Articulating User Preferences in Many-Objective Problems by Sampling the Weighted Hypervolume

Anne Auger, Johannes Bader, Dimo Brockhoff, Eckart Zitzler

HAL Id: hal-00431271

https://hal.archives-ouvertes.fr/hal-00431271

Submitted on 11 Nov 2009

Articulating User Preferences in Many-Objective Problems by Sampling the Weighted Hypervolume

Anne Auger[*]   Johannes Bader[†]

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

Dimo Brockhoff[†]   Eckart Zitzler[†]

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

ABSTRACT

The hypervolume indicator has become popular in recent years both for performance assessment and to guide the search of evolutionary multiobjective optimizers. Two critical research topics can be emphasized with respect to hypervolume-based search: (i) the hypervolume indicator inherently introduces a specific preference and the question is how arbitrary user preferences can be incorporated; (ii) the exact calculation of the hypervolume indicator is expensive and efficient approaches to tackle many-objective problems are needed. In two previous studies, we addressed both issues independently: a study proposed the *weighted hypervolume indicator* with which user-defined preferences can be articulated; other studies exist that propose to estimate the hypervolume indicator by *Monte-Carlo sampling*. Here, we combine these two approaches for the first time and extend them, i.e., we present an approach of sampling the weighted hypervolume to incorporate user-defined preferences into the search for problems with many objectives. In particular, we propose weight distribution functions to stress extreme solutions and to define preferred regions of the objective space in terms of so-called preference points; sampling them allows to tackle problems with many objectives. Experiments on several test functions with up to 25 objectives show the usefulness of the approach in terms of decision making and search.

Categories and Subject Descriptors

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

General Terms

Algorithms

Keywords

Hypervolume indicator, preference articulation, Monte Carlo sampling

1. INTRODUCTION

The hypervolume indicator has been proposed about a decade ago to compare the performance of multiobjective optimization algorithms [26]. It measures the quality of a set of solutions quantitatively as the "size of the space covered" [26]. Equivalent definitions based on the Lebesgue measure [19], or the attainment function [23] have been proposed later. The hypervolume indicator has also gained interest as a selection criterion within multiobjective evolutionary algorithms in recent years [17, 21, 4, 23, 3], especially if the number of objectives is large [3]. The main reason for the hypervolume's popularity is its property of refining the Pareto dominance relation: whenever a solution set is dominating another, the hypervolume indicator value of the former is strictly larger than the one of the latter [27]—resulting in the fact that a set that optimizes the hypervolume indicator covers the entire Pareto front [15]. Up to now, no unary indicator is known that is not based on the hypervolume and that has the refinement property.

Clearly, using the hypervolume indicator for guiding search introduces a certain preference. We investigated this recently by characterizing the distribution of μ points that maximize the hypervolume indicator [2]. One question that is of special interest in practice is whether the inherent preference of the hypervolume indicator can be changed to arbitrary user preferences, e.g., towards extreme solutions or towards so-called preference points[1]. There already exist various studies that focus on the issue of preference articulation in EMO, in particular integrating preferences such as priorities, goals, and preference points [5, 11, 22]. However, none of these methods leads to a Pareto-compliant relation—to obtain such relations we proposed a generalized version of the hypervolume indicator where arbitrary user preferences can be incorporated by stressing certain regions of the objective space in terms of a weight distribution function [23].

In the same paper [23], we also propose three exemplary weight distribution functions to incorporate user preferences towards the extremes of the front and towards predefined preference points for biobjective problems. The study is only preliminary and inhibits two drawbacks: (i) the proposed weight distribution function for articulating preference points is not easily extendable to more than two objectives and (ii) the exact computation of the hypervolume

---

[1]Instead of the standard term *reference point*, see for example [20], we use the term *preference point* throughout this paper to reduce the likelihood of confusion with the hypervolume's reference point.

indicator is expensive if the number of objectives is high[2].

In this paper, we tackle these two drawbacks by *estimating* the weighted hypervolume indicator by Monte Carlo sampling and propose according weight distribution functions that can be used for incorporating user preferences for an arbitrary number of objectives. In particular, we

- introduce the approach of sampling the weighted hypervolume indicator based on the work for the standard hypervolume indicator in [3] to avoid the exponential running time of the hypervolume indicator,
- propose two weight distributions to articulate preferences towards extremes and towards predefined preference points which can be arbitrarily combined and applied to problems with any number of objectives and
- show the potentials of the new approach experimentally for several test problems with up to 25 objectives by means of both visual inspection and statistical tests.

The paper is organized as follows. We start with a brief review of the weighted hypervolume indicator of [23] (Sec. 2). Section 3 presents the general ideas of sampling the weighted hypervolume after which two examples of weight distributions and their combination are introduced in Sec. 4. An extensive experimental validation is presented in Sec. 5 before Sec. 6 concludes the paper.

2. THE WEIGHTED HYPERVOLUME

Throughout this study, we assume without loss of generality, that k objective functions $f_i : X \mapsto \mathbb{R}$ $(1 \leq i \leq k)$ need to be minimized that map solutions x from the decision space X to their objective vector $f(x) = (f_1(x), \ldots, f_k(x))$. Furthermore, we assume the underlying dominance structure is given by the weak Pareto dominance relation where a solution $a \in X$ *weakly dominates* another solution $b \in X$, written $a \preceq b$, if and only if $f_i(a) \leq f_i(b)$ holds for all $1 \leq i \leq k$. We call a solution $x^* \in X$ *Pareto optimal* if there is no other $x \in X$ that *dominates* x^*, i.e., $\nexists x \in X : x \preceq x^* \wedge x^* \npreceq x$. The set of all Pareto optimal solutions is called *Pareto set* and its image *Pareto front*.

In this paper, we consider the following optimization goal: find a set A of μ compromise solutions such that the weighted hypervolume indicator value $I_H^w(A)$ of [23] is maximized. To this end, we recall the definition of the weighted hypervolume indicator $I_H^w(A)$ for a solution set $A \subseteq X$ from [23]:

$$I_H^w(A) = \int_{\vec{l}}^{\vec{r}} w(\vec{z}) \cdot \alpha_A(\vec{z}) d\vec{z} \qquad (1)$$

where $\alpha_A(\vec{z}) = \begin{cases} 1 & \text{if } f(A) \preceq \{\vec{z}\} \\ 0 & \text{else} \end{cases}$ is the attainment function[3] of a solution set $A \subseteq X$, $\vec{z} \in \mathbb{R}^k$ and $w : \mathbb{R}^k \to \mathbb{R}_{>0}$ is a weight distribution function where without loss of generality[4] $\int_{\vec{l}}^{\vec{r}} w(\vec{z}) d\vec{z} = 1$; the limits $\vec{l} \in \mathbb{R}^k$ and $\vec{r} \in \mathbb{R}^k$ of the integral are parameters of the hypervolume, where usually $\vec{l}$ is set to $(-\infty, \ldots, -\infty)$ and $\vec{r}$ is called the *reference point* of the indicator. We will see in the following section, that these

---

[2]Meanwhile, the #$\mathcal{P}$-hardness proof in [6] has theoretically shown that the hypervolume computation is indeed exponential in the number of objectives unless $\mathcal{P} = \mathcal{NP}$.

[3]Where $\preceq$ is generalized to sets as follows: $A \preceq B$ iff $\forall b \in B : \exists a \in A : a \preceq b$

[4]If $\int_{\vec{l}}^{\vec{r}} w(\vec{z}) d\vec{z} \neq 1$ we can normalize w by $\int_{\vec{l}}^{\vec{r}} w(\vec{z}) d\vec{z}$ without changing the ordering of solutions with respect to the hypervolume indicator.

points do not need to be fixed when the weighted hypervolume is sampled according to Gaussian distributions where the integral over the entire objective space is estimated.

3. SAMPLING THE HYPERVOLUME INDICATOR

Monte Carlo sampling is a well-known and easy-to-use approach to solve problems numerically by using random numbers. Monte Carlo sampling is used within several application areas such as atomic physics or finance. However, its most popular field of application is the computation of integrals [7].

3.1 Sampling the Non-Weighted Hypervolume

Using Monte Carlo methods to evaluate the hypervolume indicator is not new. Everson et al. [14] sampled the standard hypervolume for performance assessment whereas Bader and Zitzler [3] proposed to sample the hypervolume within multiobjective search. The basic idea behind sampling the hypervolume is to draw N independent random samples $X_1, \ldots, X_N$ from the objective space uniformly at random within a defined (hyper-)box and use the fraction of dominated objective vectors as an estimate for the hypervolume indicator value $I_H(A)$. The attainment function α_A of a solution set $A \subseteq X$ formalizes the domination of objective vectors by assigning 1 to an objective vector if it is dominated by A and 0 otherwise. The unweighted hypervolume indicator can, thus, be estimated as[5]

$$I_H(A) \approx \frac{1}{N} \sum_{k=1}^{N} \alpha_A(X_k) \ . \qquad (2)$$

As indicated in [3], the space from which the objective vectors are sampled should be as small as possible to avoid unnecessary samples, e.g., the (hyper-)box could have the hypervolume's reference point as one corner and the ideal point as the opposite one. When used during the selection step of a hypervolume-based multiobjective evolutionary algorithms, e.g., in SMS-EMOA [4] or MO-CMA-ES[18], the boxes can be much smaller if only the *hypervolume loss* of a single solution has to be estimated, see [3] for details.

3.2 Sampling the Weighted Hypervolume

The straightforward way of approximating the weighted hypervolume indicator of Eq. 1 is to sample $X_1, \ldots, X_N$ uniformly at random as before and use

$$I_H^w(A) \approx \frac{1}{N} \sum_{k=1}^{N} w(X_k) \alpha_A(X_k)$$

as an estimate of the weighted hypervolume. In this approach, however, the precision of the estimation heavily depends on the weight distribution w: if the support of w is small, the number of samples N needs to be large to have a reliable estimation. Using Hoeffding's inequality [16], one can show that the length of a confidence interval for a given confidence level is proportional to the supremum of w. In the

---

[5]The hypervolume indicator satisfies $I_H(A) = E(\alpha_A(X))$, where X is a random variable uniformly distributed on the objective space. Therefore from the strong law of large numbers the right-hand side of Eq. 2 converges to $I_H(A)$ for the number of samples N to infinity.

extreme case of a dirac "function" as suggested in [23] this would result in an infinite length for the confidence interval.

Therefore, we propose a different approach to sample the weighted hypervolume indicator here. Since w is positive and $\int_l^{\vec{r}} w(\vec{z})d\vec{z} = 1$, the weight function w is a density function. In principle, any density function can be used as w. For an efficient way of sampling, however, we choose w in the following such that we are able to sample efficiently random numbers distributed according to w. For this reason, multivariate normal distributions and exponential distributions will be used for sampling non-uniformly.

To give the explicit expression of the Monte Carlo estimator let X^w denote a random variable admitting w as probability density function[6]. Let $X_1^w, \ldots, X_N^w$ be N independent samples of random variables distributed as X^w. The weighted hypervolume $I_H^w(A)$ corresponds to the expected value of the random variable $\alpha_A(X^w)$ and can be approximated by

$$I_H^w(A) \approx \frac{1}{N}\sum_{k=1}^{N}\alpha_A(X_k^w) \ . \tag{3}$$

This technique of sampling according to the weight distribution function instead of uniformly has the advantage that the accuracy of the estimate, i.e., the confidence interval, is independent of the weight distribution[7].

In order to sample weight distributions that are defined as a mixture of several independent distributions w_i ($1 \leq i \leq m$) as proposed in the following section, we propose to distribute the number of samples among the different distributions in the following way: a weight distribution $w(\vec{z}) = \sum_{i=1}^{m} p_i \cdot w_i(\vec{z})$ with $\sum_{1 \leq i \leq m} p_i = 1$ is estimated by sampling each of the distributions w_i independently with $N \cdot p_i$ samples and summing up the resulting estimates.

4. INTEGRATING USER PREFERENCES

In this section, we present two different weight distribution functions to express user preferences. Both distributions are continuous probability densities that enable to draw samples according to the procedure presented above. The first distribution allows to attribute importance to one objective and the second to emphasize a preference point in the objective space. We complete the section by demonstrating how any number of the two distributions can be combined, e.g., to use more than one preference point.

4.1 Stressing the Extremes

One potential preference a user might have is to optimize preferably one objective, say f_s. The corresponding weight distribution should therefore increase for decreasing values of f_s. In terms of the rest of the objectives, the weight distribution should stay constant for changing values in order

---

[6]We refer to [13] for an extensive overview how random samples from those distributions can be generated.

[7]Denoting the estimator of $I_H^w(A)$ (Eq. 3) as $\widehat{I_H^w}(A, N) = \frac{1}{N}\sum_{k=1}^N \alpha_A(X_k^w)$, Hoeffding's inequality implies that with probability larger than $1 - \alpha$, $I_H^w(A)$ belongs to $[\widehat{I_H^w}(A,N) - t_{N,\alpha}, \widehat{I_H^w}(A,N) + t_{N,\alpha}]$ where $t_{N,\alpha} = \left(\frac{8}{N}\log(2/\alpha)\right)^{1/2}$ which is independent of w and the same confidence interval than for the non-weighted case. In other words, it is not more expensive to do a Monte Carlo integration of the weighted hypervolume than for the standard hypervolume indicator.

---

not to introduce additional preference.

In [23] the authors proposed to use an exponential function as the weight distribution. Here, the same distribution is represented by the probability density function whose marginal distribution for objective f_s is an exponential distribution with rate parameter λ and whose marginal distributions of the remaining objectives is a uniform distribution:

$$w(z_1, \cdots, z_k) = \begin{cases} \left(\prod_{i \neq s}(b_i^u - b_i^l)\right)^{-1} \lambda e^{-\lambda(z_s - b_s^l)} & \vec{z} \in B \\ 0 & \vec{z} \notin B \end{cases}$$

where $B = [b_1^l, b_1^u] \times \ldots \times [b_k^l, b_k^u]$ denotes the space with non-zero probability density.

Figure 1(a) shows the weight distribution for a biobjective problem when stressing f_1 with an exponential distribution in f_1 ($\lambda = 5$) together with a uniform distribution in the interval $[0, 0.95]$ in the second objective ($B = [b_s^l, b_s^u] \times [b_2^l, b_2^u] = [0, \infty] \times [0, 0.95]$).

The spread of the distribution is inversely proportional to the parameter λ. Hence, the smaller λ the steeper the weight distribution increases at the border of the objective space and the smaller the weight farther away (see Figure 2(a) for contour plots of the exponential weight distribution for distinct values of λ).
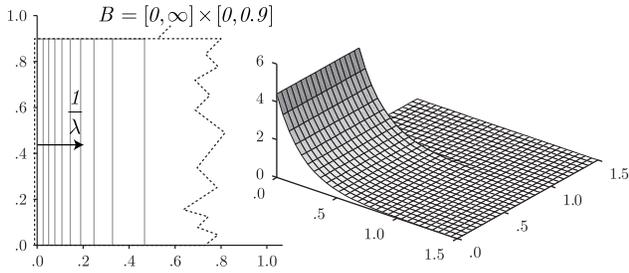
4.2 Preference Points

Another user preference is the *preference point* [20]. This point, as well as to a lesser extent the adjacent region, represents the most important part of the objective space for the user. Together with the location of the preference point, denoted by $\vec{\mu} = (\mu_1, \cdots, \mu_k)^T \in \mathbb{R}^k$, the user has to define a direction $\vec{t} = (t_1, \cdots, t_k)^T \in \mathbb{R}^k$. The solutions should preferably lie along this direction if the preference point cannot be reached or, on the contrary, even better solutions are found. The corresponding weight distribution function reflects this preference by having the largest values at the preference point and along $\vec{t}$ while decreasing fast perpendicular to $\vec{t}$. To this end, [23] proposes a bivariate ridge-like function that cannot be easily extended to an arbitrary number of objectives. Therefore, we propose to use the following multivariate Gaussian distribution here, which allows an efficient sampling according to Eq. 3 and which can be used for any number of objectives. Besides $\vec{\mu}$ and $\vec{t}$, let $\sigma_\varepsilon, \sigma_t \in \mathbb{R}$ denote standard deviations of the distribution. Then the following probability density function describes a multivariate normal distribution centered at $\vec{\mu}$

$$w(\vec{z}) = \frac{1}{(2\pi)^{k/2}|C|^{1/2}}e^{-\frac{1}{2}(\vec{z}-\vec{\mu})^T C^{-1}(\vec{z}-\vec{\mu})}$$

where the covariance matrix $C := \sigma_\varepsilon^2 \mathbf{I} + \sigma_t^2 \vec{t}\vec{t}^T/\|\vec{t}\|^2$ is non-singular with orthogonal eigenvectors $\vec{t}, \vec{t}_2, \cdots, \vec{t}_k$ where the vectors $\vec{t}_2, \ldots, \vec{t}_k$ can be taken from an arbitrary orthogonal basis of the hyperplane orthogonal to $\vec{t}$. The eigenvalues associated to $\vec{t}, \vec{t}_2, \cdots, \vec{t}_k$ are $\sigma_\varepsilon^2 + \sigma_t^2, \sigma_\varepsilon^2, \cdots, \sigma_\varepsilon^2$; $|C|$ denotes the determinant of C.

The equidensity contours of the distributions are ellipsoids whose principal axis are $\vec{t}, \vec{t}_2, \cdots, \vec{t}_k$, see Fig. 1(b). The lengths of the axes are given by the two standard deviations (i) σ_t for the axis spanned by $\vec{t}$ and (ii) σ_ε for the remaining $k - 1$ axes perpendicular to $\vec{t}$. The larger σ_t is chosen the farther the objective vectors can lie from the preference point in direction of $\pm\vec{t}$ while they are still affected by the

(a) Stressing the first objective
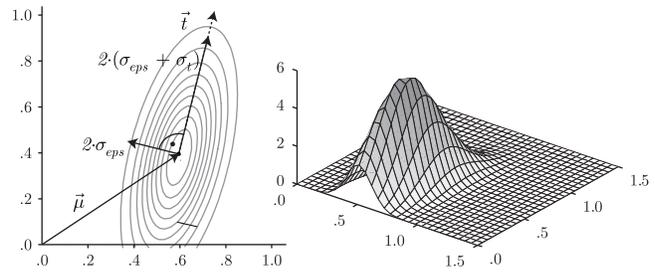
(b) Preference point at $\vec{\mu}$

Figure 1: Illustrates the exponential distribution that corresponds to stressing one objective (a) and the Gaussian distribution representing a preference point (b). The left parts of the two subplots indicate the notation used along with a contour plot at intervals of 10% of the maximum value observed (which occurs on the second axis and at $\vec{\mu}$ respectively). The right parts of the subplots show the actual value of the distribution as a third component.

weight distribution. At the same time, however, the number of samples near the Pareto front approximation decreases which reduces the accuracy of sampling.

The second variance, σ_ε, influences the extension of points close to the preference point. The smaller σ_ε, the less widespread the solutions are (see Figures 2(b) for contour plots of three different choices of σ_ε).

4.3 Combinations

A mixture of m weight distributions admitting the probability density functions $w_1(\vec{z}), \ldots, w_m(\vec{z})$ yields the distribution with density

$$w(\vec{z}) = p_1 w_1(\vec{z}) + \ldots + p_m w_m(\vec{z})$$

where the p_i are positive real numbers with $p_1 + \ldots + p_m = 1$. Though it is not possible to translate any user preference directly to a weight distribution function as in [1], a wide range of different user preferences can be represented by combining weight distributions. These are—in contrast to the weight distributions in [1]—also applicable to problems with more than two objectives. In the next section we will examine mixtures of the two distributions presented above.

5. EXPERIMENTAL VALIDATION

In order to test the approach of articulating user preferences presented in Sec. 4, the sampling strategy is employed by an existing hypervolume-based algorithm called "Hypervolume Estimation Algorithm for Multiobjective Optimization (HypE)" [3]. The application to different multiobjective test problems investigates three important aspects of the approach.

First, we investigate visually the influence of the different parameters on the distribution of the resulting Pareto front approximations for both approaches preferring preference points and extremes. In particular, we investigate for a preference point its location $\vec{\mu}$, the direction $\vec{t}$ and the influence of the standard deviations σ_ε and σ_t. When stressing extremes, we show the effects of changing the parameter λ.

Secondly, the weighted hypervolume approach is visually compared to existing reference algorithms that do not optimize any user preference explicitly for problems with more than two objectives. This demonstrates that our approach

is—in contrast to [23]—also applicable to problems involving more than two objectives.

Finally, a short statistical comparison on problems with up to 25 objectives is carried out to investigate whether the generated Pareto front approximations obtained by HypE, as a matter of fact, better fulfill the underlying user preference than Pareto fronts resulting from reference algorithms.

5.1 Experimental Setup

To maximize the weighted hypervolume, the evolutionary algorithm HypE [3] is used. While mating selection is performed uniformly, the next population is chosen in environmental selection as follows: first, the offspring and parent population are merged and allocated to fronts by non-dominated sorting. Thereafter, the worst front is removed entirely as long as the resulting population size does not fall below the desired population size. If excluding entire fronts is no longer possible and there are still too many individuals in the population, the algorithm then starts removing individuals one-by-one from the worst front. Instead of other hypervolume-based selection schemes such as the ones in the SMS-EMOA [4] or the MO-CMA-ES [18], HypE not only considers the hypervolume solely dominated by a solution as its fitness but estimates the expected loss if more than one solution is deleted from the population by means of Monte Carlo sampling. To this end, $10,000$ samples are generated according to the probability density functions presented in Sec. 4 using the corresponding built-in functions of MATLAB® version 2008a. These samples are then used to calculate a fitness value for each individual of the worst front. Iteratively, the individual with the smallest value is deleted until the desired population size is reached where the fitness values are recalculated after each removal using the same samples as in the previous step. A detailed description of the fitness calculation of HypE can be found in [3].

The evolutionary multiobjective algorithms NSGA-II [10] and IBEA [25] serve as reference algorithms. For the latter, the ε-indicator has been used since preliminary experiments showed this variant to be superior to the one using the hypervolume indicator. The parameters of IBEA are set as $\kappa = 0.05$ and $\rho = 1.1$. All algorithms are run for 100 generations. New individuals are generated by the SBX crossover
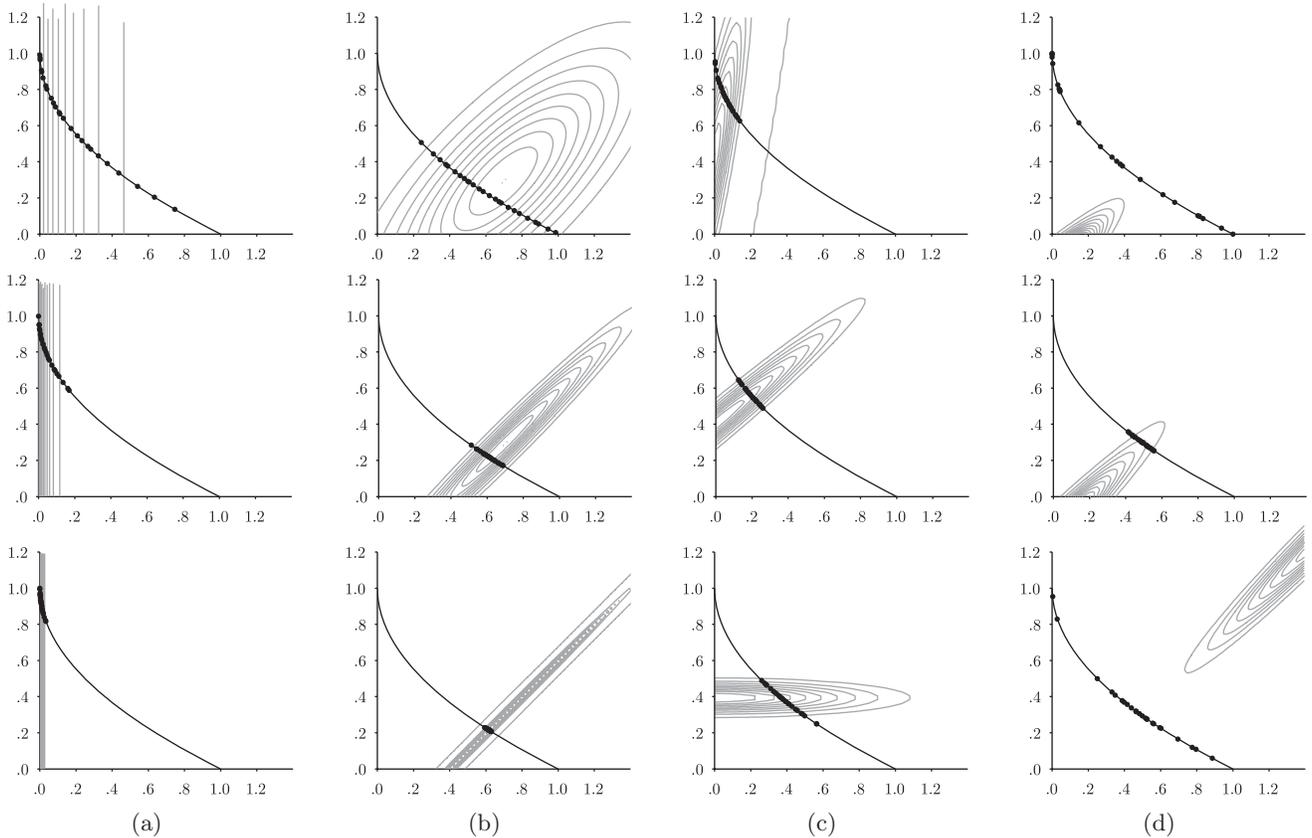
Figure 2: Shows the Pareto front approximations (dots) found by HypE using different weight distribution functions, shown as contour lines at intervals of 10% of the maximum weight value. For each column of three figures one parameter of the sampled distribution was modified, i.e., (a) the rate parameter of the exponential distribution λ, (b) the spread σ_ε, (c) the direction $\vec{t}$ and (d) the location $\vec{\mu}$ of the preference point (see text for details and the values used). The test problem is ZDT1 where the Pareto front is shown as a solid line.

operator with $\eta_c = 15$ and the variable-wise polynomial mutation operator with $\eta_m = 20$ [9]. The crossover and mutation probabilities are set to 1 and $1/20$ respectively.

For the biobjective test problems both the population size and the number of offspring are set to 25 while for more objectives these numbers are doubled. For the biobjective investigations, the test problems ZDT1 (convex Pareto front), ZDT3 (discontinuous Pareto front) [24] and DTLZ2 [12] (concave Pareto front) are utilized with 20 decision variables. For higher dimensions only DTLZ2 is employed.

5.2 Visual Inspection of Parameter Choices

In this section, the influence of different parameters on the weight distribution functions and the resulting Pareto front approximations are investigated. Unless noted otherwise, we use $\sigma_t = 0.5$, $\sigma_\varepsilon = 0.05$ and $\vec{t} = (1,1)$ when stressing a preference point and $B = [0,\infty] \times [0,3]$ when stressing the first objective ($f_s = f_1$). The weight distributions are indicated by contour lines at the intervals of 10% of the maximum value that arises. The contour lines do not reflect the actual weight but only the relative distribution thereof. We tested multiple runs for each testcase that led to similar results such that we mostly display only one run to illustrate the influence of the weight on the distribution of points.

5.2.1 Spread of the Distribution

Both proposed weight distribution functions have parameters that cause the weight to be more or less spread. Figure 2(a) shows the weight distribution and the resulting Pareto front approximation using the exponential distribution proposed in Sec. 4.1 for $\lambda = 100$ (top), $\lambda = 20$ (center) and $\lambda = 5$ (bottom). Figure 2(b) shows the distribution of points for a preference point located at $\vec{\mu} = (0.7, 0.3)$ where σ_ε is set to 0.2 (top), 0.05 (middle) and 0.01 (bottom).

5.2.2 Direction of the Distribution

By $\vec{t}$, the user can define the desired trade-off between the objectives for the case that either the preference point cannot be reached or if solutions dominating the preference point are obtained. In Fig. 2(c) the preference point is positioned at $\vec{\mu} = (0, 0.4)$ which lies below the Pareto front and can therefore not be reached. In this case, the direction $\vec{t}$ predetermines where the resulting points lie. In the topmost example, a choice of $\vec{t} = (\cos(80°), \sin(80°))$ reflects a higher preference of the first objective at the expense of the second. On the other hand, the bottom figure is obtained for $\vec{t} = (1, 0)$, i.e., values of 0.4 are preferred for the second objective and only increases of the first objective are desired. The figure in the middle presents a compromise where the focus lies close to the diagonal, $\vec{t} = (\cos(40°), \sin(40°))$.

5.2.3 Location of the Preference Point

Since the preference point can be placed both too optimistically (as in the previous section) or too pessimistically, the parameter σ_t allows to tune how far away the individuals can be from the preference point and still be influenced by it. For a fixed σ_t however, the location of the preference point has a high impact on the resulting distribution of solutions, see Fig. 2(d). If none to only a few samples are dominated by the individuals (top, $\vec{\mu} = (-1.2, -1.4)$), no pressure towards the preference point is active—in fact only non-dominated sorting operates. In this case, the preference point should be combined with a uniform distribution, e.g., as in Fig. 3(a) 90% of the samples are used for the preference point and 10% to sample uniformly in the objective space within $[0, 3] \times [0, 3]$. This causes the solutions to be distributed according to the unweighted hypervolume indicator as long as the preference point has no influence.

As soon as a couple of samples are dominated, the corresponding individuals are promoted which leads to an accumulation in that area (middle, $\vec{\mu} = (-0.3, -0.5)$). If the preference point is chosen very pessimistically (bottom, $\vec{\mu} = (1.5, 1.3)$) individuals are able to dominate all or most of the samples even if they are not located where the direction $\vec{t}$ dictates. This leads to a much ampler arrangement of solutions than expected considering the chosen σ_ε.

5.2.4 Combinations of Distributions

As demonstrated in Sec. 4, any number of weight distribution functions can be combined as a weighted sum, even assigning them different weights or focus. For example, the user might define different preference points he or she is interested in as depicted in Fig. 3(b): three preference points are positioned at $\vec{\mu} = (0.2, 0.8)$, at $\vec{\mu} = (0.5, 0.5)$ and at $\vec{\mu} = (0.8, 0.2)$. The one in the middle is declared to be the most important one by assigning the largest weight $p_2 = 0.5$, the preference points to the left and right use $p_1 = 0.2$ and $p_3 = 0.3$ respectively. As expected, in this case the points are partitioned into disjoint regions around the three preference points. 10 individuals cluster around the center where the most samples emerge, 7 are associated with the preference point on the left and 8 with the one on the right.

To promote individuals at the border of the objective space, two exponential weight distributions can be added up as in Fig. 3(c) where $\lambda = 10$ with $p_1 = 0.3$ for the first objective and $p_2 = 0.7$ for the second.

5.2.5 Comparison Between Different Problems

In addition to ZDT1, the tests of the previous sections were also carried out for other test problems, namely ZDT3 which has a discontinuous Pareto front shape, DTLZ2 and ZDT6 (both non-convex). These three test problems are much harder to optimize and neither HypE nor the reference algorithms used were able to find Pareto optimal solutions. The points are nevertheless clustered at regions with the largest weight, see Figures 3(d), 3(e) and 3(f) where one preference point with $\vec{\mu} = (0.7, 0.3)$ and $\sigma_\varepsilon = 0.1$ is used.

5.3 High-Dimensional Spaces

For illustrative reasons in the previous section we applied the sampling procedure to biobjective problems. The advantage of the method, however, is that an arbitrary number of objectives can be tackled. Figure 4 shows the Pareto front and the solutions found by different algorithms on the
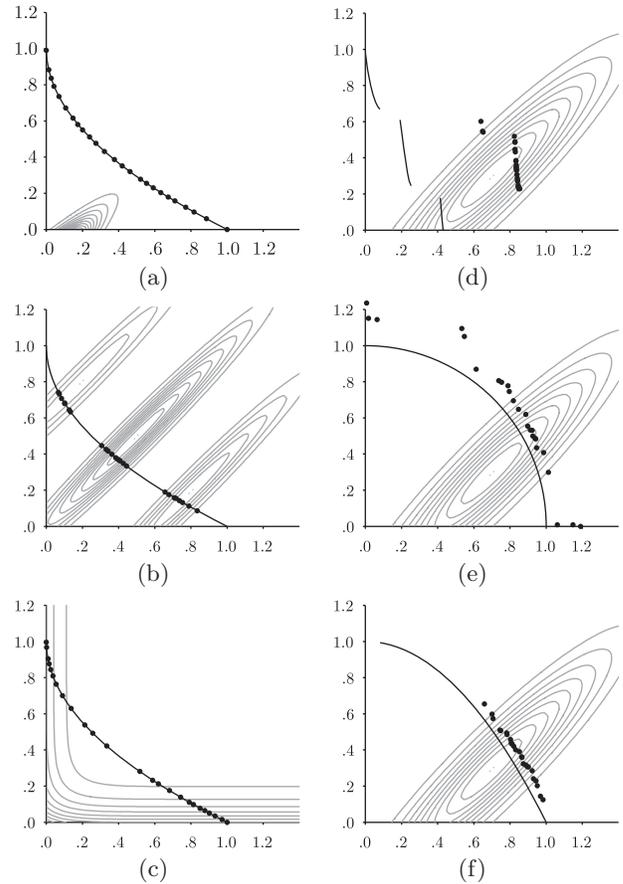

Figure 3: These figures use the same visual elements as Fig. 2 which explains them in its caption. In (a) the same preference point is used as in the upper plot of Fig. 2(d) but spending 10% of the samples on a uniform distribution. Figure (b) shows the combination of three preference points, and Fig. (c) stressing both the first and second objective. The figures on the right show the distribution of the objective vectors when applying the same preference point to different test problems, i.e., (d) ZDT3 (only the positive part shown) (e) DTLZ2 and (f) ZDT6.

DTLZ2 problem with 3 objectives. While NSGA-II and IBEA do not optimize any user defined preference, HypE uses two preference points at $\vec{\mu}_1 = (0.8, 0.2, 0.6)$ ($p_1 = 0.2$) and $\vec{\mu}_2 = (0.2, 0.9, 0.5)$ ($p_2 = 0.8$) with $\sigma_\varepsilon = 0.1$ (shown as ellipsoids). This leads to a cluster of points at each preference point.

The Pareto front approximation on DTLZ2 with 7 objectives is depicted in Fig. 5 by means of parallel coordinates plots for IBEA and HypE with $\sigma_\varepsilon = 0.05$. The plot for NSGA-II is omitted due to space limitations; it can be noted that it looks similar to the one of IBEA except that NSGA-II does not come as close to the Pareto front as IBEA, i.e., the objective values are spread between 0 and 4.5. Both IBEA and NSGA-II generate solutions at the boundary of the objective space while only the former finds solutions near the Pareto front. To get solutions near the center of the Pareto front, HypE is applied with a preference point at
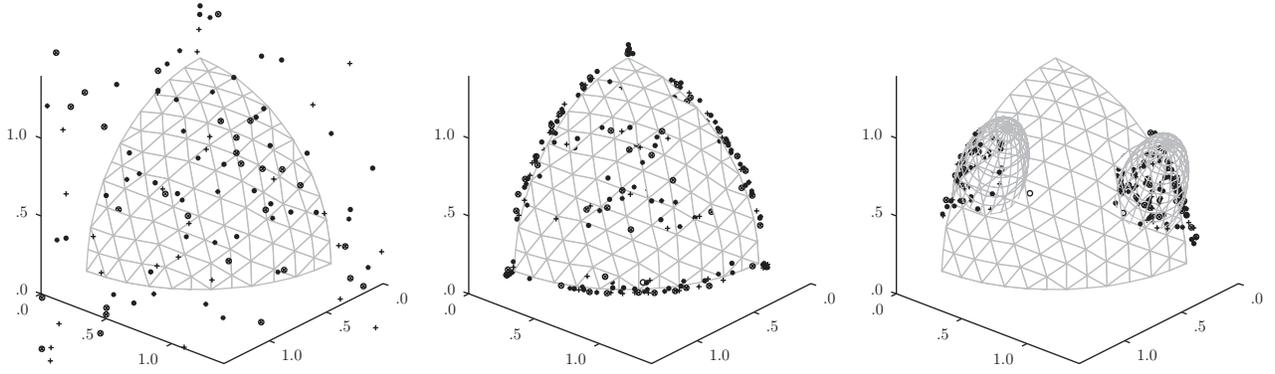
Figure 4: Pareto front approximations of five runs (depicted by different symbols) for the 3-objective DTLZ2 test problem using NSGA-II (left), IBEA (middle), and HypE with two preference points displayed as ellipsoids (right).
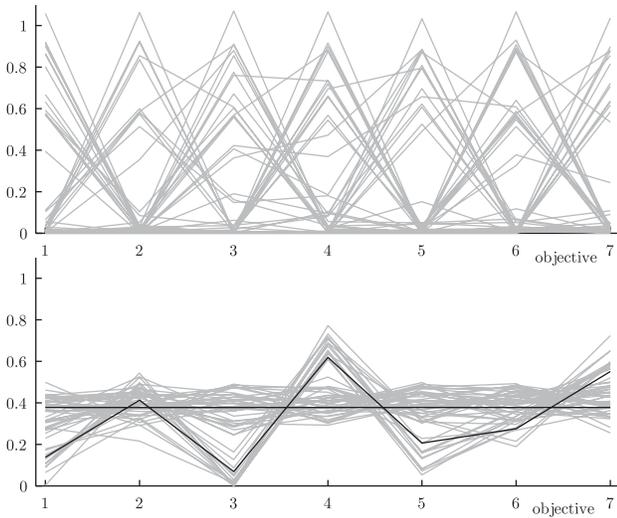

Figure 5: Distribution of points, plotted in parallel coordinates, for the 7-objective DTLZ2 test problem for IBEA (top) and HypE with two preference points (solid black lines) and $\sigma_\varepsilon = 0.05$ (bottom).

Table 1: Normalized hypervolume values using a weight distribution w_p that reflects two preference points. The mean value and twice the standard deviation of 50 runs are listed. HypE$_p$ uses the weight distribution function w_p while HypE$_{uni}$ uses a uniform weight distribution.

	NSGA-II	IBEA	HypE$_{uni}$	HypE$_p$
2d	0.443±0.321	0.658±0.156	0.487±0.202	0.910±0.065
3d	0.480±0.353	0.956±0.015	0.917±0.035	0.996±0.005
5d	0.021±0.071	0.807±0.033	0.959±0.121	0.995±0.004
10d	0.215±0.311	0.688±0.121	0.964±0.049	0.999±0.002
25d	0.171±0.280	0.613±0.322	0.826±0.152	0.999±0.003

The Pareto front approximations for 50 independent runs of both versions of HypE as well as of NSGA-II and IBEA are then compared in terms of the weighted hypervolume indicator with the weight distribution function w_p, see Table 1 where the hypervolume values are normalized to the minimum and maximum value observed for the respective number of objectives. In each case, HypE$_p$ outperforms statistically significant the other algorithms in terms of the hypervolume indicator with w_p—assessed by Kruskal–Wallis and the Conover–Inman post hoc tests with a significance level of 5% according to [8]. This indicates that applying the weighted integration technique during search will generate Pareto front approximations that score better on the corresponding hypervolume indicator than using general purpose algorithms with no user defined search direction.

6. CONCLUSIONS

This paper has described two procedures to approximate the weighted hypervolume integration developed in [23] by means of Monte Carlo sampling. This offers the possibility to incorporate user preferences in a Pareto-compliant way, where problems with an arbitrary number of objectives can be considered. Two types of user preferences have been expressed by probability density functions that ease the fast generation of samples—one stressing certain objectives and the second emphasizing a preference point. Additionally, any combination of the two is possible.

The combination of the sampling approach with HypE,

$0.3780 \cdot (1, \ldots, 1)$. A second preference point is set at a random location near the Pareto front[8]. The spread σ_ε is set for both preference points to 0.05 and the probabilities of the mixture are set as 0.8 and 0.2 respectively leading to a population of solutions grouped around the two preference points (Fig. 5).

To investigate further whether HypE actually optimizes the weight distribution used during search, two versions of HypE are run in another experiment: the first version (denoted by HypE$_{uni}$) uses a uniform weight distribution w_{uni} with reference point $r = (3, \ldots, 3)$, the second (HypE$_p$) optimizes a weight distribution w_p consisting of a combination of two preference points at $\mu_1 = (0.9, 0.1, 0.9, 0.1, \ldots)$ with $p_1 = 0.3$, $\sigma_\varepsilon = 0.05$ and at $\mu_2 = (0.1, 0.9, 0.1, 0.9, \ldots)$ with $p_2 = 0.7$, $\sigma_\varepsilon = 0.2$. Both versions together with IBEA and NSGA-II then optimized the DTLZ2 test problem with different numbers of objectives ranging from 2 to 25.

---

[8] $\vec{\mu} = (0.1377, 0.4131, 0.0688, 0.6196, 0.2065, 0.2754, 0.5507)$

an existing hypervolume based algorithm, has been applied to various test problems. It has turned out by both visual inspection and statistical tests that the generated Pareto front approximations reflect the underlying weight distribution better than methods with no user defined preference.

7. REFERENCES

[1] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Investigating and Exploiting the Bias of the Weighted Hypervolume to Articulate User Preferences. In G. Raidl et al., editors, *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] J. Branke, T. Kaußler, and H. Schmeck. Guidance in Evolutionary Multi-Objective Optimization. *Advances in Engineering Software*, 32:499–507, 2001.

[6] K. Bringmann and T. Friedrich. Approximating the Volume of Unions and Intersections of High-Dimensional Geometric Objects. In S. H. Hong, H. Nagamochi, and T. Fukunaga, editors, *International Symposium on Algorithms and Computation (ISAAC 2008)*, volume 5369 of *LNCS*, pages 436–447, Berlin, Germany, 2008. Springer.

[7] R. E. Caflisch. Monte Carlo and Quasi-Monte Carlo Methods. *Acta Numerica*, 7:1–49, 1998.

[8] W. J. Conover. *Practical Nonparametric Statistics*. John Wiley, 3 edition, 1999.

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

[10] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. In M. Schoenauer et al., editors, *Conference on Parallel Problem Solving from Nature (PPSN VI)*, volume 1917 of *LNCS*, pages 849–858. Springer, 2000.

[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 A. Abraham, R. Jain, and R. Goldberg, editors, *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, chapter 6, pages 105–145. Springer, 2005.

[13] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.

[14] R. Everson, J. Fieldsend, and S. Singh. Full Elite-Sets for Multiobjective Optimisation. In I. Parmee, editor, *Conference on adaptive computing in design and manufacture (ADCM 2002)*, pages 343–354, London, UK, 2002. Springer.

[15] M. Fleischer. The measure of Pareto optima. Applications to multi-objective metaheuristics. In C. M. Fonseca et al., editors, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *LNCS*, pages 519–533, Faro, Portugal, 2003. Springer.

[16] W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[17] S. Huband, P. Hingston, L. White, and L. Barone. An Evolution Strategy with Probabilistic Mutation for Multi-Objective Optimisation. In *Congress on Evolutionary Computation (CEC 2003)*, volume 3, pages 2284–2291, Canberra, Australia, 2003. IEEE Press.

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

[19] M. Laumanns, G. Rudolph, and H.-P. Schwefel. Approximating the Pareto Set: Concepts, Diversity Issues, and Performance Assessment. Technical Report CI-7299, University of Dortmund, 1999.

[20] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, MA, USA, 1999.

[21] M. Nicolini. A Two-Level Evolutionary Approach to Multi-criterion Optimization of Water Supply Systems. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, volume 3410 of *LNCS*, pages 736–751. Springer, 2005.

[22] 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, July 2006.

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

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

[25] E. Zitzler and S. Künzli. Indicator-Based Selection in Multiobjective Search. In X. Yao et al., editors, *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, volume 3242 of *LNCS*, pages 832–842. Springer, 2004.

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

[27] E. Zitzler, L. Thiele, and J. Bader. SPAM: Set Preference Algorithm for Multiobjective Optimization. In G. Rudolph et al., editors, *Conference on Parallel Problem Solving From Nature (PPSN X)*, volume 5199 of *LNCS*, pages 847–858. Springer, 2008.