

On The Use of Hypervolume for Diversity Measurement of Pareto Front Approximations

Shouyong Jiang

Centre for Computational Intelligence (CCI)
School of Computer Science and Informatics
De Montfort University
The Gateway, Leicester LE1 9BH, UK
Email: shouyong.jiang@email.dmu.ac.uk

Shengxiang Yang

Centre for Computational Intelligence (CCI)
School of Computer Science and Informatics
De Montfort University
The Gateway, Leicester LE1 9BH, UK
Email: syang@dmu.ac.uk

Miqing Li

Natural Computation Group
School of Computer Science
University of Birmingham
Birmingham B15 2TT, UK
Email: m.li.8@cs.bhm.ac.uk

Abstract—In multiobjective optimization, a good quality indicator is of great importance to the performance assessment of algorithms. This paper investigates the effectiveness of the widely-used hypervolume indicator, which is the only one found so far to strictly comply with the Pareto dominance. While hypervolume is of undisputed success to assess the quality of an approximation, it is sensitive to misleading cases, particularly for diversity assessment. To address this issue, this paper presents a modified hypervolume indicator based on linear projection for diversity evaluation. In addition to experimental studies to demonstrate the effectiveness of the proposed indicator, the indicator is introduced into the environmental selection of an indicator-based multiobjective optimization evolutionary algorithm. Experiments show that the proposed indicator yields more evenly-distributed approximations than the original hypervolume indicator.

I. INTRODUCTION

Multiobjective optimization involves the simultaneous optimization of several conflicting objectives. Due to multiobjectivity, the optima of a multiobjective optimization problem (MOP) is not a single solution but a set of tradeoff solutions, whose image in the objective space is known as the Pareto-optimal front (POF). Multiobjective optimization evolutionary algorithms (MOEAs) are an important class of approaches for solving MOPs, as they can obtain a set of solutions without any knowledge of problem features in a single run. This advantage has led to a fast growth of the design of MOEAs, and many MOEAs [3], [5], [9], [23], [24] have been proposed so far.

While many MOEAs are available for MOPs, a natural question arises – how to measure these algorithms. Due to the presence of conflicting objectives, evaluating the quality of approximations obtained by MOEAs that comprise multiple solutions is not an easy task. This makes performance assessment an important research topic in multiobjective optimization. Over the past 20 years, a variety of quality indicators [9], [10], [23], [14], [15] have been proposed to measure POF approximations. They focus on two aspects of the quality of an approximation, i.e., convergence and diversity. Convergence is related to the closeness of an approximation to the true POF whereas diversity in general refers to the uniformity and extensity of the approximation. Some quality indicators are specifically designed to measure either convergence or diversity, but others try to capture both aspects. Among these

indicators, hypervolume (HV) is the only one known to be strictly compliant with the Pareto dominance [25]. Due to this nice property, HV has been widely adopted as a standard offline indicator for performance assessment [5], [3], [25] and also an online selection operator for indicator-based MOEAs [5], [3].

In real industrial design a decision maker (DM) is able to consider only a few possible solutions (Pareto points). In such a context, it is important to have an even distribution and a good coverage of Pareto points from which the DM can obtain maximum information about the Pareto front at minimum computational cost. A well-distributed Pareto set can also be a good foundation for visualizing the Pareto front. It can substantially simplify the DM's work. Note that, some practitioners may appreciate a focus on regions of interest, e.g., knees and borders on the POF, while knowing about the limits of their problem. However, we believe that it is more desired to obtain a set of well-diversified Pareto points including the knees and borders so that the DM can have more alternatives when making a decision. Therefore, one may wonder how diversified a POF approximation is enough to represent the whole POF. This naturally gives rise to the development of quality indicators to measure the diversity of the POF approximation in the literature [9], [8]. Despite that, these diversity indicators often require the information of the true POF or problem-specific parameter settings to guarantee the correctness of assessment.

In this paper we focus on the use of HV for diversity assessment. Recently, HV was found to give biased evaluations in terms of diversity, which makes conclusions based on HV unreliable [15]. In view of this, this paper presents a modified HV indicator to quantify diversity. The proposed indicator is examined on different scenarios of POF approximations, showing fairly accurate measure of the approximations' diversity. Besides, the proposed indicator is also adopted as a selection operator in SMS-EMOA [5], and results demonstrate that it helps SMS-EMOA to achieve a more uniformly-distributed approximation than the original HV indicator.

The rest of this paper is organized as follows. Section 2 presents some existing diversity indicators and discusses possible limitations of HV. Section 3 describes our proposed

TABLE I
DIVERSITY PERFORMANCE INDICATORS AND THEIR CHARACTERISTICS

Indicator	Relative or Absolute	Uniformity	Extensity	Convergence	Boundary Solutions Needed	Reference Point/Set Needed	space partition, hyperboxes, or parameters Needed
G-Metric [17]	Relative	✓	✓	✓	x	x	✓
DCI [14]	Relative	✓	✓	x	x	x	✓
PCI [15]	Relative	✓	✓	✓	x	x	✓
HV [25]	Absolute	✓	✓	✓	x	point	x
IGD [23], IGD ⁺ [12]	Absolute	✓	✓	✓	✓	set	x
Δ_p [19]	Absolute	✓	✓	✓	✓	set	x
Spacing [8]	Absolute	✓	x	x	x	x	x
R_2 [7]	Absolute	✓	✓	✓	✓	set	x
ϵ -indicator [25]	Absolute	x	x	✓	✓	set	x
Δ Metric [9]	Absolute	✓	✓	x	✓	set	x
Sigma Diversity Metric [18]	Absolute	✓	✓	x	✓	set	x
DM [10]	Absolute	✓	✓	x	✓	set	✓

indicator. Experimental studies are delivered in Section 4. Section 5 concludes the paper.

II. RELATED WORK

A. Diversity Indicators

Diversity is well recognized to play a very important role in evolutionary multiobjective optimization. Throughout the paper, diversity refers to only the distribution of an approximation in the objective space, and population distribution in the decision space is beyond the scope of this work. Since it is rarely possible to obtain all (infinite) solutions on the POF, it is plausible to present DMs a ‘suitable’ finite-size approximation of the POF. Therefore, it is important that the POF approximation should be arranged in order to be a sufficiently good representation of the POF. In practice, it is often desirable that the POF approximation is well-diversified over the whole range of the POF so as to maximize the information of the POF presented to the DMs [20].

The diversity of POF approximations is often related to the uniformity and extensity of the POF approximations. Note that, a uniformly-distributed POF approximation not necessarily means that the approximation covers well the POF. As a complement to uniformity, extensity considers the coverage or the spread of the POF approximation. That is, in many cases it is desired to obtain simultaneously an evenly-distributed and well-extended POF approximation. Mathematically, these two properties can be generally formulated as $uniformity(P)$ and $extensity(P, P^*)$, where P and P^* are an POF approximation and the true POF, respectively. An instantiation of $uniformity(P)$ is Spacing [8], where equispaced approximation elements are thought to be an ideal case of even distribution. Likewise, maximum spread [11] can be considered as a realization of $extensity(P, P^*)$, where the true POF P^* should be used as a reference when mentioning the extensity of the approximation P . $uniformity(P)$ and $extensity(P, P^*)$ collectively help to reflect the diversity of the POF approximation P .

In the literature, there have been a few performance indicators for evaluating MOEAs’ diversity. Table I lists some

existing indicators that can quantify the diversity of approximations, where the characteristics of each indicator are briefly tabulated. It can be seen from the table that, the indicators can be grouped into two categories, i.e., relative and absolute indicators, according to the first characteristic. Relative indicators can make a relative comparison between different approximations, but they generally cannot reflect real approximation to the true POF because they do not use any reference set/point related to the POF. Besides, relative indicators often need to perform space (hyperbox) partition or population clustering to help evaluation, resulting in the introduction of extra parameters. If these parameters are not properly set, relative indicators may make an incorrect diversity evaluation. For example, the diversity comparison indicator (DCI) [14] is sensitive to its internal parameter (e.g., the size of hyperboxes) used for space partition. On the other hand, absolute indicators can tell how well an approximation is to the true POF, but they often require a known reference point/set. Boundary solutions on the true POF are explicitly needed in order for these indicators (except HV [25]) to make an accurate evaluation of an approximation.

It is also clear from Table I that most of the indicators can help to quantify the diversity performance (including uniformity and extensity) of POF approximations. However, they have weaknesses too. In order to analyze both uniformity and extensity, relative indicators requires space partition [17] or hyperbox construction [14] whereas absolute ones need a reference point [25] or reference set [23]. Note that, Spacing [8] is an exception which assesses diversity without the need of reference or extra parameters. But, it concentrates mainly on measuring the uniformity of an approximation and cannot reflect how extensive the approximation is over the POF.

Among these quality indicators, the inverted generational distance (IGD) [23] and HV [25] are two of the most widely used indicators for evaluating MOEAs. Both IGD and HV can in a sense assess the two aspects (i.e., uniformity and extensity) of diversity for an approximation. The accuracy of IGD depends largely on the used reference set that is uniformly sampled from the whole range of true POF. The larger the size of the reference set, the more accurate IGD

will be. This hinders its wide application onto many-objective optimization due to the exponential increase in the required size of the reference set for a good evaluation. Also, if the POF is unknown beforehand, IGD is not applicable.

Unlike IGD, HV can evaluate an approximation without much knowledge of the true POF. The computation of HV involves only a reference point instead of a reference set, so the accuracy depends only on the selected reference point, particularly when practitioners are interested in extremal or boundary POF points (these kinds of solutions are often related to the extensity of the approximation). The HV indicator measures the size of the objective space dominated by the approximated solution set S and bounded by the selected reference point $R = (R_1, \dots, R_M)^T$ that is dominated by all points of an approximation, and is computed by:

$$HV(S) = Leb(\cup_{x \in S} [f_1(x), R_1] \times \dots \times [f_M(x), R_M]) \quad (1)$$

where $Leb(A)$ is the Lebesgue measure of a set A , and M is the number of objectives.

HV has a nice property that it guarantees strict compliance regarding the Pareto dominance [25]. This property makes it preferable than other performance indicators [25], and has been widely used for assessing MOEAs' overall performance.

B. Limitations of Hypervolume

Despite its great success in assessing the overall quality of an approximation, HV has been reported to have the following limitations:

- Sensitivity to the reference point: HV is very sensitive to the setting of the reference point [2]. Different reference points may result in different evaluations.
- Bias toward certain regions: HV is biased toward knee points which make larger contributions than no-knee points [25], [15]. Thus, HV-based search methods are very likely to distribute too many solutions in knee-near regions, leading to potential diversity loss during the search.
- Expensive computation: While it is easy to compute HV for two objectives, the computation of HV grows exponentially when the number of objectives increases. This prevents the wide application of HV to many-objective optimizations.
- Inconsistency with true performance: As a result of the second limitation, HV may give an incorrect measure of the performance of an approximation even if the reference point is well placed [15].

While most studies on HV are based on empirical observations, Auger *et al.* [2] made a theoretical investigation regarding the first two limitations. They mathematically proved that the optimal distribution of μ solutions that maximizing HV depends largely on the slope of the POF. That is, HV is biased toward knee points, but not necessarily toward boundary points. They also further investigated the influence of the reference point, thereby deriving an explicit lower bound required for the reference point to ensure the good behaviour of the HV indicator.

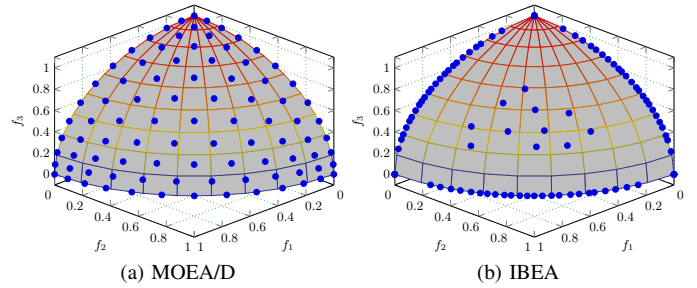


Fig. 1. An example that HV is not consistent with diversity, where true Pareto optimal solutions are obtained by MOEA/D and IBEA on the three-objective DTLZ2 without distance-related parameters.

TABLE II
HV RESULTS OF MOEA/D AND IBEA UNDER DIFFERENT REFERENCE POINTS $R = (r, r, r)^T$

r	1.01	1.05	1.10	1.20	1.50	2.00	3.00
MOEA/D	0.4441	0.5715	0.7448	1.1418	2.7888	7.4138	26.4138
IBEA	0.4453	0.5726	0.7460	1.1430	2.7900	7.4150	26.4150

To deal with the third limitation, a few efficient HV computation methods [3], [6], [21] have been proposed. This makes HV applicable to many-objective optimization. For example, in [3], the Monte Carlo simulation was used to estimate the HV value in high-dimensional cases. Despite helping speed up computation, estimated values cannot guarantee the accuracy of HV. Also, the estimation method is still computationally demanding when the number of objectives is very large.

On the other hand, the fourth limitation makes it hard to effectively distinguish different approximations. Li *et al.* [15] gave an example where inconsistency occurs when comparing MOEA/D [22] and IBEA [24] on a DTLZ problem, which is shown in Fig. 1. The figure clearly indicates MOEA/D is better than IBEA whereas the HV value in Table II contradicts this regardless of the setting of the reference point.

III. HYPERVOLUME-BASED DIVERSITY INDICATOR

As illustrated above, HV has the potential risk of making incorrect evaluations, particularly when diversity is concerned. In view of this, we concentrate on how to use HV to correctly measure the diversity of an approximation.

Auger *et al.* [2] proved a theorem that for a linear bi-objective POF, a set of μ points maximizes the HV indicator if and only if the points are equally spaced. Inspired by this, we can project the solutions of an approximation onto a linear POF and use the HV indicator to assess the diversity of the projections. Without loss of generality, the hyperplane $f_1 + f_2 + \dots + f_M = 1$ is considered where the approximation should be projected. The projection process is described in Algorithm 1. The projection method can nicely inherit the distribution structure from the original approximation.

Since HV here is used only for diversity evaluation, we call the new indicator HV_d . Figure 2 illustrates how HV_d works for evaluating two different distributions. According to the theorem by Auger *et al.* [2], the approximation in Fig. 2(a) will have a higher HV_d value than that in Fig. 2(b), which

Algorithm 1 Projection of A POF Approximation

- 1: **Input:** S (normalized POF approximation), $f_1 + \dots + f_M = 1$ (projection plane)
- 2: **Output:** \hat{S} (translated POF approximation).
- 3: **for** each member $s \in S$ **do**
- 4: **for** $j = 1 : M$ **do**
- 5: $\hat{f}_j(s) = f_j(s) + \frac{1 - \sum_{i=1}^M f_i(s)}{M}$;
- 6: **end for**
- 7: Save s with the translated objectives into \hat{S} ;
- 8: **end for**

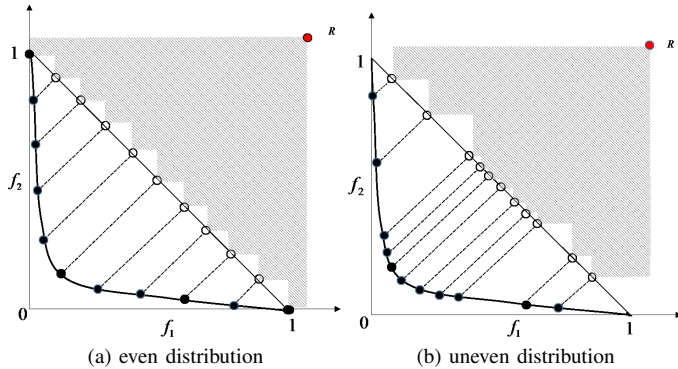


Fig. 2. Illustration of the idea behind the hypervolume-based diversity indicator. The approximations (black dotted), projections (circled), and reference points (red dotted) are presented. The hypervolume of the projections corresponds to the hatched area.

TABLE III
HV_d RESULTS OF MOEA/D AND IBEA UNDER DIFFERENT REFERENCE POINTS $R = (r, r, r)^T$

r	1.01	1.05	1.10	1.20	1.50	2.00	3.00
MOEA/D	0.9453	1.0989	1.3069	1.7796	3.7031	8.9546	29.8289
IBEA	0.9012	1.0547	1.2626	1.7349	3.6574	8.9072	29.7781

is consistent with the real diversity performance of the two considered approximations.

IV. EXPERIMENTAL STUDY

A. Evaluation of Artificial Approximations on Linear POFs

In this subsection, artificial approximations on linear POFs are used to assess the two aspects of the diversity of HV_d, i.e., extensity and uniformity in tri-objective cases. Although Auger *et al.* [2] has theoretically proved that equispaced solutions on linear POFs can maximize the HV value in bi-objective cases, but whether their theorem is applicable to tri-objective cases or not still remains unknown. For this reason, we empirically study the relationship between HV and the distribution of linear POF approximations in tri-objective cases. Note that, when the POF is linear, HV_d is equivalent to HV.

Now, let us look back to the earlier illustrated example of Fig. 1. The HV_d results of two algorithms are presented in Table III, which correctly distinguish the algorithms regardless of the choice of the reference point. This further confirms the effectiveness of HV_d.

Figures 3 and 4 present artificial examples with different scenarios of extensity and uniformity, respectively. The testing results HV_d of these examples are given for each scenario. Clearly, an approximation with a better extensity is rewarded a higher HV_d value. Similar observations are obtained in the evaluation of uniformity: the more uniform an approximation is, the better HV_d it will have. Thus, the artificial approximations show that HV_d is able to effectively evaluate and distinguish approximations with different extensities and uniformities in three-objective cases.

B. Evaluation of Real Approximations on Nonlinear POFs

To verify the effectiveness of HV_d, evaluating real approximations and nonlinear POF geometries is needed. Next, we apply HV_d to the diversity assessment of six MOEAs, i.e., NSGA-II [9], AR [4], IBEA [24], DMO [1], TDEA [13], and AR+Grid [16], on the three-objective DTLZ2 problem. These MOEAs were also used to evaluate a diversity indicator introduced by Li *et al.* [14].

The POF approximations of the six algorithms are graphically presented in Fig. 5, where the display order is arranged according to the algorithms' visual diversity. This figure illustrates that all the algorithms are well-converged but with different solution distributions. Then, HV and HV_d are used to quantify the quality of the six approximations. The reference point of HV and HV_d is placed on $(1.01, 1.01, 1.01)^T$, which is slightly larger than the nadir point of the POF of DTLZ2. All the approximations need to be normalized before being assessed.

Table IV gives the HV and HV_d values of the six approximations. The superscript of each value indicates the corresponding algorithm's rank under the corresponding indicator. Algorithms obtaining the highest values are given rank one whereas those having the worst values are ranked last.

Table IV shows that HV is unable to make a fair judgement for algorithms' diversity performance. For example, HV indicates IBEA is better than NSGA-II and DMO, which is inconsistent with the approximation plots given in Fig. 5. The inconsistency may result in a misleading conclusion for these three algorithms. On the other hand, HV_d can correctly rank the compared algorithms according to their diversity. This indicates HV_d is effective for algorithms' performance evaluation and comparison.

It is worth noting that, despite the advantage of diversity evaluation, HV_d cannot quantify the convergence aspect of an approximation. However, HV and HV_d can be used jointly to discriminate algorithms, although each has its own drawback(s). Taking NSGA-II and DMO as examples, both HV and HV_d values in Table IV show the former is better than the latter. On the basis of this, we can conclude that NSGA-II obtains a better approximation than DMO in terms of both diversity and convergence.

C. Integration of HV_d into SMS-EMOA

The previous experiments have validated the effectiveness of HV_d for assessing POF approximations' diversity. It will

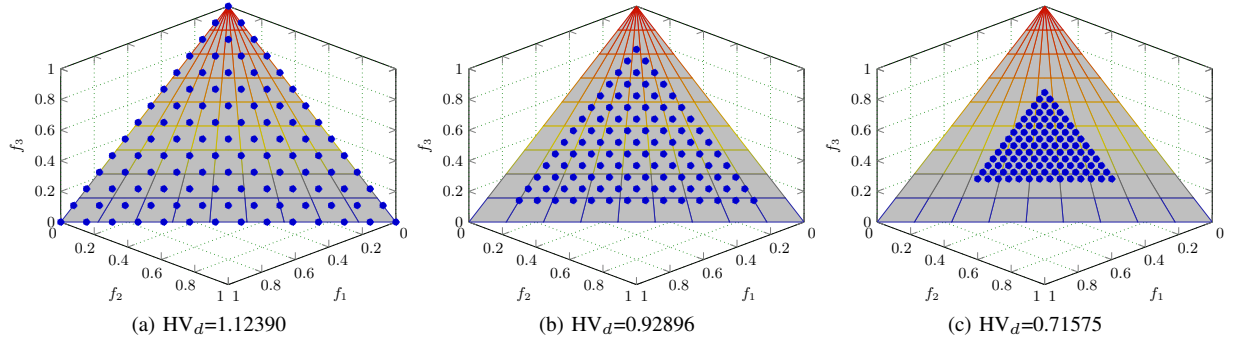


Fig. 3. HV_d testing results of artificial examples with different POF extensities. All points are located on the POF $f_1 + f_2 + f_3 = 1$.

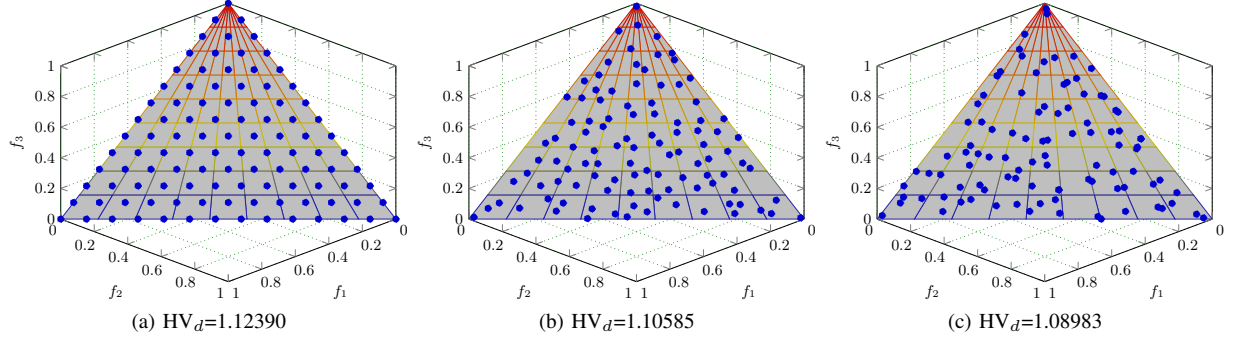


Fig. 4. HV_d results of artificial examples with different uniformities. All points are located on the POF $f_1 + f_2 + f_3 = 1$.

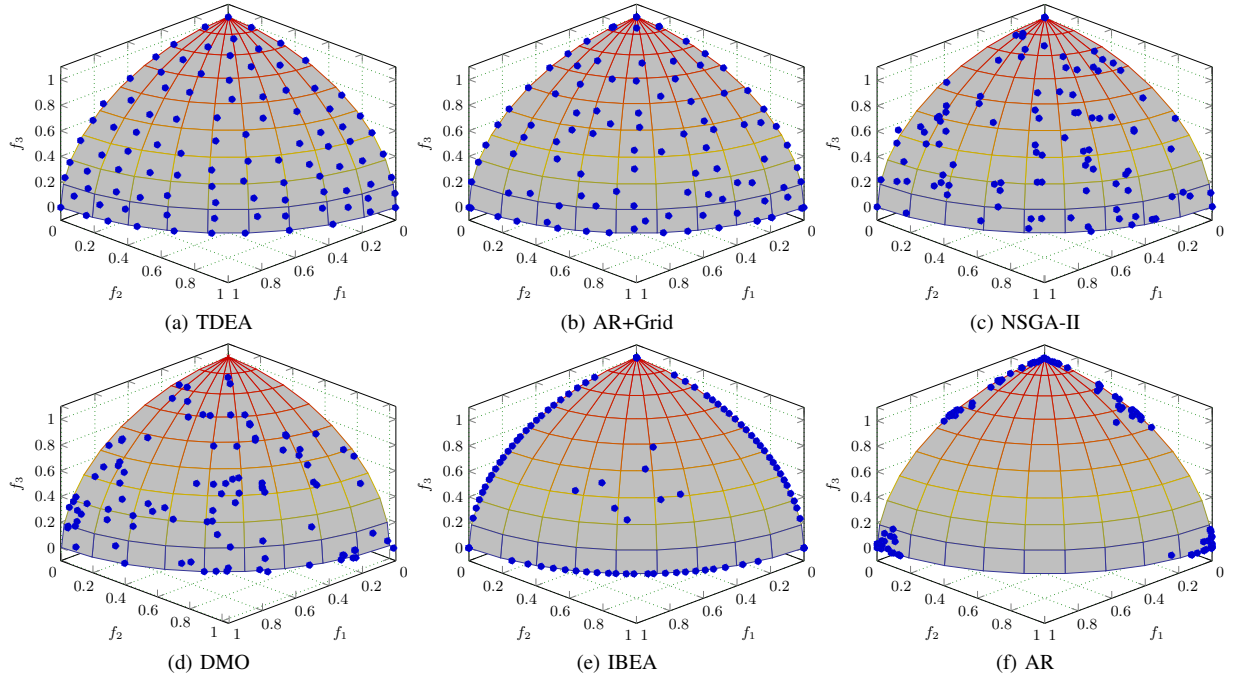


Fig. 5. HV_d results of POF approximations obtained by six algorithms on the three-objective DTLZ2.

TABLE IV
THE PERFORMANCE OF SIX ALGORITHMS MEASURED BY HV AND HV_d

Indicator	TDEA	AR+Grid	NSGA-II	DMO	IBEA	AR
HV	0.77977 ⁽²⁾	0.77982 ⁽¹⁾	0.74223 ⁽⁴⁾	0.70132 ⁽⁵⁾	0.77950 ⁽³⁾	0.56255 ⁽⁶⁾
HV_d	1.21451 ⁽¹⁾	1.20662 ⁽²⁾	1.17038 ⁽³⁾	1.15582 ⁽⁴⁾	1.15006 ⁽⁵⁾	0.78778 ⁽⁶⁾

Algorithm 2 SMS-EMOA

```
1: Input:  $N$  (population size)
2: Output:  $P$  (a POF approximation).
3: Create an initial population  $P$ ;
4: while stopping criterion not met do
5:   for  $i = 1 : N$  do
6:     Generate one offspring  $q$  by variation operators on
       two random parents from  $P$ ;
7:      $Q = P \cup q$ ;
8:      $\{R_1, \dots, R_l\} = \text{fast-nondominated-sort}(Q)$ ;
9:      $\Delta_H(s, R_l) = HV(R_l) - HV(R_l \setminus \{s\}), \forall s \in R_l$ ;
10:     $r = \text{argmin}_{s \in R_l} \Delta_H(s, R_l)$ ;
11:     $P = Q \setminus \{r\}$ ;
12:   end for
13: end while
14: Output  $P$ ;
```

be interesting to use HV_d as a quality indicator in the environmental selection of MOEAs. A straightforward way is to integrate HV_d into SMS-EMOA [5], a steady-state MOEA originally based on the HV indicator. The basic framework of SMS-EMOA is presented in Algorithm 2. It can be seen that SMS-EMOA uses nondominated sorting [9] to rank the population into different fronts and the HV indicator to discard a solution from the last front R_l considered. Here, the hypervolume contribution of a solution $s \in R_l$ is the difference $\Delta_H(s, R_l) = HV(R_l) - HV(R_l \setminus \{s\})$ between the HV of R_l and that of R_l without s . In a word, HV acts like a density estimator in environmental selection.

Here, we would like to replace HV with HV_d in SMS-EMOA while computing the hypervolume difference and see how this can affect the quality of the resulting approximations. Thus, the hypervolume difference is modified as $\Delta_H(s, R_l) = HV_d(R_l) - HV_d(R_l \setminus \{s\})$. Originally, SMS-EMOA was tested on the $EBN(\gamma)$ problem [5], which is mathematically described as follows:

$$EBN(\gamma) : \begin{cases} f_1(x) = (n^{-1} \sum_{i=1}^n |x_i|)^\gamma \\ f_2(x) = (n^{-1} \sum_{i=1}^n |1 - x_i|)^\gamma \end{cases} \quad (2)$$

where n is the number of decision variables and $x = (x_1, \dots, x_n) \in [0, 1]^n$. The parameter γ controls the curvature of the POF of $EBN(\gamma)$, thereby yielding different POF characteristics.

Despite these nice features, $EBN(\gamma)$ can only generate Pareto-optimal solutions and is not applicable for examining an algorithm's convergence performance. For this reason, we generalize the definition of $EBN(\gamma)$, and therefore construct a variant, called $EBN^*(\gamma)$, as follows:

$$EBN^*(\gamma) : \begin{cases} f_1(x) = (k^{-1} \sum_{i=1}^k |x_i|)^\gamma + g(x) \\ f_2(x) = (k^{-1} \sum_{i=1}^k |1 - x_i|)^\gamma + g(x) \\ g(x) = \sum_{i=k+1}^n (x_i - \frac{1}{i})^2 \end{cases} \quad (3)$$

where $k \leq n$ is a parameter that controls the difficulty related to diversity maintenance, and $g(x)$ is a function controlling the difficulty related to convergence. $k = n$ means that $EBN^*(\gamma)$

degenerates to the original $EBN(\gamma)$. The search space remains the same as $EBN(\gamma)$.

The population size ($N = 100$), the maximum number of function evaluations ($MaxFE = 2000$), and other key parameters remain the same as in [5]. SMS-EMOA with HV and HV_d are tested in the following scenarios of $EBN^*(\gamma)$. The reference point used in HV and HV_d is set by the worst value of each objective plus 0.5.

1) *Scenario I*: In this scenario, we focus only on the effect of different HV variants on POF diversity. Correspondingly, $(k, n) = (2, 2)$ is used in $EBN^*(\gamma)$.

Figure 6 shows a comparison between HV-based and HV_d -based SMS-EMOAs for optimizing the $EBN(\gamma)$ problem. The HV-based SMS-EMOA can distribute solutions uniformly only when the POF is linear. In the case of nonlinear POFs, SMS-EMOA with HV tends to favour knee regions, and very few solutions are placed on the boundaries. Also, when the POF is extremely shaped, e.g., $\gamma = 0.2$ or 4, extremal solutions are likely to be discarded in the HV-based selection. In contrast, SMS-EMOA based on HV_d is able to yield a set of uniformly-distributed solutions regardless of POF shapes, and extremal solutions are also well maintained.

2) *Scenario II*: In this scenario, we would like to study the convergence aspect of different HV variants on population diversity and convergence. Thus, $(k, n) = (2, 10)$ is used in $EBN^*(\gamma)$.

Figure 7 plots the POF approximations obtained by HV-based and HV_d -based SMS-EMOAs. Clearly, both HV and HV_d can help SMS-EMOA to converge toward the POF, and HV_d has advantage over HV in uniformly distributing solutions across the POF. In contrast to Scenario I, Scenario II illustrates that the convergence-related function $g(x)$ in $EBN^*(\gamma)$ affects more or less the extensity performance, particularly when the POF is irregularly shaped. As a result, both HV-based and HV_d -based SMS-EMOAs miss some boundary points, but clearly the former misses more than the latter.

3) *Scenario III*: In this scenario, we complicate diversity maintenance by increasing the value of k in $EBN^*(\gamma)$. That is, $(k, n) = (5, 10)$ is used in $EBN^*(\gamma)$.

Figure 8 presents the POF approximations on Scenario III. Clearly, the variation of k in $EBN^*(\gamma)$ influences much the performance of HV and HV_d . Specifically, compared with the plots in Scenario I and Scenario II, those displayed in Scenario III do not cover the POF very well, particularly on boundary regions. Despite that, it is very clear that HV_d -based selection performs better than HV-based selection in terms of extensity and uniformity.

The above comparisons demonstrate that HV_d can be applied in the environmental selection of indicator-based MOEAs to achieve an even distribution of approximations.

V. CONCLUSIONS

When evaluating and comparing MOEAs' performance, it is important that the adopted quality indicator is able to give a true measure of their obtained approximations. This paper has discussed the effectiveness of the popular HV indicator,

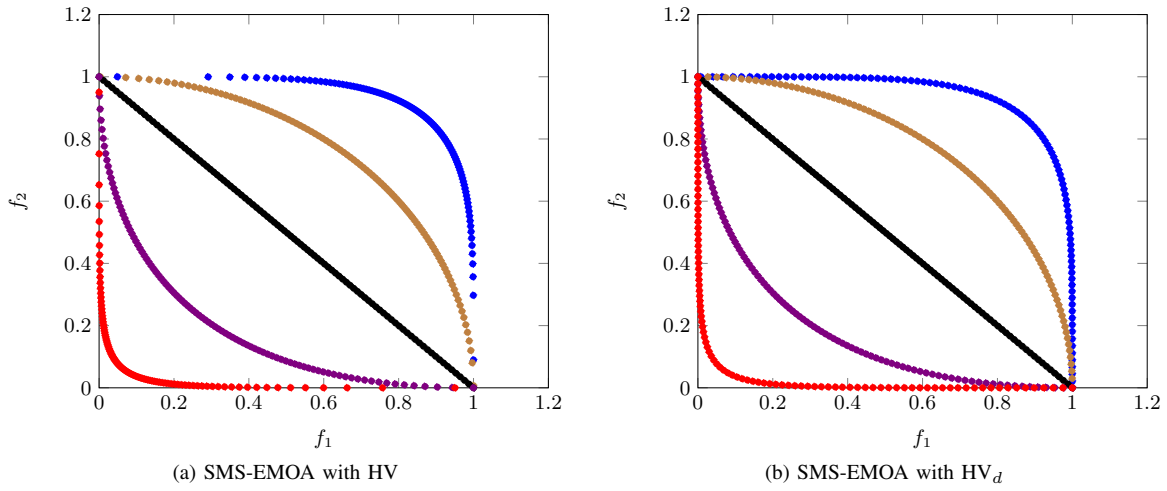


Fig. 6. A comparison of HV-based and HV_d -based SMS-EMOAs on scenario I. Approximations plotted are EB(0.2) (blue), EB(0.5) (brown), EB(1.0) (black), EB(2.0) (violet), and EB(4.0) (red).

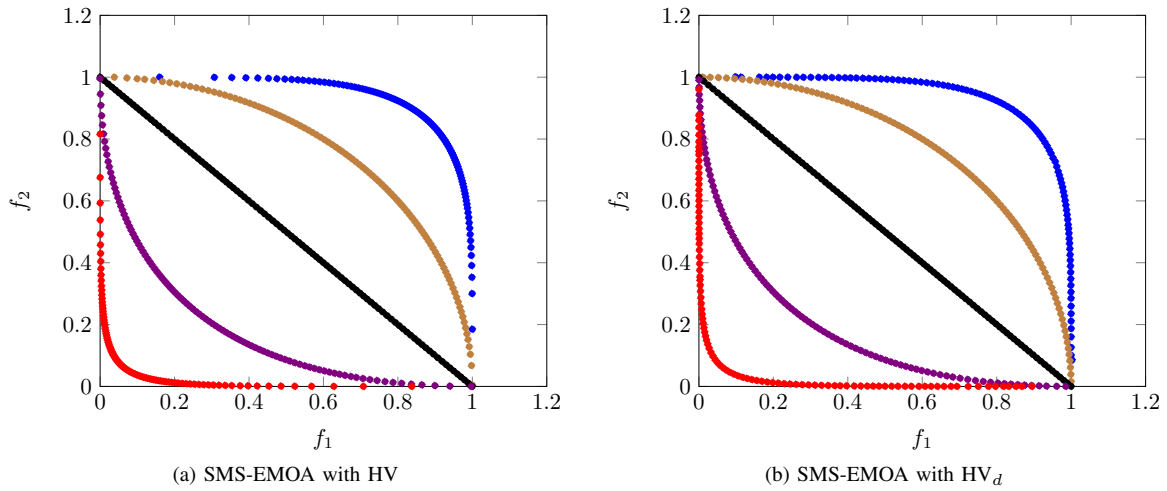


Fig. 7. A comparison of HV-based and HV_d -based SMS-EMOAs on scenario II. Approximations plotted are EB(0.2) (blue), EB(0.5) (brown), EB(1.0) (black), EB(2.0) (violet), and EB(4.0) (red).

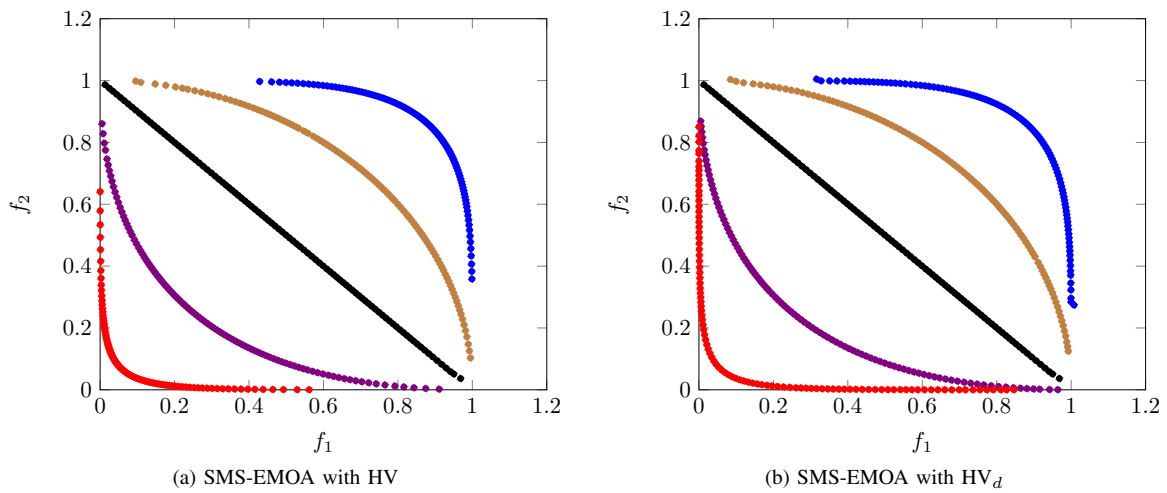


Fig. 8. A comparison of HV-based and HV_d -based SMS-EMOAs on scenario III. Approximations plotted are EB(0.2) (blue), EB(0.5) (brown), EB(1.0) (black), EB(2.0) (violet), and EB(4.0) (red).

which is the only metric known to be strictly compliant with the Pareto dominance. One drawback of HV is that it is likely to bias diversity assessment of an POF approximation. In view of this, we have suggested a modified HV indicator, i.e., HV_d , to measure MOEAs' diversity performance.

The proposed HV_d indicator has been verified on both artificial and real approximations with different solution distributions. Experimental results have demonstrated the effectiveness and correctness of HV_d for diversity evaluation. Furthermore, HV_d has been introduced into an indicator-based MOEA, i.e., SMS-EMOA, to act as a selection indicator. Compared with HV, HV_d helps SMS-EMOA to yield a more uniform distribution and a better extensivity of POF approximations.

A future study includes the sensitivity analysis of the reference point in HV_d . It will be also interesting to further investigate the behavior of HV_d on many-objective optimization.

ACKNOWLEDGMENTS

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) of U.K. under Grant EP/K001310/1.

REFERENCES

- [1] S. F. Adra and P. J. Fleming, "Diversity management in evolutionary many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 2, pp. 183–195, 2011.
- [2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, "Theory of the hypervolume indicator: Optimal μ -distributions and the choice of the reference point," in *Foundations of Genetic Algorithms (FOGA 2009)*, New York, 2009, pp. 87–102.
- [3] 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.
- [4] P. J. Bentley, and J. P. Wakefield, "Finding acceptable solutions in the Pareto-optimal range using multiobjective genetic algorithms," in *Soft Computing in Engineering Design and Manufacturing*, 1998, pp. 231–240.
- [5] 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.
- [6] L. Bradstreet, L. While, and L. Barone, "A fast incremental hypervolume algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 714–723, 2008.
- [7] D. Brockhoff, T. Wagner, and H. Trautmann, "On the properties of the R2 indicator," in *Proc. Ann. Conf. Genetic Evol. Comput. (GECCO 2012)*, 2012, pp. 465–472.
- [8] C. Coello Coello, G. B. Lamont, and D. A. van Veldhuizen, *Evolutionary Algorithms for Solving Multiobjective Problems*, 2nd Eds., Springer, New York, 2007.
- [9] K. Deb, S. Agrawal, A. Pratap, 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.
- [10] K. Deb and S. Jain, "Running performance metrics for evolutionary multi-objective optimization," in *Simulated Evolution and Learning (SEAL)*, 2002, pp. 13–20.
- [11] C. K. Goh and K. C. Tan, "An investigation on noisy environments in evolutionary multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 11, no. 3, pp. 354–381, 2007.
- [12] H. Ishibuchi, H. Masuda, H., Y. Tanigaki, and Y. Nojima, "Modified Distance calculation in generational distance and inverted generational distance," in *Evolutionary Multi-Criterion Optimization*, Lecture Notes in Computer Science, vol. 9019, 2015, pp. 110–125.
- [13] I. Karahan and M. Koksalan, "A territory defining multiobjective evolutionary algorithm and preference incorporation," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 4, pp. 636–664, 2010.
- [14] 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.
- [15] M. Li, S. Yang, and X. Liu, "A performance comparison indicator for Pareto front approximations in many-objective optimization," in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation (GECCO'15)*, 2015, pp. 703–710.
- [16] M. Li, J. Zheng, K. Li, Q. Yuan, and R. Shen, "Enhancing diversity for average ranking method in evolutionary many-objective optimization," in *Proceedings of the 11th International Conference on Parallel Problems Solving from Nature (PPSN XI)*, Lecture Notes in Computer Science, vol. 6238, 2010, pp. 647–656.
- [17] G. Lizarraga-Lizarraga, "On the evaluation of the quality of nondominated sets," Ph.D. dissertation, Center Res. Math. A.C., Computer Science Area, Guanajuato, Mexico, 2009.
- [18] S. Mostaghim and J. Teich, "A new approach on many objective diversity measurement," in *Dagstuhl Proceedings number 04461*, Dagstuhl, Germany, 2004, pp. 1–15.
- [19] O. Schütze, X. Esquivel, A. Lara, A., and C. A. Coello Coello, "Using the averaged Hausdorff distance as a performance measure in evolutionary multi-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 4, pp. 504–522, 2012.
- [20] H. Trautmann, G. Rudolph, C. Dominguez-Medina, and O. Schütze, "Finding evenly spaced Pareto fronts for three-objective optimization problems," in *EVOLVE—A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation II*, vol. 175, 2013, pp. 89–105.
- [21] L. While, L. Bradstreet, and L. Barone, "A fast way of calculating exact hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 1, pp. 86–95, 2012.
- [22] 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.
- [23] Q. Zhang, A. Zhou, and Y. Jin, "RM-MEDA: A regularity model-based multiobjective estimation of distribution algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, pp. 41–63, 2008.
- [24] E. Zitzler and S. Kunzli, "Indicator-based selection in multiobjective search," in *Proceedings of the 8th International Conference on Parallel Problems Solving from Nature (PPSN VIII)*, Lecture Notes in Computer Science, vol. 3242, 2004, pp. 832–842.
- [25] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, 2003.