On the Properties of the R2 Indicator

Dimo Brockhoff, Tobias Wagner, Heike Trautmann

On the Properties of the $R2$ Indicator

Dimo Brockhoff
INRIA Lille Nord-Europe
Dolphin Team
Villeneuve d'Ascq, France
dimo.brockhoff@inria.fr

Tobias Wagner
Institute of Machining Technology
TU Dortmund
Dortmund, Germany
wagner@isf.de

Heike Trautmann
Statistics Department
TU Dortmund, Germany
trautmann@statistik.uni-dortmund.de

ABSTRACT

In multiobjective optimization, set-based performance indicators are commonly used to assess the quality of a Pareto front approximation. Based on the scalarization obtained by these indicators, a performance comparison of multiobjective optimization algorithms becomes possible. The $R2$ and the Hypervolume (HV) indicator represent two recommended approaches which have shown a correlated behavior in recent empirical studies. Whereas the HV indicator has been comprehensively analyzed in the last years, almost no studies on the $R2$ indicator exist. In this paper, we thus perform a comprehensive investigation of the properties of the $R2$ indicator in a theoretical and empirical way. The influence of the number and distribution of the weight vectors on the optimal distribution of μ solutions is analyzed. Based on a comparative analysis, specific characteristics and differences of the $R2$ and HV indicator are presented.

Categories and Subject Descriptors

G.1.6 [**Mathematics of Computing**]: Numerical Analysis—*Optimization*

General Terms

Algorithms

Keywords

Performance assessment, Hypervolume indicator, $R2$ indicator, Multiobjective optimization

1. INTRODUCTION

Multiobjective optimization comprises the optimization of a vector of m objective functions $\vec{f} = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$. We assume minimization tasks in the following, as maximization problems can easily be transferred to those by negation. A solution $\vec{x}$ in the decision space $\mathbb{S}$ is said to dominate a solution $\vec{y} \in \mathbb{S}$ ($\vec{x} \prec \vec{y}$), iff $\vec{f}(\vec{x})$ is better than or equal to $\vec{f}(\vec{y})$ in all objectives ($\forall i \in \{1, \ldots, m\} : f_i(\vec{x}) \leq f_i(\vec{y})$) and better than $\vec{f}(\vec{y})$ in at least one ($\exists j \in \{1, \ldots, m\} : f_j(\vec{x}) < f_j(\vec{y})$). Only the first condition is necessary for $\vec{x}$ to weakly dominate $\vec{y}$ ($\vec{x} \preceq \vec{y}$). The aim of multiobjective optimization is to detect the Pareto optimal set $P = \{\vec{x} \in \mathbb{S} \mid \nexists\, \vec{x}' \in \mathbb{S} : \vec{x}' \prec \vec{x}\}$. The Pareto front is the corresponding mapping $\vec{f}(P)$ to the objective space $\mathbb{R}^m$.

Evolutionary multiobjective algorithms are a specific class of solvers for multiobjective problems. They basically aim at approximating the true Pareto front of the problem at hand by minimizing the distance to the true Pareto front (convergence) and simultaneously covering all its parts (spread) in a well-defined way (distribution) [12]. Several performance indicators were introduced [14, 16] to assess one or all of these quality aspects of Pareto front approximations.

The Hypervolume (HV) [15] and the $R2$ indicator [7] are two recommended approaches which simultaneously evaluate all desired aspects of a Pareto front approximation. The HV is so far the only known indicator which fulfills the property of strict monotonicity, i.e., the indicator value $I(A)$ of a set A that dominates the set B has to be higher than the indicator value $I(B)$ for the set B, assuming that the indicator is to be maximized. In contrast, the $R2$ indicator is only weakly monotonic, i.e., $I(A) \geq I(B)$ in case A weakly dominates B. Nevertheless, the $R2$ indicator is often preferred over the HV for two reasons. On the one hand, the runtime of the HV is exponential with respect to the number of objectives m [4]. On the other hand, the distributions obtained using the HV are biased towards the knee regions of the Pareto front [1]. The $R2$ indicator is assumed to result in a more uniform distribution.

As monotonicity of both, the HV and the $R2$ indicator, is guaranteed [16], a similar behavior of the indicators can be expected. However, as the preferred distributions seem to be different, the degree of similarity might vary in different settings. In [10, 11], optimal parameterizations of the multiobjective optimizer SMS-EMOA [5] were investigated. Correlation structures between the evaluations of Pareto fronts based on different quality indicators were analyzed in order to select the most suitable indicator for the used Sequential Parameter Optimization (SPO) approach. It was shown that the HV and $R2$ indicator are highly correlated on a huge number of randomly generated populations in different dimensions. Pearson's correlation coefficient [9] takes a statistically significant value of 0.76. However, the indicators are different enough that they led to statistically distinguishable results when used individually in the SPO framework.

In this paper, we aim at a more detailed understanding of the properties of the $R2$ indicator. We perform an analysis about *how* it differs from the HV indicator for which several theoretical properties are already known [1, 2]. Furthermore, the finite sets of μ solutions maximizing the $R2$ indicator among all feasible sets of μ solutions—the so-called *optimal μ-distributions of the $R2$ indicator*—are generated for standard test functions with different Pareto front shapes. The effect of the distribution of the weight vectors is analyzed by alternatively investigating a uniform coverage of the angle space in addition to the standard uniform coverage of the weight space.

The paper is organized as follows. Details of the indicators are given in Sec. 2. Properties of the $R2$ indicator are discussed in Sec. 3, followed by an analysis and comparison of the optimal μ-distributions for both, the $R2$ and the HV indicator, in Sec. 4. Finally, conclusions are drawn and an outlook is given in Sec. 5.

2. *R2* AND HV INDICATORS

The R indicator family [7] is based on utility functions which map a vector $\vec{y} \in \mathbb{R}^m$ to a scalar utility value $u \in \mathbb{R}$ for assessing the relative quality of two Pareto front approximation sets.

DEFINITION 1. *For a set U of general utility functions, a probability distribution p on U, and a reference set R, the $R2$ indicator of a solution set A is defined as the expected utility [7]*

$$\begin{aligned} R2(R,A,U,p) &= \int_{u\in U} \max_{r\in R}\{u(r)\}p(u)du \\ &\quad - \int_{u\in U} \max_{a\in A}\{u(a)\}p(u)du \ . \end{aligned}$$

DEFINITION 2. *For a discrete and finite set U and a uniform distribution p over U, the original $R2$ indicator can be written as [14]*

$$R2(R,A,U) = \frac{1}{|U|}\sum_{u\in U}\left(\max_{r\in R}\{u(r)\} - \max_{a\in A}\{u(a)\}\right).$$

Since the first summand is a constant if we assume R to be constant, we delete the first summand and call the resulting unary indicator also $R2$ for simplicity.

DEFINITION 3. *For a constant reference set, the R2 indicator can be defined as a unary indicator*

$$R2(A,U) = -\frac{1}{|U|}\sum_{u\in U}\max_{a\in A}\{u(a)\} \ .$$

Note that we assume R to be constant throughout the paper and will only refer to Definition 3 when we use the term $R2$ indicator.

Different choices of the required utility function exist, e.g., based on weighted sum and weighted Tchebycheff functions or a combination of both. As suggested in [7], we use the standard weighted Tchebycheff function $u(z) = u_\lambda(\vec{z}) = -\max_{j\in\{1\dots m\}} \lambda_j|z_j^* - z_j|$ within the $R2$ indicator in the following where $\lambda = (\lambda_1,\dots,\lambda_m) \in \Lambda$ is a given weight vector and z^* is a utopian point[1].

[1]An objective vector that is not dominated by any feasible search point.

DEFINITION 4. *The $R2$ indicator of a solution set A for a given set of weights Λ and a utopian point z^* is defined as*

$$\begin{aligned} R2(A,\Lambda,z^*) &= \frac{1}{|\Lambda|} - \sum_{\lambda\in\Lambda}\max_{a\in A}\{-\max_{j\in\{1,\dots,m\}}\{\lambda_j|z_j^* - a_j|\}\} \\ &= \frac{1}{|\Lambda|}\sum_{\lambda\in\Lambda}\min_{a\in A}\{\max_{j\in\{1,\dots,m\}}\{\lambda_j|z_j^* - a_j|\}\} \ . \end{aligned}$$

Usually, the weight vectors are chosen uniformly distributed over the weight space [14], for example for $m = 2$ objectives,

$$\Lambda_{w,k} = \left(0,1;\frac{1}{k-1},1-\frac{1}{k-1};\frac{2}{k-1},1-\frac{2}{k-1};\dots;1,0\right).$$

denotes k uniformly distributed weights in the space $[0,1]^2$. Alternatively, a uniform distribution over the angle space

$$\Phi_k = \left(0;\frac{\pi}{(2(k-1)};\frac{2\pi}{2(k-1)};\dots;\frac{\pi}{2}\right)$$

can be considered for $m = 2$ objectives. In this case, rays from the utopian point with different slopes $\tan(\varphi)$ for $\varphi \in \Phi_k$ are constructed and then mapped to weight vectors

$$\Lambda_{\varphi,k} = \left\{ \left(\frac{1}{\tan(\varphi)+1},\frac{\tan(\varphi)}{\tan(\varphi)+1}\right) \,\middle|\, \varphi\in\Phi_k \right\}$$

by normalization.[2] A weight vector corresponding to a ray which directly crosses a point $a \in A$ is defined as λ^a.

In addition to an overall quality assessment of an approximation set A, the contribution of an individual solution to the $R2$ value can be of interest.

DEFINITION 5. *The contribution of a solution $a \in A$ to the $R2$ indicator is defined as*

$$C_{R2}(a,A,\Lambda,z^*) = R2(A,\Lambda,z^*) - R2(A\setminus\{a\},\Lambda,z^*)$$

which can be written as

$$\begin{aligned} C_{R2}(a,A,\Lambda,z^*) &= \frac{1}{|\Lambda|}\sum_{\lambda\in\Lambda}\left(\min_{b\in A}\left\{\max_{j\in\{1\dots,m\}}\left\{\lambda_j|z_j^* - b_j|\right\}\right\}\right. \\ &\quad \left. - \min_{b\in A\setminus\{a\}}\left\{\max_{j\in\{1\dots,m\}}\left\{\lambda_j|z_j^* - b_j|\right\}\right\}\right). \end{aligned}$$

In the remainder, we only consider the bi-objective case, i.e., $m = 2$. We furthermore assume that the Pareto front is given by a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that every point $(x, y = f(x))$ on the front can be uniquely described by a real value $x \in \mathbb{R}$. Finding the optimal μ-distribution then reduces to a μ-dimensional optimization problem

$$(x_1^\mu,\dots,x_\mu^\mu) = \operatorname*{argmin}_{(x_1,\dots,x_\mu)\in\mathbb{R}^\mu} R2(\{(x_i,f(x_i)) \,|\, 1\le i\le\mu\},\Lambda,z^*)$$

where we refer to the resulting set of x-values $(x_1^\mu,\dots,x_\mu^\mu)$ as optimal μ-distribution [1].

The HV indicator [15, 14] describes the hypervolume of a bounded space dominated by an approximation set.

DEFINITION 6. *Based on an anti-optimal reference point r, the HV indicator of an approximation set A is defined as*

$$HV(A,r) = \mathcal{L}\Big(\bigcup_{a\in A}\{a' | a \prec a' \prec r\}\Big)$$

with $\mathcal{L}(.)$ denoting the Lebesgue measure of a set.

[2]The normalized weight vector to the slope (x, y) is given by $(\frac{1/y}{1/x+1/y}, \frac{1/x}{1/x+1/y})$ as long as $x, y \neq 0$, and by $(0, 1)$ if $x = 0$ or $(1, 0)$ if $y = 0$. In our case, $x = 1$ and $y = \tan(\varphi)$.

Analogously to Definition 5, the HV-contribution of an individual solution reflects the influence of a single point on the quality of the approximation set.

DEFINITION 7. *The HV contribution of a solution $a \in A$ is defined as*

$$C_{HV}(a, A, r) = HV(A, r) - HV(A \setminus \{a\}, r).$$

As mentioned in Sec. 1, the HV indicator is the only known performance indicator which is strictly compliant with the dominance relation. Implementations and a comprehensive overview of the state-of-the art research w.r.t. the HV indicator are provided on [3].

3. PROPERTIES OF THE $R2$ INDICATOR

In order to interpret the values of a quality indicator during the performance assessment of multiobjective optimizers, it is crucial to understand the inherent preferences expressed by the choice of this indicator. To be more precise, it is a fundamental information to know which solution sets of size μ achieve the maximum indicator values among all possible sets of a given size μ [1]. Only with this knowledge, it is then possible to interpret the resulting indicator values of different multiobjective optimizer outcomes also absolutely instead of only relatively as the *achieved* indicator values can be compared with the *maximum achievable* values. For the optimum sets of size μ, the term *optimal μ-distribution* was introduced and such sets have already been characterized theoretically for the standard, as well as for the weighted, HV indicator [1, 2]. Before we investigate the concrete optimal μ-distributions on specific test functions also for the $R2$ indicator in Sec. 4, we first prove some general theoretical statements about the $R2$ indicator.

3.1 Locality of the $R2$ Indicator Contribution

At first, we prove that for bi-objective problems the optimal placement of a point according to the $R2$ indicator only depends on its two neighbors and only on a subset of the weight vectors. Note that also for the HV indicator such a locality property has been proven [1]. Before formalizing the general locality property of the $R2$ indicator, let us prove two preliminary statements.

LEMMA 1. *Given a specific weight vector $\lambda \in \Lambda$ and a specific solution $a \in A$ where A only contains nondominated solutions. Then all points $b \in A$ to the left of a (to the right of a) are resulting in a worse Tchebycheff utility than a if $D_1 = \lambda_1|z_1^* - a_1| < D_2 = \lambda_2|z_2^* - a_2|$ (if $D_1 > D_2$).*

PROOF. All nondominated points b that are to the left of a have an f_1 value that is smaller than a ($b_1 < a_1$) and their f_2 value is larger than the one of a ($b_2 > a_2$). These points lie in the filled box in the left plot of Fig. 1—otherwise, b would not be nondominated with respect to a and lie to the left. In case that $D_1 = \lambda_1|z_1^* - a_1| < D_2 = \lambda_2|z_2^* - a_2|$, this means that for all other points b

$$\lambda_1|z_1^* - b_1| < \lambda_1|z_1^* - a_1| < \lambda_2|z_2^* - a_2| < \lambda_2|z_2^* - b_2|$$

and thus for the Tchebycheff utility function, it holds that $\max\{\lambda_1|z_1^* - b_1|, \lambda_2|z_2^* - b_2|\} = \lambda_2|z_2^* - b_2| > \lambda_2|z_2^* - a_2| = \max\{\lambda_1|z_1^* - a_1|, \lambda_2|z_2^* - a_2|\}$, i.e., a is better since we minimize the maximum weighted component. The proof of the other case $D_1 > D_2$ follows a symmetric argumentation and is visualized in the right plot of Fig. 1. □

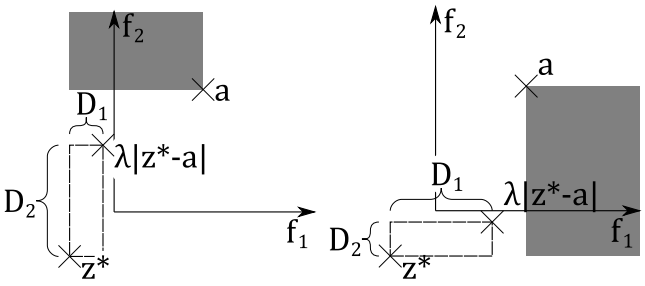

Figure 1: Sketch of idea for the proof of Lemma 1.

As a special case, we can investigate the specific weight vector λ^a that corresponds to a ray starting in the utopian point z^* and going through a (cf. Sec. 2).

LEMMA 2. *For each solution $a \in A$, there exists a specific weight $\lambda^a = (\lambda_1^a, \lambda_2^a)$ with $\lambda_1^a = \frac{|a_2 - z_2^*|}{|a_1 - z_1^*| + |a_2 - z_2^*|}$ and $\lambda_2^a = 1 - \lambda_1^a$ such that all other solutions $b \in A$ which do not dominate the point a have a worse weighted Tchebycheff value $\max\{\lambda_1^a|b_1 - z_1^*|, \lambda_2^a|b_2 - z_2^*|\}$ compared to a.*

PROOF. The proof follows the same ideas than the previous one. For $\lambda_1^a = \frac{|a_2 - z_2^*|}{|a_1 - z_1^*| + |a_2 - z_2^*|}$ and $\lambda_2^a = 1 - \lambda_1^a = \frac{|a_1 - z_1^*|}{|a_1 - z_1^*| + |a_2 - z_2^*|}$, it follows that $D_1 = \lambda_1^a|a_1 - z_1^*| = \lambda_2^a|a_2 - z_2^*| = D_2$. But then, for all solutions a' that do not dominate a and are left of a, $a_1' < a_1$ and $a_2' > a_2$ holds, such that $\lambda_1^a|a_1' - z_1^*| < \lambda_1^a|a_1 - z_1^*| = \lambda_2^a|a_2 - z_2^*| < \lambda_2^a|a_2' - z_1^*|$ which is equivalent to the fact that also the max in the weighted Tchebycheff for a' is larger and thus worse than the weighted Tchebycheff function for a. With a symmetric argument, we can prove that also all solutions a' that do not dominate a and are to the right of it are worse. □

As a result of the above lemmas, it follows that the optimal placement of a point according to the $R2$ indicator, given that all other points are already placed optimally, only depends on its neighbor(s) as well as on a subset of weight vectors.

COROLLARY 1. *The optimal placement of a point x_i with $x_{i-1}^\mu \leq x_i \leq x_{i+1}^\mu$ that maximizes the $R2$ indicator, given that all other points $x_1^\mu, \ldots, x_{i-1}^\mu, x_{i+1}^\mu, \ldots, x_\mu^\mu$ are already known, only depends on x_{i-1}^μ and x_{x+1}^μ and the weight vectors between these two points. Likewise, the optimal placements of x_1 and x_μ only depend on their single neighbor.*

PROOF. According to the above lemmas, for all weight vectors to the left of x_{i-1}^μ, x_{i-1}^μ is better than x_i and, analogously, for all weight vectors to the right of x_{i+1}^μ, x_{i+1}^μ is better than x_i. Consequently, the optimal placement of x_i only depends on these vectors. □

3.2 Optimal μ-Distributions for the $R2$ Indicator

With the above lemmas, we can now prove a few general statements about the optimal distribution of μ points with respect to the $R2$ indicator. In particular for the special case of more points than weight vectors ($\mu \geq |\Lambda|$), the optimal μ-distributions turn out to contain the intersection points of the rays corresponding to the weight vectors with the Pareto front.

THEOREM 1. *In the case that $\mu \geq |\Lambda|$, optimal μ-distributions for the $R2$ indicator with given weight vectors Λ contain all intersection points $\{a \in A | \lambda^a \in \Lambda\}$ between the rays defined by the weights and the actual Pareto front.*

PROOF. As we have seen in Lemma 2, no nondominated solution is better with respect to a given weight vector than the solution actually lying on the corresponding ray. Together with the weak monotonicity of the $R2$ indicator, this means that for a given weight vector, no other (feasible) solution can have a better $R2$ indicator contribution for this weight than the intersection between the corresponding ray and the Pareto front. Assuming $\mu \geq |\Lambda|$, the optimal μ-distribution consequently will include all those intersection points. □

Note here that the optimal μ-distribution is unique in the case $\mu = |\Lambda|$ and that all rays have to intersect with the Pareto front. In the case where μ is larger than the number N of intersection points between the rays and the Pareto front, the $R2$ indicator has no influence on $\mu - N$ points. On the opposite, we can show that the optimal μ-distributions in the case $\mu < |\Lambda|$ are also not always unique.

THEOREM 2. *In case $\mu < |\Lambda|$, there are examples where the optimal μ-distributions are not unique.*

PROOF. Let us investigate an example with $\mu = 3$ points and $N = 4$ weights. The problem is DTLZ1 [6], i.e., the front is described by $f(x) = 0.5 - x$. Let the four weights be $\lambda_1 = (\frac{1}{7}, \frac{6}{7})$, $\lambda_2 = (\frac{3}{7}, \frac{4}{7})$, $\lambda_3 = (\frac{4}{7}, \frac{3}{7})$, and $\lambda_4 = (\frac{6}{7}, \frac{1}{7})$. With the utopian point in $(-0.1, -0.1)$, this gives the scenario depicted in Fig. 2.

Let us assume for the moment that the two extreme points are used and that we have to place the third point in between them. We can then compute the $R2$ indicator value with respect to the weight vectors λ_2 and λ_3 dependent on the x-value of the third point while the $R2$ indicator contributions with respect to λ_1 and λ_4 are fixed to their optimal value by the other two points lying on the extremes. With simple calculus, we obtain

$$\begin{aligned} R2(x) &= \min\left\{\max\left\{\frac{3}{7}|-0.1-x|, \frac{4}{7}|-0.6+x|\right\}, \frac{18}{70}\right\} \\ &\quad + \min\left\{\max\left\{\frac{4}{7}|-0.1-x|, \frac{3}{7}|-0.6+x|\right\}, \frac{18}{70}\right\} \\ &= \min\left\{\max\left\{\frac{3}{70}+\frac{3}{7}x, \frac{12}{35}-\frac{4}{7}x\right\}, \frac{9}{35}\right\} \\ &\quad + \min\left\{\max\left\{\frac{2}{35}+\frac{4}{7}x, \frac{9}{35}-\frac{3}{7}x\right\}, \frac{9}{35}\right\} \end{aligned}$$

for the $R2$ indicator value depending on the x-value of the third point with $0 \leq x \leq 0.5$. The previous equation can be rewritten as

$$R2(x) = \begin{cases} \frac{36}{70} - \frac{3}{7}x & \text{for } 0 \leq x < 0.15 \\ 0.6 - x & \text{for } 0.15 \leq x < 0.2 \\ 0.4 & \text{for } 0.2 \leq x \leq 0.3 \\ 0.1 + x & \text{for } 0.3 < x \leq 0.35 \\ 0.3 + \frac{3}{7}x & \text{for } 0.35 < x \leq 0.5 \end{cases}$$

with the minimum value of 0.4 reached for $x \in [0.2, 0.3]$, see also the righthand plot of Fig. 2. Hence, the optimal 3-distribution is not unique in this case and not even a finite number of distinct optimal 3-distributions exist, but infinitely many solution sets are optimal.

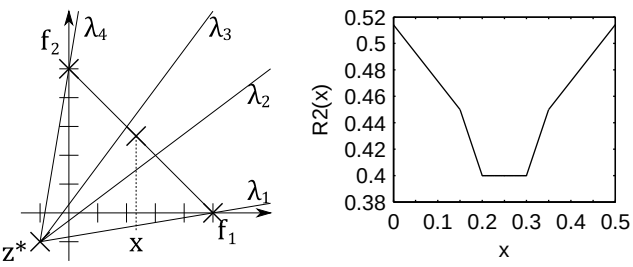

Figure 2: An example where the optimal μ-distribution for the $R2$ indicator is not unique: scenario (left) and corresponding $R2$ indicator (w.r.t. λ_2 and λ_3) when the third point moves from $x = 0$ on the left of the front to $x = 0.5$ on the right of the front (right).

Let us briefly comment on the leftmost and rightmost solutions and the above assumption to place them on the boundaries of the Pareto front. Due to symmetry reasons, we only consider the placement of the leftmost point. With x_2^μ in the optimal interval $[0.2, 0.3]$, we know that the placement of the leftmost point x_1^μ is only influenced by λ_3 and λ_4. Hence, we can compute the contribution of x_1^μ which results, again with simple calculus, in $\frac{6}{7}x_1^\mu + \frac{18}{70}$ and which is minimized for $x_1^\mu = 0$. □

4. APPROXIMATIONS OF OPTIMAL μ-DISTRIBUTIONS FOR THE $R2$ INDICATOR

The above theoretical results already gave some insights into the optimal distribution of μ points regarding the $R2$ indicator. In this section, these insights are enhanced with empirical observations with respect to the optimal μ-distributions of this indicator and an analysis of the difference to the optimal μ-distributions of the HV indicator. To accomplish this, we chose the three established test problems ZDT1 [13], DTLZ1, and DTLZ2 [6] with different Pareto front shapes—convex, linear (as a special case of convexity), and concave. Their exact Pareto front definitions are $f(x) = 1 - \sqrt{x}$ (ZDT1), $f(x) = 0.5 - x$ (DTLZ1), and $f(x) = \sqrt{1 - x^2}$ (DTLZ2). On each of these test problems, we approximated the optimal μ-distributions for $\mu = 10$ and the ideal point $(0, 0)$ as utopian using CMA-ES [8] on the ten respective x values. To also analyze the effect of the number of weight vectors N and to approximate the integral in Definition 1 which resembles an infinite number of weight vectors, the different multiples $N = \{\mu, 2\mu, 3\mu, 4\mu, 5\mu, 10\mu, 20\mu, 50\mu, 100\mu\}$ were considered. The same holds for different weight distributions by using uniform distributions in weight and angle space. On each combination of test problem, number of weight vectors, and weight distribution, 10 restarts of the CMA-ES are performed in order to have a more accurate estimate of the optimum distribution and to also analyze the variability in the results.

For a smart initialization, we used the theoretical result of Theorem 1. Based on the given distribution, μ weight vectors were generated and the corresponding intersections between the rays and the Pareto front were computed using standard calculus. The intersections' x-coordinates were then used as initial mean vector for the CMA-ES runs.

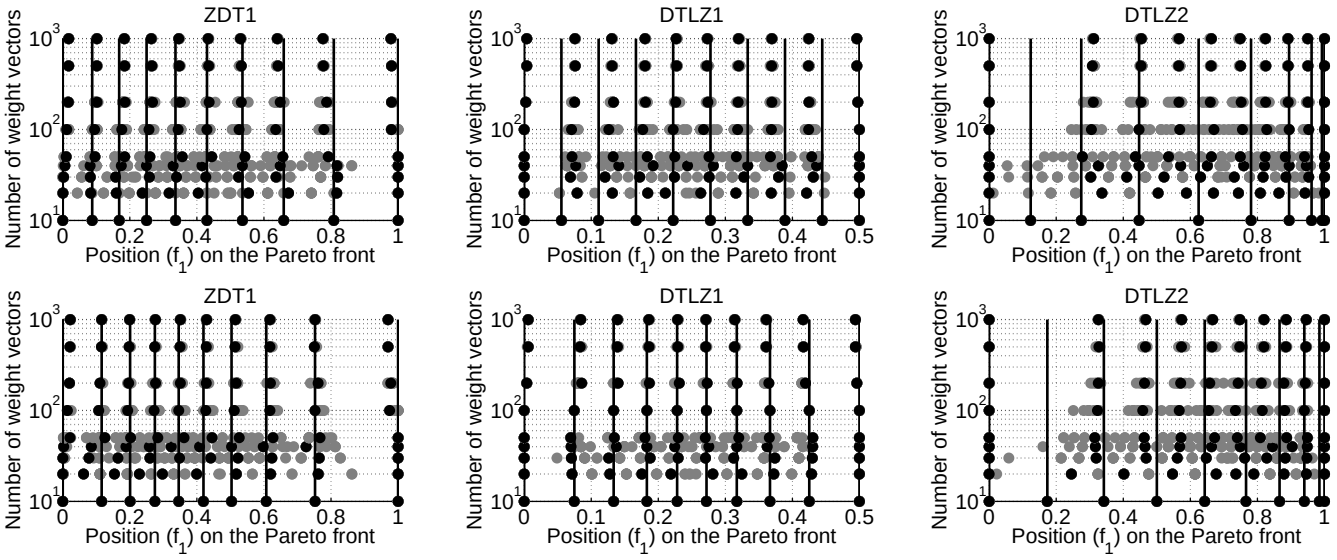

Figure 3: Positions (f_1-values) of the approximated optimal 10-distributions of the $R2$ indicator with increasing number of weight vectors uniformly distributed in weight space (top) and in angle space (bottom). Three different front shapes are considered: convex (ZDT1, $f_2 = 1 - \sqrt{f_1}$), linear (DTLZ1, $f_2 = 0.5 - f_1$), and concave (DTLZ2, $f_2 = \sqrt{1 - f_1^2}$). The final results of each CMA-ES run are depicted by gray dots. The run obtaining the best R2 value is highlighted in black. The positions of the initial points are indicated by the black lines. The utopian point was chosen as $(0, 0)$.

4.1 Effect of the Number of Weight Vectors

The results of the different runs of the CMA-ES are shown in Fig. 3. The positions of the initial solutions, i.e., the optimum distribution for $N = \mu$, is highlighted using black vertical lines. The final positions after each restart of the CMA-ES are depicted using gray dots. The best result for each setting is printed in black.

As a general observation, it can be stated that the variance of the results is higher for a lower number of weight vectors. In particular for $N < 10\mu$, it can hardly be distinguished between the result distributions of the different positions. For $N \geq 50\mu$, in contrast, the results of the CMA-ES hardly show any variance. It seems like each intersection point being a local optimum position, making the optimization hard for fewer weight vectors and higher distances between these local optima. With an increasing number of weights, the local basins become so small that the CMA-ES can easily jump from one to another.

With respect to the approximated optimal positions, the increase of N shows no significant effect for the convex Pareto front of ZDT1 and—given a uniform distribution in angle space—for the linear front of DTLZ1. The final points lie all close to the black lines indicating the optimal position for $N = \mu$. For uniform weights on DTLZ1, the points tend to become denser in the center of the Pareto front, as they all move towards the center from their initial position. Interestingly, for DTLZ2 it is shown that the optimal positions of the points extremely change with increasing N, i.e., for more than 100 weight vectors the distribution of the points is highly biased towards the right part of the front. No points but the extreme point at $f_1 = 0$ are placed within the interval $[0, 0.3]$. This may be caused by the possibility to strongly improve f_2 without deteriorating f_1 too much (cf. Fig. 5).

Even for the scenarios, in which the optimum distribution changes with increasing N, this distribution stabilizes after a specific number $N_{\text{threshold}}$ of weight vectors is exceeded. In our bi-objective experiments, this threshold lies at about $N_{\text{threshold}} = 10\mu = 100$. For higher N, the actual integral seems to be well approximated by the discretization by means of the sum over a finite set of N weight vectors.

4.2 Comparing the Hypervolume and the $R2$ Indicators

Despite of the empirically shown correlated behavior [10, 11], the indicators show specific preferences regarding the optimal distribution of the solutions on the Pareto front. For this purpose, it is analyzed how the R2 indicator would rearrange the points of the optimal μ-distribution w.r.t. the HV indicator and vice versa. In Fig. 4, the results for $\mu = 10$ on the considered test problems are shown. Starting from the optimal points on the x-axis (black dots), the contributions to the other indicator in case the least-contributing point (see Definitions 5 and Definition 7) is shifted to the respective x are shown on the y-axis.

For the optimal μ-distributions of the HV indicator, it becomes obvious that the R2 contributions tend to shift the points to the center of the front in order to reach a denser distribution in this region. The opposite tendency can be seen for the contributions to the HV indicator starting from the optimal μ-distribution of the R2 indicator. For the linear front of DTLZ1, for which the contributions are symmetric, the size of this shift decreases towards the center. This result is consistent for both kinds of weight distributions in the R2 indicator.

We now analyze the differences between the optimal distributions based on Fig. 5, in which we plot the 10-optimal distributions in objective space. For illustrating these dif-

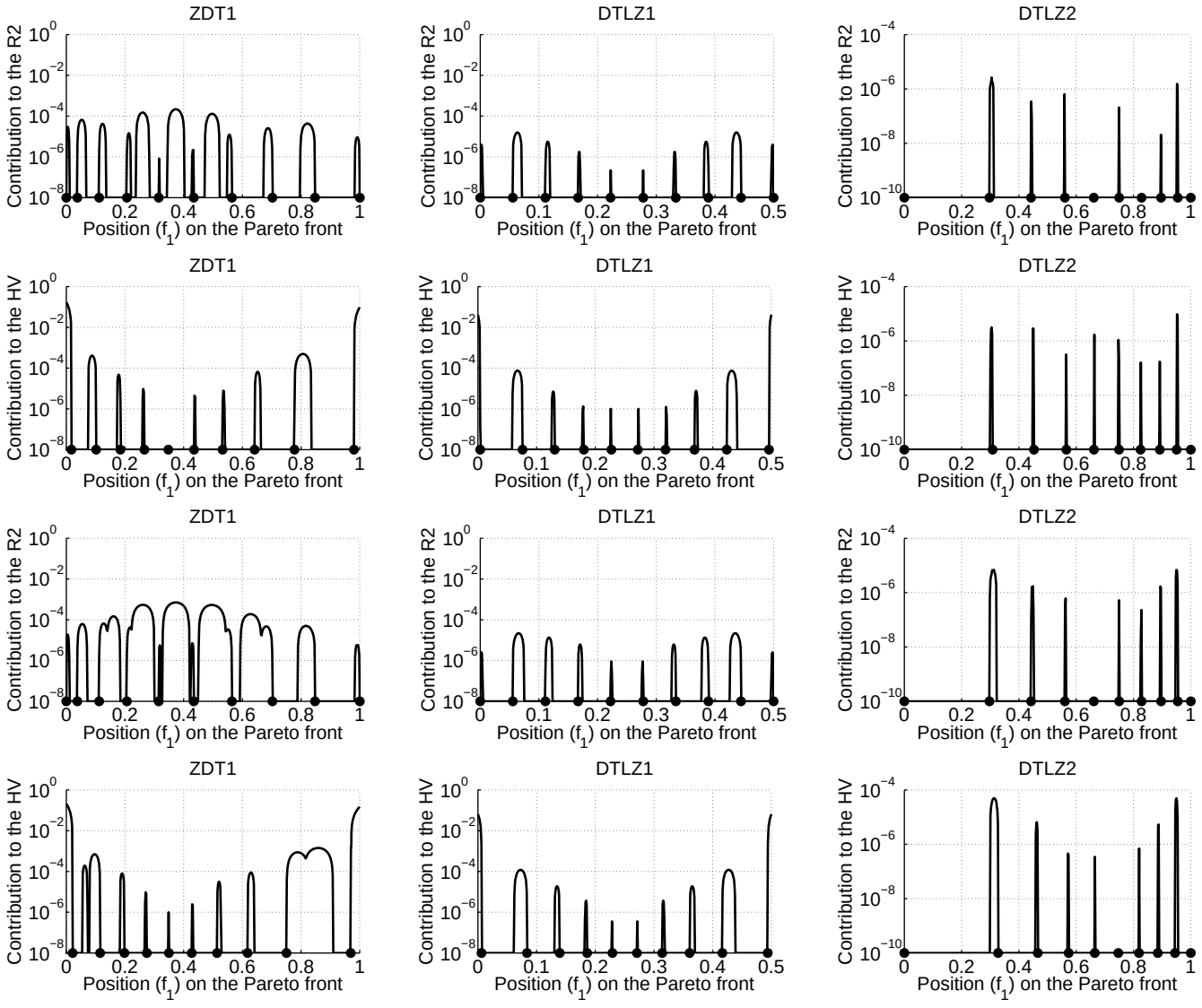

Figure 4: Contributions to either the R2 or the hypervolume (log-scale) when shifting the f_1 values of the optimal 10-distribution of the respective other indicator. The indicator to which the contributions are computed is indicated on the y-axis. The R2 indicator is calculated using 500 weights uniformly distributed in weight space (two lines on the top) or in angle space (two lines on the bottom) and the utopian point (0,0). The hypervolume is bounded using the reference point (11,11). Three different front shapes are considered: convex (ZDT1, $f_2 = 1 - \sqrt{f_1}$), linear (DTLZ1, $f_2 = 0.5 - f_1$), and concave (DTLZ2, $f_2 = \sqrt{1 - f_1^2}$).

ferences between the approximations of the optimal μ-distributions for the $R2$ indicator and the theoretically proven optimal ones for the HV indicator, we plot the individual differences between the x-values of corresponding points, i.e., $x^{\mu}_{1,R2} - x^{\mu}_{1,HV}$, $x^{\mu}_{2,R2} - x^{\mu}_{2,HV}$, ..., from left to right (on the bottom of Fig. 5). What can be seen is that in all cases, the optimal μ-distributions of the $R2$ indicators have a tendency towards the middle of the front, i.e., the distances between neighboring points compared to the HV indicator are larger at the extremes and smaller in the middle of the front. These distances turn out to be higher in angle space than in weight space. Sine-shaped curves of difference values result in the bottom of Fig. 5. It is caused by the higher influence of balanced weight vectors, i.e., vectors having almost similar weight components for each objective.

To understand the increase of influence for these weight vectors, consider the linear front of DTLZ1, $N = 101$ weight vectors, and a significantly lower number of points. For a uniform distribution in weight space, we know (cf. initial distributions of Fig. 3) that the intersection points of the vectors with the linear front are also uniformly distributed. Due to the selection of the best solution for each weight vector of the R2 indicator, each point covers a subset of weight vectors in its direct neighborhood. Thereby, the optimum point $(0.5, 0)$ for the extreme weight vector $(0, 1)$ has a value of $\max(0 \cdot 0.5, 1 \cdot 0.0) = 0$. Even for the

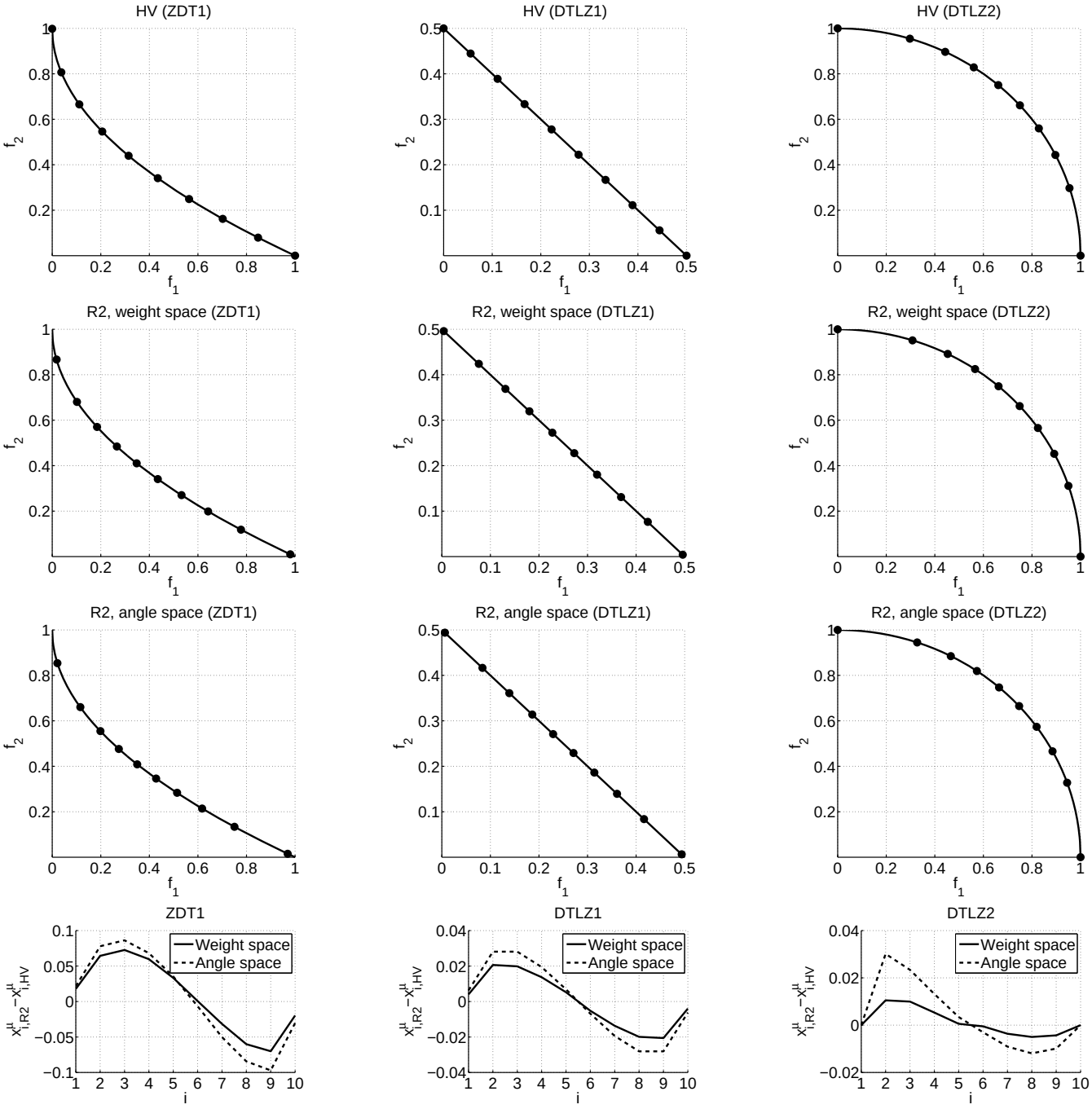

Figure 5: Comparison between approximations of the optimal 10-distributions for both $R2$ indicators and the theoretically known optimal 10-distributions for the hypervolume indicator on the ZDT1 (convex), DTLZ1 (linear) and DTLZ2 (concave) Pareto fronts. In the R2 indicator, 500 weight vectors and the utopian point $(0, 0)$ are used. On the top, the distributions are plotted in objective space, whereas the distances between the f_1-values of corresponding points in each distribution are plotted in the second row.

next weight vector $(0.01, 0.99)$, this point obtains a value of $\max(0.01 \cdot 0.5, 0.99 \cdot 0.0) = 0.005$ compared to the optimum value of $\max(0.01 \cdot 0.495, 0.99 \cdot 0.005) = 0.00495$. In contrast, the optimum point $(0.25, 0.25)$ for the balanced weight vector $(0.5, 0.5)$ has a value of $\max(0.5 \cdot 0.25, 0.5 \cdot 0.25) = 0.125$. For the next weight vector $(0.49, 0.51)$, the corresponding value of this point is $\max(0.49 \cdot 0.25, 0.51 \cdot 0.25) = 0.1275$ compared to the optimum value of $\max(0.49 \cdot 0.255, 0.51 \cdot 0.245) = 0.12495$. Whereas we loose 0.0005 at the extremes, we loose 0.00255 in the center—the loss is more than five times greater in this region. As a consequence, the density of points becomes higher to compensate for this effect.

5. CONCLUSIONS

As addition to the hypervolume (HV) indicator, the $R2$ indicator has been recommended as an indicator for performance assessment of multiobjective optimizers [16, 14]. Although the $R2$ indicator is only weakly monotonic with respect to the dominance relation, while the HV indicator is strictly monotonic, it has been reported that both indicators show a correlated behavior in practice [11].

In this paper, we introduced a unary, completely equivalent, version of the $R2$ indicator and investigated its properties and its differences to the HV indicator. Such kind of study was performed for the first time in this detail. It assisted in obtaining a deeper understanding of those popular indicators. In particular, it was shown that in the bi-objective case the $R2$ indicator has a bias towards the center of the Pareto front when compared to the HV indicator, which we demonstrated to be caused by the maximum operation of the Tchebycheff function. This finding is in contrast to the common assumption of a more uniform coverage obtained by the $R2$ indicator. We proved that the optimal placement of a point according to the $R2$ indicator only depends on its two neighbors and a subset of the weight vectors. Furthermore, approximations of the optimal μ-distributions for the $R2$ indicator based on uniform weight distributions in weight as well as in angle space were generated for standard test functions with different Pareto front shapes. They showed different characteristics compared to the optimal μ-distributions of the HV indicator [1], whereby the bias towards the center is even more pronounced for the uniform distribution in angle space.

Although the comparisons between the two indicators obtained important insights into their specific characteristics, several open questions need to be addressed in future studies. Besides investigating other problems with different, in particular discontinuous front shapes, one important aspect is the scaling of the indicators with the number of objectives. It is expected that the differences between the indicators become larger with three and more objectives, as it is known that the HV indicator tends towards the extremes of the Pareto front if the reference point is far enough away. In contrast, a uniform distribution of the $R2$ indicator's weight vectors should still have the effect of focusing on the center of the front, as observed in the bi-objective case. Also additional theoretical results on the optimal μ-distributions for the $R2$ indicator would be valuable, e.g., characterizing the exact optimal placements of μ solutions depending on the number and distribution of the weight vectors in terms of a density as in [1, 2]. These results may improve the design of multi-objective optimization algorithms, as well as their performance assessment.

6. ACKNOWLEDGMENTS

This paper is based on investigations of the project D5 "Synthesis and multi-objective model-based optimization of process chains for manufacturing parts with functionally graded properties" as part of the collaborative research center SFB/TR TRR 30, which is kindly supported by the Deutsche Forschungsgemeinschaft (DFG).

7. REFERENCES

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

[2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Hypervolume-based Multiobjective Optimization: Theoretical Foundations and Practical Implications. *Theoretical Computer Science*, 425:75–103, 2012.

[3] N. Beume et al. Hypervolume, 2 2012. `http://www.hypervolume.org`.

[4] N. Beume, C. M. Fonseca, M. Lopez-Ibanez, L. Paquete, and J. Vahrenhold. On the Complexity of Computing the Hypervolume Indicator. *IEEE Transactions on Evolutionary Computation*, 13(5):1075–1082, 2009.

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

[6] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable Multi-Objective Optimization Test Problems. In *Congress on Evolutionary Computation (CEC 2002)*, pages 825–830. IEEE Press, 2002.

[7] M. P. Hansen and A. Jaszkiewicz. Evaluating The Quality of Approximations of the Non-Dominated Set. Technical report, Institute of Mathematical Modeling, Technical University of Denmark, 1998. IMM Technical Report IMM-REP-1998-7.

[8] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[9] G. W. Snedecor and W. G. Cochran. *Statistical Methods*. Iowa State University Press, Iowa, 1989.

[10] S. Wessing. Towards Optimal Parameterizations of the S-Metric Selection Evolutionary Multi-Objective Algorithm. Diploma Thesis, Technical Report TR 09-06, TU Dortmund, 2009.

[11] S. Wessing and B. Naujoks. Sequential parameter optimization for multi-objective problems. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE, 2010.

[12] E. Zitzler. *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, ETH Zurich, Switzerland, 1999.

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

[14] E. Zitzler, J. Knowles, and L. Thiele. Quality Assessment of Pareto Set Approximations. In *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 373–404. Springer, 2008.

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

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