A kernel-based indicator for multi/many-objective optimization

Cai, Xinye; Xiao, Yushun; Li, Zhenhua; Sun, Qi; Xu, Hanchuan; Li, Miqing; Ishibuchi, Hisao

DOI:
10.1109/TEVC.2021.3105565

A Kernel-Based Indicator for Multi/Many-Objective Optimization

Xinye Cai, *Member, IEEE,* Yushun Xiao, Zhenhua Li, *Member, IEEE,* Qi Sun, Hanchuan Xu, Miqing Li, *Member, IEEE,* Hisao Ishibuchi *Fellow, IEEE*

***Abstract*—How to evaluate Pareto front approximations generated by multi/many-objective optimizers is a critical issue in the field of multiobjective optimization. Currently, there exist two types of comprehensive quality indicators (i.e., volume-based and distance-based indicators). Distance-based indicators, such as Inverted Generational Distance (IGD), are usually computed by summing up the distance of each reference point to its nearest solution. Their high computational efficiency leads to their prevalence in many-objective optimization. However, in the existing distance-based indicators, the distributions of the solution sets are usually neglected, leading to their lacks of ability to well distinguish between different solution sets. This phenomenon may become even more severe in high-dimensional space. To address such an issue, a kernel-based indicator (KBI) is proposed as a comprehensive indicator. Different from other distance-based indicators, a kernel-based maximum mean discrepancy is adopted in KBI for directly measuring the difference that can characterize the convergence, spread and uniformity of two sets, i.e., the solution set and reference set, by embedding them in Reproducing Kernel Hilbert Space (RKHS). As a result, KBI not only reflects the distance between the solution set and the reference set, but also can reflect the distribution of the solution set itself. In addition, to maintain the desirable weak Pareto compliance property of KBI, a nondominated set reconstruction approach is also proposed to shift the original solution set. The detailed theoretical and experimental analysis of KBI is provided in this paper. The properties of KBI have also been analyzed by the optimal μ-distribution.**

***Index Terms*—Multiobjective optimization; many-objective optimization, quality evaluation, kernel method.**

I. Introduction

In real-world optimization problems, multiple conflicting objectives usually need to be optimized simultaneously. Different from a single-objective optimization problem, the goal of a multiobjective optimization problem (MOP) is to obtain a set of Pareto optimal solutions, which represents the trade-offs between objectives. These Pareto optimal solutions are usually called a *Pareto Set* (PS); and its projection in the objective space is usually called a *Pareto Front* (PF) [1]. A good PF approximation apparently can provide useful information for the decision makers for choosing their preferred solution.

For MOPs, how to evaluate the quality of the high-dimensional PF approximations is a critical issue [2]. Most quality indicators [3], [4], [5] are designed for evaluating approximations in some or all of the following aspects: 1) convergence [6], 2) spread [7] (i.e., coverage [8] or extensity [9]) and 3) uniformity [10]. The latter two are combined and usually called diversity of a solution set [3], [9]. In addition, the Pareto compliance is a desirable property for multi/many-objective indicators. The lack of such a property in indicators may result in unfair evaluation results [11].

Over several decades, a considerable number of comprehensive indicators that measure both convergence and diversity of a PF approximation have been proposed in the field of evolutionary multiobjective optimization [12], [13], [14], [15]. In general, these comprehensive indicators can be classified into two types: volume-based indicators and distance-based indicators [11]. The former type of indicators measures the size of the volume determined by PF approximation under consideration with some specification. Hypervolume (HV) [16] is a prevalent representative of volume-based indicators. However, high computational cost of calculating HV makes it difficult to use for high-dimensional PFs [17]. The proper choice of the reference point is also a tedious task, which will largely affect the ability of HV to correctly distinguish the quality of approximations [18].

The latter type measures the distance of the reference PF to the considered solution set. Thus distance-based indicators (such as Inverted Generational Distance (IGD) [19], IGD$^+$ [20] and Δ_p [21]) usually can be calculated very efficiently, leading to their prevalence in many-objective optimization. However, they somehow suffer from shortcomings that may severely limit their further applications [12].

1) The existing distance-based indicators may fail to distinguish between approximations as they are computed by summing up the distances between each reference point with its nearest solution, without the consideration of the distribution of the solution set itself. Fig. 1 shows such an example in a two-dimensional space, where IGD/IGD$^+$/Δ_p values of solution sets A and B are the

Xinye Cai, Yushun Xiao, Zhenhua Li and Qi Sun are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 210016, China, and also with the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210023, China (e-mail: xinye@nuaa.edu.cn, Shaw_yushun@nuaa.edu.cn, zhenhua.li@nuaa.edu.cn, sunshinesq@163.com).

Hanchuan Xu is with the Department of Computing, Harbin Institute of Technology.

Miqing Li is with Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA) in the School of Computer Science at the University of Birmingham.

Hisao Ishibuchi is with Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China.

This work was supported in part by the National Natural Science Foundation of China (NSFC) under grant 61732006, 62072234 and 61473241, by the National Key R&D Program of China under grant 2018YFB1402900, by the Natural Science Foundation of Jiangsu Province of China under grant BK20181288 and by China Postdoctoral Science Foundation under grant 2015M571751.

The code of KBI is available in: https://github.com/xinyecai/KBI

same, as each of the three reference points has the same distance to multiple solutions. However, we can visualize that the diversity of A is better than that of B (their convergence is the same). Such a phenomenon becomes even more severe in high-dimensional space [11].

2) Most attention has been put in designing new indicators or indicator-based algorithms in the literature. However, the validation of the indicator properties is often overlooked in the existing studies.

In this paper, we propose a new distance-based comprehensive indicator that addresses the aforementioned issues. The major contributions of this work can be summarized as follows:

1) Maximum Mean Discrepancy (MMD), which originally measures the difference of two distributions [22], can also be applied as a measure of the difference between two sets [23]. Based on it, a kernel-based indicator (KBI) is proposed as a comprehensive quality indicator. Different from IGD (or its variants) that sums up the distance between each reference point with its nearest solution, KBI computes the upper bounded difference that can characterize convergence, spread and uniformity between the solution set and the reference set by embedding them in Reproducing Kernel Hilbert Space (RKHS). As a result, KBI not only reflects the distance between the solution set and the reference set, but also can reflect the distribution of the solution set itself, thus being able to well-distinguish between different solution sets, as shown in Fig. 1.
2) The optimal μ-distribution [24] is introduced for validating the preferences and the properties of KBI. The experimental results on the optimal μ-distributions A_I of KBI has shown its effectiveness and uniqueness compared with other comprehensive indicators (see Section IV. E).

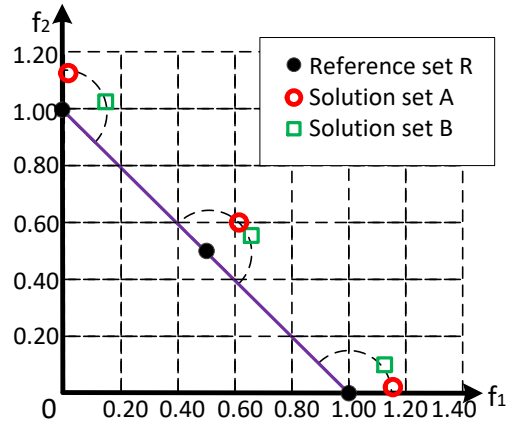


Fig. 1: An example of approximations in terms of IGD/IGD$^+$/Δ_p and KBI. Solution set A can be visualized better than B in terms of diversity. IGD/IGD$^+$/Δ_p(A) = IGD/IGD$^+$/Δ_p(B) = 0.1, KBI(A) = 0.041 $<$ KBI(B) = 0.0466.

The rest of this paper is organized as follows. Related studies on the proposed indicator are introduced in Section II. Section III elaborates the proposed kernel-based indicator (KBI). In Section IV, the systematic experiments are conducted to verify the effectiveness of KBI. The optimal μ-distributions of KBI, compared with other comprehensive indicators, are also investigated in this section. Finally this paper is concluded in Section V where some future research topics are suggested.

II. Background

A. Basic Definitions

A multiobjective optimization problem (MOP) can be defined as follows:

$$\begin{aligned} \min \quad & F(x) = (f_1(x), \ldots, f_m(x)) \\ \text{s.t.} \quad & x \in \Omega \end{aligned} \tag{1}$$

where Ω is the *decision space*, $F : \Omega \to R^m$ consists of m real-valued objective functions. Note that an MOP is usually called a many-objective optimization problem (MaOPs) when $m > 3$.

Let $a, b \in R^m$, a is said to *dominate* b, denoted by $a \prec b$, if and only if $a_i \leq b_i$ for every $i \in \{1, \ldots, m\}$ and $a_j < b_j$ for at least one index $j \in \{1, \ldots, m\}$; a is said to *weakly dominate* b, denoted by $a \preceq b$, if and only if $a_i \leq b_i$ for every $i \in \{1, \ldots, m\}$[1] [11]. Given a set S in R^m, a solution in S is called non-dominated in S if no other solution in S dominates it. A solution $x^* \in \Omega$ is *Pareto-optimal* if $F(x^*)$ is non-dominated in the attainable objective set. $F(x^*)$ is then called a *Pareto-optimal (objective) vector*. In other words, any improvement in one objective of a Pareto optimal solution must lead to deterioration in at least another objective. The set of all the Pareto-optimal points is called the *Pareto set* (*PS*) and the set of all the Pareto-optimal objective vectors is called the *Pareto front* (*PF*) [1].

The dominance relationship can be extended to solution sets [25], as follows. Set A is said to *dominate* set B (denoted as $A \prec B$) if every solution $b \in B$ is dominated by at least one solution $a \in A$, i.e., $\forall b \in B, \exists a \in A, a \prec b$. Similarly, set A is said to *weakly dominate* B (denoted as $A \preceq B$) if every solution $b \in B$ is *weakly dominated* by at least one solution $a \in A$, i.e., $\forall b \in B, \exists a \in A, a \preceq b$.

A quality indicator is said to be *strictly Pareto compliant* [25], [11], if and only if $\forall A, B : A \preceq B \wedge B \npreceq A \implies I(A) < I(B)$; where $I(.)$ is a mapping from an objective vector to an indicator value. Similarly, a quality indicator is said to be *weakly Pareto compliant* [25], [11], if and only if $\forall A, B : A \preceq B \implies I(A) \leq I(B)$.

B. Related studies on quality indicators

As two representative volume-based indicators, Hypervolume (HV) [16] and R2 [26] have been widely used in the field of multiobjective optimization. Their definitions are given as follows.

- Hypervolume (HV) [16]: Let $r^* = (r_1^*, r_2^*, ..., r_m^*)^T$ be a reference point in the objective space that is dominated by all solutions in a PF approximation set S. HV measures

[1] In the case of maximization, the inequality signs should be reversed.

the volume of the objective space dominated by the solutions in S and bounded by r^*.

$$HV(S) = VOL(\bigcup_{y \in S}[y_1, r_1^*] \times ...[y_m, r_m^*]) \quad (2)$$

where $VOL(\bullet)$ indicates the Lebesgue measure. The value of HV can be computed as the union volume of the hypercubes determined by the solution set S and the reference point r^*. The larger the HV value is, the better the quality of a solution set S is. The main advantage of HV is its strictly Pareto compliance property [27]. However, it is difficult to compute the exact value of HV for a large solution set with many objectives, although some fast computational methods have been proposed for approximating HV [28], [29], [30], [31], [32]. For example, Sharpe-Ratio [33], [34] is such an indicator with interesting properties and not as time consuming as hypervolume indicator. Another weakness of HV is its favor of very non-uniform solution sets on a highly nonlinear Pareto front no matter what the reference point is, possibly leading to unfair comparison results [35]. It is worth to mention that some ideas on using kernels for approximating Hypervolume already exist in the literature [36], [37]. Their weakness is also similar to that of Hypervolume.

- R2 [26]: The indicator of R2, integrated the decision maker's preference, was proposed to assess the quality of two individual sets separately. Having the standard weighted Tchebycheff function with a particular reference point z^*, the indicator can be used to assess the quality of a single individual set against z^*[38]. Given an approximation set S, a set of weight vectors W, and the standard Tchebycheff aggregation function, the R2 indicator can be defined as follow.

$$R2(S, W, z^*) = \frac{1}{|W|}\sum_{w \in W}\min_{y \in S}\{\max\{w_i(y_i - z_i^*)\} \quad (3)$$

where $i \in \{1, \ldots, m\}$, $w = (w_1, \ldots, w_m)$ is a weight vector; z^* is the ideal point. A smaller R2 value indicates that a solution set S is closer to the reference point. The R2 indicator possesses a desirable weak Pareto compliance property. However, the R2 indicator, using a set of weight vectors W, naturally has a low accuracy, which may result in its lack of ability to distinguish different solution sets.

Beside the above volume-based indicators, the following distance-based quality indicators have also been prevalently used in multi/many-objective optimization community.

- Inverted Generational Distance (IGD) [19]: Let S be the set of solutions obtained by an multiobjective optimizer, and P^* be the set of points uniformly sampled over the true PF. The value of IGD is calculated as:

$$IGD(S, P^*) = \frac{\sum_{y \in P^*} dist(y, S)}{|P^*|} \quad (4)$$

where $dist(y, S)$ denotes the Euclidean distance between a point $y \in P^*$ and its nearest neighbor in S, and $|P^*|$ is the cardinality of P^*. IGD calculates an average minimum distance from each point in P^* to the nearest solution in S, which measures both convergence and diversity of a solution set S. The smaller the IGD value is, the better the quality of S is. The drawbacks of IGD include its lack of Pareto compliance property and the sensitivity to the dominated solutions, both of which may lead to biased comparison results. In addition, the use of Euclidean distance may result in similar evaluation results of different solution sets in a high-dimensional space [11].
- Modified Inverted Generational Distance (IGD$^+$) [20]: Let S be a set of solutions obtained by an optimizer, and P^* be a set of points uniformly sampled over the true PF. The only difference between IGD(S, P^*) and IGD$^+$(S, P^*) lies in the distance calculation. In the minimization problems, the distance calculation for IGD$^+$ is

$$dist^+(y, S) = \sqrt{\sum_{i=1}^{m}(\max\{y_i - z_i, 0\})^2} \quad (5)$$

where $dist^+(y, S)$ is the modified distance between a point $y \in P^*$ and its nearest neighbor $z \in S$. This distance modification ensures that IGD$^+$ is weakly Pareto compliant whereas the original IGD is Pareto non-compliant. Like IGD, the smaller the IGD$^+$ value is, the better the quality of S is. Similar to IGD, One drawback of IGD$^+$ is its lack of ability to distinguish different solution sets, as shown in Fig. 1.
- $I_{\epsilon+}$ [39]: Let S be the set of solutions obtained by a multiobjective optimizer, and P^* be the set of reference points uniformly sampled over the true PF. The value $I_{\epsilon+}$ of two points $y \in P^*$ and $s \in S$ is given as follows.

$$I_{\epsilon+}^{vec}(s, y) = \max_{i \in \{1,..m\}}\{s_i - y_i\} \quad (6)$$

where the $I_{\epsilon+}^{vec}(s, y)$ value is the minimal shift such that s weakly dominates y. It should be noted that $I_{\epsilon+}^{vec}(s, y) \neq I_{\epsilon+}^{vec}(y, s)$ in Eq.(6). The $I_{\epsilon+}$ indicator can be computed as follows.

$$I_{\epsilon+}(S) = \max_{y \in P^*}\{\min_{s \in S}\{I_{\epsilon+}^{vec}(s, y)\}\} \quad (7)$$

where the $I_{\epsilon+}$ value of S is the minimal shift such that each reference point $y \in P^*$ is weakly dominated by at least one solution $s \in S$. The smaller the $I_{\epsilon+}$ value is, the better quality of S is. However, the $I_{\epsilon+}^{vec}(s, y)$ value depends on one particular objective of one particular solution in each set, the indicator may ignore a significant amount of sets' difference. It was demonstrated in the computational experiments in [40] that two different solution sets often have the same value of $I_{\epsilon+}$.
- Δ_p [21]: "Averaged Hausdorff distance" is adopted to calculate the distance between the obtained solution set S and the reference set R, which evaluates both conver-

gence and diversity as follows [41], [42].

$$\begin{aligned}\Delta_p(S,R) &= \max(GD_p(S,R), IGD_p(S,R)) \\ &= \max((\frac{1}{|S|}\sum_{s\in S} dist(s,R)^p)^{\frac{1}{p}}, \\ &(\frac{1}{|R|}\sum_{r\in R} dist(r,S)^p)^{\frac{1}{p}})\end{aligned} \tag{8}$$

where $dist(s,R)$ is the Euclidean distance between a solution $s \in S$ and its nearest point in reference set R, and $|S|$ is the cardinality of S, $dist(r,S)$ is the Euclidean distance between a solution $r \in R$ and its nearest point in solution set S, and $|R|$ is the cardinality of R. The smaller the Δ_p value is, the better the quality of S for approximating the whole PF is. Like IGD/IGD$^+$, it also lacks of ability to distinguish different solution sets, as shown in Fig. 1.

C. Distance metrics between two sets

In the literature, there exist a number of metrics for measuring the difference between two sets. The most popular ones include Kullback-Leibler (KL) divergence [43], maximum mean discrepancy (MMD) [23], Wasserstein distance [44], and Hausdorff distance [45]. Among them, KL divergence technically is not a distance metric, as it neither is symmetric nor satisfies triangle inequality [46]. Wasserstein distance is too computational expensive to use for designing an indicator. Although the variant of Hausdorff distance [47] is able to satisfy all axioms of a distance metric, one clear disadvantage of Hausdorff distance and its variants are their negligence of the distributions of the solution sets, which leads to the fact that they may not be able to well distinguish between different solution sets in some cases, as explained in Fig. 1. In this paper, MMD is adopted to design a comprehensive indicator for multiobjective/many-objective optimization.

III. The proposed indicator

In this section, the intuitions of why maximum mean discrepancy works for measuring the convergence and diversity of a solution set are explained, followed by its computation with the kernel function. After that, the algorithm of using MMD for computing KBI is specified in detail. The computational complexity and weak Pareto compliance of KBI are also given in this section. Finally, KBI is compared with other indicators in terms of desirable properties.

A. Why and how MMD works?

Maximum Mean Discrepancy (MMD), originally used for measuring the difference of two distributions [22], can also be applied to measuring the difference between two sets [23].

Let $S = \{s_1, ..., s_u\}$ be a solution set and $R = \{r_1, ..., r_v\}$ be a reference set. *Maximum Mean Discrepancy (MMD)* [23], [48] can be used for computing the difference of these two sets by embedding them in *Reproducing Kernel Hilbert Space* (*RKHS*). Let $\mathcal{G}$ be a class of functions $g : \mathcal{Y} \to \mathbb{R}$. MMD between R and S can be estimated as follows.

$$\begin{aligned}\mathrm{MMD}(\mathcal{G},R,S) &:= \sup_{g\in\mathcal{G}}(\mathrm{E}_R[g(r)] - \mathrm{E}_S[g(s)]) \\ &= \sup_{g\in\mathcal{G}}(\frac{1}{u}\sum_{i=1}^{u} g(r_i) - \frac{1}{v}\sum_{j=1}^{v} g(s_j)).\end{aligned} \tag{9}$$

where sup (*supremum*) denotes the maximum difference between the two sets. When g is a first-order function, the above equation actually compares the means of the two sets, i.e., $\mathrm{E}_R[r]$ and $\mathrm{E}_S[s]$, which can be interpreted as the convergence of the solution set S to the reference set R (case 1, as shown in Fig. 2a) or the spread between S and R (case 2, as shown in Fig. 2b). When $\mathrm{E}_R[r] = \mathrm{E}_S[s]$ and g is a second-order function, the above equation compares the variance of the two sets, i.e., $\mathrm{E}_R[r^2]$ and $\mathrm{E}_S[s^2]$, which can be interpreted as the comparison of the spread between S and R (case 3, as shown in Fig. 2c). When $\mathrm{E}_R[r] = \mathrm{E}_S[s]$ and g is a third-order function, the above equation compares the skewness of the two sets, i.e., $\mathrm{E}_R[r^3]$ and $\mathrm{E}_S[s^3]$, which, to some extent, can be interpreted as the comparison between the uniformity for S and R (case 4, as shown in Fig. 2d). This indicates that MMD can be used to design a comprehensive indicator that characterizes both convergence and diversity of a solution set S by comparing it with the reference set R.

Furthermore, the kernel function can be added into Eq. 9 for practical computation as follows. Let $\mathcal{H}$ be a complete inner product space (i.e., a Hilbert Space) of functions $g : \mathcal{Y} \to \mathbb{R}$, where $\mathcal{Y}$ is a nonempty compact set. $\mathcal{H}$ is called a ***RKHS*** if for all $y \in \mathcal{Y}$, the linear point evaluation functional mapping $g \to g(y)$ exists and is continuous. In this case, $g(y)$ can be expressed as an *inner product*:

$$g(y) = \langle \phi(y), g\rangle_{\mathcal{H}} \tag{10}$$

where $\phi : \mathcal{Y} \to \mathcal{H}$ is known as the *feature space map* from y to $\mathcal{H}$. The inner product between two feature maps is called the *kernel*, $k(y,y') := \langle \phi(y), \phi(y')\rangle_{\mathcal{H}}$.

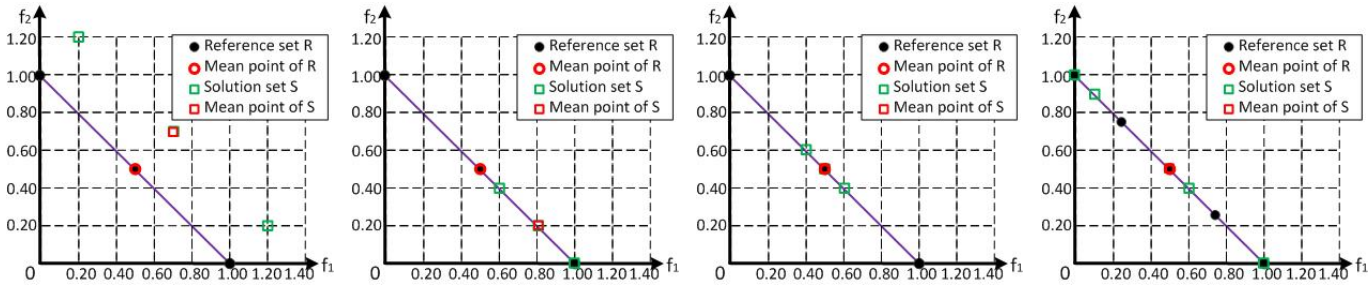
Moreover, $\mu_r := \mathrm{E}_r[\phi(y)]$ is the mean of $\phi(y)$ in feature space (assume that it exists) [49]. Then, MMD can be written as:

$$\mathrm{MMD}(\mathcal{G},R,S) = ||\mu_R - \mu_S||_{\mathcal{H}} \tag{11}$$

From the Eq. 9- 11, MMD can be reformulated as:

$$\begin{aligned}\mathrm{MMD}(\mathcal{G},R,S) &= \sup_{||g||_{\mathcal{H}}\le 1} \mathrm{E}_R[g(r)] - \mathrm{E}_S[g(s)] \\ &= \sup_{||g||_{\mathcal{H}}\le 1} \mathrm{E}_R[\langle\phi(r),g\rangle_{\mathcal{H}}] - \mathrm{E}_S[\langle\phi(s),g\rangle_{\mathcal{H}}] \\ &= \sup_{||g||_{\mathcal{H}}\le 1} \langle\mu_R - \mu_S, g\rangle_{\mathcal{H}} \\ &= ||\mu_R - \mu_S||_{\mathcal{H}} \\ &= [\langle\mu_R-\mu_S, \mu_R-\mu_S\rangle_{\mathcal{H}}]^{\frac{1}{2}} \\ &= [\langle\mu_R,\mu_R\rangle_{\mathcal{H}} + \langle\mu_S,\mu_S\rangle_{\mathcal{H}} - 2\langle\mu_R,\mu_S\rangle_{\mathcal{H}}]^{\frac{1}{2}}\end{aligned} \tag{12}$$

where sup (*supremum*) denotes the maximum difference between the two sets. $||.||$ is the *norm* for computing the distance between the solution set and the true PF in *feature space*. As a



(a) Case 1: When g is a first-order function, MMD can be interpreted as the convergence of the solution set S to the reference set R.

(b) Case 2: When g is a first-order function, MMD can also be interpreted as the spread of the solution set S, compared with the reference set R.

(c) Case 3: $\mathrm{E}_R[r] = \mathrm{E}_S[s]$ and g is a second-order function; MMD can be interpreted as the comparison of the spread between S and R.

(d) Case 4: $\mathrm{E}_R[r] = \mathrm{E}_S[s]$ and g is a third-order function; MMD can be interpreted as the comparison between the uniformity for S and R.

Fig. 2: The intuitions of why MMD can characterize both convergence and diversity of a solution set S by comparing it with the reference set R.

result, the proposed KBI is able to evaluate the comprehensive quality of a solution set in terms of convergence, spread and uniformity.

As $\mu_R := \frac{1}{v}\sum_{i=1}^{v}\phi(r_i)$, $\mu_S := \frac{1}{u}\sum_{i=1}^{u}\phi(s_i)$ and $k(y,y') := \langle\phi(y),\phi(y')\rangle_{\mathcal{H}}$ [23], [48], [49], an empirical estimate of MMD can be rewritten as follows.

$$\begin{aligned}\mathrm{MMD}(\mathcal{G},R,S) =&[\frac{1}{v^2}K(R,R') + \frac{1}{u^2}K(S,S') \\ &- \frac{2}{uv}K(R,S)]^{\frac{1}{2}} \\ =&[\frac{1}{u^2}\sum_{i\neq j}^{v}k(r_i,r_j) + \frac{1}{v^2}\sum_{i\neq j}^{u}k(s_i,s_j) \\ &- \frac{2}{uv}\sum_{i,j=1}^{u,v}k(r_i,s_j)]^{\frac{1}{2}}.\end{aligned} \tag{13}$$

where $K(R,S)$ is the summation of all $k(r,s)$.

When $|S| = |R|$, the above Eq. 13 can be rewritten as follows.

$$\begin{aligned}MMD(R,S) =&(\frac{1}{|R|^2}K(R,R') + \frac{1}{|R|^2}K(S,S') \\ &- \frac{2}{|R|^2}K(R,S))^{\frac{1}{2}}\end{aligned} \tag{14}$$

It can be obtained from Eq. 13 and Eq. 14 $MMD(R,S) = 0$, when $R = S$; and MMD has a positive value. It can also be observed that the distance between the solution set and the reference set is quantified by MMD in terms of $K(R,S)$, and the distribution of the solution set itself is quantified by MMD in terms of $K(S,S')$, which explains why it has the ability of well-distinguishing different solution sets compared with other distance-based indicators, as shown in Fig. 1.

In addition, the induction of g in MMD usually requires a universal kernel [23], [49], [36]. As *Gaussian Kernel function* contains the least prior knowledge, it is safer to use when the distributions of the solution sets are not known in advance, as follows.

$$k(y,y') = exp(-\frac{||y-y'||^2}{2\sigma^2}) \tag{15}$$

where σ is a standard deviation parameter for controlling the range of Gaussian distribution. More information on MMD can be referred to in [23].

B. Algorithms for computing KBI

Based on MMD specified above, the KBI value of a solution set S is calculated by Algorithm 1 using a reference set R and the parameter σ of the Gaussian kernel function.

In Algorithm 1, KBI is divided into two major steps: 1) shift-based nondominated set reconstruction and 2) computing Gaussian-Kernel and KBI. In the following sections, each step is specified in detail.

Algorithm 1: Pseudo code for computing KBI

Input : A reference set R ;
A solution set S ;
The parameter: σ;
Output: KBI

$[R,S] = Normalization(R,S)$;
$MMD(R,S,\sigma) = 0$;
`Step 1: Shift-Based Non-Dominated Set Reconstruction:`
$S = SNSR(S,R)$;
`Step 2: Computing Kernel:`
`//Compute MMD based on Eq.13.`
$KBI = MMD(R,S,\sigma) = [\frac{1}{|R|^2}K(R,R,\sigma) + \frac{1}{|S|^2}K(S,S,\sigma) - \frac{2}{|R|\cdot|S|}K(R,S,\sigma)]^{\frac{1}{2}}$;
return KBI;

1) Shift-Based NonDominated Set Reconstruction(SNSR): Inspired from [20], [50], [51], a shift-based nondominated set reconstruction (SNSR) is proposed to ensure the weak Pareto compliance property of KBI. In Step 1 of Algorithm 1, SNSR is called for obtaining a nondominated set S, as shown in Algorithm 2.

The procedures of Algorithm 2 are described as follows. In Step 1, a reference subset $R_{wds} \in R$, each of which dominates at least one solution $s \in S$, is firstly obtained. The difference set R_s between the original reference set R and R_{wds} is considered as the reference set to be shifted. In Step 2, for each solution $r \in R_s$, its nearest solution $s \in S$ is obtained. Then, a direction vector d between r and s is computed as $d = max\{s-r, 0\}$. The solution s' can be obtained by shifting the reference point r with the direction vector d and it is finally merged into S.

Algorithm 2: Shift-based nondominated set reconstruction ($SNSR$)

Input : The solution set: S;
The reference set: R;
Output: S;
Step 1: Initialization:
// R_{wds} is a reference subset.
$R_{wds} = \{r | r \preceq s, \forall r \in R, \exists s \in S\}$;
// R_s is the reference set to be shifted.
$R_s = R \backslash R_{wds}$;
Step 2: Shift the reference points:
foreach $r \in R_s$ **do**
  $s = argmin_{s \in S} dist(s, r)$;
  $//d = [d_1, \cdots, d_m]$ is a direction vector.
  $d = max\{s - r, 0\}$;
  $s' = r + d$;
  $S = S \cup \{s'\}$;
end
return S;

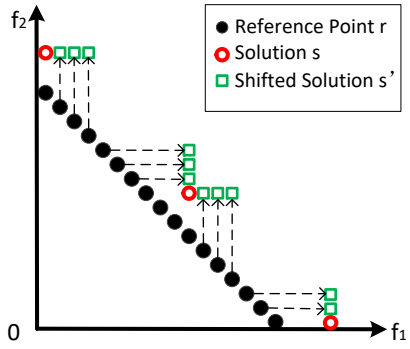


Fig. 3: An example of Shift-based nondominated set reconstruction (SNSR).

Fig.3 shows an example of reconstructing the shift-based nondominated set S. Each reference point $r \in R_s$ has been shifted to s' (marked by green "□"). Each shifted solution s' is weakly dominated by its nearest solution s (marked by red "○").

2) Computing KBI: In Step 2 of Algorithm 1, KBI is evaluated based on the maximum mean discrepancy (Eq.14), which is computed by calling Algorithm 3.

In Algorithm 3, $K(P, Q, \sigma)$ is firstly initialized to 0. Then, for each $p \in P$ and $q \in Q$, $k(p, q, \sigma)$ is computed based on Eq. (15). After that, $K(P, Q, \sigma)$ is obtained by summing up all $k(p, q, \sigma)$.

C. Computational Complexity of KBI

The average computational complexity of SNSR (Step 2 in Algorithm 1) is $O(|R| * |S|)$. Step 3 of Algorithm 1 requires $O(|R|^2)$ to compute the kernel values. As the number of the reference points (i.e., $|R|$) is usually equal or larger than that of solutions in the approximation (i.e., $|S|$), the total computational complexity of KBI is $O(|R|^2)$.

Algorithm 3: Computing Kernel (K)

Input : The set: P;
The set: Q;
The parameter: σ;
Output: $K(P, Q, \sigma)$;
Step 1: Initialization:
$K(P, Q, \sigma) = 0$
Step 2: Computing Kernel:
foreach $p \in P$ **do**
  foreach $q \in Q$ **do**
    //Compute $k(p, q, \sigma)$ based on Eq. 15.
    $k(p, q, \sigma) = exp(-\frac{||p-q||^2}{2\sigma^2})$;
    $K(P, Q, \sigma) = K(P, Q, \sigma) + k(p, q, \sigma)$;
  end
end
return $K(P, Q, \sigma)$;

D. Weak Pareto Compliance of KBI

KBI is weakly Pareto compliant (i.e., $\forall A, B : A \preceq B \implies I(A) \leq I(B)$). This indicates that KBI(A) $\leq$ KBI(B) always holds whenever $A \preceq B$ holds between two nondominated sets A and B. The proof of the weakly Pareto compliant property of KBI can be found in the supplementary material.

E. Comparisons with other indicators

The major properties of our proposed KBI and other indicators are summarized in Table I. It can be observed that the dominated or duplicate solutions have unwanted effect on IGD/$I_{\epsilon+}$/Δ_p. The distance-based indicators including IGD/IGD+/$I_{\epsilon+}$/Δ_p and the volume-based indicator including R2, lack of the ability to distinguish different solution sets. The computational complexity of HV grows exponentially with the increase in the number of objectives, which makes it computationally expensive for many-objective optimization. IGD and Δ_p are Pareto non-compliant. As a comprehensive quality indicator, only KBI possesses all the above desirable properties.

IV. Experiments and Discussions

In this section, we conduct computational experiments for

- validating KBI on artificial PFs;
- validating KBI on PF approximations;
- validating the optimal μ distributions of KBI;
- investigating the effects of σ on KBI;

A. Experimental Setup

In the experimental studies, KBI is verified on both artificial PFs and PF approximations delivered by five multi/many-objective optimizers (NSGA-II [52], IBEA [53], PAES [54], GrEA [55], NSGA-III [56] and MaOEA/D-2ADV [57]) on the DTLZ and WFG benchmark problems [58], [59]. For each algorithm, the population size is set to the size of the reference vector set, which is obtained by uniformly sampling on a unit

TABLE I: The comparisons between KBI and other comprehensive indicators.

Indicator	HV	R2	IGD	IGD$^+$	$I_{\epsilon+}$	Δ_p	KBI
No effects for dominated/ duplicate solutions	✓	✓		✓			✓
Effective for high dimensional PFs		✓	✓	✓	✓	✓	✓
Ability to distinguish different solution sets	✓						✓
Computational effort	exponential in m	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Pareto Compliant	strictly	weakly		weakly	weakly		weakly

simplex. This is usually called Das and Dennis's systematic approach [60]. In this approach, the reference vector size is

$$N = \binom{m-1}{H+m-1} \tag{16}$$

where $H > 0$ is the number of divisions along each objective coordinate and m is the number of objectives. However, as pointed out in [56], the direct use of Das and Dennis's approach may not be appropriate for $m > 6$. Instead, a two-layer direction vector generation method [56], [61] has been usually used for MaOPs with more than 6 objectives.

In the experimental studies, KBI is compared with other five indicators including HV [16], IGD [19], IGD$^+$ [20], $I_{\epsilon+}$ [39] and R2 [26]. For HV, the reference point is set to 1.1 times of the maximum values of PF approximations. For IGD/IGD$^+$/KBI, the reference set is uniformly sampled on PF. The size of the PF approximations (i.e., $|S|$), the size of the reference set (i.e., $|R|$) for different problems, and the parameter σ in KBI on m-objective optimization problems are listed in Table II. It is worth noting that the σ values are set to 1 for various problems based on the sensitivity test in Section IV. F.

TABLE II: The approximation/reference set size and the parameter σ for MaOPs.

the number of objectives m	3	5	10
the approximation size $\|S\|$	120	126	276
the parameter σ	1	1	1
reference set size $\|R\|$ for DTLZ2/DTLZ2^{-1}	9870	8855	7007
reference set size $\|R\|$ for DTLZ7	10000	10000	19683
reference set size $\|R\|$ for WFG1	9870	8855	7007
reference set size $\|R\|$ for WFG2	7425	7932	7007

B. KBD on Artificial PF Approximations

Three groups of artificial PF approximations are generated and presented in Figs. 4- 6. In the first group, solutions in each artificial PF approximations are uniformly distributed on the same hyperplane $f_1 + f_2 + f_3 = 1$ with a different range of objective values. Fig. 4a shows a uniformly distributed PF approximation with the objective values ranging in $[0, 1]$. Fig. 4b shows a uniformly distributed PF approximation with the objective values ranging in $[0.1, 0.8]$. Fig. 4c shows a uniformly distributed pf approximation with the objective values ranging in $[0.2, 0.6]$. It can be clearly seen in Fig. 4d that the three artificial PF approximations are all located on the same hyperplane $f_1 + f_2 + f_3 = 1$. It can be observed from Figs. 4a- 4c that the approximations with larger spread have smaller (better) KBI values, which indicates that KBI can accurately reflect the spread of PF approximations.

In the second group, solutions in different PF approximations are uniformly distributed with different convergence levels. Figs. 5a - 5c show the generated three artificial PF approximations, which are uniformly distributed on the hyperplanes $f_1 + f_2 + f_3 = 1$, $f_1 + f_2 + f_3 = 0.8$ and $f_1 + f_2 + f_3 = 0.5$, respectively. It can be observed that KBI values decrease for the artificial PF approximations with better convergence levels. Obviously, KBI can correctly reflect the convergence of PF approximations.

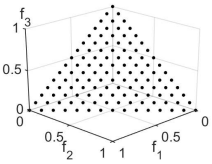
In the third group, four artificial PF approximations are distributed over the entire PF, although their uniformities are quite different from each other. These four artificial PF approximations are generated in the following way. The solutions are firstly uniformly generated on the plane $f_1 + f_2 + f_3 = 1$, ranging in $[0, 1]$, as shown in Fig. 4a. After that, by randomly selecting 30%, 50%, 70% or 100% of solutions and disturb them with randomly generated noise on the same hyperplane, approximations with different uniformities are obtained, as shown in Figs. 6a, 6b, 6c and 6d. It can be observed from these figures that the artificial PF approximations with better uniformities have smaller (better) KBI values. Clearly, KBI is able to correctly reflect the uniformity of PF approximations.

C. KBI on solution sets for regular PFs

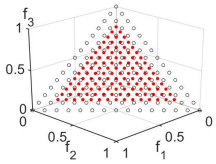
DTLZ2 [58] and WFG1 [59] , whose PFs are regular, have been selected for verifying the effectiveness of KBI. DTLZ2^{-1} has the inverted front of DTLZ2. For DTLZ2/DTLZ2^{-1} and WFG1, it is relatively easy for all the optimizers to converge to their PFs. This characteristic can help to test the diversity performance of approximations obtained by the different algorithms. KBI, IGD, IGD$^+$, HV, $I_{\epsilon+}$ and R2 values/ranks of solution sets obtained by the six algorithms on DTLZ2/DTLZ2^{-1}/WFG1 are presented in Table III.

Approximations delivered by the five many-objective optimizers on tri-objective DTLZ2 are plotted in Fig. 7. Their performance, in terms of KBI/IGD/HV/IGD$^+$/$I_{\epsilon+}$/R2 values, are given in Table III. For five-objective DTLZ2, the parallel coordinate plots of five approximations are plotted in Fig. 8. Their corresponding KBI values are presented in Table III.

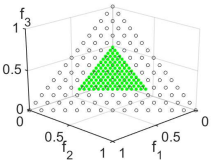
For the tri-objective DTLZ2, it can be observed from Fig. 7 that the approximation obtained by NSGA-III is uniformly distributed on the whole PF, thus it has the best (lowest) KBI



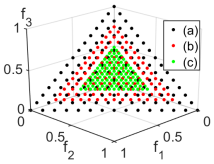
(a) Every objective value of the PF approximation ranges in $[0, 1]$ and its KBI is 0.000000.



(b) Every objective value of the PF approximation ranges in $[0.1, 0.8]$ and its KBI is 0.010639.

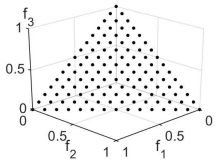


(c) Every objective value of the PF approximation ranges in $[0.2, 0.6]$ and its KBI is 0.027228.

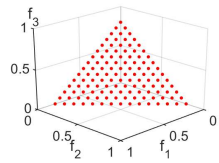


(d) The PF approximations in Figs. 4a, 4b and 4c are located on the hyperplane $f_1 + f_2 + f_3 = 1$.

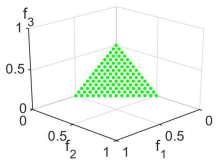
Fig. 4: Three artificial PF approximations with different spread, uniformly located on the same hyperplane $f_1 + f_2 + f_3 = 1$.



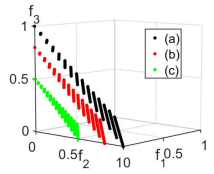
(a) The PF approximation is uniformly located on the hyperplane $f_1 + f_2 + f_3 = 1$ and its KBI is 0.000000.



(b) The PF approximation is uniformly located on the hyperplane $f_1 + f_2 + f_3 = 0.8$ and its KBI is 0.038138.

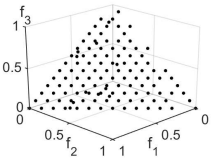


(c) The PF approximation is uniformly located on the hyperplane $f_1 + f_2 + f_3 = 0.5$ and its KBI is 0.095652.

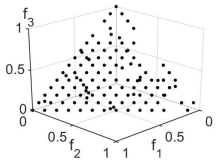


(d) The PF approximations in Figs. 5a, 5b and 5c are uniformly located on three different hyperplanes thus they have different convergence levels.

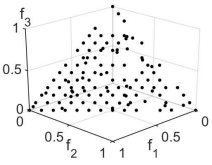
Fig. 5: Three artificial PF approximations uniformly located on three different hyperplanes.



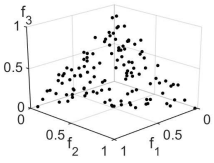
(a) 30% of solutions are replaced with randomly generated solutions on the same hyperplane. Its KBI is 0.001287.



(b) 50% of solutions are replaced with randomly generated solutions on the same hyperplane. Its KBI is 0.002574.

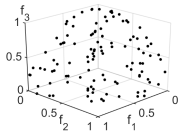


(c) 70% of solutions are replaced with randomly generated solutions on the same hyperplane. Its KBI is 0.003025.

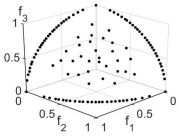


(d) 100% of solutions are replaced with randomly generated solutions on the same hyperplane. Its KBI is 0.007761.

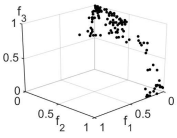
Fig. 6: Artificial PF approximations with different uniformities located on the hyperplane $f_1 + f_2 + f_3 = 1$.



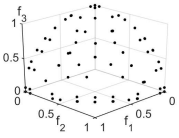
(a) KBI=0.010376 (5), NSGA-II



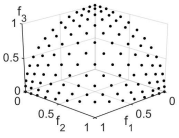
(b) KBI=0.008351 (2), IBEA



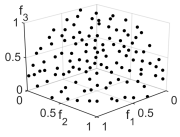
(c) KBI=0.059429 (6), PAES



(d) KBI=0.009480 (4), GrEA

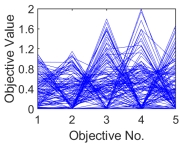


(e) KBI=0.006145 (1), NSGA-III

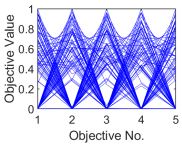


(f) KBI=0.008376 (3), MaOEA/D-2ADV

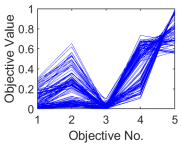
Fig. 7: The nondominated sets obtained by the six algorithms and the corresponding ranks in terms of KBI values on the tri-objective DTLZ2.



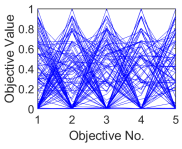
(a) KBI=0.036642 (5), NSGA-II



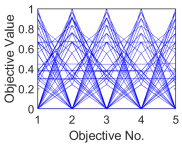
(b) KBI=0.009722 (1), IBEA



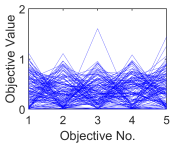
(c) KBI=0.082184 (6), PAES



(d) KBI=0.012632 (3), GrEA



(e) KBI=0.011059 (2), NSGA-III

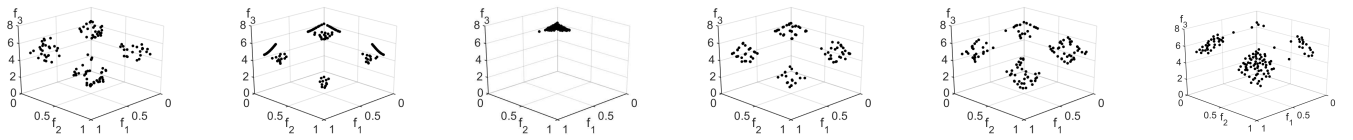


(f) KBI=0.020862 (4), MaOEA/D-2ADV

Fig. 8: The nondominated sets obtained by the six algorithms and the corresponding ranks in terms of KBI values on the five-objective DTLZ2.

TABLE III: KBI, IGD, IGD$^+$, HV, $I_{\epsilon+}$ and R2 values/ranks of solution sets obtained by six algorithms on DTLZ2/DTLZ2^{-1}/WFG1.

instances (m-obj)	indicator	NSGA-II	IBEA	PAES	GrEA	NSGA-III	MaOEA/D-2ADV
DTLZ2 (3-obj)	KBI	$1.04E-02$ (5)	$8.35E-03$ (2)	$5.94E-02$ (6)	$9.48E-03$ (4)	$6.15E-03$ (1)	$8.38E-03$ (3)
	IGD	$6.37E-02$ (3)	$8.26E-02$ (5)	$2.97E-01$ (6)	$7.72E-02$ (4)	$4.68E-02$ (1)	$5.61E-02$ (2)
	HV	$7.09E-01$ (5)	$7.57E-01$ (1)	$4.74E-01$ (6)	$7.25E-01$ (4)	$7.53E-01$ (2)	$7.41E-01$ (3)
	IGD$^+$	$3.47E-02$ (5)	$1.87E-02$ (1)	$1.96E-01$ (6)	$3.02E-02$ (4)	$1.96E-02$ (2)	$2.63E-02$ (3)
	$I_{\epsilon+}$	$1.08E-01$ (5)	$6.40E-02$ (1)	$7.09E-01$ (6)	$1.07E-01$ (4)	$7.09E-02$ (2)	$7.61E-02$ (3)
	R2	$9.33E-02$ (4)	$9.18E-02$ (2)	$1.42E-01$ (6)	$9.33E-02$ (5)	$9.17E-02$ (1)	$9.20E-01$ (3)
DTLZ2 (5-obj)	KBI	$3.66E-02$ (5)	$9.72E-03$ (1)	$8.22E-02$ (6)	$1.26E-02$ (3)	$1.11E-02$ (2)	$2.09E-02$ (4)
	IGD	$3.03E-01$ (5)	$2.36E-01$ (4)	$6.75E-01$ (6)	$1.99E-01$ (2)	$1.95E-01$ (1)	$2.04E-01$ (3)
	HV	$7.17E-01$ (5)	$1.30E-00$ (1)	$4.30E-01$ (6)	$1.27E-00$ (3)	$1.28E-00$ (2)	$1.10E-00$ (4)
	IGD$^+$	$2.34E-01$ (5)	$6.33E-02$ (1)	$4.55E-01$ (6)	$7.68E-02$ (3)	$7.19E-02$ (2)	$1.36E-01$ (4)
	$I_{\epsilon+}$	$3.49E-01$ (5)	$1.74E-01$ (1)	$6.01E-01$ (6)	$1.86E-01$ (2)	$1.87E-01$ (3)	$3.06E-01$ (4)
	R2	$2.29E-02$ (4)	$1.84E-02$ (2)	$1.30E-01$ (6)	$1.88E-02$ (3)	$1.83E-02$ (1)	$2.51E-02$ (5)
DTLZ2 (10-obj)	KBI	$1.01E-01$ (6)	$2.23E-02$ (2)	$4.04E-02$ (5)	$2.31E-02$ (3)	$1.56E-02$ (1)	$3.59E-02$ (4)
	IGD	$1.35E-00$ (6)	$7.90E-01$ (5)	$6.81E-01$ (4)	$5.00E-01$ (2)	$4.22E-01$ (1)	$5.68E-01$ (3)
	HV	$4.60E-03$ (6)	$2.09E-00$ (2)	$1.29E-00$ (4)	$1.90E-00$ (3)	$2.51E-00$ (1)	$1.07E-00$ (5)
	IGD$^+$	$1.19E-00$ (6)	$3.64E-01$ (3)	$4.55E-01$ (4)	$2.95E-01$ (2)	$1.79E-01$ (1)	$4.92E-01$ (5)
	$I_{\epsilon+}$	$8.76E-01$ (6)	$5.84E-01$ (5)	$4.96E-01$ (4)	$4.18E-01$ (2)	$2.03E-01$ (1)	$4.64E-01$ (3)
	R2	$2.31E-02$ (5)	$9.31E-03$ (2)	$3.93E-02$ (6)	$9.53E-03$ (3)	$8.28E-03$ (1)	$1.42E-02$ (4)
WFG1 (3-obj)	KBI	$2.98E-01$ (3)	$1.98E-01$ (1)	$4.62E-01$ (6)	$2.81E-01$ (2)	$3.45E-01$ (4)	$3.51E-01$ (5)
	IGD	$1.04E-00$ (3)	$7.07E-01$ (1)	$1.68E-00$ (6)	$9.79E-01$ (2)	$1.23E-00$ (5)	$1.21E-00$ (4)
	HV	$3.02E+01$ (2)	$3.83E+01$ (1)	$1.13E+01$ (6)	$2.99E+01$ (3)	$2.33E+01$ (5)	$2.48E+01$ (4)
	IGD$^+$	$1.02E-00$ (3)	$6.94E-01$ (1)	$1.62E-00$ (6)	$9.65E-01$ (2)	$1.18E-00$ (4)	$1.20E-00$ (5)
	$I_{\epsilon+}$	$1.14E-00$ (3)	$8.00E-01$ (1)	$1.64E-00$ (6)	$1.09E-00$ (2)	$1.30E-00$ (5)	$1.23E-00$ (4)
	R2	$4.14E-01$ (2)	$3.24E-01$ (1)	$6.29E-01$ (6)	$4.22E-01$ (3)	$4.68E-01$ (4)	$5.07E-01$ (5)
WFG1 (5-obj)	KBI	$3.26E-01$ (3)	$2.15E-01$ (1)	$3.82E-01$ (6)	$2.64E-01$ (2)	$3.32E-01$ (5)	$3.32E-01$ (4)
	IGD	$1.86E-00$ (4)	$1.27E-00$ (1)	$2.20E-00$ (6)	$1.53E-00$ (2)	$1.86E-00$ (3)	$1.87E-00$ (5)
	HV	$2.30E+03$ (3)	$3.12E+03$ (1)	$1.53E+03$ (6)	$2.62E+03$ (2)	$1.97E+03$ (4)	$1.94E+03$ (5)
	IGD$^+$	$1.83E-00$ (3)	$1.22E-00$ (1)	$2.13E-00$ (6)	$1.49E-00$ (2)	$1.84E-00$ (4)	$1.86E-00$ (5)
	$I_{\epsilon+}$	$1.45E-00$ (5)	$9.28E-01$ (1)	$1.59E-00$ (6)	$1.11E-00$ (2)	$1.23E-00$ (3)	$1.32E-00$ (4)
	R2	$3.78E-01$ (3)	$2.77E-01$ (1)	$4.64E-01$ (6)	$3.48E-01$ (2)	$4.39E-01$ (4)	$4.44E-01$ (5)
WFG1 (10-obj)	KBI	$2.60E-01$ (4)	$2.36E-01$ (1)	$2.73E-01$ (6)	$2.52E-01$ (2)	$2.67E-01$ (5)	$2.56E-01$ (3)
	IGD	$2.98E-00$ (3)	$2.74E-00$ (1)	$3.21E-00$ (6)	$2.91E-00$ (2)	$3.13E-00$ (5)	$2.99E-00$ (4)
	HV	$2.06E+09$ (3)	$2.31E+09$ (1)	$1.90E+09$ (6)	$2.06E+09$ (2)	$2.03E+09$ (4)	$1.99E+09$ (5)
	IGD$^+$	$2.87E-00$ (4)	$2.62E-00$ (1)	$3.01E-00$ (6)	$2.79E-00$ (2)	$2.94E-00$ (5)	$2.82E-00$ (3)
	$I_{\epsilon+}$	$1.58E-00$ (4)	$1.31E-00$ (2)	$1.81E-00$ (6)	$1.30E-00$ (1)	$1.73E-00$ (5)	$1.40E-00$ (3)
	R2	$3.44E-01$ (2)	$3.26E-01$ (1)	$3.58E-01$ (6)	$3.49E-01$ (4)	$3.47E-01$ (3)	$3.52E-01$ (5)
DTLZ2^{-1} (3-obj)	KBI	$1.05E-02$ (5)	$7.38E-03$ (1)	$4.99E-02$ (6)	$8.45E-03$ (3)	$1.04E-02$ (4)	$7.63E-03$ (2)
	IGD	$6.17E-02$ (2)	$8.00E-02$ (4)	$2.00E-01$ (6)	$8.01E-02$ (5)	$6.76E-02$ (3)	$5.31E-02$ (1)
	HV	$6.88E-01$ (5)	$7.14E-01$ (2)	$3.78E-01$ (6)	$7.05E-01$ (3)	$6.96E-01$ (4)	$7.20E-01$ (1)
	IGD$^+$	$3.79E-02$ (5)	$2.84E-02$ (2)	$1.89E-01$ (6)	$3.23E-02$ (3)	$3.64E-02$ (4)	$2.58E-02$ (1)
	$I_{\epsilon+}$	$1.17E-01$ (5)	$7.65E-02$ (2)	$2.82E-01$ (6)	$9.40E-02$ (4)	$9.04E-02$ (3)	$7.44E-02$ (1)
	R2	$1.26E-01$ (5)	$1.24E-01$ (2)	$1.65E-01$ (6)	$1.25E-01$ (4)	$1.25E-01$ (3)	$1.23E-01$ (1)
DTLZ2^{-1} (5-obj)	KBI	$2.64E-02$ (4)	$1.58E-02$ (2)	$8.33E-02$ (6)	$1.38E-02$ (1)	$2.64E-02$ (5)	$1.77E-02$ (3)
	IGD	$2.46E-01$ (3)	$2.79E-01$ (5)	$5.69E-01$ (6)	$2.11E-01$ (2)	$2.50E-01$ (4)	$2.02E-01$ (1)
	HV	$9.26E-02$ (4)	$1.80E-01$ (2)	$3.26E-04$ (6)	$1.83E-01$ (1)	$7.79E-02$ (5)	$1.64E-01$ (3)
	IGD$^+$	$1.74E-01$ (4)	$1.25E-01$ (3)	$5.29E-01$ (6)	$1.01E-01$ (1)	$1.83E-01$ (5)	$1.17E-01$ (2)
	$I_{\epsilon+}$	$3.14E-01$ (4)	$2.42E-01$ (3)	$6.04E-01$ (6)	$1.85E-01$ (1)	$3.50E-01$ (5)	$2.01E-01$ (2)
	R2	$1.50E-01$ (5)	$1.49E-01$ (4)	$2.06E-01$ (6)	$1.38E-01$ (1)	$1.46E-01$ (3)	$1.40E-01$ (2)
DTLZ2^{-1} (10-obj)	KBI	$2.38E-02$ (4)	$1.45E-02$ (2)	$9.04E-02$ (6)	$1.28E-02$ (1)	$1.93E-02$ (3)	$2.59E-02$ (5)
	IGD	$4.84E-01$ (1)	$5.29E-01$ (4)	$1.14E+00$ (6)	$5.87E-01$ (5)	$5.06E-01$ (3)	$4.88E-01$ (2)
	HV	$5.08E-05$ (4)	$8.53E-04$ (2)	$0.00E+00$ (6)	$1.26E-03$ (1)	$3.37E-04$ (3)	$3.29E-05$ (5)
	IGD$^+$	$3.51E-01$ (4)	$2.77E-01$ (1)	$1.07E+00$ (6)	$2.83E-01$ (2)	$3.08E-01$ (3)	$3.67E-01$ (5)
	$I_{\epsilon+}$	$4.44E-01$ (2)	$5.17E-01$ (4)	$9.02E-01$ (6)	$5.22E-01$ (5)	$4.50E-01$ (3)	$3.71E-01$ (1)
	R2	$1.61E-01$ (1)	$1.73E-01$ (3)	$2.38E-01$ (6)	$1.86E-01$ (5)	$1.74E-01$ (4)	$1.66E-01$ (2)

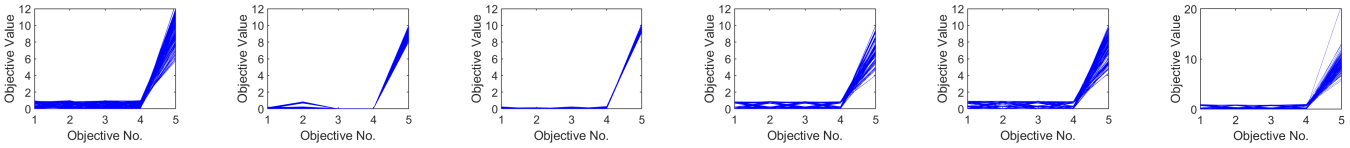


(a) KBI=0.025633 (4), NSGA-II (b) KBI=0.006393 (1), IBEA (c) KBI=0.318480 (6), PAES (d) KBI=0.011279 (2), GrEA (e) KBI=0.024524 (3), NSGA-III (f) KBI=0.067828 (5), MaOEA/D-2ADV

Fig. 9: The nondominated sets obtained by six algorithms and the corresponding ranks in terms of KBI values on tri-objective DTLZ7.

TABLE IV: KBI, IGD, IGD$^+$, HV, $I_{\epsilon+}$ and R2 values/ranks of solution sets obtained by the six algorithms on DTLZ7 and WFG2.

instances (m-obj)	indicator	NSGA-II	IBEA	PAES	GrEA	NSGA-III	MaOEA/D-2ADV
DTLZ7 (3-obj)	KBI	$2.56E-02$ (4)	$6.39E-03$ (1)	$3.18E-01$ (6)	$1.13E-02$ (2)	$2.45E-02$ (3)	$6.78E-02$ (5)
	IGD	$1.13E-01$ (4)	$1.11E-01$ (3)	$1.15E-00$ (6)	$7.50E-02$ (1)	$1.06E-01$ (2)	$3.19E-01$ (5)
	HV	$1.38E-00$ (4)	$1.63E-00$ (1)	$8.65E-01$ (6)	$1.57E-00$ (2)	$1.40E-00$ (3)	$9.43E-01$ (5)
	IGD$^+$	$9.98E-02$ (4)	$2.58E-02$ (1)	$1.01E-00$ (6)	$4.47E-02$ (2)	$9.43E-02$ (3)	$3.09E-01$ (5)
	$I_{\epsilon+}$	$3.38E-01$ (4)	$9.16E-02$ (1)	$3.01E-00$ (6)	$1.45E-01$ (2)	$3.17E-01$ (3)	$5.70E-01$ (5)
	R2	$9.88E-01$ (4)	$9.23E-01$ (1)	$1.87E-00$ (6)	$9.46E-01$ (2)	$9.68E-01$ (3)	$1.08E-00$ (5)
DTLZ7 (5-obj)	KBI	$1.05E-01$ (4)	$2.07E-01$ (5)	$3.74E-01$ (6)	$1.79E-02$ (1)	$2.64E-02$ (2)	$8.27E-02$ (3)
	IGD	$7.82E-01$ (4)	$1.48E-00$ (5)	$2.21E-00$ (6)	$2.78E-01$ (1)	$3.23E-01$ (2)	$6.57E-01$ (3)
	HV	$3.11E-01$ (6)	$1.52E-00$ (3)	$1.36E-00$ (4)	$2.25E-00$ (1)	$2.05E-00$ (2)	$8.19E-01$ (5)
	IGD$^+$	$7.67E-01$ (4)	$1.19E-00$ (5)	$2.00E-00$ (6)	$1.12E-01$ (1)	$1.88E-01$ (2)	$6.42E-01$ (3)
	$I_{\epsilon+}$	$2.54E-00$ (4)	$4.79E-00$ (5)	$5.94E-00$ (6)	$7.95E-01$ (1)	$1.02E-00$ (2)	$2.52E-00$ (3)
	R2	$1.16E-00$ (3)	$1.60E-00$ (5)	$1.84E-00$ (6)	$8.15E-01$ (1)	$8.60E-01$ (2)	$1.59E-00$ (4)
DTLZ7 (10-obj)	KBI	$1.16E-00$ (5)	$5.32E-01$ (4)	$3.63E-01$ (3)	$3.62E-02$ (1)	$7.92E-02$ (2)	$1.18E-00$ (6)
	IGD	$1.57E+01$ (5)	$5.93E-00$ (4)	$4.05E-00$ (3)	$8.55E-01$ (1)	$1.29E-00$ (2)	$1.62E+01$ (6)
	HV	$0.00E-00$ (5)	$1.55E-00$ (3)	$2.21E-00$ (2)	$1.23E-00$ (4)	$2.30E-00$ (1)	$0.00E-00$ (6)
	IGD$^+$	$1.57E+01$ (5)	$5.75E-00$ (4)	$3.86E-00$ (3)	$6.50E-01$ (1)	$9.88E-01$ (2)	$1.62E+01$ (5)
	$I_{\epsilon+}$	$2.45E+01$ (5)	$1.46E+01$ (4)	$1.27E+01$ (3)	$3.40E-00$ (1)	$8.07E-00$ (2)	$2.51E+01$ (6)
	R2	$2.95E-00$ (5)	$1.94E-00$ (4)	$1.75E-00$ (3)	$8.24E-01$ (1)	$1.28E-00$ (2)	$2.99E-00$ (6)
WFG2 (3-obj)	KBI	$3.15E-02$ (5)	$2.34E-02$ (2)	$5.10E-02$ (6)	$2.16E-02$ (1)	$2.54E-02$ (3)	$2.62E-02$ (4)
	IGD	$2.04E-01$ (4)	$2.75E-01$ (6)	$2.59E-01$ (5)	$1.95E-01$ (3)	$1.55E-01$ (1)	$1.58E-01$ (2)
	HV	$5.78E+01$ (3)	$5.90E+01$ (1)	$5.40E+01$ (6)	$5.84E+01$ (2)	$5.77E+01$ (4)	$5.74E+01$ (5)
	IGD$^+$	$1.05E-01$ (5)	$3.23E-02$ (1)	$1.83E-01$ (6)	$6.38E-02$ (2)	$8.47E-02$ (3)	$8.49E-02$ (4)
	$I_{\epsilon+}$	$3.47E-01$ (6)	$7.16E-02$ (1)	$3.15E-01$ (5)	$1.50E-01$ (2)	$2.00E-01$ (4)	$1.99E-01$ (3)
	R2	$1.60E-01$ (5)	$1.48E-01$ (1)	$1.82E-01$ (6)	$1.55E-01$ (2)	$1.59E-01$ (3)	$1.59E-01$ (4)
WFG2 (5-obj)	KBI	$8.04E-02$ (5)	$3.36E-02$ (1)	$9.64E-02$ (6)	$3.71E-02$ (2)	$4.39E-02$ (4)	$4.19E-02$ (3)
	IGD	$7.91E-01$ (6)	$4.91E-01$ (3)	$7.44E-01$ (5)	$5.26E-01$ (4)	$4.70E-01$ (2)	$4.45E-01$ (1)
	HV	$5.77E+03$ (2)	$5.79E+03$ (1)	$4.94E+03$ (6)	$5.69E+03$ (3)	$5.69E+03$ (4)	$5.69E+03$ (5)
	IGD$^+$	$5.47E-01$ (6)	$1.81E-01$ (1)	$5.36E-01$ (5)	$2.00E-01$ (2)	$2.57E-01$ (4)	$2.51E-01$ (3)
	$I_{\epsilon+}$	$8.60E-01$ (6)	$3.95E-01$ (2)	$5.94E-01$ (5)	$3.51E-01$ (1)	$4.67E-01$ (4)	$4.32E-01$ (3)
	R2	$5.41E-02$ (2)	$5.57E-02$ (3)	$1.03E-01$ (6)	$6.00E-02$ (5)	$5.19E-02$ (1)	$5.62E-02$ (4)
WFG2 (10-obj)	KBI	$7.91E-02$ (5)	$3.38E-02$ (1)	$1.32E-01$ (6)	$5.23E-02$ (3)	$4.92E-02$ (2)	$6.08E-02$ (4)
	IGD	$1.50E-00$ (5)	$1.05E-00$ (1)	$1.80E-00$ (6)	$1.22E-00$ (4)	$1.16E-00$ (3)	$1.13E-00$ (2)
	HV	$8.00E+09$ (3)	$8.43E+09$ (1)	$6.13E+09$ (6)	$8.02E+09$ (2)	$7.87E+09$ (4)	$7.63E+09$ (5)
	IGD$^+$	$8.41E-01$ (5)	$3.63E-01$ (1)	$1.52E-00$ (6)	$5.91E-01$ (3)	$5.62E-01$ (2)	$6.28E-01$ (4)
	$I_{\epsilon+}$	$8.06E-01$ (5)	$3.73E-01$ (1)	$1.04E-00$ (6)	$5.09E-01$ (3)	$6.46E-01$ (4)	$4.53E-01$ (2)
	R2	$3.37E-02$ (3)	$2.49E-02$ (1)	$1.07E-01$ (6)	$3.25E-02$ (2)	$3.69E-02$ (4)	$5.35E-02$ (5)



(a) KBI=0.104890 (4), NSGA-II (b) KBI=0.206909 (5), IBEA (c) KBI=0.374074 (6), PAES (d) KBI=0.017905(1), GrEA (e) KBI=0.026395 (2), NSGA-III (f) KBI=0.082793 (3), MaOEA/D-2ADV

Fig. 10: The nondominated sets obtained by six algorithms and the corresponding ranks in terms of KBI values on five-objective DTLZ7.

value (0.006145). IBEA ranks the second in terms of KBI value (0.008351) as the boundaries of its obtained approximation is well-spread but not as uniform as the one obtained by NSGA-III. The KBI value of MaOEA/D-2ADV (0.008376) is better than that of GrEA (0.009480). This is consistent with the observations in Fig. 7d and 7f that the approximation obtained by MaOEA/D-2ADV is more extensive than that of GrEA. As shown in Figs. 7a and 7c, the approximation obtained by NSGA-II is not uniformly distributed and the approximation obtained by PAES has a worse distribution than that of NSGA-II. Thus, the KBI value of NSGA-II (0.010376) is the second largest and PAES (0.059429) is the largest.

For the five-objective DTLZ2 whose PF is in $[0,1]^5$, it can be observed from Fig. 8 that the approximation obtained by the IBEA is closest to the real PF, thus it has the best KBI value (0.009722). NSGA-III ranks the second in terms of KBI value (0.011059) as its distribution is not as extensive as the one obtained by IBEA. GrEA ranks the third in terms of KBI value (0.012632) as its distribution is not as uniform as the ones obtained by IBEA and NSGA-III. As show in Figs. 8f and 8a, the approximation obtained by MaOEA/D-2ADV does not converge well, and NSGA-II converges even worse. Thus, the KBI value of MaOEA/D-2ADV (0.020862) is better than that of NSGA-II (0.036642). In addition, it can be observed from Fig. 8c that the approximation obtained by PAES is different from the true PF, which leads to the worst KBI value (0.082184).

D. KBI on solution sets for irregular PFs

The effectiveness of KBI on solution sets for irregular PFs is further verified in this section. DTLZ7 [58] is a typical irregular PF, consisting of $2m-1$ disconnected segments, which can be either convex or concave. Similarly, WFG2 [59] also has a disconnected PF, which is scalable with the number of objectives. KBI, IGD, IGD$^+$, HV, $I_{\epsilon+}$ and R2 values/ranks of solution sets obtained by the six algorithms

TABLE V: Values and ranks in terms of six indicators for their approximately optimal μ-distributions A_I obtained by each quality indicator I on DTLZ2.

indicator	A_{HV}	A_{IGD}	A_{IGD+}	$A_{I_{\epsilon+}}$	A_{R2}	A_{KBI}
HV	$6.89E-01$ (1)	$5.89E-01$ (6)	$6.87E-01$ (2)	$6.43E-01$ (5)	$6.75E-01$ (3)	$6.67E-01$ (4)
IGD	$1.54E-01$ (4)	$1.04E-01$ (1)	$1.56E-01$ (5)	$1.24E-01$ (2)	$1.31E-01$ (3)	$1.80E-01$ (6)
IGD$^+$	$4.19E-02$ (2)	$5.88E-02$ (6)	$4.16E-02$ (1)	$5.16E-02$ (5)	$4.47E-02$ (3)	$4.57E-02$ (4)
$I_{\epsilon+}$	$1.16E-01$ (2)	$1.18E-01$ (4)	$1.26E-01$ (5)	$9.91E-02$ (1)	$1.17E-01$ (3)	$1.70E-01$ (6)
R2	$9.08E-02$ (3)	$1.05E-01$ (6)	$9.10E-02$ (2)	$9.64E-02$ (5)	$9.07E-02$ (1)	$9.43E-02$ (4)
KBI	$1.46E-02$ (3)	$1.54E-02$ (5)	$1.40E-02$ (2)	$1.60E-02$ (6)	$1.52E-02$ (4)	$1.04E-02$ (1)

TABLE VI: Kendall rank τ of the six indicators on DTLZ2.

	HV	IGD	IGD$^+$	$I_{\epsilon+}$	R2	KBI
HV	1.00	-0.47	0.87	-0.07	0.60	0.33
IGD	-0.47	1.00	-0.60	0.47	-0.33	-0.87
IGD$^+$	0.87	-0.60	1.00	-0.07	0.73	0.47
$I_{\epsilon+}$	0.07	0.47	-0.07	1.00	-0.07	-0.60
R2	0.60	-0.33	0.73	-0.07	1.00	0.20
KBI	0.33	-0.87	0.47	-0.60	0.20	1.00

on DTLZ7/WFG2 are presented in Table IV.

For tri-objective DTLZ7, it can be observed from Fig. 9 that the boundary of the approximation obtained by IBEA is well-distributed on the whole PF, thus IBEA has the best (lowest) KBI value (0.006393). GrEA ranks the second in terms of its KBI value (0.011279) as its boundary is not as uniform as the one obtained by IBEA. As shown in Fig. 9e and Fig. 9a, the approximations obtained by NSGA-III and NSGA-II are distributed extensively but non-uniformly. The KBI values of NSGA-III (0.024524) and NSGA-II (0.025633) really reflect such a phenomenon. As shown in Fig. 9f, the boundary of the approximation obtained by MaOEA/D-2ADV is not uniformly-distributed. Thus, it has the worse KBI value (0.067828) than NSGA-II. PAES has the worst KBI value (0.318480), as it only covers the top part of PF, as shown in Fig. 9c.

For the five-objective DTLZ7, it can be observed from Fig. 10 that the closest approximation to the true PF is obtained by GrEA, thus it has the best KBI value (0.017905). NSGA-III ranks the second in terms of the KBI value (0.026395) as its approximation is not as well-distributed as the one obtained by GrEA. MaOEA/D-2ADV ranks the third in terms of KBI value (0.082793) as its approximation is not as uniform as the one obtained by both GrEA and NSGA-III. As shown in Fig. 10a, the approximation obtained by NSGA-II is uniformly-distributed. The spread of the approximation obtained by IBEA in Fig. 10b is even worse than that obtained by NSGA-II. Thus, the KBI value of NSGA-II (0.104890) is better than IBEA (0.206909). In addition, it can be observed from Fig. 10c that the approximation obtained by the PAES covers only a part of PF, which leads to the worst KBI value (0.371074).

E. The optimal μ-distributions for KBI

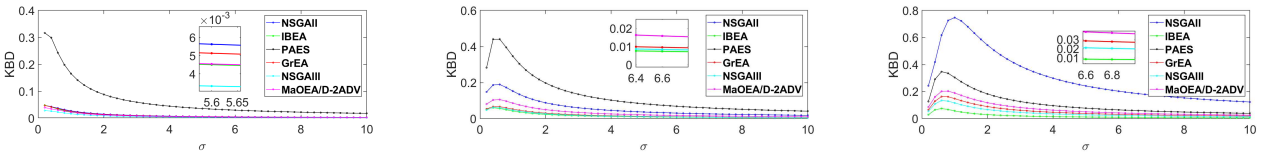
The effectiveness of KBI on MaOPs with different types of PFs has been verified in Section IV-B, IV-C and IV-D. In this section, the optimal μ-distribution [62] is used to further investigate the properties of KBI, compared with other indicators. The optimal μ-distribution reflects the geometrical distributions of μ solutions that maximize the corresponding indicator. Similar to [62], the approximately optimal μ-distribution for each indicator is obtained by running algorithm L-SHADE [63]. In this section, the approximately optimal μ-distribution for a quality indicator I is denoted as A_I (e.g. A_{KBI}).

The values and ranks in terms of the six indicators for the approximately optimal μ-distributions A_I obtained by a quality indicator I on DTLZ2 have been summarized in Table V. One obvious observation consistent with our intuition is that the approximately optimal μ-distribution A_I always ranks the first in terms of the quality indicator I. However, the ranks of A_I in terms of other indicators are different from each other. This observation suggests that the properties of different indicators are different.

To further investigate the correlation of the six indicators, the Kendall rank τ [64], which measures the correlation of two indicators, have been evaluated for each indicator pair on DTLZ2, as shown in Table VI. The value of τ ranges from -1 to 1. A positive τ value means that the two indicators are consistent with each other while a negative value represents their results are conflicting with each other. It can be observed from Table VI that the Kendall rank τ of KBI are quite different from that of the other indicators. The τ values between KBI and IGD/$I_{\epsilon+}$ are -0.87/-0.60, indicating that KBI and IGD/$I_{\epsilon+}$ are negatively correlated with each other. The τ value between KBI and HV/IGD$^+$/R2 are 0.33/0.47/0.20, indicating that KBI and HV/IGD$^+$/R2 are weakly correlated with each other. However, none of the other indicators is strongly correlated with KBI, which verifies its uniqueness. Due to the page limit, the Kendall ranks of the six indicators on other test problems are shown in the supplementary material.

F. Sensitivity test of parameter σ

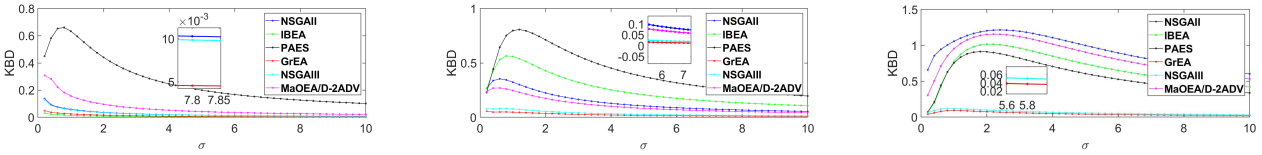
In this section, we investigate the effects of the width parameter σ on KBI. Fig. 11 shows the values of KBI for various σ values on the tri-, five- and ten-objective DTLZ2/DTLZ7. It can be observed that the ranks of the six algorithms in terms of KBI remain unchanged over a wide range of σ values. This indicates that KBI is not sensitive to the parameter σ. Another interesting observation from Fig. 11 is that KBI values always firstly increase then decrease with the increase of the σ value. The compared algorithms have the largest gaps in term of their KBI values when σ value is around 1 for all the problems except for ten-objective DTLZ7, which implies that KBI has the best performance in terms of distinguishing different approximations when the σ value is around 2 for ten-objective DTLZ7 and 1 for all the other test problems. Based on the above analysis, the σ value is suggested to be set to 1 for the users of KBI.



(a) KBI of the approximations obtained by the six algorithms on tri-objective DTLZ2 with different σ values.

(b) KBI of the approximations obtained by the six algorithms on five-objective DTLZ2 with different σ values.

(c) KBI of the approximations obtained by the six algorithms on ten-objective DTLZ2 with different σ values.



(d) KBI of the approximations obtained by the six algorithms on tri-objective DTLZ7 with different σ values.

(e) KBI of the approximations obtained by the six algorithms on five-objective DTLZ7 with different σ values.

(f) KBI of the approximations obtained by the six algorithms on ten-objective DTLZ7 with different σ values.

Fig. 11: KBI values of the six approximations with different σ values.

V. Conclusion

In this paper, we proposed a new performance indicator to evaluate PF approximations obtained by multiobjective optimizers. In the proposed KBI, the kernel-based distance is used for measuring the difference between two solution sets. A shift-based nondominated set reconstruction (SNSR) was also proposed for maintaining the weak Pareto compliance property of KBI. Through computational experiments, it was demonstrated that KBI can measure both the convergence and the diversity of solution sets in a consistent manner with our intuition. It is also shown that KBI does not have a large similarity to any other indicators. One interesting research direction is to use KBI for solution subset selection. It is also interesting to use KBI in indicator-based algorithms.

References

[1] K. Miettinen, *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic Publishers, 1999.

[2] Z. He and G. G. Yen, "Visualization and performance metric in many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 20, pp. 386–402, June 2016.

[3] T. Okabe, Y. Jin, and B. Sendhoff, "A critical survey of performance indices for multi-objective optimisation," in *IEEE Congress on Evolutionary Computation CEC'03*, vol. 2, pp. 878–885, 2003.

[4] E. Zitzler, J. Knowles, and L. Thiele, "Quality assessment of Pareto set approximations," in *Multiobjective Optimization*, pp. 373–404, Springer, 2008.

[5] G. G. Yen and Z. He, "Performance metric ensemble for multiobjective evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 1, pp. 131–144, 2014.

[6] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms — A comparative case study," in *Parallel Problem Solving from Nature (PPSN)*, pp. 292–301, Springer, 1998.

[7] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results.," *Evolutionary Computation*, vol. 8, no. 2, pp. 173 – 195, 2000.

[8] X. Cai, H. Sun, and Z. Fan, "A diversity indicator based on reference vectors for many-objective optimization," *Information Sciences*, vol. 430, pp. 467–486, 2018.

[9] M. Li, S. Yang, and X. Liu, "Diversity comparison of Pareto front approximations in many-objective optimization," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2568–2584, 2014.

[10] J. R. Schott, "Fault tolerant design using single and multicriteria genetic algorithm optimization," Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 1995.

[11] M. Li and X. Yao, "Quality evaluation of solution sets in multiobjective optimisation: A survey," *ACM Computing Surveys*, vol. 52, no. 2, pp. 26:1–26:38, 2019.

[12] A. L. Jaimes and C. A. C. Coello, "Study of preference relations in many-objective optimization," in *Proceedings of the 11th Annual conference on Genetic and Evolutionary Computation 2009*, pp. 611–618, Association for Computing Machinery, July 2009.

[13] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated Hypervolume," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1653–1669, 2007.

[14] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[15] H. Wang, Y. Jin, and X. Yao, "Diversity assessment in many-objective optimization," *IEEE Transactions on Cybernetics*, vol. 47, no. 6, pp. 1510–1522, 2017.

[16] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

[17] J. Bader and E. Zitzler, "HypE: An algorithm for fast Hypervolume-based many-objective optimization," *Evolutionary Computation*, vol. 19, no. 1, pp. 45–76, 2011.

[18] H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "How to specify a reference point in Hypervolume calculation for fair performance comparison," *Evolutionary Computation*, vol. 26, no. 3, pp. 411 – 440, 2018.

[19] P. A. Bosman and D. Thierens, "The balance between proximity and diversity in multiobjective evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 174–188, 2003.

[20] H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima, "Modified distance calculation in generational distance and inverted generational distance," in *8th International Conference on Evolutionary Multi-Criterion Optimization, EMO 2015*, pp. 110–125, 2015.

[21] O. Schütze, X. Esquivel, A. Lara, and C. A. C. Coello, "Using the averaged hausdorff distance as a performance measure in evolutionary multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 4, pp. 504–522, 2012.

[22] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[23] A. Smola, A. Gretton, L. Song, and B. Schölkopf, *A Hilbert Space Embedding for Distributions*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[24] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, "Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications," *Theoretical Computer Science*, vol. 425, pp. 75 – 103, 2012.

[25] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. D. Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, 2003.

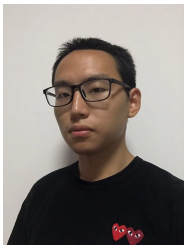
[26] M. P. Hansen and A. Jaszkiewicz, *Evaluating the quality of approximations to the non-dominated set*. IMM, Department of Mathematical Modelling, Technical Universityof Denmark, 1994.

[27] E. Zitzler, D. Brockhoff, and L. Thiele, "The Hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted in-

tegration," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 862–876, Springer, 2007.
[28] W. Cox and L. While, "Improving and extending the HV4D algorithm for calculating Hypervolume exactly," in *Australasian Joint Conference on Artificial Intelligence*, pp. 243–254, Springer, 2016.
[29] C. Priester, K. Narukawa, and T. Rodemann, "A comparison of different algorithms for the calculation of dominated Hypervolumes," in *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pp. 655–662, ACM, 2013.
[30] L. M. S. Russo and A. P. Francisco, "Quick Hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 481–502, 2014.
[31] T. Watanabe, T. Tatsukawa, and A. Oyama, "On the fast Hypervolume calculation method," in *Evolutionary Computation*, pp. 965–969, 2010.
[32] L. While, L. Bradstreet, and L. Barone, "A fast way of calculating exact Hypervolumes," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 1, pp. 86–95, 2012.
[33] A. P. Guerreiro and C. M. Fonseca, "An analysis of the Hypervolume sharpe-ratio indicator," *European Journal of Operational Research*, vol. 283, no. 2, pp. 614–629, 2020.
[34] I. Yevseyeva, A. P. Guerreiro, M. T. M. Emmerich, and C. M. Fonseca, "A portfolio optimization approach to selection in multiobjective evolutionary algorithms," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 672–681, Springer, 2014.
[35] M. Li, S. Yang, and X. Liu, "A performance comparison indicator for Pareto front approximations in many-objective optimization," in *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2015, Madrid, Spain, July 11-15, 2015*, pp. 703–710, 2015.
[36] M. T. M. Emmerich, A. H. Deutz, and I. Yevseyeva, "On reference point free weighted Hypervolume indicators based on desirability functions and their probabilistic interpretation," *Procedia Technology*, vol. 16, pp. 532–541, 2014.
[37] J. G. Falcón-Cardona, M. T. M. Emmerich, and C. A. C. Coello, "On the construction of pareto-compliant quality indicators," in *2019 The Genetic and Evolutionary Computation Conference (GECCO)*, (New York, NY, USA), pp. 2024–2027, 2019.
[38] D. Brockhoff, T. Wagner, and H. Trautmann, "On the properties of the R2 indicator," in *Proceedings of the 14th annual conference on Genetic and evolutionary computation*, pp. 465–472, ACM, 2012.
[39] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. D. Fonseca, "Performance assessment of multiobjective optimizers: an analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, 2003.
[40] H. Ishibuchi, H. Masuda, and Y. Nojima, "A study on performance evaluation ability of a modified inverted generational distance indicator," in *GECCO*, pp. 695–702, ACM, 2015.
[41] G. Rudolph, H. Trautmann, S. Sengupta, and O. Schütze, "Evenly spaced Pareto front approximations for tricriteria problems based on triangulation," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 443–458, Springer, 2013.
[42] A. Menchaca-Mendez and C. A. C. Coello, "GD-MOEA: A new multi-objective evolutionary algorithm based on the generational distance indicator," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 156–170, Springer, 2015.
[43] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
[44] I. Olkin and F. Pukelsheim, "The distance between two random vectors with given dispersion matrices," *Linear Algebra Application*, vol. 48, pp. 257–263, 1982.
[45] R. J.-B. W. R. Tyrrell Rockafellar, *Variational Analysis*. Springer-Verlag, 2005.
[46] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Annals of Mathematical Statistics*, vol. 3, pp. 141–158, 1975.
[47] J. M. Bogoya, A. Vargas, O. Cuate, and O. Schütze, "A (p, q)-averaged hausdorff distance for arbitrary measurable sets," *Mathematical and Computational Applications*, vol. 23, no. 3, p. 51, 2018.
[48] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *NIPS*, pp. 513–520, MIT Press, 2006.
[49] K. M. Borgwardt, A. Gretton, M. J. Rasch, H. P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by Kernel Maximum Mean Discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. 49–57, 2006.
[50] 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 *2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM)*, pp. 170–177, Dec 2014.
[51] M. Li, S. Yang, and X. Liu, "Shift-based density estimation for Pareto-based algorithms in many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 348–365, 2014.
[52] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
[53] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *International Conference on Parallel Problem Solving from Nature*, pp. 832–842, Springer, 2004.
[54] J. Knowles and D. Corne, "The Pareto archived evolution strategy: A new baseline algorithm for Pareto multiobjective optimisation," in *Congress on Evolutionary Computation (CEC99)*, vol. 1, pp. 98–105, 1999.
[55] S. Yang, M. Li, X. Liu, and J. Zheng, "A grid-based evolutionary algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 5, pp. 721–736, 2013.
[56] 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 Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.
[57] X. Cai, Z. Mei, and Z. Fan, "A decomposition-based many-objective evolutionary algorithm with two types of adjustments for direction vectors," *IEEE Transactions on Cybernetics*, vol. 48, no. 8, pp. 2335–2348, 2018.
[58] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary multiobjective optimization*, pp. 105–145, Springer, 2005.
[59] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, 2006.
[60] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM Journal on Optimization*, vol. 8, no. 3, pp. 631–657, 1998.
[61] K. Li, K. Deb, Q. Zhang, and S. Kwong, "An evolutionary many-objective optimization algorithm based on dominance and decomposition.," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 5, pp. 694–716, 2015.
[62] R. Tanabe and H. Ishibuchi, "An analysis of quality indicators using approximated optimal distributions in a three-dimensional objective space," *IEEE Transactions on Evolutionary Computation*, press online,2020.
[63] R. Tanabe and A. S. Fukunaga, "Improving the search performance of shade using linear population size reduction," in *2014 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1658–1665, 2014.
[64] A. Liefooghe and B. Derbel, "A correlation analysis of set quality indicator values in multiobjective optimization," in *2016 The Genetic and Evolutionary Computation Conference (GECCO)*, pp. 581–588, 2016.

Xinye Cai (M'10) received the BSc in information engineering from Huazhong University of Science and Technology, China in 2004, the MSc in electronic engineering from University of York, UK, in 2006 and the PhD in electrical engineering from Kansas State University, U.S, in 2009, respectively.

He is an Associate Professor at the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), China. His current research interests include optimization, machine learning and their applications. He is currently leading an Intelligent Optimization Research Group in NUAA, who has won the Evolutionary Many-Objective Optimization Competition at the Congress of Evolutionary Computation 2017. Dr. Cai is an Associate Editor of Swarm and Evolutionary Computation.

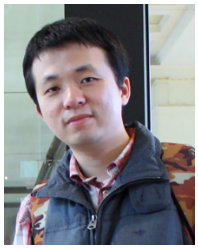


Yushun Xiao is currently pursuing Ph.D. degree in computer science with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include multiobjective optimization, data mining, and kernel method.

Zhenhua Li received the B.S. degree in mathematics from Zhengzhou University, Zhengzhou, China, in 2009 and the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, in 2018. He is currently a Lecturer with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His current research interests include evolutionary optimization and machine learning.

Qi Sun received his Master degree in Software Engineering from the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China, in 2021. He is is currently pursuing Ph.D. degree in computer science with the College of Computer Science and Technology, Southeast University, China. His research interests include multiobjective optimization, bilevel optimization, cloud computing.

Hanchuan Xu received his B.SC., M.Sc. and Ph.D. degrees in computer science and technology from the Harbin Institute of Technology, Harbin, China, in 1999, 2003 and 2011, respectively. He is currently working at the department of Computing, Harbin Institute of Technology, Harbin, China. He is the author or co-author over more than 20 academic papers. His current research interests include service computing, software service engineering, enterprises computing.



Miqing Li received the Ph.D. degree in computer science from the Department of Computer Science, Brunel University London, U.K. in 2015.

Dr Miqing Li is currently a lecturer with the University of Birmingham, U.K. His research is principally on multi-objective optimisation, where he works on developing population-based randomised algorithms (mainly evolutionary algorithms) for both general challenging problems (e.g. many-objective optimisation, constrained optimisation, multi-modal optimisation, robust optimisation, expensive optimisation) and practical multi-objective problems in other fields (e.g. software engineering, system engineering, product disassembly, post-disaster response, neural architecture search, and reinforcement learning). He is the founding chair of IEEE CIS Task Force on Many-Objective Optimisation.

Hisao Ishibuchi (M'93–SM'10–F'14) received the B.S. and M.S. degrees in precision mechanics from Kyoto University, Kyoto, Japan, in 1985 and 1987, respectively, and the Ph.D. degree in computer science from Osaka Prefecture University, Sakai, Osaka, Japan, in 1992. He was with Osaka Prefecture University in 1987-2017. Since 2017, he is a Chair Professor at Southern University of Science and Technology, China. His research interests include fuzzy rule-based classifiers, evolutionary multi-objective and many-objective optimization, memetic algorithms, and evolutionary games.

Dr. Ishibuchi was the IEEE CIS Vice-President for Technical Activities in 2010-2013 and the Editor-in-Chief of the IEEE Computational Intelligence Magazine in 2014-2019. Currently he is an AdCom member of the IEEE CIS (2014-2019, 2021-2023), a Program Co-Chair of SSCI 2022 in Singapore, and a General Co-Chair of WCCI 2024 in Yokohama, Japan.