Expected Hypervolume Improvement Is a Particular Hypervolume Improvement

Jingda Deng[1], **Jianyong Sun**[1*], **Qingfu Zhang**[2, 3], **Hui Li**[1]

[1]School of Mathematics and Statistics, Xi'an Jiaotong University
[2]Department of Computer Science, City University of Hong Kong
[3]The City University of Hong Kong Shenzhen Research Institute
{jddeng, jy.sun}@xjtu.edu.cn, qingfu.zhang@cityu.edu.hk, lihui10@mail.xjtu.edu.cn

Abstract

Multi-objective Bayesian optimization (MOBO) aims to optimize multiple competing objective functions in the expensive-to-evaluate scenario. The Expected Hypervolume Improvement (EHVI) is a commonly used acquisition function for MOBO and shows a good performance. However, the computation of EHVI becomes challenging as the number of objective functions grows. In this paper, we revisit the formulation of EHVI, as well as its multi-point counterpart qEHVI, and derive much simpler analytic expressions for them. The main contributions of this paper include: (1) first formulating EHVI as a particular hypervolume improvement, and thus immediately obtaining a formal proof of its NP-hardness, faster algorithms in both theory and practice, and more results on its derivatives; (2) first obtaining the analytic expressions of qEHVI for any $q > 1$ and $m \geq 2$ where m is the number of objectives; and (3) demonstrating the advantages of our formulation over existing exact and approximation methods for computing EHVI and qEHVI through a large number of numerical experiments.

Introduction

In many real-world optimization problems, the objective function evaluation requires costly or time-consuming simulations in physic or on computer. For these expensive problems, Bayesian optimization (BO) is popular and efficient (Garnett 2023). When there are multiple expensive-to-evaluate objective functions to be optimized simultaneously, the problem may become more challenging, for example, finding the optimal parameters in different engineering design problems (Tanabe and Ishibuchi 2020) and seeking the optimal hyper-parameters in the neural network structure design (Binois et al. 2020). Efficient multi-objective Bayesian optimization (MOBO) algorithms are desired to solve these multi-objective expensive optimization problems.

In this paper, we consider the maximization of a continuous vector-valued function $\mathbf{f}(\mathbf{x}) = (\mathbf{f}_1(\mathbf{x}), \dots, \mathbf{f}_m(\mathbf{x})) : \mathbb{R}^d \to \mathbb{R}^m$ over a feasible set $\mathcal{X} \subset \mathbb{R}^d$, where each of the objective function $\mathbf{f}_i(\mathbf{x})$ is expensive to evaluate with an unknown analytic expression and gradient. To generate promising candidate solutions to evaluate, a typical MOBO algorithm builds a probabilistic surrogate model based on evaluated solutions, always a multi-output Gaussian process (GP) using kernels for vector-valued functions (Rasmussen 2003; Álvarez, Rosasco, and Lawrence 2012), to approximate the unknown black-box objective functions, and then selects candidate solutions guided by an acquisition function.

As a widely used acquisition function in BO, expected improvement (EI) has been adapted for MOBO and one well-known adaptation is the expected hypervolume improvement (EHVI) (Emmerich, Giannakoglou, and Naujoks 2006). EHVI combines the concepts of EI and the hypervolume indicator (HV), which is a scalar performance indicator to measure the quality of solution sets in the field of multi-objective optimization (Zitzler and Thiele 1999). EHVI has shown a good performance in a number of applications (Couckuyt, Deschrijver, and Dhaene 2014; Yang et al. 2019b), and there are some effective variants such as qEHVI for batch optimization (Daulton, Balandat, and Bakshy 2020) and LogEHVI for high-precision optimum search (Ament et al. 2023).

However, the application of EHVI in MOBO is limited due to its complicated computation. The currently best algorithm for computing EHVI scales exponentially with respect to the number of objectives (Yang et al. 2019a). As shown in the numerical experiments, it becomes intractable when the number of objectives $m > 4$ and the size of the non-inferior set $n > 50$. As a result, the difficulty in the computation makes optimizing EHVI itself not an easy task.

To address the difficulties in the computation of EHVI and its variants, we investigate more efficient analytic expressions for them in this work. Table 1 compares the complexities of our formulations for qEHVI ($q \geq 1$) and state-of-the-art results.[1] The contributions of this work are summarized as follows:

- We prove that EHVI can be expressed as a particular hypervolume improvement (**Theorem 1**). As the hypervolume improvement concept has been adequately investigated in the field of multi-objective optimization, this new formulation provides many results both in theory and in practice (**Corollary 1-3**);

*Corresponding author.

[1]In the complexities, m and q are assumed to be constant, and the $\tilde{O}$ notation hides polynomials of $\log n$.

Algorithm	KMAC (Yang et al. 2019a)		DBB (Daulton, Balandat, and Bakshy 2020)		This paper	
Dimension	$m = 2, 3$	$m \geq 4$	$m = 2, 3$	$m \geq 4$	$m = 2, 3$	$m \geq 4$
Compute EHVI	$\Theta(n \log n)$ time $O(n)$ space	$O(n^{\lfloor \frac{m}{2} \rfloor})$ time $O(n^{\lfloor \frac{m}{2} \rfloor})$ space	The same method as KMAC		$\Theta(n \log n)$ time $O(n)$ space	$\boldsymbol{\tilde{O}(n^{\frac{m}{3}})}$ **time** $\boldsymbol{O(n)}$ **space**
Compute qEHVI ($q > 1$)	N/A		Monte-Carlo method with the same complexities of DBB for EHVI		**Exact method** with the same complexities of our method for EHVI	

Table 1: Complexity comparison between our formulations and state-of-the-art results

- Using a similar strategy, we formulate qEHVI ($q > 1$) as a sum of a constant number (assuming q to be a constant) of particular hypervolume improvements (**Theorem 2**). This is the first analytic expression of qEHVI for any $q > 1$. We also obtain many theoretical results and a practical algorithm for qEHVI based on this expression (**Corollary 4-6**);
- We conduct a number of experiments to validate the correctness and efficiency of our formulation. The results suggest that our implementation outperforms the existing fastest implementation for EHVI and qEHVI when $q \leq 4$. For $q > 4$, our implementation stands out when the number of objectives are large enough.

Background

Bayesian Optimization

In a GP-based single-objective BO algorithm, given a set of evaluated solutions $\mathcal{D} = \{(\mathbf{x}^{(i)}, f(\mathbf{x}^{(i)}))\}_{i=1}^{K}$, the GP model provides a posterior distribution $f(\mathbf{x}|\mathcal{D}) \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$ for any candidate solution $\mathbf{x} \in \mathcal{X}$, where $\mu(\mathbf{x}), \sigma^2(\mathbf{x})$ are calculated based on some kernel function. The quality of the candidate $\mathbf{x}$ is then estimated by an acquisition function. The upper confidence bound (UCB), possibility of improvement (PoI) and EI are some representative acquisition functions. EI is quite popular due to the milestone algorithm EGO (Jones, Schonlau, and Welch 1998). EI measures the expected improvement of $f(\mathbf{x})$ to the best function value f^* in all evaluated solutions depended on $\mathcal{D}$. The definition of EI is given below, where $\mathbf{x}$ is omitted for brevity.

Definition 1 (EI)**.** *Given a normal random variable* $y \sim \mathcal{N}(\mu, \sigma^2), f^* \in \mathbb{R}$. EI *is defined by:*

$$\begin{aligned} \text{EI}(f^*, \mu, \sigma^2) &= \mathbb{E}_{y \sim \mathcal{N}(\mu, \sigma^2)}\left[(y - f^*)_+\right] \\ &= \sigma\phi\left(\frac{\mu - f^*}{\sigma}\right) + (\mu - f^*)\Phi\left(\frac{\mu - f^*}{\sigma}\right) \end{aligned} \tag{1}$$

where $(\cdot)_+$ *means the* $\max\{\cdot, 0\}$ *operation,* $\phi(z)$ *is the probability density function (p.d.f.) of the standard normal distribution, and* $\Phi(z)$ *is the corresponding cumulative distribution function (c.d.f.).*

The qEI, a.k.a. multi-point EI, is introduced in (Ginsbourger, Riche, and Carraro 2008). It is an extension of EI for batch optimization. It aims to measure the expected improvement to f^* over q candidates:

Definition 2 (qEI)**.** *Given a q-D normal random variable* $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), f^* \in \mathbb{R}$. qEI *is defined by:*

$$q\text{EI}(f^*, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathbb{E}_{\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}\left[\left(\max_{1 \leq i \leq q}\{\mathbf{y}_i\} - f^*\right)_+\right]. \tag{2}$$

The expectation in qEI is not easy to compute as q candidates are correlated. The following lemma is based on an algorithm for qEI in (Chevalier and Ginsbourger 2013):[2]

Lemma 1 (Complexity of qEI)**.** *The computation of* qEI *requires* $O(q)$ *evaluations of the c.d.f. of q-variate normal distributions and* $O(q^2)$ *evaluations of the c.d.f. of* $(q-1)$*-variate normal distributions.*

For multivariate normal distributions, we denote by $\xi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{z})$ the p.d.f. of a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and by $\Phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{z})$ the corresponding c.d.f..

Multi-Objective Bayesian Optimization

Related Works Similar to single-objective BO, MOBO may build m GP models[3] for objective functions and obtain a posterior distribution $\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x}))$ for any candidate $\mathbf{x} \in \mathcal{X}$. However, it is not always possible to compare vectors in high-dimensional objective space. Thus, there is no direct way to measure the quality of a candidate, as in single-objective BO.

To address this problem, many works consider to aggregate multiple objective functions into a single one and define acquisition functions based on the aggregated function, such as the Tchebycheff scalarization method assisted with UCB (Lin et al. 2022), the multiplication of multiple EIs (Gaudrie et al. 2020), the knowledge hypervolume gradient (Daulton, Balandat, and Bakshy 2023), EHVI (Emmerich, Giannakoglou, and Naujoks 2006), ER2I (Deutz, Emmerich, and Yang 2019), ETI (Zhang et al. 2010; Zhao and Zhang 2023) and DirEGO (Zhao and Zhang 2024). A survey on the infill criteria for multi-objective Bayesian optimization are given in (Emmerich, Yang, and Deutz 2020).

[2] If the max sign in Equation (2) is replaced by min, the method in (Chevalier and Ginsbourger 2013) is also applicable and **Lemma 1** still holds. We denote it by qMinEI in this case, which will be used in our formulation of qEHVI.

[3] In this work, we follow the common assumption in MOBO that the objective functions are mutually independent. For correlated problems, there are very a few researches limited to two objectives (Yang et al. 2023).

Among them, EHVI has attracted many interests as it is based on HV, which is a widely used performance indicator to measure the quality of solution sets. Moreover, EHVI-based MOBO algorithms have shown good performances on many applications (Couckuyt, Deschrijver, and Dhaene 2014; Yang et al. 2019b; Daulton, Balandat, and Bakshy 2020; Ament et al. 2023).

HV and EHVI Intuitively, for a maximization problem, the objective values of a good solution set should approach $+\infty$ as close as possible. Thus, it should also cover as more volume below as possible. Following this intuition, HV was proposed as an indicator to measure the quality of a solution set by the volume covered by the set.

Definition 3 (HV). *Given a set of points $A = \{\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(n)}\} \subset \mathbb{R}^m$, a reference point $\mathbf{r} \in \mathbb{R}^m$, the HV of A with respect to $\mathbf{r}$ is defined by:*

$$\mathrm{HV}(A, \mathbf{r}) = \lambda_m(\{\mathbf{z} \in \mathbb{R}^m | \mathbf{z} \geq \mathbf{r} \wedge \exists 1 \leq i \leq n, \mathbf{f}^{(i)} \geq \mathbf{z}\}). \tag{3}$$

where $\lambda_m(\cdot)$ means the measure of an m-D set.

The use of the reference point $\mathbf{r}$ is to bound below the region covered by A. Without loss of generality, we assume in the following that $\mathbf{r} = \mathbf{0}$ and $\mathbf{f}^{(i)} \in \mathbb{R}^m_+ := (0, +\infty)^m, \forall 1 \leq i \leq n$. In this way, $\mathrm{HV}(A, \mathbf{r})$ is simplified as $\mathrm{HV}(A)$. Besides, HV can be expressed as an m-D integral:

$$\mathrm{HV}(A) = \int_{\mathbb{R}^m} \mathbf{1}_{\cup_{i=1}^n [\mathbf{0}, \mathbf{f}^{(i)}]}(\mathbf{z}) \mathrm{d}\mathbf{z}$$

where $\mathbf{1}_G(\mathbf{z})$ is the indicator function of a set G.

HV is equipped with many good properties such as the strict compliance with the Pareto optimality and the ease of use without much prior knowledge of the distribution of the optimal solution set (Li and Yao 2019). Therefore, it has been widely used as a performance indicator to compare the performance of multi-objective optimization algorithms. However, the computation of HV has been proved to be #**P**-hard (Bringmann and Friedrich 2010). Despite the theoretical bottleneck, there are many efficient algorithms for exactly or approximately computing HV.

Based on the concept of HV, the hypervolume improvement (HVI) is the improvement of HV after some new points are added into a point set.

Definition 4 (HVI). *Given $A \subset \mathbb{R}^m_+, B \subset \mathbb{R}^m_+$, the HVI of B with respect to A is defined by:*

$$\mathrm{HVI}(B, A) = \mathrm{HV}(A \cup B) - \mathrm{HV}(A). \tag{4}$$

Applying the idea of EI to HVI, the concept of EHVI is originally suggested in (Emmerich, Giannakoglou, and Naujoks 2006).

Definition 5 (EHVI). *Given $A \subset \mathbb{R}^m_+$ and a posterior distribution of a candidate $\mathbf{x}$ with mean $\boldsymbol{\mu}(\mathbf{x})$ and covariance $\boldsymbol{\Sigma}(\mathbf{x})$, the* EHVI *of $\mathbf{x}$ with respect to A is defined by:*

$$\begin{aligned}&\mathrm{EHVI}(\mathbf{x}, A; \boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x})) = \\ &\quad \mathbb{E}_{\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x}))} \left[\mathrm{HVI}(\{\mathbf{f}(\mathbf{x})\}, A)\right].\end{aligned} \tag{5}$$

Similarly to EI, it is convenient to drop $\mathbf{x}$ and write:

$$\mathrm{EHVI}(A; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathbb{E}_{\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \left[\mathrm{HVI}(\{\mathbf{y}\}, A)\right].$$

As mentioned before, the application of EHVI meets the difficulty in its computation. Currently, the fastest algorithm for EHVI is the KMAC algorithm developed in (Yang et al. 2019a). This algorithm runs in $O(n \log n)$ time and $O(n)$ space for $m = 2, 3$ and in $O(n^{\lfloor \frac{m}{2} \rfloor})$ time and $O(n^{\lfloor \frac{m}{2} \rfloor})$ space for $m \geq 4$ (assuming m to be a constant). To our knowledge, the implementation in BoTorch (Balandat et al. 2020) is the only available implementation for KMAC.

Variants of EHVI Some variants of EHVI have been designed for different purpose. qEHVI aiming to batch optimization is an extension of EHVI like qEI to EI (Daulton, Balandat, and Bakshy 2020). The formal definition of qEHVI is given below.

Definition 6 (qEHVI). *Given $A \subset \mathbb{R}^m_+$ and a posterior distribution of q candidates $\{\mathbf{x}^{(i)}\}_{i=1}^q$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (should not be confused with those in* EHVI*), the q*EHVI *of $\{\mathbf{x}^{(i)}\}_{i=1}^q$ with respect to A is defined by:*

$$\begin{aligned}&q\mathrm{EHVI}(A; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \\ &\mathbb{E}_{\{\mathbf{y}^{(i)}\}_{i=1}^q \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \left[\mathrm{HVI}(\{\mathbf{y}^{(i)}\}_{i=1}^q, A)\right].\end{aligned} \tag{6}$$

Similar to qEI, $\mathbf{y}^{(i)}$'s in qEHVI are correlated and difficult to handle. Until now, there is no exact algorithm for qEHVI when $q > 1$. The approximation method based on Monte-Carlo sampling has been designed for qEHVI, as well as its gradient, in (Daulton, Balandat, and Bakshy 2020). According to the numerical experiments, the approximated qEHVI can obtain a competitive performance to EHVI.

LogEHVI and qLogEHVI are proposed in (Ament et al. 2023). Noting that the gradient of traditional EI-like acquisition functions may vanish in the case of a small variance, the logarithm of the traditional EI, named LogEI, is proposed in (Ament et al. 2023). As an application of LogEI in MOBO, LogEHVI shows a significant advantage over EHVI in the numerical experiments. Furthermore, qEHVI is revised as qLogEHVI in the same work.

For completeness, we point out that there are also several variants of EHVI such as the truncated EHVI (Yang et al. 2016) and the combination between the Tchebycheff scalarization method and EHVI (Zhao and Zhang 2024).

EHVI As a Hypervolume Improvement

In this section, we give the main result of this paper, including the formulation of EHVI as a simple hypervolume improvement, and its applications both in theory and practice.

Theorem 1. *Given $A = \{\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(n)}\} \subset \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^m$, $\boldsymbol{\Sigma} \in \mathcal{S}^{m \times m}_{++}$ in the* EHVI *problem, define $\tilde{A} = \{\tilde{\mathbf{f}}^{(1)}, \ldots, \tilde{\mathbf{f}}^{(n)}\} \subset \mathbb{R}^m$, $\tilde{\mathbf{r}} \in \mathbb{R}^m$ by*

$$\tilde{\mathbf{r}}_j = \mathrm{EI}(0, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{jj}), \forall 1 \leq j \leq m \tag{7}$$

and

$$\tilde{\mathbf{f}}^{(i)}_j = \tilde{\mathbf{r}}_j - \mathrm{EI}(\mathbf{f}^{(i)}_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{jj}), \forall 1 \leq i \leq n, 1 \leq j \leq m. \tag{8}$$

Then,

$$\mathrm{EHVI}(A; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{HVI}(\{\tilde{\mathbf{r}}\}, \tilde{A}). \tag{9}$$

Proof. The EHVI can be written as a repeated integral:

$$\mathrm{EHVI}(A;\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_{\mathbb{R}^m} \mathrm{HVI}(\{\mathbf{z}\},A)\xi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})\mathrm{d}\mathbf{z}$$
$$= \int_{\mathbb{R}^m}\left\{\int_{\mathbb{R}^m} \mathbf{1}_{[\mathbf{0},\mathbf{z}]\setminus\cup_{i=1}^n[\mathbf{0},\mathbf{f}^{(i)}]}(\mathbf{w})\mathrm{d}\mathbf{w}\right\}\xi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})\mathrm{d}\mathbf{z}.$$

For the indicator function, we have:

$$\begin{aligned}&\mathbf{1}_{[\mathbf{0},\mathbf{z}]\setminus\cup_{i=1}^n[\mathbf{0},\mathbf{f}^{(i)}]}(\mathbf{w})\\ &=\begin{cases}1, & \mathbf{w}\in[\mathbf{0},\mathbf{z}]\setminus\cup_{i=1}^n[\mathbf{0},\mathbf{f}^{(i)}]\\ 0, & \text{otherwise}\end{cases}\\ &=\begin{cases}1, & \mathbf{z}\in[\mathbf{w},+\boldsymbol{\infty}),\mathbf{w}\in[\mathbf{0},+\boldsymbol{\infty})\setminus\cup_{i=1}^n[\mathbf{0},\mathbf{f}^{(i)}]\\ 0, & \text{otherwise.}\end{cases}\end{aligned} \tag{10}$$

Further, the order of the repeated integral is exchangeable as $[\mathbf{0},\mathbf{z}]\setminus\cup_{i=1}^n[\mathbf{0},\mathbf{f}^{(i)}]$ is a measurable set and $\xi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})$ is a nonnegative measurable function. After exchanging the integration order, it gives:

$$\int_{\Psi}\left\{\int_{[\mathbf{w},+\boldsymbol{\infty})}\xi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})\mathrm{d}\mathbf{z}\right\}\mathrm{d}\mathbf{w} \tag{11}$$

where $\Psi := [\mathbf{0},+\boldsymbol{\infty})\setminus\cup_{i=1}^n[\mathbf{0},\mathbf{f}^{(i)}]$. Due to the mutual independence between objective functions, it further equals:

$$\begin{aligned}&\int_{\Psi}\left\{\prod_{i=1}^m\int_{w_i}^{+\infty}\xi_{\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_{ii}}(z_i)\mathrm{d}z_i\right\}\mathrm{d}\mathbf{w}\\ &=\int_{\Psi}\prod_{i=1}^m\left\{\Phi\left(\frac{\boldsymbol{\mu}_i-w_i}{\sqrt{\boldsymbol{\Sigma}_{ii}}}\right)\mathrm{d}w_i\right\}.\end{aligned} \tag{12}$$

Define $\tilde{\mathbf{w}}\in\mathbb{R}^m$ such that for $1\le i\le m$,

$$\begin{aligned}\tilde{w}_i &= \int_{w_i}^{+\infty}\Phi\left(\frac{\boldsymbol{\mu}_i-z_i}{\sqrt{\boldsymbol{\Sigma}_{ii}}}\right)\mathrm{d}z_i\\ &=\sqrt{\boldsymbol{\Sigma}_{ii}}\phi\left(\frac{\boldsymbol{\mu}_i-w_i}{\sqrt{\boldsymbol{\Sigma}_{ii}}}\right)+(\boldsymbol{\mu}_i-w_i)\Phi\left(\frac{\boldsymbol{\mu}_i-w_i}{\sqrt{\boldsymbol{\Sigma}_{ii}}}\right)\\ &=\mathrm{EI}\left(w_i,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_{ii}\right).\end{aligned} \tag{13}$$

Then, we have $\mathrm{d}\tilde{w}_i = -\Phi\left(\frac{\boldsymbol{\mu}_i-w_i}{\sqrt{\boldsymbol{\Sigma}_{ii}}}\right)\mathrm{d}w_i$. As $\tilde{w}_i$ is a strictly decreasing function of w_i, we can transform the coordinate system in Equations (12) from $\mathbf{w}$ to $\tilde{\mathbf{w}}$. Recall the definition of $\tilde{\mathbf{r}}$ in Equation (7):

$$\tilde{\mathbf{r}}_j = \mathrm{EI}(0,\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_{jj}) = \int_0^{+\infty}\Phi\left(\frac{\boldsymbol{\mu}_i-w_i}{\sqrt{\boldsymbol{\Sigma}_{ii}}}\right)\mathrm{d}w_i,$$

which is exactly corresponding to the lower bound 0 of Ψ. Similarly for $\tilde{\mathbf{f}}^{(i)}$ in Equation (8), and we have $\tilde{\mathbf{r}}_j \ge \tilde{\mathbf{f}}_j^{(i)}, \forall 1\le i\le n$. Then, when the coordinate system is transformed to $\tilde{\mathbf{w}}$, the integration region Ψ is transformed as:

$$\tilde{\Psi} := [\mathbf{0},\tilde{\mathbf{r}}]\setminus\cup_{i=1}^n[\mathbf{0},\tilde{\mathbf{f}}^{(i)}].$$

Finally, the proof finishes by:

$$\mathrm{EHVI}(A;\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_{\tilde{\Psi}}\prod_{i=1}^m\mathrm{d}\tilde{w}_i = \lambda_m(\tilde{\Psi}) = \mathrm{HVI}(\{\tilde{\mathbf{r}}\},\tilde{A}). \tag{14}$$

□

Based on **Theorem 1** and the latest results on HVI, we can immediately obtain the following corollaries:

Corollary 1. EHVI *can be computed in* $\Theta(n\log n)$ *time and* $O(n)$ *space in the cases of* $m=2,3$*, and in* $\tilde{O}(n^{\frac{m}{3}})$ *time and* $O(n)$ *space in the case of* $m\ge 4$*.*

Proof. Assuming that the value of $\Phi(\cdot)$ is computed in constant time, the coordinates of $\tilde{\mathbf{r}}$ and points in $\tilde{A}$ in **Theorem 1** can be constructed in $O(mn)$ time. Then, this corollary follows from the complexity of the best known algorithms for HV (Beume et al. 2009; Chan 2013). □

Corollary 2. EHVI *is #**P**-hard and approximating it by a factor of* $2^{m^{1-\epsilon}}$ *is **NP**-hard for any* $\epsilon>0$*.*

Proof. **Theorem 1** has shown that an algorithm for computing the HVI of a single point ($\tilde{\mathbf{r}}$ in Equation (9)) leads to an algorithm for EHVI, after at most $O(mn)$ extra operations. The opposite construction also exists, as $\mathrm{EI}(z,\mu,\sigma^2)$ is an invertible function of z. Thus, the **NP**-hardness of EHVI and HVI of a single point is the same.

In **Theorem 1** in (Bringmann and Friedrich 2012), it is proved that computing the HVI of a single point (a.k.a. the hypervolume contribution) is #**P**-hard and approximating it by a factor of $2^{m^{1-\epsilon}}$ is **NP**-hard for any $\epsilon>0$ (assuming n and m are variables). Therefore, the same results also apply to EHVI. □

Corollary 3. *The first and second derivatives of* EHVI *with respect to* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *can be computed in* $O(n\log n)$ *time in the cases of* $m=2,3$*.*

Proof. The derivatives of HV in the cases of $m=2,3$ are given in (Emmerich and Deutz 2012; Deutz, Emmerich, and Wang 2023), with a time complexity of $O(n\log n)$. Thus, the derivatives of EHVI with respect to $\boldsymbol{\mu},\boldsymbol{\Sigma}$, and furthermore $\mathbf{x}$, can be computed using the chain rule based on Equation (9). For $m>3$, the derivatives of EHVI are also computable using the methods for HV in the above two literature. □

We would like to make the following comments:

- Figure 1 illustrates the idea of our formulation of EHVI in the case of $m=2$. The expectation in the objective space (left) is calculated based on the HVI of EI values (right) of $\mathbf{f}^{(i)}$ and $\mathbf{r}$ (i.e., $\tilde{\mathbf{f}}^{(i)}$ and $\tilde{\mathbf{r}}$ in Equation (9)).
- **Theorem 1** reveals a direct relationship between computing HV and EHVI. All the algorithms for computing HV are also applicable to EHVI. Although the cost of computing HV also suffers from its #**P**-hardness, there have been a number of practical algorithms for efficient HV computation, such as HBDA (Lacour, Klamroth, and Fonseca 2017), HV4D$^+$ (Guerreiro and Fonseca 2018), WFG (While, Bradstreet, and Barone 2012) and QHV-II (Jaszkiewicz 2018). Among them, QHV-II has shown the state-of-the-art performance in high-dimensional objective space.

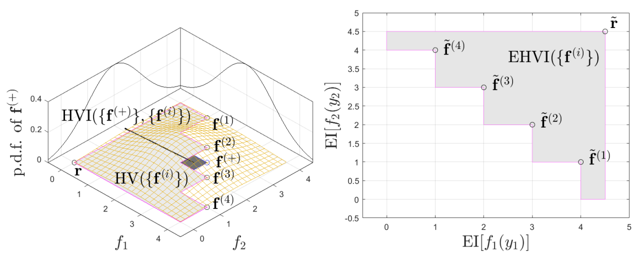

Figure 1: Illustration of the original formulation (left) and our formulation of EHVI (right) in the case of $m = 2$

- Compared with the existing fastest exact algorithm KMAC (Yang et al. 2019a), **Corollary 1** improves the worst-case complexity for computing EHVI both in time and in space for any $m \geq 4$. In addition, our formulation is proven exactly equal to EHVI, while the formulations in (Wagner et al. 2010; Svenson 2011) are not because their do not calculate EI but the expectations of predictive distributions and they use a fixed reference point.
- To our knowledge, there is no formal proof stating that EHVI is #**P**-hard, although this has been widely accepted as EHVI is based on HV. **Corollary 2** fills this gap.
- According to **Corollary 2**, it is unlikely to approximate EHVI within a desirable error bound in polynomial time. Specifically, the approximation method in (Daulton, Balandat, and Bakshy 2020) relies on the exact computation of HVI and thus needs exponential running time, whereas the approximation methods in (Rahat et al. 2022; Pang et al. 2023) rely on fixed-number Monte-Carlo sampling for approximating HVI so that they may not approximate EHVI within a desirable error bound.
- Although a fully polynomial-time approximation method hardly exists, practical approximation methods can be designed based on Equation (9). Specifically, advanced approximation methods for HV can be applied for $\mathrm{HVI}(\{\tilde{\mathbf{r}}\}, \tilde{A})$ in the equation, for example, the FPRAS in (Bringmann and Friedrich 2010), the coordinate transformation method in (Deng and Zhang 2019), and learning-based algorithms in (Shang et al. 2023; Boelrijk, Ensing, and Forré 2023). In fact, these methods may provide unsatisfactory results only in extreme cases.'
- **Corollary 3** transforms the problem of derivatives of EHVI as derivatives of HV. Applying results of derivatives of HV, we can obtain better results than those investigated in (Yang et al. 2019b; Daulton, Balandat, and Bakshy 2020).

qEHVI As a Sum of Hypervolume Improvements

In this section, we present the results for qEHVI parallel to those for EHVI:

Theorem 2. *Given* $A = \{\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(n)}\} \subset \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^{qm}$, $\boldsymbol{\Sigma} \in \mathcal{S}_{++}^{qm \times qm}$ *in the* qEHVI *problem. For each index set* $I \subset \{1, \ldots, q\}$*, define* $\hat{A}(I) = \{\hat{\mathbf{f}}^{(1)}(I), \ldots, \hat{\mathbf{f}}^{(n)}(I)\} \subset \mathbb{R}^m, \hat{\mathbf{r}}(I) \in \mathbb{R}^m$ *by*

$$\hat{\mathbf{r}}_j(I) = q\mathrm{MinEI}(0, \boldsymbol{\mu}_I^{(j)}, \boldsymbol{\Sigma}_{I \times I}^{(j)}), \forall 1 \leq j \leq m \tag{15}$$

and for $1 \leq i \leq n, 1 \leq j \leq m$,

$$\hat{\mathbf{f}}_j^{(i)}(I) = \hat{\mathbf{r}}_j(I) - q\mathrm{MinEI}(\mathbf{f}_j^{(i)}, \boldsymbol{\mu}_I^{(j)}, \boldsymbol{\Sigma}_{I \times I}^{(j)}), \tag{16}$$

where $\boldsymbol{\mu}_I^{(j)}, \boldsymbol{\Sigma}_{I \times I}^{(j)}$ *are the mean and covariance matrix of* $(\mathbf{y}_j^{(i)})_{i \in I}$ *(see Equation (6)), respectively. Then,*

$$q\mathrm{EHVI}(A; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{I \subset \{1, \ldots, q\}} (-1)^{|I|+1} \mathrm{HVI}(\{\hat{\mathbf{r}}(I)\}, \hat{A}(I)). \tag{17}$$

Proof. See **Appendix A.1**. □

Based on **Theorem 2** and the latest results on HVI and qEI, we can immediately obtain the following corollaries:

Corollary 4. qEHVI *can be computed in* $\Theta(n \log n)$ *time and* $O(n)$ *space in the cases of* $m = 2, 3$ *and* $\tilde{O}(n^{\frac{m}{3}})$ *time and* $O(n)$ *space in the case of* $m \geq 4$.

Proof. Assuming that q is a constant and the value of the c.d.f. of multivariate normal distributions is computed in constant time, the coordinates of $\hat{\mathbf{r}}(I)$ and points in $\hat{A}(I)$ in **Theorem 2** can be constructed in $O(mn)$ time based on **Lemma 1**. Then, this corollary follows from the complexity of the best known algorithms for HV (Beume et al. 2009; Chan 2013). □

Corollary 5. qEHVI *is #**P**-hard and approximating it by a factor of* $2^{m^{1-\epsilon}}$ *is **NP**-hard for any* $\epsilon > 0$.

Proof. This corollary can be proved analogously to **Corollary 2** based on the fact that $q\mathrm{MinEI}(z, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a strictly decreasing function of z. □

Corollary 6. *The gradient of* qEHVI *with respect to* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *can be computed in* $O(n \log n)$ *time in the cases of* $m = 2, 3$.

Proof. This corollary can be proved analogously to **Corollary 3** based on the fact that the gradient of qEI can be computed with $O(q^{q-k+1})$ evaluations of the c.d.f. of k-variate normal distributions for $q - 3 \leq k \leq q$, as proved in (Marmin, Chevalier, and Ginsbourger 2015). □

We would like to make the following comments:

- **Theorem 2** is the first result of the analytic expression of qEHVI when $q > 1$. As qMinEI reduces to EI when $q = 1$, it is easily seen that qEHVI also reduces to EHVI when $q = 1$.
- In the analysis of **Corollary 4**, q is assumed to be constant. However, the cost of computing c.d.f.'s may become an issue when q is comparable to m. To be clear, the cost of computing HV is $\tilde{O}(n^{\frac{m}{3}})$, while the cost of computing c.d.f. of $\frac{q}{2}$-variate normal distributions is $O(mnq^{\frac{3}{2}}2^q)$ (See **Appendix A.2**). Thus, the cost of computing c.d.f.'s may dominate when $q = \Omega(m \log n)$. The influence of q will be investigated in the numerical experiments.

m	n	q	b
$2,3,4,5,6,7,8$	$10,20,40,80$	$1,2,4,6$	51

Table 2: The settings of (m, n, q, b) in the experiments

- Due to the difficulty caused by q in computing qEHVI, there are two strategies when designing approximation methods for qEHVI. One is approximating c.d.f.'s of multivariate normal distributions and then computing an exact HVI value, and the other vice versa. The former can ensure the approximation error and it is more efficient when q is large, while the latter is more practical when n and m are large.

Experiments

In this section, we experimentally examine the correctness and efficiency of our formulations for EHVI and qEHVI.

Settings

The platform is the BoTorch platform (Balandat et al. 2020) on 64-bit Windows with Intel 2.60GHz Core i5-13500H processor and 16GB of RAM. We implement our method in a Tensor manner of BoTorch, and compare it with the exact and approximation methods for EHVI and qEHVI in BoTorch. The tested algorithms include:

- qEHVI-HVI[4]: the implementation of our formulation for qEHVI ($q \geq 1$), where the exact values of HV are computed by QHV-II (Jaszkiewicz 2018);
- DBB[5]: the exact method for EHVI in (Daulton, Balandat, and Bakshy 2020). It should be noted that DBB cannot compute qEHVI when $q > 1$;
- qDBB-MC[5]: the quasi Monte-Carlo based method for qEHVI ($q \geq 1$) in (Daulton, Balandat, and Bakshy 2020) using default 128 samples.

Following researches in the field of EHVI computation (Yang et al. 2019a), we generate artificial random point sets of different sizes n as the point set A in the qEHVI ($q \geq 1$) problem. Given a set of evaluated solutions $\mathcal{D} = \{\mathbf{x}^{(i)}, \mathbf{f}(\mathbf{x}^{(i)})\}_{i=1}^{K}$ randomly sampled by BoTorch, a multi-task GP model is trained. Then, we randomly generate different batches of candidates $\{\mathbf{x}^{(i)}\}_{i=1}^{q}$ and obtain the posterior distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. With these data, the above algorithms are applied for computing $q\text{EHVI}(A; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The experimental settings of the number of objectives m, the number of points n in A, the batch size q, and the number of batches b are presented in Table 2. For each group of (m, n, q), we generate b different batches of test samples $\{\mathbf{x}^{(i)}\}_{i=1}^{q}$. Larger values of (m, n, q) are not tested because at least one tested algorithm spends too much time.

For EHVI, we verify our implementation for computing its value and gradient. The gradient of EHVI is computed by the chain rule based on Equation (9):

$$\frac{\partial \text{EHVI}(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\partial \text{HVI}(\hat{\mathbf{r}}_j, \hat{A})}{\partial (\hat{\mathbf{r}}_j, \hat{\mathbf{f}}_j^{(i)})} \frac{\partial (\hat{\mathbf{r}}_j, \hat{\mathbf{f}}_j^{(i)})}{\partial \mathbf{x}}$$

where $\frac{\partial \text{HVI}(\hat{\mathbf{r}}_j, \hat{A})}{\partial (\hat{\mathbf{r}}_j, \hat{\mathbf{f}}_j^{(i)})}$ is computed based on the algorithm in (Deutz, Emmerich, and Wang 2023),[6] and $\frac{\partial (\hat{\mathbf{r}}_j, \hat{\mathbf{f}}_j^{(i)})}{\partial \mathbf{x}}$ is provided by BoTorch via the auto-differentiation technique.

For qEHVI ($q > 1$), we verify our implementation for computing its value. Since there is no implementation for computing the c.d.f. of multivariate normal distributions in BoTorch, as well as its gradient, qMinEI cannot be computed in a Tensor manner of BoTorch. Instead, we use the `scipy.stats.multivariate_normal` class for qMinEI. Due to the same reason, the gradient of qEHVI based on our formulation is complex to be implemented in BoTorch. Therefore, we do not test the gradient of qEHVI in the experiments.

To compare the running time of algorithms, we compute the average and standard deviation over 51 batches, and conduct Wilcoxon rank sum test with a significance level of 0.01, where the symbols '$+$', '$\approx$' and '$-$' denote that the result of one algorithm is significantly faster, statistically similar and significantly slower to our qEHVI-HVI, respectively.

Results and Analysis

Due to the space limitation, in the main text, we only present the results for computing EHVI and the gradient of EHVI when $n = 10$ in Table 3, while the other results are given in **Appendix B**. We have the following observations:

- The relative difference between our implementation and the exact method DBB never exceeds 10^{-13}, not only for EHVI but also for the gradient of EHVI (in terms of the distance). Therefore, our implementation is correct.
- In terms of the running time, all the tested algorithms are similar when $m = 2$. This is because they share the same time complexity of $O(n \log n)$ in the case of $m = 2$. For larger values of m, our implementation is generally faster than DBB and qDBB-MC, while DBB and qDBB-MC cost similar time. As m and n increase, the advantage of our implementation is more significant. Specifically, it is **hundreds times faster** than DBB and qDBB-MC for computing EHVI when $m \geq 7$ and $n = 10$. This is because our implementation is based on QHV-II, an advanced algorithm for computing HV, while DBB and qDBB-MC relies on the box decomposition method, which is time-consuming and space-consuming in high-dimensional objective spaces.
- In general, the approximation method qDBB-MC with 128 samples has a satisfactory 3%-10% approximation error. However, when $m \leq 4$, the performance of qDBB-MC with 128 samples is unstable. To further examine the quality of qDBB-MC, we also test qDBB-MC with 1280 and 12800 samples on $n = 10$ test sets. The

[4]Our codes and the supplementary materials are available at https://github.com/Ksrma/EHVI-HVI.

[5]Downloaded from https://github.com/pytorch/botorch.

[6]Downloaded from https://github.com/wangronin/HypervolumeDerivatives.

m			2	3	4	5	6	7	8
EHVI	Time	DBB	0.0027 (–) (0.0002)	0.020 (–) (0.002)	0.058 (–) (0.016)	0.203 (–) (0.018)	1.10 (–) (0.04)	6.98 (–) (0.07)	34.14 (–) (0.19)
		qDBB-MC	0.0040 (–) (0.0006)	0.0237 (–) (0.0024)	0.063 (–) (0.044)	0.169 (–) (0.010)	1.06 (–) (0.03)	6.79 (–) (0.06)	33.82 (–) (0.18)
		qEHVI-HVI	**0.0023** (0.0006)	**0.0054** (0.0004)	**0.010** (0.001)	**0.015** (0.003)	**0.015** (0.002)	**0.038** (0.004)	**0.063** (0.007)
	Error	qDBB-MC	0.2120 (0.1730)	0.0660 (0.0502)	0.060 (0.059)	0.034 (0.028)	0.030 (0.021)	0.031 (0.029)	0.017 (0.016)
		qEHVI-HVI	<2e-14	<1e-14	<1e-14	<1e-14	<1e-14	<2e-14	<3e-14
Gradient of EHVI	Time	DBB	0.0040 (–) (0.0003)	0.022 (–) (0.002)	0.083 (–) (0.015)	0.180 (–) (0.032)	1.14 (–) (0.03)	7.13 (–) (0.11)	34.64 (–) (0.14)
		qDBB-MC	0.0054 (–) (0.0007)	0.0244 (–) (0.0017)	0.061 (–) (0.007)	0.186 (–) (0.009)	1.07 (–) (0.03)	6.86 (–) (0.05)	33.87 (–) (0.15)
		qEHVI-HVI	**0.0036** (0.0003)	**0.0091** (0.0006)	**0.034** (0.003)	**0.067** (0.025)	**0.108** (0.018)	**0.185** (0.070)	**0.441** (0.192)
	Error	qDBB-MC	0.2274 (0.1310)	0.0886 (0.0527)	0.049 (0.028)	0.085 (0.041)	0.042 (0.021)	0.033 (0.020)	0.025 (0.011)
		qEHVI-HVI	<7e-15	<6e-15	<6e-15	<2e-14	<7e-15	<1e-14	<7e-15

Table 3: Results in the case of $n = 10$ on the average running time and standard deviation (in brackets) of computing EHVI and the gradient of EHVI. The fastest method is shown in **bold**. The error of qDBB-MC is the average relative error between qDBB-MC and the exact method DBB, while that of qEHVI-HVI is the maximum relative error between qEHVI-HVI and DBB

results are given in **Appendix B**. It is seen that 1280 samples are sufficient to qDBB-MC for approximating EHVI and the gradient of EHVI.

The results of computing qEHVI ($q > 1$) are also presented in **Appendix B**. As there is no other exact method for qEHVI, we compare our qEHVI-HVI with qDBB-MC using 128, 1280, and 12800 samples, respectively. We have the following observations:

- The relative errors between qEHVI-HVI and qDBB-MC become smaller when qDBB-MC uses more samples. Therefore, the correctness of our implementation is convinced. Besides, it is found that 128 samples are not enough for qDBB-MC to provide a satisfactory approximation of qEHVI.
- In terms of the running time, the performance of qDBB-MC for approximating qEHVI is similar to that of approximating EHVI, while the performance of qEHVI-HVI is quite different as q increases. When $q = 2$, qEHVI-HVI is slightly slower than qDBB-MC when $m = 2$ but generally faster than qDBB-MC in the cases of $m \geq 3$. For $q = 4$ or 6, qEHVI-HVI is already very slow when $m = 2$ and $n = 10$, although it may become more efficient than qDBB-MC when m is large enough. This result validates the analysis in the previous section: The cost of computing c.d.f.'s of multivariate normal distributions will dominate when q is large.
- In summary, qEHVI-HVI and qDBB-MC have their own strengths for qEHVI ($q > 1$). The running time of qEHVI-HVI does not increase much as m and n increase, while the running time of qDBB-MC does not increase much as q increases. Of course, it should be noted again that qEHVI-HVI is an exact method whereas qDBB-MC has a Monte-Carlo integration error.

Conclusion

In this work, we have investigated the fundamental concept expected hypervolume improvement (EHVI) in multi-objective Bayesian optimization. We prove that EHVI is by nature a particular hypervolume improvement after a coordinate system transformation. This proof reveals a direct relationship between EHVI and HV, followed by the hardness proof of EHVI in theory and more efficient methods for computing EHVI and its derivatives in practice. Further, the multi-point version of EHVI, qEHVI, is expressed as the sum of a constant number of hypervolume improvements. This formulation not only first provides the analytic expression of qEHVI, but also leads to many theoretical and practical results of qEHVI parallel to those of EHVI. In the numerical experiments, the new formulations of EHVI and qEHVI are verified and show significantly better performance than the existing fastest (deterministic or random) algorithms on a number of random point sets.

In the future, we will continue to explore applications of the new formulations to other variants of EHVI such as the truncated EHVI (Yang et al. 2016), LogEHVI, and qLogEHVI (Ament et al. 2023), and more expensive multi-objective optimization problems such as the correlated problems (Yang et al. 2023), constrained problems, and noisy problems (Daulton, Balandat, and Bakshy 2020).

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 11991023, 62076197, 62276223, 62072364, in part by the Research Grants Council of the Hong Kong Special Administrative Region, China (GRF Project No. 11215622, 11215723, 11208121), and in part by the Key Basic Research Foundation of Shenzhen, China (JCYJ20220818100005011).

References

Álvarez, M. A.; Rosasco, L.; and Lawrence, N. D. 2012. Kernels for Vector-Valued Functions: A Review. *Foundations and Trends® in Machine Learning*, 4(3): 195–266.

Ament, S.; Daulton, S.; Eriksson, D.; Balandat, M.; and Bakshy, E. 2023. Unexpected Improvements to Expected Improvement for Bayesian Optimization. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023*.

Balandat, M.; Karrer, B.; Jiang, D. R.; Daulton, S.; Letham, B.; Wilson, A. G.; and Bakshy, E. 2020. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*.

Beume, N.; Fonseca, C. M.; López-Ibáñez, M.; Paquete, L.; and Vahrenhold, J. 2009. On the Complexity of Computing the Hypervolume Indicator. *IEEE Transactions on Evolutionary Computation*, 13(5): 1075–1082.

Binois, M.; Picheny, V.; Taillandier, P.; and Habbal, A. 2020. The Kalai-Smorodinsky solution for many-objective Bayesian optimization. *Journal of Machine Learning Research*, 21: 150:1–150:42.

Boelrijk, J.; Ensing, B.; and Forré, P. 2023. Multi-objective optimization via equivariant deep hypervolume approximation. In *The Eleventh International Conference on Learning Representations, ICLR 2023*. OpenReview.net.

Bringmann, K.; and Friedrich, T. 2010. Approximating the volume of unions and intersections of high-dimensional geometric objects. *Computational Geometry*, 43(6-7): 601–610.

Bringmann, K.; and Friedrich, T. 2012. Approximating the least hypervolume contributor: NP-hard in general, but fast in practice. *Theoretical Computer Science*, 425: 104–116.

Chan, T. M. 2013. Klee's Measure Problem Made Easy. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 410–419. IEEE.

Chevalier, C.; and Ginsbourger, D. 2013. Fast Computation of the Multi-Points Expected Improvement with Applications in Batch Selection. In *Learning and Intelligent Optimization - 7th International Conference, LION 7, Revised Selected Papers*, volume 7997 of *Lecture Notes in Computer Science*, 59–69. Springer.

Couckuyt, I.; Deschrijver, D.; and Dhaene, T. 2014. Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. *Journal of Global Optimization*, 60(3): 575–594.

Daulton, S.; Balandat, M.; and Bakshy, E. 2020. Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*.

Daulton, S.; Balandat, M.; and Bakshy, E. 2023. Hypervolume Knowledge Gradient: A Lookahead Approach for Multi-Objective Bayesian Optimization with Partial Information. In *International Conference on Machine Learning, ICML 2023*, volume 202 of *Proceedings of Machine Learning Research*, 7167–7204. PMLR.

Deng, J.; and Zhang, Q. 2019. Approximating Hypervolume and Hypervolume Contributions Using Polar Coordinate. *IEEE Transactions on Evolutionary Computation*, 23(5): 913–918.

Deutz, A. H.; Emmerich, M. T. M.; and Wang, H. 2023. The Hypervolume Indicator Hessian Matrix: Analytical Expression, Computational Time Complexity, and Sparsity. In *Evolutionary Multi-Criterion Optimization - 12th International Conference, EMO 2023, Proceedings*, volume 13970 of *Lecture Notes in Computer Science*, 405–418. Springer.

Deutz, A. H.; Emmerich, M. T. M.; and Yang, K. 2019. The Expected R2-Indicator Improvement for Multi-objective Bayesian Optimization. In *Evolutionary Multi-Criterion Optimization - 10th International Conference, EMO 2019, Proceedings*, volume 11411 of *Lecture Notes in Computer Science*, 359–370. Springer.

Emmerich, M. T. M.; and Deutz, A. H. 2012. Time Complexity and Zeros of the Hypervolume Indicator Gradient Field. In *EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation III [EVOLVE 2012, selection of extended papers]*, volume 500 of *Studies in Computational Intelligence*, 169–193. Springer.

Emmerich, M. T. M.; Giannakoglou, K. C.; and Naujoks, B. 2006. Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4): 421–439.

Emmerich, M. T. M.; Yang, K.; and Deutz, A. H. 2020. Infill Criteria for Multiobjective Bayesian Optimization. In Bartz-Beielstein, T.; Filipic, B.; Korosec, P.; and Talbi, E., eds., *High-Performance Simulation-Based Optimization*, volume 833 of *Studies in Computational Intelligence*, 3–16. Springer.

Garnett, R. 2023. *Bayesian Optimization*. Cambridge University Press.

Gaudrie, D.; Riche, R. L.; Picheny, V.; Enaux, B.; and Herbert, V. 2020. Targeting solutions in Bayesian multi-objective optimization: sequential and batch versions. *Annals of Mathematics and Artificial Intelligence*, 88(1-3): 187–212.

Ginsbourger, D.; Riche, R. L.; and Carraro, L. 2008. A Multi-points Criterion for Deterministic Parallel Global Optimization based on Gaussian Processes. Technical Report hal-00260579, Ecole Nationale Supérieure des Mines de Saint-Etienne.

Guerreiro, A. P.; and Fonseca, C. M. 2018. Computing and Updating Hypervolume Contributions in Up to Four Dimensions. *IEEE Transactions on Evolutionary Computation*, 22(3): 449–463.

Jaszkiewicz, A. 2018. Improved quick hypervolume algorithm. *Computers & Operations Research*, 90: 72–83.

Jones, D. R.; Schonlau, M.; and Welch, W. J. 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4): 455–492.

Lacour, R.; Klamroth, K.; and Fonseca, C. M. 2017. A box decomposition algorithm to compute the hypervolume indicator. *Computers & Operations Research*, 79: 347–360.

Li, M.; and Yao, X. 2019. Quality Evaluation of Solution Sets in Multiobjective Optimisation: A Survey. *ACM Computing Surveys*, 52(2): 26:1–26:38.

Lin, X.; Yang, Z.; Zhang, X.; and Zhang, Q. 2022. Pareto Set Learning for Expensive Multi-Objective Optimization. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022*.

Marmin, S.; Chevalier, C.; and Ginsbourger, D. 2015. Differentiating the Multipoint Expected Improvement for Optimal Batch Design. In *Machine Learning, Optimization, and Big Data - First International Workshop, MOD 2015, Revised Selected Papers*, volume 9432 of *Lecture Notes in Computer Science*, 37–48. Springer.

Pang, Y.; Wang, Y.; Zhang, S.; Lai, X.; Sun, W.; and Song, X. 2023. An Expensive Many-Objective Optimization Algorithm Based on Efficient Expected Hypervolume Improvement. *IEEE Transactions on Evolutionary Computation*, 27(6): 1822–1836.

Rahat, A. A. M.; Chugh, T.; Fieldsend, J. E.; Allmendinger, R.; and Miettinen, K. 2022. Efficient Approximation of Expected Hypervolume Improvement Using Gauss-Hermite Quadrature. In *Parallel Problem Solving from Nature - PPSN XVII - 17th International Conference, PPSN 2022, Proceedings, Part I*, volume 13398 of *Lecture Notes in Computer Science*, 90–103. Springer.

Rasmussen, C. E. 2003. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning, ML Summer Schools 2003, Revised Lectures*, volume 3176 of *Lecture Notes in Computer Science*, 63–71. Springer.

Shang, K.; Chen, W.; Liao, W.; and Ishibuchi, H. 2023. HV-Net: Hypervolume Approximation Based on DeepSets. *IEEE Transactions on Evolutionary Computation*, 27(4): 1154–1160.

Svenson, J. 2011. *Computer experiments: multiobjective optimization and sensitivity analysis*. Ph.D. diss., Ohio State Univ.

Tanabe, R.; and Ishibuchi, H. 2020. An easy-to-use real-world multi-objective optimization problem suite. *Applied Soft Computing*, 89: 106078.

Wagner, T.; Emmerich, M.; Deutz, A. H.; and Ponweiser, W. 2010. On Expected-Improvement Criteria for Model-based Multi-objective Optimization. In *Parallel Problem Solving from Nature - PPSN XI, 11th International Conference, Proceedings, Part I*, volume 6238 of *Lecture Notes in Computer Science*, 718–727. Springer.

While, L.; Bradstreet, L.; and Barone, L. 2012. A Fast Way of Calculating Exact Hypervolumes. *IEEE Transactions on Evolutionary Computation*, 16(1): 86–95.

Yang, K.; Chen, K.; Affenzeller, M.; and Werth, B. 2023. A New Acquisition Function for Multi-objective Bayesian Optimization: Correlated Probability of Improvement. In *Companion Proceedings of the Conference on Genetic and Evolutionary Computation, GECCO 2023, Companion Volume*, 2308–2317. ACM.

Yang, K.; Deutz, A. H.; Yang, Z.; Bäck, T.; and Emmerich, M. T. M. 2016. Truncated expected hypervolume improvement: Exact computation and application. In *IEEE Congress on Evolutionary Computation, CEC 2016*, 4350–4357. IEEE.

Yang, K.; Emmerich, M. T. M.; Deutz, A. H.; and Bäck, T. 2019a. Efficient Computation of Expected Hypervolume Improvement Using Box Decomposition Algorithms. *Journal of Global Optimization*, 75(1): 3–34.

Yang, K.; Emmerich, M. T. M.; Deutz, A. H.; and Bäck, T. 2019b. Multi-Objective Bayesian Global Optimization using expected hypervolume improvement gradient. *Swarm and Evoluationary Computation*, 44: 945–956.

Zhang, Q.; Liu, W.; Tsang, E. P. K.; and Virginas, B. 2010. Expensive Multiobjective Optimization by MOEA/D With Gaussian Process Model. *IEEE Transactions on Evolutionary Computation*, 14(3): 456–474.

Zhao, L.; and Zhang, Q. 2023. Exact Formulas for the Computation of Expected Tchebycheff Improvement. In *IEEE Congress on Evolutionary Computation, CEC 2023*, 1–8. IEEE.

Zhao, L.; and Zhang, Q. 2024. Hypervolume-Guided Decomposition for Parallel Expensive Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation*, 28(2): 432–444.

Zitzler, E.; and Thiele, L. 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4): 257–271.