

Investigating and Exploiting the Bias of the Weighted Hypervolume to Articulate User Preferences

Anne Auger* Johannes Bader†

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

Dimo Brockhoff† Eckart Zitzler†

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

ABSTRACT

Optimizing the hypervolume indicator within evolutionary multiobjective optimizers has become popular in the last years. Recently, the indicator has been generalized to the weighted case to incorporate various user preferences into hypervolume-based search algorithms. There are two main open questions in this context: (i) how does the specified weight influence the distribution of a fixed number of points that maximize the weighted hypervolume indicator? (ii) how can the user articulate her preferences easily without specifying a certain weight distribution function?

In this paper, we tackle both questions. First, we theoretically investigate optimal distributions of μ points that maximize the weighted hypervolume indicator. Second, based on the obtained theoretical results, we propose a new approach to articulate user preferences within biobjective hypervolume-based optimization in terms of specifying a desired density of points on a predefined (imaginary) Pareto front. Within this approach, a new exact algorithm based on dynamic programming is proposed which selects the set of μ points that maximizes the (weighted) hypervolume indicator. Experiments on various test functions show the usefulness of this new preference articulation approach and the agreement between theory and practice.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search

General Terms

Algorithms

1. INTRODUCTION

In the context of evolutionary multiobjective optimization, the hypervolume indicator is a set quality measure being a refinement of the Pareto dominance, i.e., whenever a Pareto set approximation entirely dominates another one, then the measure of the former will be larger. Hence, max-

imizing the hypervolume indicator is, in general, in accordance to the Pareto dominance relation [15, 13, 20]. In recent years, this significant property has made the indicator popular as a selection criterion in evolutionary multiobjective optimizers [4, 14, 18].

An important question in this context concerns the bias of the hypervolume indicator when introduced into search. In a recent study [2], this bias has been investigated theoretically by characterizing the distribution of μ points that maximize the hypervolume indicator, so-called *optimal μ -distributions*. When hypervolume-based algorithms are employed, the inherent bias of the hypervolume indicator is often unwanted but rather the incorporation of the user's own preference into search is desired. Several approaches for articulating these user preferences are known from the literature, e.g., by defining preference points [11], specifying preferred search directions [10] or defining linear minimum and maximum tradeoffs [6]. For a general overview of articulating user preferences, we refer to [16, 7, 17].

Recently, a general framework to incorporate user preferences into hypervolume-based search algorithms has been proposed [18]. The so-called *weighted hypervolume indicator* allows stressing certain regions of the objective space by applying a weight distribution. It has been shown for three different weight distribution functions that optimizing the weighted hypervolume indicator results in solutions clustered in regions with higher weight whereas regions with low weight contain only a few solutions. Further work improved the applicability of this weighted hypervolume approach to problems with many objectives by means of Monte Carlo sampling [1].

Two main questions arise in this context: (1) what is the bias of the *weighted* hypervolume indicator, i.e., how can the distribution of μ points maximizing the weighted hypervolume be characterized? (2) how can the user articulate and integrate her preferences easily into the search without defining the weight distribution functions by hand?

Both questions are tackled in this paper. First, we investigate the optimal distribution of μ points optimizing the weighted hypervolume indicator; generalizing the results for the unweighted case presented in [2] to the weighted case. Second, we propose a new approach to articulate user preferences by defining a reference front and a desired density of points on that front. This user preference is automatically transferred into a weight distribution function for the weighted hypervolume of [18], the optimization of which guarantees that the resulting distribution of solutions on the actual front follows the desired density.

This paper is organized as follows: In Sec. 3 we generalize the theoretical results on optimal μ -distributions from [2] to the weighted hypervolume. Based on the theoretical results on optimal μ -distributions, we propose in Sec. 4 a new approach for articulating user preferences in terms of defining a desired density of points on a predefined front for biobjective problems. In Sec. 4.2 we incorporate the weighted hypervolume into search and propose a new exact algorithm for selecting the best q individuals from a set of p solutions that maximize the overall (weighted) hypervolume by means of dynamic programming for biobjective problems. In Sec. (5), we apply the new approach to several test problems for scenarios with known and unknown Pareto fronts.

2. PRELIMINARIES

In this study we consider biobjective optimization problems $\mathcal{F} : X \rightarrow Z$. The two objective functions $(\mathcal{F}_1(x), \mathcal{F}_2(x)) = \mathcal{F}(x)$ map the *decision space* X to the *objective space* $Z \subseteq \mathbb{R}^2$ and are, without loss of generality, to be minimized. Let \preceq denote the *weak Pareto dominance relation* defined as $x \preceq y$ if and only if $\mathcal{F}_1(x) \leq \mathcal{F}_1(y)$ and $\mathcal{F}_2(x) \leq \mathcal{F}_2(y)$. The *Pareto-optimal set* P_s then consists of all solutions $x^* \in X$, such that $\nexists x \in X$ with $x \preceq x^*$ and $x^* \not\preceq x$. The image of P_s under \mathcal{F} is called *Pareto front* or *front* for short. For the sake of simplicity, we describe the Pareto front in terms of a function f mapping the image of the Pareto set under \mathcal{F}_1 —which is presumed to be a closed interval $[x_{min}, x_{max}]$ —onto the image of the Pareto set under \mathcal{F}_2 , see Fig. 1(a).

Throughout this study, we assume the optimization goal of maximizing the weighted hypervolume indicator of [18], i.e., finding a set A of μ solutions that maximizes the weighted hypervolume. A set of μ points maximizing the unweighted hypervolume indicator has been denoted in [2] as an *optimal μ -distribution* and we use the same term here for the weighted case as well. For simplicity, we define the weighted hypervolume indicator only for sets of objective vectors instead for solution sets $A \subseteq X$ as it was already done in [18] and [2]. For a set of μ biobjective objective vectors $S = \{(x_1^\mu, y_1^\mu), \dots, (x_\mu^\mu, y_\mu^\mu)\}$, we define the weighted hypervolume indicator $I_{H,w}^\mu$ according to [18] and as a generalization of [2] as

$$I_{H,w}^\mu(S) := \sum_{i=1}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \left(\int_{y_i^\mu}^{y_0^\mu} w(x, y) dy \right) dx \quad (1)$$

where $x_{\mu+1}^\mu = r_1$, $y_0^\mu = r_2$, and $\vec{r} = (r_1, r_2)$ is the reference point of the hypervolume indicator, see Fig. 1(a). Note that for points $(x_i, f(x_i))$ that all lie on the front described by f , Eq. 1 results in a μ -dimensional function $I_{H,w}^\mu(x_1^\mu, \dots, x_\mu^\mu)$ that is equivalent to Eq. 2 in [2] if we choose $w(x, y) = 1$.

3. THEORETICAL BACKGROUND

The theoretical results presented in this section build the basis of the new preference articulation approach proposed later on in this paper. To this end, we investigate the influence of the weighted hypervolume indicator on optimal μ -distributions by deriving the limit density when μ goes to infinity. We then build on this result and provide for a desired density the corresponding weight that will allow a user to articulate his preferences.

3.1 Limit Density of Optimal μ -Distributions for the Weighted Hypervolume

In [2], we argued that, besides for linear fronts, it is very difficult to determine the optimal μ -distributions for arbitrary fronts exactly and therefore investigated optimal μ -distributions in the limit case of μ to ∞ . This limit is expressed as a density—approximating the percentage of points in a particular segment of the front. Here, we generalize this approach so as to obtain the density of points associated to the weighted hypervolume. Before we do so, note that optimal μ -distributions exist in the weighted case under the same assumption as for the unweighted case in [2, Theorem 1]: if the function f describing the Pareto front is continuous, there exists (at least) one set of μ points maximizing the weighted hypervolume¹.

We assume without loss of generality that $x_{min} = 0$ and that $f : x \in [0, x_{max}] \mapsto f(x)$ with $f(x_{max}) = 0$ [2, Fig. 5]. We also assume that f is continuous within $[0, x_{max}]$, is differentiable and that its derivative is a continuous function f' defined in the interval $]0, x_{max}[$. An optimal μ distribution is defined as a set of μ points maximizing the weighted hypervolume indicator $I_{H,w}^\mu((x_1^\mu, \dots, x_\mu^\mu))$. However, instead of maximizing the weighted hypervolume indicator $I_{H,w}^\mu$, it is easy to see that, since $r_1 r_2$ is constant, one can equivalently minimize

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

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

$$\sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \int_0^{f(x_i^\mu)} w(x, y) dy dx - \int_0^{x_{max}} \int_0^{f(x)} w(x, y) dy dx \quad (2)$$

is equivalent to maximizing the weighted hypervolume indicator (Fig. 1(d)).

For a fixed integer μ , we consider a sequence of μ ordered points in $[0, x_{max}]$, $x_1^\mu, \dots, x_\mu^\mu$ that lie on the Pareto front. We assume that the sequence converges—when μ goes to ∞ —to a density $\delta(x)$ that is regular enough. Formally, the density in $x \in [0, x_{max}]$ is defined as the limit of the number of points contained in a small interval $[x, x+h[$ normalized by the total number of points μ when both μ goes to ∞ and h to 0, i.e., $\delta(x) = \lim_{\mu \rightarrow \infty} \lim_{h \rightarrow 0} \left(\frac{1}{\mu h} \sum_{i=1}^{\mu} \mathbf{1}_{[x, x+h[}(x_i^\mu) \right)$. As explained above, maximizing the weighted hypervolume is equivalent to minimizing Eq. 2, which is also equivalent to minimizing

$$E_\mu = \mu \left[\sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \left(\int_0^{f(x_i^\mu)} w(x, y) dy \right) dx - \int_0^{x_{max}} \left(\int_0^{f(x)} w(x, y) dy \right) dx \right] \quad (3)$$

¹The proof is similar to the one in [2]: the μ -dimensional function in Eq. 1 is continuous and upper bounded by the hypervolume of the entire front, i.e. $\int_{x_{min}}^{r_1} \int_{f(x_{max})}^{r_2} \mathbf{1}_{y > f(x)} w(x, y) dy dx$.

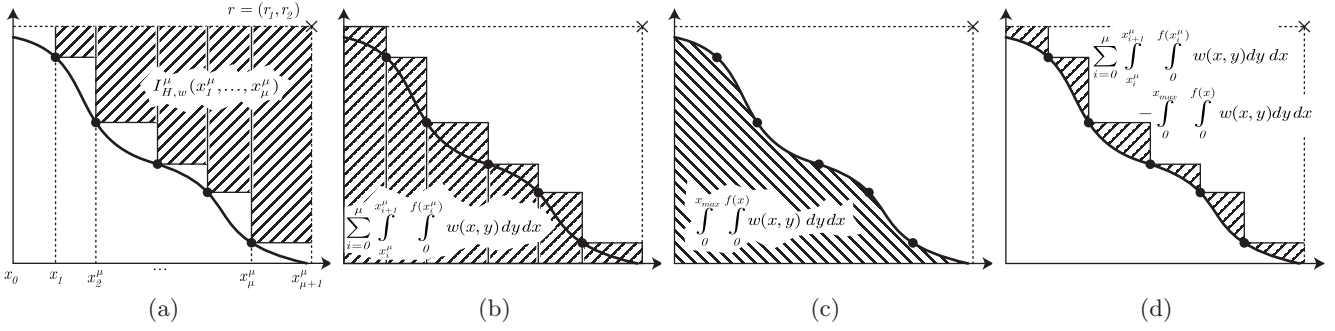


Figure 1: Illustration of the idea behind deriving the optimal density: Instead of maximizing the weighted hypervolume indicator $I_{H,w}^\mu((x_1^\mu, \dots, x_\mu^\mu))$ (a), one can minimize (b) which is equivalent to minimize the integral between the attainment surface of the solution set and the front itself (c) which can be expressed with the help of the integral of f (d).

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 $E(\delta) := \lim_{\mu \rightarrow \infty} E_\mu$. The first step therefore consists in computing the limit of E_μ .

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

$$\mu \sup \left(\left(\sup_{0 \leq i \leq \mu-1} |x_{i+1}^\mu - x_i^\mu| \right), |x_{max} - x_\mu^\mu| \right) \rightarrow c$$

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

$$E(\delta) := -\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)w(x, f(x))}{\delta(x)} dx . \quad (4)$$

PROOF. For the sake of readability, this proof has been sent to the appendix. \square

The limit density of the μ -distribution, as explained before, is minimizing $E(\delta)$. It remains therefore to find the density minimizing $E(\delta)$. This optimization problem is posed in a functional space, the Banach space $L^2(0, x_{max})$ and is also a constraint problem since the density δ has to satisfy the constraint $J(\delta) := \int_0^{x_{max}} \delta(x) dx = 1$. The constraint optimization problem (P) that needs to be solved is summarized in:

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

THEOREM 1. *The density solution of the constraint optimization problem (P) equals*

$$\delta(x) = \frac{\sqrt{-f'(x)w(x, f(x))}}{\int_0^{x_{max}} \sqrt{-f'(x)w(x, f(x))} dx} . \quad (5)$$

PROOF. This proof is similar to the proof of Theorem 7 in [2] where $-f'$ needs to be replaced everywhere by $-f'w$. \square

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

3.2 Deriving the Weight from Given Density

In the previous section, we characterized the density $\delta(x)$ of points that maximize the weighted hypervolume indicator for a given weight distribution $w(x, y)$ and front shape $f(x)$. Here, we use the result from the opposite direction: given user-defined preferences, expressed by a density³, we are interested in the corresponding weight distribution. This allows to model user preferences in a concise manner by optimizing the weighted hypervolume indicator as we will see in the following.

Instead of specifying the density $\delta(x)$, which is the density projected onto the x -axis, it is more convenient for the user to define the density δ_F of points on the front itself, which results from $\delta(x)$ by dividing it by the norm of the tangent at the corresponding point on the front, i.e., $\delta_F(x) = \delta(x) / \sqrt{1 + f'(x)^2}$. Transforming the density this way, we see that the density on the front $\delta_F(x)$ is proportional to $\sqrt{\frac{-f'(x)w(x, f(x))}{1 + f'(x)^2}}$ which implies that

$$w(x, f(x)) \propto \frac{1 + f'(x)^2}{-f'(x)} \cdot \delta_F(x)^2 . \quad (6)$$

Note that the weight is a strictly positive finite function if $-f'(x)$ is and that it peaks to infinity if the derivative of f either goes to zero or to $-\infty$.

Since it is more convenient to define the desired density as a function $\delta_F^\phi(\phi)$ of the angle ϕ rather than as a function of x , in the following we use $\delta_F(x) = \delta_F^\phi(\arctan(f(x)/x))$.

The weight distribution is only defined for points $(x, f(x))$ on the front. In order to make the weight useful for search, it is extended to the rest of the objective space in the following section.

4. ARTICULATING USER PREFERENCES

This section builds on the theoretical results from the previous section and proposes an approach to articulate user preferences in hypervolume-based evolutionary algorithms. First we present a way to generalize the weight distribution of Eq. 6 to the entire objective space. Second, we propose a method to perform environmental selection based

³Practically speaking, the user can define the desired density for example with the help of a graphical user interface in terms of piecewise linear or polynomial functions. For the user, this might be more convenient than specifying a weight distribution function over the entire objective space.

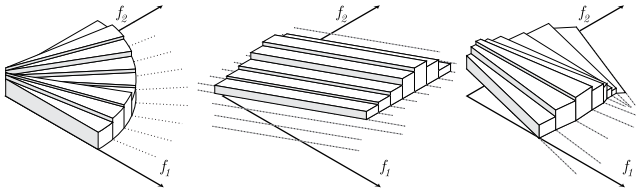


Figure 2: Different arrangements of stripes to represent a weight distribution function: radial (left), parallel (middle) or freely (right).

on the weight distribution derived. Thereby, the weighted hypervolume—and at the same time the user’s preference—is optimized.

4.1 Extending the Weight Distribution to the Entire Objective Space

In principle, any non-negative function can be chosen as the extension of w to points not on the front. For practical applications, however, a few things need to be considered:

1. The number of solutions is finite which leads to contributions that extend beyond the front. If the weight varies greatly in this region, the resulting distribution of points might be a bad approximation of δ_F^w .
2. During search, the solutions are not mapped to the Pareto front where the weight $w(x, f(x))$ is defined. In those cases, the resulting distribution should possibly meet the user’s expectations and not hinder convergence to the front.
3. The corresponding hypervolume indicator should be calculable in reasonable time.

To address all three issues, we propose to use a two-dimensional piecewise constant weight distribution function, where straight lines separate the objective space into stripes S_i of constant weight denoted by w_i , see Fig. 2⁴. A stripe S_i usually covers more than one Pareto-optimal point such that the weight w_i is not defined uniquely. Instead, we set the weight w_i to the mean

$$w_i = \frac{\int_{\gamma_i} w(s) ds}{\int_{\gamma_i} ds}, \quad \gamma_i : \{x \in \mathbb{R}^2 | (x, f(x)) \in S_i\} \mapsto \mathbb{R}^2, \quad (7)$$

where the curve γ_i denotes the part of the Pareto front inside S_i . Since the weight usually varies along γ_i , the weight according to (7) does not reflect (6) precisely. However, the approximation by w_i can be made arbitrarily accurate by increasing the number of stripes.

Besides facilitating the calculation of the hypervolume indicator, the stripes have the advantage of the weight remaining constant if a solution is moving along the direction of the stripe. Many different arrangements of stripes are possible, see Fig. 2. Here, we propose using rays starting from the ideal point $\bar{z}^I = (z_1^I, z_2^I)$ at equally separated angles ranging from 0° to 90° to separate the stripes

$$S_i^{rays} = \left\{ (x, y) | \phi_i \leq \arctan\left(\frac{x - z_1^I}{y - z_2^I}\right) < \phi_{i+1} \right\} \quad (8)$$

where $\phi_i = (i-1)/n \cdot 90^\circ$ and $1 \leq i \leq n$. For example, on the left-hand side of Fig. 2 nine rays and $z^I = (0, 0)$ are used.

⁴In this setting, the corresponding hypervolume calculation can be done easily by determining intersections of straight lines and calculating the area of triangles and trapezoids.

4.2 Optimizing the Hypervolume in 2D

The weighted hypervolume indicator represented by the stripes proposed before, shall be used as the optimization criterion of an evolutionary algorithm in order to find solutions that meet the corresponding user preference as good as possible. Several evolutionary algorithms aiming at maximizing the hypervolume indicator in their environmental selection step which can be formulated as solving the hypervolume subset selection problem (HSSP) [3]: given a set of solutions A and $0 \leq q \leq |A|$, find a subset $A^* \subseteq A$ with $|A^*| = q$, such that the weighted hypervolume indicator of A^* is maximal.

While for more than two objectives the HSSP problem is expected to be difficult and for this reason greedy heuristics are used to tackle the HSSP, e.g., in [4, 14, 18], we propose an efficient exact algorithm for the case of 2 objectives here—using the fact that the hypervolume contribution of an objective vector only depends on its two adjacent neighbors⁵. Exploiting this property, *dynamic programming* [8] can be used to solve the problem exactly in time $\mathcal{O}(|A|^3)$ as opposed to $\mathcal{O}(|A|^2)$ for the greedy approach by combining solutions of smaller subproblems P_c^{t-1} in a bottom-up fashion to solutions for larger subproblems P_c^t : for a fixed solution $a_c \in A$ and a $t \in \{0, \dots, |A|\}$, we define the subproblem P_c^t as finding the set $A_c^t \subseteq A$ of t solutions maximizing the hypervolume such that A_c^t contains a_c and in addition, only elements a_k lying to the right of a_c , i.e., $f_1(a_c) \leq f_1(a_k)$.

Obviously, $\{a_c\}$ is the solution for P_c^1 . According to the above made statement, the solution for P_c^t with $t > 1$ can now be easily found when considering the unions of $\{a_c\}$ with the solutions of all P_k^{t-1} with $f_1(a_c) \leq f_1(a_k)$ and taking the resulting solution set with the highest hypervolume. Once the solutions for $t = q$ are determined, the subset which then has the largest hypervolume corresponds to the solution to the overall problem. Algorithm 1 shows the pseudo code of the procedure where sets S_c^t of indices instead of the sets A_c^t are considered for clarity. We illustrate the algorithm by means of an example:

EXAMPLE 1. Consider six objective vectors o_1 to o_6 of which one likes to choose those $q = 3$ that maximize the hypervolume, see Fig. 3. In the first stage (a), the optimal subsets of size 1 and their hypervolume value are calculated (Lines 1 and 2 in Alg. 1). Please note that some subsets do not exist or will not be used to build the overall solution and can therefore be neglected (dashes).

In the next stages, the subsets of size $t = 2$ to q (Lines 3-13) are determined for all individuals o_c (Lines 4-12). To this end, the hypervolume of combining o_c with any subset to its right of size $t - 1$ (Lines 6-8) are calculated. For example, in Fig. 3(b) o_3 is combined with the subset S_5^1 to form $S_3^2 = \{3, 5\}$ with hypervolume $h_3^2 = 127$. In this way, all subsets of size 2 (c) and then of size 3 (d) are determined.

Reaching $t = q$, the optimal solution to the overall problem corresponds to the set with the largest hypervolume, in this example $S_1^3 = \{1, 3, 5\}$ with value $h_1^3 = 142$ (Line 14).

Note that the advantage of the exact algorithm over often used greedy approaches for HSSP is that it overcomes the non-convergence of greedy algorithms, see [5] for details.

⁵For example, in Figure 3(a) the hypervolume contribution of o_2 is bounded by o_1 and o_3 but not by o_4 , o_5 or o_6 . This in turn means, that the increase in hypervolume, when adding o_2 to any subset whose left-most element is o_3 , is equal.

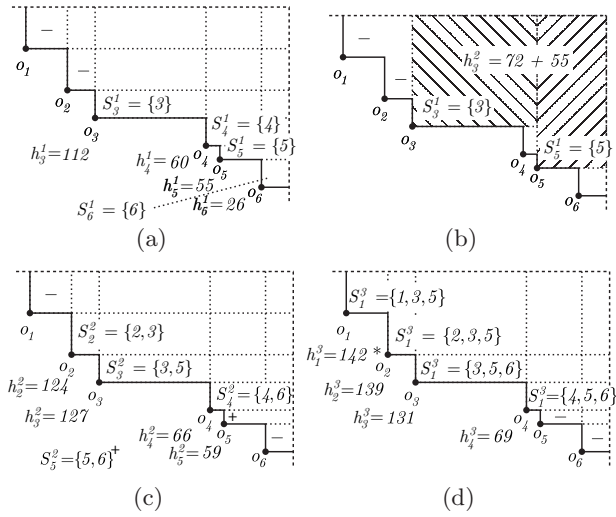


Figure 3: Three out of six objective vectors need to be selected. Algorithm 1 starts by calculating subsets of size 1 (a). Then the results (b) are combined to sets of size of 2 (c) and finally of size 3 (d).

5. EXPERIMENTS

In this section, the theoretical concept to model (Sec. 3) and to optimize (Sec. 4) user preferences by a weighted hypervolume indicator are examined by simulations. To this end, the algorithm and the test problems used are presented first. After that, the Pareto front approximations obtained by the algorithm are visually compared to the expected density that corresponds to the user's preference.

5.1 Experimental Setup

The approach derived in Sec. 4 is applied to different test problems and user preferences. Unless noted otherwise, the latter is modeled by 90 stripes starting from the origin which corresponds to a resolution of 1° , see Eq. 8. The reference point of the hypervolume is set to (1000, 1000) to ensure the extremes in the optimal μ -distribution if possible [2].

The evolutionary algorithm in use first generates an initial population of α individuals that code potential solutions as real vectors. Thereafter, μ parent individuals are selected uniformly at random to generate λ offspring using the SBX crossover operator and the variable-wise polynomial mutation operator [9]. The recombination and mutation probabilities were set according to [12]. After that, non-dominated sorting and front-wise removal is carried out as in [18]. From the set F of equally ranked solutions, the omission of which would result in a population P' that is smaller than α , the $q = \alpha - |P'|$ solutions computed by Alg. 1 that solve the HSSP on F are taken to fill up the population to $|P'| = \alpha$. The above steps are repeated for N generations. In this study $N = 1,000$ and $\alpha = \lambda = \mu = 50$. The main difference between the described algorithm and standard hypervolume-based algorithms such as SIBEA [18] is the environmental selection step where Alg. 1 is used instead of a greedy strategy to solve the HSSP.

As test problems, ZDT1 (front shape $f(x) = 1 - \sqrt{x}$), ZDT2 ($f(x) = 1 - x^2$), and ZDT3 (discontinuous front) [19] as well as DTLZ2 ($f(x) = \sqrt{1 - x^2}$) [12] are used. For some simulations, the objective functions are offset or scaled.

Algorithm 1 HSSP-SOLVER(O, q, r, C_H^w)

Require: A matrix $O := (o_{i,j})_{p \times 2}$, where the rows represent the objective vectors sorted in ascending order according to the first objective; subset size $q \geq 1$; reference point $r = (r_1, r_2)$. $C_H^w(\vec{l}, \vec{u})$ returns the weighted hypervolume of the rectangle from \vec{l} to \vec{u} .

Ensure: Returns a set S that references rows of O that maximize the weighted hypervolume.

- 1: $h_i^1 \leftarrow h_i := C_H^w((o_{i,1}, o_{i,2}), r) \quad \forall 1 \leq i \leq p$
- 2: $S_c^1 \leftarrow \{c\} \quad \forall q \leq c \leq p \quad \triangleright$ optimal subsets
- 3: **for** $t = 2$ to q **do** \triangleright bottom-up approach
- 4: **for** $c = q - t + 1$ to $p - t + 1$ **do** \triangleright subproblem
- 5: $l \leftarrow (0, \dots, 0)$
- 6: **for** $d = c + 1$ to $p - t + 1$ **do**
- 7: $l_d = h_d^{t-1} + C_H^w((o_{c,1}, o_{c,2}), (o_{d,1}, r_2))$
- 8: **end for**
- 9: $m \leftarrow \arg \max_i l_i$
- 10: $S_c^t \leftarrow \{c\} \cup S_m^{t-1} \quad \triangleright$ merge c with S_m and ...
- 11: $h_c^t \leftarrow l_m \quad \triangleright$... update hypervolume
- 12: **end for**
- 13: **end for**
- 14: $m \leftarrow \arg \max_i h_i^q \quad \triangleright$ pick best subset
- 15: **return** $S_m^q \quad \triangleright$ solution to overall problem

Table 1: Densities used for the experiments.

Figure	Density
4(a)	$\delta_F^\phi(\phi) \propto 1 - (\phi/90^\circ)^3$
4(b)	$\delta_F^\phi(\phi) \propto \begin{cases} 1, & \phi \in \{30^\circ, 50^\circ, [70^\circ, 90^\circ]\} \\ 0.1, & \text{otherwise} \end{cases}$
4(c)	$\delta_F^\phi(\phi) \propto 3 \cdot (5 \cdot 10^{-3} + (\phi - 45^\circ)^2)^{-1} + (5 \cdot 10^{-3} + (\phi - 10^\circ)^2)^{-1}$
4(d)	as in 4(b)
4(e)	$\delta_F^\phi(\phi) \propto 1$
4(f)	$\delta_F^\phi(\phi) \propto \begin{cases} 3, & \phi \in [78.3^\circ, 83.8^\circ] \\ 1, & \phi \in [39.7^\circ, 54.8^\circ] \\ 2, & \phi \in [14.9^\circ, 33.3^\circ] \\ 1 \cdot 10^{-3}, & \text{otherwise} \end{cases}$
5(a) to 5(f)	as in 4(c)

5.2 Results

In the following, the distribution of points obtained when using different weight functions and test problems are presented. To this end, a graphical representation is used that shows multiple elements (see Fig. 4(a)): (i) the solutions found (\bullet) together with the Pareto front (\curvearrowright); (ii) the front, specified by the user, that is used to calculate the weight rays (\curvearrowright) if differing from the Pareto front; (iii) the weight per ray proportional to the length of gray bars (▬) and at intervals of 10° the direction of the rays (\curvearrowright); (iv) the density desired by the user (\curvearrowright) as a function of the angle in polar coordinates with 0 lying at the radius where the weight bars start (e.g., at $r = 1.2$ for Fig. 4(a)); and (v) the obtained density represented by the solutions in terms of a histogram (\curvearrowright) in the same coordinate system than the desired density.

5.2.1 Test Problems with Known Pareto Fronts

If the Pareto front is known, e.g., for test problems, the weight rays directly follow from Eq. 6 and 7. Figures 4(a) to 4(c) show the results on ZDT1 for (a) stressing the second objective (b) desiring a piecewise constant density and, (c) preferring two preference points. Analytic descriptions of

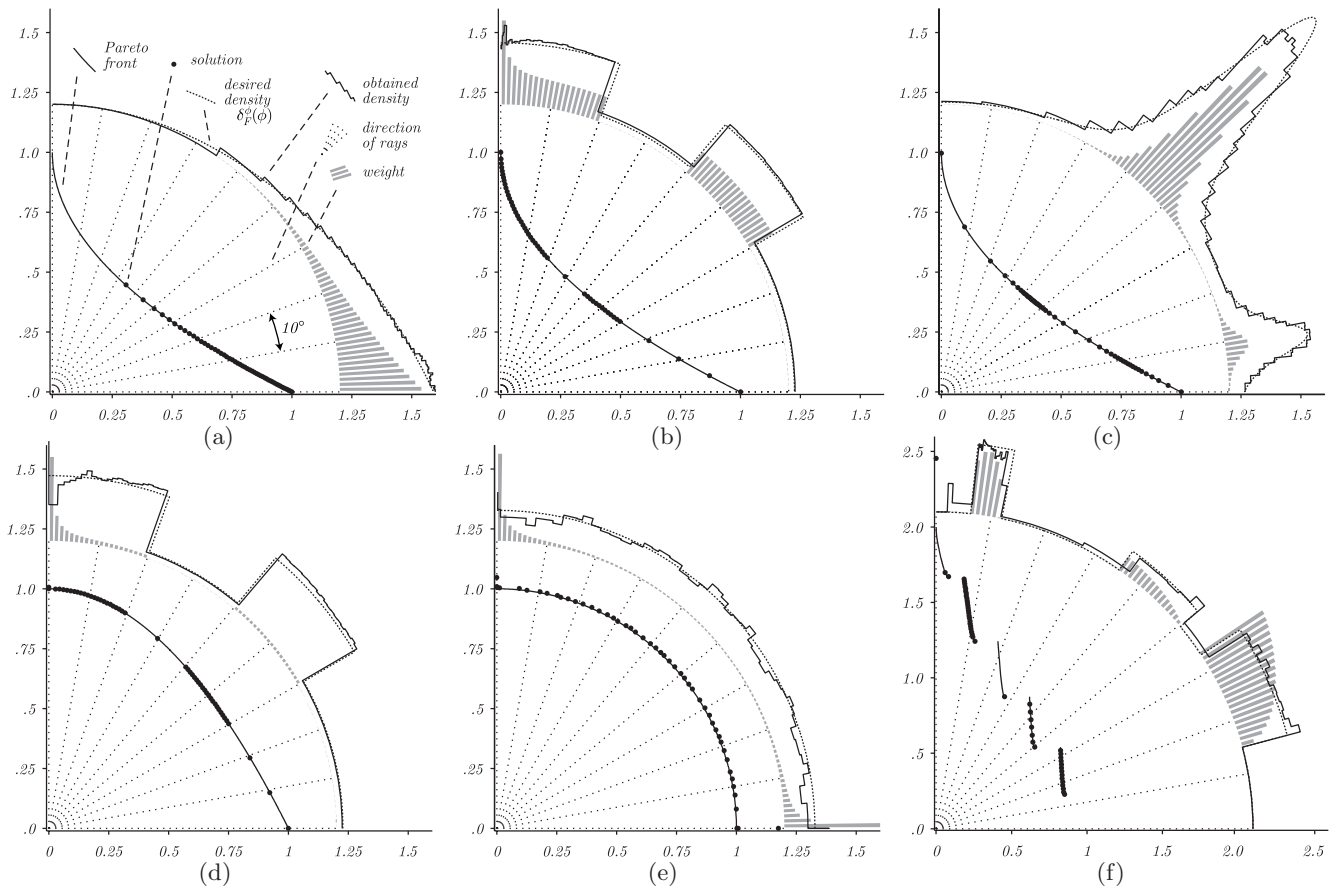


Figure 4: Shows the solutions found together with the Pareto front for different test problems, i.e., ZDT1 (a)-(c), ZDT2 (d), DTLZ2 (e) and ZDT3 (f). The underlying densities (listed in Tab. 1) are shown as a function of the angle, as well as the corresponding weight and the obtained density (see text for details).

the desired densities are listed in Tab. 1.

Figures 4(d) to 4(f) show: (d) the same weight as in (b) but for ZDT2, (e) trying to obtain a uniform density on DTLZ2, and (f) focusing on different parts of the discontinuous front of ZDT3. As one can see, the distribution of solutions found closely matches the predetermined density. Difficulties arise mainly where the front is very steep or flat and therefore a high weight is necessary to compensate this, e.g., at the extremes of DTLZ2 in Fig. 4(e).

5.2.2 The Typical Case of Unknown Fronts

If the front is unknown like in practice, one has to assume a front shape or find an approximate front shape by running an algorithm with respect to the unweighted hypervolume. For simplicity, we assume the front to be linear here. Then, the obtained density does not correspond to the desired density if the real front differs from the assumption or if the front is not reached. In Fig. 5(a) to 5(c) the weight distribution is derived for an assumed linear front, but the real front was (a) convex, (b) concave, and (c) concave while being scaled. In the latter example, the density is not affected by the scaling due to the radial nature of the weight rays.

If the assumed front lies at a different position than the real front, the obtained density differs from the desired one, see Fig. 5(d) and (e), even if the shape is correct as in (d).

5.2.3 Number of Weight Rays

Setting the number of rays to 90 does not lead to any visually noticeable impreciseness. However, if the resolution is decreased by using only a small number of rays, e.g., 15 in Fig. 5(f), the density obtained clearly reflects the step like character of the weight distribution.

6. CONCLUSIONS

The hypervolume indicator has an inherent characteristic in terms of how a set of μ solutions, optimizing the hypervolume indicator, is distributed on the Pareto front, which has been expressed in terms of density in [2]. However, the bias can be modified by applying weight distribution functions as shown in [18]. Here, the two precursor studies have been combined to derive the density for the weighted case. On the one hand, this allows to determine the bias of the weighted hypervolume in a concise manner. On the other hand, the dual viewpoint has been taken to determine the weight distribution for a given density and front shape. This enables to express user preferences in a very precise way by a unary indicator.

To utilize the indicator for search, an approximation of the weight distribution function by means of piecewise constant rays has been suggested. It allows simplifying the hypervolume calculation while maintaining desirable properties. Additionally, an algorithm based on dynamic programming

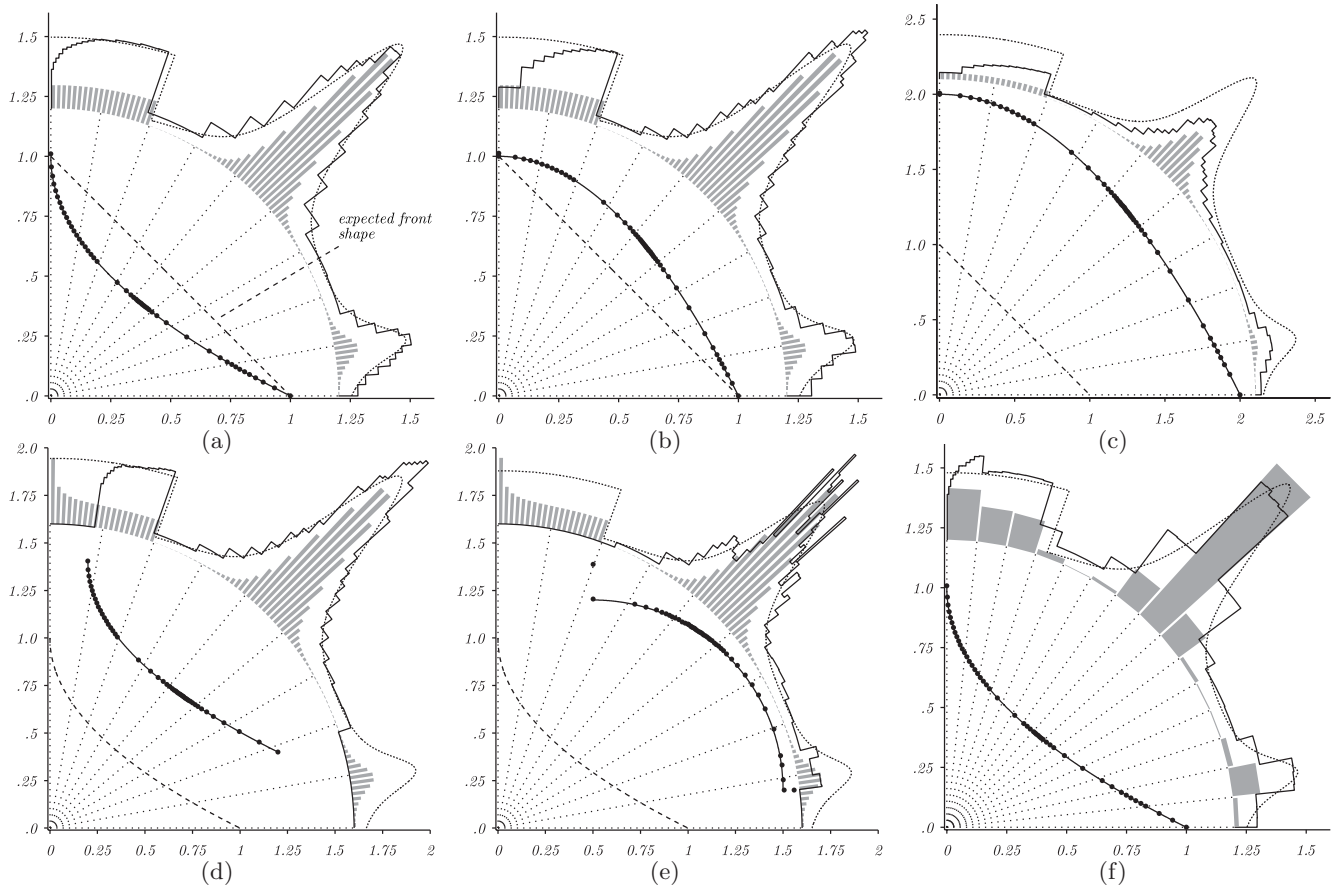


Figure 5: See Fig. 4(b) for a legend to the visual elements. Additionally, the front used to calculate the weight is depicted. The test problems are (modified versions) of ZDT1 for (a), (d) and (f); ZDT2 (b); and DTLZ2 (e). The underlying density is the same in all six cases (see Tab. 1), consult the text for details.

that selects solutions yielding maximal hypervolume value has been proposed. A visual comparison of the obtained Pareto front approximations with the expected density—provided by the user—has shown the practical usefulness of the theoretical concept as well as the efficiency of the proposed algorithm. The approach is restricted to biobjective problems, though. Therefore, we should aim at extending it to more objectives in future studies. The approach has also to show its usefulness in real-world applications.

APPENDIX

Proof of Lemma 1

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

$$\int_0^{x_{max}} \frac{|f'(x)w(x, f(x))|}{|\delta(x)|} dx \leq \frac{\int_0^{x_{max}} (f'(x)w(x, f(x)))^2 dx}{\int_0^{x_{max}} (1/\delta(x))^2 dx} \quad (9)$$

and since $x \rightarrow f'(x)w(x, f(x)) \in L^2(0, x_{max})$ and $\frac{1}{\delta} \in L^2(0, x_{max})$, the right-hand side of Eq. 9 is finite and Eq. 4 is well-defined.

Step 1. In a first step we are going to prove that E_μ defined in Eq. 3 satisfies

$$E_\mu = \mu \sum_{i=0}^{\mu} \left(-\frac{1}{2} f'(x_i^\mu) w(x_i^\mu, f(x_i^\mu)) (x_{i+1}^\mu - x_i^\mu)^2 + O((x_{i+1}^\mu - x_i^\mu)^3) \right) \quad (10)$$

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

$\int_0^{x_{max}} \int_0^{f(x)} w(x, y) dy dx$ into $\sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \int_0^{f(x_i^\mu)} w(x, y) dy dx$, where we have used the fact that $\int_{x_{max}}^{x_{\mu+1}^\mu} \int_0^{f(x)} w(x, y) dy dx = 0$. Using the right hand side of the previous equation in Eq. 3, we find that

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

and thus

$$E_\mu = \mu \sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \int_{f(x)}^{f(x_i^\mu)} w(x, y) dy dx \quad (11)$$

At the first order, we have that

$$\int_{f(x)}^{f(x_i^\mu)} w(x, y) dy = w(x_i^\mu, f(x_i^\mu)) (f(x_i^\mu) - f(x)) + O((x - x_i^\mu)^2) \quad (12)$$

Since f is differentiable, we can use a Taylor approximation of f in each interval $[x_i^\mu, x_{i+1}^\mu]$ and write $f(x) = f(x_i^\mu) + f'(x_i^\mu)(x - x_i^\mu) + O((x - x_i^\mu)^2)$, which thus implies that

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

and thus the left hand side of Eq. 12 becomes

$$-w(x_i^\mu, f(x_i^\mu)) f'(x_i^\mu) (x - x_i^\mu) + O((x - x_i^\mu)^2)$$

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

$$\int_{x_i^\mu}^{x_{i+1}^\mu} \int_{f(x)}^{f(x_i^\mu)} w(x, y) dy dx = -0.5 w(x_i^\mu, f(x_i^\mu)) f'(x_i^\mu) (x_{i+1}^\mu - x_i^\mu)^2 + O((x_{i+1}^\mu - x_i^\mu)^3)$$

Summing up for $i = 0$ to $i = \mu$, multiplying by μ and using Eq. 11 we obtain Eq. 10, 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

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

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

$$\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)w(x, f(x))}{\delta(x)} dx = \frac{1}{2} \sum_{i=0}^{\mu} \int_{x_i^\mu}^{x_{i+1}^\mu} \frac{f'(x)w(x, f(x))}{\delta(x)} dx \quad (13)$$

For μ to ∞ , the assumption $\mu \sup((\sup_{0 \leq i \leq \mu-1} |x_{i+1}^\mu - x_i^\mu|), |x_{max} - x_\mu^\mu|) \rightarrow c$ implies that the distance between two consecutive points $|x_{i+1}^\mu - x_i^\mu|$ as well as $|x_\mu^\mu - x_{max}|$ converges to zero. Let $x \in [0, x_{max}]$ and let us define for a given μ , $\varphi(\mu)$ as the index of the points such that $x_{\varphi(\mu)}^\mu$ and $x_{\varphi(\mu)+1}^\mu$ surround x , i.e., $x_{\varphi(\mu)}^\mu \leq x < x_{\varphi(\mu)+1}^\mu$. Since we assume that δ is continuous, a first order approximation of $\delta(x)$ is $\delta(x_{\varphi(\mu)}^\mu)$, i.e. $\delta(x) = \delta(x_{\varphi(\mu)}^\mu) + O(x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu)$ and therefore by integrating between $x_{\varphi(\mu)}^\mu$ and $x_{\varphi(\mu)+1}^\mu$ we obtain

$$\int_{x_{\varphi(\mu)}^\mu}^{x_{\varphi(\mu)+1}^\mu} \delta(x) dx = \delta(x_{\varphi(\mu)}^\mu)(x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu) + O((x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu)^2) \quad (14)$$

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

$$\mu \int_{x_{\varphi(\mu)}^\mu}^{x_{\varphi(\mu)+1}^\mu} \delta(x) dx = 1 + O((x_{\varphi(\mu)+1}^\mu - x_{\varphi(\mu)}^\mu)) \quad (15)$$

Using Eq. 14 and Eq. 15, we thus have

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

Therefore for every i we have that

$$\frac{1}{\delta(x_i^\mu)} = \mu(x_{i+1}^\mu - x_i^\mu) + O(\mu(x_{i+1}^\mu - x_i^\mu)^2). \quad (16)$$

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

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

Injecting Eq. 16 in the previous equation, we obtain

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

Multiplying by $1/2$ and summing up for i from 0 to μ and using Eq. 13 we obtain

$$\frac{1}{2} \int_0^{x_{max}} \frac{f'(x)w(x, f(x))}{\delta(x)} dx = -E_\mu + \sum_{i=0}^{\mu} O(\mu(x_{i+1}^\mu - x_i^\mu)^3) \quad (17)$$

Let us define Δ_μ as $\sup((\sup_{0 \leq i \leq \mu-1} |x_{i+1}^\mu - x_i^\mu|), |x_{max} - x_\mu^\mu|)$. By assumption, we know that $\mu\Delta_\mu$ converges to a positive constant c . The last term of Eq. 17 satisfies

$$|\sum_{i=0}^{\mu} O(\mu(x_{i+1}^\mu - x_i^\mu)^3)| \leq K\mu^2(\Delta_\mu)^3$$

where $K > 0$. Since $\mu\Delta_\mu$ converges to c , $(\mu\Delta_\mu)^2$ converges to c^2 . With Δ_μ converges to 0, we therefore have that $\mu^2\Delta_\mu^3$ converges to 0. Taking the limit in Eq. 17 we therefore obtain

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

A. REFERENCES

- [1] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Articulating User Preferences in Many-Objective Problems by Sampling the Weighted Hypervolume. In *Genetic and Evolutionary Computation Conference (GECCO 2009)*. ACM, 2009.
- [2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the Hypervolume Indicator: Optimal μ -Distributions and the Choice of the Reference Point. In *Foundations of Genetic Algorithms (FOGA 2009)*. ACM, 2009.
- [3] J. Bader and E. Zitzler. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. TIK Report 286, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Nov. 2008.
- [4] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume. *European Journal on Operational Research*, 181:1653–1669, 2007.
- [5] N. Beume, B. Naujoks, M. Preuss, and G. R. amd T. Wagner. Effects of 1-Greedy \mathcal{S} -Metric-Selection on Innumerably Large Pareto Fronts. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, pages 21–35. Springer, 2009.
- [6] J. Branke, T. Kaufler, and H. Schmeck. Guidance in Evolutionary Multi-Objective Optimization. *Advances in Engineering Software*, 32:499–507, 2001.
- [7] C. A. Coello Coello. Handling Preferences in Evolutionary Multiobjective Optimization: A Survey. In *Congress on Evolutionary Computation (CEC 2000)*, pages 30–37. IEEE Press, 2000.
- [8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2nd edition, 2001.
- [9] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, Chichester, UK, 2001.
- [10] K. Deb and A. Kumar. Interactive Evolutionary Multi-Objective Optimization and Decision-Making using Reference Direction Method. In *Genetic and Evolutionary Computation Conference (GECCO 2007)*, pages 781–788. ACM, 2007.
- [11] K. Deb, J. Sundar, U. B. Rao N., and S. Chaudhuri. Reference Point Based Multi-Objective Optimization Using Evolutionary Algorithms. *Int. Journal of Computational Intelligence Research*, 2(3):273–286, 2006.
- [12] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable Test Problems for Evolutionary Multi-Objective Optimization. In *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, chapter 6, pages 105–145. Springer, 2005.
- [13] M. Fleischer. The measure of Pareto optima. Applications to multi-objective metaheuristics. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 519–533, Faro, Portugal, 2003. Springer.
- [14] C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, 2007.
- [15] J. Knowles and D. Corne. On Metrics for Comparing Non-Dominated Sets. In *Congress on Evolutionary Computation (CEC 2002)*, pages 711–716. IEEE Press, 2002.
- [16] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, MA, USA, 1999.
- [17] L. Rachmawati and D. Srinivasan. Preference Incorporation in Multi-objective Evolutionary Algorithms: A Survey. In *Congress on Evolutionary Computation (CEC 2006)*, pages 962–968. IEEE Press, 2006.
- [18] E. Zitzler, D. Brockhoff, and L. Thiele. The Hypervolume Indicator Revisited: On the Design of Pareto-compliant Indicators Via Weighted Integration. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, pages 862–876, Berlin, 2007. Springer.
- [19] E. Zitzler, K. Deb, and L. Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, 2000.
- [20] E. Zitzler, L. Thiele, and J. Bader. On Set-Based Multiobjective Optimization (Revised Version). TIK Report 300, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Dec. 2008.