

# Performance assessment of multiobjective optimizers: an analysis and review

**Report****Author(s):**

Zitzler, Eckart; Thiele, Lothar; Laumanns, Marco; Fonseca, Carlos M.; Grunert da Fonseca, Viviane

**Publication date:**

2002-06

**Permanent link:**

<https://doi.org/10.3929/ethz-a-004363410>

**Rights / license:**

In Copyright - Non-Commercial Use Permitted

**Originally published in:**

TIK Report 139

# Performance Assessment of Multiobjective Optimizers: An Analysis and Review

Eckart Zitzler<sup>1</sup>, Lothar Thiele<sup>1</sup>, Marco Laumanns<sup>1</sup>,  
Carlos M. Fonseca<sup>2</sup>, and Viviane Grunert da Fonseca<sup>2</sup>

<sup>1</sup> Computer Engineering and Networks Laboratory (TIK)  
Department of Information Technology and Electrical Engineering  
Swiss Federal Institute of Technology (ETH) Zurich, Switzerland  
Email: {zitzler, thiele, laumanns}@tik.ee.ethz.ch

<sup>2</sup>ADEEC and ISR (Coimbra)  
Faculty of Sciences and Technology  
University of Algarve, Portugal  
Email: cmfonsec@ualg.pt, vgrunert@csi.fct.ualg.pt

TIK-Report No. 139  
Institut für Technische Informatik und Kommunikationsnetze, ETH Zürich  
Gloriastrasse 35, ETH-Zentrum, CH-8092 Zürich, Switzerland

June 26, 2002

## Abstract

An important issue in multiobjective optimization is the quantitative comparison of the performance of different algorithms. In the case of multiobjective evolutionary algorithms, the outcome is usually an approximation of the Pareto-optimal front, which is denoted as an approximation set, and therefore the question arises of how to evaluate the quality of approximation sets. Most popular are methods that assign each approximation set a vector of real numbers that reflect different aspects of the quality. Sometimes, pairs of approximation sets are considered too. In this study, we provide a rigorous analysis of the limitations underlying this type of quality assessment. To this end, a mathematical framework is developed which allows to classify and discuss existing techniques.

# 1 Introduction

With many multiobjective optimization problems, knowledge about the Pareto-optimal front helps the decision maker in choosing the best compromise solution. For instance, when designing computer systems, engineers often perform a so-called design space exploration to learn more about the trade-off surface. Thereby, the design space is reduced to the set of optimal trade-offs: a first step in selecting an appropriate implementation.

However, generating the Pareto-optimal front can be computationally expensive and is often infeasible, because the complexity of the underlying application prevents exact methods from being applicable. Evolutionary algorithms (EAs) are an alternative: they usually do not guarantee to identify optimal trade-offs but try to find a good approximation, i.e., a set of solutions that are (hopefully) not too far away from the optimal front. Various multiobjective EAs are available, and certainly we are interested in the technique that provides the best approximation for a given problem. For this reason, comparative studies are conducted, e.g., [26][22][19]; they aim at revealing strengths and weaknesses of certain approaches and at identifying the most promising algorithms. This, in turn, leads to the question of how to compare the performance of multiobjective optimizers.

The notion of performance includes both the quality of the outcome as well as the computational resources needed to generate this outcome. Concerning the latter aspect, it is common practice to keep the number of fitness evaluations or the overall runtime constant—in this sense, there is no difference between single and multiobjective optimization. As to the quality aspect, however, there is a difference. In single-objective optimization, we can define quality by means of the objective function: the smaller (or larger) the value, the better the solution. In contrast, it is not clear what quality means in the presence of several optimization criteria: closeness to the optimal front, coverage of a wide range of diverse solutions, or other properties? Therefore, it is difficult to define appropriate quality measures for approximations of the Pareto-optimal front, and as a consequence

graphical plots have been used to compare the outcomes of multiobjective EAs until recently, as Van Veldhuizen points out [21].

Progress, though, has been made and meanwhile several studies can be found in the literature that address the problem of comparing approximations of the trade-off surface in a quantitative manner. Most popular are unary quality measures, i.e., the measure assigns each approximation set a number that reflects a certain quality aspect, and usually a combination of them is used, e.g., [22][4]. Other methods are based on binary quality measures, which assign numbers to pairs of approximation sets, e.g., [26][9]. A third, and conceptually different approach, is the attainment function approach [8], which consists of estimating the probability of attaining arbitrary goals in objective space from multiple approximation sets. Despite of this variety, it has remained unclear up to now how the different measures are related to each other and what their advantages and disadvantages are. Accordingly, there is no common agreement on which measure(s) should be used.

Recently, a few studies have been carried out to clarify this situation. Hansen and Jaszekiewicz [9] studied and proposed some quality measures that induce a linear ordering on the space of possible approximations—on the basis of assumptions about the decision maker’s preferences. They first introduced three different “outperformance” relations for multiobjective optimizers and then investigated whether the measures under consideration are compliant with these relations. The basic question they considered was: whenever an approximation is better than another according to an “outperformance” relation, does the comparison method also evaluate the former as being better (or at least not worse) than the latter? More from a practical point of view, Knowles, Corne, and Oates [12] compared the information provided by different assessment techniques on two database management applications. Later, Knowles [14] and Knowles and Corne [13] discussed and contrasted several commonly used quality measures in the light of Hansen and Jaszekiewicz’s approach as well as according to other criteria such as, e.g., sensitivity to scaling. They showed that about

one third of the investigated quality measures are not compliant with any of the "outperformance" relations introduced by Hansen and Jaszkievicz.

This paper takes a different perspective that allows a more rigorous analysis and classification of comparison methods. In contrast to [9], [14], and [13], we focus on the statements that can be made on the basis of the information provided by quality measures. Is it, for instance, possible to conclude from the quality "measurements" that an approximation  $A$  is undoubtedly better than approximation  $B$  in the sense that  $A$ , loosely speaking, entirely dominates  $B$ ? This is a crucial issue in any comparative study, and implicitly most papers in this area rely on the assumption that this property is satisfied for the measures used. To investigate quality measures from this perspective, a formal framework will be introduced that substantially goes beyond Hansen and Jaszkievicz's approach as well as that of Knowles and Corne; e.g., it will enable us to consider combinations of quality measures and to prove theoretical limitations of unary quality measures, both issues not addressed in [9], [14], and [13]. In detail, we will show that

- there exists no unary quality measure that is able to indicate *whether* an approximation  $A$  is better than an approximation  $B$ ;
- the above statement even holds if we consider a finite combination of unary measures;
- most existing quality measures that have been proposed to indicate *that*  $A$  is better than  $B$  at best allow to infer that  $A$  is not worse than  $B$ , i.e.,  $A$  is better than or incomparable to  $B$ ;
- unary measures being able to detect *that*  $A$  is better than  $B$  exist, but their use is in general restricted;
- binary quality measures overcome the limitations of unary measures and, if properly designed, are capable of indicating *whether*  $A$  is better than  $B$ .

Furthermore, we will review existing quality measures in the light of this framework and discuss them

also from a practical point of view. Note that we focus on the comparison of approximations of the Pareto-optimal front rather than on algorithms, i.e., we assume that for each multiobjective EA only one run is performed. In the case of multiple runs, the distribution of the indicator values would have to be considered instead of the values themselves; this important issue will not be addressed in the present paper.

## 2 Theoretical Framework

Before analyzing and classifying quality measures, we must clarify the concepts we will be dealing with: what is the outcome of a multiobjective optimizer, when is an outcome considered to be better than another, what is a quality measure, what is a comparison method, etc.? These terms will be formally defined in this section.

### 2.1 Approximation Sets

The scenario considered in this paper involves an arbitrary optimization problem with  $n$  objectives, which are, without loss of generality, all to be minimized. We will use the symbol  $Z$  to denote the space of all possible solutions to the problem with respect to the objective values;  $Z$  is also called the *objective space* and each element of  $Z$  is referred to as *objective vector*. Here, we will use the terms objective vector and solution interchangeably.

We consider the most general case, in which all objectives are considered to be equally important—no additional knowledge about the problem is available. The only assumption we make is that a solution  $z^1$  is preferable to another solution  $z^2$  if  $z^1$  is at least as good as  $z^2$  in all objectives *and* better with respect to at least one objective. This is commonly known as the concept of *Pareto dominance*, and we also say  $z^1$  *dominates*  $z^2$ . The dominance relation induces a partial order on the search space, so that we can define an optimal solution to be one that is not dominated by any other solution. However, several such solutions, which are denoted as *Pareto optimal*, may

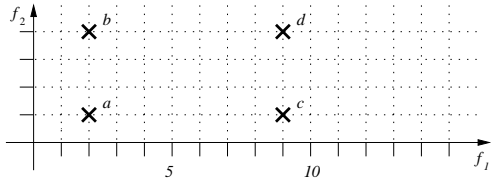


Figure 1: Examples of dominance relations on objective vectors. Assuming that two objectives are to be minimized, it holds that  $a \succ b$ ,  $a \succ c$ ,  $a \succ d$ ,  $b \succ d$ ,  $c \succ d$ ,  $a \succ\succeq d$ ,  $a \succeq a$ ,  $a \succeq b$ ,  $a \succeq c$ ,  $a \succeq d$ ,  $b \succeq b$ ,  $b \succeq d$ ,  $c \succeq c$ ,  $c \succeq d$ ,  $d \succeq d$ , and  $b \parallel c$ .

exist as two objective vectors can be incomparable to each other: each is superior to the other in some objectives and inferior in other objectives. Fig. 1 visualizes these concepts and also gives some examples for other common relations on pairs of objective vectors. Table 2 comprises a summary of the relations used in this paper.

The vast majority of papers in the area of evolutionary multiobjective optimization is concerned with the problem of how to identify the Pareto-optimal solutions or, if this is infeasible, to generate good approximations of them. Taking this as the basis of our study, we here consider the outcome of a multiobjective EA (or other heuristic) as a set of incomparable solutions, or formally [9]:

**Definition 1 (Approximation set)** *Let  $A \subseteq Z$  be a set of objective vectors.  $A$  is called an approximation set if any element of  $A$  does not dominate or is not equal to any other objective vector in  $A$ . The set of all approximation sets is denoted as  $\Omega$ .*

The motivation behind this definition is that all solutions dominated by any other solution outputted by the optimization algorithm are of no interest and therefore can be discarded. This will simplify the considerations in the following sections.

Note that the above definition does not comprise any notion of quality. We are certainly not interested in *any* approximation set, but we want the EA to generate a *good* approximation set. The ultimate

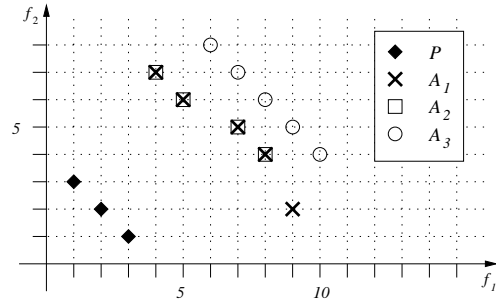


Figure 2: Outcomes of three hypothetical algorithms for a two-dimensional minimization problem. The corresponding approximation sets are denoted as  $A_1$ ,  $A_2$ , and  $A_3$ ; the Pareto-optimal front  $P$  consist of three objective vectors. Between  $A_1$ ,  $A_2$ , and  $A_3$ , the following dominance relations hold:  $A_1 \succ A_3$ ,  $A_2 \succ A_3$ ,  $A_1 \succ\succeq A_3$ ,  $A_1 \succeq A_1$ ,  $A_1 \succeq A_2$ ,  $A_1 \succeq A_3$ ,  $A_2 \succeq A_2$ ,  $A_2 \succeq A_3$ ,  $A_3 \succeq A_3$ ,  $A_1 \triangleright A_2$ ,  $A_1 \triangleright A_3$ , and  $A_2 \triangleright A_3$ .

goal is to identify the so-called *Pareto-optimal* front, that is the set of all Pareto-optimal solutions. This aim, however, is usually not achievable. Moreover, it is impossible to exactly describe what a good approximation is in terms of a number of criteria such as closeness to the Pareto-optimal front, diversity, etc.—this will be shown in Section 3.1. However, we can make statements about the quality of approximation sets in comparison to other approximation sets.

Consider, e.g., the outcomes of three hypothetical algorithms as depicted in Fig. 2. Solely on the basis of Pareto dominance, one can state that  $A_1$  and  $A_2$  are both superior to  $A_3$  as any solution in  $A_3$  is dominated by at least one solution in  $A_1$  and  $A_2$ . Furthermore,  $A_1$  can be considered superior to  $A_2$  as it contains all solutions in  $A_2$  and another solution not included in  $A_2$ , although this statement is weaker than the previous one. Accordingly, we will distinguish four levels of superiority in this paper as defined in Table 1:  $A$  strictly dominates  $B$  ( $A \succ\succeq B$ ),  $A$  dominates  $B$  ( $A \succ B$ ),  $A$  is better than  $B$  ( $A \triangleright B$ ), and  $A$  weakly dominates  $B$  ( $A \succeq B$ ), where  $A \succ\succeq B \Rightarrow A \succ B \Rightarrow A \triangleright B \Rightarrow A \succeq B$ .

relation	objective vectors		approximation sets	
strictly dominates	$\mathbf{z}^1 \succ\succ \mathbf{z}^2$	$\mathbf{z}^1$ is better than $\mathbf{z}^2$ in all objectives	$A \succ\succ B$	every $\mathbf{z}^2 \in B$ is strictly dominated by at least one $\mathbf{z}^1 \in A$
dominates	$\mathbf{z}^1 \succ \mathbf{z}^2$	$\mathbf{z}^1$ is not worse than $\mathbf{z}^2$ in all objectives and better in at least one objective	$A \succ B$	every $\mathbf{z}^2 \in B$ is dominated by at least one $\mathbf{z}^1 \in A$
better			$A \triangleright B$	every $\mathbf{z}^2 \in B$ is weakly dominated by at least one $\mathbf{z}^1 \in A$ and $A \neq B$
weakly dominates	$\mathbf{z}^1 \succeq \mathbf{z}^2$	$\mathbf{z}^1$ is not worse than $\mathbf{z}^2$ in all objectives	$A \succeq B$	every $\mathbf{z}^2 \in B$ is weakly dominated by at least one $\mathbf{z}^1 \in A$
incomparable	$\mathbf{z}^1 \parallel \mathbf{z}^2$	neither $\mathbf{z}^1$ weakly dominates $\mathbf{z}^2$ nor $\mathbf{z}^2$ weakly dominates $\mathbf{z}^1$	$A \parallel B$	neither $A$ weakly dominates $B$ nor $B$ weakly dominates $A$

Table 1: Relations on objective vectors and approximation sets considered in this paper. The relations  $\prec$ ,  $\prec\prec$ ,  $\triangleleft$ , and  $\preceq$  are defined accordingly, e.g.,  $\mathbf{z}^1 \prec \mathbf{z}^2$  is equivalent to  $\mathbf{z}^2 \succ \mathbf{z}^1$  and  $A \triangleleft B$  is defined as  $B \triangleright A$ .

Weak dominance ( $A \succeq B$ ) means that any solution in  $B$  is weakly dominated by a solution in  $A$ . However, this does not rule out equality, because  $A \succeq A$  for all approximation sets  $A \in \Omega$ . In this case, one cannot say that  $A$  is better than  $B$ . Instead, the relation  $\triangleright$  can be used as it represents the most general and weakest form of superiority. It requires that an approximation set is at least as good as another approximation set ( $A \succeq B$ ), while the latter is not as good as the former ( $B \not\succeq A$ ), roughly speaking. In the example,  $A_1$  is better than  $A_2$  and  $A_3$ , and  $A_2$  is better than  $A_3$ . This definition of superiority is the one implicitly used in most papers in the field. The next level of superiority, the  $\succ$  relation, is a straight-forward extension of Pareto dominance to approximation sets. It does not allow that two solutions in  $A$  and  $B$  are equal and therefore is stricter than what we usually require. As mentioned above,  $A_1$  and  $A_2$  dominate  $A_3$ , but  $A_1$  does not dominate  $A_2$ . Strict dominance stands for the highest level of superiority and means an approximation set is superior to another approximation set in the sense that for any solution in the latter there exists a solution in the former that is better in all objectives. In Fig. 2,  $A_1$  strictly dominates  $A_3$ , but  $A_2$  does not as the objective vector  $(10, 4)$  is not strictly dominated by any objective vector in  $A_2$ .

## 2.2 Comparison Methods

Quality measures have been introduced to compare the outcomes of multiobjective optimizers in a quantitative manner. Certainly, the simplest comparison method would be to check whether an outcome is better than another with respect to the three dominance relations  $\triangleright$ ,  $\succ$ , and  $\succ\succ$ . We have demonstrated this in the context of the discussion of Fig. 2. The reason, however, why quality measures have been used is to be able to make more precise statements:

- If one algorithm is better than another, can we express how much better it is?
- If no algorithm can be said to be better than the other, are there certain aspects in which respect we can say the former is better than the latter?

Hence, the key question when designing quality measures is how to best summarize approximation sets by means of a few characteristic numbers—similarly to statistics where the mean, the standard deviation, etc. are used to describe a probability distribution in a compact way. It is unavoidable to lose information by such a reduction, and the crucial point is not to lose the information one is interested in.

There are many examples of quality measures in the literature. Some aim at measuring the distance of an approximation set to the Pareto-optimal front: Van Veldhuizen [21], e.g., calculated for each solution

in the approximation set under consideration the Euclidean distance to the closest Pareto-optimal objective vector and then took the average over all of these distances. Other measures try to capture the diversity of an approximation set, e.g., the chi-square-like deviation measure used by Srinivas and Deb [18]. A further example is the hypervolume measure which considers the volume of the objective space dominated by an approximation set [26]. In these three cases, an approximation set is assigned a real number which is meant to reflect (certain aspects of) the quality of an approximation set. Alternatively, one can assign numbers to pairs of approximation sets. Zitzler and Thiele [26], e.g., introduced the coverage function which gives for a pair  $(A, B)$  of approximation sets the fraction of solutions in  $B$  that are weakly dominated by one or more solutions in  $A$ .

In summary, we can state that quality measures map approximation sets to the set of real numbers. The underlying idea is to quantify quality differences between approximation sets by applying common metrics (in the mathematical sense) to the resulting real numbers. This observation enables us to formally define what a quality measure is; however, we will use the term “quality indicator” in the following as “measure” is often used with different meanings.

**Definition 2 (Quality indicator)** *An  $m$ -ary quality indicator  $I$  is a function  $I : \Omega^m \rightarrow \mathbb{R}$ , which assigns each vector  $(A_1, A_2, \dots, A_m)$  of  $m$  approximation sets a real value  $I(A_1, \dots, A_m)$ .*

The measures discussed above are examples for unary and binary quality indicators; however, in principle a quality indicator can take an arbitrary number of arguments. Thereby, also other comparison methods that explicitly account for multiple runs and involve statistical testing procedures [7][11][8] can be expressed within this framework. Furthermore, often not a single indicator but rather a combination of different quality indicators is used in order to assess approximation sets. Van Veldhuizen and Lamont [22], for instance, applied a combination  $\mathbf{I} = (I_{GD}, I_S, I_{ONVG})$  of three indicators where  $I_{GD}(A)$  denotes the average distance of solutions in

$A$  to the Pareto-optimal front,  $I_S(A)$  measures the variance of distances between neighboring solutions in  $A$ , and  $I_{ONVG}(A)$  gives the number of elements in  $A$ . Accordingly, the combination (or quality indicator vector)  $\mathbf{I}$  can be regarded as a function that assigns each approximation set a triple of real numbers.

Quality indicators, though, need interpretation. In particular, we would like to formally describe statements such as “if and only if  $I_{GD}(A) = 0$ , then all solutions in  $A$  have zero distance to the Pareto-optimal front  $P$  and therefore  $A \subseteq P$  and also  $B \not\subseteq A$  for any  $B \in \Omega$ ”. To this end, we introduce two concepts. A pseudo-Boolean function  $E$  maps vectors of real numbers to Booleans. In the above example, we would define  $E(I_{GD}(A)) := (I_{GD}(A) = 0)$ , i.e.,  $E$  is true if and only if  $I_{GD}(A) = 0$ . Such a combination of one or more quality indicators  $\mathbf{I}$  and a Boolean function  $E$  is also called a comparison method  $C_{\mathbf{I},E}$ . In the example, the comparison method  $C_{I_{GD},E}$  based on  $I_{GD}$  and  $E$  would be defined as  $C_{I_{GD},E}(A, B) = E(I_{GD}(A))$ , and the conclusion is that  $C_{I_{GD},E}(A, B) \Leftrightarrow A \subseteq P \wedge B \not\subseteq A$ . In the following, we will focus on comparison methods that i) consider two approximation sets only and ii) use either only unary or only binary indicators (cf. Fig. 3).

**Definition 3 (Comparison method)** *Let  $A, B \in \Omega$  be two approximation sets,  $\mathbf{I} = (I_1, I_2, \dots, I_k)$  a combination of quality indicators, and  $E : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \{\text{false}, \text{true}\}$  a Boolean function which takes 2 real vectors of length  $k$  as arguments. If all indicators in  $\mathbf{I}$  are unary, the comparison method  $C_{\mathbf{I},E}$  defined by  $\mathbf{I}$  and  $E$  is a Boolean function of the form*

$$C_{\mathbf{I},E}(A, B) = E(\mathbf{I}(A), \mathbf{I}(B))$$

where  $\mathbf{I}(A') = (I_1(A'), I_2(A'), \dots, I_k(A'))$  for  $A' \in \Omega$ . If  $\mathbf{I}$  contains only binary indicators, the comparison method  $C_{\mathbf{I},E}$  is defined as

$$C_{\mathbf{I},E}(A, B) = E(\mathbf{I}(A, B), \mathbf{I}(B, A))$$

where  $\mathbf{I}(A', B') = (I_1(A', B'), I_2(A', B'), \dots, I_k(A', B'))$  for  $A', B' \in \Omega$ .

Whenever we will specify a particular comparison method  $C_{\mathbf{I},E}$ , we will write  $E :=$

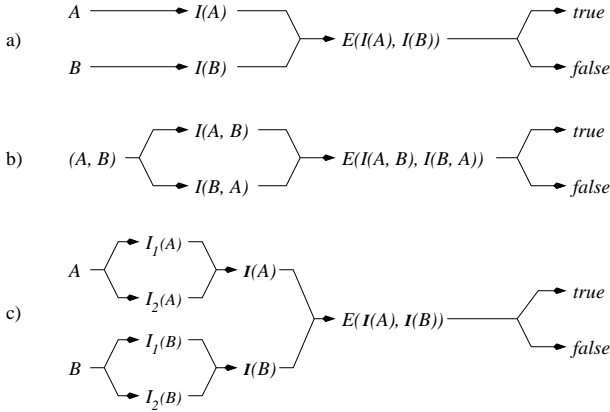


Figure 3: Illustration of the concept of a comparison method for a single unary quality indicator (a), a single binary quality indicator (b), and a combination of two unary quality indicators (c). In cases (a) and (b), first the indicator  $I$  is applied to the two approximation sets  $A, B$ . The resulting two real values are passed to the Boolean function  $E$ , which defines the outcome of the comparison. In case (c), each of the two indicators is applied to  $A$  and  $B$ , and the resulting two indicator values are combined in a vector  $\mathbf{I}(A)$  and  $\mathbf{I}(B)$  respectively. Afterwards, the Boolean function  $E$  decides the outcome of the comparison on the basis of these two real vectors.

$\langle \text{expression} \rangle$  instead of  $E(\dots) \Leftrightarrow \langle \text{expression} \rangle$  in order to improve readability. For instance,  $E := (I_1(A) > I_1(B))$  means that  $E((I_1(A), I_2(A), \dots, I_k(A)), (I_1(B), I_2(B), \dots, I_k(B)))$  is true if and only if  $I_1(A) > I_1(B)$ , given a combination of  $k$  unary indicators.

Definition 3 may appear overly formal for describing what a comparison method basically is, and furthermore it does not specify the actual conclusion (what does it mean if  $C_{I,E}(A, B)$  is true?). As we will see in the following, however, it provides a sound basis for studying the power of quality indicators—the power of indicating relationships (better, incomparable, etc.) between approximation sets.

## 2.3 Linking Comparison Methods and Dominance Relations

The goal of a comparative study is to reveal differences in performance between multiobjective optimizers, and the strongest statement we can make in this context is that an algorithm outperforms another one. Independently of what definition of “outperformance” we use, it always should be compliant with the most general notion in terms of the  $\triangleright$ -relation, i.e., the statement “algorithm  $a$  outperforms algorithm  $b$ ” should also imply that the outcome  $A$  of the first method is better than the outcome  $B$  of the second method ( $A \triangleright B$ ).<sup>1</sup> More accurate assessments may be possible if preference information is given [9], however, most studies assume that additional knowledge is not available, i.e., all objectives are to be considered equally important.

In this paper, we are interested in the question what conclusions can be drawn with respect to the dominance relations listed in Table 1 on the basis of a comparison method  $C_{I,E}$ . If  $C_{I,E}(A, B)$  is a sufficient condition for, e.g.,  $A \triangleright B$ , then this comparison method is capable of indicating that  $A$  is better than  $B$ , i.e.,  $C_{I,E}(A, B) \Rightarrow A \triangleright B$ . If  $C_{I,E}(A, B)$  is in addition a necessary condition for  $A \triangleright B$ , then the comparison method even indicates whether  $A$  is better than  $B$ , i.e.,  $C_{I,E}(A, B) \Leftrightarrow A \triangleright B$ . In the following, we will use the terms compatibility and completeness in order to characterize a comparison method in the above manner.

### Definition 4 (Compatibility and completeness)

Let  $\blacktriangleright$  be an arbitrary binary relation on approximation sets. The comparison method  $C_{I,E}$  is denoted as  $\blacktriangleright$ -compatible if either for any  $A, B \in \Omega$

$$C_{I,E}(A, B) \Rightarrow A \blacktriangleright B$$

or for any  $A, B \in \Omega$

$$C_{I,E}(A, B) \Rightarrow B \blacktriangleright A$$

The comparison method  $C_{I,E}$  is denoted as  $\blacktriangleright$ -

<sup>1</sup>Recall that we assume that only a single optimization run is performed per algorithm.



complete if either for any  $A, B \in \Omega$

$$A \triangleright B \Rightarrow C_{I,E}(A, B)$$

or for any  $A, B \in \Omega$

$$B \triangleright A \Rightarrow C_{I,E}(A, B)$$

To illustrate this terminology, let us go back to the example depicted in Fig. 2 and consider the following binary indicator  $I_\epsilon$ , which is inspired by concepts presented in [15]:

**Definition 5 (Binary  $\epsilon$ -indicator)** *Suppose without loss of generality a minimization problem with  $n$  positive objectives, i.e.,  $Z \subseteq \mathbb{R}^{+n}$ . An objective vector  $\mathbf{z}^1 = (z_1^1, z_2^1, \dots, z_n^1) \in Z$  is said to  $\epsilon$ -dominate another objective vector  $\mathbf{z}^2 = (z_1^2, z_2^2, \dots, z_n^2) \in Z$ , written as  $\mathbf{z}^1 \succepsilon_\epsilon \mathbf{z}^2$ , if and only if*

$$\forall 1 \leq i \leq n : z_i^1 \leq \epsilon \cdot z_i^2$$

for a given  $\epsilon > 0$ . We define the binary  $\epsilon$ -indicator  $I_\epsilon$  as

$$I_\epsilon(A, B) = \inf_{\epsilon \in \mathbb{R}} \{ \forall \mathbf{z}^2 \in B \exists \mathbf{z}^1 \in A : \mathbf{z}^1 \succepsilon_\epsilon \mathbf{z}^2 \}$$

for any two approximation sets  $A, B \in \Omega$ .

The  $\epsilon$ -indicator gives the factor by which an approximation set is worse than another with respect to all objectives, or to be more precise:  $I_\epsilon(A, B)$  equals the minimum factor  $\epsilon$  such that for any solution in  $B$  there is at least one solution in  $A$  that is not worse by a factor of  $\epsilon$  in all objectives.<sup>2</sup> In practice, the  $\epsilon$  value can be calculated as

$$I_\epsilon(A, B) = \max_{\mathbf{z}^2 \in B} \min_{\mathbf{z}^1 \in A} \max_{1 \leq i \leq n} \frac{z_i^1}{z_i^2}$$

<sup>2</sup>In the same manner, an additive  $\epsilon$ -indicator  $I_{\epsilon+}$  can be defined:

$$I_{\epsilon+}(A, B) = \inf_{\epsilon \in \mathbb{R}} \{ \forall \mathbf{z}^2 \in B \exists \mathbf{z}^1 \in A : \mathbf{z}^1 \succepsilon_{\epsilon+} \mathbf{z}^2 \}$$

where  $\mathbf{z}^1 \succepsilon_{\epsilon+} \mathbf{z}^2$  if and only if

$$\forall 1 \leq i \leq n : z_i^1 \leq \epsilon + z_i^2$$

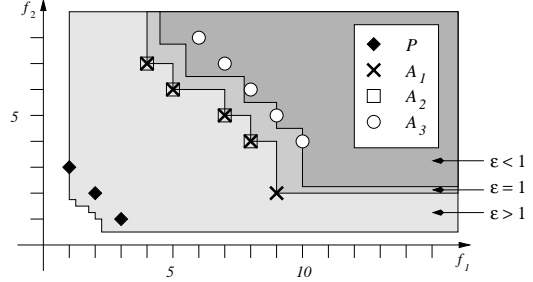


Figure 4: The dark-shaded area depicts the subspace that is  $\epsilon$ -dominated by the solutions in  $A_1$  for  $\epsilon = \frac{9}{10}$ ; the medium-shaded area represents the subspace weakly dominated by  $A_1$  (equivalent to  $\epsilon = 1$ ); the light-shaded area refers to the subspace  $\epsilon$ -dominated by the solutions in  $A_1$  for  $\epsilon = 4$ . Note that the areas are overlapping, i.e., the medium-shaded area includes the dark-shaded one, and the light-shaded area includes both of the other areas.

For instance,  $I_\epsilon(A_1, A_2) = 1$ ,  $I_\epsilon(A_1, A_3) = \frac{9}{10}$ , and  $I_\epsilon(A_1, P) = 4$  in our previous example (cf. Fig. 4). In the single-objective case,  $I_\epsilon(A, B)$  simply is the ratio between the two objective values represented by  $A$  and  $B$ .

Now, what comparison methods can be constructed using the  $\epsilon$ -indicator? Consider, e.g., the Boolean function  $E := (I_\epsilon(B, A) > 1)$ . The corresponding comparison method  $C_{I_\epsilon, E}$  is  $\triangleright$ -complete as  $A \triangleright B$  implies that that  $I_\epsilon(B, A) > 1$ . On the other hand,  $C_{I_\epsilon, E}$  is not  $\triangleright$ -compatible as  $A \parallel B$  also implies that  $I_\epsilon(B, A) > 1$ . If we choose a slightly modified Boolean function  $F := (I_\epsilon(A, B) \leq 1 \wedge I_\epsilon(B, A) > 1)$ , then we obtain a comparison method  $C_{I_\epsilon, F}$  that is both  $\triangleright$ -compatible and  $\triangleright$ -complete. The differences between the two comparison methods are graphically depicted in Fig. 5.

In the remainder of this paper, we will theoretically study and classify quality indicators using the above framework. Given a particular quality indicator (or a combination of several indicators), we will investigate whether there exists a Boolean function such that the resulting comparison method is compatible and in addition complete with respect to the vari-

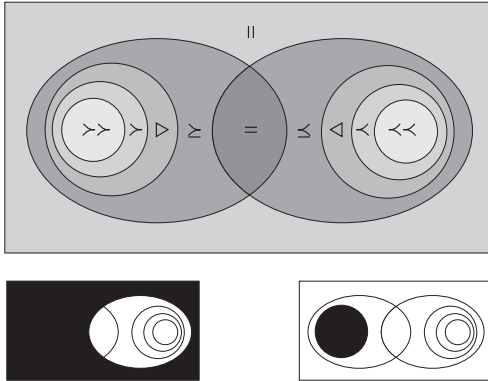


Figure 5: Top: Partitioning of the set of ordered pairs  $(A, B) \in \Omega^2$  of approximation sets into (overlapping) subsets induced by the different dominance relations; each subset labeled with a certain relation  $\blacktriangleright$  contains those pairs  $(A, B)$  for which  $A \blacktriangleright B$ . Note that this is only a schematic representation, e.g., there are no pairs  $(A, B)$  with  $A \succeq B$ ,  $A \not\succeq B$ , and  $A \neq B$ . Bottom: The black area stands for those ordered pairs  $(A, B)$  for which  $I_\epsilon(B, A) > 1$  (left) resp.  $I_\epsilon(A, B) \leq 1 \wedge I_\epsilon(B, A) > 1$  (right).

ous dominance relations. That is we determine how powerful existing quality indicators are in terms of their capability of indicating *that* or *whether*  $A \triangleright B$ ,  $A \succ B$ ,  $A \parallel B$ , etc. The next section is devoted to unary quality indicators, while binary indicators will be discussed in Section 4.

### 3 Comparison Methods Based on Unary Quality Indicators

Unary quality indicators are most commonly used in the literature; what makes them attractive is their capability of assigning quality values to an approximation set independent of other sets under consideration. They have limitations, though, and there are differences in the power of existing indicators as will be shown in the following.

#### 3.1 Limitations

Naturally, many studies have attempted to capture the multiobjective nature of approximation sets by deriving distinct indicators for the distance to the Pareto-optimal front and the diversity within the approximation front. Therefore, the question arises whether in principle there exists such a combination of, e.g., two indicators—one for distance, one for diversity—such that we can detect *whether* an approximation set is better than another. Such a combination of indicators, applicable to any type of problem, would be ideal because then any approximation set could be characterized by two real numbers that reflect the different aspects of the overall quality. The variety among the indicators proposed suggests that this goal is, at least, difficult to achieve. The following theorem shows that in general it cannot be achieved.

**Theorem 1** *Suppose an optimization problem with  $n \geq 2$  objectives where the objective space is  $Z = \mathbb{R}^n$ . Then, there exists no comparison method  $C_{\mathbf{I}, E}$  based on a finite combination  $\mathbf{I}$  of unary quality indicators that is  $\triangleright$ -compatible and  $\triangleright$ -complete at the same time, i.e.,*

$$C_{\mathbf{I}, E}(A, B) \Leftrightarrow A \triangleright B$$

for any approximation sets  $A, B \in \Omega$ .

That is for any combination  $\mathbf{I}$  of a finite number of unary quality indicators we cannot find a Boolean function  $E$  such that the corresponding comparison method is  $\triangleright$ -compatible *and*  $\triangleright$ -complete. Or in other words: the number of criteria, that determine what a good approximation set is, is infinite.

We only sketch the proof here, the details can be found in the appendix. First, we need the following fundamental results from set theory [10]:

- $\mathbb{R}$ ,  $\mathbb{R}^k$ , and any open interval  $(a, b)$  in  $\mathbb{R}$  resp. hypercube  $(a, b)^k$  in  $\mathbb{R}^k$  have the same cardinality, denoted as  $2^{\aleph_0}$ , i.e., there is a bijection from any of these sets to any other;
- If a set  $S$  has cardinality  $2^{\aleph_0}$ , then the cardinality of the power set  $\mathcal{P}(S)$  of  $S$  is  $2^{2^{\aleph_0}}$ , i.e., there is

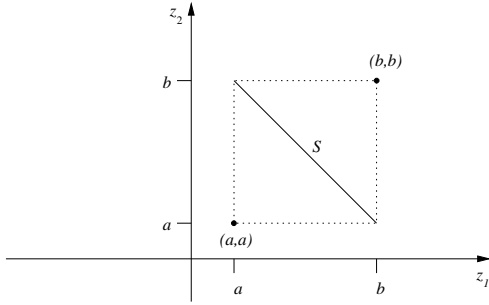


Figure 6: Illustration of the construction used in Theorem 1 for a two-dimensional minimization problem. We consider an open rectangle  $(a, b)^2$  and define an open line  $S$  within. For  $S$  holds that any two objective vectors contained are incomparable to each other, and therefore any subset  $A \subseteq S$  is an approximation set.

no injection from  $\mathcal{P}(S)$  to any set of cardinality  $2^{N_0}$ .

As we consider the most general case where  $Z = \mathbb{R}^n$ , we can construct a set  $S$  (cf. Fig. 6) such that any two points contained are incomparable to each other. Accordingly, any subset  $A$  of  $S$  is an approximation set and the power set of  $S$ , the cardinality of which is  $2^{2^{N_0}}$ , is exactly the set of all approximation sets  $A \subseteq S$ . We will then show that any two approximation sets  $A, B \subseteq S$  with  $A \neq B$  must differ in at least one of the  $k$  indicator values. Therefore, an injection from a set of cardinality  $2^{2^{N_0}}$  to  $\mathbb{R}^k$  is required, which finally leads to a contradiction.

Note that Theorem 1 also holds (i) if we only assume that  $Z$  contains an open hypercube in  $\mathbb{R}^n$  for which  $C_{I,E}$  has the desired property, and (ii) if we consider any other relation from Table 1 (for  $\parallel$  and  $\succeq$  it follows directly from Theorem 1, for  $\succ$  and  $\succ\succeq$  the proof has to be slightly modified).

Given this result, one may ask under which conditions the construction of such a comparison method is possible. For instance, such a comparison method exists if we allow an infinite number of indicators. The empirical attainment function [8], when applied

to single approximation sets, can be understood as a combination of  $|Z|$  unary indicators, where  $|Z|$  denotes the cardinality of  $Z$ . If  $Z = \mathbb{R}^n$ , then this combination comprises an infinite number of unary indicators. On its basis, a  $\triangleright$ -compatible and  $\triangleright$ -complete comparison method can be constructed.

The situation also changes, if we require that each approximation set contains at maximum  $l$  objective vectors.

**Corollary 1** *Let  $Z = \mathbb{R}^n$ . It exists a unary indicator  $I$  and a Boolean function  $E$  such that*

$$C_{I,E}(A, B) \Leftrightarrow A \triangleright B$$

for any  $A, B \in \Omega$  with  $|A|, |B| \leq l$ .

*Proof.* Without loss of generality we restrict ourselves to  $Z = (0, 1)^n$  in the proof. The indicator  $I$  is constructed as follows:

$$I(A) = 0.d_1^1 d_2^1 \dots d_l^1 d_1^2 d_2^2 \dots d_l^2 d_1^3 \dots$$

where  $d_j^i$  denotes the  $i$ th digit after the decimal point of the  $j$ th element in  $A$ . If  $A$  contains less than  $l$  elements, the first element is duplicated as many times as necessary. Accordingly, there is an injective function  $R$  that maps each real number in  $(0, 1)$  to an approximation set. If we define  $E$  as  $E := (R(I(A)) \triangleright R(I(B)))$ , the corresponding comparison method  $C_{I,E}$  has the desired properties.  $\square$

The theorem, however, is rather of theoretical than of practical use. The indicator constructed in the proof is able to indicate *whether*  $A$  is better than  $B$ , but it does not express how much better it is—this is one of the motives for using quality indicators. What we actually want is to apply a metric to the indicator values. Therefore, a reasonable requirement for a useful combination of indicators may be that if  $A$  is better than or equal to  $B$ , then  $A$  is at least as good as  $B$  with respect to all  $k$  indicators, i.e.:

$$A \succeq B \Rightarrow (\forall 1 \leq i \leq k : I_i(A) \geq I_i(B))$$

That this condition holds is an implicit assumption made in many studies. If we now restrict the size of

the approximation sets to  $l$  and assume an indicator combination with the above property, can we then detect *whether*  $A$  is better than  $B$ ? To answer this question, we will investigate a slightly reformulated statement, namely

$$A \succeq B \Leftrightarrow (\forall 1 \leq i \leq k : I_i(A) \geq I_i(B))$$

as this is equivalent to

$$A \triangleright B \Leftrightarrow \left( \begin{array}{l} \forall 1 \leq i \leq k : I_i(A) \geq I_i(B) \\ \exists 1 \leq j \leq k : I_j(A) > I_j(B) \end{array} \right) \wedge$$

Furthermore, we will only consider the simplest case where  $l = 1$ , i.e., each approximation set consists of a single objective vector.

**Theorem 2** *Suppose an optimization problem with  $n \geq 2$  objectives where the objective space is  $Z = \mathbb{R}^n$ . Let  $\mathbf{I} = (I_1, I_2, \dots, I_k)$  be a combination of  $k$  unary quality indicators and  $E := (\forall 1 \leq i \leq k : I_i(\{\mathbf{z}^1\}) \geq I_i(\{\mathbf{z}^2\}))$  a Boolean function such that*

$$C_{\mathbf{I}, E}(\{\mathbf{z}^1\}, \{\mathbf{z}^2\}) \Leftrightarrow \mathbf{z}^1 \succeq \mathbf{z}^2$$

for any pair of objective vectors  $\mathbf{z}^1, \mathbf{z}^2 \in Z$ . Then, the number of indicators is greater than or equal to the number of objectives, i.e.,  $k \geq n$ .

*Proof.* See appendix.

This theorem is a formalization of what is intuitively clear: we cannot reduce the dimensionality of the objective space without losing information. We need at least as many indicators as objectives to be able to detect *whether* an objective vector weakly dominates or dominates another objective vector. As a consequence, a fixed number of unary indicators is not sufficient for problems of arbitrary dimensionality even if we consider sets containing a single objective vector only.

In summary, we can state that the power of unary quality indicators is restricted. Theorem 1 proves that there does not exist any comparison methods based on unary indicators that is  $\triangleright$ -compatible and  $\triangleright$ -complete at the same time. This rules out also other combinations, Table 2 shows which. It reveals that the best we can achieve is either

compatibility	completeness						
	none	$\succ\succ$	$\succ$	$\triangleright$	$\not\succ\succ$	$\not\succ$	$\not\triangleright$
$\succ\succ$	+	-	-	-	-	-	-
$\succ$	+	?	-	-	-	-	-
$\triangleright$	+	?	?	-	-	-	-
$\not\succ\succ$	+	+	+	+	-	?	?
$\not\succ$	+	+	+	+	-	-	?
$\not\triangleright$	+	+	+	+	-	-	-

Table 2: Overview of possible compatibility/completeness combinations with unary quality indicators. A minus means there is no comparison method  $C_{\mathbf{I}, E}$  that is compatible regarding the row-relation and complete regarding the column-relation. A plus indicates that such a comparison method is known, while a question mark stands for a combination for which it is unclear whether a corresponding comparison method exists.

$\succ\succ$ -compatibility without any completeness, or  $\not\triangleright$ -compatibility in combination with  $\triangleright$ -completeness. That means we either can make strong statements (“ $A$  strongly dominates  $B$ ”) for only a few pairs  $A \triangleright B$ ; or we can make weaker statements (“ $A$  is not worse than  $B$ ”, i.e.,  $A \succeq B$  or  $A \parallel B$ ) for all pairs  $A \triangleright B$ .

## 3.2 Classification

We now will review existing unary quality indicators according to the inferential power of the comparison methods that can be constructed on their basis:  $\triangleright$ -compatible,  $\not\triangleright$ -compatible, and not compatible with any relation listed in Table 2. Table 3 provides an overview of the various indicators discussed here. In this context, we would also like to point out the relationships between the dominance relations, e.g.,  $\succ\succ$ -compatibility implies  $\triangleright$ -compatibility,  $\not\triangleright$ -compatibility implies  $\not\succ$ -compatibility, and  $\triangleright$ -completeness implies  $\succ\succ$ -completeness.

### 3.2.1 $\triangleright$ -Compatibility

The use of  $\triangleright$ -compatible comparison methods based on unary indicators is restricted according to Theorem 2: in order to detect dominance between objec-

tive vectors at least as many indicators as objectives are required. Hence, it is not surprising that, to our best knowledge, no  $\triangleright$ -compatible comparison methods have been proposed in the literature; their design, though, is possible:

- Suppose a minimization problem and let

$$\begin{aligned} I_1^{HC}(A) &= \sup_{a \in \mathbf{R}} \{ \{(a, a, \dots, a)\} \triangleright A \} \\ I_2^{HC}(A) &= \inf_{b \in \mathbf{R}} \{ \{(b, b, \dots, b)\} \triangleleft A \} \end{aligned}$$

We assume that  $Z$  is bounded, i.e.,  $I_1^{HC}(A)$  and  $I_2^{HC}(A)$  always exists. As illustrated in Fig. 7, the two indicator values characterize a hypercube that contains all objective vectors in  $A$ . If we define the indicator  $\mathbf{I}_{HC} = (I_1^{HC}, I_2^{HC})$  and the Boolean function  $E$  as  $E := (I_2^{HC}(A) < I_1^{HC}(B))$ , then the comparison method  $C_{\mathbf{I}_{HC}, E}$  is  $\triangleright$ -compatible.

- Suppose a minimization problem and let

$$I_i^O(A) = \inf_{a \in \mathbf{R}} \{ \forall (z_1, \dots, z_n) \in A : z_i \leq a \}$$

for  $1 \leq i \leq n$  and

$$I_{n+1}^O(A) = \begin{cases} 0 & \text{if } A \text{ contains two or} \\ & \text{more elements} \\ 1 & \text{else} \end{cases}$$

The idea behind these indicators is similar to above. We consider the smallest hyperrectangle that entirely encloses  $A$ . This hyperrectangle comprises exactly one point  $O$  that is weakly dominated by all members in  $A$ ; in the case of a two-dimensional minimization problem, it is the upper right corner of the enclosing rectangle (cf. Fig. 7). We see that  $I_1^O, \dots, I_n^O$  are the coordinates of this point  $O$ .  $I_{n+1}^O$  serves to distinguish between single objective vectors and larger approximation sets. Let  $\mathbf{I}_O = (I_1^O, \dots, I_{n+1}^O)$  and define the Boolean function  $E$  as  $E := (\forall 1 \leq i \leq n+1 : I_i^O(A) < I_i^O(B))$ . Then, the comparison method  $C_{\mathbf{I}_O, E}$  is  $\triangleright$ -compatible; it detects dominance between an approximation set and those objective vectors that are dominated by all members of this approximation set.

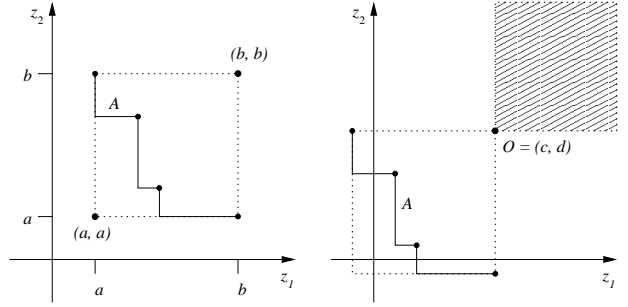


Figure 7: Two indicators capable of indicating that  $A \triangleright B$  for some  $A, B \in \Omega$ . On the left hand side, it is depicted how the  $\mathbf{I}_{HC}$  indicator defines a hypercube around an approximation set  $A$ , where  $I_1^{HC}(A) = a$  and  $I_2^{HC}(A) = b$ . The right picture is related to the  $\mathbf{I}_O$  indicator: for any objective vector in the shaded area we can detect that it is dominated by the approximation set  $A$ . Here,  $I_1^O(A) = c$ ,  $I_2^O(A) = d$ , and  $I_3^O(A) = 0$ .

Note that both comparison methods are even  $\succ\succ$ -compatible, but neither is complete with regard to any dominance relation.

Moreover, some unary indicators can also be used to design a  $\triangleright$ -compatible comparison method if the Pareto-optimal front  $P$  is known. Consider, e.g., the following unary  $\epsilon$ -indicator  $I_{\epsilon 1}$  that is based on the binary  $\epsilon$ -indicator from Definition 5:

$$I_{\epsilon 1}(A) = I_{\epsilon}(A, P)$$

Obviously,  $I_{\epsilon 1}(A) = 1$  implies  $A = P$ . Thus, in combination with the Boolean function  $E := (I_{\epsilon 1}(A) = 1 \wedge I_{\epsilon 1}(B) > 1)$  a comparison method can be defined that is  $\triangleright$ -compatible and detects that  $A$  is better than  $B$  for all pairs  $A, B \in \Omega$  with  $A = P$  and  $B \neq P$ . The same construction can be made for some other indicators, e.g., the hypervolume indicator, as well. Nevertheless, these comparison methods are only applicable if some of the algorithms under consideration can actually generate the Pareto-optimal front.

### 3.2.2 $\not\prec$ -Compatibility

Consider the above unary  $\epsilon$ -indicator  $I_{\epsilon 1}$ . For any pair  $A, B \in \Omega$  it holds

$$A \succ B \Rightarrow I_{\epsilon 1}(A) < I_{\epsilon 1}(B)$$

and (which follows from this)

$$I_{\epsilon 1}(A) < I_{\epsilon 1}(B) \Rightarrow A \not\prec B \Rightarrow A \not\prec B$$

Therefore, the comparison method  $C_{I_{\epsilon 1}, E}$  with  $E := (I_{\epsilon 1}(A) < I_{\epsilon 1}(B))$  is  $\not\prec$ -compatible and  $\succ$ -complete, but neither  $\triangleright$ - nor  $\succ$ -complete. That is whenever  $A \succ B$ , we will be able to state that  $A$  is not worse than  $B$ . On the other hand, there are cases  $A \succ B$  for which this conclusion cannot be drawn, although  $A$  is actually not worse than  $B$ . The same holds for the two indicators proposed by [6] and [1]. We will not discuss these in detail and only remark that the following example can be used to show that both indicators in combination with the Boolean function  $E := (I(A) < I(B))$  are not  $\succ$ -complete (and  $\triangleright$ -complete): the Pareto-optimal front is  $P = \{(1, 1)\}$ , and  $A = \{(4, 2)\}$  and  $B = \{(4, 3)\}$ .

The hypervolume indicator  $I_H$  [26][24] is the only unary indicator we are aware of that is capable of detecting that  $A$  is not worse than  $B$  for all pairs  $A \triangleright B$ . It gives the hypervolume of that portion of the objective space that is weakly dominated by an approximation set  $A$ .<sup>3</sup> We notice that from  $A \triangleright B$  follows that  $I_H(A) > I_H(B)$ ; the reason is that  $A$  must contain at least one objective vector that is not weakly dominated by  $B$ , thus, a certain portion of the objective space is dominated by  $A$  but not by  $B$ . This observation implies both  $\not\prec$ -compatibility and  $\triangleright$ -completeness.

Van Veldhuizen [21] suggested an indicator, the error ratio  $I_{ER}$ , on the basis of which a  $\not\prec$ -compatible (but not  $\not\prec$ -compatible) comparison method can be defined.  $I_{ER}(A)$  gives the ratio of Pareto-optimal objective vectors to all objective vectors in the approximation set  $A$ . Obviously, if  $I_{ER}(A) > 0$ , i.e.,  $A$  contains at least one Pareto-optimal solution, then there

<sup>3</sup>Note that  $Z$  has to be bounded, i.e., there must exist a hypercube in  $\mathbb{R}^n$  that encloses  $Z$ . If this requirement is not fulfilled, it can be easily achieved by an appropriate transformation.

exists no  $B \in \Omega$  with  $B \succ A$ . On the other hand, if  $A$  consist of only a single Pareto-optimal point, then  $I_{ER}(A) \geq I_{ER}(B)$  for all  $B \triangleright A$ ; if  $B$  contains not only Pareto-optimal points, then  $I_{ER}(A) > I_{ER}(B)$ . Therefore,  $C_{(I_{ER}, E)}$  with  $E := (I_{ER}(A) > I_{ER}(B))$  is not  $\not\prec$ -compatible. However, if we consider just the total number (rather than the ratio) of Pareto-optimal points in the approximation set, we obtain  $\not\prec$ -compatibility. This also holds for the indicator used in [20], which gives the ratio of the number of Pareto-optimal solutions in  $A$  to the cardinality of the Pareto-optimal front. Nevertheless, the power of these comparison methods is limited because none of them is complete with respect to any dominance relation.

### 3.2.3 Incompatibility

Section 3.1 has revealed the difficulties when trying to separate the overall quality of approximation sets into distinct aspects. Nevertheless, it would be desirable if we could look at certain criteria such as diversity separately, and accordingly several authors suggested formalizations of specific aspects by means of unary indicators. However, we have to be aware that often these indicators do *in general* neither indicate *that*  $A \triangleright B$  nor  $A \not\prec B$ .

One class of indicators that do not allow any conclusions to be drawn regarding the dominance relationship between approximation sets is represented by the various diversity indicators [18][17][24][16][3][23]. If we consider a pair  $A, B \in \Omega$  with  $A \triangleright B$ , in general the indicator value of  $A$  can be less or greater than or even equal to the value assigned to  $B$  (for the diversity indicators referenced above). Therefore, the comparison methods based on these indicators are neither compatible nor complete with respect to any dominance relation or complement of it. For a more detailed discussion of some of the above indicators, the interested reader is referred to [14].

The same holds for the three indicators proposed in [21]: overall nondominated vector generation  $I_{ONVG}$ , generational distance  $I_{GD}$ , and maximum Pareto front error  $I_{ME}$ . The first just gives the number of

indicator	name / reference	Boolean function	compatibility	completeness
$I_{HC}$	enclosing hypercube indicator / Section 3.2.1	$I_2^{HC}(A) < I_1^{HC}(B)$	$\succ\succ$	-
$I_O$	objective vector indicator / Section 3.2.1	$I_i^O(A) > I_i^O(B)$	$\succ\succ$	-
$I_H$	hypervolume indicator / [26]	$I_H(A) > I_H(B)$	$\not\succeq$	$\triangleright$
$I_W$	average best weight combination / [6]	$I_W(A) < I_W(B)$	$\not\succeq$	$\succ\succ$
$I_D$	distance from reference set / [1]	$I_D(A) < I_D(B)$	$\not\succeq$	$\succ\succ$
$I_{\epsilon 1}$	unary $\epsilon$ -indicator / Section 3.2.2	$I_{\epsilon 1}(A) < I_{\epsilon 1}(B)$	$\not\succeq$	$\succ\succ$
$I_{PF}$	fraction of Pareto-optimal front covered / [20]	$I_{PF}(A) > I_{PF}(B)$	$\not\succeq$	-
$I_P$	number of Pareto points contained / Section 3.2.2	$I_P(A) > I_P(B)$	$\not\succeq$	-
$I_{ER}$	error ratio / [21]	$I_{ER}(A) > 0$	$\not\succeq$	-
$I_{CD}$	chi-square-like deviation indicator / [18]	$I_{CD}(A) < I_{CD}(B)$	-	-
$I_S$	spacing / [17]	$I_S(A) < I_S(B)$	-	-
$I_{ONVG}$	overall nondominated vector generation / [21]	$I_{ONVG}(A) > I_{ONVG}(B)$	-	-
$I_{GD}$	generational distance / [21]	$I_{GD}(A) < I_{GD}(B)$	-	-
$I_{ME}$	maximum Pareto front error / [21]	$I_{ME}(A) < I_{ME}(B)$	-	-
$I_{MS}$	maximum spread / [24]	$I_{MS}(A) > I_{MS}(B)$	-	-
$I_{MD}$	minimum distance between two solutions / [16]	$I_{MD}(A) > I_{MD}(B)$	-	-
$I_{CE}$	coverage error / [16]	$I_{CE}(A) < I_{CE}(B)$	-	-
$I_{DU}$	deviation from uniform distribution / [3]	$I_{DU}(A) < I_{DU}(B)$	-	-
$I_{OS}$	Pareto spread / [23]	$I_{OS}(A) > I_{OS}(B)$	-	-
$I_A$	accuracy / [23]	$I_A(A) > I_A(B)$	-	-
$I_{NDC}$	number of distinct choices / [23]	$I_{NDC}(A) > I_{NDC}(B)$	-	-
$I_{CL}$	cluster / [23]	$I_{CL}(A) < I_{CL}(B)$	-	-

Table 3: Overview of unary indicators. Each entry corresponds to a specific comparison method defined by the indicator and the Boolean function in that row. With respect to compatibility and completeness, not all relations are listed but only the strongest as, e.g.,  $\succ\succ$ -compatibility implies  $\triangleright$ -compatibility (cf. Section 3.2).

elements in the approximation set, and it is obvious that it does not provide sufficient information to conclude  $A \triangleright B$ ,  $A \not\succeq B$ , etc. Why this also applies to the other two, both distance indicators, will only be sketched here. Assume a two-dimensional minimization problem for which the Pareto-optimal front  $P$  consists of the two objective vectors  $(1, 0)$  and  $(0, 10)$ . Now, consider the three sets  $A = \{(2, 5)\}$ ,  $B = \{(3, 9)\}$ , and  $C = \{(10, 10)\}$ . For both distance indicators holds  $I(B) < I(A) < I(C)$ , but  $A \succ\succ B \succ\succ C$ , provided that Euclidean distance is considered. Thus, we cannot conclude whether one set is better or worse than another by just looking at the order of the indicator values. A similar argument as for the generational distance applies to the coverage error indicator presented in [16]; the only difference is that the coverage error denotes the minimum distance to the Pareto-optimal front instead of the average distance.

Finally, one can ask whether it is possible to combine several indicators for which no  $\not\succeq$ -compatible comparison method exists in such a way that the resulting indicator vector allows to detect that  $A$  is not worse than  $B$ . Van Veldhuizen and Lamont [22], for instance, used generational distance and overall nondominated vector generation in conjunction with the diversity indicator of [17], while Deb et al. [4] applied a similar combination of diversity and distance indicators. Other examples can be found in, e.g., [2] and [16]. As in all of these cases counterexamples can be constructed that show the corresponding comparison methods to be not  $\not\succeq$ -compatible, the above question remains open and is not investigated in more depth here.

## 4 Comparison Methods Based on Binary Quality Indicators

Binary quality indicators can be used to overcome the difficulties with unary indicators. However, they also have a drawback: when we compare  $t$  algorithms using a single binary indicator, we obtain  $t(t-1)$  distinct indicator values—in contrast to the  $t$  values in the case of a unary indicator. This renders the analysis and the presentation of the results more difficult. Nevertheless, Theorem 1 suggests that this is in the nature of multiobjective optimization problems.

### 4.1 Limitations

In principle, there are no such theoretical limitations of binary indicators as for unary indicators. For instance, the indicator

$$I(A, B) = \begin{cases} 4 & A \succ\!\succ B \\ 3 & A \succ B \\ 2 & A \triangleright B \\ 1 & A = B \\ 0 & \text{else} \end{cases}$$

allows to construct comparison methods compatible and complete with regard to any of the dominance relations. However, this usually does not hold for existing, practically useful binary indicators, in particular for those indicators that are, as Knowles and Corne [13] denote it, *symmetric*, i.e.,  $I(A, B) = c - I(B, A)$  for a constant  $c$ . Although, symmetric indicators are attractive as only half the number of indicator values has to be considered in comparison to a general binary indicator, their inferential power is restricted as we will show in the following.

Without loss of generality, suppose that  $c = 0$ , i.e.,  $I(A, B) = -I(B, A)$ ; otherwise consider the transformation  $I'(A, B) = c/2 - I(A, B)$ . The question is whether we can construct a  $\triangleright$ -compatible and  $\triangleright$ -complete comparison method based on this indicator; according to the discussion in Section 3.1, we assume that  $E =: (I(A, B) > I(B, A))$ .

**Theorem 3** *Let  $I$  be a binary indicator with  $I(A, B) = -I(B, A)$  for  $A, B \in \Omega$  and  $E$  a Boolean*

*function with  $E =: I(A, B) > I(B, A)$ . If the corresponding comparison method  $C_{I,E}$  is  $\triangleright$ -compatible and  $\triangleright$ -complete, then  $I(A, B) = 0$  for all  $A, B \in \Omega$  with  $A = B$  or  $A \parallel B$ .*

*Proof.* Let  $A, B \in \Omega$ . From  $A \triangleright B \Leftrightarrow I(A, B) > I(B, A)$  follows that  $A \not\triangleright B \Leftrightarrow I(A, B) \leq I(B, A)$  and therefore  $A \parallel B \vee A = B \Leftrightarrow A \not\triangleright B \wedge B \not\triangleright A \Leftrightarrow I(A, B) = I(B, A)$ . From the symmetry  $I(A, B) = -I(B, A)$  then follows that  $A \parallel B \vee A = B$  is equivalent to  $I(A, B) = 0$ .  $\square$

A consequence of this theorem is that a symmetric, binary indicator, for which  $A \triangleright B \Leftrightarrow I(A, B) > I(B, A)$ , can detect whether  $A$  is better than  $B$ , but not whether  $A \succeq B$ ,  $A \parallel B$ , or  $A = B$ . On the other hand, it follows from  $I(A, B) \neq 0$  for a pair  $A \parallel B$  that  $C_{I,E}$  cannot be  $\triangleright$ -compatible, if it is  $\triangleright$ -complete. We will use this result in the following discussion of existing binary indicators.

### 4.2 Classification

In contrast to unary indicators, only a few binary indicators can be found in the literature. We will classify them according to the criterion whether a corresponding comparison method exists that is  $\blacktriangleright$ -compatible and  $\blacktriangleright$ -complete with regard to a specific relation  $\blacktriangleright$ .

As mentioned in Section 2.2, Zitzler and Thiele [26] suggested the coverage indicator  $I_C$  where  $I_C(A, B)$  gives the fraction of solutions in  $B$  that are weakly dominated by at least one solution in  $A$ .  $I_C(A, B) = 1$  is equivalent to  $A \succeq B$  ( $A$  weakly dominates  $B$ ) and therefore comparison methods  $C_{I_C,E}$  compatible and complete with regard to the  $\triangleright$ ,  $\succeq$ ,  $\parallel$ , and  $=$  relations can be constructed. Furthermore, with  $E := (I_C(A, B) = 1 \wedge I_C(B, A) = 0)$  we obtain a comparison method  $C_{I_C,E}$  that is  $\succ$ -compatible and  $\succ$ -complete.

Hansen and Jaszkievicz [9] proposed three symmetric, binary indicators  $I_{R_1}$ ,  $I_{R_2}$ , and  $I_{R_3}$  that are based on a set of utility functions. The utility functions can be used to formalize and incorporate pref-



erence information; however, if no additional knowledge is available, Hansen and Jaszkievicz suggest to use a set of weighted Tchebycheff utility functions. In this case, the resulting comparison methods are in general  $\triangleright$ -complete but not  $\triangleright$ -compatible as Theorem 3 applies ( $I(A, B)$  can be greater or less than 0 if  $A \parallel B$ ). Accordingly, these indicators in general do not allow to construct a comparison method that is both compatible and complete with respect to any of the relations in Table 1. However, it has to be emphasized here that these indicators have been designed with regard to the incorporation of preference information.

In [24], a binary version  $I_{H2}$  of the hypervolume indicator  $I_H$  [26] was proposed; the same indicator was used in [12].  $I_{H2}(A, B)$  is defined as the hypervolume of the subspace that is weakly dominated by  $A$  but not by  $B$ . From  $I_{H2}(A, B) = 0$  follows that  $B \succeq A$  and therefore, as with the coverage indicator, comparison methods  $C_{I_{H2}, E}$  compatible and complete regarding the  $\triangleright$ ,  $\succeq$ ,  $\parallel$ , and  $=$  relations are possible. However, there exists no  $\succ$ -compatible and  $\succ$ -complete or  $\succ\succ$ -compatible and  $\succ\succ$ -complete comparison method solely based on the binary hypervolume indicator.

Knowles and Corne [11] presented a comparison method based on the study by Fonseca and Fleming [7]. Although designed for the statistical analysis of multiple optimization runs, the method can be formulated in terms of an  $m$ -ary indicator  $I_{LI}$  if only one run is performed per algorithm or the algorithms are deterministic. We here restrict ourselves to the case  $m = 2$  as all of the following statements also hold for  $m > 2$ . A user-defined set of lines in the objective space, all of them passing the origin and none of them perpendicular to any of the axes, forms the scaffolding of Knowles and Corne’s approach. First, for each line the intersections with the attainment surfaces [7] defined by the approximation sets under consideration are calculated. The intersections are then sorted according to their distance to the origin, and the resulting order defines a ranking of the approximation sets with respect to this line. If only two approximation sets are considered, then  $I_{LI}(A, B)$  gives the fraction of the lines

for which  $A$  is ranked higher than  $B$ . Accordingly, the most significant outcome would be  $I_{LI}(A, B) = 1$  and  $I_{LI}(B, A) = 0$ . However, this method strongly depends on the choice of the lines, and certain parts of the attainment surface are not sampled. Therefore, in the above case either  $A$  is better than  $B$  or both approximation are incomparable to each other. As a consequence, the comparison method  $C_{I_{LI}, E}$  with  $E := (I_{LI}(A, B) = 1 \wedge I_{LI}(B, A) = 0)$  is in the general case not  $\triangleright$ -compatible; however, it is  $\not\triangleright$ -compatible and  $\succ\succ$ -complete.

Finally, we have shown already in Section 2.3 that a  $\triangleright$ -compatible and  $\triangleright$ -complete comparison method exists for the  $\epsilon$ -indicator. The case  $I_\epsilon(A, B) \leq 1$  is equivalent to  $A \succeq B$  and the same statements as for the coverage and the binary hypervolume indicators hold. Furthermore, the comparison method  $C_{I_\epsilon, E}$  with  $E := (I_\epsilon(A, B) < 1)$  is  $\succ\succ$ -compatible and  $\succ\succ$ -complete.

Table 4 summarizes the results of this section. Note that it only contains information about comparison methods that are both compatible and complete with respect to the different dominance relations.

## 5 Discussion

### 5.1 Summary of Results

We have proposed a mathematical framework to study quality assessment methods for multiobjective optimizers. Starting with the assumption that the outcome of a multiobjective EA is a set of incomparable solutions, a so-called approximation set, we have introduced several dominance relations on approximation sets. These relations represent a formal description of what we intuitively understand by one approximation set being better than another. The term quality indicator has been used to capture the notion of a quality measure, and a comparison method has been defined as a combination of quality indicators and a pseudo-Boolean function that evaluates the indicator values. Furthermore, we have discussed two properties of comparison methods, namely compatibility and completeness, which characterize

ind.	name / reference	relation					
		$\succ\succ$	$\succ$	$\triangleright$	$\succeq$	$=$	$\parallel$
$I_\epsilon$	epsilon indicator / Section 2.2	$I_\epsilon(A, B) < 1$	-	$I_\epsilon(A, B) \leq 1$ $I_\epsilon(B, A) > 1$	$I_\epsilon(A, B) \leq 1$	$I_\epsilon(A, B) = 1$ $I_\epsilon(B, A) = 1$	$I_\epsilon(A, B) > 1$ $I_\epsilon(B, A) > 1$
$I_{\epsilon+}$	additive epsilon indicator / Section 2.2	$I_{\epsilon+}(A, B) < 0$	-	$I_{\epsilon+}(A, B) \leq 0$ $I_{\epsilon+}(B, A) > 0$	$I_{\epsilon+}(A, B) \leq 0$	$I_{\epsilon+}(A, B) = 0$ $I_{\epsilon+}(B, A) = 0$	$I_{\epsilon+}(A, B) > 0$ $I_{\epsilon+}(B, A) > 0$
$I_C$	coverage / [26]	-	$I_C(A, B) = 1$ $I_C(B, A) = 0$	$I_C(A, B) = 1$ $I_C(B, A) < 1$	$I_C(A, B) = 1$	$I_C(A, B) = 1$ $I_C(B, A) = 1$	$0 < I_C(A, B) < 1$ $0 < I_C(B, A) < 1$
$I_{H2}$	binary hypervolume indicator / [24]	-	-	$I_{H2}(A, B) > 0$ $I_{H2}(B, A) = 0$	$I_{H2}(A, B) \geq 0$ $I_{H2}(B, A) = 0$	$I_{H2}(A, B) = 0$ $I_{H2}(B, A) = 0$	$I_{H2}(A, B) > 0$ $I_{H2}(B, A) > 0$
$I_{R1}$	utility function indicator $R1$ / [9]	-	-	-	-	-	-
$I_{R2}$	utility function indicator $R2$ / [9]	-	-	-	-	-	-
$I_{R3}$	utility function indicator $R3$ / [9]	-	-	-	-	-	-
$I_{LI}$	lines of intersection / [11]	-	-	-	-	-	-

Table 4: Overview of binary indicators. A minus means that in general there is no comparison method  $C_{I,E}$  based on the indicator  $I$  in the corresponding row that is compatible and complete regarding the relation in the corresponding column. Otherwise, an expression is given that describes an appropriate Boolean function  $E$ .

the relationship between comparison methods and dominance relations. On the basis of this framework, existing comparison methods have been analyzed and discussed. The key results are:

- Unary quality indicators, i.e., quality measures that summarize an approximation set in terms of a real number, are *in general* not capable of indicating *whether* an approximation set is better than another—also if several of them are used. This even holds if we consider approximation sets containing a single objective vector only.
- Existing unary indicators at best allow to infer *that* an approximation set is not worse than another, e.g., the distance indicator by Czyzak and Jaszkiwicz [1], the hypervolume indicator by Zitzler and Thiele [26], or the unary  $\epsilon$ -indicator presented in this paper. However, with many unary indicators and also combinations of unary indicators no statement about the relation between the corresponding approximation sets can be made. That is, although an approximation set  $A$  may be evaluated better than an approxi-

mation set  $B$  with respect to all of the indicators,  $B$  can actually be superior to  $A$  with respect to the dominance relations. This holds especially for the various diversity measures and also for some of the distance indicators proposed in the literature.

- We have given two examples demonstrating that comparison methods based on unary indicators can be constructed such that  $A$  can be recognized as being better than  $B$  for some approximation sets  $A, B$ . It has also been shown that the practical use of this type of indicator is naturally restricted.
- Binary indicators, which assign real numbers to ordered pairs of approximation sets, in principle do not possess the theoretical limitations of unary indicators. The binary  $\epsilon$ -indicator proposed in this paper, e.g., is capable of detecting *whether* an approximation set is better than another. However, not all existing binary indicators have this property. Furthermore, it has to be mentioned that the greater inferential power

comes along with additional complexity: in contrast to unary indicators, the number of indicator values to be considered is not linear but quadratic in the number of approximation sets.

## 5.2 Conclusions

This study has shown that in general the quality of an approximation set cannot be completely described by a (finite) set of distinct criteria such as diversity and distance. Hence, binary quality indicators represent the lowest level of representation on which it is still possible to detect whether an algorithm performs better than another in terms of the quality of the outcomes. On the other hand, this does not mean that unary quality indicators are generally useless. In conjunction with a  $\triangleright$ -compatible and  $\triangleright$ -complete comparison method, they can be used to further differentiate between incomparable approximation sets and to focus on specific, usually problem-dependent aspects. However, we have to be aware that they often represent preference information and therefore for each problem the assumptions and knowledge exploited should be clearly specified. A more detailed discussion of this issue can be found in [9].

Moreover, we have studied quality indicators only for one, but essential criterion: the inferential power. Certainly, there are many other aspects according to which comparison methods can be investigated, e.g., the computational effort, the sensitivity to scaling, the requirement to have knowledge about the Pareto-optimal front, etc. Several such aspects are studied in [14] and [13]. The coverage indicator [25] represents an example where these additional considerations come into play. Although being capable of detecting dominance between approximation sets, it does not provide additional information if, e.g.,  $A$  dominates  $B$  and  $B$  dominates  $C$  (“how much better is  $A$  than  $B$  with respect to  $C$ ?”); furthermore, the indicator values are often difficult to interpret if the two approximation sets under consideration are incomparable. In the light of this discussion, the binary  $\epsilon$ -indicator defined in Section 2.2 possesses several desirable features. It represents a natural extension to the evaluation of approximation schemes

in theoretical computer science [5] and gives the factor by which an outcome is worse than another. In addition to that, it is cheap to compute.

Finally, the stochasticity of multiobjective EAs is another issue that has to be addressed. Multiple optimization runs require the application of statistical tests, and in principle there are two ways to incorporate these tests in a comparison method: the statistical testing procedure can be included in the indicator functions or in the Boolean function. Knowles and Corne’s approach [11] belongs to the first category, while Van Veldhuizen and Lamont’s study [22] is an example for the second category. The attainment function method proposed by Grunert da Fonseca, Fonseca, and Hall [8] can be expressed in terms of an infinite number of indicators and therefore falls in the second category. However, in contrast to [22] and [11] this method is able to detect *whether* an approximation set is better than another. To investigate in more depth how all these approaches are related to each other is the subject of ongoing research.

## Appendix

*Proof of Theorem 1.* Let us suppose that such a comparison method  $C_{I,E}$  exists where  $I = (I_1, I_2, \dots, I_k)$  is a combination of  $k$  unary quality indicators and  $E$  a corresponding Boolean function  $\mathbb{R}^{2k} \rightarrow \{\text{false}, \text{true}\}$ . Furthermore, assume, without loss of generality, that the first two objectives are to be minimized (otherwise the definition of the following set  $S$  has to be modified accordingly).

Choose  $a, b \in \mathbb{R}$  with  $a < b$ , and consider  $S = \{(z_1, z_2, \dots, z_n) \in Z ; a < z_i < b, 1 \leq i \leq n \wedge z_2 = b + a - z_1\}$ ; obviously, for any  $\mathbf{z}^1, \mathbf{z}^2 \in Z$  either  $\mathbf{z}^1 = \mathbf{z}^2$  or  $\mathbf{z}^1 \parallel \mathbf{z}^2$ , because  $z_1^1 > z_1^2$  implies  $z_2^1 < z_2^2$ . Furthermore, let  $\Omega_S \subseteq \Omega$  denote the set of approximations sets  $A \in \Omega$  with  $A \subseteq S$ .

As  $S \in \Omega$  and any subset of an approximation set is again an approximation set,  $\Omega_S$  is identical to the power set  $\mathcal{P}(S)$  of  $S$ . In addition, there is an injection  $f$  from the open interval  $(a, b)$  to  $S$  with  $f(r) = (r, b + a - r, (b + a)/2, (b + a)/2, \dots, (b + a)/2)$ , it follows that the cardinality of  $S$  is at least  $2^{\aleph_0}$ . As a consequence,

the cardinality of  $\Omega_S$  is at least  $2^{2^{n_0}}$ .

Now, we will use Lemma 1 (see below): it shows that for any  $A, B \in \Omega_S$  with  $A \neq B$  the quality indicator values differ, i.e.,  $I_i(A) \neq I_i(B)$  for at least one indicator  $I_i$ ,  $1 \leq i \leq k$ . Therefore, there must be an injection from  $\Omega_S$  to  $\mathbb{R}^k$ , the codomain of  $\mathbf{I}$ . This means there is an injection from a set of cardinality  $2^{2^{n_0}}$  (or greater) to a set of cardinality  $2^{n_0}$ . From this absurdity, it follows that such a comparison method  $C_{\mathbf{I}, E}$  cannot exist.  $\square$

**Lemma 1** *Let  $Z = \{(z_1, z_2, \dots, z_n) \in \mathbb{R}^n; a < z_i < b, 1 \leq i \leq n\}$  be an open hypercube in  $\mathbb{R}^n$  with  $n \geq 2$ ,  $a, b \in \mathbb{R}$ , and  $a < b$ . Furthermore, assume there exists a combination of unary quality indicators  $\mathbf{I} = (I_1, I_2, \dots, I_k)$  and a Boolean function  $E$  such that for any approximation sets  $A, B \in \Omega$ :*

$$C_{\mathbf{I}, E}(A, B) \Leftrightarrow A \triangleright B$$

*Then, for all  $A, B \in \Omega$  with  $A \neq B$  there is at least one quality indicator  $I_i$  with  $1 \leq i \leq k$  such that  $I_i(A) \neq I_i(B)$ .*

*Proof.* Let  $A, B \in \Omega$  be two arbitrary approximation sets with  $A \neq B$ . First note that  $C_{\mathbf{I}, E}(A, B)$  implies  $C_{\mathbf{I}, E}(B, A)$  is false (and vice versa) as  $A \triangleright B$  implies  $B \not\triangleright A$ . If  $A \triangleright B$  or  $B \triangleright A$ , then  $I_i(A) \neq I_i(B)$  for at least one  $1 \leq i \leq k$  because otherwise  $C_{\mathbf{I}, E}(A, B) = C_{\mathbf{I}, E}(B, A) = C_{\mathbf{I}, E}(A, A)$  would be false. If  $A \parallel B$ , there are two cases: (1) both  $A$  and  $B$  contain only a single objective vector, or (2) either set consists of more than one element.

**Case 1:** Choose  $\mathbf{z} \in Z$  with  $A \parallel \{\mathbf{z}\}$  and  $B \parallel \{\mathbf{z}\}$  (such an objective vector exists as  $Z$  is an open hypercube in  $\mathbb{R}^n$ ). Then  $A \cup \{\mathbf{z}\} \triangleright A$  and  $A \cup \{\mathbf{z}\} \parallel B$ , and from the former follows  $C_{\mathbf{I}, E}(A \cup \{\mathbf{z}\}, A)$  is true. Accordingly,  $I_i(A) \neq I_i(B)$  for at least one  $1 \leq i \leq k$  because otherwise  $C_{\mathbf{I}, E}(A \cup \{\mathbf{z}\}, B) = C_{\mathbf{I}, E}(A \cup \{\mathbf{z}\}, A)$  would be true which contradicts  $A \cup \{\mathbf{z}\} \parallel B$ .

**Case 2:** Assume, without loss of generality, that  $A$  contains more than one objective vector, and

choose  $\mathbf{z} \in A$  with  $\{\mathbf{z}\} \parallel B$  (such an element must exist as  $A \parallel B$ ). Then,  $A \triangleright \{\mathbf{z}\}$ , which implies that  $C_{\mathbf{I}, E}(A, \{\mathbf{z}\})$ . Now suppose  $I_i(A) = I_i(B)$  for all  $1 \leq i \leq k$ ; it follows that  $C_{\mathbf{I}, E}(B, \{\mathbf{z}\}) = C_{\mathbf{I}, E}(A, \{\mathbf{z}\})$  is true which is a contradiction to  $B \parallel \{\mathbf{z}\}$ .

In summary, all cases ( $A \triangleright B$ ,  $B \triangleright A$ , and  $A \parallel B$ ) imply that  $I_i(A) \neq I_i(B)$  for at least one  $1 \leq i \leq k$ .  $\square$

*Proof of Theorem 2.* We will exploit the fact that in  $\mathbb{R}$  the number of disjoint open intervals  $(a, b) = \{z \in \mathbb{R} ; a < z < b\}$  with  $a < b$  is countable [10]; in general, this means that  $\mathbb{R}^k$  contains only countably many disjoint open hyperrectangles  $(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_k, b_k) = \{(z_1, z_2, \dots, z_k) \in \mathbb{R}^k ; a_i < z_i < b_i, 1 \leq i \leq k\}$  with  $a_i < b_i$ . The basic idea is that whenever fewer indicators than objectives are available, uncountably many disjoint open hyperrectangles arise—a contradiction. Furthermore, we will show a slightly modified statement, which is more general: if  $Z$  contains an open hypercube  $(u, v)^n$  with  $u < v$  such that for any  $\mathbf{z}^1, \mathbf{z}^2 \in (u, v)^n$ :

$$(\forall 1 \leq i \leq k : I_i(\{\mathbf{z}^1\}) \geq I_i(\{\mathbf{z}^2\})) \Leftrightarrow \mathbf{z}^1 \succeq \mathbf{z}^2$$

then  $k \geq n$ .

Without loss of generality assume a minimization problem in the following. We will argue by induction.

$n = 2$ : Let  $a, b \in (u, v)$  with  $a < b$  and consider the incomparable objective vectors  $(a, b)$  and  $(b, a)$ . If  $k = 1$ , then either  $I_1(\{(a, b)\}) \geq I_1(\{(b, a)\})$  or vice versa; this leads to a contradiction to  $(a, b) \not\preceq (b, a)$  and  $(b, a) \not\preceq (a, b)$ .

$n - 1 \rightarrow n$ : Suppose  $n > 2$ ,  $k < n$  and that the statement holds for  $n - 1$ . Choose  $a, b \in (u, v)$  with  $a < b$ , and consider the  $n - 1$  dimensional open hypercube  $S_c = \{(z_1, z_2, \dots, z_{n-1}, c) \in (u, v)^n ; a < z_i < b, 1 \leq i \leq n - 1\}$  for an arbitrary  $c \in (u, v)$ .

First, we will show that  $I_i(\{(b, \dots, b, c)\}) < I_i(\{(a, \dots, a, c)\})$  for all  $1 \leq i \leq k$ . Assume  $I_i(\{(b, \dots, b, c)\}) \geq I_i(\{(a, \dots, a, c)\})$

for any  $i$ . If  $I_i(\{(b, \dots, b, c)\}) > I_i(\{(a, \dots, a, c)\})$ , then  $(a, \dots, a, c) \not\preceq (b, \dots, b, c)$ , which yields a contradiction. If  $I_i(\{(b, \dots, b, c)\}) = I_i(\{(a, \dots, a, c)\})$ , then  $I_i(\{z\}) = I_i(\{(b, \dots, b, c)\})$  for all  $z \in S_c$ , because  $(a, \dots, a, c) \succeq z$  if  $z \in S_c$ . Then for any  $z^1, z^2 \in S_c$  it holds

$$\forall 1 \leq j \leq k, j \neq i : \\ I_j(\{z^1\}) \geq I_j(\{z^2\}) \Leftrightarrow z^1 \succeq z^2$$

which contradicts the assumption that for any  $n - 1$  dimensional open hypercube in  $\mathbb{R}^{n-1}$  at least  $n - 1$  indicators are necessary. Therefore,  $I_i(\{(b, \dots, b, c)\}) < I_i(\{(a, \dots, a, c)\})$ .

Now, we consider the image of  $S_c$  in indicator space. The vectors  $\mathbf{I}(\{(b, \dots, b, c)\})$  and  $\mathbf{I}(\{(a, \dots, a, c)\})$  determine an open hyperrectangle  $H_c = \{(y_1, y_2, \dots, y_k) \in \mathbb{R}^k ; I_i(\{(b, \dots, b, c)\}) < y_i < I_i(\{(a, \dots, a, c)\}), 1 \leq i \leq k\}$  where  $\mathbf{I}(z) = (I_1(z), I_2(z), \dots, I_k(z))$ .  $H_c$  has the following properties:

1.  $H_c$  is open in all  $k$  dimensions as for all  $1 \leq i \leq k$ :  $\inf\{y_i ; (y_1, y_2, \dots, y_k) \in H_c\} = I_i(\{(b, \dots, b, c)\}) < I_i(\{(a, \dots, a, c)\}) = \sup\{y_i ; (y_1, y_2, \dots, y_k) \in H_c\}$ .
2.  $H_c$  contains an infinite number of elements.
3.  $H_c \cap H_d = \emptyset$  for any  $d \in (u, v), d > c$ : assume  $\mathbf{y} \in H_c \cap H_d$ ; then  $\mathbf{I}(\{(a, \dots, a, c)\}) \geq \mathbf{y} \geq \mathbf{I}(\{(b, \dots, b, d)\})$ , which yields a contradiction as  $(a, \dots, a, c) \not\preceq (b, \dots, b, d)$ .

Since  $c$  was arbitrarily chosen within  $(u, v)$ , there are uncountably many disjoint open hyperrectangles of dimensionality  $k$  in the  $k$  dimensional indicator space. This contradiction implies that  $k \geq n$ .  $\square$

## Acknowledgments

The authors would like to thank Wade Ramey for the tip about the proof of Theorem 2. The re-

search has been supported by the Swiss National Science Foundation (SNF) under the ArOMA project 2100-057156.99/1 and by the Portuguese Foundation for Science and Technology under the POCTI programme (Project POCTI/MAT/10135/98), co-financed by the European Regional Development Fund.

## References

- [1] P. Czyzak and A. Jaszkievicz. Pareto simulated annealing—a metaheuristic for multiobjective combinatorial optimization. *Multi-Criteria Decision Analysis*, 7:34–47, 1998.
- [2] Prabuddha De, Jay B. Ghosh, and Charles E. Wells. Heuristic estimation of the efficient frontier for a bi-criteria scheduling problem. *Decision Sciences*, 23:596–609, 1992.
- [3] Kalyanmoy Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, Chichester, UK, 2001.
- [4] Kalyanmoy Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In M. Schoenauer et al., editors, *Parallel Problem Solving from Nature – PPSN VI*, pages 849–858, Berlin, 2000. Springer.
- [5] Thomas Erlebach, Hans Kellerer, and Ulrich Pferschy. Approximating multi-objective knapsack problems. In Frank K. H. A. Dehne, Jörg-Rüdiger Sack, and Roberto Tamassia, editors, *Proceedings of the Seventh International Workshop on Algorithms and Data Structures (WADS 2001)*, pages 210–221, Berlin, Germany, 2001. Springer.
- [6] Henrik Esbensen and Ernest S. Kuh. Design space exploration using the genetic algorithm. In *IEEE International Symposium on Circuits and Systems (ISCAS'96)*, volume 4, pages 500–503, Piscataway, NJ, 1996. IEEE Press.

- [7] Carlos M. Fonseca and Peter J. Fleming. On the performance assessment and comparison of stochastic multiobjective optimizers. In Hans-Michael Voigt, Werner Ebeling, Ingo Rechenberg, and Hans-Paul Schwefel, editors, *Fourth International Conference on Parallel Problem Solving from Nature (PPSN-IV)*, pages 584–593, Berlin, Germany, 1996. Springer.
- [8] Viviane Grunert da Fonseca, Carlos M. Fonseca, and Andreia O. Hall. Inferential performance assessment of stochastic optimisers and the attainment function. In E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, editors, *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, volume 1993 of *Lecture Notes in Computer Science*, pages 213–225, Berlin, 2001. Springer-Verlag.
- [9] Michael P. Hansen and Andrzej Jaskiewicz. 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.
- [10] Karel Hrbacek and Thomas Jech. *Introduction to Set Theory*. Marcel Dekker, Inc., New York, 1999.
- [11] J. D. Knowles and D. W. Corne. Approximating the nondominated front using the pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172, 2000.
- [12] J. D. Knowles, D. W. Corne, and M. J. Oates. On the assessment of multiobjective approaches to the adaptive distributed database management problem. In M. Schoenauer et al., editors, *Parallel Problem Solving from Nature – PPSN VI*, pages 869–878, Berlin, 2000. Springer.
- [13] Joshua Knowles and David Corne. On metrics for comparing non-dominated sets. In *Congress on Evolutionary Computation (CEC 2002)*, pages 711–716, Piscataway, NJ, 2002. IEEE Press.
- [14] Joshua D. Knowles. *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. PhD thesis, Department of Computer Science, University of Reading, UK, 2002.
- [15] Marco Laumanns, Lothar Thiele, Kalyanmoy Deb, and Eckart Zitzler. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary Computation*, 10(3), 2002.
- [16] Serpil Sayin. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Math. Program., Ser. A* 87, pages 543–560, 2000.
- [17] J. Schott. Fault tolerant design using single and multicriteria genetic algorithm optimization. Master’s thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 1995.
- [18] N. Srinivas and Kalyanmoy Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221–248, 1994.
- [19] K. C. Tan, T. H. Lee, and E. F. Khor. Evolutionary algorithms for multi-objective optimization: Performance assessments and comparisons. In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, pages 979–986, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea, 27-30 May 2001. IEEE Press.
- [20] E. L. Ulungu, J. Teghem, Ph. Fortemps, and D. Tuyttens. Mosa method: a tool for solving multiobjective combinatorial optimization problems. *Journal of Multi-Criteria Decision Analysis*, 8:221–236, 1999.
- [21] David A. Van Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Graduate School of Engineering of the Air Force Institute of Technology, Air University, June 1999.

- [22] David A. Van Veldhuizen and Gary B. Lamont. On measuring multiobjective evolutionary algorithm performance. In A. Zalzal and R. Eberhart, editors, *Congress on Evolutionary Computation (CEC 2000)*, volume 1, pages 204–211, Piscataway, NJ, 2000. IEEE Press.
- [23] Jin Wu and Shapour Azarm. Metrics for quality assessment of a multiobjective design optimization solution set. *Transactions of the ASME, Journal of Mechanical Design*, 123:18–25, March 2001.
- [24] Eckart Zitzler. *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, 1999. TIK-Schriftenreihe Nr. 30, Diss ETH No. 13398, Shaker Verlag, Aachen, Germany.
- [25] Eckart Zitzler and Lothar Thiele. An evolutionary algorithm for multiobjective optimization: The strength pareto approach. Technical Report 43, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich, Switzerland, May 1998.
- [26] Eckart Zitzler and Lothar Thiele. Multiobjective optimization using evolutionary algorithms — a comparative case study. In Agoston E. Eiben, Thomas Bäck, Marc Schoenauer, and Hans-Paul Schwefel, editors, *Fifth International Conference on Parallel Problem Solving from Nature (PPSN-V)*, pages 292–301, Berlin, Germany, 1998. Springer.

## Addendum (July 3, 2002)

**Section 3.2.2, 1st paragraph:** The statement

$$I_{\epsilon_1}(A) < I_{\epsilon_1}(B) \Rightarrow A \not\prec\prec B \Rightarrow A \not\prec B$$

is wrong. The  $\not\prec$ -compatibility follows from

$$A \triangleright B \Rightarrow I_{\epsilon_1}(A) \leq I_{\epsilon_1}(B)$$

which implies

$$I_{\epsilon_1}(A) < I_{\epsilon_1}(B) \Rightarrow A \not\prec B$$

**Section 3.2.3** Independently of this study, Knowles and Corne [14][13] have also shown the incompatibility of the following indicators:  $I_S$ ,  $I_{ME}$ ,  $I_{DU}$ , and  $I_{ONVG}$ .