Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

BI-OBJECTIVE BAYESIAN OPTIMIZATION WITH TRANSFORMED ADDITIVE GAUSSIAN PROCESSES

Caroline M. Kerfonta[1], Qiong Zhang[1] and Margaret M. Wiecek[1]

[1]School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC, USA

ABSTRACT

Simulation is often used to optimize the performance of an engineering or scientific process. The input-output relation of a simulation model can be black-box functions and expensive to evaluate. Bayesian optimization, popularly used to improve efficiency in searching for optimal input settings, has been extended to bi-objective optimization problems. However, the common challenge in bi-objective Bayesian optimization is the computational challenge of maximizing the expected hypervolume improvement (EHI) to search for the input point to evaluate. In this paper, we utilize the transformed additive Gaussian process to simplify the objectives to additive functions for each dimension of the input space. Under this model assumption, the maximization of EHI over the entire decision space is reduced to the maximization of the resulting EHI functions for each dimension separately. We demonstrate the performances of the proposed method through numerical comparisons of the bi-objective Bayesian optimization with the traditional Gaussian process model.

1 INTRODUCTION

Simulation is a popular approach used to optimize engineering and scientific design (Lee et al. 2007; Zhang 2008; Fliege and Xu 2011; Hunter and McClosky 2016; Hunter et al. 2019; de Castro et al. 2022). For complex engineering design problems, running the corresponding simulation codes is often time-consuming. Also, the input and output relationship of the simulation model may not be expressed in a closed form to enable gradient-based optimization approaches. Thus, Bayesian optimization is often used to solve these problems by efficiently guiding new experiments through running the simulation models (Frazier 2018). Bayesian optimization approaches contain two key components: a computationally inexpensive surrogate model of the simulation and an acquisition function to select input points for the follow-up experiments. In terms of surrogate models, Gaussian process is a common choice (e.g., Jones et al. (1998)). A popular choice of the acquisition function is the expected improvement (EI, e.g, Qin et al. (2017)) to the target optimization problem over the input space. Then the input point for the new evaluation is given as a maximizer of EI over the input space.

Some engineering design problems involve two or more objectives. For multi-objective problems, the solution improvement can not be simply defined as for the single objective problems. In the literature of multi-objective optimization, the quality of the solution given by an optimization approach can be quantified by a variety of performance indicators, see various concepts given by Audet et al. (2021). Among them, the hypervolume indicator is a popular performance indicator to measure the quality of the solutions to multi-objective optimization problems (Knowles et al. 2003). Hence, in the literature on multi-objective Bayesian optimization (Hunter et al. 2019), a popular choice for the acquisition function is the expected hypervolume improvement (EHI), which is the expected difference between the hypervolume indicators computed after and before adding a new simulation evaluation (Emmerich et al. 2011). Even though EHI is widely used, there are some computational challenges to maximizing this acquisition function. When the dimension of the decision space increases, the optimization of EHI becomes much more complex, requires a longer computing time, and takes more memory space to complete (Hupkens et al. 2015). There are a number of existing approaches proposed to reduce the complexities of maximizing EHI. Yang et al. (2019)

consider bi-objective Bayesian optimization and provide an algorithm to evaluate the gradient function of EHI to enhance the optimization of EHI for searching new input points. A truncated form of the EHI calculation to optimize computing time is also proposed by Yang et al. (2016). This approach uses prior knowledge about the objective function values to more efficiently compute EHI. Since the gradient of EHI can be extremely complex and difficult to compute, Daulton et al. (2020) provide a method for computing EHI by using the gradients of the Monte Carlo estimator via auto-differentiation. This method simplifies the calculation of EHI by using first-order and quasi-second-order methods.

In this paper, we provide an algorithm for bi-objective Bayesian optimization to solve the computational challenges of maximizing EHI. Differently from perspectives in current literature, we approximate the objective functions to additive functions for each dimension of the decision space using the transformed additive Gaussian processes (Lin and Joseph 2020). Then the maximization of EHI over the entire decision space becomes the maximization of multiple univariate EHI functions, i.e., the number of EHI functions is equal to the number of dimensions of the input space. In our implementation, we maximizes univariate EHI functions for each dimension of the input space through enumerating on a set of discrete points. We compare the numerical performance of the proposed approach with the bi-objective Bayesian optimization under the classical Gaussian process model.

This paper is organized as follows. Section 2 gives a review of related background, including single-objective Bayesian optimization with Gaussian processes and EI, bi-objective optimization and the hypervolume indicator, and bi-objective Bayesian optimization with Gaussian processes and EHI. Section 3 proposes our method, which uses transformed additive Gaussian process (TAG) models as the surrogate models and derives the EHI expression under this model assumption. Section 4 provides a numerical study that demonstrates the advantages of the proposed method. Section 5 concludes the paper and points out directions for future improvement of the proposed method.

2 BACKGROUND

In this section, we review related background approaches including Bayesian optimization with Gaussian processes and EI, bi-objective optimization, the hypervolume indicator, and bi-objective Bayesian optimization with classical Gaussian processes and EHI.

2.1 Bayesian Optimization

Let $y(\boldsymbol{x})$ be a scalar deterministic black-box function with inputs $\boldsymbol{x} = (x_1, \ldots, x_d)^\top \in \mathscr{R}^d$. The goal of Bayesian optimization is to solve problems in the form:

$$\min_{\boldsymbol{x} \in \mathscr{X}} \quad y(\boldsymbol{x}), \tag{1}$$

where $\mathscr{X}$ is the domain of the decision. Bayesian optimization (Frazier 2018) is comprised of two parts: the surrogate statistical model and the acquisition function. The surrogate model is fitted by the data collected from evaluating the function $y(\boldsymbol{x})$ and informs the posterior distribution of $y(\boldsymbol{x})$ given the data. Under the distribution of $y(\boldsymbol{x})$, the acquisition function is optimized to select a new input point from $\mathscr{X}$. A generic Bayesian optimization algorithm is shown in Algorithm 1 for a budget of N function evaluation. The acquisition function is optimized in step 3, and the surrogate model is updated in step 4. Bayesian optimization algorithms could differ in different target optimization problems (e.g., single or multiple objectives), surrogate statistical models and acquisition functions.

We illustrate Bayesian optimization in detail using the Gaussian process as the surrogate model and the EI as the acquisition function. The Gaussian process model is a popular choice of surrogate models (Sacks et al. 1989). We assume that the function value $y(\boldsymbol{x})$ is a realization of the stochastic process

$$Y(\boldsymbol{x}) = \mu + Z(\boldsymbol{x}), \tag{2}$$

where μ is the unknown deterministic mean value, and $Z(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathscr{X}$ is mean zero Gaussian process with variance σ^2 and correlation function $R(\cdot)$, e.g., the Gaussian correlation function (Sacks et al. 1989)

Algorithm 1 Generic Bayesian Optimization Algorithm

1: Provide a prior distribution of $y(\boldsymbol{x})$
2: **for** $n = 0, 1, \ldots, N-1$ **do**
3: Compute acquisition function based on the distribution of $y(\boldsymbol{x})$ and find $\boldsymbol{x}_{n+1}$ that maximizes the acquisition over $\boldsymbol{x} \in \mathscr{X}$
4: Collect a new data point by evaluating $y(\boldsymbol{x})$ at $\boldsymbol{x}_{n+1}$ and update the distribution of $y(\boldsymbol{x})$ as the posterior distribution of $y(\boldsymbol{x})$ given all existing data.
5: **end for**
Return the input point with the smallest posterior mean as the optimal decision.

for two inputs points $\boldsymbol{x}$ and $\boldsymbol{x}'$ is given by

$$R\left(\boldsymbol{x}-\boldsymbol{x}'\right) = \exp\left(-\sum_{i=1}^{d} \theta_i (x_i - x_i')^2\right), \tag{3}$$

where θ_i's are the correlation parameters. Under this model assumption, given two input points $\boldsymbol{x}, \boldsymbol{x}' \in \mathscr{X}$, we have that

$$Y(\boldsymbol{x}) \sim N(\mu, \sigma^2) \quad \text{and} \quad \text{Cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) = \sigma^2 R(\boldsymbol{x}-\boldsymbol{x}').$$

Assume that we have evaluated the function $y(\boldsymbol{x})$ at n input points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, and collected the outputs $\boldsymbol{y}_n = (y(\boldsymbol{x}_1), \ldots, y(\boldsymbol{x}_n))^\top$. Under the Gaussian process assumption, the conditional distribution of $Y(\boldsymbol{x})$ given $\boldsymbol{y}_n$ at a new input point $\boldsymbol{x}$ (see for example, Jones et al. (1998)) is expressed by :

$$Y(\boldsymbol{x}) \mid \boldsymbol{y}_n \sim N(\hat{y}(\boldsymbol{x}), s^2(\boldsymbol{x})), \tag{4}$$

with mean

$$\hat{y}(\boldsymbol{x}) = \hat{\mu} + \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1} (\boldsymbol{y}_n - \mathbf{1}\hat{\mu}) \tag{5}$$

and variance

$$s^2(\boldsymbol{x}) = \sigma^2 \left(1 - \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}) - \frac{(1 - \mathbf{1}^\top \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}))^2}{\mathbf{1}^\top \boldsymbol{R}^{-1} \mathbf{1}}\right) \tag{6}$$

where $\mathbf{1}$ is a vector of ones of size n, $\hat{\mu} = \mathbf{1}^\top \boldsymbol{R}^{-1} \boldsymbol{y}_n / \mathbf{1}^\top \boldsymbol{R}^{-1} \mathbf{1}$ is the best unbiased linear estimator of μ, $\boldsymbol{r}(\boldsymbol{x}) = (R(\boldsymbol{x}-\boldsymbol{x}_1), \ldots, R(\boldsymbol{x}-\boldsymbol{x}_n))^\top$, and $\boldsymbol{R}$ be the $n \times n$ correlation matrix with the (i,j)-th element $R(\boldsymbol{x}_i - \boldsymbol{x}_j)$.

EI, as a popular choice for the acquisition function (Jones et al. 1998), measures the expected improvement to the target optimization problem at input point $\boldsymbol{x}$. Let $y_n^{\min} = \min_{1 \le i \le n} y(\boldsymbol{x}_i)$ be the current minimum value at step n. Given by Jones et al. (1998), the expected improvement of input point $\boldsymbol{x}$ is defined by:

$$\text{EI}_n(\boldsymbol{x}) = \text{E}\left[\max\left(y_n^{\min} - Y(\boldsymbol{x}), 0\right)\right], \tag{7}$$

where the expectation is taken with respect to the conditional distribution of $Y(\boldsymbol{x})$ given $\boldsymbol{y}_n$ in (4). Then, EI can be expressed by:

$$\text{EI}_n(\boldsymbol{x}) = \left(y_n^{\min} - \hat{y}(\boldsymbol{x})\right) \Phi\left(\frac{y_n^{\min} - \hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}\right) + s(\boldsymbol{x}) \phi\left(\frac{y_n^{\min} - \hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}\right), \tag{8}$$

where $\hat{y}(\boldsymbol{x})$ and $s(\boldsymbol{x})$ are the mean and standard deviation of $Y(\boldsymbol{x}) | \boldsymbol{y}_n$ in (4), and $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function and the probability density function of the standard normal distribution, respectively.

2.2 Bi-objective Optimization

Let $y_1(\boldsymbol{x})$ and $y_2(\boldsymbol{x})$ be two scalar-valued deterministic black-box functions with input $\boldsymbol{x} \in \mathscr{R}^d$. The bi-objective optimization problem is defined by

$$\begin{aligned} \min \quad & \boldsymbol{y}(\boldsymbol{x}) = [y_1(\boldsymbol{x}), y_2(\boldsymbol{x})] \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathscr{X} \end{aligned} \tag{9}$$

where $\mathscr{X} \subset \mathscr{R}^d$ is a feasible set. Euclidean spaces $\mathscr{R}^d$ and $\mathscr{R}^2$ are respectively referred to as the decision space and the objective (outcome) space. Given two input points $\boldsymbol{x}, \boldsymbol{x}' \in \mathscr{X}$, the outcome $\boldsymbol{y}(\boldsymbol{x})$ is said to dominate the outcome $\boldsymbol{y}(\boldsymbol{x}')$, denoted as $\boldsymbol{y}(\boldsymbol{x}) \prec \boldsymbol{y}(\boldsymbol{x}')$, if and only if $y_l(\boldsymbol{x}) \leq y_l(\boldsymbol{x}')$ for $l = 1,2$ and $y_l(\boldsymbol{x}) < y_l(\boldsymbol{x}')$ for $l = 1$ or $l = 2$. If $y_l(\boldsymbol{x}) \leq y_l(\boldsymbol{x}')$ for $l = 1,2$, the outcome $\boldsymbol{y}(\boldsymbol{x})$ is said to weakly dominate the outcome $\boldsymbol{y}(\boldsymbol{x}')$, denoted as $\boldsymbol{y}(\boldsymbol{x}) \preceq \boldsymbol{y}(\boldsymbol{x}')$. An outcome point is said to be nondominated if there does not exist another outcome point dominating it. To solve problem (9), we apply the classical concept of Pareto-optimality, that is to identify the set of all input points whose images are nondominated outcomes (Ehrgott 2005). The Pareto set is defined by

$$\mathscr{P} = \{\boldsymbol{y}(\boldsymbol{x}) \in \mathscr{R}^2 : \boldsymbol{x} \in \mathscr{X} \mid \nexists\, \boldsymbol{x}' \in \mathscr{X} \quad \text{s.t.} \quad \boldsymbol{y}(\boldsymbol{x}') \prec \boldsymbol{y}(\boldsymbol{x})\}. \tag{10}$$

Since the Pareto set is typically not available in a closed form even if the objective and constraint functions are available, various computational approaches and algorithms have been developed to provide approximation of different types (Ruzika and Wiecek 2005; Herzel et al. 2021). Due to the diversity of these methods and the resulting sets being exactly or only approximating the Pareto set, many ways to measure their quality have been proposed (Faulkenberg and Wiecek 2010; Audet et al. 2021). One way to measure the quality of the computed Pareto set is the hypervolume indicator (e.g., Knowles et al. (2003), Guerreiro et al. (2021)). Given a reference point $\boldsymbol{t} = (t_1, t_2) \in \mathscr{R}^2$, the hypervolume indicator $\mathrm{H}(\mathscr{P}, \boldsymbol{t})$ is the two-dimensional Lebesgue measure Λ of the region weakly dominated by $\mathscr{P}$ and bounded above by the reference point $\boldsymbol{t}$, i.e.,

$$\mathrm{H}(\mathscr{P}, \boldsymbol{t}) = \Lambda\left(\{\boldsymbol{v} \in \mathscr{R}^2 \mid \nexists\, \boldsymbol{p} \in \mathscr{P} : \boldsymbol{p} \preceq \boldsymbol{v} \text{ and } \boldsymbol{v} \preceq \boldsymbol{t}\}\right). \tag{11}$$

The hypervolume improvement (HVI) (Hupkens et al. 2015) can be used to assess the effect of evaluating the vector-valued objective function $\boldsymbol{y}(\boldsymbol{x})$ at a new input point $\boldsymbol{x} \in \mathscr{X}$:

$$\mathrm{HVI}\left(\boldsymbol{y}(\boldsymbol{x}), \mathscr{P}, \boldsymbol{t}\right) = \mathrm{H}\left(\mathscr{P} \cup \{\boldsymbol{y}(\boldsymbol{x})\}, \boldsymbol{t}\right) - \mathrm{H}\left(\mathscr{P}, \boldsymbol{t}\right). \tag{12}$$

2.3 Bi-objective Bayesian Optimization with EHI

In bi-objective Bayesian optimization, the expected value of the hypervolume improvement (EHI) in (12) is a commonly used acquisition function (Emmerich et al. 2011). Given that the two objectives $y_1(\boldsymbol{x})$ and $y_2(\boldsymbol{x})$ are modeled by two independent Gaussian processes $Y_1(\boldsymbol{x})$ and $Y_2(\boldsymbol{x})$ as in (2), we provide the expression of EHI.

Assume that we evaluated the two objectives at n input points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, and collected the outputs $\boldsymbol{y}_{n,1} = (y_1(\boldsymbol{x}_1) \ldots, y_1(\boldsymbol{x}_n))^\top$ and $\boldsymbol{y}_{n,2} = (y_2(\boldsymbol{x}_1) \ldots, y_2(\boldsymbol{x}_n))^\top$. Following (4), we have

$$Y_1(\boldsymbol{x}) | \boldsymbol{y}_{n,1} \sim N(\hat{y}_1(\boldsymbol{x}), \hat{s}_1^2(\boldsymbol{x})) \quad \text{and} \quad Y_2(\boldsymbol{x}) | \boldsymbol{y}_{n,2} \sim N(\hat{y}_2(\boldsymbol{x}), \hat{s}_2^2(\boldsymbol{x})), \tag{13}$$

where $\hat{y}_l(\boldsymbol{x})$ and $\hat{s}_l^2(\boldsymbol{x})$ for $l = 1,2$ are the conditional means and variances, which can be calculated using (5) and (6). Under the assumption that $Y_1(\boldsymbol{x})$ and $Y_2(\boldsymbol{x})$ are independent, $Y_1(\boldsymbol{x}) | \boldsymbol{y}_{n,1}$ and $Y_2(\boldsymbol{x}) | \boldsymbol{y}_{n,2}$ are also independent. Note that the two Gaussian processes $Y_1(\boldsymbol{x})$ and $Y_2(\boldsymbol{x})$ can have different parameters and correlation functions. Given a reference point $\boldsymbol{t}$, EHI is defined by

$$\mathrm{EHI}_n\left(\boldsymbol{x}; \mathscr{P}, \boldsymbol{t}\right) = \mathrm{E}\left\{\mathrm{HVI}\left((Y_1(\boldsymbol{x}), Y_2(\boldsymbol{x})), \mathscr{P}, \boldsymbol{t}\right)\right\}, \tag{14}$$

where HVI is given by (12) with the new objective function evaluation $(Y_1(\boldsymbol{x}), Y_2(\boldsymbol{x}))$ and the expectation is taken with respect to the conditional distributions in (13). By maximizing $\text{EHI}_n(\boldsymbol{x}; \mathscr{P}, \boldsymbol{t})$ over $\boldsymbol{x} \in \mathscr{X}$, we find an input point for the new function evaluation. Given an approximation of the Pareto set $\mathscr{P}$ by a set of nondominated points, EHI can be calculated empirically with a closed-form expression (Hupkens et al. 2015). We denote the nondominated points by $\boldsymbol{y}^j = (y_1^j, y_2^j) = (y_1(\boldsymbol{x}_{i_j}), y_2(\boldsymbol{x}_{i_j}))$ for $j = 1, \ldots, p$ with $\{i_1, \ldots, i_p\} \subset \{1, \ldots, n\}$. Without loss of generality, we assume that those nondominated points are ordered based on the values of the first objective, i.e., $y_1(\boldsymbol{x}_{i_1}) < y_1(\boldsymbol{x}_{i_2}) < \ldots < y_1(\boldsymbol{x}_{i_p})$. Then the approximation of the Pareto set $\mathscr{P}$ is denoted by

$$\mathscr{P}^{(n)} = \left\{\boldsymbol{y}^0, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^p, \boldsymbol{y}^{p+1}\right\}, \tag{15}$$

where $\boldsymbol{y}^0 = (-\infty, t_2)$ and $\boldsymbol{y}^{p+1} = (t_1, -\infty)$ with t_1 and t_2 being the two coordinates of the reference point $\boldsymbol{t} = (t_1, t_2)$. Following Emmerich et al. (2011), EHI in (14) can be calculated empirically based on the conditional distributions in (13):

$$\begin{aligned}
\text{EHI}_n\left(\boldsymbol{x}; \mathscr{P}^{(n)}, \boldsymbol{t}\right) &= \text{E}\left\{\text{HVI}\left((Y_1(\boldsymbol{x}), Y_2(\boldsymbol{x})), \mathscr{P}^{(n)}, \boldsymbol{t}\right)\right\} \\
&= \sum_{j=1}^{p+1} \left(y_1^{j-1} - y_1^j\right) \cdot \Phi\left(\frac{y_1^j - \hat{y}_1(\boldsymbol{x})}{\hat{s}_1(\boldsymbol{x})}\right) \cdot \Psi\left(y_2^j, y_2^j, \hat{y}_2(\boldsymbol{x}), \hat{s}_2(\boldsymbol{x})\right) \\
&+ \sum_{j=1}^{p+1} \left(\Psi\left(y_1^{j-1}, y_1^{j-1}, \hat{y}_1(\boldsymbol{x}), \hat{s}_1(\boldsymbol{x})\right) - \Psi\left(y_1^{j-1}, y_1^j, \hat{y}_1(\boldsymbol{x}), \hat{s}_1(\boldsymbol{x})\right)\right) \\
&\cdot \Psi\left(y_2^j, y_2^j, \hat{y}_2(\boldsymbol{x}), \hat{s}_2(\boldsymbol{x})\right),
\end{aligned} \tag{16}$$

where $\Psi(a, b, \mu, \sigma) = \sigma\phi\left(\frac{b-\mu}{\sigma}\right) + (a-\mu)\Phi\left(\frac{b-\mu}{\sigma}\right)$. It is worth noting that, the evaluation of this function is computationally non-trivial as demonstrated by Emmerich et al. (2016) and Yang et al. (2019). Since the key properties (such as concavity or convexity) of this function can not be justified, the number of function evaluations of EHI required to obtain a high-quality solution is not negligible for median to high dimensional decision spaces.

3 PROPOSED METHOD

We propose a bi-objective Bayesian optimization algorithm using transformed additive Gaussian processes (TAG) (Lin and Joseph 2020) as the surrogate model. TAG approximates each of the two objective functions by additive functions, each additive term associated with one-dimensional decision variable. Under this model assumption, maximizing the acquisition function EHI given by (14) can be simplified to maximize the EHI's for each decision variable. In this section, we introduce using TAG as a surrogate model for bi-objective problems, develop the EHI functions under the TAG model assumption, and summarize our algorithm implementation.

3.1 Bi-objective Surrogate with Transformed Additive Gaussian Processes

Transformed additive Gaussian processes aim to simplify a complex response to additive functions. Consider the bi-objective problem in (9). As noted in Section 2.3, we assume that $y_1(\boldsymbol{x})$ and $y_2(\boldsymbol{x})$ are the realizations of two independent transformed additive Gaussian processes $Y_1(\boldsymbol{x})$ and $Y_2(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathscr{X} \subset \mathscr{R}^d$. Equivalently, following Lin and Joseph (2020), we have

$$g_{\lambda_l}(Y_l(\boldsymbol{x})) = \mu_l + \sum_{k=1}^{d} Z_{lk}(x_k) + \varepsilon_l, \quad \text{for} \quad l = 1, 2. \tag{17}$$

We explain the notation in turn as follows. The function $g_{\lambda_l}(\cdot)$ is the Box-Cox transformation (Box and Cox 1964):

$$g_{\lambda_l}(y) = \begin{cases} \frac{y^{\lambda_l}-1}{\lambda_l} & \lambda_l \neq 0 \\ \log y & \lambda_l = 0, \end{cases}, \quad \text{for} \quad l = 1,2, \tag{18}$$

where λ_l is the parameter of the transformation. The deterministic means of TAG are denoted by μ_l's. For $k = 1,\ldots,d$ and $l = 1,2$, $Z_{lk}(x_k)$'s are mutually independent mean-zero Gaussian processes with variances τ_{lk}^2's, and correlation functions $R_{lk}(\cdot)$'s. The additive noise $\varepsilon_l \sim N(0,\sigma_l^2)$ is independent of $Z_{lk}(x_k)$'s.

Under this model assumption, the correlation of $g_{\lambda_l}(Y_l(\boldsymbol{x}))$ and $g_{\lambda_l}(Y_l(\boldsymbol{x}'))$ for $\boldsymbol{x},\boldsymbol{x}' \in \mathscr{X}$ is

$$R_l(\boldsymbol{x}-\boldsymbol{x}') = \sum_{k=1}^{d} \omega_{lk} R_{lk}\left(x_k - x_k'\right), \tag{19}$$

where $\tau_l^2 = \sum_{k=1}^{d} \tau_{lk}^2$ and $\omega_{lk} = \tau_{lk}^2/\tau_l^2$. An example of the correlation function is given by (3), which is associated with some unknown correlation parameters. In Lin and Joseph (2020), empirical Bayes estimation and non-linear optimization toolboxes are used to estimate those unknown parameters, including τ_l, ω_{lk}, λ_l and correlation parameters, which are referred to as hyperparameters of TAG.

Assume that we collected outputs $\boldsymbol{y}_{n,1} = (y_1(\boldsymbol{x}_1)\ldots,y_1(\boldsymbol{x}_n))^\top$ and $\boldsymbol{y}_{n,2} = (y_2(\boldsymbol{x}_1)\ldots,y_2(\boldsymbol{x}_n))^\top$. Let $g_{\lambda_l}(\boldsymbol{y}_{n,l}) = (g_{\lambda_l}(y_l(\boldsymbol{x}_1)),\ldots,g_{\lambda_l}(y_l(\boldsymbol{x}_n)))^\top$ for $l = 1,2$. For $l = 1,2$ and $k = 1,\ldots,d$, we have that

$$\begin{pmatrix} g_{\lambda_l}(\boldsymbol{y}_{n,l}) - \mu_l \mathbf{1} \\ Z_{lk}(x_k) \end{pmatrix} \sim N_{n+1}\left[\begin{pmatrix} \mathbf{0}_n \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_l^2 \boldsymbol{R}_l + \sigma_l^2 \boldsymbol{I} & \tau_{lk}^2 \boldsymbol{r}_{lk}(x_k) \\ \tau_{lk}^2 \boldsymbol{r}_{lk}(x_k)^\top & \tau_{lk}^2, \end{pmatrix}\right], \tag{20}$$

where $\mathbf{0}_n$ is the $n \times 1$ vector of zeros, $\boldsymbol{I}$ is the $n \times n$ identify matrix, $\boldsymbol{r}_{lk}(x_k) = (R_{lk}(x_k - x_{1k}),\ldots,R_{lk}(x_k - x_{nk}))^\top$ and $\boldsymbol{R}_l$ is the $n \times n$ correlation matrix with the (i,j)-th element $R_l(\boldsymbol{x}_i - \boldsymbol{x}_j)$ given by (19). Using the conditional distribution for the normal distribution (e.g., Eaton (2007)), we can derive the conditional distribution of $Z_{lk}(x_k)$ given $\boldsymbol{y}_{n,l}$:

$$Z_{lk}(x_k)|\boldsymbol{y}_{n,l} \sim N\left(\hat{z}_{lk}(x_k), s_{lk}^2(x_k)\right) \quad \text{for} \quad l = 1,2, \tag{21}$$

with

$$\hat{z}_{lk}(x_k) = \omega_{lk} \boldsymbol{r}_{lk}(x_{lk})^\top \left(\boldsymbol{R}_l + \delta_l \boldsymbol{I}\right)^{-1} \left(g_{\lambda_l}(\boldsymbol{y}_{n,l}) - \mu_l \mathbf{1}\right),$$

and

$$s_{lk}^2(x_k) = \tau_l^2 \omega_{lk} \left(1 - \boldsymbol{r}_{lk}(x_{lk})^\top \left(\boldsymbol{R}_l + \delta_l \boldsymbol{I}\right)^{-1} \omega_{lk} \boldsymbol{r}_{lk}(x_{lk})\right),$$

where $\delta_l = \sigma_l^2/\tau_l^2$ and μ_l can be replaced by $\hat{\mu}_l = \mathbf{1}^\top \left(\boldsymbol{R}_l + \delta_l \boldsymbol{I}\right)^{-1} g_{\lambda_l}(\boldsymbol{y}_{n,l}) / \mathbf{1}^\top \left(\boldsymbol{R}_l + \delta_l \boldsymbol{I}\right)^{-1} \mathbf{1}$ as a plug-in estimator. The conditional distribution of $Z_{lk}(x_k)$ will then be used to develop EHIs for each dimension of the decision variable.

3.2 Expected Hypervolume Improvement with Transformed Additive Gaussian Processes

Based on the conditional distributions in (21), we simplify the EHI in (14) to the expected hypervolume improvements for the bi-objective problems for each dimension of the decision variable. Provided by the approximation of TAG, we have

$$g_{\lambda_l}\left(y_l(\boldsymbol{x})\right) \approx \mu_l + \sum_{k=1}^{d} z_{lk}\left(x_k\right) \quad \text{for} \quad l = 1,2, \tag{22}$$

where $z_{lk}(x_k)$ is a realization of $Z_{lk}(x_k)$. Assume that the decision space $\mathscr{X} = \mathscr{X}_1 \times \ldots \times \mathscr{X}_d$. Under this approximation, we obtain a surrogate optimization problem of the the original bi-objective optimization problem in (9)

$$\min_{x_k \in \mathscr{X}_k} \quad \boldsymbol{z}_k(x_k) = \left[z_{1k}\left(x_k\right), z_{2k}\left(x_k\right)\right] \quad \text{for} \quad k = 1,\ldots,d \tag{23}$$

Let $\mathscr{P}_k$ be the Pareto set of the k-th problem in (23), i.e.,

$$\mathscr{P}_k = \{\boldsymbol{z}_k(x_k) \in \mathscr{R}^2 \mid \nexists\, \boldsymbol{z}_k(x'_k) \in \mathscr{R}^2 : \boldsymbol{z}_k(x'_k) \prec \boldsymbol{z}_k(x_k)\} \tag{24}$$

for $k = 1,\ldots,d$. Under TAG, we can define the EHI for the problem in (23) by:

$$\mathrm{EHI}_{n,k}(x_k; \mathscr{P}_k, \boldsymbol{t}_k) = \mathrm{E}\left\{\mathrm{HVI}\left((Z_{1k}(x_k), Z_{2k}(x_k)), \mathscr{P}_k, \boldsymbol{t}_k\right)\right\}, \tag{25}$$

where the expectation is taken with respect to the conditional distribution of $Z_{lk}(x_k)$ in (21). As a result, instead of maximizing $\mathrm{EHI}_n(\boldsymbol{x}; \mathscr{P}, \boldsymbol{t})$ over $\boldsymbol{x} \in \mathscr{X}$, a new input point is given by maximizing $\mathrm{EHI}_{n,k}(x_k; \mathscr{P}_k, \boldsymbol{t}_k)$ with one dimensional decision variable $x_k \in \mathscr{X}_k$ for $k = 1,\ldots,d$. To be specific, the new design point is generated by selecting $\boldsymbol{x}_{new} = \left(x_1^*, \ldots, x_d^*\right)$ with

$$x_k^* \in \max_{x \in \mathscr{X}_k} \mathrm{EHI}_{n,k}(x_k; \mathscr{P}_k, \boldsymbol{t}_k) \quad \text{for} \quad k = 1,\ldots,d.$$

Given an approximation of the Pareto set $\mathscr{P}_k$, EHI can attain a close form expression as in (16) using the conditional distribution in (21). In the next subsection, we describe our detailed implementation of the proposed algorithm.

3.3 The Proposed Algorithm

We detail our implementation of the proposed bi-objective Bayesian optimization algorithm in this subsection. The surrogate model is a TAG model in (17) implemented by the R package `TAG` (Lin and Joseph 2021). The calculation of EHI depends on the approximation of the Pareto set in (15). We use the R package `ecr` (Bossek 2017) to approximate Pareto set. Given responses $\boldsymbol{y}_{m,1}$ and $\boldsymbol{y}_{m,2}$ of size m, the approximations of the Pareto sets of $\mathscr{P}$ in (10) and $\mathscr{P}_k$ in (24) are denoted by $\mathscr{P}^{(m)}$ and $\mathscr{P}_k^{(m)}$, respectively. We summarize our implementation in Algorithm 2. We first generate an initial dataset of size n (line 1 in Algorithm 2), and then add one more data point at each step until we exhaust the budget of function evaluation N (line 2-10 in Algorithm 2). In each step, we fit a TAG model for each objective function. For simplicity, the maximization of $\mathrm{EHI}_{n,k}(x_k; \mathscr{P}_k, \boldsymbol{t}_k)$ is done by evaluating q random candidate points in $\mathscr{X}_k$ (line 4-9 in Algorithm 2). In our implementation, we set $q = 10$ and generate a $q \times d$ random Latin hypercube design using the R package `lhs` (Carnell 2022) with k-th column being the candidate points in $\mathscr{X}_k$ (line 4 in Algorithm 2). At each step, we compute the hypervolume indicator in (11) based on the approximated Pareto set $\mathscr{P}^{(m)}$ (line 10 in Algorithm 2).

The hyperparameters (such as λ_l's and τ_l's etc) in TAG are not updated in every step in our implementation. To save computation time and resources, we will refit those hyperparameters in every ten steps. The size of Latin hypercube design in line 4 of Algorithm 2 can be adjusted according to the computational requirement of the implementation of the algorithm. Also note that, for convenience, we use the enumeration approach to maximize univariate EHI functions in our implementation. In practice, alternative univariate optimization algorithms can be applied here to improve the efficiency and accuracy.

4 NUMERICAL RESULTS

In this section, we compare the proposed algorithm as described in Section 3.3 with the bi-objective Bayesian optimization procedure implemented with the function `GParetoptim` in the R package `GPareto` (Binois and Picheny 2019). This algorithm uses independent Gaussian processes as a surrogate model for the two objectives, and uses EHI in (16) as the acquisition function. To distinguish between the two approaches, our method is denoted by "Transformed Additive GP", and the method in `GPareto` is referred to as "Classical GP". The purpose of the numerical comparison is to explore the situations that the proposed algorithm shows competitive performances or advantages by incorporating the additive assumption. Therefore, "Classical GP" with original EHI is the key benchmark without the additive assumption.

Algorithm 2 Bi-objective Bayesian Optimization with TAG

Given an initial dataset with inputs $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ and the outputs $\boldsymbol{y}_{n,1}$ and $\boldsymbol{y}_{n,2}$, compute an initial Pareto set approximation $\mathscr{P}^{(n)}$ and compute $\mathrm{H}(\mathscr{P}^{(n)},\boldsymbol{t})$.
for $m = n, n+1,\ldots,N$ **do**
  Fit the TAG models based on $\boldsymbol{y}_{m,1}$ and $\boldsymbol{y}_{m,2}$ and $\boldsymbol{x}$, find $\mathscr{P}_k^{(m)}$ for $k = 1,\ldots,d$.
  Generate a $q \times d$ matrix D as a random Latin Hypercube design with ij-th entries D_{ij}.
  for $k = 1,\ldots,d$ **do**
    Evaluate $\mathrm{EHI}_{n,k}\left(D_{ik}; \mathscr{P}_k^{(m)}, \boldsymbol{t}_k\right)$ for $i = 1,\ldots,q$.
    Select point D_k^* that maximizes $\mathrm{EHI}_{n,k}\left(D_{ik}; \mathscr{P}_k^{(m)}, \boldsymbol{t}_k\right)$ over $i = 1,\ldots,q$.
  end for
  Set new input point $\boldsymbol{x}^{(m+1)} = \left(D_1^*,\ldots,D_d^*\right)$ and obtain $\boldsymbol{y}(\boldsymbol{x}^{(m+1)})$.
  By including the new response, update $\boldsymbol{y}_{m,1} \to \boldsymbol{y}_{m+1,1}$, $\boldsymbol{y}_{m,2} \to \boldsymbol{y}_{m+1,2}$, and $\mathscr{P}^{(m)} \to \mathscr{P}^{(m+1)}$. Compute $\mathrm{H}(\mathscr{P}^{(m+1)},\boldsymbol{t})$.
end for
Return $\mathrm{H}(\mathscr{P}^{(m)},\boldsymbol{t})$ for $m = n, n+1,\ldots,N$.

First, we generate initial data using the R package `lhs` (Carnell 2022). Next, we fit an initial surrogate model to the data. For our method, the surrogate model is a TAG model as in (17). The classical GP approach will use the Gaussian processes as the surrogate models using the R package (Roustant et al. 2012) `DiceKriging`. The reference point $\boldsymbol{t}$ is chosen as a point that bounds the outcome space of the bi-objective problem when $0 \le x_k \le 1$ for $k = 1,\ldots,d$. In our numerical examples, we careful select a common reference point $\boldsymbol{t}$ for all $\boldsymbol{t}_k$'s. We sequentially add new points one by one using EHI for 20 steps. In each step, we compute $\mathrm{H}(\mathscr{P},\boldsymbol{t})$ in (11) based on approximations of $\mathscr{P}$ following (Fonseca et al. 2006).

We replicate the whole procedure 30 times (with 30 different random initial datasets) for both methods. For the returned hypervolume values of each method in each step, we compute their average and use 10^{th} and 90^{th} quantiles to construct a confidence band. For our method, in the R package `TAG`, the default range of λ_l in (18) is -2 to 2, we use this default range for the examples unless otherwise noted.

4.1 Example 1

We construct bi-objective functions as follows:

$$y_l(\boldsymbol{x}) = \exp\left(\Sigma_{k=1}^{d}\left(x_k - a_l\right)^2 + c_l x_1 x_2\right) \quad \text{for} \quad l = 1,2, \tag{26}$$

which can be transformed to additive functions using the Box-Cox transformation (18) for $\lambda_l = 0$ and $c_l = 0$.

We first study the impact of the dimension of the decision space on the proposed method as compared to the classical GP approach. We consider the performance of the proposed algorithm under different sizes of d. We specify three cases with parameters of (26) in Table 1, where the dimension of input spaces changes from 4 to 8. To ensure there are a sufficient number of initial points for $d = 8$, we use 25 initial data points in this study. The results are depicted in Figure 1. As the dimension of the decision space increases, the advantage of the proposed method is greater in temrs of both mean and variance of the approximated hypervolume indicators.

Next, we study the performance of the proposed method when the objective functions have an interaction term and are not strictly transformed to additive functions with one-dimensional input. We consider the performance of the proposed algorithm under different values of c_l. We specify three cases with parameters of (26) in Table 2, where the value of c_2 changes from 0 to 0.5. For this example, 15 initial data points are used. The results are depicted in Figure 2. The results show that when the interaction term is negligible as

Table 1: Table of parameter settings of (26) for the three cases in Figure 1.

Cases	(1)			(2)			(3)		
l	d	a_l	c_l	d	a_l	c_l	d	a_l	c_l
1	4	0.3	0	6	0.3	0	8	0.3	0
2	4	0.4	0	6	0.4	0	8	0.4	0

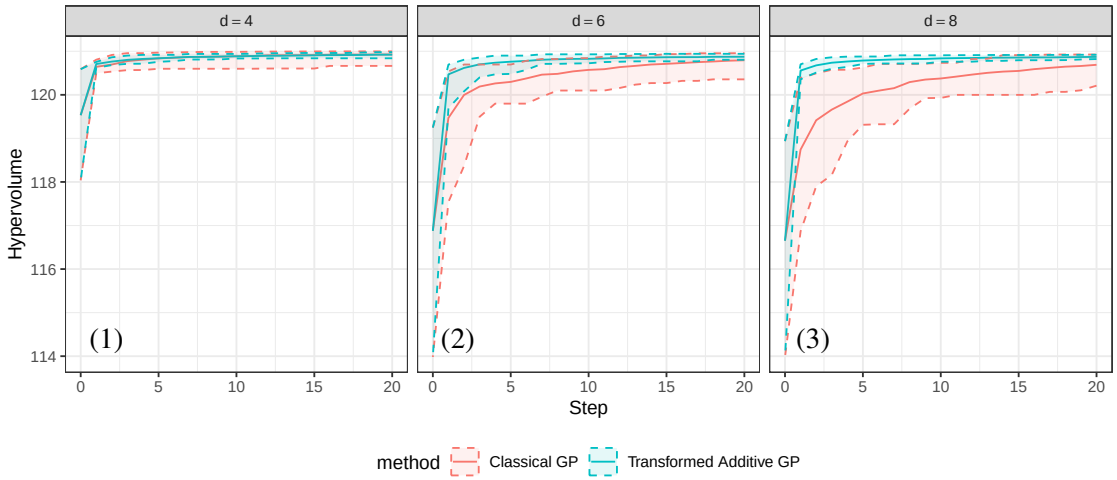

Figure 1: The mean, 10-th and 90-th quantiles of approximated hypervolume over 30 replications and 20 steps for optimizing the objective functions in (26) with the parameter settings in Table 1.

$c_2 = 0$ or 0.01, our method is more robust than Classical GP in early stages. However, when the interaction term increases to $c_2 = 0.5$, classical GP slightly outperforms the proposed algorithm in terms of robustness.

Table 2: Table of parameter settings of (26) for the three cases in Figure 2

Cases	(1)			(2)			(3)		
l	d	a_l	c_1	d	a_l	c_1	d	a_l	c_1
1	4	0.3	0	4	0.3	0	4	0.3	0
2	4	0.4	0	4	0.4	0.01	4	0.4	0.5

4.2 Example 2

We use the example FES1 from the literature of bi-objective optimization (Fieldsend et al. 2003) to demonstrate the performance of the proposed method. The original problem of FES1 is given below:

$$\text{FES1} = \begin{cases} y_1(\boldsymbol{x}) = \sum_{i=1}^{4} \left| x_i - \frac{\exp\{\left(\frac{i}{4}\right)^2\}}{3} \right|^{0.5}, \\ y_2(\boldsymbol{x}) = \sum_{i=1}^{4} \left(x_i - 0.5\cos\left(\frac{10\pi i}{4}\right) - 0.5 \right)^2. \end{cases} \tag{27}$$

For a comparison purpose, we construct a modified version of FES1 by taking the exponential values of the original objectives:

$$\text{FES1-modified} = \begin{cases} y_1(\boldsymbol{x}) = \exp\{\sum_{i=1}^{4} \left| x_i - \frac{\exp\{\left(\frac{i}{4}\right)^2\}}{3} \right|^{0.5}\} \\ y_2(\boldsymbol{x}) = \exp\{\sum_{i=1}^{4} \left(x_i - 0.5\cos\left(\frac{10\pi i}{4}\right) - 0.5 \right)^2\} \end{cases} \tag{28}$$

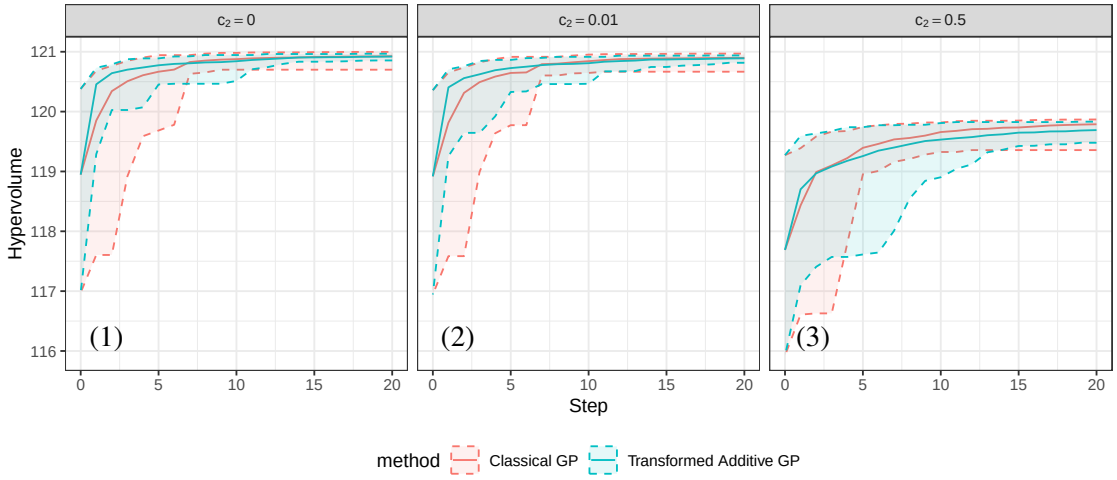

Figure 2: The mean, 10-th and 90-th quantiles of approximated hypervolume over 30 replications and 20 steps for optimizing the objective functions in (26) with the parameter settings in Table 2.

FES1 in (27) is an additive function, whereas FES1-modified in (28) can be transformed into an additive function. For FES1, we restrict λ_l for $l = 1, 2$ in TAG to be between 0.5 and 1.5 since the existing procedure does not estimate it properly. For this example, 20 initial data points are used. The results are depicted in Figure 3. Classical GP is more robust than our method when the function is originally additive without transformation. However, when the additive objective functions are not originally additive (for FES1-modified), our method outperforms the classical GP method in the early stages.

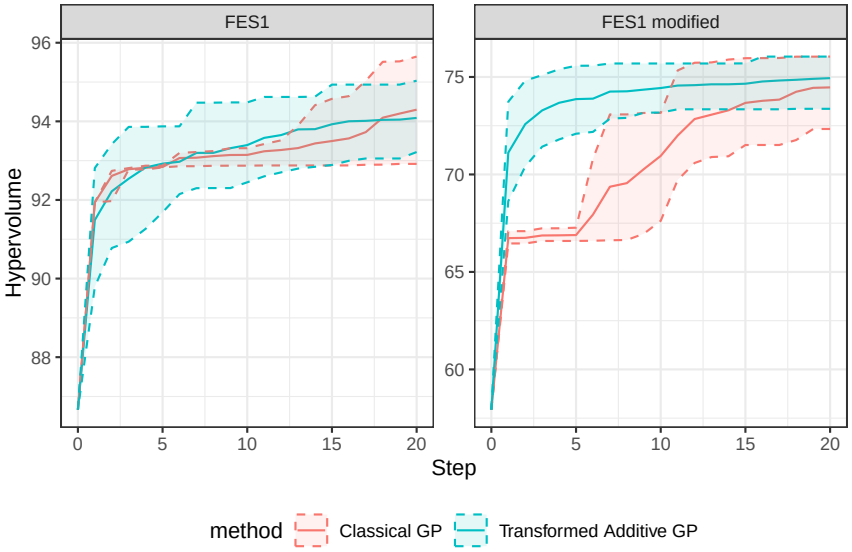

Figure 3: The mean, 10-th and 90-th quantiles of approximated hypervolume over 30 replications and 20 steps for the two problems in (27) and (28).

5 CONCLUSION

We have proposed a method for bi-objective Bayesian optimization using TAG as the surrogate model to simplify the optimization of EHI by decomposing the decision space into one-dimensional additive subspaces. Through a numerical study, we compare the performances of the proposed method with a classical bi-objective Bayesian optimization algorithm under different situations. The proposed algorithm can be further enhanced by improving the parameter tuning under the TAG framework, utilizing better algorithms in maximizing univariate EHI functions and integrating other existing approaches into the proposed algorithm to optimize EHI. Also, motivated by the promising computational results, it can be beneficial to develop theoretical justification of the proposed method. We aim to overcome the limitation of the current work in future and extend the method to the multi-objective case.

ACKNOWLEGMENTS

This work was supported by Clemson University's Virtual Prototyping of Autonomy Enabled Ground Systems (VIPR-GS), under Cooperative Agreement W56HZV-21-2-0001 with the US Army DEVCOM Ground Vehicle Systems Center (GVSC).

DISCLAIMERS

All opinions, conclusions and findings wherein are those of the authors and may not be those of the affiliated institutions. DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited. OPSEC # 8423.

REFERENCES

Audet, C., J. Bigeon, D. Cartier, and S. L. Digabel. 2021. "Performance Indicators in Multiobjective Optimization". *European Journal of Operational Research* 292:397–422.

Binois, M. and V. Picheny. 2019. "GPareto: An R Package for Gaussian-Process-Based Multi-Objective Optimization and Analysis". *Journal of Statistical Software* 89(8):1–30.

Bossek, J. 2017. "ecr 2.0: A Modular Framework for Evolutionary Computation in R". In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '17, 1187–1193.

Box, G. E. P. and D. R. Cox. 1964. "An Analysis of Transformations". *Journal of the Royal Statistical Society. Series B (Methodological)* 26(2):211–252.

Carnell, R. 2022. *lhs: Latin Hypercube Samples*. R Package Version 1.1.6.

Daulton, S., M. Balandat, and E. Bakshy. 2020. "Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization". *Advances in Neural Information Processing Systems* 33:9851–9864.

de Castro, P., H. Stewart, C. Turner, M. Wiecek, G. Hartman, D. Rizzo, , *et al*. 2022. "Decomposition and Coordination to Support Tradespace Analysis for Ground Vehicle Systems". Technical report, SAE Technical Paper.

Eaton, M. L. 2007. "Multivariate Statistics: A Vector Space Approach". *Lecture Notes-Monograph Series* 53:i–512.

Ehrgott, M. 2005. *Multicriteria Optimization*, Volume 491. Springer Science & Business Media.

Emmerich, M., K. Yang, A. Deutz, H. Wang and C. M. Fonseca. 2016. "A multicriteria generalization of Bayesian global optimization". *Advances in stochastic and deterministic global optimization*:229–242.

Emmerich, M. T., A. H. Deutz, and J. W. Klinkenberg. 2011. "Hypervolume-Based Expected Improvement: Monotonicity Properties and Exact Computation". In *2011 IEEE Congress of Evolutionary Computation (CEC)*, 2147–2154. IEEE.

Faulkenberg, S. L. and M. M. Wiecek. 2010. "On the Quality of Discrete Representations in Multiple Objective Programming". *Optimization and Engineering* 11:423–440.

Fieldsend, J. E., R. M. Everson, and S. Singh. 2003. "Using Unconstrained Elite Archives for Multiobjective Optimization". *IEEE Transactions on Evolutionary Computation* 7(3):305–323.

Fliege, J. and H. Xu. 2011. "Stochastic Multiobjective Optimization: Sample Average Approximation and Applications". *Journal of Optimization Theory and Applications* 151:135–162.

Fonseca, C. M., L. Paquete, and M. López-Ibánez. 2006. "An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator". In *2006 IEEE International Conference on Evolutionary Computation*, 1157–1163. IEEE.

Frazier, P. I. 2018. "A Tutorial on Bayesian Optimization". *arXiv preprint arXiv:1807.02811*.

Guerreiro, A. P., C. M. Fonseca, and L. Paquete. 2021. "The Hypervolume Indicator: Computational Problems and Algorithms". *ACM Computing Surveys (CSUR)* 54(6):1–42.

Herzel, A., S. Ruzika, and C. Thielen. 2021. "Appoximation Methods for Multiobjective Optimization Problems: A Survey". *INFORMS Journal on Computing* 33(4):1284–1299.

Hunter, S. R., E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara *et al*. 2019. "An Introduction to Multiobjective Simulation Optimization". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 29(1):1–36.

Hunter, S. R. and B. McClosky. 2016. "Maximizing Quantitative Traits in the Mating Design Problem via Simulation-Based Pareto Estimation". *IIE Transactions* 48(6):565–578.

Hupkens, I., A. H. Deutz, K. Yang, and M. T. Emmerich. 2015. "Faster Exact Algorithms for Computing Expected Hypervolume Improvement.". In *EMO (2)*, 65–79.

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

Knowles, J. D., D. W. Corne, and M. Fleischer. 2003. "Bounded Archiving Using the Lebesgue Measure". In *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*, Volume 4, 2490–2497. IEEE.

Lee, L. H., C. U. Lee, and Y. P. Tan. 2007. "A Multi-Objective Genetic Algorithm for Robust Flight Scheduling Using Simulation". *European Journal of Operational Research* 177(3):1948–1968.

Lin, L.-H. and V. R. Joseph. 2020. "Transformation and Additivity in Gaussian Processes". *Technometrics* 62(4):525–535.

Lin, L.-H. and V. R. Joseph. 2021. *TAG: Transformed Additive Gaussian Processes*. R Package Version 0.5.1.

Qin, C., D. Klabjan, and D. Russo. 2017. "Improving the Expected Improvement Algorithm". *Advances in Neural Information Processing Systems* 30.

Roustant, O., D. Ginsbourger, and Y. Deville. 2012. "DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization". *Journal of Statistical Software* 51(1):1–55.

Ruzika, S. and M. M. Wiecek. 2005. "Approximation Methods in Multiobjective Programming". *Journal of Optimization Theory and Applications* 126(3):473–501.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn. 1989. "Design and Analysis of Computer Experiments". *Statistical Science* 4(4):409 – 423.

Yang, K., A. Deutz, Z. Yang, T. Back and M. Emmerich. 2016. "Truncated Expected Hypervolume Improvement: Exact Computation and Application". In *2016 IEEE Congress on Evolutionary Computation (CEC)*, 4350–4357. IEEE.

Yang, K., M. Emmerich, A. Deutz, and T. Bäck. 2019. "Efficient computation of expected hypervolume improvement using box decomposition algorithms". *Journal of Global Optimization* 75:3–34.

Yang, K., M. Emmerich, A. Deutz, and T. Bäck. 2019. "Multi-Objective Bayesian Global Optimization Using Expected Hypervolume Improvement Gradient". *Swarm and Evolutionary Computation* 44:945–956.

Zhang, H. 2008. "Multi-Objective Simulation-Optimization for Earthmoving Operations". *Automation in Construction* 18(1):79–86.

AUTHOR BIOGRAPHIES

CAROLINE M. KERFONTA obtained her Ph.D in Mathematical Sciences at Clemson University in May 2024. Her research interests include Bayesian Optimization, experimental design, and statistical applications for manufacturing and engineering, particularly aerospace. Her email address is ckerfon@clemson.edu.

QIONG ZHANG is an Associate Professor of Statistics at Clemson University. Her research interests include the interface between information collection and statistical modeling, experimental design, modeling, and uncertainty quantification for computer experiments, statistical design for web experiments, and data analytics for engineering applications. Her email address is qiongz@clemson.edu.

MARGARET M. WIECEK is Professor of Mathematical Sciences in the School of Mathematical and Statistical Sciences and has held joint professorship with the Department of Mechanical Engineering at Clemson University. Her research area includes theory, methodology, and applications of mathematical programming with special interest in multiobjective optimization and decision-making. Part of her work is interdisciplinary since she has introduced new multiobjective optimization concepts and methods into engineering optimization to enrich the field of automotive and structural design. She has served as the President of the MCDM Section of INFORMS. In recognition of her lifetime contributions, she has been awarded the MCDM Gold Medal by the International Society on MCDM. Her email address is wmalgor@clemson.edu.