A Critical Survey of Performance Indices for Multi-Objective Optimisation

Tatsuya Okabe
Honda Research Institute Europe
Carl-Legien Strasse 30,
63073 Offenbach/M, Germany
tatsuya.okabe@honda-ri.de

Yaochu Jin
Honda Research Institute Europe
Carl-Legien Strasse 30,
63073 Offenbach/M, Germany
yaochu.jin@honda-ri.de

Bernhard Sendhoff
Honda Research Institute Europe
Carl-Legien Strasse 30,
63073 Offenbach/M, Germany
bernhard.sendhoff@honda-ri.de

Abstract- A large number of methods for solving multi-objective optimisation (MOO) problems have been developed. To compare these methods rigorously, or to measure the performance of a particular MOO algorithm quantitatively, a variety of performance indices (PIs) have been proposed. This paper provides an overview of the various PIs and attempts to categorise them into a certain number of classes according to their properties. Comparative studies have been conducted using a group of artificial solution sets and a group of solution sets obtained by various MOO solvers to show the advantages and disadvantages of the PIs. The comparative studies show that many PIs may be misleading in that they fail to truly reflect the quality of solution sets. Thus, it may not be a good practice to evaluate the performance of MOO solvers based on PIs only.

1 Introduction

Various performance indices (PIs) [1] for measuring the quality of Pareto-optimal sets have been proposed to compare the performance of different multi-objective optimisation (MOO) algorithms. Usually, the quality of a Pareto-optimal set can be assessed from three aspects. First, the number of Pareto-optimal solutions in the set. Second, the accuracy of the solutions in the set, i.e., the closeness of the solutions to the theoretical Pareto-front. In case the theoretical Pareto-front is not known, only the relative closeness can be obtained. Finally, the distribution and spread of the solutions. Distribution and spread are two very closely related facets, yet they are not completely the same.

Review of PIs can be found in [Coe01, Deb01, Kno02, Tan02, Zit02, Zit03]. In this paper, we aim to provide a survey and categorisation of the PIs, including the most recent results, according to the aspect they account for. Compared to the existing reviews, we concentrate more on the performance of the PIs, i.e., whether the PIs can truly evaluate the quality of a Pareto-optimal set or correctly compare different sets. Empirical studies on a group of artificial solution sets of a simple problem and a group of Pareto-optimal solution sets of a benchmark problem show that many PIs often fail to properly reflect the quality of a solution set. Thus, performance evaluation of different MOO solvers based on PIs only may be misleading.

In this paper, only PIs for measuring static quality of Pareto-optimal solution sets are investigated. PIs for measuring run-time performance [Bae96, Deb02b, Hoo98, Sch95, Tan02, Vel98], PIs where a decision-maker's preferences are involved [Esb96, Han98], and PIs for averaging solution sets from different runs [Fon96, Kno00] are not considered in this paper.

Different notations and definitions have been used and various assumptions have been made in defining PIs. For the sake of clarity, we try to use uniform notations and definitions throughout the paper. Besides, we assume that we are dealing with minimisation problems for all PIs without loss of generality. PIs that have been defined for maximisation problems are modified accordingly.

The rest of the paper is organised as follows. In Section 2, relevant notations and definitions are reviewed. Section 3 describes the cardinality-based PIs, followed by those for measuring accuracy based on distance and volume, respectively. In Section 5, the PIs for distribution, spread or both are given. Empirical comparative studies and discussions are presented in Section 6. The conclusions of the paper are summarised in Section 7.

2 Terminology

Parameter Space and Fitness Space The space, of which the vector of design parameters, $\vec{x}$, is a subset, is known as *Parameter Space*, notated as R^n , where n is the number of design parameters. The design parameters are projected onto a space of dimension M, R^M, by the objective functions, where M is the number of objective functions. This space is termed as *Fitness Space* in evolutionary multi-objective optimisation.

Feasible Region and Infeasible Region The set of all admissible solutions to an optimisation problem is called *Feasible Region*, Ω. In the feasible region, all constraints are

[1] Performance indices are often referred to in the literature as *metrics*. However, *metric* is a well-defined terminology in mathematics and many PIs in MOO do not necessarily satisfy the conditions for a metric. To avoid confusions, we use performance indices (PIs) instead of metrics in this paper.

satisfied. The rest, i.e. $R^n \backslash \Omega$, is called *Infeasible Region*.

Weak Dominance, Dominance and Strong Dominance A design vector $\vec{a}$ is said to *weakly dominate* another vector $\vec{b}$, notated as $\vec{a} \succeq \vec{b}$ or $\vec{b} \preceq \vec{a}$ iff $\forall\ i; f_i(\vec{a}) \leq f_i(\vec{b})$ where f_i is the i-th objective function. Vector $\vec{a}$ *dominates* vector $\vec{b}$, notated as $\vec{a} \succ \vec{b}$ or $\vec{b} \prec \vec{a}$, iff $\forall\ i; f_i(\vec{a}) \leq f_i(\vec{b})$ and $\exists\ i; f_i(\vec{a}) < f_i(\vec{b})$. Vector $\vec{a}$ *strongly dominates* $\vec{b}$, notated as $\vec{a} \succ\succ \vec{b}$ or $\vec{b} \prec\prec \vec{a}$, iff $\forall\ i; f_i(\vec{a}) < f_i(\vec{b})$.

Pareto Optimality A design vector $\vec{x} \in \Omega$ is said to be *Pareto Optimal* if $\nexists \vec{\omega} \in \Omega; \vec{\omega} \succ \vec{x}$. The union of all Pareto optimal design vectors $\vec{x} \in \Omega$ are called *Pareto Optimal Front* or *Pareto Optimal Set*. We denote Pareto Optimal Front in the parameter space as P_{true} and in the fitness space as PF_{true} [Vel99].

Pareto Optimal Solution Set It is generally impossible to get P_{true} or PF_{true}. A finite number of Pareto-optimal solutions for the approximation of P_{true} or PF_{true} are termed as *Pareto Optimal Solution Set* P.

Solution Set and Non-dominated Solution Set The set of solutions found by an optimiser is known as *Solution Set* S. The solutions in S that are not dominated by others in the set define the *Non-dominated Solution Set* S_N. Since in most cases, only non-dominated solutions will be generated, we will not distinguish S and S_N hereafter unless explicitly indicated.

Reference Set In many cases, the Pareto optimal set is unknown. Thus, to evaluate the solution set S, the user needs to specify an artificial or desired solution set. This set is termed as *Reference Set* R.

Good (Utopian) Point and Bad Point A *Good Point*, P_G, is defined as $P_G = [f_1^g, f_2^g, ...]\ \forall i : f_i^g < f_i^s$, and a *Bad Point*, P_B, is defined as $P_B = [f_1^b, f_2^b, ...]\ \forall i : f_i^s < f_i^b$.

3 Cardinality-based PIs

The simplest cardinality-based PI is to count the number of solutions in S, which is termed *Overall Non-dominated Vector Generation ($ONVG$)* [Sch95, Vel99, Vel00]. $|S|$ should not be too small and not too large, where $|.|$ means the number of components in the set.

Overall Non-dominated Vector Generation Ratio ($ONVGR$) has been proposed to calculate the ratio $ONVGR(S,P) = |S|/|P|$ [Vel99, Vel00]. In order to monitor $ONVG$ and $ONVGR$ during the optimisation, *Generational Non-dominated Vector Generation ($GNVG$)* and *Generational Non-dominated Vector Generation Ratio ($GNVGR$)* have also been proposed [Vel99, Vel00]. Both can be expressed mathematically by replacing S by $S(t)$. Here, t is the generation index. Additionally, *Non-dominated Vector Addition (NVA)* is proposed to reflect the change of the number of solutions in each generation, $NVA(S,t) = GNVG(S,t) - GNVG(S,t-1)$ [Vel99, Vel00]. A PI for measuring the percentage of non-dominated solutions (S_N) in S has been considered in [Tan01, Tan02].

Ideally, the obtained Pareto-optimal solutions are accurate, well-distributed and widely spread. If we know a Pareto optimal solution set P or a reference set R, the ratio of found solutions against the ideal or reference set is a reasonable PI. For this purpose, *Error Ratio (ER)* [Vel99, Vel99b] and *Ratio of the Reference Points Found ($C1_R$)* have been proposed [Czy98, Han98]. The ER means the ratio of solutions against S that are not in P and $C1_R$ indicates the ratio of found solutions in R. If $R = P$ and $|R| = |P| = |S|$, then we can get the following relation: $ER(S,P) = 1.0 - C1_R(S,P)$.

In the above PIs, only solutions that are included in R or P can contribute to the PIs. This is very strict and in fact, useful solutions which are not the member of R or P can also be generated, e.g., solutions that are not dominated by those in R or P. In *Ratio of Non-dominated Points by the Reference Set ($C2_R$)*, the number of non-dominated solutions are counted [Han98]. The $C2_R$ index can be expressed as:

$$C2_R(S,R) = |\{s \in S; \nexists r \in R : r \succ s\}|/|S|. \quad (1)$$

When we compare two solution sets obtained by different optimisers, we can use *Coverage of Two Sets (C)* [Zit98, Zit99, Zit99b]:

$$C(S_1,S_2) = |\{s_2 \in S_2; \exists s_1 \in S_1 : s_1 \succeq s_2\}|/|S_2|. \quad (2)$$

It is worth mentioning that $C(S_1,S_2)$ does not have to be equal to $1 - C(S_2,S_1)$. Thus, both $C(S_1,S_2)$ and $C(S_2,S_1)$ should be given. In order to show these results, box plot is often adopted.

It can be found that the definition of $C2_R$ and C are closely related:

$$C(S_1,S_2) = 1.0 - C2_R(S_2,S_1) + \frac{|s_2 \in S_2; \exists s_1 \in S_1 : s1 = s2|}{|S_2|}. \quad (3)$$

The PIs in this category have several drawbacks. For example, no information about the accuracy and distribution of the solutions is contained in $ONVG$, $ONVGR$, $GNVG$, $GNVGR$, and NVA. Besides, all PIs in this class are insensitive to small improvements. For example, as pointed out in [Zit99b] , the C index often fails to indicate the relative quality of two solution sets. An illustration is given in Figure 1. It can be seen that $C(S_1,S_2)$ and $C(S_2,S_1)$ have the same values, i.e. 0.5, although the quality of S_1 and S_2 are significantly different.

4 Accuracy PIs

4.1 Distance-based Accuracy PIs

In this subsection, PIs for accuracy based on distance are shown, which can again be largely divided into two sub-

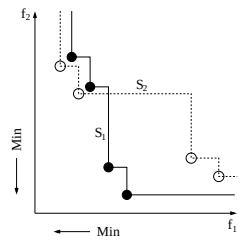


Figure 1: The potential problem of the C index.

classes, i.e. the distance from S to P (or R) or from P (or R) to S. The difference of the two sub-classes originates from the fact that the number of solutions in S and P (or R) may be different.

- The distance from S to P (or R). A natural accuracy PI is to calculate the average distance from S to P. For this purpose, *Generational Distance (GD)* has been suggested [Vel98, Vel99]:

$$GD(S,P) = \left(\sum_{i=1}^{|S|} d_i^q\right)^{1/q} / |S|, \tag{4}$$

$$d_i = \min_{p \in P} \left\{ \sqrt{\sum_{k=1}^{M} (f_k(s_i) - f_k(p))^2} \right\}, \tag{5}$$

  where M is the number of objectives. When $q = 2$, d_i is the Euclidean distance. In [Deb00], a γ index has been proposed, where q equals 1 in Equation (4). It is recommended that a large number of solutions in P should be used [Deb01]. A generalised form of γ, termed as $\mathcal{M}_1^*$, has been suggested in [Zit99b, Zit00], where d_i is replaced by a more generalised distance. Zitzler also has proposed $\mathcal{M}_1$ which is defined on the parameter space.

- The distance from P (or R) to S. An alternative is to calculate the average distance from P or R to S. Since it is generally difficult to get P, R is often used instead of P. Schott proposed *Seven Points Average Distance ($SPAD$)* [Sch95]. As a reference solution set R, the following 7 points are generated: $(0, \max(f_2))$, $(0, 2 * \max(f_2)/3)$, $(0, \max(f_2)/3)$, $(0, 0)$, $(\max(f_1), 0)$, $(\max(f_1)/3, 0)$, $(2 * \max(f_1)/3, 0)$. Thus $SPAD$ is defined as

$$SPAD(R,S) = \frac{1}{|R|} \sum_{i=1}^{|R|} \min_{s \in S} \sqrt{\sum_{k=1}^{M} |f_k(r_i) - f_k(s)|^2}. \tag{6}$$

  Similar to GD, *Average Distance from Reference Set ($DIST1$)* has been suggested [Czy98], which can be expressed as:

$$Dist1(R,S) = \frac{1}{|R|} \sum_{i=1}^{|R|} \left\{ \min_{s \in S} \{c(r_i, s)\} \right\}, \tag{7}$$

$$c(r_i, s) = \max_{j=1,2,\ldots,M} \{0, w_j \left(f_j(s) - f_j(r_i)\right)\}, \tag{8}$$

  where w_j is the reciprocal number of range of objective f_j in the reference set R. It differs from GD in how the distance is calculated, from where the distance is calculated and whether normalisation is used or not.

  Instead of calculating the average distance, the *Maximum Pareto Front Error ($MPFE$)* [Vel99], considers the maximal approximation error:

$$MPFE(P,S) = \max_{p \in P} \sqrt{\min_{s \in S} \left(\sum_{k=1}^{M} |f_k(s) - f_k(p)|^2 \right)} \tag{9}$$

  Similarly, the *Worst Distance from Reference Set ($DIST2$)* has also been proposed in [Czy98]:

$$DIST2(R,S) = \max_{r \in R} \left\{ \min_{s \in S} \{c(s, r)\} \right\}. \tag{10}$$

  This PI gives information about the biggest distance from $r \in R$ to the closest solution in S.

The PIs in this class consider the distance between P (or R) and S. A prerequisite is that either P or R must be given, which may not be trivial for some real-world applications. Finally, it should be pointed out that although they are categorised as accuracy PIs, they can also be influenced by the spread and distribution of P and R, which will be discussed later.

4.2 Volume-based Accuracy PIs

In this class, we present the PIs for accuracy based on volume, which means the size of the area that is dominated by S. The basic idea is that the larger the area the solutions can dominate in the fitness space, the better [Zit98, Zit99b].

Zitzler proposed *Size of the Dominated Space H* [2], often termed as *Hyperarea* or *Hypervolume* [Zit98, Zit99], as shown in Figure 2. A normalised H has also been defined by $\prod_{i=1}^{M}(f_i^{max} - f_i^{min})$. Veldhuizen [Vel99] use $H(P, O')$ to normalise H. This PI is called *Hyperarea Ratio (HR)*, which gives us the information about the difference between S and P.

To address the problem of the C index, the *Coverage Difference of Two Set (D)* has been proposed [Zit99b], refer to Figure 3. The index $D(S_1, S_2)$ quantifies the size of the area which is dominated by S_1 but is not by S_2.

The *Hyperarea Difference (HD)* has been suggested in [Wu01]. Since it is difficult to know P, good point P_G and

[2] Originally, Zitzler used S, however to maintain a consistent notation, we chose H.

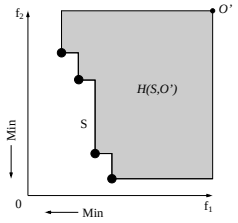


Figure 2: The definition of the H index in the minimisation problem. H is the area generated by Solution Set S and the defined origin O', which needs to be specified.

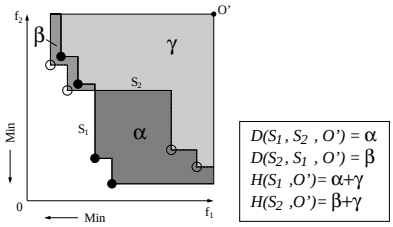


Figure 3: The definition of D and H indices.

bad point P_B are used instead of P. The dominated area is calculated on the normalised fitness by P_G and P_B.

The drawbacks of the PIs in this class is that O' needs to be given and that convex parts of the Pareto front are more preferable than concave parts [Vel99]. If O' is far from S, their sensitivity to improvement decreases.

5 Distribution and Spread PIs

5.1 Distribution PIs

Roughly speaking, PIs for distribution can be calculated based on distance or niche.

- Distance-based. The first PI to be described in this sub-class is *Spacing (SP)* proposed by Schott [Sch95]:

$$SP(S) = \sqrt{\frac{1}{|S-1|}\sum_{i=1}^{|S|}(d_i - \bar{d})^2}, \tag{11}$$

$$d_i = \min_{s_k \in S \wedge s_k \neq s_i} \sum_{m=1}^{M} |f_m(s_i) - f_m(s_k)|, \tag{12}$$

where $\bar{d}$ is the average of d_i. It is worth noticing that the distance is calculated by the sum of absolute differences along each axis. Additionally, only the shortest distance from each point is used without sorting. Furthermore, the distance is summed up from $i = 1$ to $i = |S|$. However, this PI may be misleading. Consider the following solution set, $A(0,10), B(1,9), C(9,1), D(10,0)$. To calculate d_i in SP, the distance between *A* and *B*, and *C* and *D* are used twice. Nevertheless, the distance between *B* and *C* is not used.

A more natural distance-based distribution PI is Δ' proposed in [Deb00] [3] At first, the Euclidean distance d_i between consecutive solutions in S are calculated. Therefore, the average of distance $\bar{d}$ is calculated. Finally, $\Delta'(S)$ is calculated according to the following equation:

$$\Delta'(S) = \sum_{i=1}^{|S|-1} \frac{|d_i - \bar{d}|}{|S|-1}. \tag{13}$$

- Niche-based. The PIs in this sub-class are based on the concept of niching. The number of solutions in each niche is used to calculate the PIs.

Zitzler proposed the $\mathcal{M}_2^*$ index which takes into account both the distribution and the number of non-dominated solutions [Zit99b, Zit00]:

$$\mathcal{M}_2^*(S) = \frac{1}{|S|-1}\sum_{s_1 \in S} \mid \{s_2 \in S \mid ||s_1 - s_2|| > \sigma\} \mid, \tag{14}$$

where σ is the niche radius. This PI is defined on the fitness space. A PI has also been proposed on the parameter space, which is denoted as $\mathcal{M}_2$.

The *Uniform Distribution (UD)* index has been suggested to measure the distribution of non-dominated solutions [Tan01, Tan02] as follows:

$$UD(S) = 1/(1 + D_{nc}), \tag{15}$$

where D_{nc} is the standard deviation of niche count of the overall set of non-dominated solutions S:

$$D_{nc} = \sqrt{\frac{\sum_{i=1}^{|S|}\left(nc(s_i) - \frac{1}{|S|}\sum_{i=1}^{|S|} nc(s_i)\right)^2}{|S|-1}}, \tag{16}$$

where $nc(s_i)$ denotes the niche count of i-th solution in S: $nc(s_i) = |s_j \in S : ||s_i - s_j|| < \sigma| - 1$.

It is found that both $\mathcal{M}_2^*$ and UD may be misleading. Consider again the example as in SP. If σ is set to 0.5, then $\mathcal{M}_2^*$ gets a value of 4, and UD equals 1.0, both reaching the best value. However, the distribution is not good at all. According to our evaluation, Δ' seems to be only one that can reflect the distribution correctly. Unfortunately, Δ' cannot be used for MOO problems with more than 2 objectives because consecutive sorting is involved.

[3] Originally, they have introduced the PI including spread information. However, they have also explained the PI without spread. See the footnote in [Deb00].

5.2 Spread PIs

In this subsection, we explain the PIs for measuring the spread of S.

Maximum Spread ($\mathcal{M}_3^*$) proposed in [Zit99b, Zit00] shows the distance between the boundary solutions in S. They have also proposed $\mathcal{M}_3$ on the parameter space. In [Deb01], normalised maximum spread $\mathcal{M}_3^*$ has been introduced, where the fitness is normalised with the range in P before calculating $\mathcal{M}_3^*$. A similar PI called *Overall Pareto Spread (OS)* is adopted in [Wu01], which can be expressed as follows:

$$OS(S, P_G, P_B) = \prod_{i=1}^{M} OS_k(S, P_G, P_B), \tag{17}$$

where, $OS_k(S, P_G, P_B)$ is called k^{th} *Objective Pareto Spread (OS_K)*:

$$OS_K(S, P_G, P_B) = \frac{|\max_{s \in S} f_k(s) - \min_{s \in S} f_k(s)|}{|f_k(P_B) - f_k(P_G)|}. \tag{18}$$

Maximum Spread and Overall Pareto Spread are conceptually very similar. At first, the hypervolume defined by the minimum value and the maximum value along each axis is calculated. For $\mathcal{M}_3^*$, the length of the diagonal line is assigned as $\mathcal{M}_3^*$. For OS, the size of the hypervolume is calculated. This value is divided by the total size of the hypervolume defined by P_G and P_B. Thus, only the representative values of the hypervolume are different.

5.3 Distribution and Spread PIs

The PIs in this class include information about both spread and distribution, which can be based on distance, niching or entropy.

- Distance-based. The Δ' PI has been extended to include spread information, which is now termed Δ index [Deb01, Deb02].

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{|S|-1} |d_i - \bar{d}|}{d_f + d_l + (|S| - 1)\bar{d}}, \tag{19}$$

  where d_f and d_l are the Euclidean distances between the extreme solutions and the boundary solutions of S. $\bar{d}$ is the average of all distances $d_i, i \in [1, |S| - 1]$.

- Niche-based. A PI called *Chi-Square-Like Deviation Measure* ι has been suggested [Deb89, Sri94].

$$\iota = \sqrt{\sum_{i=1}^{|P|+1} \left(\frac{n_i - \bar{n}_i}{\sigma_i}\right)^2}, \tag{20}$$

  where n_i is the number of solutions in S that are located in the circle generated around each solution in P, $\bar{n}_i$ is the expected number of solutions when the ideal solution set can be obtained, and σ_i for $i = 1, 2, ..., |P|$ is suggested as follows: $\sigma_i^2 = \bar{n}_i \left(1 - \frac{1}{|P|}\right)$. The $(|P| + 1)$-th sub-region is defined as the other feasible region, which is the shadowed region in Figure 4. Here, $\bar{n}_{|P|+1}$ and $\sigma_{|P|+1}$ are defined as: $\bar{n}_{|P|+1} = 0$ and $\sigma_{|P|+1}^2 = \sum_{i=1}^{|P|} \sigma_i^2 = |S| \left(1 - \frac{1}{|P|}\right)$.

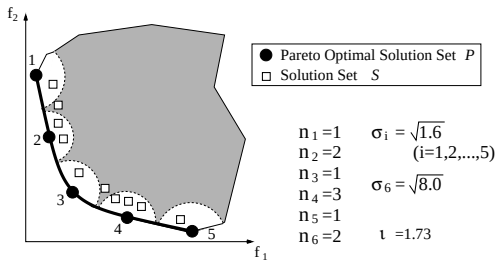


Figure 4: The image of ι.

A PI termed *Number of Distinct Choices ($NDC\mu$)* has been suggested in [Wu01]. A normalised fitness space defined by P_G and P_B is divided into a number of small hyper-squares with the length μ; $(0 < \mu \leq 1)$. The number of squares with solutions in S is then counted. In addition, a PI called *Cluster ($CL\mu$)* to show the average number of solutions in each small hyper-square has been also considered in [Wu01]:

$$CL\mu(S, P_G, P_B, \mu) = \frac{|S|}{NDC\mu(S, P_G, P_B, \mu)}. \tag{21}$$

- Entropy-based. A PI for spread, EN, based on Shannon's entropy has been proposed in [Far02]. The basic idea is that each solution point provides some information about its neighbourhood modelled by a Gaussian distribution. A *Density Function* has also been calculated by the sum of all Gaussian distributions from all solution points. The peaks and valleys of density function correspond to the dense areas and the sparse areas, respectively. A desirable solution set should have a "uniform" [4] density function which was evaluated with Shannon's entropy.

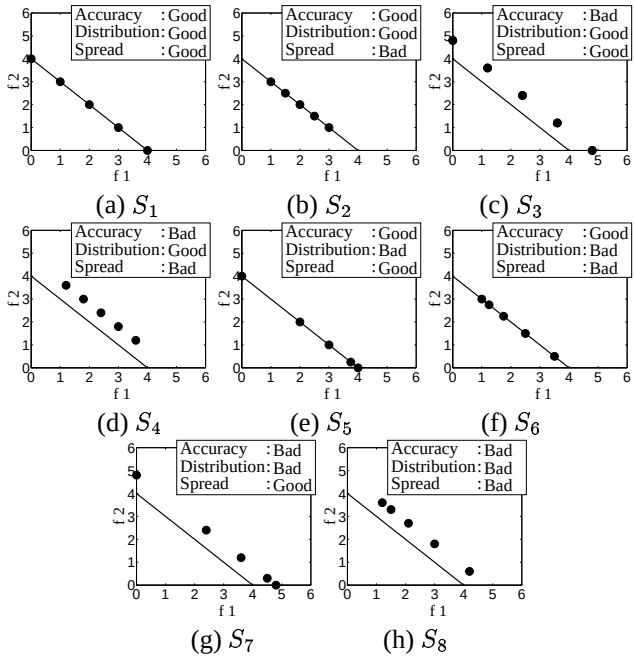
As mentioned in Section 5.1, Δ works only for 2 objectives. In contrast, the PIs ι, NDC_μ, and CL_μ can easily be extended to more than 2 objectives. But the distribution in the same niche cannot be taken into account. In [Far03], the EN is extended to deal with problems with 3 objectives and the boundary influence is compensated. One problem is the distortion of the distribution by Gram-Schmidt orthogonalisation, refer to [Far03].

[4] Originally, the term "flat" was used for the description of the density function, however, statistically the term "uniform" seems to be more appropriate.

6 Empirical Comparisons and Discussions

6.1 Artificial Solution Sets

In this section, we compare the results of PIs using a number of artificial solution sets, which are shown in Figure 5. In the figure, the solid line denotes the artificial Pareto front. For clarity, the properties of each solution set are shown together with the solutions.



(a) S_1 (b) S_2 (c) S_3

(d) S_4 (e) S_5 (f) S_6

(g) S_7 (h) S_8

Figure 5: Artificial solution sets for empirical comparison.

In the calculation, S_1 is used as P and R. Additionally, P_G, P_B and O' are set to $(-1,-1)$, $(5,5)$ and $(5,5)$, respectively. Niche radius is selected to be 0.7 and μ in $NDCU$ and CLU is 0.2. The external points in Δ are $(0,4)$ and $(4,0)$. The calculation results are shown in Table 1. In the table, the quality ranks according to the corresponding PI are listed, i.e. if the value is 1, this solution set is evaluated as the best according to the corresponding PI, and a rank of 8 is the worst.

It can be seen that the cardinality-based PIs fail to determine the correct order of S_3, S_4, S_7, and S_8. Furthermore, the order of other solution sets cannot be correctly determined by $C2_R$. As mentioned before, these PIs cannot distinguish solution sets which are far from P and thus a correct ranking cannot be achieved for these solution sets. From the accuracy point of view, S_1, S_2, S_5 and S_6 are better than other sets. In this sense, the accuracy PIs are performing well. As previously discussed, PIs for accuracy are also influenced by the distribution and spread of P or R. If the distribution and the spread of S are similar to P or R,

Table 1: Empirical comparison of several PIs with the artificial and simple solution sets. The ranks of solution sets are shown. Tied PIs have the same ranks.

PIs	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
ER,$C1_R$	1	3	5	5	2	4	5	5
$C2_R$	1	1	5	5	1	1	5	5
GD,γ,$\mathcal{M}_1^*$	1	3	7	7	2	4	6	5
$SPAD$	1	6	3	8	2	5	4	7
$DIST1$	1	3	5	8	2	4	6	7
$MPFE$	1	3	5	5	2	3	5	5
$DIST2$	1	3	2	6	3	3	8	6
H,HR,HD	1	4	7	6	2	3	8	5
SP,Δ'	1	1	1	1	7	5	8	6
$\mathcal{M}_2^*$,UD	1	1	1	1	5	5	5	5
$\mathcal{M}_3^*$,OS,OS_1,OS_2	3	8	1	7	3	6	1	5
Δ	1	4	2	3	4	8	7	6
ι	1	3	3	3	2	7	3	7
NDC_μ,CL_μ	1	1	1	8	1	5	5	5
EN	4	8	1	6	3	7	2	5

the values of these PIs become better.

As recommended in [Deb00], the number of solutions in P should be large enough. To investigate the influence of number of solutions in P on the γ index, additional artificial solutions evenly distributed on the artificial Pareto Front are generated. The computation results are shown in Figure 6. From these results, it can be seen that a larger number of solutions in P or R reduce the influence of the spread and the distribution.

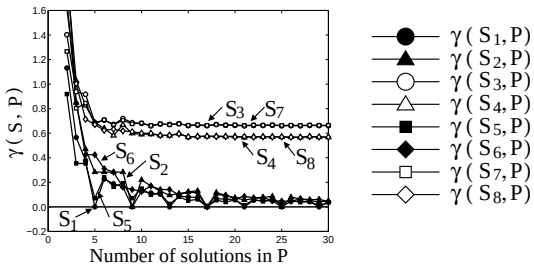


Figure 6: γ index against the number of solutions in P.

In comparing the group ($S4$, $S8$) with the group ($S3$, $S7$), the influence of the boundary points still exist because of the distance between both boundary points on P and S. Although $SPAD$ is categorised as an accuracy PI, it turns outs that $SPAD$ tends to assign a higher rank to solution sets with a wider spread than those with a better accuracy. With regard to the volume-based accuracy PIs, the group (S_1, S_2, S_5, S_6) is better than the other solutions. Note that these PIs also include the influence of distribution and spread. The influence is more complex than the distance-based accuracy PIs. The strength of the influence changes according to the location of O' and the distribution of solu-

tions. From the results, it seems that SP [5] and Δ' are able to reflect the distribution of the solution sets correctly. The rank obtained from Δ seems confusing. It seems that PIs of S_1, S_2, S_3, and S_4 perform better than others. However, it is very hard to distinguish the influence of distribution from that of spread. The PIs based on the niche, i.e. ι, $\mathcal{M}_2^*$ and UD, show a lot of incomparable cases. This is caused by the insensitivity of the niche count. For one solution set, the niche count can only take the value of zero or one.

The ranks from the PIs for the spread show a good agreement with the test cases. However, a high value of these PIs does not imply that the solution set has high quality from the accuracy and distribution point of view. By comparing S_1 with S_3, this is clear.

It can also be found that the results of NDC_μ and CL_μ are very sensitive to μ. If μ is too large or too small, many solution sets will become incomparable. However, to specify an appropriate μ is not easy. From the result of EN, it can be seen that EN prefers widely spread solutions. The group (S_1,S_3,S_5,S_7) has higher values than others, although the order is confusing.

6.2 Real Solution Sets

In this section, we compare solution sets for the 50-dimensional Schaffer's function generated by seven widely used MOO optimisers. Details of the MOO optimisers are omitted here because they are not of interest in the present paper. The obtained solution sets are shown in Figure 7.

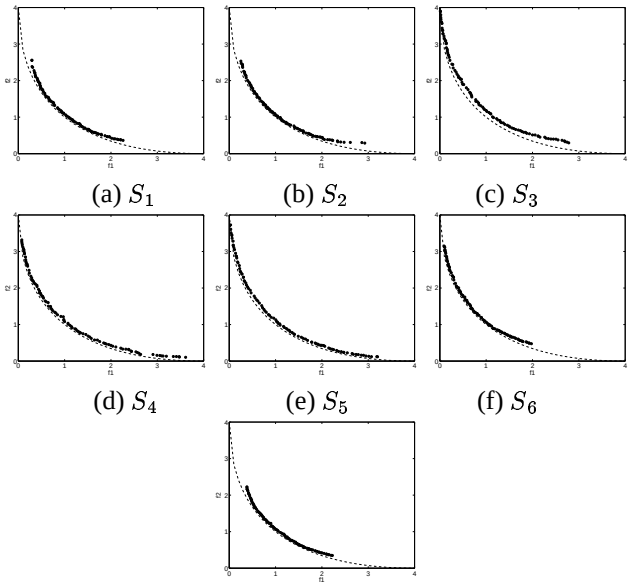
With these solution sets, we calculate the value of PIs. According to the values, we determine the rank of the optimisers. The results are shown in Table 2. Similarly, a rank of 1 indicates that the performance of the optimiser is best, and a rank of 7 that it is worst, according to the PI used. The following parameters have been used in the calculation:

P and R	:	500 ideal solutions for ER, $C1_R$, GD, γ, $\mathcal{M}_1^*$, HR 100 ideal solutions for $ONVGR$, $C2_R$, $DIST1$, $MPFE$, $DIST2$, ι
Good point	:	$P_G = (-0.5, -0.5)$
Bad point	:	$P_B = (5.0, 5.0)$
O' Point	:	$O' = (5.0, 5.0)$
External points	:	$P_E = (0.0, 4.0), (4.0, 0.0)$
Number of division	:	$\mu = 0.01$
Niche radius	:	$\sigma = 0.0325$.

First of all, it is very difficult to determine which optimiser is best because the ranks of a solution set are completely different based on different PIs.

In general, PIs that count the number of solutions, such as $ONGV$ and $ONVGR$ are too sensitive to be practical. No useful information has been revealed based on ER, $C1_R$ and $C2_R$. Neither can one draw any conclusion on the solution quality according to the PIs for accuracy based on the distance, $SPAD$ and $DIST1$, which are apparently strongly influenced by the spread of solution sets. Besides, $MPFE$ is also strongly influenced by the most-right solution and other solutions have no influence at all. The PIs for accuracy based on the volume tend to prefer more spread solution sets than more accurate solution sets.

[5] Although we pointed out that SP is misleading in Section 5.1, we can not observe the problem with these solution sets.



(a) S_1 (b) S_2 (c) S_3 (d) S_4 (e) S_5 (f) S_6 (g) S_7

Figure 7: Real solution sets for the test function.

The PIs for distribution can be divided into two types. The SP and Δ', based on the distance, seem to prefer the less spread solution sets because they produce smaller values if a solution set shrinks. In contrast, $\mathcal{M}_2^*$ and UD seem to prefer more spread solutions because the niche radius is fixed. Thus, if a PI is dedicated to the uniformity of the distribution, then the niche radius should be scaled to the spread of the solutions. The PIs for spread seem to reflect spread of the solution set correctly. Wider spread solutions have a better value and vice versa. The PIs for distribution and spread, i.e. Δ, ι, NDC_μ, CL_μ and EN, seem to prefer wider spread solution sets because the influence of the end points is very large.

7 Conclusions

From our empirical evaluations and discussions, it can be seen that no single existing PI is able to account for all aspects of the quality of solution sets for MOO problems. Some PIs are even quite misleading in certain cases. Thus, one should be very careful in evaluating and comparing the performance of MOO optimisers according to performance

Table 2: Empirical comparison of several PIs with real solution sets. The ranks of solution sets are shown. Tied PIs have the same ranks.

PIs	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$ONVG, ONVGR$	7	1	1	1	6	5	1
ER,$C1_R$,$C2_R$	1	1	1	1	1	1	1
GD	5	6	7	4	3	1	2
γ,$\mathcal{M}_1^*$	5	3	7	6	4	1	2
$SPAD$	6	4	3	1	2	5	7
$DIST1$	6	4	3	2	1	5	7
$MPFE$	5	6	7	4	1	2	3
$DIST2$	5	3	4	1	2	7	6
H,HR,HD	6	4	3	2	1	5	7
SP	3	2	5	6	7	4	1
Δ'	4	2	5	7	6	3	1
$\mathcal{M}_2^*$	7	3	2	1	6	5	4
UD	4	7	1	3	2	5	6
$\mathcal{M}_3^*$,OS	6	4	3	2	1	5	7
OS_1	5	4	3	1	2	6	7
OS_2	6	5	2	3	1	4	7
Δ	7	4	1	3	2	5	6
ι	3	5	1	6	2	4	6
NDC_μ	6	4	1	3	2	5	6
CL_μ	6	4	2	3	1	5	7
EN	6	5	1	3	2	4	7

indices only. This is particularly important when there is little information about the shape of the true Pareto-optimal front, as in most real-world applications.

We should also note that some of the existing PIs work only for bi-objective optimisation problems. The following PIs can theoretically be applied to problems with more than three objectives: cardinality-based PIs, distance-based accuracy PIs except for $SPAD$, volume-based accuracy PIs, niche-based distribution PIs, spread PIs, and entropy-based PIs. For MOO problems with higher number of objectives, a proper performance index for the distribution based measures seems more difficult. Although measures of uniformity can easily be defined in higher dimensions for a finite number of data points, they are not unique. Thus, it becomes harder - even for humans - to define what type of definition is favoured.

Bibliography

[Bae96] T. Bäck (1996) "Evolutionary Algorithms in Theory and Practice", Oxford University Press.

[Coe01] C. A. Coello Coello, D. A. Van Veldhuizen and G. B. Lamont (2001) "Evolutionary Algorithms for Solving Multi-Objective Problems", Kluwer Academic Publishers.

[Czy98] P. Czyzak and A. Jaszkiewicz (1998) "Pareto Simulated Annealing - A Metaheuristic Technique for Multiple Objective Combinatorial Optimization", Journal of Multi-Criteria Decision Analysis, vol. 7, p.p. 34-47.

[Deb89] K. Deb (1989) "Genetic Algorithms in Multi-Modal Function Optimization", Master Thesis, Tuscaloosa, AL: University of Alabama.

[Deb00] K. Deb, S. Agrawal, A. Pratap and T. Meyarivan (2000) "A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II", Parallel Problem Solving from Nature - PPSN VI, p.p. 849-858.

[Deb01] K. Deb (2001) "Multi-Objective Optimization using Evolutionary Algorithms", John Wiley & Sons, LTD.

[Deb02] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan (2002) "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II", IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, p.p. 182-197.

[Deb02b] K. Deb and S. Jain (2002) "Running Performance Metrics for Evolutionary Multi-Objective Optimization", Indian Institute of Technology, Kanpur: Kanpur Genetic Algorithms Laboratory (KanGAL), Technical Report 2002004.

[Esb96] H. Esbensen and E. S. Kuh (1996) "Design Space Exploration Using the Genetic Algorithm", IEEE International Symposium on Circuits and Systems (ISCAS'96), vol. 4, p.p. 500-503.

[Far02] A. Farhang-Mehr and S. Azarm (2002) "Diversity Assessment of Pareto Optimal Solution Sets: An Entropy Approach", Proceedings of IEEE Congress on Evolutionary Computation - CEC 2002, p.p. 723-728.

[Far03] A. Farhang-Mehr and S. Azarm (2003) "An Information-Theoretic Metric for Assessing Multi-Objective Optimization Solution Set Quality", ASME, Transactions, Journal of Mechanical Design (to appeare).

[Fon96] C. M. Fonseca and P. J. Fleming (1996) "On the Performance Assessment and Comparison of Stochastic Multiobjective Optimizers", Parallel Problem Solving from Nature - PPSN IV, p.p. 584-593.

[Han98] M. P. Hansen and A. Jaszkiewicz (1998) "Evaluating the Quality of Approximations to the Non-dominated Set", Institute of Mathematical Modeling, Technical University of Denmark, Denmark, IMM Technical Report IMM-REP-1998-7.

[Hoo98] H. H. Hoos and T. Stuetzle (1998) "Evaluating Las Vegas Algorithms - Pitfalls and Remedies", Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, p.p. 238-245.

[Kno00] J. D. Knowles, D. W. Corne and M. J. Oates (2000) "On the Assesment of Multiobjective Approaches to the Adaptive Distributed Database Management Problem", Parallel Problem Solving from Nature - PPSN IV, p.p. 869-878.

[Kno02] J. Knowles and D. Corne (2002) "On Metrics for Comparing Nondominated Sets", Proceedings of IEEE Congress on Evolutionary Computation - CEC 2002, p.p. 711-716.

[Sch95] J. R. Schott (1995) "Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization", Master Thesis, Boston, MA: Department of Aeronautics and Astronautics, Massachusetts Institute of Technology.

[Sri94] N. Srinivas and K. Deb (1994) "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms", Evolutionary Computation, vol. 2, no. 3, p.p. 221-248.

[Tan01] K. C. Tan, T. H. Lee and E. F. Khor (2001) "Evolutionary Algorithms for Multi-Objective Optimization: Perfomance Assesments and Comparisons", Proceedings of IEEE Congress on Evolutionary Computation - CEC 2001, vol. 2, p.p. 979-986.

[Tan02] K. C. Tan, T. H. Lee and E. F. Khor (2002) "Evolutionary Algorithms for Multi-Objective Optimization: Perfomance Assesments and Comparisons", Artificial Intelligence Review, 17, p.p. 253-290.

[Vel98] D. A. Van Veldhuizen and G. B. Lamont (1998) "Evolutionary Computation and Convergence to a Pareto Front", Late Breaking Papers on the Genetic Programmming 1998 Conference, p.p. 221-228.

[Vel99] D. A. Van Veldhuizen (1999) "Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations", Faculty of the Graduate School of Engineering of the Air Force Institute of Technology, Air University, Dissertation AFIT/DS/ENG/99-01.

[Vel99b] D. A. Van Veldhuizen and G. B. Lamont (1999) "Multiobjective Evolutionary Algorithm Test Suites", Proceedings of the 1999 ACM Symposium on Applied Computing, p.p. 351-357.

[Vel00] D. A. Van Veldhuizen and G. B. Lamont (2000) "On Measuring Multiobjective Evolutionary Algorithm Performance", Proceedings of IEEE Congress on Evolutionary Computation - CEC 2000, p.p. 204-211.

[Wu01] J. Wu and S. Azarm (2001) "Metrics for Quality Assessment of a Multiobjective Design Optimization Solution Set", Transactions of the ASME, Journal of Mechanical Design, vol. 123, p.p. 18-25.

[Zit98] E. Zitzler and L. Thiele (1998) "Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study", Parallel Problem Solving from Nature - PPSN V, p.p. 292-301.

[Zit99] E. Zitzler and L. Thiele (1999) "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach", IEEE Transactions on Evolutionary Computation, vol. 3, no. 4, p.p. 257-271.

[Zit99b] E. Zitzler (1999) "Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications", Zuerich, Switzerland: Swiss Federal Institute of Technology (ETH), Dissertation ETH No. 13398.

[Zit00] E. Zitzler, K. Deb and L. Thiele (2000) "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results", Evolutionary Computation, vol. 8, no. 2, p.p. 173-195.

[Zit02] E. Zitzler, M. Laumanns, L. Thiele, C. M. Fonseca and V. Grunert da Fonseca (2002) "Why Quality Assessment of Multiobjective Optimizers Is Difficult", Proceedings of the Genetic and Evolutionary Computation Conference - GECCO 2002, p.p. 666-673.

[Zit03] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca and V. Gurnert da Fonseca (2003) "Performance Assessment of Multiobjective Optimizers: An Analysis and Review", IEEE Transactions on Evolutionary Computation 7(2), p.p. 117-132.