

Preference Articulation by Means of the R2 Indicator

Tobias Wagner, Heike Trautmann, Dimo Brockhoff

► **To cite this version:**

Tobias Wagner, Heike Trautmann, Dimo Brockhoff. Preference Articulation by Means of the R2 Indicator. Evolutionary Multi-criterion Optimization (EMO 2013), Mar 2013, Sheffield, United Kingdom. 7811, pp.81-95, 2013, Lecture Notes in Computer Science Volume. <10.1007/978-3-642-37140-0_10>. <hal-00807867>

HAL Id: hal-00807867

<https://hal.inria.fr/hal-00807867>

Submitted on 4 Apr 2013

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.

Preference Articulation by Means of the $R2$ Indicator^{*}

Tobias Wagner¹ Heike Trautmann² Dimo Brockhoff³

¹ Institute of Machining Technology (ISF), TU Dortmund University, Germany
wagner@isf.de

² Statistics Department, TU Dortmund University, Germany
trautmann@statistik.tu-dortmund.de

³ DOLPHIN Team, INRIA Lille - Nord Europe, Villeneuve d'Ascq, France
dimo.brockhoff@inria.fr

Abstract. In multi-objective optimization, set-based performance indicators have become the state of the art for assessing the quality of Pareto front approximations. As a consequence, they are also more and more used within the design of multi-objective optimization algorithms. The $R2$ and the Hypervolume (HV) indicator represent two popular examples. In order to understand the behavior and the approximations preferred by these indicators and algorithms, a comprehensive knowledge of the indicator's properties is required. Whereas this knowledge is available for the HV, we presented a first approach in this direction for the $R2$ indicator just recently. In this paper, we build upon this knowledge and enhance the considerations with respect to the integration of preferences into the $R2$ indicator. More specifically, we analyze the effect of the reference point, the domain of the weights, and the distribution of weight vectors on the optimization of μ solutions with respect to the $R2$ indicator. By means of theoretical findings and empirical evidence, we show the potentials of these three possibilities using the optimal distribution of μ solutions for exemplary setups.

1 Introduction

Evolutionary multi-objective algorithms usually approximate the complete Pareto front of a problem in a single run. This is in contrast to classical MCDM approaches, which often apply sequential or hierarchical optimization runs to accomplish this task. In the beginnings of multi-objective optimization, three requirements for the set approximating the true Pareto front were defined: minimization of the distance (convergence), coverage of the extremes (spread), and a good representation of the actual shape (distribution) of the Pareto front [18]. In order to evaluate one or more of these requirements, several performance indicators were introduced [20, 22]. In particular, the Hypervolume (HV) [21] and the $R2$ indicator [11] are two recommended approaches which simultaneously evaluate all these desired aspects. Whereas the HV is a set-based quality indicator by definition, the R indicator family allows Pareto front approximations to be assessed based on a set of utility functions. Thereby, it is possible to find out which of the sets is better for specific preferences encoded in the weight vectors of the utility

^{*} This is an author version of the EMO'2013 paper published by Springer Verlag. The final publication is available at www.springerlink.com.

functions. A unary set-based quality indicator utilizing the mean utility over the weight vectors was proposed just recently [8]. In this paper, the former ideas of assessing regions of specific preferences are thus transferred to the unary variant of the indicator.

Many popular optimization algorithms are based on set-based quality indicators [5, 14, 17]. In order to learn about the outcome expected from these algorithms, the preferences and bias introduced by the choice of the indicator have to be understood. For the HV indicator, empirical studies [6] and theoretical results [1, 2] do already exist. With respect to the $R2$ indicator, we showed that the indicator tends to place the points more concentrated in the central region of the Pareto front than the HV 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 in the bi-objective case [8]. Here, we use these insights in order to define and analyze the preferences that can be introduced by a targeted choice of the reference point and the set of weight vectors. For the latter, particularly the influence of the covered domain and the density of the weight vectors on the optimal distributions of μ solutions are assessed. Thereby, we show that it is possible to restrict the search to a subregion and to adjust the focus of the distribution on the center or on the extremes of the Pareto front.

The terms and concepts required for understanding the methodological contributions are provided in the following section 2. Based on these auxiliary means, methods for integrating preferences into the $R2$ indicator are presented in section 3. The conceptual thoughts behind the methods are validated using empirical evidence. This is done in section 4. In the final section 5, the results are summarized and an outlook on further potentials for research on the $R2$ indicator is provided.

2 Foundations

Throughout the paper, we consider, without loss of generality, the simultaneous minimization of k objective functions $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ ($1 \leq j \leq k$) with respect to the *Pareto-dominance relation*. Since we are interested in optimal distributions of objective vectors, we will further neglect the corresponding decision variables in \mathbb{R}^n . Hence, we will use the terms solutions and objective vectors interchangeably in the following. We say a solution $x \in \mathbb{R}^k$ dominates a solution $y \in \mathbb{R}^k$ if $\forall j : x_j \leq y_j$ and $\exists j : x_j < y_j$. The solutions that are non-dominated by any other feasible solution are called Pareto-optimal, and we call the entire set of Pareto-optimal solutions *Pareto front*.

We can formulate the search for a Pareto-optimal solution as a single-objective problem, for example, by means of the achievement scalarizing function (ASF, [16])

$$u_{\mathbf{w}}(\mathbf{y}) = \max_j w_j (y_j - r_j)$$

where $\mathbf{r} = (r_1, \dots, r_k)$ is a reference point⁴ and $\mathbf{w} = (w_1, \dots, w_k)$ a weight vector. Each Pareto-optimal solution a is associated with a weight vector $\mathbf{w}^a = (\frac{\beta}{a_1 - i_1}, \dots, \frac{\beta}{a_k - i_k})$ such that the minimization of the ASF with weight \mathbf{w}^a yields the solution a , with $\beta > 0$

⁴ Often, the ideal or a utopian point, is used as reference point, i.e., an objective vector typically better than all feasible solutions.

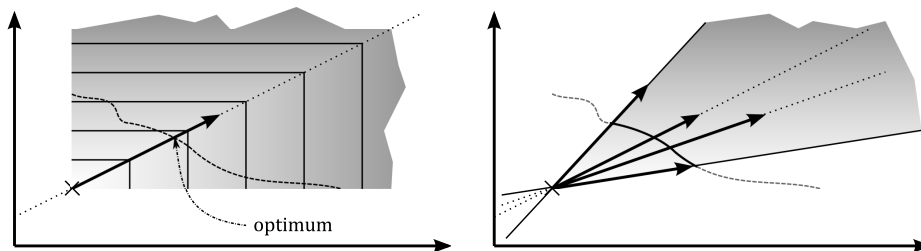


Fig. 1. Left: Illustration of the target direction (dotted line) through the reference point (cross) corresponding to the target vector \mathbf{t} (black arrow) as well as some lines of equal function value $u_{\mathbf{w},\mathbf{t}}(x) = \text{const}$. Lighter colors in the background indicate a smaller value of the achievement scalarizing function. The optimal solution (tip of dash-dotted arrow) lies at the intersection of the target direction with the Pareto front (dashed line). **Right:** Illustration of a target cone (gray area), given by a set of four weight/target vectors (black arrows). The reference point is denoted again as a cross and the attainable part of the Pareto front is depicted as the solid black part of the gray dashed Pareto front.

being a normalization factor such that $\sum_j w_j = 1$. In case that the reference point is dominating a or dominated by a , the weight vector will be strictly positive. For a proof of this case, see for example [16, Theorem 2.3.4] or [8, Lemma 2].

This result about the ASF can be interpreted also graphically. Given a so-called *target direction* $\mathbf{r} + m\mathbf{t}$ through the reference point \mathbf{r} in direction of the *target vector* $\mathbf{t} = (t_1, \dots, t_k)$, the minimization of the ASF $u_{\mathbf{w},\mathbf{t}}(\mathbf{y}) = \max_j w_j^{\mathbf{t}}(y_j - r_j)$ with the weights $\mathbf{w}^{\mathbf{t}} = (\beta/t_1, \dots, \beta/t_k)$ with $\beta = 1/(\sum_j 1/t_j)$ results in the indifference contours shown in the lefthand plot of Fig. 1 and finally in the optimal solution that is lying at the intersection between the target direction and the Pareto front. In the example of Fig. 1, the target vector is $\mathbf{t} = (2, 1)$ with a *slope* of $t_2/t_1 = 1/2$ while the corresponding normalized weight vector is $\mathbf{w} = \frac{1}{1/t_1+1/t_2} \cdot (1/t_1, 1/t_2) = (1/3, 2/3)$.

Definition 1 Given a target vector $\mathbf{t} = (t_1, \dots, t_k)$ or its associated weight vector $\mathbf{w} = (\beta/t_1, \dots, \beta/t_k)$ with $\beta = 1/(\sum_j 1/t_j)$, we call a feasible solution mapped to the minimal value of the ASF $\max_j w_j(y_j - r_j)$ an optimal solution for the target vector \mathbf{t} .

Note that we will, in the following, use interchangeably either the weight vector $\mathbf{w} = (w_1, \dots, w_k)$ or the target direction $\mathbf{t} = (1/w_1, \dots, 1/w_k)$ to define an ASF problem. In the case of bi-objective problems, we also allow weight vectors of the form $(0, 1)$ and $(1, 0)$ for which the corresponding target directions are $(1, 0)$ and $(0, 1)$ respectively. Furthermore, to make the text more readable, we will also denote any vector pointing in the target direction as *target vector*, cf. again Fig. 1.

Definition 2 Given a target direction $\mathbf{r} + m\mathbf{t}$ through the reference point \mathbf{r} , we call any vector $m\mathbf{t}$ mutually different from the null vector (i.e. $m > 0$) a target vector for the corresponding target direction.

In case we are interested in finding more than one Pareto-optimal solution, we have to either perform several independent optimization runs for different weight or target

vectors or we can optimize the achievement scalarizing function for several weight/target vectors simultaneously. The $R2$ indicator [8,11] is based exactly on this concept⁵. In the following, we consider the unary $R2$ indicator of [8] where the weighted Tchebycheff function is replaced by the more general ASF—later results on the optimal distribution of μ solutions also hold for the indicator investigated in [8].

Definition 3 Given a set of N weight vectors $\mathbf{W} = \{\mathbf{w}^1, \dots, \mathbf{w}^N\} \subset \mathbb{R}^k$ and a reference point $\mathbf{r} \in \mathbb{R}^k$, the unary $R2$ indicator assigns the following value to a set $S \subset \mathbb{R}^k$ of solutions

$$R2(S, \mathbf{W}, \mathbf{r}) = \frac{1}{N} \sum_{\mathbf{w} \in \mathbf{W}} \min_{s \in S} \max_j w_j (s_j - r_j) . \quad (1)$$

Note that a smaller $R2$ value corresponds to a better solution set and that, following [8], we only consider the unary $R2$ indicator, which does not require a reference set, here. This indicator is popular in multi-objective optimization due to its runtime in $O(Nk|S|)$ which is linear in the number of weights, the objective dimension, and the number of solutions in the set S . In this context, however, it is still unclear how many weight vectors are required for sufficiently covering the weight space. First suggestions for bi-objective problems can be found in [8], but the scaling with number of objectives k is still an open issue. As the volume of the space increases exponentially with k , it would be plausible that the same holds for N , making the indicator as expensive as the hypervolume. In the case of the $R2$ indicator, and more generally, if more than one target direction is involved in the optimization, we generalize the idea of the target direction to *target cones*, see also the righthand illustration in Fig. 1.

Definition 4 The minimum cone including the target directions $\mathbf{t}^i \in \mathbf{T}$ related to all weight vectors $\mathbf{w}^i \in \mathbf{W}$ is denoted as target cone.

The target cone therefore defines the region of interest defined by the reference point \mathbf{r} and the set of weight vectors \mathbf{W} (or the corresponding set of target vectors \mathbf{T}). For bi-objective problems, we can prove that the solution set which is optimal with respect to the $R2$ indicator lies within the target cone for a suitable set of weight vectors. In order to simplify the proof and to clarify what suitable means, we define what we call a *free target vector* and prove a small technical lemma beforehand.

Definition 5 Given the $R2$ indicator with reference point $\mathbf{r} \in \mathbb{R}^k$ and weight vectors $\mathbf{W} \subset \mathbb{R}^k$, we call a target vector \mathbf{t} corresponding to a weight vector $\mathbf{w}^t \in \mathbf{W}$ free with respect to a solution set $S \subseteq \mathbb{R}^k$ iff S does not contain an optimal solution for \mathbf{t} .

Lemma 1. If a solution $s \in S$ of a solution set S lies outside the target cone and has a positive contribution to the bi-objective $R2$ indicator defined by the reference point $\mathbf{r} \in \mathbb{R}^2$ and a set of weight vectors $\mathbf{W} \subset \mathbb{R}^2$, i.e., if $R2(S, \mathbf{W}, \mathbf{r}) > R2(S \setminus \{s\}, \mathbf{W}, \mathbf{r})$, then s has also a positive contribution to the closest extreme target vector.

⁵ Originally, the indicator was introduced as a binary indicator and for no specific utility function, see [11] for details.

Proof. Let us assume without loss of generality that s is lying to the upper left of the target cone and denote by \mathbf{t}' the leftmost target vector, defined by $\mathbf{w}^{\mathbf{t}'} \in \mathbf{W}$.

Since s lies on the top left of any target vector \mathbf{t} defined by a weight vector $\mathbf{w}^{\mathbf{t}} = (w_1^{\mathbf{t}}, w_2^{\mathbf{t}}) \in \mathbf{W}$, we have that $w_1^{\mathbf{t}}(s_1 - r_1) < w_2^{\mathbf{t}}(s_2 - r_2)$ (I). Moreover, we know that $w_1^{\mathbf{t}'} \geq w_1^{\mathbf{t}}$ and $w_2^{\mathbf{t}'} \leq w_2^{\mathbf{t}}$ holds for any weight vector $\mathbf{w}^{\mathbf{t}} = (w_1^{\mathbf{t}}, w_2^{\mathbf{t}}) \in \mathbf{W}$ because \mathbf{t}' is the leftmost target vector (II).

We prove the lemma by contradiction. Assume that s has a positive contribution to a target vector \mathbf{t} associated with $\mathbf{w}^{\mathbf{t}} = (w_1^{\mathbf{t}}, w_2^{\mathbf{t}}) \neq (w_1^{\mathbf{t}'}, w_2^{\mathbf{t}'})$, i.e., $\max\{w_1^{\mathbf{t}}(s_1 - r_1), w_2^{\mathbf{t}}(s_2 - r_2)\} < \max\{w_1^{\mathbf{t}'}(s_1 - r_1), w_2^{\mathbf{t}'}(s_2 - r_2)\}$ for any other solution $x \in S$. On the other hand, assume also that s has no positive contribution to the leftmost target vector \mathbf{t}' , i.e., that there exists a solution $a \in S \setminus \{s\}$ such that $\max\{w_1^{\mathbf{t}'}(a_1 - r_1), w_2^{\mathbf{t}'}(a_2 - r_2)\} < \max\{w_1^{\mathbf{t}'}(s_1 - r_1), w_2^{\mathbf{t}'}(s_2 - r_2)\}$. Then with (I) and (II), we can show a contradiction to the above assumption that s has a positive contribution to the target vector associated with the weight vector $\mathbf{w}^{\mathbf{t}} = (w_1^{\mathbf{t}}, w_2^{\mathbf{t}})$:

$$\begin{aligned}
& \max\{w_1^{\mathbf{t}'}(a_1 - r_1), w_2^{\mathbf{t}'}(a_2 - r_2)\} < \max\{w_1^{\mathbf{t}'}(s_1 - r_1), w_2^{\mathbf{t}'}(s_2 - r_2)\} \\
& \stackrel{(I)}{\implies} \max\{w_1^{\mathbf{t}'}(a_1 - r_1), w_2^{\mathbf{t}'}(a_2 - r_2)\} < w_2^{\mathbf{t}'}(s_2 - r_2) \\
& \implies w_1^{\mathbf{t}'}(a_1 - r_1) < w_2^{\mathbf{t}'}(s_2 - r_2) \text{ and } w_2^{\mathbf{t}'}(a_2 - r_2) < w_2^{\mathbf{t}'}(s_2 - r_2) \\
& \stackrel{(II)}{\implies} w_1^{\mathbf{t}}(a_1 - r_1) < w_2^{\mathbf{t}}(s_2 - r_2) \text{ and } w_2^{\mathbf{t}}(a_2 - r_2) < w_2^{\mathbf{t}}(s_2 - r_2) \\
& \implies \max\{w_1^{\mathbf{t}}(a_1 - r_1), w_2^{\mathbf{t}}(a_2 - r_2)\} < w_2^{\mathbf{t}}(s_2 - r_2) \\
& \stackrel{(I)}{\implies} \max\{w_1^{\mathbf{t}}(a_1 - r_1), w_2^{\mathbf{t}}(a_2 - r_2)\} \stackrel{!}{<} \max\{w_1^{\mathbf{t}}(s_1 - r_1), w_2^{\mathbf{t}}(s_2 - r_2)\} \quad \square
\end{aligned}$$

Now, we are finally able to prove that in the bi-objective case, the optimal solution set of size μ that minimizes the $R2$ indicator fully lies within the target cone.

Theorem 1. *Given a bi-objective optimization problem and a set of weight vectors $\mathbf{W} \subset \mathbb{R}^2$ for the achievement scalarization functions within the $R2$ indicator with reference point $\mathbf{r} \in \mathbb{R}^2$. Then, all objective vectors of a solution set $S \subseteq \mathbb{R}^k$ with $|S| = \mu$ solutions that minimizes the $R2$ indicator lie within the target cone defined by \mathbf{W} if $\mu \leq |\mathbf{W}|$ and if at least μ target vectors defined by the weight vectors in \mathbf{W} intersect with the Pareto front.*

Proof. For the case $\mu = |\mathbf{W}|$, we refer to the proof of Theorem 1 in [8] and prove the case $\mu < |\mathbf{W}|$ by contradiction. To this end, let us assume that the solution set $S \subseteq \mathbb{R}^k$ minimizes the $R2$ indicator with respect to \mathbf{W} and \mathbf{r} and that the solution $s \in S$ is lying outside the target cone. We distinguish two cases: Either s has no contribution to the $R2$ indicator, i.e., $R2(S, \mathbf{W}, \mathbf{r}) = R2(S \setminus \{s\}, \mathbf{W}, \mathbf{r})$ (case 1) or s has a positive contribution to the $R2$ indicator, i.e., $R2(S, \mathbf{W}, \mathbf{r}) > R2(S \setminus \{s\}, \mathbf{W}, \mathbf{r})$ (case 2).

Case 1: If s itself has no contribution to the $R2$ indicator, we can replace s by an optimal solution s^* of a free target vector \mathbf{t}^* which exists due to the pigeonhole principle (we presupposed at least as many intersections of target vectors with the Pareto front as there are solutions in S). Then, the $R2$ indicator for $(S \setminus \{s\}) \cup \{s^*\}$ is larger than for S which is a contradiction to the assumed optimality of S .

Case 2: Let us now assume that solution s , which lies outside of the target cone, has a positive contribution to the $R2$ indicator. Then, we know from Lemma 1 that s has also a positive contribution to the closest extreme target vector t' of the $R2$ indicator. Hence, t' must be a free target vector and we can replace s by the optimal solution s' with respect to t' and improve the overall $R2$ indicator value which contradicts the assumed optimality of S . \square

Note that in case the number of points in the set S is larger than the number of target vectors, the optimal solution sets of size μ can contain solutions outside the target cone due to solutions with no contribution to the $R2$ indicator [8]. As the above theoretical investigations do only show qualitative results, but do not allow concrete solution sets to be proven to correspond to an optimal $R2$ indicator value, we investigate those optimal solution sets of size μ , also called optimal μ -distributions [2], in the following by means of numerical approximations while changing the location of the reference point, the target cone, and the distribution of the target directions within the cone.

3 Integrating Preferences into the $R2$ Indicator

In this section, the concept of the target cone (Definition 4) is further elaborated. We will discuss how the target cone is modified by changing the position of the reference point and by restricting the weight space covered by the weight vectors in \mathbf{W} . In addition, the effect of the density of the weight vectors in \mathbf{W} on the distribution of target directions within the target cone is discussed to allow the preferences to be further refined.

3.1 Position of the Reference Point

In the previous section, the special role of the reference vector \mathbf{r} as intersection of all target directions became obvious. By moving the reference point, the target cones are moved accordingly. This is shown in Figure 2. By changing the position of the reference point from $\mathbf{r} = (0, 0)^T$ (left) to $\mathbf{r} = (0.2, 0.1)^T$ (right), the focused region on the exemplary Pareto front $f_2 = 0.5 - f_1$ (DTLZ1, [10]) is narrowed significantly.

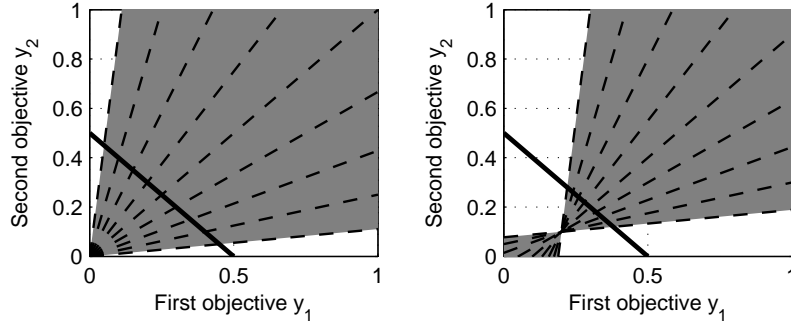


Fig. 2. Moving the target cone (gray area) by changing the position of the reference point from $\mathbf{r} = (0, 0)^T$ (left) to $\mathbf{r} = (0.2, 0.1)^T$ (right). The dashed lines correspond to the target directions and the Pareto front of DTLZ1 is indicated by the thick black line.

3.2 Restriction of the Weight Space

The target cone is defined as the envelope of the target directions (cf. Definition 4). As a consequence, it can be narrowed by restricting the components of the normalized weight vectors $\mathbf{w} \in \mathbf{W}$ to subintervals of $[0, 1]$. This is shown in Figure 3.

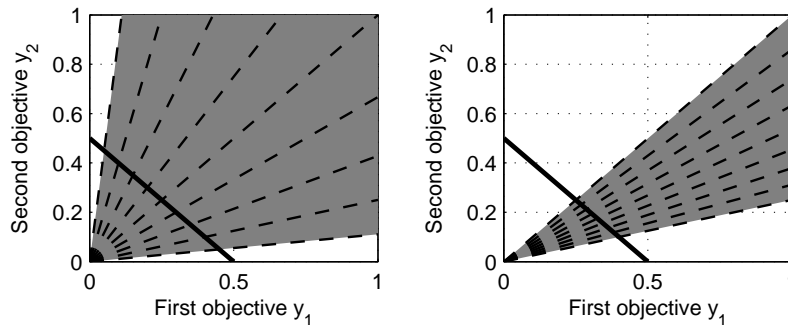


Fig. 3. Narrowing the target cone (gray area) by restricting the first component of the normalized weight vectors from $w_1 \in [0.1, 0.9]$ (left) to $w_1 \in [0.2, 0.5]$ (right). The dashed lines correspond to the target directions and the Pareto front of DTLZ1 is indicated by the thick black line.

3.3 Density of the Weight Vector Distribution

So far, the standard approach is to distribute the weight vectors uniformly within $[0, 1]$, i.e., $w_1 = 0, 0.01, \dots, 0.99, 1$ and $w_2 = 1 - w_1$ for 101 weight vectors, see e. g. [15]. However, the weight vector distribution influences the optimal distribution of solution sets regarding $R2$. In order to obtain additional flexibility, we propose Algorithm 1 as an exemplary method to generate weight vector distributions which express increased preferences regarding the extremes of the front. These kinds of weight vector distributions might be desired due to the fact that the optimal distributions of μ solutions regarding $R2$ result in more centered point distributions than the HV indicator if the weights are chosen uniformly in the weight space [8]. To accomplish this shifted focus, a power transformation with exponent γ is implemented. With increasing $\gamma > 1$, the initial uniform distribution is more and more skewed. To stick with a symmetric distribution, this skewing is only performed on weights $w \leq 0.5$ which are then mirrored along $w = 0.5$ to obtain the weight components $w > 0.5$. This approach on the one hand increases the effect of the transformation, but on the other hand requires a rescaling for still covering the whole domain of weight vectors. The entire rescaling is provided in Algorithm 1. Fig. 4 shows the weight vector distributions resulting for exemplary values of the skewing factor γ .

4 Results

In the remainder of this paper, we investigate experimentally how the previously described ways to incorporate preferences into the $R2$ indicator change its bias. More

Algorithm 1 Generate Weight Vectors

```

function GENERATEWEIGHTVECTORS( $\gamma, n$ )
     $\triangleright \gamma$ : skew factor,  $n$ : number of vectors, must be odd
    weights.x  $\leftarrow$  sequence(0, 0.5, stepsize= 1/( $n - 1$ ))  $\triangleright$  distribute uniformly
    weights.x  $\leftarrow$  0.5 $^{-\gamma+1}$ (weights.x) $^\gamma$   $\triangleright$  skew and rescale weights
    diffs  $\leftarrow$  1 - reverse(weights.x[1, . . . , ( $n - 1$ )/2])  $\triangleright$  Mirroring differences along  $w = 0.5$ 
    weights.x  $\leftarrow$  concatenate(weights.x, diffs)  $\triangleright$  Build symmetric distribution
    return weights.x
end function

```

concretely, we approximate the solution sets of μ points that minimize the $R2$ indicator among all sets of μ points, the so-called optimal μ -distributions, see [2, 8]. Results are obtained by means of standard numerical optimization algorithms for several well-known bi-objective test functions. As we know from theoretical investigations [8], the optimal μ -distributions of the $R2$ indicator lie on the Pareto front if at least μ target vectors of the indicator intersect with the Pareto front. Hence, we are interested in finding the positions of μ points (y_1^i, y_2^i) ($1 \leq i \leq \mu$) on the Pareto front such that the $R2$ indicator of these points is minimal. It is easy to see that this optimization problem is only of dimension μ due to the fact that the points (y_1^i, y_2^i) have to lie on the Pareto front and are therefore dependent variables [2], for instance $y_2^i = 0.5 - y_1^i$ for DTLZ1.

4.1 Position of the Reference Point

For the validation of the analytical thoughts of subsection 3.1, a simple experiment was performed. Using the experimental setup of a former study [8], the position of the reference point was changed and a corresponding optimal μ -distribution for the $R2$ indicator was empirically determined by optimizing the above mentioned y_1^i values with the CMA-ES [3]. Thereby, the standard setup of the recent MATLAB implementation was used [13]. The results of the best of ten replications on the DTLZ1 test function featuring the linear Pareto front $y_2^i = 0.5 - y_1^i$ are shown in Fig. 5 for solution sets of

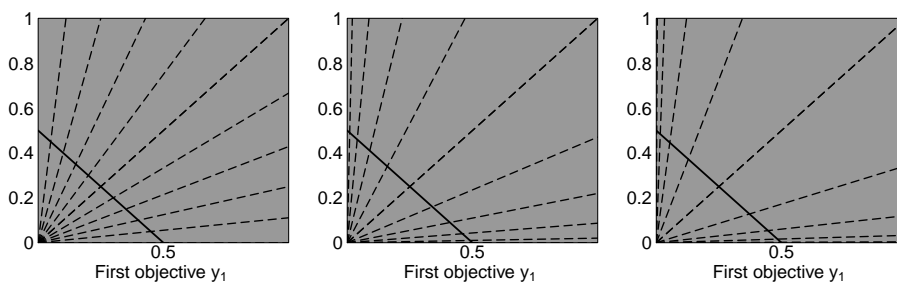


Fig. 4. Different weight vector distributions for eleven weight vectors: Uniform ($\gamma = 1$, left), $\gamma = 2$ (middle) and $\gamma = 3$. The dashed lines correspond to the corresponding target directions. The Pareto front of DTLZ1 is indicated by the solid black line.

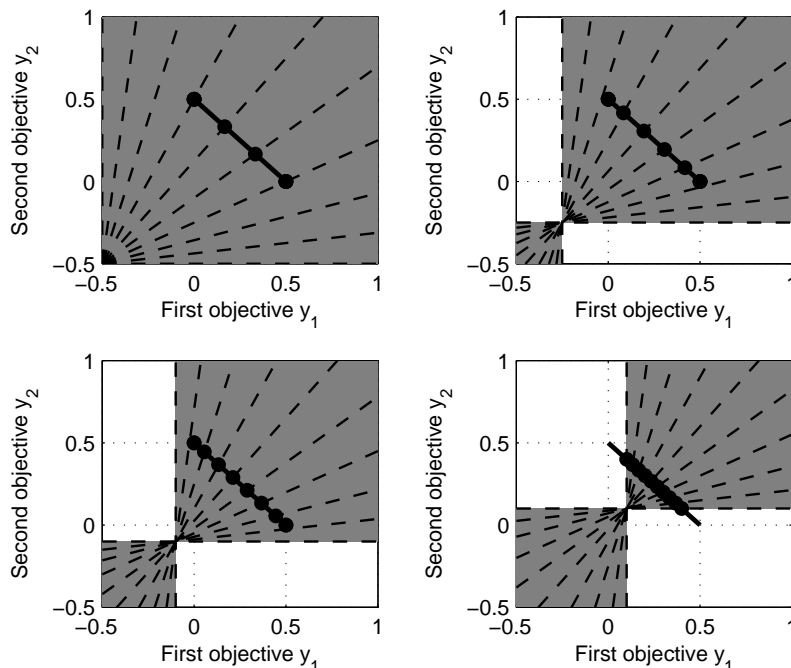


Fig. 5. Experimental results for moving the target cone (gray area) defined by 10 uniformly distributed weight vectors by changing the position of the reference point. Shown are $\mathbf{r} = (-0.5, -0.5)^T$ (upper left), $\mathbf{r} = (-0.25, -0.25)^T$ (upper right), $\mathbf{r} = (-0.1, -0.1)^T$ (bottom left), and $\mathbf{r} = (0.1, 0.1)^T$ (bottom right). The dashed lines correspond to the target directions. The best result of the CMA-ES optimization for $\mu = 10$ solutions is depicted using black dots. The Pareto front of DTLZ1 is indicated by the thick black line.

size $\mu = 10$ and $N = 10$ target directions. As the findings with respect to the movement of the target cone and the distribution of the individuals are general, the choice of the linear front does not represent a restriction. For other fronts with more complex shapes, qualitatively the same results can be expected.

The optimal solutions are located at the intersections between the target directions and the Pareto front. Although the positions of $\mu - N$ solutions are not uniquely determined from a theoretical point of view in the case when less than μ target directions intersect with the Pareto front [8], the used initialization of the CMA-ES result in clusters of solutions at the extremes. As a consequence, one can only distinguish as many different points in the optimal μ -distributions of Fig. 5 as there are intersections between target directions and the Pareto front. Hence, the number of inner mutual solutions decreases with increasing distance to the reference point. If the reference point, however, is moved inside the interval of component values within the front, a focus on the respective region can be realized (bottom right plot of Fig. 5).

The fact that not all μ solutions are uniquely defined in the above examples with $N = \mu = 10$ target vectors, i.e., that they potentially are even dominated in the opti-

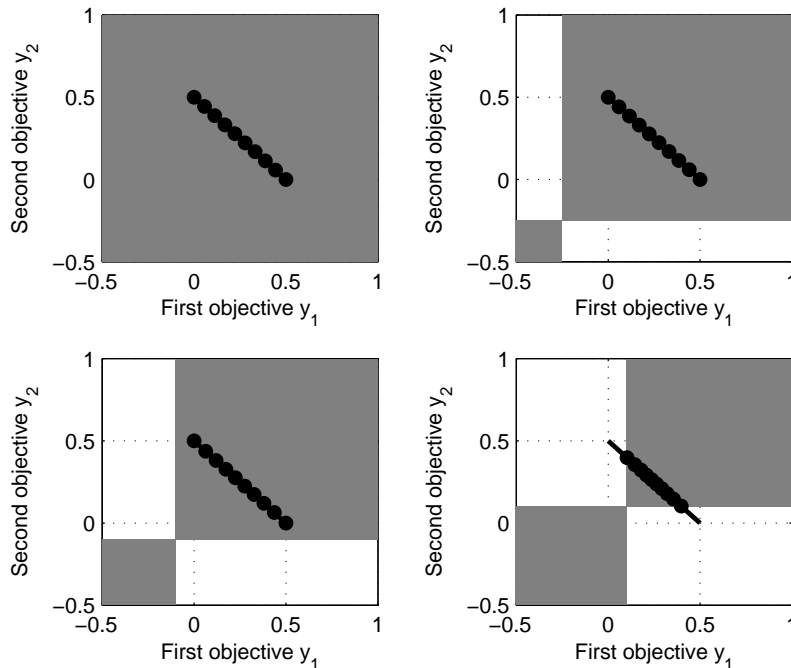


Fig. 6. Experimental results for moving the target cone (gray area) defined by 1000 uniformly distributed weight vectors by changing the position of the reference point. Shown are $\mathbf{r} = (-0.5, -0.5)^T$ (upper left), $\mathbf{r} = (-0.25, -0.25)^T$ (upper right), $\mathbf{r} = (-0.1, -0.1)^T$ (bottom left), and $\mathbf{r} = (0.1, 0.1)^T$ (bottom right). The best result of the CMA-ES optimization for $\mu = 10$ solutions is depicted using black dots. The Pareto front of DTLZ1 is indicated by the thick black line.

mal μ -distributions, is usually not desired, as in these cases, not all solutions provide additional information about the shape of the Pareto front. In particular, if the true ideal point of the front is not known a priori, it is hard to specify an appropriate reference point allowing the whole front to be covered. A simple solution to this problem is to increase the number of weight vectors. In this case, there are enough intersecting points to locate all individuals on different locations of the Pareto front. This is shown in Figure 6 which only differs from Figure 5 by an increase of the number of weight vectors from $N = 10$ to $N = 1000$. The three cases where the reference point is outside the interval of component values within the front are now resulting in almost the same distribution of individuals.

4.2 Density of the Weight Vector Distribution

Further experiments were conducted in order to experimentally check how the optimal μ -distributions regarding $R2$ are influenced by varying the underlying weight vector distribution. More specifically, it is probable that by shifting the density of the weight

vectors away from the center of the Pareto front, the optimal μ -distributions for the $R2$ indicator become more similar to the respective ones for the HV indicator based on former results presented in [8]. Algorithm 1 forms the basis for generating the required weight vector distributions using $\gamma \in \{1, 1.5, 2, \dots, 4.5, 5\}$.

The size μ of the focused solution set is chosen as 10 to be in line with preceding studies and to ensure a meaningful visualization of the results. The number of weight vectors is set to 501 as former experiments in [8] indicated this as a threshold for generating sufficiently robust optimization results. Three different test functions representing different kinds of Pareto front shapes are addressed, i.e., ZDT1 (convex, [19]), DTLZ1 (linear, [10]) and DTLZ2 (concave, [10]). The definitions of the Pareto fronts are $y_2^i = 1 - \sqrt{y_1^i}$ (ZDT1), $y_2^i = 0.5 - y_1^i$ (DTLZ1), and $y_2^i = \sqrt{1 - (y_1^i)^2}$ (DTLZ2). On each of these test problems, we approximated the optimal μ -distributions using the reference point $(0, 0)^T$. To accomplish this, ten CMA-ES runs with a population size of five and an offspring size of 10 as recommended in [12] enhanced by one L-BFGS-B [9] run were conducted. As starting population, the optimal μ -distributions regarding $R2$ based on uniformly distributed weight vectors [8] were used.

The experimental results are presented in Fig. 7. In order to allow for a visual comparison, the optimal μ -distribution for the hypervolume indicator [1] for $\mu = 10$ solutions is depicted at the top of each figure. It can be clearly seen that the optimal positions of the resulting solutions follow the shifts induced to the weight vectors by increasing γ . For a coarser density of weight vectors at the center of the front, the positions of the optimized solutions indicate movements towards the extremes of the front. Even distributions similar to the HV-optimal ones result which is in line with our expectations. The perturbations in the trends for increasing γ are due to the high problem difficulty for the solvers which sporadically produce slightly suboptimal results, especially for DTLZ2.

4.3 Restriction of the Weight Space

Additionally, experiments were carried out to visualize the effect indicated by Theorem 1 that all solutions of the optimal μ -distributions regarding $R2$ lie within the target cone defined by weight vectors $\mathbf{w}^i \in \mathbf{W}$. By restricting the weight space to predefined intervals (see Sec. 3.2), the solutions y^i are located in between the intersections of the outmost weight vectors with the front. However, as general theoretical results for more than one separate interval in weight space were not yet derived—the optimal μ -distributions depend on the proximity of these intervals—we experimentally analyzed this situation.

The experimental setup coincides with the respective settings of the previous section with respect to the test functions and optimization algorithms considered. However, in the current setup, the 501 weight vectors are uniformly distributed within predefined intervals. Specifically, the situation of two separate intervals is addressed. Thereby, the number of weight vectors is split equally to both intervals. Regardless of this, the initial populations of the CMA-ES are filled by 10 uniformly spaced points on the y_1 -axis within the interval defined by the two outmost weight vectors over all considered intervals.

Figure 8 presents the corresponding experimental results. In line with Theorem 1, the optimal μ -distributions regarding $R2$ are located within the intervals defined by

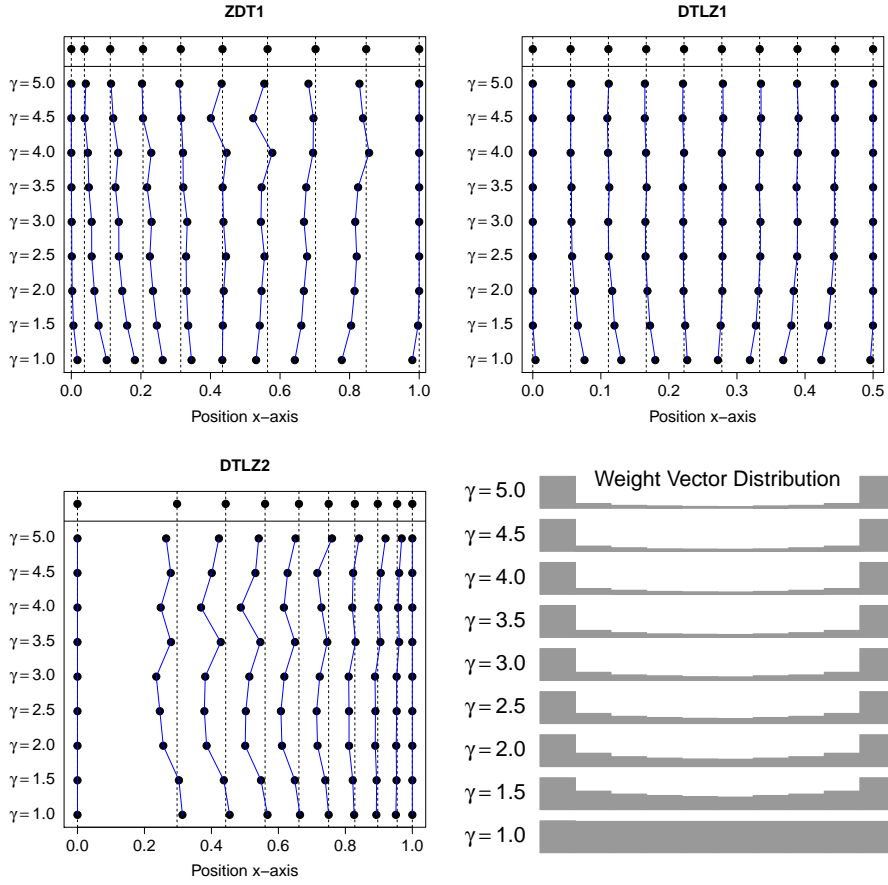


Fig. 7. Experimental results for weight vector distributions with varying density. The best result of the 11 optimization runs is depicted using black dots. The parameter γ of the underlying weight vector distributions is given on the y -axes of the figures. At the top line, the corresponding optimal μ -distribution for the HV is shown for reference. The respective weight vector distributions for different parameter levels γ are visualized using histograms in the lower right plot.

the target cones resulting from the restriction of the weight space. Additionally, as the two separate intervals considered are not close to each other, solutions are concentrated within the two individual target cones in a defined way. Although the points, especially for DTLZ2 and ZDT1, appear to be distributed quite skewed towards the edges of the displayed intervals, the corresponding Euclidean distances in two-dimensional space are much more homogeneous. This is due to the curvature characteristics of the respective fronts with either very small or very high gradients towards the extremes in each dimension and the fact that the plots, as shown here, present only the projections of the solutions onto the y_1 -axis of the first objective function.

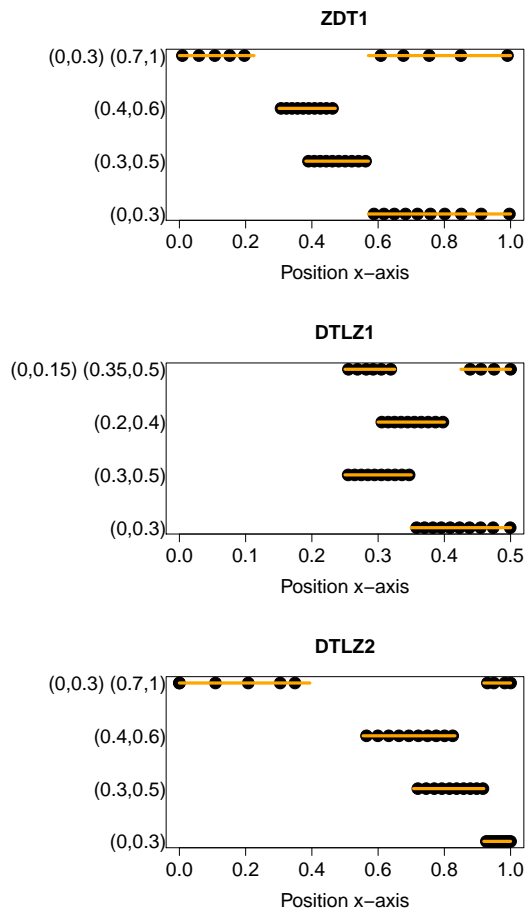


Fig. 8. Experimental results for restricted weight spaces. The best approximation regarding $R2$ out of 11 runs is visualized by black dots. 501 weight vectors are uniformly distributed in the region(s) given on the y -axis of the plot. In case two intervals are considered, the number of weight vectors is split equally to both intervals. The resulting interval(s) on the x -axis bounded by the intersections of the outmost weight vectors with the Pareto front are visualized in orange.

5 Conclusions

In this paper, three variants for integrating preferences into the $R2$ indicator were introduced. These variants exploit that the optimal μ -distributions for the $R2$ -indicator are affected by moving the reference point, by restricting the weight space, as well as by skewing the weight vector distribution. In addition, a sound theoretical background for the first two variants was provided by proving a theorem which relates the location of the optimal μ distribution to the target cone of the weight vectors of the $R2$ indicator. Moreover, experiments were conducted to visually illustrate that the choice

and the target cone of the weight vector distribution heavily influences the optimal μ -distribution for the $R2$ indicator. This is of particular interest, as the standard approach is still to uniformly distribute the weight vectors over their complete domain without being aware of the resulting implications. While the results on the one hand demand for a cautious choice of the weight vector distribution, it thereby becomes possible to take into account preferences regarding the distribution of the points on the Pareto front approximation.

The reference point, as a parameter of $R2$, was found to have an influence which is not negligible. Especially, in case it is chosen too close to the Pareto front, i.e., in case it does not dominate the whole Pareto front, the optimal μ -distributions regarding $R2$ will never cover the whole extent of the Pareto front, independent from the number of weight vectors. On the other hand, the number of weight vectors which intersect the Pareto front decreases with increasing distance to the front. Thus, the recommendation is to apply a very conservative reference point (far better than the approximated Pareto front) and to use more weight vectors the higher the uncertainty of the location of the actual ideal point. This increases the probability that a sufficiently high number of weight vectors intersects the complete Pareto front.

Regarding future work, several research directions are worth to be explored. In addition to further quantitative theoretical analyses of concrete optimal μ -distributions for the $R2$ indicator in the bi-objective case, it will be important to investigate the optimal distributions also for problems with more than two objective functions or other related indicator such as the one in [7]. Furthermore, the question arises how the optimal μ -distributions for the (preference-based) $R2$ indicator can be actually obtained in practice. A simple $R2$ -indicator-based selection within an evolutionary multi-objective optimization algorithm similar to known HV-indicator-based algorithms, such as [4, 5], would be a first step towards this goal.

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 and the project B4 of the Collaborative Research Center SFB 823, which are kindly supported by the Deutsche Forschungsgemeinschaft (DFG). 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”.

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. A. Auger and N. Hansen. A Restart CMA Evolution Strategy With Increasing Population Size. In *Congress on Evolutionary Computation (CEC 2005)*, pages 1769–1776. IEEE Press, 2005.

4. J. Bader and E. Zitzler. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. *Evolutionary Computation*, 19(1):45–76, 2011.
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. N. Beume, B. Naujoks, M. Preuss, G. Rudolph, and T. Wagner. Effects of 1-Greedy \mathcal{S} -Metric-Selection on Innumerably Large Pareto Fronts. In M. Ehrgott et al., editors, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, volume 5467 of *LNCS*, pages 21–35. Springer, 2009.
7. B. Bozkurt, J. W. Fowler, E. S. Gel, B. Kim, M. Köksalan, and J. Wallenius. Quantitative Comparison of Approximate Solution Sets for Multicriteria Optimization Problems with Weighted Tchebycheff Preference Function. *Operations Research*, 58(3):650–659, 2010.
8. D. Brockhoff, T. Wagner, and H. Trautmann. On the Properties of the $R2$ Indicator. In *Genetic and Evolutionary Computation Conference (GECCO 2012)*, pages 465–472, 2012.
9. R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16:1190–1208, 1994.
10. 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.
11. M. P. Hansen and A. Jaszkwicz. 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.
12. N. Hansen. The cma evolution strategy: A comparing review. In J. Lozano, P. Larraaga, I. Inza, and E. Bengoetxea, editors, *Towards a New Evolutionary Computation*, volume 192 of *Studies in Fuzziness and Soft Computing*, pages 75–102. Springer, 2006.
13. N. Hansen. *CMA Evolution Strategy Source Code*, 2012. http://www.lri.fr/~hansen/cmaes_inmatlab.html.
14. C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, 2007.
15. J. Knowles. ParEGO: A Hybrid Algorithm With On-Line Landscape Approximation for Expensive Multiobjective Optimization Problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66, 2005.
16. K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, MA, USA, 1999.
17. Q. Zhang and H. Li. MOEA/D: A multi-objective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.
18. E. Zitzler. *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, ETH Zurich, Switzerland, 1999.
19. E. Zitzler, K. Deb, and L. Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, 2000.
20. E. Zitzler, J. Knowles, and L. Thiele. Quality Assessment of Pareto Set Approximations. In *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 373–404. Springer, 2008.
21. 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.
22. 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.