

R2 Indicator Based Multiobjective Search

Dimo Brockhoff, Tobias Wagner, Heike Trautmann

► **To cite this version:**

Dimo Brockhoff, Tobias Wagner, Heike Trautmann. R2 Indicator Based Multiobjective Search. *Evolutionary Computation*, Massachusetts Institute of Technology Press (MIT Press), 2015, 23 (3), pp.369-395. <http://www.mitpressjournals.org/doi/10.1162/EVCO_a_00135>. <10.1162/EVCO_a_00135>. <hal-01329559>

HAL Id: hal-01329559

<https://hal.inria.fr/hal-01329559>

Submitted on 10 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

R2 Indicator Based Multiobjective Search*

Dimo Brockhoff

dimo.brockhoff@inria.fr

DOLPHIN team, INRIA Lille - Nord Europe, 59650 Villeneuve d'Ascq, France

Tobias Wagner

wagner@isf.de

Institute of Machining Technology, TU Dortmund University, Germany

Heike Trautmann

trautmann@wi.uni-muenster.de

Information Systems and Statistics, Münster University, Germany

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 extended version of our previous conference 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. Furthermore, the $R2$ indicator is integrated into an indicator-based steady-state evolutionary multiobjective optimization algorithm (EMOA). It is shown that the so-called $R2$ -EMOA can accurately approximate the optimal distribution of μ solutions regarding $R2$.

Keywords

Multiobjective optimization, performance assessment, $R2$ indicator, $R2$ -EMOA, indicator-based search, environmental selection.

1 Introduction

Multiobjective optimization comprises the optimization of vectors $\vec{a} = (a_1, \dots, a_m)$ of m objectives. We assume minimization tasks in the following, as maximization problems can easily be transferred to those by negation. A solution \vec{a} in the objective space \mathbb{R}^m is said to dominate a solution $\vec{b} \in \mathbb{R}^m$ ($\vec{a} \prec \vec{b}$), iff \vec{a} is better than or equal to \vec{b} in all objectives ($\forall i \in \{1, \dots, m\} : a_i \leq b_i$) and better than \vec{b} in at least one ($\exists j \in \{1, \dots, m\} : a_j < b_j$). Only the first condition is necessary for \vec{a} to weakly dominate \vec{b} ($\vec{a} \preceq \vec{b}$). The aim of multiobjective optimization is to detect the Pareto optimal front $\mathcal{P} = \{\vec{a} \in \mathbb{R}^m \mid \nexists \vec{b} \in \mathbb{R}^m : \vec{b} \prec \vec{a}\}$.

Evolutionary multiobjective algorithms (EMOAs) 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 (distrib-

*This is the authors' version of the paper of the same title, published in the journal *Evolutionary Computation*, 23(2):369–395. The copyright is with MIT Press.

bution) (Zitzler, 1999). Several performance indicators were introduced (Zitzler et al., 2008, 2003) to assess one or all of these quality aspects of Pareto front approximations.

The Hypervolume (HV) (Zitzler and Thiele, 1998) and the $R2$ indicator (Hansen and Jaszkiewicz, 1998) are two recommended approaches (Knowles et al., 2006) which simultaneously evaluate all desired aspects of a Pareto front approximation. The HV 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 . In practice, this makes typically no difference for problems with continuous variables as it is unlikely that two solutions with the exact same objective values exist. Hence, the $R2$ indicator has been considered as an alternative for the HV for two reasons. On the one hand, the runtime of the HV is expected to increase exponentially with respect to the number of objectives m as its computation is known to be #P-hard (Bringmann and Friedrich, 2008). The best known algorithms to exactly compute the hypervolume indicator are all exponential in the number of objectives n , although there was some important recent progress: an $O(n^{(m+2)/3})$ time algorithm (Bringmann, 2012) beaten for 6, 5, and 4 objective functions by an $O(n^{(m-1)/2} \log(n))$ bound (Yildiz and Suri, 2012) and most recently a proof of an $O(n^{m/3} \text{polylog } n)$ upper bound (Chan, 2013) with m being the number of objectives. On the other hand, the distributions obtained using the HV are biased towards the knee regions of the Pareto front (Auger et al., 2009b). Note that in practice, the first argument against the hypervolume indicator is nowadays less convincing as, for example, in the 2- and 3-objective case algorithms running in time $O(n \log n)$ are known (Beume et al., 2009) and the hypervolume indicator (and single solutions' contributions to it) can be efficiently approximated by Monte Carlo sampling (Bader and Zitzler, 2011; Bringmann and Friedrich, 2009). Despite this point, we believe that there is a strong interest in an in-depth investigation of the $R2$ indicator and its relation to the HV.

As monotonicity of both, the HV and the $R2$ indicator, is guaranteed (Zitzler et al., 2003), a similar behavior of the indicators can be expected. Ishibuchi et al. even proposed a scalarization-based approach equal to the $R2$ indicator for approximating the hypervolume of an approximation set (Ishibuchi et al., 2009, 2010; Tsukamoto et al., 2011). However, as the preferred distributions seem to be different, the degree of similarity might vary in different settings. In (Wessing, 2009) and (Wessing and Naujoks, 2010), optimal parameterizations of the multiobjective optimizer SMS-EMOA (Beume et al., 2007) 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 (Snedecor and Cochran, 1989) 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, which is a revised and extended version of a former conference paper (Brockhoff et al., 2012), we aim at a more detailed understanding of the properties of the $R2$ indicator. We perform an analysis about *how* the $R2$ indicator differs from the HV indicator for which several theoretical properties are already known (Auger et al., 2009b, 2012). More specifically, we show that in the bi-objective case the $R2$ indicator

tends to place the points more concentrated in the central region of the Pareto front than the HV indicator and 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, 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 approximated for standard test functions with different Pareto front shapes. We show that those have different characteristics compared to the optimal μ -distributions of the HV indicator (Auger et al., 2009b). They turn out to be not unique in general and to depend on the number and distribution of the weight vectors used. In addition to the results presented in (Brockhoff et al., 2012), this paper also considers test problems with disconnected Pareto fronts within the experiments. 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 influence of the weight vector distribution and the placement of the ideal point is investigated in detail in (Wagner et al., 2013) where it is shown that the optimal μ -distributions of the $R2$ indicator heavily depends on the weight vector distribution, the domain of the weight space as well as the position of the ideal point.

Furthermore, we integrated the $R2$ indicator into the general framework of indicator-based EMOAs. The contribution to the $R2$ indicator is used as secondary selection criterion within a respective steady state version, denoted as $R2$ -EMOA in the following. In particular, we investigate whether the approximated optimal μ -distributions of the $R2$ indicator can be accurately reproduced by the greedy $R2$ -EMOA. Though preliminary experiments with the $R2$ -EMOA have been already presented in (Trautmann et al., 2013) on few test functions, we here analyze in more detail whether the $R2$ -EMOA is able to converge to the optimal μ -distributions for the $R2$ indicator in particular on problems with disconnected Pareto fronts. The results are also compared with the standard SMS-EMOA of Beume et al. (2007).

To be precise, the paper was enhanced and improved w.r.t. (Brockhoff et al., 2012) and (Trautmann et al., 2013) regarding the following aspects: a) additional investigations on test problems with disconnected Pareto fronts, b) an increased accuracy of the approximation of the optimal μ -distributions w.r.t. the $R2$ indicator, c) improvements of the comparison plots of weight and angle space, d) a more detailed analysis whether the optimal μ -distributions can actually be sufficiently approximated by a specifically designed EMOA based on $R2$ -indicator based selection ($R2$ -EMOA), e) a revised notation, and f) a few technical corrections (see Sec. 6).

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. The $R2$ -EMOA and related experiments are presented in Sec. 5. Finally, conclusions are drawn and an outlook is given in Sec. 6.

1.1 Notation

In this paper, we focus on optimal distributions in the objective space. The parameter vectors in the decision space are therefore omitted. Objective vectors are referred using lowercase letters \vec{a}, \vec{b}, \dots from the beginning of the alphabet. In cases where the positions on the bi-objective Pareto front are directly optimized, we refer to the respective coordinates of the first objective using x_1, \dots, x_μ for the μ points of the discrete approximation set. The corresponding values of the second objective with respect to the analytical equation of the Pareto front $f : \mathbb{R} \rightarrow \mathbb{R}$ are denoted as $y_1 = f(x_1), \dots, y_\mu = f(x_\mu)$.

2 R2 and HV Indicators

The R indicator family (Hansen and Jaszkiwicz, 1998) is based on utility functions which map a vector $\vec{a} \in \mathbb{R}^m$ to a scalar utility value $u \in \mathbb{R}$ for assessing the relative quality of two Pareto front approximation sets.

Definition 1 (Hansen and Jaszkiwicz (1998)). *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*

$$R2(R, A, U, p) = \int_{u \in U} \max_{\vec{r} \in R} \{u(\vec{r})\} p(u) du - \int_{u \in U} \max_{\vec{a} \in A} \{u(\vec{a})\} p(u) du .$$

Definition 2 (Zitzler et al. (2008)). *For a discrete and finite set U and a uniform distribution p over U , the original $R2$ indicator can be written as*

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

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 (Brockhoff et al. (2012)). *For a constant reference set, the $R2$ indicator can be defined as a unary indicator*

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

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 by Hansen and Jaszkiwicz (1998), we use the standard weighted Tchebycheff function $u(\vec{a}) = u_\lambda(\vec{a}) = -\max_{j \in \{1, \dots, m\}} \lambda_j |r_j^* - a_j|$ within the $R2$ indicator in the following where $\lambda = (\lambda_1, \dots, \lambda_m) \in \Lambda$ is a given weight vector and \vec{r}^* is a utopian point*.

Definition 4 (Brockhoff et al. (2012)). *The $R2$ indicator of a solution set A for a given set of weights Λ and a utopian point \vec{r}^* is defined as*

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

Usually, the weight vectors are chosen uniformly distributed over the weight space (Zitzler et al., 2008), 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

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

*An objective vector that is not dominated by any feasible search point.

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

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

by normalization.[†] A weight vector corresponding to a ray which directly crosses a point $\vec{a} \in A$ is denoted as $\lambda^{\vec{a}}$.

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 $\vec{a} \in A$ to the $R2$ indicator is defined as*

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

which can be written as

$$C_{R2}(\vec{a}, A, \Lambda, \vec{r}^*) = \frac{1}{|\Lambda|} \sum_{\lambda \in \Lambda} \left(\min_{\vec{b} \in A} \left\{ \max_{j \in \{1, \dots, m\}} \{ \lambda_j |r_j^* - b_j | \} \right\} - \min_{\vec{b} \in A \setminus \{\vec{a}\}} \left\{ \max_{j \in \{1, \dots, m\}} \{ \lambda_j |r_j^* - b_j | \} \right\} \right) .$$

In the remainder, we only consider the bi-objective case, i.e., $m = 2$. Finding the optimal μ -distribution then reduces to a μ -dimensional optimization problem

$$(x_1^\mu, \dots, x_\mu^\mu) = \underset{(x_1, \dots, x_\mu) \in \mathbb{R}^\mu}{\operatorname{argmin}} R2(\{(x_i, f(x_i)) \mid 1 \leq i \leq \mu\}, \Lambda, \vec{r}^*)$$

where we refer to the resulting set of x -values $(x_1^\mu, \dots, x_\mu^\mu)$ as optimal μ -distribution (Auger et al., 2009b; Brockhoff et al., 2012).

The HV indicator (Zitzler and Thiele, 1998; Zitzler et al., 2008) describes the hypervolume of a bounded space dominated by an approximation set.

Definition 6. *Based on an anti-optimal reference point \vec{p} , the HV indicator of an approximation set A is defined as*

$$HV(A, \vec{p}) = \mathcal{L} \left(\bigcup_{\vec{a} \in A} \{ \vec{b} \mid \vec{a} \prec \vec{b} \prec \vec{p} \} \right)$$

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

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 $\vec{a} \in A$ is defined as*

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

As mentioned in Sec. 1, the HV indicator is a performance indicator which is strictly compliant with the Pareto dominance relation. Implementations and a comprehensive overview of the state-of-the-art research w.r.t. the HV indicator are provided for example by Beume et al. (2012) and Bringmann (2012).

[†]The normalized weight vector to the ray $y = r_1^* + tx$ is given by $(\frac{t}{t+1}, \frac{1}{t+1})$ for $t \in]0, \infty[$, and by $(1, 0)$ if $t \rightarrow \infty$ or $(0, 1)$ if $t = 0$.

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 (Auger et al., 2009b). 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 (Auger et al., 2009b, 2012). 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 (Auger et al., 2009b). Before formalizing the general locality property of the $R2$ indicator, let us prove two preliminary statements.

Lemma 1 (Brockhoff et al. (2012)). *Given a specific weight vector $\lambda \in \Lambda$ and a specific solution $\vec{a} \in A$ where A only contains nondominated solutions. Then all points $\vec{b} \in A$ to the left of \vec{a} (to the right of \vec{a}) are resulting in a worse Tchebycheff utility than \vec{a} if $D_1 = \lambda_1|r_1^* - a_1| < D_2 = \lambda_2|r_2^* - a_2|$ (if $D_1 > D_2$).*

Proof. All nondominated points \vec{b} that are to the left of \vec{a} have an f_1 value that is smaller than the one of \vec{a} ($b_1 < a_1$) and their f_2 value is larger than the one of \vec{a} ($b_2 > a_2$). These points lie in the filled box in the left plot of Fig. 1—otherwise, \vec{b} would not be nondominated with respect to \vec{a} and lies to the left. In case that $D_1 = \lambda_1|r_1^* - a_1| < D_2 = \lambda_2|r_2^* - a_2|$, this means that for all other points \vec{b}

$$\lambda_1|r_1^* - b_1| < \lambda_1|r_1^* - a_1| < \lambda_2|r_2^* - a_2| < \lambda_2|r_2^* - b_2|$$

and thus for the Tchebycheff utility function, it holds that $\max\{\lambda_1|r_1^* - b_1|, \lambda_2|r_2^* - b_2|\} = \lambda_2|r_2^* - b_2| > \lambda_2|r_2^* - a_2| = \max\{\lambda_1|r_1^* - a_1|, \lambda_2|r_2^* - a_2|\}$, i.e., \vec{a} is better since we minimize the maximum weighted component. The proof of the other case $D_1 > D_2$ follows a symmetric argument and is visualized in the right plot of Fig. 1. \square

As a special case, we can investigate the specific weight vector $\lambda^{\vec{a}}$ that corresponds to a ray starting in the utopian point \vec{r}^* and going through \vec{a} (cf. Sec. 2).

Lemma 2 (Brockhoff et al. (2012)). *For each solution $\vec{a} \in A$, there exists a specific weight $\lambda^{\vec{a}} = (\lambda_1^{\vec{a}}, \lambda_2^{\vec{a}})$ with $\lambda_1^{\vec{a}} = \frac{|a_2 - r_2^*|}{|a_1 - r_1^*| + |a_2 - r_2^*|}$ and $\lambda_2^{\vec{a}} = 1 - \lambda_1^{\vec{a}}$ such that all other solutions $\vec{b} \in A$ which do not dominate the point \vec{a} have a worse weighted Tchebycheff value $\max\{\lambda_1^{\vec{a}}|b_1 - r_1^*|, \lambda_2^{\vec{a}}|b_2 - r_2^*|\}$ compared to \vec{a} .*

Proof. The proof follows the same ideas than the previous one. For $\lambda_1^{\vec{a}} = \frac{|a_2 - r_2^*|}{|a_1 - r_1^*| + |a_2 - r_2^*|}$ and $\lambda_2^{\vec{a}} = 1 - \lambda_1^{\vec{a}} = \frac{|a_1 - r_1^*|}{|a_1 - r_1^*| + |a_2 - r_2^*|}$, it follows that $D_1 = \lambda_1^{\vec{a}}|a_1 - r_1^*| = \lambda_2^{\vec{a}}|a_2 - r_2^*| = D_2$.

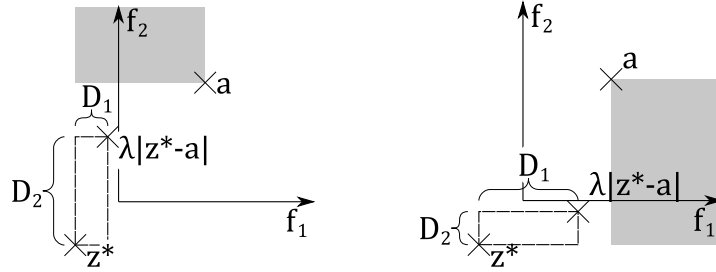


Figure 1: Sketch of idea for the proof of Lemma 1.

But then, for all solutions \vec{b} that do not dominate \vec{a} and are left of \vec{a} , $b_1 < a_1$ and $b_2 > a_2$ holds, such that $\lambda_1^{\vec{a}}|b_1 - r_1^*| < \lambda_1^{\vec{a}}|a_1 - r_1^*| = \lambda_2^{\vec{a}}|a_2 - r_2^*| < \lambda_2^{\vec{a}}|b_2 - r_2^*|$ which is equivalent to the fact that also the max in the weighted Tchebycheff for \vec{b} is larger and thus worse than the weighted Tchebycheff function for \vec{a} . With a symmetric argument, we can prove that also all solutions \vec{b} that do not dominate \vec{a} and are to the right of it are worse. \square

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 (Brockhoff et al. (2012)). *The optimal placement of a point with $x_{i-1}^\mu \leq x_i \leq x_{i+1}^\mu$ that maximizes the $R2$ indicator, given that all other points $x_1^\mu, \dots, x_{i-1}^\mu, x_{i+1}^\mu, \dots, x_\mu^\mu$ are already known, only depends on x_{i-1}^μ and x_{i+1}^μ and the weight vectors targeting the areas of the Pareto front 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. \square

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 (Brockhoff et al. (2012)). *In the case that $\mu \geq |\Lambda|$, optimal μ -distributions for the $R2$ indicator with given weight vectors Λ contain all intersection points $\{\vec{a} \in A | \lambda^{\vec{a}} \in \Lambda\}$ 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. \square

Note here that the optimal μ -distribution is unique in the case $\mu = |\Lambda|$ and given 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 (Brockhoff et al. (2012)). *In some cases where $\mu < |\Lambda|$, the optimal μ -distribution is not unique, i.e., more than one optimal μ -distribution might exist.*

Proof. Let us investigate an example with $\mu = 3$ points and $N = 4$ weights. The problem is DTLZ1 (Deb et al., 2002), 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 $\bar{r}^* = (-0.1, -0.1)$, 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

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

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.

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 $x_1^\mu = 0$. \square

Note that in the previous example, both the Pareto front and the weight vectors are symmetric with respect to the angle bisection of the first coordinate system quadrant

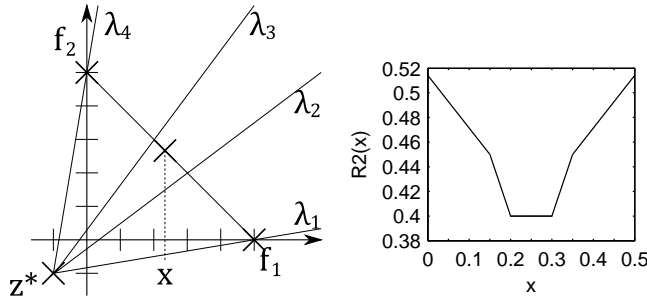


Figure 2: An example where the optimal μ -distribution for the $R2$ indicator is not unique: scenario (left) and corresponding $R2$ indicator (w.r.t. λ_2 and λ_3) when the third point moves from $x = 0$ on the left of the front to $x = 0.5$ on the right of the front (right).

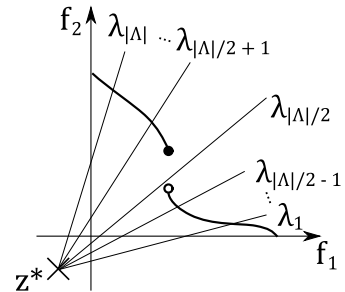


Figure 3: An example with non-existing μ -distributions where the number of weight vectors is even.

and it is furthermore *not* proven that, in general, more than one optimal μ -distribution can occur for any choice of $\mu < |\Lambda|$. Even more fundamentally, the question is open under which assumptions at least one optimal μ -distribution *exists*—for any choice of μ and $|\Lambda|$. One simple example for the non-existence of optimal μ -distributions is given below.

Example 1. Consider the case $\mu = |\Lambda|$, μ even, and an upper semi-continuous function $f(x) : x \mapsto f_2(x)$ describing the Pareto-optimal points with objective vector $(x, f_2(x))$ and which has a discontinuity. Let furthermore the $|\Lambda|$ weight vectors and the utopian point be defined such that all but one weight vectors define a weighted Tchebycheff problem that intersects with the front, described by f . The other weight vector shall define a weighted Tchebycheff problem, the direction of which goes through the discontinuity, see Fig. 3 for an illustration. Then, no optimal μ -distribution for the $R2$ indicator exist: Placing $\mu - 1$ points at the intersections of the weighted Tchebycheff problems with the Pareto front results in the minimal possible term for $\mu - 1$ of the summands in the $R2$ indicator of Def. 4—leaving room for only one point to be placed optimally. In order to also minimize the last summand, the last point can approach the discontinuity of f from the right, resulting in smaller and smaller $R2$ indicator values. But the minimal value can not be achieved because of the upper semi-continuity of f . Approaching the discontinuity from the other side will always result in even larger $R2$ indicator values.

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 consider the Pareto fronts of the three established test problems ZDT1 (Zitzler et al., 2000), DTLZ1, and DTLZ2 (Deb et al., 2002) because these fronts feature different 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 true ideal point $(0, 0)$ as utopian \vec{r}^* using the CMA-ES (Hansen and Ostermeier, 2001) on the ten respective x values. Other reference points \vec{r}^* were not considered as the effect of moving \vec{r}^* has been discussed in a recent

paper (Wagner et al., 2013). Other points than the true ideal point would either result in a focus on particular areas of the front or in a reduction of intersection points between target vectors and the front which is not desired in this experiment.

To also analyze the effect of the number of weight vectors and to approximate the integral in Definition 1 which resembles an infinite number of weight vectors, the different multiples $|\Lambda| = \{\mu, 2\mu, 3\mu, 4\mu, 5\mu, 10\mu, 20\mu, 50\mu, 100\mu\}$ 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 independent runs 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.

In addition, two test problems with disconnected Pareto fronts were considered, i.e. ZDT3 (Zitzler et al., 2000) and DTLZ7 (Deb et al., 2002). The respective experimental setup in principle coincides with the previously described experiments. The ideal point for ZDT3 is set to $\bar{r}^* = (0, -0.7735)$, and the levels of considered weight vector numbers are extended by $|\Lambda| = \{200\mu, 500\mu, 1000\mu\}$ as the number of rays intersecting with the Pareto front is smaller than $|\Lambda|$ due to the gaps between the disconnected parts of the Pareto front.

Due to the disconnected Pareto fronts, the search for the optimal μ -distributions is restricted to several disconnected intervals which imposes additional difficulties on the optimization process. In order to circumvent this situation, the intervals are artificially connected within the optimization process. The sum SL of the individual interval lengths is taken and each interval represents its proportion of SL in the connected space. The CMA-ES operates on the so-transformed decision space, i.e. $[0, 1]^\mu$. Within the fitness function calculation, the transformed solutions are mapped back to the original intervals, as it is of course the case for the final solutions of the CMA-ES optimization. The initial solutions are determined analogously on the transformed space. Due to the complex shape of the fronts and the additional transformation, an analytical solution is no longer possible. Therefore, binary search is used to calculate the intersections of the weight vectors with the objective space boundary.

4.1 Effect of the Number of Weight Vectors

The results of the different runs of the CMA-ES are shown in Figs. 4 and 5. The figures on the left result from a uniform distribution in weight space while the respective ones on the right are based on uniformly distributing weight vectors in angle space. The positions of the initial solutions, i.e., the optimum distribution for $|\Lambda| = \mu$, is highlighted using black vertical lines with the exception of ZDT3 and DTLZ7. As in this case the initial solution sets do not represent the optimal 10-distribution regarding $R2$ for 10 weight vectors, no vertical lines are added in the corresponding plots. The final positions after each run 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 for the test problems with continuous Pareto fronts the variance of the results is higher for a lower number of weight vectors. In particular for $|\Lambda| < 10\mu$, it can hardly be distinguished between the result

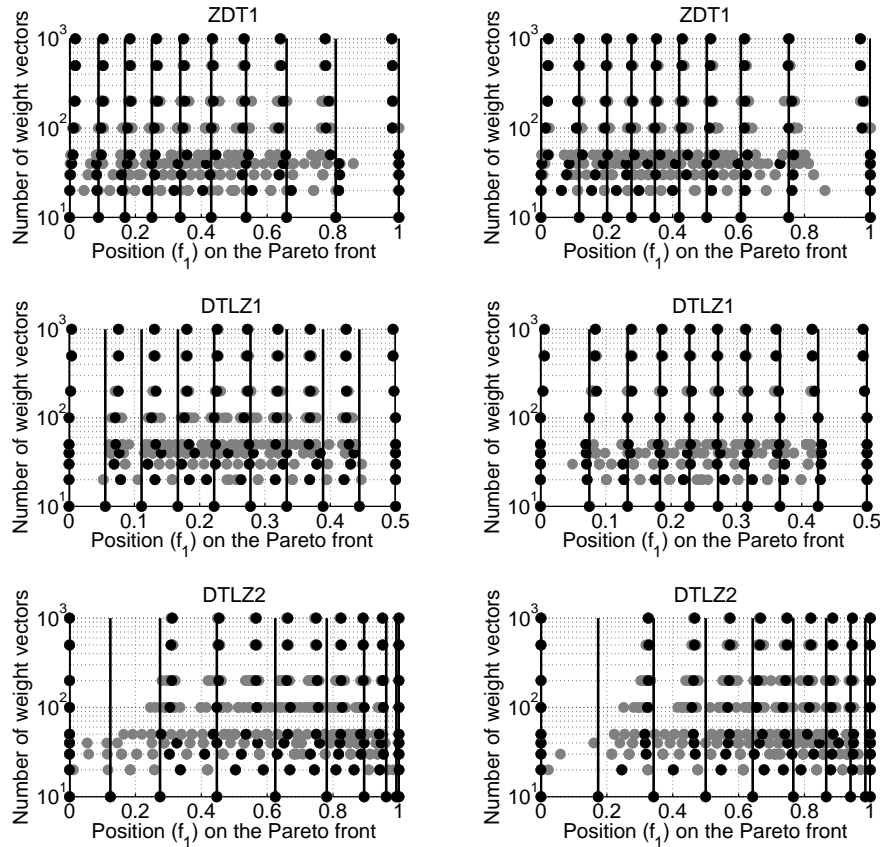


Figure 4: Positions (f_1 -values) of the approximated optimal 10-distributions on the connected fronts. On the y-axis of each figure, the corresponding number of weight vectors is provided. Each horizontal group of points thus relates to one of the considered sub-problems. The weight vectors are either uniformly distributed in weight space (left) or in angle space (right). 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 R_2 value is highlighted in black. The positions of the initial points are indicated by the vertical black lines. The utopian point was chosen as $\bar{r}^* = (0, 0)$.

distributions of the different positions. For $|\Lambda| \geq 50\mu$, 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. In contrast, for ZDT3 and DTLZ7 the variability of the result distributions unfortunately does not decrease significantly for $|\Lambda| > 50$, though the variability of the result distributions is much higher for $|\Lambda| \leq 50\mu$.

With respect to the approximated optimal positions, the increase of $|\Lambda|$ shows no significant effect for the convex Pareto front of ZDT1 and—given a uniform distribu-

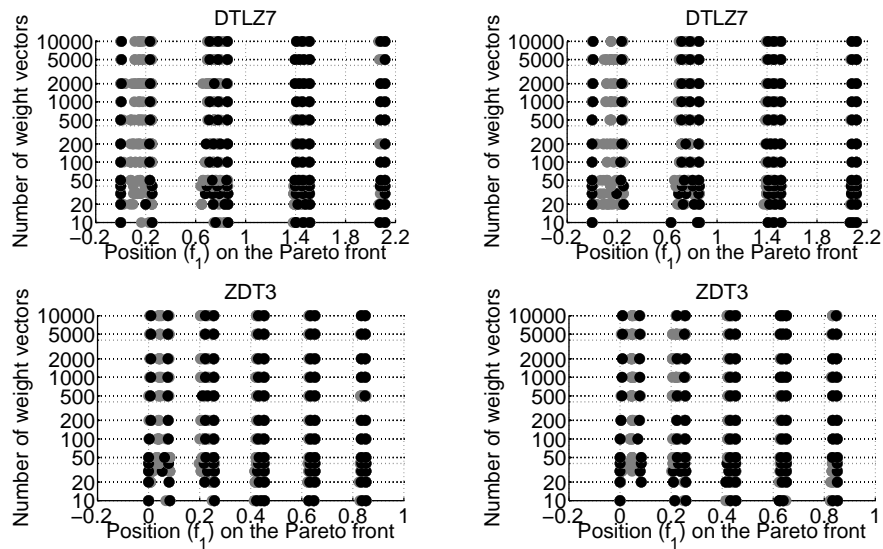


Figure 5: Positions (f_1 -values) of the approximated optimal 10-distributions on the disconnected fronts. On the y-axis of each figure, the corresponding number of weight vectors is provided. Each horizontal group of points thus relates to one of the considered subproblems. The weight vectors are either uniformly distributed in weight space (left) or in angle space (right). Two different front shapes are considered: DTLZ7 and ZDT3. The final results of each CMA-ES run are depicted by gray dots. The run obtaining the best R_2 value is highlighted in black. The utopian point was chosen as $\vec{r}^* = (0, 0)$ for DTLZ7 and $\vec{r}^* = (0, -0.7735)$ for ZDT3.

tion 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. 8).

Due to the fact that for ZDT3 and DTLZ7, the optimization has to deal with discontinuities we continue to see a large variance within the 10 CMA-ES runs for increasing number of weight functions. Especially in the two leftmost front parts, we can see this behavior in Fig. 5. When the CMA-ES has, for example, three solutions within the first front part, it seems to be good to place one of them in the middle of it while the other two should lie on the borders. However, in the more unlikely case (due to the initialization) that only two solutions map to the first front part, the R_2 indicator is better (black dots). What can also be observed from Fig. 5 is that, besides the case of $\mu = 10$ where more solutions exist than weight vectors intersect with the front, the number of weight vectors seems to have much less effect on the optimal μ -distribution than for the problems with connected fronts.

In all problems, and even for scenarios in which the optimum distribution changes

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

4.2 Comparing the Hypervolume and the $R2$ Indicators

Despite of the empirically shown correlated behavior (Wessing, 2009; Wessing and Naujoks, 2010), 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. 6 and 7, the results for $\mu = 10$ on the considered test problems are shown. Based on the optimal points on the x -axis (black dots), the contributions to the other indicator are shown on the y -axis. The latter are derived by analyzing the situation where the least-contributing point is shifted to the respective x ; the contributions along the shift direction are considered. Recall that the contributions of a solution to the specific indicators are provided in Definitions 5 and 7. By this means, structural differences of the two indicators in the optimal placement of solutions can be investigated. Moreover, the approach allows for the assessment of the idea of Ishibuchi and al. (Ishibuchi et al., 2010; Tsukamoto et al., 2011) to approximate the HV by means of the $R2$ indicator.

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. For the disconnected fronts, the contributions mainly indicate a replacement of solutions from one region to another. The shifts within a region are rather small.

We now analyze the differences between the optimal distributions based on Fig. 8, in which we plot the 10-optimal distributions in objective space. For illustrating these differences 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., $x_{1,R2}^\mu - x_{1,HV}^\mu$, $x_{2,R2}^\mu - x_{2,HV}^\mu, \dots$, from left to right (second last row of Fig. 8). In order to have resulting symmetric plots if the front itself is symmetric and to be able to compare results between different problems, we also plot the differences between the distributions' $\mu = 10$ points in the coordinate system along the front instead of their projection on the f_1 axis (last row in Fig. 8). The arc length along the front between the point $(0, f(0))$ and a point $(a, f(a))$ is thereby computed as $l_f(a) = \int_0^a \sqrt{1 + (f'(x))^2} dx$ where $f : x \mapsto f(x)$ describes the front shape and f' is its derivative. For the three test problems of Fig. 8,

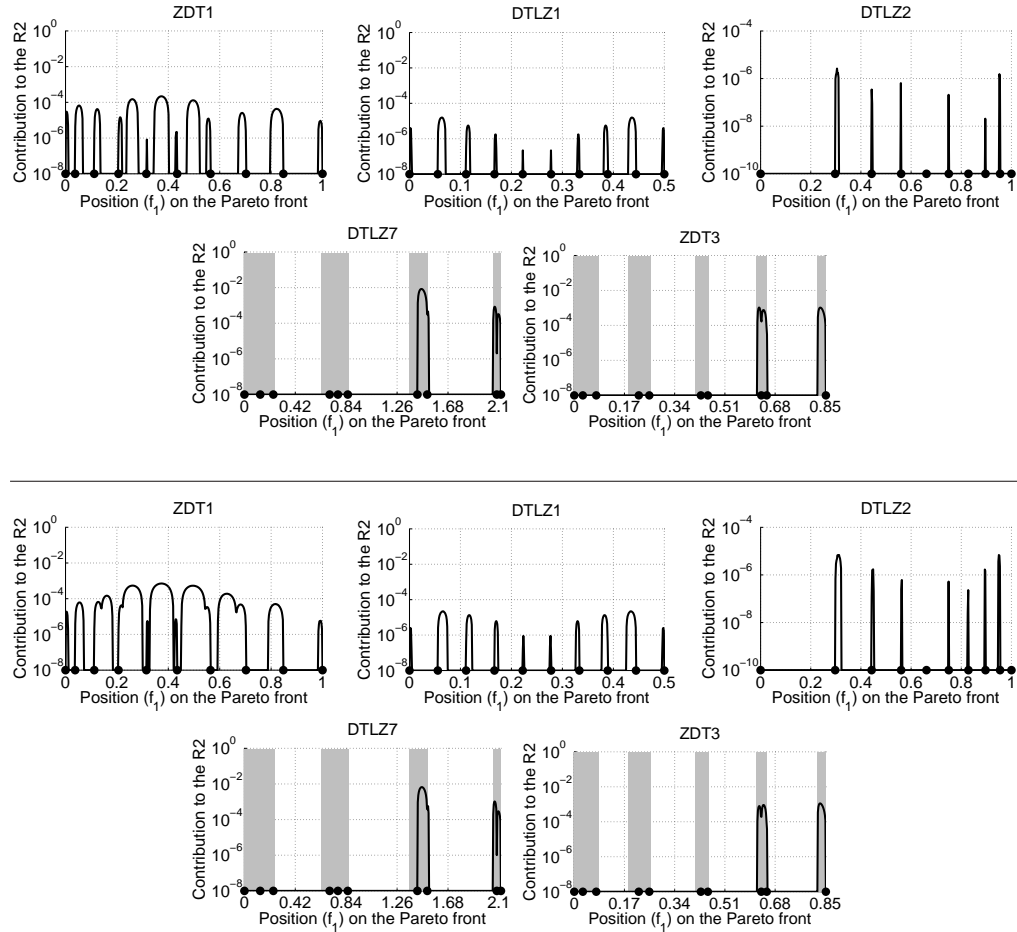


Figure 6: Contributions to the $R2$ indicator (log-scale) when shifting the f_1 values of the optimal 10-distribution of the HV indicator. The $R2$ indicator is calculated using 500 weights uniformly distributed in weight space (two lines on the top) or in angle space (two lines on the bottom) and the utopian point $(0, 0) - (0, -0.7735)$ for ZDT3. Five different front shapes are considered: convex (ZDT1, $f_2 = 1 - \sqrt{f_1}$), linear (DTLZ1, $f_2 = 0.5 - f_1$), and concave (DTLZ2, $f_2 = \sqrt{1 - f_1^2}$), as well as the disconnected fronts of DTLZ7 and ZDT3. For the latter, the f_1 -values resulting in Pareto-optimal solutions are highlighted in gray.

we can give analytical expressions for l_f :

$$\begin{aligned}
 l_{\text{DTLZ1}}(a) &= \int_0^a \sqrt{2} dx = \sqrt{2}a \\
 l_{\text{DTLZ2}}(a) &= \int_0^a \frac{1}{\sqrt{1-x^2}} dx = \arcsin(a) \text{ and} \\
 l_{\text{ZDT1}}(a) &= \int_0^a \sqrt{1 + \frac{1}{4x}} dx = \frac{1}{8} \left(4x \sqrt{\frac{1}{x} + 4} + \log \left(4x \left(\sqrt{\frac{1}{x} + 4} + 2 \right) + 1 \right) \right) .
 \end{aligned}$$

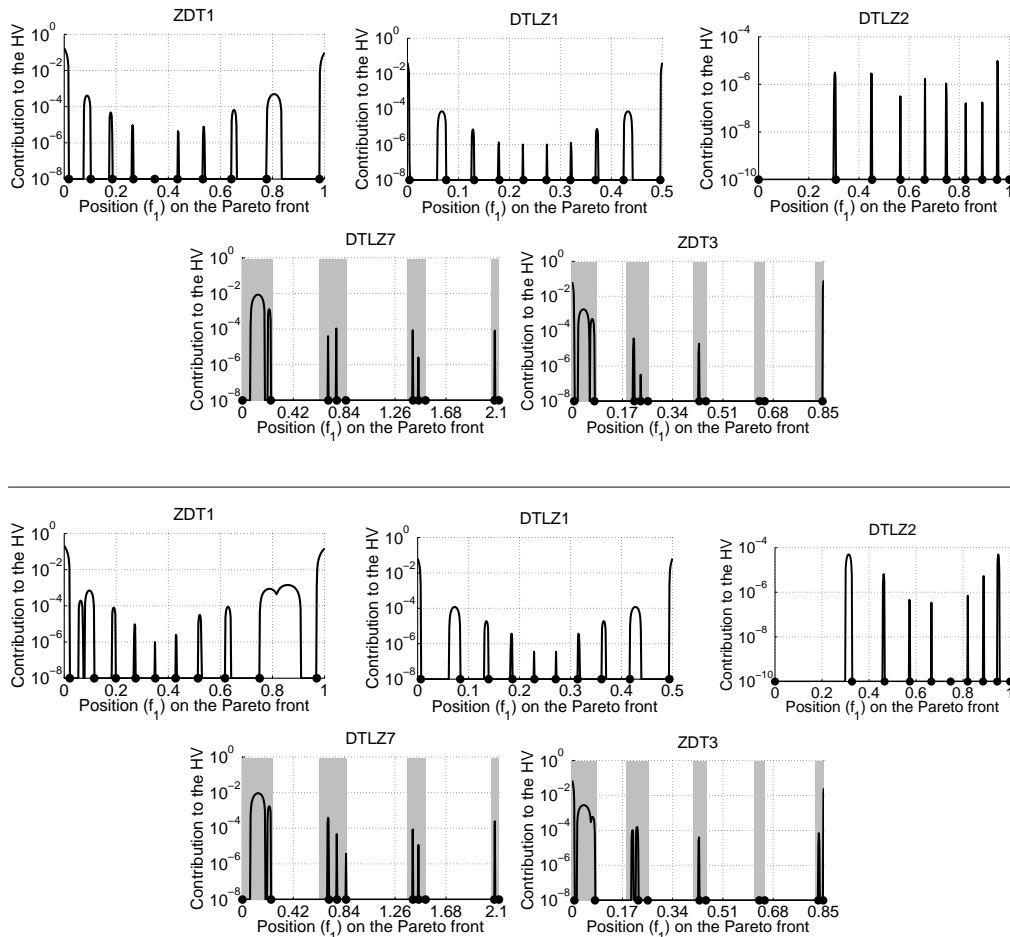


Figure 7: Contributions to hypervolume (log-scale) when shifting the f_1 values of the optimal 10-distribution of the R_2 indicator. The hypervolume is bounded using the reference point (11,11). The R_2 indicator is calculated using 500 weights uniformly distributed in weight space (two lines on the top) or in angle space (two lines on the bottom) and the utopian point $(0, 0) - (0, -0.7735)$ for ZDT3. Five different front shapes are considered: convex (ZDT1, $f_2 = 1 - \sqrt{f_1}$), linear (DTLZ1, $f_2 = 0.5 - f_1$), and concave (DTLZ2, $f_2 = \sqrt{1 - f_1^2}$), as well as the disconnected fronts of DTLZ7 and ZDT3. For the latter, the f_1 -values resulting in Pareto-optimal solutions are highlighted in gray.

All values are normalized by the total length of the front.

What can be seen from Fig. 8 is that in all cases, the optimal μ -distributions of the R_2 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. 8. 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

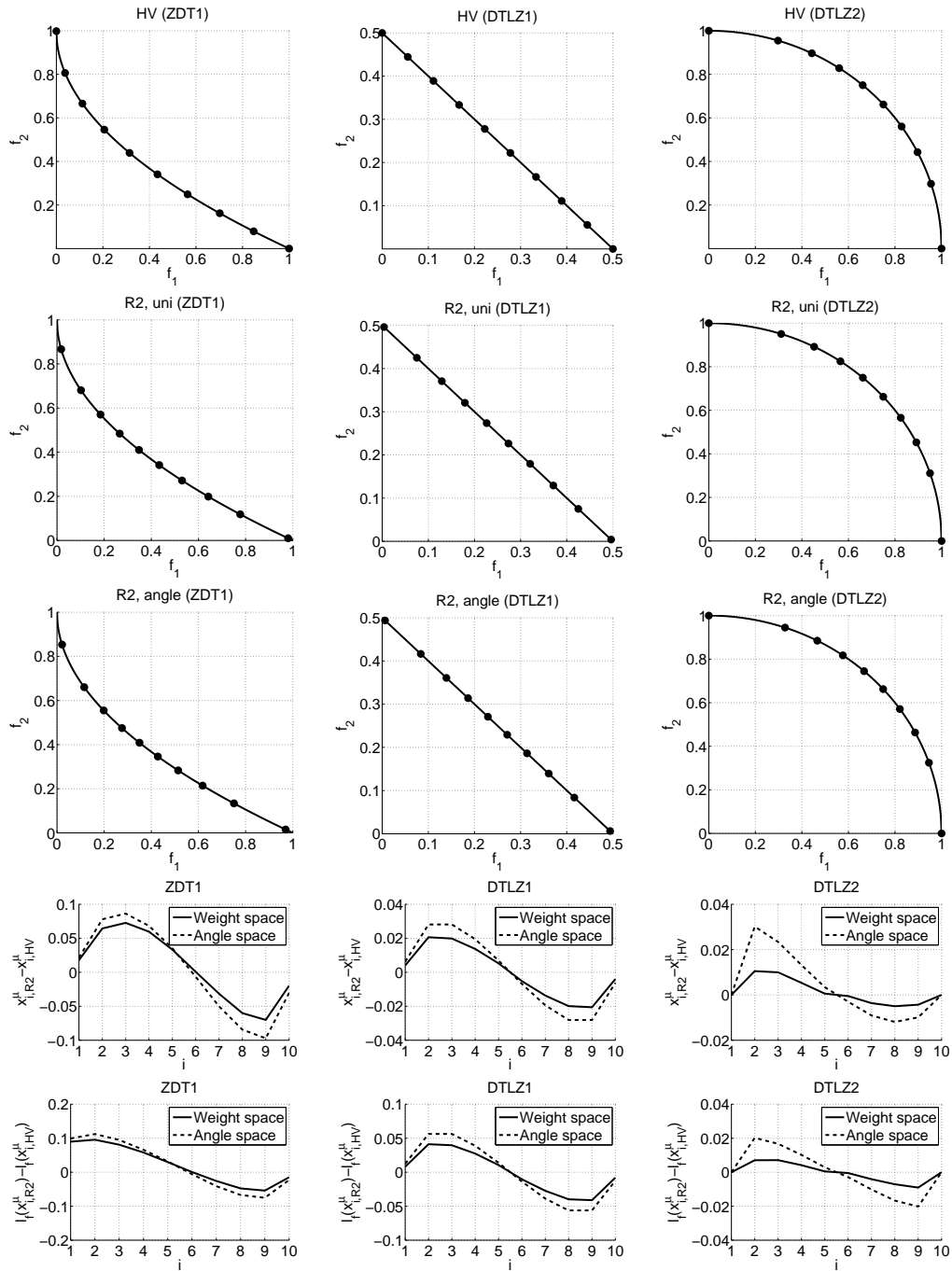


Figure 8: Comparison between approximations of the optimal 10-distributions for both $R2$ indicators and the theoretically known optimal 10-distributions for the HV indicator on the ZDT1 (convex), DTLZ1 (linear) and DTLZ2 (concave) Pareto fronts (top three rows). In the $R2$ indicator, 500 weight vectors and the utopian point $(0, 0)$ are used. Below the distances between the $R2$ indicator distributions' positions and the optimal HV distributions in f_1 -space (second last row) and along the front (last row) are shown.

linear front of DTLZ1, $|\Lambda| = 101$ weight vectors, and a significantly lower number of points. For a uniform distribution in weight space, we know (cf. initial distributions of Fig. 4) 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 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 lose 0.0005 at the extremes, we lose 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.

Note also that the bottom row of Fig. 8 suggests that the absolute differences of the optimal μ -distributions between a uniform distribution of the weights in weight and a uniform distribution in angle space seem to be independent of the shape of the Pareto front. On the other hand, we can observe that, for both weight distributions, the differences to the optimal μ -distribution of the HV indicator become smaller “the more concave” the Pareto front is.

Fig. 9 visualizes the corresponding results for ZDT3 and DTLZ7. As a general observation it can be seen that the locations of the solutions have a tendency towards the right interval borders. This is caused by two reasons:

1. The points at the right border dominate all intersection points between the Pareto boundary and the target vectors in the gap-region. Even for the points on the upper left border of the next front segment, they provide almost the same value of the Tchebycheff scalarization with respect to the corresponding target vectors ,
2. The curvature of each segment monotonically decreases from left to right.

The latter is particularly interesting, as a higher density of points with slopes close to one has already been shown for the HV (Auger et al., 2009b). Indeed, differences in the point distributions between the HV and the *R2* indicator with equally distributed weight vectors in angle and weight space only occur if single points switch from one interval to an adjacent one while the positions of the remaining points stay the same. For DTLZ7, even the two different kinds of weight vector distributions lead to a very similar optimization result. It can already be observed in Fig. 5 that though there is variability in the optimization results even for large numbers of weight vectors, the best solution found almost stays the same for $|\Lambda| > 50$. Note that the coordinate system change to plot the differences along the front as in Fig. 8 is more difficult for the problems with disconnected fronts. It might also not bring additional insights as none of the fronts is symmetric such that we therefore do not show a similar plot here.

The disconnected fronts also allow the effect of the relative position of the reference point \bar{r}^* to the Pareto front to be assessed. As the central three Pareto front segments of ZDT3 are of comparable length and shape, they can be considered as results of different experiments with changing relative positions. A significant effect of the relative position can be observed. A third point is put to the central segment. In turn, the outer

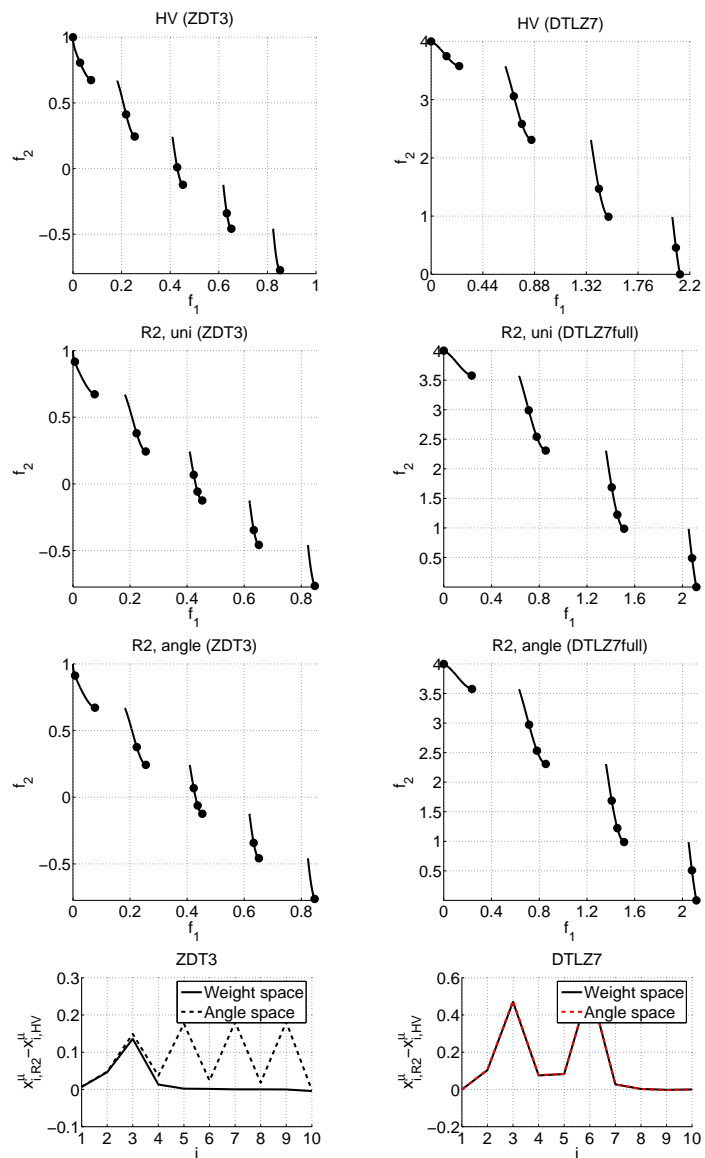


Figure 9: Comparison between approximations of the optimal 10-distributions for both $R2$ indicators and the theoretically known optimal 10-distributions for the HV indicator on the ZDT3 and DTLZ7 Pareto fronts (top three rows). In the $R2$ indicator, 500 weight vectors and the utopian point $(0, -0.7735)$ resp. $(0, 0)$ are used. Below the distances between the $R2$ indicator distributions' positions and the optimal HV distributions in x -space are shown.

Algorithm 1 Pseudocode of the *R2*-EMOA.

- 1: draw multiset P with μ elements $\in \mathbb{R}^n$ at random
 - 2: **repeat**
 - 3: generate offspring $\vec{z} \in \mathbb{R}^n$ from P by variation
 - 4: $P = P \cup \{\vec{z}\}$
 - 5: non-dominated sorting:
 build ranking R_1, \dots, R_h from P
 - 6: $\forall \vec{a} \in R_h : r(\vec{a}) = R2(R_h \setminus \{\vec{a}\}; \Lambda; \vec{r}^*)$
 - 7: $\vec{a}^* = \operatorname{argmin}\{r(\vec{a}) : \vec{a} \in R_h\}$
 - 8: $P = P \setminus \{\vec{a}^*\}$
 - 9: **until** stopping criterion fulfilled
-

segment is only covered by a single point. This reflects the higher density due to the balanced weight vectors already observed for the linear front.

5 The *R2*-EMOA: An EMOA With *R2* Indicator Based Selection

When investigating the optimal μ -distributions for the *R2* indicator, the immediate question arises whether they can be reached by an actual optimization algorithm. For the (weighted) HV indicator, Auger et al. (2009a) already showed that the algorithm W-HypE can approximate the optimal μ -distributions. Here, we will investigate how the basic indicator-based selection scheme that is used in many state-of-the-art EMOAs, such as in the SMS-EMOA (Beume et al., 2007) or in the MO-CMA-ES (Igel et al., 2007), can be translated to the *R2* indicator. In (Trautmann et al., 2013), we already presented preliminary experiments on this topic but with the slightly different focus on integrating preferences of a decision maker into the search algorithm and also only for three test functions with connected Pareto fronts. In the following, we will present the idea behind the *R2*-indicator based selection in the *R2*-EMOA, discuss some implementation details, and show an experimental comparison between the *R2*-EMOA, the SMS-EMOA, and the optimal μ -distributions for the *R2* indicator on several well-known test problems with connected and disconnected Pareto fronts.

5.1 *R2* Indicator Based Selection

Algorithm 1 shows the pseudocode of the *R2*-EMOA with a simple *R2* indicator based steady state selection. After a random initialization of the population and the generation of one new offspring per iteration, the *R2*-EMOA uses standard non-dominated sorting and deletes the solution with the worst rank that has the smallest *R2* indicator loss, or, in other words, the deletion of which results in the smallest *R2* indicator value of the remaining population. The utopian point \vec{r}^* and the set of weight vectors Λ are direct parameters of the algorithm which makes it possible to integrate various preferences of a decision maker into the search as we have shown in a preliminary study (Trautmann et al., 2013). Here, we present results of a MATLAB implementation which is further optimized for speed and therefore runs similarly fast as the MATLAB implementation of the SMS-EMOA (Beume et al., 2007) for the tested instances and parameter settings.

5.2 Implementation Details of the R2-EMOA

The above described R2-EMOA has been implemented in MATLAB on the basis of the SMS-EMOA implementation of Kretschmar and Wagner, available via Beume et al. (2012) which also served as a reference algorithm[‡]. To optimize for speed, the weighted Tchebycheff value for each solution and weight vector is only calculated once and stored throughout the algorithm. Furthermore, instead of calculating the R2 value for all the solution sets $R_h \setminus \{\vec{a}\}$, as in the pseudocode of Algorithm 1, we determine the worst solution within the set R_h directly. For each solution $\vec{a} \in R_h$, we thereby calculate the sum of differences in Tchebycheff values between solution \vec{a} and the *second best* solution for each weight vector over all the weight vectors, for which \vec{a} itself is resulting in the *best* Tchebycheff function value, see also Def. 5. The overall worst solution in R_h is then the one with the smallest sum.

For the following experiments, we use a population size of $\mu = 10$ in order to be able to compare the R2-EMOA results with the approximated optimal μ -distributions of the previous experiments. As variation operators, polynomial mutation and SBX crossover are used with the standard parameter setting from (Beume et al., 2012) (variable crossover probability: 0.9, distribution index for crossover: 15, variable mutation probability: $1/(\text{number of variables})$, distribution index for mutation: 20, variable swap probability: 0.5). Online convergence detection (Wagner and Trautmann, 2010; Wagner et al., 2011) is turned off and we run each algorithm 10 times with independently chosen random seeds. Each run is stopped after 150,000 function evaluations. According to whether the weights are chosen uniformly in weight space or in angle space, we distinguish between the two algorithms R2-EMOA^{uniweight} and R2-EMOA^{uniangle}. In all cases, the number of weights is 500. As test problems, we use again the above DTLZ1, DTLZ2, DTLZ7, ZDT1, and ZDT3 functions with 6, 11, 21, 30, and 30 decision variables as suggested in (Zitzler et al., 2000), (Deb et al., 2002), and (Deb et al., 2001) (for DTLZ7). For the DTLZ7 problem, and different to the original publication, we restrict the variables to the interval $[0, 2.116426807]$ instead of the interval $[0, 1]$ in order to have a Pareto front with 4 disconnected front parts as in the results on the optimal μ -distributions above. The initial population is drawn by sampling within the feasible search space uniformly at random. The ideal point is $(0, 0)$ except for the ZDT3 problem where it is chosen as $(0, -0.773500)$. The SMS-EMOA chooses the reference point of the HV indicator adaptively as in (Beume et al., 2007).

5.3 Comparison Baseline: Approximations of Optimal μ -Distributions

In order to compare the two R2-EMOA variants with the SMS-EMOA, we generated the best known approximation of the optimal μ -distribution for each test problem and 500 weight vectors in the following way. First, 10 independent CMA-ES runs with maximally 3 restarts each were started similar to the ones in Sec. 4. Second, a simple local search in terms of MATLAB's `fmincon` was run for each test problem. And last, the CMA-ES approximations as well as the final R2-EMOA populations after 150,000 function evaluations were further refined by running the same `fmincon` local search with the CMA-ES and R2-EMOA outputs as starting solutions. The best of all those approximations of the optimal μ -distribution was taken as a baseline in the following experimental comparison.

[‡]The source code of the R2-EMOA is available for download at <http://inriadortmund.gforge.inria.fr/r2emoa/>. It contains also a greedy version of the above steady-state selection as a generalization to the case of $\lambda \geq 1$ offspring per iteration.

5.4 Experimental Comparison

When looking at the differences between the $R2$ indicator values obtained by the optimization algorithms and the $R2$ indicator value of the best approximation of the optimal μ -distribution for $\mu = 10$ in Fig. 10, we can make two main observations.

First, the algorithm optimizing the $R2$ indicator used also for the performance assessment (with weights uniformly distributed in either weight or angle space) results most of the time in the smallest differences to the approximation of the optimal one. This is actually expected as the algorithm optimizes the same indicator that is used for the comparison. The only exceptions are the runs for the discontinuous problems ZDT3 and DTLZ7 and uniform weights in angle space. Here, several algorithms also show a large variance, both in the $R2$ indicator differences (Fig. 10) and in objective space (Fig. 11), especially for the DTLZ7 problem. The main reason for this seems to be the fact that not all algorithm runs cover all disconnected front parts. In particular for the DTLZ7 problem, it seems to be more difficult to produce a solution on a not yet covered front part than to (locally) optimize the $R2$ indicator by moving solutions within an already covered part of the Pareto front.

The second observation is that the $R2$ -EMOA is able to find solution sets that are close to the best known approximation of the optimum. In other words, the local search of `fmincon` did only allow for an additional gain in $R2$ indicator value between around 10^{-6} and 10^{-8} . Note that except for the case of DTLZ1 and uniform weight (where the stand-alone `fmincon` run resulted in the best $R2$ value), it was always the refinements of the $R2$ -EMOA populations which defined the best known approximations.

6 Conclusions

As addition to the hypervolume (HV) indicator, the $R2$ indicator has been recommended as an indicator for performance assessment of multiobjective optimizers (Zitzler et al., 2003, 2008). 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 (Wessing and Naujoks, 2010).

In this paper, which is a revised and extended version of a former conference paper (Brockhoff et al., 2012), 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 (see the discussion in Sec. 4.2). 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 (Auger et al., 2009b), whereby the bias towards the center is even more pronounced for the uniform distribution in angle space.

A steady-state EMOA based on the $R2$ indicator contributions as secondary selection criterion, denoted as $R2$ -EMOA, was also introduced. It was experimentally

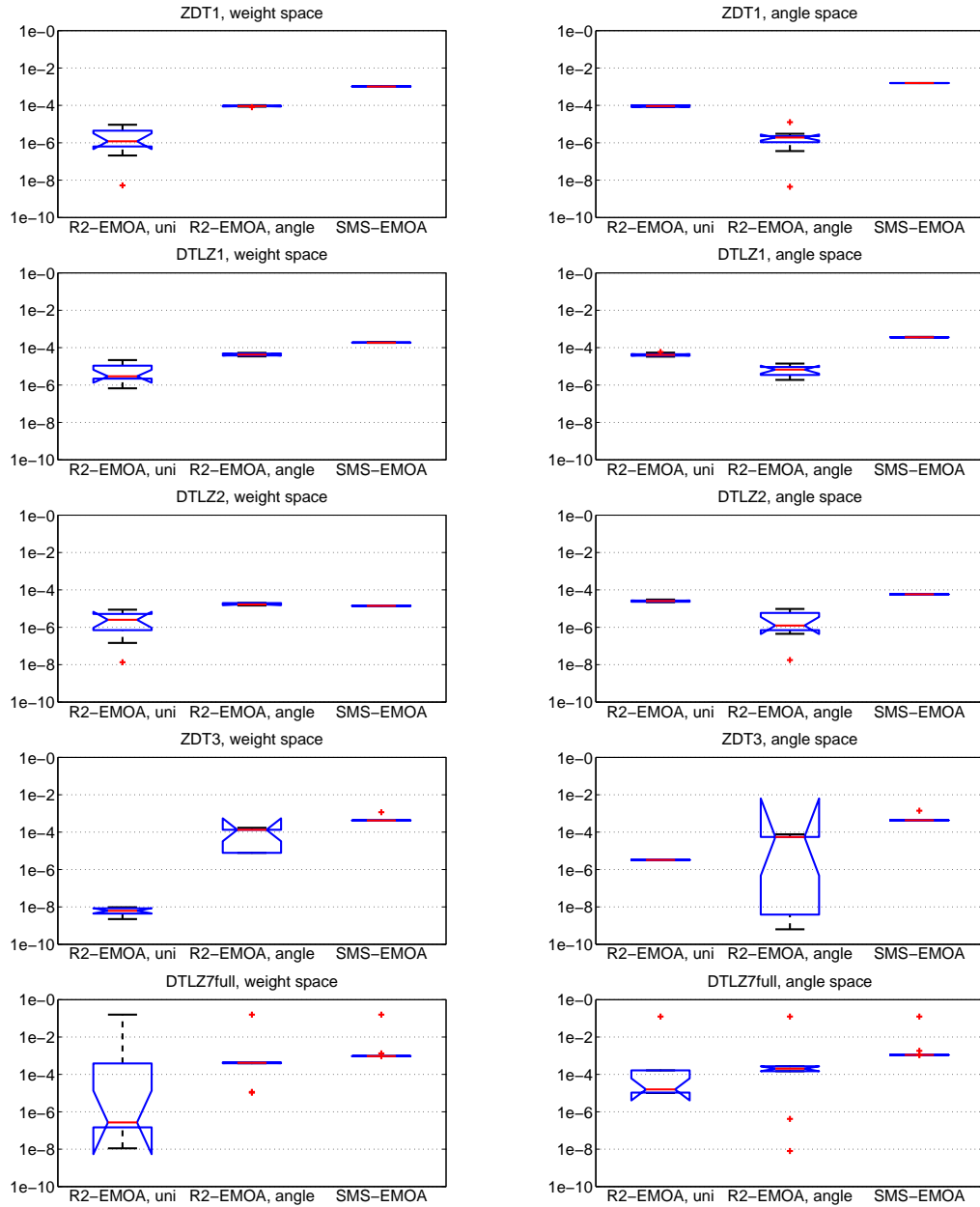


Figure 10: Box plots showing the differences in the $R2$ indicator (left: weight space, right: angle space) between the obtained population of the $R2$ -EMOA with weight vectors uniformly chosen in weight space, the $R2$ -EMOA with weights uniformly chosen in angle space, and the SMS-EMOA to an approximation of the optimal $R2$ value. Results are shown for (from top to bottom) ZDT1, DTLZ1, DTLZ2, ZDT3, and DTLZ7. For details, see the text.

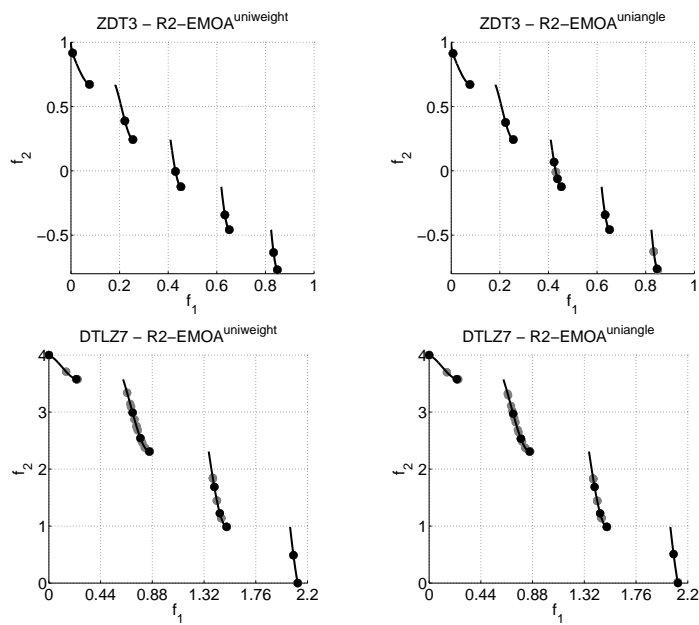


Figure 11: Final populations of the $R2$ -EMOA^{uniweight} (left) and the $R2$ -EMOA^{uniangle} (right) for the problems with disconnected Pareto fronts: ZDT3 (top) and DTLZ7 (bottom). In black the best population with respect to the $R2$ indicator with weights uniform in weight space (left) and uniform in angle space (right).

shown that the $R2$ -EMOA successfully approximates the optimal 10-distributions regarding $R2$ on the considered test functions. It turned out that disconnected Pareto fronts impose a much higher challenge on the $R2$ -EMOA optimization process than continuous fronts. However, for all test problems in most cases similar approximations could be evolved compared to the approximations of the optimal 10-distributions conducted by exploiting the knowledge of the true Pareto front shape. In future studies, the potential of other multiobjective optimization approaches beyond the steady-state concept will be investigated and additionally compared to other EMOA optimizing the $R2$ -indicator such as MOEA/D (Zhang and Li, 2007) and MSOPS (Hughes, 2007).

Although the comparisons between the two indicators obtained important insights into their specific characteristics, several open questions need to be addressed in future studies beyond the previously mentioned aspects. One important issue 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 (Auger et al., 2009b, 2012). These results may further improve the design of multi-objective optimization algorithms, as well as their performance assessment.

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). Heike Trautmann acknowledges support by the European Center of Information Systems (ERCIS). In addition, the authors acknowledge support by the French national research agency (ANR) within the Modèles Numérique project "NumBBO - Analysis, Improvement and Evaluation of Numerical Blackbox Optimizers" (ANR-12-MONU-0009).

Corrections With Respect to the GECCO Version

Note that with respect to the previous version of the paper (Brockhoff et al., 2012), the following two errors have been corrected.

- Definition 4 now reads $-\frac{1}{|\Lambda|} \sum_{\lambda \in \Lambda} \dots$ and not $\frac{1}{|\Lambda|} - \sum_{\lambda \in \Lambda} \dots$
- The conversion from slopes $(1, \tan(\phi))$ into weights now reads "rays from the utopian point with different slopes $\tan(\varphi)$ for $\varphi \in \Phi_k$ are constructed and then mapped to weight vectors

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

by normalization" and, correspondingly, the second footnote on page 2 has been corrected to "The normalized weight vector to the slope (x, y) is given by $(\frac{1/x}{1/x+1/y}, \frac{1/y}{1/x+1/y})$ as long as $x, y \neq 0$, and by $(1, 0)$ if $x = 0$ or $(0, 1)$ if $y = 0$. In our case, $x = 1$ and $y = \tan(\varphi)$."

References

- Auger, A., Bader, J., Brockhoff, D., and Zitzler, E. (2009a). Investigating and Exploiting the Bias of the Weighted Hypervolume to Articulate User Preferences. In Raidl, G. et al., editors, *Genetic and Evolutionary Computation Conference (GECCO 2009)*, pages 563–570. ACM.
- Auger, A., Bader, J., Brockhoff, D., and Zitzler, E. (2009b). Theory of the Hypervolume Indicator: Optimal μ -Distributions and the Choice of the Reference Point. In *Foundations of Genetic Algorithms (FOGA 2009)*, pages 87–102. ACM.
- Auger, A., Bader, J., Brockhoff, D., and Zitzler, E. (2012). Hypervolume-based Multiobjective Optimization: Theoretical Foundations and Practical Implications. *Theoretical Computer Science*, 425:75–103.
- Bader, J. and Zitzler, E. (2011). HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. *Evolutionary Computation*, 19(1):45–76.
- Beume, N. et al. (2012). Hypervolume. <http://www.hypervolume.org>.
- Beume, N., Fonseca, C. M., Lopez-Ibanez, M., Paquete, L., and Vahrenhold, J. (2009). On the Complexity of Computing the Hypervolume Indicator. *IEEE Transactions on Evolutionary Computation*, 13(5):1075–1082.
- Beume, N., Naujoks, B., and Emmerich, M. (2007). SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669.
- Bringmann, K. (2012). Bringing order to special cases of klee's measure problem. In *Mathematical Foundations of Computer Science (MFCS 2013)*, pages 207–218. Springer.
- Bringmann, K. and Friedrich, T. (2008). Approximating the Volume of Unions and Intersections of High-Dimensional Geometric Objects. In Hong, S. H., Nagamochi, H., and Fukunaga, T., editors, *International Symposium on Algorithms and Computation (ISAAC 2008)*, pages 436–447. Springer.
- Bringmann, K. and Friedrich, T. (2009). Approximating the Least Hypervolume Contributor: NP-hard in General, But Fast in Practice. In Ehr Gott, M. et al., editors, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, volume 5467 of LNCS, pages 6–20. Springer.
- Brockhoff, D., Wagner, T., and Trautmann, H. (2012). On the Properties of the R2 Indicator. In *Genetic and Evolutionary Computation Conference (GECCO 2012)*, pages 465–472.
- Chan, T. (2013). Klee's Measure Problem Made Easy. In *Foundations of Computer Science (FOCS 2013)*, pages 410–419.
- Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. (2001). Scalable Test Problems for Evolutionary Multi-Objective Optimization. TIK Report 112, Computer Engineering and Networks Laboratory (TIK), ETH Zurich.
- Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. (2002). Scalable Multi-Objective Optimization Test Problems. In *Congress on Evolutionary Computation (CEC 2002)*, pages 825–830. IEEE Press.
- Hansen, M. P. and Jaszkiwicz, A. (1998). Evaluating The Quality of Approximations of the Non-Dominated Set. Technical report, Institute of Mathematical Modeling, Technical University of Denmark. IMM Technical Report IMM-REP-1998-7.
- Hansen, N. and Ostermeier, A. (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195.
- Hughes, E. J. (2007). MSOPS-II: A general-purpose many-objective optimiser. In *Congress on Evolutionary Computation (CEC 2007)*, pages 3944–3951. IEEE Press.
- Igel, C., Hansen, N., and Roth, S. (2007). Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28.

- Ishibuchi, H., N. Tsukamoto, Y. S., and Nojima, Y. (2009). Hypervolume approximation using achievement scalarizing functions for evolutionary many-objective optimization. In *Proc. IEEE Congress on Evolutionary Computation (CEC '09)*, pages 530–537.
- Ishibuchi, H., N. Tsukamoto, Y. S., and Nojima, Y. (2010). Indicator-based evolutionary algorithm with hypervolume approximation by achievement scalarizing functions. In *Proc. 12th Annual Conference on Genetic and Evolutionary Computation (GECCO '10)*, pages 527–534.
- Knowles, J., Thiele, L., and Zitzler, E. (2006). A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. TIK Report 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich.
- Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods*. Iowa State University Press, Iowa.
- Trautmann, H., Wagner, T., and Brockhoff, D. (2013). R2-EMOA: Focused Multiobjective Search Using R2-Indicator-Based Selection. In *Learning and Intelligent Optimization Conference (LION 2013)*, pages 70–74.
- Tsukamoto, N., Sakane, Y., Nojima, Y., and Ishibuchi, H. (2011). Incorporation of hypervolume approximation with scalarizing functions into indicator-based evolutionary multiobjective optimization algorithms. *Transactions of the Institute of Systems, Control and Information Engineers*, 23:165–177.
- Wagner, T. and Trautmann, H. (2010). Online convergence detection for evolutionary multi-objective algorithms revisited. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE.
- Wagner, T., Trautmann, H., and Brockhoff, D. (2013). Preference Articulation by Means of the R2 Indicator. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2013)*, pages 81–95.
- Wagner, T., Trautmann, H., and Marti, L. (2011). A taxonomy of online stopping criteria for multi-objective evolutionary algorithms. In Takahashi, R., Deb, K., Wanner, E., and Greco, S., editors, *Evolutionary Multi-Criterion Optimization*, pages 16–30. Springer.
- Wessing, S. (2009). Towards Optimal Parameterizations of the S-Metric Selection Evolutionary Multi-Objective Algorithm. Diploma Thesis, Technical Report TR 09-06, TU Dortmund.
- Wessing, S. and Naujoks, B. (2010). Sequential parameter optimization for multi-objective problems. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE.
- Yildiz, H. and Suri, S. (2012). On Klee’s measure problem for grounded boxes. In *Symposium on Computational Geometry (SoGC 2012)*, pages 111–120. ACM.
- Zhang, Q. and Li, H. (2007). MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731.
- Zitzler, E. (1999). *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, ETH Zurich, Switzerland.
- Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195.
- Zitzler, E., Knowles, J., and Thiele, L. (2008). Quality Assessment of Pareto Set Approximations. In *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 373–404. Springer.
- Zitzler, E. and Thiele, L. (1998). Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In *Conference on Parallel Problem Solving from Nature (PPSN V)*, pages 292–301.
- Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Grunert da Fonseca, V. (2003). Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132.