

A Study on Performance Evaluation Ability of a Modified Inverted Generational Distance Indicator

Hisao Ishibuchi

Osaka Prefecture University
Sakai, Osaka, Japan 599-8531
+81-72-254-9350
hisaoi@cs.osakafu-u.ac.jp

Hiroyuki Masuda

Osaka Prefecture University
Sakai, Osaka, Japan 599-8531
+81-72-254-9198
hiroyuki.masuda@ci.cs.osakafu-u.ac.jp

Yusuke Nojima

Osaka Prefecture University
Sakai, Osaka, Japan 599-8531
+81-72-254-9198
nojima@cs.osakafu-u.ac.jp

ABSTRACT

The inverted generational distance (IGD) has been frequently used as a performance indicator for many-objective problems where the use of the hypervolume is difficult. However, since IGD is not Pareto compliant, it is possible that misleading Pareto incompliant results are obtained. Recently, a simple modification of IGD was proposed by taking into account the Pareto dominance relation between a solution and a reference point when their distance is calculated. It was also shown that the modified indicator called IGD^+ is weakly Pareto compliant. However, actual effects of the modification on performance comparison have not been examined. Moreover, IGD^+ has not been compared with other distance-based weakly Pareto compliant indicators such as the additive epsilon indicator and the D1 indicator (i.e., IGD with the weighted achievement scalarizing function). In this paper, we examine the effect of the modification by comparing IGD^+ with IGD for multiobjective and many-objective problems. In computational experiments, we generate a large number of ordered pairs of non-dominated solution sets where one is better than the other. Two solution sets in each pair are compared by the above-mentioned performance indicators. We examine whether each indicator can correctly say which solution set is better between them.

Categories and Subject Descriptors

Mathematics of computing ~ Evolutionary algorithms

Keywords

Evolutionary multiobjective optimization, solution set comparison, many-objective optimization, inverted generational distance.

1. INTRODUCTION

Many-objective optimization has been a hot topic in the field of evolutionary multiobjective optimization (EMO). Since the well-known Pareto dominance-based EMO algorithms such as SPEA [30], SPEA2 [28] and NSGA-II [7] do not always work well on many-objective optimization problems, a number of new EMO algorithms and modifications of existing ones have been proposed in the literature [13], [22]. Difficulties in evolutionary many-

objective optimization are related to various aspects such as the search for Pareto optimal solutions, the approximation of the Pareto front by a limited number of solutions, the presentation of obtained solutions to the decision maker, the choice of a single final solution, and the monitoring of the search behavior of EMO algorithms. It is also difficult to evaluate an obtained set of non-dominated solutions of a many-objective optimization problem.

In the EMO community, the hypervolume [29] has been used to evaluate a set of obtained non-dominated solutions. This is mainly because no other Pareto compliant unary indicator is known [27], [31]. The hypervolume has also been used to evaluate the fitness of each solution in indicator-based EMO algorithms such as SMS-EMOA [2] and HypE [1]. High performance of those algorithms has been reported for many-objective problems [1], [23]. However, it is not easy to use the hypervolume when the number of objectives is large (e.g., more than 10) and the number of non-dominated solutions is large (e.g., more than 1000). This is because a long computation time is still needed to calculate the hypervolume of a large number of non-dominated solutions in a high-dimensional objective space whereas some fast algorithms have been proposed for efficient calculation [4], [24], [18].

Recently the inverted generational distance (IGD) has often been used to evaluate solution sets of many-objective problems [17], [25]. To the best of our knowledge, the term "inverted generational distance (IGD)" was first used in 2004 [5], [20]. Similar ideas have been used in the literature since 1998 [6], [9] where the weighted achievement scalarizing function was used for distance calculation from each reference point to the nearest solution. This type of IGD was referred to as D1 in [9] and denoted as $D1_R$ in [9], [16]. However, the Euclidian distance has been usually used in IGD for performance evaluations in the literature since 2003 [3], [14]. One advantage of IGD as a performance indicator over the hypervolume is its computational efficiency. Since IGD is based on the distance calculation between two points in the objective space, its calculation is usually very efficient even in the case of many-objective optimization.

One well-known flaw of the Euclidian distance-based IGD in comparison with the hypervolume is the lack of the Pareto compliant property [27], [31]. This flaw is illustrated in Fig. 1 where two solution sets $A = \{a_1, a_2, a_3\}$, $B = \{b_1, b_2, b_3\}$ and a reference point set $Z = \{z_1, z_2, z_3\}$ are given in a two-dimensional objective space. Since every solution in B is dominated by at least one solution in A , we can say that A is a better solution set than B from the viewpoint of Pareto dominance. However, the dominated solution set B is evaluated as being better than A if we use the Euclidian distance-based IGD in Fig. 1 for comparing A and B .

Another difficulty of IGD is its dependency on the specification of a reference point set. In Fig. 2, we show another example with

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

GECCO '15, July 11 - 15, 2015, Madrid, Spain

© 2015 ACM. ISBN 978-1-4503-3472-3/15/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2739480.2754792>

two solution sets A, B and a reference point set Z . The solution set A is evaluated as being better in Fig. 2 by the Euclidian distance-based IGD. In Fig. 3, four reference points z_4, z_5, z_6 and z_7 are added to Z in the example in Fig. 2. After this modification of Z , the solution set B is evaluated as being better by IGD in Fig. 3 while A is evaluated as being better in Fig. 2.

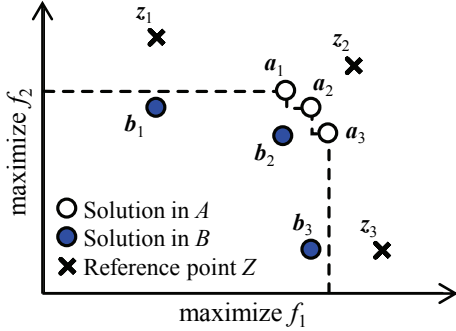


Figure 1. Pareto incompliant example where two solution sets $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3\}$ are evaluated by IGD using a reference point set $Z = \{z_1, z_2, z_3\}$.

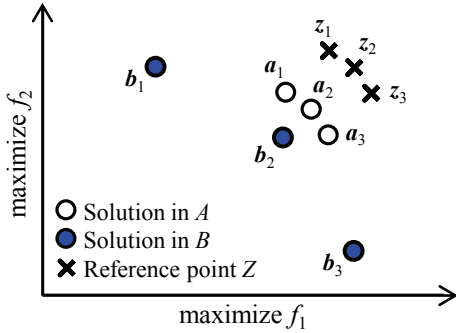


Figure 2. Another example with two solution sets $A = \{a_1, a_2, a_3\}$, $B = \{b_1, b_2, b_3\}$, and a reference point set $Z = \{z_1, z_2, z_3\}$.

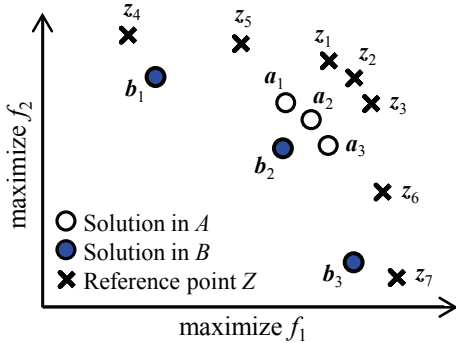


Figure 3. Addition of four reference points to Z in Fig. 2.

To remedy these difficulties of IGD, a simple modification of the distance calculation has been proposed in [11]. The proposed idea, which is explained in Fig. 4 for the solution set $A = \{a_1, a_2, a_3\}$, is to calculate the distance from each reference point z_i to the dominated region by the solution set instead of the distance to the nearest solution. The modified indicator, which is called IGD^+ , is the average length of the three dotted lines in Fig. 4. It should be noted that the two solution sets A, B and the reference point set Z in Fig. 4 are the same as those in Fig. 1. As we explained in Fig. 1, the dominated solution set B was evaluated as being better than A by IGD. Such a Pareto incompliant result is not obtained by IGD^+ in Fig. 4 where the solution set A is evaluated as being better than

B . If we use IGD^+ in Fig. 2 and Fig. 3, the solution set A is evaluated as being better than B . It should be noted that B is evaluated as being better than A in Fig. 3 by IGD. It was also shown that IGD^+ is weakly Pareto compliant in [12].

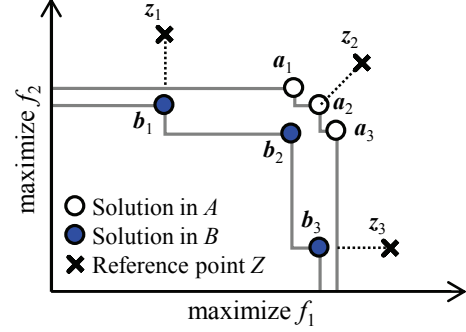


Figure 4. Illustration of the basic idea of IGD^+ .

Whereas the effect of using IGD^+ instead of IGD is clear in Figs. 1-4, its practical usefulness is still unclear. This is because these examples were intentionally created for illustration purposes. For example, the diversities of the two solution sets A and B are unusually different in each figure. The solution set A has a similar diversity to the reference point set Z in Fig. 2 whereas B is similar to Z in the other figures. All examples in [11], [12] are somewhat similar to these examples in Figs. 1-4 intentionally generated for illustrating the flaws of IGD. Similar situations to Figs. 1-4 are not likely to happen frequently in performance comparison of non-dominated solution sets obtained by EMO algorithms.

The aim of this paper is to examine the effect of using IGD^+ instead of IGD for performance comparison of EMO algorithms on many-objective test problems. We also compare IGD^+ with other weakly Pareto compliant unary indicators such as the additive epsilon [31] and D1 [9].

This paper is organized as follows. In Section 2, we explain IGD, IGD^+ , the additive epsilon and D1. The concept of the Pareto compliance [27] is also explained in Section 2. In Section 3, we report experimental results where pairs of non-dominated solution sets are searched from multiobjective knapsack problems and DTLZ2 with 2-10 objectives. In each pair of solution sets, one set is better than the other in the sense of Pareto dominance (more specifically, with respect to the relation “better” in [31]). Using those pairs of solution sets, the discrimination ability of each indicator is examined. Experimental results demonstrate the effect of using IGD^+ instead of IGD. High discrimination ability of D1 is also demonstrated. Finally this paper is concluded in Section 4.

2. MULTI-OBJECTIVE OPTIMIZATION

In this section, we briefly explain a multiobjective optimization problem, Pareto dominance relation between solutions, Pareto dominance relation between solution sets, performance indicators, and the concept of Pareto compliance.

2.1 Multiobjective Optimization Problems

Let us consider the following m -objective maximization problem:

$$\text{Maximize } \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})), \quad (1)$$

$$\text{subject to } \mathbf{x} \in X, \quad (2)$$

where $\mathbf{f}(\mathbf{x})$ is an m -dimensional objective vector, $f_i(\mathbf{x})$ is its i -th objective to be maximized ($i = 1, 2, \dots, m$), \mathbf{x} is the decision vector, and X is the feasible region in the decision space.

2.2 Pareto Dominance between Solutions

Let \mathbf{x} and \mathbf{y} be two feasible solutions of the maximization problem in (1)-(2). When the following relations hold, we say that \mathbf{x} is dominated by \mathbf{y} :

$$\forall i, f_i(\mathbf{x}) \leq f_i(\mathbf{y}) \text{ and } \exists j, f_j(\mathbf{x}) < f_j(\mathbf{y}). \quad (3)$$

This Pareto dominance relation is denoted as $\mathbf{x} < \mathbf{y}$:

$$\forall i, f_i(\mathbf{x}) \leq f_i(\mathbf{y}) \text{ and } \exists j, f_j(\mathbf{x}) < f_j(\mathbf{y}) \Leftrightarrow \mathbf{x} < \mathbf{y}. \quad (4)$$

When the following relation holds, we say that \mathbf{x} is weakly dominated by \mathbf{y} :

$$\forall i, f_i(\mathbf{x}) \leq f_i(\mathbf{y}). \quad (5)$$

This weak Pareto dominance relation is denoted as $\mathbf{x} \preceq \mathbf{y}$:

$$\forall i, f_i(\mathbf{x}) \leq f_i(\mathbf{y}) \Leftrightarrow \mathbf{x} \preceq \mathbf{y}. \quad (6)$$

If a feasible solution \mathbf{x} is not dominated by any other feasible solutions of the multiobjective problem in (1)-(2), \mathbf{x} is called a Pareto optimal solution. Its projection $\mathbf{f}(\mathbf{x})$ in the objective space is also often referred to as a Pareto optimal solution. The set of all Pareto optimal solutions is the Pareto optimal solution set. The projection of all Pareto optimal solutions (i.e., the Pareto optimal solution set) to the objective space is called the Pareto front.

If a feasible solution \mathbf{x} is not dominated by any other feasible solutions in a solution set A , \mathbf{x} is called a non-dominated solution in the solution set A . If all solutions in A are non-dominated solutions in A , A is called a non-dominated solution set.

2.3 Pareto Dominance between Solution Sets

Let us denote two non-dominated solution sets by $A = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|A|}\}$ and $B = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{|B|}\}$ where $|A|$ and $|B|$ are the number of solutions in A and B , respectively. These two non-dominated solution sets are also denoted using the corresponding solutions in the objective space as $A = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{|A|}\}$ and $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{|B|}\}$ where $\mathbf{a}_i = \mathbf{f}(\mathbf{x}_i)$, $i = 1, 2, \dots, |A|$ and $\mathbf{b}_j = \mathbf{f}(\mathbf{y}_j)$, $j = 1, 2, \dots, |B|$.

In Hansen & Jaszkiewicz [9] and Zitzler et al. [31], relations between solution sets were discussed for solution set comparison. In this paper, we use the Pareto dominance relation, the weak Pareto dominance relation, and the relation ‘‘better’’ between two solution sets.

The Pareto dominance relation $A < B$, which means that the non-dominated solution set $A = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|A|}\}$ is dominated by $B = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{|B|}\}$, was defined as follows [31]:

$$\forall \mathbf{x}_i \in A, \exists \mathbf{y}_j \in B: \mathbf{x}_i < \mathbf{y}_j \Leftrightarrow A < B. \quad (7)$$

The weak Pareto dominance relation $A \preceq B$, which means that A is weakly dominated by B , was defined as follows [31]:

$$\forall \mathbf{x}_i \in A, \exists \mathbf{y}_j \in B: \mathbf{x}_i \preceq \mathbf{y}_j \Leftrightarrow A \preceq B. \quad (8)$$

This weak Pareto dominance relation holds even when the two solution sets are the same (i.e., $A = B$). In order to represent the case where the weak Pareto dominance relation holds for different non-dominated solution sets A and B (i.e., the case where B is better than A), the relation ‘‘better’’ was proposed as follows [31]:

$$A \preceq B \text{ and } A \neq B \Leftrightarrow A < B. \quad (9)$$

From these definitions, the following relation holds:

$$A < B \Rightarrow A < B \Rightarrow A \preceq B. \quad (10)$$

The left-hand side of (9) can be rewritten as

$$\forall \mathbf{x}_i \in A, \exists \mathbf{y}_j \in B: \mathbf{x}_i \preceq \mathbf{y}_j \text{ and } \exists \mathbf{x}_i \in A, \exists \mathbf{y}_j \in B: \mathbf{x}_i < \mathbf{y}_j. \quad (11)$$

That is, at least one solution \mathbf{y}_j in the solution set B is needed to be better than a solution \mathbf{x}_i in A in order that B is better than A .

2.4 Performance Indicators

Since the Euclidean distance in the objective space has been used in the generational distance (GD) and the inverted generational distance (IGD) in recent studies on evolutionary many-objective optimization [17], [25], we use the Euclidean distance whereas GD was originally proposed in a more general form [21]. For further discussions on GD and IGD, see Schuetz et al. [19].

Each of GD and IGD evaluates the distance between a solution set and the Pareto front in a different manner. We assume that a non-dominated reference point set $Z = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{|Z|}\}$ is given in the objective space as an approximation of the Pareto front. The GD of a non-dominated solution set $A = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{|A|}\}$ is the average distance from each solution to the nearest reference point:

$$GD(A) = \frac{1}{|A|} \sum_{i=1}^{|A|} \min_{z_j \in Z} d(\mathbf{a}_i, \mathbf{z}_j), \quad (12)$$

where $d(\mathbf{a}_i, \mathbf{z}_j)$ is the Euclidean distance between \mathbf{a}_i and \mathbf{z}_j .

The IGD of $A = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|A|}\}$ is the average distance from each reference point to the nearest solution:

$$IGD(A) = \frac{1}{|Z|} \sum_{j=1}^{|Z|} \min_{\mathbf{a}_i \in A} d(\mathbf{a}_i, \mathbf{z}_j). \quad (13)$$

In Fig. 5 (a), we show the contour lines of the Euclidean distance from the reference point \mathbf{z} in the objective space for the case of a two-objective maximization problem. In Fig. 5 (a), a solution \mathbf{a} is dominated by a solution \mathbf{b} . However, if the Euclidean distance is used, the solution \mathbf{a} has a shorter distance to the reference point \mathbf{z} than the solution \mathbf{b} . In this case, a singleton solution set $A = \{\mathbf{a}\}$ is evaluated as being better than another singleton solution set $B = \{\mathbf{b}\}$ by GD and IGD with a singleton reference point set $Z = \{\mathbf{z}\}$.

In IGD^+ , the distance between a reference point $\mathbf{z} = (z_1, z_2, \dots, z_m)$ and a solution $\mathbf{a} = (a_1, a_2, \dots, a_m)$ is calculated in the objective space for our m -objective maximization problem as follows:

$$d_{IGD^+}(\mathbf{a}, \mathbf{z}) = \sqrt{\sum_{k=1}^m (\max\{z_k - a_k, 0\})^2}. \quad (14)$$

If the solution \mathbf{a} is dominated by the reference point \mathbf{z} , $d_{IGD^+}(\mathbf{a}, \mathbf{z})$ is the same as the Euclidean distance $d(\mathbf{a}, \mathbf{z})$ since $a_k \leq z_k$ holds for all k 's in (14). When \mathbf{a} and \mathbf{z} are non-dominated with each other, only those objectives with $a_k \leq z_k$ are used in the distance calculation in (14). That is, if the solution \mathbf{a} is better than the reference point \mathbf{z} with respect to some objectives, those objectives are not used in the distance calculation in (14). The contour lines of the distance from \mathbf{z} measured by (14) are shown in Fig. 5 (b). In the dominated region by \mathbf{z} (i.e., in the lower-left region of \mathbf{z}), the contour lines of $d_{IGD^+}(\mathbf{a}, \mathbf{z})$ in Fig. 5 (b) are the same as those of the Euclidean distance in Fig. 5 (a). However, the contour lines of $d_{IGD^+}(\mathbf{a}, \mathbf{z})$ in Fig. 5 (b) are parallel to an axis of the objective space in the non-dominated region with respect to \mathbf{z} . As a result, the singleton solution set $B = \{\mathbf{b}\}$ is correctly evaluated as being better than $A = \{\mathbf{a}\}$ by IGD^+ with $Z = \{\mathbf{z}\}$ in Fig. 5 (b).

In the D1 indicator [9], the weighted achievement scalarizing function was used to measure the distance from the reference point $\mathbf{z} = (z_1, z_2, \dots, z_m)$ to the solution $\mathbf{a} = (a_1, a_2, \dots, a_m)$ as

$$d_{D1_R}(\mathbf{a}, \mathbf{z}) = \max_k \{\lambda_k (z_k - a_k)\}, \quad (15)$$

where λ_k is a weight for the k th objective defined as $\lambda_k = 1/R_k$ with R_k being a range of the k th objective in the reference point set Z [9]. The difference among IGD, IGD⁺ and D1 is only in their distance calculation. The D1 indicator can be written in the same manner as IGD and IGD⁺ as follows:

$$D1_R(A) = \frac{1}{|Z|} \sum_{j=1}^{|Z|} \min_{\mathbf{a}_i \in A} d_{D1_R}(\mathbf{a}_i, \mathbf{z}_j). \quad (16)$$

The contour lines of the distance in the D1 indicator are shown in Fig. 5 (c) for $(\lambda_1, \lambda_2) = (1, 1)$. In Fig. 5 (c), $B = \{\mathbf{b}\}$ is correctly evaluated as being better than $A = \{\mathbf{a}\}$ by D1 with $Z = \{\mathbf{z}\}$.

Whereas all IGD, IGD⁺ and D1 evaluate a solution set A using the average distance from each reference point to its nearest solution over all reference points, the additive epsilon indicator I_{ε^+} uses the maximum distance in the following manner [31]:

$$I_{\varepsilon^+}(A, Z) = \inf_{\varepsilon} \{\forall \mathbf{z}_j \in Z, \exists \mathbf{a}_i \in A: \mathbf{a}_i + \varepsilon \succeq \mathbf{z}_j\}, \quad (17)$$

where ε is an m -dimensional vector of $\varepsilon: \varepsilon = (\varepsilon, \varepsilon, \dots, \varepsilon)$.

For the case of $Z = \{\mathbf{z}_1, \mathbf{z}_2\}$, the contour lines of $I_{\varepsilon^+}(\{\mathbf{a}\}, \{\mathbf{z}_1, \mathbf{z}_2\})$ are written in Fig. 5 (d) as bold red lines. Thin lines and dotted lines can be viewed as distances from \mathbf{z}_1 and \mathbf{z}_2 measured by the weighted achievement scalarizing function with $(\lambda_1, \lambda_2) = (1, 1)$, respectively. The red bold lines in Fig. 5 (d) correspond to the larger distance from \mathbf{z}_1 and \mathbf{z}_2 . It should be noted that the additive epsilon was originally proposed as a binary performance indicator to compare two solution sets A and B as $I_{\varepsilon^+}(A, B)$ [31].

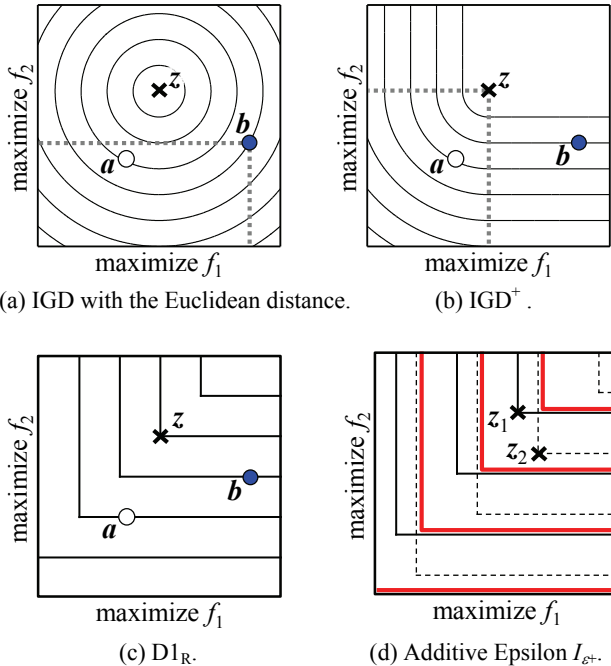


Figure 5. Contour lines of the distance from the reference point \mathbf{z} (\mathbf{z}_1 and \mathbf{z}_2 in (d)) in each indicator.

2.5 Pareto Compliance

In [9], [27], [31], comparison of solution sets has been discussed in detail. Let us assume that we have two non-dominated solution sets A and B . A unary performance indicator $I(\cdot)$ is called “Pareto compliant” when the following relation holds for arbitrary pairs of non-dominated solution sets A and B [27]:

$$\text{If } A \triangleleft B \text{ then } I(A) > I(B), \quad (18)$$

where we assume that smaller values of $I(\cdot)$ means better solution sets as in the case of all the performance indicators explained in the previous subsection. Whenever a non-dominated solution set B is better than another non-dominated solution set A in the sense of the better relation “ \triangleleft ” in Section 2, a Pareto compliant indicator always correctly says that B is better than A . For the best of our knowledge, only the hypervolume is known as a Pareto compliant unary indicator [27].

As shown in Fig. 5 (a), GD and IGD are not Pareto compliant. They cannot always correctly say which is better between an arbitrary pair of two non-dominated solution sets A and B with $A \triangleleft B$. For some pairs of A and B with $A \triangleleft B$, GD and IGD incorrectly say that A is better than B . Such a Pareto incompliant evaluation result is not obtained from IGD⁺, D1 and I_{ε^+} . That is, the following relation holds for arbitrary pairs of non-dominated solution sets A and B when IGD⁺, D1 and I_{ε^+} are used as $I(\cdot)$:

$$\text{If } A \triangleleft B \text{ then } I(A) \geq I(B). \quad (19)$$

For arbitrary pairs of non-dominated solution sets A and B with $A \triangleleft B$, IGD⁺, D1_R and I_{ε^+} never say that A is better than B . However, for some pairs of A and B with $A \triangleleft B$, these indicators say that the two solution sets A and B have the same performance: $I(A) = I(B)$ whereas $A \triangleleft B$. That is, the two solution sets are evaluated as being equivalent while $A \triangleleft B$ holds. This issue will be further discussed in the next section.

In this paper, we use the additive epsilon I_{ε^+} as a unary indicator in order to compare it with other unary indicators such as IGD⁺ and D1. However, as we have already mentioned, I_{ε^+} was originally proposed as a binary indicator to compare two non-dominated solution sets A and B in [31] where both $I_{\varepsilon^+}(A, B)$ and $I_{\varepsilon^+}(B, A)$ were used to compare A and B . When the additive epsilon is used as a binary indicator, it is Pareto compliant [31].

3. COMPUTATIONAL EXPERIMENTS

In our computational experiments, we first search for a number of ordered pairs of non-dominated solution sets (A, B) where $A \triangleleft B$ holds between A and B . Then each performance indicator is applied to A and B to check whether it can correctly say that B is better than A . As test problems, we use DTLZ2 with a continuous concave Pareto front and multiobjective knapsack problems with a convex Pareto front. These two test problems are used because it was reported in [15] that the shape of the Pareto front (i.e., concave or convex) had a dominant effect on the consistency among performance indicators.

3.1 Two-Objective DTLZ2

As many-objective test problems, the DTLZ test suite [8] has been used in the literature. We apply NSGA-II to a two-objective 11-variable DTLZ2 problem 100 times under the following setting:

String: Real-number string of length 11,

Crossover: SBX with the probability 0.8,

Mutation: PM with the probability $1/n$ where n is the string length,

Population size: 100,
Termination condition: 500 generations.

From a single run of NSGA-II, a series of 500 non-dominated solution sets $\{S_1, S_2, \dots, S_{500}\}$ is obtained where S_t is a set of non-dominated solutions in the t th population. From the series of these 500 solution sets, we extract an ordered series of solution sets with respect to the relation better \triangleleft in the following manner.

Let the first solution set S_1 be the current solution set. The current solution set S_1 is compared with the next solution set S_2 to examine whether $S_1 \triangleleft S_2$ holds. If $S_1 \triangleleft S_2$ holds, the current solution set S_1 is replaced with S_2 , and the new current solution set S_2 is compared with the next solution set S_3 to examine whether $S_2 \triangleleft S_3$ holds. If $S_1 \triangleleft S_2$ does not hold, S_2 is removed from the series of the solution sets, and the current solution set S_1 is compared with the next solution set S_3 . The comparison of the current solution set S_i with the next solution set S_j is iterated until the last solution set S_{500} is examined. If $S_i \triangleleft S_j$ holds, the current solution set S_i is replaced with S_j , and S_j is compared with the next solution set S_{j+1} . If $S_i \triangleleft S_j$ does not hold, S_j is removed, and S_i is compared with S_{j+1} . In this manner, we extract a series of ordered non-dominated solution sets.

Let N be the number of obtained ordered non-dominated solution sets. Those solution sets are renumbered as S_1, S_2, \dots, S_N where $S_1 \triangleleft S_2 \triangleleft \dots \triangleleft S_N$ holds. From this series, we extract $(N-1)$ pairs of solution sets (S_t, S_{t+1}) with $S_t \triangleleft S_{t+1}$ where $t=1, 2, \dots, N-1$.

This procedure is applied to each of 100 runs of NSGA-II on the two-objective DTLZ2 problem. Then we have 2193 pairs of ordered non-dominated solution sets (A_i, B_i) with $A_i \triangleleft B_i$, $i=1, 2, \dots, 2193$. Two solution sets A_i and B_i in each of the obtained 2193 pairs are compared using each performance indicator. The reference point set Z in each indicator is generated by uniformly sampling 200 points along the known true Pareto front.

Experimental results are summarized in Table 1. Table 1 shows the percentage of correctly evaluated pairs ($I(A_i) > I(B_i)$), pairs incorrectly evaluated as being equivalent ($I(A_i) = I(B_i)$), and pairs with totally inconsistent evaluations ($I(A_i) < I(B_i)$).

In Table 1, about 2% of evaluation results by GD and IGD are Pareto incompliant: A_i is evaluated as being better than B_i whereas B_i is better than A_i . Those pairs are correctly evaluated by IGD⁺. That is, the modification of the distance calculation in IGD⁺ improves the quality of IGD as a performance indicator. Good results are also obtained from D1 with the weighted achievement scalarizing function for the distance calculation. In Table 1, about 9% of pairs are incorrectly evaluated as being equivalent by I_{ϵ^+} . This may be because $I_{\epsilon^+}(A, Z)$ calculates the maximum distance between the two non-dominated solution sets A and Z whereas the other indicators calculate the average distance between them.

Table 1. Results on the 2193 pairs of ordered non-dominated solution sets of the two-objective DTLZ2 problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	98.04%	0.00%	1.96%
IGD	97.95%	0.00%	2.05%
IGD ⁺	100.00%	0.00%	0.00%
D1	100.00%	0.00%	0.00%
Additive ϵ	91.24%	8.97%	0.00%

3.2 Many-Objective DTLZ2

In the same manner as in the previous subsection, we apply NSGA-II to DTLZ2 with 4, 6, 8 and 10 objectives. After 100 runs on each test problem, we obtain the following number of pairs of ordered non-dominated solution sets for each test problem:

145 pairs for the four-objective DTLZ2 with 13 variables,
0 pairs for the six-objective DTLZ2 with 15 variables,
0 pairs for the eight-objective DTLZ2 with 17 variables,
0 pairs for the ten-objective DTLZ2 with 19 variables.

The number of obtained pairs of ordered non-dominated solution sets severely decreases from 2193 for the two-objective DTLZ2 problem by increasing the number of objectives. This is because (i) most solutions become non-dominated with each other by increasing the number of objectives and (ii) each population has a large diversity throughout the 500 generations from a random initial population to the last population at the 500th generation.

Since no pair of ordered non-dominated solution sets is obtained for many-objective DTLZ2 test problems with 6-10 objectives, we only report experimental results on the four-objective DTLZ2 problem here. The reference point set is generated by randomly sampling 1000 points on the entire Pareto front. Experimental results are summarized in Table 2 in the same manner as in Table 1. In Table 2, the same result is obtained by IGD, IGD⁺ and D1.

Table 2. Results on the 145 pairs of ordered non-dominated solution sets of the four-objective DTLZ2 problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	97.93%	0.00%	2.07%
IGD	100.00%	0.00%	0.00%
IGD ⁺	100.00%	0.00%	0.00%
D1	100.00%	0.00%	0.00%
Additive ϵ	93.79%	6.21%	0.00%

3.3 Two-Objective Knapsack Problem

Multiobjective knapsack problems have been frequently used as discrete test problems in the EMO community since Zitzler & Thiele [30]. We use the two-objective 500-item knapsack problem in [30] in our computational experiment. NSGA-II is applied to this test problem 100 times under the following setting:

String: Binary string of length 500,
Crossover: Uniform crossover with the probability 0.8,
Mutation: Bit-flip mutation with the probability 1/500,
Constraint handling: Greedy repair in [30],
Population size: 100,
Termination condition: 2000 generations.

From 100 runs of NSGA-II, we obtain 47993 pairs of ordered non-dominated solution sets (A_i, B_i) with $A_i \triangleleft B_i$. As the reference points, we use all the known 1427 Pareto optimal solutions. Evaluation results of the 47993 pairs of ordered non-dominated solution sets are summarized in Table 3.

In Table 3, many inconsistent results (18.20%) are obtained by GD. IGD also generate inconsistent results for 5.21% pairs of solution sets. Such an inconsistency is removed by the use of IGD⁺. This observation shows the effect of the modification of the distance calculation. However, about 17.6% pairs of A_i and B_i with $A_i \triangleleft B_i$ are evaluated as being equivalent by IGD and IGD⁺.

Table 3. Results on the 47993 pairs of ordered non-dominated solution sets of the two-objective 500-item knapsack problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	81.80%	0.00%	18.20%
IGD	77.22%	17.58%	5.21%
IGD ⁺	82.41%	17.59%	0.00%
D1	99.96%	0.04%	0.00%
Additive ε	14.83%	85.17%	0.00%

One question in Table 3 is the difference in the number of pairs with $I(A_i) = I(B_i)$ between IGD⁺ and D1. As we have already shown in Fig. 5, these two indicators have similar contour lines of the distance from a reference point (see Fig. 5 (b) and Fig. 5 (c)).

In Fig. 6, we show an example of two non-dominated solution sets A_i and B_i with $A_i \triangleleft B_i$, which are correctly evaluated by D1 as $D1_R(A_i) > D1_R(B_i)$ and incorrectly evaluated by IGD⁺ as $IGD^+(A_i) = IGD^+(B_i)$. The difference between these two solution sets is only two solutions highlighted by a red circle in Fig. 6 (a) and Fig. 6 (b). These two solutions are contributed to the calculation of $D1_R(A_i)$ and $D1_R(B_i)$. However, they are not contributed to $IGD^+(A_i)$ and $IGD^+(B_i)$, which leads to $IGD^+(A_i) = IGD^+(B_i)$.

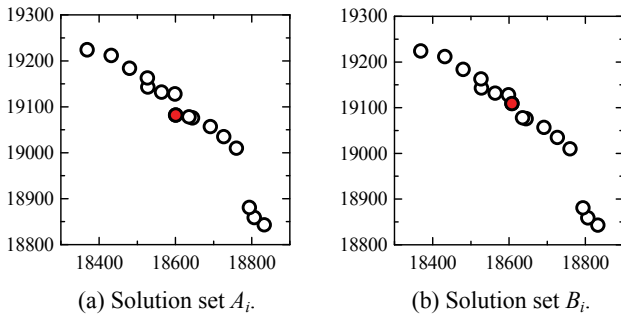


Figure 6. Two non-dominated solution sets A_i and B_i for which $A_i \triangleleft B_i$, $D1_R(A_i) > D1_R(B_i)$ and $IGD^+(A_i) = IGD^+(B_i)$ hold.

Fig. 6 suggests that more solutions are used in the calculation of D1 than that of IGD⁺. To examine this issue, we calculate the average number of solutions in A_i and B_i used for the calculation of each indicator in Table 3. The following results are obtained:

- GD: 33.4 solutions (all solutions in A_i and B_i),
- IGD: 13.3 solutions,
- IGD⁺: 13.3 solutions,
- D1: 33.0 solutions,
- Additive ε : 1.0 solution.

Since GD is the average distance over all solutions, 33.4 is the average number of solutions in A_i and B_i . Since the maximum distance is calculated in the additive ε , I_{ε^+} depends on only a single solution. Whereas the other indicators calculate the average distance, more solutions are used for the calculation of D1 than IGD and IGD⁺. This may be the reason for the difference in the percentages of $I(A_i) = I(B_i)$ between IGD⁺ and D1 in Table 3.

The next question is why more solutions are used for the calculation of D1 than IGD⁺. The reason can be explained by the difference in the distance calculation between these two indicators. In Fig. 7, we explain the dependency of the nearest solution on the distance calculation in each performance measure. In Fig. 7 (b), a different solution is chosen as the nearest solution for each

reference point (i.e., all the three solutions in Fig. 7 (b) are used in the calculation of D1). However, in Fig. 7 (a), the solution a_2 is not chosen as the nearest solution for any reference points (a_3 is chosen as the nearest solution for z_2 and z_3). As a result, only the two solutions a_1 and a_3 are used in the calculation of IGD⁺. As illustrated in Fig. 7 (a), a non-dominated solution at a locally concave location is not likely to be selected as the nearest solution of a reference point when the Euclidean distance is used and the reference points are not close to the obtained solutions.

To examine this issue, we show in Fig. 8 a Voronoi diagram for the three solutions in Fig. 7 using the Euclidean distance and the weighted achievement scalarizing function. In Fig. 8 (a), the solution a_2 has a triangular territory in its right-upper region. Only when a reference point is in this triangular territory, a_2 is selected as its nearest solution. The same solution a_2 has a belt-like territory in its right-upper direction in Fig. 8 (b).

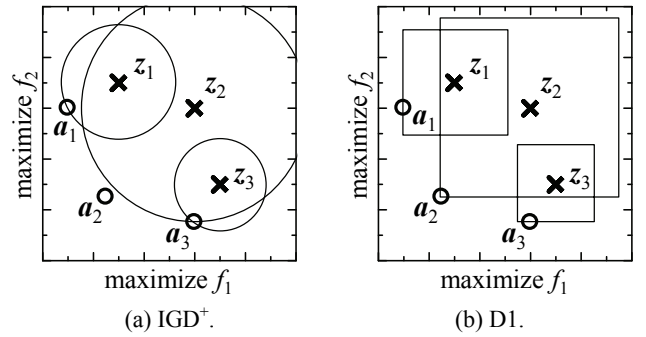


Figure 7. Contour line corresponding to the distance from each reference point to its nearest solution.

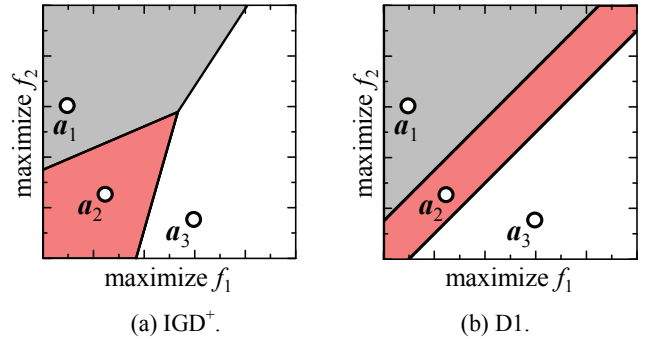


Figure 8. Voronoi diagram for the three solutions in Fig. 7.

3.4 Many-Objective Knapsack Problems

In the same manner as in the previous subsection, we perform computational experiments on 500-item multiobjective knapsack problems with 4-10 objectives. We use the same four-objective 500-item problem as in [30]. Multiobjective knapsack problems with more than four objectives are generated by adding objectives to the four-objective problem. The added objectives are randomly generated as in [30]. Pairs of ordered non-dominated solution sets are selected from 100 runs on each test problem. Reference points are specified by choosing non-dominated solutions among the merged population of all the selected pairs of solution sets and the final populations of 100 runs on each test problem.

In Table 4, we show the number of obtained pairs of ordered non-dominated solution sets (A_i, B_i) with $A_i \triangleleft B_i$, and the number of reference points in the reference point set for each test problem. In

Tables 5-8, experimental results are summarized in the same manner as in Table 3. From the comparison of Table 3 with Tables 5-8, we can see that much better results are obtained in Tables 5-8 than Table 3. Among the examined indicators, the best results are obtained by D1.

Table 4. The number of pairs of ordered non-dominated solution sets and the number of reference points.

Problem	(A_i, B_i)	Reference points
Two-Objective	47993	1427
Four-Objective	4704	2259
Six-Objective	1913	3780
Eight-Objective	1030	4938
Ten-Objective	624	5084

Table 5. Results on the four-objective knapsack problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	98.15%	0.00%	1.85%
IGD	99.30%	0.60%	0.11%
IGD ⁺	99.36%	0.64%	0.00%
D1	100.00%	0.00%	0.00%
Additive ε	53.93%	46.07%	0.00%

Table 6. Results on the six-objective knapsack problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	99.53%	0.00%	0.47%
IGD	99.53%	0.47%	0.00%
IGD ⁺	99.53%	0.47%	0.00%
D1	100.00%	0.00%	0.00%
Additive ε	71.20%	28.80%	0.00%

Table 7. Results on the eight-objective knapsack problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	99.32%	0.00%	0.68%
IGD	98.74%	1.26%	0.00%
IGD ⁺	98.54%	1.46%	0.00%
D1	100.00%	0.00%	0.00%
Additive ε	78.54%	21.46%	0.00%

Table 8. Results on the ten-objective knapsack problem.

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	99.36%	0.00%	0.64%
IGD	96.96%	3.04%	0.00%
IGD ⁺	96.63%	3.37%	0.00%
D1	100.00%	0.00%	0.00%
Additive ε	81.09%	18.91%	0.00%

4. DISCUSSIONS AND CONCLUSIONS

We examined the distance calculation in the IGD indicator. We also compared distance-based indicators through computational experiments and obtained the following observations:

- (i) Inconsistent evaluation results by IGD were actually obtained and removed by using IGD⁺ instead of IGD.
- (ii) Best evaluation results with respect to the Pareto compliance were obtained by D1 among GD, IGD, IGD⁺, D1 and I_{ε^+} .

The first observation supports the use of IGD⁺ instead of IGD. This observation is consistent with theoretical discussions in [12]. The use of IGD⁺ instead of IGD was also suggested in [11] by comparing them with respect to the sensitivity to the specification of reference points. The second observation may need further studies using other test problems and other EMO algorithms. This is because the best indicator depends on the setting of experiments. Whereas we used the Voronoi diagram in Fig. 8 to explain the best results by D1 in Table 3 (and Fig. 7), the same diagram can be used to explain the advantage of IGD⁺ over D1 if the reference points are much closer to the three solutions in Fig. 7. Actually a different observation can be obtained from the same experiment on a different test problem. In Table 9, we show experimental results on a two-objective WFG1 problem [10]. The best results are obtained from IGD⁺ in Table 9. Table 10 shows experimental results on the same WFG1 problem by MOEA/D with the weighted Techebycheff function [26]. In Table 10, the best results are obtained from IGD⁺. However, good results are also obtained by all indicators. These observations suggest the necessity of further studies of each indicator.

Table 9. Results on the 11782 pairs of ordered non-dominated solution sets of a two-objective WFG1 problem (NSGA-II).

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	83.58%	0.00%	16.42%
IGD	96.92%	0.87%	2.21%
IGD ⁺	99.13%	0.87%	0.00%
D1	98.50%	1.50%	0.00%
Additive ε	50.53%	49.47%	0.00%

Table 10. Results on the 1672 pairs of ordered non-dominated solution sets of a two-objective WFG1 problem (MOEA/D).

Indicator	Correct $I(A_i) > I(B_i)$	Equivalent $I(A_i) = I(B_i)$	Inconsistent $I(A_i) < I(B_i)$
GD	98.56%	0.00%	1.44%
IGD	98.56%	0.00%	1.44%
IGD ⁺	100.00%	0.00%	0.00%
D1	99.70%	0.30%	0.00%
Additive ε	98.92%	1.08%	0.00%

5. REFERENCES

- [1] Bader, J. and Zitzler, E. 2011. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation* 19, 1 (Spring 2011) 45-76.
- [2] Beume, N., Naujoks, B., and Emmerich, M. 2007. SMS-EMOA: multiobjective selection based on dominated hypervolume. *European Journal of Operational Research* 180, 3 (September 2007) 1653-1669.
- [3] Bosman, P.A.N. and Thierens, D. 2003. The balance between proximity and diversity in multiobjective evolutionary algorithms. *IEEE Trans. on Evolutionary Computation* 7, 2 (April 2003) 174-188.

- [4] Bradstreet, L., While, L., and Barone, L. 2008. A fast incremental hypervolume algorithm. *IEEE Trans. on Evolutionary Computation* 12, 6 (December 2008) 714-723.
- [5] Coello, C.A.C. and Sierra, M.R. 2004. A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. *Lecture Notes in Computer Science 2972: MICAI 2004* (April 2004) 688-697.
- [6] Czyzak, P. and Jaszkiwicz, A. 1998. Pareto simulated annealing - A metaheuristic technique for multiple-objective combinatorial optimization. *Journal of Multi-Criteria Decision Analysis* 7, 1 (January 1998) 34-47.
- [7] Deb, K., Paratop, A., Agarwal, S., and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evolutionary Computation* 6, 2 (April 2002) 182-197.
- [8] Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. 2002. Scalable multi-objective optimization test problems. *Proc. of 2002 Congress on Evolutionary Computation* (Honolulu, USA, May 12-17, 2002) 825-830.
- [9] Hansen, M.P. and Jaszkiwicz, A. 1998. Evaluating the quality of approximations of the non-dominated set. *Technical Report IMM-REP-1998-7*, Technical University of Denmark.
- [10] Huband, S., Hingston, P., Barone, L., and While, L. 2006. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Trans. on Evolutionary Computation* 10, 5 (October 2006) 477-506.
- [11] 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. *Proc. of 2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making* (Orlando, USA, December 9-12, 2014) 170-177.
- [12] Ishibuchi, H., Masuda, H., Tanigaki, Y., and Nojima, Y. 2015. Modified distance calculation in generational distance and inverted generational distance. *Lecture Notes in Computer Science 9019: EMO 2015, Part II* (Springer, Berlin, March-April 2015) 110-125.
- [13] Ishibuchi, H., Tsukamoto, N., and Nojima Y. 2008. Evolutionary many-objective optimization: A short review. *Proc. of 2008 IEEE Congress on Evolutionary Computation* (Hong Kong, China, June 1-6, 2008) 2424-2431.
- [14] Ishibuchi, H., Yoshida, T., and Murata, T. 2003. Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling. *IEEE Trans. on Evolutionary Computation* 7, 2 (April 2003) 204-223.
- [15] Jiang, S., Ong, Y., Zhang, J., and Feng, L. 2014. Consistencies and contradictions of performance metrics in multiobjective optimization. *IEEE Trans. on Cybernetics* 44, 12 (December 2014) 2391-2404.
- [16] Knowles, J.D. and Corne, D.W. 2002. On metrics for comparing non-dominated sets. *Proc. of 2002 Congress on Evolutionary Computation* (Honolulu, May 12-17, 2002) 711-716.
- [17] Li, M., Yang, S., and Liu, X. Shift-based density estimation for Pareto-based algorithms in many-objective optimization. *IEEE Trans. on Evolutionary Computation* 18, 3 (June 2014) 348-365.
- [18] Russo, L.M.S. and Francisco, A.P. 2014. Quick hypervolume. *IEEE Trans. on Evolutionary Computation* 18, 4 (August 2014) 481-502.
- [19] Schuetze, O., Esquivel, X., Lara, A., and Coello, C.A.C. 2012. Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Trans. on Evolutionary Computation* 16, 4 (August 2012) 504-522.
- [20] Sierra, M.R. and Coello, C.A.C. 2004. A new multi-objective particle swarm optimizer with improved selection and diversity mechanisms. *Technical Report of CINESTAV-IPN* (November 23, 2004).
- [21] Van Veldhuizen D.A. 1999. Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations. Ph.D. Thesis, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, USA (1999)
- [22] von Lüken, C., Barán, B., and Brizuela, C. 2014. A survey on multi-objective evolutionary algorithms for many-objective problems. *Computational Optimization and Applications* 53, 3 (July 2014) 707-756.
- [23] Wagner, T., Beume, N., and Naujoks, B. 2007. Pareto-, aggregation-, and indicator-based methods in many-objective optimization. *Lecture Notes in Computer Science 4403: Evolutionary Multi-Criterion Optimization - EMO 2007*, (Springer, Berlin, March 2007), 742-756.
- [24] While, L., Bradstreet, L., and Barone, L. 2012. A fast way of calculating exact hypervolumes. *IEEE Trans. on Evolutionary Computation* 16, 1 (February 2012) 86-95.
- [25] Yang, S., Li, M., Liu, X., and Zheng, J. 2013. A grid-based evolutionary algorithm for many-objective optimization. *IEEE Trans. on Evolutionary Computation* 17, 5 (October 2013) 721-736.
- [26] Zhang, Q. and Li, H. 2007. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. on Evolutionary Computation* 11, 6 (December 2007) 712-731.
- [27] Zitzler, E., Brockhoff, D., and Thiele, L. 2007. The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. *Lecture Notes in Computer Science 4403: EMO 2007* (Springer, Berlin, March 2007) 862-876.
- [28] Zitzler, E., Laumanns, M., and Thiele, L. 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK-Report*, No. 103, ETH Zurich.
- [29] Zitzler, E. and Thiele, L. 1998. Multiobjective optimization using evolutionary algorithms - A comparative case study. *Lecture Notes in Computer Science 1498: Parallel Problem Solving from Nature - PPSN V* (September 1998) 292-301.
- [30] Zitzler, E. and Thiele, L. 1999. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Trans. on Evolutionary Computation* 3, 4 (November 1999) 257-271.
- [31] 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 Trans. on Evolutionary Computation* 7, 2 (April 2003) 117-132.

