On the Construction of Pareto-Compliant Combined Indicators

J. G. Falcón-Cardona jfalcon@computacion.cs.cinvestav.mx
Computer Science Department, CINVESTAV-IPN, Mexico City, 07360, Mexico

M. T. M. Emmerich m.t.m.emmerich@liacs.leidenuniv.nl
LIACS, Leiden University, Leiden, 2333, The Netherlands

C. A. Coello Coello ccoello@cs.cinvestav.mx
Computer Science Department, CINVESTAV-IPN, Mexico City, 07360, Mexico
Basque Center for Applied Mathematics (BCAM) & Ikerbasque, Spain

https://doi.org/10.1162/evco_a_00307

Abstract

The most relevant property that a quality indicator (QI) is expected to have is Pareto compliance, which means that every time an approximation set strictly dominates another in a Pareto sense, the indicator must reflect this. The hypervolume indicator and its variants are the only unary QIs known to be Pareto-compliant but there are many commonly used weakly Pareto-compliant indicators such as R2, IGD^+, and ϵ^+. Currently, an open research area is related to finding new Pareto-compliant indicators whose preferences are different from those of the hypervolume indicator. In this article, we propose a theoretical basis to combine existing weakly Pareto-compliant indicators with at least one being Pareto-compliant, such that the resulting combined indicator is Pareto-compliant as well. Most importantly, we show that the combination of Pareto-compliant QIs with weakly Pareto-compliant indicators leads to indicators that inherit properties of the weakly compliant indicators in terms of optimal point distributions. The consequences of these new combined indicators are threefold: (1) to increase the variety of available Pareto-compliant QIs by correcting weakly Pareto-compliant indicators, (2) to introduce a general framework for the combination of QIs, and (3) to generate new selection mechanisms for multiobjective evolutionary algorithms where it is possible to achieve/adjust desired distributions on the Pareto front.

Keywords

Performance indicators, Pareto compliance, multiobjective optimization indicator-based selection.

1 Introduction

The quality assessment of Pareto front approximations[1] (also known as approximation sets) has been a critical factor to compare multi-objective evolutionary algorithms (MOEAs) (Coello Coello et al., 2007). When assessing an approximation set, three quality aspects have been commonly considered: convergence towards the Pareto front, spread, and diversity of solutions (Zitzler et al., 2000). The first evaluation method consisted in qualitative comparisons by plotting the approximation sets (Horn et al., 1994). However, a visual comparison is insufficient when the number of objective functions,

[1] A Pareto front approximation is a finite set of objective vectors that aims to represent a Pareto front.

Manuscript received: 19 December 2019; revised: 7 March 2021, 14 January 2022, 25 January 2022; accepted: 11 February 2022.

Table 1: Relations on approximation sets based on Pareto dominance relations (Zitzler et al., 2003; Guerreiro and Fonseca, 2020).

Relation	Description	Name
$\mathcal{A} \preceq \mathcal{B}$	For every element $\vec{b} \in \mathcal{B}$, there is at least an element $\vec{a} \in \mathcal{A}$ such that $\vec{a} \preceq \vec{b}$.	Weak dominance
$\mathcal{A} \lhd \mathcal{B}$	$\mathcal{A} \preceq \mathcal{B}$ and $\mathcal{B} \npreceq \mathcal{A}$.	Strict dominance
$\mathcal{A} \prec \mathcal{B}$	$\mathcal{A} \neq \emptyset$ and for every element $\vec{b} \in \mathcal{B}$, there is at least an element $\vec{a} \in \mathcal{A}$ such that $\vec{a} \prec \vec{b}$.	Strict elementwise dominance
$\mathcal{A} \prec\prec \mathcal{B}$	$\mathcal{A} \neq \emptyset$ and for every element $\vec{b} \in \mathcal{B}$, there is at least an element $\vec{a} \in \mathcal{A}$ such that $\vec{a} \prec\prec \vec{b}$.	Strong dominance

MOEAs, and the cardinality of the approximation sets increases. To overcome this issue, researchers extended the Pareto dominance relation and its variants (which give a general notion of optimality) to be applied on approximation sets. In this regard, given two objective vectors $\vec{u}, \vec{v} \in \mathbb{R}^m$, $\vec{u}$ Pareto dominates $\vec{v}$ (denoted as $\vec{u} \prec \vec{v}$) if and only if $u_i \leq v_i$ for all $i = 1, 2, \ldots, m$ and there exists at least an index $j \in \{1, 2, \ldots, m\}$ such that $u_j < v_j$. In case that $u_i \leq v_i$ for all i, it is said that $\vec{u}$ weakly Pareto dominates $\vec{v}$ (denoted as $\vec{u} \preceq \vec{v}$). If $u_i < v_i$ for all i, $\vec{u}$ strongly Pareto dominates $\vec{v}$ (denoted as $\vec{u} \prec\prec \vec{v}$). The extended Pareto dominance relations are shown in Table 1. An important drawback of these set-based binary relations is their impossibility to take into account the spread and diversity of solutions. To alleviate the issues of the two previous comparison methods, quality indicators (QIs) were proposed as a quantitative methodology focused on measuring the three main quality aspects previously indicated (Veldhuizen, 1999; Zitzler et al., 2003; Jiang et al., 2014; Liefooghe and Derbel, 2016; Li and Yao, 2019).

Quality indicators are set functions that assign a real value to one or more approximation sets simultaneously, according to specific preference information (Zitzler et al., 2003; Li and Yao, 2019). Mathematically, an l-ary QI is a function $I : \mathcal{A}_1 \times \cdots \times \mathcal{A}_l \to \mathbb{R}$, where each $\mathcal{A}_j \subset \mathbb{R}^m$, $j = 1, \ldots, l$ is a non-empty approximation set. Due to its mathematical definition, they impose a total order on the set Ψ of all approximation sets related to a multiobjective optimization problem. Hence, this property makes QIs a remarkable option to compare the performance of MOEAs. In the specialized literature, there are several QIs that aim to assess convergence, spread and uniformity of approximation sets (Li and Yao, 2019). QIs focused on measuring convergence have a noteworthy relevance since they have been used to assess the performance of MOEAs and also to design their selection mechanisms (Falcón-Cardona and Coello Coello, 2020). Regarding the assessment of MOEAs, *Pareto compliance* is a critical property of convergence QIs[2] that allows them to reflect the order imposed by the $\lhd$-relation (see Table 1). It is worth noting that throughout the years, the term *Pareto compliance* has been commonly used. However, Hansen and Jaszkiewicz (1998) firstly named this property as *compatibility* with an outperformance relation; Zitzler et al. (2003) denoted these indicators as $\lhd$-*complete*; and, finally, Zitzler, Knowles, et al. (2008) refined the term as *strict monotonicity*. In the following, we define the Pareto compliance and the weak Pareto compliance

[2]From this point onwards, convergence QIs will be denoted just as QIs.

properties, assuming, without loss of generality, that a lower indicator value implies a better quality.

Property 1 (Pareto compliance): *Given two approximation sets $\mathcal{A}$ and $\mathcal{B}$, a unary indicator I is $\lhd$-compliant (Pareto-compliant) if $\mathcal{A} \lhd \mathcal{B} \Rightarrow I(\mathcal{A}) < I(\mathcal{B})$.*

Property 2 (Weak Pareto compliance): *Given two approximation sets $\mathcal{A}$ and $\mathcal{B}$, a unary indicator I is weakly $\lhd$-compliant (weakly Pareto-compliant) if $\mathcal{A} \lhd \mathcal{B} \Rightarrow I(\mathcal{A}) \leq I(\mathcal{B})$.*

Pareto compliance implies that every time an approximation set $\mathcal{A}$ strictly dominates another set $\mathcal{B}$, the indicator I will reflect this situation by assigning a lower indicator value to $\mathcal{A}$. In contrast, weak Pareto compliance indicates that $\mathcal{A}$ and $\mathcal{B}$ at least have the same quality. If the QI does not reflect the order imposed by the extended dominance relations, then it is denoted as a non Pareto-compliant indicator. The importance of Pareto-compliant indicators when assessing MOEAs lies in their impossibility of contradicting the structure of order imposed by the $\lhd$-relation (Zitzler et al., 2003). Hence, when comparing MOEAs, Pareto compliance avoids the generation of misleading conclusions regarding the use of Pareto dominance.

Among the plethora of convergence QIs currently available, the hypervolume indicator (HV) is the only unary QI that is Pareto-compliant (Zitzler, 1999; Zitzler et al., 2003, Zitzler, Knowles, et al., 2008). The HV measures the extent of the volume dominated by an approximation set and bounded by a user-supplied reference point that should be dominated by all points in the Pareto front approximation. For nonlinear Pareto front shapes, the set of size μ that approximates the solution to the HV-based subset selection problem presents a non-uniform distribution of objective vectors (Shang, Ishibuchi, He, et al., 2020). The consequences of these non-uniform optimal μ-distributions are twofold: (1) HV penalizes uniform Pareto front approximations in comparison to certain non-uniform approximation sets, and (2) MOEAs that use HV-based selection mechanisms, produce non-uniform approximation sets. To improve the uniformity of the optimal μ-distributions of HV, the weighted HV (Zitzler et al., 2007), logarithmic HV (Friedrich et al., 2011), free HV (Emmerich et al., 2014), and the transformation-based HV (Shang, Ishibuchi, Nan, et al., 2020), which are all variants of HV, have been proposed. Moreover, these variants preserve the Pareto compliance property of HV. Additionally, some other QIs have been proposed, having different preferences to those of the HV but being weakly Pareto-compliant or non Pareto-compliant. For instance, the most noteworthy weakly Pareto-compliant QIs are R2 (Brockhoff et al., 2012), the Inverted Generational Distance plus (IGD^+) (Ishibuchi et al., 2015), and the unary additive ϵ indicator (ϵ^+) (Zitzler et al., 2003), while IGD (Coello Coello and Cruz Cortés, 2005) and Generational Distance (Veldhuizen, 1999) are non-Pareto-compliant indicators.

Due to the imperative need to propose new Pareto-compliant QIs whose preferences are significantly different to those of the HV, we propose in this article a methodology to generate new Pareto-compliant QIs. It is worth noting that in this article we provide both a theoretical and an experimental extension of the work introduced by Falcón-Cardona et al. (2019). Under this methodology, we combine one or more weakly Pareto-compliant indicators with at least one Pareto-compliant QI, using an order-preserving combination function. Under these conditions, we demonstrate that the combined indicators preserve the Pareto compliance property. Additionally, we show through preference analysis and the approximate optimal μ-distributions, that our framework allows to create Pareto-compliant QIs with different preferences to those of the HV in two ways: (1) by exploiting the conflict that sometimes exists between the preferences of indicators, such that the combined indicator shows intermediate

preferences, and (2) by keeping the original preferences of the weakly Pareto-compliant QIs but using a correcting factor derived from the Pareto-compliant QI being used. Overall, the contributions of this article are the following:

1. We provide guidelines for the construction of new Pareto-compliant QIs whose preferences are essentially different from those of the HV.
2. Our proposed framework allows correcting weakly Pareto-compliant QIs, such as R2, IGD^+, and ϵ^+, so that they can become Pareto-compliant.
3. We provide an empirical study of the optimal μ-distribution of solutions generated by a steady-state MOEA based on some (selected) new Pareto-compliant QIs.

The remainder of the article is organized as follows. Section 2 defines the quality indicators that we use throughout the article. The previous related work is described in Section 3. Our mathematical framework for the construction of new Pareto-compliant QIs is introduced in Section 4. The experimental results using the combined indicators is presented in Section 5. Finally, the main conclusions and future work are described in Section 6.

2 Background

In this article, we consider, without loss of generality, multiobjective optimization problems (MOPs) for minimization which are defined as follows (Coello Coello et al., 2007):

$$\min_{\vec{x}\in\Omega} \left\{ f(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \ldots, f_m(\vec{x}))^T \right\},$$

where $\vec{x} = (x_1, \ldots, x_n)^T \in \Omega$ is a vector of n decision variables, and $\Omega \subseteq \mathbb{R}^n$ is the decision space. $f(\vec{x})$ is a vector of $m \geq 2$ objective functions $f_i : \Omega \to \mathbb{R}, \forall i \in \{1, 2, \ldots, m\}$ where some or all of them are mutually in conflict. Since there is no single decision vector $\vec{x}$ whose objective vector $f(\vec{x})$ minimizes all objective functions simultaneously, the goal is to find the so-called Pareto optimal solutions whose images in the objective function space represent the best possible trade-offs among the objective functions. A decision vector $\vec{x} \in \Omega$ is Pareto optimal if there is no other $\vec{y} \in \Omega$ such that $f(\vec{x}) \prec f(\vec{y})$. The set of all Pareto optimal solutions P^* is called Pareto set and its image, given by $PF^* = f(P^*)$, is known as the Pareto front.

In the following, we provide the mathematical definitions of HV, R2, IGD^+, and ϵ^+, which are extensively used throughout the article. In all cases, let $\mathcal{A}$ denote an approximation set and $\mathcal{Z}$ be a reference set.

Definition 1 (Hypervolume indicator): *Let Λ denote the Lebesgue measure in $\mathbb{R}^m$, the hypervolume indicator (HV) is defined as follows:*

$$\text{HV}(\mathcal{A}, \vec{z}_{ref}) = \Lambda\left(\bigcup_{\vec{a}\in\mathcal{A}} \{\vec{x} \mid \vec{a} \prec \vec{x} \prec \vec{z}_{ref}\}\right),$$

where $\vec{z}_{ref} \in \mathbb{R}^m$ is a reference point which should be dominated by all points in $\mathcal{A}$.

HV measures the extent of volume jointly dominated by the points in $\mathcal{A}$ and bounded by $\vec{z}_{ref}$. Currently, HV and the closely related logarithmic HV (Friedrich et al., 2011), the weighted HV (Auger et al., 2009), the free HV (Emmerich et al., 2014), and the transformation-based HV (Shang, Ishibuchi, Nan, et al., 2020) are the only Pareto-compliant QIs known so far. The two main drawbacks of the HV are the following. First,

under NP $\neq$ P, its computational cost increases super-polynomially with the number of objective functions (Bringmann and Friedrich, 2009, 2010). The other issue is related to $\vec{z}_{ref}$, since the preferences of HV strongly depend on it (Auger et al., 2009; Ishibuchi, Imada, et al., 2017). In other words, the specification of the reference is dependent on the Pareto front shape. It has been shown that the distribution of points is often concentrated on the boundary and in knee-point regions.

Definition 2 (Unary R2 indicator): *The unary R2 indicator is defined as follows:*

$$\mathrm{R2}(\mathcal{A}, W) = \frac{1}{|W|} \sum_{\vec{w} \in W} \min_{\vec{a} \in \mathcal{A}} \{u_{\vec{w}}(a)\},$$

where W is a set of m-dimensional weight vectors and $u_{\vec{w}} : \mathbb{R}^m \to \mathbb{R}$ is a utility function, parameterized by $\vec{w} \in W$, that assigns a real value to each solution vector.

The R2 indicator is a convergence-diversity QI that measures the average minimum utility values of the approximation set with respect to a set of weight vectors (Hansen and Jaszkiewicz, 1998; Brockhoff et al., 2012). Its computational cost is in $\Theta(m|W| \cdot |\mathcal{A}|)$. Unlike the hypervolume indicator, the time complexity of R2 grows linearly with the number of objectives. Its time complexity is, however, proportional to the number of weight vectors,[3] which has to grow exponentially in size, if the number of objectives increases and the same resolution of sampling is desired. A major conceptual difference with respect to the hypervolume indicator is that the R2 indicator does not require an anti-optimal reference point. Instead, it works with an ideal or utopian reference point. In many applications involving, for instance, error or cost minimization, there is a natural choice for an ideal point, but it is difficult to define an anti-optimal reference point. So, in such cases, the R2 indicator could be a better choice than the hypervolume.

A problem, however, arises due to the fact that the R2 indicator is not Pareto-compliant, and it is only weakly Pareto-compliant (Hansen and Jaszkiewicz, 1998; Brockhoff et al., 2012). This makes it possible that a set might have equal R2 indicator values than another set, although it is dominated in the set order, or that sets degenerate if this indicator is used as a guide in a Pareto optimization process. One might argue that these are rare cases, as they always involve shared coordinate values among points, and in most cases, the R2 indicator works well when comparing sets. In fact, in continuous unconstrained optimization, such cases might occur with low probability, but they are relatively likely to occur in continuous optimization and in cases in which box constraints are introduced.

Definition 3 (Inverted Generational Distance plus): *The IGD$^+$, for minimization, is defined as follows:*

$$\mathrm{IGD}^+(\mathcal{A}, \mathcal{Z}) = \frac{1}{|\mathcal{Z}|} \sum_{\vec{z} \in \mathcal{Z}} \min_{\vec{a} \in \mathcal{A}} d^+(\vec{a}, \vec{z}),$$

where $d^+(\vec{a}, \vec{z}) = \sqrt{\sum_{k=1}^{m} (\max\{a_k - z_k, 0\})^2}$.

Ishibuchi et al. (2015) proposed IGD$^+$ as an improvement of the IGD indicator (Coello Coello and Cruz Cortés, 2005). Both QIs measure convergence and diversity of

[3]The Simplex-Lattice-Design method is usually employed to construct the set of weight vectors (Das and Dennis, 1998). Using this method, the number of weight vectors is the following combinatorial number: $N = C_{m-1}^{H+m-1}$, where $H \in \mathbb{N}$ is a user-supplied parameter that determines the number of divisions of the space, and m is the number of objectives.

solutions simultaneously. However, IGD^+ is weakly Pareto-compliant while IGD is not Pareto-compliant (Bezerra et al., 2017). IGD^+ measures the average distance from the reference set to the dominated space of the approximation set. Its computational cost is $\Theta(m|\mathcal{Z}| \cdot |\mathcal{A}|)$. A critical aspect is how to specify the reference set when no information is available about PF^* (Ishibuchi et al., 2014).

Definition 4 (Unary ϵ^+ indicator): *Mathematically, it is defined as follows:*

$$\epsilon^+(\mathcal{A}, \mathcal{Z}) = \max_{\vec{z} \in \mathcal{Z}} \min_{\vec{a} \in \mathcal{A}} \max_{1 \leq i \leq m} \{z_i - a_i\}.$$

The unary ϵ^+-indicator gives the minimum distance by which a Pareto front approximation needs to or can be translated in all dimensions at once in objective space such that a reference set is weakly dominated. In consequence, this QI exclusively measures convergence to PF^* and it is weakly Pareto-compliant. A remarkable aspect is that ϵ^+ does not require any parameters but, as in the case of IGD^+, a reference set has to be supplied. Additionally, ϵ^+ is not very sensitive to local changes in the solutions in $\mathcal{A}$ (Bringmann et al., 2011).

Recently, the combination of the high interest in solving MOPs with more than three objective functions and the expensive calculation of the HV have promoted the utilization of weakly and non-Pareto-compliant QIs in spite of their clear drawbacks when comparing MOEAs (Deb and Jain, 2014; Yuan et al., 2016; Tian et al., 2018; Li et al., 2018). In the following, we analyze, using two examples, why weakly and non-Pareto-compliant QIs are not good enough for comparing MOEAs (Zitzler, Knowles, et al., 2008 also discussed these issues). First, Figure 1 shows two approximation sets $\mathcal{A}$ and $\mathcal{B}$, where $\mathcal{A} \lhd \mathcal{B}$, for which the HV, IGD^+, and IGD[4] values are calculated. It is worth noting that a lower value of IGD^+ and IGD means higher quality in contrast to the HV which aims to maximize the dominated volume. Due to its Pareto compliance, the HV effectively shows that $\mathcal{A}$ strictly dominates $\mathcal{B}$ since $HV(\mathcal{A}, \vec{z}_{ref}) = 0.781875$ and $HV(\mathcal{B}, \vec{z}_{ref}) = 0.671250$. On the other hand, IGD^+ cannot reflect that $\mathcal{A} \lhd \mathcal{B}$ since the same IGD^+ value (0.125) is assigned to both approximation sets. In contrast, IGD determines that $\mathcal{B}$ is better because $IGD(\mathcal{B}, \mathcal{Z}) = 0.125$ and $IGD(\mathcal{A}, \mathcal{Z}) = 0.167705$, even though it is strictly dominated by $\mathcal{A}$. Regarding the second example, let's consider a QI which we will call "Zero indicator". It is defined as $Z : \Psi \to \mathbb{R}$ with $Z \equiv 0$. Clearly, for every $\mathcal{A}, \mathcal{B} \in \Psi$ such that $\mathcal{A} \lhd \mathcal{B}$, it implies $Z(\mathcal{A}) = Z(\mathcal{B})$, i.e., Z is weakly Pareto-compliant. Although indicators such as R2, IGD^+ and ϵ^+ are more complex than Z in a mathematical sense, all of them are only weakly Pareto-compliant as the Zero indicator. Based on the discussed examples, we can see that is not enough to construct weakly and non Pareto-compliant QIs, since they can lead to misleading conclusions when comparing approximation sets (Zitzler, Knowles, et al., 2008). Consequently, this is a motivation to construct new Pareto-compliant QIs.

3 Previous Related Work

In this section, we briefly describe the previous work done in the direction of the employment of multiple QIs. Additionally, we first provide a discussion on the use of different terms to denote the Pareto compliance property.

[4]Given an approximation set $\mathcal{A}$ and a reference set $\mathcal{Z}$, the IGD indicator is defined as follows: $IGD(\mathcal{A}, \mathcal{Z}) = \frac{1}{|\mathcal{Z}|} \left(\sum_{\vec{z} \in \mathcal{Z}} \min_{\vec{a} \in \mathcal{A}} d(\vec{a}, \vec{z})^p \right)^{1/p}$, where d is the Euclidean distance and $p > 0$ is a user-supplied parameter, usually set to 2.

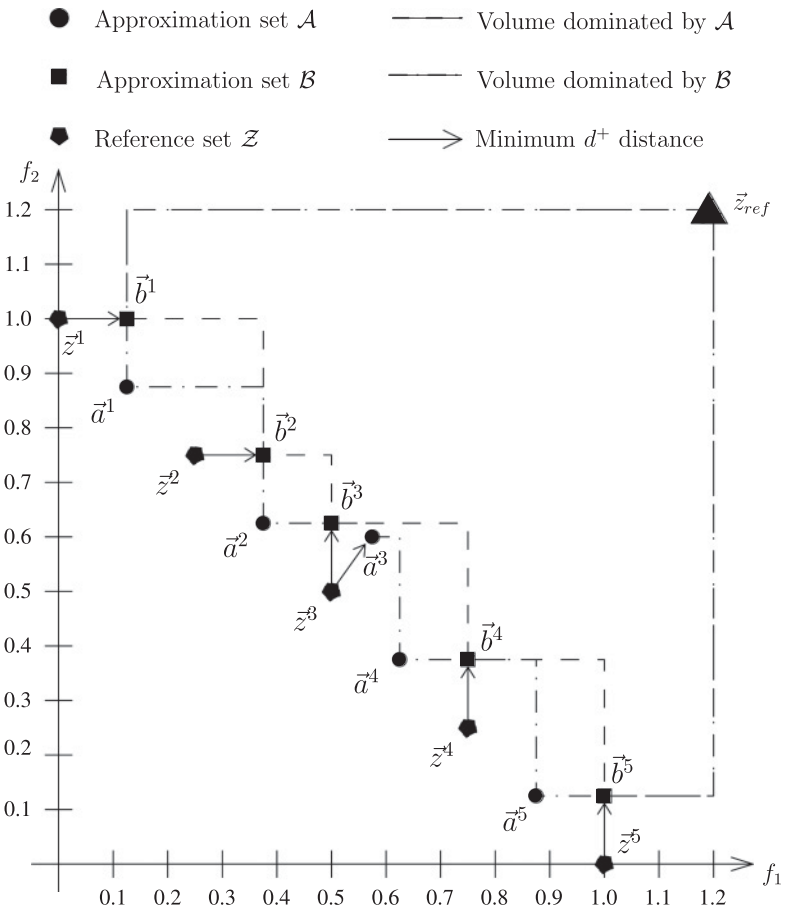

Figure 1: Let $\mathcal{A} = \{(0.125, 0.875), (0.375, 0.625), (0.575, 0.6), (0.625, 0.375), (0.875, 0.125)\}$, $\mathcal{B} = \{(0.125, 1), (0.375, 0.75), (0.5, 0.625), (0.75, 0.375), (1, 0.125)\}$, and $\mathcal{Z} = \{(0, 1), (0.25, 0.75), (0.5, 0.5), (0.75, 0.25), (1, 0)\}$. Even though $\mathcal{A} \lhd \mathcal{B}$, IGD prefers $\mathcal{B}$ and IGD^+ assigns the same quality to both sets, HV prefers $\mathcal{A}$ since it is Pareto-compliant.

3.1 Pareto Compliance

Pareto compliance (defined in Property 1) is a characteristic of convergence QIs which allows them to reflect the order structure imposed by the $\lhd$-relation defined in Table 1. Over the years, this property has been named in different ways, namely, *compatibility*, *completeness*, and *strict monotonicity*. The origin of this property dates back to the work of Hansen and Jaszkiewicz (1998). In that paper, the authors defined the general notion of an *outperformance relation*[5] $\mathcal{R}$ as a set-based binary order relation to compare approximation sets. Based on the general outperformance relation, Hansen and Jaszkiewicz defined the term *compatibility*, which is currently known as Pareto compliance, as follows:

[5]Based on the general outperformance relation, Hansen and Jaszkiewicz defined four relations where three of them were based on Pareto dominance and the remaining one was focused on utility functions (Miettinen, 1999; Pescador-Rojas et al., 2017).

Definition 5: *Compatibility with an outperformance relation (Hansen and Jaszkiewicz, 1998). A comparison method R is compatible with an outperformance relation $\mathcal{R}$ if for each pair of approximation sets $\mathcal{A}$ and $\mathcal{B}$, such that $\mathcal{A}\mathcal{R}\mathcal{B}$, R will evaluate $\mathcal{A}$ as being better than $\mathcal{B}$.*

Definition 6: *Weak compatibility with an outperformance relation (Hansen and Jaszkiewicz, 1998). A comparison method R is weakly compatible with an outperformance relation $\mathcal{R}$ if for each pair of approximation sets $\mathcal{A}$ and $\mathcal{B}$, such that $\mathcal{A}\mathcal{R}\mathcal{B}$, R will evaluate $\mathcal{A}$ as being not worse than $\mathcal{B}$.*

The comparison method R in the above definitions can be replaced by a quality indicator. Hence, assuming that a lower indicator value implies better quality, then $\mathcal{A}\mathcal{R}\mathcal{B} \Rightarrow I(\mathcal{A}) < \mathcal{I}(\mathcal{B})$ and $\mathcal{A}\mathcal{R}\mathcal{B} \Rightarrow I(\mathcal{A}) \leq \mathcal{I}(\mathcal{B})$ imply that I is compatible and weakly compatible, respectively. In case that $\mathcal{R}$ is replaced by $\lhd$, we obtain Properties 1 and 2.

Following the definitions of Hansen and Jaszkiewicz, Zitzler et al. (2003) denoted the compatibility as *completeness* with respect to an arbitrary binary relation $\mathcal{R}$. However, the main diference was the consideration of an indicator vector $\vec{I} = (I_1, \ldots, I_k)$ instead of a single QI. Moreover, the comparison method R was redefined as a function $C_{\vec{I},E} : \mathcal{A} \times \mathcal{B} \rightarrow \{\text{true, false}\}$, where $E : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \{\text{true, false}\}$ is an interpretation function. This newly defined comparison method aims to determine by a Boolean value if $\mathcal{A}$ is better than $\mathcal{B}$. The comparison method $C_{\vec{I},E} = E(\vec{I}(\mathcal{A}), \vec{I}(\mathcal{B}))$ is denoted as $\mathcal{R}$-complete if either for any $\mathcal{A}, \mathcal{B} \in \Psi$, it holds $\mathcal{A}\mathcal{R}\mathcal{B} \Rightarrow C_{\vec{I},E}(\mathcal{A}, \mathcal{B})$. This definition extended the compatibility of Hansen and Jaszkiewicz to consider multiple QIs when comparing approximation sets. We can easily derive the compatibility definition from the completeness one.

Finally, Zitzler, Knowles, et al. (2008) employed the term *strict monotonicity* to denote a QI for which the following holds: $\forall \mathcal{A}, \mathcal{B} \in \Psi : \mathcal{A} \prec \mathcal{B} \Rightarrow I(\mathcal{A}) < I(\mathcal{B})$, where $\prec$ corresponds to the strict elementwise dominance in Table 1. According to Zitzler et al. (2003), $\mathcal{A} \prec \mathcal{B} \Rightarrow \mathcal{A} \lhd \mathcal{B}$, thus, for these subset of approximation sets for which $\prec$ holds, we can employ the $\lhd$-relation as it is stated in the compatibility and completeness definitions.

3.2 Combination of Multiple Indicators

Even though quality indicators have been widely employed by the evolutionary multiobjective optimization community, the exploration of quantitative methods using multiple QIs has been scarce. In this section, we discuss some previous works in which the *combination of multiple indicators* was discussed. Particularly, we focus on the works of Zitzler et al. (2003), Knowles et al. (2006), Zitzler, Knowles, et al. (2008), and Zitzler et al. (2010).

When assessing MOEAs, it is a common strategy to employ several QIs to characterize different quality aspects of approximation sets. In this regard, Zitzler et al. (2003) proposed to analyze such combinations or, more formally, quality indicator vectors $\vec{I} = (I_1, \ldots, I_k)$ to better interpret the results, based on multiple quality aspects, when comparing two approximation sets. To this aim, the authors proposed to use comparison methods based on interpretation functions E (defined in the previous section) to analyze the indicator vectors. Depending on the several possibilities to define E, different claims could be produced when comparing two approximation sets. However, no theoretical properties about $\vec{I}$ or the interpretations functions were provided.

Knowles et al. (2006) briefly explained that the combination of indicators, ideally using Pareto-compliant ones, could lead to powerful interpretations in contrast to employing a single QI. However, no mathematical definition was provided to support this

claim. In a subsequent work, Zitzler, Knowles, et al. (2008) explained that the combination of indicators could allow us to overcome the difficult situation of finding an ideal indicator, that is, a QI being Pareto-compliant, scaling invariant, and cheap to compute. To this aim, one has to look for a way to combine the resulting indicator values. Hence, they were the first to suggest the use of a sequence of indicators to evaluate approximation sets.

Zitzler et al. (2010) mathematically defined a multi-indicator preference relation $\mathcal{R}_S$ that sequentially applies preference relations based on a single quality indicator. $\mathcal{R}_S$ utilizes a sequence $S = (\preceq_{I_1}, \preceq_{I_2}, \ldots, \preceq_{I_k})$, where each $\preceq_{I_j}$ is a preference relation based on the indicator I_j that compares the indicator values of two given approximation sets. The backbone of this proposal is to create a chain of refinements of indicator preferences to compare two given approximation sets and, in this way, break ties (i.e., when a QI cannot decide which approximation is better). Based on this idea, the authors proposed an algorithm for the evaluation of $\mathcal{A}$ and $\mathcal{B}$ similar to the non-dominated sorting algorithm (Deb et al., 2002). First, an index j is set to 1 to use $\preceq_{I_1}$, i.e., the value $I_{j=1}$ is calculated for both $\mathcal{A}$ and $\mathcal{B}$. If $I_j(\mathcal{A}) = I_j(\mathcal{B})$, then j is increased to point to the next indicator-based preference relation in the sequence if it exists. In case that $\preceq_{I_j}$ claims an approximation set with better quality, the decision is returned. In consequence, the use of these indicator-based preference relations could increase the sensitivity of an order relation by solving cases of incomparability. Furthermore, Zitzler et al. proposed that the first $1 \leq t < k$ indicators are weakly Pareto-compliant and the remaining ones are Pareto-compliant. This way, the Pareto-compliant QIs could be employed to refine the preferences of the weakly Pareto-compliant indicators.

It is worth noting at this point what is the main focus of these multi-indicator-based preference relations. Zitzler, Thiele, et al. (2008) and Zitzler et al. (2010) employed these preference relations to construct the Set Preference Algorithm for Multiobjective Optimization (SPAM) that uses populations as individuals instead of single decision vectors. The reason to use populations as individuals is related to the utilization of the indicator-based preference relations that require a population (a set of objective vectors) as an input parameter. Internally, SPAM manages three populations: the main population P, the randomly mutated population P', and the heuristically mutated population P''. P is compared with P' and P'', using the indicator-based preference relations, to determine the one that possibly replaces P. In case of using HV in the sequence of indicators, the high computational cost of HV-based MOEAs such as the $\mathcal{S}$-Metric Selection Evolutionary Multiobjective Algorithm (SMS-EMOA) (Beume et al., 2007) is avoided because it is necessary to calculate only two HV values at each generation if necessary, that is, when HV needs to refine the preferences. However, SPAM needs more function evaluations than a usual MOEA due to the utilization of three populations.

From the above discussion, it is clear that the indicator-based preference relations show remarkable advantages to derive set-based MOEAs such as SPAM. Furthermore, another utilization of these binary relations could be the comparison of the Pareto front approximations generated by different MOEAs. This is possible since the indicator-based preference relations are basically a lexicographic order based on a sequence $S = (\preceq_{I_1}, \preceq_{I_2}, \ldots, \preceq_{I_k})$. Thus, they impose a preorder[6] on the set of all approximation sets. In spite of the possibility of using the indicator-based preference relations to

[6]As long as we are concerned with how approximation sets are ordered, the indicator-based preference relations are not anti-symmestric since different approximation sets may have equal indicator values.

compare MOEAs, there are some issues that should be pointed out. The preference relation on its own, only determines what is the relationship between two given approximation sets but it does not numerically indicate the difference in quality between them. Even though if we inspect the indicator values produced by the preference relation, there is a possibility that some of the comparisons do not employ the same $\preceq_{I_j} \in S$. Hence, we would have performance evaluations in a different scale of measure.

In the next section, we present a generalization of the multi-indicator-based preference relations proposed by Zitzler et al. (2010). Our proposal is a framework to combine multiple QIs in order to define new unary indicators. The resulting combined unary indicator merges the preferences of all the baseline QIs, defining new preferences and a common numerical field of comparisons to determine the difference in quality between Pareto front approximations. Another important aspect is that the user can specify the relative importance of all the baseline QIs in the combination. Furthermore, this framework allows the construction of new Pareto-compliant indicators by combining as many weakly Pareto-compliant indicators as needed, but including at least one Pareto-compliant QI to retain the Pareto compliance property. Due to the use of at least one Pareto-compliant QI, we refine the preferences of the weakly Pareto-compliant indicators in a similar way as the one proposed by Zitzler et al.

4 New Pareto-Compliant Indicators

In this section, we propose a systematic framework for combining QIs, following the basic idea introduced by Falcón-Cardona et al. (2019). Additionally, we provide the mathematical argumentation to ensure that when combining QIs with specific properties, the resulting combined indicator will be Pareto-compliant. This leads not only to new types of indicators but also proves to be a way to create new Pareto-compliant indicators with very different properties than those of the HV in terms of the distribution of points that they favor, and in terms of the parameters provided by the user. In the following, we present the mathematical framework for the combination of QIs. Let $\mathcal{A}$ be an approximation set in Ψ.

Definition 7 (Combination function): *A combination function* $C : \mathbb{R}^k \to \mathbb{R}$ *assigns a real value to a vector* $\vec{I}(\mathcal{A}) = (I_1(\mathcal{A}), I_2(\mathcal{A}), \ldots, I_k(\mathcal{A}))$, *where each* $I_j(\mathcal{A})$ *is the value of the* j^{th} *unary indicator.*

Definition 8 (Combined Indicator): *Given an indicator vector* $\vec{I}(\mathcal{A}) = (I_1(\mathcal{A}), I_2(\mathcal{A}), \ldots, I_k(\mathcal{A}))$ *and a combination function* C, *a combined indicator* $\mathcal{I}(\mathcal{A})$ *is defined as follows:* $\mathcal{I}(\mathcal{A}) = C(\vec{I}(\mathcal{A}))$.

Regarding Definition 8, a combined indicator $\mathcal{I}$ is a QI as well (since $\mathcal{I} : \Psi \to \mathbb{R}$) but it requires the combination function C to transform an indicator vector $\vec{I}(\mathcal{A})$ to a real value. Figure 2 shows how to map Pareto front approximations (in objective space) to the quality space $Q \subseteq \mathbb{R}^k$, where each axis corresponds to a specific indicator. Then, the indicator vectors in Q are assigned a real value, using the combination function C. Based on the above definitions, nothing can be said about the properties of $\mathcal{I}$ at this point. Hence, for getting more important theoretical results, we should say something about the properties of each I_j, $j = 1, \ldots, k$ and the combination function C. We are interested in analyzing the Pareto compliance of $\mathcal{I}$. Based on Properties 1 and 2, we construct a special vector of indicators that is necessary for the refinement of the combined indicator model.

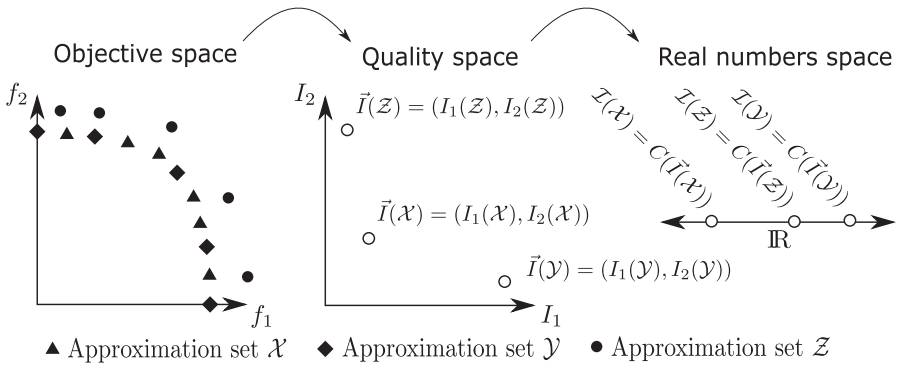

Figure 2: The objective space contains the approximation sets $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ that are mapped to the quality space using an indicator vector. The points $\vec{I}(\mathcal{X})$, $\vec{I}(\mathcal{Y})$, and $\vec{I}(\mathcal{Z})$ in quality space are then transformed each to a single real value by the combination function $C : \mathbb{R}^2 \to \mathbb{R}$ to generate the real values $\mathcal{I}(\mathcal{X})$, $\mathcal{I}(\mathcal{Y})$, and $\mathcal{I}(\mathcal{Z})$.

Definition 9 (Compliant Indicator Vector): *$\vec{I}(\mathcal{A}) = (I_1(\mathcal{A}), I_2(\mathcal{A}), \ldots, I_k(\mathcal{A})) \in Q$ is called a compliant indicator vector (CIV) if $\forall j = 1, \ldots, k$, I_j is weakly Pareto-compliant and there exists at least one index $t \in \{1, \ldots, k\}$ such that I_t is Pareto-compliant. $Q \subseteq \mathbb{R}^k$ is denoted as the quality space.*

For the following theorem, let us assume, without loss of generality, that the unary indicators $I_1, \ldots, I_k$ are to be minimized.

Theorem 1 (Construction of Pareto-compliant combined indicators): *Let $I_1, \ldots, I_k$ be unary indicators that form a compliant indicator vector $\vec{I}$. A combined indicator $\mathcal{I}$, based on the combination function $C(\vec{I})$, is $\lhd$-compliant (Pareto-compliant) if C has the order-preserving property:*

$$\forall \vec{u}, \vec{v} \in Q, \vec{u} \prec \vec{v} \Rightarrow C(\vec{u}) < C(\vec{v}).$$

Proof: Consider two approximation sets $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \lhd \mathcal{B}$ and let $\vec{I}$ be a CIV. Then, $\mathcal{A} \lhd \mathcal{B} \Rightarrow \vec{I}(\mathcal{A}) \prec \vec{I}(\mathcal{B})$ because the Pareto-compliant indicators get better and the weakly Pareto-compliant ones get better or stay equal. Moreover, by definition $\vec{I}(\mathcal{A}) \prec \vec{I}(\mathcal{B}) \Rightarrow C(\vec{I}(\mathcal{A})) < C(\vec{I}(\mathcal{B}))$. Hence, by transitivity of $\Rightarrow$ and considering that $\mathcal{I}(\mathcal{A}) = C(\vec{I}(\mathcal{A}))$ and $\mathcal{I}(\mathcal{B}) = C(\vec{I}(\mathcal{B}))$, it holds $\mathcal{A} \lhd \mathcal{B} \Rightarrow \mathcal{I}(\mathcal{A}) < \mathcal{I}(\mathcal{B})$, that is, $\mathcal{I}$ is Pareto-compliant. □

Theorem 1 provides a sufficient condition for constructing Pareto-compliant combined indicators on the basis of compliant indicator vectors. In other words, a combined indicator preserves the Pareto compliance property because of the use of order-preserving combination functions.

Remark 1: The condition of Theorem 1 is suffcient but not necessary. For instance, given $\vec{I}(\mathcal{A}) = (I_1(\mathcal{A}), I_2(\mathcal{A}), \ldots, I_k(\mathcal{A}))$ where I_1 is Pareto-compliant and the I_j, $j = 2, \ldots, k$ are not Pareto-compliant, the "combined" indicator $\mathcal{I}(\mathcal{A}) = C(\vec{I}(\mathcal{A})) = I_1(\mathcal{A})$ is also Pareto-compliant. Hence, there is a large number of possibilities to construct combined and compliant indicators.

There exist many combination functions that have the property of Theorem 1. In this article, we focus on order-preserving utility functions $u : \mathbb{R}^k \to \mathbb{R}$ (Miettinen, 1999; Pescador-Rojas et al., 2017). A utility function (UF) is a model of the decision maker

preferences that assigns to each k-dimensional vector a utility value. Thus, a combination function C can be defined in terms of these functions. Generally, UFs employ a convex weight vector $\vec{w} \in [0,1]^k$ such that $\sum_{i=1}^{k} w_i = 1$, $w_i \geq 0$. However, for the combination of QIs, we need $w_i > 0$ for all $i = 1, \ldots, k$ such that all QIs contribute to the combined indicator value. Based on the above, we denote $u_{\vec{w}}(\vec{I}(\mathcal{A}))$ as a Pareto-compliant utility indicator (PCUI) if u is order preserving and $\vec{I}$ is a CIV.

5 Experimental Results

Throughout this section, we analyze six newly created PCUIs, emphasizing the distribution of points in different Pareto fronts that they prefer. To define the PCUIs, we focused our attention on two well-known utility functions that are order-preserving, namely, the weighted sum function (WS) and a slightly modified augmented Tchebycheff function (ATCH) (Pescador-Rojas et al., 2017). Let $\vec{I} \in \mathbb{R}^k$ be an indicator vector and $\vec{w} \in (0,1)^k$ be a convex weight vector. In the following, we define WS and ATCH as $\mathrm{WS}_{\vec{w}}(\vec{I}) = \sum_{i=1}^{k} w_i I_i$ and $\mathrm{ATCH}_{\vec{w}}(\vec{I}) = \max_{i=1,\ldots,k}\{w_i I_i\} + \alpha \sum_{i=1}^{k} I_i$, respectively. Regarding ATCH, we modified its original definition by not considering the absolute value of the term $w_i x_i$ such that the function is order-preserving in the whole $\mathbb{R}^k$ and $\alpha > 0$ is a user-supplied parameter. Since a PCUI requires all its baseline QIs to be minimized, we exclusively consider $-$HV in their definition. The proposed PCUIs, defined in the following, correspond to Pareto-compliant versions of the indicators R2, IGD^+, and ϵ^+:

- $\mathrm{WS}_{\vec{w}}(-\mathrm{HV}, \mathrm{R2})$ and $\mathrm{ATCH}_{\vec{w}}(-\mathrm{HV}, \mathrm{R2})$;
- $\mathrm{WS}_{\vec{w}}(-\mathrm{HV}, \mathrm{IGD}^+)$ and $\mathrm{ATCH}_{\vec{w}}(-\mathrm{HV}, \mathrm{IGD}^+)$;
- $\mathrm{WS}_{\vec{w}}(-\mathrm{HV}, \epsilon^+)$ and $\mathrm{ATCH}_{\vec{w}}(-\mathrm{HV}, \epsilon^+)$.

In the following sections, we are interested in understanding the properties of the adopted PCUIs as performance assessment functions to compare MOEAs' performance. To this aim, we proposed two main experiments. The first experiment aims to compare how the PCUIs and their baseline QIs rank several state-of-the-art MOEAs that produce Pareto front approximations with different specific distributions of objective vectors (we employed the Lamé and Mirror superspheres problems proposed by Emmerich and Deutz (2007)). The comparison is based on measuring the correlation of preferences between all the adopted PCUIs and QIs, following the methodology of Liefooghe and Derbel (2016). On the other hand, the second experiment analyzes the approximate optimal μ-distributions related to PCUIs to reinforce the preference information. To obtain the approximate optimal μ-distributions, we set up a steady-state MOEA similar to the $\mathcal{S}$-Metric Selection Evolutionary Multiobjective Algorithm (SMS-EMOA) (Beume et al., 2007), that uses the PCUIs as part of its density estimator. The proposed algorithm, denoted as PCUI-EMOA, is tested on MOPs from the Deb-Thiele-Laumanns-Zitzler (DTLZ) (Deb et al., 2005), Walking-Fish-Group (WFG) (Huband et al., 2006) test suites, and their minus versions, DTLZ^{-1} and WFG^{-1} (Ishibuchi, Setoguchi, et al., 2017), respectively. It is worth noting that we turned off all the search difficulties of these problems to avoid issues related to the PCUI-EMOA's performance.

5.1 Weight Vector Effect

Before studying the properties of the newly created PCUIs, we need to punctualize the relationship between the order-preserving functions and the PCUIs and we also need

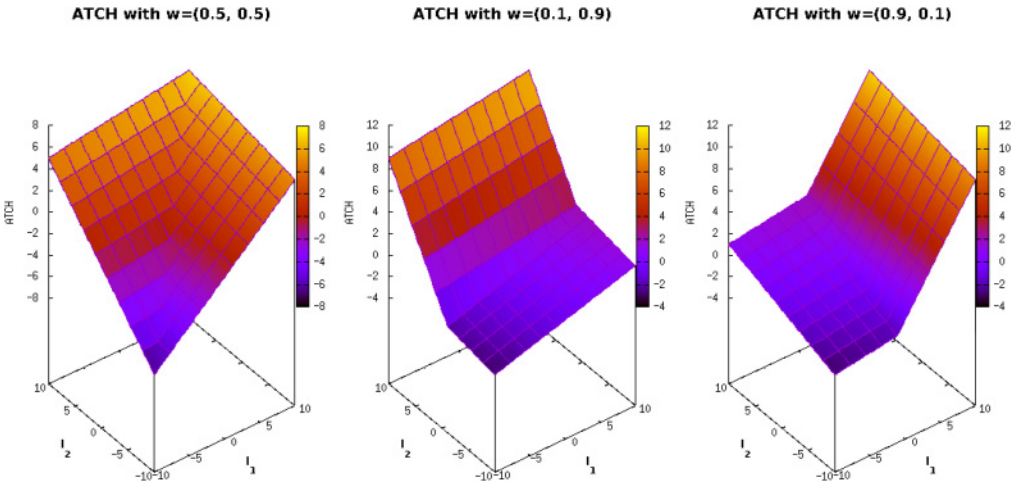

Figure 3: Landscapes of ATCH function varying the weight vector $\vec{w} = (w_1, w_2)$.

to clarify what is the effect of the combination weight vector $\vec{w}$ that the PCUIs require. First, PCUIs are invariant to the indicator scales because of the order-preserving combination function u. Regardless of the order-preserving function being used, it ensures that if $\vec{x} \prec \vec{y}$, then $u(\vec{x}) < u(\vec{y})$. However, if $\vec{x}$ and $\vec{y}$ are mutually nondominated, we cannot say what will be the relation between $u(\vec{x})$ and $u(\vec{y})$ unless we know the definition of u. In consequence, each u expresses specific preferences when dealing with non-dominated solutions and such preferences depend on the landscape of u (see Figure 3). On the other hand, PCUIs require a weight vector $\vec{w} = (w_1, \ldots, w_k)$ where $w_i > 0$ for all $i = 1, \ldots, k$. Each w_j assigns a relative importance to its associated indicator I_j. Hence, this can control the impact of each indicator to the final preferences of the PCUI. Figure 3 shows the landscape of the ATCH function for three different settings of $\vec{w}$. This image shows that depending on the setting of $\vec{w}$, the preferences of the PCUIs may change, that is, the total ordering induced by the order-preserving functions. In specific cases where the preferences of the baseline QIs are in conflict, the weight vector could lead to exploit the trade-off between them. In other words, when all w_i are equal, the PCUI preferences are the intermediate point between the preferences of its baseline QIs. For the other two cases in Figure 3, the preferences of the PCUI will be biased to the indicator having the greatest w_i. If we assume that a PCUI is integrated in the selection mechanism of a MOEA, we could control the final distribution of points by defining $\vec{w}$.

5.2 Analysis of Preferences

In this experiment, we are interested in analyzing how similarly the six adopted PCUIs and their baseline QIs rank different Pareto front approximations corresponding to the Lamé and Mirror superspheres problems with unimodal difficulty (Emmerich and Deutz, 2007). The correlation analysis is based on the methodology proposed by Liefooghe and Derbel (2016) where the main focus is the comparison of *rankings* of approximation sets obtained within each QI. The reason to use the Lamé superspheres is their scalability in the objective space and the controlling of the Pareto front shapes. Regarding the latter, a parameter γ controls the Pareto front geometry. For Lamé problems, $\gamma \in (0, 1)$, $\gamma = 1$, and $\gamma > 1$ correspond to convex, linear, and concave Pareto front

geometries. In case of the Mirror problems, $\gamma \in (0, 1)$ corresponds to a concave geometry while $\gamma > 1$ is related to convex shapes. For both Lamé and Mirror problems, we employed $\gamma = 0.25, 0.50, 0.75, 1.00, 1.50, 2.00, 6.00$ for 2, 3, and 4 objective functions.

Following the med-Q sampling methodology of Liefooghe and Derbel where a black-box MOEA (they employed NSGA-II) produces Pareto front approximations, we selected MOEAs that exhibit particular distribution characteristics to have a representative sample of the set Ψ. For each test instance, 30 independent executions were produced by each MOEA, where all the algorithms shared the same settings. For all the objective functions, the size of all the produced approximation sets was 120. To avoid performance issues of the selected MOEAs, all of them used Pareto optimal solutions as their initial population.[7] Hence, during the execution time, the MOEAs explore these initial solutions to impose their own preferences. The adopted MOEAs are classified in five classes as follows:

- Indicator-based MOEAs: SMS-EMOA (Beume et al., 2007), MOMBI2 (Hernández Gómez and Coello Coello, 2015), IGD^+-MaOEA (Falcón-Cardona and Coello Coello, 2018), and Δ_p-MaOEA[8].
- Pareto-based MOEAs: NSGA-II (Deb et al., 2002) and SPEA2 (Zitzler et al., 2001).
- Reference set-based MOEAs: NSGA-III (Deb and Jain, 2014).
- Decomposition-based MOEAs: MOEA/D (Zhang and Li, 2007).
- Image analysis-based MOEAs: MOVAP (Hernández Gómez et al., 2016).

Regarding the assessment of the Pareto front approximations generated by the adopted MOEAs, we used the following settings on the QIs and PCUIs. We set $\vec{z}_{ref} = (1.12, \ldots, 1.12)$ to calculate HV for all test instances. A set of convex weight vectors (constructed by the Simplex-Lattice-Design method; Das and Dennis, 1998) was employed as the set W for R2. The reference sets for IGD^+ and ϵ^+ are constructed by collecting all the Pareto front approximations and, then, obtaining a set of 120 nondominated solutions using a subset selection based on the Riesz s-energy as proposed by Falcón-Cardona et al. (2020). For all the six adopted PCUIs, the weight vector was set as $\vec{w} = (0.0001, 0.9999)$, where 0.0001 is the weight associated with the HV and 0.9999 is related to the weakly Pareto-compliant QI. This setting was employed to mostly preserve the preferences of the weakly Pareto-compliant QIs while producing Pareto-compliant results due to the use of the HV as a correction factor.

We aim to correlate the rankings of MOEAs within each indicator, that is, by how much the PCUIs and QIs rank the MOEAs (i.e., the characteristic Pareto front approximations) similarly. For each test instance and QI, the MOEAs are ranked by their mean indicator value (as proposed by Liefooghe and Derbel (2016), using the med-Q sampling). The ranks of MOEAs are then analyzed for correlation with the remaining QIs

[7]To generate the initial populations, we executed the algorithms several times to gather in and archive the best-found solutions. Then, each time an algorithm runs, it randomly selects a subset of the archive as its initial population.

[8]We proposed this algorithm based on the framework of IGD^+-MaOEA (Falcón-Cardona and Coello Coello, 2018) but using the Δ_p indicator as the density estimator.

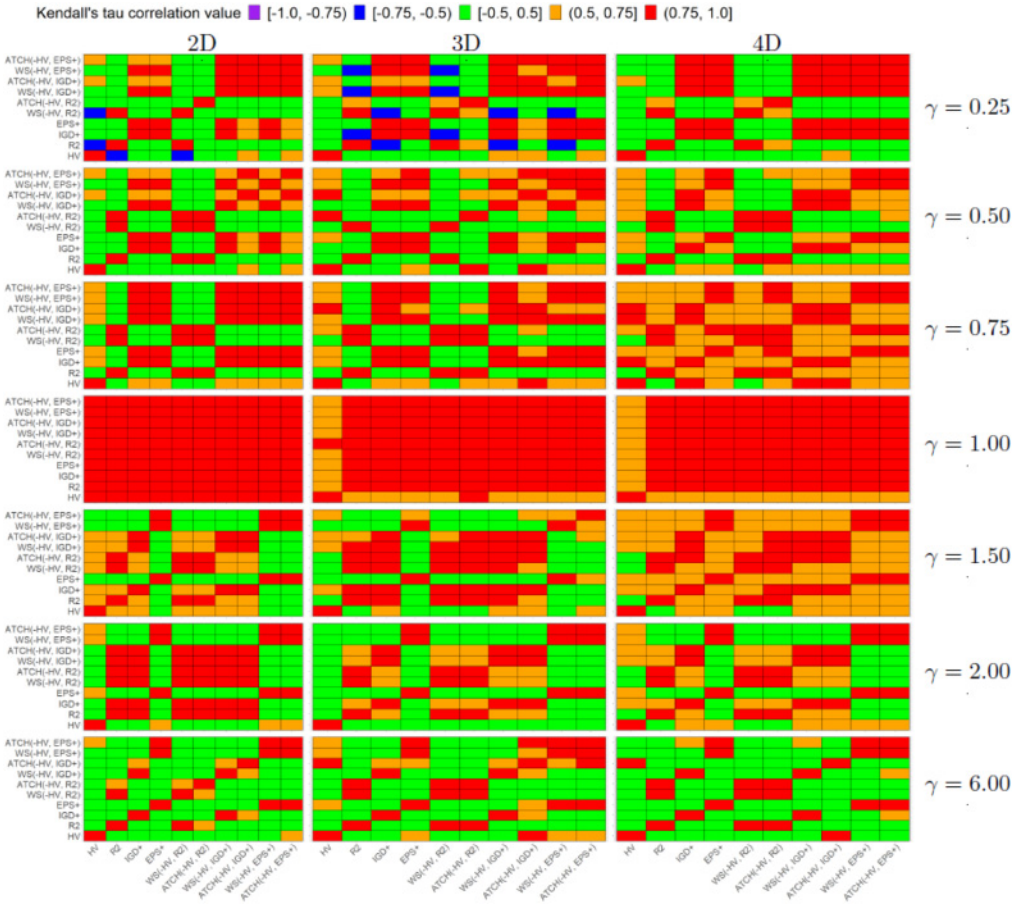

Figure 4: Heatmap Kendall rank correlation τ for each pair of quality indicators and each Lamé problem on different dimensions of the objective space.

using the Kendall's τ correlation coefficient. It is worth emphasizing that Kendall's τ quantifies the difference between the proportion of concordant and discordant pairs among all possible pairwise MOEAs. Since $\tau \in [-1, 1]$, where $\tau = -1$ means perfect disagreement and $\tau = 1$ means perfect agreement of ranks, we decided to create intervals of τ values in order to represent them using Heatmaps. Such intervals are the following: $[-1, -0.75)$, $[-0.75, -0.5)$, $[-0.5, 0.5]$, $(0.5, 0.75]$, and $(0.75, 1]$. It is worth noting that we consider $\tau \in [-0.5, 0.5]$ as a result of no correlation between two different rankings. Figures 4 and 5 show the τ values based on the intervals, using heatmaps for all the adopted Lamé and Mirror test instances, respectively.

5.2.1 Correlation between PCUIs and Baseline QIs

Regarding the correlation analysis on Lamé problems in Figure 4, we have the following conclusions. For problems with $\gamma = 0.75, 1.00, 1.50$, the PCUIs are consistently correlated with their baseline QIs (the correlation is stronger with the weakly Pareto-compliant indicator), regardless of the dimension of the objective space. Regarding $\gamma = 1.00$, the Kendall's τ is in $(0.75, 1.0]$ in almost all comparisons although for three- and

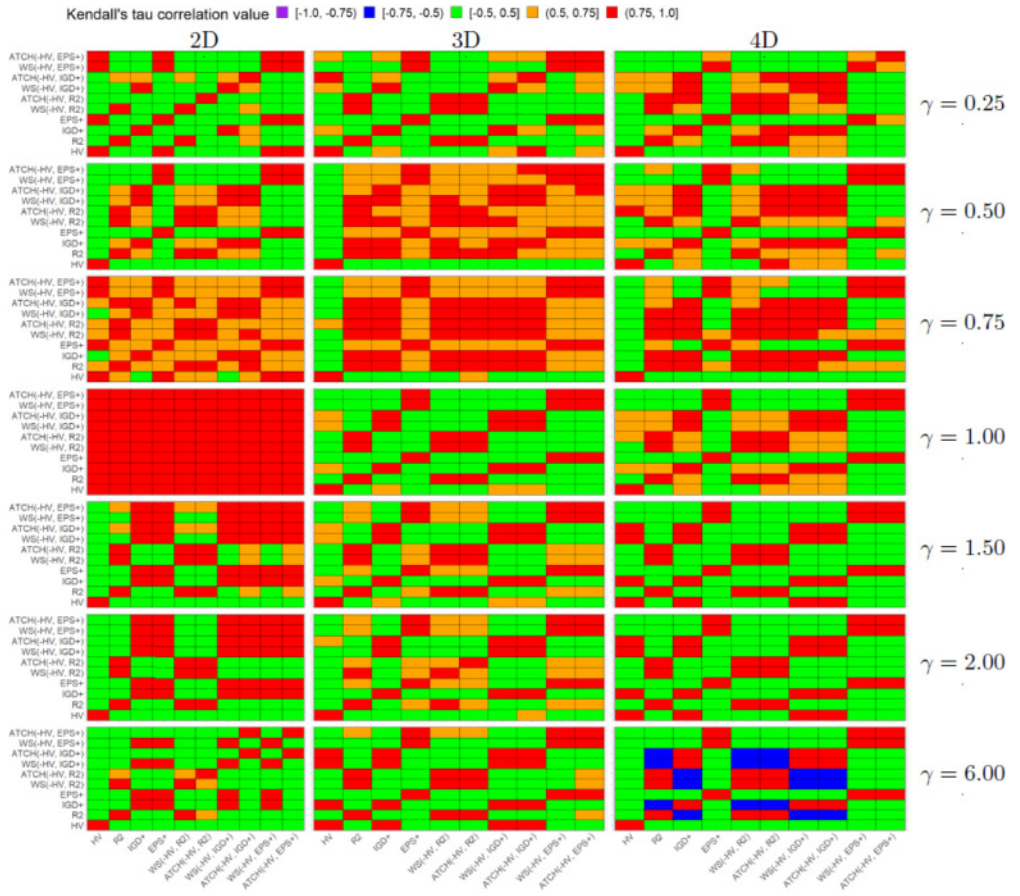

Figure 5: Heatmap Kendall rank correlation τ for each pair of quality indicators and each Mirror problem on different dimensions of the objective space.

four-objective MOPs, the correlation with all of the QIs with HV gets weaker ($\tau \in (0.5, 0.75]$). This behavior can be explained from the studies where it is stated that indicators such as HV, R2, IGD$^+$, and ϵ^+ have strongly correlated preferences for MOPs with linear Pareto front shapes (Jiang et al., 2014; Liefooghe and Derbel, 2016; Falcón-Cardona and Coello, 2019). For both highly convex and highly concave Pareto fronts, the correlation is strongly positive with the weakly Pareto-compliant QI (i.e., R2, IGD$^+$, and ϵ^+) and, overall, there is no significant correlation with HV (i.e., $\tau \in [-0.5, 0.5]$). For three- and four-objective MOPs, all the PCUIs present consistently a strong correlation with their baseline weakly Pareto-compliant QIs. In contrast, they do not exhibit a significant correlation with HV in most cases except for $\gamma = 0.75, 1.00, 1.50$. From the results, we have a preference bias towards the weakly Pareto-compliant indicators. This is an effect of giving more importance to these indicators, that is, using $w_2 = 0.9999$. Hence, the preferences of the PCUIs are very similar to those of their baseline weakly Pareto-compliant indicators. In other words, the PCUIs exhibit different preferences to those of HV but they still are Pareto-compliant regardless of the Pareto front shape and the dimensionality of the objective space.

For the Mirror problems in Figure 5, we have some similar results. However, in these problems which represent inverted Pareto front shapes of the Lamé problems, the PCUIs tend to be even more correlated to their baseline weakly Pareto-compliant QIs. The only exception is the two-objective Mirror problem with $\gamma = 1.00$ where all the PCUIs and QIs are positively correlated. This result is consistent with that of the two-objective Lamé problem, using the same γ value due to the same Pareto front shape. For all the other test instances, the preferences of the PCUIs are different to those of HV because in almost all cases, the τ correlation value is in the interval $[-0.5, 0.5]$. A clear example of the previous claim is the 4-objective Mirror problem with $\gamma = 6.00$ where no PCUI is correlated with HV while they present a Kendall's $\tau \in (0.75, 1.0]$ with respect of their baseline weakly Pareto-compliant QIs.

In summary, there are two important points to emphasize. The use of a setting of $\vec{w}$ that gives more importance to the weakly Pareto-compliant QI in the combination, will promote that the PCUIs inherit its preferences, leaving HV just as a correction factor to make the PCUI Pareto-compliant. This is a remarkable result since we can create new Pareto-compliant QIs but having preferences mostly different to those of HV. Hence, we can increase the number of Pareto-compliant indicators that researchers can employ for different comparison situations (e.g., when comparing MOEAs that can produce evenly distributed solutions). The second point relates to the design of new selection mechanisms where the distribution of solutions can be adjusted depending on the QIs to combine. In other words, PCUIs could be employed to manipulate the distribution properties of MOEAs while maintaining the Pareto compliance property.

5.2.2 Correlation between PCUIs

We analyzed the correlation between the preferences of all PCUIs to ensure that the combination does produce different indicators. Concerning both the Lamé and Mirror problems, the correlation analysis indicates that the PCUIs based on the same weakly Pareto-compliant QI are strongly correlated between them. In consequence, the use of WS or ATCH is basically producing the same PCUI although they have different landscapes. In the next section, we analyze that in some cases, the preferences of indicators are in conflict which promotes the formation of a Pareto front in the Quality Space. Hence, the PCUI could exploit this trade-off to show preferences similar to those of its baseline QIs or biased to one of them.

Another remarkable conclusion is that the preferences of PCUIs based on a different weakly Pareto-compliant QI are, in general, independent. Hence, each class of PCUIs are presenting distinct preferences. This is explained by the analysis of the correlation between R2-IGD$^+$, R2-ϵ^+, and IGD$^+$-ϵ^+ that are mostly independent as shown in Figures 4 and 5. Additionally, due to the use of $\vec{w} = (0.0001, 0.9999)$, each PCUI inherits the preferences of its weakly Pareto-compliant QI. Hence, the PCUI will behave in a similar way to its weakly Pareto-compliant QI but maintaining the Pareto compliance property.

5.2.3 Pareto Fronts in Quality Space

In objective space, we find Pareto fronts that represent the solution to a MOP. These Pareto fronts are formed due to the conflict among objective functions. In Quality Space (see Figure 2), it is also possible to find Pareto fronts when the preferences of an indicator are in conflict with the preferences of other QI. Based on the correlation analysis previously explained, we found that when there is independence of preferences between two QIs or when the preferences are negatively correlated (as in the case of $-$HV and R2 for the Lamé problem with $\gamma = 0.25$ in 2D), a Pareto front in the Quality Space Q is formed.

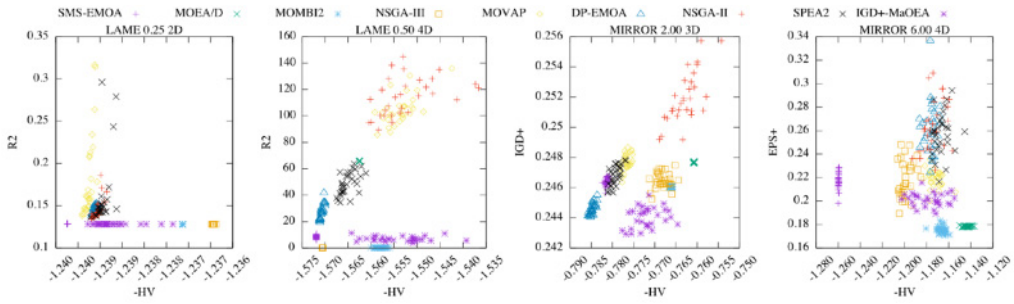

Figure 6: From left to right, it is shown the Quality Spaces: $-$HV vs R2 for Lamé $\gamma = 0.25$ 2D, $-$HV vs R2 for Lamé $\gamma = 0.50$ 4D, $-$HV vs IGD$^+$ for Mirror $\gamma = 2.00$ 3D, and $-$HV vs ϵ^+ for Mirror $\gamma = 6.00$ 4D. All cases tend to show a Pareto front in Quality Space. We use $-$HV to show the QIs in each plot for minimization.

Figure 6 shows four examples where it is possible to see the tendency to a Pareto front in quality space. These plots present the indicator vectors associated with each execution of the adopted MOEAs in the correlation study for a specific test instance. Since we are minimizing HV, R2, IGD$^+$, and ϵ^+ to use the PCUIs, it is possible to see that all plots introduce convex Pareto front shapes. Hence, this fact supports the observation that there is no critical difference when using WS or ATCH for constructing PCUIs. The rest of the cases of independence on both heatmaps in Figures 4 and 5 present convex Pareto fronts. In case a PCUI is employed in the selection mechanism of a MOEA, a compromise between the indicators will be found, resulting in new distributions on the Pareto fronts that represent the solution to a MOP. In conclusion, this result supports the fact that PCUIs could be employed to better control the diversity of a MOEA but maintaining the Pareto compliance.

5.3 Analysis of Optimal μ-Distributions

In this section, we investigate the optimal μ-distributions of the selected PCUIs. To this aim, we considered the framework of SMS-EMOA that uses a density estimator (DE) based on HV but, in our case, a PCUI is employed in the DE. Algorithm 1 presents the general framework of our proposed PCUI-EMOA whose main loop is in lines 2 to 12. At each generation, a new solution is created using genetic operators and, then, this newly created solution is added to the population P to create the temporary population Q. Then, in line 5, a set of ranks $R_1, \ldots, R_t$ are created using the nondominated sorting algorithm (Deb et al., 2002), where R_t has the worst solutions according to the Pareto dominance relation. If R_t has more than one solution, the individual contributions to the PCUI are computed to delete the worst-contributing solution in line 11. Finally, the Pareto front approximation is returned when the stopping criterion is fulfilled.

We focused our attention on studying the approximate optimal μ-distributions produced by PCUI-EMOA in comparison with those of four steady-state MOEAs based on the indicators HV, R2, IGD$^+$, and ϵ^+, that is, SMS-EMOA, R2-EMOA, IGD$^+$-MaOEA, and ϵ^+-MaOEA. The latter is similar to IGD$^+$-MaOEA. Regarding PCUI-EMOA, we employed the six PCUIs of the previous section. Since all the adopted indicator-based MOEAs (IB-MOEAs) share the same structure, the parameters settings are the following. For all objective functions, the population size is 120. All MOEAs use simulated binary crossover and polynomial-based mutation as their genetic operators (Deb et al.,

Algorithm 1 PCUI-EMOA general framework

Require: PCUI $u_{\vec{w}}(\vec{I})$, where $\vec{I}(I_1, I_2, \ldots, I_k)$.

Ensure: Approximation to the Pareto front

```
Randomly initialize population P
while stopping criterion is not fulfilled do
    q ← Variation(P)
    Q ← P ∪ {q}
    {R_1, ..., R_t} ← NondominatedSorting(Q)
    if |R_t| > 1 then
        p⃗_worst = arg max_{p⃗∈R_t} {u_w⃗(I⃗(R_t)) − u_w⃗(I⃗(R_t \ {p}))}
    else
        Let p⃗_worst be the sole solution in R_t
    end if
    P ← Q \ {p⃗_worst}
end while
return P
```

2002), where, for all cases, the crossover probability is set to 0.9, the mutation probability is $1/n$ (n is the number of decision variables), and both the crossover and mutation distribution indexes are set to 20. PCUI-EMOA employs the combination vector as $\vec{w} = (0.5, 0.5)$ to look for the knee point on the Pareto front in quality space, that is, to generate distributions similar to both baseline QIs. We tested the adopted MOEAs on 14 MOPs from the benchmarks DTLZ, WFG, DTLZ^{-1}, and WFG^{-1} for 2, 3, and 4 objective functions. We employed the problems DTLZ1, DTLZ2, DTLZ5, DTLZ7, WFG1, WFG2, WFG3, and their minus versions. It is worth noting that we turned off all the difficulties of these MOPs to avoid biasing the results such that we can observe the approximate optimal μ-distributions of the selected MOEAs. We adopted these MOPs since they possess Pareto fronts with different geometries, namely, linear, concave, convex, degenerate, mixed, disconnected, correlated with the simplex shape, and not correlated with it (Ishibuchi, Setoguchi, et al., 2017).

The study proposed here is focused on determining the similarity between the approximate optimal μ-distributions produced by the six PCUI-EMOAs and those of the selected IB-MOEAs. For each test instance, the MOEAs were executed $N = 30$ independent times. Thus, each one produced N approximation sets for each MOP. We investigate the similarity between two sets of approximation sets produced by two MOEAs, using a similarity measure based on the Hausdorff distance that we propose in the following:

Definition 10 (One-sided Hausdorff-based similarity measure): *Given two sets* $\mathbb{A} = \{A_1, \ldots, A_N\}$ *and* $\mathbb{B} = \{B_1, \ldots, B_N\}$, *each one consisting of N Pareto front approximations, the one-sided Hausdorff-based similarity measure S is given as follows:*

$$S(\mathbb{A}, \mathbb{B}) = \frac{1}{N} \sum_{i=1}^{N} median(A_i, \mathbb{B}),$$

where $median(A_i, \mathbb{B})$ *computes all the Hausdorff–Pompeiu distances from* A_i *to every element in* $\mathbb{B}$ *and returns the median value. The median is used here to prevent outliers.*

$S(\mathbb{A}, \mathbb{B})$ calculates the similarity of $\mathbb{A}$ to (most) elements in $\mathbb{B}$. It is worth noting that S is a similarity measure, not a metric (distance function), as it is not symmetrical. It attains small values if every set in $\mathbb{A}$ is similar to at least half of the sets in $\mathbb{B}$. It can, however, be the case that $\mathbb{B}$ contains a number of sets that by no means resemble sets in $\mathbb{A}$. In that sense it is not symmetrical. To make it symmetrical, one would have to compute also $S(\mathbb{B}, \mathbb{A})$ and take the maximum of these two values, as it is done in the Hausdorff–Pompeiu Distance (on the level of points sets). In consequence, we propose the two-sided Hausdorff-based similarity measure which is symmetric.

Definition 11 (Two-sided Hausdorff-based similarity measure): *Given two sets* $\mathbb{A} = \{A_1, \ldots, A_N\}$ *and* $\mathbb{B} = \{B_1, \ldots, B_N\}$*, each one consisting of N Pareto front approximations, the two-sided Hausdorff-based similarity measure* $\mathcal{H}$ *is given as follows:*

$$\mathcal{H}(\mathbb{A}, \mathbb{B}) = \max(S(\mathbb{A}, \mathbb{B}), S(\mathbb{B}, \mathbb{A})).$$

In the case that we are given three sets of approximation sets $\mathbb{A}, \mathbb{B}$, and $\mathbb{C}$ and we would like to know if $\mathbb{A}$ is similar to $\mathbb{B}$, to $\mathbb{C}$, to both, or to none of them, a classification function is required. Such classifier is given as follows.

Definition 12 (Classifier): *Given three sets of approximation sets* $\mathbb{A}$, $\mathbb{B}$, *and* $\mathbb{C}$ *and a threshold* $\epsilon > 0$*, the classifier function is given as follows:*

$$C_\epsilon(\mathcal{H}(\mathbb{A}, \mathbb{B}), \mathcal{H}(\mathbb{A}, \mathbb{C})) = \begin{cases} -1, & \mathcal{H}(\mathbb{A}, \mathbb{B}) \leq \epsilon \wedge \mathcal{H}(\mathbb{A}, \mathbb{C}) > \epsilon \\ 0, & \mathcal{H}(\mathbb{A}, \mathbb{B}) \leq \epsilon \wedge \mathcal{H}(\mathbb{A}, \mathbb{C}) \leq \epsilon \\ 1, & \mathcal{H}(\mathbb{A}, \mathbb{B}) > \epsilon \wedge \mathcal{H}(\mathbb{A}, \mathbb{C}) > \epsilon \\ 2, & \mathcal{H}(\mathbb{A}, \mathbb{C}) \leq \epsilon \wedge \mathcal{H}(\mathbb{A}, \mathbb{B}) > \epsilon \end{cases}$$

where -1 *means that* $\mathbb{A}$ *is exclusively similar to* $\mathbb{B}$*; 0 means that* $\mathbb{A}$ *is similar to both* $\mathbb{B}$ *and* $\mathbb{C}$*; 1 means that* $\mathbb{A}$ *is not similar to* $\mathbb{B}$ *nor* $\mathbb{C}$*; and, 2 means that* $\mathbb{A}$ *is exclusively similar to* $\mathbb{C}$*.*

Based on the classification function, we analyzed the similarities between the approximation sets produced by the PCUI-EMOAs and their corresponding IB-MOEAs that use the baseline indicators for the construction of the PCUI. Table 2 shows the results for all the considered test instances using $\epsilon = \sqrt{m}/10$. Since all PCUI-EMOAs use $\vec{w} = (0.5, 0.5)$ as the combination weight vector for the order-preserving utility functions, our hypothesis is that the Pareto front approximations should be similar to both IB-MOEAs that employ the baseline indicators. This hypothesis is true for several cases related to the DTLZ and DTLZ^{-1} problems. Nevertheless, for most of the WFG and WFG^{-1} problems, the PCUI-EMOA tends to produce approximation sets with particular distributions that are not similar to the baseline IB-MOEAs. This fact could be explained by the independence of preferences between HV and the weakly Pareto-compliant indicators on these MOPs. Considering the linear problems DTLZ1 and DTLZ1^{-1}, it is clear that in most cases the PCUI-EMOAs produce approximation sets similar to the IB-MOEAs using their baseline indicators. This result is explained by the correlation analysis of Section 5.2 where in almost all cases HV, R2, IGD$^+$, and ϵ^+ are strongly correlated. The most important observation is related to the PCUI-EMOAs based on $\text{WS}_{\vec{w}}(-\text{HV}, \text{R2})$ and $\text{ATCH}_{\vec{w}}(-\text{HV}, \text{R2})$. On the one hand, SMS-EMOA produces uniformly distributed solutions in convex and linear Pareto fronts and there is a bias towards the knee and boundaries of concave Pareto fronts. Additionally, SMS-EMOA presents good results in degenerate problems such as DTLZ5 and WFG3. On the other

Table 2: Distribution similarities between each PCUI-EMOA and the IB-MOEAs based on the indicators HV, R2, IGD^+, and ϵ^+. For each test instance, it is shown if the distribution of the PCUI-EMOA is similar to one or other baseline indicator, to both or none of them.

MOP	Dim.	$WS_{\vec{w}}(-\text{HV}, \text{R2})$	$ATCH_{\vec{w}}(-\text{HV}, \text{R2})$	$WS_{\vec{w}}(-\text{HV}, \text{IGD}^+)$	$ATCH_{\vec{w}}(-\text{HV}, \text{IGD}^+)$	$WS_{\vec{w}}(-\text{HV}, \epsilon^+)$	$ATCH_{\vec{w}}(-\text{HV}, \epsilon^+)$
DTLZ1	2	Both	Both	Both	Both	Both	Both
	3	Both	Both	Both	Both	Both	Both
	4	Both	Both	Both	Both	Both	Both
$DTLZ1^{-1}$	2	Both	Both	Both	Both	Both	Both
	3	Both	Both	Both	Both	Both	Both
	4	HV	HV	Both	Both	HV	None
DTLZ2	2	Both	Both	Both	Both	Both	Both
	3	R2	R2	Both	Both	Both	Both
	4	R2	R2	None	None	None	None
$DTLZ2^{-1}$	2	HV	HV	Both	Both	Both	Both
	3	HV	HV	None	None	None	None
	4	None	None	None	None	None	None
DTLZ5	2	Both	Both	Both	Both	Both	Both
	3	Both	Both	Both	Both	Both	Both
	4	HV	HV	IGD^+	IGD^+	ϵ^+	ϵ^+
$DTLZ5^{-1}$	2	HV	HV	Both	Both	Both	Both
	3	HV	HV	IGD^+	None	None	None
	4	None	None	None	None	None	None
DTLZ7	2	Both	Both	Both	Both	Both	Both
	3	None	None	IGD^+	None	None	None
	4	None	None	None	None	None	None
$DTLZ7^{-1}$	2	Both	Both	Both	Both	Both	Both
	3	R2	R2	Both	Both	Both	Both
	4	None	None	None	ϵ^+	ϵ^+	None
WFG1	2	None	None	None	None	None	None
	3	None	None	None	None	None	None
	4	None	None	IGD^+	IGD^+	ϵ^+	None
$WFG1^{-1}$	2	R2	R2	None	None	None	None
	3	None	None	None	None	None	None
	4	None	None	None	None	None	None
WFG2	2	None	None	None	None	None	None
	3	None	None	None	None	None	None
	4	None	None	None	None	None	None

Table 2: Continued.

MOP	Dim.	$WS_{\vec{w}}(-HV, R2)$	$ATCH_{\vec{w}}(-HV, R2)$	$WS_{\vec{w}}(-HV, IGD^+)$	$ATCH_{\vec{w}}(-HV, IGD^+)$	$WS_{\vec{w}}(-HV, \epsilon^+)$	$ATCH_{\vec{w}}(-HV, \epsilon^+)$
$WFG2^{-1}$	2	None	None	None	None	None	None
	3	R2	R2	None	None	None	None
	4	None	None	None	None	None	None
WFG3	2	None	None	None	None	None	None
	3	Both	Both	Both	Both	Both	Both
	4	None	None	IGD^+	IGD^+	ϵ^+	ϵ^+
$WFG3^{-1}$	2	R2	R2	None	None	None	None
	3	Both	Both	Both	Both	Both	Both
	4	HV	HV	Both	Both	HV	HV

hand, R2-EMOA (Brockhoff et al., 2015) does not produce uniformly distributed solutions in convex Pareto fronts, but it does in linear and concave ones. Regarding degenerate MOPs, R2-EMOA does not produce good results since its weight vectors do not completely intersect the Pareto front shape. Hence, SMS-EMOA and R2-EMOA have specific strengths and weaknesses depending on the MOP being tackled. We refer to *strengths* and *weaknesses* of an IB-MOEA (and, more specifically, to the underlying QI) as its capacity to generate (or failure to generate) Pareto front approximations with good diversity. For instance, SMS-EMOA does not produce a uniform distribution in DTLZ2 (which has a concave Pareto front) because of the inner preferences of the HV while R2-EMOA can produce uniformly distributed solutions in this problem.

Regarding DTLZ2 having three and four objective functions and concave Pareto fronts, it is possible to see that the distribution of the PCUI-EMOAs based on $WS_{\vec{w}}(-HV, R2)$ and $ATCH_{\vec{w}}(-HV, R2)$ are similar to the preferences of R2, that is, R2-EMOA. When we analyze the minus version $DTLZ2^{-1}$ for the same objective functions, the distributions are similar to those of SMS-EMOA. This also happens for $DTLZ5^{-1}$ 3D which is inverted convex where the distributions are similar to those of SMS-EMOA as well. Hence, we have empirical evidence on the compensation of weaknesses of one indicator with the strengths of the other baseline indicator when employing PCUI-EMOA. Figure 7 shows some examples of this remarkable compensation.

5.4 Computational Complexity

Theorem 1 proposes the combination of as many weakly Pareto-compliant QIs as required with at least one Pareto-compliant indicator. Currently, the only Pareto-compliant QI available is the hypervolume indicator whose computational cost increases super-polynomially with the number of objective functions (Bringmann and Friedrich, 2009, 2010). In terms of computational cost, it is clear that the runtime complexity of HV dominates those of R2, IGD^+, and ϵ^+, which implies that a PCUI and,

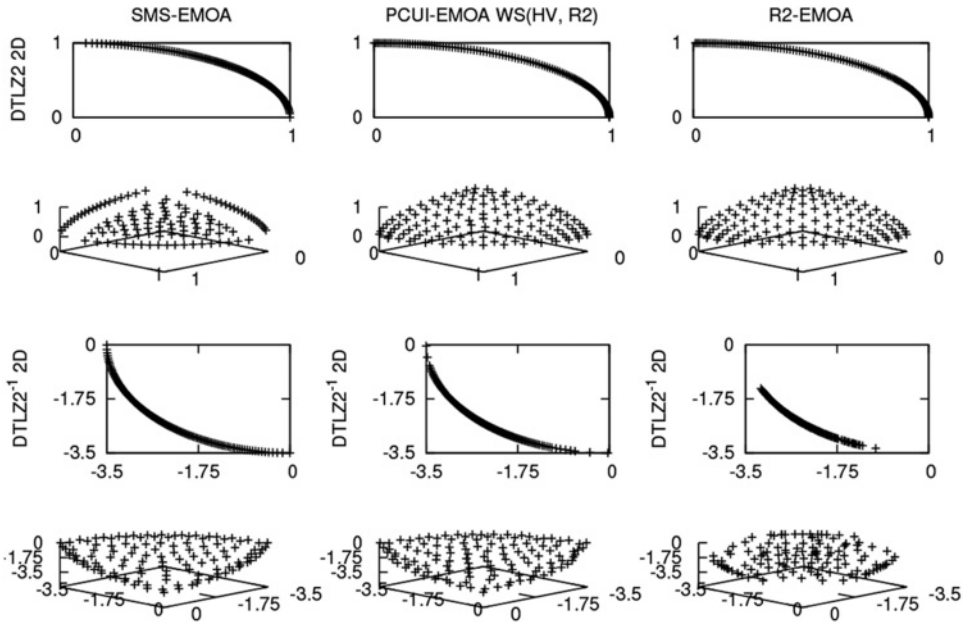

Figure 7: Pareto fronts that show the compensation of weaknesses of one indicator with the strengths of other when coupled to PCUI-EMOA.

in general, all the QIs constructed by Theorem 1 inherit the cost of HV. Hence, one may argue that this a clear disadvantage of the proposed method. However, we need to emphasize that even though we showed that a PCUI can be utilized to guide the evolutionary process of a MOEA (in fact, we used PCUI-EMOA to study the approximate optimal μ-distributions of PCUIs), the primary focus of PCUIs is to increase the number of available Pareto-compliant indicators for performance assessment. In consequence, when using a PCUI to compare MOEAs, one may use fast HV calculation methods such as the Walking-Fish-Group (WFG) algorithm (While et al., 2012) that significantly reduces the computational time to get exact HV values. Although following Theorem 1 allows the construction of new Pareto-compliant QIs, this does not overcome the runtime complexity of the HV, and it does not theoretically increase the overall computational cost.[9] Additionally, these combined indicators provide different preferences to those of the HV and they are controlled by the user when setting the weight vector $\vec{w}$. Hence, we argue that we obtain more advantages than drawbacks by using our proposed framework.

At this point, it is worth comparing the PCUIs and the multi-indicator-based preference relations proposed by Zitzler et al. (2010) in terms of computational cost. Due to the sequential application of the indicator-based preference relations ($\preceq_{I_j}$), it is likely that a multi-indicator-based preference relation avoids the computational cost of calculating HV whenever a cheaper QI could solve the comparison. In contrast, PCUIs do always require the calculation of HV in furtherance of obtaining extra flexibility to

[9]Even in practice, the calculation of as many weakly Pareto-compliant QIs as required does not aggregate a considerable overhead when computing a PCUI value.

control how the preferences are combined. Moreover, from the discussion throughout this article, it is possible to visualize two distinct application paths. On the one hand, PCUIs are useful for performance assessment of MOEAs, considering the Pareto compliance property and offering different preferences to those of HV. On the other hand, the multi-indicator-based preference relations are a promising direction to construct set-based MOEAs as in the case of SPAM (Zitzler et al., 2010). In consequence, each approach is specialized in different areas of evolutionary multiobjective optimization. Nevertheless, it is worth noting that PCUIs and, in general, our proposed framework for the combination of multiple QIs represent a generalization of the multi-indicator-based preferences relations proposed by Zitzler et al. (2010).

6 Conclusions and Future Work

In this article, we proposed to construct new Pareto-compliant indicators by combining existing QIs, under specific conditions. In order to ensure the Pareto compliance property, it is mandatory to combine at least one Pareto-compliant indicator (such as the hypervolume indicator) with as many weakly Pareto-compliant QIs as required, by using an order-preserving function. Following this framework, we proposed six new Pareto-compliant QIs based on the combination of HV with R2, IGD^+, and ϵ^+, using the weighted sum and augmented Tchebycheff utility functions as order-preserving combination functions. We studied the properties of these six PCUIs from a preferences and approximate optimal μ-distributions perspective. Our experimental results showed that the proposed PCUIs do exhibit different preferences to those of the HV which implies that they can be used as an alternative to the HV to assess performance of MOEAs while ensuring the Pareto compliance property. These results were supported by an analysis of approximate optimal μ-distributions. Furthermore, when using a steady-state MOEA based on a PCUI, we observed that the non-uniform distributions produced by one of the baseline QIs of the PCUI could be compensated by the uniform distribution strengths of another baseline QI. It is worth emphasizing that the main focus of PCUIs is the performance assessment of MOEAs. In this direction, we discussed that even though a PCUI requires the calculation of the HV, the main advantage of PCUIs lies on their different preferences with respect to those of the HV. Additionally, PCUIs have the same theoretical runtime complexity as the HV while allowing to draw conclusions supported by the Pareto compliance property. As part of our future work, we aim to provide a mechanism that adapts the combination vector depending on the geometrical features of the MOP. Moreover, we aim to theoretically analyze the properties of PCUIs in order to define in which cases they can be used to provide better information. Finally, an analysis considering MOPs with more than four objective functions is also a task that we would like to undertake as part of our future work.

Acknowledgments

The first author acknowledges support from CINVESTAV-IPN to pursue graduate studies in computer science and the 2018 IEEE CIS Graduate Student Research Grant. The third author gratefully acknowledges support from CONACyT grant no. 2016-01-1920 (*Investigación en Fronteras de la Ciencia 2016*) and from a SEP-Cinvestav grant (proposal no. 4). He was also partially supported by the Basque Government through the BERC 2022–2025 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017-0718.

References

Auger, A., Bader, J., Brockhoff, D., and Zitzler, E. (2009). Theory of the hypervolume indicator: Optimal $\{\mu\}$-distributions and the choice of the reference point. In *Proceedings of the Tenth ACM SIGEVO Workshop on Foundations of Genetic Algorithms*, pp. 87–102.

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. 10.1016/j.ejor.2006.08.008

Bezerra, L. C. T., López-Ibáñez, M., and Stützle, T. (2017). An empirical assessment of the properties of inverted generational distance on multi- and many-objective optimization. In *Proceedings of the 9th International Conference on Evolutionary Multi-Criterion Optimization*, pp. 31–45.

Bringmann, K., and Friedrich, T. (2009). Approximating the least hypervolume contributor: NP-hard in general, but fast in practice. In *Fifth International Conference on Evolutionary Multi-Criterion Optimization*, pp. 6–20. Lecture Notes in Computer Science, Vol. 5467. 10.1007/978-3-642-01020-0_6

Bringmann, K., and Friedrich, T. (2010). Approximating the volume of unions and intersections of high-dimensional geometric objects. *Computational Geometry—Theory and Applications*, 43(6–7):601–610. 10.1016/j.comgeo.2010.03.004

Bringmann, K., Friedrich, T., Neumann, F., and Wagner, M. (2011). Approximation-guided evolutionary multi-objective optimization. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pp. 1198–1203.

Brockhoff, D., Wagner, T., and Trautmann, H. (2012). On the properties of the $R2$ indicator. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 465–472.

Brockhoff, D., Wagner, T., and Trautmann, H. (2015). R2 indicator-based multiobjective search. *Evolutionary Computation*, 23(3):369–395. 10.1162/EVCO_a_00135, PubMed: 24983593

Coello Coello, C. A., and Cruz Cortés, N. (2005). Solving multiobjective optimization problems using an artificial immune system. *Genetic Programming and Evolvable Machines*, 6(2):163–190. 10.1007/s10710-005-6164-x

Coello Coello, C. A., Lamont, G. B., and Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems*. 2nd ed. New York: Springer.

Das, I., and Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657. 10.1137/S1052623496307510

Deb, K., and Jain, H. (2014). An evolutionary many-objective optimization algorithm using reference-point-based-nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601. 10.1109/TEVC.2013.2281535

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA–II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197. 10.1109/4235.996017

Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. (2005). Scalable test problems for evolutionary multiobjective optimization. In A. Abraham, L. Jain, and R. Goldberg (Eds.), *Evolutionary multiobjective optimization. Theoretical advances and applications*, pp. 105–145. New York: Springer.

Emmerich, M. T., and Deutz, A. H. (2007). Test problems based on lamé superspheres. In *Fourth International Conference on Evolutionary Multi-Criterion Optimization*, pp. 922–936. Lecture Notes in Computer Science, Vol. 4403. 10.1007/978-3-540-70928-2_68

Emmerich, M. T., Deutz, A. H., and Yevseyeva, I. (2014). On reference point free weighted hypervolume indicators based on desirability functions and their probabilistic interpretation. *Procedia Technology*, 16:532–541. CENTERIS 2014 - Conference on ENTERprise Information Systems/ProjMAN 2014 - International Conference on Project MANagement/HCIST 2014 - International Conference on Health and Social Care Information Systems and Technologies. 10.1016/j.protcy.2014.10.001

Falcón-Cardona, J. G., and Coello, C. A. C. (2019). Convergence and diversity analysis of indicator-based multi-objective evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 524–531.

Falcón-Cardona, J. G., and Coello Coello, C. A. (2018). Towards a more general many-objective evolutionary optimizer. In *2018 Parallel Problem Solving from Nature*.

Falcón-Cardona, J. G., and Coello Coello, C. A. (2020). Indicator-based multi-objective evolutionary algorithms: A comprehensive survey. *ACM Computing Surveys*, 53(2):1–35.

Falcón-Cardona, J. G., Emmerich, M. T. M., and Coello Coello, C. A. (2019). On the construction of Pareto-compliant quality indicators. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO)*, pp. 2024–2027.

Falcón-Cardona, J. G., Ishibuchi, H., and Coello Coello, C. A. (2020). Riesz s-energy-based reference sets for multi-objective optimization. In *IEEE Congress on Evolutionary Computation*, pp. 1–8.

Friedrich, T., Bringmann, K., Voß, T., and Igel, C. (2011). The logarithmic hypervolume indicator. In *Proceedings of the 2011 ACM/SIGEVO Foundations of Genetic Algorithms*, pp. 81–92.

Guerreiro, A. P., and Fonseca, C. M. (2020). An analysis of the hypervolume Sharpe-ratio indicator. *European Journal of Operational Research*, 283(2):614–629. 10.1016/j.ejor.2019.11.023

Hansen, M. P., and Jaszkiewicz, A. (1998). *Evaluating the quality of approximations to the non-dominated set*. Technical Report IMM-REP-1998-7. Technical University of Denmark.

Hernández Gómez, R., and Coello Coello, C. A. (2015). Improved metaheuristic based on the $R2$ indicator for many-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 679–686.

Hernández Gómez, R., Coello Coello, C. A., and Alba Torres, E. (2016). A multi-objective evolutionary algorithm based on parallel coordinates. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 565–572.

Horn, J., Nafpliotis, N., and Goldberg, D. E. (1994). A niched Pareto genetic algorithm for multiobjective optimization. In *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, Vol. 1, pp. 82–87. 10.1109/ICEC.1994.350037

Huband, S., Hingston, P., Barone, L., and While, L. (2006). A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506. 10.1109/TEVC.2005.861417

Ishibuchi, H., Imada, R., Setoguchi, Y., and Nojima, Y. (2017). Reference point specification in hypervolume calculation for fair comparison and efficient search. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 585–592.

Ishibuchi, H., Masuda, H., Tanigaki, Y., and Nojima, Y. (2014). Difficulties in specifying reference points to calculate the inverted generational distance for many-objective optimization problems. In *2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making*, pp. 170–177.

Ishibuchi, H., Masuda, H., Tanigaki, Y., and Nojima, Y. (2015). Modified distance calculation in generational distance and inverted generational distance. In *Proceedings of the 8th*

International Conference on Evolutionary Multi-Criterion Optimization, pp. 110–125. Lecture Notes in Computer Science, Vol. 9019. 10.1007/978-3-319-15892-1_8

Ishibuchi, H., Setoguchi, Y., Masuda, H., and Nojima, Y. (2017). Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes. *IEEE Transactions on Evolutionary Computation*, 21(2):169–190. 10.1109/TEVC.2016.2587749

Jiang, S., Ong, Y.-S., Zhang, J., and Feng, L. (2014). Consistencies and contradictions of performance metrics in multiobjective optimization. *IEEE Transactions on Cybernetics*, 44(12):2391–2404. 10.1109/TCYB.2014.2307319, PubMed: 25415945

Knowles, J., Thiele, L., and Zitzler, E. (2006). A tutorial on the performance assessment of stochastic multiobjective optimizers. Report 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland. Revised version.

Li, F., Cheng, R., Liu, J., and Jin, Y. (2018). A two-stage R2 indicator based evolutionary algorithm for many-objective optimization. *Applied Soft Computing*, 67:245–260. 10.1016/j.asoc.2018.02.048

Li, M., and Yao, X. (2019). Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys*, 52(2):26:1–26:38.

Liefooghe, A., and Derbel, B. (2016). A correlation analysis of set quality indicator values in multiobjective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 581–588.

Miettinen, K. (1999). *Nonlinear multiobjective optimization*. Boston: Kluwer Academic Publishers.

Pescador-Rojas, M., Hernández Gómez, R., Montero, E., Rojas-Morales, N., Riff, M.-C., and Coello Coello, C. A. (2017). An overview of weighted and unconstrained scalarizing functions. In *Proceedings of the 9th International Conference on Evolutionary Multi-Criterion Optimization*, pp. 499–513. Lecture Notes in Computer Science, Vol. 10173. 10.1007/978-3-319-54157-0_34

Shang, K., Ishibuchi, H., He, L., and Pang, L. M. (2020). A survey on the hypervolume indicator in evolutionary multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, pp. 1–20.

Shang, K., Ishibuchi, H., Nan, Y., and Chen, W. (2020). Transformation-based hypervolume indicator: A framework for designing hypervolume variants. In *2020 IEEE Symposium Series on Computational Intelligence*, pp. 157–164.

Tian, Y., Cheng, R., Zhang, X., Cheng, F., and Jin, Y. (2018). An indicator-based multiobjective evolutionary algorithm with reference point adaptation for better versatility. *IEEE Transactions on Evolutionary Computation*, 22(4):609–622. 10.1109/TEVC.2017.2749619

Veldhuizen, D. A. V. (1999). Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations. PhD thesis, Department of Electrical and Computer Engineering. Graduate School of Engineering. Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

While, L., Bradstreet, L., and Barone, L. (2012). A fast way of calculating exact hypervolumes. *IEEE Transactions on Evolutionary Computation*, 16(1):86–95. 10.1109/TEVC.2010.2077298

Yuan, Y., Xu, H., Wang, B., and Yao, X. (2016). A new dominance relation-based evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 20(1):16–37. 10.1109/TEVC.2015.2420112

Zhang, Q., and Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731. 10.1109/TEVC.2007.892759

Zitzler, E. (1999). Evolutionary algorithms for multiobjective optimization: Methods and applications. PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Zitzler, E., Brockhoff, D., and Thiele, L. (2007). The hypervolume indicator revisited: On the design of Pareto-compliant indicator via weighted integration. In *Proceedings of the 4th International Conference on Evolutionary Multi-Criterion Optimization*, pp. 862–876. Lecture Notes in Computer Science, Vol. 4403. 10.1007/978-3-540-70928-2_64

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2):173–195. 10.1162/106365600568202, PubMed: 10843520

Zitzler, E., Knowles, J., and Thiele, L. (2008). Quality assessment of Pareto set approximations. In J. Branke, K. Deb, K. Miettinen, and R. Slowinski (Eds.), *Multiobjective optimization. Interactive and evolutionary approaches*, pp. 373–404. Lecture Notes in Computer Science, Vol. 5252. Berlin: Springer. 10.1007/978-3-540-88908-3_14

Zitzler, E., Laumanns, M., and Thiele, L. (2001). SPEA2: Improving the strength Pareto evolutionary algorithm. In *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, pp. 95–100.

Zitzler, E., Thiele, L., and Bader, J. (2008). SPAM: Set Preference Algorithm for Multiobjective Optimization. In *Parallel Problem Solving from Nature*, pp. 847–858. Lecture Notes in Computer Science, Vol. 5199.

Zitzler, E., Thiele, L., and Bader, J. (2010). On set-based multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 14(1):58–79. 10.1109/TEVC.2009.2016569

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and da Fonseca, V. G. (2003). Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132. 10.1109/TEVC.2003.810758