A Performance Comparison Indicator for Pareto Front Approximations in Many-Objective Optimization

Miqing Li
Department of Computer Science
Brunel University London
Uxbridge UB8 3PH, U. K.
miqing.li@brunel.ac.uk

Shengxiang Yang
Centre for Computational Intelligence
De Montfort University
Leicester LE1 9BH, U. K
syang@dmu.ac.uk

Xiaohui Liu
Department of Computer Science
Brunel University London
Uxbridge UB8 3PH, U. K.
xiaohui.liu@brunel.ac.uk

ABSTRACT

Increasing interest in simultaneously optimizing many objectives (typically > 3) of problems leads to the emergence of various many-objective algorithms in the evolutionary multiobjective optimization (EMO) field. However, in contrast to the development of algorithm design, how to assess many-objective algorithms has received scant concern. Many performance indicators are designed in principle for any number of objectives, but in practice are invalid or infeasible to be used in many-objective optimization. In this paper, we explain the difficulties that popular performance indicators face and propose a performance comparison indicator (PCI) to assess Pareto front approximations obtained by many-objective algorithms. PCI evaluates the quality of approximation sets with the aid of a reference set constructed by themselves. The points in the reference set are divided into many clusters, and the proposed indicator estimates the minimum moves of solutions in the approximation sets to weakly dominate these clusters. PCI has been verified both by an analytic comparison with several well-known indicators and by an empirical test on four groups of Pareto front approximations with different objective numbers and problem characteristics.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.2.8 [**Problem Solving, Control Methods, and Search**]: Heuristic Methods

Keywords

Multi-objective optimization, many-objective optimization, performance assessment indicator

1. INTRODUCTION

Over the past decade, many-objective optimization has attracted great attention and gradually developed into an independent research topic in the evolutionary multiobjective optimization (EMO) field [9, 17]. In the last five years, a wide variety of many-objective optimization techniques have been proposed, ranging from the modification of Pareto-based algorithms [22] and the introduction of decomposition-based and indicator-based algorithms [1, 5] to the design of new optimizers specially for many-objective problems [11, 31, 32, 34].

However, in contrast to the development of algorithm design, how to assess and compare many-objective optimizers has received scant concern [21]. Many performance indicators are designed in principle for any number of objectives, but in practice are invalid or infeasible when assessing Pareto front approximations (i.e., solution sets obtained by population-based search algorithms) with high dimensions. In general, the difficulties of performance indicators in assessing high-dimensional Pareto front approximations can be summarized as follows.

- *Ineffectiveness of Pareto dominance.* Since the portion of the space that a solution dominates decreases exponentially with the number of objectives, most solutions in different approximation sets are likely to be incomparable under the criterion of Pareto dominance. This can lead to the ineffectiveness of many performance indicators based on the comparison of solutions' Pareto dominance relation, such as the *coverage* [36] and G-metric [26].

- *Rapid increase of time or space requirement.* The time or space required by some performance indicators increases exponentially with the number of objectives, such as the *hypervolume* [36], *hyperarea radio* [29], and *diversity measure* [6]. This can affect their application in many-objective optimization.

- *Difficulties of parameter setting.* Purshouse and Fleming [27] have shown that the *sweet-spot* of algorithm parameter setting that produces good results could shrink markedly in many-objective optimization. This could also apply to parameter setting of performance indicators. In general, the sensitiveness of assessment results to an indicator's parameter(s) increases with the number of objectives. Two high-dimensional approximation sets could return completely contrary results when assessed by an indicator with slightly different settings of its parameter(s) [21].

- *Difficulties of the substitution of Pareto front.* Many

performance indicators require a reference set as a substitution of the Pareto front, such as GD [29] and IGD [2]. However, to represent a higher-dimensional Pareto front, exponentially more points are needed. As shown by Ishibuchi et al [14], insufficient points can easily lead to an inaccurate or even misleading assessment result.

In this paper, we propose a performance comparison indicator (PCI) to assess high-dimensional Pareto front approximations, attempting to overcome the above difficulties. PCI can be partially viewed as an integration of two effective performance indicators, IGD^+ [16] and the *binary ϵ-indicator* [37]. Like IGD^+, when comparing an individual p with an individual q, PCI only considers the objectives where p is inferior to q. Like the binary ϵ-indicator, PCI considers the comparison of multiple approximation sets, providing relative quality for these sets. This can avoid theoretical limitations of unary performance indicators [37]. Moreover, PCI uses a reference set constructed by the tested approximation sets themselves, rather than one specified by the user. More importantly, unlike many performance indicators (such as IGD, IGD^+, and the *average Hausdorff Distance* [28]) which compare each point in the reference set with the approximation set, the proposed indicator divides points in the reference set into many clusters and estimates the minimum move of the approximation sets to weakly dominate these clusters.

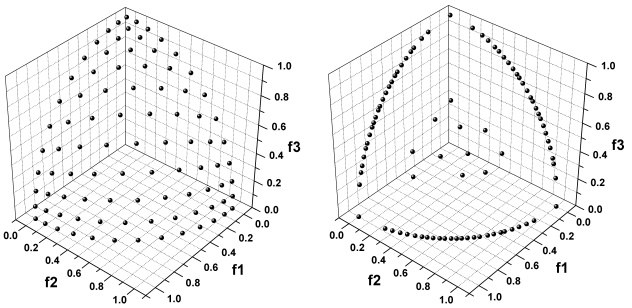
Note that this paper focuses on the comparison of Pareto front approximations rather than on search algorithms, that is, we assume that for each algorithm only one run is performed. In addition, without loss of generality, we consider minimization MOPs throughout this paper. The rest of this paper is organized as follows. Related work is reviewed in Section 2. Section 3 is devoted to the description of the proposed indicator. Analytic and empirical studies are carried out in Section 4 and Section 5, respectively. Finally, Section 6 concludes the paper.

2. RELATED WORK

Despite the difficulties in assessing high-dimensional Pareto front approximations, some effort has been made. Jaimes and Coello [18] measured the Tchebycheff distance of an approximation set to the "knee" of the Pareto front in many-objective optimization. This measurement is based on the assumption that, in the absence of particular preference information, the decision maker may have more interest in the knee of a problem's Pareto front. Recently, Li et al [21] proposed an indicator to compare the diversity of approximation sets. They put all approximation sets into a grid environment and calculated the contribution of each set to those nonempty hyperboxes. A problem in these indicators is that they only focus on one particular aspect of the performance of approximation sets, failing to provide a general evaluation of the sets' performance.

On the other hand, some classic performance indicators which measure the overall quality of approximation sets have been frequently used to test and compare many-objective optimizers [10, 24, 30]. Examples are the hypervolume [36], ϵ-indicator [37], and IGD [2] (also its variation IGD^+ [16]). Next, we will briefly discuss their pros and cons, especially when assessing high-dimensional approximation sets.

The hypervolume (HV) indicator [36] calculates the volume of the objective space enclosed by an approximation set and a reference point. HV has good theoretical properties and can give a general evaluation of the set, but its computational complexity increases exponentially with the number of objectives. These lead to its dominant application in the 2- or 3-objective problems. While the Monte Carlo method can largely reduce the time cost [1, 3], how to choose a proper reference point in the HV calculation is an important issue [19] and its difficulty also increases with the number of objectives. In addition, the HV indicator generally prefers the knee and boundary points of the Pareto front to well-distributed ones. Figure 1 gives such an example, where we consider two sets of Pareto optimal solutions[1] obtained by two well-known EMO algorithms, MOEA/D [33] and IBEA [35], on test problem DTLZ2 [8]. Table 1 gives the HV results of these two solution sets with various settings of the reference point. As shown, the solution set with better distribution (obtained by MOEA/D) has a worse (lower) HV evaluation value than its competitor, regardless of the choice of the reference point.



(a) MOEA/D (b) IBEA

Figure 1: An example that HV prefers the knee and boundary points on the Pareto front, where two sets of Pareto optimal solutions on DTLZ2 are obtained by MOEA/D and IBEA. The solution set with better distribution (obtained by MOEA/D) has a worse (lower) HV result, as given in Table 1.

Table 1: HV results of the two sets in Figure 1 under different reference points. The range of DTLZ2's Pareto front is $[0, 1]$ for all objectives.

Reference point	MOEA/D	IBEA
(1.0, 1.0, 1.0)	4.1413E–1	4.1525E–1
(1.1, 1.1, 1.1)	7.4484E–1	7.4596E–1
(1.2, 1.2, 1.2)	1.1418E+0	1.1430E+0
(1.4, 1.4, 1.4)	2.1578E+0	2.1590E+0
(1.7, 1.7, 1.7)	4.3268E+0	4.3280E+0
(2.0, 2.0, 2.0)	7.4138E+0	7.4150E+0
(2.5, 2.5, 2.5)	1.5039E+1	1.5040E+1
(3.0, 3.0, 3.0)	2.6414E+1	2.6415E+1

The unary ϵ-indicator [37] measures the minimum factor ϵ for an approximation set such that any point in a reference set is ϵ-dominated [20] (additively or multiplicatively) by at least one solution in the approximation set. One weakness in the ϵ-indicator is that its evaluation value is only related to one particular solution in an approximation set, which

[1]Here, the number of DTLZ2's decision variables n is set to $m - 1$ (m is the number of objectives) to ensure that all solutions produced by the algorithms are Pareto optimal.

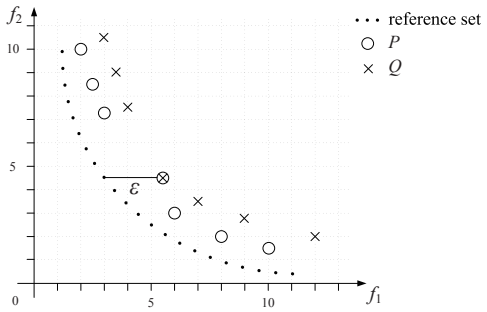


Figure 2: An example that the unary additive ϵ-indicator fails to distinguish between two approximation sets. P and Q have the same evaluation result ($\epsilon = 2.5$).

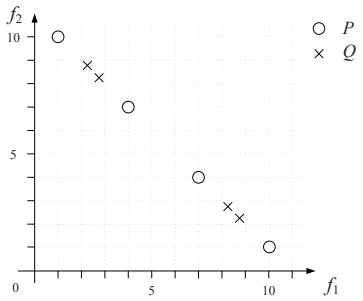


Figure 3: An example that IGD and IGD$^+$ fails to reflect the performance difference between approximation sets, where the reference set is constructed by the approximation sets themselves. P and Q have the same IGD and IGD$^+$ evaluation results (0.884 and 0.625 respectively).

could lead to an inaccurate estimation of the set's performance. Figure 2 gives an example that the unary additive ϵ-indicator fails to distinguish between approximation sets (P and Q). As can be seen from the figure, P clearly outperforms Q, but the two sets have the same evaluation result.

Zitzler et al [37] also presented a binary ϵ-indicator and stated its desirable features, such as being free from the limitations of unary performance indicators, having a low computational cost, and representing natural extension to the evaluation of approximation schemes in theoretical computer science. However, the ϵ-indicator (both unary and binary) only considers one particular objective of the problem (i.e., the objective where the considered approximation sets have the biggest difference), ignoring the difference on the others. This unavoidably leads to an information loss, especially for optimization problems with many objectives. Consider two 10-objective solutions, $p = (0, 0, 0, ..., 0, 1)$ and $q = (1, 1, 1, ..., 1, 0)$. p performs better on nine objectives and q does better only on the last objective, but they have same evaluation result ($\epsilon = 1$).

IGD [2] calculates the average Euclidean distance from each point in a reference set (a substitution of the Pareto front) to its closest solution in an approximation set. This overcomes the ϵ-indicator's problem of only returning one particular objective value of one particular solution. However, one weakness of IGD is its non-compliance with the Pareto dominance relation. Ishibuchi et al [16] presented several examples that an approximation set P could obtain a worse IGD result than an approximation set Q even if P Pareto-dominates Q. This even happens when the reference set is exactly the Pareto front. In fact, a reference set with sufficient, well-distributed points can only reduce the possibility of this misleading result, rather than eliminating it completely.

To solve this problem, Ishibuchi et al [15, 16] proposed a modified IGD (IGD$^+$), only considering the superior objective values of the reference point to the solution of the approximation set in their distance calculation. This can enable the indicator to be weakly Pareto compliant[2] [16]. However, like IGD, IGD$^+$ also needs a reference set specified by the user. While IGD$^+$ alleviates the severe sensitivity of IGD to the reference set, different reference sets can cause IGD$^+$ to prefer different approximation sets. This may lead to inconsistent evaluation results among these sets. In fact, how to specify a proper preference set in many-objective optimization is a challenging issue [15]; this also applies to artificial test functions with a known Pareto front, given that exponentially increasing points are needed to accurately represent a higher-dimensional Pareto front.

For an MOP with an unknown Pareto front, a practical method of constructing a reference set is to use the nondominated solutions of all solutions in the considered approximation sets. This method is widely used in many-objective optimization [12, 15]. However, in this method, there is no information about the location and distribution of the points in the reference set. Some points could be far away from the Pareto front, like the *dominance resistant solutions*[3] [8, 13] which are preserved more likely in a higher-dimensional objective space. Some points could be located closely or even overlapping, and this will result in some areas overcrowded and some others empty. Overall, such a reference set with an arbitrary distribution of nondominated solutions can significantly decrease the accuracy of the evaluation result. Figure 3 gives an example that with the reference set constructed by the tested approximation sets, IGD and IGD$^+$ fail to reflect the performance difference between the sets. As shown, the points of the reference set are not well distributed, which leads to the result that P with uniformly-distributed solutions has the same evaluation value as Q whose solutions concentrate in two small areas.

3. PROPOSED METHOD

As indicated before, the binary ϵ-indicator directly considers the performance difference of approximation sets, but only returns one particular objective value of one particular solution; IGD$^+$ (or IGD) considers the whole approximation set and multiple (or all) objectives, but requires a reference set which is hard to specify in many-objective optimization.

[2]For a weakly Pareto compliant indicator I, if a set P weakly dominates a set Q, then $I(P)$ is not worse than $I(Q)$ [37].

[3]Dominance resistant solutions in a set are the solutions with a quite poor value in at least one of the objectives but with (near) optimal values in some others, and thus are nondominated in the set.

So, an intuitive idea of developing an indicator is to consider their combination, that is, using the calculation method of IGD^+ to directly compare two approximation sets (i.e., two approximation sets are viewed as mutual reference sets in the IGD^+ calculation). This indicator, though, also suffers from the effect of a poorly-distributed reference set, e.g., still failing to distinguish between the approximation sets in Figure 3. In fact, since each point (regardless of its location) in the reference set is equally treated in the evaluation, the areas with many similar (or duplicate) points have more effect than the same-size ones with few points on the evaluation result. This naturally leads such an indicator to prefer an approximation set having similar distribution with the considered reference set. In addition, since a binary indicator manipulates only two approximation sets, many comparisons are required when more than two approximation sets are involved ($\binom{n}{2}$ comparisons for n sets).

Given the above, this paper presents a performance comparison indicator (PCI), capable of assessing multiple approximation sets (any number) in a single run. PCI constructs a reference set by using all the tested Pareto front approximations and assesses each approximation on the basis of this reference set. Rather than dealing with each point in the reference set, PCI divides all points into many clusters according to their distribution (the method of the division will be explained later), and considers the relationship between the clusters and solutions in the approximation set. Specifically, PCI measures the minimum move distance of one solution to weakly dominate[4] all points in a cluster, as defined as follows.

Definition 1 (Dominance Distance of a point to a point set): Let p be a point and Q be a set of points $\{q_1, q_2, ..., q_k\}$ ($k \geq 1$). The dominance distance $D(p, Q)$ of p to Q is defined as the minimum move of p in the objective space such that p weakly dominates all points in Q. It can be calculated as

$$D(p,Q) = \sqrt{\sum_{i=1}^{m}(p^{(i)} - d(p^{(i)}, Q))^2} \quad (1)$$

$$d(p^{(i)}, Q) = \begin{cases} \min\{q_1^{(i)}, q_2^{(i)}, ..., q_k^{(i)}\}, & \text{if } p^{(i)} > \min\{q_1^{(i)}, q_2^{(i)}, ..., q_k^{(i)}\} \\ p^{(i)}, & \text{otherwise} \end{cases}$$

where $p^{(i)}$ denotes the objective value of point p in the ith objective and m is the number of objectives.

Dominance distance $D(p, Q)$ only considers the objectives where Q performs better than p (i.e., the best value of the points in Q is better than that of p), regardless of the advantage of p over Q. This can lead to the indicator free from the effect of poorly-converged reference points, such as the dominance resistant solutions. The range of $D(p, Q)$ is from 0 to ∞; the smaller the better. If p performs slightly worse than Q in only few objectives, $D(p, Q)$ will be small. Only when p weakly dominates all points in Q, $D(p, Q) = 0$.

Note that when Q has only one point q, the dominance distance of p to Q becomes $D(p, q)$, i.e., the minimum move of p to weakly dominate q in the objective space. This method has already been used in the area to measure the difference between two solutions. IGD^+ used this measure to replace the Euclidean distance in the IGD calculation, making the indicator compliant with weak Pareto dominance. Li et al [22] adopted a similar principle to improve solutions' density estimation, enabling Pareto-based algorithms to be suitable for many-objective problems. In fact, $D(p, Q)$ can also be viewed as the dominance distance of p to a point, q_{ideal}, constructed by the best value of each objective for all k points in Q.

In contrast to IGD and IGD^+ which consider each point in the reference set, PCI replaces a set of similar points with an imaginary (ideal) point constructed by themselves. This avoids the overlapping effect of similar points and also takes into account the superiority of these points on each objective. In fact, the dominance distance of a point to a set of points is larger than or equal to the maximum dominance distance of the point to all members in the set, but is smaller than their summation, i.e., $\max\{D(p, q_1), ..., D(p, q_k)\} <= D(p, Q) < D(p, q_1) + ... + D(p, q_k)$. The proof is omitted here for lack of space.

Given that the dominance distance of a point to a cluster reflects the advantage of the cluster over the point, one may think to use this dominance distance to consider all clusters in the reference set to assess an approximation set. However, this evaluation may lead to an inaccurate result when in a cluster there is more than one solution from the assessed approximation set. To explain this, we introduce the following definition.

Definition 2 (Dominance Distance of a point set to another): Let P be a set of points and Q be a set of points. The dominance distance $D(P, Q)$ of P to Q is defined as the minimum total distance of the move of points of P in the objective space such that for any point $q \in Q$, there is at least one point $p \in P$ that weakly dominates q.

The dominance distance of two sets reflects their performance difference. If P (weakly) dominates Q, then $D(P, Q) = 0$.

In the proposed indicator, since the reference set consists of all the approximation sets, a cluster can contain points from different approximation sets. Let a cluster C be comprised of P and Q, where $P = \{p_1, ..., p_i\}$ is from the assessed approximation set and $Q = \{q_1, ..., q_j\}$ from other approximation sets. Apparently, $D(P, C) = D(P, Q)$. When $i = 1$, $D(P, C)$ is the dominance distance of p_1 to the ideal point of C. When $i \geq 2$, $D(P, C)$ could be smaller than $\min\{D(p_1, C), ..., D(p_i, C)\}$.

Figure 4 gives an example to illustrate this situation. Consider the dominance distance of three solution sets P_1 ,P_2, and P_3 to three clusters C_1 ,C_2, and C_3, respectively in the figure, where $P_1 \in C_1, P_2 \in C_2, P_3 \in C_3, P = P_1 \cup P_2 \cup P_3$. For P_1 which has only one solution, $D(P_1, C_1)$ is the dominance distance of the solution to the cluster, ($D(P_1, C_1) = (0.5^2 + 0.5^2)^{0.5} \approx 0.707$). For P_2 which has two solutions, $D(P_2, C_2)$ is the dominance distance of the upper solution to the ideal point of the two Q solutions, and this is smaller than the minimum of the dominance distance of the two single P_2 solutions to the cluster ($0.559 < \min\{1.031, 1.25\}$). Cluster C_3 gives an extreme situation where no move of the P_3 solutions is needed to weakly dominate all solutions in the cluster (i.e., $D(P_3, C_3) = 0$), but the dominance distance of the two single P_3 solutions to the cluster is 1.0.

From the above, it is clear that for a cluster $C = P \cup Q$ ($P = \{p_1, ..., p_i\}, i \geq 2, Q = \{q_1, ..., q_j\}$), the required move of P to weakly dominate C can be (much) smaller than the

[4]For two solutions p and q, p weakly dominates q ($p \preceq q$) if and only if p is not worse than q in all objectives.

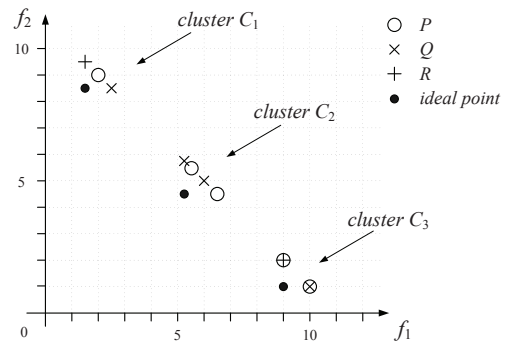


Figure 4: An example that the dominance distance of a set of solutions to a cluster can be smaller than the minimum of their single dominance distance to the cluster. For three sets P_1 ,P_2, and P_3 ($P_1 \in C_1, P_2 \in C_2, P_3 \in C_3, P = P_1 \cup P_2 \cup P_3$), their dominance distance to C_1, C_2 and C_3 is 0.707, 0.559 and 0.0, respectively, while the minimum of their single solution's dominance distance to C_1, C_2 and C_3 is 0.707, 1.031 and 1.0, respectively.

minimum move of any one point in P to weakly dominate C. This is because the points in C can be "divided and conquered" by multiple solutions in P. However, how to determine $D(P, C)$ (i.e., the minimum move of $\{p_1, ..., p_i\}$ to weakly dominate $\{q_1, ..., q_j\}$) could be quite time consuming; there are i^j possibilities for $p_1, p_2, ..., p_i$ to divide $q_1, q_2, ..., q_j$. Here, we approximately measure it by

$$D'(P,C) = \max\{\min\{D(p_1,q_1),\ldots,D(p_i,q_1)\}, \ldots, \min\{D(p_1,q_j),\ldots,D(p_i,q_j)\}\} \quad (2)$$

when $i \geq 2$. This only requires $i \times j$ comparisons. Although $D'(P, C) \leq D(P, C)$, their difference is generally slight when the size of C is small. In the example of Figure 4, $D'(P_2, C_2) = 0.5 < D(P_2, C_2) = 0.559$ and $D'(P_3, C_3) = D(P_3, C_3) = 0$.

Algorithm 1 gives the main procedure of the proposed indicator. It is necessary to mention here that we normalize approximation sets before the evaluation, in order to make PCI be able to deal with MOPs with non-commensurable objective functions. Specifically, the range of an MOP's Pareto front is used to implement the normalization if it is available; otherwise, the boundary of the constructed reference set is done. From Algorithm 1, it can be seen that PCI treats the clusters differently. For a cluster where the number of solutions from the assessed approximation set is less than two, PCI takes account of the minimum move of one solution in the approximation set to weakly dominate the cluster (step 8), on the basis of Eq. (1). Otherwise, PCI estimates the minimum move of the set's solutions in the cluster to weakly dominate the cluster by Eq. (2) (step 10).

An important issue in the proposed indicator is to cluster the points in the reference set before assessing approximation sets (step 2). Here, we use a greedy method to stepwise merge points according to their dominance distance. Algorithm 2 gives the details of clustering. A threshold σ is required in this clustering. We set σ to be the interval (in the sense of the dominance distance) of two neighboring

Algorithm 1 Performance Comparison Indicator (PCI)

Require: $P^1, P^2, \ldots, P^{pn}$ (tested approximation sets)
1: $S \leftarrow NondominanceSelection(P^1, P^2, \ldots, P^{pn})$ /* Finding out the nondominated solutions of the mixed set consisting of all the approximation sets */
2: $(C_1, C_2, ..., C_{cn}) \leftarrow Clustering(S)$ /* Clustering the points in S according to Algorithm 2 */
3: **for all** $P^i \in \{P^1, P^2, \ldots, P^{pn}\}$ **do**
4: PCI$(P^i) \leftarrow 0$
5: **for all** $C_j \in \{C_1, C_2, \ldots, C_{cn}\}$ **do**
6: $P^i_j \leftarrow P^i \cap C_j$
7: **if** $|P^i_j| < 2$ **then**
8: PCI$(P^i) \leftarrow$ PCI$(P^i) + \min\{D(p^i_1, C_j), ..., D(p^i_n, C_j)\}$ /* Finding out the minimum dominance distance of one solution in $P^i = \{p^i_1, ..., p^i_n\}$ to C_j */
9: **else**
10: PCI$(P^i) \leftarrow$ PCI$(P^i) + D'(P^i_j, C_j)$ /* Estimating the dominance distance of P^i_j to C_j */
11: **end if**
12: **end for**
13: PCI$(P^i) \leftarrow$ PCI$(P^i)/cn$
14: **end for**
15: **return** PCI(P^1), PCI$(P^2), \ldots,$ PCI(P^{pn})

Algorithm 2 $Clustering(S)$

Require: $S = \{s_1, s_2, \ldots, s_{sn}\}$, σ (threshold)
1: $(p_1, p_2, ..., p_{pn}) \leftarrow FindSortPair(S, \sigma)$ /* Find all pairs of solutions in S satisfying that $\max\{D(s_i, s_j), D(s_j, s_i)\} \leq \sigma$ where s_i and s_j are the two solution of a pair, and then sort the pairs in the ascending order according to $\max\{D(s_i, s_j), D(s_j, s_i)\}$ */
2: $C_1 \leftarrow s_1, C_2 \leftarrow s_2, ..., C_{sn} \leftarrow s_{sn}$ /* Cluster initialization */
3: **for all** $p \in \{p_1, p_2, \ldots, p_{pn}\}$ **do**
4: $(C_m, C_n) \leftarrow Locate(p)$ /* Let C_m and C_n be clusters which the two solutions of pair p fall into */
5: **if** $m \neq n$ **then**
6: **if** $\forall s_i \in C_m, s_j \in C_n : D(s_i, s_j) < \sigma \wedge D(s_j, s_i) < \sigma$ **then**
7: $C_m \leftarrow C_m \cup C_n$
8: $C_n \leftarrow \emptyset$ /* Merging two clusters if the dominance distance of any two solutions in them is not larger than σ */
9: **end if**
10: **end if**
11: **end for**
12: **return** $\{C_1, C_2, ..., C_{cn}\} : |C_i| > 0$

points in the normalized hyperplane with ideal distribution of N points, where N is the size of the reference set. In this case, $\sigma = 1/h$ and $N = \binom{m-1+h}{m-1}$, where h denotes the divisions along each objective and m is the number of objectives. Since $(h+m-1) \times (h+m-2) \times ... \times (h+1) \approx (h+m/2)^{m-1}$, σ can be approximately obtained by

$$\sigma \approx \frac{1}{\sqrt[m-1]{N(m-1)!} - (m/2)} \quad (3)$$

4. COMPARISON WITH POPULAR PERFORMANCE INDICATORS

Table 2 summarizes the properties of PCI and some popular performance indicators. As seen, PCI and the ϵ-indicator perform similarly in terms of many aspects, such as neither specifying a reference set nor requiring parameter setting in the evaluation. Only difference between them shown in the table is that PCI is capable of dealing with more than two

Table 2: Properties of some performance indicators

Indicators	Convergence	Diversity	Number of sets	Time complexity	Pareto compliant	Reference set specification	Parameter setting
Coverage [36]	√		binary	quadratic	weakly	no	no
DCI [21]		√	arbitrary	quadratic	no	no	yes
HV [36]	√	√	unary	exponential	yes	no	yes
ϵ-indicator [37]	√	√	unary/binary	quadratic	weakly/yes	no	no
IGD [2]	√	√	unary	quadratic	no	yes	no
IGD^+ [16]	√	√	unary	quadratic	weakly	yes	no
PCI	√	√	arbitrary	quadratic	weakly or yes*	no	no

*PCI is Pareto compliant when two approximation sets are considered, but weakly Pareto compliant when more sets are involved. The proof is omitted due to the limit of the space.

Table 3: Evaluation results of PCI and the peer indicator (HV, ϵ-indicator, or IGD^+) on the approximation set instances in Figures 1–3. The reference point 1.1 **is used in the HV calculation of Figure 1's instance. A better result is highlighted in boldface.**

Two sets	Peer indicator	PCI
Fig. 1(a) vs Fig. 1(b)	0.7448 vs **0.7460** (HV)	**0.0159** vs 0.0170
P vs Q in Fig. 2	2.5 vs 2.5 (ϵ-indicator)	**0.0000** vs 0.1204
P vs Q in Fig. 3	0.625 vs 0.625 (IGD^+)	**0.0648** vs 0.0926

approximation sets in a single run. In fact, one significant weakness of the ϵ-indicator to other performance indicators, as indicated before, is its return only being the difference on one particular objective of one particular solution for two approximation sets.

Like IGD^+, PCI only considers the inferiority of solutions in the evaluation. However, one difference is that PCI considers the comparison between a point to a point set (or two point sets), which could be viewed as a more general case of two-points comparison in IGD^+. In addition, since the reference set in PCI consists of the tested approximation sets, the evaluation result of one approximation set depends entirely on its performance difference with other sets. In contrast, in IGD^+ the reference set is specified, and the evaluation result of the approximation sets not only depends on their performance difference but also could be largely affected by the reference set (such as its distribution).

Finally, Table 3 gives the evaluation results of PCI and the peer indicators on the approximation sets in Figures 1–3. As shown, compared with HV, ϵ-indicator and IGD^+, the proposed indicator can accurately reflect the difference of solution sets in terms of uniformity, convergence and diversity, respectively.

5. EXPERIMENTAL STUDIES

In this section, we verify the proposed indicator by assessing Pareto front approximations obtained by six established multiobjective algorithms, NSGA-II [7], AR [4, 25], IBEA [35], MOEA/D-TCH [33], MOEA/D-PBI [33], and SPEA2+SDE [22]. These algorithms are representative in terms of both convergence and diversity in dealing with many-objective problems.

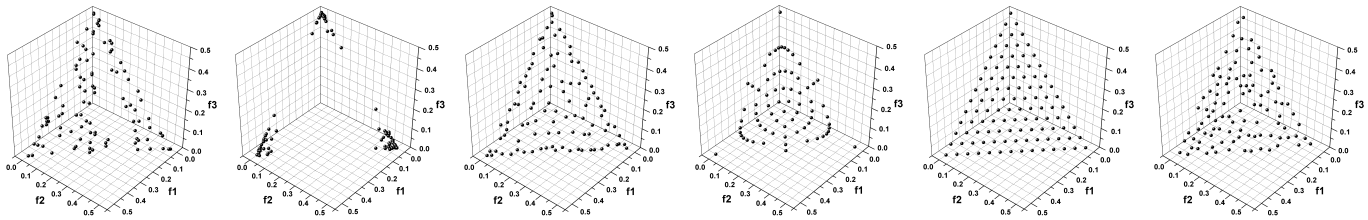
A crossover probability $p_c = 1.0$ and a mutation probability $p_m = 1/n$ (where n is the number of decision variables) are used. The crossover and mutation operators are simulated binary crossover (SBX) and polynomial mutation with both distribution indexes 20. The population size is set to 100 (or approximately 100 for MOEA/D due to its property [33]) and the termination criterion is $30,000$ evaluations.

To start with, we consider a tri-objective MOP, DTLZ1 [8], whose Pareto front is the positive part of a hyperplane satisfying $f_1 + f_2 + f_3 = 0.5$. Here, we ease the difficulty of the problem by using $g = \sum_{i=m}^{n}(x_i - 0.5)$ in DTLZ1 [8] in order to focus on the diversity verification of the indicator.

Figure 5 shows the approximation sets of the six algorithms as well as the corresponding PCI result. As can be seen from the figure, the solutions of MOEA/D-PBI are perfectly distributed over the whole Pareto front, thus having the best evaluation result. Although the solution sets of SPEA2+SDE, IBEA, and NSGA-II cover the triangle, their uniformity is different, which is consistent with the PCI results. The solution set of MOEA/D-TCH is of great regularity, but fails to cover the boundary region of the Pareto front, thus leading to a worse PCI than the above four sets. The solutions of AR only concentrate around three extreme points and obtain the worst PCI value.

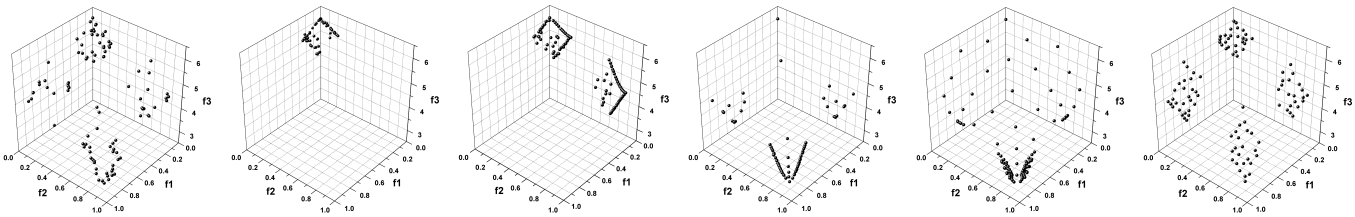
To examine the proposed indicator on MOPs with an irregular Pareto front, we introduce the test problem DTLZ7 [8]. The tri-objective DTLZ7 has a disconnect Pareto front consisting of four regions with both convex and concave shapes. Figure 6 gives the approximation sets and the evaluation results. As shown, the set of SPEA2+SDE and NSGA-II is located in the four optimal regions, with a nearly equal number of solutions. The solutions of the two MOEA/D algorithms are mainly distributed in the bottom region of the Pareto front, and IBEA and AR fail to find all the four regions. It is clear that the PCI results confirm the observations. The solution set with a lower PCI value means that it performs better regarding the combined performance in finding multiple Pareto optimal regions and in maintaining solutions' uniformity and diversity in each region.

Next, we consider a four-objective MOP, Rectangle problem [23]. The Rectangle problem has two interesting features: 1) its Pareto optimal solutions lie in a rectangle in the two-variable decision space and 2) they are similar (in the sense of Euclidean geometry) to their images in the four-dimensional objective space. These make the visual examination of Pareto front approximations feasible by observing their behavior in the decision space. Figure 7 shows the approximation sets and the evaluation results on an instance of the Rectangle problem where the search range and Pareto optimal range of x_1 and x_2 are $[-0.2, 1.2]$ and $[0, 1]$, respectively. As seen, the performance of these approximation sets is consistent with the PCI results. SPEA2+SDE and IBEA have a set of well-converged and well-distributed solutions, thus having better PCI results than the other four algorithms. IBEA obtains a slightly worse PCI value than SPEA2+SDE, with more solutions being located onto the boundary of the optimal region. MOEA/D-TCH, NSGA-II and MOEA/D-PBI perform similarly in terms of con-



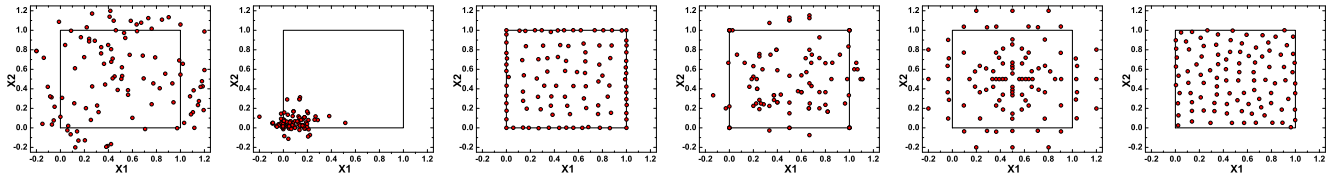
(a) NSGA-II PCI = 0.0417 (b) AR PCI = 0.1619 (c) IBEA PCI = 0.0305 (d) MOEA/D-TCH PCI = 0.0434 (e) MOEA/D-PBI PCI = 0.0270 (f) SPEA2+SDE PCI = 0.0291

Figure 5: Approximation sets of the six algorithms and their PCI result on the tri-objective DTLZ1.



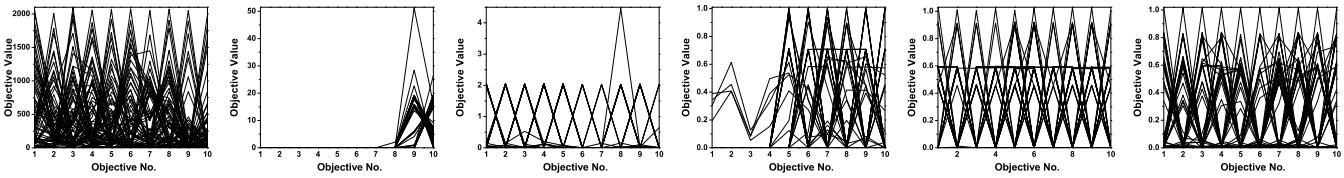
(a) NSGA-II PCI = 0.0346 (b) AR PCI = 0.2564 (c) IBEA PCI = 0.1045 (d) MOEA/D-TCH PCI = 0.0638 (e) MOEA/D-PBI PCI = 0.0483 (f) SPEA2+SDE PCI = 0.0163

Figure 6: Approximation sets of the six algorithms and their PCI result on the tri-objective DTLZ7.



(a) NSGA-II PCI = 0.0692 (b) AR PCI = 0.4147 (c) IBEA PCI = 0.0506 (d) MOEA/D-TCH PCI = 0.0666 (e) MOEA/D-PBI PCI = 0.0798 (f) SPEA2+SDE PCI = 0.0498

Figure 7: Approximation sets of the six algorithms in the two-variable decision space and their PCI result on the four-objective Rectangle problem, where the Pareto optimal solutions in the decision space are similar to their images in the objective space in the sense of Euclidean geometry.



(a) NSGA-II PCI = 236.07 (b) AR PCI = 11.820 (c) IBEA PCI = 1.0058 (d) MOEA/D-TCH PCI = 0.2218 (e) MOEA/D-PBI PCI = 0.0936 (f) SPEA2+SDE PCI = 0.0845

Figure 8: Parallel coordinate plot of approximation sets of the six algorithms and their PCI result on the ten-objective DTLZ3.

vergence and diversity. Among them, MOEA/D-TCH has fewer solutions out of the optimal rectangle and MOEA/D-PBI has many solutions concentrated into the center of the rectangle. This leads to the difference of their PCI results. All solutions of AR are located around the lower left corner of the rectangle and thus have the worst evaluation result.

Finally, the 10-objective DTLZ3 is used to verify PCI in assessing approximation sets for high-dimensional challenging MOPs. DTLZ3 has a vast number of local optimal fronts $(3^{10}-1)$ and a global one satisfying $f_1^2+f_2^2+...+f_m^2=1$ in the range $f_1, f_2, ..., f_m \in [0,1]$. Figure 8 shows the approximation sets by parallel coordinates as well as the corresponding evaluation results. As shown, the solutions of only three algorithms SPEA2+SDE, MOEA/D-PBI, and MOEA/D-TCH can converge into the optimal front. SPEA2+SDE and MOEA/D-PBI achieve a good balance between convergence and diversity, while MOEA/D-TCH fails to cover the whole Pareto front. Almost all solutions of IBEA converge into the boundary of a local optimum ($f_1^2+...+f_{10}^2=2$), and most of AR's solutions concentrate around two boundary points of a local optimum ($f_1^2+...+f_{10}^2=20$). NSGA-II completely fails to approach the optimal front, with the upper boundary of its solutions exceeding 2000 on each objective. The PCI results clearly reflect the performance of the approximation

sets. A set with a lower DCI value means that it performs better with respect to the tradeoff between convergence and diversity.

6. CONCLUSIONS

This paper presents a performance comparison indicator (PCI) to assess Pareto front approximations of population-based search algorithms. PCI evaluates the relative quality of approximation sets to a reference set which is constructed by themselves. The points in the reference set are clustered according to their distribution, and PCI estimates the minimum moves of one solution (or a set of solutions) in one approximation set to weakly dominate these clusters. The proposed indicator can be practical, especially in many-objective optimization, given its combined evaluation of convergence and diversity, (weak) compliance with Pareto dominance, applicability for any number of approximation sets, no need of a specified reference set, and no requirement of parameter setting in the assessment.

7. REFERENCES

[1] J. Bader and E. Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evol. Comput.*, 19(1):45–76, 2011.

[2] P. A. N. Bosman and D. Thierens. The balance between proximity and diversity in multiobjective evolutionary algorithms. *IEEE Tran. Evol. Comput.*, 7(2):174–188, 2003.

[3] K. Bringmann, T. Friedrich, C. Igel, and T. Voß. Speeding up many-objective optimization by Monte Carlo approximations. *Artificial Intelligence*, 204:22–29, 2013.

[4] D. W. Corne and J. D. Knowles. Techniques for highly multiobjective optimisation: some nondominated points are better than others. In *Proc. 9th Genetic Evol. Comput. Conf. (GECCO)*, pages 773–780, 2007.

[5] K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Tran. Evol. Comput.*, 18(4):577–601, 2014.

[6] K. Deb and S. Jain. Running performance metrics for evolutionary multi-objective optimization. Technical Report 2002004, KanGAL, Indian Institute of Technology, 2002.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Tran. Evol. Comput.*, 6(2):182–197, 2002.

[8] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable test problems for evolutionary multiobjective optimization. In A. Abraham, L. Jain, and R. Goldberg, editors, *Evolutionary Multiobjective Optimization. Theoretical Advances and Applications*, pages 105–145. Berlin, Germany: Springer, 2005.

[9] P. Fleming, R. Purshouse, and R. Lygoe. Many-objective optimization: An engineering design perspective. In *Proc. 3rd Int. Conf. Evol. Multi-Criterion Optimiz. (EMO)*, pages 14–32. 2005.

[10] D. Hadka and P. Reed. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. *Evol. Comput.*, 20(3):423–452, 2012.

[11] Z. He, G. G. Yen, and J. Zhang. Fuzzy-based Pareto optimality for many-objective evolutionary algorithms. *IEEE Tran. Evol. Comput.*, 18(2):269–285, 2014.

[12] E. J. Hughes. MSOPS-II: A general-purpose many-objective optimiser. In *Proc. IEEE Congr. Evol. Comput. (CEC)*, pages 3944–3951, 2007.

[13] K. Ikeda, H. Kita, and S. Kobayashi. Failure of Pareto-based MOEAs: does non-dominated really mean near to optimal? In *Proce. IEEE Congr. Evol. Comput. (CEC)*, volume 2, pages 957–962, 2001.

[14] H. Ishibuchi, N. Akedo, and Y. Nojima. Behavior of multi-objective evolutionary algorithms on many-objective knapsack problems. *IEEE Tran. Evol. Comput.*, 19(2):264–283, 2015.

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

[16] H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima. Modified distance calculation in generational distance and inverted generational distance. In *Evol. Multi-Criterion Optimiz.*, pages 110–125, 2015.

[17] H. Ishibuchi, N. Tsukamoto, and Y. Nojima. Evolutionary many-objective optimization: A short review. In *Proce IEEE Congr Evol. Comput. (CEC)*, pages 2419–2426, 2008.

[18] A. L. Jaimes and C. A. Coello Coello. Study of preference relations in many-objective optimization. In *Proce. 11th Genetic Evol. Comput. Conf. (GECCO)*, pages 611–618, 2009.

[19] J. Knowles and D. Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Tran. Evol. Comput.*, 7(2):100–116, 2003.

[20] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.*, 10(3):263–282, Sept. 2002.

[21] M. Li, S. Yang, and X. Liu. Diversity comparison of Pareto front approximations in many-objective optimization. *IEEE Trans. Cybern.*, 44(12):2568–2584, 2014.

[22] M. Li, S. Yang, and X. Liu. Shift-based density estimation for Pareto-based algorithms in many-objective optimization. *IEEE Tran. Evol. Comput.*, 18(3):348–365, 2014.

[23] M. Li, S. Yang, and X. Liu. A test problem for visual investigation of high-dimensional multi-objective search. In *Proc IEEE Congr Evol. Comput. (CEC)*, pages 2140–2147, 2014.

[24] M. Li, S. Yang, X. Liu, and R. Shen. A comparative study on evolutionary algorithms for many-objective optimization. In *Proc. 7th Int. Conf. Evol. Multi-Criterion Optimiz. (EMO)*, pages 261–275, 2013.

[25] M. Li, J. Zheng, K. Li, Q. Yuan, and R. Shen. Enhancing diversity for average ranking method in evolutionary many-objective optimization. In *Proc. Parallel Problem Solving From Nature (PPSN XI)*, pages 647–656, 2010.

[26] G. Lizárraga-Lizárraga, A. Hernández-Aguirre, and S. Botello-Rionda. G-Metric: an M-ary quality indicator for the evaluation of non-dominated sets. In *Proc. 10th Genetic Evol. Comput.*, pages 665–672, 2008.

[27] R. C. Purshouse and P. J. Fleming. On the evolutionary optimization of many conflicting objectives. *IEEE Tran. Evol. Comput.*, 11(6):770–784, 2007.

[28] O. Schütze, X. Esquivel, A. Lara, and C. A. C. Coello. Using the averaged Hausdorff distance as a performance measure in evolutionary multi-objective optimization. *IEEE Tran. Evol. Comput.*, 16(4):504–522, 2012.

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

[30] T. Wagner, N. Beume, and B. Naujoks. Pareto-, aggregation-, and indicator-based methods in many-objective optimization. In *Proc. 4th Int. Conf. Evol. Multi-Criterion Optimiz. (EMO)*, pages 742–756. 2007.

[31] R. Wang, R. C. Purshouse, and P. J. Fleming. Preference-inspired co-evolutionary algorithms for many-objective optimisation. *IEEE Tran. Evol. Comput.*, 17(4):474–494, 2013.

[32] S. Yang, M. Li, X. Liu, and J. Zheng. A grid-based evolutionary algorithm for many-objective optimization. *IEEE Tran. Evol. Comput.*, 17(5):721–736, 2013.

[33] Q. Zhang and H. Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Tran. Evol. Comput.*, 11(6):712–731, 2007.

[34] X. Zhang, Y Tian, and Y. Jin. A knee point driven evolutionary algorithm for many-objective optimization. *IEEE Tran. Evol. Comput.*, 2014, in press.

[35] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Proc. Int. Conf. Parallel Problem Solving from Nature (PPSN)*, pages 832–842. 2004.

[36] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Tran. Evol. Comput.*, 3(4):257–271, 1999.

[37] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Tran. Evol. Comput.*, 7(2):117–132, 2003.