Hypervolume-based Multiobjective Optimization: Theoretical Foundations and Practical Implications

Anne Auger, Johannes Bader, Dimo Brockhoff, Eckart Zitzler

Hypervolume-based Multiobjective Optimization: Theoretical Foundations and Practical Implications

Anne Auger[a], Johannes Bader[b], Dimo Brockhoff[a,c], Eckart Zitzler[b]

[a] *TAO Team INRIA Saclay—Île-de-France, LRI Paris Sud University, 91405 Orsay Cedex, France*
firstname.lastname@inria.fr
[b] *Computer Engineering and Networks Lab, ETH Zurich, 8092 Zurich, Switzerland*
firstname.lastname@tik.ee.ethz.ch
[c] *corresponding author; currently at LIX, École Polytechnique, Palaiseau, France*
brockho@lix.polytechnique.fr

Abstract

In recent years, indicator-based evolutionary algorithms, allowing to implicitly incorporate user preferences into the search, have become widely used in practice to solve multiobjective optimization problems. When using this type of methods, the optimization goal changes from optimizing a set of objective functions simultaneously to the single-objective optimization goal of finding a set of μ points that maximizes the underlying indicator. Understanding the difference between these two optimization goals is fundamental when applying indicator-based algorithms in practice. On the one hand, a characterization of the inherent optimization goal of different indicators allows the user to choose the indicator that meets her preferences. On the other hand, knowledge about the sets of μ points with optimal indicator values—so-called *optimal μ-distributions*—can be used in performance assessment whenever the indicator is used as a performance criterion. However, theoretical studies on indicator-based optimization are sparse.

One of the most popular indicators is the weighted hypervolume indicator. It allows to guide the search towards user-defined objective space regions and at the same time has the property of being a refinement of the Pareto dominance relation with the result that maximizing the indicator results in Pareto-optimal solutions only. In previous work, we theoretically investigated the *unweighted* hypervolume indicator in terms of a characterization of optimal μ-distributions and the influence of the hypervolume's reference point for general bi-objective optimization problems. In this paper, we generalize those results to the case of the *weighted* hypervolume indicator. In particular, we present general investigations for finite μ, derive a limit result for μ going to infinity in terms of a density of points and derive lower bounds (possibly infinite) for placing the reference point to guarantee the Pareto front's extreme points in an optimal μ-distribution. Furthermore, we state conditions about the slope of the front at the extremes such that there is no finite reference point that allows to include the extremes in an optimal μ-distribution—contradicting previous belief that a reference point chosen just above the nadir point or the objective space boundary is sufficient for obtaining the extremes. However, for fronts where there exists a finite reference point allowing to obtain the extremes, we show that for μ to infinity, a reference point that is slightly worse in all objectives than the nadir point is a sufficient choice. Last, we apply the theoretical results to problems of the ZDT, DTLZ, and WFG test problem suites.

Key words: multiobjective optimization, evolutionary algorithms, hypervolume indicator, reference point, optimal μ-distributions

1. Introduction

Multiobjective optimization aims at optimizing several criteria simultaneously. In the last decades, evolutionary algorithms have been shown to be well-suited for those problems in practice (Deb, 2001; Coello Coello et al., 2007). A recent trend is to use quality indicators to turn a multiobjective optimization problem into a single-objective one by optimizing the quality indicator itself. An *indicator-based algorithm* uses a specific quality indicator to assign every individual a single-objective fitness—most of the time proportional to the *indicator loss*, a measure of how much the quality indicator decreases if the corresponding individual is removed from the population. Instead of optimizing the objective functions directly, indicator-based algorithms therefore aim at finding a set of solutions that maximizes

the underlying quality indicator and a fundamental question is whether these two optimization goals coincide or how they differ. In practice, the population size of indicator-based algorithms is usually finite, i.e., equal to $\mu \in \mathbb{N}$, and the optimization goal changes to finding a set of μ solutions optimizing the quality indicator[1]. We call such a set an *optimal μ-distribution for the given indicator* generalizing the definition given by Auger et al. (2009b). In this case, the additional questions arise how the number of points μ influences the optimization goal and to which set of μ objective vectors the optimal μ-distribution is mapped, i.e., which search bias is introduced by changing the optimization goal. Ideally, the optimal μ-distribution for an indicator only contains Pareto-optimal points and an increase in μ covers more and more points until the entire Pareto front is covered if μ approaches infinity. It is clear that in general, two different quality indicators yield a priori two different optimal μ-distributions, or in other words, introduce a different search bias. This has for instance been shown experimentally by Friedrich et al. (2009) for the multiplicative ε-indicator and the hypervolume indicator.

The hypervolume indicator and its weighted version (Zitzler et al., 2007) are particularly interesting indicators since they are refinements of the Pareto dominance relation (Zitzler et al., 2010)[2]. Thus, an optimal μ-distribution for these indicators contains only Pareto-optimal solutions and the set (probably unbounded in size) that maximizes the (weighted) hypervolume indicator covers the entire Pareto front (Fleischer, 2003). Many other quality indicators do not have this fundamental property. It explains the success of the hypervolume indicator as quality indicator applied to environmental selection of indicator-based evolutionary algorithms such as ESP (Huband et al., 2003), SMS-EMOA (Beume et al., 2007b), MO-CMA-ES (Igel et al., 2007), or HypE (Bader and Zitzler, 2011). Nevertheless, it has been argued that using the (weighted) hypervolume indicator to guide the search introduces a certain bias. Interestingly, several contradicting beliefs about this bias have been reported in the literature which we will discuss later on in more detail (see Sec. 3). They range from stating that *convex regions may be preferred to concave regions* to the argumentation that *the hypervolume is biased towards boundary solutions*. In the light of those contradicting beliefs, a thorough investigation of the effect of the hypervolume indicator on optimal μ-distributions is necessary.

Another important issue when dealing with the hypervolume indicator is the choice of the reference point, a parameter, both the unweighted and the weighted hypervolume indicator depend on. The influence of this reference point on optimal μ-distributions has not been fully understood, especially for the weighted hypervolume indicator, and only rules-of-thumb exist on how to choose the reference point in practice. In particular, it could not be observed from practical investigations how the reference point has to be set to ensure to find the extremes of the Pareto front. Several authors recommend to use the corner of a space that is a little bit larger than the actual objective space as the reference point (Knowles, 2005; Beume et al., 2007b). For performance assessment, others recommend to use the estimated nadir point as the reference point (Purshouse and Fleming, 2003; Purshouse, 2003; Hughes, 2005). Also here, theoretical investigations are highly needed to assist in practical applications.

First theoretical studies on optimal μ-distributions for the (unweighted) hypervolume indicator and the choice of its reference point have been published in an earlier work by the authors (Auger et al., 2009b). The theoretical analyses resulted in a better understanding of the search bias the hypervolume indicator introduces and in theoretically founded recommendations on where to place the reference point in the case of two objectives. In particular, some beliefs about the indicator's search bias could be disproved and others confirmed, the optimal μ-distributions for linear Pareto fronts were characterized exactly (see also (Brockhoff, 2010)), and lower bounds on the reference point's objective values that allow to include the extremes of the Pareto front in certain cases have been given. Recently, a specific result of Auger et al. (2009b) has been already generalized to the weighted hypervolume indicator (Auger et al., 2009a) and another exact result for specific Pareto fronts have been provided (Friedrich et al., 2009).

In this paper, we extend *all* results by Auger et al. (2009b) to the weighted case and provide a general theory of the weighted hypervolume indicator in terms of both the inherently introduced search bias and the choice of the reference point. In particular, we

- characterize the sets of μ points that maximize the (weighted) hypervolume indicator; besides general investigations for finite μ, we derive a limit result for μ going to infinity in terms of a density of points. The presented results for the weighted hypervolume indicator comply with the results for the unweighted case (Auger et al.,

[1] Sometimes, the population size might not be fixed, e.g., when deleting all dominated solutions, but the maximum number of simultaneously considered solutions is typically upper bounded by a constant μ.

[2] Other studies introduced the equivalent terms of being *compatible* or *compliant* with the Pareto dominance relation (Knowles and Corne, 2002; Zitzler et al., 2003).

2009b). Furthermore, we

- investigate the influence of the reference point on optimal μ-distributions, i.e., we derive lower bounds for the objective values of the reference point (possibly infinite) for guaranteeing the Pareto front's extreme points in an optimal μ-distribution and investigate cases where the extremes are never contained in such a set; these results generalize the work by Auger et al. (2009b) to the weighted hypervolume indicator. In addition, we
- prove, in case the extremes can be obtained, that for any reference point dominated by the nadir point—with any small but positive distance between the two points—there is a finite number of points μ_0 (possibly large in practice) such that for all $\mu > \mu_0$, the extremes are included in optimal μ-distributions. Last, we
- apply the theoretical results to linear Pareto fronts (Auger et al., 2009b; Brockhoff, 2010) and to benchmark problems of the ZDT (Zitzler et al., 2000), DTLZ (Deb et al., 2005b), and WFG (Huband et al., 2006) test problem suites resulting in recommended choices of the reference point including numerical and sometimes analytical expressions for the resulting density of points on the front.

The paper is structured as follows. First, we recapitulate the basics of the (weighted) hypervolume indicator and introduce the notations and definitions needed in the remainder of the paper (Sec. 2). Then, we consider the bias of the weighted hypervolume indicator in terms of optimal μ-distributions. After characterizing optimal μ-distributions for a finite number of solutions (Sec. 3.1), we derive results on the density of points if the number of points goes to infinity (Sec. 3.2). Section 4 investigates the influence of the reference point on optimal μ-distributions especially on the extremes. The application of the results to test problems is presented in Sec. 5, and Sec. 6 concludes the paper.

2. The Hypervolume Indicator: General Aspects and Notations

Throughout this study we consider, without loss of generality, minimization problems where k objective functions $\mathcal{F}_i : X \to Z$, $1 \leq i \leq k$ have to be minimized simultaneously. The vector function $\mathcal{F} := (\mathcal{F}_1, \ldots, \mathcal{F}_k)$ thereby maps each solution x in the decision space X to its corresponding objective vector $\mathcal{F}(x)$ in the objective space $\mathcal{F}(X) = Z \subseteq \mathbb{R}^k$. Furthermore, we assume that the underlying dominance structure is given by the weak Pareto dominance relation $\preceq$ which is defined between arbitrary solution pairs. We say $x \in X$ *weakly dominates* $y \in X$ if for all $1 \leq i \leq k$, $\mathcal{F}_i(x) \leq \mathcal{F}_i(y)$ and write $x \preceq y$. This weak Pareto dominance relation is generalized to sets of solutions in the following straightforward manner: we say a set A of solutions weakly dominates another solution set B if for all $b \in B$ there exists an $a \in A$ such that $a \preceq b$. The *Pareto(-optimal) set* P_s consists of all solutions $x^* \in X$, such that there is no $x \in X$ that satisfies $x \preceq x^*$ and $x^* \not\preceq x$. The image of P_s under $\mathcal{F}$ is called *Pareto(-optimal) front* or *front* for short. We also use the weak Pareto dominance relation notation $\preceq$ among objective vectors, i.e., for two objective vectors $x = (x_1, \ldots, x_k), y = (y_1, \ldots, y_k) \in \mathbb{R}^k$ we define $x \preceq y$ if and only if for all $1 \leq i \leq k : x_i \leq y_i$.

In the following, in order to simplify notations[3], we define the indicators for *sets of objective vectors* $A \subseteq \mathbb{R}^k$ instead for *solution sets* $A' \subseteq X$ as it was already done before (Zitzler et al., 2007; Auger et al., 2009b). The weighted hypervolume indicator $I_{H,w}(A, r)$ for a set of objective vectors $A \subseteq Z$ is then the weighted Lebesgue measure of the set of objective vectors weakly dominated by the solutions in A that at the same time weakly dominate a so-called reference point $r \in Z$ (Bader and Zitzler, 2011)[4]:

$$I_{H,w}(A, r) = \int_{\mathbb{R}^k} w(z) \mathbf{1}_{H(A,r)}(z) dz \tag{1}$$

where $H(A, r) := \{z \in Z \mid \exists a \in A : a \preceq z \preceq r\}$, $\mathbf{1}_{H(A,r)}(z)$ is the characteristic function of $H(A, r)$ that equals 1 iff $z \in H(A, r)$ and 0 otherwise, and $w : \mathbb{R}^k \to \mathbb{R}_{>0}$ is a strictly positive weight function integrable on any bounded set, i.e., $\int_{B(0,\gamma)} w(z)dz < \infty$ for any $\gamma > 0$, where $B(0, \gamma)$ is the open ball centered in 0 and of radius γ. In other words, we assume that the measure associated to w is σ-finite[5]. Throughout the paper, the notation I_H refers to the non-weighted

[3] Considering an indicator on solution sets introduces the possibility of solutions that map to the same objective vector. Adding such a so-called *indifferent solution* to a solution set does not affect the set's hypervolume indicator value but the consideration of such solutions makes the text less readable if we want to state the results formally correct.

[4] Instead of a reference set as by Bader and Zitzler (2011), we consider one reference point only as in earlier publications (Zitzler et al., 2007).

[5] Several results presented in this paper also hold if the weight is strictly positive almost everywhere, i.e., it can be 0 for null sets. However, we decided to consider only strictly positive weights to keep the proofs simple.

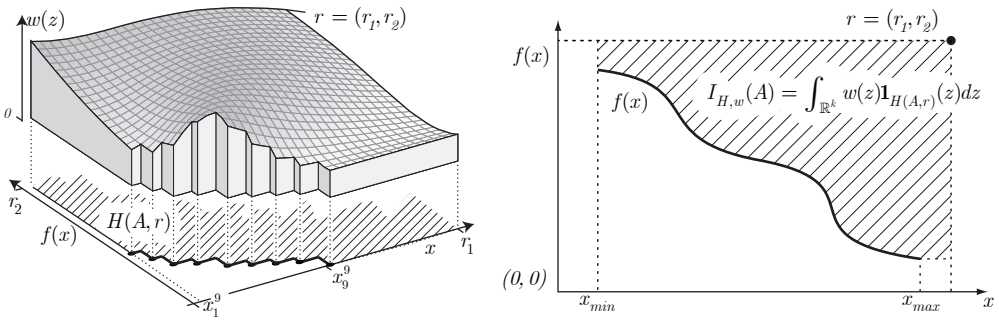


Figure 1: The hypervolume indicator $I_{H,w}(A)$ corresponds to the integral of a weight function $w(z)$ over the set of objective vectors that are weakly dominated by a solution set A and in addition weakly dominate the reference point r (hatched areas). On the left, the set A consists of nine objective vectors whereas on the right, the infinite set A can be described by a function $f : [x_{\min}, x_{\max}] \to \mathbb{R}$. The left-hand plot shows an example of a weight function $w(z)$, where for all objective vectors z that are not dominated by A or not enclosed by r the function w is not plotted, such that the weighted hypervolume indicator corresponds to the volume of the gray shape.

hypervolume where the weight is 1 everywhere, and we will explicitly use the term non-weighted hypervolume for I_H while the weighted hypervolume indicator $I_{H,w}$ is, for simplicity, referred to as hypervolume.

The left-hand plot of Fig. 1 illustrates the hypervolume $I_{H,w}$ for a bi-objective problem. The three-objective plot shows the objective values of nine points on the first two axes and the weight function w on the third axis. The hypervolume indicator $I_{H,w}(A)$ for the set A of nine points equals the integral of the weight function over the objective space that is weakly dominated by the set A and which weakly dominates the reference point $r = (r_1, r_2)$.

In what follows, we consider bi-objective problems. The Pareto front can thus be described by a one-dimensional function f mapping the image of the Pareto set under the first objective $\mathcal{F}_1$ onto the image of the Pareto set under the second objective $\mathcal{F}_2$,

$$f : x \in D \mapsto f(x) \ ,$$

where D denotes the image of the Pareto set under the first objective. D can be, for the moment, either a finite or an infinite set. An illustration is given in the right-hand plot of Fig. 1 where the function f describing the front has a domain of $D = [x_{\min}, x_{\max}]$.

Example 1. *Consider the bi-objective problem DTLZ2 (Deb et al., 2005b) which is defined as*

$$\begin{aligned} \textit{minimize} \quad & \mathcal{F}_1(d) = (1 + g(d_M)) \cos(d_1 \pi/2) \\ \textit{minimize} \quad & \mathcal{F}_2(d) = (1 + g(d_M)) \sin(d_1 \pi/2) \\ & g(d_M) = \sum_{d_i \in d_M} (d_i - 0.5)^2 \\ \textit{subject to} \quad & 0 \le d_i \le 1 \textit{ for } i = 1, \ldots n \end{aligned} \tag{2}$$

where d_M denotes a subset of the decision variables $d = (d_1, \ldots, d_n) \in [0, 1]^n$ with $g(d_M) \ge 0$. The Pareto front is reached for $g(d_M) = 0$. Hence, the Pareto-optimal points have objective vectors $(\cos(d_1\pi/2), \sin(d_1\pi/2))$ *with* $0 \le d_1 \le 1$ *which can be rewritten as points* $(x, f(x))$ *with* $f(x) = \sqrt{1 - x^2}$ *and* $x \in D = [0, 1]$*, see Fig. 9(f).*

Since f represents the shape of the trade-off surface, we can conclude that, for minimization problems, f is strictly monotonically decreasing in D[6]. The coordinates of a point belonging to the Pareto front are given as a pair $(x, f(x))$ with $x \in D$ and therefore, a point is entirely determined by the function f and the first coordinate $x \in D$. For μ points on the Pareto front, we denote their first coordinates as $(x_1, \ldots, x_\mu)$. Without loss of generality, it is assumed that $x_i \le x_{i+1}$, for $i = 1, \ldots, \mu - 1$ and for notation convenience, we set $x_{\mu+1} := r_1$ and $f(x_0) := r_2$ where r_1 and r_2 are the first and second coordinate of the reference point (see Figure 2). The weighted hypervolume enclosed by these points

[6]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 $x_1, x_2 \in D$ such that, without loss of generality, $x_1 < x_2$ and $f(x_1) \le f(x_2)$, i.e., $(x_1, f(x_1))$ dominates $(x_2, f(x_2))$.

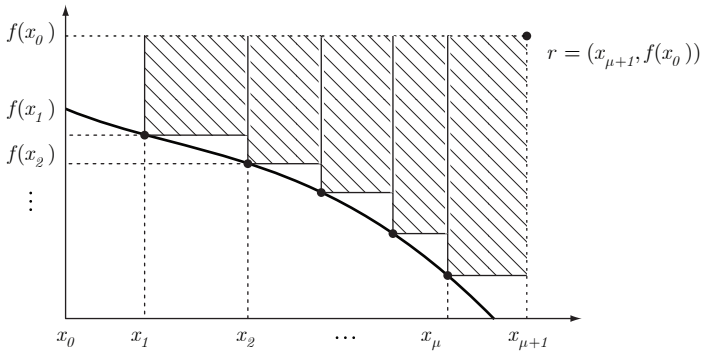


Figure 2: Computation of the hypervolume indicator for μ solutions $(x_1, f(x_1)), \ldots, (x_\mu, f(x_\mu))$ and the reference point $r = (r_1, r_2)$ in the bi-objective case as defined in Eq. 3 and Eq. 4 respectively.

can be decomposed into μ components, each corresponding to the integral of the weight function w over a rectangular area (see Figure 2). The resulting weighted hypervolume writes:

$$I_{H,w}((x_1, \ldots, x_\mu)) := \sum_{i=1}^{\mu} \int_{x_i}^{x_{i+1}} \left(\int_{f(x_i)}^{f(x_0)} w(x, y) dy \right) dx \ . \tag{3}$$

When the weight function equals one everywhere, one retrieves the expression for the (non-weighted) hypervolume (Auger et al., 2009b)

$$I_H((x_1, \ldots, x_\mu)) := \sum_{i=1}^{\mu} (x_{i+1} - x_i)(f(x_0) - f(x_i)) \ . \tag{4}$$

Indicator-based evolutionary algorithms that aim at optimizing a unary indicator $I : 2^X \to \mathbb{R}$ such as the hypervolume transform a multiobjective problem into the single-objective one consisting of finding a set of points maximizing the respective indicator I. In practice, the cardinality of these sets of points is usually upper bounded by a constant μ, typically the population size. Generalizing the definition by Auger et al. (2009b), we define an *optimal μ-distribution* as a set of μ points maximizing I.

Definition 1 (Optimal μ-distribution)**.** *For $\mu \in \mathbb{N}$ and a unary indicator I, a set of μ points maximizing I is called an optimal μ-distribution for I.*

The rest of the paper is devoted to understand optimal μ-distributions for the hypervolume indicator in the bi-objective case. The x-coordinates of an optimal μ-distribution for the hypervolume $I_{H,w}$ will be denoted $(\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu)$ and will thus satisfy

$$I_{H,w}((\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu)) \geq I_{H,w}((x_1, \ldots, x_\mu)) \text{ for all } (x_1, \ldots, x_\mu) \in D \times \ldots \times D \ .$$

Note, that the optimal μ-distribution might not be unique, and $(\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu)$ therefore refers to *one* optimal μ-distribution. The corresponding value of the hypervolume will be denoted $\overline{I_{H,w}^\mu}$, i.e., $\overline{I_{H,w}^\mu} = I_{H,w}((\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu))$.

Remark 1. *Looking at Eq. 3 and Eq. 4, we see that for a fixed f, a fixed weight w, and a fixed reference point, the problem of finding a set of μ points maximizing the weighted hypervolume amounts to finding the solution of a μ-dimensional single-objective maximization problem, i.e., optimal μ-distributions are the solution of a single objective problem of μ variables.*

3. Characterization of Optimal μ-Distributions for Hypervolume Indicators

Several contradicting beliefs about the bias introduced by the hypervolume indicator have been reported in the literature. For example, Zitzler and Thiele (1998) stated that, when optimizing the hypervolume in maximization problems, "convex regions may be preferred to concave regions", which has been also stated by Lizarraga-Lizarraga et al. (2008) later on, whereas Deb et al. (2005a) argued that "[...] the hyper-volume measure is biased towards

the boundary solutions". Knowles and Corne (2003) observed that a local optimum of the hypervolume indicator "seems to be 'well-distributed'" which was also confirmed empirically (Knowles et al., 2003; Emmerich et al., 2005). Beume et al. (2007b), in addition, state several properties of the hypervolume's bias: (i) optimizing the hypervolume indicator focuses on knee points; (ii) the distribution of points on the extremes is less dense than on knee points; (iii) only linear front shapes allow for equally spread solutions; and (iv) extremal solutions are maintained. In the light of these contradicting statements, a thorough characterization of optimal μ-distributions for the hypervolume indicator is necessary. Especially for the weighted hypervolume indicator, the bias of the indicator and the influence of the weight function w on optimal μ-distributions in particular has not been fully understood.

In this section, we first prove the existence of optimal μ-distributions for lower semi-continuous fronts, we show the monotonicity in μ of the hypervolume associated with optimal μ-distributions, and derive necessary conditions satisfied by optimal μ-distributions. In a second part, we derive the density associated with optimal μ-distributions when μ grows to infinity.

3.1. Finite Number of Points

3.1.1. Existence of Optimal μ-Distributions

Before to further investigate optimal μ-distributions for $I_{H,w}$, we establish a setting ensuring their existence. We will from now on assume that D is a closed interval that we denote $[x_{\min}, x_{\max}]$ such that f writes:

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

A function is lower semi-continuous if for all x_0, $\liminf_{x \to x_0} f(x) \geq f(x_0)$. If f is decreasing (which is the case when f describes a Pareto front), lower semi-continuous is equivalent to continuity to the right. As shown in the following theorem, a sufficient setting for the existence of optimal distributions is the lower semi-continuity of f.

Theorem 1 (Existence of optimal μ-distributions)**.** *Let $\mu \in \mathbb{N}$, if the function f describing the Pareto front is lower semi-continuous, there exists (at least) one set of μ points maximizing the hypervolume.*

Proof. We are going to prove that $I_{H,w}$ is upper semi-continuous if f is lower semi-continuous, and then apply the Extreme Value Theorem. Since $I_{H,w}$ is the sum of μ functions $g(x_i, x_{i+1})$ where $g(\alpha,\beta) = \int_\alpha^\beta \left(\int_{f(\alpha)}^{f(x_0)} w(x,y)dy \right) dx$, we will prove the upper semi-continuity of $g(x_i, x_{i+1})$ for $(x_i, x_{i+1}) \in [x_{\min}, x_{\max}]$. This will imply the upper semi-continuity of $I_{H,w}$ (Bourbaki, 1989, p 362). Let $(x_i, x_{i+1}) \in [x_{\min}, x_{\max}]$ and let $(x_i^n, x_{i+1}^n)_{n\in\mathbb{N}}$ converging to (x_i, x_{i+1}). We will now prove that $\limsup g(x_i^n, x_{i+1}^n) \leq g(x_i, x_{i+1})$ (see Knapp, 2005, p 481). Since

$$\limsup_{n\to\infty} g(x_i^n, x_{i+1}^n) = \limsup_{n\to\infty} \int\int \mathbf{1}_{[x_i^n, x_{i+1}^n]}(x) \mathbf{1}_{[f(x_i^n), f(x_0)]}(y) w(x,y) dy dx \ ,$$

and $\mathbf{1}_{[x_i^n, x_{i+1}^n]}(x)\mathbf{1}_{[f(x_i), f(x_0)]}(x)w(x,y) \leq \mathbf{1}_{[x_{\min}, x_{\max}]}(x)\mathbf{1}_{[f(x_{\max}), f(x_0)]}(x)w(x,y)$ we can use the (Reverse) Fatou Lemma (Knapp, 2005, p 252) that implies $\limsup g(x_i^n, x_{i+1}^n) \leq \int\int \limsup \mathbf{1}_{[x_i^n, x_{i+1}^n]}(x)\mathbf{1}_{[f(x_i^n), f(x_0)]}(y)w(x,y)dydx$. Since f is lower semi-continuous, $\liminf f(x_i^n) \geq f(x_i)$ holds which is equivalent to $\limsup(f(x_0) - f(x_i^n)) = f(x_0) - \liminf f(x_i^n) \leq f(x_0) - f(x_i)$. Hence, $\limsup \mathbf{1}_{[f(x_i^n), f(x_0)]}(y) \leq \mathbf{1}_{[f(x_i), f(x_0)]}(y)$ and thus

$$\limsup_{n\to\infty} g(x_i^n, x_{i+1}^n) \leq \int\int \mathbf{1}_{[x_i, x_{i+1}]}(x)\mathbf{1}_{[f(x_i), f(x_0)]}(y)w(x,y)dydx = g(x_i, x_{i+1}) \ .$$

We have proven the upper semi-continuity of g which implies the upper semi-continuity of $I_{H,w} : [x_{\min}, x_{\max}]^\mu \to \mathbb{R}$. Given that $[x_{\min}, x_{\max}]^\mu$ is compact, we can imply from the Extreme Value Theorem that there exists a set of μ points maximizing the hypervolume indicator. □

Note that, in case of bi-objective *maximization* problems, the lower semi-continuity of f has to be changed into upper semi-continuity which has been proven recently for the unweighted hypervolume (Bringmann and Friedrich, 2010). Note also that the previous theorem states the existence but not the uniqueness, which cannot be guaranteed in general. With this respect, we would like to mention that the question of uniqueness is related loosely to another property of the hypervolume which is not discussed here but has high importance in practice: For indicator-based

algorithms and the analysis of their convergence speed, it is highly important whether *local* optima are observed during the search. This property is, however, defined within the decision space X and especially depends on the mapping between the decision space and the objective space which is not taken into account in this study.

Furthermore, if the front is not semi-continuous, optimal μ-distributions might not exist. In the following proposition, we construct an example of a front where this is the case, i.e., where there is *no* optimal μ-distribution for $\mu = 1$.

Proposition 1. *Let $r = (r_1, r_1)$ be a reference point with $r_1 > 1.2$. Consider the front $f_{\text{ce}} : [0, 1] \to [0, 1.2]$ with*

$$f_{\text{ce}}(x) = \begin{cases} 1 - x + 0.2 & \text{if } x \leq \frac{1}{2}, \\ 1 - x & \text{if } x \in]\frac{1}{2}, 1] \end{cases} .$$

Then f does not admit an optimal 1-distribution for the unweighted hypervolume.

Proof. Consider first the linear front $f : x \in [0, 1] \to [0, 1], x \mapsto 1 - x$. Here, the optimal 1-distribution is the point $(0.5, 0.5)$ with a corresponding hypervolume value of $\gamma = (r_1 - \frac{1}{2})(r_1 - \frac{1}{2})$ [7]. Consider now $h(x) = f_{\text{ce}}(x)$ for all $x \in [0, 1]$ except for $x = 0.5$ where $h(x) = 0.5$. Then, h is continuous to the right and thus lower semi-continuous. Hence, according to Theorem 1 it admits an optimal 1-distribution. In addition, remark that the hypervolume contribution for any $x \in [0, 0.5[$ is strictly smaller for h than for f and equal for $x \in [0.5, 1]$. Thus $(0.5, 0.5)$ is also the optimal 1-distribution of h with hypervolume γ. However, for f_{ce}, the hypervolume contribution is strictly smaller than for f for $x \in [0, 0.5]$ and equal for $x \in]0.5, 1]$ with a gap at 0.5 such that γ cannot be reached for any point in $[0, 1]$ though one has values arbitrary close from it for x arbitrary close from 0.5 to the right. □

We have chosen $\mu = 1$ in the previous proposition for the sake of simplicity, however, such a counter-example can be generalized for arbitrary μ by following the same idea. Let us also note that, lower semi-continuity is not a necessary condition for the existence of optimal μ-distributions: if we simply introduce the discontinuity of the function f_{ce} in the previous proposition somewhere in $]0, 0.5[$ instead of at $x = 0.5$, the optimal 1-distribution would exist (and be located at $x = 0.5$) though the function describing the front is not lower semi-continuous.

3.1.2. Strict Monotonicity of Hypervolume in μ for Optimal μ-Distributions

The following proposition establishes that the hypervolume of optimal $(\mu + 1)$-distributions is strictly larger than the hypervolume of optimal μ-distributions. This result is a generalization of (Auger et al., 2009b, Lemma 1).

Proposition 2. *Let $D \subseteq \mathbb{R}$, possibly finite and $f : x \in D \mapsto f(x)$ describe a Pareto front. Let μ_1 and $\mu_2 \in \mathbb{N}$ with $\mu_1 < \mu_2$, then*

$$\overline{I_{H,w}^{\mu_1}} < \overline{I_{H,w}^{\mu_2}}$$

holds if D 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 proposition, it suffices to show the inequality for $\mu_2 = \mu_1 + 1$. Assume $D_{\mu_1} = \{\overline{x}_1^{\mu_1}, \ldots, \overline{x}_{\mu_1}^{\mu_1}\}$ with $\overline{x}_i^{\mu} \in \mathbb{R}$ is the set of x-values of the objective vectors of the optimal μ_1-distribution for $I_{H,w}$ with a hypervolume value of $\overline{I_{H,w}^{\mu_1}}$ if the Pareto front is described by f. Since D contains at least $\mu_1 + 1$ elements, the set $D \backslash D_{\mu_1}$ is not empty and we can pick any $x_{\text{new}} \in D \backslash D_{\mu_1}$ that is not contained in the optimal μ_1-distribution for $I_{H,w}$ and for which $f(x_{\text{new}})$ is defined. Let $x_r := \min\{x | x \in D_{\mu_1} \cup \{r_1\}, x > x_{\text{new}}\}$ 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 $f_l := \min\{r_2, \{f(x) | x \in D_{\mu_1}, x < x_{\text{new}}\}\}$ 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 $H_{\text{new}} := [x_{\text{new}}, x_r[\times [f(x_{\text{new}}), f_l[$ are weakly dominated by the new point $(x_{\text{new}}, f(x_{\text{new}}))$ 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 $x_{\text{new}} > x_r$ and $f_l > f(x_{\text{new}})$ and the weight w is strictly positive which gives $\overline{I_{H,w}^{\mu_1}} < \overline{I_{H,w}^{\mu_2}}$. □

[7] In case $\mu = 1$ and $f(x) = 1 - x$, we can easily compute the maximum of the hypervolume $I_{H,w}(x) = (r_1 - x)(r_1 - (1 - x)) = r_1^2 - r_1 + x - x^2$ of the single point at x by computing the derivative of $I_{H,w}(x)$ and setting it to zero: $I'_{H,w}(x) = 1 - 2x = 0$.

3.1.3. Characterization of Optimal μ-Distributions for Finite μ

In this section, we derive a general result to characterize optimal μ-distributions for the hypervolume indicator if μ is finite. The result holds under the assumption that the front f is differentiable and is a direct application of the fact that solutions of a maximization problem that do not lie on the boundary of the search domain are stationary points, i.e., points where the gradient is zero.

Theorem 2 (Necessary conditions for optimal μ-distributions for $I_{H,w}$)**.** *If f is continuous and differentiable and $(\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu)$ are the x-coordinates of an optimal μ-distribution for $I_{H,w}$, then for all $\overline{x}_i^\mu$ with $\overline{x}_i^\mu > x_{\min}$ and $\overline{x}_i^\mu < x_{\max}$*

$$f'(\overline{x}_i^\mu) \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} w(x, f(\overline{x}_i^\mu))dx = \int_{f(\overline{x}_{i-1}^\mu)}^{f(\overline{x}_i^\mu)} w(\overline{x}_i^\mu, y)dy \tag{5}$$

holds where f' denotes the derivative of f, $f(\overline{x}_0^\mu) = r_2$ and $\overline{x}_{\mu+1}^\mu = r_1$.

Proof. The proof idea is simple: optimal μ-distributions maximize the μ-dimensional function $I_{H,w}$ defined in Eq. 3 and should therefore satisfy necessary conditions for local extrema of a μ-dimensional function stating that the coordinates of local extrema either lie on the boundary of the domain (here $x_{\min}$ or $x_{\max}$) or satisfy that the partial derivative with respect to this coordinate is zero. Hence, we see that the partial derivatives of $I_{H,w}$ have to be computed. This step is quite technical and is presented in Appendix 7.1 on page 22 together with the full proof of the theorem. □

The previous theorem proves an implicit relation between the points of an optimal μ-distribution. However, in certain cases of weights, this implicit relation can be made explicit as illustrated first on the example of the weight function $w(x, y) = \exp(-x)$, aiming at favoring points with small values along the first objective.

Example 2. *If $w(x, y) = \exp(-x)$, Eq. 5 simplifies into the explicit relation*

$$f'(\overline{x}_i^\mu)(e^{-\overline{x}_i^\mu} - e^{-\overline{x}_{i+1}^\mu}) = e^{-\overline{x}_i^\mu}(f(\overline{x}_i^\mu) - f(\overline{x}_{i-1}^\mu)) \ . \tag{6}$$

Another example where the relation is explicit is given for the unweighted hypervolume I_H that we can obtain as a corollary of the previous theorem and which coincides with a previous result (Auger et al., 2009b, Proposition 1).

Corollary 1. *(Necessary condition for optimal μ-distributions for I_H) If f is continuous, differentiable and $(\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu)$ are the x-coordinates of an optimal μ-distribution for I_H, then for all $\overline{x}_i^\mu$ with $\overline{x}_i^\mu > x_{\min}$ and $\overline{x}_i^\mu < x_{\max}$*

$$f'(\overline{x}_i^\mu)(\overline{x}_{i+1}^\mu - \overline{x}_i^\mu) = f(\overline{x}_i^\mu) - f(\overline{x}_{i-1}^\mu) \tag{7}$$

holds where f' denotes the derivative of f, $f(\overline{x}_0^\mu) = r_2$ and $\overline{x}_{\mu+1}^\mu = r_1$.

Proof. The proof follows immediately from setting $w = 1$ in Eq. 5. □

Remark 2. *Corollary 1 implies that the points of an optimal μ-distribution for I_H are linked by a second order recurrence relation. Thus, in this case, finding optimal μ-distributions for I_H does not correspond to solving a μ-dimensional optimization problem as stated in Remark 1 but to a 2-dimensional one. The same remark holds for $I_{H,w}$ and $w(x, y) = \exp(-x)$ as can be seen in Eq. 6.*

The previous corollary can also be used to characterize optimal μ-distributions for certain Pareto fronts more generally as the following example shows.

Example 3. *Consider a linear Pareto front, i.e., a front that can be formally defined as $f : x \in [x_{\min}, x_{\max}] \mapsto \alpha x + \beta$ where $\alpha < 0$ and $\beta \in \mathbb{R}$. Then, it follows immediately from Corollary 1 and Eq. 7 that the optimal μ-distribution for I_H maps to objective vectors with equal distances between two neighbored solutions (see also Theorem 7 in Sec. 5.1):*

$$\alpha\left(\overline{x}_{i+1}^\mu - \overline{x}_i^\mu\right) = f(\overline{x}_i^\mu) - f(\overline{x}_{i-1}^\mu) = \alpha(\overline{x}_i^\mu - \overline{x}_{i-1}^\mu)$$

for $i = 2, \ldots, \mu - 1$. Note that this result coincides with earlier results for linear fronts with slope $\alpha = -1$ (Beume et al., 2007a) or the even more specific case of a front of shape $f(x) = 1 - x$ (Emmerich et al., 2007).

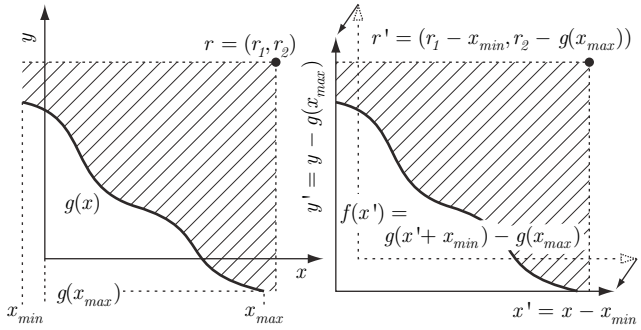


Figure 3: Every continuous front $g(x)$ (left) can be described by a function $f : x' \in [0, x'_{max}] \mapsto f(x')$ with $f(x'_{max}) = 0$ (right) by a simple translation.

3.2. Number of Points Going to Infinity

Besides for simple fronts, like the linear one, Eq. 5 and Eq. 7 cannot be easily exploited to derive optimal μ-distributions explicitly. However, one is interested in knowing how the hypervolume indicator influences the spread of points on the front and in characterizing the bias introduced by the hypervolume. To reply to these questions, we will assume that the number of points μ grows to infinity and derive the density of points associated with optimal μ-distributions for the hypervolume indicator.

We assume without loss of generality that $x_{\min} = 0$ and that $f : x \in [0, x_{\max}] \mapsto f(x)$ with $f(x_{\max}) = 0$ (Fig. 3). We also assume that f is continuous within $[0, x_{\max}]$, differentiable, and that its derivative is a continuous function f' defined in the interval $]0, x_{\max}[$. Instead of maximizing the weighted hypervolume indicator $I_{H,w}$, it is easy to see that, since $r_1 r_2$ is constant, one can equivalently minimize

$$r_1 r_2 - I_{H,w}((x_1, \ldots, x_\mu)) = \sum_{i=0}^{\mu} \int_{x_i}^{x_{i+1}} \int_0^{f(x_i)} w(x, y)\, dy\, dx$$

with $x_0 = 0$, $f(x_0) = r_2$, and $x_{\mu+1} = r_1$ (see Fig. 4(b)). If we subtract the area below the front curve, i.e., the integral $\int_0^{x_{\max}} \left(\int_0^{f(x)} w(x, y) dy \right) dx$ of constant value (Fig. 4(c)), we see that minimizing

$$\sum_{i=0}^{\mu} \int_{x_i}^{x_{i+1}} \int_0^{f(x_i)} w(x, y)\, dy\, dx - \int_0^{x_{\max}} \int_0^{f(x)} w(x, y)\, dy\, dx \tag{8}$$

is equivalent to maximizing the weighted hypervolume indicator (Fig. 4(d)).

For a fixed integer μ, we now consider a sequence of μ ordered points $\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu$ in $[0, x_{\max}]$ 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., $\delta(x) = \lim_{\substack{\mu \to \infty \\ h \to 0}} \left(\frac{1}{\mu h} \sum_{i=1}^{\mu} \mathbf{1}_{[x, x+h[}(\overline{x}_i^\mu) \right)$. As explained above, maximizing the weighted hypervolume is equivalent to minimizing Eq. 8, which is also equivalent to minimizing

$$E_\mu = \mu \left[\sum_{i=0}^{\mu} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \left(\int_0^{f(\overline{x}_i^\mu)} w(x, y) dy \right) dx - \int_0^{x_{\max}} \left(\int_0^{f(x)} w(x, y) dy \right) dx \right] , \tag{9}$$

where we have multiplied Eq. 8 by μ to obtain a quantity that will converge to a limit when μ goes to ∞. Indeed Eq. 8 converges to 0 when μ increases. We now conjecture that the equivalence between minimizing E_μ and maximizing the hypervolume also holds for μ going to infinity. 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.

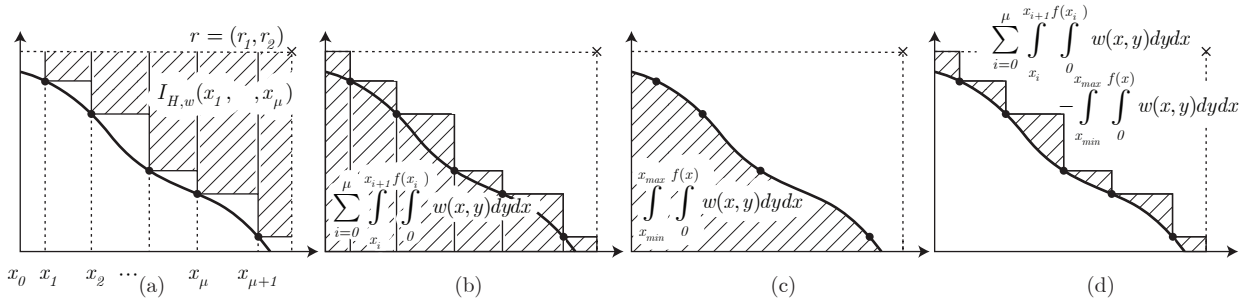


Figure 4: Illustration of the idea behind deriving the optimal density: Instead of maximizing the weighted hypervolume indicator $I_{H,w}((x_1, \ldots, x_\mu))$ (a), one can minimize the shaded area in (b) which is equivalent to minimizing the integral between the attainment surface of the solution set and the front itself which can be expressed with the help of the integral of f (d).

Lemma 1. *If f is continuous, differentiable with the derivative f' continuous, if $x \mapsto w(x, f(x))$ is continuous, if $\overline{x}_1^\mu, \ldots, \overline{x}_\mu^\mu$ converge to a continuous density δ, with $\frac{1}{\delta} \in L^2(0, x_{\max})$[8], and $\exists\, c \in \mathbb{R}^+$ such that*

$$\mu \sup\left(\left(\sup_{0 \le i \le \mu-1} |\overline{x}_{i+1}^\mu - \overline{x}_i^\mu|\right), |x_{\max} - \overline{x}_\mu^\mu|\right) \to c$$

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

$$E(\delta) := -\frac{1}{2} \int_0^{x_{\max}} \frac{f'(x) w(x, f(x))}{\delta(x)} dx \ . \tag{10}$$

Proof. For the technical proof, we refer to Appendix 7.2 on page 23. □

The limit density of a μ-distribution for $I_{H,w}$, as explained before, minimizes $E(\delta)$. It remains therefore to find the density which minimizes $E(\delta)$. This optimization problem is posed in a functional space and is also a constrained problem since the density δ has to satisfy the constraint $J(\delta) := \int_0^{x_{\max}} \delta(x) dx = 1$. The constraint optimization problem (P) that needs to be solved is summarized in:

$$\begin{aligned} &\text{minimize } E(\delta) \\ &\text{subject to } J(\delta) = 1 \ . \end{aligned} \tag{P}$$

In a similar way than Theorem 7 in (Auger et al., 2009b) where $-f'$ needs to be replaced everywhere by $-f'w$[9], we find that the density solution of the constraint optimization problem (P) equals

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

For $x_{\min} \neq 0$, the density reads

$$\delta(x) = \frac{\sqrt{-f'(x) w(x, f(x))}}{\int_{x_{\min}}^{x_{\max}} \sqrt{-f'(x) w(x, f(x))} dx} \ . \tag{11}$$

Remark 3. *The previous density corresponds to the density of points of the front projected onto the x-axis, however, if one is interested into the density on the front δ_F[10] one has to normalize the result from Eq. 11 by the norm of the*

[8] $L^2(0, x_{\max})$ is a functional space (Banach space) defined as the set of all functions whose square is integrable in the sense of the Lebesgue measure.

[9] Note that in (Auger et al., 2009b, Theorem 7) and its proof, the density should belong to $L^2(0, x_{\max})$ but also, $1/\delta \in L^2(0, x_{\max})$.

[10] The density on the front gives for any curve on the front (a piece of the front) C, the proportion of points of the optimal μ-distribution (for μ to infinity) contained in this curve by integration on the curve: $\int_C \delta_F ds$. Since we know that for any parametrization of C, say $t \in [a, b] \to \gamma(t) \in \mathbb{R}^2$, we have $\int_C \delta_F ds = \int_a^b \delta_F(\gamma(t)) \|\gamma'(t)\|_2 dt$, we can for instance use the natural parametrization of the front given by $\gamma(t) = (t, f(t))$ giving $\|\gamma'(t)\|_2 = \sqrt{1 + f'(t)^2}$ that therefore implies that $\delta(x) = \delta_F(x) \sqrt{1 + f'(x)^2}$. Note that we do a small abuse of notation writing $\delta_F(x)$ instead of $\delta_F(\gamma(x)) = \delta_F((x, f(x)))$.

tangent for points of the front, i.e., $\sqrt{1 + f'(x)^2}$*. Therefore, the density on the front is*

$$\delta_F(x) = \frac{\sqrt{-f'(x)w(x, f(x))}}{\int_{x_{\min}}^{x_{\max}} \sqrt{-f'(x)w(x, f(x))}dx} \frac{1}{\sqrt{1 + f'(x)^2}}. \tag{12}$$

Example 4. *Let us consider the test problem ZDT2 (Zitzler et al., 2000, see also Fig. 9) the Pareto front of which can be described by* $f(x) = 1 - x^2$ *with* $x_{\min} = 0$ *and* $x_{\max} = 1$ *and* $f'(x) = -2x$ *(Auger et al., 2009b). Considering the unweighted case, the density on the x-axis according to Eq. 11 is* $\delta(x) = \frac{3}{2}\sqrt{x}$ *and the density on the front according to Eq. 12 is* $\delta_F(x) = \frac{3}{2}\frac{\sqrt{x}}{\sqrt{1+4\,x^2}}$*, see Fig. 9 for an illustration.*

To summarize, we have seen that the density follows as a limit result from the fact that the integral between the attainment function of the solution set with μ points and the front itself (Fig. 4(d)) has to be minimized and the optimal μ-distribution for $I_{H,w}$ and a finite number of points converges to the density when μ increases. Furthermore, we can conclude that the proportion of points of an optimal μ-distribution with x-values within a certain interval $[a, b]$ converges to $\int_a^b \delta(x)dx$ if the number of points μ goes to infinity. How this relates to practice will be presented in Sec. 5 where analytical and experimental results on the density for specific well-known test problems are shown.

Instead of applying the results to specific test functions, the above results on the hypervolume indicator can also be interpreted in a broader sense: From (11), we know that it is only the weight function and the slope of the front that influences the density of the points of an optimal μ-distribution—contrary to several prevalent beliefs as stated in the beginning of this section. Since the density of points does not depend on the position on the front but only on the gradient and the weight at the respective point, the density close to the extreme points of the front can be very high or very low—it only depends on the front shape. Section 4.1.1 will even present conditions under which the extreme points will never be included in an optimal μ-distribution for $I_{H,w}$—in contrast to the statement by Beume et al. (2007b). In the unweighted case, we observe that the density has its maximum for front parts where the tangent has a gradient of -45° (see also Auger et al., 2009b). Therefore, and compliant with the statement by Beume et al. (2007b), optimizing the unweighted hypervolume indicator stresses so-called knee-points—parts of the Pareto front decision makers believe to be interesting regions (Das, 1999; Branke et al., 2004). However, choosing a non-constant weight can highly change the distribution of points and makes it possible to include several user preferences into the search. The new result in (11) now explains *how* the distribution of points changes: for a fixed front, it is the square root of the weight that is directly reflected in the optimal density.

4. Influence of the Reference Point on the Extremes

Clearly, optimal μ-distributions for $I_{H,w}$ are in some way influenced by the choice of the reference point r as the definition of $I_{H,w}$ in Eq. 3 depends on r and it is well-known from experiments that the reference point can influence the outcomes of multiobjective evolutionary algorithms drastically (Knowles et al., 2003). How in general, the outcomes of hypervolume-based algorithms are influenced by the choice of the reference point, however, has not been investigated from a theoretical perspective. In particular, it could not be observed from practical investigations how the reference point has to be set to ensure to find the extremes of the Pareto front.

In practice, mainly rules-of-thumb exist on how to choose the reference point. 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 (Knowles, 2005) or a box that is larger by an additive term of 1 than the extremal objective values obtained (Beume et al., 2007b). In various publications where the hypervolume indicator is used for performance assessment, the reference point is chosen as the nadir point[11] of the investigated solution set (Purshouse and Fleming, 2003; Purshouse, 2003; Hughes, 2005), while others recommend a rescaling of the objective values everytime the hypervolume indicator is computed (Zitzler and Künzli, 2004).

In this section, we ask the question of how the choice of the reference point influences optimal μ-distributions and theoretically investigate in particular whether there exists a choice for the reference point that implies that the

[11] In our notation, the nadir point equals $(x_{\max}, f(x_{\min}))$, i.e., is the smallest objective vector that is weakly dominated by all Pareto-optimal points.

extremes of the Pareto front are included in optimal μ-distributions. The presented results generalize the statements by Auger et al. (2009b) to the weighted hypervolume indicator and give insights into how the reference point should be chosen if the weight function does not equal 1 everywhere. Our main result, stated in Theorem 4 and Theorem 5, shows that for continuous and differentiable Pareto fronts we can give implicit lower bounds on the $\mathcal{F}_1$ and $\mathcal{F}_2$ value for the reference point (possibly infinite depending on f and w) such that all choices above this lower bound ensure the existence of the extremes in an optimal μ-distribution for $I_{H,w}$. For the special case of the unweighted hypervolume indicator, these lower bounds turn into explicit lower bounds (Corollaries 2 and 3). Moreover, Sec. 4.1.1 shows that it is necessary to have a finite derivative on the left extreme and a non-zero one on the right extreme to ensure that the extremes are contained in an optimal μ-distribution. 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 or the border of the objective space to obtain the extremes as indicated above. A new result (Theorem 6), not covered by Auger et al. (2009b), shows that a point slightly worse than the nadir point in all objectives starts to become a good choice for the reference point as soon as μ is large enough.

Before we present the results, 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 the optimal μ-distribution, we assume that the reference point is dominated by all Pareto-optimal solutions, i.e., $r_1 > x_{\max}$ and $r_2 > f(x_{\min})$.

4.1. Finite Number of Points

For the moment, we assume that the number of points μ is finite and provide necessary and sufficient conditions for finding a finite reference point such that the extremes are included in any optimal μ-distribution for $I_{H,w}$. In Sec. 4.2, we later on derive further results in case μ goes to infinity.

4.1.1. Fronts for Which It Is Impossible to Have the Extremes

A previous belief was that choosing the reference point of the hypervolume indicator in a way, such that it is dominated by all Pareto-optimal points, is enough to ensure that the extremes can be reached by an indicator-based algorithm aiming at maximizing the hypervolume indicator. The main reason for this belief is that with such a choice of reference point, the extremes of the Pareto front always have a positive contribution to the overall hypervolume indicator and should be therefore chosen by the algorithm's environmental selection. However, theoretical investigations revealed that we cannot always ensure that the extreme points of the Pareto front are contained in an optimal μ-distribution for the unweighted hypervolume indicator (Auger et al., 2009b). In particular, a necessary condition to have the left (resp. right) extreme included in optimal μ-distributions is to have a finite (resp. non-zero) derivative on the left extreme (resp. right extreme). The following theorem generalizes this result and shows that also for the weighted hypervolume indicator, the same necessary condition holds.

Theorem 3. *Let μ be a positive integer. Assume that f is continuous on $[x_{\min}, x_{\max}]$, non-increasing, differentiable on $]x_{\min}, x_{\max}[$ and that f' is continuous on $]x_{\min}, x_{\max}[$ and that the weight function w is continuous and positive. If $\lim_{x \to x_{\min}} f'(x) = -\infty$, the left extremal point of the front is never included in an optimal μ-distribution for $I_{H,w}$. Likewise, if $f'(x_{\max}) = 0$, the right extremal point of the front is never included in an optimal μ-distribution for $I_{H,w}$.*

Proof. The idea behind the proof is to assume the extreme point to be contained in an optimal μ-distribution and to show a contradiction. In particular, the gain and loss in hypervolume if the extreme point is shifted can be computed analytically. A limit result for the case that $\lim_{x \to x_{\min}} f'(x) = -\infty$ (and $f'(x_{\max}) = 0$ respectively) shows that one can always increase the overall hypervolume indicator value if the outmost point is shifted, see also Fig. 11. For the technical details, including a technical lemma, we refer to Appendix 7.3 on page 25. □

Example 5. *Consider the test problem ZDT1 (Zitzler et al., 2000) with a Pareto front described by $f(x) = 1 - \sqrt{x}$ with $x_{\min} = 0$ and $x_{\max} = 1$, see Figure 9(a). The derivative $f'(x) = -1/(2\sqrt{x})$ equals $-\infty$ at the left extreme $x_{\min}$ and the left extreme is therefore never included in an optimal μ-distribution for $I_{H,w}$ according to Theorem 3.*

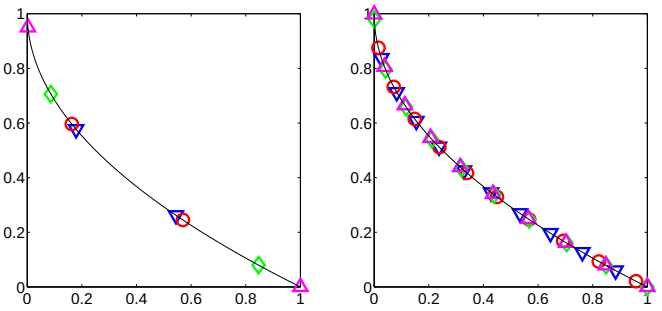


Figure 5: Influence of the choice of the reference point $r = (r_1, r_2)$ on optimal 2- (left) and optimal 10-distributions on the ZDT1 problem, in particular on the left extreme. Shown are the best approximations found within 100 CMA-ES runs for $r = (1.01, 1.01)$ ($\triangledown$), $r = (1.1, 1.1)$ ($\bigcirc$), $r = (2, 2)$ ($\Diamond$), and $r = (11, 11)$ ($\triangle$). Note that according to theory, the left extreme is never included in optimal μ-distributions and the lower bound on r_1 to ensure the right extreme is $\mathcal{R}_1 = 3$ (Auger et al., 2009b).

Although one should keep the previous result in mind when using the hypervolume indicator, the fact that the extreme can never be obtained in the cases of Theorem 3 is less restrictive in practice. Due to the continuous search space for most of the test problems, no algorithm will obtain a specific solution exactly—and the extreme in particular—and if the number of points is high enough, a solution close to the extreme[12] will be found also by hypervolume-based algorithms. However, if the number of points is low the choice of the reference point is crucial and choosing it too close to the nadir point will massively change the optimal μ-distribution as can be seen exemplary for the ZDT1 problem in Fig. 5[13]. Moreover, when using the weight function in the weighted hypervolume indicator to model preferences of the user towards certain regions of the objective search, one should pay attention to this fact by increasing the weight drastically close to such extremes if they are desired, see (Auger et al., 2009a) for examples.

4.1.2. Lower Bound for Choosing the Reference Point for Obtaining the Extremes

We have seen in the previous section that if the limit of the derivative of the front at the left extreme equals $-\infty$ (resp. if the derivative of the front at the right extreme equals zero) there is no choice of reference point that allows to have the extremes included in optimal μ-distributions for $I_{H,w}$. We assume now that the limit of the derivative of the front at the left extreme is finite (resp. the derivative of the front at the right extreme is not zero) and investigate conditions ensuring that there exists (finite) reference points ensuring to have the extremes in the optimal μ-distributions.

Lower Bound for Left Extreme.

Theorem 4 (Lower bound for left extreme)**.** *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 $]x_{\min}, x_{\max}[$ and $\lim_{x \to x_{\min}} -f'(x) < \infty$. If there exists a $\mathcal{K}_2 \in \mathbb{R}$ such that for all $x_1 \in]x_{\min}, x_{\max}]$*

$$\int_{f(x_1)}^{\mathcal{K}_2} w(x_1, y)dy > -f'(x_1) \int_{x_1}^{x_{\max}} w(x, f(x_1))dx \ , \tag{13}$$

then for all reference points $r = (r_1, r_2)$ such that $r_2 \geq \mathcal{K}_2$ and $r_1 > x_{\max}$, the leftmost extremal point is contained in optimal μ-distributions for $I_{H,w}$. In other words, defining $\mathcal{R}_2$ as

$$\mathcal{R}_2 = \inf\{\mathcal{K}_2 \textit{ satisfying Eq. } 13\} \ , \tag{14}$$

the leftmost extremal point is contained in optimal μ-distributions if $r_2 > \mathcal{R}_2$, and $r_1 > x_{\max}$.

Proof. This proof is presented in Appendix 7.4 on page 26. □

[12] Although the distance of solutions to the extremes might be sufficiently small in practice also for the scenario of Theorem 3, the theoretical result shows that for a finite μ, we cannot expect that the solutions approach the extremes arbitrarily close.

[13] The shown approximations of the optimal μ-distribution have been obtained by using the algorithm CMA-ES (Hansen and Kern, 2004, version 3.40beta with standard settings) to solve the 2-dimensional optimization problem of Remark 2 with the two leftmost points as variables and a boundary handling with penalties if the leftmost or rightmost point is outside $[x_{\min}, x_{\max}]$ (population size 20, best result over 100 runs shown).

Remark 4. *The previous theorem states only an implicit condition for $\mathcal{K}_2$ and it is not always obvious whether a finite $\mathcal{K}_2$ with the stated properties exists. There are different reasons for a non-existence of a finite $\mathcal{K}_2$—although we assume that $\lim_{x\to x_{\min}} -f'(x) < \infty$. One reason can be the fact that $f'(x_1)$ is infinite for some $x_1 \in]x_{\min}, x_{\max}]$ such that the right-hand side of Eq. 13 is not finite and therefore $\mathcal{K}_2$ cannot be finite as well. Example 6, however, shows an example where $f'(x_1) = -\infty$ for an $x_1 \in]x_{\min}, x_{\max}]$ and $\mathcal{K}_2$ is still finite. Another possible reason for the non-existence of a finite $\mathcal{K}_2$ can be a choice of w such that the left-hand side of Eq. 13 is always smaller than the right-hand side—even assuming that w is continuous does not prevent such a choice of w.*

We will now apply the previous theorem to the unweighted hypervolume and prove an *explicit* lower bound for setting the reference point so as to have the left extreme. This results recovers (Auger et al., 2009b, Theorem 2).

Corollary 2 (Lower bound for left extreme)**.** *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 $[x_{\min}, x_{\max}[$. Let us assume that $\lim_{x\to x_{\min}} -f'(x) < \infty$. If*

$$\mathcal{R}_2 = \sup\{f'(x)(x - x_{\max}) + f(x) : x \in]x_{\min}, x_{\max}]\} \tag{15}$$

is finite, then the leftmost extremal point is contained in optimal μ-distributions for I_H if the reference point $r = (r_1, r_2)$ is such that r_2 is strictly larger than $\mathcal{R}_2$ and $r_1 > x_{\max}$.

Proof. The proof is presented in Appendix 7.5 page 28. □

Example 6. *Consider again the DTLZ2 test function from Example 1 with $f(x) = \sqrt{1 - x^2}$ and $f'(x) = -\frac{x}{\sqrt{1-x^2}}$ where $x_{\min} = 0$ and $x_{\max} = 1$. Assume $w = 1$, i.e., the unweighted hypervolume indicator I_H. We see that $f'(x_{\max}) = -\infty$ but nevertheless, $\mathcal{R}_2$ is finite according to Eq. 15, namely*

$$\mathcal{R}_2 = \sup\left\{-\frac{x}{\sqrt{1 - x^2}}(x - x_{\max}) + \sqrt{1 - x^2} : x \in]x_{\min}, x_{\max}]\right\} = \sqrt{6\sqrt{3} - 9} \approx 1.18 \ ,$$

which can be obtained for example with a computer algebra system such as Maple.

Lower Bound for Right Extreme.
We now turn to the case of the right extreme and address the same question as for the left extreme: assuming that $f'(x_{\max}) \neq 0$, can we find an explicit lower bound for the first coordinate of the reference point ensuring that the right extreme is included in optimal μ-distributions? The following result holds.

Theorem 5 (Lower bound for right extreme)**.** *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 $]x_{\min}, x_{\max}[$ and $f'(x_{\max}) \neq 0$. If there exists a $\mathcal{K}_1 \in \mathbb{R}$ such that for all $x_\mu \in [x_{\min}, x_{\max}[$*

$$-f'(x_\mu) \int_{x_\mu}^{\mathcal{K}_1} w(x, f(x_\mu))dx > \int_{f(x_\mu)}^{f(x_{\min})} w(x_\mu, y)dy \ , \tag{16}$$

then for all reference points $r = (r_1, r_2)$ such that $r_1 \geq \mathcal{K}_1$ and $r_2 > f(x_{\min})$, the rightmost extremal point is contained in optimal μ-distributions. In other words, defining $\mathcal{R}_1$ as

$$\mathcal{R}_1 = \inf\{\mathcal{K}_1 \textit{ satisfying Eq. } 16\} \ , \tag{17}$$

the rightmost extremal point is contained in optimal μ-distributions if $r_1 > \mathcal{R}_1$, and $r_2 > f(x_{\min})$.

Proof. This proof is presented in Appendix 7.6 on page 28. □

We will now apply the previous theorem to the unweighted hypervolume and prove an explicit lower bound for setting the reference point so as to have the right extreme. This results recovers (Auger et al., 2009b, Theorem 2).

Corollary 3 (Lower bound for right extreme)**.** *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 $]x_{\min}, x_{\max}]$. If*

$$\mathcal{R}_1 = \sup\left\{x + \frac{f(x) - f(x_{\min})}{f'(x)} : x \in [x_{\min}, x_{\max}[\right\} \tag{18}$$

is finite, then the rightmost extremal point is contained in optimal μ-distributions for I_H if the reference point $r = (r_1, r_2)$ is such that $r_1 > \mathcal{R}_1$ and $r_2 > f(x_{\min})$.

Proof. The proof is presented in Appendix 7.7 page 29. □

4.2. Number of Points Going to Infinity

The lower bounds we have derived for the reference point such that the extremes are included are independent of μ. It can be seen in the proof that those bounds are not tight if μ is larger than 2. Deriving tight bounds is, however, difficult because it would require to know for a given μ where the second point of optimal μ-distributions is located. It can be certainly achieved in the linear case (see (Brockhoff, 2010)), but it might be impossible in more general cases. However, we want to investigate now how μ influences the choice of the reference point so as to have the extremes. In this section, we will denote $\mathcal{R}_1^{\text{Nadir}}$ and $\mathcal{R}_2^{\text{Nadir}}$ the first and second coordinates of the nadir point, namely $\mathcal{R}_1^{\text{Nadir}} = x_{\max}$ and $\mathcal{R}_2^{\text{Nadir}} = f(x_{\min})$.

We will prove that for any reference point dominated by the nadir point, there exists a μ_0 such that for all μ larger than μ_0, optimal μ-distributions associated to this reference point include the extremes in case the extremes can be contained in optimal μ-distributions, i.e., if $-f'(x_{\min}) < \infty$ and $f'(x_{\max}) < 0$. Before, we establish a lemma saying that if there exists a reference point R^1 allowing to have the extremes, then all reference points R^2 dominated by this reference point R^1 will also allow to have the extremes.

Lemma 2. *Let $R^1 = (r_1^1, r_2^1)$ and $R^2 = (r_1^2, r_2^2)$ be two reference points with $r_1^1 < r_1^2$ and $r_2^1 < r_2^2$. If both extremes are included in optimal μ-distributions for $I_{H,w}$ associated with R^1 then both extremes are included in optimal μ-distributions for $I_{H,w}$ associated with R^2.*

Proof. The proof is presented in Appendix 7.8 page 29. □

Theorem 6. *Let us assume that f is continuous, differentiable with f' continuous on $[x_{\min}, x_{\max}]$, $f'(x_{\max}) < 0$, and w is bounded, i.e., there exists $W > 0$ such that $w(x, y) \leq W$ for all (x, y). For all $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2_{>0}$,*

1. *there exists a μ_1 such that for all $\mu \geq \mu_1$, and any reference point R dominated by the nadir point such that $R_2 \geq \mathcal{R}_2^{\text{Nadir}} + \varepsilon_2$, the left extreme is included in optimal μ-distributions,*
2. *there exists a μ_2 such that for all $\mu \geq \mu_2$, and any reference point R dominated by the nadir point such that $R_1 \geq \mathcal{R}_1^{\text{Nadir}} + \varepsilon_1$, the right extreme is included in optimal μ-distributions.*

Proof. The proof is presented in Appendix 7.9 page 30. □

As a corollary, we obtain the following result for obtaining both extremes simultaneously:

Corollary 4. *Let us assume that f is continuous, differentiable with f' continuous on $[x_{\min}, x_{\max}]$, $f'(x_{\max}) < 0$, and w is bounded, i.e., there exists a $W > 0$ such that $w(x, y) \leq W$ for all (x, y). For all $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2_{>0}$, there exists a $\mu_0 \in \mathbb{N}$ such that for μ larger than μ_0 and for all reference points weakly dominated by $(\mathcal{R}_1^{\text{Nadir}} + \varepsilon_1, \mathcal{R}_2^{\text{Nadir}} + \varepsilon_2)$, both the left and right extremes are included in optimal μ-distributions.*

Proof. The proof is straightforward taking for μ_0 the maximum of μ_1 and μ_2 in Theorem 6. □

Theorem 6 and Corollary 4 state that for bi-objective Pareto fronts which are continuous on the interval $[x_{\min}, x_{\max}]$ and a bounded weight, we can expect to have the extremes in optimal μ-distributions for any reference point dominated by the nadir point if μ is large enough, i.e., larger than μ_0. Unfortunately, the proof does not allow to state how large μ_0 has to be chosen for a given reference point but it is expected that μ_0 depends on the reference point as well as on the front shape and w. Recently, for linear Pareto fronts, this dependency could be shown explicitly (Brockhoff, 2010) and we will briefly summarize this result in the following.

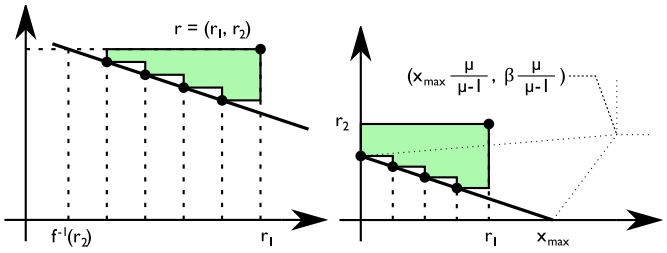


Figure 6: Optimal μ-distribution for $\mu = 4$ points and the unweighted hypervolume indicator if the reference point is not dominated by the extreme points of the Pareto front (Theorem 8, left) and in the most general case (Theorem 9, right) for a front with slope $f'(x) = \alpha = -\frac{1}{3}$. The dotted lines in the right plot limit the regions where the leftmost point, the rightmost point, or both are included in the optimal μ-distributions for $\mu = 4$ (see also Fig. 7).

5. Application to Multiobjective Test Problems

Besides being used within indicator-based algorithms, the hypervolume indicator has been also frequently used for performance assessment when comparing multiobjective optimizers—mainly because of its refinement property (Zitzler et al., 2010) and its resulting ability to map both information about the proximity of a solution set to the Pareto front and about the set's spread in objective space into a single scalar. Also here, knowing the optimal μ-distribution and its corresponding hypervolume value for certain test problems is crucial. On the one hand, knowing the largest hypervolume value obtainable by μ solutions allows to compare the achieved hypervolume values of different algorithms not only relatively but also absolutely in terms of the difference between the achieved and the achievable hypervolume value. On the other hand, only knowing the actual optimal μ-distributions for a certain test problem allows to investigate whether hypervolume-based algorithms really converge to their inherent optimization goal (or get stuck in local optima of (3) and (4)) which has not been investigated yet. In this section, we therefore apply the theoretical concepts derived in Sections 3 and 4 to several known test problems. First, we recapitulate results from (Auger et al., 2009b) and (Brockhoff, 2010) in Sec. 5.1 and investigate optimal μ-distributions for the unweighted hypervolume indicator I_H in case of a linear Pareto front. Then, we apply the results to the test function suites ZDT, DTLZ, and WFG in Sec. 5.2.

5.1. Linear Fronts

In this section, we have again a closer look at linear Pareto fronts, i.e., fronts 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. (2007a) (and later on Emmerich et al. (2007) for a more restricted front of shape $f(x) = 1 - x$) already proved that a set of μ points maximizes the unweighted hypervolume if and only if the points are equally spaced. However, the used proof techniques do not allow to state where the leftmost and rightmost point have to be placed in order to maximize the hypervolume with respect to a certain reference point—an assumption that later results do not require (Auger et al., 2009b). We will recapitulate those recent results briefly and in particular show for linear fronts of arbitrary slope, how the—in this case unique—optimal μ-distribution for I_H looks like without making assumptions on the positions of extreme solutions.

First of all, we formalize the result of Example 3 that, as a direct consequence of Corollary 1, the distance between two neighbored solutions is constant for arbitrary linear fronts:

Theorem 7. *If the Pareto front is a (connected) line, the optimal μ-distribution with respect to the unweighted hypervolume indicator is such that the distance is the same between all neighbored solutions.*

Proof. Applying Eq. 7 to $f(x) = \alpha x + \beta$ implies that $\alpha(\overline{x}^\mu_{i+1} - \overline{x}^\mu_i) = f(\overline{x}^\mu_i) - f(\overline{x}^\mu_{i-1}) = \alpha(\overline{x}^\mu_i - \overline{x}^\mu_{i-1})$ for $i = 2, \ldots, \mu - 1$ and therefore the distance between consecutive points of the optimal μ-distribution for I_H 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 such that there exists (a unique) $\overline{x}^\mu_0 \in [x_{\min}, x_{\max}]$ with $\overline{x}^\mu_0 = f^{-1}(r_2)$, the exact position of the optimal μ-distribution for I_H on the linear front can be determined, see also the left plot of Fig. 6:

Theorem 8. *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, the optimal μ-distribution with respect to the unweighted hypervolume indicator is unique and satisfies for all $i = 1, \ldots, \mu$*

$$\overline{x}^\mu_i = f^{-1}(r_2) + \frac{i}{\mu + 1} \cdot (r_1 - f^{-1}(r_2)) \quad . \tag{19}$$

Proof. From Eq. 7 and the previous proof we know that $\alpha\left(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i}\right) = f(\overline{x}^{\mu}_{i}) - f(\overline{x}^{\mu}_{i-1}) = \alpha(\overline{x}^{\mu}_{i} - \overline{x}^{\mu}_{i-1})$, for $i = 1, \ldots, \mu$ while $f(\overline{x}^{\mu}_{0}) = r_2$ and $\overline{x}^{\mu}_{\mu+1} = r_1$ are defined as in Corollary 1; in other words, the distances between $\overline{x}^{\mu}_{i}$ and its two neighbors $\overline{x}^{\mu}_{i-1}$ and $\overline{x}^{\mu}_{i+1}$ are the same for each $1 \leq i \leq \mu$. Therefore, the points $(\overline{x}^{\mu}_{i})_{1 \leq i \leq \mu}$ partition the interval $[\overline{x}^{\mu}_{0}, \overline{x}^{\mu}_{\mu+1}]$ into $\mu + 1$ sections of equal size and we obtain Eq. 19. □

Although Theorem 8 proves the exact unique positions of the μ points maximizing the unweighted hypervolume indicator in the restricted case where the reference point r is not dominated by the extremes of the front, the result can be used to obtain the exact distributions also in the most general case for any reasonable[14] choice of the reference point and any $\mu \in \mathbb{N}$ if the linear front is defined in the interval $[0, x_{\max}]$ (Brockhoff, 2010)[15].

Theorem 9 (Brockhoff (2010))**.** *Given $\mu \in \mathbb{N}_{\geq 2}$, $\alpha \in \mathbb{R}_{<0}$, $\beta \in \mathbb{R}_{>0}$, and a linear Pareto front $f(x) = \alpha x + \beta$ within $[0, x_{\max} = -\frac{\beta}{\alpha}]$, the unique optimal μ-distribution $(\overline{x}^{\mu}_{1}, \ldots, \overline{x}^{\mu}_{\mu})$ for the unweighted hypervolume indicator I_H with reference point $(r_1, r_2) \in \mathbb{R}^2_{>0}$ can be described by*

$$\overline{x}^{\mu}_{i} = f^{-1}(F_l) + \frac{i}{\mu + 1}\left(F_r - f^{-1}(F_l)\right) \tag{20}$$

for all $1 \leq i \leq \mu$ where

$$F_l = \min\left\{r_2,\ \frac{\mu+1}{\mu}\beta - \frac{1}{\mu}f(r_1),\ \frac{\mu}{\mu-1}\beta\right\} \quad \text{and} \quad F_r = \min\left\{r_1,\ \frac{\mu+1}{\mu}x_{\max} - \frac{1}{\mu}f^{-1}(r_2),\ \frac{\mu}{\mu-1}x_{\max}\right\}$$

if the reference point is dominated by at least one Pareto-optimal point.

Proof. The proof idea is the following. We can elongate the linear front beyond $x_{\min}$ and $x_{\max}$ and use the result of Theorem 8 to obtain the optimal placement dependent on r_1 and r_2—keeping in mind that all points are restricted to the interval $[x_{\min}, x_{\max}]$. In case r_1 and r_2 are too far away from the nadir point $(x_{\max}, \beta)$ such that Theorem 8 gives us $\overline{x}^{\mu}_{1} < x_{\min}$ or $\overline{x}^{\mu}_{\mu} > x_{\max}$, we have to make sure that these constraints are fulfilled by restricting the values F_l and F_r in Eq. 20 accordingly. For the details, we refer to (Brockhoff, 2010) due to space limitations. □

Right from the technicalities in the proof of Theorem 9 we see for which choices of the reference point the left and/or the right extreme are contained in the optimal μ-distribution.

Corollary 5. *Given $\mu \in \mathbb{N}_{\geq 2}$, $\alpha \in \mathbb{R}_{<0}$, $\beta \in \mathbb{R}_{>0}$, and a linear Pareto front $f(x) = \alpha x + \beta$ within $[0, x_{\max} = -\frac{\beta}{\alpha}]$,*

- *the left extreme point $(0, \beta)$ is included in the optimal μ-distribution for the unweighted hypervolume indicator if the reference point $(r_1, r_2) \in \mathbb{R}^2_{>0}$ lies above the line $L(x) = \frac{\mu+1}{\mu}\beta - \frac{1}{\mu}f(x) = \beta - \frac{\alpha}{\mu}x$ or if $r_2 > \frac{\mu}{\mu-1}\beta$ and*
- *the right extreme point $(x_{\max}, 0)$ is included if the reference point lies below the line $R(x) = \frac{\mu+1}{\mu}x_{\max} - \frac{1}{\mu}f^{-1}(x) = -\alpha\mu x - \mu\beta$ or if $r_1 > \frac{\mu}{\mu-1}x_{\max}$.*

Figure 7 gives an example for the front $f(x) = 2 - \frac{x}{3}$ and shows the regions within which the reference point ensures the left and/or the right extreme of the front for various choices of μ. Note that in the specific case of linear Pareto fronts, we not only know that the reference point to obtain both extremes approaches the nadir point if μ goes to infinity as proven in Sec. 4.2 but with the previous corollary, we also know *how fast* this happens.

As pointed out before, we do not know in general whether an optimal μ-distribution for a given indicator is unique or not. The example of a linear front is a case where we can ensure the uniqueness due to the concavity of the hypervolume indicator (Beume et al., 2009). Note also that besides for linear fronts, only one front shape is known so far for which we can also determine optimal μ-distributions exactly: for front shapes of the form $f(x) = \beta/x$ with $\beta > 1$, $x_{\min} = -\beta$, and $x_{\max} = -1$ and when the reference point is in $(0, 0)$ (Friedrich et al., 2009). On the other hand, even in the case of convex Pareto fronts, examples are known where the hypervolume indicator is not concave anymore and therefore the uniqueness of optimal μ-distributions is not known (Beume et al., 2009).

[14] Again, choosing the reference point such that it dominates Pareto-optimal points does not make sense as no solution will have positive hypervolume contributions.

[15] Assuming $x_{\min} = 0$ is not a restriction as the result for other choices of $x_{\min}$ can be derived by a simple coordinate transformation.

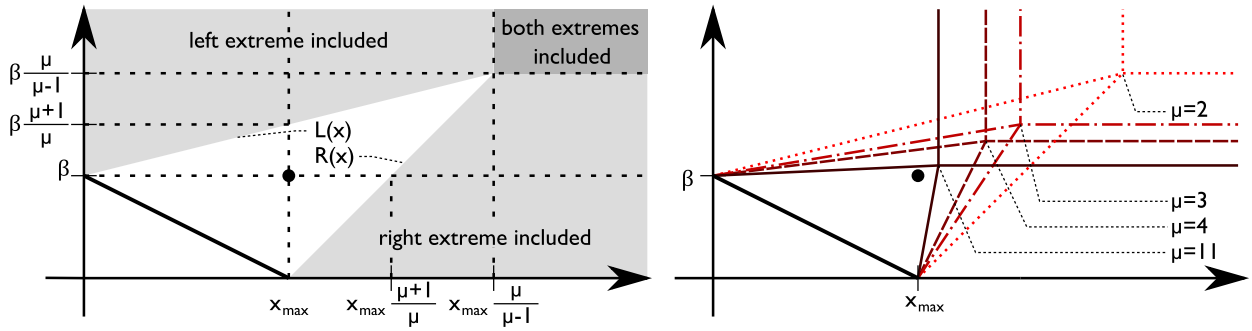


Figure 7: Influence of the reference point on the extremes for problems with linear Pareto fronts: the left plot shows the different regions within which the reference point ensures one (light gray), both (dark gray) or none (white) of the extremes in the optimal μ-distribution for $\mu = 2$ and the example front of $f(x) = 2 - \frac{x}{3}$. The right plot shows the borders of these regions for $\mu = 2$ (dotted), $\mu = 3$ (dash-dotted), $\mu = 4$ (dashed), and $\mu = 11$ (solid) for the same front. For clarity, the nadir point is shown as a black circle.

5.2. Test Function Suites ZDT, DTLZ, and WFG

In this section, we apply the presented results to problems in the ZDT (Zitzler et al., 2000), the DTLZ (Deb et al., 2005b), and the WFG (Huband et al., 2006) test function suites. All results are derived for the unweighted case of I_H, but they can also be derived for any other weight function $w(x, y) \neq 1$. In particular, we derive the function $f(x)$ describing the Pareto front and its derivative $f'(x)$ which directly leads to the density $\delta_F(x)$ with constant C. Furthermore, we derive a lower bound $\mathcal{R}$ for the choice of the reference point such that the extremes are included and compute an approximation of the optimal μ-distribution for $\mu = 20$ points. For the latter, the approximation schemes as proposed by Auger et al. (2009b) are used to get a precise picture for a given μ[16]. The densities and the lower bounds $\mathcal{R}$ for the reference point are obtained by the commercial computer algebra system Maple 12.0.

Figure 8 summarizes the results on the density and the lower bounds for the reference point for all investigated problems whereas we refer to the appendix for more detailed derivations (Appendix 7.10 presents the ZDT, Appendix 7.11 the DTLZ, and Appendix 7.12 the WFG results). Moreover, Fig. 9 shows a plot of the Pareto front, the obtained approximation of an optimal μ-distribution for $\mu = 20$, and the derived density $\delta_F(x)$ (as the hatched area on top of the front $f(x)$) for all investigated test problems.

The presented results show that for several of the considered test problems, analytical results for the density and the lower bounds for the reference point can be given easily—at least if a computer algebra system such as Maple is used. Otherwise, numerical results can be provided that approximate the mathematical results with an arbitrary high precision (up to the machine precision) which also holds for the approximations of the optimal μ-distributions shown in Fig. 9. Note that in the latter case, the approximation schemes used do not guarantee that the actual maximum of Eq. 3 and Eq. 4 is found as already discussed by Auger et al. (2009b). However, the distributions shown in Fig. 9 have been cross-checked by using the robust stochastic search optimizer CMA-ES (Hansen and Kern, 2004) in a similar manner as for the plots in Fig. 5. Moreover, the resulting optimal μ-distributions are independent of the starting conditions of the approximation schemes which is a strong indicator that the distributions found are indeed good approximations of the optimal distributions of μ points (Auger et al., 2009b).

Last, we give an additional interpretation of the density results: the density not only gives information about the bias of the hypervolume indicator for a given front, but can also be used to assess the number of solutions to be expected on a given segment of the front, as the following example illustrates.

Example 7. *Consider again ZDT2 as in Example 4. We would like to answer the question what is the fraction of points r_F of an optimal μ-distribution with the first and second objective being smaller or equal* 0.5 *and* 0.95 *respectively, see the highlighted front part in Figure 10. From $f^{-1}(y) = \sqrt{1-y}$ and $f^{-1}(0.95) = \sqrt{0.05}$ follows, that for the considered*

[16]For the test suites ZDT and DTLZ, additional approximations of the optimal μ-distribution for other typical numbers of points can be downloaded at `http://www.tik.ee.ethz.ch/sop/mudistributions`.

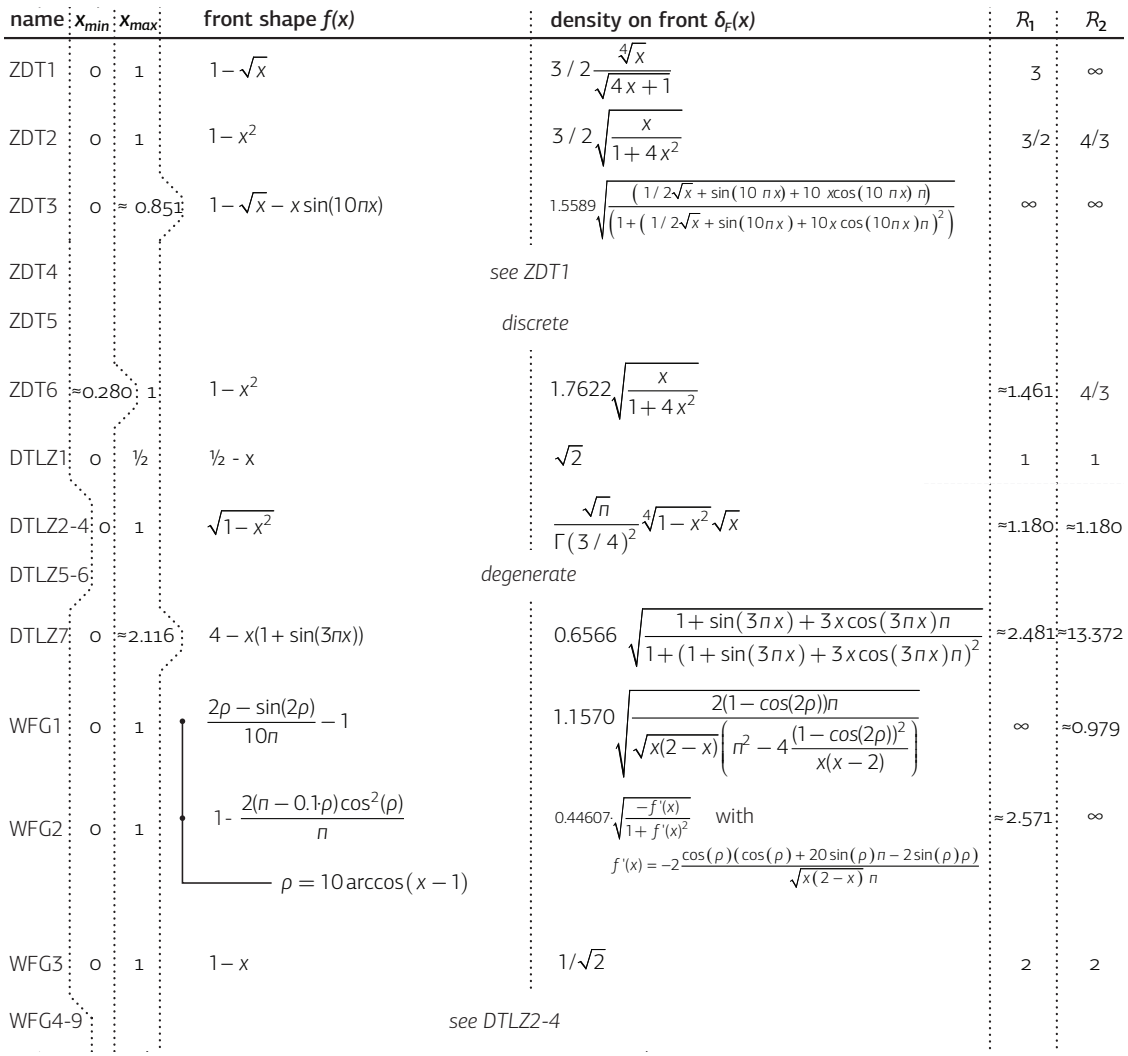


Figure 8: Lists for all ZDT, DTLZ, and WFG test problems and the unweighted hypervolume indicator I_H: (i) the Pareto front as $x \in [x_{\min}, x_{\max}] \mapsto f(x)$, (ii) the density $\delta_F(x)$ on the front according to Eq. 12, and (iii) a lower bound $\mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2)$ of the reference point to obtain the extremes (Eq. 18 and 15 respectively).

front segment $x \in [\sqrt{0.05}, 0.5]$ *holds. Using* $\delta(x)$ *given in Example 4 and integrating over* $[\sqrt{0.05}, 0.5]$ *yields:*

$$r_F = \int_{\sqrt{0.05}}^{0.5} \delta(x)dx = \int_{\sqrt{0.05}}^{0.5} \frac{3}{2}\sqrt{x}dx = \frac{1}{4}\sqrt{2} - 0.05^{3/4} \approx 24.78\% \ .$$

The same result can be obtained by taking the line integral of the density on the front over the considered front segment. Let $\delta_F^s(x, f(x)) := \delta_F(x)$ *denote the density on the front for a given point* $(x, f(x))$*, then* $r_F = \int_\gamma \delta_F^s(s)ds = \int_a^b \delta_F^s(\gamma(t))\ \|\dot{\gamma}(t)\|_2\ dt$ *where the path* γ *denotes the considered line segment on the front, i.e.,* $\gamma : [a = \sqrt{0.05}, b = 0.5] \to \mathbb{R}^2$*,* $t \mapsto (t, 1-t^2)$*. With* $\|\dot{\gamma}(t)\|_2 = \sqrt{1+f'(t)^2}$ *and* $\delta_F(\gamma(t)) = \delta_F(t)$ *we have* $r_F = \int_{\sqrt{0.05}}^{0.5} \delta_F(t)\sqrt{1+f'(t)^2}dt = \int_{\sqrt{0.05}}^{0.5} \delta(t)dt \approx 24.78\%$*. Note that for the approximated optimal* μ*-distribution of a finite number of* $\mu = 100$ *points*[17] *we obtained* 24 *points in the considered line segment, which is close to the predicted percentage of* $r_F = 24.78\%$*.*

[17] see `http://www.tik.ee.ethz.ch/sop/download/supplementary/testproblems/zdt2/data/mu100.txt`

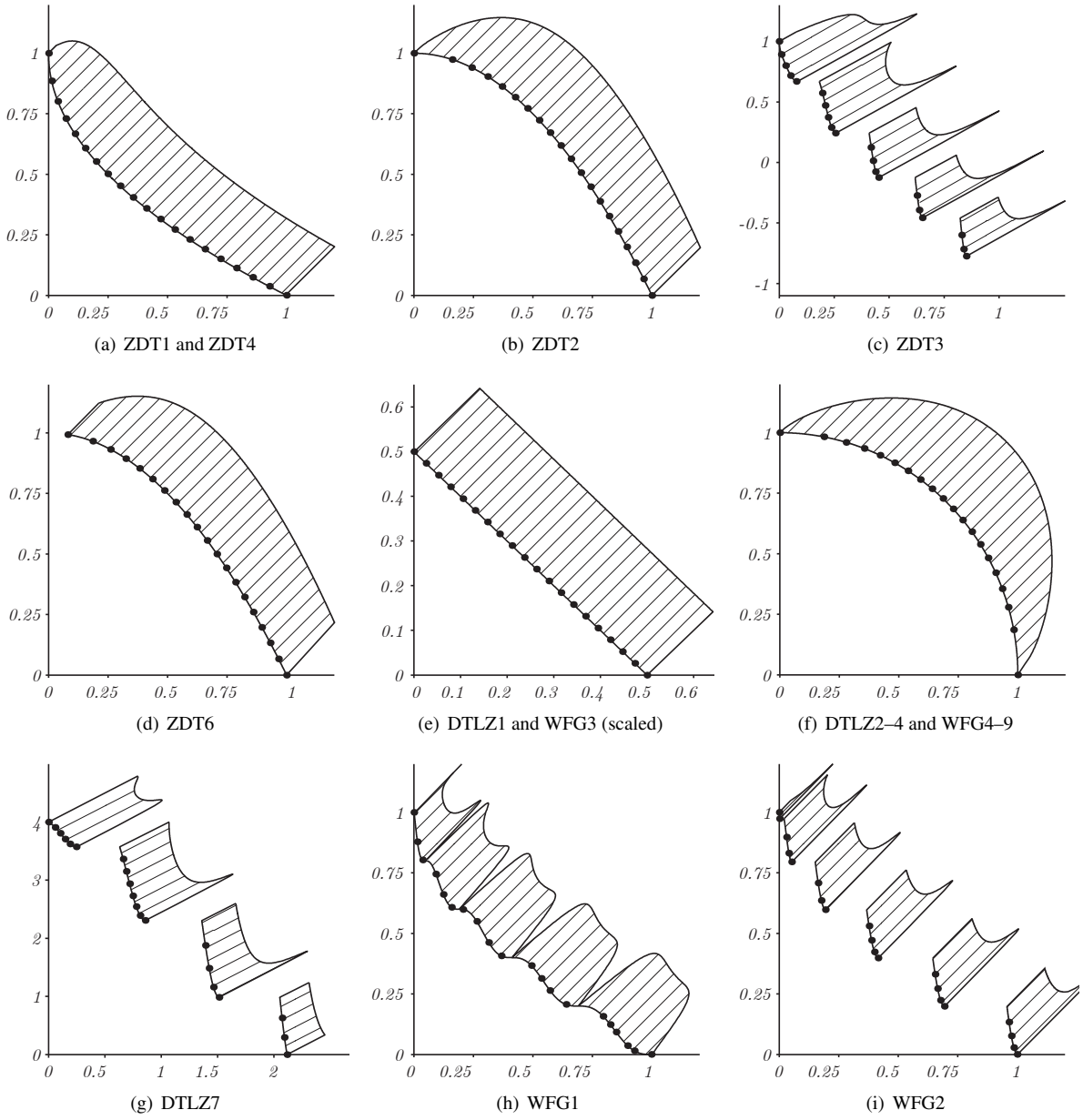


Figure 9: Pareto front shape $f(x)$, approximate optimal distribution of 20 points (black dots) for the unweighted hypervolume indicator, and the density $\delta_F(x)$ (hatched area) for different test problems.

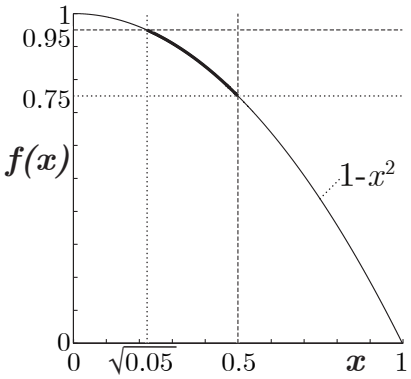


Figure 10: The density of points $\delta(x)$ and $\delta_F(x)$ can be used to assess the number of points to be expected in a given part of the front. The plot shows the thick line segment of the Pareto-front of ZDT2 for which $f(x) \leq 0.95$ and $x \leq 0.5$ hold, see Example 7.

6. Conclusions

Indicator-based evolutionary algorithms transform a multiobjective optimization problem into a single-objective one that corresponds to finding a set of μ points that maximizes the underlying quality indicator. Theoretically understanding these so-called *optimal μ-distributions* for a given indicator is a fundamental issue both for performance assessment of multiobjective optimizers and for the decision which indicator to take for the optimization in practice such that the search bias introduced by the indicator meets the user's preferences.

In this paper, we theoretically characterize optimal μ-distributions for the weighted hypervolume indicator in case of bi-objective problems. The results generalize previous work on the unweighted hypervolume indicator and are, in addition, applied to several known test problems. In particular, we investigate the sets of μ points that maximize the weighted hypervolume indicator and, besides general investigations for finite μ, we derive a limit result for μ going to infinity in terms of a density of points. Furthermore, we investigate the influence of the reference point on optimal μ-distributions, i.e., we derive lower bounds for placing the reference point (possibly infinite) for guaranteeing the Pareto front's extreme points in an optimal μ-distribution and investigate cases where the extremes are never contained in an optimal μ-distribution. In addition, we show that the belief, the best choice for the reference point corresponds to a point that is slightly worse than the nadir point in all objectives, can be founded theoretically if the number of points goes to infinity. Last, we apply the theoretical results to problems of the ZDT, DTLZ, and WFG test problem suites resulting in recommended choices of the reference point including numerical and sometimes analytical expressions for the resulting density of points on the front.

We believe the results presented in this paper are important for several reasons. On the one hand, we prove that several previous beliefs about the bias of the hypervolume indicator and the choice of the reference point to obtain the extremes of the front have been wrong. On the other hand, the results on optimal μ-distributions are highly useful in performance assessment if the hypervolume indicator is used as a quality measure. For the first time, approximations of optimal μ-distributions for finite μ allow to compare the outcome of indicator-based evolutionary algorithms to the actual optimization goal. Moreover, the actual hypervolume indicator of optimal μ-distributions (or the approximations we provide) offers a way to interpret the obtained hypervolume indicator values in an absolute fashion as the hypervolume of an optimal μ-distribution is a better estimate of the best achievable hypervolume than the hypervolume of the entire Pareto front. Last, we would like to mention that the presented results for the weighted hypervolume indicator also provide a basis for a better understanding of how to articulate user preferences with the weighted hypervolume indicator in terms of the question on how to choose the weight function in practice.

Acknowledgments

Dimo Brockhoff has been supported by the French national research agency (ANR) within the SYSCOMM project ANR-08-SYSC-017. The authors would like to thank Cyril Furtlehner for several helpful discussions regarding the general approach in Section 3.2.

7. Appendix

7.1. Proof of Theorem 2 stated on page 8

Before to prove the result, we rewrite Eq. 3 (page 5) in the following way

$$I_{H,w}(x_1, \ldots, x_\mu) = \sum_{i=1}^{\mu} g(x_i, x_{i+1}) \ , \tag{21}$$

where g is the 2-dimensional function defined as

$$g(\alpha, \beta) = \int_{\alpha}^{\beta} \left(\int_{f(\alpha)}^{f(x_0)} w(x, y) dy \right) dx \ . \tag{22}$$

The derivation of the gradient of $I_{H,w}$ thus relies on computing the partial derivatives of g. The following lemma gives the expressions of the partial derivatives of g:

Lemma 3. *Let w be a weight function for the weighted hypervolume indicator $I_{H,w}$ and $f : [x_{\min}, x_{\max}] \to \mathbb{R}$ be a continuous and differentiable function describing a 2-dimensional Pareto front. Let g be defined as $g(\alpha, \beta) = \int_{\alpha}^{\beta} \left(\int_{f(\alpha)}^{f(x_0)} w(x, y) dy \right) dx$ where $f(x_0) = r_2$. Then,*

$$\partial_1 g(\alpha, \beta) = -f'(\alpha) \int_{\alpha}^{\beta} w(x, f(\alpha)) dx - \int_{f(\alpha)}^{f(x_0)} w(\alpha, y) dy \tag{23}$$

$$\partial_2 g(\alpha, \beta) = \int_{f(\alpha)}^{f(x_0)} w(\beta, y) dy \tag{24}$$

Proof. To compute the first partial derivative of g, we need to compute the derivative of the function $g_1 : \alpha \to g(\alpha, \beta)$. Let us define $\gamma(l, m) = \int_{f(m)}^{f(x_0)} w(l, y) dy$ such that $g_1(\alpha) = \int_{\alpha}^{\beta} \gamma(x, \alpha) dx$. Define $K(\bar{x}, \bar{y}) = \int_{\bar{x}}^{\beta} \gamma(x, \bar{y}) dx$ and be $\Phi : \alpha \in \mathbb{R} \to (\alpha, \alpha) \in \mathbb{R}^2$. Then $g_1(\alpha) = K \circ \Phi(\alpha)$ such that we can apply the chain rule to find the derivative of g_1. Since g_1 maps $\mathbb{R}$ into $\mathbb{R}$, the differential of g_1 in α applied in h equals the derivative of g_1 in alpha times h. We thus have that for any $h \in \mathbb{R}$

$$g_1'(\alpha) h = (D_\alpha g_1)(h) = D_{\Phi(\alpha)} K \circ D_\alpha \Phi(h) \tag{25}$$

where $D_\alpha \Phi$ (resp. $D_{\Phi(\alpha)} K$) are the differential of Φ (resp. K) in α (resp. $\Phi(\alpha)$). We therefore need to compute the differentials of Φ and K. Since Φ is linear, $D_\alpha \Phi = \Phi$ and thus

$$D_\alpha \Phi(h) = (h, h) \ . \tag{26}$$

Moreover, the differential of K can be expressed with the partial derivatives of K, i.e., $D_{(\bar{x}, \bar{y})} K(h_1, h_2) = (\nabla K) \cdot (h_1, h_2)$ where ∇ is the vector differential operator $\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right) = (\partial_1, \ldots, \partial_n)$ and $(h_1, h_2) \in \mathbb{R}^2$. Hence,

$$D_{(\bar{x}, \bar{y})} K(h_1, h_2) = \partial_1 K(\bar{x}, \bar{y}) \, h_1 + \partial_2 K(\bar{x}, \bar{y}) \, h_2.$$

We thus need to compute the partial derivatives of K. From the fundamental theorem of calculus, $\partial_1 K(\bar{x}, \bar{y}) = -\gamma(\bar{x}, \bar{y})$. Besides, $\partial_2 K(\bar{x}, \bar{y}) = \int_{\bar{x}}^{\beta} \partial_2 \gamma(x, \bar{y}) dx$ and therefore

$$D_{(\bar{x}, \bar{y})} K(h_1, h_2) = -\gamma(\bar{x}, \bar{y}) h_1 + \left(\int_{\bar{x}}^{\beta} \partial_2 \gamma(x, \bar{y}) dx \right) h_2.$$

Applying again the fundamental theorem of calculus to compute the second partial derivative of γ, we find that

$$\partial_2 \gamma(x, \bar{y}) = -f'(\bar{y}) w(x, f(\bar{y}))$$

and thus

$$D_{(\bar{x}, \bar{y})} K(h_1, h_2) = \left(- \int_{f(\bar{y})}^{f(x_0)} w(\bar{x}, y) dy \right) h_1 + \left(\int_{\bar{x}}^{\beta} -f'(\bar{y}) w(x, f(\bar{y})) dx \right) h_2. \tag{27}$$

Combining Eq. 27 and Eq. 26 in Eq. 25 we obtain

$$\partial_1 g(\alpha,\beta) = g_1'(\alpha) = -f'(\alpha)\int_\alpha^\beta w(x,f(\alpha))dx - \int_{f(\alpha)}^{f(x_0)} w(\alpha,y)dy$$

which gives Eq. 23.

To compute the second partial derivative of g, we need to compute, for any α, the derivative of the function $g_2 : \beta \to g(\alpha,\beta)$. The function g_2 can be rewritten as $g_2 : \beta \to \int_\alpha^\beta \theta(x)dx$ where $\theta(x) = \int_{f(\alpha)}^{f(x_0)} w(x,y)dy$. Therefore, from the fundamental theorem of calculus we have that $\partial_2 g(\alpha,\beta) = g_2'(\beta) = \theta(\beta)$ and thus

$$\partial_2 g(\alpha,\beta) = \int_{f(\alpha)}^{f(x_0)} w(\beta,y)dy \ .$$

□

We are now ready to prove Theorem 2

Proof. From the first order necessary optimality conditions, we know that if $(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu)$ maximizes Eq. 3, then either $\overline{x}_i^\mu$ belongs to $]x_{\min},x_{\max}[$ and the i-th partial derivative of $I_{H,w}(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu)$ equals zero in $\overline{x}_i^\mu$, or $\overline{x}_i^\mu$ belongs to the boundary of $[x_{\min},x_{\max}]$, i.e., $\overline{x}_i^\mu = x_{\min}$ or $\overline{x}_i^\mu = x_{\max}$. Therefore, we need to compute the partial derivatives of $I_{H,w}$. From Eq. 21, we have that $\partial_1 I_{H,w}(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu) = \partial_1 g(\overline{x}_1^\mu,\overline{x}_2^\mu)$ and from Lemma 3 we therefore obtain that

$$\partial_1 I_{H,w}(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu) = -f'(\overline{x}_1^\mu)\int_{\overline{x}_1^\mu}^{\overline{x}_2^\mu} w(x,f(\overline{x}_1^\mu))dx - \int_{f(\overline{x}_1^\mu)}^{f(\overline{x}_0^\mu)} w(\overline{x}_1^\mu,y)dy$$

and thus if $\overline{x}_1^\mu \neq x_{\min}$ and $\overline{x}_1^\mu \neq x_{\max}$. By setting the previous equation to zero, we obtain

$$-f'(\overline{x}_1^\mu)\int_{\overline{x}_1^\mu}^{\overline{x}_2^\mu} w(x,f(\overline{x}_1^\mu)dx = \int_{f(\overline{x}_1^\mu)}^{f(\overline{x}_0^\mu)} w(\overline{x}_1^\mu,y)dy \ .$$

For $2 \leq i \leq \mu$, $\partial_i I_{H,w}(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu) = \partial_2 g(\overline{x}_{i-1}^\mu,\overline{x}_i^\mu) + \partial_1 g(\overline{x}_i^\mu,\overline{x}_{i+1}^\mu)$. Using Lemma 3 we obtain

$$\partial_i I_{H,w}(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu) = \int_{f(\overline{x}_{i-1}^\mu)}^{f(\overline{x}_0^\mu)} w(\overline{x}_i^\mu,y)dy - f'(\overline{x}_i^\mu)\int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} w(x,f(\overline{x}_i^\mu))dx - \int_{f(\overline{x}_i^\mu)}^{f(\overline{x}_0^\mu)} w(\overline{x}_i^\mu,y)dy \ .$$

Gathering the first and last term of the right-hand side, we obtain

$$\partial_i I_{H,w}(\overline{x}_1^\mu,\dots,\overline{x}_\mu^\mu) = \int_{f(\overline{x}_{i-1}^\mu)}^{f(\overline{x}_i^\mu)} w(\overline{x}_i^\mu,y)dy - f'(\overline{x}_i^\mu)\int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} w(x,f(\overline{x}_i^\mu))dx \tag{28}$$

and thus if $\overline{x}_{i+1}^\mu \neq x_{\min}$ and $\overline{x}_{i+1}^\mu \neq x_{\max}$, by setting the previous equation to zero, we obtain

$$\int_{f(\overline{x}_{i-1}^\mu)}^{f(\overline{x}_i^\mu)} w(\overline{x}_i^\mu,y)dy = f'(\overline{x}_i^\mu)\int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} w(x,f(\overline{x}_i^\mu))dx \ .$$

□

7.2. Proof of Lemma 1 stated on page 9

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

$$\int_0^{x_{\max}} \frac{|f'(x)w(x,f(x))|}{|\delta(x)|}dx \leq \sqrt{\int_0^{x_{\max}} (f'(x)w(x,f(x)))^2 dx \int_0^{x_{\max}} (1/\delta(x))^2 dx} \tag{29}$$

and since $x \to f'(x)w(x, f(x)) \in L^2(0, x_{\max})$ and $\frac{1}{\delta} \in L^2(0, x_{\max})$, the right-hand side of Eq. 29 is finite and Eq. 10 is well-defined. The proof is divided into two steps. First, we rewrite E_μ and, in a second step, the limit result is derived by using this new characterization of E_μ.

Step 1. In a first step we are going to prove that E_μ defined in Eq. 9 satisfies

$$E_\mu = \mu \sum_{i=0}^{\mu} \Big(-\tfrac{1}{2} f'(\overline{x}_i^\mu) w(\overline{x}_i^\mu, f(\overline{x}_i^\mu))(\overline{x}_{i+1}^\mu - \overline{x}_i^\mu)^2 + O((\overline{x}_{i+1}^\mu - \overline{x}_i^\mu)^3) \Big) \ . \tag{30}$$

To this end, we elongate the front to the right such that f equals $f(x_{\max}) = 0$ for $x \in [x_{\max}, \overline{x}_{\mu+1}^\mu]$. Like that, we can decompose $\int_0^{x_{\max}} \int_0^{f(x)} w(x,y) dy dx$ into $\sum_{i=0}^{\mu} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \int_0^{f(x)} w(x,y)\, dy\, dx$, while using the fact that $\int_{x_{\max}}^{\overline{x}_{\mu+1}^\mu} \int_0^{f(x)} w(x,y) dy dx = 0$. Using the right-hand side of the previous equation in Eq. 9, we find that

$$E_\mu = \mu \left[\sum_{i=0}^{\mu} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \left(\int_0^{f(\overline{x}_i^\mu)} w(x,y)\, dy \right) dx - \sum_{i=0}^{\mu} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \left(\int_0^{f(x)} w(x,y)\, dy \right) dx \right] = \mu \sum_{i=0}^{\mu} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \int_{f(x)}^{f(\overline{x}_i^\mu)} w(x,y)\, dy\, dx \ . \tag{31}$$

At the first order, we have that

$$\int_{f(x)}^{f(\overline{x}_i^\mu)} w(x,y) dy = w(\overline{x}_i^\mu, f(\overline{x}_i^\mu))(f(\overline{x}_i^\mu) - f(x)) + O((x - \overline{x}_i^\mu)) \ . \tag{32}$$

Since f is differentiable, we can use a Taylor approximation of f in each interval $[\overline{x}_i^\mu, \overline{x}_{i+1}^\mu]$ and write $f(x) = f(\overline{x}_i^\mu) + f'(\overline{x}_i^\mu)(x - \overline{x}_i^\mu) + O((x - \overline{x}_i^\mu)^2)$, which thus implies that $f(\overline{x}_i^\mu) - f(x) = -f'(\overline{x}_i^\mu)(x - \overline{x}_i^\mu) + O((x - \overline{x}_i^\mu)^2)$ and thus the left-hand side of Eq. 32 becomes $-w(\overline{x}_i^\mu, f(\overline{x}_i^\mu)) f'(\overline{x}_i^\mu)(x - \overline{x}_i^\mu) + O((x - \overline{x}_i^\mu)^2)$. By integrating the previous equation between $\overline{x}_i^\mu$ and $\overline{x}_{i+1}^\mu$ we obtain

$$\int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \int_{f(x)}^{f(\overline{x}_i^\mu)} w(x,y)\, dy\, dx = -\frac{1}{2} w(\overline{x}_i^\mu, f(\overline{x}_i^\mu)) f'(\overline{x}_i^\mu)(\overline{x}_{i+1}^\mu - \overline{x}_i^\mu)^2 + O((\overline{x}_{i+1}^\mu - \overline{x}_i^\mu)^3) \ .$$

Summing up for $i = 0$ to $i = \mu$, multiplying by μ and using Eq. 31, we obtain Eq. 30, which concludes Step 1.

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

$$\frac{1}{2} \sum_{i=0}^{\mu-1} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \frac{f'(x) w(x, f(x))}{\delta(x)} dx + \frac{1}{2} \int_{\overline{x}_\mu^\mu}^{x_{\max}} \frac{f'(x) w(x, f(x))}{\delta(x)} dx \ .$$

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

$$\frac{1}{2} \int_0^{x_{\max}} \frac{f'(x) w(x, f(x))}{\delta(x)} dx = \frac{1}{2} \sum_{i=0}^{\mu} \int_{\overline{x}_i^\mu}^{\overline{x}_{i+1}^\mu} \frac{f'(x) w(x, f(x))}{\delta(x)} dx \tag{33}$$

For μ to ∞, the assumption $\mu \sup((\sup_{0 \le i \le \mu-1} |\overline{x}_{i+1}^\mu - \overline{x}_i^\mu|), |x_{\max} - \overline{x}_\mu^\mu|) \to c$ implies that the distance between two consecutive points $|\overline{x}_{i+1}^\mu - \overline{x}_i^\mu|$ as well as $|\overline{x}_\mu^\mu - x_{\max}|$ 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 $\overline{x}_{\varphi(\mu)}^\mu$ and $\overline{x}_{\varphi(\mu)+1}^\mu$ surround x, i.e., $\overline{x}_{\varphi(\mu)}^\mu \le x < \overline{x}_{\varphi(\mu)+1}^\mu$. Since we assume that δ is continuous, a first order approximation of $\delta(x)$ is $\delta(\overline{x}_\mu^\mu)$, i.e., $\delta(x) = \delta(\overline{x}_{\varphi(\mu)}^\mu) + O(\overline{x}_{\varphi(\mu)+1}^\mu - \overline{x}_{\varphi(\mu)}^\mu)$ and therefore by integrating between $\overline{x}_{\varphi(\mu)}^\mu$ and $\overline{x}_{\varphi(\mu)+1}^\mu$ we obtain

$$\int_{\overline{x}_{\varphi(\mu)}^\mu}^{\overline{x}_{\varphi(\mu)+1}^\mu} \delta(x) dx = \delta(\overline{x}_{\varphi(\mu)}^\mu)(\overline{x}_{\varphi(\mu)+1}^\mu - \overline{x}_{\varphi(\mu)}^\mu) + O(\overline{x}_{\varphi(\mu)+1}^\mu - \overline{x}_{\varphi(\mu)}^\mu)^2) \ . \tag{34}$$

Moreover by definition of the density δ, $\int_{\overline{x}_{\varphi(\mu)}^\mu}^{\overline{x}_{\varphi(\mu)+1}^\mu} \delta(x) dx$ approximates the number of points contained in the interval $[\overline{x}_{\varphi(\mu)}^\mu, \overline{x}_{\varphi(\mu)+1}^\mu[$ (i.e., one) normalized by μ:

$$\mu \int_{\overline{x}_{\varphi(\mu)}^\mu}^{\overline{x}_{\varphi(\mu)+1}^\mu} \delta(x) dx = 1 + O((\overline{x}_{\varphi(\mu)+1}^\mu - \overline{x}_{\varphi(\mu)}^\mu)) \ . \tag{35}$$

Using Eq. 34 and Eq. 35, we thus have $1/\delta(\overline{x}^{\mu}_{\varphi(\mu)}) = \mu(\overline{x}^{\mu}_{\varphi(\mu)+1} - \overline{x}^{\mu}_{\varphi(\mu)}) + O(\mu(\overline{x}^{\mu}_{\varphi(\mu)+1} - \overline{x}^{\mu}_{\varphi(\mu)})^2)$. Therefore for every i we have that

$$\frac{1}{\delta(\overline{x}^{\mu}_{i})} = \mu(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i}) + O(\mu(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i})^2) \ . \tag{36}$$

Since $x \to f'(x)w(x, f(x))/\delta(x)$ is continuous, we also obtain

$$\int_{\overline{x}^{\mu}_{i}}^{\overline{x}^{\mu}_{i+1}} \frac{f'(x)w(x, f(x))}{\delta(x)} dx = \frac{f'(\overline{x}^{\mu}_{i})w(\overline{x}^{\mu}_{i}, f(\overline{x}^{\mu}_{i}))}{\delta(\overline{x}^{\mu}_{i})}(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i}) + O((\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i})^2) \ .$$

Injecting Eq. 36 in the previous equation, we obtain

$$\int_{\overline{x}^{\mu}_{i}}^{\overline{x}^{\mu}_{i+1}} \frac{f'(x)w(x, f(x))}{\delta(x)} dx = \mu f'(\overline{x}^{\mu}_{i})w(\overline{x}^{\mu}_{i}, f(\overline{x}^{\mu}_{i}))(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i})^2 + O(\mu(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i})^3) \ .$$

Multiplying by $1/2$ and summing up for i from 0 to μ and using Eq. 30 and Eq. 33, we obtain

$$\frac{1}{2}\int_{0}^{x_{\max}} \frac{f'(x)w(x, f(x))}{\delta(x)} = -E_{\mu} + \sum_{i=0}^{\mu} O(\mu(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i})^3) \ . \tag{37}$$

Let us define Δ_{μ} as $\sup((\sup_{0\leq i\leq \mu-1} |\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i}|), |x_{\max} - \overline{x}^{\mu}_{\mu}|)$. By assumption, we know that $\mu\Delta_{\mu}$ converges to a positive constant c. The last term of Eq. 37 satisfies

$$\left| \sum_{i=0}^{\mu} O(\mu(\overline{x}^{\mu}_{i+1} - \overline{x}^{\mu}_{i})^3) \right| \leq K\mu^2(\Delta_{\mu})^3$$

where $K > 0$. Since $\mu\Delta_{\mu}$ converges to c, $(\mu\Delta_{\mu})^2$ converges to c^2. With Δ_{μ} converging to 0, we therefore have that $\mu^2\Delta^3_{\mu}$ converges to 0. Taking the limit in Eq. 37, we therefore obtain

$$-\frac{1}{2}\int_{0}^{x_{\max}} \frac{f'(x)w(x, f(x))}{\delta(x)} dx = \lim_{\mu\to\infty} E_{\mu} \ .$$

□

7.3. Proof of Theorem 3 stated on page 12

Before to state and prove Theorem 3, we need to establish a technical lemma.

Lemma 4. *Let us assume that f is continuous on $[x_{\min}, x_{\max}]$ and differentiable on $]x_{\min}, x_{\max}[$. Let $x_2 \in]x_{\min}, r_1]$ and let us define the function $\Theta : [0, x_{\max} - x_{\min}] \to \mathbb{R}$ as*

$$\Theta(\varepsilon) = \int_{x_{\min}+\varepsilon}^{x_2} \left(\int_{f(x_{\min}+\varepsilon)}^{f(x_{\min})} w(x, y)dy \right) dx$$

and $\Gamma : [0, x_2 - x_{\min}] \to \mathbb{R}$ as

$$\Gamma(\varepsilon) = \int_{x_{\min}}^{x_{\min}+\varepsilon} \left(\int_{f(x_{\min})}^{r_2} w(x, y)dy \right) dx \ .$$

If w is continuous, positive and $\lim_{x\to x_{\min}} f'(x) = -\infty$ then for any $r_2 > f(x_{\min})$

$$\lim_{\varepsilon\to 0} \frac{\Theta(\varepsilon)}{\Gamma(\varepsilon)} = +\infty \ .$$

Proof. The limits of Θ and Γ for ε converging to 0 equal 0. We will therefore apply the l'Hôpital rule to compute $\lim_{\varepsilon\to 0}\frac{\Theta(\varepsilon)}{\Gamma(\varepsilon)}$. First of all, note that since f is differentiable on $]x_{\min}, x_{\max}[$, Θ and Γ are differentiable on $]0, x_{\max}-x_{\min}]$. Moreover, we see that $\Theta(\varepsilon) = g(x_{\min}+\varepsilon, x_2)$ where g is defined in Eq. 22 except for the change from $f(\overline{x}_0^\mu)$ to $f(x_{\min})$. The proof of Lemma 3, however, does not change if we exchange the constant $f(\overline{x}_0^\mu)$ to the constant $f(x_{\min})$ and we deduce that

$$\Theta'(\varepsilon) = -f'(x_{\min}+\varepsilon)\int_{x_{\min}+\varepsilon}^{x_2} w(x, f(x_{\min}+\varepsilon))dx - \int_{f(x_{\min}+\varepsilon)}^{f(x_{\min})} w(x_{\min}+\varepsilon, y)dy \ .$$

From the fundamental theorem of calculus, we also have that

$$\Gamma'(\varepsilon) = \int_{f(x_{\min})}^{r_2} w(x_{\min}+\varepsilon, y)dy \ .$$

From the l'Hôpital rule, we deduce that

$$\lim_{\varepsilon\to 0}\frac{\Theta(\varepsilon)}{\Gamma(\varepsilon)} = \lim_{\varepsilon\to 0}\frac{\Theta'(\varepsilon)}{\Gamma'(\varepsilon)} \ . \tag{38}$$

By continuity of w, we deduce that

$$\lim_{\varepsilon\to 0}\Gamma'(\varepsilon) = \lim_{\varepsilon\to 0}\int_{f(x_{\min})}^{r_2} w(x_{\min}+\varepsilon, y)dy = \int_{f(x_{\min})}^{r_2} w(x_{\min}, y)dy$$

and by continuity of f and w, we deduce that

$$\lim_{\varepsilon\to 0}\int_{x_{\min}+\varepsilon}^{x_2} w(x, f(x_{\min}+\varepsilon))dx = \int_{x_{\min}}^{x_2} w(x, f(x_{\min}))dx \qquad \text{and} \qquad \lim_{\varepsilon\to 0}\int_{f(x_{\min}+\varepsilon)}^{f(x_{\min})} w(x_{\min}+\varepsilon, y)dy = 0 \ .$$

Therefore $\lim_{\varepsilon\to 0}\Theta'(\varepsilon) = \lim_{\varepsilon\to 0} -f'(x_{\min}+\varepsilon)\cdot\int_{x_{\min}}^{x_2} w(x, f(x_{\min}))dx = +\infty$ because x_2 is fixed, i.e., independent of ε, and therefore, the integral is constant. By Eq. 38 we obtain the result. □

Now, we are ready to prove Theorem 3.

Proof. We first prove the result for the left extreme. We denote $\overline{x}_1^\mu$ and $\overline{x}_2^\mu$ the two leftmost points of an optimal μ-distribution for $I_{H,w}$ if $\mu \geq 2$. In case of $\mu = 1$, let $\overline{x}_1^\mu$ be the optimal position of the (single) point. In this case, the contribution of $\overline{x}_1^\mu$ in the first dimension extends to the reference point, which we represent by setting $\overline{x}_2^\mu = r_1$ such that from now on, we can assume $\mu \geq 2$. We assume that $\lim_{x\to x_{\min}} f'(x) = -\infty$ and that $\overline{x}_1^\mu = x_{\min}$ in order to get a contradiction. Let $I_{H,w}(x_{\min})$ be the hypervolume solely dominated by the point $x_{\min}$. If we shift $\overline{x}_1^\mu$ to the right by $\varepsilon > 0$ (see Figure 11), then the new hypervolume contribution $I_{H,w}(x_{\min}+\varepsilon)$ satisfies

$$I_{H,w}(x_{\min}+\varepsilon) = I_{H,w}(x_{\min}) + \int_{x_{\min}+\varepsilon}^{\overline{x}_2^\mu}\int_{f(x_{\min}+\varepsilon)}^{f(x_{\min})} w(x,y)dydx - \int_{x_{\min}}^{x_{\min}+\varepsilon}\int_{f(x_{\min})}^{r_2} w(x,y)dydx \ .$$

Identifying x_2 with $\overline{x}_2^\mu$ in the definition of Θ in Lemma 4, the previous equation can be rewritten as

$$I_{H,w}(x_{\min}+\varepsilon) = I_{H,w}(x_{\min}) + \Theta(\varepsilon) - \Gamma(\varepsilon) \ .$$

From Lemma 4, for any $r_2 > f(x_{\min})$, there exists an $\varepsilon > 0$ such that $\frac{\Theta(\varepsilon)}{\Gamma(\varepsilon)} > 1$ and thus $\Theta(\varepsilon) - \Gamma(\varepsilon) > 0$. Thus, for any $r_2 > f(x_{\min})$, there exists an ε such that $I_{H,w}(x_{\min}+\varepsilon) > I_{H,w}(x_{\min})$ and thus $I_{H,w}(x_{\min})$ is not maximal which contradicts the fact that $\overline{x}_1^\mu = x_{\min}$. In a similar way, we can prove the result for the right extreme. □

7.4. Proof of Theorem 4 stated on page 13

The proof of the theorem requires to establish a technical proposition. We have assumed that the reference point is dominated by the Pareto front, i.e., at least $r_1 > x_{\max}$ and $r_2 > f(x_{\min})$. Let us consider a set of points on the front and the hypervolume contribution of the leftmost point $P_1 = (x_1, f(x_1))$ (see Figure 12). This hypervolume contribution is a function of x_1 itself, x_2, the x-coordinate of the second leftmost point, and r_2, the second coordinate of the reference

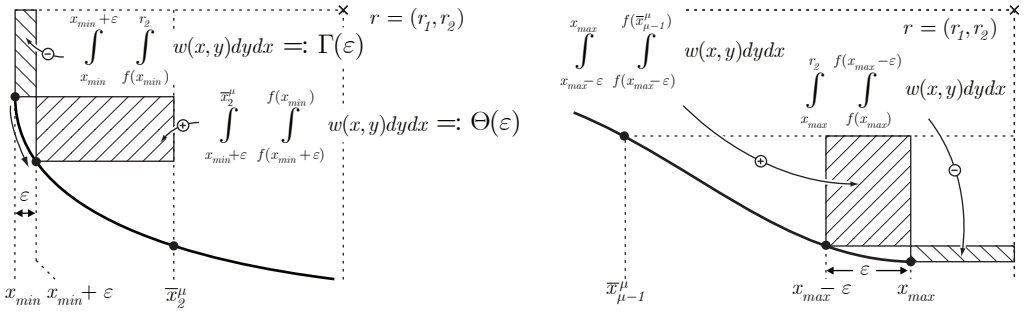
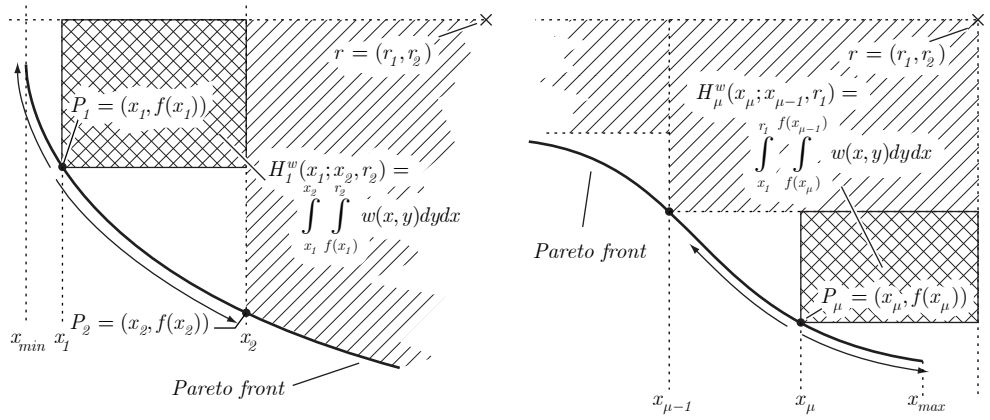


Figure 11: If the function $f(x)$ describing the Pareto front has an infinite derivative at its left extreme, the leftmost Pareto-optimal point at $x_{\min}$ will never coincide with the leftmost point $\overline{x}_1^\mu$ of an optimal μ-distribution for $I_{H,w}$ (left); similarly, if the derivative is zero at the right extreme, the rightmost Pareto-optimal point at $x_{\max}$ will never coincide with the rightmost point $\overline{x}_\mu^\mu$ (right). The reason is in both cases that for any finite r_1, and r_2 respectively, there exists an $\varepsilon > 0$, such that the dominated space gained ($\oplus$) when moving $\overline{x}_1^\mu$ from $x_{\min}$ to $x_{\min} + \varepsilon$, and $\overline{x}_\mu^\mu$ from $x_{\max}$ to $x_{\max} - \varepsilon$ respectively, is larger than the space no longer dominated ($\ominus$).



(a) left extreme (b) right extreme

Figure 12: Shows the notation and formula to compute the hypervolume contributions of the leftmost and rightmost point P_1 and P_μ respectively.

point. 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 $H_1^w(x_1; x_2, r_2)$ and reads

$$H_1^w(x_1; x_2, r_2) = \int_{x_1}^{x_2} \int_{f(x_1)}^{r_2} w(x,y)dydx \ . \tag{39}$$

The following proposition establishes a key property of the function H_1^w.

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

Proof. Assume that $H_1^w(x_1; x_{\max}, r_2)$ is maximal for $x_1 = x_{\min}$, i.e., $H_1^w(x_{\min}; x_{\max}, r_2) \geq H_1^w(x_1; x_{\max}, r_2)$, for all $x_1 \in]x_{\min}, x_{\max}]$. Let $\{D_1, \ldots, D_5\}$ denote the weighted hypervolume indicator values of different non-overlapping rectangular areas shown in Fig. 13. Then for all x_1 in $]x_{\min}, x_{\max}]$, $H_1^w(x_{\min}; x_{\max}, r_2) \geq H_1^w(x_1; x_{\max}, r_2)$ 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 $H_1^w(x_{\min}; x_2, r_2) \geq H_1^w(x_1; x_2, r_2)$. Hence, $H_1^w(x_1; x_2, r_2)$ is also maximal for $x_1 = x_{\min}$ for any choice $x_2 \in]x_1, x_{\max}]$. □

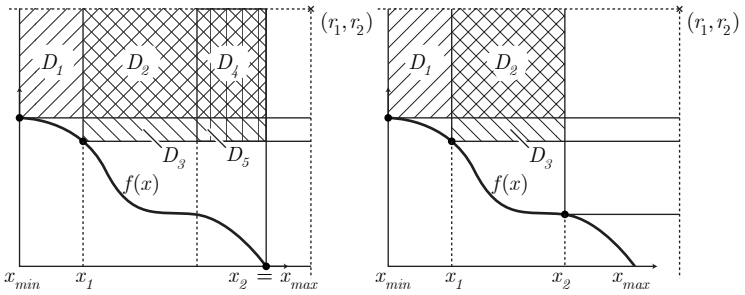


Figure 13: If the hypervolume indicator is larger for the choice of $x_1 = x_{\min}$ than when choosing $x_1 > x_{\min}$ if $x_2 = x_{\max}$ (left-hand side), it is also larger for $x_1 = x_{\min}$ for any $x_2 > x_1$ (right-hand side).

We are now ready to prove Theorem 4.

Proof of Theorem 4. Let x_1 and x_2 denote the x-coordinates of the two leftmost points $P_1 = (x_1, f(x_1))$ and $P_2 = (x_2, f(x_2))$. Then the hypervolume contribution of P_1 is given by Eq. 39. To prove that P_1 is the extremal point $(x_{\min}, f(x_{\min}))$, we need to prove that $x_1 \in [x_{\min}, x_2] \mapsto H_1^w(x_1; x_2, r_2)$ is maximal for $x_1 = x_{\min}$. By using Proposition 3, we know that if we prove that $x_1 \to H_1^w(x_1; x_{\max}, r_2)$ is maximal for $x_1 = x_{\min}$ then we will also have that $H_1^w : x_1 \in [x_{\min}, x_2] \mapsto H_1^w(x_1; x_2, r_2)$ is maximal for $x_1 = x_{\min}$. Therefore we will now prove that $x_1 \to H_1^w(x_1; x_{\max}, r_2)$ is maximal for $x_1 = x_{\min}$. To do so, we will show that $\frac{dH_1^w(x_1; x_{\max}, r_2)}{dx_1} \neq 0$ for all $x_{\min} < x_1 \leq x_{\max}$. According to Lemma 3, the derivative of the hypervolume contribution of P_1 is

$$\frac{dH_1^w(x_1; x_{\max}, r_2)}{dx_1} = -f'(x_1) \int_{x_1}^{x_{\max}} w(x, f(x_1))dx - \int_{f(x_1)}^{r_2} w(x_1, y)dy$$

Hence, by choosing $r_2 > \mathcal{K}_2$ according to Theorem 4, $\frac{dH_1^w(x_1; x_{\max}, r_2)}{dx_1} \neq 0$. □

7.5. Proof of Corollary 2 stated on page 14

Proof. We replace $w(x, y)$ in Eq. 13 of Theorem 4 by 1 and obtain that if there exists a $\mathcal{K}_2 \in \mathbb{R}$ such that

$$\forall x_1 \in]x_{\min}, x_{\max}] : \mathcal{K}_2 - f(x_1) > -f'(x_1)(x_{\max} - x_1), \tag{40}$$

then for any $r_2 \geq \mathcal{K}_2$, the leftmost extreme is included. The previous equation writes $\mathcal{K}_2 > f(x_1) - f'(x_1)(x_{\max} - x_1)$ for all $x_1 \in]x_{\min}, x_{\max}]$. However $-f'(x_1)(x_{\max} - x_1) = f'(x_1)(x_1 - x_{\max})$. Therefore Eq. 40 writes as

$$\forall x_1 \in]x_{\min}, x_{\max}] : \mathcal{K}_2 > f(x_1) + f'(x_1)(x_1 - x_{\max}) \ . \tag{41}$$

Since $\mathcal{K}_2$ has to be larger than the right-hand side of Eq. 41 for all x_1 in $]x_{\min}, x_{\max}]$, it has to be larger than the supremum of $f(x_1) + f'(x_1)(x_1 - x_{\max})$ for x_1 in $]x_{\min}, x_{\max}]$ and thus

$$\mathcal{K}_2 > \sup\{f(x_1) + f'(x_1)(x_1 - x_{\max}) : x \in]x_{\min}, x_{\max}]\} \ . \tag{42}$$

Defining $\mathcal{R}_2$ as the infimum over $\mathcal{K}_2$ satisfying Eq. 42 results in Eq. 15 which concludes the proof. □

7.6. Proof of Theorem 5 stated on page 14

Before to present the proof, we consider the hypervolume contribution of the rightmost point:

$$H_\mu^w(x_\mu; x_{\mu-1}, r_1) = \int_{x_\mu}^{r_1} \int_{f(x_\mu)}^{f(x_{\mu-1})} w(x, y)dydx \tag{43}$$

Similar to Proposition 3 we can establish the following proposition:

Proposition 4. *If $x_\mu \to H_1^w(x_\mu; x_{\min}, r_1)$ is maximal for $x_\mu = x_{\max}$, then for any $x_\mu \in [x_{\min}, x_{\mu-1}[$ the contribution $H_\mu^w(x_\mu; x_{\mu-1}, r_1)$ is maximal for $x_\mu = x_{\max}$ too.*

We are now ready to prove Theorem 5.

Proof of Theorem 5. Let x_μ and $x_{\mu-1}$ denote the x-coordinates of the two rightmost points $P_\mu = (x_\mu, f(x_\mu))$ and $P_{\mu-1} = (x_{\mu-1}, f(x_{\mu-1}))$. Then the hypervolume contribution of P_μ is given by Eq. 43. To prove that P_μ is the extremal point $(x_{\max}, f(x_{\max}))$, we need to prove that $x_\mu \in [x_{\mu-1}, x_{\max}] \mapsto H^w_\mu(x_\mu; x_{\mu-1}, r_1)$ is maximal for $x_\mu = x_{\max}$. By using Proposition 4, we know that if we prove that $x_\mu \to H^w_\mu(x_\mu; x_{\min}, r_1)$ is maximal for $x_\mu = x_{\max}$ then we will also have that $H^w_\mu : x_\mu \in [x_{\mu-1}, x_{\max}] \mapsto H^w_\mu(x_\mu; x_{\mu-1}, r_1)$ is maximal for $x_\mu = x_{\max}$. Therefore, we will now prove that $x_\mu \to H^w_\mu(x_\mu; x_{\min}, r_1)$ is maximal for $x_\mu = x_{\max}$. To do so, we will show that $\frac{dH^w_\mu(x_\mu; x_{\min}, r_1)}{dx_\mu} \neq 0$ for all $x_{\min} \leq x_\mu < x_{\max}$. According to Lemma 3, the derivative of the hypervolume contribution of P_μ is

$$\frac{dH^w_\mu(x_\mu; x_{\min}, r_1)}{dx_\mu} = -f'(x_\mu) \int_{x_\mu}^{r_1} w(x, f(x_\mu))dx - \int_{f(x_\mu)}^{f(x_{\min})} w(x_\mu, y)dy \ .$$

Hence, by choosing $r_1 > \mathcal{K}_1$ according to Theorem 5, $\frac{dH^w_\mu(x_\mu; x_{\min}, r_1)}{dx_\mu} \neq 0$. □

7.7. Proof of Corollary 3 stated on page 15

Proof. We replace $w(x, y)$ in Eq. 16 of Theorem 5 by 1 and obtain that if there exists a $\mathcal{K}_1 \in \mathbb{R}$ such that $-f'(x_\mu)(\mathcal{K}_1 - x_\mu) > (f(x_{\min}) - f(x_\mu))$ holds for all $x_\mu \in [x_{\min}, x_{\max}[$, then for every $r_1 \geq \mathcal{K}_1$, the rightmost extreme is included in optimal μ-distributions for I_H. The previous inequality writes

$$\forall x_\mu \in [x_{\min}, x_{\max}[: \mathcal{K}_1 > (f(x_\mu) - f(x_{\min}))/f'(x_\mu) + x_\mu \ . \tag{44}$$

Since $\mathcal{K}_1$ has to be larger than the right-hand side of Eq. 44 for all x_μ in $[x_{\min}, x_{\max}[$, it has to be larger than the supremum of the right-hand side of Eq. 44 for x_μ in $[x_{\min}, x_{\max}[$ and thus

$$\mathcal{K}_1 > \sup \left\{ x + \frac{f(x) - f(x_{\min})}{f'(x)} : x \in [x_{\min}, x_{\max}[\right\} \ . \tag{45}$$

Defining $\mathcal{R}_1$ as the infimum over $\mathcal{K}_1$ satisfying Eq. 45 results in Eq. 18 which concludes the proof. □

7.8. Proof of Lemma 2 stated on page 15

Proof. Let us denote the leftmost and the rightmost point of an optimal μ-distribution for $I_{H,w}$ as $\overline{x}^\mu_1(R)$ and $\overline{x}^\mu_\mu(R)$ respectively when the hypervolume indicator is computed with respect to a reference point R. By assumption, $\overline{x}^\mu_1(R^1) = x_{\min}$ and $\overline{x}^\mu_\mu(R^1) = x_{\max}$. Assume, in order to get a contradiction, that $\overline{x}^\mu_1(R^2) > x_{\min}$ (i.e., the leftmost point of the optimal μ-distribution for $I_{H,w}$ and R^2 is not the left extreme) and assume that $\overline{x}^\mu_\mu(R^2) = x_{\max}$ for the moment. Let us denote $\overline{I^\mu_{H,w}}(R^2)$ the hypervolume associated with an optimal μ-distribution for $I_{H,w}$ computed with respect to the reference point R^2 (and $\overline{I^\mu_{H,w}}(R^1)$ accordingly for R^1). We decompose $\overline{I^\mu_{H,w}}(R^2)$ in the following manner (see Figure 14)

$$\overline{I^\mu_{H,w}}(R^2) = A_1 + A_2 + A_3 \tag{46}$$

where A_1 is the hypervolume (computed with respect to w) enclosed in between the optimal μ-distribution associated with R^2 and the reference point R^1, A_2 is the hypervolume (computed with respect to w) enclosed in the rectangle whose diagonal extremities are R^2 and $(\overline{x}^\mu_1(R^2), r^1_2)$ and A_3 is the hypervolume (again with respect to w) enclosed in the rectangle with diagonal $[(r^1_1, f(x_{\max})), (r^2_1, r^1_2)]$. Consider now an optimal μ-distribution for $I_{H,w}$ associated with the reference point R^1 and denote this optimal μ-distribution $(\overline{x}^\mu_1(R^1), \ldots, \overline{x}^\mu_\mu(R^1))$. The weighted hypervolume enclosed by this set of points and R^2 equals $\overline{I^\mu_{H,w}}(R^1) + A_2 + A'_2 + A_3$ where A'_2 is the hypervolume (computed with respect to w) enclosed in the rectangle whose diagonal is $[(x_{\min}, r^1_2), (\overline{x}^\mu_1(R^2), r^2_2)]$ (Fig. 14). By definition of $\overline{I^\mu_{H,w}}(R^2)$ we have that

$$\overline{I^\mu_{H,w}}(R^2) \geq \overline{I^\mu_{H,w}}(R^1) + A_2 + A'_2 + A_3 \ . \tag{47}$$

However, since $\overline{I^{\mu}_{H,w}}(R^1)$ is the maximal hypervolume value possible for the reference point R^1 and a set of μ points, we have that $A_1 \leq \overline{I^{\mu}_{H,w}}(R^1)$ and thus with Eq. 47 that $\overline{I^{\mu}_{H,w}}(R^2) \geq A_1 + A_2 + A'_2 + A_3$. From Eq. 46, we deduce that

$$\overline{I^{\mu}_{H,w}}(R^2) \geq \overline{I^{\mu}_{H,w}}(R^2) + A'_2 \ . \tag{48}$$

Since we have assumed that $\overline{x}^{\mu}_1(R^2) > x_{\min}$ and that $r^2_2 > r^1_2$, we have $A'_2 > 0$. And thus, Eq. 48 implies that $\overline{I^{\mu}_{H,w}}(R^2) > \overline{I^{\mu}_{H,w}}(R^2)$, which contradicts our assumption. In a similar way, we show a contradiction if we assume that both $\overline{x}^{\mu}_1(R^2) > x_{\min}$ and $\overline{x}^{\mu}_{\mu}(R^2) < x_{\max}$, i.e., if both extremes are not contained in an optimal μ-distribution for $I_{H,w}$ and the reference point R^2. Also the proof for the right extreme is similar. □

7.9. Proof of Theorem 6 stated on page 15

Proof. Let us fix $\varepsilon_2 \in \mathbb{R}_{>0}$ and let $R = (R_1, R_2) = (r_1, \mathcal{R}^{\text{Nadir}}_2 + \varepsilon_2)$ for r_1 arbitrarily chosen with $r_1 \geq \mathcal{R}^{\text{Nadir}}_1$. The optimal μ-distributions for $I_{H,w}$ and the reference point R obviously depend on μ. Let $\overline{x}^{\mu}_2(R)$ denote the second point of an optimal μ-distribution for $I_{H,w}$ when R is chosen as reference point. We know that for μ to infinity, $\overline{x}^{\mu}_2(R)$ converges to $x_{\min}$. Also, because f' is continuous on $[x_{\min}, x_{\max}]$, the extreme value theorem implies that there exists $\theta > 0$ such that $|f'(x)| \leq \theta$ for all $x \in [x_{\min}, x_{\max}]$. Since f' is negative we therefore have

$$\forall x \in [x_{\min}, x_{\max}] : -f'(x) \leq \theta \ . \tag{49}$$

In order to prove that the leftmost point of an optimal μ-distribution is $x_{\min}$, it is enough to show that the first partial derivative of $I_{H,w}$ is non-zero on $]x_{\min}, \overline{x}^{\mu}_2(R)]$. According to Eq. 3 and Lemma 3, the first partial derivative of $I_{H,w}((\overline{x}^{\mu}_1, \ldots, \overline{x}^{\mu}_{\mu}))$ equals (we omit the dependence in R for the following equations)

$$\begin{aligned}
\partial_1 I_{H,w} &= -f'(\overline{x}^{\mu}_1) \int_{\overline{x}^{\mu}_1}^{\overline{x}^{\mu}_2} w(x, f(\overline{x}^{\mu}_1))dx - \int_{f(\overline{x}^{\mu}_1)}^{R_2} w(\overline{x}^{\mu}_1, y)dy \\
&= \left(-f'(\overline{x}^{\mu}_1)\right) \int_{x_{\min}}^{\overline{x}^{\mu}_2} w(x, f(\overline{x}^{\mu}_1))dx - \left(-f'(\overline{x}^{\mu}_1)\right) \int_{x_{\min}}^{\overline{x}^{\mu}_1} w(x, f(\overline{x}^{\mu}_1))dx - \int_{f(\overline{x}^{\mu}_1)}^{\mathcal{R}^{\text{Nadir}}_2} w(\overline{x}^{\mu}_1, y)dy - \int_{\mathcal{R}^{\text{Nadir}}_2}^{\mathcal{R}^{\text{Nadir}}_2 + \varepsilon_2} w(\overline{x}^{\mu}_1, y)dy \ .
\end{aligned} \tag{50}$$

Since the second and third summand are non-positive due to w being strictly positive we have

$$\leq \left(-f'(\overline{x}^{\mu}_1)\right) \int_{x_{\min}}^{\overline{x}^{\mu}_2} w(x, f(\overline{x}^{\mu}_1))dx - \int_{\mathcal{R}^{\text{Nadir}}_2}^{\mathcal{R}^{\text{Nadir}}_2 + \varepsilon_2} w(\overline{x}^{\mu}_1, y)dy \tag{51}$$

and because $w \leq W$ and with Eq. 49, Eq. 51 can be upper bounded by

$$\leq \theta W(\overline{x}^{\mu}_2 - x_{\min}) - \int_{\mathcal{R}^{\text{Nadir}}_2}^{\mathcal{R}^{\text{Nadir}}_2 + \varepsilon_2} w(\overline{x}^{\mu}_1, y)dy \ . \tag{52}$$

Since $\overline{x}^{\mu}_2$ converges to $x_{\min}$ for μ to infinity, and $-\int_{\mathcal{R}^{\text{Nadir}}_2}^{\mathcal{R}^{\text{Nadir}}_2 + \varepsilon_2} w(\overline{x}^{\mu}_1, y)dy < 0$ we deduce that there exists μ_1 such that for all μ larger than μ_1, Eq. 52 is strictly negative and thus for all μ larger than μ_1, the first partial derivative of $I_{H,w}$ is non zero, i.e., $\overline{x}^{\mu}_1 = x_{\min}$. With Lemma 2 we deduce that all reference points dominated by R will also allow to obtain the left extreme.

We will now follow the same steps for the right extreme. Let us fix $\varepsilon_1 \in \mathbb{R}_{>0}$ and let $R = (\mathcal{R}^{\text{Nadir}}_1 + \varepsilon_1, r_2)$ for $r_2 \geq \mathcal{R}^{\text{Nadir}}_2$. Following the same steps for the right extreme, we need to prove that the μ-th partial derivative of $I_{H,w}$ is non zero for all $\overline{x}^{\mu}_{\mu} \in [\overline{x}^{\mu}_{\mu-1}, x_{\max}[$. According to Eq. 28,

$$\begin{aligned}
\partial_{\mu} I_{H,w}(\overline{x}^{\mu}_1, \ldots, \overline{x}^{\mu}_{\mu}) &= -\int_{f(\overline{x}^{\mu}_{\mu})}^{f(\overline{x}^{\mu}_{\mu-1})} w(\overline{x}^{\mu}_{\mu}, y)dy - f'(\overline{x}^{\mu}_{\mu}) \int_{\overline{x}^{\mu}_{\mu}}^{\mathcal{R}^{\text{Nadir}}_1 + \varepsilon_1} w(x, f(\overline{x}^{\mu}_{\mu}))dx \\
&\geq -W(f(\overline{x}^{\mu}_{\mu-1}) - f(\overline{x}^{\mu}_{\mu})) - f'(\overline{x}^{\mu}_{\mu}) \int_{\overline{x}^{\mu}_{\mu}}^{\mathcal{R}^{\text{Nadir}}_1 + \varepsilon_1} w(x, f(\overline{x}^{\mu}_{\mu}))dx
\end{aligned} \tag{53}$$

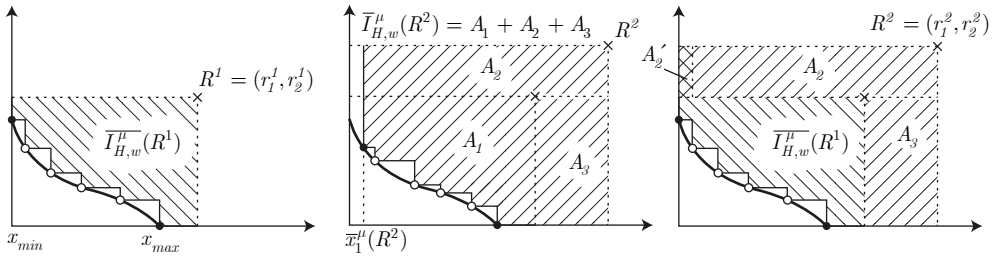


Figure 14: If the optimal distribution of μ points contains the extremes (left-hand side), then after increasing the reference point from R^1 to R^2 the extremes are still included in the optimal μ-distribution (right-hand side). This can be proven by contradiction (middle).

and since $\overline{x}^{\mu}_{\mu} \leq \mathcal{R}_1^{\text{Nadir}}$, we obtain

$$\geq - W(f(\overline{x}^{\mu}_{\mu-1}) - f(\overline{x}^{\mu}_{\mu})) - f'(\overline{x}^{\mu}_{\mu}) \int_{\mathcal{R}_1^{\text{Nadir}}}^{\mathcal{R}_1^{\text{Nadir}}+\varepsilon_1} w(x, f(\overline{x}^{\mu}_{\mu}))dx \tag{54}$$

By continuity of f and the fact that both $\overline{x}^{\mu}_{\mu}$ and $\overline{x}^{\mu}_{\mu-1}$ converge to $x_{\max}$ the term $W(f(\overline{x}^{\mu}_{\mu-1}) - f(\overline{x}^{\mu}_{\mu}))$ converges to zero. Since $-f'(\overline{x}^{\mu}_{\mu}) \int_{\mathcal{R}_1^{\text{Nadir}}}^{\mathcal{R}_1^{\text{Nadir}}+\varepsilon_1} w(x, f(\overline{x}^{\mu}_{\mu}))dx$ is strictly positive, we deduce that there exists μ_2 such that for all $\mu \geq \mu_2$, $\partial_{\mu} I_{H,w}(\overline{x}^{\mu}_1, \ldots, \overline{x}^{\mu}_{\mu})$ is strictly positive and thus for all μ larger than μ_2 the μ-th partial derivative of $I_{H,w}$ is non zero, i.e., $\overline{x}^{\mu}_{\mu} = x_{\max}$. With Lemma 2 we deduce that all reference points dominated by R allow to obtain the right extreme. □

7.10. Results for the ZDT Test Function Suite

There exist six ZDT test problems—ZDT1 to ZDT6—of which ZDT5 has a discrete Pareto front and is therefore excluded from our investigations (Zitzler et al., 2000). In the following, let $d = (d_1, \ldots, d_n) \in \mathbb{R}^n$ denote the decision vector of n real-valued variables. The shapes of the Pareto fronts as stated below follow from the definition of the objectives including a function $g(d)$ and the fact that the Pareto front is obtained by setting $g(d) = 1$.

ZDT1 From Example 5, we recapitulate the front shape of ZDT1 as $f(x) = 1 - \sqrt{x}$ with $x_{\min} = 0$ and $x_{\max} = 1$, see Figure 9(a). From $f'(x) = -1/(2\sqrt{x})$ the density on the front according to Eq. 12 is $\delta_F(x) = \frac{3\sqrt[4]{x}}{2\sqrt{4x+1}}$. Since $f'(x_{\min}) = -\infty$, the left extreme is never included as stated already in Example 5. The lower bound of the reference point $\mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2)$ to have the right extreme, according to Eq. 18, equals $\mathcal{R}_1 = \sup_{x \in]x_{\min}, x_{\max}]} x + \frac{1-\sqrt{x}-1}{-1/(2\sqrt{x})} = \sup_{x \in]0,1]} 3x = 3$.

ZDT2 From Example 4, we recapitulate the front shape of ZDT2 as $f(x) = 1 - x^2$ with $x_{\min} = 0$ and $x_{\max} = 1$ and the density of $\delta_F(x) = \frac{3\sqrt{x}}{2\sqrt{1+4x^2}}$ (see Fig. 9(b)). The lower bounds for the reference point $\mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2)$ to obtain the extremes are according to the equations Eq. 18 and Eq. 15 $\mathcal{R}_1 = \sup_{x \in]x_{\min}, x_{\max}]} x + \frac{1-x^2-1}{-2x} = \sup_{x \in]0,1]} \frac{3}{2}x = \frac{3}{2}$ and $\mathcal{R}_2 = \sup_{x \in [x_{\min}, x_{\max}[} -2x \cdot (x-1) + 1 - x^2 = \sup_{x \in [0,1[} 2x - 3x^2 + 1 = \frac{4}{3}$ respectively.

ZDT3 Due to the sine-function in the definition of ZDT3's second objective, the front is discontinuous where $f : D \to [-1, 1],\ x \mapsto 1 - \sqrt{x} - x \cdot \sin(10\pi x)$ where $D = [0, 0.0830] \cup (0.1823, 0.2578] \cup (0.4093, 0.4539] \cup (0.6184, 0.6525] \cup (0.8233, 0.8518]$ is derived numerically. Hence $x_{\min} = 0$ and $x_{\max} = 0.8518$. The density is

$$\delta_F(x) = C \cdot \sqrt{\tfrac{1}{2\sqrt{x}} + \sin(10\pi x) + 10\pi x \cos(10\pi x)} \Big/ \sqrt{1 + \left(\tfrac{1}{2\sqrt{x}} + \sin(10\pi x) + 10\pi x \cos(10\pi x)\right)^2} \text{ with } C \approx 1.5589$$

where $x \in D$ and $\delta_F(x) = 0$ otherwise. Figure 9(c) shows the Pareto front and the density. Since $f'(x_{\min}) = -\infty$ and $f'(x_{\max}) = 0$, the left and right extremes are never included.

ZDT4 The Pareto front of ZDT4 is again given by $f(x) = 1 - \sqrt{x}$. Hence, the density and the choice of the reference point is the same as for ZDT1.

ZDT6 The Pareto front of ZDT6 is $f : [x_{\min}, x_{\max}] \to [0, 1],\ x \mapsto 1 - x^2$ with $x_{\min} \approx 0.2808$ and $x_{\max} = 1$, see Fig. 9(d). Hence, the Pareto front coincides with the one of ZDT2 except for $x_{\min}$ which is shifted slightly to the right.

From this, it follows that also the density is the same except for a constant factor, i.e., $\delta_F(x)$ is larger than the density for ZDT2 by a factor of ≈ 1.25. For the lower bound $\mathcal{R}$ of the reference point, we obtain

$$\mathcal{R}_1 = \sup_{x\in]x_{\min},x_{\max}]} x + \frac{1-x^2-(1-{x_{\min}}^2)}{-2x} = \sup_{x\in]0.2808,1]} \frac{{x_{\min}}^2-3x^2}{-2x} = \frac{3-{x_{\min}}^2}{2} \approx 1.461 \quad \text{and}$$

$$\mathcal{R}_2 = \sup_{x\in[x_{\min},1[} -2x(x-x_{\max})+1-x = \sup_{x\in[x_{\min},1[} 2x-3x^2+1 = \frac{4}{3} \ .$$

Hence, the lower bound $\mathcal{R}_2$ is the same as for ZDT2, but $\mathcal{R}_1$ differs slightly from ZDT2.

7.11. Results for the DTLZ Test Function Suite

The DTLZ test suite offers seven test problems which can be scaled to any number of objectives (Deb et al., 2005b). For the bi-objective variants, DTLZ5 and DTLZ6 are degenerated, i.e., the Pareto fronts consist of only a single point and are not examined in the following. For the definitions of the problems, we refer to (Deb et al., 2005b) and only state the shapes of the Pareto fronts which can be obtained by setting $g(d) = 0$ similar to the ZDT problems.

DTLZ1 The Pareto front of DTLZ1 is described by $f(x) = 1/2 - x$ with $x_{\min} = 0$ and $x_{\max} = 1/2$, see Fig. 9(e). According to Eq. 12, we have $\delta_F(x) = \sqrt{2}$. A lower bound for the reference point is given by $\mathcal{R}_1 = \sup_{x\in]0,1/2]} 1-x = 1$ and $\mathcal{R}_2 = \mathcal{R}_1$ for symmetry reasons.

DTLZ2 From Example 1, we recapitulate the front shape of $f(x) = \sqrt{1-x^2}$ with $x_{\min} = 0$ and $x_{\max} = 1$, see Fig. 9(f). According to Eq. 12, the density on the front is $\delta_F(x) = \sqrt{\pi x}\sqrt[4]{1-x^2}\,\big/\Gamma(3/4)^2$ where Γ denotes the gamma-function, i.e., $\Gamma(3/4) \approx 1.225$. A lower bound for the reference point is given by

$$\mathcal{R}_1 = \sup_{x\in]x_{\min},x_{\max}]} x + \frac{\sqrt{1-x^2}-\sqrt{1-{x_{\min}}^2}}{-x/\sqrt{1-x^2}} = \sup_{x\in]0,1]} \frac{\sqrt{1-x^2}-1+2x^2}{x} = 1/2\left(\sqrt{3}-1\right)3^{3/4}\sqrt{2} \approx 1.18$$

and for symmetry reasons $\mathcal{R}_2 = \mathcal{R}_1$.

DTLZ3 The problem formulation of DTLZ3 is the same as for DTLZ2 except for the function $g(d)$. However, the Pareto front is formed by the same decision vectors as for DTLZ2 and the fronts of DTLZ2 and DTLZ3 are identical. Hence, also the density and the choice of the reference point are the same as for DTLZ2.

DTLZ4 In DTLZ4, the same functions as in DTLZ2 are used with an additional meta-variable mapping $m : [0,1] \to [0,1]$ of the decision variables, i.e., the decision variable $m(d_i) = d_i^\alpha$ is used instead of the original decision variable d_i in the formulation of the DTLZ2 function. This transformation does not affect the shape of the Pareto front and the results on optimal μ-distributions for the unweighted hypervolume indicator again coincide with the ones for DTLZ2.

DTLZ7 The Pareto front of DTLZ7 is discontinuous and described by the function $f : D \to [0,4]$, $x \mapsto 4 - x(1 + \sin(3\pi x))$ where $D = [0, 0.2514] \cup (0.6316, 0.8594] \cup (1.3596, 1.5148] \cup (2.0518, 2.1164]$ which is derived numerically, see Fig. 9(g). Hence, $x_{\min} = 0$ and $x_{\max} \approx 2.1164$. The derivative of $f(x)$ is $f'(x) = -1 - \sin(3\pi x) - 3\pi x\cos(3\pi x)$ and the density therefore is $\delta_F(x) = C \cdot \sqrt{1+\sin(3\,\pi\,x)+3\pi x\cos(3\,\pi\,x)}\,\Big/\sqrt{1+(1+\sin(3\,\pi\,x)+3\pi x\cos(3\,\pi\,x))^2}$ with $C \approx 0.6566$. For $\mathcal{R}$, we find $\mathcal{R}_1 \approx 2.481$ and $\mathcal{R}_2 \approx 13.3720$.

7.12. Results for the WFG Test Function Suite

The WFG test suite offers nine test problems which can be scaled to any number of objectives. In contrast to DTLZ and ZDT, the problem formulations are build using an arbitrary number of so-called *transformation functions*. We abstain from quoting these functions here and refer the interested reader to (Huband et al., 2006). The resulting Pareto front shape is determined by parameterized *shape functions* h_i mapping $[0,1]$ to the range $[0,1]$. All test functions WFG4 to WFG9 share the same shape functions and are therefore examined together in the following.

WFG1 For WFG1, the shape functions are convex and mixed respectively which leads to the Pareto front $f(x) = \frac{2\rho-\sin(2\rho)}{10\pi} - 1$ with $\rho = 10\arccos(1-x)$, $x_{\min} = 0$ and $x_{\max} = 1$, see Fig. 9(h). The density becomes

$$\delta_F(x) = C \cdot \sqrt{\frac{2\,(1-\cos(2\rho))\,\pi}{\sqrt{x\,(2-x)}\left(\pi^2 - 4\,\frac{(1-\cos(2\rho))^2}{x(x-2)}\right)}}$$

with $C \approx 1.1569$. Since $\lim_{x \to x_{\max}} f'(x_{\max}) = 0$ the rightmost extreme point is never included in an optimal μ-distribution for $I_{H,w}$. For the choice of $\mathcal{R}_2$ the analytical expression is very long and therefore omitted. A numerical approximation leads to $\mathcal{R}_2 \approx 0.9795$.

WFG2 For WFG2, the shape functions are convex and discontinuous respectively which leads to the discontinuous Pareto front $f : D \to [0, 1]$, $x \mapsto 1 - 2\,\frac{(\pi - 0.1\rho)\cos^2(\rho)}{\pi}$ where $\rho = \arccos(x - 1)$, and with a numerically derived domain $D = [0, 0.0021] \cup (0.0206, 0.0537] \cup (0.1514, 0.1956] \cup (0.3674, 0.4164] \cup (0.6452, 0.6948] \cup (0.9567, 1]$, $x_{\min} = 0$ and $x_{\max} = 1$, see Fig. 9(i). The density becomes

$$\delta_F(x) = C \cdot \frac{\sqrt{-f'(x)}}{\sqrt{1 + f'(x)^2}} \quad \text{with } C \approx 0.44607 \text{ and } f'(x) = -2\,\frac{\cos(\rho)\,(\cos(\rho) + 20\,\sin(\rho)\,\pi - 2\,\sin(\rho)\,\rho)}{\sqrt{x\,(2 - x)}\pi}$$

for all $x \in D$ and $\delta_F(x) = 0$ otherwise. Again, $f'(0) = -\infty$ such that the leftmost extreme point is never included in an optimal μ-distribution for $I_{H,w}$. For the rightmost extreme one finds $\mathcal{R}_1 \approx 2.571$.

WFG3 For WFG3, the shape functions are both linear—leading to the linear Pareto front $f(x) = 1 - x$ with $x_{\min} = 0$ and $x_{\max} = 1$. Hence, the density is $\delta_F(x) = 1/\sqrt{2}$, see Fig. 9(e) for a scaled version of this Pareto front. For the choice of the reference point the same arguments as for DTLZ1 hold, which leads to $\mathcal{R} = (2, 2)$.

WFG4 to WFG9 For the six remaining test problems WFG4 to WFG9, the shape functions h_1 and h_2 are both concave—resulting in a spherical Pareto front $f(x) = \sqrt{1 - x^2}$ with $x_{\min} = 0$ and $x_{\max} = 1$. Hence, the Pareto front coincides with the front of DTLZ2 and also the density and the choice of the reference point are the same.

References

Auger, A., Bader, J., Brockhoff, D., Zitzler, E., 2009a. Investigating and Exploiting the Bias of the Weighted Hypervolume to Articulate User Preferences. In: Raidl, G., et al. (Eds.), Genetic and Evolutionary Computation Conference (GECCO 2009). ACM, New York, NY, USA, pp. 563–570.

Auger, A., Bader, J., Brockhoff, D., Zitzler, E., 2009b. Theory of the Hypervolume Indicator: Optimal μ-Distributions and the Choice of the Reference Point. In: Foundations of Genetic Algorithms (FOGA 2009). ACM, New York, NY, USA, pp. 87–102.

Bader, J., Zitzler, E., 2011. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. Evolutionary Computation 19 (1), 45–76.

Beume, N., Fonseca, C. M., Lopez-Ibanez, M., Paquete, L., Vahrenhold, J., Dec. 2007a. On the Complexity of Computing the Hypervolume Indicator. Tech. Rep. CI-235/07, University of Dortmund.

Beume, N., Naujoks, B., Emmerich, M., 2007b. SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume. European Journal of Operational Research 181 (3), 1653–1669.

Beume, N., Naujoks, B., Preuss, M., Rudolph, G., Wagner, T., 2009. Effects of 1-Greedy $\mathcal{S}$-Metric-Selection on Innumerably Large Pareto Fronts. In: Ehrgott, M., et al. (Eds.), Conference on Evolutionary Multi-Criterion Optimization (EMO 2009). Vol. 5467 of LNCS. Springer, pp. 21–35.

Bourbaki, N., 1989. Elements of Mathematics: General Topology (Chapter 1–4), 2nd Edition. Springer.

Branke, J., Deb, K., Dierolf, H., Osswald, M., 2004. Finding Knees in Multi-objective Optimization. In: Yao, X., et al. (Eds.), Conference on Parallel Problem Solving from Nature (PPSN VIII). Vol. 3242 of LNCS. Springer, pp. 722–731.

Bringmann, K., Friedrich, T., 2010. The Maximum Hypervolume Set Yields Near-optimal Approximation. In: Branke, J., et al. (Eds.), Genetic and Evolutionary Computation Conference (GECCO 2010). ACM, pp. 511–518.

Brockhoff, D., 2010. Optimal μ-Distributions for the Hypervolume Indicator for Problems With Linear Bi-Objective Fronts: Exact and Exhaustive Results. In: Deb, K., et al. (Eds.), Simulated Evolution and Learning (SEAL 2010). Vol. 6457 of LNCS. Springer, pp. 24–34.

Coello Coello, C. A., Lamont, G. B., Van Veldhuizen, D. A., 2007. Evolutionary Algorithms for Solving Multi-Objective Problems. Springer, Berlin, Germany.

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

Deb, K., 2001. Multi-Objective Optimization Using Evolutionary Algorithms. Wiley, Chichester, UK.

Deb, K., Mohan, M., Mishra, S., Winter 2005a. Evaluating the ϵ-Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-Optimal Solutions. Evolutionary Computation 13 (4), 501–525.

Deb, K., Thiele, L., Laumanns, M., Zitzler, E., 2005b. Scalable Test Problems for Evolutionary Multi-Objective Optimization. In: Abraham, A., Jain, R., Goldberg, R. (Eds.), Evolutionary Multiobjective Optimization: Theoretical Advances and Applications. Springer, Ch. 6, pp. 105–145.

Emmerich, M., Beume, N., Naujoks, B., 2005. An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. In: Conference on Evolutionary Multi-Criterion Optimization (EMO 2005). Vol. 3410 of LNCS. Springer, pp. 62–76.

Emmerich, M., Deutz, A., Beume, N., 2007. Gradient-Based/Evolutionary Relay Hybrid for Computing Pareto Front Approximations Maximizing the S-Metric. In: Hybrid Metaheuristics. Vol. 4771 of LNCS. Springer, pp. 140–156.

Fleischer, M., 2003. The Measure of Pareto Optima. Applications to Multi-Objective Metaheuristics. In: Fonseca, C. M., et al. (Eds.), Conference on Evolutionary Multi-Criterion Optimization (EMO 2003). Vol. 2632 of LNCS. Springer, Faro, Portugal, pp. 519–533.

Friedrich, T., Horoba, C., Neumann, F., 2009. Multiplicative Approximations and the Hypervolume Indicator. In: Raidl, G., et al. (Eds.), Genetic and Evolutionary Computation Conference (GECCO 2009). ACM, pp. 571–578.

Hansen, N., Kern, S., 2004. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In: Yao, X., et al. (Eds.), Conference on Parallel Problem Solving from Nature (PPSN VIII). Vol. 3242 of LNCS. Springer, Berlin, Germany, pp. 282–291.

Huband, S., Hingston, P., Barone, L., While, L., 2006. A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit. IEEE Transactions on Evolutionary Computation 10 (5), 477–506.

Huband, S., Hingston, P., White, L., Barone, L., 2003. An Evolution Strategy with Probabilistic Mutation for Multi-Objective Optimisation. In: Congress on Evolutionary Computation (CEC 2003). Vol. 3. IEEE Press, Canberra, Australia, pp. 2284–2291.

Hughes, E. J., 2005. Evolutionary Many-Objective Optimisation: Many Once or One Many? In: Congress on Evolutionary Computation (CEC 2005). IEEE Press, pp. 222–227.

Igel, C., Hansen, N., Roth, S., 2007. Covariance Matrix Adaptation for Multi-objective Optimization. Evolutionary Computation 15 (1), 1–28.

Knapp, A. W., 2005. Basic Real Analysis, 1st Edition. Birkhäuser.

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

Knowles, J., Corne, D., 2002. On Metrics for Comparing Non-Dominated Sets. In: Congress on Evolutionary Computation (CEC 2002). IEEE Press, Piscataway, NJ, pp. 711–716.

Knowles, J., Corne, D., 2003. Properties of an Adaptive Archiving Algorithm for Storing Nondominated Vectors. IEEE Transactions on Evolutionary Computation 7 (2), 100–116.

Knowles, J. D., Corne, D. W., Fleischer, M., 2003. Bounded Archiving using the Lebesgue Measure. In: Congress on Evolutionary Computation (CEC 2003). IEEE Press, Canberra, Australia, pp. 2490–2497.

Lizarraga-Lizarraga, G., Hernandez-Aguirre, A., Botello-Rionda, S., 2008. G-Metric: an M-ary quality indicator for the evaluation of non-dominated sets. In: Genetic And Evolutionary Computation Conference (GECCO 2008). ACM, New York, NY, USA, pp. 665–672.

Purshouse, R. C., 2003. On the Evolutionary Optimisation of Many Objectives. Ph.D. thesis, The University of Sheffield.

Purshouse, R. C., Fleming, P. J., 2003. An Adaptive Divide-and-Conquer Methodology for Evolutionary Multi-criterion Optimisation. In: Fonseca, C., et al. (Eds.), Conference on Evolutionary Multi-Criterion Optimization (EMO 2003). No. 2632 in LNCS. Springer, pp. 133–147.

Zitzler, E., Brockhoff, D., Thiele, L., 2007. The Hypervolume Indicator Revisited: On the Design of Pareto-compliant Indicators Via Weighted Integration. In: Obayashi, S., et al. (Eds.), Conference on Evolutionary Multi-Criterion Optimization (EMO 2007). Vol. 4403 of LNCS. Springer, Berlin, pp. 862–876.

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

Zitzler, E., Künzli, S., 2004. Indicator-Based Selection in Multiobjective Search. In: Yao, X., et al. (Eds.), Conference on Parallel Problem Solving from Nature (PPSN VIII). Vol. 3242 of LNCS. Springer, pp. 832–842.

Zitzler, E., Thiele, L., 1998. Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In: Conference on Parallel Problem Solving from Nature (PPSN V). Vol. 1498 of LNCS. Amsterdam, pp. 292–301.

Zitzler, E., Thiele, L., Bader, J., 2010. On Set-Based Multiobjective Optimization. IEEE Transactions on Evolutionary Computation 14 (1), 58–79.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., Grunert da Fonseca, V., 2003. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. IEEE Transactions on Evolutionary Computation 7 (2), 117–132.