ON SETTING THE REFERENCE POINT FOR THE HYPERVOLUME INDICATOR IN MULTI-OBJECTIVE BAYESIAN OPTIMISATION

Anonymous authors
Paper under double-blind review

ABSTRACT

The Hypervolume (HV) indicator is widely used in Bayesian optimisation for expensive multi-objective optimisation problems, such as in Expected Hypervolume Improvement (EHVI). A key issue in HV-based methods is the specification of the reference point, which critically influences optimisation performance. In situations where an estimate of the problem's true Pareto front is unavailable (typically the case in real-world applications), a common strategy is to dynamically update the reference point according to the boundaries of the observed Pareto front. However, different regions of the Pareto front often exhibit different levels of reachability during the optimisation process. For example, certain boundary regions may be particularly difficult to discover. As such, selecting the reference point dynamically based on the currently observed Pareto front may bias the search and hinder the identification of the entire Pareto front, as solutions beyond the reference point contribute (approximately) zero to the EHVI/HV value. In this paper, we propose a simple method to address this issue. Specifically, we adaptively adjust the HV reference point to shift the search focus between the boundary and interior regions of the Pareto front, aiming to identify the boundaries first and then the interior regions. We validate the proposed method on a diverse set of 69 benchmark and real-world problems. Extensive experimental results demonstrate that it consistently outperforms various reference-point specification strategies.

1 INTRODUCTION

Bayesian Optimisation (BO) is a sample-efficient optimisation framework that iteratively fits a probabilistic surrogate model to the objective function and selects new evaluation points via an acquisition function that balances exploration and exploitation. However, many real-world problems involve multiple, often conflicting objectives rather than a single objective. This setting leads to Multi-Objective Bayesian Optimisation (MOBO). Unlike single-objective optimisation, where there is typically a single optimal solution, multi-objective optimisation yields a set of trade-off solutions known as the Pareto front. The goal of an MOBO method is therefore to efficiently identify a set of solutions that provide a good approximation to the Pareto front under a limited evaluation budget.

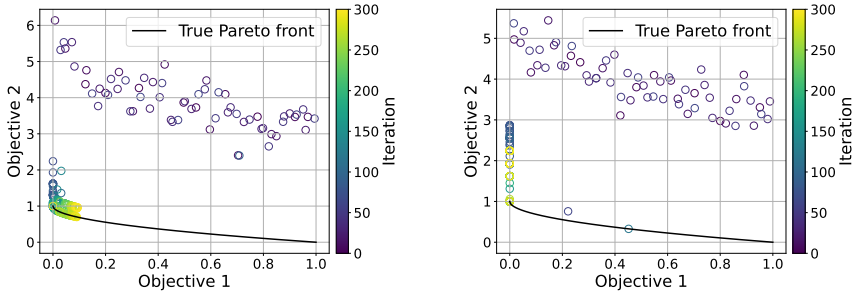
To achieve this, several strategies have been developed in MOBO (Emmerich et al., 2019; Wang et al., 2023b; Karl et al., 2023). One of them is to map the quality of a solution set to a scalar value using a performance indicator, thereby enabling the direct application of single-objective BO methods. In the field of multi-objective optimisation, a range of such indicators have been proposed to evaluate solution set quality in terms of both proximity to the Pareto front (i.e., convergence) and the diversity of solutions (see, e.g., Li & Yao, 2019). Among them, the Hypervolume (HV) indicator (Zitzler & Thiele, 1998) is arguably the most widely used due to its appealing characteristics; for example, it is the only known strictly Pareto-compliant indicator for comparing solution sets, the true Pareto front attains the maximum HV value, and it does not require a representation or approximation of the true Pareto front (Zitzler et al., 2007; Shang et al., 2020; Guerreiro et al., 2021).

Consequently, many BO methods attempt to incorporate the HV indicator into their optimisation process (Ponweiser et al., 2008; Lin et al., 2022; Daulton et al., 2020; 2023; Wang et al., 2024;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Zhang, 2024). Among them, Expected Hypervolume Improvement (EHVI) (Emmerich et al., 2006), which quantifies the expected gain in HV, is one of the most frequently adopted and extensively studied formulations. EHVI has demonstrated strong performance across a range of settings (Rahat et al., 2022; Deng et al., 2025), including sequential optimisation (Emmerich et al., 2006; Rahat et al., 2022), batch optimisation (Yang et al., 2019c; Wada & Hino, 2019; Daulton et al., 2020), preference-aware optimisation (Yang et al., 2016; Feliot et al., 2018), constrained optimisation (Abdolshah et al., 2018; de Winter et al., 2022), and noisy environments (Daulton et al., 2021). In addition, substantial efforts have also been devoted to improving its computational efficiency (Couckuyt et al., 2014; Hupkens et al., 2015; Zhao et al., 2018; Yang et al., 2019a;b; Rahat et al., 2022; Deng et al., 2025; Li & Zhan, 2025).

An important issue in EHVI, and more broadly, any HV-based methods, lies in how to set the reference point in the HV calculation (Ishibuchi et al., 2018; Guerreiro et al., 2021). The HV of a solution set is defined as the volume of the region dominated by the set and bounded by a reference point, and the effectiveness of HV-based methods relies heavily on how this reference point is specified. In many MOBO studies, a predefined, fixed point, often chosen based on prior problem knowledge (e.g., an estimate of the range of the true Pareto front), is used (Daulton et al., 2020; 2021; 2023; de Winter et al., 2022; Wang et al., 2024). However, in real-world applications, such prior knowledge is typically unavailable before the optimisation process begins. Moreover, different EHVI-based methods may favour different settings, and a reference point that performs well for a method may not be suitable for another. In fact, even in the case that the range of the true Pareto front is known, using a fixed reference point may not be ideal (we will show this in our experiment in Section 5).



(a) $1.1\times$ observed Pareto front (b) Maximum observed objective space

Figure 1: Failure of EHVI (Ament et al., 2023) in finding the entire Pareto front of the benchmark problem ZDT1 (Zitzler et al., 2000) over 300 evaluations (colour bar), under two common HV reference point settings. The $2(d+1)$ initial points are sampled using a Sobol sequence (Sobol, 1967).

A natural alternative is to adapt the reference point during the optimisation process. This motivates dynamic strategies in which the reference point is updated based on the information accumulated so far. A seemingly plausible approach is to derive the reference point from the currently observed Pareto front and slightly relax its boundaries (e.g., by expanding each dimension to 1.1 times the normalised observed range (Li et al., 2022)). Such an approach has been widely adopted in HV-based methods for general multi-objective optimisation (e.g., Beume et al., 2007; Bader & Zitzler, 2011) as well as MOBO methods (Daulton et al., 2020; Tu et al., 2022). However, when it comes to optimising EHVI, this approach may cause the search to become confined to a subregion of the Pareto front. One reason is that the reference point is set only slightly worse than the currently observed Pareto front, and any solution that does not dominate the reference point contributes zero or very small HV/EHVI value (with EHVI typically remaining slightly non-zero due to predictive uncertainty). However, the current Pareto front may only reflect a small part of the true Pareto front, resulting in the search hard, if not impossible, to expand towards other regions. Figure 1(a) illustrates such a situation on the widely used bi-objective benchmark problem ZDT1 (Zitzler et al., 2000). When the HV reference point is set to 1.1 times the observed Pareto front, the solutions

obtained by a state-of-the-art EHVI algorithm (i.e., logEHVI (Ament et al., 2023)[1]) concentrate on a small portion (the lower-left region) of the true Pareto front.

A straightforward remedy to the above issue is to place the reference point sufficiently far from the Pareto front, for example, by using the observed objective ranges. This has been implemented in HV-based MOBO methods (de Winter et al., 2022; Deng et al., 2025). However, such a setting is detrimental to solution diversity since a distant reference point tends to place substantially greater emphasis on boundary solutions (Ishibuchi et al., 2018; Li & Yao, 2019). Moreover, this strategy may still fail to approximate the entire Pareto front. This is because in many problems (e.g., ZDT1), certain regions of the true Pareto front are considerably easier to reach than others (as will be discussed in detail in Section 3.1). Figure 1(b) presents such a situation. As shown, when the observed objective bounds are used as the reference point, most solutions obtained by EHVI still concentrate along the leftmost boundary of the objective space, with only one or two reaching the true Pareto front.

In light of the above observations, this paper proposes a novel HV reference point specification strategy for EHVI. The proposed approach divides the optimisation process into two phases through different reference point specifications. In the first phase, the search is focussed on identifying boundary solutions of the Pareto front for each objective. Once these boundaries are settled, the second phase is then to progressively fill in the intermediate regions between them. The switch between the two phases, as well as on different objectives in the first phase, is adaptively determined based on the feedback from the search process. The main contributions of this work are summarised as follows.

- We examine commonly used reference point specifications in EHVI-based MOBO, and illustrate, through the EHVI landscape over the objective space, why they can fail on certain problems.
- We propose a simple reference point specification strategy that targets different regions of the Pareto front at different phases.
- To evaluate the proposed strategy, we conduct extensive experiments on a diverse set of 69 multi-objective benchmark and real-world problems, including comparisons with existing fixed and dynamic strategies, parameter sensitivity analysis.

2 PRELIMINARY

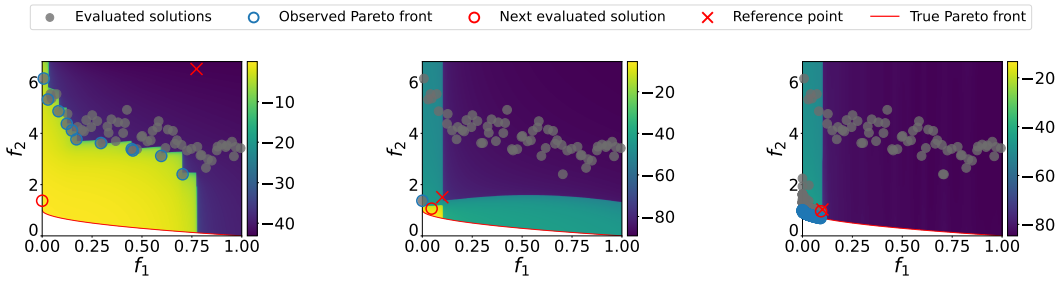
2.1 BASIC CONCEPTS

Bayesian optimisation (BO) is a sequential framework for optimising expensive black-box functions. Consider the problem $\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$, where $\mathcal{X} \subset \mathbb{R}^d$ is the decision space, f is an unknown objective function and each evaluation of $f(\mathbf{x})$ is costly. Suppose after T evaluations we have collected $\mathcal{D}_T = \{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^T$. BO fits a probabilistic surrogate, typically a Gaussian process (GP), whose posterior provides the predictive mean and uncertainty at each candidate point. Based on this model, BO uses an acquisition function $a_T(\mathbf{x})$ to determine where to evaluate next. The acquisition function balances exploration and exploitation by favouring points that are either uncertain or likely to yield better function values. The next query point is selected as $\mathbf{x}_{T+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} a_T(\mathbf{x})$.

Multi-objective optimisation (MOO) aims to optimise multiple objective functions simultaneously. Formally, we consider $\min_{\mathbf{x}\in\mathcal{X}} \ \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, where $\mathbf{f} : \mathcal{X} \to \mathbb{R}^m$ maps a decision vector to m objective values. For $\mathbf{y}, \mathbf{y}' \in \mathbb{R}^m$, $\mathbf{y}$ dominates $\mathbf{y}'$ if $y_i \leq y'_i$ for all i and $y_j < y'_j$ for at least one j. A feasible point $\mathbf{x}^\star$ is Pareto optimal if $\mathbf{f}(\mathbf{x}^\star)$ is not dominated by any other feasible objective vector; the resulting objective vectors form the Pareto front.

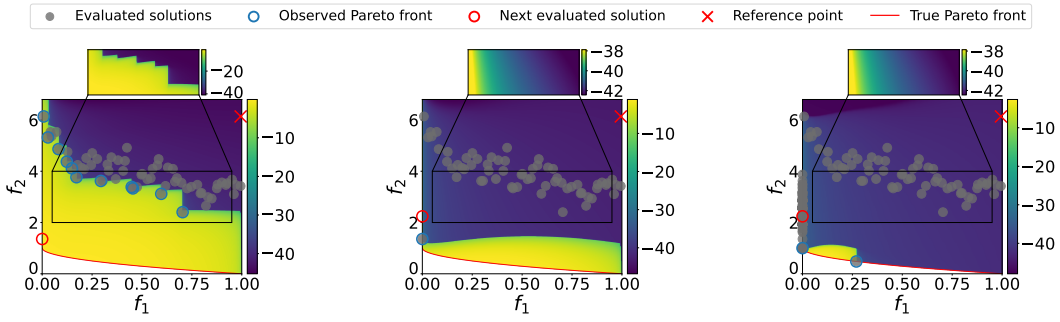
Hypervolume (HV) (Zitzler & Thiele, 1998) is a widely used performance indicator in multi-objective optimisation. Given a set of non-dominated objective vectors $\mathcal{Y} = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}\}$ and a reference point $\mathbf{r} \in \mathbb{R}^m$, the hypervolume is defined as the Lebesgue measure of the region domi-

[1]We use the well-established logEHVI (Ament et al., 2023) as an exemplary algorithm throughout the paper and refer to it as EHVI for simplicity.



(a) 1st iteration (after initial sampling) (b) Iteration 2 (c) Final iteration

Figure 2: EHVI search on ZDT1 using a reference point set to 1.1 times the estimated nadir point of the observed Pareto front. The panels show the first, second, and final iterations under a budget of 300 evaluations. Brighter colours indicate larger EHVI values. Grey dots represent evaluated solutions, blue circles represent the observed Pareto front, the red circle indicates the next evaluation, and the red curve is the true Pareto front.



(a) 1st iteration (after initial sampling) (b) 2nd iteration (c) Last iteration

Figure 3: When the HV reference point is the maximum objective value observed: three snapshots (1st, 2nd and final iterations) of the search process of EHVI on the ZDT1 problem under 300 evaluations, along with the EHVI landscape over the objective space (brighter colours indicating better EHVI values). The inset at the top of each plot shows a zoom-in EHVI landscape of the region enclosed by the rectangle for clearer visualisation of local EHVI values (note the difference of colour bar scale from the main plot).

nated by $\mathcal{Y}$ and bounded by $\mathbf{r}$. Formally, it can be written as $\mathrm{HV}(\mathcal{Y}, \mathbf{r}) = \lambda\left(\bigcup_{\mathbf{y}\in\mathcal{Y}}[\mathbf{y}, \mathbf{r}]\right)$, where $\lambda(\cdot)$ denotes the Lebesgue measure, and $[\mathbf{y}, \mathbf{r}]$ represents the hyper-rectangle defined by $\mathbf{y}$ and $\mathbf{r}$.

Expected hypervolume improvement (EHVI) (Emmerich et al., 2006) is the expected HVI under the surrogate posterior. Let $\mathcal{P}_T$ denote the current Pareto front obtained from the observed data $\mathcal{D}_T$, and let $\mathbf{r} \in \mathbb{R}^m$ be a reference point. For a candidate point $\mathbf{x}$, its hypervolume improvement (HVI) is defined as $\mathrm{HVI}(\mathbf{x} \mid \mathcal{P}_T, \mathbf{r}) = \mathrm{HV}(\mathcal{P}_T \cup \{\mathbf{f}(\mathbf{x})\}, \mathbf{r}) - \mathrm{HV}(\mathcal{P}_T, \mathbf{r})$.Since the true objective value $\mathbf{f}(\mathbf{x})$ is unknown before evaluation, EHVI computes the expectation of the hypervolume improvement under the posterior distribution induced by the surrogate model: $\alpha_{\mathrm{EHVI}}(\mathbf{x}) = \mathbb{E}\big[\mathrm{HVI}(\mathbf{x} \mid \mathcal{P}_T, \mathbf{r})\big]$. The next evaluation point is selected by maximising $\alpha_{\mathrm{EHVI}}(\mathbf{x})$, which encourages sampling points that are expected to yield the largest increase in dominated hypervolume.

2.2 Reference point setting in HV-based acquisition functions

In HV-based acquisition functions, the reference point plays a fundamental role, as it directly determines the dominated region used to measure the improvement. Different choices of the reference point can significantly influence both the magnitude of HV improvement and the search focus of the algorithm.

Many methods in MOBO use a predefined, fixed reference point throughout the optimisation process. In some works, the reference point is specified by the decision maker to reflect user preferences (Feliot et al., 2018). In many other studies, fixed reference points are directly specified for benchmark problems (Couckuyt et al., 2014; Yang et al., 2019b;a;c; Daulton et al., 2020; 2021; 2022; Wang et al., 2024). However, even for the same benchmark problem, different studies often adopt substantially different reference point settings. For example, for the ZDT1 problem, Daulton et al. (2021) set the reference point to $[1.1, 1.1]$, while Yang et al. (2019b) used $[11, 11]$, and Yang et al. (2016) used $[15, 15]$. Such inconsistencies make experimental results difficult to compare and may lead to different search behaviours of HV-based methods.

Another line of work adopts dynamic strategies that update the reference point during optimisation. One approach bases the reference point on the current Pareto front, typically setting it slightly worse than the estimated nadir point (Ponweiser et al., 2008; Seada & Deb, 2015; Li et al., 2018). Alternatively, other methods define the reference point using the observed objective space, for example by taking the worst value in each objective dimension among the evaluated points (de Winter et al., 2022; Deng et al., 2025).

3 Proposed Method

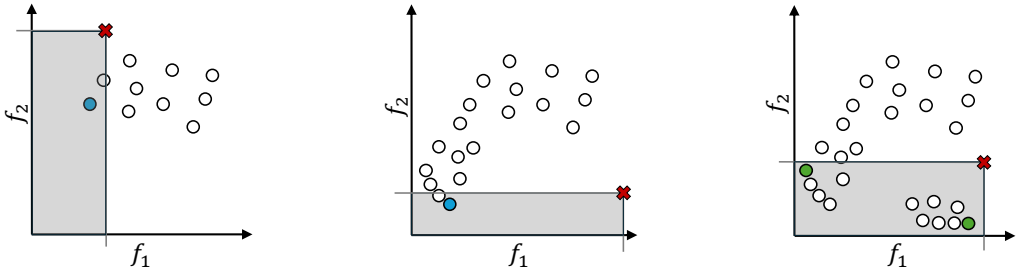
In this section, we first present the motivation why there is a need for a new HV reference point specification in EHVI-based methods, namely, why existing ones do not work. Then, we describe the proposed method.

3.1 Motivation

The most common way of setting the reference point in HV-based multi-objective optimisation algorithms is to consider a point slightly worse than the nadir point of the Pareto front found during the optimisation process (Beume et al., 2007; Bader & Zitzler, 2011; Daulton et al., 2020; Tu et al., 2022), e.g., 1.1 times the normalised range of the observed Pareto front (Li et al., 2022). However, the observed Pareto front typically fails to represent the true Pareto front, especially at the beginning of the search, and may cover only a small portion of it. In such a case, the solution that maximises the EHVI can only come from those that dominate the reference point, since only such solutions can potentially contribute the HV value of the solution set under the reference point.

Figure 2 shows three snapshots (1st, 2nd and last iterations) of the search process of EHVI (Ament et al., 2023) on ZDT1 under 300 evaluations, along with the EHVI landscape over the objective space (details of how to construct the EHVI landscape are provided in Section A of the appendix), where the HV reference point is set to 1.1 times the observed Pareto front. As can be seen from the Figure 2(a), before the first iteration of EHVI, the reference point (red cross), according to the observed Pareto front (from initial sampling), is at the upper-right corner of the space. However, after the first iteration in which the algorithm finds a very good solution that dominates all previous solutions, the reference point moves to the lower-left corner (just slightly worse than the found solution), leading to only a very small area having a high EHVI (bright colour, see Figure 2(b)). This continues until the search ends (Figure 2(c)).

To deal with this issue, one may wonder how about placing the reference point sufficiently far from the Pareto front, for example, setting it to the observed worst objective values. This has been implemented in several HV-based MOBO methods (de Winter et al., 2022; Deng et al., 2025). However, again, different regions of the objective space often have different levels of reachability during optimisation. Taking the problem ZDT1 as an example, solutions in the leftmost boundary of the objective space (i.e., with the optimal f_1 value) are much easier to reach than other Pareto front solutions. This can be seen from the EHVI landscape in Figure 3, where three snapshots (1st, 2nd and last iterations) of the search process of EHVI on ZDT1 under 300 evaluations for this reference point setting are shown. Interestingly, as can be seen in Figure 3(b), even though there is a much better area for EHVI in the bottom of the graph, the algorithm (using L-BFGS-B (Zhu et al., 1997) with 10 restarts, following the common practice (Balandat et al., 2020)) cannot identify it. This is because the leftmost area is much easier to reach and the gradient-based search always ends up there – see the zoom-in EHVI landscape in the inset at the top of each figure (note the difference of colour bar scale from the main plot), where left areas are always better than right ones. This

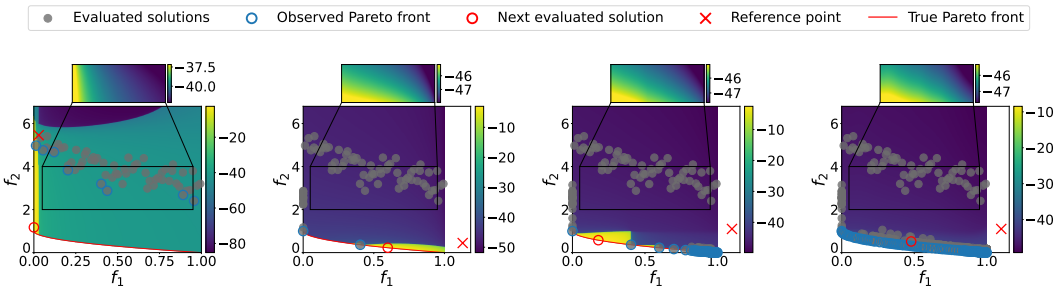


(a) Phase 1: Start of the focus search on f_1

(b) Phase 1: Switch to the focus search on f_2

(c) Phase 2: Start of the balanced search

Figure 4: An illustration of the two phases of the proposed reference point specification on a bi-objective minimisation scenario, where circle points represent evaluated solutions and the red cross represents the reference point. In the first phase (plots (a) and (b)), the reference point is set based on the best value (denoted by blue point) on the respective objective, whereas in the second phase (plot (c)), the reference point is set based on the two boundaries found.



(a) Focus search on f_1

(b) Focus search on f_2

(c) Balanced search

(d) Final iteration

Figure 5: Our method: four snapshots of the EHVI search process on the ZDT1 problem under a budget of 300 evaluations using the proposed HV reference point specification strategy. The EHVI landscape is shown over the objective space, with brighter colours indicating higher EHVI values. The inset at the top of each plot enlarges the region enclosed by the rectangle to provide a clearer visualisation of local EHVI values. Note that its colour-bar scale differs from that of the main plot.

phenomenon persists until the final iteration, as shown in Figure 3(c), where most solutions found are concentrated in the boundary region (i.e., $f_1 = 0$), though occasionally a solution may escape this landscape-induced bias.

3.2 Our Reference Point Specification

As seen above, different regions of the objective space exhibit varying levels of reachability. To address this, we propose a two-phase HV reference point specification strategy that adaptively shifts the search focus across different parts of the Pareto front. In the first phase, we specify the reference point to focus on the search on the boundary regions of the Pareto front. Specifically, for one objective that we want to identify the boundary, we set the reference point value based on its best value observed, while for the remaining objective(s), we set the value based on their maximum values observed. In this case, the search will focus on "pushing down" solutions on a single objective. We adaptively alternate this for different objectives (the details will be described later on). When the boundaries of all objectives have been sufficiently explored, we switch to the second phase by adopting a balanced reference point based on the current Pareto front (i.e., slightly worse than the nadir point of the observed Pareto front, which is the same as many studies mentioned previously).

Figure 4 illustrates this process. As shown, Phase 1 encourages a focused search and aims to identify the best solution in respective boundary regions. After this, in Phase 2 the search becomes globally oriented, seeking to fill sparsely populated regions of the Pareto front (by maximising the EHVI value). It is worth noting that Phase 1 provides an additional benefit beyond identifying the extreme

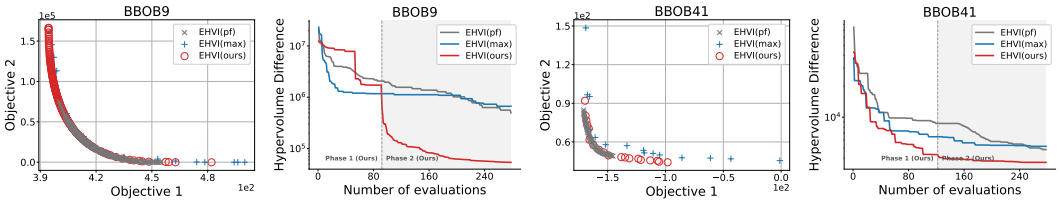


Figure 6: Pareto fronts and hypervolume-difference trajectories of three reference-point strategies on BBOB9 and BBOB41 under 300 evaluations in a typical run. The hypervolume difference is the gap to the best HV obtained (i.e., by all the non-dominated solutions), the lower the better. Dashed lines mark the start of Phase 2 of the proposed method.

objective values for setting the HV reference point for Phase 2. It also produces a relatively dense set of solutions in the boundary regions of the Pareto front, as illustrated in Figure 4(c).

To understand the EHVI landscape under the proposed strategy, Figure 5 shows the four snapshots (start of Phase 1 for each objective, start of Phase 2, and the final iteration) of the search process of EHVI on ZDT1 under 300 evaluations. As can be seen in the figure, the setting of the reference point can predominately affect the EHVI landscape. When the reference point is set for the focus search on objective f_1 (Figure 5(a)), only a slender area in the leftmost has high EHVI values. This is the same for the focus search on objective f_2. Note that this setting affects the EHVI landscape on other areas in the objective space. Looking at the inset at Figure 5(b), the EHVI value increases from right-top to left-bottom, which is unlike Figure 3(b). As such, the solutions located on the lower part of the true Pareto front can be identified, as can be seen in Figure 5(c).

Now we formally give how the reference point is set. In the first phase, we first randomly select one of the m objectives, denoted by f_k, as the objective to focus on. The corresponding reference point $\boldsymbol{r}^{\text{bias}_k} = (r_1^{\text{bias}_k}, \ldots, r_m^{\text{bias}_k})$ is defined as:

$$r_i^{\text{bias}_k} = \begin{cases} \min\limits_{\mathbf{x}\in\mathcal{X}_T} f_i(\mathbf{x}) + \dfrac{\max\limits_{\mathbf{x}\in\mathcal{X}_T} f_i(\mathbf{x}) - \min\limits_{\mathbf{x}\in\mathcal{X}_T} f_i(\mathbf{x})}{N}, & i = k, \\ \max\limits_{\mathbf{x}\in\mathcal{X}_T} f_i(\mathbf{x}) + \dfrac{\max\limits_{\mathbf{x}\in\mathcal{X}_T} f_i(\mathbf{x}) - \min\limits_{\mathbf{x}\in\mathcal{X}_T} f_i(\mathbf{x})}{N}, & i \neq k, \end{cases} \tag{1}$$

where $i \in \{1, 2, \cdots, m\}$, $\mathcal{X}_T = \{\mathbf{x}^{(t)}\}_{t=1}^{T}$ denotes the set of evaluated solutions up to iteration T, and N denotes the number of initial sampling points.

In Phase 2, the reference point becomes the same as the settings in many HV-based methods (i.e., 1.1 times of the range of the observed Pareto front (Daulton et al., 2020; Li et al., 2022; Tu et al., 2022; Wang et al., 2023a):

$$r_i^{\text{normal}} = \max_{\mathbf{x}\in\mathcal{P}} f_i(\mathbf{x}) + 0.1(\max_{\mathbf{x}\in\mathcal{P}} f_i(\mathbf{x}) - \min_{\mathbf{x}\in\mathcal{P}} f_i(\mathbf{x})), \tag{2}$$

where $i = 1, \ldots, m$, and $\mathcal{P}$ is the observed Pareto front.

As mentioned earlier, a key issue in the proposed reference point specification strategy is determining when to switch from the first phase to the second phase, as well as when to switch between objectives during the first phase. Here, we consider an adaptive mechanism. That is, in the first phase, if no good solution is found over a number of consecutive iterations (i.e., no HV improvement under the considered reference point), we switch to another objective. Here, the number of consecutive iterations, denoted by s, is set to 10 (we investigate the sensitivity of the proposed method to this parameter in the experiment in Section 5). The switch from the first phase to the second phase is triggered when no HV improvement is observed for $m \times s$ consecutive iterations, where m is the number of objectives in the optimisation problem. The pseudocode of the proposed method is provided in Section B of the appendix.

4 Experimental Design

Compared Methods. Under the EHVI framework (Ament et al., 2023), we compare the proposed method with two dynamic baselines: EHVI(pf), whose reference point is based on the 1.1-scaled

Table 1: Mean HV values and standard deviations (in parentheses) obtained by our method and two peer methods across 69 benchmark and real-world problems, using a budget of 300 evaluations over 30 independent runs. Parenthesized powers of ten in the column headings scale both the mean and standard deviation. Bold indicates the statistically best method. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than our method, respectively.

Method	BBOB1(10^4)	BBOB2(10^9)	BBOB3(10^8)	BBOB4(10^7)	BBOB5(10^5)	BBOB6(10^4)	BBOB7(10^4)	BBOB8(10^3)
EHVI(PF)	1.39(0.19)$^-$	1.22(0.08)$^-$	1.08(0.04)$^-$	4.28(0.24)$^-$	**3.45(0.02)**$^+$	7.90(0.25)$^-$	1.59(0.15)$^-$	1.35(0.09)$^\sim$
EHVI(MAX)	1.50(0.00)$^-$	1.42(0.00)$^-$	1.13(0.01)$^-$	4.51(0.04)$^-$	3.43(0.02)$^-$	8.29(0.06)$^-$	1.58(0.08)$^-$	1.11(0.09)$^-$
EHVI(Ours)	**1.50(0.00)**	**1.42(0.00)**	**1.13(0.03)**	**4.58(0.01)**	3.44(0.02)	**8.32(0.02)**	**1.72(0.11)**	**1.35(0.05)**
Method	BBOB9(10^7)	BBOB10(10^3)	BBOB11(10^{13})	BBOB12(10^{11})	BBOB13(10^{10})	BBOB14(10^9)	BBOB15(10^7)	BBOB16(10^9)
EHVI(PF)	1.85(0.14)$^-$	7.61(0.45)$^-$	5.69(0.04)$^-$	2.23(0.13)$^-$	1.66(0.14)$^-$	1.27(0.48)$^-$	2.73(1.10)$^-$	2.15(0.04)$^-$
EHVI(MAX)	1.90(0.02)$^-$	8.10(0.63)$^\sim$	5.68(0.02)$^-$	1.96(0.17)$^-$	1.13(0.26)$^-$	1.50(0.06)$^-$	**3.60(0.58)**$^\sim$	2.07(0.06)$^-$
EHVI(Ours)	**1.98(0.00)**	**8.22(0.41)**	**5.73(0.01)**	**2.36(0.02)**	**1.73(0.07)**	**1.73(0.04)**	3.55(1.05)	**2.17(0.04)**
Method	BBOB17(10^7)	BBOB18(10^9)	BBOB19(10^8)	BBOB20(10^{10})	BBOB21(10^{10})	BBOB22(10^9)	BBOB23(10^7)	BBOB24(10^7)
EHVI(PF)	2.37(0.17)$^-$	**8.32(0.20)**$^\sim$	5.15(0.61)$^-$	1.57(0.01)$^\sim$	3.71(0.10)$^-$	**1.66(0.03)**$^+$	1.25(0.06)$^-$	**3.16(0.11)**$^\sim$
EHVI(MAX)	2.13(0.20)$^-$	5.66(0.76)$^-$	**6.91(0.20)**$^\sim$	1.55(0.02)$^-$	3.57(0.06)$^-$	1.60(0.03)$^-$	1.23(0.03)$^-$	2.57(0.20)$^-$
EHVI(Ours)	**2.53(0.16)**	8.28(0.20)	6.90(0.38)	**1.57(0.01)**	**3.80(0.07)**	1.64(0.03)	**1.29(0.02)**	3.14(0.10)
Method	BBOB25(10^7)	BBOB26(10^8)	BBOB27(10^7)	BBOB28(10^7)	BBOB29(10^8)	BBOB30(10^7)	BBOB31(10^6)	BBOB32(10^5)
EHVI(PF)	0.99(0.12)$^-$	6.47(0.59)$^-$	2.42(0.45)$^-$	9.15(0.48)$^-$	1.22(0.11)$^-$	2.04(0.05)$^-$	**2.84(0.19)**$^\sim$	5.51(0.39)$^\sim$
EHVI(MAX)	**1.13(0.06)**$^\sim$	5.29(0.68)$^-$	2.77(0.31)$^\sim$	5.70(1.60)$^-$	1.12(0.06)$^-$	**2.17(0.06)**$^+$	2.31(0.31)$^-$	**5.52(0.49)**$^\sim$
EHVI(Ours)	1.11(0.11)	**6.86(0.15)**	**2.77(0.30)**	**9.52(0.35)**	**1.23(0.03)**	2.11(0.08)	2.82(0.18)	5.46(0.58)
Method	BBOB33(10^7)	BBOB34(10^7)	BBOB35(10^6)	BBOB36(10^5)	BBOB37(10^6)	BBOB38(10^4)	BBOB39(10^6)	BBOB40(10^5)
EHVI(PF)	6.51(1.70)$^-$	0.89(0.14)$^-$	**4.45(0.02)**$^+$	1.94(0.07)$^\sim$	1.15(0.04)$^-$	**7.35(0.28)**$^+$	7.55(0.46)$^-$	1.23(0.13)$^-$
EHVI(MAX)	4.55(1.30)$^-$	**1.17(0.09)**$^\sim$	4.39(0.04)$^\sim$	1.87(0.07)$^-$	1.12(0.04)$^-$	7.17(0.30)$^\sim$	5.71(0.82)$^-$	1.37(0.10)$^\sim$
EHVI(Ours)	**8.03(0.75)**	1.14(0.15)	4.40(0.04)	**1.97(0.04)**	**1.17(0.05)**	7.10(0.46)	**8.01(0.31)**	**1.37(0.12)**
Method	BBOB41(10^4)	BBOB42(10^4)	BBOB43(10^3)	BBOB44(10^6)	BBOB45(10^4)	BBOB46(10^5)	BBOB47(10^3)	BBOB48(10^6)
EHVI(PF)	2.82(0.05)$^-$	4.24(0.14)$^-$	**4.62(0.16)**$^+$	1.20(0.09)$^\sim$	1.30(0.12)$^-$	2.09(0.08)$^\sim$	**5.34(0.37)**$^\sim$	2.87(0.55)$^-$
EHVI(MAX)	2.83(0.04)$^-$	4.09(0.12)$^-$	4.29(0.25)$^\sim$	1.05(0.07)$^-$	**1.47(0.07)**$^\sim$	2.05(0.10)$^-$	4.46(0.74)$^-$	2.42(0.52)$^-$
EHVI(Ours)	**2.88(0.04)**	**4.39(0.12)**	4.35(0.27)	**1.21(0.08)**	1.44(0.12)	**2.13(0.07)**	5.29(0.49)	**3.15(0.48)**
Method	BBOB49(10^4)	BBOB50(10^1)	BBOB51(10^5)	BBOB52(10^3)	BBOB53(10^8)	BBOB54(10^5)	BBOB55(10^3)	ZDT1(10^{-1})
EHVI(PF)	**2.11(0.36)**$^\sim$	**5.90(1.20)**$^+$	2.33(0.24)$^-$	1.08(0.10)$^+$	1.39(0.05)$^-$	7.12(0.62)$^-$	3.31(0.90)$^\sim$	5.23(1.50)$^-$
EHVI(MAX)	1.91(0.35)$^\sim$	3.73(1.40)$^-$	2.31(0.20)$^-$	**1.11(0.09)**$^+$	1.31(0.09)$^-$	7.19(0.70)$^-$	**3.61(0.69)**$^\sim$	4.58(2.20)$^-$
EHVI(Ours)	1.95(0.39)	5.15(1.24)	**2.51(0.20)**	0.96(0.17)	**1.46(0.02)**	**7.90(0.40)**	3.37(0.96)	**8.68(0.01)**
Method	ZDT2(10^{-1})	ZDT3(10^{-1})	ZDT4(10^{-4})	ZDT6(10^0)	DTLZ1(10^4)	DTLZ2(10^{-1})	DTLZ3(10^6)	DTLZ4(10^{-1})
EHVI(PF)	0.88(0.00)$^-$	7.57(0.95)$^\sim$	**4.96(27.00)**$^\sim$	3.14(0.56)$^\sim$	**1.08(0.49)**$^\sim$	**6.95(0.19)**$^+$	1.85(0.48)$^-$	**6.94(0.24)**$^+$
EHVI(MAX)	1.30(0.70)$^-$	**8.06(0.82)**$^\sim$	0.00(0.00)$^\sim$	**3.25(0.40)**$^\sim$	0.94(0.43)$^\sim$	1.39(0.73)$^-$	2.07(0.37)$^\sim$	1.43(0.78)$^-$
EHVI(Ours)	**2.97(0.10)**	7.92(0.63)	0.00(0.00)	3.19(0.44)	1.06(0.57)	5.92(1.59)	**2.15(0.38)**	6.29(1.13)
Method	DTLZ5(10^{-1})	DTLZ6(10^{-2})	DTLZ7(10^0)	4-Bar(10^1)	Vessel(10^9)	Summary		
EHVI(PF)	**1.33(0.01)**$^\sim$	**9.02(1.60)**$^+$	0.41(0.01)$^-$	**5.22(0.00)**$^+$	8.64(0.08)$^\sim$	11/18/40		
EHVI(MAX)	0.61(0.10)$^-$	6.75(0.81)$^-$	0.78(0.19)$^-$	5.21(0.06)$^-$	8.51(0.10)$^-$	2/19/48		
EHVI(Ours)	1.32(0.02)	8.12(1.65)	**1.14(0.22)**	5.22(0.00)	**8.65(0.07)**			

observed Pareto-front range, and EHVI(max), which uses the componentwise maximum observed objective values. We also compare with EHVI(true pf), which uses a fixed reference point based on the 1.1-scaled true Pareto-front range (Li et al., 2022).

Tested Problems. We consider 69 problems from the ZDT (Zitzler et al., 2000), BBOB-biobj (Brockhoff et al., 2022) and DTLZ (Deb et al., 2005) suites, together with two real-world problems: four-bar truss design (Cheng & Li, 1999) and pressure vessel design (Kannan & Kramer, 1994).

Budget and Statistical Validation. Each method is run 30 times with a budget of 300 evaluations. Statistical significance against the proposed method is assessed using the Wilcoxon rank-sum test (Wilcoxon, 1992) at $\alpha = 0.05$.

Performance Indicator. HV is used to compare all methods. For the evaluation, we use the commonly used setting in quality evaluation of multi-objective optimisation (Li & Yao, 2019). That is, we collect all solutions obtained by all methods over 30 independent runs, extract their Pareto front, and set the reference point to $1.1\times$ the Pareto front.

Implementation details are provided in Section C of the appendix.

5 Experimental Results

Comparison with common HV reference point specifications. We begin our evaluation by comparing our proposed method, denoted by EHVI(ours), with two commonly used dynamic reference point methods, namely EHVI(pf) and EHVI(max), as described in Section 4. Table 1 shows the experimental results (mean and standard deviation) of the HV obtained by our method and the two peer methods across 69 benchmark and real-world problems. As can be seen in the table, our method demonstrates a clear performance advantage: it performs statistically better than EHVI(pf) and EHVI(max) on 40 and 48 out of 69 problems, respectively, and is worse only on 11 and 2 out of 69 problems, respectively. Figure 6 shows the obtained Pareto fronts and the hypervolume-difference trajectories of the three methods on two representative problems, BBOB9 and BBOB41. The HV difference is the gap between a method's current HV and the HV of the pooled non-dominated solutions from all runs of the three methods, hence the lower the better. As can be seen from Figure 6, our method outperforms EHVI(pf) and EHVI(max), with more diversified solutions. The HV-difference trajectories illustrate the intended two-phase behaviour of our method. During Phase 1, the hypervolume gain is relatively limited because the method focuses on exploring the boundary regions of the Pareto front. Using the boundary information obtained in Phase 1, the method explores the central region in Phase 2 and rapidly converges towards the Pareto front, as reflected by the faster reduction in the HV difference of the trajectories.

Comparison with a fixed reference point based on the true Pareto front. We next investigate whether the proposed method remains competitive when the range of the true Pareto front is known and can be used to specify the reference point. We denote this oracle method as EHVI(true pf). Detailed results, including the mean and standard deviation of the HV obtained by both methods across the 69 benchmark and real-world problems, are reported in Table D.1 of the appendix. As shown in Table D.1, EHVI(ours) is statistically better than, statistically indistinguishable from, and statistically worse than EHVI(true pf) on 20, 30, and 19 problems, respectively. Thus, despite not having access to the true Pareto front, our adaptive strategy achieves competitive overall performance compared with the oracle setting.

Sensitivity Analysis. The only parameter required in the proposed method is the timing to switch from Phase 1 to Phase 2, as well as between objectives in Phase 1. In our experiments, we set $s = 10$ by default, and we now evaluate several alternative values, including 1, 5, 15 and 20, to investigate the sensitivity of our method to it. Table D.2 in the appendix reports the statistical comparisons over the 69 problems. As can be seen, the performance of the proposed method under $s = 5, 10, 15$ is mostly incomparable – the results with $s = 10$ are statistically equivalent to $s = 5$ and $s = 15$ in 54 and 60 out of the 69 problems. Consider the two extreme settings. When $s = 1$, the switching decision is triggered after only one iteration, which makes the behaviour similar to EHVI(pf). On the other hand, when $s = 20$, the switching may be too slow, which can reduce exploration in the intermediate region. Overall, these results indicate that the performance of the proposed method is robust to the switch parameter within a wide range of settings [5,15].

6 Conclusion

Proper specification of the reference point is crucial for HV-based acquisition functions in multi-objective Bayesian optimisation. We proposed a two-phase HV reference point specification strategy that adaptively shifts the search focus across different regions of the Pareto front. By dynamically adjusting the reference point, the method mitigates the imbalance in reachability of different regions without requiring prior knowledge of the true Pareto front. We evaluated the proposed strategy on 69 benchmark and real-world problems and showed their effectiveness compared with existing fixed and dynamic methods. A follow-up work is to investigate the proposed method in higher-dimensional search spaces. Currently, this study only considered problems with up to 30 decision variables (according to the problem default settings). However, we expect that higher-dimensional problems may benefit more due to the focus search on different regions at different phases.

REFERENCES

Majid Abdolshah, Alistair Shilton, Santu Rana, Sunil Gupta, and Svetha Venkatesh. Expected hypervolume improvement with constraints. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3238–3243. IEEE, 2018.

Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 20577–20612. Curran Associates, Inc., 2023.

Johannes Bader and Eckart Zitzler. Hype: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary computation*, 19(1):45–76, 2011.

Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems 33*, 2020. URL http://arxiv.org/abs/1910.06403.

Nicola Beume, Boris Naujoks, and Michael Emmerich. Sms-emoa: Multiobjective selection based on dominated hypervolume. *European journal of operational research*, 181(3):1653–1669, 2007.

Dimo Brockhoff, Anne Auger, Nikolaus Hansen, and Tea Tušar. Using well-understood single-objective functions in multiobjective black-box optimization test suites. *Evolutionary computation*, 30(2):165–193, 2022.

Franklin Y Cheng and Xiang-Song Li. Generalized center method for multiobjective engineering optimization. *Engineering Optimization*, 31(5):641–661, 1999.

Ivo Couckuyt, Dirk Deschrijver, and Tom Dhaene. Fast calculation of multiobjective probability of improvement and expected improvement criteria for pareto optimization. *Journal of Global Optimization*, 60(3):575–594, 2014.

Sam Daulton, Maximilian Balandat, and Eytan Bakshy. Hypervolume knowledge gradient: a lookahead approach for multi-objective bayesian optimization with partial information. In *International Conference on Machine Learning*, pp. 7167–7204. PMLR, 2023.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. *Advances in neural information processing systems*, 33:9851–9864, 2020.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement. *Advances in neural information processing systems*, 34:2187–2200, 2021.

Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Multi-objective bayesian optimization over high-dimensional search spaces. In *Uncertainty in Artificial Intelligence*, pp. 507–517. PMLR, 2022.

Roy de Winter, Bas van Stein, and Thomas Bäck. Multi-point acquisition function for constraint parallel efficient multi-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 511–519, 2022.

Kalyanmoy Deb, Lothar Thiele, Marco Laumanns, and Eckart Zitzler. Scalable test problems for evolutionary multiobjective optimization. In *Evolutionary multiobjective optimization: theoretical advances and applications*, pp. 105–145. Springer, 2005.

Jingda Deng, Jianyong Sun, Qingfu Zhang, and Hui Li. Expected hypervolume improvement is a particular hypervolume improvement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 16217–16225, 2025.

Michael T. M. Emmerich, Kyriakos C. Giannakoglou, and Boris Naujoks. Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439, 2006. doi: 10.1109/TEVC.2005.859463.

Michael TM Emmerich, Kaifeng Yang, and André H Deutz. Infill criteria for multiobjective Bayesian optimization. In *High-performance simulation-based optimization*, pp. 3–16. Springer, 2019.

Paul Feliot, Julien Bect, and Emmanuel Vazquez. User preferences in bayesian multi-objective optimization: The expected weighted hypervolume improvement criterion. In *Proceedings of the 4th International Conference on Machine Learning, Optimization, and Data Science (LOD 2018)*, pp. 533–544. Springer, 2018. doi: 10.1007/978-3-030-13709-0_45.

Jacob Gardner, Geoff Pleiss, Kilian Q Weinberger, David Bindel, and Andrew G Wilson. Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Andreia P Guerreiro, Carlos M Fonseca, and Luís Paquete. The hypervolume indicator: Computational problems and algorithms. *ACM Computing Surveys (CSUR)*, 54(6):1–42, 2021.

Iris Hupkens, André Deutz, Kaifeng Yang, and Michael Emmerich. Faster exact algorithms for computing expected hypervolume improvement. In *international conference on evolutionary multi-criterion optimization*, pp. 65–79. Springer, 2015.

Hisao Ishibuchi, Ryo Imada, Yu Setoguchi, and Yusuke Nojima. How to specify a reference point in hypervolume calculation for fair performance comparison. *Evolutionary computation*, 26(3): 411–440, 2018.

BK Kannan and Steven N Kramer. An augmented lagrange multiplier based method for mixed integer discrete continuous optimization and its applications to mechanical design. *Journal of mechanical design*, 116(2):405–411, 1994.

Florian Karl, Tobias Pielok, Julia Moosbauer, Florian Pfisterer, Stefan Coors, Martin Binder, Lennart Schneider, Janek Thomas, Jakob Richter, Michel Lang, et al. Multi-objective hyperparameter optimization in machine learning—an overview. *ACM Transactions on Evolutionary Learning and Optimization*, 3(4):1–50, 2023.

Mei Li and Dawei Zhan. Pointwise expected hypervolume improvement for expensive multi-objective optimization. *Journal of Global Optimization*, 91(1):171–197, 2025.

Miqing Li and Xin Yao. Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys (CSUR)*, 52(2):1–38, 2019.

Miqing Li, Tao Chen, and Xin Yao. How to evaluate solutions in pareto-based search-based software engineering: A critical review and methodological guidance. *IEEE Transactions on Software Engineering*, 48(5):1771–1799, 2022.

Zheng Li, Xinyu Wang, Shilun Ruan, Zhaojun Li, Changyu Shen, and Yan Zeng. A modified hypervolume based expected improvement for multi-objective efficient global optimization method: Z. li et al. *Structural and Multidisciplinary Optimization*, 58(5):1961–1979, 2018.

Xi Lin, Zhiyuan Yang, Xiaoyuan Zhang, and Qingfu Zhang. Pareto set learning for expensive multi-objective optimization. *Advances in neural information processing systems*, 35:19231–19247, 2022.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Wolfgang Ponweiser, Tobias Wagner, Dirk Biermann, and Markus Vincze. Multiobjective optimization on a limited budget of evaluations using model-assisted s-metric selection. In *International conference on parallel problem solving from nature*, pp. 784–794. Springer, 2008.

Alma Rahat, Tinkle Chugh, Jonathan Fieldsend, Richard Allmendinger, and Kaisa Miettinen. Efficient approximation of expected hypervolume improvement using gauss-hermite quadrature. In *International Conference on Parallel Problem Solving from Nature*, pp. 90–103. Springer, 2022.

Haitham Seada and Kalyanmoy Deb. A unified evolutionary optimization procedure for single, multiple, and many objectives. *IEEE Transactions on Evolutionary Computation*, 20(3):358–369, 2015.

Ke Shang, Hisao Ishibuchi, Linjun He, and Lie Meng Pang. A survey on the hypervolume indicator in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 25(1):1–20, 2020.

Ilya M Sobol. The distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.

Ben Tu, Axel Gandy, Nikolas Kantas, and Behrang Shafei. Joint entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9922–9938. Curran Associates, Inc., 2022.

Takashi Wada and Hideitsu Hino. Bayesian optimization for multi-objective optimization and multi-point search. *arXiv preprint arXiv:1905.02370*, 2019.

Bing Wang, Hemant Kumar Singh, and Tapabrata Ray. Adjusting normalization bounds to improve hypervolume based search for expensive multi-objective optimization. *Complex & Intelligent Systems*, 9(2):1193–1209, 2023a.

Hao Wang, Kaifeng Yang, and Michael Affenzeller. Probability distribution of hypervolume improvement in bi-objective bayesian optimization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 52002–52018. PMLR, 2024.

Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. Recent advances in Bayesian optimization. *ACM computing surveys*, 55(13s):1–36, 2023b.

Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in Statistics: Methodology and Distribution*, pp. 196–202. Springer, 1992.

Kaifeng Yang, Andre Deutz, Zhiwei Yang, Thomas Back, and Michael Emmerich. Truncated expected hypervolume improvement: Exact computation and application. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 4350–4357. IEEE, 2016.

Kaifeng Yang, Michael Emmerich, André Deutz, and Thomas Bäck. Efficient computation of expected hypervolume improvement using box decomposition algorithms. *Journal of Global Optimization*, 75(1):3–34, 2019a.

Kaifeng Yang, Michael Emmerich, André Deutz, and Thomas Bäck. Multi-objective bayesian global optimization using expected hypervolume improvement gradient. *Swarm and evolutionary computation*, 44:945–956, 2019b.

Kaifeng Yang, Pramudita Satria Palar, Michael Emmerich, Koji Shimoyama, and Thomas Bäck. A multi-point mechanism of expected hypervolume improvement for parallel multi-objective bayesian global optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 656–663, 2019c.

Qiuyi Richard Zhang. Optimal scalarizations for sublinear hypervolume regret. *Advances in Neural Information Processing Systems*, 37:39963–39999, 2024.

Guang Zhao, Raymundo Arroyave, and Xiaoning Qian. Fast exact computation of expected hypervolume improvement. *arXiv preprint arXiv:1812.07692*, 2018.

Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.

Eckart Zitzler and Lothar Thiele. Multiobjective optimization using evolutionary algorithms—a comparative case study. In *International conference on parallel problem solving from nature*, pp. 292–301. Springer, 1998.

Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary computation*, 8(2):173–195, 2000.

Eckart Zitzler, Dimo Brockhoff, and Lothar Thiele. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In *International conference on evolutionary multi-criterion optimization*, pp. 862–876. Springer, 2007.

A Details of the EHVI Landscape Construction

Landscape construction in the objective space. To construct the EHVI landscape for the ZDT1 problem, we discretise the objective space by sampling $f_1 \in [0, 1]$ and $f_2 \in [0, 7]$ with a step size of 0.01, resulting in $70,000$ points over the objective space. For each objective pair (f_1, f_2), we recover the corresponding decision variables and generate 100 feasible solutions whose mean values satisfy the ZDT1 formulation. The EHVI value associated with (f_1, f_2) is then computed as the average acquisition value over these 100 sampled solutions.

ZDT1 formulation and sampling procedure. For the ZDT1 problem (30D), the objectives are defined as

$$f_1(x) = x_1,$$
$$g(x) = 1 + 9 \cdot \frac{1}{29} \sum_{i=2}^{30} x_i,$$
$$f_2(x) = g(x) \left(1 - \sqrt{\frac{x_1}{g(x)}}\right).$$

Given (f_1, f_2), we set $x_1 = f_1$ and recover the corresponding value of g analytically from the above equation. This determines the mean value

$$a = \frac{g - 1}{9}$$

of $(x_2, \ldots, x_{30})$. Since only the mean value of these variables is constrained, the mapping from (f_1, f_2) to the remaining decision variables is not unique. To account for this non-uniqueness, we uniformly sample 100 feasible decision vectors satisfying

$$\frac{1}{29} \sum_{i=2}^{30} x_i = a \quad \text{and} \quad x_i \in [0, 1].$$

Noting that we observe that the 100 sampled solutions result in very similar EHVI values. This can be attributed to the GP lengthscales, which show similar values for $x_2, \ldots, x_{30}$ but a markedly different scale for x_1.

B PSEUDOCODE OF THE PROPOSED METHOD

Algorithm B.1 Reference Point Switching Strategy

Input: switch parameter s, evaluation budget T
Initialize dataset $\mathcal{D}_N$ with N initial points
Sample objective index $k \sim \mathrm{Uniform}(\{1, \dots, m\})$
Set counters $c_k \leftarrow 0$, $C \leftarrow 0$
for $t = N, \dots, T-1$ **do**
  if $C < ms$ **then**
    $\boldsymbol{r} \leftarrow \boldsymbol{r}^{\mathrm{bias}_k}(\mathcal{D}_t)$
  else
    $\boldsymbol{r} \leftarrow \boldsymbol{r}^{\mathrm{normal}}(\mathcal{D}_t)$
  end if
  $\mathbf{x}_{t+1} \leftarrow \arg\max_{\mathbf{x} \in \mathcal{X}} \alpha_{\mathrm{EHVI}}(\mathbf{x}; \boldsymbol{r})$
  Evaluate $\mathbf{y}_{t+1} = \boldsymbol{f}(\mathbf{x}_{t+1})$
  Update $\mathcal{D}_{t+1}$
  if $C < ms$ **then**
    if no improvement is observed **then**
      $c_k \leftarrow c_k + 1, \quad C \leftarrow C + 1$
    else
      $c_k \leftarrow 0, \quad C \leftarrow 0$
    end if
    if $c_k \geq s$ **then**
      $k \leftarrow (k \bmod m) + 1$
      $c_k \leftarrow 0$
    end if
  end if
end for
Output: Non-dominated set $\mathcal{P}_T$

C EXPERIMENT SETTINGS

All experiments were executed as Slurm batch jobs on a CPU-based high-performance computing cluster. Each job consisted of a single task allocated two CPU cores and 16 GB of memory. The compute nodes were equipped with Intel Xeon Platinum 8570 processors and ran Red Hat Enterprise Linux 8.10 on the x86_64 architecture. No GPU was requested or used; all computations were performed on CPUs.

The experiments were implemented in Python 3.11.14 within a Conda environment. The main software dependencies were BoTorch 0.16.1 (Balandat et al., 2020), GPyTorch 1.15.1 (Gardner et al., 2018), PyTorch 2.10.0+cpu (Paszke et al., 2019), NumPy 2.4.2, and SciPy 1.17.0. BoTorch, together with GPyTorch and PyTorch, was used for Gaussian process modelling, acquisition-function evaluation, and automatic differentiation.

In all experiments, we employed Gaussian process (GP) surrogate models with a radial basis function (RBF) kernel with automatic relevance determination (ARD). The initial design points were generated using Sobol low-discrepancy sequences (Sobol, 1967), with an initial sample size of $2(d+1)$. We optimised all acquisition functions using L-BFGS-B (Zhu et al., 1997) with 10 restarts and a maximum of 200 iterations, following common practice (Balandat et al., 2020).

The code is available in an anonymised repository for reproducibility: https://anonymous.4open.science/r/mobo-refpoint-switch-B41F.

D DETAILED EXPERIMENTAL RESULTS

Table D.1: The statistical results (mean and standard deviation) of the HV obtained by the two methods across the 69 benchmark and real-world problems, under a budget of 300 evaluations over 30 independent runs. Methods that are statistically the best are highlighted in **bold**. The symbols "$+$", "$\sim$" and "$-$" indicate that the method is statistically better than, equivalent to and worse than our method, respectively.

Method	BBOB1(10^4)	BBOB2(10^9)	BBOB3(10^8)	BBOB4(10^7)	BBOB5(10^5)	BBOB6(10^4)	BBOB7(10^4)	BBOB8(10^3)
EHVI(true pf)	1.50(0.00)$^-$	**1.42(0.00)**$^\sim$	**1.13(0.01)**$^+$	4.56(0.02)$^-$	**3.44(0.02)**$^\sim$	8.27(0.05)$^-$	1.63(0.08)$^-$	1.13(0.12)$^-$
EHVI(ours)	**1.50(0.00)**	1.42(0.00)	1.13(0.03)	**4.58(0.01)**	3.44(0.02)	**8.32(0.02)**	**1.72(0.11)**	**1.35(0.05)**
Method	BBOB9(10^7)	BBOB10(10^3)	BBOB11(10^{13})	BBOB12(10^{11})	BBOB13(10^{10})	BBOB14(10^9)	BBOB15(10^7)	BBOB16(10^9)
EHVI(true pf)	1.93(0.01)$^-$	**8.44(0.26)**$^+$	**5.73(0.01)**$^\sim$	2.29(0.03)$^-$	1.36(0.15)$^-$	**1.74(0.02)**$^\sim$	**4.64(0.30)**$^+$	**2.18(0.03)**$^\sim$
EHVI(ours)	**1.98(0.00)**	8.22(0.41)	5.73(0.01)	**2.36(0.02)**	**1.73(0.07)**	1.73(0.04)	3.55(1.05)	2.17(0.04)
Method	BBOB17(10^7)	BBOB18(10^9)	BBOB19(10^8)	BBOB20(10^{10})	BBOB21(10^{10})	BBOB22(10^9)	BBOB23(10^7)	BBOB24(10^7)
EHVI(true pf)	2.52(0.11)$^\sim$	7.66(0.29)$^-$	**6.97(0.16)**$^\sim$	1.57(0.00)$^-$	3.72(0.03)$^-$	1.60(0.02)$^-$	1.25(0.02)$^-$	3.11(0.08)$^\sim$
EHVI(ours)	**2.53(0.16)**	**8.28(0.20)**	6.90(0.38)	**1.57(0.01)**	**3.80(0.07)**	**1.64(0.03)**	**1.29(0.02)**	**3.14(0.10)**
Method	BBOB25(10^7)	BBOB26(10^8)	BBOB27(10^7)	BBOB28(10^7)	BBOB29(10^8)	BBOB30(10^7)	BBOB31(10^6)	BBOB32(10^5)
EHVI(true pf)	**1.21(0.05)**$^+$	6.41(0.19)$^-$	2.70(0.36)$^\sim$	9.06(0.38)$^-$	**1.24(0.01)**$^\sim$	**2.25(0.04)**$^+$	**2.88(0.17)**$^\sim$	**5.99(0.45)**$^+$
EHVI(ours)	1.11(0.11)	**6.86(0.15)**	**2.77(0.30)**	**9.52(0.35)**	1.23(0.03)	2.11(0.08)	2.82(0.18)	5.46(0.58)
Method	BBOB33(10^7)	BBOB34(10^7)	BBOB35(10^6)	BBOB36(10^5)	BBOB37(10^6)	BBOB38(10^4)	BBOB39(10^6)	BBOB40(10^5)
EHVI(true pf)	7.35(0.57)$^-$	**1.17(0.15)**$^\sim$	**4.44(0.02)**$^+$	**2.00(0.03)**$^+$	**1.18(0.03)**$^\sim$	**7.25(0.33)**$^\sim$	7.17(0.41)$^-$	**1.41(0.04)**$^+$
EHVI(ours)	**8.03(0.75)**	1.14(0.15)	4.40(0.04)	1.97(0.04)	1.17(0.05)	7.10(0.46)	**8.01(0.31)**	1.37(0.12)
Method	BBOB41(10^4)	BBOB42(10^4)	BBOB43(10^3)	BBOB44(10^6)	BBOB45(10^4)	BBOB46(10^5)	BBOB47(10^3)	BBOB48(10^6)
EHVI(true pf)	2.87(0.04)$^\sim$	**4.45(0.11)**$^\sim$	4.26(0.25)$^\sim$	**1.27(0.03)**$^+$	**1.47(0.06)**$^\sim$	**2.16(0.08)**$^\sim$	**5.38(0.36)**$^\sim$	**3.32(0.25)**$^\sim$
EHVI(ours)	**2.88(0.04)**	4.39(0.12)	**4.35(0.27)**	1.21(0.08)	1.44(0.12)	2.13(0.07)	5.29(0.49)	3.15(0.48)
Method	BBOB49(10^4)	BBOB50(10^1)	BBOB51(10^5)	BBOB52(10^3)	BBOB53(10^8)	BBOB54(10^5)	BBOB55(10^3)	ZDT1(10^{-1})
EHVI(true pf)	1.87(0.33)$^\sim$	**5.20(1.07)**$^\sim$	**2.69(0.11)**$^+$	**1.06(0.10)**$^\sim$	1.42(0.03)$^-$	7.57(0.57)$^-$	**3.47(0.89)**$^\sim$	**8.70(0.01)**$^+$
EHVI(ours)	**1.95(0.39)**	5.15(1.24)	2.51(0.20)	0.96(0.17)	**1.46(0.02)**	**7.90(0.40)**	3.37(0.96)	8.68(0.01)
Method	ZDT2(10^{-1})	ZDT3(10^{-1})	ZDT4(10^{-4})	ZDT6(10^0)	DTLZ1(10^4)	DTLZ2(10^{-1})	DTLZ3(10^6)	DTLZ4(10^{-1})
EHVI(true pf)	**3.01(0.35)**$^+$	7.86(0.86)$^\sim$	0.00(0.00)$^\sim$	**3.52(0.06)**$^+$	0.91(0.49)$^\sim$	**7.51(0.18)**$^+$	**2.31(0.45)**$^\sim$	**7.50(0.24)**$^+$
EHVI(ours)	2.97(0.10)	**7.92(0.63)**	**0.00(0.00)**	3.19(0.44)	**1.06(0.57)**	5.92(1.59)	2.15(0.38)	6.29(1.13)
Method	DTLZ5(10^{-1})	DTLZ6(10^{-2})	DTLZ7(10^0)	4-Bar(10^1)	Vessel(10^9)	Summary		
EHVI(true pf)	**1.35(0.01)**$^+$	5.81(2.60)$^-$	1.08(0.20)$^\sim$	**5.23(0.00)**$^+$	**8.71(0.01)**$^+$	19/30/20		
EHVI(ours)	1.32(0.02)	**8.12(1.65)**	**1.14(0.22)**	5.22(0.00)	8.65(0.07)			

Table D.2: The statistical results (mean and standard deviation) of the HV obtained by the five methods across the 69 benchmark and real-world problems, under a budget of 300 evaluations over 30 independent runs. Methods that are statistically the best are highlighted in **bold**. The symbols "+", "∼" and "−" indicate that the method is statistically better than, equivalent to and worse than our method, respectively.

Method	BBOB1(10^4)	BBOB2(10^9)	BBOB3(10^8)	BBOB4(10^7)	BBOB5(10^5)	BBOB6(10^4)	BBOB7(10^4)	BBOB8(10^3)
EHVI (stag1)	1.49(0.00)$^-$	1.35(0.10)$^-$	1.10(0.05)$^-$	4.42(0.21)$^-$	**3.45(0.01)**$^+$	8.05(0.20)$^-$	1.53(0.20)$^-$	1.37(0.04)$^\sim$
EHVI (stag5)	1.50(0.00)$^\sim$	1.42(0.00)$^\sim$	1.13(0.03)$^\sim$	**4.58(0.01)**$^\sim$	3.45(0.02)$^\sim$	**8.33(0.02)**$^\sim$	1.53(0.27)$^-$	1.35(0.06)$^\sim$
EHVI (stag15)	**1.50(0.00)**$^\sim$	**1.42(0.00)**$^\sim$	1.13(0.02)$^\sim$	4.58(0.01)$^\sim$	3.35(0.22)$^-$	8.31(0.05)$^\sim$	1.68(0.15)$^\sim$	1.36(0.05)$^\sim$
EHVI (stag20)	1.50(0.00)$^\sim$	1.42(0.00)$^\sim$	**1.13(0.02)**$^\sim$	4.57(0.02)$^\sim$	3.32(0.22)$^-$	8.18(0.37)$^-$	1.72(0.12)$^\sim$	**1.37(0.05)**$^\sim$
EHVI (stag10)	1.50(0.00)	1.42(0.00)	1.13(0.03)	4.58(0.01)	3.44(0.02)	8.32(0.02)	**1.72(0.11)**	1.35(0.05)
Method	BBOB9(10^7)	BBOB10(10^3)	BBOB11(10^{13})	BBOB12(10^{11})	BBOB13(10^{10})	BBOB14(10^9)	BBOB15(10^7)	BBOB16(10^9)
EHVI (stag1)	1.95(0.06)$^-$	7.71(0.43)$^-$	5.70(0.05)$^-$	2.28(0.12)$^-$	1.70(0.06)$^\sim$	1.37(0.42)$^\sim$	2.55(1.18)$^-$	2.15(0.04)$^\sim$
EHVI (stag5)	1.98(0.00)$^\sim$	**8.36(0.53)**$^\sim$	**5.73(0.01)**$^\sim$	2.36(0.02)$^\sim$	1.74(0.03)$^\sim$	**1.76(0.04)**$^+$	3.03(1.06)$^-$	2.16(0.04)$^\sim$
EHVI (stag15)	1.98(0.00)$^\sim$	8.03(0.49)$^\sim$	5.73(0.01)$^\sim$	2.36(0.02)$^\sim$	**1.75(0.04)**$^\sim$	1.71(0.05)$^\sim$	**4.05(0.68)**$^\sim$	2.17(0.04)$^\sim$
EHVI (stag20)	1.98(0.00)$^\sim$	7.89(0.84)$^\sim$	5.72(0.02)$^\sim$	2.36(0.02)$^\sim$	1.73(0.06)$^\sim$	1.69(0.07)$^-$	4.03(0.57)$^\sim$	2.17(0.04)$^\sim$
EHVI (stag10)	**1.98(0.00)**	8.22(0.41)	5.73(0.01)	**2.36(0.02)**	1.73(0.07)	1.73(0.04)	3.55(1.05)	**2.17(0.04)**
Method	BBOB17(10^7)	BBOB18(10^9)	BBOB19(10^8)	BBOB20(10^{10})	BBOB21(10^{10})	BBOB22(10^9)	BBOB23(10^7)	BBOB24(10^7)
EHVI (stag1)	2.37(0.19)$^-$	8.31(0.24)$^\sim$	5.23(0.58)$^-$	1.57(0.01)$^\sim$	3.71(0.19)$^-$	1.65(0.08)$^+$	1.26(0.04)$^-$	**3.20(0.08)**$^+$
EHVI (stag5)	2.48(0.17)$^\sim$	8.32(0.33)$^\sim$	6.54(0.70)$^-$	**1.57(0.00)**$^\sim$	**3.82(0.03)**$^\sim$	**1.66(0.03)**$^\sim$	1.26(0.05)$^-$	3.18(0.08)$^\sim$
EHVI (stag15)	**2.53(0.14)**$^\sim$	8.31(0.20)$^\sim$	6.92(0.50)$^\sim$	1.57(0.00)$^\sim$	3.76(0.07)$^-$	1.61(0.05)$^-$	**1.30(0.02)**$^\sim$	3.09(0.16)$^\sim$
EHVI (stag20)	2.50(0.11)$^\sim$	**8.32(0.15)**$^\sim$	**7.04(0.08)**$^+$	1.57(0.00)$^\sim$	3.72(0.09)$^-$	1.60(0.06)$^-$	1.29(0.03)$^\sim$	3.03(0.16)$^-$
EHVI (stag10)	2.53(0.16)	8.28(0.20)	6.90(0.38)	1.57(0.01)	3.80(0.07)	1.64(0.03)	1.29(0.02)	3.14(0.10)
Method	BBOB25(10^7)	BBOB26(10^8)	BBOB27(10^7)	BBOB28(10^7)	BBOB29(10^8)	BBOB30(10^7)	BBOB31(10^6)	BBOB32(10^5)
EHVI (stag1)	0.97(0.11)$^-$	6.72(0.44)$^\sim$	2.40(0.42)$^-$	9.09(0.47)$^-$	1.24(0.04)$^\sim$	2.05(0.05)$^-$	**2.87(0.19)**$^\sim$	5.13(0.52)$^\sim$
EHVI (stag5)	1.03(0.14)$^\sim$	**6.90(0.09)**$^\sim$	2.58(0.46)$^\sim$	9.18(0.49)$^-$	**1.25(0.03)**$^+$	2.06(0.07)$^\sim$	2.85(0.18)$^\sim$	5.34(0.43)$^\sim$
EHVI (stag15)	1.09(0.09)$^\sim$	6.80(0.18)$^\sim$	2.65(0.52)$^\sim$	9.23(0.63)$^\sim$	1.19(0.07)$^\sim$	2.13(0.09)$^\sim$	2.87(0.18)$^\sim$	**5.67(0.61)**$^\sim$
EHVI (stag20)	1.10(0.10)$^\sim$	6.75(0.31)$^\sim$	2.60(0.45)$^\sim$	9.10(1.04)$^\sim$	1.12(0.08)$^-$	**2.18(0.09)**$^+$	2.72(0.21)$^\sim$	5.63(0.47)$^\sim$
EHVI (stag10)	**1.11(0.11)**	6.86(0.15)	**2.77(0.30)**	**9.52(0.35)**	1.23(0.03)	2.11(0.08)	2.82(0.18)	5.46(0.58)
Method	BBOB33(10^7)	BBOB34(10^7)	BBOB35(10^6)	BBOB36(10^5)	BBOB37(10^6)	BBOB38(10^4)	BBOB39(10^6)	BBOB40(10^5)
EHVI (stag1)	6.42(1.42)$^-$	0.89(0.14)$^-$	**4.46(0.02)**$^+$	1.97(0.06)$^\sim$	1.17(0.04)$^\sim$	7.18(0.33)$^\sim$	7.32(0.54)$^-$	1.19(0.12)$^-$
EHVI (stag5)	7.05(1.70)$^\sim$	1.12(0.15)$^\sim$	4.43(0.02)$^+$	**2.01(0.03)**$^+$	**1.18(0.03)**$^\sim$	**7.20(0.32)**$^\sim$	7.67(0.53)$^-$	1.33(0.12)$^\sim$
EHVI (stag15)	7.82(1.03)$^\sim$	**1.14(0.12)**$^\sim$	4.21(0.30)$^-$	1.93(0.06)$^-$	1.15(0.05)$^\sim$	7.17(0.44)$^\sim$	7.91(0.34)$^\sim$	1.26(0.19)$^-$
EHVI (stag20)	7.75(1.03)$^\sim$	1.14(0.14)$^\sim$	3.90(0.37)$^-$	1.85(0.12)$^-$	1.12(0.07)$^-$	7.13(0.52)$^\sim$	7.77(0.45)$^-$	1.26(0.12)$^-$
EHVI (stag10)	**8.03(0.75)**	1.14(0.15)	4.40(0.04)	1.97(0.04)	1.17(0.05)	7.10(0.46)	**8.01(0.31)**	**1.37(0.12)**
Method	BBOB41(10^4)	BBOB42(10^4)	BBOB43(10^3)	BBOB44(10^6)	BBOB45(10^4)	BBOB46(10^5)	BBOB47(10^3)	BBOB48(10^6)
EHVI (stag1)	2.81(0.05)$^-$	4.31(0.16)$^\sim$	**4.58(0.29)**$^+$	1.18(0.11)$^\sim$	1.31(0.13)$^-$	2.09(0.08)$^\sim$	5.14(0.39)$^\sim$	2.82(0.53)$^\sim$
EHVI (stag5)	2.86(0.04)$^\sim$	**4.41(0.15)**$^\sim$	4.46(0.41)$^\sim$	1.16(0.11)$^\sim$	1.41(0.13)$^\sim$	2.13(0.09)$^\sim$	**5.41(0.36)**$^\sim$	3.13(0.35)$^\sim$
EHVI (stag15)	2.86(0.05)$^\sim$	4.33(0.18)$^\sim$	4.29(0.41)$^\sim$	1.22(0.07)$^\sim$	1.46(0.09)$^\sim$	2.08(0.08)$^-$	5.28(0.51)$^\sim$	**3.18(0.36)**$^\sim$
EHVI (stag20)	2.84(0.04)$^-$	4.19(0.23)$^-$	4.40(0.24)$^\sim$	**1.22(0.08)**$^\sim$	**1.46(0.08)**$^\sim$	2.08(0.12)$^\sim$	4.87(0.49)$^-$	3.14(0.40)$^\sim$
EHVI (stag10)	**2.88(0.04)**	4.39(0.12)	4.35(0.27)	1.21(0.08)	1.44(0.12)	**2.13(0.07)**	5.29(0.49)	3.15(0.48)
Method	BBOB49(10^4)	BBOB50(10^1)	BBOB51(10^5)	BBOB52(10^3)	BBOB53(10^8)	BBOB54(10^5)	BBOB55(10^3)	ZDT1(10^{-1})
EHVI (stag1)	**2.05(0.31)**$^\sim$	5.26(1.20)$^\sim$	2.37(0.23)$^\sim$	**1.04(0.12)**$^\sim$	1.41(0.03)$^-$	6.85(0.84)$^-$	3.45(0.69)$^\sim$	**8.70(0.01)**$^+$
EHVI (stag5)	1.86(0.39)$^\sim$	**5.46(1.40)**$^\sim$	2.44(0.24)$^\sim$	0.96(0.17)$^\sim$	1.45(0.03)$^\sim$	7.33(0.83)$^-$	**3.55(0.66)**$^\sim$	8.69(0.01)$^+$
EHVI (stag15)	1.76(0.55)$^\sim$	4.74(1.35)$^\sim$	2.54(0.19)$^\sim$	0.92(0.23)$^\sim$	1.46(0.01)$^\sim$	7.71(0.55)$^\sim$	2.76(1.14)$^\sim$	8.67(0.01)$^-$
EHVI (stag20)	1.65(0.48)$^\sim$	4.37(1.53)$^\sim$	**2.55(0.21)**$^\sim$	0.90(0.25)$^\sim$	1.45(0.02)$^\sim$	7.74(0.43)$^\sim$	3.07(1.16)$^\sim$	8.67(0.01)$^-$
EHVI (stag10)	1.95(0.39)	5.15(1.24)	2.51(0.20)	0.96(0.17)	**1.46(0.02)**	**7.90(0.40)**	3.37(0.96)	8.68(0.01)
Method	ZDT2(10^{-1})	ZDT3(10^{-1})	ZDT4(10^{-4})	ZDT6(10^0)	DTLZ1(10^4)	DTLZ2(10^{-1})	DTLZ3(10^6)	DTLZ4(10^{-1})
EHVI (stag1)	2.77(0.63)$^-$	7.17(0.72)$^-$	0.00(0.00)$^\sim$	3.24(0.42)$^\sim$	**1.25(0.51)**$^\sim$	**7.00(0.16)**$^+$	2.04(0.52)$^\sim$	**6.99(0.10)**$^+$
EHVI (stag5)	**3.05(0.12)**$^\sim$	7.79(0.89)$^\sim$	**70.00(383.00)**$^\sim$	3.13(0.54)$^\sim$	1.12(0.43)$^\sim$	6.92(0.24)$^+$	2.13(0.41)$^\sim$	6.80(0.29)$^+$
EHVI (stag15)	2.74(0.70)$^\sim$	7.88(0.99)$^\sim$	0.00(0.00)$^\sim$	**3.34(0.40)**$^\sim$	1.10(0.52)$^\sim$	5.18(2.37)$^\sim$	2.30(0.51)$^\sim$	4.62(2.46)$^-$
EHVI (stag20)	2.87(0.54)$^\sim$	**7.97(1.02)**$^\sim$	0.00(0.00)$^\sim$	3.21(0.49)$^\sim$	1.10(0.43)$^\sim$	4.11(2.36)$^-$	**2.40(0.42)**$^\sim$	4.81(1.86)$^-$
EHVI (stag10)	2.97(0.10)	7.92(0.63)	0.00(0.00)	3.19(0.44)	1.06(0.57)	5.92(1.59)	2.15(0.38)	6.29(1.13)
Method	DTLZ5(10^{-1})	DTLZ6(10^{-2})	DTLZ7(10^0)	4-Bar(10^1)	Vessel(10^9)	Summary		
EHVI (stag1)	**1.33(0.01)**$^+$	**8.55(1.58)**$^\sim$	0.90(0.21)$^-$	**5.23(0.00)**$^+$	8.61(0.17)$^\sim$	10/29/30		
EHVI (stag5)	1.32(0.01)$^\sim$	8.25(1.59)$^\sim$	1.10(0.23)$^\sim$	5.22(0.00)$^+$	8.66(0.05)$^\sim$	8/54/7		
EHVI (stag15)	1.32(0.02)$^\sim$	8.30(1.64)$^\sim$	1.07(0.25)$^\sim$	5.22(0.00)$^\sim$	8.66(0.05)$^\sim$	0/60/9		
EHVI (stag20)	1.25(0.24)$^-$	8.37(1.50)$^\sim$	**1.17(0.16)**$^\sim$	5.22(0.01)$^-$	**8.67(0.02)**$^\sim$	2/47/20		
EHVI (stag10)	1.32(0.02)	8.12(1.65)	1.14(0.22)	5.22(0.00)	8.65(0.07)			

E AI USE STATEMENT

Generative AI tools were used to assist with and polish the writing of this manuscript. All AI-assisted revisions were carefully reviewed and verified by the authors, who take full responsibility for the final content.