This is a repository copy of *A new diversity performance indicator for many-objective optimisation problems*.

A new Diversity Performance Indicator for Many - Objective Optimisation Problems

Kai Eivind Wu
Department of ACSE
University of Sheffield
Sheffield, UK
kewu1@sheffield.ac.uk

George Panoutsos
Department of ACSE
University of Sheffield
Sheffield, UK
g.panoutsos@sheffield.ac.uk

Abstract— Current performance indicators for assessing the diversity of many-objective optimisation approximations are often underperforming as the number of objectives increases, particularly for complex optimisation problems. In this article, a new pure unary diversity indicator is proposed, Inverse Ratio of Net Avertence angle (IRNA), which is formulated by minimising the sum of the included angles between approximation set and a set of reference vectors. It is achieved by effectively rotating the reference vectors system in all dimensions simultaneously with an optimised spatial angle. Any potential systematic bias in included angles is removed, and the highest possible diversity score of a solution set is obtained. Numerical results from evaluating performance on synthetic solutions on a unit simplex plane and benchmark functions of MaF show that the proposed performance indicator IRNA is more sensitive to capturing diversity changes as the number of objectives increases compared to other popular indicators.

Keywords— Many-objective optimisation, Performance indicator, Diversity, Reference vectors, Benchmark testing

I. INTRODUCTION

Various metaheuristic methodologies have been developed to address the research challenge of many-objective optimisation problems (MaOP) [1]. It appears that there is no single methodology superior to all the rest in solving the plethora of MaOPs [1]. Performance Indicators (PIs) are crucial for assessing the quality of MaOP algorithms' approximations, as mathematical guarantees for diversity and convergence properties and global optima are incredibly challenging to derive [2]. PIs may also play a central role in guiding the evolution of the solution set toward optimality. It is most commonly accepted that the quality of the Pareto Front (PF) approximation is determined by its three major characteristics: convergence, distribution (or evenness) and extent (or spread), where the last two jointly describe its diversity property [2]–[4]. Convergence indicators have been formulated, for instance, by measuring and comparing the distance between individuals and the nominal ideal point, which is an auxiliary point at which all objectives are at their minimum value. Diversity indicators are more challenging to develop since PF's actual shape and distribution are unknown a priori and challenging to be described via a limited number of discrete points in a high dimensional objective space.

Several design principles, such as cardinality, distance, hypervolume, dominance, and included angles, have been utilised to develop PIs [4]. PIs based on included angles are formulated by measuring the included angles between the vectors of candidate solutions and reference vectors [3], [5]. Pure diversity indicators have been developed in this way [3], [5], [6]. One major challenge remains that candidate solutions may have a systematic bias in included angles with reference vectors which may cause inconsistent scores of indicators and hinders derivation of meaningful metrics. This challenge could be mitigated by minimising the included angles between reference vectors and candidates through rotation of the reference vector system to remove eventual systematic bias in data. The proposed PI's underpinning idea is to search for an optimised diversity score for each of two competing approximation sets by rotating the reference vector system with an optimal angle; hence the systematic bias in both data sets can be removed. A new unary diversity indicator is proposed. A reference vector-based pure diversity indicator is expressed with the Inverse Ratio of Net Avertence angles (IRNA), as detailed in Section 3. Moreover, alternative assessment methods for comparing and contrasting efficacy and efficiency among several performance indicators in terms of monotonicity and sensitivity are suggested and demonstrated numerically through MaF benchmark functions in 3, 5, 7 and 10 objectives.

The rest of the paper is organised as follows: Section II includes an overview of state of the art in performance indicators and the creation of reference vectors in MaOP problems, focusing on PIs using reference vectors. Section III provides a detailed description of the proposed IRNA for evaluation of the diversity of MaOP cases. Section IV includes numerical studies of IRNA, where the effectiveness is assessed against two well-established indicators in dealing with synthetic and benchmark problems. Section V concludes the article.

II. PERFORMANCE INDICATORS

A. Performance Indicators in many-objective optimisation problems

Numerous performance indicators have been designed for the assessment of MaOP approximation sets. Earlier comprehensive surveys on PIs and their properties can be found in the literature [2, 7-8].

PIs in MaOP can be grouped into three main categories: those which primarily evaluate convergence [9], those mainly assessing diversity [10], and those measuring both convergence and diversity simultaneously [11].

Representative PIs for assessing convergence only include the Generational distance (GD) [12] and GD$^+$ [13], which are distance-based PIs. PI for diversity is further divided into subgroups. Those primarily evaluating spread, those for mainly measuring uniformity and those for assessing both spread and uniformity. An example of commonly adopted PIs of this kind is Δ_p [14]. Some are

newly developed, e.g. PD [2] and Coverage over Pareto Front (CPF) [3].

PIs for measuring the combined performance of convergence and diversity are most commonly used, e.g., Epsilon indicator (ϵ-indicator) [15], Inverted generational distance (IGD), IGD$^+$ [13], Hypervolume (HV) [16], and *R*-metric [17].

Deb et al. [6] suggest a Δ metric, measuring the extent of spread and distribution achieved by the approximations. The main drawback of the indicator is the computational cost when scaling up in higher dimensions (for MaOPs).

Mostaghim and Teich [11] propose Sigma Diversity Metric (SDM) to evaluate diversity by calculating angular positions of solutions in the objective space. A limitation of this method is that there is a systematic bias in angular locations of the approximations relative to σ reference lines.

Deb and Jain [18] suggest a diversity measure (DM), which measures the diversity against a reference set. Here, solutions are projected on an $(m-1)$-dimensional hyperplane with hyper-boxes. The indicator value is proportional to the number of hyper-boxes containing both a reference solution and a candidate solution. Several challenges exist [19], including its dependence on a reference set, computationally high cost in creating hyper-boxes, and the determination of neighbouring hyper-boxes in high dimensional objective space.

Li et al. [19] propose a pure diversity comparison indicator (DCI) to assess the relative diversity of two or more Pareto front approximations in many-objective optimisation by counting the number of solutions in a grid covering the objective space. No reference set is required for DCI calculation. But the method is sensitive to the number of divisions chosen in the grid. Li et al. [20] suggest a parameter-less performance comparison indicator (PCI) to assess both convergence and diversity of approximations using a reference set constructed by dividing the approximation set into clusters and calculating the minimum moves of solutions to dominate these clusters weakly. The merit of PCI is that it does not require any prior reference set. However, PCI depends on determining the number of data groups utilised in the evaluation, which leads to a change of indicator value.

Using reference vectors to create a diversity score for contrasting two competing approximation sets has been implemented successfully in the past [3], [5], [7].

Cai et al. [6] define a diversity indicator (DIR) using reference vectors by identifying candidate solutions' systematic deviations away from the reference vectors. The mean values and variances of the so-called coverage vectors that stores the number of reference vectors linked to each candidate solution are found. A major demerit of the method is its inability to deal with local clusters of candidate solutions, which cannot be solved by increasing the number of reference vectors, which largely influences the indicator value. Moreover, the regional groups of data are not easily detected in high dimensional MaOPs a priori.

Most recently, Tian et al. [3] propose a pure diversity PI named CPF by first projecting a solution set to the (m-1)-dimensional unit simplex plane and then to a unit hypercube. The hypervolume of the predicted solution set is found as the score for its diversity. A significant disadvantage of the method is that partially PF coverage is enlarged if it locates higher than the unit simplex plane when projected onto it then cast to a unit hypercube. In the opposite case, the coverage is shrunk when projected on to the unit simplex plane. See section IV for details.

B. Creation of system of reference vectors

The Dan and Dennis method [21] is mainly used to generate reference vectors, which creates uniformly spaced vectors only on a normalised hyper-plane – an (m−1)-dimensional unit simplex plane to all objective axes which have an intercept of one on each axis.

Deb and Jain [22] suggest using two layers of reference points with each of the smaller p, p1 for boundary layer and p2 for the inside layer. The total number of reference lines is dramatically reduced to a manageable (for MaOPs) level.

Tian et al. [23] proposed to generate reference vectors on known true PFs starting with the points created by Das and Dennis method on the unit simplex plane and projecting them to the actual Pareto fronts. However, the final distribution of the reference points generated in this way is not uniform.
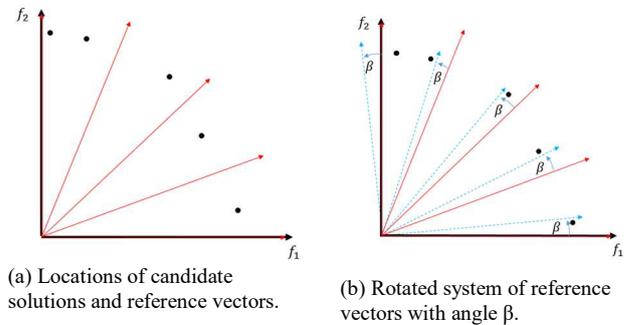

(a) Locations of candidate solutions and reference vectors.

(b) Rotated system of reference vectors with angle β.

Fig. 1. A 2D schematic view shows the rotation of reference vectors by β, where the sum of included angles is minimised to attain the minimal sum of net avertence angles.

III. FORMULATION OF THE PROPOSED DIVERSITY INDICATOR

A. Definition of diversity indicator - Inverted Ratio of Net Avertence angles (IRNA)

Possible systematic bias in diversity measurement may exist when using included angles to a predefined set of reference vectors to formulate PI for diversity. See Fig. 1(a), where each candidate solution has a similar angle difference from its closest reference vector; Erroneously, a diversity score based on these angles is inevitably low. It is even so when assessing approximations of high dimensional MaOPs since the number of solutions is scarce compared with the need to cover the problem's high dimensionality. One way to improve the formulation is by introducing a rotating reference vector system to remove eventual systematic bias in avertence angles between approximation sets and reference vectors. By rotating the reference plane with an optimised angle β, the sum of angle difference is decreased maximumly, and an optimal diversity score is obtained. See Fig. 1(b). When two approximation sets contrast in diversity, the comparison should be made based on each's optimal diversity score. The minimised Inverted Ratio of Net Avertence angles (IRNA) is formulated as a pure diversity indicator defined as the sum of unity minus the ratio of the actual included angle to maximum possible

included angle between individual candidate solutions and the reference vector. See Eq. (1).

$$IRNA = \frac{1}{N}\sum_{k=1}^{N}(1 - \frac{1}{\gamma^k}\theta_a^{(k,min)}) \qquad (1)$$

in which $\theta_a^{(k,min)}$ is the minimised avertence angle between candidate solution k and its nearest reference vector. It is illustrated in a 3D situation as an example shown in Fig. 2.
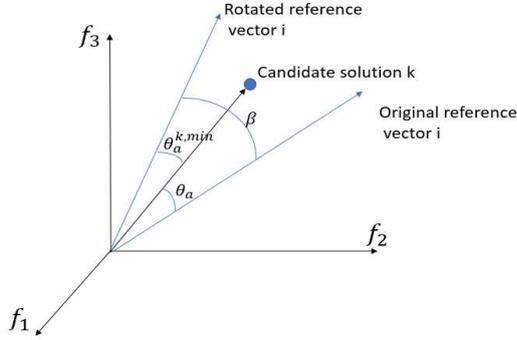

Fig. 2. A 3D schematic view of the rotation of reference vector by β, where the original included angle θ_a is minimised to attain the minimal avertence angle $\theta_a^{(k,min)}$.

γ^k is 1/2 of the included angle between two adjacent reference vectors for candidate k. N is the number of candidate solutions.

The range of the IRNA value is between 0 and 1; a higher score for an approximation set indicates better diversity.

B. Relationship between the included angle of two vectors in m dimensional space and their projections on planes of pairwise coordinate axes

The optimised spatial included angle $\theta_a^{(k,min)}$ is nontrivial to be calculated directly. See Fig. 2. One way to find the angle is by decomposing all involved angles onto respective 2D planes where arithmetic operations can be done. The resultant angle is found based on the net components.

Their rotational projections can express included angle formed by two arbitrarily located spatial vectors in high dimensional space onto respective 2D planes.
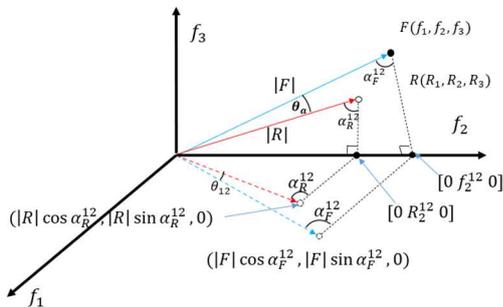

Fig. 3. Rotation about f2 axis, the rotational projection of the spatial angle onto f1-f2 plane is θ_{12}.

The included angle can be proven numerically to be given as:

$$\theta_a = \sqrt{(\theta_{12})^2 + (\theta_{23})^2 ... + (\theta_{ij})^2 ... + (\theta_{m1})^2} \qquad (2)$$

in which θ_a is the avertence angle between two spatial vectors in an m-dimensional space and can be expressed by θ_{12}, θ_{23}, ..., $\theta_{(m-1)m}$ and θ_{m1}, where θ_{ij}, $i \in (1,m), j = i + 1$ $and\ j = 1\ when\ i = m$, are angles of the rotational projections of the vectors about the axis of 2, 3, ..., m and 1 respectively onto planes formed on 1-2, 2-3, ..., m-1 axes and there are m components in total. Fig. 3 illustrates a 3D case of finding θ_{12}.

Assuming:

$$\alpha_R \approx \frac{r_2}{|R|}, \qquad \alpha_F \approx \frac{f_2}{|F|} \qquad (3)$$

$$\theta_{12} = \arccos\left(\frac{[|R|\cdot\cos\alpha_R \ \ |R|\cdot\sin\alpha_R \ \ 0][|F|\cdot\cos\alpha_F \ \ |F|\cdot\sin\alpha_F \ \ 0]}{|R||F|}\right) \qquad (4)$$

Similarly, all other θ_{ij} are found by rotating R and F about f_j axis accordingly. Included angles in high dimensional space can be added or subtracted by first projecting them onto the same respective planes, and the arithmetic operations are done on the projected components. The partial results after component-wise addition or subtraction are enumerated back to the resultant spatial angle searched.

C. Calculation of IRNA

The proposed IRNA is computed in steps as follows:

Find the components β_{ij} of rotation angle β in the various pairwise coordinate planes by minimisation of the expression:

$$\beta_{ij} = argmin_{\beta_{ij},\ \substack{i=1,...,m \\ j=i+1,...,m-i+1}} \sum_{k=1}^{N}\sum_{i=1}^{m}\sum_{j=i+1}^{m-i+1}\left|\theta_{ij}^{(k)} - \beta_{ij}\right| \qquad (5)$$

in which N is the number of candidate solutions.

Find the components of net avertence angle $\theta_{ij}^{k,net}$

$$\theta_{ij}^{(k,min)} = \theta_{ij}^{(k)} - \beta_{ij}, \qquad (6)$$

for $i = 1, ..., m,\ j = i + 1, ..., m - i + 1,\ k = 1, ..., N$

Eq.2 is the vector sum of avertence angle $\theta_a^{(k,min)}$. IRNA is finally computed using Eq. 1.

Algorithm 1 Pseudo Code for IRNA
Input: X (approximation set)
Output: IRNA
1: $X \leftarrow$ Apply non-dominant sorting (X)
2: $X_1 \leftarrow$ Normalise the solution (X)
3: $R \leftarrow$ Build reference plane based on Das and Dennis's approach.
4: $\gamma \leftarrow$ Calculate the minimum angle between each reference line.
5: $X_2 \leftarrow$ Assign solutions to the closest reference line (X_1, R)
6: $\theta_{app} \leftarrow$ calculate separate angle value for each solution (X_2)
7: $\theta \leftarrow$ angle correction (θ_{app})
8: $\beta \leftarrow$ a single optimisation process is carried out to obtain optimal angle(θ, γ, R)
9: IRNA $\leftarrow$ calculate diversity value (β)

Algorithm 1 depicts the pseudo-code for the calculation of IRNA. It starts with the import of population (X) of *n* candidate solutions. The algorithm first eliminates dominated solutions (line 1) and then normalises the remaining data (line 2). The upper and lower limit of the

(a) Fully systematically constructed
HV: 0.8417 ($\pm$1.129e-16)
CPF: 0.7175 ($\pm$2.258e-16)
IRNA: 1.000 ($\pm$4.517e-16)

b) 10% randomly chosen and discarded.
HV: 0.8367 ($\pm$2.392e-4)
CPF: 0.7064 ($\pm$1.275-2)
IRNA: 0.9011 ($\pm$5.729e-8)

c) 25% randomly chosen and discarded.
HV: 0.8297 ($\pm$1.000e-3)
CPF: 0.6875 ($\pm$2.176e-3)
IRNA: 0.7473 ($\pm$8.185e-8)

d) 50% randomly chosen and discarded.
HV: 0.8112 ($\pm$2.944e-3)
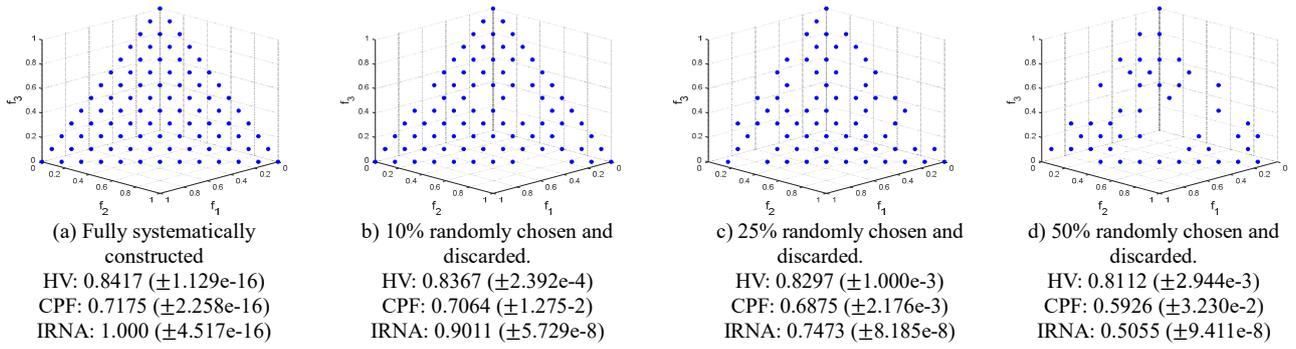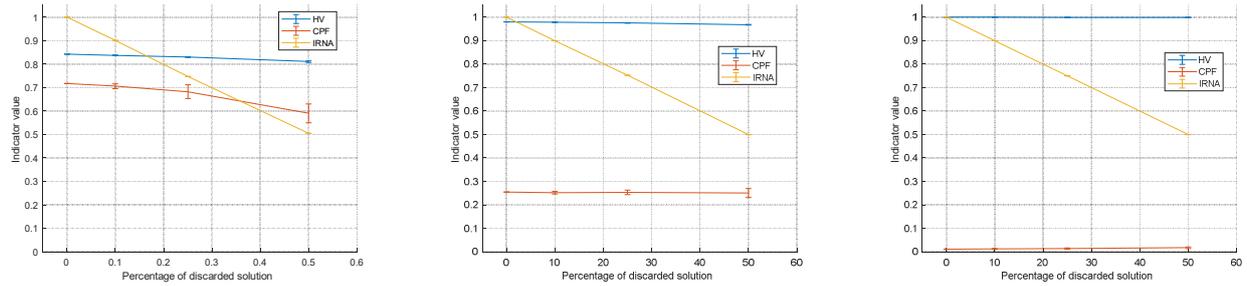CPF: 0.5926 ($\pm$3.230e-2)
IRNA: 0.5055 ($\pm$9.411e-8)

Fig. 4. A typical set of constructed candidate solutions on the normalised unit simplex plane is displayed where a specific portion of randomly chosen solutions are discarded. (30 repeated tests are conducted. The Mean value and standard deviation are shown.)

(a) Analysis done on the normalised unit simplex plane of 3 obj.

(b) Analysis done on the normalised unit simplex plane of 5 obj.

(c) Analysis done on the normalised unit simplex plane of 10 obj.

Fig. 5. The quality indicator values are shown versus the percentage of discarded solutions on the normalised unit simplex plane.

normalisation range can either be decided by the decision-maker or found using the maximum and the minimum value from the data sets. Systematic reference vectors are created based on the method of reference point generation proposed by Das and Dennis's (line 3) approach. The minimal angles between the reference vectors are calculated and stored as variables γ (line 4). Each candidate solution is assigned to its closest reference line based on the size of included angles (Line 5), and IRNA is calculated using the obtained included angles (Lines 6-9).

IV. NUMERICAL STUDIES ON IRNA

The effectiveness of IRNA is assessed through four numerical studies. Firstly, IRNA, HV and CPF are applied on a set of synthetic PFs of uniformly spaced candidate solutions on the unit simplex plane and compared on four different synthetic candidate solution sets; candidate solutions are iteratively and randomly removed from the solution set. The consistency of all metrics is evaluated as diversity decreases (by design) and as the number of objective functions increases. (Section IV.A). Secondly, IRNA, HV and CPF are compared in the evaluation of PFs of MaF Benchmarks functions [24]. (Section IV.B). Thirdly, IRNA, HV and CPF are applied to evaluate approximated PF of MaF Benchmark problems using NSGA III [22] for 3, 5 and 10 objectives (Section IV.C). Finally, IRNA, HV and CPF are used to track the PF performance during optimisation (Section IV.D).

A. Assessment on synthetic candidate solutions on unit simplex plane

The purpose of the tests is to assess the monotonicity and sensitivity of IRNA to known proportional changes of diversity. This test method has been used successfully in

earlier studies, e.g. in [2], to assess performance indicators. Four different cases are studied, a) fully systematically constructed candidate solutions on the unit simplex plane, b) 90% systematically constructed while 10% of candidate solutions are randomly discarded, c) 75% systematically constructed and 25% randomly discarded, and d) 50% systematically constructed and 50% randomly discarded. Fig. 4 displays a typical set of candidate solutions created on the unit simplex plane with 100% systematic creation while 0%, 10%, 25% and 50% are randomly discarded. The competing diversity indicators are calculated and shown in Fig. 5. The calculation has been repeated 20 times to account for the problem's stochastic nature and provide statistics to understand the resulting performance.

As it can be seen on the values of PIs, HV and CPF start with non-unity value for although 100% perfect diversity and reduces in an unproportionate fashion with the further reduction of diversity. IRNA begins with a value equaling 1.0 (designed property). It reduces proportionately with increased amounts of randomly discarded solutions. It shows that IRNA effectively captures the monotonic decrease in diversity.

B. Evaluation on true PF of Benchmark MaFs

HV, CPF and IRNA are tested on theoretical Pareto Front of Benchmark problems MaF. Fig. 6 depicts the indicator values to various true PF of MaF Benchmarks of 1, 2, 3, 5, 6, 7, 10 and 11. Results of MaF 4, 8, 9, 12 and 13 are omitted in this study due to space limitations; MaF 4 has PF of badly scaled, standard shape form and consists of an inverse partial hypercube. At the same time, PF of MaF 8 and 9 are degenerated and are expressed as functions of two decision variables. These are not studied in this paper. MaF 12 and 13 have the same PF shape as that of MaF 5. MaF 14

(a) MaF1
HV= 0.24521 CPF= 0.91035
IRNA= 0.27419

(b) MaF2
HV= 0.21542 CPF= 0.8148
IRNA= 0.3946

(c) MaF3
HV= 0.97024 CPF= 0.70751
IRNA= 0.41757

(d) MaF5
HV= 0.58721 CPF= 0.74412
IRNA= 0.99998

(e) MaF6
HV= 0.20212 CPF= 0.63619
IRNA= 0.050262

(f) MaF7
HV= 0.28874 CPF= 0.81041
IRNA= 0.26204,

(g) MaF10
HV= 0.95671 CPF= 0.82621
IRNA= 0.9874

(h) MaF11
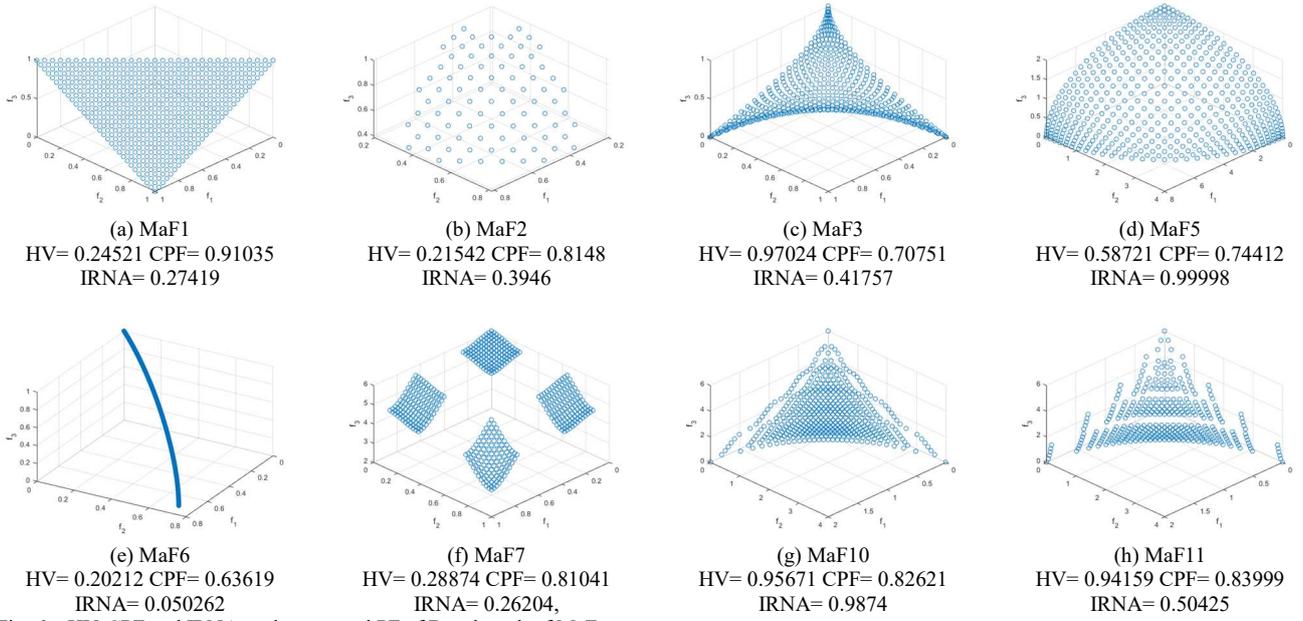HV= 0.94159 CPF= 0.83999
IRNA= 0.50425

Fig. 6. HV, CPF and IRNA evaluate actual PF of Benchmark of MaFs.

has the same PF form as MaF1, and MaF 15 has the same PF shape as MaF 4. Both are dedicated to large scale problems, which are not the subject of this article.

Fig. 6(a) shows PF of MaF1, which consists of a partial simplex plane of non-unity. Its partial coverage of the objective space is reflected in the HV and IRNA values. CPF is proportional to the coverage of the projected PF on a unit simplex plane. The partial PF is located higher up than the unit simplex plane. When projected onto the unit simplex plane, the projection covers – erroneously – a larger proportion of the unit simplex plane, resulting in a large CPF value.

Fig. 6 (b) displays the PF of MaF2, a partial sphere. HV and IRNA have reasonable low values since the PF covers only partially the objective space, while the CPF value is too high for the same reason as it is explained for MaF1.

Fig. 6 (c) displays the PF of MaF3, which forms a convex plane. The PF is calculated based on a mathematical formula that creates unevenly distributed candidate solutions. However, this yields high HV values since many solutions are concentrated along the boundaries favoured by HV [19]. The value of CPF is lower than HV because it fails to cover the simplex plane completely when it is projected to the plane in parallel. In other words, CPF shows lower diversity estimation erroneously on PF of convex shapes, which locate lower than the simplex plane. IRNA expresses the averaged uniformity of an approximation set. The PF solutions shown in Fig. 6 (c) are found using a mathematical expression valid for the benchmark [25] and the result is unevenly distributed although seemingly densely populated. Hence, the sum of inclusion angles in IRNA calculation is averaged over, by dividing with the total number of solutions, the resulted value of IRNA is relatively low. Subsequently, the same benchmark is used with NSGA III; the IRNA PI is much higher, as in Fig. 7(c). This is because a uniformly distributed set of reference points is utilised on the search for solutions.

Fig. 6(d) shows the PF of MaF5, which is a sphere. PF covers the objective space fully, which is reflected in high IRNA value. When projected on a unit simplex plane, the

solutions have uneven distribution, which results in lower CPF values. HV favours candidates locating on the boundaries, where such solutions are not overwhelmingly present as the case for MaF3; hence a low HV value is reached.

Fig. 6(e) shows the PF of MaF6, which consists of a degenerated PF shape of an arc, with low coverage of the objective space, which gives low IRNA value since IRNA expresses the coverage distribution solution on objective space. The value of HV is also low since there are only two extremal solutions and otherwise only intermediate solutions of the concave type, which result in low HV [19]. CPF results in too high value in this case.

Fig. 6(f) shows the PF of MaF7 that covers partially the objective space, which is reflected on HV and IRNA values. Similarly, as stated above, CPF covers an erroneously large proportion of the unit simplex plane, which results in a large CPF value.

Fig. 6(g) depicts the PF of MaF10 that covers the objective space fully. This is correctly reflected on HV, CPF and IRNA values.

Fig. 6(h) shows the PF of MaF11 that covers only partially the objective space, but HV and CPF have relatively high values, while the partial coverage is reflected only on the value of IRNA. When PF is projected onto the unit simplex plane, it covers a large proportion of it, which results in a very high CPF value, erroneously. Solutions at knee points and boundary areas contribute more to the HV value in the case of a PF of convex shape[19], and there are many such points in the solution, which lead to high HV value.

C. Evaluation on PF of MaF Benchmark problems with the increasing number of objectives

Comparisons among HV, CPF and IRNA when applied on PF approximations calculated using NSGA III for 3, 5, 7 and 10 objectives of MaF1-7 and 10-11 are performed. See Fig. 7. The purpose of the tests is to examine further the proposed indicator's robustness and monotonicity as the number of objectives increases. With almost the same
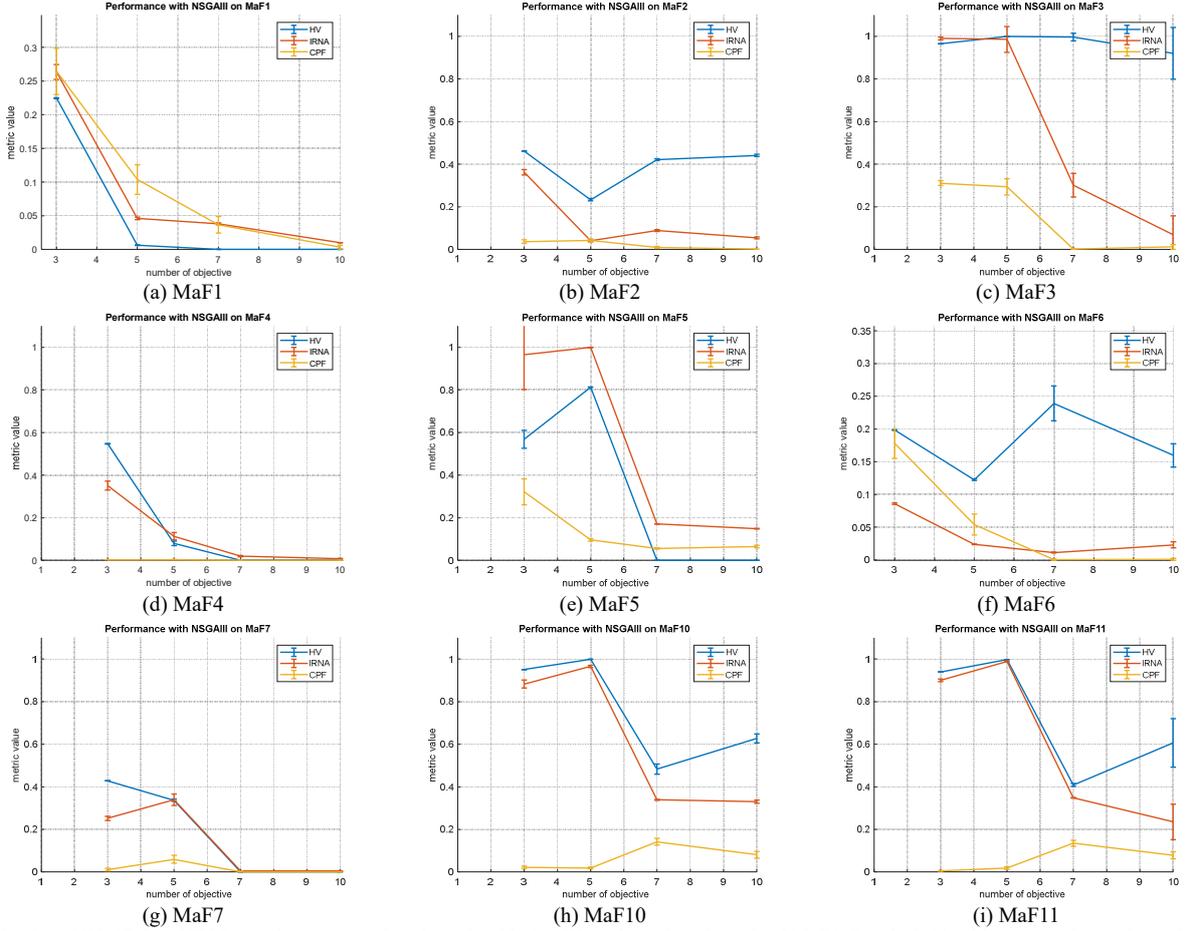
Fig. 7. HV, CPF and IRNA evaluate approximations that NSGA III analyses benchmark of MaF1-7 and 10-11 with mean value and standard deviation after 30 independent runs and solutions.

number of candidate solutions in analysis, it is expected that the diversity of solutions will reduce as the number of objective functions increases since larger objective space will be covered by the same number of candidates [19].

TABLE I.   NUMBERS OF ITERATIONS AND CANDIDATE SOLUTIONS

Number of objectives	Number of evaluations	Number of solutions
3	200,000	210
5	500,000	210
7	500,000	210
10	500,000	275

The algorithmic parameters adopted in NSGA III are based on default values acquired from PlatEMO version 2.7 [25]. The number of iterations and number of candidate solutions is listed in Table 1. Reference vectors are generated using Das and Dennis. Each Benchmark problem with a specific number of objectives is calculated 30 times. Approximation sets are evaluated by PIs and shown in their mean values and standard deviations.

Fig. 7(a) displays the change of HV, CPF and IRNA indicator values applied on approximations of MaF1 for an increasing number of objectives. All three indicators behave as expected, i.e., values decrease monotonically with the increase of objectives, but HV's values are extremely low on 7 and 10 objectives. The volume above the PF reduces drastically with the increasing number of objectives. This

can also be reasoned as follows: simplex planes can be in general expressed as in Eq. 7.

$$f_1 + f_2 + \cdots + f_m = a_m \tag{7}$$

in which a_m is the value of the objective function at which the simplex plane coincides with f_i axis. ($f_i = a_m$ while $f_j = 0$, for all $j \neq i$). For midpoint on PF: $f_1 = f_2 = \ldots = f_m$, we have:

$$f_1 = f_2 = \ldots = f_m = {a_m}/{m} \tag{8}$$

Its distance to the Ideal point r_m is:

$$r_m = \sqrt{(f_1)^2 + (f_2)^2 + \cdots + (f_m)^2} = {a_m}/{\sqrt{m}} \tag{9}$$

a_m in MaF1 increases more rapidly than $\sqrt{m}$ with increasing m causing r_m increases rapidly, and HV reduces drastically.

In MaF2, HV breaks monotonicity, which value increases from 5 objectives to 7 and 10 objectives. PF of a hypercube is governed by:

$$(f_1)^2 + (f_2)^2 + \cdots + (f_m)^2 = r^2 \tag{10}$$

in which r is the radius of the hypercube, and in this case, r < 1.0. See Fig. 6(b). The volume of the hypercube is
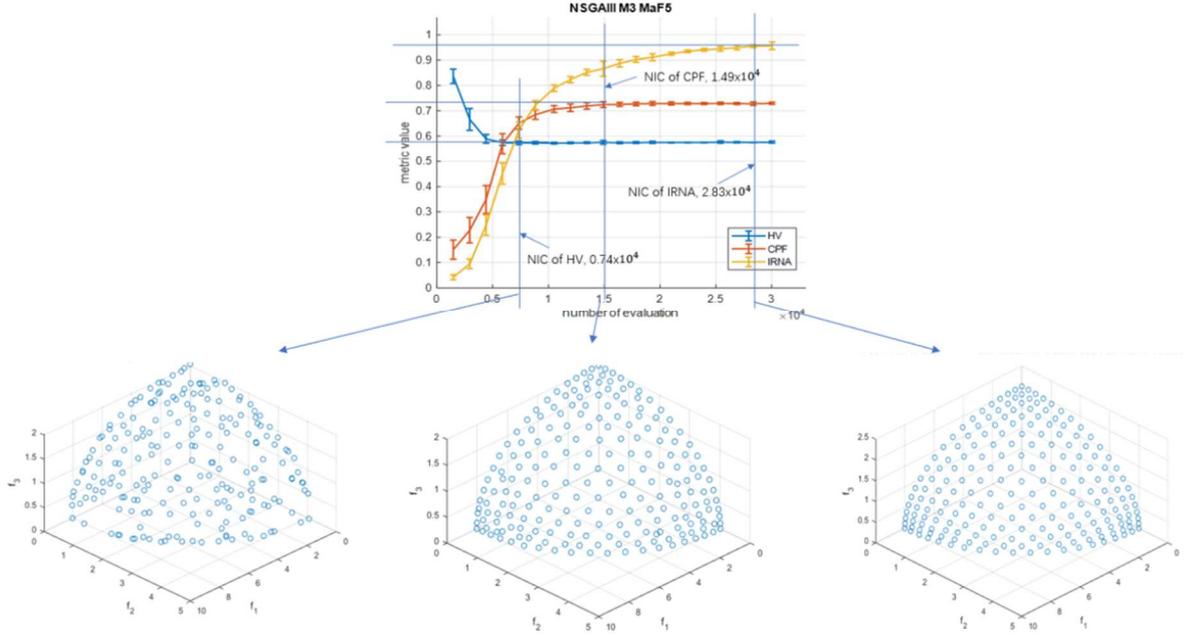
(a) NSGAIII M3 MaF5 converge 25%
HV: 0.56478 CPF: 0.60497 IRNA: 0.56138

(b) NSGAIII M3 MaF5 converge 50%
HV: 0.57224 CPF: 0.73351 IRNA: 0.88431

(c) NSGAIII M3 MaF5 converge 95%
HV: 0.57509 CPF: 0.73410 IRNA: 0.96261

Fig. 8. It visually compares solutions at a different nominal number of iterations at convergence (NIC) on contrasted PIs. The result is based on 30 independent runs, and the mean and standard deviation are shown.

proportional to r^m and HV is proportional to $1/r^m$, thus increases with the increasing number of objectives.

When MaF3 is concerned, see Fig. 7(c), HV value keeps high for all cases of the number of objectives, which is caused by the fact that HV value favours boundary points and knee points on PF of convex shape, which gives high HV in all cases of the number of objectives. CPF has too low value in 7 and 10 objectives, while IRNA changes monotonically with the number of objectives.

Monotonic change of HV and IRNA with the number of objectives are also observed in results of MaF4, see Fig. 7(d). MaF4 is a partial convex hypercube or partial inverse hypercube. CPF, in discussed earlier, has a projection on the simplex plane that is 'shrunk', hence covers a falsely small portion of the plane.

In MaF5, IRNA has a full score in diversity in 3 and 5 objective cases. Its values decrease as the number of objectives goes up because PF is under-represented by the available number of candidate solutions. The values of CPF and HV are very low.

Non-monotonic behaviour is observed in HV for MaF6, which has a degenerate PF shape, with an increasing number of objectives.

PF of MaF7 is of partial type and consists of several 'flakes' in the objective space. Only HV shows a monotonic change from 3 to 5 objectives, while all three indicators have near-zero values for 7 and 10 objective cases.

When results of MaF10 and MaF11 are concerned, only IRNA displays nearly monotonic behaviour. See Figs: 7 (e) and 7(f).

In summary, we conclude that IRNA exhibits overall satisfactory monotonicity in all benchmark functions tested. In general, it performs more reliable than both HV and CPF, which behave erratically in some cases.

D. Tracking the performance of approximations

It is logical to assume that convergence and diversity of PF approximations are both very low in the early stages of optimisation (low number of iterations), while both improve gradually as the optimisation algorithm progresses. Capturing and tracking this progress is crucial for understanding (and guiding via feedback) the performance of optimisation algorithms; hence consistent and sensitive performance indicators are important.

Benchmarks of MaF 5, 6 and 7 are investigated using NSGA III, where approximations are evaluated by the indicators HV, CPF and IRNA to examine each PI's performance. A total of 30 independent runs per benchmark function per chosen number of objective functions, have been carried out, and the mean value and standard deviation are calculated to take the stochastic effect on the approximation into account in the analyses. PF of MaF5 consists of a hypercube, i.e., the PF covers the objective space fully. MaF6 comprises a PF of a pure arc, which is of degenerate type. MaF7 is made of several 'flakes' in the objective space and represents a typical PF of a partial sort. These benchmarks are selected to represent three main PF categories (full coverage, degenerative, partial). The resulting numerical simulations demonstrate the differences in consistency, sensitivity, and monotonicity of the indicators.

The number of iterations at convergence (NIC) for each PI is used here as a measure for its judgement on the completion of iteration process. When analysing approximation sets at various iteration stages using the same algorithm, the diversity indicator that results in the largest NIC (with the best diversity displayed) is most sensitive to detect diversity changes of solutions. This is explained and demonstrated visually in Fig. 8 via monitoring the convergence process of MaF5 with three objective functions. At NIC of HV, diversity is still low. See Fig.8(a). At NIC of CPF, the diversity is improved, see Fig.8(b). Only
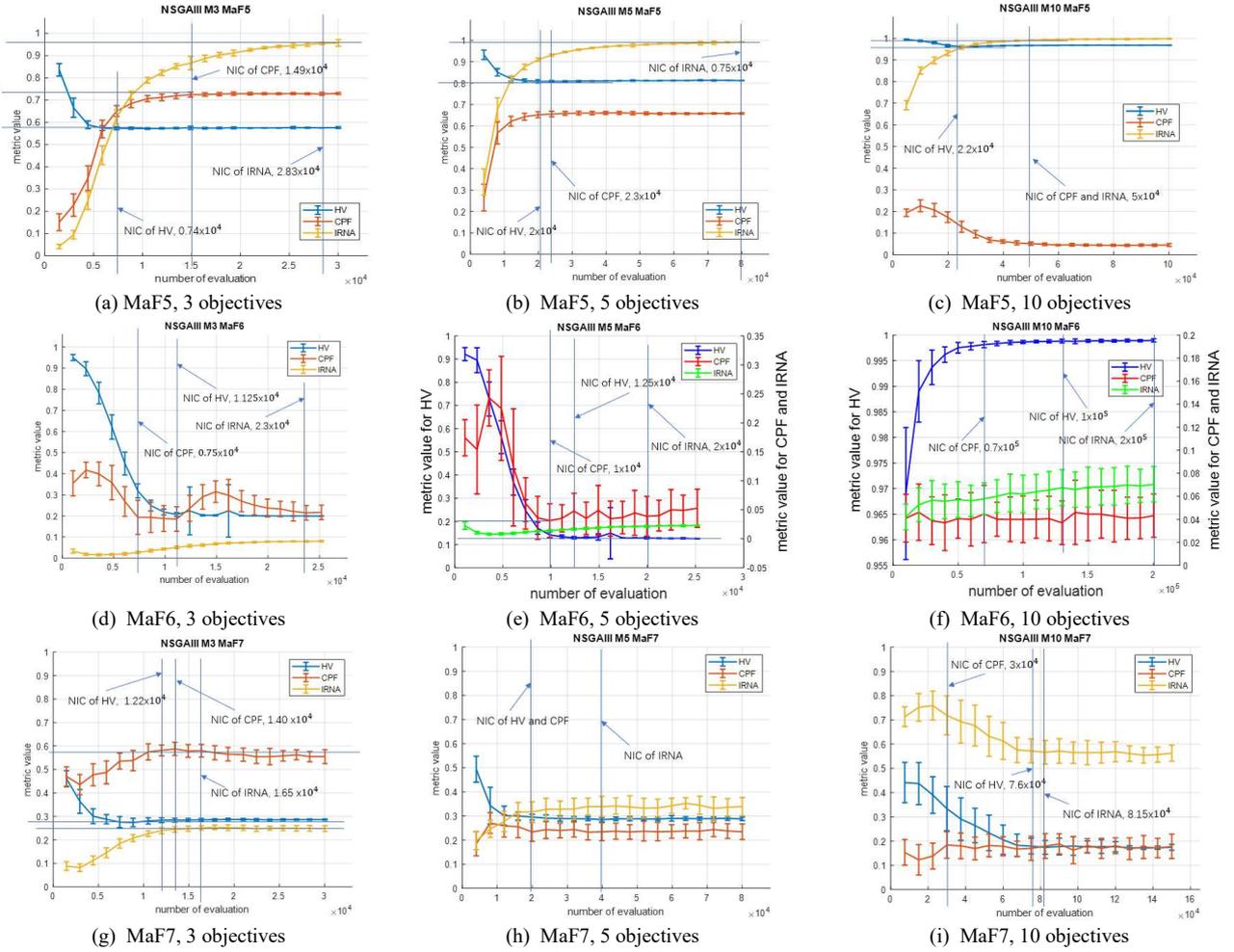
Fig. 9 Approximations over time for 30 independent runs of Benchmark of MaF5-7 are obtained using NSGA III. The solutions are evaluated by HV, CPF and IRNA, and the results are shown in mean value and standard deviation.

when the NIC of IRNA is reached, the diversity becomes superior. See Fig.8(c). It is valid for all three cases tested, i.e., MaF5, MaF6 and MaF7. Due to space limitations, only the result of MaF5 is shown in this article.

Fig. 9(a) shows the same approximations of MaF5 in three objectives, using NSGA III and evaluated by HV, CPF and IRNA. CPF and IRNA increase gradually and monotonically as number of evaluations goes up. But HV decreases gradually and monotonically to a stable level. Besides, HV has the lowest NIC followed by CPF while IRNA obtains the highest NIC, which means that IRNA is most sensitive to capturing diversity change, followed by CPF. The same trend is observed in approximations of MaF5 of 5 objectives. See Fig. 9(b), CPF and IRNA increase gradually and monotonically as the evaluations go up. At the same time, HV decreases gradually and monotonically to a stable level, and the IRNA value has the largest NIC. Similar is the case in 10 objectives, except that the CPF value decreases with the increasing number of iterations and has the same NIC value as IRNA. See Fig. 9(c).

IRNA can also be used to describe the diversity of the degenerated type of PF of MaF6. Fig. 9(d) depicts the approximation process over time of MaF6 with three objective functions. Once again, the IRNA value increases monotonically until a stable level is reached, indicating the end of approximation process. IRNA has the largest NIC value among the three indicators. It is also true for MaF6 in

5 and 10 objectives. See Figs. 9(e) and 9(f). The high standard deviation of CPF and IRNA value in these cases should be noted, indicating strong variations in the results.

Better sensitivity of IRNA is also demonstrated using approximations of PF of partial types, e.g., that of MaF7. See Figs. 9(g) for 3 objectives, 9(h) for 5 objectives and 9(i) for 10 objectives, respectively. In 3 and 5 objective cases, the IRNA amount displays monotonic increasing behaviour and maintains the three indicators' largest NIC value. But the situation is changed in the case of 10 objectives. IRNA still varies almost monotonically with the increasing number of iterations but decreasing toward a stable level. This is caused by the larger spread of non-converged solutions in the early stage of the approximation process (PF of partial coverage).

We conclude that IRNA exhibits a balanced sensitivity and monotonicity performance and yields arguably a more robust indicator than HV and CPF in the benchmark functions tested.

V. CONCLUSION

A new pure diversity indicator, the Inverted Ratio of Net Avertence angles (IRNA), is introduced. The proposed performance indicator is empirically tested for its efficacy on solutions of known diversity (synthetic data) constructed on a unit simplex plane and approximations of 3, 5, 7 and 10 objectives of Benchmark problems MaF1-7 and 10-13.

MOEA algorithm NSGA III is used to reach approximations on Benchmark problems, IRNA is used against the commonly used HV and the more recent CPF. The novelty of the proposed diversity score yields by rotating the reference vector system with an optimal spatial angle. This rotation results in removing any potential systemic bias in included angles in data of approximations so that impartial scores of the diversity of approximation sets are obtained. Numerical results and analysis show that IRNA yields an overall more balanced performance. IRNA is more sensitive and monotonically proportionate in capturing diversity changes than HV and CPF indicators in the synthetic data problems, in actual PFs of MaF benchmark problems, and in cases when the number of objectives increases above three (for many-objective problems). Moreover, this robust performance is also observed in tracking the progress of algorithms during optimisation.

Two new alternative methods in assessing sensitivity and monotonicity of performance indicators are used. One is studying variations when evaluating an increasing number of objectives while keeping the number of candidate solutions constant; the other examines their value changes at various iteration stages up to final convergence.

The proposed IRNA is assessed here against popular performance indicators to provide a first insight. Towards further research, comparisons need to be made against a broader range of state-of-the-art performance indicators, additional coverage of types of benchmark problems in terms of dimensionality and complexity, as well as real life problems.

REFERENCES

[1] K. Li, R. Wang, T. Zhang, and H. Ishibuchi, "Evolutionary Many-Objective Optimization: A Comparative Study of the State-of-the-Art," *IEEE Access*, vol. 6, pp. 26194–26214, 2018, doi: 10.1109/ACCESS.2018.2832181.

[2] H. Wang, Y. Jin, and X. Yao, "Diversity Assessment in Many-Objective Optimisation," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1510–1522, 2017, doi: 10.1109/TCYB.2016.2550502.

[3] Y. Tian, R. Cheng, X. Zhang, M. Li, and Y. Jin, "Diversity Assessment of Multi-Objective Evolutionary Algorithms: Performance Metric and Benchmark Problems [Research Frontier]," *IEEE Comput. Intell. Mag.*, vol. 14, no. 3, pp. 61–74, Aug. 2019, doi: 10.1109/MCI.2019.2919398.

[4] N. Riquelme, C. Von Lücken, and B. Baran, "Performance metrics in multi-objective optimisation," in *2015 Latin American Computing Conference (CLEI)*, 2015, pp. 1–11, doi: 10.1109/CLEI.2015.7360024.

[5] E. Zitzler, J. Knowles, and L. Thiele, "Quality assessment of Pareto set approximations," in *Multiobjective Optimisation Interactive and Evolutionary Approaches*, vol. 5252 LNCS, Springer Berlin Heidelberg, 2008, pp. 373–404.

[6] X. Cai, H. Sun, and Z. Fan, "A diversity indicator based on reference vectors for many-objective optimisation," *Inf. Sci. (NY).*, vol. 430–431, pp. 467–486, 2018, doi: 10.1016/j.ins.2017.11.051.

[7] C. Audet, J. Bigeon, D. Cartier, and S. Le, "Performance indicators in multiobjective optimisation," *Eur. J. Oper. Res.*, pp. 1–39, 2018.

[8] M. Li and X. Yao, "Quality Evaluation of Solution Sets in Multiobjective Optimisation," *ACM Comput. Surv.*, vol. 52, no. 2, pp. 1–38, May 2019, doi: 10.1145/3300148.

[9] D. A. Van Veldhuizen and G. B. Lamont, "On measuring multiobjective evolutionary algorithm performance," in *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512)*, 2000, vol. 1, pp. 204–211 vol.1, doi: 10.1109/CEC.2000.870296.

[10] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," *IEEE Trans. Evol. Comput.*, vol. 10, no. 1, pp. 29–38, 2006, doi: 10.1109/TEVC.2005.851275.

[11] S. Mostaghim and J. Teich, "A New Approach on Many Objective Diversity Measurement," *Pract. Approaches to Multi-Objective Optim.*, pp. 1–15, 2005.

[12] D. a. Van Veldhuizen and G. B. Lamont, "Evolutionary Computation and Convergence to a Pareto Front," *Late Break. Pap. Genet. Program. 1998 Conf.*, pp. 221–228, 1998, [Online]. Available: http://www.lania.mx/~ccoello/EMOO/vanvel2.ps.gz.

[13] H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima, "Modified Distance Calculation in Generational Distance and Inverted Generational Distance," A. Gaspar-Cunha, C. Henggeler Antunes, and C. C. Coello, Eds. Cham: Springer International Publishing, 2015, pp. 110–125.

[14] O. Schutze, X. Esquivel, A. Lara, and C. A. C. Coello, "Using the Averaged Hausdorff Distance as a Performance Measure in Evolutionary Multiobjective Optimization," *IEEE Trans. Evol. Comput.*, vol. 16, no. 4, pp. 504–522, Aug. 2012, doi: 10.1109/TEVC.2011.2161872.

[15] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimisers: an analysis and review," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 117–132, 2003, doi: 10.1109/TEVC.2003.810758.

[16] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," *IEEE Trans. Evol. Comput.*, vol. 3, no. 4, pp. 257–271, 1999, doi: 10.1109/4235.797969.

[17] K. Li, K. Deb, and X. Yao, "R-Metric: Evaluating the Performance of Preference-Based Evolutionary Multiobjective Optimization Using Reference Points," *IEEE Trans. Evol. Comput.*, vol. 22, no. 6, pp. 821–835, Dec. 2018, doi: 10.1109/TEVC.2017.2737781.

[18] K. Deb and S Jain, "Running Performance Metrics for Evolutionary Multi-Objective Optimization," *Kangal Rep.*, vol. 2002004, pp. 13–20, 2002, doi: 2002004.

[19] M. Li, S. Yang, and X. Liu, "Diversity comparison of Pareto front approximations in many-objective optimisation," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2568–2584, 2014, doi: 10.1109/TCYB.2014.2310651.

[20] M. Li, S. Yang, and X. Liu, "A performance comparison indicator for Pareto front approximations in many-objective optimisation," *GECCO 2015 - Proc. 2015 Genet. Evol. Comput. Conf.*, pp. 703–710, 2015, doi: 10.1145/2739480.2754687.

[21] J. E. Dennis, "NORMAL-BOUNDARY INTERSECTION : A NEW METHOD FOR GENERATING THE PARETO SURFACE IN NONLINEAR," *Soc. Ind. Appl. Mathematics*, vol. 8, no. 3, pp. 631–657, 1998, doi: 10.1137/S1052623496307510.

[22] K. Deb and H. Jain, "An Evolutionary many-objective Optimisation Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 577–601, 2014, doi: 10.1109/TEVC.2013.2281535.

[23] Y. Tian, X. Xiang, X. Zhang, R. Cheng, and Y. Jin, "Sampling Reference Points on the Pareto Fronts of Benchmark Multi-Objective Optimization Problems," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, Jul. 2018, pp. 1–6, doi: 10.1109/CEC.2018.8477730.

[24] R. Cheng *et al.*, "A benchmark test suite for evolutionary many-objective optimisation," *Complex Intell. Syst.*, vol. 3, no. 1, pp. 67–81, Mar. 2017, doi: 10.1007/s40747-017-0039-7.

[25] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: A MATLAB Platform for Evolutionary Multi-Objective Optimization [Educational Forum]," *IEEE Comput. Intell. Mag.*, vol. 12, no. 4, pp. 73–87, Nov. 2017, doi: 10.1109/MCI.2017.2742868.