $\int_{x_{min}}^{r_1} \int_{f(x_{max})}^{r_2} \mathbf{1}_{\{y > f(x)\}} dy dx$ where we prolong f for $x > x_{max}$ by setting $f(x)$ to the constant value $f(x_{max})$. Therefore, from the Mean Value Theorem there exists a set of μ points maximizing the hypervolume indicator. $\square$

Note that the previous theorem states the existence but not the uniqueness, which is not true in general but will be for linear fronts (and certain choices of the reference point), as proven later in the paper. A set of points maximizing the hypervolume whose existence is proven in the previous theorem will be called an *optimal μ-distribution*. The associated value of the hypervolume is denoted as $\overline{I_H^\mu}$.

On the basis of this notation, the research question of this paper can be reformulated as follows:

> How are the vectors $(x_1^\mu, \ldots, x_\mu^\mu)$ characterized that provide the maximum hypervolume given a reference point r and a front shape f?

This issue will be addressed in the following sections.

3. PRELIMINARY RESULTS

This section presents preliminary results on optimal μ-distributions. The first result is about how the hypervolume associated with optimal μ-distributions increases with μ. This result is in particular useful for the proof of Corollary 1.

LEMMA 1. *Let $X \subseteq \mathbb{R}$ and $f : X \mapsto f(X)$ describe the Pareto front. Let μ_1 and $\mu_2 \in \mathbb{N}$ with $\mu_1 < \mu_2$, then*

$$\overline{I_H^{\mu_1}} < \overline{I_H^{\mu_2}}$$

holds if X contains at least $\mu_1 + 1$ elements x_i for which $x_i < r_1$ and $f(x_i) < r_2$ holds.

PROOF. To prove the lemma, it suffices to show the inequality for $\mu_2 = \mu_1 + 1$. Assume $D_{\mu_1} = \{x_1^{\mu_1}, \ldots, x_{\mu_1}^{\mu_1}\}$ with $x_i^\mu \in \mathbb{R}$ is the set of x-values of the objective vectors of an optimal μ_1-distribution for the Pareto front defined by f with a hypervolume value of $\overline{I_H^{\mu_1}}$. Since X contains at least $\mu_1 + 1$ elements, the set $X \backslash D_{\mu_1}$ is not empty and we can pick any $x_{new} \in X \backslash D_{\mu_1}$ that is not contained in the optimal μ_1-distribution and for which $f(x_{new})$ is defined. Let $x_r := \min\{x | x \in D_{\mu_1} \cup \{r_1\}, x > x_{new}\}$ be the closest element of D_{μ_1} to the right of x_{new} (or r_1 if x_{new} is larger than all elements of D_{μ_1}). Similarly, let $f_l := \min\{r_2, \{f(x) | x \in D_{\mu_1}, x < x_{new}\}\}$ be the function value of the closest element of D_{μ_1} to the left of x_{new} (or r_2 if x_{new} is smaller than all elements of D_{μ_1}). Then, all objective vectors within $H_{new} := [x_{new}, x_r) \times [f_l, f(x_{new}))$ are (weakly) dominated by the new point $(x_{new}, f(x_{new}))$ but are not dominated by any objective vector given by D_{μ_1}. Furthermore, H_{new} is not a null set (i.e. has a strictly positive measure) since $x_{new} > x_r$ and $f_l > f(x_{new})$ which gives $\overline{I_H^{\mu_1}} < \overline{I_H^{\mu_2}}$. $\square$

The next proposition is a central result of the paper stating that the x-coordinates of a set of μ points have to necessarily satisfy a recurrence relation (Eq. 3) in order to be an optimal μ-distribution. The key idea for the derivation is that, given three consecutive points on the Pareto front, moving the middle point will only affect that hypervolume contribution that is solely dedicated to this point (the joint hypervolume contributions remain fixed). Consequently, to belong to an optimal μ-distribution, the hypervolume component solely attributed to the middle point has to be maximal.

PROPOSITION 1. **(Necessary condition for optimal μ-distributions)** *If f is continuous, differentiable and $(x_1^\mu, \ldots, x_\mu^\mu)$ denote the x-coordinates of a set of μ points maximizing the hypervolume indicator, then for all $x_{min} < x_i^\mu < x_{max}$*

$$f'(x_i^\mu)(x_{i+1}^\mu - x_i^\mu) = f(x_i^\mu) - f(x_{i-1}^\mu), \ i = 1 \ldots \mu \quad (3)$$

where f' denotes the derivative of f, $f(x_0^\mu) = r_2$ and $x_{\mu+1}^\mu = r_1$.

PROOF. The position of a point $(x_i^\mu, f(x_i^\mu))$ between its neighbors $(x_{i-1}^\mu, f(x_{i-1}^\mu))$ and $(x_{i+1}^\mu, f(x_{i+1}^\mu))$ only influences the hypervolume with respect to the objective space that is solely dominated by $(x_i^\mu, f(x_i^\mu))$. The volume of the single hypervolume contribution of the point $(x_i^\mu, f(x_i^\mu))$ is denoted as H_i (see Fig 3) and equals

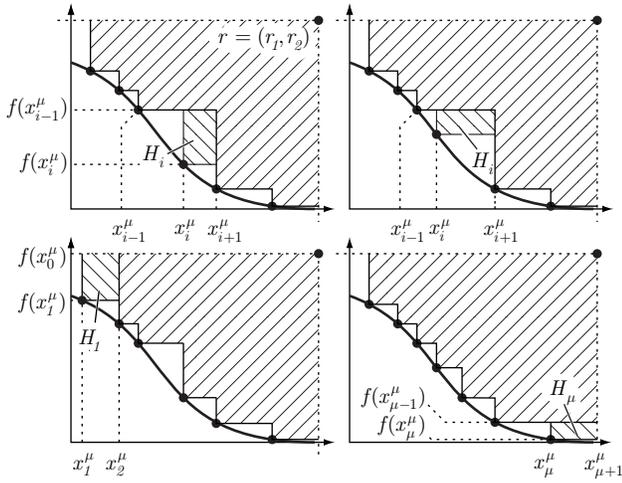$$H_i = (x_{i+1}^\mu - x_i^\mu)(f(x_{i-1}^\mu) - f(x_i^\mu)) \ .$$

Figure 3: Hypervolume contribution H_i of the point $(x_i^\mu, f(x_i^\mu))$ for various i. Please note that moving $x_i^\mu \in [x_{i-1}^\mu, x_{i+1}^\mu]$ only affects H_i but not the rest of the hypervolume (upper figures).

For an optimal μ-distribution, each of the hypervolume contributions H_i has to be maximal with respect to x_i^μ since otherwise one would be able to increase $\overline{I_H^\mu}$ by moving x_i^μ. Since the hypervolume contribution becomes zero for x_i^μ at the boundary of $[x_{i-1}^\mu, x_{i+1}^\mu]$, the maximum must lie in the interior of the domain. Therefore, the necessary condition holds that the derivative of H_i with respect to x_i^μ is zero or x_i^μ is an endpoint of f, i.e., either $x_i^\mu = x_{min}$ or $x_i^\mu = x_{max}$. The derivative of H_i with respect to x_i^μ equals

$$H_i'(x_i^\mu) = -x_{i+1}^\mu f'(x_i^\mu) - f(x_{i-1}^\mu) + f(x_i^\mu) + x_i^\mu f'(x_i^\mu)$$

Reorganizing the terms and setting $H_i'(x_i^\mu)$ to zero, we obtain Eq. 3. $\square$

The previous proposition implies that the points of optimal μ-distributions are linked with Eq. 3. In particular, the μ points are entirely determined as soon as x_1^μ and x_2^μ are set. Hence, finding the points of an optimal μ-distribution, i.e., maximizing the seemingly μ-dimensional function (2), turns out to be a two dimensional optimization problem.

A first corollary from Lemma 1 and Proposition 1 is that an optimal point is either on an extreme of the Pareto front or cannot be a stationary point of f, i.e., a point where the derivative of f equals zero.

COROLLARY 1. *If $x_i^\mu, i = 2 \ldots \mu - 1$ is a point of a set of μ points maximizing the Hypervolume indicator and x_i^μ is not an endpoint of the Pareto front, then*

$$f'(x_i^\mu) \neq 0 \ .$$

PROOF. We will prove this result by contradiction. Assume that there exists $i_0 \in \{2 \ldots, \mu - 1\}$ such that $f'(x_{i_0}^\mu) = 0$, then with Eq. 3 we have $f(x_{i_0}^\mu) = f(x_{i_0-1}^\mu)$. The strict monotonicity of f implies that $x_{i_0}^\mu = x_{i_0-1}^\mu$ and therefore $\overline{I_H^\mu} = \overline{I_H^{\mu-1}}$ which contradicts Lemma 1. $\square$

Note that since f is strictly monotone, the only possible interior stationary points are saddle points. Therefore, the previous corollary states that the points $x_2^\mu, \ldots, x_{\mu-1}^\mu$ cannot be saddle points.

4. ON THE CHOICE OF THE REFERENCE POINT

Optimal μ-distributions are the solutions of the maximization problem in Eq. 2 that depends on the choice of the reference point. We ask now the question of how the choice of the reference point is influencing optimal μ-distributions and investigate in particular whether there exists a choice of the reference point that implies that the extremes of the Pareto front are included in optimal μ-distributions. We prove in Theorem 2 (resp. Theorem 3) that if the derivative of the Pareto front at the left extreme (resp. right extreme) is infinite (resp. is zero), there is no choice of reference point that will allow that the left (resp. right) extreme of the front is included in optimal μ-distributions. This result contradicts the common belief that it is sufficient to choose the reference point slightly above and to the right to the nadir point to obtain the extremes.

Moreover, when the derivative is finite at the left extreme and non-zero at the right extreme we prove an explicit lower bound (possibly infinite) for the choice of the reference point ensuring that any reference point above this lower bound guarantees that the extremes of the front are included in optimal μ-distributions.

Recall that $r = (r_1, r_2)$ denotes the reference point and $y = f(x)$ with $x \in [x_{min}, x_{max}]$ represents the Pareto front where therefore $(x_{min}, f(x_{min}))$ and $(x_{max}, f(x_{max}))$ are the left and right extremal points. Since we want that all Pareto-optimal solutions have a contribution to the hypervolume of the front in order to be possibly part of optimal μ-distributions, we assume that the reference point is dominated by all Pareto-optimal solutions, i.e. $r_1 \geq x_{max}$ and $r_2 \geq f(x_{min})$.

THEOREM 2. *Let μ be an integer larger or equal 2. Assume that f is continuous on $[x_{min}, x_{max}]$, non-increasing, differentiable on $]x_{min}, x_{max}[$ and that f' is continuous on $]x_{min}, x_{max}[$. If $\lim_{x \to x_{min}} -f'(x) < +\infty$, let*

$$\mathcal{R}_2 := \sup_{x \in [x_{min}, x_{max}[} \{f'(x)(x - x_{max}) + f(x)\} \ , \quad (4)$$

where the supremum in the previous equation is possibly infinite. When $\mathcal{R}_2$ is finite, the leftmost extremal point is contained in optimal μ-distributions if the reference point $r = (r_1, r_2)$ is such that r_2 is strictly larger than $\mathcal{R}_2$.

Moreover, if $\lim_{x \to x_{min}} -f'(x) = +\infty$, the left extremal point of the front is never included in optimal μ-distributions.

For the sake of readability of the section, the proof of the above theorem, as well as the following proof are in the appendix.

In a similar way, we derive the following theorem for including the rightmost Pareto-optimal point $(x_{max}, f(x_{max}))$ into optimal μ-distributions.

THEOREM 3. *Let μ be an integer larger or equal 2. Assume that f is continuous on $[x_{min}, x_{max}]$, non-increasing, differentiable on $]x_{min}, x_{max}[$ and that f' is continuous and strictly negative on $]x_{min}, x_{max}]$. Let*

$$\mathcal{R}_1 := \sup_{x \in]x_{min}, x_{max}]} \left\{ x + \frac{f(x) - f(x_{min})}{f'(x)} \right\} \ , \quad (5)$$

where the supremum in the previous equation is possibly infinite. When $\mathcal{R}_1$ is finite, the rightmost extremal point is contained in optimal μ-distributions if the reference point $r = (r_1, r_2)$ is such that r_1 is strictly larger than $\mathcal{R}_1$.

problem	r_1	r_2
ZDT1, ZDT4	3	$+\infty$
ZDT2	3/2	4/3
ZDT3	$+\infty$	5.4460
ZDT 6	$\frac{3}{2} - \frac{(\arctan(9\,\pi))^2}{72\pi^2} \approx 1.497$	4/3
DTLZ1	1	1
DTLZ2, 3, and 4	$\frac{3^{3/4}(\sqrt{3}-1)}{\sqrt{2}} \approx 1.180$	$\frac{3^{3/4}(\sqrt{3}-1)}{\sqrt{2}} \approx 1.180$
DTLZ7	≈ 13.372	≈ 5.3054

Table 1: The values indicated in column "r_1" (resp. "r_2") correspond to the bound in Theorem 2 and 3 for ZDT [24] and DTLZ [8] test problems. Choosing the first (resp. second) coordinate of the reference point strictly larger than this value ensures that the left (resp. right) extreme is included in optimal μ-distributions. When a value is infinite it means that there is no choice of the reference point allowing to include the extreme. Note that ZDT5 (discrete problem) as well as DTLZ5 and DTLZ6 (bi-objective Pareto front contains only one point) are not shown here.

If $f'(x_{max}) = 0$, the right extremal point is never included in optimal μ-distributions.

In the following example we illustrate how Theorem 2 and 3 can be used on test problems.

EXAMPLE 1   (ZDT1 [24]). For $f(x) = 1 - \sqrt{x}$ and $x_{min} = 0, x_{max} = 1$, $f'(x) = -\frac{1}{2\sqrt{x}}$ and therefore

$$\lim_{x \to x_{min}} -f'(x) = +\infty \ .$$

According to Theorem 2, the leftmost Pareto-optimal point is never included in optimal μ-distributions. In addition, we have $f'(x) < 0$ for all $x \in]0,1]$. Let us compute $\mathcal{R}_1$ defined in Eq. 5:

$$\mathcal{R}_1 = \sup_{x \in]0,1]} \{x + \frac{1 - \sqrt{x} - 1}{-1/2\sqrt{x}}\} = \sup_{x \in]0,1]} \{x + 2\} = 3 \ .$$

From Theorem 3, we therefore know that the right extreme is included if $r_1 > 3$.

Table 1 shows the results also for other test problems.

5. EXACT DISTRIBUTION FOR LINEAR FRONTS

In this section, we have a closer look at linear Pareto fronts, i.e., fronts pictured as straight lines that can be formally defined as $f : x \in [x_{min}, x_{max}] \mapsto \alpha x + \beta$ where $\alpha < 0$ and $\beta \in \mathbb{R}$. For linear fronts with slope $\alpha = -1$, Beume et al. [2] already proved that a set of μ points maximizes the hypervolume if and only if the points are equally spaced. However, their method does not allow to state where the leftmost and rightmost points have to be placed in order to maximize the hypervolume with respect to a certain reference point; furthermore, the approach cannot be generalized to arbitrary linear fronts with other slopes than -1. The same result of equal distances between points that maximize the hypervolume has been shown with a different technique in [10] for the front $f(x) = 1 - x$. Although the proof
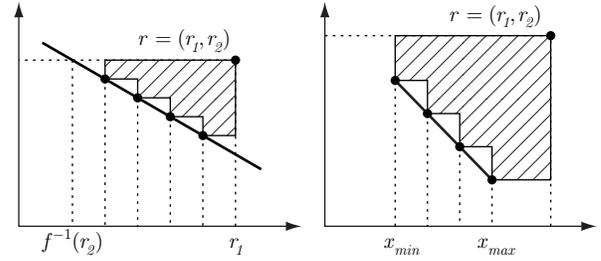

Figure 4: Optimal μ-distribution for $\mu = 4$ points if the reference point is not dominated by the extreme points of the Pareto front (Theorem 5, left) and if the reference point ensures that the extreme points are included in the optimal μ-distribution (Theorem 6, right).

technique used in [10] could be generalized to arbitrary linear fronts, the provided result again only holds under the assumption that both the leftmost and rightmost point is fixed. Therefore, the question of where μ points have to be placed on linear fronts to maximize the hypervolume indicator without any assumption on the extreme points is still not answered.

Within this section, we show for linear fronts of arbitrary slope, how optimal μ-distributions look like while making no assumptions on the positions of extreme solutions. First of all, we see as a direct consequence of Proposition 1 that the distance between two neighbored solutions is constant for arbitrary linear fronts:

THEOREM 4. If the Pareto front is a (connected) line, optimal μ-distributions are such that the distance is the same between all neighbored solutions.

PROOF. Applying Eq. 3 to $f(x) = \alpha x + \beta$ implies that

$$\alpha \left(x_{i+1}^\mu - x_i^\mu \right) = f(x_i^\mu) - f(x_{i-1}^\mu) = \alpha(x_i^\mu - x_{i-1}^\mu)$$

for $i = 2, \ldots, \mu - 1$ and therefore the distance between consecutive points of optimal μ-distributions is constant. □

Moreover, in case the reference point is not dominated by the extreme points of the Pareto front, i.e., $r_1 < x_{max}$ and r_2 is set such that there exists (a unique) $x_0^\mu \in [x_{min}, x_{max}]$ with $x_0^\mu = f^{-1}(r_2)$, there exists a unique optimal μ-distribution that can be determined exactly, see also the left plot of Fig. 4:

THEOREM 5. If the Pareto front is a (connected) line and the reference point (r_1, r_2) is not dominated by the extremes of the Pareto front, there exists a unique optimal μ-distribution satisfying for all $i = 1, \ldots, \mu$

$$x_i^\mu = f^{-1}(r_2) + \frac{i}{\mu + 1} \cdot (r_1 - f^{-1}(r_2)) \ . \qquad (6)$$

PROOF. From Eq. 3 and the previous proof we know that

$$\alpha \left(x_{i+1}^\mu - x_i^\mu \right) = f(x_i^\mu) - f(x_{i-1}^\mu) = \alpha(x_i^\mu - x_{i-1}^\mu) \ ,$$

for $i = 1, \ldots, \mu$ if we define $f(x_0^\mu) = r_2$ and $x_{\mu+1}^\mu = r_1$ as in Proposition 1; in other words, the distances between x_i and its two neighbors x_{i-1} and x_{i+1} are the same for each $1 \leq i \leq \mu$. Therefore, the points $(x_i^\mu)_{1 \leq i \leq \mu}$ partition the interval $[x_0^\mu, x_{\mu+1}^\mu]$ into $\mu + 1$ sections of equal size and we obtain Eq. 6. □

In the light of Section 4, the next issue we investigate is the choice of the reference point ensuring that optimal μ-distributions contain the extremes of the front. Since $f'(x) = \alpha$ with $0 < \alpha < +\infty$, we see that $\mathcal{R}_2$ and $\mathcal{R}_1$ defined in Theorems 2 and 3 are finite and thus we now that there exists a choice of reference point ensuring to obtain the extremes. The following theorem states the corresponding lower bounds for the reference point and specifies the optimal μ-distribution associated with such a choice. The right plot in Fig. 4 illustrates the optimal μ-distribution for $\mu = 4$ points.

THEOREM 6. *If the Pareto front is a (connected) line $x \in [x_{min}, x_{max}] \mapsto \alpha x + \beta$ with $\alpha < 0$, $\beta \in \mathbb{R}$ and the reference point $r = (r_1, r_2)$ is such that r_1 is strictly larger than $2x_{max} - x_{min}$ and r_2 is strictly larger than $2\alpha x_{min} - \alpha x_{max} + \beta$, there exists a unique optimal μ-distribution. This optimal μ-distribution includes the extremes of the front and for all $i = 1, \ldots, \mu$*

$$x_i^{\mu} = x_{min} + \frac{i-1}{\mu-1}\left(x_{max} - x_{min}\right) \quad . \tag{7}$$

PROOF. From Eq. 4, r_2 strictly larger than

$$\sup_{x \in [x_{min}, x_{max}[}\{\alpha(x - x_{max}) + \alpha x + \beta\},$$

i.e., r_2 strictly larger than $2\alpha x_{min} - \alpha x_{max} + \beta$, ensures the presence of x_{min} in optimal μ-distributions which yields $x_i^{\mu} = x_{min}$ for $i = 1$. From Eq. 5, r_1 strictly larger than $\sup_{x \in]x_{min}, x_{max}]} x + \frac{\alpha x + \beta - \alpha x_{min} - \beta}{\alpha}$, i.e., r_1 strictly larger than $\sup_{x \in]x_{min}, x_{max}]} 2x - x_{min} = 2x_{max} - x_{min}$, ensures the presence of x_{max} in optimal μ-distributions and thus $x_i^{\mu} = x_{max}$ for $i = \mu$. Using the same argument as in Theorem 5, we prove the uniqueness of optimal μ-distributions, the equal distances between the points and therefore Eq. 7. $\square$

As pointed out in the beginning, we do not know in general whether optimal μ-distributions are unique or not. Theorem 5 and 6 are two settings where we can ensure the uniqueness.

6. OPTIMAL DENSITY

Except for simple fronts like the linear one, it is difficult (presumably impossible) to determine precisely optimal μ-distributions. However, in this section, we determine the distributions of points on arbitrary fronts, when their number goes to infinity. These distributions are going to be characterized in terms of density that can be used to approximate, for a finite μ, the percentage of points in a particular segment of the front. Moreover, it allows to quantify in a rigorous manner the bias of the hypervolume with respect to a given front. Our main result, stated in Theorem 7, is that the density is proportional to the square root of the negative of the derivative of the front.

Although the results in this section hold for arbitrary Pareto front shapes that can be described by a continuous function $g : x \in [x_{min}, x_{max}] \mapsto g(x)$ (lefthand plot of Fig. 5), we will, without loss of generality, consider only fronts $f : x \in [0, x_{max}] \mapsto f(x)$ with $f(x_{max}) = 0$. The reason is that an arbitrary front shape $g(x)$ can be easily transformed into the latter type by translating it, i.e., by introducing a new coordinate system $x' = x - x_{min}$ and $y' = y - g(x_{max})$ with $y' = f(x')$ describing the front in the new coordinate system (righthand plot of Fig. 5).
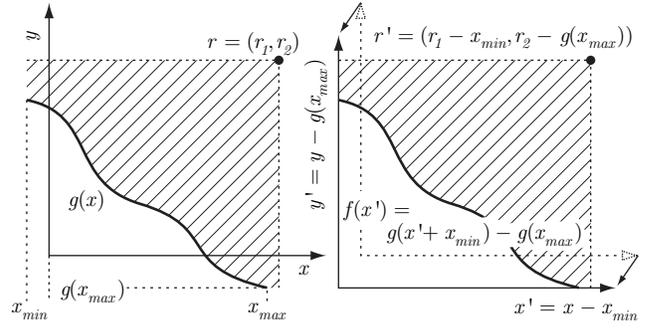

Figure 5: Every continuous Pareto front $g(x)$ (left) can be described by a function $f : x' \in [0, x'_{max}] \mapsto f(x')$ with $f(x'_{max}) = 0$ (right) by a simple translation.

This translation is not affecting the hypervolume indicator which is computed relatively to the (also translated) reference point $r' = (r_1 - x_{min}, r_2 - g(x_{max}))$.

In addition to assuming f to be continuous within the entire domain $[0, x_{max}]$, we assume that f is differentiable and that its derivative is a continuous function f' defined in the interval $]0, x_{max}[$. For a certain number of points μ, which later on will go to infinity, we would like to compute optimal μ-distributions, i.e., the positions of the μ points $(x_1^{\mu}, f(x_1^{\mu})), \ldots, (x_{\mu}^{\mu}, f(x_{\mu}^{\mu}))$ such that the hypervolume indicator $I_H^{\mu}((x_1^{\mu}, \ldots, x_{\mu}^{\mu}))$ is maximized, see the upper left plot in Fig. 6. Instead of maximizing the hypervolume indicator I_H^{μ}, it is easy to see that, since $r_1 r_2$ is constant, one can equivalently minimize

$$r_1 r_2 - I_H^{\mu}((x_1^{\mu}, \ldots, x_{\mu}^{\mu})) = \sum_{i=0}^{\mu}\left(x_{i+1}^{\mu} - x_i^{\mu}\right) f(x_i^{\mu})$$

with $x_0^{\mu} = 0$, $f(x_0^{\mu}) = r_2$, and $x_{\mu+1}^{\mu} = r_1$ as it is illustrated in the upper right plot of Fig. 6. If we subtract the area below the front curve, i.e., the integral $\int_0^{x_{max}} f(x)dx$ of constant value (lower left plot in Fig. 6), we see that minimizing

$$\sum_{i=0}^{\mu}\left(x_{i+1}^{\mu} - x_i^{\mu}\right) f(x_i^{\mu}) - \int_0^{x_{max}} f(x)dx \tag{8}$$

is equivalent to maximizing the hypervolume indicator $I_H^{\mu}((x_1^{\mu}, \ldots, x_{\mu}^{\mu}))$, cf. the lower right plot of Fig. 6.

For any integer μ, we now consider a sequence of μ ordered points in $[0, x_{max}]$, $x_1^{\mu}, \ldots, x_{\mu}^{\mu}$ that lie on the Pareto front. We assume that the sequence converges—when μ goes to ∞—to a density $\delta(x)$ that is regular enough. Formally, the density in $x \in [0, x_{max}]$ is defined as the limit of the number of points contained in a small interval $[x, x + h[$ normalized by the total number of points μ when both μ goes to ∞ and h to 0, i.e.,

$$\delta(x) = \lim_{\substack{\mu \to \infty \\ h \to 0}}\left(\frac{1}{\mu h}\sum_{i=1}^{\mu}\mathbf{1}_{[x, x+h[}(x_i^{\mu})\right) \quad . \tag{9}$$

In the following, we want to characterize the density associated with points $x_1^{\mu}, \ldots, x_{\mu}^{\mu}$ that are maximizing the hypervolume indicator. As discussed above, maximizing the hypervolume is equivalent to minimizing Eq. 8. We here
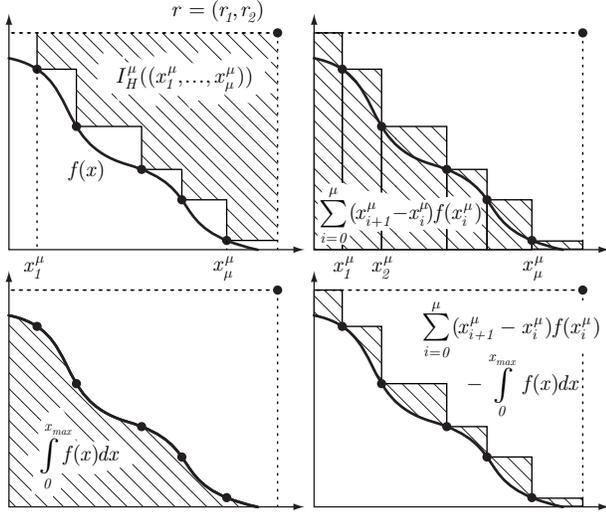
Figure 6: Illustration of the idea behind deriving the optimal density: Instead of maximizing the hypervolume indicator $I_H^\mu((x_1^\mu, \ldots, x_\mu^\mu))$ (upper left), one can minimize $r_1 r_2 - I_H$ (upper right) which is equivalent to minimize the integral between the attainment surface of the solution set and the front itself (lower right) which can be expressed with the help of the integral of f (lower left).

argue that for a fixed number of points μ, minimizing

$$E_\mu = \mu \left(\sum_{i=0}^{\mu} (x_{i+1}^\mu - x_i^\mu)(f(x_i^\mu)) - \int_0^{x_{max}} f(x)dx \right) \quad (10)$$

is also equivalent to maximizing the hypervolume such that we conjecture that the equivalence between minimizing E_μ and maximizing the hypervolume also holds for μ going to infinity.

We now heuristically deduce[2] that the limit density of Eq. 9 will minimize the limit of E_μ in Eq. 10. Therefore, our proof consists of two steps: (1) compute the limit of E_μ when μ goes to ∞. This limit is going to be a function of a density δ. (2) Find the density δ that minimizes $E(\delta) := \lim_{\mu \to \infty} E_\mu$. The first step therefore consists in computing the limit of E_μ. Let us first recall some definitions from integration theory. A function $g : [0, x_{max}] \to \mathbb{R}$ is said to be integrable if $\int_0^{x_{max}} |g(x)|dx$ is finite (usually denoted $\int_0^{x_{max}} |g(x)|dx < +\infty$). The set of functions g that are integrable is a (Banach) vector space denoted $L^1(0, x_{max})$. Another Banach vector space is the set of functions whose square is in $L^1(0, x_{max})$, this space is denoted $L^2(0, x_{max})$. From the Cauchy-Schwarz inequality, $L^2(0, x_{max})$ is included in $L^1(0, x_{max})$[3].

LEMMA 2. *If f is continuous, differentiable with the derivative f' continuous, if $x_1^\mu, \ldots, x_\mu^\mu$ converge to a continuous*

---

[2]"Heuristically deduce" means, we cannot prove it completely rigorously. Doing so would involve functional analysis concepts that are far beyond the scope of this paper.

[3]The Cauchy-Schwarz inequality states that $(\int_0^{x_{max}} |fg|)^2 \le (\int_0^{x_{max}} f^2)(\int_0^{x_{max}} g^2)$ and therefore setting $g = 1$, $(\int_0^{x_{max}} |f|)^2 \le (\int_0^{x_{max}} f^2)$, i.e. if $f \in L^2(0, x_{max})$ then $f \in L^1(0, x_{max})$.

density δ, with $\frac{1}{\delta} \in L^2(0, x_{max})$, and $\exists c \in \mathbb{R}^+$ such that

$$\mu \sup(\sup_{0 \le i \le \mu-1} |x_{i+1}^\mu - x_i^\mu|, |x_{max} - x_\mu^\mu|) \to c \quad (11)$$

then E_μ converges for $\mu \to \infty$ to

$$E(\delta) := -\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)}{\delta(x)} dx \quad . \quad (12)$$

For the sake of readability of the section, the proof of the previous lemma, as well as the following have been sent in the appendix.

Note that the assumption in Eq. 11 characterizes the convergence of the μ points to the density and is needed in the proof.

As explained before, the limit density of an optimal μ-distribution is minimizing $E(\delta)$. It remains therefore to find the density minimizing $E(\delta)$. This optimization problem is posed in a functional space, the Banach space $L^2(0, x_{max})$ and is also a constraint problem since the density δ has to satisfy the constraint

$$J(\delta) := \int_0^{x_{max}} \delta(x)dx = 1 \quad .$$

The constraint optimization problem (P) that needs to be solved is summarized in:

$$\begin{aligned} & \text{minimize } E(\delta), \delta \in L^2(0, x_{max}) \\ & \text{subject to } J(\delta) = 1 \end{aligned} \quad \text{(P)}$$

THEOREM 7. *The limit density of points maximizing the hypervolume is a solution of the constraint optimization problem (P) and equals*

$$\delta(x) = \frac{\sqrt{-f'(x)}}{\int_0^{x_{max}} \sqrt{-f'(x)}dx} \quad .$$

REMARK 1. *In order to get the density δ_F for points of the front $(x, f(x))_{x \in [0, x_{max}]}$ and not on the projection onto the x-axis, one has to normalize the previous result by the norm of the tangent for points of the front, i.e., $\sqrt{1 + f'(x)^2}$. Therefore the density on the front is proportional to:*

$$\delta_F(x) \propto \frac{\sqrt{-f'(x)}}{\sqrt{1 + f'(x)^2}} \quad (13)$$

As we have seen, the density follows as a limit result from the fact that the area between the attainment function of the solution set with μ points and the front itself (lower right plot of Fig. 6) has to be minimized and an optimal μ-distribution for finite points converges to the density when μ increases. It also follows that the number of points of an optimal μ-distribution with x-values in a certain interval $[a, b]$ converges to $\int_a^b \delta(x)dx$ if μ goes to infinity. In the next section, we will show experimentally that the density can be used as an approximation of optimal μ-distributions not only for a large μ but also for reasonably small numbers of points.

Besides plotting the density to understand the bias of the hypervolume indicator for specific fronts, the results above also allow a more general statement on the hypervolume indicator. From Theorem 7, we know that the density of points only depends on the slope of the front and not on whether the front is convex or concave in contrast to prevalent belief [26, 18]. Figure 7 illustrates this dependency between the
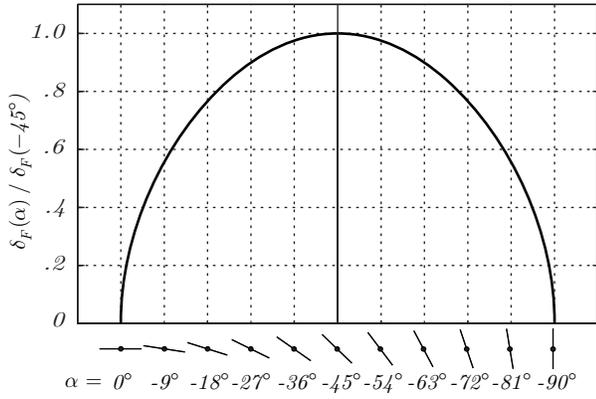
Figure 7: Shows the density at different slopes of the Pareto front according to Eq. 13, expressed as the angle $\alpha = \arctan(f'(x))$ the front makes with the positive x-axis, . Note that the density is normalized such that $\delta_F(-45°) = 1$.

density and the slope of the front. We observe that the density has its maximum for front parts where the tangent has a gradient of -45°. Therefore, optimizing the hypervolume indicator stresses so-called knee-points—parts of the Pareto front decision makers believe to be interesting regions [5, 4, 6]. Extreme points are not generally preferred as claimed in [7], since the density of points does not depend on the position on the front but only on the gradient at the respective point. The gradient varies for different locations on the front except for the linear case, so does the spacing between points. Hence, the distributions of points can differ highly from a well-distributed set of points suggested in [16].

7. APPROXIMATING OPTIMAL μ-DISTRIBUTIONS

Deriving the density of points allowed us to investigate the bias of the hypervolume in general as we have seen in the previous section, but the question remains what happens for a finite, especially small number of points. To answer this question, we carry out an experimental study using two local search strategies for approximating optimal μ-distributions for finite μ. These search strategies serve to (i) investigate optimal μ-distributions for small μ on various well-known test problems and (ii) show that optimal μ-distributions can be approximated accurately enough by the density already for reasonably small μ.

Algorithms for finding optimal μ-distributions. To find an approximation of an optimal distribution of μ points, given by their x-values $(x_1^\mu, \ldots, x_\mu^\mu)$, we propose a simple hill climber (Algorithm 1) based on Eq. 3. It starts with an initial distribution that follows the distribution function of Theorem 7 (Line 2). Such a distribution is obtained by

$$(x_1^\mu, \ldots, x_\mu^\mu) \text{ s.t. } \int_{x_{min}}^{x_i^\mu} \delta(x)dx = \frac{i - 0.5}{\mu} \qquad (14)$$

which ensures that the cumulated density between two adjacent points is constant. Additionally, the two points $x_0^\mu = x_{min}$ with $f(x_0^\mu) = r_2$ and $x_{\mu+1}^\mu = r_1$ are added to simplify

---

Algorithm 1 Optimal μ-Distribution Version 1

1: **procedure** DISTRIBUTEV1$(\mu, (x_{min}, x_{max}), (r_1, r_2), f)$
2:     $(x_0^\mu, \ldots, x_{\mu+1}^\mu) \leftarrow InitialDist(\mu, (x_{min}, x_{max}), f)$
3:     $v \leftarrow CalculateHypervolume((x_0^\mu, \ldots, x_{\mu+1}^\mu), f)$
4:     $v_{old} \leftarrow -1$
5:     **while** $v - v_{old} > \varepsilon$ **do**
6:         **for** $i \leftarrow 1, \mu$ **do**
7:             $x_i^\mu \leftarrow OptimalPosition(x_{i-1}^\mu, x_{i+1}^\mu, f)$
8:         **end for**
9:         $v_{old} \leftarrow v$
10:        $v \leftarrow CalculateHypervolume((x_0^\mu, \ldots, x_{\mu+1}^\mu), f)$
11:    **end while**
12:    **return** $(x_1^\mu, \ldots, x_\mu^\mu)$
13: **end procedure**

---

the handling of the extreme points. The resulting distribution of points follows the density function of Theorem 7 with all points in order from left to right, i.e., $x_i^\mu < x_j^\mu$ $\forall\, 0 \leq i < j \leq \mu + 1$.

After the points have been initialized, the contribution to the hypervolume is maximized for all points successively by placing each point according to Eq. 3 (Lines 6 to 8). This ensures that either the hypervolume is increased in the current step or the corresponding point is already placed optimally with respect to its neighbors.

Since changing the position of one point might change other points that were previously considered as optimally placed to be suboptimally placed, the procedure is repeated as long as the improvement to the hypervolume indicator is larger than a user defined threshold ε (usually the precision of software implementation).

Unfortunately, the optimal position of x_i^μ according to Eq. 3 cannot be determined analytically for some test problems, e.g., DTLZ2 [8]. In those cases, a modified version of the hill climbing procedure is used (see Algorithm 2). After creating an initial distribution of points as in Algorithm 1 (Line 2), the position of each point is again modified separately one after another. But in contrast to Algorithm 1, a new position $x_i'^\mu$ is randomly determined by adding a Gaussian distributed value centered around the previous value. Initially, the variance σ^2 of this random variable is set to a large value such that big changes are attempted. For discontinuous front shapes (like DTLZ7 [8]) the position of the point is alternatively set to any value between x_{min} and x_{max}. This enables jumps of points to different parts of the front even in the later stage of the algorithms. The x-value x_i^μ is set to the new position $x_i'^\mu$ and the current hypervolume v_{old} is updated to v only if the hypervolume indicator increases; otherwise, the modified position is discarded (Lines 16 to 20).

If for no point an improvement of the hypervolume could be realized, the variance σ^2 is decreased by 5%. As soon as the variance is smaller than ε (defined as in Algorithm 1), the current distribution $(x_1^\mu, \ldots, x_\mu^\mu)$ is returned.

Note that both presented algorithms do not guarantee to find optimal μ-distributions. However, the experiments presented in the following show that the point distributions found by the two algorithms are converging to the density if μ goes to infinity like optimal μ-distributions theoretically converges to the density. Testing the algorithms using multiple random initial distributions instead of according to Eq. 14 within the $InitialDist(\cdot)$ function in both algorithms

---
Algorithm 2 Optimal μ-Distribution Version 2

1: **procedure** DISTRIBUTEV2$(\mu, (x_{min}, x_{max}), (r_1, r_2), f)$
2:    $(x_0^\mu, \ldots, x_{\mu+1}^\mu) \leftarrow InitialDist(\mu, (x_{min}, x_{max}), f)$
3:    $\sigma \leftarrow x_{max} - x_{min}$
4:    $v_{old} \leftarrow CalculateHypervolume((x_0^\mu, \ldots, x_{\mu+1}^\mu), f)$
5:    **while** $\sigma > \varepsilon$ **do**
6:       $improvement \leftarrow$ false
7:       **for** $i \leftarrow 1, \mu$ **do**
8:          $(x'^\mu_0, \ldots, x'^\mu_{\mu+1}) \leftarrow (x_0^\mu, \ldots, x_{\mu+1}^\mu)$
9:          $J \leftarrow$ either 0 or 1 (with prob. 1/2 each)
10:         **if** J is 1 **then**
11:            $x'^\mu_i \leftarrow N(x'^\mu_i, \sigma^2)$ $\triangleright$ offset by GRV
12:         **else**
13:            $x'^\mu_i \leftarrow U(x_{min}, x_{max})$ $\triangleright$ Jump
14:         **end if**
15:         $v \leftarrow CalculateHypervolume((x_0^\mu, \ldots, x_{\mu+1}^\mu), f)$
16:         **if** $v > v_{old}$ **then**
17:            $v_{old} \leftarrow v$
18:            $(x_0^\mu, \ldots, x_{\mu+1}^\mu) \leftarrow (x'^\mu_0, \ldots, x'^\mu_{\mu+1})$
19:            $improvement \leftarrow$ true
20:         **end if**
21:       **end for**
22:       **if** $improvement$ is false **then**
23:         $\sigma \leftarrow 0.95 \cdot \sigma$
24:       **end if**
25:    **end while**
26:    **return** $(x_1^\mu, \ldots, x_\mu^\mu)$
27: **end procedure**
---

always lead to the same final distribution, although the convergence turned out to be slower. This is a strong indicator that the distributions found are indeed good approximations of the optimal distributions of μ points.

We would also like to mention that every general purpose search heuristic, e.g., an evolution strategy, could be used for approximating optimal μ-distributions. In preliminary experiments for example, we employed the CMA-ES of [12] to derive optimal μ-distributions. It turned out that the problem-specific algorithms are much faster than the CMA-ES. Even when using the representation as a two-dimensional problem (see Sec. 3), the CMA-ES needed significantly more time than Algorithms 1 and 2 to derive similar hypervolume values. Nevertheless, the distributions found by CMA-ES where consistent with the distributions found by Algorithm 1 and Algorithm 2.

Approximating optimal μ-distributions. With the help of the algorithms proposed above, we now investigate optimal μ-distributions for concrete test problems where the Pareto fronts are known. Figure 8 shows the best found μ-distribution for the test problems of the ZDT [24] and DTLZ [8] test function suites exemplary for $\mu = 50$—more results can be found on the supplementary web page `http://www.tik.ee.ethz.ch/sop/muDistributions`. Furthermore, Table 2 provides the corresponding front shapes and derived densities. The optimal μ-distributions for ZDT1, ZDT2, and DTLZ1 have been approximated by Algorithm 1 whereas the μ-distributions for ZDT3, DTLZ2 and DTLZ7 have been computed with Algorithm 2 because Eq. 3 cannot be solved analytically or the front is not continuous. The reference point has been set to (15,15) such that the extreme points are contained in optimal μ-distributions if possible.

The experimentally derived optimal μ-distributions for $\mu = 50$ qualitatively show the same results as the theoretically predicted density: more points can be found at front regions that have a gradient of $-45°$, front parts that have a very high or very low gradient are less crowded. In addition, the equi-distance results for linear fronts (Theorems 5 and 6) can be observed for the linear DTLZ1 front.

Convergence to the density. From Section 6 we know that the empirical density associated with optimal μ-distributions, i.e., the normalized histogram, converges to the density when μ goes to infinity and therefore that the limit density of Theorem 7 approximates the normalized histogram. Here, we investigate experimentally the quality of the approximation.

To this end, we compute approximations of the optimal μ-distribution exemplary for the ZDT2 problem for $\mu = 10$, $\mu = 100$, and $\mu = 1,000$ obtained with Algorithm 1. The reference point has been set to $(15, 15)$ as before. Figure 9 shows both the experimentally observed histogram of the μ points on the front and the comparison between the theoretically derived density and the obtained experimental approximation thereof. By visual inspection, we see that the histogram associated with the found μ-distributions converges quickly to the density and that for $\mu = 1,000$ points, the theoretically derived density gives already a sufficient description of the finite optimal μ-distribution.

8. DISCUSSION AND CONCLUSIONS

This study provides rigorous results on the question of how optimal Pareto front approximations of finite size μ look like when the hypervolume indicator is maximized. Most surprising might be the fact that the hypervolume indicator is insensitive to the way the front is bend (convex or concave) which contradicts previous assumptions [27, 18]. As we show, it is not the front shape itself but only the slope of the front that determines the distribution of points on the front. This implies that when optimizing the standard hypervolume indicator, an evenly distributed solution set can be obtained if and only if the front shape is a straight line.

Furthermore, the question of how to choose the reference point in order to obtain extremes of the front, remaining unsolved for several years, can now be answered by our theoretical results. The explicit lower bound provided in our theorems will hopefully help practitioners.

Although the results presented here hold for two objectives only, we assume that they generalize to an arbitrary number of objectives. The extension of the proposed mathematical framework is the subject of future research.

Acknowledgments

9. REFERENCES

[1] J. Bader and E. Zitzler. HypE: Fast Hypervolume-Based Multiobjective Search Using Monte Carlo Sampling. TIK Report 286, Computer
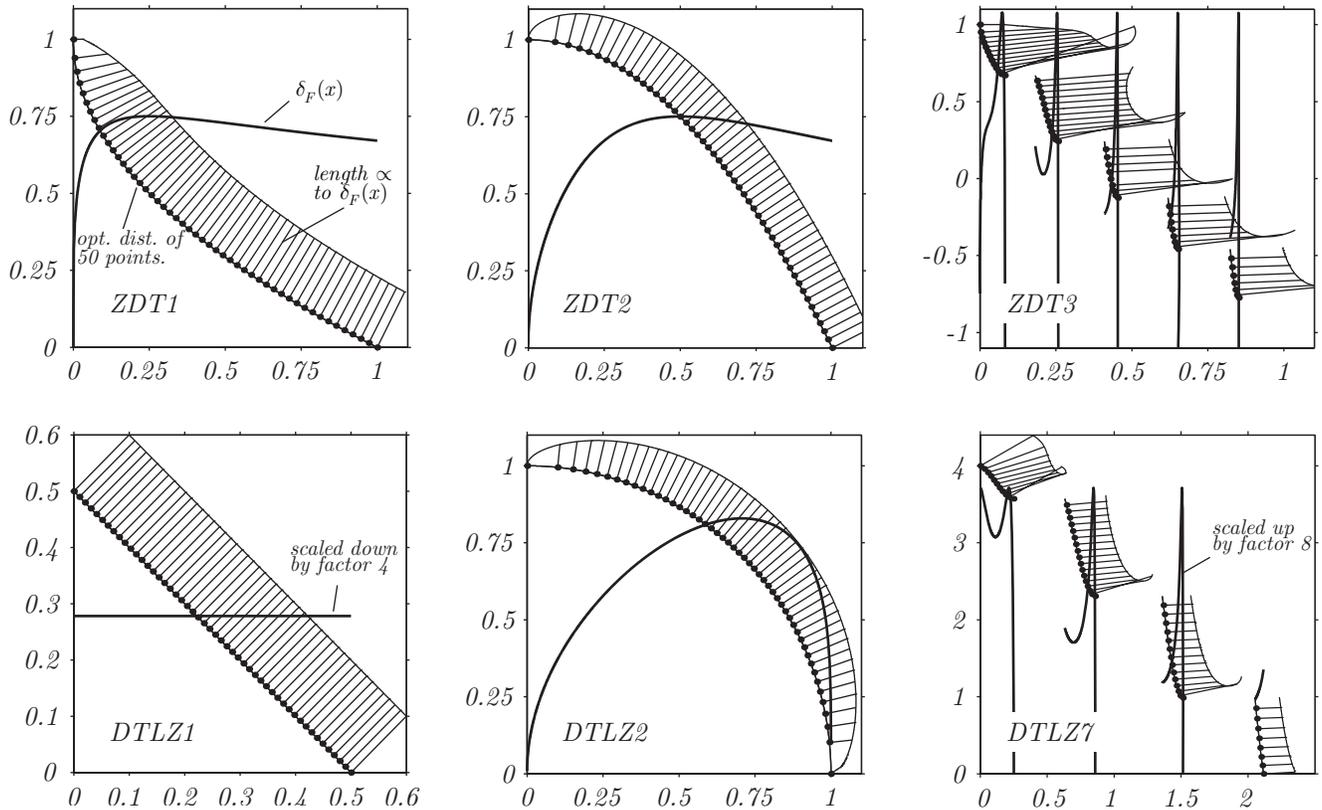
Figure 8: Density for six front shapes of the ZDT and DTLZ test problem suites. In addition to the optimal μ-distributions for a finite $\mu = 50$, the density is plotted in two different ways. On the one hand, the density is shown as a function of the first objective ($\delta_F(x)$) and, on the other hand, the density is plotted with respect to the front itself: starting from the approximation of the optimal finite set of points on the front, lines perpendicular to the front indicate the value of the density which is proportional to the lines' lengths.

Engineering and Networks Laboratory, ETH Zurich, Nov. 2008.

[2] N. Beume, C. M. Fonseca, M. Lopez-Ibanez, L. Paquete, and J. Vahrenhold. On the Complexity of Computing the Hypervolume Indicator. Technical Report CI-235/07, University of Dortmund, Dec. 2007.

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

[4] J. Branke, K. Deb, H. Dierolf, and M. Osswald. Finding Knees in Multi-objective Optimization. In *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, pages 722–731. Springer, 2004.

[5] I. Das. On Characterizing the "Knee" of the Pareto Curve Based on Normal-Boundary Intersection. *Structural and Multidisciplinary Optimization*, 18(2–3):107–115, 1999.

[6] K. Deb. Current Trends in Evolutionary Multi-Objective Optimization. *International Journal for Simulation and Multidisciplinary Design Optimization*, 1:1–8, 2007.

[7] K. Deb, M. Mohan, and S. Mishra. Evaluating the ε-Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of

Pareto-Optimal Solutions. *Evolutionary Computation*, 13(4):501–525, 2005.

[8] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable Test Problems for Evolutionary Multi-Objective Optimization. In A. Abraham, R. Jain, and R. Goldberg, editors, *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, chapter 6, pages 105–145. Springer, 2005.

[9] M. Emmerich, N. Beume, and B. Naujoks. An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, pages 62–76. Springer, 2005.

[10] M. Emmerich, A. Deutz, and N. Beume. Gradient-Based/Evolutionary Relay Hybrid for Computing Pareto Front Approximations Maximizing the S-Metric. In *Hybrid Metaheuristics*, pages 140–156. Springer, 2007.

[11] M. Fleischer. The measure of Pareto optima. Applications to multi-objective metaheuristics. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 519–533. Springer, 2003.

[12] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In
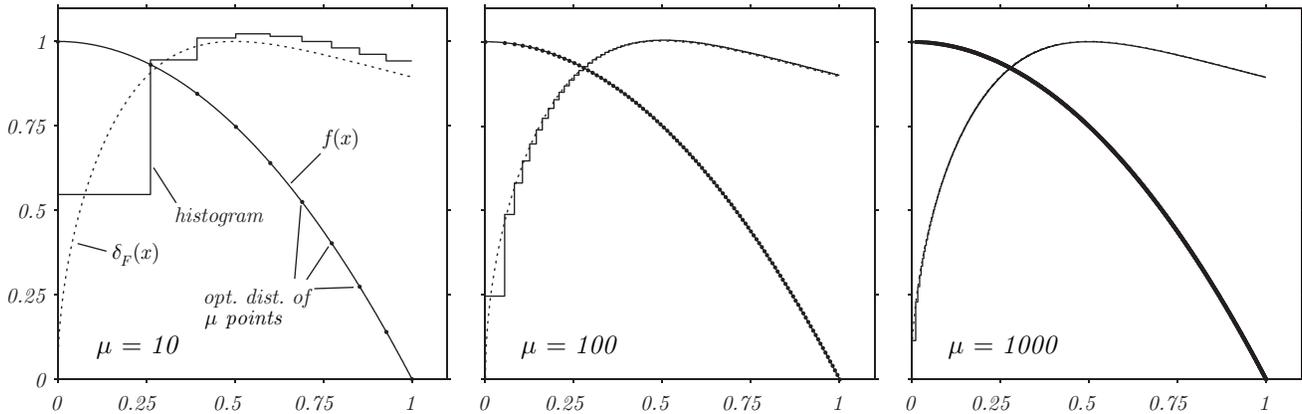
Figure 9: Comparison between the experimental density of points (shown as dots on the front and as a step function to compare with the theoretical density) and the theoretical prediction ($\delta_F(x)$) (dashed line) of $\mu = 10$ points (left), $\mu = 100$ (middle) and $\mu = 1000$ (right) on the 2-objective ZDT2 problem.

X. Y. et al., editor, *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, volume 3242 of *LNCS*, pages 282–291, Berlin, Germany, 2004. Springer.

[13] E. J. Hughes. Evolutionary Many-Objective Optimisation: Many once or one many? In *Congress on Evolutionary Computation (CEC 2005)*, pages 222–227. IEEE Press, 2005.

[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. ParEGO: A Hybrid Algorithm With On-Line Landscape Approximation for Expensive Multiobjective Optimization Problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66, 2005.

[16] 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.

[17] J. D. Knowles, D. W. Corne, and M. Fleischer. Bounded Archiving using the Lebesgue Measure. In *Congress on Evolutionary Computation (CEC 2003*, pages 2490–2497. IEEE Press, 2006.

[18] 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 2008)*, pages 665–672. ACM, 2008.

[19] R. C. Purshouse. *On the Evolutionary Optimisation of Many Objectives*. PhD thesis, The University of Sheffield, 2003.

[20] R. C. Purshouse and P. J. Fleming. An Adaptive Divide-and-Conquer Methodology for Evolutionary Multi-criterion Optimisation. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 133–147. Springer, 2003.

[21] T. Wagner, N. Beume, and B. Naujoks. Pareto-, Aggregation-, and Indicator-based Methods in Many-objective Optimization. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, pages 742–756. Springer, 2007.

[22] E. Zeidler. *Applied Functional Analysis: Main Principles and Their Applications.* Applied Mathematical Sciences 109. Springer, 1995.

[23] E. Zitzler, D. Brockhoff, and L. Thiele. The Hypervolume Indicator Revisited: On the Design of Pareto-compliant Indicators Via Weighted Integration. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, pages 862–876. Springer, 2007.

[24] E. Zitzler, K. Deb, and L. Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, 2000.

[25] E. Zitzler and S. Künzli. Indicator-Based Selection in Multiobjective Search. In *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, pages 832–842. Springer, 2004.

[26] E. Zitzler and L. Thiele. An Evolutionary Approach for Multiobjective Optimization: The Strength Pareto Approach. Technical Report 43, Computer Engineering and Networks Laboratory, ETH Zurich, May 1998.

[27] E. Zitzler and L. Thiele. Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In *Conference on Parallel Problem Solving from Nature (PPSN V)*, pages 292–301, 1998.

[28] 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.

Table 2: Pareto front shapes and density for the test problems used throughout this paper. For problems marked with a star, the used description is derived numerically. If not defined explicitly, we assume $0 \leq x \leq 1$.

| Problem	front description	density
bi-objective sphere	$f(x) = \left((b-a) - x^{1/\alpha}\right)^{\alpha}$	$\delta(x) = C \cdot \sqrt{(b - a - x^{1/\alpha})^{\alpha-1} \cdot x^{\frac{\alpha-a}{\alpha}}}$
ZDT1, ZDT4 [24]	$f(x) = 1 - \sqrt{x}$	$\delta(x) = \dfrac{3}{4x^{1/4}}$
ZDT2 [24]	$f(x) = 1 - x^2$ for $x \in [0,1]$	$\delta(x) = \dfrac{3}{2}\sqrt{x}$
ZDT3* [24]	$f(x) = 1 - \sqrt{x} - x \cdot \sin(10\pi x)$	$\delta(x) = 1.5609 \cdot \sqrt{\dfrac{1}{2\sqrt{x}} + \sin(10\pi x) + 10\pi x \cos(10\pi x)}$
	for all $x \in F$ where $F = [0, 0.0830015349] \, \cup \,]0.1822287280, 0.2577623634] \, \cup \,]0.4093136748, 0.4538821041] \, \cup$ $]0.6183967944, 0.6525117038] \, \cup \,]0.8233317983, 0.8518328654]$	
ZDT6 [24]	$f(x) = 1 - x^2$ for $x \in [\frac{\arctan(9\pi)}{6\pi}, 1] \approx [0.08146, 1]$	$\delta(x) = C \cdot \sqrt{x}$ with $C = \frac{3}{2}\left(1 - \frac{\arctan(9\pi)}{6\pi}^{3/2}\right)^{-1} \approx 1.53570$
DTLZ1 [8]	$f(x) = \dfrac{1}{2} - x$	$\delta(x) = 1$
DTLZ2, DTLZ3, DTLZ4 [8]	$f(x) = \sqrt{1 - x^2}$	$\delta(x) = 1.1803 \cdot \sqrt{\dfrac{x}{\sqrt{1-x^2}}}$
DTLZ7* [8]	$f(x) = 4 - x(1 + \sin(3\pi x))$	$\delta(x) = 0.6566 \cdot \sqrt{1 + \sin(3\pi x) + 3\pi x \cos(3\pi x)}$
	for all $x \in F$ where $F = [0, 0.2514118361] \, \cup \,]0.6316265307, 0.8594008566] \, \cup \,]1.3596178368, 1.5148392681] \, \cup$ $]2.0518383519, 2.1164268079]$	

APPENDIX

A. PROOFS OF SECTION 4
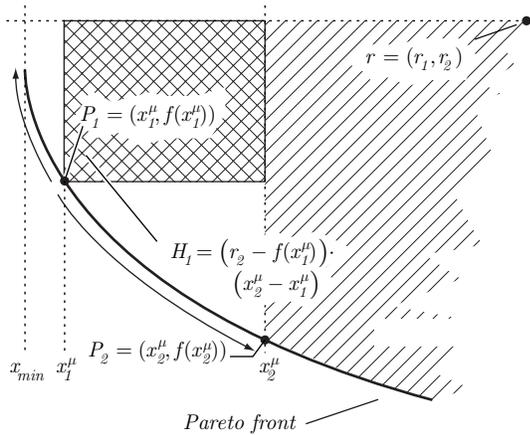
Figure 10: Shows the notation and formula to compute the hypervolume contribution H_1 of the leftmost point P_1.

Proof of Theorem 2

THEOREM 2. *Let μ be an integer larger or equal 2. Assume that f is continuous on $[x_{min}, x_{max}]$, non-increasing, differentiable on $]x_{min}, x_{max}[$ and that f' is continuous on*

$]x_{min}, x_{max}[$. *If $\lim_{x \to x_{min}} -f'(x) < +\infty$, let*

$$\mathcal{R}_2 := \sup_{x \in [x_{min}, x_{max}[} \{f'(x)(x - x_{max}) + f(x)\} \, , \quad (4)$$

where the supremum in the previous equation is possibly infinite. When $\mathcal{R}_2$ is finite, the leftmost extremal point is contained in optimal μ-distributions if the reference point $r = (r_1, r_2)$ is such that r_2 is strictly larger than $\mathcal{R}_2$.

Moreover, if $\lim_{x \to x_{min}} -f'(x) = +\infty$, the left extremal point of the front is never included in optimal μ-distributions.

The proof of the theorem requires to establish two technical lemmas. Let us assume the reference point is dominated by the Pareto front, i.e., at least $r_1 > x_{max}$ and $r_2 > f(x_{min})$. Consider a set of points on the front and let the hypervolume contribution of the leftmost point be $P_1 = (x_1, f(x_1))$ (see Figure 10). This is a function of x_1, the x-coordinate of the second leftmost point x_2, and the second coordinate of the reference point r_2. For a fixed x_2, r_2, the hypervolume contribution of the leftmost point with coordinate $x_1 \in [x_{min}, x_2[$ is denoted $H_1(x_1; x_2, r_2)$ and reads

$$H_1(x_1; x_2, r_2) = (x_2 - x_1)(r_2 - f(x_1)) \, . \quad (15)$$

The following lemma establishes a key property of the function H_1.

LEMMA 3. *If $x_1 \to H_1(x_1; x_{max}, r_2)$ is maximal for $x_1 = x_{min}$, then for any $x_2 \in]x_1, x_{max}]$, $x_1 \to H_1(x_1; x_2, r_2)$ is maximal for $x_1 = x_{min}$, too.*

PROOF. Assume that $H_1(x_1; x_{max}, r_2)$ is maximal for $x_1 = x_{min}$, i.e.,

$$H_1(x_{min}; x_{max}, r_2) \geq H_1(x_1; x_{max}, r_2), \forall x_1 \in]x_{min}, x_{max}]$$
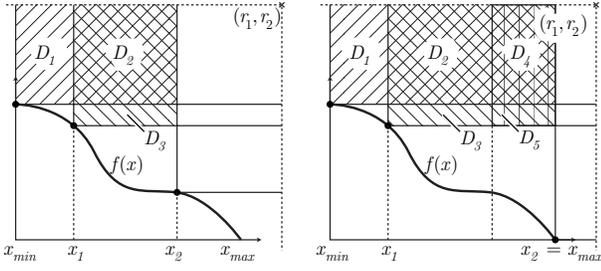
Figure 11: Illustration for proof of Lemma 3.

and let $\{D_1, \ldots, D_5\}$ denote the hypervolume indicator values of different non-overlapping rectangular areas shown in Fig. 11. Then for all x_1 in $]x_{min}, x_{max}]$,

$$H_1(x_{min}; x_{max}, r_2) \geq H_1(x_1; x_{max}, r_2)$$

can be rewritten using $D_1, \ldots, D_5$ as

$$D_1 + D_2 + D_4 \geq D_2 + D_3 + D_4 + D_5$$

which in turn implies that $D_1 + D_2 \geq D_2 + D_3 + D_5$. Since $D_5 \geq 0$, we have that $D_1 + D_2 \geq D_2 + D_3$, which corresponds to $H_1(x_{min}; x_2, r_2) \geq H_1(x_1; x_2, r_2)$. Hence, $H_1(x_1; x_2, r_2)$ is also maximal for $x_1 = x_{min}$ for any choice $x_2 \in]x_1, x_{max}]$. $\square$

LEMMA 4. *If* $\lim_{x \to x_{min}} f'(x) = -\infty$, *for any* $r_2(> f(x_{min}))$

$$\lim_{\varepsilon \to 0} \frac{(f(x_{min}) - f(x_{min} + \varepsilon))(x_2 - (x_{min} + \varepsilon))}{\varepsilon(r_2 - f(x_{min}))} = +\infty$$

PROOF. In order to prove the lemma, we only need to show that

$$\lim_{\varepsilon \to 0} \frac{f(x_{min}) - f(x_{min} - \varepsilon)}{\varepsilon} = +\infty \qquad (16)$$

since the remaining term converges to the constant $\frac{x_2 - x_{min}}{r_2 - f(x_{min})}$. To prove Eq. 16, we take a sequence (ε_n) with $\varepsilon_n > 0$ that converges to 0 and we show that

$$\lim_{\varepsilon_n \to 0} \frac{f(x_{min}) - f(x_{min} + \varepsilon_n)}{\varepsilon_n} = +\infty \ .$$

Thanks to the Mean Value Theorem, for all ε_n there exists a $c_n \in]x_{min}, x_{min} + \varepsilon_n[$ such that

$$\frac{f(x_{min}) - f(x_{min} + \varepsilon_n)}{\varepsilon_n} = -f'(c_n) \ .$$

When ε_n converges to 0, c_n converges to x_{min}. Moreover, $f'(x_{min}) = -\infty$ means that for every sequence θ_n that goes to x_{min} for $n \to \infty$, $f'(\theta_n)$ converges to $-\infty$. Therefore, $-f'(c_n)$ and as a consequence also $\frac{f(x_{min}) - f(x_{min} + \varepsilon_n)}{\varepsilon_n}$ converges to $+\infty$. $\square$

We are now ready to prove Theorem 2.

PROOF OF THEOREM 2. Let x_1 and x_2 denote the x-coordinates of the two leftmost points $P_1 = (x_1, f(x_1))$ and $P_2 = (x_2, f(x_2))$ as in the previous proof. Then the hypervolume contribution of P_1 is given by Eq. 15. To prove that P_1 is the extremal point $(x_{min}, f(x_{min}))$, we need to prove that $x_1 \in [x_{min}, x_2] \mapsto H_1(x_1; x_2, r_2)$ is maximal for $x_1 = x_{min}$. By using Lemma 3, we know that if we prove that $x_1 \to H_1(x_1; x_{max}, r_2)$ is maximal for $x_1 = x_{min}$ then we will also have that $H_1 : x_1 \in [x_{min}, x_2] \mapsto H_1(x_1; x_2, r_2)$
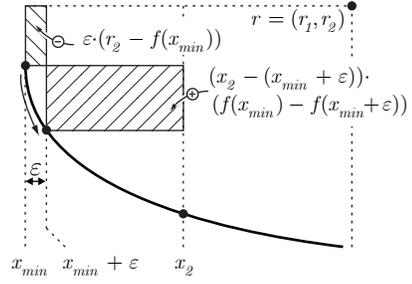

Figure 12: If the function f describing the Pareto front has an infinite derivative at its left extreme, the leftmost Pareto-optimal point will never coincide with the leftmost point x_1^μ of an optimal μ-distribution; since for any finite r_2 there exists an $\varepsilon > 0$, such that dominated space gained ($\oplus$) when moving x_1^μ from x_{min} to $x_{min} + \varepsilon$ is larger than the space no longer dominated ($\ominus$).

is maximal for $x_1 = x_{min}$. Therefore we will now prove that $x_1 \to H_1(x_1; x_{max}, r_2)$ is maximal for $x_1 = x_{min}$. To do so, we will show that the derivative of $H_1(x_1; x_{max}, r_2)$ never equals zero for all $x_{min} < x_1 \leq x_{max}$. The derivative of $H_1(x_1; x_{max}, r_2)$ equals $f(x_1) - r_2 + f'(x_1)(x_1 - x_{max})$ such that choosing r_2 strictly larger than $\mathcal{R}_2$ ensures that the derivative of $H_1(x_1; x_{max}, r_2)$ never equals zero.

Assume now $\lim_{x \to x_{min}} f'(x) = -\infty$ and that $x_1 = x_{min}$ in order to get a contradiction. Let $I_H(x_{min})$ be the hypervolume solely dominated by the point x_{min}. If we shift x_1 to the right by $\varepsilon > 0$ (see Figure 12 for the illustration and notations) then the new hypervolume contribution $I_H(x_{min} + \varepsilon)$ satisfies

$$I_H(x_{min} + \varepsilon) = I_H(x_{min})$$
$$+ (f(x_{min}) - f(x_{min} + \varepsilon))(x_2 - (x_{min} + \varepsilon))$$
$$- (r_2 - f(x_{min}))\varepsilon$$

From Lemma 4, we know that for any r_2, there exists $\varepsilon > 0$ such that

$$\frac{(f(x_{min}) - f(x_{min} + \varepsilon))(x_2 - (x_{min} + \varepsilon))}{(r_2 - f(x_{min}))\varepsilon} > 1$$

or equivalently

$$(f(x_{min}) - f(x_{min} + \varepsilon))(x_2 - (x_{min} + \varepsilon)) - (r_2 - f(x_{min}))\varepsilon > 0$$

which in turn implies that for any r_2 there exists an ε such that $I_H(x_{min} + \varepsilon) > I_H(x_{min})$ which contradicts the fact that $I_H(x_{min})$ is maximal and therefore that $x_1 = x_{min}$. $\square$

Proof of Theorem 3

THEOREM 3. *Let μ be an integer larger or equal 2. Assume that f is continuous on $[x_{min}, x_{max}]$, non-increasing, differentiable on $]x_{min}, x_{max}[$ and that f' is continuous and strictly negative on $]x_{min}, x_{max}]$. Let*

$$\mathcal{R}_1 := \sup_{x \in]x_{min}, x_{max}]} \left\{ x + \frac{f(x) - f(x_{min})}{f'(x)} \right\} \ , \qquad (5)$$

where the supremum in the previous equation is possibly infinite. When $\mathcal{R}_1$ is finite, the rightmost extremal point is contained in optimal μ-distributions if the reference point $r = (r_1, r_2)$ is such that r_1 is strictly larger than $\mathcal{R}_1$.

If $f'(x_{max}) = 0$, the right extremal point is never included in optimal μ-distributions.

PROOF. The proof of Theorem 3 is similar to the proof of Theorem 2 where—instead of $H_1(x_1; x_2, r_2)$—we consider the hypervolume contribution $H_\mu(x_\mu; x_{\mu-1}, r_1)$ of the rightmost point x_μ that equals $(r_1 - x_\mu)(f(x_{\mu-1}) - f(x_\mu))$ where $x_{\mu-1}$ is the x-coordinate of the neighbor point of the rightmost point. Since the proof is similar we only sketch the main points:
First of all, similarly to Lemma 3, if $x_\mu \to H_\mu(x_\mu; x_{min}, r_1)$ is maximal for $x_\mu = x_{max}$, then for any $x_{\mu-1} \in [x_{\min}, x_\mu[$, the contribution $H_\mu(x_\mu; x_{\mu-1}, r_1)$ is maximal for $x_\mu = x_{max}$, too.
Second, similarly to the proof of Theorem 2, we need to show that the derivative of $x_\mu \to H_\mu(x_\mu; x_{min}, r_1)$ is never zero. The derivative of $H_\mu(x_\mu; x_{min}, r_1)$ equals $(f(x_\mu) - f(x_{min})) + f'(x_\mu)(x_\mu - r_1)$ such that setting r_1 strictly larger than $\mathcal{R}_1$ ensures that the derivative of $H_\mu(x_\mu; x_{min}, r_1)$ is never zero.
Third, assuming that $f'(x_{max})$ equals zero, the last point of Theorem 3 follows by contradiction by assuming that $x_\mu = x_{max}$ and showing that then one can increase the hypervolume by moving x_μ to the left. $\square$

B. PROOFS OF SECTION 6

Proof of Lemma 2

LEMMA 2. *If f is continuous, differentiable with the derivative f' continuous, if $x_1^\mu, \ldots, x_\mu^\mu$ converge to a continuous density δ, with $\frac{1}{\delta} \in L^2(0, x_{max})$, and $\exists c \in \mathbb{R}^+$ such that*

$$\mu \sup(\sup_{0 \leq i \leq \mu-1} |x_{i+1}^\mu - x_i^\mu|, |x_{max} - x_\mu^\mu|) \to c \qquad (11)$$

then E_μ converges for $\mu \to \infty$ to

$$E(\delta) := -\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)}{\delta(x)} dx \ . \qquad (12)$$

PROOF. Let us first note that the Cauchy-Schwarz inequality implies that

$$\int_0^{x_{max}} |f'(x)/\delta(x)| dx \leq \sqrt{\int_0^{x_{max}} f'(x)^2 dx \int_0^{x_{max}} (1/\delta(x))^2 dx} \ (17)$$

and since $f' \in L^2(0, x_{max})$ and $\frac{1}{\delta} \in L^2(0, x_{max})$, the right-hand side of Eq. 17 is finite and Eq. 12 is well-defined.

Step 1. In a first step we are going to prove that E_μ defined in Eq. 10 satisfies

$$E_\mu = \mu \sum_{i=0}^{\mu} \left(-\frac{1}{2} f'(x_i^\mu)(x_{i+1}^\mu - x_i^\mu)^2 + O((x_{i+1}^\mu - x_i^\mu)^3) \right) \ (18)$$

To this end, we elongate the front to the right such that f equals $f(x_{max}) = 0$ for $x \in [x_{max}, x_{\mu+1}^\mu]$. Like that, we can decompose $\int_0^{x_{max}} f(x) dx$ as

$$\int_0^{x_{max}} f(x) dx + \int_{x_{max}}^{x_{\mu+1}^\mu} f(x) dx = \sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} f(x) dx \ .$$

which can be rewritten as

$$\int_0^{x_{max}} f(x) dx = \sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} f(x) dx \qquad (19)$$

because $f(x) = 0$ in the interval $[x_{max}, x_{\mu+1}^\mu]$ and therefore $\int_{x_{max}}^{x_{\mu+1}^\mu} f(x) dx = 0$. Since f is differentiable, we can use a Taylor approximation of f in each interval $[x_i^\mu, x_{i+1}^\mu]$ and write

$$f(x) = f(x_i^\mu) + f'(x_i^\mu)(x - x_i^\mu) + O((x - x_i^\mu)^2) \ .$$

By integrating the previous equation between x_i^μ and x_{i+1}^μ we obtain

$$\int_{x_i^\mu}^{x_{i+1}^\mu} f(x) dx = f(x_i^\mu)(x_{i+1}^\mu - x_i^\mu) +$$
$$\frac{1}{2} f'(x_i^\mu)(x_{i+1}^\mu - x_i^\mu)^2 + O((x_{i+1}^\mu - x_i^\mu)^3) \quad (20)$$

Summing up for $i = 0$ to $i = \mu$ and using Eq. 19 we obtain

$$\int_0^{x_{max}} f(x) dx = \sum_{i=0}^{\mu} f(x_i^\mu)(x_{i+1}^\mu - x_i^\mu)$$
$$+ \sum_{i=0}^{\mu} \frac{1}{2} f'(x_i^\mu)(x_{i+1}^\mu - x_i^\mu)^2 + O((x_{i+1}^\mu - x_i^\mu)^3) \quad (21)$$

Hence, by definition of E_μ (Eq. 10) we obtain Eq. 18, which concludes Step 1.

Step 2. We now decompose $\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)}{\delta(x)} dx$ into

$$\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)}{\delta(x)} dx = \frac{1}{2} \sum_{i=0}^{\mu-1} \int_{x_i^\mu}^{x_{i+1}^\mu} \frac{f'(x)}{\delta(x)} dx$$
$$+ \frac{1}{2} \int_{x_\mu^\mu}^{x_{max}} \frac{f'(x)}{\delta(x)} dx$$

For the sake of convenience in the notations, for the remainder of the proof, we redefine $x_{\mu+1}^\mu$ as x_{max} such that the previous equation becomes

$$\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)}{\delta(x)} dx = \frac{1}{2} \sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \frac{f'(x)}{\delta(x)} dx \qquad (22)$$

For μ to ∞, the assumption $\mu \sup(\sup_{0 \leq i \leq \mu-1} |x_{i+1}^\mu - x_i^\mu|, |x_\mu^\mu - x_{max}|) \to c$ implies that the distance between two consecutive points $|x_{i+1}^\mu - x_i^\mu|$ as well as $|x_\mu^\mu - x_{max}|$ converges to zero. Let $x \in [0, x_{max}]$ and let us define for a given μ, $\varphi(\mu)$ as the index of the points such that $x_{\varphi(\mu)}^\mu$ and $x_{\varphi(\mu)+1}^\mu$ surround x:

$$x_{\varphi(\mu)}^\mu \leq x < x_{\varphi(\mu)+1}^\mu \ .$$

Since we assume that δ is continuous, a first order approximation of $\delta(x)$ is $\delta(x_{\varphi(\mu)}^\mu)$, i.e. $\delta(x) = \delta(x_{\varphi(\mu)}^\mu) + O(x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu)$ and therefore by integrating between $x_{\varphi(\mu)}^\mu$ and $x_{\varphi(\mu)+1}^\mu$ we obtain

$$\int_{x_{\varphi(\mu)}^\mu}^{x_{\varphi(\mu)+1}^\mu} \delta(x) dx = \delta(x_{\varphi(\mu)}^\mu)(x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu)$$
$$+ O((x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu)^2) \quad (23)$$

Moreover by definition of the density δ, $\int_{x_{\varphi(\mu)}^\mu}^{x_{\varphi(\mu)+1}^\mu} \delta(x) dx$ approximates the number of points contained in the interval

$[x^\mu_{\varphi(\mu)}, x^\mu_{\varphi(\mu)+1}[$ (i.e. one) normalized by μ:

$$\mu \int_{x^\mu_{\varphi(\mu)}}^{x^\mu_{\varphi(\mu)+1}} \delta(x)dx = 1 + O((x^\mu_{\varphi(\mu)+1} - x^\mu_{\varphi(\mu)})) \ . \quad (24)$$

Using Eq. 23 and Eq. 24, we thus have

$$\frac{1}{\delta(x^\mu_{\varphi(\mu)})} = \mu(x^\mu_{\varphi(\mu)+1} - x^\mu_{\varphi(\mu)}) + O(\mu(x^\mu_{\varphi(\mu)+1} - x^\mu_{\varphi(\mu)})^2) \ .$$

Therefore for every i we have that

$$\frac{1}{\delta(x^\mu_i)} = \mu(x^\mu_{i+1} - x^\mu_i) + O(\mu(x^\mu_{i+1} - x^\mu_i)^2) \ . \quad (25)$$

Since f'/δ is continuous, we also obtain

$$\int_{x^\mu_i}^{x^\mu_{i+1}} \frac{f'(x)}{\delta(x)}dx = \frac{f'(x^\mu_i)}{\delta(x^\mu_i)}(x^\mu_{i+1} - x^\mu_i) + O((x^\mu_{i+1} - x^\mu_i)^2)$$

Injecting Eq. 25 in the previous equation, we obtain

$$\int_{x^\mu_i}^{x^\mu_{i+1}} \frac{f'(x)}{\delta(x)}dx = \mu f'(x^\mu_i)(x^\mu_{i+1} - x^\mu_i)^2 + O(\mu(x^\mu_{i+1} - x^\mu_i)^3)$$

Multiplying by $1/2$ and summing up for i from 0 to μ and using Eq. 22 we obtain

$$\frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)} = -E_\mu + \sum_{i=0}^{\mu} O(\mu(x^\mu_{i+1} - x^\mu_i)^3) \quad (26)$$

Let us define Δ_μ as $\sup(\sup_{0 \le i \le \mu-1} |x^\mu_{i+1} - x^\mu_i|, |x_{max} - x^\mu_\mu|)$. By assumption, we know that $\mu\Delta_\mu$ converges to a positive constant c. The last term of Eq. 26 satisfies

$$|\sum_{i=0}^{\mu} O(\mu(x^\mu_{i+1} - x^\mu_i)^3)| \le K\mu^2(\Delta_\mu)^3$$

where $K > 0$. Since $\mu\Delta_\mu$ converges to c, $(\mu\Delta_\mu)^2$ converges to c^2. With Δ_μ converges to 0, we therefore have that $\mu^2\Delta_\mu^3$ converges to 0. Taking the limit in Eq. 26 we therefore obtain

$$-\frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)}dx = \lim_{\mu\to\infty} E_\mu$$

$\square$

Proof of Theorem 7

THEOREM 7. *The limit density of points maximizing the hypervolume is a solution of the constraint optimization problem (P) and equals*

$$\delta(x) = \frac{\sqrt{-f'(x)}}{\int_0^{x_{max}} \sqrt{-f'(x)}dx} \ .$$

PROOF. We first need to compute the differential of E with respect to the density δ, denoted by $DE_\delta(h)$. Let $h \in L^2(0, x_{max})$. Then,

$$E(\delta + h) = -\frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x) + h(x)}dx$$
$$= -\frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)\left(1 + \frac{h(x)}{\delta(x)}\right)}dx \ .$$

Due to the Taylor expansion of $\frac{1}{1+y} \overset{y\to 0}{=} 1 - y + o(y)$ this equals

$$E(\delta + h) = -\frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)}\left(1 - \frac{h(x)}{\delta(x)} + o\left(\|h(x)\|\right)\right)dx$$
$$= -\frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)}dx + \frac{1}{2}\int_0^{x_{max}} \frac{f'(x)h(x)}{\delta(x)^2}dx$$
$$- \frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)}o\left(\|h(x)\|\right)dx$$
$$= E(\delta) + \frac{1}{2}\int_0^{x_{max}} \frac{f'(x)h(x)}{\delta(x)^2}dx + o\left(\|h(x)\|\right) \ .$$

Since $h \to \frac{1}{2}\int_0^{x_{max}} \frac{f'h}{\delta^2}dx$ is linear (in h), we know from differential calculus that

$$DE_\delta(h) = \frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)^2}h(x)dx \ .$$

In a similar way,

$$J(\delta + h) = \int_0^{x_{max}} (\delta(x) + h(x))\,dx$$
$$= \int_0^{x_{max}} \delta(x)dx + \int_0^{x_{max}} h(x)dx$$
$$= J(\delta) + \int_0^{x_{max}} h(x)dx$$

and as $h \to \int_0^{x_{max}} h(x)dx$ is linear, the differential of J equals

$$DJ_\delta(h) = \int_0^{x_{max}} h(x)dx \ .$$

From the the Lagrange multiplier theorem for Banach spaces [22], we know that there exists a $\lambda \in \mathbb{R}$ such that the solution of P satisfies

$$\forall h : DE_\delta(h) + \lambda DJ_\delta(h) = 0$$

that can be rewritten as

$$\forall h : \frac{1}{2}\int_0^{x_{max}} \frac{f'(x)}{\delta(x)^2}h(x)dx + \lambda\int_0^{x_{max}} h(x)dx = 0$$

or

$$\forall h : \int_0^{x_{max}} \left(\frac{1}{2}\frac{f'(x)}{\delta(x)^2} + \lambda\right)h(x)dx = 0 \ . \quad (27)$$

Since a solution for P has to satisfy Eq. 27 for all h, we know for the choice of $h(x) = \frac{1}{2}\frac{f'(x)}{\delta(x)^2} + \lambda$ that

$$\int_0^{x_{max}} \left(\frac{1}{2}\frac{f'(x)}{\delta(x)^2} + \lambda\right)^2 dx = 0$$

holds which in turn implies that $\frac{1}{2}\frac{f'}{\delta^2} + \lambda = 0$ or in other words that

$$\delta(x) = \sqrt{-f'(x)}/\sqrt{2\lambda}$$

where the constant λ is still to be determined. We know that δ is a density and needs therefore to satisfy that $\int_0^{x_{max}} \delta(x)dx = 1$. Then, we can determine the missing $\sqrt{2\lambda}$ from

$$1 = \int_0^{x_{max}} \delta(x)dx = \int_0^{x_{max}} \frac{\sqrt{-f'(x)}}{\sqrt{2\lambda}}dx$$
$$= \frac{1}{\sqrt{2\lambda}}\int_0^{x_{max}} \sqrt{-f'(x)}dx$$

as $\sqrt{2\lambda} = \int_0^{x_{max}} \sqrt{-f'(x)}dx$ which yields then

$$\delta(x) = \frac{\sqrt{-f'(x)}}{\int_0^{x_{max}} \sqrt{-f'(x)}dx} \quad .$$

□