



Adjusting normalization bounds to improve hypervolume based search for expensive multi-objective optimization

Bing Wang¹ · Hemant Kumar Singh¹ · Tapabrata Ray¹

Received: 18 July 2021 / Accepted: 11 November 2021 / Published online: 7 December 2021
© The Author(s) 2021

Abstract

When solving expensive multi-objective optimization problems, surrogate models are often used to reduce the number of true evaluations. Based on predictions from the surrogate models, promising candidate solutions, also referred to as infill solutions, can be identified for evaluation to expedite the search towards the optimum. This infill process in turn involves optimization of certain criteria derived from the surrogate models. In this study, predicted hypervolume maximization is considered as the infill criterion for expensive multi/many-objective optimization. In particular, we examine the effect of normalization bounds on the performance of the algorithm building on our previous study on bi-objective optimization. We propose a more scalable approach based on “surrogate corner” search that shows improved performance where some of the conventional techniques face challenges. Numerical experiments on a range of benchmark problems with up to 5 objectives demonstrate the efficacy and reliability of the proposed approach.

Keywords Normalization · Surrogate assisted optimization · Corner search

Introduction and background

Multi-objective optimization problems (MOPs) are frequently encountered in practical applications, wherein two or more conflicting performance criteria need to be optimized simultaneously [26]. The problems with more than three objectives are also commonly referred to as many-objective optimization problems (MaOPs) [22]. Furthermore, such problems can often be expensive in nature [7], when the performance estimate of each candidate solution incurs a significant expense. Commonly, the expense referred to is computational, involving time-consuming numerical simulations. However, the term is also applicable to other scenarios, such as where physical experiments or financial cost is involved in the evaluation. Evidently, the key

challenge in solving expensive optimization problems stems from the fact that a very limited number of candidate designs can be evaluated during the course of search. To achieve competitive results within such a small number of evaluations, researchers have often resorted to the use of surrogate models [4,13]; also referred to as metamodels [11]. Surrogate-assisted optimization (SAO) attempts to build approximations for the response function based on a small number of truly evaluated solutions. The predictions from the surrogate model are then primarily used to guide the search, with only occasional calls to the true evaluations. The newly evaluated solutions in turn help to train more accurate surrogate functions for subsequent search.

The promising solutions selected using the surrogate models for true evaluations are referred to as infill solutions. Typically, an infill solution is obtained by optimizing certain criterion derived from surrogates [20]. In the context of single-objective optimization, a simple and effective scheme is selecting optimal solutions based on the predicted mean values [4,27], the so-called ‘believer’ model [12]. Another widely used infill criterion is the expected improvement (EI), implemented within the efficient global optimization algorithm (EGO) [20]. The EI criterion takes both the mean prediction and the associated uncertainty of the surro-

The authors would like to acknowledge the Discovery Project (DP190102591) grant from the Australian Research Council.

✉ Bing Wang
bing.wang@adfa.edu.au

Hemant Kumar Singh
h.singh@adfa.edu.au

Tapabrata Ray
t.ray@adfa.edu.au

¹ The University of New South Wales, Canberra, Australia

gate model into consideration. Likewise, in the context of MOPs, some of the reported infill criteria include scalarized objective functions [21], hypervolume (HV) improvement, expected HV improvement, Euclidean distance improvement, maxmin distance improvement, etc. [36]. In this paper, we focus on a steady-state algorithmic framework that uses predicted HV improvement as the infill criterion.

Hypervolume (HV), also referred to as Lebesgue measure, was originally proposed for comparing the performance of multi-objective evolutionary algorithms (MOEAs) [38]. It computes the volume dominated by a given set of non-dominated solutions, bounded by a reference point R . It attempts to capture both convergence and diversity of the obtained non-dominated set and is one of the few known unary measures that are Pareto-compliant. This property implies that a solution (set) that dominates another solution (set) will have a better (higher) HV. Thus, the theoretical Pareto front (PF) has the maximum HV in the objective space. Given these interesting and desirable properties, HV has attracted significant attention from both theoretical and benchmarking perspectives [1,29,30]. Furthermore, it has become a prominent approach for environmental selection in indicator-based evolutionary algorithms [2,3,35] for solving MOPs.

There are some important considerations while utilizing HV for online/offline use. One of the concern relates to the user-specified reference point R with respect to which the dominated volume is calculated [17,30]. The common practice in this regard is to position R slightly beyond (dominated by) the nadir point Z^N . This allows for the HV contributions of the extremities of the non-dominated set to be included, while not being significantly higher than the rest of the non-dominated set. Another critical step in the application of HV is the normalization of the objective values. In MOPs, the range of different objectives can be in different orders of magnitude. Consequently, the HV calculated in the raw objective space can be biased towards the certain objective(s). To manage these two aspects simultaneously, objective vectors are first normalized between 0 and 1 using certain normalization bounds, typically defined by the ideal and nadir points in the current non-dominated set. Then, for an M -objective problem, the reference point is set to $R = 1.1^M = (1.1, 1.1, \dots, 1.1)$, as in, e.g. [35] to compute and identify the next solution(s) that have high HV contribution. Some recent works have also discussed the impact of the the reference point specification [17,30] and provided further insights and recommendations regarding its implications in benchmarking.

In MOEAs, normalization plays an important role [15], especially for estimating and enforcing diversity measures among the solutions. Normalization has been studied in

various frameworks such as NSGA-III [5], MOEA/D [14], and DBEA [32]. Normalization evidently helps in balancing the scales of different objectives. However, it also comes with a downside, that the global perspective on where these solutions lie in the raw objective space may become lost during environmental selection. This can affect the performance of the algorithm in terms of convergence and/or diversity of the population, as will be discussed in the next section. Although the effects of normalization have been discussed in the above studies, the focus was not on expensive MOPs. In this paper, we take a step towards addressing the above research gap, in the context of optimization where the sampling is derived from HV-based infill criterion. This research is motivated by our observation that the use of surrogate models in expensive MOPs provides an opportunity to adjust the normalization bounds to achieve better search performance. Thus, we are interested in developing ways that could provide better normalization bounds to aid the search, and in particular overcome the conceptual shortcomings of some of the standard normalization methods currently used.

This study builds upon the ideas we introduced in our previous conference paper [34]. In [34], the discussion and development of the methodology were limited to bi-objective scenarios. The main proposal was to obtain the two extremities of the PF based on the search on surrogate models, and use them to supplement the current non-dominated front to establish the normalization bounds. The underlying principle remains the same in this work, but we make the following extended contributions:

- We utilize a more generalized method for finding the extremities (“surrogate corners”) to adjust the normalization bounds for $M \geq 2$ objectives. The implicit assumptions specific to solving bi-objective problems [34] are thus removed.
- Since the number of corners grows with the number of objectives, we introduce additional selection mechanisms to reduce the number of true evaluations directed towards corner search in lieu of infill search.
- Lastly, we conduct numerical experiments on problems with up to 5 objectives with a limited budget to demonstrate that the proposed improvements lead to more consistency in achieving competitive results compared to other standard methods.

Following this introduction, the basic idea and algorithmic framework is outlined in Sect. 2, followed by numerical experiments and discussion in Sect. 3. Concluding remarks and future work are discussed in Sect. 4.

Methodology

General idea

As mentioned previously, a common approach to identify normalization bounds is based on the limits of the current non-dominated (ND) set of solutions. The technique is simple, intuitive and often effective. However, a potential drawback of this procedure is that if the current ND front is biased towards a certain part of the PF, the resulting infill search then also tends to be limited in exploring that part of the PF. This could potentially lead to delay in exploring other regions, or in the worst case, yield only a small subset of the PF as the final output. This situation is visualized in Fig. 1a. In this snapshot of the search, the ND front is concentrated towards one end of the PF. If normalization is conducted using the bounds of this ND front, along with the standard reference point specification (1.1^M), the infill search is bounded by the region enclosed by the black dotted lines. Since the resulting infill point (in the predicted space) will end up only from this region, the HV contribution of any point outside these bounds with respect to R is zero. Consequently, the improvement in the quality of the current solution set is likely to be small.

An alternative explanation of the shortcoming is demonstrated in Fig. 1a is that the current estimates of ideal and nadir points (Z^I and Z^N , respectively) are not accurate relative to the true ideal and nadir points (Z^{*I} , Z^{*N}). The main idea behind the proposed method is therefore to improve the estimation of these points to come up with adjusted normalization bounds. We propose to do so by searching for the extremities of the PF based on the surrogate models. The search on the surrogate models itself does not add to the cost of true evaluations. Only after the surrogate models have identified the extreme points, such solutions would need to undergo true evaluation before being added to the current ND set to identify the new normalization bounds. This process will consume some (small number of) true evaluations and hence does not come at a zero cost. Hence, for the strategy to be useful, the cost directed towards evaluating the extreme points should be (a) reduced to the extent possible, and (b) should eventuate in bringing relative benefits to the search.

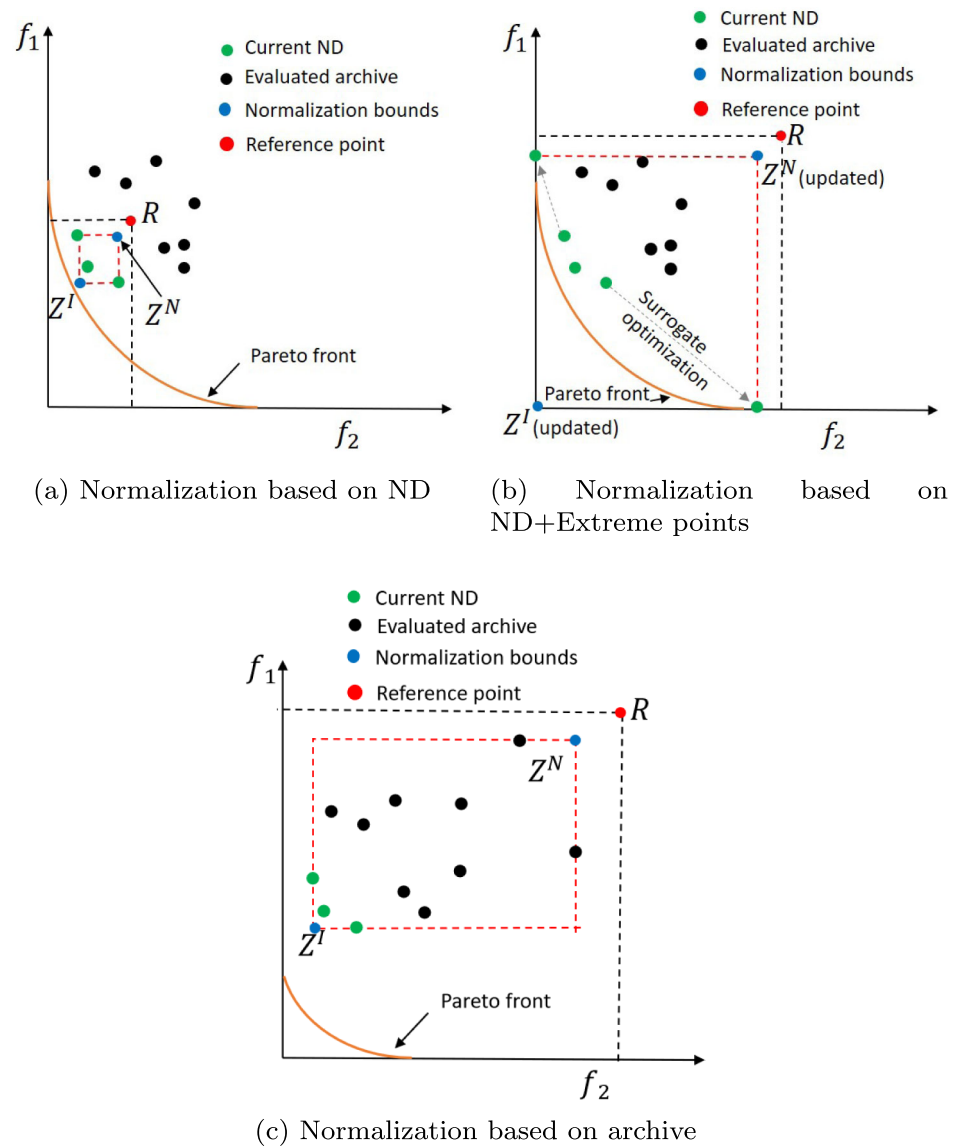
Since the number of such additional evaluations due to corner point evaluations is not disproportionately high in the above paradigm, we hypothesize that surrogate-assisted optimization can particularly benefit from such normalization adjustment. This is in stark contrast for attempting the same process in a non-surrogate optimization frameworks, where obtaining the extreme points themselves could take up a substantial portion of the computational budget [32]. The above discussion is visualized in Fig. 1b. Starting from the current ND front, we search for an extreme point along each

objective using its surrogate model. Subject to the models being good representations of the true functions, the extreme points on PF are identified, and estimation of nadir and ideal points is improved. This process yields updated normalization bounds (Z^I (updated) and Z^N (updated)) and the updated reference point. This process expands the search space of infill search, and has the potential of uncovering more diverse solutions on the PF. Of course, it is anticipated that the exact extremities may not be identified in one shot; instead, the search may need to be invoked more than once during the overall optimization process to progressively approach the true values. Additionally, the surrogates are not likely to be accurate from the beginning for generic non-linear functions; but rather start from a coarse approximation that becomes better over time as more solutions are evaluated. However, even with relatively coarse models, the surrogates may be able to capture general global trend, leading the search to global optimal regions. Where this doesn't happen, the evaluation of the identified corner point instead brings benefit in improving the surrogate models for subsequent iterations.

Also worth noting is another type of normalization scheme that is based on the whole archive of the solutions evaluated so far [23,36]. The strategy may be of interest when the overall numbers of solutions evaluated are small, which is a scenario consistent with expensive MOPs. The archive-based normalization bounds are less likely to be localized compared to the ND-based bounds. However, it may suffer from a different issue. When the evaluated solutions are located far away from the true PF, the solutions that may not be near the PF start to contribute significantly to the HV. Consequently, the HV maximization for infill search does not translate to the selection pressure required to emphasize the diversity on the true PF itself. This situation is illustrated in Fig. 1c, where the infill search ends up with significantly expanded search area beyond the PF. As a result, solutions that lie beyond the true PF boundaries could be suggested as infill, adversely affecting the convergence (and diversity) of final PF approximation.

The proposed approach intends to overcome the above-mentioned limitations to yield a more consistent performance across a range of problems. In our previous attempt [34], we applied the extreme point search on each objective function independently to adjust the bounds, which works well for bi-objective problems due to a deterministic number ($=2$) of extremities. Beyond two objectives, the definition of extreme points becomes somewhat subjective, hence implementation of the method needs some further considerations. To make the approach more generic, we present here a modified version of the Pareto corner search [31] and introduce further improvements to obtain the adjusted bounds. In the following sections, we introduce the surrogate corner search that is used for normalization adjustment.

Fig. 1 Illustration of normalization schemes



Corner points in multi-objective problems

A unique feature for bi-objective problems is that for a given non-dominated front (including the PF), the extremities (also referred to as corners [31]) have a clear and intuitive geometric interpretation. One extremity has the lowest value of f_1 and the highest value of f_2 within the set, and the reverse is true for the other extremity. This situation is shown in Fig. 2a. Thus, the two extremities can in principle be identified by minimizing one objective at a time independently. The only caveat is that if an objective function has multiple global optima, then the second objective needs to be subsequently considered to pick the correct extremity.

For MOPs with $M \geq 3$, the above does not hold. Minimizing the objectives independently may not lead to unique points on PF that can be considered as corners. Moreover,

minimizing a given objective may not result in an extreme value for any of the remaining objectives. Consider, for example, the 3-objective PF shown in Fig. 2b, where minimizing f_3 alone can lead to any of the solutions in the PF that lie on the $f_1 - f_2$ plane. Out of these, only the two extremes can be intuitively perceived as geometric corners. Figure 2c on the other hand shows a case where there is a unique point corresponding to minimum f_3 (and similarly, other objectives), which can be considered as a corner.

A formal, though restrictive, definition of corner point was provided in [31] in the context of dimensionality reduction. It suggested that if the minimization of a subset of $k < M$ objectives led to a single point, then the point can be considered as a corner. Thus a corner solution can minimize anywhere from 1 to $M - 1$ objectives, leading to up to total $2^M - 1$ possible combinations. Evidently, as the number of

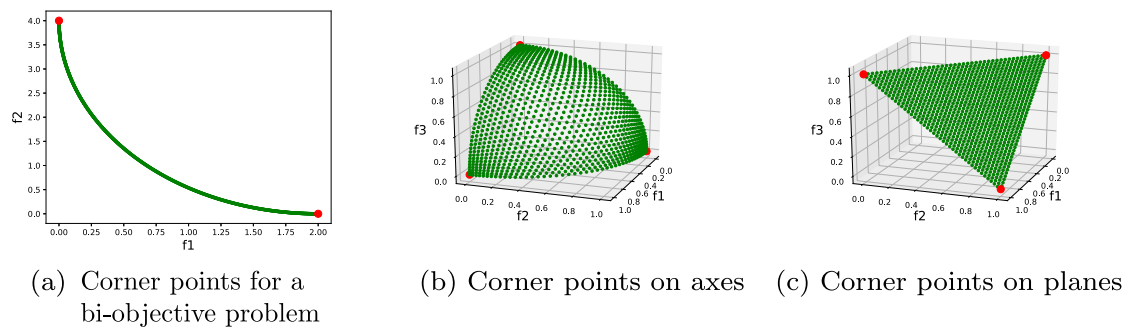


Fig. 2 Some examples of corner points (shown in red)

objectives increase, searching for all combinations becomes tedious and impractical. Mostly, the common MOPs have much fewer corners, for e.g., there are M corners for most non-degenerate M -objective problems in DTLZ and WFG benchmarks [8,16]. Considering this, two specific subsets of corners were deemed to be of interest in [31]: *Type-1 corners*, that only minimize one objective value (Fig. 2c), and *Type-2 corners*, that minimize $M - 1$ objectives simultaneously (Fig. 2b).

It is additionally worth noting that theoretically any simplex with an arbitrary number of vertices (e.g., a star/square) can be projected on these PF surfaces, comprising potential PF shapes that could have any number of perceived “corners”. Such cases are not covered (and it is subjective whether they should be considered as corners) in the definitions discussed above. Nonetheless, given that in the surrogate corner search we are primarily concerned with objective value bounds rather than internal geometric features of the PF, they still seem sufficient for the task at hand.

Corner search scheme

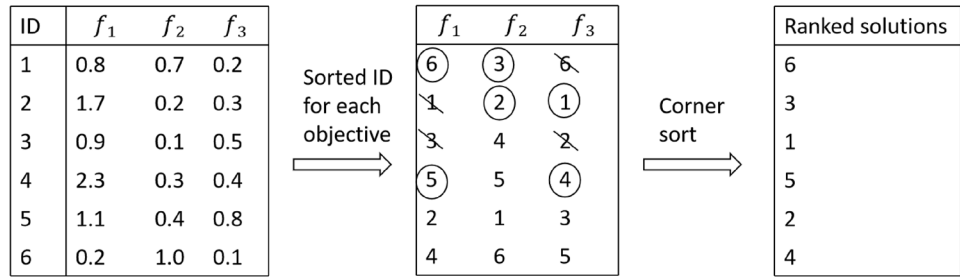
From the above discussions, it is clear that as the number of objectives grow, independent search along the objectives may lead to increasing possibility of identifying solutions that are far from the intended corners. This will in turn lead to normalization bounds that are not suitable. Hence, the search for the corners need to be conducted concurrently so that the resulting solutions are likely to be closer to the true PF extremities. Therefore, in this paper, we adapt the corner sort ranking from [31] to simultaneously search for corners using a single evolutionary algorithm (EA) that can be used to establish the normalization bounds. An illustration is given shortly, but we would point out an important change in the ranking compared to [31]. That is, we do not consider the L_2 norm of $M - 1$ objectives during the sorting process; which was originally intended to search for Type-2 corners. The reason is that our application context is expensive problems where the evaluation budget needs to be spent carefully. Even though corner search has the potential to improve normaliza-

tion bounds, too many evaluations spent on it will leave much fewer of them for the infill search, reducing the anticipated benefits of the improved normalization bounds. To strike a balance between the true evaluations spent on corner evaluations and infill evaluations, we exclude the ranking based on Type-2 corners; thus targeting a maximum of M possible corners of Type-1. The details of the adapted corner search is presented below.

The general goal of corner search discussed here is to identify Type-1 corners as defined in Sect. 2.2, i.e., solutions that minimize each objective individually. An intuitive approach could be running a single-objective optimization sequentially for each objective, however the search for one objective’s optimum will not have any knowledge from the points explored for the other objective when using such approach. Instead, the corner search aims to achieve this with a single population in a more efficient and scalable manner. The ranking scheme plays an essential role in enforcing the preferences for the corner points in this method. The ranking can be better explained with an illustrative example shown in Fig. 3.

In Fig. 3, the sample population for a 3-objective problem has six solutions with their IDs marked as 1–6. The intent of sorting is that for each objective, a better solution is ranked higher in the final list. Since we have multiple objectives, this is done in a sequential manner while avoiding duplication. The left block shows the 6 solutions and their objective values. First, each objective is sorted in ascending order independently (minimization is assumed). The sorted IDs of solutions are listed for each objective as shown in the middle block. Next, top solution of each objective is picked in turn (e.g. from left to right) to form the final sorted list. Note that each time a solution ID is picked based on a given objective, this ID is deleted from the sorted solutions available for the rest of the objectives. The solutions selected following this sequence are marked with a circle in the middle block. From the sorted list for f_1 , solution #6 is selected and for f_2 , the solution #3 is selected. For objective f_3 , since the solution #6 has already been picked (crossed out), the top among the remaining solutions, #1, is selected. Now in the

Fig. 3 Corner sort procedure implemented in this study



final sorted list, we have three solutions #6, #3, #1. Next we restart from objective f_1 . Since the solutions #1 and #3 are already picked in the last round (crossed out), the available solutions start from #5, so solution #5 is selected for f_1 . For objective f_2 , the next best solution is #2 and it is still available, so it is selected for objective f_2 . Next, for objective f_3 , since #2 is already picked in the last step, available solutions start from #4, therefore #4 is selected. All solutions have a unique rank now (in the sequence they were picked), hence the sorting procedure terminates here. The final sorted order is shown in the right block in Fig. 3.

With the above as the ranking process, the corner-search EA based on the surrogate models is run with a population size N evolving over G generations (this part incurs no true evaluations). Then, from the final population, the corners need to be selected for true evaluations. The selection strategies are discussed in the next section.

Corner selection for true evaluation

The most straight forward way is to pick up top M solutions directly from the ND front of the final population. This baseline strategy consumes exactly M true evaluations. However, this number can be reduced. In some cases, the corners obtained by optimizing different objectives can return the same (or similar) solutions. For example, in the Fig. 4, red corner points can assist in preserving low f_1 values or low f_2 values. In this case, there is no need to evaluate corner solutions that are extremely close to each other. Instead, the ND front of the final corner search EA population is analyzed using Silhouette analysis [28] to identify the optimal number of clusters in the set. Clustering is conducted using the K-means method [24]. In silhouette analysis, we calculate the silhouette score for each possible cluster number (from 1 to M). Silhouette score calculates the separation distance between the clusters for each cluster number. The highest score determines the optimal cluster number. From this number of clusters, we can decide how many solutions should be extracted for true evaluation. In each cluster, the solution that has the highest rank in the corner sort is selected. Thus, for the case shown in Fig. 4, even though $M = 3$, only two solu-

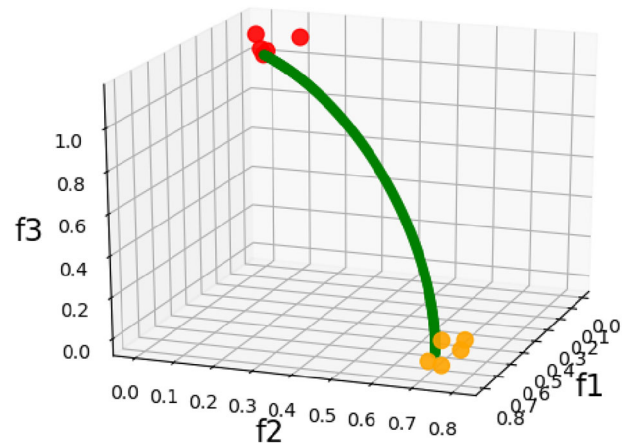


Fig. 4 Reducing the number of corners using Silhouette analysis

tions, one from each of the clusters at the two ends of the PF, will be selected for evaluation.

Another approach to reduce the candidate solutions before evaluation is to observe their location with respect to the current reference point. Recall that the main objective of identifying these solutions is to adjust the range of selection to appropriate bounds. Therefore, if the predicted values of the solution do not affect the current bounds enclosed by the reference point, then their evaluation will not provide the related benefit. This consideration is illustrated in Fig. 5. The surrogate corners may often be located close to and along the edge of the current normalization boundary. We should consider the benefit they bring in terms of HV improvement, especially with a higher number of objectives. As shown in Fig. 5, the surrogate corner under reference bound (black line) will have a much less HV contribution (purple shade) than a normal infill point (green shade). Therefore, such surrogate corner is ignored. However, surrogate corner beyond the reference bound (black line), has better potential in terms of adjusting the normalization bound. We only evaluate such corners. Thus in this case, out of the two potential corners identified via corner search, only the one outside the box enclosed by R is evaluated.

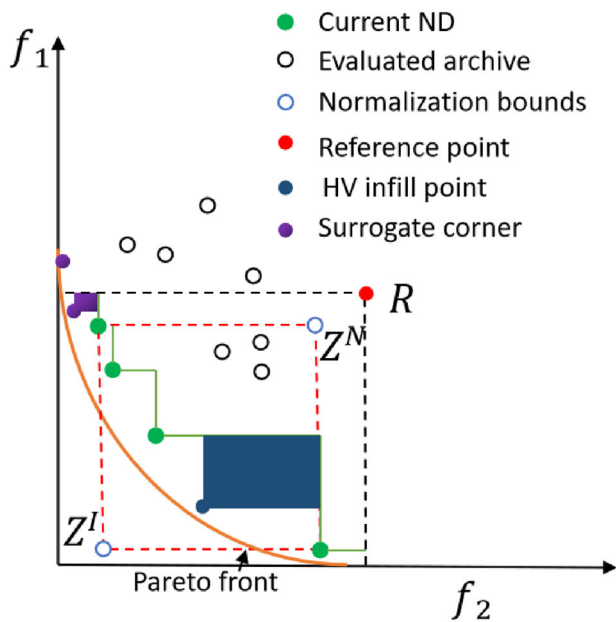


Fig. 5 Selection of corners based on their location relative to R

Algorithm implementation

The pseudo-code of the overall framework is presented in Algo. 1. The main algorithm starts with the generation of an initial set of designs using the Latin Hypercube Sampling (LHS), a widely used method for design of experiments. The designs are then truly evaluated and form the initial archive \mathcal{A} . Surrogate models are built for each objective function based on the archive of solutions. This archive \mathcal{A} of all truly evaluated solutions is updated in each iteration as the new solution(s) get evaluated. Corners are identified by corner search outlined in Algo. 2. Importantly, note that the corner search is not needed in every iteration. It is only invoked after an infill point is found that is better than the current ideal point (Z^I) in any of the objectives, as indicated in Algo. 1, line 5. The obtained corners are then selectively evaluated using the components discussed above (we discuss three variants in numerical experiments). All evaluated solutions are then added to the archive. The maximum and minimum values among the existing non-dominated solutions are used as the normalization bounds. Using reference point $R = 1.1^M$, an HV maximization is conducted to obtain the infill solution, which is then evaluated and added to the archive. The surrogate models, archive and number of evaluations are updated, and the process repeats until the evaluation budget is exhausted.

In the above process, we have used Kriging [25] as the surrogate model for approximating the objective functions in Lines 3, 12 and 17 of Algo. 1. Kriging is one of the widely used surrogate models for approximating generic non-linear functions. For HV maximization via infill identification in

Algorithm 1 Proposed HV infill-based algorithm

Input: Max. number of function evaluations FE_{max} , Initial sample size N , normalization bounds search related parameters, Infill search related parameters
Output: PF approximation

- 1: Initialize archive of evaluation solutions, $\mathcal{A} = \emptyset$, No. of evaluations $FE = 0$.
- 2: Sample N initial designs & evaluate them. Update \mathcal{A} , FE .
- 3: Build surrogate models $\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2 \dots \mathcal{K}_M\}$
- 4: **while** ($FE < FE_{max}$) **do**
- 5: **if** First iteration or infill point is better than current ideal on any objective **then**
- 6: Search for surrogate corner solutions $\mathbf{E}_1, \mathbf{E}_2 \dots \mathbf{E}_S$ with Algorithm 2
- 7: **for** $\mathbf{E} \in \{\mathbf{E}_1, \mathbf{E}_2 \dots \mathbf{E}_S\}$ **do**
- 8: Selectively evaluate \mathbf{E}
- 9: **end for**
- 10: Update \mathcal{A} , FE
- 11: Set Z^I, Z^N based on non-dominated solutions $\mathcal{N} \subseteq \mathcal{A}$
- 12: Update \mathcal{K}
- 13: **end if**
- 14: Identify infill solution \mathbf{x}_{HV} by maximising HV with reference point R
- 15: Evaluate \mathbf{x}_{HV} . Update \mathcal{A} , FE
- 16: Set Z^I, Z^N based on non-dominated solutions $\mathcal{N} \subseteq \mathcal{A}$
- 17: Update \mathcal{K}
- 18: **end while**
- 19: Output the non-dominated solutions $\mathcal{N} \subseteq \mathcal{A}$ as the PF approximation

Algorithm 2 Corner search for identifying corner points

Input: Surrogate models $\mathcal{K}_1, \mathcal{K}_2, \dots \mathcal{K}_M$, population size N_I , number of generations G_I , evolution parameters
Output: Corner solutions $\mathbf{E}_1, \mathbf{E}_2 \dots \mathbf{E}_S$; $S \leq M$

Initialize population P_I of size N_I .

for $g = 1$ to G_I **do**

Apply evolutionary operator to generate child population C_I of size N_I

Apply corner sort on $P_I \cup C_I$

Select the top N_I solutions in the above sorted list for the next generation P_I

end for

In the last population's ND front, apply Silhouette analysis to determine the number of clusters S .

In each cluster select the solution with the best corner sort rank

Return these selected corner points $\mathbf{E}_1, \mathbf{E}_2 \dots \mathbf{E}_S$

Line 14 of Algo. 1, we have used differential evolution (DE) [33] consistent with the previous work [34]. For the corner search in Algo 2, we use the same approach as done in PCSEA [31] and use simulated binary crossover (SBX) [9] and polynomial mutation (PM) [10] as the variation operators. For removing the dominance resistance solutions, for each objective, the tolerance $\epsilon = 1e - 5$ is used as done in [34]. The code for the proposed method can be obtained from the authors' website http://mdolab.net/research_resources.html

Numerical experiments

Benchmark problems used

To demonstrate the performance of the proposed approach, numerical experiments are conducted on a range of test problems widely used in the evolutionary multi-objective optimization domain. These include problems from the ZDT [37], DTLZ [8], WFG [16] and MAF [6], inverted DTLZ [19] and minus DTLZ [18] series, to cover a wide range of problem characteristics and PF shapes. The settings used for the experiments are listed in Table 1. In our preceding work [34], we only considered bi-objective problem suits, one of which was the ZDT series that is not scalable beyond 2 objectives. Hence, we have added MAF series to complement DTLZ and WFG for $M \geq 3$. For consistency with our previous paper, we keep the 2-objective experiments on the same problem set (ZDT, DTLZ, WFG) and include additional inverted and minus DTLZ series (iDTLZ and mDTLZ for short) problems. For 3-objective and 5-objective problems, we have used DTLZ (standard, inverted and minus series), WFG and MAF series. Since MAF1 is designed from inverted DTLZ1, to eliminate duplication, we exclude inverted DTLZ1 for 3-objective and 5-objective and keep MAF1. Some adjustments are made in the ruggedness and bias parameters of the problems so that they are not too difficult for approximation; as the focus of this study is on normalization bounds rather than improving surrogate accuracy. Accordingly, we have set ruggedness parameter = 2π and scaling factor for $g(\mathbf{x}) = 1$ for DTLZ1, DTLZ3 and bias parameter $b_{poly} = 0.5$ for WFG. For MAF3 and MAF4 problems, the ruggedness parameter is reduced to π from 20π , scaling factor in g is reduced to 1 from 100. Moreover, the type of surrogate (Kriging) is kept the same for all compared strategies, so that the limitations in approximation, where applicable, affect all strategies in a similar manner. There are also three problems (DTLZ7, MAF2, MAF6), which are not extendable to 5-objective and

therefore, they have not been considered in the 5-objective experiments.

Algorithm and experimental settings

The Kriging model used in this study is based on the implementation of the DACE toolbox [25] in Python. The regression function is set to be zero-order polynomial, while the correlation model is set to be Gaussian. Kriging surrogate is chosen since it is extensively used in surrogate-assisted optimization studies. The infill strategy is based on the predicted values (“believer” approach) for all the compared strategies. Other infill strategies, such as expected improvement, are excluded in comparisons as they involve additional factors (such as uncertainties) whose influence is not the subject of the current study. HV is used as the metric for performance comparison of the final solutions, in the objective space normalized by the true ideal and nadir points with the reference point as 1.1^M . A total of 29 independent runs are conducted for each problem to observe the statistical behavior of performance. The performance is compared across six different normalization methods. These include the two standard methods (ND and archive) and our method from the preceding work [34]. As for the proposed method, we consider three variants, by gradually activating the selective evaluation and Silhouette analysis to ascertain their contribution in the overall performance. For reference, the six compared methods are listed below. The main comparison is how each of the four extreme point/corner point-based methods compare with the standard approach of using ND-based or full archive-based approach for identifying the normalization bounds.

- $NormR_{ND}$, normalization bounds based on current non-dominated solutions.
- $NormR_A$, normalization bounds based on full archive.

Table 1 Settings for numerical experiments

	FEs.	Initialization	D (no. of variables)	M (no. of objectives)
Problem-specific settings				
ZDT	200	$0.5 \times FE$	6	2
DTLZ	200/300/400	$0.5 \times FE$	6	2/3/5
WFG	200/300/400	$0.5 \times FE$	6	2/3/5
MAF	-/300/400	$0.5 \times FE$	6	-/3/5
	Pop. size	Gen.	Evolutionary parameters	
Evolutionary search settings				
Corner search	100	100	0.8 (crossover rate) 20 (crossover index)	0.2 (mutation rate) 30 (mutation index)
Infill search	100	100	0.8 (crossover rate)	0.8 (scaling factor)

- $NormR_{NDE}$, normalization bounds based on augmented ND solutions with extreme points identified from independent objective search, the method proposed in the preceding work [34].
- $NormR_{NDC}(S1)$, the proposed method that augments current ND solutions with corner points to define the normalization bounds. In this variant, all top M corners in the final population of the corner search are evaluated.
- $NormR_{NDC}(S2)$ the proposed method that augments current ND solutions with selected corner points to define the normalization bounds. Selective evaluation (as discussed in Sect. 2.4) is applied on the top M solutions in the final population of the surrogate corner search. Thus, among these M points, only those that are non-dominated with respect to the current ND front and lie outside the reference point bounds based on predicted values are evaluated. Silhouette analysis is not used in this variant.
- $NormR_{NDC}(S3)$ the final version of the proposed method (described in Algo. 1 and 2), that augments current ND solutions with selected corner points to define the normalization bounds. Both Silhouette analysis and selective evaluation are activated in this variant, in that order.

Results and discussion

The results obtained for 2-, 3- and 5- objective problems are summarized in Tables 2, 3 and 4, respectively. In the following sections, we discuss the obtained results from three different perspectives. These include the proposed method compared with the conventional approaches (archive and ND front based normalization), performance comparison between the surrogate-based corner search and the extreme point search proposed in our previous conference paper, and the internal comparison between the three variants (S1–S3) of the proposed approach.

Comparison with conventional normalization methods (archive- and ND-based)

In our preliminary work [34], we had highlighted three typical problem characteristics that influenced the normalization performance: initialization position (i.e., whether the initial solutions are close to or far from the PF), coverage of the PF (i.e., whether the initial solutions already provide a good estimate of ideal by encompassing the PF range), and optimization behavior on the surrogate model (i.e., whether the search on the surrogate models led to a significantly better estimate of the extreme points). Both the conventional approaches, i.e., archive-based and ND-based normalization, face challenges on problems with certain characteristics.

Below, we look into whether the proposed surrogate corner search can provide a more consistent and competitive performance across these problems across varying number of objectives.

- Comparison with archive-based normalization:

- For the archive-based normalization, it typically faces challenges when initialized solutions span a much larger range of objective values compared to the PF. As discussed earlier, the implication of using a reference point far away is that the solutions that are not on the PF may have a high contribution, which takes away the selection pressure to drive the solutions to a well converged and diverse PF approximation. On the other hand, if the initialized solutions are close to the PF and span its range well, it has the advantage of identifying good nadir and ideal approximation without additional effort. We observed that the typical scenario reflecting the first situation is common in DTLZ (except DTLZ7) problems and MAF problems. The representative cases of the latter scenario are common in the WFG, ZDT and DTLZ7 problems. From Tables 2, 3 and 4, we observe that the proposed $NormR_{NDC}$ (S3) either outperforms or achieves equivalent results compared to the archive-based method ($NormR_A$) in nearly all cases. This verifies that for the cases where the initial population was far away, the proposed approach was able to provide tighter bounds for the HV infill search by bringing the reference point much closer to the true Z^I and Z^N .
- To verify the above case where $NormR_A$ faced challenges, we observe the search behavior on DTLZ1 closely in Fig. 6. Figure 6a shows the initial solutions for the 2-objective DTLZ1, its normalization bounds (orange box), and the position of the reference point. Evidently, due to the distribution of initial solutions, reference point is placed far away from its PF. In contrast, Fig. 6b shows the normalization bounds and location of the reference point suggested by the proposed strategy ($NormR_{NDC}$) using the same initial solutions. It is clear that the performance of the latter is likely to be better than $NormR_A$ strategy. It is expected that the convergence of $NormR_A$ will be inferior in this case, which indeed is the case as can be seen from Fig. 6c.
- The two scenarios discussed above are further visualized using 3-objective problems in Figs. 7 and 8. In Fig. 7 for DTLZ7 problem, initial population covers the PF range well, and the solutions are not too far away from the PF. Consequently, the archive-based normalization has the advantage on covering

Table 2 Median HV performance in test problems with 2 objectives across 29 independent runs

Problems	$NormR_A$ (archive)	$NormR_{ND}$ (ND)	$NormR_{NDE}$ [34]	$NormR_{NDC}$ (S1)	$NormR_{NDC}$ (S2)	$NormR_{NDC}$ (S3)
mDTLZ1	0.7011	0.7014	0.7020 $\uparrow_1 \approx_2$	0.7009 $\approx_1 \approx_2$	0.7012 $\approx_1 \approx_2$	0.7016 $\approx_1 \approx_2$
mDTLZ2	0.9915	0.9915	0.9915 $\downarrow_1 \downarrow_2$	0.9915 $\downarrow_1 \downarrow_2$	0.9915 $\approx_1 \downarrow_2$	0.9915 $\approx_1 \approx_2$
mDTLZ3	0.9908	0.991	0.9906 $\approx_1 \downarrow_2$	0.9905 $\downarrow_1 \downarrow_2$	0.9905 $\approx_1 \downarrow_2$	0.9906 $\approx_1 \downarrow_2$
mDTLZ4	0.7918	0.8265	0.8131 $\approx_1 \approx_2$	0.7941 $\approx_1 \approx_2$	0.8179 $\approx_1 \approx_2$	0.8335 $\approx_1 \approx_2$
iDTLZ1	0.2881	0.7007	0.7005 $\uparrow_1 \approx_2$	0.7002 $\uparrow_1 \downarrow_2$	0.7005 $\uparrow_1 \approx_2$	0.7006 $\uparrow_1 \approx_2$
iDTLZ2	0.9715	0.9866	0.9855 $\uparrow_1 \downarrow_2$	0.9857 $\uparrow_1 \downarrow_2$	0.9862 $\uparrow_1 \approx_2$	0.9857 $\uparrow_1 \downarrow_2$
ZDT1	0.8649	0.4646	0.8664 $\uparrow_1 \uparrow_2$	0.8659 $\uparrow_1 \uparrow_2$	0.8659 $\uparrow_1 \uparrow_2$	0.8661 $\uparrow_1 \uparrow_2$
ZDT2	0.5360	0.5386	0.5387 $\uparrow_1 \uparrow_2$	0.5385 $\uparrow_1 \downarrow_2$	0.5386 $\uparrow_1 \approx_2$	0.5386 $\uparrow_1 \approx_2$
ZDT3	0.7196	0.2973	0.7193 $\approx_1 \uparrow_2$	0.6500 $\downarrow_1 \uparrow_2$	0.7025 $\downarrow_1 \uparrow_2$	0.7201 $\approx_1 \uparrow_2$
ZDT6	0.4624	0.5105	0.4961 $\approx_1 \approx_2$	0.4470 $\approx_1 \downarrow_2$	0.4736 $\approx_1 \approx_2$	0.5049 $\uparrow_1 \approx_2$
DTLZ1	0.3111	0.7008	0.7005 $\uparrow_1 \approx_2$	0.7002 $\uparrow_1 \downarrow_2$	0.7003 $\uparrow_1 \downarrow_2$	0.7004 $\uparrow_1 \downarrow_2$
DTLZ2	0.4095	0.4198	0.4196 $\uparrow_1 \downarrow_2$	0.4195 $\uparrow_1 \downarrow_2$	0.4196 $\uparrow_1 \downarrow_2$	0.4197 $\uparrow_1 \downarrow_2$
DTLZ3	0.0563	0.4155	0.4137 $\uparrow_1 \downarrow_2$	0.4135 $\uparrow_1 \downarrow_2$	0.4134 $\uparrow_1 \downarrow_2$	0.4133 $\uparrow_1 \downarrow_2$
DTLZ4	0.1478	0.1522	0.2140 $\uparrow_1 \uparrow_2$	0.1838 $\approx_1 \approx_2$	0.2043 $\uparrow_1 \approx_2$	0.1989 $\uparrow_1 \uparrow_2$
DTLZ7	0.5427	0.3534	0.5447 $\uparrow_1 \uparrow_2$	0.5446 $\uparrow_1 \uparrow_2$	0.5447 $\uparrow_1 \uparrow_2$	0.5447 $\uparrow_1 \uparrow_2$
WFG1	0.1893	0.1900	0.1841 $\downarrow_1 \downarrow_2$	0.1761 $\downarrow_1 \downarrow_2$	0.1872 $\approx_1 \approx_2$	0.1867 $\approx_1 \approx_2$
WFG2	0.6617	0.6520	0.6695 $\approx_1 \approx_2$	0.6776 $\uparrow_1 \uparrow_2$	0.6506 $\approx_1 \approx_2$	0.6569 $\approx_1 \approx_2$
WFG3	0.6727	0.6655	0.6742 $\approx_1 \uparrow_2$	0.6723 $\approx_1 \uparrow_2$	0.6718 $\approx_1 \uparrow_2$	0.6729 $\approx_1 \uparrow_2$
WFG4	0.3476	0.3418	0.3463 $\approx_1 \uparrow_2$	0.349 $\approx_1 \uparrow_2$	0.3474 $\approx_1 \uparrow_2$	0.3468 $\approx_1 \uparrow_2$
WFG5	0.3646	0.3627	0.3634 $\downarrow_1 \approx_2$	0.362 $\downarrow_1 \approx_2$	0.3639 $\downarrow_1 \approx_2$	0.3635 $\downarrow_1 \approx_2$
WFG6	0.3644	0.3543	0.3629 $\approx_1 \approx_2$	0.3596 $\approx_1 \approx_2$	0.3587 $\approx_1 \approx_2$	0.3604 $\approx_1 \approx_2$
WFG7	0.3164	0.3129	0.3197 $\approx_1 \approx_2$	0.3254 $\approx_1 \uparrow_2$	0.3191 $\approx_1 \approx_2$	0.3222 $\approx_1 \uparrow_2$
WFG8	0.2189	0.2103	0.2177 $\approx_1 \approx_2$	0.2096 $\approx_1 \approx_2$	0.2141 $\approx_1 \approx_2$	0.2144 $\approx_1 \approx_2$
WFG9	0.3575	0.3578	0.3445 $\approx_1 \approx_2$	0.3542 $\approx_1 \approx_2$	0.3518 $\approx_1 \approx_2$	0.3526 $\approx_1 \approx_2$
Summary	Compared to $NormR_A$		$\uparrow_1 10 \downarrow_1 3$	$\uparrow_1 9 \downarrow_1 5$	$\uparrow_1 9 \downarrow_1 2$	$\uparrow_1 10 \downarrow_1 1$
	Compared to $NormR_{ND}$		$\uparrow_2 7 \downarrow_2 6$	$\uparrow_2 6 \downarrow_2 10$	$\uparrow_2 5 \downarrow_2 5$	$\uparrow_2 7 \downarrow_2 5$

Wilcoxon rank-sum test is used to test significance with 0.05 significance level. The footnote 1 or 2 refers to the comparison is conducted with either the archive-based normalization method or the ND-based normalization method, respectively

all the four patches of PF, since it can estimate the bounds well from the beginning. The proposed method does not have the scope of improving this performance significantly via corner search, but we can see that it achieves similar PF convergence and diversity (Fig. 7d–f). Notably, the ND-based normalization ($NormR_{ND}$) performs much worse here, which will be discussed shortly. The second scenario occurs for problem MAF6 in Fig. 8, wherein it is observed that the initial solutions are far from PF. The performance of the archive-based normalization method suffers in this case (for similar reasons as in Fig. 6), while the proposed method maintains its competitive performance.

- Comparison with ND-based normalization:
 - Though ND-based normalization generally performs well, it faces challenges in the scenario where the

initial non-dominated population is concentrated in localized region rather than spanning a good range relative to the PF. In such cases, $NormR_{ND}$ is not able to easily expand the range to obtain other parts of the PF and the diversity is severely compromised. This is illustrated for DTLZ7 in Fig. 7b, where the final solutions cover only a part of the PF. This is because during the initialization, although the solutions in the archive spans a wide range, the non-dominated solutions were concentrated only around two of the patches. The surrogate corner search in Fig. 7d is able to overcome this situation by supplementing the ND set with other corners spanning a larger range which in turn improved its performance significantly.

- The performance of the proposed method is also very consistent compared to the ND-based normalization across different number of objectives as

Table 3 Median HV performance in test problems with 3 objective across 29 independent runs

Problems	$NormR_A$ (archive)	$NormR_{ND}$ (ND)	$NormR_{NDE}$ [34]	$NormR_{NDC}$ (S1)	$NormR_{NDC}$ (S2)	$NormR_{NDC}$ (S3)
mDTLZ1	0.3009	0.3007	0.3002 $\approx_1 \approx_2$	0.2992 $\downarrow_1 \downarrow_2$	0.3010 $\approx_1 \approx_2$	0.3004 $\approx_1 \approx_2$
mDTLZ2	0.7381	0.7068	0.7375 $\downarrow_1 \uparrow_2$	0.7375 $\downarrow_1 \uparrow_2$	0.7378 $\downarrow_1 \uparrow_2$	0.7378 $\downarrow_1 \uparrow_2$
mDTLZ3	0.7333	0.7325	0.7317 $\downarrow_1 \approx_2$	0.7328 $\downarrow_1 \approx_2$	0.7324 $\downarrow_1 \approx_2$	0.7325 $\downarrow_1 \approx_2$
mDTLZ4	0.2140	0.2222	0.2225 $\approx_1 \approx_2$	0.2384 $\approx_1 \approx_2$	0.2511 $\approx_1 \approx_2$	0.1981 $\approx_1 \approx_2$
iDTLZ2	0.7213	0.7348	0.7289 $\uparrow_1 \downarrow_2$	0.7272 $\uparrow_1 \downarrow_2$	0.7268 $\uparrow_1 \downarrow_2$	0.7269 $\uparrow_1 \downarrow_2$
MAF1	0.2992	0.3074	0.3070 $\uparrow_1 \downarrow_2$	0.3057 $\uparrow_1 \downarrow_2$	0.3066 $\uparrow_1 \downarrow_2$	0.3068 $\uparrow_1 \downarrow_2$
MAF2	0.7284	0.7286	0.7264 $\downarrow_1 \downarrow_2$	0.7259 $\downarrow_1 \downarrow_2$	0.7285 $\approx_1 \approx_2$	0.7285 $\approx_1 \approx_2$
MAF3	0.7881	0.8744	0.8631 $\approx_1 \approx_2$	0.8208 $\approx_1 \downarrow_2$	0.7863 $\approx_1 \approx_2$	0.8585 $\approx_1 \approx_2$
MAF4	0.5525	0.7227	0.6817 $\uparrow_1 \downarrow_2$	0.6695 $\uparrow_1 \downarrow_2$	0.6607 $\uparrow_1 \downarrow_2$	0.6814 $\uparrow_1 \downarrow_2$
MAF5	0.1404	0.1404	0.1489 $\approx_1 \approx_2$	0.1479 $\approx_1 \approx_2$	0.1402 $\approx_1 \approx_2$	0.1412 $\approx_1 \approx_2$
MAF6	0.0039	0.2515	0.2493 $\uparrow_1 \approx_2$	0.2469 $\uparrow_1 \approx_2$	0.2496 $\uparrow_1 \approx_2$	0.2489 $\uparrow_1 \approx_2$
DTLZ1	0.2263	1.0953	1.0963 $\uparrow_1 \approx_2$	1.0937 $\uparrow_1 \downarrow_2$	1.0944 $\uparrow_1 \approx_2$	1.0953 $\uparrow_1 \approx_2$
DTLZ2	0.7163	0.7633	0.7632 $\uparrow_1 \approx_2$	0.7620 $\uparrow_1 \downarrow_2$	0.7631 $\uparrow_1 \approx_2$	0.7632 $\uparrow_1 \approx_2$
DTLZ3	0.0265	0.737	0.7387 $\uparrow_1 \approx_2$	0.7366 $\uparrow_1 \approx_2$	0.7362 $\uparrow_1 \approx_2$	0.7375 $\uparrow_1 \approx_2$
DTLZ4	0.2176	0.2101	0.2094 $\approx_1 \approx_2$	0.2559 $\approx_1 \uparrow_2$	0.2419 $\approx_1 \uparrow_2$	0.2322 $\approx_1 \approx_2$
DTLZ7	0.5757	0.4774	0.5815 $\uparrow_1 \uparrow_2$	0.5813 $\uparrow_1 \uparrow_2$	0.5814 $\uparrow_1 \uparrow_2$	0.5815 $\uparrow_1 \uparrow_2$
WFG1	0.3493	0.3382	0.3433 $\approx_1 \approx_2$	0.3315 $\approx_1 \approx_2$	0.3329 $\approx_1 \approx_2$	0.3441 $\approx_1 \approx_2$
WFG2	1.1949	1.0197	1.0345 $\downarrow_1 \approx_2$	1.0382 $\downarrow_1 \uparrow_2$	1.0338 $\downarrow_1 \uparrow_2$	1.0234 $\downarrow_1 \approx_2$
WFG3	0.5190	0.5193	0.5171 $\approx_1 \approx_2$	0.5164 $\approx_1 \approx_2$	0.5194 $\approx_1 \approx_2$	0.5190 $\approx_1 \approx_2$
WFG4	0.6183	0.6200	0.6155 $\approx_1 \downarrow_2$	0.6151 $\approx_1 \downarrow_2$	0.6174 $\approx_1 \approx_2$	0.6231 $\approx_1 \approx_2$
WFG5	0.6549	0.6509	0.6465 $\downarrow_1 \downarrow_2$	0.6538 $\approx_1 \approx_2$	0.6540 $\approx_1 \approx_2$	0.6529 $\approx_1 \approx_2$
WFG6	0.6686	0.6707	0.6586 $\downarrow_1 \downarrow_2$	0.6642 $\approx_1 \downarrow_2$	0.6686 $\approx_1 \approx_2$	0.6700 $\approx_1 \approx_2$
WFG7	0.5850	0.5746	0.5739 $\approx_1 \approx_2$	0.5786 $\approx_1 \approx_2$	0.5779 $\approx_1 \approx_2$	0.5813 $\approx_1 \approx_2$
WFG8	0.3954	0.3956	0.4020 $\approx_1 \approx_2$	0.4050 $\approx_1 \approx_2$	0.4021 $\approx_1 \approx_2$	0.4103 $\approx_1 \uparrow_2$
WFG9	0.5561	0.5714	0.5567 $\approx_1 \approx_2$	0.5418 $\approx_1 \downarrow_2$	0.5509 $\approx_1 \downarrow_2$	0.5623 $\approx_1 \approx_2$
Summary	Compared to $NormR_A$		$\uparrow_1 8 \downarrow_1 6$	$\uparrow_1 8 \downarrow_1 5$	$\uparrow_1 8 \downarrow_1 3$	$\uparrow_1 8 \downarrow_1 3$
	Compared to $NormR_{ND}$		$\uparrow_2 2 \downarrow_2 7$	$\uparrow_2 4 \downarrow_2 11$	$\uparrow_2 4 \downarrow_2 4$	$\uparrow_2 3 \downarrow_2 3$

Wilcoxon rank-sum test is used to test significance with 0.05 significance level. The footnote 1 or 2 refers to the comparison is conducted with either the archive-based normalization method or the ND-based normalization method, respectively

evidenced from Tables 2 and 3. In Table 2 for 2-objectives, the last column has 7 problems where $NormR_{NDC}$ (S3) outperforms $NormR_{ND}$ significantly. There are 5 problems where the proposed method performs worse. However, it is worth noting that although there are cases where the proposed method performs worse than the ND-based normalization, the gap in performance is marginal. For example, in the DTLZ1 problem in Table 2, the median HV of the ND-based method is 0.7008, while the median HV of the proposed method in the last column is 0.7004. In contrast, for the cases where $NormR_{ND}$ faces challenge, such as DTLZ7 in Table 2, the gap is significantly larger.

- In addition, one can observe from Tables 3 and 4 that $NormR_{NDC}$ (S3) outperforms $NormR_{ND}$ in only a few problems, while being equivalent in others. How-

ever, this does not indicate that the performance of the proposed approach gets worse with increasing number of objectives. In fact, the main takeaway is that for $NormR_A$ and $NormR_{ND}$ method, there are both typical scenarios where they suffer disadvantages. The proposed strategy is expected to mitigate these disadvantages in both scenarios and generate rather stable performance across a range of problems, rather than always outperforming the other two strategies. In problems with a higher number of objectives, the performance is equivalent in the majority of instances. Overall, it can be clearly seen that for the scenarios that are challenging for $NormR_{ND}$, the proposed approach is able to adapt and maintain its competitive performance.

- Among the problems with a higher number of objectives, there are fewer that exhibit biased initial ND

Table 4 Median HV performance in test problems with 5 objective across 29 independent runs

Problems	$NormR_A$ (archive)	$NormR_{ND}$ (ND)	$NormR_{NDE}$ [34]	$NormR_{NDC}$ (S1)	$NormR_{NDC}$ (S2)	$NormR_{NDC}$ (S3)
mDTLZ1	0.0214	0.0213	0.0213 ↓ ₁ ≈ ₂	0.0214 ≈ ₁ ≈ ₂	0.0214 ≈ ₁ ≈ ₂	0.0213 ≈ ₁ ≈ ₂
mDTLZ2	0.2329	0.2323	0.2321 ↓ ₁ ≈ ₂	0.2321 ↓ ₁ ≈ ₂	0.2326 ↓ ₁ ↑ ₂	0.2326 ↓ ₁ ↑ ₂
mDTLZ3	0.2305	0.2303	0.2293 ↓ ₁ ↓ ₂	0.2293 ↓ ₁ ↓ ₂	0.2304 ≈ ₁ ≈ ₂	0.2304 ≈ ₁ ≈ ₂
mDTLZ4	0.0065	0.0063	0.0043 ↓ ₁ ↓ ₂	0.0056 ≈ ₁ ≈ ₂	0.0067 ≈ ₁ ≈ ₂	0.0063 ≈ ₁ ≈ ₂
iDTLZ2	0.2189	0.2288	0.2163 ↓ ₁ ↓ ₂	0.2188 ≈ ₁ ↓ ₂	0.2148 ↓ ₁ ↓ ₂	0.2134 ↓ ₁ ↓ ₂
MAF1	0.0182	0.0216	0.0213 ↑ ₁ ↓ ₂	0.0212 ↑ ₁ ↓ ₂	0.0214 ↑ ₁ ↓ ₂	0.0216 ↑ ₁ ≈ ₂
MAF3	1.5865	1.59	1.5921 ↑ ₁ ≈ ₂	1.5907 ↑ ₁ ≈ ₂	1.5896 ↑ ₁ ≈ ₂	1.5898 ↑ ₁ ≈ ₂
MAF4	0.1473	0.224	0.1613 ↑ ₁ ↓ ₂	0.1632 ↑ ₁ ↓ ₂	0.1575 ↑ ₁ ↓ ₂	0.1495 ↑ ₁ ↓ ₂
MAF5	0.5568	0.5662	0.6225 ↑ ₁ ↑ ₂	0.6637 ↑ ₁ ↑ ₂	0.5664 ≈ ₁ ≈ ₂	0.5844 ≈ ₁ ≈ ₂
DTLZ1	1.3406	1.5294	1.5272 ↑ ₁ ≈ ₂	1.5293 ↑ ₁ ≈ ₂	1.5241 ↑ ₁ ≈ ₂	1.5317 ↑ ₁ ≈ ₂
DTLZ2	1.1930	1.2663	1.2835 ↑ ₁ ↑ ₂	1.2820 ↑ ₁ ↑ ₂	1.2854 ↑ ₁ ↑ ₂	1.2853 ↑ ₁ ↑ ₂
DTLZ3	0.8338	1.1742	1.2322 ↑ ₁ ↑ ₂	1.2540 ↑ ₁ ↑ ₂	1.2552 ↑ ₁ ↑ ₂	1.2567 ↑ ₁ ↑ ₂
DTLZ4	0.6440	0.6128	0.6351 ≈ ₁ ↑ ₂	0.6678 ≈ ₁ ↑ ₂	0.6294 ≈ ₁ ≈ ₂	0.6083 ≈ ₁ ≈ ₂
WFG1	0.2172	0.1979	0.1547 ≈ ₁ ≈ ₂	0.3349 ↑ ₁ ↑ ₂	0.2737 ↑ ₁ ↑ ₂	0.2732 ↑ ₁ ↑ ₂
WFG2	1.5756	1.5729	1.5706 ↓ ₁ ↓ ₂	1.5705 ≈ ₁ ↓ ₂	1.5749 ≈ ₁ ≈ ₂	1.5754 ≈ ₁ ≈ ₂
WFG3	0.3572	0.3625	0.3695 ↑ ₁ ≈ ₂	0.3643 ↑ ₁ ≈ ₂	0.3679 ↑ ₁ ↑ ₂	0.3638 ≈ ₁ ≈ ₂
WFG4	1.0582	1.0612	1.0505 ≈ ₁ ↓ ₂	1.0558 ≈ ₁ ≈ ₂	1.0569 ≈ ₁ ≈ ₂	1.0688 ≈ ₁ ≈ ₂
WFG5	1.0765	1.0721	1.0626 ≈ ₁ ≈ ₂	1.0777 ≈ ₁ ≈ ₂	1.0615 ≈ ₁ ≈ ₂	1.0608 ≈ ₁ ≈ ₂
WFG6	1.1989	1.2186	1.2024 ≈ ₁ ↓ ₂	1.2063 ↑ ₁ ↓ ₂	1.2183 ↑ ₁ ≈ ₂	1.2176 ↑ ₁ ≈ ₂
WFG7	0.9865	0.9853	0.9889 ≈ ₁ ≈ ₂	0.9816 ≈ ₁ ≈ ₂	0.9840 ≈ ₁ ≈ ₂	0.9855 ≈ ₁ ≈ ₂
WFG8	0.8229	0.8273	0.8176 ≈ ₁ ≈ ₂	0.8182 ≈ ₁ ≈ ₂	0.8162 ≈ ₁ ≈ ₂	0.8271 ≈ ₁ ≈ ₂
WFG9	0.8900	0.904	0.8864 ≈ ₁ ≈ ₂	0.8771 ≈ ₁ ≈ ₂	0.8591 ↓ ₁ ↓ ₂	0.8860 ≈ ₁ ≈ ₂
Summary	Compared to $NormR_A$		↑ ₁ 8 ↓ ₁ 6	↑ ₁ 10 ↓ ₁ 2	↑ ₁ 9 ↓ ₁ 3	↑ ₁ 8 ↓ ₁ 2
	Compared to $NormR_{ND}$		↑ ₂ 4 ↓ ₂ 8	↑ ₂ 5 ↓ ₂ 6	↑ ₂ 5 ↓ ₂ 4	↑ ₂ 4 ↓ ₂ 2

Wilcoxon rank sum test is used to test significance with 0.05 significance level. The footnote 1 or 2 refers to the comparison is conducted with either the archive-based normalization method or the ND-based normalization method

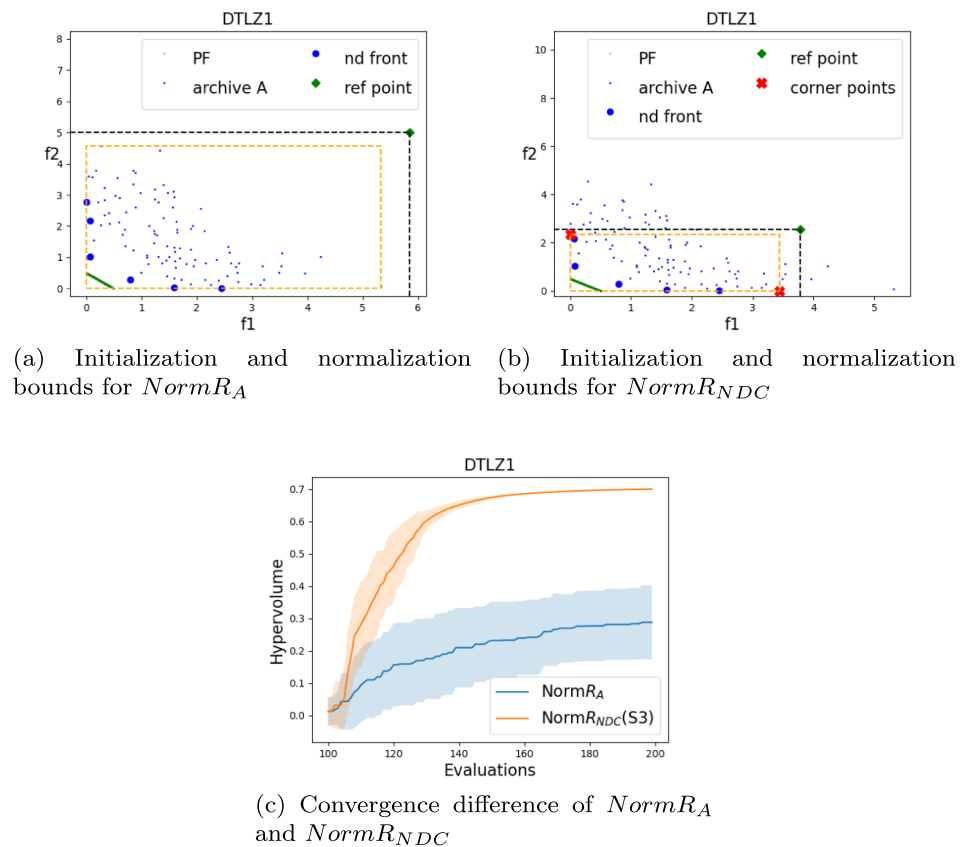
front; unlike the ZDT problems in the 2-objective benchmarks (Table 2). Therefore, the performance of $NormR_{ND}$ and $NormR_{NDC}$ (S3) tend to be similar. For the 5-objective test problems, $NormR_{NDC}$ (S3) performs better than $NormR_{ND}$ for more(4) problems, and worse for fewer (2) problems. For 3-objective cases, it has the same number of better and worse cases (3 each). From the perspective of the experiment setting, 5-objective problems have 200 solutions to train its surrogates while there were 150 solutions for 3-objective problems. Since the number of variables was 6 in both cases, we expect more accurate surrogates to be generated for the 5-objective case. The better surrogate accuracy is likely to lead to offer benefit to surrogate-assisted corner search. However, a crucial thing to note here is that the above numbers are relatively small in comparison to the size of the problem set. For a majority of problems in both 3-objective (19 out of 25) and 5-objective (16 out of 22) sets, the strategies actually perform similar. This

is consistent with the principles of experiment design. The main intent of the proposed strategy is to provide competent (best or close to best) performance when either $NormR_{ND}$ or $NormR_A$ face challenging scenarios such as those discussed above.

Proposed approach compared to previous extreme point search method

The proposed method in this paper is a more generalized version of the extreme point-based normalization ($Norm_{NDE}$) proposed in our previous work [34]. Thus, in this section we compare the performance between the two. In general, the proposed method $NormR_{NDC}$ (S3) is more consistent across different problem settings than $NormR_{NDE}$. This can be observed from the number of cases where the two normalization adjustment methods perform worse than the ND-based normalization method $NormR_{ND}$. The reason to choose comparison with $NormR_{ND}$ is that for a higher number of objectives, there are fewer test problems that pose

Fig. 6 DTLZ1 example of search behavior difference



challenges to $NormR_{ND}$. As a baseline, $NormR_{ND}$ performs better than the archive-based normalization method. Therefore, using comparison with the ND-based normalization method can allow us to better illustrate the difference in performance. In Table 2, the two normalization adjustment methods perform the same, they both have 6 ($NormR_{NDE}$) and 5 problems ($NormR_{NDC}(S3)$) performing worse than the ND-based normalization method and in 7 problems they perform better. In Table 3, last row, it can be seen that for 3-objective problems, $NormR_{NDE}$ performs worse in 7 problems when compared with $NormR_{ND}$, while the proposed method $NormR_{NDC}(S3)$ performs worse in only 3 problems. As for Table 4, these numbers are 8 and 2, respectively. This verifies that the proposed improvements make the method more scalable to higher number of objectives and provides a better overall performance..

Internal comparison between the variants of proposed approach (S1–S3).

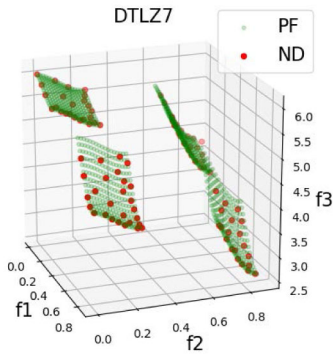
Following the comparisons with the previous approaches, in this section we observe the impact of the selective evaluation and Silhouette analysis in the proposed approach. It can be observed from the Tables 2, 3 and 4 that these mechanisms progressively help to improve the consistency of the pro-

posed surrogate assisted corner search based normalization. In Table 2, compared to the ND-based normalization (last row), and with adding selective evaluation, the cases of worse performance reduces from 10 (S1) to 5 (S2, S3). With increasing number of objectives, the impact of Silhouette analysis also becomes more prominent. In Tables 3 and 4, the numbers of worse cases compared to the ND-based normalization are further reduced when the Silhouette analysis is active. For example, in Table 3, compared to $NormR_{ND}$, the number of worse cases reduces from 11 (S1) to 4 (S2), to 3 (S3). In Table 4, this number reduces from 6 (S1) to 4 (S2), to 2 (S3). Similar trend can also be observed for comparisons with the archive-based normalization method. Overall, these observations suggest the contribution of these modules on the general performance of the proposed method.

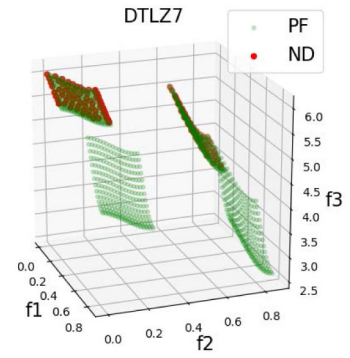
Convergence behavior for some representative problems

To supplement the few cases discussed in the above sections, we provide below a few more representative problems to show the convergence behavior. For the cases of Fig. 9a, b, $NormR_A$ provides good initial bounds (not too far away from the true PF), and is therefore among the better strategies in terms of convergence across all considered. For Fig. 9c, d on the other hand, the case is reversed. The archive of

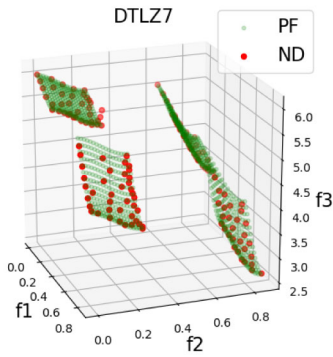
Fig. 7 Typical scenario observed for 3-objective DTLZ7 problem



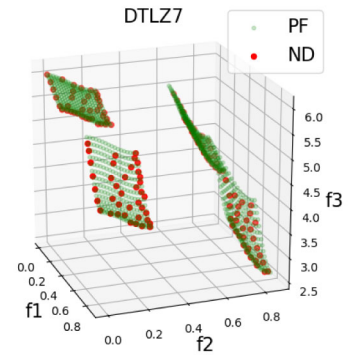
(a) Median result for $NormR_A$



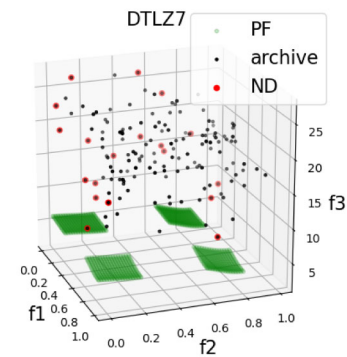
(b) Median result for $NormR_{ND}$



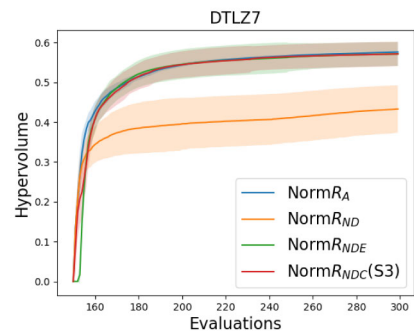
(c) Median result for $NormR_{NDE}$



(d) Median result for $NormR_{NDc} (S3)$

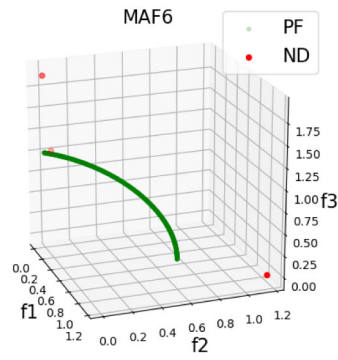


(e) Initialisation

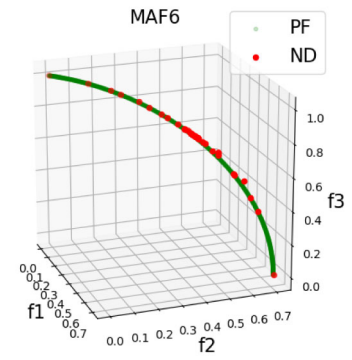


(f) Mean HV convergence

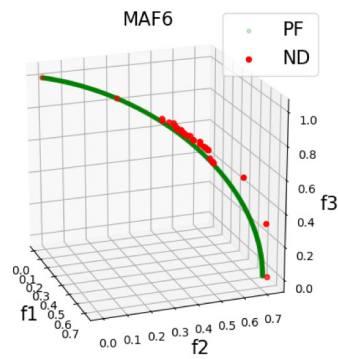
Fig. 8 Typical scenario observed for 3-objective MAF6 problem



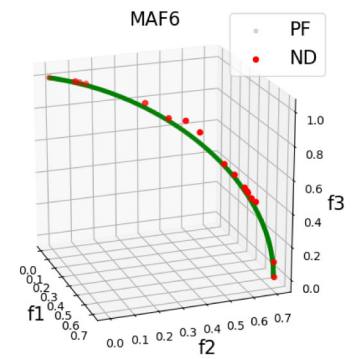
(a) Median final result for $NormR_A$



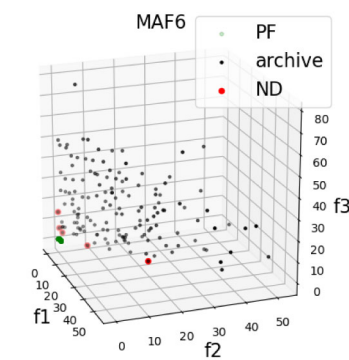
(b) Median final result for $NormR_{ND}$



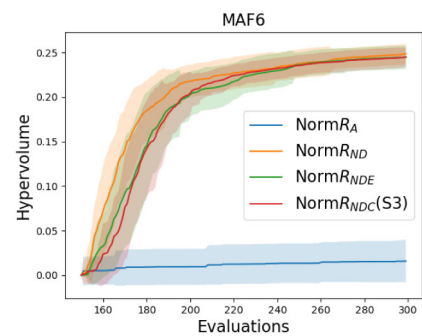
(c) Median final result for $NormR_{NDE}$



(d) Median final result for $NormR_{NDC}(S3)$

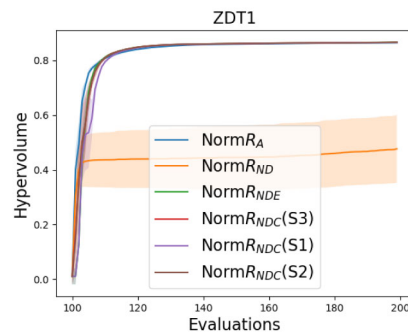


(e) Initialisation

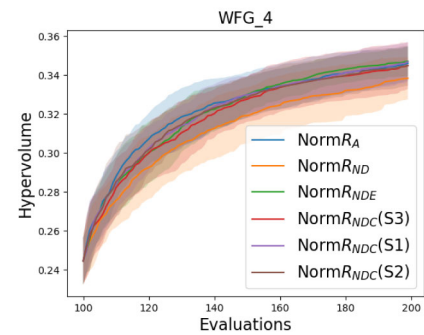


(f) Mean HV convergence

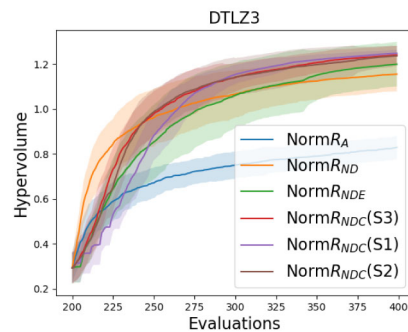
Fig. 9 Mean HV convergence for selected test problems using all strategies



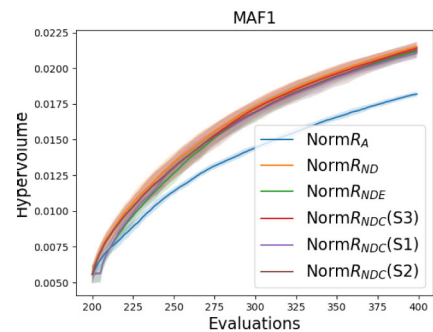
(a) HV mean convergence for 2-objective ZDT1



(b) HV mean convergence for 2-objective WFG4



(c) HV mean convergence for 5-objective DTLZ3



(d) HV mean convergence for 5-objective MAF1

initial solutions span a much bigger range in the objective space compared to the PF, and hence the reference point is set far away, affecting the performance adversely. As for $NormR_{ND}$, it can be seen to struggle in Fig. 9a, where the initial set of ND solutions are biased and consequently could not provide a good reference point.

Concluding remarks

In this paper, we presented an approach for adjusting the normalization bounds for HV-infill based search to solve expensive MOPs. Towards this end, we incorporated a modified version of corner search on the surrogate models, followed by further selection mechanisms to reduce the computational budget directed towards corner evaluations while maintaining the associated benefits. The numerical experiments demonstrate the consistency of the proposed approach. For the problems where the ND-based or archive-based schemes face challenges, it was observed that the proposed scheme can adapt and maintain its competitive performance. Moreover, compared to the extreme point search approach proposed in our previous work [34], the performance of the proposed approach scales better with an increasing number of objectives. In the future, we plan to further improve the per-

formance by incorporating other infill criteria and strategies to deal with problems with extreme convex and non-convex PF shapes.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Auger A, Bader J, Brockhoff D, Zitzler E (2012) Hypervolume-based multiobjective optimization: theoretical foundations and practical implications. *Theor Comput Sci* 425:75–103

2. Bader J, Zitzler E (2011) HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evol Comput* 19(1):4
3. Beume N, Naujoks B, Emmerich M (2007) SMS-EMOA: Multi-objective selection based on dominated hypervolume. *Eur J Oper Res* 181:1653–1669
4. Bhattacharjee KS, Singh HK, Ray T (2018) Multiple surrogate-assisted many-objective optimization for computationally expensive engineering design. *J Mech Design* 140(5) (2018)
5. Blank J, Deb K, Roy PC (2019) Investigating the normalization procedure of NSGA-III. In: *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, New York, pp 229–240 (2019)
6. Cheng R, Li M, Tian Y, Zhang X, Yang S, Jin Y, Yao X (2017) A benchmark test suite for evolutionary many-objective optimization
7. Chugh T, Sindhya K, Hakanen J, Miettinen K (2019) A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms. *Soft Comput* 23(9):3137–3166
8. Deb K (2005) *Multi-objective optimization using evolutionary algorithms*. Wiley, New York
9. Deb K, Agrawal RB et al (1995) Simulated binary crossover for continuous search space. *Complex Syst* 9(2):115–148
10. Deb K, Agrawal S (1999) A niched-penalty approach for constraint handling in genetic algorithms. In: *Artificial Neural Nets and Genetic Algorithms*. Springer, New York, pp 235–243
11. Deb K, Hussein R, Roy PC, Toscano-Pulido G (2018) A taxonomy for metamodeling frameworks for evolutionary multiobjective optimization. *IEEE Trans Evol Comput* 23(1):104–116
12. Ginsbourger D, Le Riche R, Carraro L (2010) Kriging is well-suited to parallelize optimization. In: *Computational intelligence in expensive optimization problems*. Springer, New York, pp 131–162
13. Habib A, Singh HK, Chugh T, Ray T, Miettinen K (2019) A multiple surrogate assisted decomposition-based evolutionary algorithm for expensive multi/many-objective optimization. *IEEE Trans Evol Comput* 23(6):1000–1014
14. He L, Ishibuchi H, Trivedi A, Srinivasan D (2020) Dynamic normalization in MOEA/D for multiobjective optimization. In: *IEEE Congress on Evolutionary Computation (CEC)*
15. He L, Ishibuchi H, Trivedi A, Wang H, Nan Y, Srinivasan D (2021) A survey of normalization methods in multiobjective evolutionary algorithms. *IEEE Trans Evol Comput*
16. Huband S, Hingston P, Barone L, While L (2006) A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Trans Evol Comput* 10(5):477–506
17. Ishibuchi H, Imada R, Setoguchi Y, Nojima Y (2018) How to specify a reference point in hypervolume calculation for fair performance comparison. *Evol Comput* 26(3):411–440
18. Ishibuchi H, Setoguchi Y, Masuda H, Nojima Y (2017) Performance of decomposition-based many-objective algorithms strongly depends on pareto front shapes. *IEEE Trans Evol Comput* 21(2):169–190. <https://doi.org/10.1109/TEVC.2016.2587749>
19. Jain H, Deb K (2014) An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part ii: Handling constraints and extending to an adaptive approach. *IEEE Trans Evol Comput* 18(4):602–622. <https://doi.org/10.1109/TEVC.2013.2281534>
20. Jones RR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Global Optim* 13:4
21. Knowles J (2006) ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Trans Evol Comput* 10(1):50–66
22. Li B, Li J, Tang K, Yao X (2015) Many-objective evolutionary algorithms: a survey. *ACM Comput Surv (CSUR)* 48(1):1–35
23. Liu Y, Ishibuchi H, Yen GG, Nojima Y, Masuyama N, Han Y (2020) On the normalization in evolutionary multi-modal multi-objective optimization. In: *IEEE Congress on Evolutionary Computation (CEC)*
24. Lloyd S (1982) Least squares quantization in PCM. *IEEE Trans Inf Theory* 28(2):129–137
25. Lophaven SN, Nielsen HB, Sondergaard J (2002) *A Matlab kriging toolbox*. Technical University of Denmark, Tech. rep
26. Miettinen K (2012) *Nonlinear multiobjective optimization*, vol. 12. Springer, New York
27. Rehbach F, Zaefferer M, Naujoks B, Bartz-Beielstein T (2020) Expected improvement versus predicted value in surrogate-based optimization. In: *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pp 868–876
28. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
29. Shang K, Ishibuchi H, He L, Pang LM (2020) A survey on the hypervolume indicator in evolutionary multiobjective optimization. *IEEE Trans Evol Comput* 25(1):1–20
30. Singh HK (2020) Understanding hypervolume behavior theoretically for benchmarking in evolutionary multi/many-objective optimization. *IEEE Trans Evol Comput* 24(3):603–610
31. Singh HK, Isaacs A, Ray T (2011) A pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems. *IEEE Trans Evol Comput* 15(4):539–556
32. Singh HK, Yao X (2017) Improvement of reference points for decomposition based multi-objective evolutionary algorithms. In: *Asia-Pacific conference on Simulated Evolution And Learning (SEAL)*, pp 284–296
33. Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optim* 11(4):341–359
34. Wang B, Singh HK, Ray T (2021) Investigating normalization bounds for hypervolume-based infill criterion for expensive multi-objective optimization. In: *Evolutionary Multi-Criterion Optimization: 11th International Conference, EMO 2021, Shenzhen, China, March 28–31, Proceedings*, vol 11. Springer, New York, pp 519–530 (2021)
35. Zapotecas-Martínez S, López-Jaimes A, García-Nájera A (2019) Libea: A lebesgue indicator-based evolutionary algorithm for multi-objective optimization. *Swarm Evol Comput* 44:404–419
36. Zhan D, Cheng Y, Liu J (2017) Expected improvement matrix-based infill criteria for expensive multiobjective optimization. *IEEE Trans Evol Comput* 21(6):956–975
37. Zitzler E, Deb K, Thiele L (2000) Comparison of multiobjective evolutionary algorithms: empirical results. *IEEE Trans Evol Comput* 8(2):173–195
38. Zitzler E, Thiele L (1998) Multiobjective optimization using evolutionary algorithms—a comparative case study. In: *International conference on parallel problem solving from Nature*, pp 292–301

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.