Multi-fidelity Bayesian Optimization in Engineering Designs

Bach Do * and Ruda Zhang †
University of Houston, Houston, TX 77204-4003, USA

Resided at the intersection of multi-fidelity optimization (MFO) and Bayesian optimization (BO), MF BO has found a niche in solving expensive engineering design optimization problems, thanks to its advantages in incorporating physical and mathematical understandings of the problems, saving resources, addressing exploitation-exploration trade-off, considering uncertainty, and processing parallel computing. The increasing number of works dedicated to MF BO suggests the need for a comprehensive review of this advanced optimization technique. In this paper, we survey recent developments of two essential ingredients of MF BO: Gaussian process (GP) based MF surrogates and acquisition functions. We first categorize the existing MF modeling methods and MFO strategies to locate MF BO in a large family of surrogate-based optimization and MFO algorithms. We then exploit the common properties shared between the methods from each ingredient of MF BO to describe important GP-based MF surrogate models and review various acquisition functions. By doing so, we expect to provide a structured understanding of MF BO. Finally, we attempt to reveal important aspects that require further research for applications of MF BO in solving intricate yet important design optimization problems, including constrained optimization, high-dimensional optimization, optimization under uncertainty, and multi-objective optimization.

I. Introduction

ENGINEERING design is an iterative process to develop a product or a system for a specific purpose. At the beginning of the conventional design process, engineers determine product specifications via a set of performance metrics. The iterative design process starts with an initial design, followed by comprehensive analyses to evaluate the associated performance metrics. This allows engineers to assess whether the current design satisfies the established specifications. If any specification is deemed invalid or in need of refinement, engineers make necessary changes to the design based on their intuition and experience, and all information gathered from trial designs [1].

Design optimization replaces the conventional design process to accelerate the design cycle and obtain better products [2, 3]. Since it necessitates a formal mathematical formulation of the optimization problem, design optimization encapsulates the performance metrics in an objective to be achieved, design variables to be changed, and constraints to

*Postdoctoral fellow, Department of Civil and Environmental Engineering, bdo3@uh.edu.
†Assistant Professor, Department of Civil and Environmental Engineering, rudaz@uh.edu (Corresponding author).

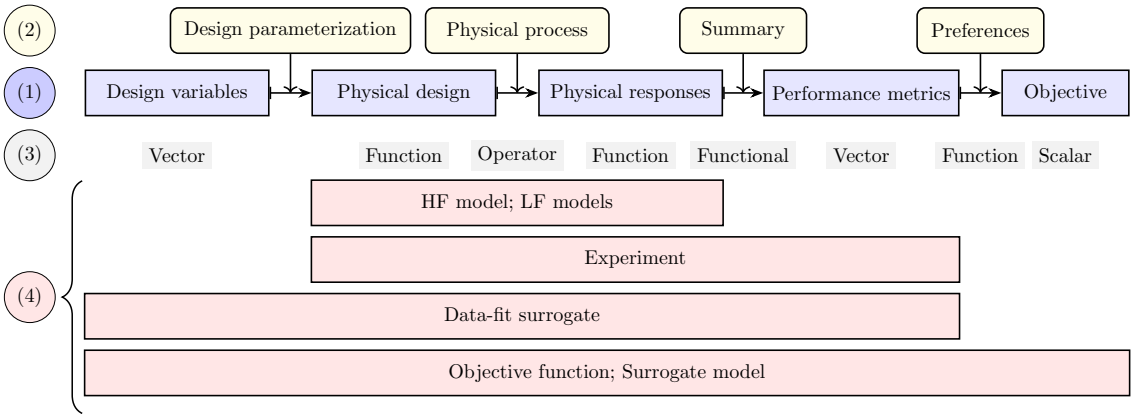

Fig. 1 Mathematical modeling for engineering design optimization: (1) fundamental objects, (2) relations, (3) mathematical abstractions, and (4) other concepts in engineering design optimization.

be met. Once the problem is formulated, changes in designs are automatically made via an optimization algorithm that is systematic and requires less intervention from engineers.

We distinguish the following two concepts in optimization: mathematical optimization and numerical optimization. *Mathematical optimization* is the problem of finding a set of points where a real-valued function attains its minimum (or maximum) value. This function is called the objective function, and its domain may be prescribed by a set of equality or inequality constraints, each determined by a separate function. *Numerical optimization* is the use of numerical algorithms to solve optimization problems. Common optimization algorithms can be broadly split into two groups, depending on whether they require derivatives of the objective function or not: derivative-based methods such as line search and trust region methods, and derivative-free methods such as population-based and direct search methods.

Consider the following engineering design optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}} \ f(\mathbf{x}) \\ &\text{s.t.} \ \ \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^d$ is the vector of d design (input) variables selected in some bounded domain $\mathcal{X}$, and $f(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}$ is the objective (output) function.

Solving problem (1) via numerical optimization requires two critical elements: (i) the mathematical modeling of the problem and (ii) an optimization algorithm. The first aspect arises because objective functions in engineering design are often regarded as black boxes that are only queried at specified points of design variables for their values or derivatives. Selecting optimization algorithms in the second aspect depends on how the objective can be evaluated.

The mathematical modeling of problem (1) entails the utilization of a mathematical model derived from our decent understanding of governing equations of the physical process underlying the relationship between the design objective

and design variables; see Fig. 1. The expression of this mathematical model is the objective function $f(\mathbf{x})$. Unfortunately, solving the governing equations in engineering design is often analytically intractable. Thus, *experiments* are considered as the ground truth to measure certain aspects of the physical process. However, they are subject to uncertainties in design manufacturing, operational control, and measurement errors, which must be considered in experimental data analysis. Oftentimes the mathematical model is implemented via its computational counterparts, such as finite element (FE) models [4], which simulate the physical process under hypothetical scenarios.

Depending on their prediction qualities and costs, computational models for the physical process can be classified into high- and low-fidelity models; see Fig. 1. A *high-fidelity model* (HF model) is a computational model that predicts the performance metrics with an accuracy level sufficient for a basic understanding of the underlying physical process and evaluation of the objective function. We denote $f_{\mathrm{H}}(\mathbf{x})$ as the objective function value obtained from experiments or an HF model. We consider both data from experiments and evaluations of an HF model as HF data. In comparison, a *low-fidelity model* (LF model) is a computational model that is designated for applications where a lower level of accuracy is acceptable for evaluating the objective function. The objective function and data from predictions of an LF model are $f_{\mathrm{L}}(\mathbf{x})$ and LF data, respectively. The advantage of the LF model is that it can be much more computationally efficient than its HF counterpart, both in space and time costs. Examples of LF models include simplified-physics approximations and projection-based reduced order models (ROMs), as described in Peherstorfer et al. [5].

If the objective function is expensive to evaluate, the optimization process may fail to find a good design given a limited budget on resources. This may be attributed to, for example, the collection of HF data. In this case, we may consider using a *surrogate model* (i.e., metamodel or model of model), that is, a computationally efficient approximation of the objective function. In the context of engineering design and throughout this paper, we distinguish surrogate models from LF models. A surrogate model approximates the objective function, while an LF model approximates an HF model of the physical process. The data-fit models as defined in Peherstorfer et al. [5], when applied to numerical optimization, usually approximate mappings with low-dimensional input and output such as the objective function; therefore they are surrogates rather than LF models.

Once the mathematical modeling of the objective function is settled, we can consider the selection of an optimization algorithm. Ideally, a good optimization algorithm possesses three properties: robustness, accuracy, and efficiency [6]. However, these goals may conflict and trade-offs between them should be addressed. A prudent approach is to select optimization algorithms based on how the objective function is accessed; see Fig. 2.

Due to their high accuracy level, HF data are expected to be used in the design optimization process. However, obtaining HF data discourages the direct application of any optimization algorithms because it requires conducting time-consuming experiments or running HF computational models, thereby reducing efficiency. For example, one crash simulation reported by Ford Motor Company required 36-160h to complete [7]. In another example, achieving high-precision results in a structural analysis required 23 days to capture the elastoplastic dynamic responses of a

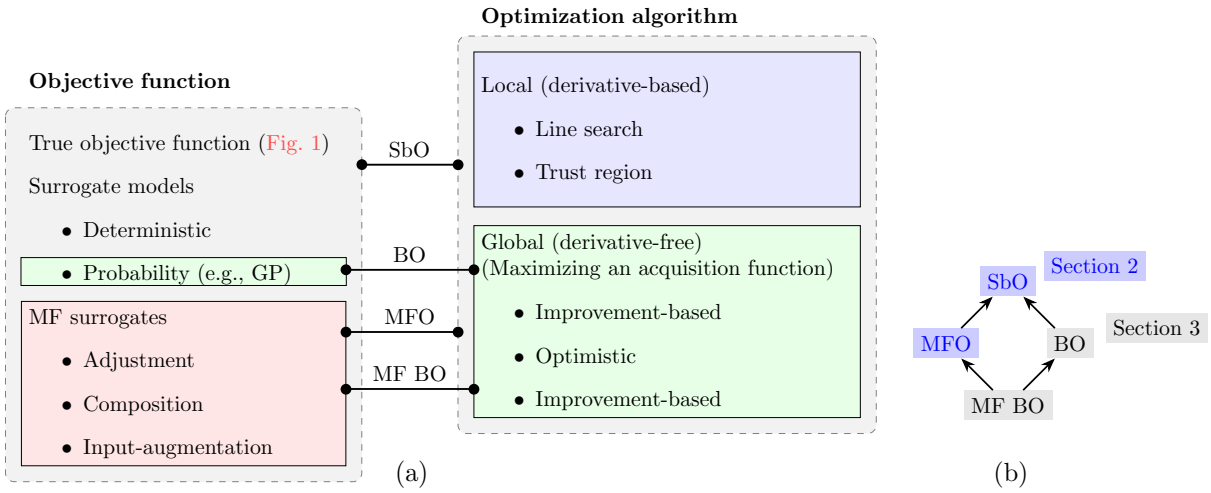

Fig. 2 (a) Distinction between SbO, MFO, BO, and MF BO; (b) Hierarchy of these optimization concepts.

5-story steel frame under a ground motion of 5s [8]. As a successful approach to addressing the efficiency-accuracy trade-off, *surrogate-based optimization* (SbO) approximates the objective function $f(\mathbf{x})$ by a surrogate model [7, 9–12]. The surrogate can be a data-fit model constructed by fitting the parameters of an interpolation, regression, or machine learning model, given pairs of design variables and performance metrics or the overall score derived from HF data, which favors accuracy. Alternatively, it can be induced from an LF model used as an approximation of the physical process in Fig. 1, which, however, favors efficiency. We can classify surrogates into deterministic surrogates, such as response surfaces and most neural networks, and probabilistic surrogates, such as Gaussian process (GP) [13]. We can also classify surrogates into local and global surrogates. A local surrogate is to approximate a black-box objective function in a restricted region of the design variable space, e.g., the neighborhood of a design variable vector. A global surrogate approximates all predictions of that black-box objective function.

To further balance the accuracy-efficiency trade-off in SbO, we can use *multi-fidelity optimization* (MFO) that leverages both LF models and HF data to construct an MF surrogate for the objective function [14–17]. From this definition, it is clear that MFO is a type of SbO. The MF surrogate can be constructed, for example, by adjusting an LF model-induced surrogate to HF data via adjustment methods [18–20]. It can also be constructed by composing an LF model-induced surrogate with an input-input mapping or an output-output mapping [21–23].

Bayesian optimization (BO) is a global optimization technique that assumes a prior probabilistic model on the objective function, and combines it with the available data to guide the optimization process. BO is well suited for small- to moderate-dimensional optimization problems that involve expensive-to-evaluate objective functions [24–26]. Most of the time, a GP is used as the prior model in BO, due to its tractability and flexibility. Typically, the GP prior is specified simply by a constant mean function and a covariance function with a closed form and a few hyperparameters. Given the dataset at each optimization iteration, the posterior model (i.e., the prior model conditioned on the dataset) determines

an acquisition function reflecting an optimization policy under uncertainty, which is then numerically maximized to find the next design point. A derivative-free, global optimization algorithm is usually used for this purpose. The GP posterior at each iteration serves as a surrogate for the objective function, and therefore BO is a type of SbO. BO tends to favor accuracy if it constructs the GP posterior based on HF data. The details of BO and its applications are deferred until Section III.B.

Multi-fidelity Bayesian optimization (MF BO) is the intersection of MFO and BO. More specifically, MF BO includes LF models within a prior probabilistic model for the objective function, and combines this prior with HF data to guide the optimization process.

MF BO is applicable whenever (i) the objective function is expensive to evaluate and (ii) LF models are available or can be constructed. In fact, we advocate that MF BO should be applied whenever these two conditions are true because it can use both data and engineering knowledge efficiently. BO is a well-recognized approach to optimizing expensive objective functions that makes efficient use of the available HF data and therefore reduces the number of HF data points to ensure solution accuracy. MFO opens up the optimization black box in engineering design, and not only exploits the structure of mathematical modeling details but also uses our knowledge about the underlying physical processes. Combining BO and MFO further accelerates engineering design optimization processes.

In recent years, there has been a remarkable increase in the number of successful applications of MF BO across diverse domains of engineering design [see e.g., 23, 27–30]. Additionally, the evolving landscapes of MFO and especially BO, with their continuous advancements in addressing intricate yet important design optimization problems [26, 31], hold a promise for further enhancing the capabilities of MF BO. Thus, there is a need for a comprehensive review of MF BO to meet the demand for future development progress and widespread applications of this powerful optimization method.

Our goal in this review is twofold.

- First, we comprehensively review the existing techniques from two essential ingredients of MF BO: the GP-based MF surrogates and acquisition functions. To accomplish this, we exploit the common properties shared between the techniques from each ingredient. With this approach, we expect to provide a structured understanding of MF BO.
- Second, we provide critical research topics on solving intricate yet important design optimization problems, aimed at expanding the horizon of the existing MF BO algorithms.

We organize the rest of this paper as follows.

- Section II surveys two basic elements of MFO: MF modeling methods and MFO strategies.
- Section III provides an overview of MF BO.
- Section IV describes important GP-based MF surrogates.
- Section V surveys various acquisition functions of BO and how we can modify them for use of MF BO.

- Section VI reviews the extensions of BO and MFO to address intricate yet important design optimization problems, including constrained optimization, high-dimensional optimization, optimization under uncertainty, and multi-objective optimization.
- Section VII summarizes and concludes this paper.

II. Multi-fidelity optimization

This section starts with a brief overview of SbO (Section II.A). It then surveys two elements of MFO: MF surrogates (Section II.B) and MFO strategies (Section II.C).

A. Surrogate-based optimization

The development of SbO in engineering design optimization has received a substantial boost since the work *Design and Analysis of Computer Experiments* (DACE) by Sacks et al. [32]. In DACE, a cost-effective surrogate model was used to estimate the output of an expensive computer code. The best linear unbiased predictor was obtained by minimizing the mean squared error (MSE) of the predictor. DACE also introduced several classical adaptive design criteria to the experimental design process, such as the integrated MSE, maximum MSE, and expected posterior entropy.

In a similar manner, since engineering design optimization often involves running expensive computational models to predict the performance metrics constituting the design objective, the idea is to use a surrogate for the objective function or the problem itself. This leads to applications of SbO in engineering design optimization. Depending on the characteristics of surrogates, SbO can be developed based on either local or global surrogates.

The use of local surrogates is a classical approach of SbO. This approach is classified into local function and local problem approximation methods [33]. The former guides the optimization process using local surrogates for the objective function at current design points. Taylor series approximations often serve as the local surrogates. The latter replaces the original optimization problem with a sequence of easier-to-solve subproblems, or attempts to reduce the number of constraints or the number of design variables.

The use of global surrogates, such as response surfaces and most neural networks, to facilitate the optimization process dates back to the 1980s [33]. At that time, there was a limited number of applications of this approach in engineering design optimization. However, the early 1990s witnessed many applications of response surface methods, first introduced by [34], to single-discipline and multi-discipline design optimization problems [35]. The community also soon recognized a major limitation of response surface methods, i.e., their applicability to small-dimensional problems due to the curse of dimensionality. Consequently, attention began to shift away from response surfaces toward alternative approximation methods such as radial basis functions [36], support vector regression [37], Kriging [38, 39], reduced order models [40], and ensembles [41, 42].

There exist invaluable surveys that provide comprehensive insights into the use of SbO methods in engineering design

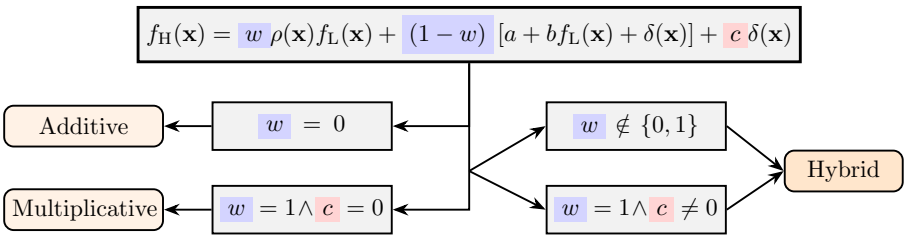

Fig. 3 Additive, multiplicative, and hybrid adjustment MF surrogates.

optimization. Queipo et al. [9] discussed the fundamental issues arising in surrogate-based analysis and optimization, including the design of experiments, construction of surrogate models, model selection, and sensitivity analysis. Wang and Shan [7], from a practitioner's perspective, provided an overview of how global surrogate modeling (i.e., metamodeling) techniques can support engineering design optimization. Simpson et al. [11] conducted a detailed review of the development of approximation methods in multi-discipline design optimization from 1980 to 2008. Forrester and Keane [12] surveyed several state-of-the-art surrogate models and their applications in optimization. Viana et al. [17] reviewed the progression of metamodeling techniques in multi-discipline design optimization.

The use of MFO in engineering design was initiated by Haftka [43] with a multiplicative MF model (see Section II.B). The objective is to extend the application range of HF responses by adjusting the associated LF responses. During the optimization process, the multiplicative MF model is imposed *consistency conditions*, which require that the HF response and its derivatives at a given design point must align with those of the corresponding scaled LF response [44]. If the HF derivatives are not available, we can use the consistency in responses only [45].

Early attempts of MFO relied on local approximations with deterministic MF surrogates valid in the neighborhood of the current point of design variables. Applications of local MFO methods were found in supersonic transport design [46], aerodynamic wing design [44], and airfoil design [47]. Since the seminal works by Huang et al. [15] and Forrester et al. [16], the attention of MFO has been shifted to global approximations with probabilistic MF surrogates valid for a large region of the design variable space. As we will see in Section II.C, a majority of recent applications of MFO in engineering design have been from global methods.

B. Multi-fidelity surrogates

There are typically three classes of MF modeling methods: adjustment, composition, and input augmentation.

Adjustment. The class of adjustment MF methods constructs MF surrogates that use adjustment factors to update LF models for explaining HF data. Most adjustment MF surrogates to approximate $f_{\mathrm{H}}(\mathbf{x})$ take the following form:

$$f_{\mathrm{H}}(\mathbf{x}) = w\rho(\mathbf{x}) f_{\mathrm{L}}(\mathbf{x}) + (1-w)\,[a + b f_{\mathrm{L}}(\mathbf{x}) + \delta(\mathbf{x})] + c\delta(\mathbf{x}). \tag{2}$$

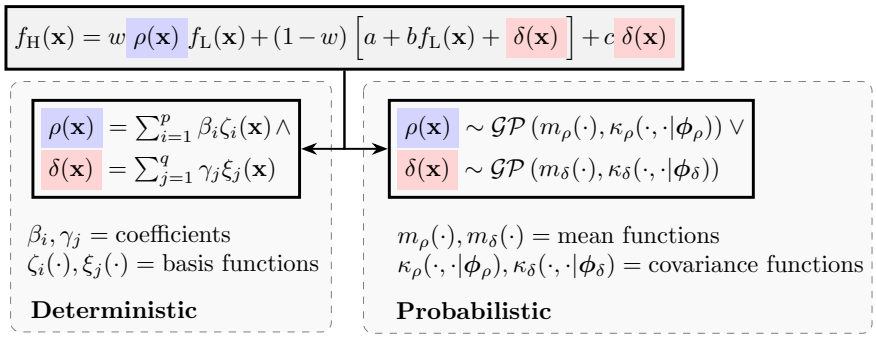

Fig. 4 Deterministic versus probabilistic adjustment MF surrogates.

Here we abuse the notations $f_{\mathrm{H}}(\mathbf{x})$ and $f_{\mathrm{L}}(\mathbf{x})$ to denote the surrogates that approximate the objective function predictions of HF and LF models, respectively. w is a weighting factor, which can be adjusted during the optimization process [47]. $\rho(\mathbf{x})$ is an adjustment coefficient function indicating that the adjusted LF response can vary in different regions of the design variable space [43]. a, b, and c are three unknown constants. $\delta(\mathbf{x})$ is a discrepancy function [18].

Note that the form in Eq. (2) does not include cases when multiple LF models are available. However, we can use it to recover almost all MF surrogates derived from multiple hierarchical LF models listed in a recent review of MF modeling methods [48].

Based on the values of w and c, we can further classify adjustment MF surrogates into three types as shown in Fig. 3:

- Additive, for $w = 0$.
- Multiplicative, for $w = 1$ and $c = 0$.
- Hybrid additive/multiplicative, for either $w \notin \{0, 1\}$, or $w = 1$ and $c \neq 0$.

Table 1 enumerates important published works on each type of adjustment MF surrogates. We see that, a substantial portion of these works has relied on additive adjustment MF surrogates.

Apart from the form in Eq. (2), adjustment MF surrogates can also be described using weighted average models when there exist a total of m LF models, such that [41]

$$f_{\mathrm{H}}(\mathbf{x}) = \sum_{i=1}^{m} w_i f_i(\mathbf{x}) + \delta(\mathbf{x}), \tag{3a}$$

$$f_{\mathrm{H}}(\mathbf{x}) = \sum_{i=1}^{m} \rho_i(\mathbf{x}) f_i(\mathbf{x}) + \delta(\mathbf{x}), \tag{3b}$$

where $f_i(\mathbf{x})$ denotes a surrogate for approximating the prediction of MF model i, w_i is a weight value corresponding to $f_i(\mathbf{x})$ with $\sum_{i=1}^{m} w_i = 1$, and $\rho_i(\mathbf{x})$ is a weight function corresponding to $f_i(\mathbf{x})$ with $\sum_{i=1}^{m} \rho_i(\mathbf{x}) = 1$. We use Eq. (3a) when the weights are fixed in different regions of the design variable space. We use Eq. (3b) when we describe the correlation between $f_{\mathrm{H}}(\mathbf{x})$ and $f_i(\mathbf{x})$ at any values of the design variables.

Table 1 Published works on additive, multiplicative, hybrid adjustment MF surrogates.

Type	Reference
Additive	Forrester et al. [16], Viana et al. [17], Kennedy and O'Hagan [18], Han and Görtz [19], Lewis and Nash [49], Gano et al. [50], Viana et al. [51], Palar et al. [52], Zhang et al. [53], Fernández-Godino et al. [54], Song et al. [55], Kou and Zhang [56], Meng and Karniadakis [57], Durantin et al. [58], Teichert and Garikipati [59], Yan and Zhou [60], Leary et al. [61], Xiong et al. [62], Kuya et al. [63], Toal and Keane [64], Keane [65], Goh et al. [66], Zheng et al. [67], de Baar et al. [68], Park et al. [69], Zhang et al. [70], Rokita and Friedmann [71], Xiao et al. [72], Jiang et al. [73], Serani et al. [74], Zhou et al. [75], Shu et al. [76], Kaps et al. [77], Toal [78], Xu et al. [79], Ribeiro et al. [80], Peng et al. [81], Wiangkham et al. [82]
Multiplicative	Haftka [43], Chang et al. [83], Goldfeld et al. [84], Hino et al. [85], Sun et al. [86]
Hybrid	Gratiet and Garnier [20], Gano et al. [47, 87], Han et al. [88], Tyan et al. [89], Nguyen et al. [90], Hu et al. [91], Absi and Mahadevan [92], Rumpfkeil and Beran [93], Bryson and Rumpfkeil [94], Rumpfkeil and Beran [95], Wang et al. [96], Qian and Wu [97], Gratiet [98], Parussini et al. [99], Hao et al. [100], Ji et al. [101], Cheng et al. [102]

Table 2 Published works on deterministic and probabilistic adjustment MF surrogates.

Type	Reference
Deterministic	Palar et al. [52], Zhang et al. [53], Fernández-Godino et al. [54], Song et al. [55], Kou and Zhang [56], Durantin et al. [58], Yan and Zhou [60], Han et al. [88], Tyan et al. [89], Nguyen et al. [90], Rumpfkeil and Beran [93, 95], Wang et al. [96], Liu et al. [103]
Probabilistic	Forrester et al. [16], Kennedy and O'Hagan [18], Han and Görtz [19], Gratiet and Garnier [20], Xiong et al. [62], Kuya et al. [63], Toal and Keane [64], Keane [65], Goh et al. [66], Park et al. [69], Xiao et al. [72], Serani et al. [74], Qian and Wu [97], Gratiet [98], Parussini et al. [99], Ji et al. [101], Allaire and Willcox [104], Zhou et al. [105]

Table 3 Published works on composition MF methods.

Type	Reference
Input-input mapping	Bandler et al. [21, 22], Koziel et al. [106], Robinson et al. [107], Tao et al. [108]
Output-output mapping	Perdikaris et al. [23], Zheng et al. [109], Zhou et al. [110], Jiang et al. [111], Cutajar et al. [112], Hebbal et al. [113], Li et al. [114]

Based on how $\rho(\mathbf{x})$ and $\delta(\mathbf{x})$ are modeled, we further classify adjustment MF surrogates into deterministic and probabilistic ones; see Fig. 4. Table 2 lists several published works on each of these classes.

Deterministic adjustment MF surrogates described in Fig. 4 model both $\rho(\mathbf{x})$ and $\delta(\mathbf{x})$ as linear combinations of a finite number of basis functions $\zeta_i(\mathbf{x})$ and $\xi_j(\mathbf{x})$, respectively. Popular basis functions include monomials [53, 54], neural networks [56], radial basis functions [58, 89, 103], and orthogonal polynomial functions [52]. The combination coefficients and hyperparameters associated with the basis functions are determined by the least squares or regularized least squares approach. In comparison, probabilistic adjustment MF surrogates describe $\rho(\mathbf{x})$ or $\delta(\mathbf{x})$ as a stochastic process, which is often a GP characterized by a mean function and a covariance function. Several GP-based MF surrogates can be found in [115]. The mathematical expressions of common GP-based MF surrogates are deferred until Section IV.

Note that the aforementioned classifications do not conflict. This means, an additive (multiplicative, or hybrid) adjustment MF surrogate can be either deterministic or probabilistic.

Composition. The class of composition methods consists of input-input and output-output mapping MF techniques. MF surrogates from input-input mapping techniques often take the form of $f_{\mathrm{H}}(\mathbf{x}) \approx f_{\mathrm{L}}(g(\mathbf{x}))$, where $g(\cdot)$ maps the input space of the HF model to that of the LF model. Given an initial set of HF design variable values and the corresponding output values, the set of corresponding LF design variable values is found by parameter extraction and a deterministic mapping $g(\cdot)$ can be found iteratively by defining $g^j(\cdot)$ at iteration j as a linear combination of some predefined and fixed basis functions [21]. An appropriate mapping $g(\cdot)$ is found when $\|\mathbf{f}_{\mathrm{H}}(\mathbf{x}_{\mathrm{H}}) - \mathbf{f}_{\mathrm{L}}(\mathbf{x}_{\mathrm{L}})\| \leq \varepsilon$, for a small positive constant ε. Meanwhile, MF surrogates from output-output mapping techniques take the form of $f_{\mathrm{H}}(\mathbf{x}) \approx h\left(f_{\mathrm{L}}(\mathbf{x})\right)$, where $h(\cdot)$ maps the output space of the LF model to that of the HF model. The idea is to convert the original many-to-one mapping relationship between the space of HF design variables and the space of HF objective function to a one-to-one mapping relationship between certain regions of the spaces of LF and HF objective functions. This is possible because the values of LF objective function $f_{\mathrm{L}}(\mathbf{x}_{\mathrm{H}})$ at HF design variable values can be obtained easily at a low computational cost. Moreover, a composition MF surrogate can be deterministic or probabilistic depending on how we define the mappings $g(\cdot)$ and $h(\cdot)$. The reader is referred to Table 3 for a list of published works related to the composition MF modeling methods.

Input augmentation. The class of input-augmentation MF modeling methods describes MF surrogates as functions that depend not only on design variables but also on fidelity variables. This differs from the class of adjustment MF surrogates in Eqs. (2) and (3) which are functions of the design variables only. The fidelity variables are indeed categorical but can be assumed to be continuous for computational efficiency [116]. If categorical fidelity variables are considered, non-continuous covariance functions are proposed for constructing GP-based MF surrogates. Section IV.F

Table 4 Published work on MFO algorithms.

Type	Reference
Derivative-based	Alexandrov et al. [14, 44], Gano et al. [47], Robinson et al. [107], Alexandrov et al. [117], March et al. [118], Elham [119], Bryson and Rumpfkeil [120], De et al. [121], Wu et al. [122], Zhang et al. [123]
Model-then-optimize	Viana et al. [51], Leusink et al. [124], Singh et al. [125], Yang et al. [126]
Model-and-optimize (MF BO only)	Huang et al. [15], Forrester et al. [16], Tran et al. [28], Zhang et al. [70], Serani et al. [74], Shu et al. [76], Ribeiro et al. [80], Hao et al. [100], Kandasamy et al. [116], Sóbester et al. [127], Perdikaris and Karniadakis [128], Chen et al. [129], Pang et al. [130], Amrit et al. [131], Bonfiglio et al. [132, 133], Ghoreishi and Allaire [134], Bailly and Bailly [135], Shi et al. [136], Kontogiannis et al. [137], Tran et al. [138], Ruan et al. [139], Nachar et al. [140], Fiore et al. [141], Wu et al. [142], He et al. [143], Zhang et al. [144], Khatamsaz et al. [145], Sacher et al. [146], Wu et al. [147], Renganathan et al. [148], Kishi et al. [149], Cheng et al. [150], Foumani et al. [151], Huang et al. [152], Grassi et al. [153], Fiore and Mainini [154], Shintani et al. [155], Lin et al. [156], Winter et al. [157]

provides different input-augmentation MF modeling methods for both continuous and categorical fidelity variables. Additionally, most input-augmentation MF surrogates that exist in the literature are probabilistic and have been predominantly used in BO.

C. Multi-fidelity optimization strategies

MFO strategies take advantage of MF surrogates and/or their derivatives to accelerate the optimization process. Each MFO strategy is characterized by its way of updating the design points and MF surrogates during the optimization process. There are three approaches to MFO strategies: derivative-based, model-then-optimize, and model-and-optimize. Notable published works on the algorithms of each approach are listed in Table 4.

Derivative-based approach. This approach uses a deterministic MF surrogate and its derivatives to inform the step length and the search direction. The optimization is often via the trust-region framework that solves a trust-region subproblem to find the search direction [6]. In particular, we first restrict the step size (i.e., define the trust region) for a reliable local model, and then find the search direction within the defined truss region so that it minimizes the local model. In each optimization iteration, a deterministic MF surrogate serves as the local model, while its derivatives guide the search toward a good solution to the trust-region subproblem [14, 44, 107]. The MF surrogate is also used for computing a ratio of actual to predicted improvement for checking whether the new design point obtained by solving the trust-region subproblem is accepted or not, and whether the local model agrees with the actual objective function if the new design point is accepted. This ratio also allows the trust region to adjust its size. To make the algorithm more robust, the MF surrogate is occasionally calibrated using consistency conditions which ensure the preservation of both HF response and its derivatives through the use of local MF surrogates [14, 44]. These conditions remain

valid under assumptions that the HF model is considered the ground truth, and that the HF model and its derivatives are deterministic. To further address constrained optimization problems, the MF trust-region framework is equipped with a constraint-handling technique, for example, Lagrange multiplier method [107, 118], augmented Lagrangian method [44], and penalty method with either l_1 penalty functions [44, 47] or quadratic penalty functions [119].

While the derivative-based MFO approach can handle high-dimensional optimization problems, its nature as a derivative-based approach drives it toward several limitations.

- First, the calibration of local MF surrogates using HF derivatives may hinder direct applications of the approach to engineering design problems because it is often nontrivial to extract the derivatives of quantities of interest from engineering HF models.
- Second, the approach demands a high level of expertise in optimization from practicing engineers. This is because the performance of local approximations strongly depends on a careful selection of optimization parameters underlying the trust-region framework. These parameters, including the threshold values for the ratio of actual to predicted improvement, trust-region scaling factors, and tolerance thresholds, are essential for ensuring not only the solution improvement but also the accuracy of local MF surrogates and solution quality in each iteration.
- Third, the approach may provide no insight to engineers because its derivative-based nature is generally nontransparent [7].
- Finally, the approach cannot consider imperfections inherent in an HF model.

A possible way to avoid using the local information of HF model in solving a trust-region subproblem is to adopt derivative-free trust-region algorithms [158, 159]. These algorithms are designed to determine the search direction for each iteration so that it simply satisfies the conditions of Cauchy decrease and eigenvector decrease of the local model. Additionally, derivative-free trust-region algorithms do not require fully-linear or fully-quadratic local models in all iterations as their derivative-based counterparts often do, but require fully-linear or fully-quadratic local models during a finite, uniformly bounded, number of iterations to achieve global convergence to first-order or second-order stationary point, respectively [158]. [160, 161] exploited the concept of derivative-free trust-region algorithms for developing both unconstrained and constrained MFO algorithms.

Model-then-optimize approach. As its name implies, this approach separates the construction of an MF surrogate for the costly objective function from the optimization process. It relies on a strong assumption that the MF surrogate for the objective function possesses sufficient accuracy to ensure both the feasibility and quality of candidate solutions. However, obtaining such a level of accuracy may be challenging without the support of an optimization algorithm, primarily because not all regions of the design variable space are useful for optimization. In practice, focusing on learning a high-confidence region of interest can lead to favorable optimal solutions [162].

The model-then-optimize approach is useful in certain cases. It allows an examination of how initial sampling

designs influence the solution quality. It also enables the use of population-based algorithms for solving multi-objective optimization problems [51, 124–126].

Model-and-optimize approach . This approach, also known as sequential model-based optimization approach [163], iterates between updating global MF surrogates for the costly objective function and using them to propose new design points via maximizing an acquisition function (i.e., infill criterion or figure of merit). From this definition, MF BO is a type of model-and-optimization approach. The model-and-optimize approach, via formulating the acquisition function, minimizes the need for using numerous optimization parameters as the derivative-based approach does. Factors that affect the performance of model-and-optimize algorithms include the choice of MF surrogates, the selection of acquisition functions, and the construction of initial MF surrogates. As shown in Table 4, this approach has recently attracted a large number of published works on the applications of MFO in design optimization.

III. Multi-fidelity Bayesian optimization

As illustrated in Fig. 2 and briefly defined in **??**, BO is a global optimization technique working based on constructing a probabilistic surrogate model representing our belief about an expensive-to-evaluate objective function given the data, which then defines an optimization policy via an acquisition function maximized before a new design point is selected for evaluating the objective function to update our belief. BO was originated by [164] and popularized by [165] and their work on the *Efficient Global Optimization* (EGO) algorithm. Combining BO with MFO leads to MF BO which further accelerates the optimization process.

This section starts by describing the fundamental elements of EGO (Section III.A) and BO (Section III.B), and how the two concepts connect to each other. It then elucidates the distinctive characteristics that set MF BO apart from BO (Section III.C).

A. Efficient global optimization

EGO operates through a sequence of three key steps in each iteration. Oftentimes it terminates when reaching a pre-specified upper limit on the number of iterations, which corresponds to a fixed computational budget.

In the first step, EGO uses a Kriging model [32, 166] as a stochastic process for approximating the costly objective function. The hyperparameters underlying the Kriging model are determined by maximizing model likelihood given the current training dataset [165]. This enables the derivation of the best linear unbiased predictor and the MSE of the predictor. Traditionally, in geostatistics, the best linear unbiased predictor of Kriging is found by minimizing the MSE [32, 39, 166]. However, within the context of EGO, minimizing the MSE is equivalent to maximizing the likelihood because EGO models the residual term of Kriging as a zero-mean random function [165].

In the second step, EGO formulates what is known as the expected improvement based on (i) the best design point

found in the current training dataset, (ii) the best linear unbiased predictor, and (iii) the MSE of the predictor. The concept of expected improvement (see Section V.A.1) provides a delicate balance between exploiting the information provided by the predictor and exploring uncertain regions within it. More specifically, exploitation (immediate reward) focuses on regions of the design variable space where the objective function values are expected to be small, while exploration (expected future reward) discovers regions where the predictions of the objective function are highly uncertain.

In the third step, EGO maximizes the expected improvement for a new design point that is added to the current training dataset, leading to updates of both the stochastic model and the best-observed solution in the next iteration. Notably, maximizing the expected improvement is more straightforward and computationally efficient than computing the derivatives of the costly objective function. This computational advantage contributes to the overall efficiency of EGO.

In the community of structural optimization, a Kriging model to approximate an objective function is often considered a global surrogate. However, if we view design points in the current training dataset locally, we can consider a Kriging model as a local surrogate, which knows nothing about the objective function in regions that are away from the design points. In this context, EGO is a technique that relies on a local surrogate for global optimization.

B. Bayesian optimization

While the fundamental concept behind BO closely resembles that of EGO, the nomenclature of BO is commonly used in the communities of statistics and machine learning [24–26, 167]. Oftentimes BO uses a univariate GP (see Section VII) as a surrogate for the costly objective function while an acquisition function guides the optimization process. Thus, EGO can be considered a specific instance of BO that adopts the expected improvement as an acquisition function. Many fundamental similarities shared between Kriging and GP further reinforce the connection between EGO and the broader family of BO algorithms [168].

Although we use Kriging and GP interchangeably in this work, as also observed elsewhere [see e.g., 10, 169], it is worth noting that there is a subtle difference between the two concepts. This difference lies in the degrees of human intervention for hyperparameter estimation and treatment of anisotropy [168]. While Kriging determines its hyperparameters from a frequentist perspective by minimizing the MSE of the predictor, GP determines its hyperparameters from a Bayesian perspective by maximizing the likelihood or utilizing a full Bayesian treatment for the mode of hyperparameter vector. In Kriging, the covariance between two objective function values only depends on the relative point-wise distance of the corresponding design variable vectors, which is called stationary. If such a relative distance is calculated in a Euclidean space, Kriging's assumption of intrinsic stationarity also implies isotropy, i.e., the covariance is invariant to both translation and rotation. This property requires a high degree of human intervention to prescale or rotate the coordinate system for ensuring fidelity in spatial modeling. In comparison, GP requires less human

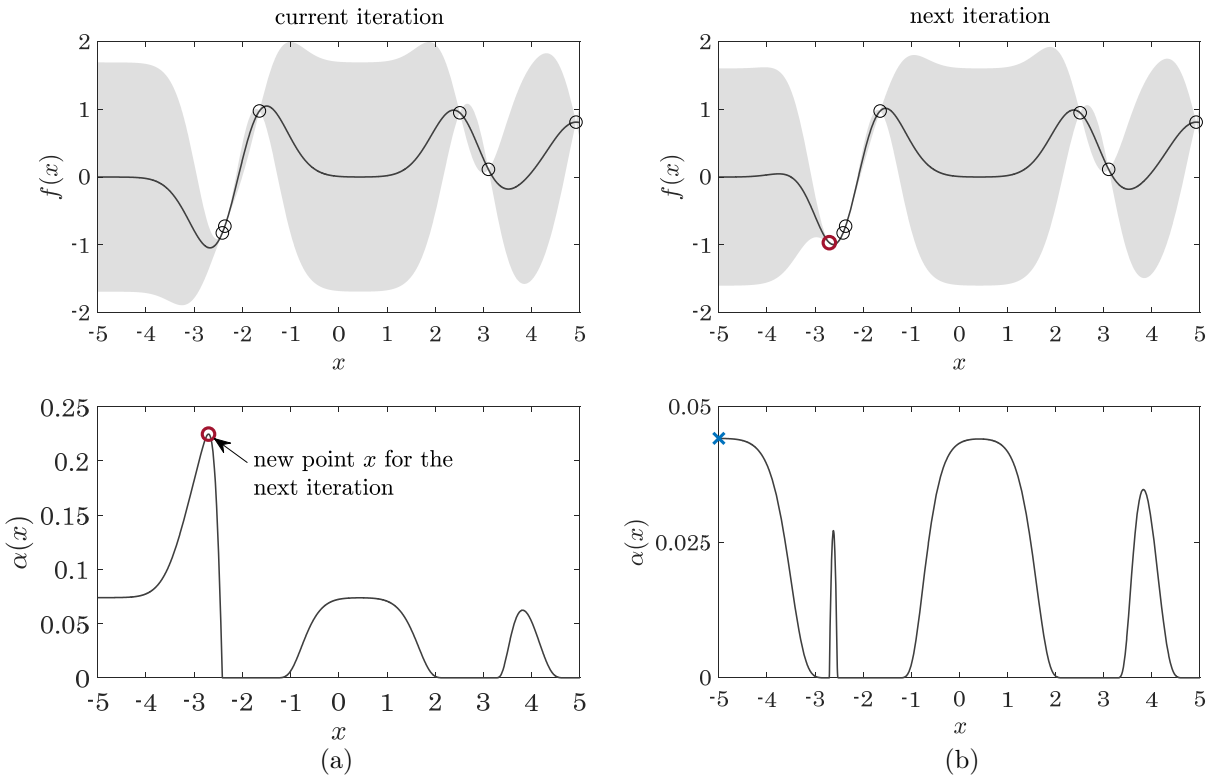

Fig. 5 Illustration of two consecutive iterations of BO for minimizing a univariate objective function $f(x)$.

intervention by using different characteristic length scales along each axis or more flexibly parameterized rotations and scales, thus allowing the use of anisotropy covariance functions. This is motivated by our expectation that a very long length scale should be used for a less influential design variable.

Figure 5 illustrates two consecutive iterations of BO for minimizing a univariate objective function $f(x)$. In Fig. 5(a), BO fits a GP model for $f(x)$ to a training dataset of six samples (top panel). Subsequently, it formulates and maximizes an acquisition function $\alpha(x)$ to identify a new design point (bottom panel). In Fig. 5(b), BO proceeds by updating the GP model and reformulating $\alpha(x)$ with the design point identified in the previous iteration.

Algorithm 1 shows a pseudo-code for solving problem (1) using the generic BO, where the objective function is the prediction $f_{\mathrm{H}}(\mathbf{x})$ of an HF model. Parameters K and N in Line 1 denote threshold values for the number of BO iterations and the number of initial training samples, respectively. $\mathcal{D}^0$ in Line 6 is the initial training dataset generated from the HF model. Note that, $\mathcal{D}^0$ can be HF data, which contain both predictions of the HF model and measures from experiments. $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ in Line 9 and $\mathcal{D}^k$ in Line 13 are the GP posterior and training dataset associated with iteration k of BO, respectively. The acquisition function $\alpha(\mathbf{x})$ is formulated in Line 10 based on $\{\mathbf{x}_{\min}, f_{\min}\}$ and $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$, which is detailed in Section V. Maximizing $\alpha(\mathbf{x})$ in Line 11 is often straightforward as it does not invoke the HF model. Thus, we replace problem (1), which is difficult and expensive to solve, by a series of simpler, inexpensive problems of maximizing the acquisition function.

Algorithm 1 Generic BO.

1: **Input:** $\mathcal{X}$, K, N;
2: Generate N samples of $\mathbf{x}$;
3: **for** $i = 1 : N$ **do**
4: $\quad f^i \leftarrow f_{\mathrm{H}}(\mathbf{x}^i)$; ▷ Costly step
5: **end for**
6: $\mathcal{D}^0 \leftarrow \{\mathbf{x}^i, f^i\}_{i=1}^N$;
7: $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow \min\{f^i,\, i = 1, \ldots, N\}$;
8: **for** $k = 1 : K$ **do**
9: $\quad$ Construct $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ based on $\mathcal{D}^{k-1}$;
10: $\quad$ Formulate $\alpha(\mathbf{x})$;
11: $\quad \mathbf{x}^k \leftarrow \underset{\mathbf{x}}{\operatorname{argmax}}\ \alpha(\mathbf{x})$ s.t. $\mathbf{x} \in \mathcal{X}$;
12: $\quad f^k \leftarrow f_{\mathrm{H}}(\mathbf{x}^k)$; ▷ Costly step
13: $\quad \mathcal{D}^k \leftarrow \mathcal{D}^{k-1} \cup \{\mathbf{x}^k, f^k\}$;
14: $\quad \{\mathbf{x}_{\min}, f_{\min}\} \leftarrow \min\{f_{\min}, f^k\}$;
15: **end for**
16: **return** $\{\mathbf{x}_{\min}, f_{\min}\}$.

At this point, a critical question arises: under what conditions can we ensure the existence of a unique global solution to a GP-distributed objective function $f_{\mathrm{H}}(\mathbf{x})$? Addressing this question makes the construction of the GP $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ in Line 9 of Algorithm 1 meaningful. Since the optimization problem is completely defined by the mean and covariance functions of the GP and the design domain $\mathcal{X}$, their properties decide whether a unique global solution is guaranteed. In fact, we can be certain that a global solution exists and that is unique under the following two mild assumptions: (i) the mean and covariance functions are continuous over a compact design domain $\mathcal{X}$, and (ii) there are no two unique points in the domain that can have perfectly correlated function values [170]. The reader may consult [170] for detailed discussions on the existence and uniqueness of global solutions to GP-distributed functions.

The unique ability to optimize expensive-to-evaluate objective functions, i.e., making efficient use of the available data for ensuring solution accuracy, has made BO an invaluable tool for various applications in scientific and engineering design. These applications can be found in machine learning [171], aircraft design [172], material design [30, 173], experimental design [174], material science [175], structural engineering [176], transportation [177], chemical engineering [178], electronics engineering [179], environmental engineering [180], and physics [181]. Moreover, sophisticated BO algorithms have been developed to (i) examine the performance of acquisition functions formulated from different optimization policies when optimizing under uncertainty in the objective function and (ii) extend applications of BO to a wide range of design optimization problems. The first aspect has examined the performance of BO acquisition functions derived from different formulation approaches: improvement-based [182], optimistic [183], information-based [184], and likelihood-weighted [185]. It has also explored the performance of acquisition functions from one-step look-ahead and multi-step look-ahead perspectives (see Section V.A). The second aspect has focused on developing advanced BO algorithms to solve intricate yet important design optimization problems such as constrained problems, high-dimensional problems, problems under uncertainty, and multi-objective problems (see Section VI).

These problems are difficult to solve because of their own nature and/or limitations of the generic BO. For the details of recent advances in BO, the reader may refer to comprehensive surveys and tutorials by [186], [25], [26], and [31].

C. Multi-fidelity Bayesian optimization

As illustrated in Fig. 2, MF BO is the intersection of MFO and BO, and therefore it inherits advantages of the two approaches. There are several reasons that make MF BO powerful, especially when solving engineering optimization problems:

- **Resource saving**: MF BO uses MF surrogates for a costly objective function, thus further reducing the number of HF data points by exploiting the structure of mathematical modeling details and/or using our knowledge about the underlying physical process.
- **Handling exploitation-exploration trade-off**: The optimization process can benefit from the attempts to address the exploitation-exploration trade-off by notable acquisition functions of the generic BO.
- **Robustness to noise**: Engineering optimization problems often involve noisy objective and constraint functions. While LF evaluations are less noisy than their HF counterparts, incorporating evaluations from both LF and HF models into BO can make the optimization process more robust to noise. This is justified because LF models are less detailed, therefore exhibit lower variability, whereas HF models capture more intricate details and complexities.
- **Parallelization**: It is easy to obtain LF evaluations in parallel, which further accelerates the optimization process.
- **Adaptive optimization of fidelity**: MF BO allows for adaptive selections of fidelity levels. This means, it can decide when and where to carry out HF evaluations based on the current state of knowledge about the objective function. This adaptability makes MF BO well-suited for situations when the cost or availability of HF evaluations varies across different regions of the design variable space.
- **Incorporation of non-GP models**: While the generic BO often relies on common GP models that, from a local view of design points in the training dataset, can only construct local surrogates, many LF models provide global information. The use of MF models enables BO to expand it horizon beyond the constraint of GP models.

Compared to the generic BO outlined Algorithm 1, MF BO introduces two important modifications. First, it constructs in Line 9 a GP-based MF surrogate for the objective function using LF models and HF data. One of the most popular GP-based MF surrogates is the auto-regressive model [18]; see Section IV.B.1. Second, MF BO develops in Line 10 an MF acquisition function that is capable of selecting both a new design point and the fidelity level for a computational model to be called in Line 12. The very first example of such an acquisition function is the so-called augmented expected improvement [15]; see Section V.B.2.

The pioneering works by [18] and [15] have initiated three directions of recent research and development efforts of MF BO:

- Enhancing MF surrogates; see Section IV.
- Innovating new MF acquisition functions based on the existing BO acquisition functions; see Section V.
- Applying MF BO to solving various optimization problems in science and engineering design; see Table 4.

IV. Gaussian process-based multi-fidelity surrogates

Let $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_T(\mathbf{x})]^\intercal, T \geq 2$, denote a vector of T outputs corresponding to T computational models with different fidelities used for predicting the objective function of an engineering design optimization problem. We assume that the models share the same input variables so that we do not need input variable mappings to perform information transfers between the fidelities. If the fidelities are sorted in increasing order, there exist $(T-1)$ LF predictions, and $f_T(\mathbf{x}) = f_{\mathrm{H}}(\mathbf{x})$ is the prediction of the highest-fidelity model. If there is no obvious fidelity ordering, then there is no hierarchy of the predictions.

Let $\mathbf{X}_t \in \mathbb{R}^{N_t \times d}$, $t = 1, \ldots, T$, be a set of N_t input samples associated with fidelity t and $\mathbf{F}_t \in \mathbb{R}^{N_t}$ be a set of the corresponding output values. Let $\mathbf{X} = [\mathbf{X}_1; \ldots; \mathbf{X}_T] \in \mathbb{R}^{\sum_{t=1}^{T} N_t \times d}$ and $\mathbf{F} = [\mathbf{F}; \ldots; \mathbf{F}] \in \mathbb{R}^{\sum_{t=1}^{T} N_t}$ denote sets of input and output data, where $\mathbf{X}$ and $\mathbf{F}$ concatenate matrices $\mathbf{X}_t$ and vectors $\mathbf{F}_t$, respectively. Thus, $\mathcal{D}_t = [\mathbf{X}_t, \mathbf{F}_t] \in \mathbb{R}^{N_t \times (d+1)}$ and $\mathcal{D} = [\mathcal{D}_1; \ldots; \mathcal{D}_T] \in \mathbb{R}^{\sum_{t=1}^{T} N_t \times (d+1)}$ are training datasets associated with fidelity t and all fidelities, respectively.

In this section, we describe popular GP-based MF surrogates constructed from $\mathcal{D}$ for use of MF BO. They include multivariate GP via linear model of coregionalization (Section IV.A), auto-regressive model and its variants (Section IV.B), graphical MF GP model (Section IV.C), Bayesian hierarchical model (Section IV.D), composition of GPs (Section IV.E), and input-augmentation GP-based MF surrogate models (Section IV.F). To obtain a structured understanding of these models, we attempt to exploit common properties shared between them.

A. Linear model of coregionalization

Without ordering the fidelities, consider the problem of constructing a T-variate GP surrogate to approximate the output vector $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_T(\mathbf{x})]^\intercal$, which can be described as

$$\mathbf{f}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \boldsymbol{\delta}(\mathbf{x}), \tag{4}$$

where $\mathbf{m}(\mathbf{x})$: $\mathbb{R}^d \mapsto \mathbb{R}^T$ and $\boldsymbol{\delta}(\mathbf{x})$: $\mathbb{R}^d \mapsto \mathbb{R}^T$ are mean and discrepancy vectors, respectively.

The *multivariate GP via linear model of coregionalization* (LMC) describes each element of $\boldsymbol{\delta}(\mathbf{x})$ as a linear combination of T independent zero-mean GPs with covariance functions $\kappa_1(\cdot, \cdot|\boldsymbol{\phi}_1), \ldots, \kappa_T(\cdot, \cdot|\boldsymbol{\phi}_T)$, where $\boldsymbol{\phi}_t$ is the hyperparameter vector of κ_t [see e.g., 166, 187]. Thus, the discrepancy vector can be written as $\boldsymbol{\delta}(\mathbf{x}) = \mathbf{R}\mathbf{z}(\mathbf{x})$, where $\mathbf{R} \in \mathbb{R}^{T \times T}$ consists of combination coefficients, and $\mathbf{z}(\cdot) \sim \mathcal{GP}(\mathbf{0}, \boldsymbol{\Sigma}(\cdot, \cdot))$ is a T-variate GP with $\boldsymbol{\Sigma}(\cdot, \cdot) =$

$\text{diag}\{\kappa_1(\cdot,\cdot|\boldsymbol{\phi}_1),\ldots,\kappa_T(\cdot,\cdot|\boldsymbol{\phi}_T)\}$. Accordingly, the GP prior for $\mathbf{f}(\cdot)$ in Eq. (4) reads

$$\mathbf{f}(\cdot) \sim \mathcal{GP}\left(\mathbf{m}(\cdot), \mathbf{S}(\cdot,\cdot)\right). \tag{5}$$

Here $\mathbf{S}(\cdot,\cdot) = \mathbf{R}\boldsymbol{\Sigma}(\cdot,\cdot)\mathbf{R}^\intercal = \sum_{t=1}^{T} \mathbf{C}_t \kappa_t(\cdot,\cdot|\boldsymbol{\phi}_t)$ is the so-called inter-group covariance matrix, where $\mathbf{C}_t = \mathbf{r}_t\mathbf{r}_t^\intercal$ is the t-th coregionalization matrix and $\mathbf{r}_t$ the t-th column of $\mathbf{R}$. To determine $\mathbf{R}$, $\boldsymbol{\phi}_t$, and parameters underlying $\mathbf{m}(\mathbf{x})$, we condition the GP prior in Eq. (5) on the training dataset $\mathcal{D}$, which is called the training process, as detailed in Section VII. It is worth noting that the inter-group covariance matrix $\mathbf{S}(\cdot,\cdot) \in \mathbb{R}^{T\times T}$ should not be confused with a significantly larger covariance matrix $\mathbf{K}(\mathbf{X},\mathbf{X}) \in \mathbb{R}^{\sum_{t=1}^{T} N_t \times \sum_{t=1}^{T} N_t}$ that is used for the training process and prediction equations, i.e., Eqs. (B.5) and (B.6) of Section VII.

Once trained, the LMC model can provide predictions for the outputs of all fidelities at unseen input variable vectors. An advantage of LMC is that it can use nonseparable covariance structures for describing the outputs. This results in sufficient good predictions and the capability of capturing joint uncertainty about the outputs, especially when they are different physical quantities [187]. A main disadvantage of the method is the huge computational cost it requires for training, which scales cubically with $\sum_{t=1}^{T} N_t$.

B. Auto-regressive model and variants

1. KOH auto-regressive model

If the fidelities are ordered, the surrogate for predicting the output of fidelity t ($t = 2,\ldots,T$) can be constructed from the surrogate that approximates the output of fidelity $(t-1)$ using the *auto-regressive model* (i.e., KOH model), which reads [18]

$$f_1(\mathbf{x}) = \delta_1(\mathbf{x}), \tag{6a}$$

$$f_t(\mathbf{x}) = b_{t-1} f_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad t = 2,\ldots,T, \tag{6b}$$

$$\text{cov}\left[f_t(\mathbf{x}), f_{t-1}(\mathbf{x}')|f_{t-1}(\mathbf{x})\right] = 0, \tag{6c}$$

where b_{t-1} in Eq. (6b) is an unknown correlation coefficient and $\delta_t(\cdot) \sim \mathcal{GP}\left(m_{\delta,t}(\cdot), \kappa_{\delta,t}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,t})\right)$ is the discrepancy function, which is independent of $f_{t-1}(\cdot),\ldots,f_1(\cdot)$. Here b_{t-1} and $\delta_t(\cdot)$ play similar roles as b and $\delta(\cdot)$ in Eq. (2), respectively. A common choice for the mean function of $\delta_t(\cdot)$ is $m_{\delta,t}(\mathbf{x}) = \sum_{i=1}^{p} \beta_i \zeta_i(\mathbf{x})$, where β_i and $\zeta_i(\mathbf{x})$ are combination coefficients and basis functions, respectively. The covariance function $\kappa_{\delta,t}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,t})$ is often the squared exponential covariance function [10, 18]. Other covariance functions for use of GP-based MF modeling include Matern [130] and composite covariance functions [188]. $\text{cov}\left[f_t(\mathbf{x}), f_{t-1}(\mathbf{x}')|f_{t-1}(\mathbf{x})\right]$ in Eq. (6c) denotes the covariance of two random variables $f_t(\mathbf{x})$ and $f_{t-1}(\mathbf{x}')$ given that $f_{t-1}(\mathbf{x})$ is known. The Markov property in Eq. (6c) implies that

observing $f_{t-1}(\mathbf{x}')$ provides no information for predicting $f_t(\mathbf{x})$ if $f_{t-1}(\mathbf{x})$ is observed.

For simplicity, consider two fidelities $f_1(\mathbf{x}) = f_{\mathrm{L}}(\mathbf{x})$ and $f_2(\mathbf{x}) = f_{\mathrm{H}}(\mathbf{x})$. In this case, the KOH model reads

$$f_1(\mathbf{x}) = \delta_1(\mathbf{x}), \tag{7a}$$

$$f_2(\mathbf{x}) = b_1 f_1(\mathbf{x}) + \delta_2(\mathbf{x}). \tag{7b}$$

Let $z_1(\cdot) \sim \mathcal{GP}\left(0, \kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1})\right)$ and $z_2(\cdot) \sim \mathcal{GP}\left(0, \kappa_{\delta,2}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,2})\right)$, where the covariance functions $\kappa_{\delta,1}$ and $\kappa_{\delta,2}$ are parameterized by the hyperparameter vectors $\boldsymbol{\phi}_{\delta,1}$ and $\boldsymbol{\phi}_{\delta,2}$, respectively. Eq. (7) can be rewritten as

$$f_1(\mathbf{x}) = m_{\delta,1}(\mathbf{x}) + z_1(\mathbf{x}), \tag{8a}$$

$$f_2(\mathbf{x}) = b_1 m_{\delta,1}(\mathbf{x}) + m_{\delta,2}(\mathbf{x}) + b_1 z_1(\mathbf{x}) + z_2(\mathbf{x}). \tag{8b}$$

This is equivalent to

$$\mathbf{f}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{R}\mathbf{z}(\mathbf{x}), \tag{9}$$

where

$$\mathbf{f}(\mathbf{x}) = \left[f_1(\mathbf{x}), f_2(\mathbf{x})\right]^\intercal, \tag{10a}$$

$$\mathbf{m}(\mathbf{x}) = \mathbf{R}\boldsymbol{\mu}(\mathbf{x}), \tag{10b}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 \\ b_1 & 1 \end{bmatrix} = \mathbf{I}_2 + b_1 \mathbf{e}_2 \mathbf{e}_1^\intercal, \tag{10c}$$

$$\boldsymbol{\mu}(\mathbf{x}) = \left[m_{\delta,1}(\mathbf{x}), m_{\delta,2}(\mathbf{x})\right]^\intercal, \tag{10d}$$

$$\mathbf{z}(\mathbf{x}) = \left[z_1(\mathbf{x}), z_2(\mathbf{x})\right]^\intercal, \tag{10e}$$

$$\mathbf{z}(\cdot) \sim \mathcal{GP}\left(\mathbf{0}, \boldsymbol{\Sigma}(\cdot,\cdot)\right), \tag{10f}$$

$$\boldsymbol{\Sigma}(\cdot,\cdot) = \begin{bmatrix} \kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}) & 0 \\ 0 & \kappa_{\delta,2}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,2}) \end{bmatrix}. \tag{10g}$$

Here $\mathbf{I}_2$ denotes the 2-by-2 identity matrix, $\mathbf{e}_1 = [1,0]^\intercal$, and $\mathbf{e}_2 = [0,1]^\intercal$.

Thus, the GP prior for $\mathbf{f}(\mathbf{x})$ in Eq. (9) can be written using the form in Eq. (5) with the inter-group covariance matrix

$$\begin{aligned}\mathbf{S}(\cdot,\cdot) = \mathbf{R}\boldsymbol{\Sigma}(\cdot,\cdot)\mathbf{R}^{\intercal} &= \begin{bmatrix} 1 & 0 \\ b_1 & 1 \end{bmatrix} \begin{bmatrix} \kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}) & 0 \\ 0 & \kappa_{\delta,2}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,2}) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ b_1 & 1 \end{bmatrix}^{\intercal} \\ &= \begin{bmatrix} \kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}) & b_1\kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}) \\ b_1\kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}) & b_1^2\kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}) + \kappa_{\delta,2}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,2}) \end{bmatrix},\end{aligned} \tag{11}$$

which is the inter-group covariance matrix for coKriging derived by [16].

In the general case when there exist T fidelities ($T \geq 2$), the terms of Eq. (9) are

$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_T(\mathbf{x})]^{\intercal}, \tag{12a}$$

$$\mathbf{m}(\mathbf{x}) = \mathbf{R}\boldsymbol{\mu}(\mathbf{x}), \tag{12b}$$

$$\mathbf{R} = \mathbf{R}_T \ldots \mathbf{R}_2,\ \mathbf{R}_t = \mathbf{I}_T + b_{t-1}\mathbf{e}_t\mathbf{e}_{t-1}^{\intercal},\ t = 2, \ldots, T, \tag{12c}$$

$$\boldsymbol{\mu}(\mathbf{x}) = [m_{\delta,1}(\mathbf{x}), \ldots, m_{\delta,T}(\mathbf{x})]^{\intercal}, \tag{12d}$$

where $\mathbf{I}_T$ denotes the T-by-T identity matrix, $\mathbf{e}_t$ denotes the T-dimensional standard unit vector with a 1 in the tth coordinate, and the coefficient matrix $\mathbf{R}$ is a T-by-T low triangular matrix [189].

2. Hierarchical Kriging model

[19] proposed a hierarchical Kriging model by making three modifications to the KOH model. First, they modeled the surrogate that approximates the lowest-fidelity output $f_1(\mathbf{x})$ in Eq. (6a) as a sum of an unknown constant and a stationary random process. Second, they replaced $f_{t-1}(\mathbf{x})$ in Eq. (6b) with the best linear unbiased predictor of Kriging, denoted by $\mu_{\mathrm{f},t-1}(\mathbf{x})$ [32]. Finally, they used $z_t(\cdot) \sim \mathcal{GP}\left(0, \kappa_{\mathrm{z},t}(\cdot,\cdot|\boldsymbol{\phi}_{\mathrm{z},t})\right)$ instead of $\delta_t(\cdot) \sim \mathcal{GP}\left(m_{\delta,t}(\cdot), \kappa_{\delta,t}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,t})\right)$, $t = 1, \ldots, T$. Accordingly, the *hierarchical Kriging model* reads

$$f_1(\mathbf{x}) = a + z_1(\mathbf{x}), \tag{13a}$$

$$f_t(\mathbf{x}) = b_{t-1}\mu_{\mathrm{f},t-1}(\mathbf{x}) + z_t(\mathbf{x}),\ \ t = 2, \ldots, T, \tag{13b}$$

where the unknown constant a in Eq. (13a) is described in Eq. (2). The GP prior for the hierarchical Kriging model can be derived from the model in Eq. (9) for $\mathbf{R} = \mathbf{I}_T$ and $\boldsymbol{\mu}(\mathbf{x}) = [a, b_1\mu_{\mathrm{f},1}(\mathbf{x}), \cdots, b_{T-1}\mu_{\mathrm{f},T-1}(\mathbf{x})]^{\intercal}$.

3. Recursive model

[20] made two modifications to the KOH model in Eq. (6). First, they adopted the hybrid additive/multiplicative

approach by replacing the regression coefficient b_{t-1} in Eq. (6b) with an adjustment coefficient function $\rho_{t-1}(\mathbf{x})$, which is similar to $\rho(\mathbf{x})$ in Eq. (2). Second, rather than using the GP prior $f_{t-1}(\mathbf{x})$ in Eq. (6b), they used the GP posterior $\hat{f}_{t-1}(\mathbf{x})$ constructed from the data $\mathcal{D}_{t-1}$ corresponding to fidelity $(t-1)$. Accordingly, the *recursive model* reads

$$f_1(\mathbf{x}) = \delta_1(\mathbf{x}), \tag{14a}$$

$$f_t(\mathbf{x}) = \rho_{t-1}(\mathbf{x})\hat{f}_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad t = 2, \ldots, T, \tag{14b}$$

$$\operatorname{cov}\left[f_t(\mathbf{x}), \hat{f}_{t-1}(\mathbf{x}')|\hat{f}_{t-1}(\mathbf{x})\right] = 0. \tag{14c}$$

Here $\rho_{t-1}(\mathbf{x}) = \boldsymbol{\zeta}_{t-1}^{\intercal}(\mathbf{x})\boldsymbol{\beta}_{t-1}$ (see Fig. 4) is a linear combination of a finite number of basis functions, and

$$\delta_t(\cdot) \sim \mathcal{GP}\left(m_{\delta,t}(\cdot), \kappa_{\delta,t}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,t})\right), \tag{15a}$$

$$\hat{f}_{t-1}(\cdot) \sim \mathcal{GP}\left(m_{\mathrm{f},t-1}(\cdot), \kappa_{\mathrm{f},t-1}(\cdot,\cdot)\right), \tag{15b}$$

where $m_{\mathrm{f},t-1}(\cdot)$ and $\kappa_{\mathrm{f},t-1}(\cdot,\cdot)$ are mean and covariance functions of the GP posterior $\hat{f}_{t-1}(\cdot)$, respectively.

An advantage of the recursive model is that the computational cost it requires for training the GP model that approximates the highest-fidelity output $f_T(\mathbf{x})$ is much less than that the KOH model requires, while the predictive efficiency is preserved. More specifically, the computational costs for training the recursive and KOH models are $O\left(T \times \max\{N_t^3, t = 1, \ldots, T\}\right)$ and $O(\sum_{t=1}^{T} N_t)^3$ respectively, where N_t is the sample size of data $\mathcal{D}_t$ [20].

C. Graphical multi-fidelity Gaussian process

In an attempt to generalize the KOH model, [101] developed the *graphical multi-fidelity Gaussian process* (GMGP) for cases when the hierarchy of LF models remains unclear. The GMGP approach involves constructing a directed acyclic graph (DAG) where each node represents a computational model, and each directed edge connecting two nodes represents the hierarchy of their fidelities. If the graph is in a topological ordering such that every directed edge (t', t) from t' to t, then $t' < t$ and the prediction of node t' is made before making the prediction of node t. Let T denote the root node for the highest-fidelity model. Figure 6 shows an example of a topological-ordering DAG of four nodes.

Consider node t. Let $t' \in \mathrm{Pa}(t)$ be a parent node of t. Thus, $t \in \mathrm{Ch}(t')$ is a child node of t'. Let V_{s} denote a set of all source nodes that have no lower-fidelity node, and $\bar{V}_{\mathrm{s}}$ denote a set of non-source nodes. The DAG in Figure 6, for example, has $V_{\mathrm{s}} = \{1, 2\}$ and $\bar{V}_{\mathrm{s}} = \{3, 4\}$.

GMGP models the surrogate for predicting the output at node t as a weighted sum of the surrogates predicting the outputs of its parent nodes, plus a discrepancy term $\delta_t(\cdot) \sim \mathcal{GP}\left(m_{\delta,t}(\cdot), \kappa_{\delta,t}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,t})\right)$. This can be generalized using

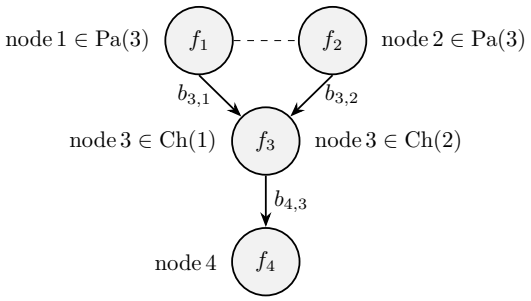

Fig. 6 An example of a directed acyclic graph for MF modeling (bottom-up descends fidelity order).

the following form:

$$\mathbf{f}(\mathbf{x}) = \mathbf{R}\boldsymbol{\mu}(\mathbf{x}) + \mathbf{R}\mathbf{z}(\mathbf{x}), \tag{16a}$$

$$\text{cov}\left[f_t(\mathbf{x}), f_{t'}(\mathbf{x}')|f_{t'}(\mathbf{x})\right] = 0. \tag{16b}$$

Here $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^T$ and $\boldsymbol{\mu}(\mathbf{x}) \in \mathbb{R}^T$ are described in Eqs. (12a) and (12d), respectively. $\mathbf{z}(\cdot) \sim \mathcal{GP}\left(\mathbf{0}, \boldsymbol{\Sigma}(\cdot,\cdot)\right)$ is a T-variate GP with $\boldsymbol{\Sigma}(\cdot,\cdot) = \text{diag}\{\kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1}), \ldots, \kappa_{\delta,T}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,T})\}$. The condition in Eq. (16b) holds for $t' \in \text{Pa}(t)$ or for $t, t' \in V_{\text{s}}$, $t \neq t'$. The coefficient matrix $\mathbf{R}$ in Eq. (16a) is described as

$$\mathbf{R} = \prod_{t=T-1}^{1} \prod_{j \in \text{Ch}(t)} \mathbf{R}_{jt}, \quad \mathbf{R}_{jt} = \mathbf{I}_T + b_{j,t}\mathbf{e}_j\mathbf{e}_t^{\intercal}, \tag{17}$$

where the correlation coefficient $b_{j,t}$ relates nodes j and t as indicated in Fig. 6. We see that, while the GMGP model is a type of the LMC model described in Section IV.A, it is more general than the KOH model described in Eq. (12) because the coefficient matrix $\mathbf{R}$ in Eq. (17) becomes that in Eq. (12c) if the fidelities of all nodes are ordered.

Inspired by [20], [101] also provided the following recursive formulation for the GMGP:

$$f_t(\mathbf{x}) = \sum_{t' \in \text{Pa}(t)} b_{t,t'}\hat{f}_{t'}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad t \in \bar{V}_{\text{s}}, \tag{18a}$$

$$\text{cov}\left[f_t(\mathbf{x}), \hat{f}_{t'}(\mathbf{x}')|\hat{f}_{t'}(\mathbf{x})\right] = 0, \quad t' \in \text{Pa}(t), \tag{18b}$$

where $\hat{f}_{t'}(\mathbf{x})$ is the GP posterior for predicting the output of node t', which is conditioned on the data associated with node t' and its ancestor nodes.

While the computational cost for training the GMGP in Eq. (16) is $O(\sum_{t=1}^{T} N_t)^3$, where N_t is the sample size of data associated with node t, that for training the model in Eq. (18) is $O\left(T \times \max\{N_t^3, t = 1, \ldots, T\}\right)$. Inference in GMGP and its applications to several numerical experiments and to emulation of heavy-ion collisions are detailed in [101].

D. Bayesian hierarchical model

Given two computational models with different fidelity for predicting an objective function, and the corresponding predictions $f_1(\mathbf{x}) = f_{\mathrm{L}}(\mathbf{x})$ and $f_2(\mathbf{x}) = f_{\mathrm{H}}(\mathbf{x})$, [97] proposed a Bayesian hierarchical model for approximating the prediction of $f_2(\mathbf{x})$. Based on the KOH model, the *Bayesian hierarchical model* adds a measurement error to the hybrid additive/multiplicative form in Eq. (2) for $w = 1$ and $c \neq 0$, such that

$$f_1(\mathbf{x}) = \delta_1(\mathbf{x}), \tag{19a}$$

$$f_2(\mathbf{x}) = \rho(\mathbf{x}) f_1(\mathbf{x}) + \delta_2(\mathbf{x}) + \varepsilon_2(\mathbf{x}). \tag{19b}$$

Here the adjustment coefficient function $\rho(\mathbf{x})$ is a GP characterized by a scalar mean m_ρ, variance parameter σ_ρ^2, and correlation function $k_\rho(\cdot,\cdot|\boldsymbol{\phi}_\rho)$. $\delta_1(\mathbf{x})$ is a GP characterized by a mean function $m_{\delta,1}(\mathbf{x})$, variance parameter $\sigma_{\delta,1}^2$, and correlation function $k_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1})$. The discrepancy function $\delta_2(\mathbf{x})$ is a GP characterized by scalar mean $m_{\delta,2}$, variance parameter $\sigma_{\delta,2}^2$, and correlation function $k_{\delta,2}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,2})$. The measurement error $\varepsilon_2(\cdot) \sim \mathcal{N}(0, \sigma_{\varepsilon,2}^2)$ is a Gaussian with zero mean and variance $\sigma_{\varepsilon,2}^2$. Thus, hyperparameters underlying the MF surrogate predicting $f_2(\mathbf{x})$ can be encapsulated in a mean parameter vector $\boldsymbol{\phi}_1$, a variance parameter vector $\boldsymbol{\phi}_2$, and a correlation parameter vector $\boldsymbol{\phi}_3$ [97]. The parameter of each vector is then assigned to a pre-specified prior distribution to enable the full Bayesian treatment.

It is computationally expensive to predict f_2 at an unseen input vector using the full Bayesian treatment because this requires the posterior samples of $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, and $\boldsymbol{\phi}_3$, while sampling the posterior of $\boldsymbol{\phi}_3$ is nontrivial as the form of its conditional distribution is irregular. To address this, [97] used the posterior samples of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ for the prediction while fixing $\boldsymbol{\phi}_3$ at its mode value. Unfortunately, finding the mode of $\boldsymbol{\phi}_3$ is still elaborate because it involves integration to estimate the associated unnormalized posterior, which in turn requires the use of sample average approximation for an approximate solution [190]. Although it is one of the important GP-based MF surrogates, the Bayesian hierarchical model may not be a good choice for use of MF BO due to its computational complexity.

E. Composition of Gaussian processes

1. Deep Gaussian processes

A deep Gaussian process (DGP) describes the mapping between the input variables and the output using a composition of GPs [191]. Figure 7 shows a general architecture of DGP with L hidden layers in which a GP relates two consecutive layers. The architecture in Fig. 7 is mathematically described as follows:

$$f(\mathbf{x}) = f_{L-1}\left(\ldots f_1(f_0(\mathbf{x}))\right) + \varepsilon_{L-1}, \tag{20}$$

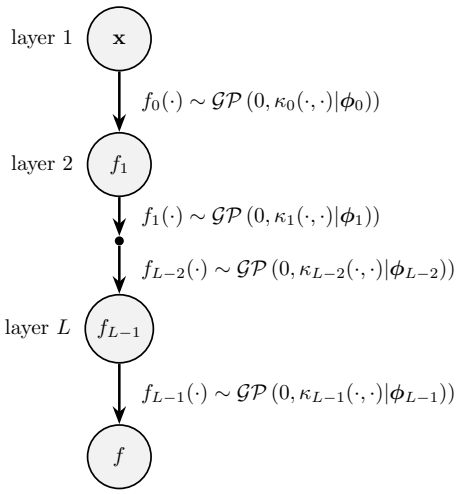

Fig. 7 General architecture of DGP.

where ε_{L-1} is often an additive zero-mean Gaussian noise corresponding to layer L, and

$$f_{l-1}(\cdot) \sim \mathcal{GP}\left(0, \kappa_{l-1}(\cdot,\cdot|\boldsymbol{\phi}_{l-1})\right), \quad l = 1, \ldots, L, \tag{21}$$

where $\kappa_{l-1}(\cdot,\cdot|\boldsymbol{\phi}_{l-1})$ denotes the covariance function parameterized by hyperparameters $\boldsymbol{\phi}_{l-1}$.

Unlike the standard GP, a DGP is capable of handling structured data that encapsulates hierarchical features, enabling it to accurately assess the similarity between pairs of data points properly. Unfortunately, learning the hyperparameters underlying DGPs is rather complicated.

2. Nonlinear auto-regressive model

Based on the KOH model and an output-output mapping technique, [23] proposed the *nonlinear auto-regressive model* for learning complex nonlinear correlations between models of variable fidelity. Let $h_t(\cdot) : \mathbb{R}^{d+1} \mapsto \mathbb{R}$ be a mapping that relates the vector of input and output of fidelity $(t-1)$ to the output of fidelity t. Let $\mathbf{z}_{t-1}(\mathbf{x}) = \left[\mathbf{x}^\intercal, \hat{f}_{t-1}(\mathbf{x})\right]^\intercal$ be the input vector of the surrogate predicting the output of fidelity t, where $\hat{f}_{t-1}(\mathbf{x})$ is the GP posterior corresponding to fidelity $(t-1)$. The nonlinear auto-regressive model reads [23]

$$f_1(\mathbf{x}) = \delta_1(\mathbf{x}), \tag{22a}$$

$$f_t(\mathbf{x}) = h_t\left(\mathbf{z}_{t-1}(\mathbf{x})\right), \quad t = 2, \ldots, T, \tag{22b}$$

$$\mathrm{cov}\left[f_t(\mathbf{x}), \mathbf{z}_{t-1}(\mathbf{x}')|\mathbf{z}_{t-1}(\mathbf{x})\right] = 0, \tag{22c}$$

where $\delta_1(\cdot) \sim \mathcal{GP}\left(0, \kappa_{\delta,1}(\cdot,\cdot|\boldsymbol{\phi}_{\delta,1})\right)$ and the condition in Eq. (22c) is to describe the Markov property in Eq. (6c).

In a probabilistic framework, the mapping $h_t(\cdot)$ in Eq. (22b) is modeled as a GP with zero mean and covariance

function $\kappa_{\mathrm{h},t}(\cdot,\cdot|\boldsymbol{\phi}_{\mathrm{h},t})$, such that

$$h_t(\cdot) \sim \mathcal{GP}\left(0, \kappa_{\mathrm{h},t}(\cdot,\cdot|\boldsymbol{\phi}_{\mathrm{h},t})\right). \tag{23}$$

Here the covariance function $\kappa_{\mathrm{h},t}(\cdot,\cdot|\boldsymbol{\phi}_{\mathrm{h},t})$ treats the input and output variables separately, as

$$\kappa_{\mathrm{h},t}((\mathbf{x},f),(\mathbf{x}',f')|\boldsymbol{\phi}_{\mathrm{h},t}) = \kappa_{\mathrm{x},t}(\mathbf{x},\mathbf{x}'|\boldsymbol{\phi}_{\mathrm{x},t})\kappa_{\mathrm{f},t}(f,f'|\boldsymbol{\phi}_{\mathrm{f},t}) + \kappa_{\delta,t}(\mathbf{x},\mathbf{x}'|\boldsymbol{\phi}_{\delta,t}), \tag{24}$$

where $\boldsymbol{\phi}_{\mathrm{h},t} = [\boldsymbol{\phi}_{\mathrm{x},t}^{\mathsf{T}}, \boldsymbol{\phi}_{\mathrm{f},t}^{\mathsf{T}}, \boldsymbol{\phi}_{\delta,t}^{\mathsf{T}}]^{\mathsf{T}}$, and $\kappa_{\mathrm{x},t}$, $\kappa_{\mathrm{f},t}$, $\kappa_{\delta,t}$ are covariance functions corresponding to $\mathbf{x}$, $f(\mathbf{x})$, and $\delta(\mathbf{x})$ of fidelity t, respectively.

We can see that, the GP surrogate for the model in Eq. (22b) is a DGP when $\delta_1(\mathbf{x})$ and $h_t(\mathbf{x})$ are GPs. However, the number of hyperparameters for this DGP is much smaller than that for the full DPG described in Section IV.E.1 [23].

For $t > 2$, we cannot analytically marginalize the likelihood function, which is described in Eq. (25). As a result, the statistical estimates of f_t at an unseen input vector for $t > 2$ are obtained via the Monte-Carlo integration.

3. Deep multi-fidelity Gaussian process

Using the DGP modeling framework, [112] proposed a *multi-fidelity deep Gaussian process* (MF DGP), where each layer of the DGP corresponds to a fidelity level, i.e., layers $2,\ldots,L$ shown in Fig. 7 correspond to fidelities $1,\ldots,T-1$, respectively. This yields the following MF DGP marginal likelihood for the surrogate that approximates the highest-fidelity output:

$$p(f_T|\mathbf{x}) = \int p(f_T|f_{T-1})\ldots p(f_1|\mathbf{x})\mathrm{d}f_1\ldots\mathrm{d}f_{T-1}, \tag{25}$$

where $p(\cdot)$ denotes the probability density function (PDF).

Since $p(f_T|\mathbf{x})$ is computationally intractable, [112] performed an approximate inference via what is called doubly stochastic variational inference method, which is a variational inference technique using sparse GPs [192].

For variational inference, let $\mathbf{Z}_{t-1} \in \mathbb{R}^{N_t\times(d+t-1)}$, $t = 1,\ldots,T$, denote the set of N_t samples of inducing input variables corresponding to the surrogate associated with fidelity t, and $\mathbf{U}_t \in \mathbb{R}^{N_t}$ denote the corresponding function values, where $p(\mathbf{U}_t) = \mathcal{N}(\boldsymbol{\mu}_{\mathrm{u},t}, \boldsymbol{\Sigma}_{\mathrm{u},t})$. The selection and optimization of inducing input variables can be found in [193]. Note that $\mathbf{Z}_{t-1} \in \mathbb{R}^{N_t\times(d+t-1)}$ because the input vector of surrogate for fidelity t consists of the input vectors and the outputs of surrogates for lower fidelities.

In the doubly stochastic variational inference method, the posterior distribution of $\{\mathbf{U}_t\}_{t=1}^{T}$ is factorized between layers; therefore, the joint posterior $p(\mathbf{F},\mathbf{U})$ is simply the product of the joint posteriors at T fidelities [192]. Since we are interested in large datasets with non-Gaussian likelihoods, we wish to find a variational joint posterior at each fidelity. By further maximizing the lower bound on the marginal likelihood, we obtain an approximate variational joint

posterior at fidelity t, which reads [192]

$$q(\mathbf{F}_t, \mathbf{U}_t) = p(\mathbf{F}_t|\mathbf{U}_t)p(\mathbf{U}_t). \tag{26}$$

Since the joint GP prior $p(\mathbf{F}_t, \mathbf{U}_t)$ and prior $p(\mathbf{U}_t)$ are Gaussians, the conditional $p(\mathbf{F}_t|\mathbf{U}_t)$ is also a Gaussian. Thus, both terms of the variational joint posterior in Eq. (26) are Gaussians. This enables marginalizing $\mathbf{U}_t$ from each fidelity level to obtain an analytical form of the variational marginal $q(\mathbf{F}_t)$, which reads

$$q(\mathbf{F}_t|\boldsymbol{\mu}_{\mathrm{u},t}, \boldsymbol{\Sigma}_{\mathrm{u},t}) = \mathcal{N}(\boldsymbol{\mu}_{\mathrm{f},t}, \boldsymbol{\Sigma}_{\mathrm{f},t}), \tag{27}$$

which is fully coupled within and between layers.

As a result, the marginal associated with the highest fidelity depends only on the marginals associated with the other fidelities, such that

$$q(f_T) = \int \prod_{t=1}^{T} q(\mathbf{F}_t|\boldsymbol{\mu}_{\mathrm{f},t}, \boldsymbol{\Sigma}_{\mathrm{f},t}) \mathrm{d}f_1 \ldots \mathrm{d}f_{T-1}, \tag{28}$$

which has no analytical form, but is computationally tractable via a sampling method.

Some applications of DGP and MF DGP to engineering design optimization can be found in [29, 113]. Nevertheless, it is worth noting that the use of DGPs or MF DGPs as surrogates for performing BO may compromise the beauty of BO because of the intricate nature of inference in DGPs, let alone non-analytical predictions from posterior DGPs.

F. Input-augmentation multi-fidelity Gaussian processes

The GP-based MF surrogates we have seen so far are built in the space of input variables, or an augmented space of input and output variables. An alternative approach, as briefly described in Section II.B, involves the construction of GP-based MF surrogates as functions of both input and fidelity variables. We call this approach input-augmentation GP-based MF modeling, and which has been used in BO predominantly. Once an input-augmentation MF surrogate has been trained, the highest-fidelity output at any set of input variables can be predicted by setting the fidelity-level variables at their highest values.

Depending on how the fidelity variables are described, we classify the input-augmentation methods into two groups. The first group relies on continuous approximations of the fidelity variables (Section IV.F.1), while the second group treats them as categorical variables (Section IV.F.2).

1. Continuous approximations

[116] proposed the method of *continuous approximations* when solving a problem of finding hyperparameters $\mathbf{x}$ for maximizing a validation accuracy $f(\mathbf{x})$. This validation accuracy depends not only on $\mathbf{x}$ but also on the number

of data points t_1, and the number of optimization iterations t_2. A pair of t_1 and t_2 corresponds to a fidelity level. Let $\mathbf{t} = [t_1, t_2]^\intercal$ denote the vector of fidelity variables, where $t_1 \in [1, N_{\max}]$, $t_2 \in [1, I_{\max}]$, and $N_{\max}$ and $I_{\max}$ are threshold values of the number of data points and the number of optimization iterations, respectively. If we use a function $g(\mathbf{t}, \mathbf{x})$ to define the validation accuracy on in a larger space, then $f(\mathbf{x})$, under continuous approximations, is a slice of $g(\mathbf{t}, \mathbf{x})$ at $\mathbf{t}$, such that

$$f(\mathbf{x}) = g(\mathbf{t}, \mathbf{x}). \tag{29}$$

To construct a GP-based MF surrogate, [116] further assigned a GP prior to $g(\mathbf{t}, \mathbf{x})$, as

$$g(\cdot) \sim \mathcal{GP}\left(0, \kappa_{\mathrm{g}}\left(\cdot, \cdot | \boldsymbol{\phi}_{\mathrm{g}}\right)\right), \tag{30}$$

and $\kappa_{\mathrm{g}}\left(\cdot, \cdot | \boldsymbol{\phi}_{\mathrm{g}}\right)$ is defined using the following point-wise product:

$$\kappa_{\mathrm{g}}\left((\mathbf{t}, \mathbf{x}), (\mathbf{t}', \mathbf{x}') | \boldsymbol{\phi}_{\mathrm{g}}\right) = \kappa_{\mathrm{t}}\left(\mathbf{t}, \mathbf{t}' | \boldsymbol{\phi}_{\mathrm{t}}\right) \kappa_{\mathrm{x}}\left(\mathbf{x}, \mathbf{x}' | \boldsymbol{\phi}_{\mathrm{x}}\right), \tag{31}$$

where $\boldsymbol{\phi}_{\mathrm{g}} = [\boldsymbol{\phi}_{\mathrm{t}}^\intercal, \boldsymbol{\phi}_{\mathrm{x}}^\intercal]^\intercal$ is the hyperparameter vector of κ_{g}.

2. Use of non-continuous covariance functions

The construction of GP-based MF surrogates for categorical fidelity variables using the input-augmentation approach requires non-continuous covariance functions $\kappa_{\mathrm{t}}\left(\mathbf{t}, \mathbf{t}' | \boldsymbol{\phi}_{\mathrm{t}}\right)$. Let $\mathbf{t} = [t_i, \ldots, t_{n_{\mathrm{t}}}]^\intercal$ denotes the vector of categorical fidelity variables, where each element t_i, $i = 1, \ldots, n_{\mathrm{t}}$, has l_i categories, i.e., $t_i \in \{t_{i,1}, \ldots, t_{i,l_i}\}$.

[194] proposed the *hypersphere decomposition* method that defines the non-continuous covariance function $\kappa_{\mathrm{t}}\left(\mathbf{t}, \mathbf{t}' | \boldsymbol{\phi}_{\mathrm{t}}\right)$ as a product of the univariate covariance functions associated with individual elements of $\mathbf{t}$. The univariate covariance function associated with each element is formulated by mapping each of its l_i categories onto a point on the surface of l_i-dimensional unit hypersphere. Accordingly, the non-continuous covariance reads

$$\kappa_{\mathrm{t}}\left(\mathbf{t}, \mathbf{t}' | \boldsymbol{\phi}_{\mathrm{t}}\right) = \prod_{i=1}^{n_t} \kappa(t_i, t_i' | \boldsymbol{\phi}_{\mathrm{t},i}), \tag{32a}$$

$$\kappa(t_i, t_i' | \boldsymbol{\phi}_{\mathrm{t},i}) = \sigma_{\mathrm{t},i}^2 \varphi(t_i)^\intercal \varphi(t_i'), \tag{32b}$$

where $\varphi(t_i) = [z_{i,0}, \ldots, z_{i,l_i}]^\intercal$ defines the hypersphere mapping, and $\sigma_{\mathrm{t},i}$ represents the cross-correlation between categories of t_i and t_i'. The determination of $[z_{i,0}, \ldots, z_{i,l_i}]^\intercal$ is detailed in [194] or [195].

[196] used a so-called *compound symmetry* covariance function for $\kappa(t_i, t_i'|\boldsymbol{\phi}_{t,i})$ in Eq. (32b), which assumes a common correlation for all categories. Accordingly, $\kappa(t_i, t_i'|\boldsymbol{\phi}_{t,i})$ reads the following parsimonious form:

$$\kappa(t_i, t_i'|\boldsymbol{\phi}_{\mathrm{t},i}) = \begin{cases} \sigma_{\mathrm{t},i}^2, & \text{for } t_i = t_i', \\ \theta\sigma_{\mathrm{t},i}^2, & \text{for } t_i \neq t_i', \end{cases} \tag{33}$$

where $0 < \theta < 1$.

Alternatively, [197] mapped the vector $\mathbf{t}$ of fidelity variables onto a *latent space* of continuous variables $\mathbf{z}(\mathbf{t})$ using a mapping matrix $\mathbf{A}$. This mapping reads

$$\mathbf{z}(\mathbf{t}) = \boldsymbol{\zeta}(\mathbf{t})\mathbf{A}, \tag{34}$$

where $\boldsymbol{\zeta}(\mathbf{t})$ is the prior vector representation of $\mathbf{t}$ which is defined using either the random initialization or the one-hot encoding technique, and the elements of $\mathbf{A}$ are found via maximum likelihood estimation. Once $\mathbf{z}(\mathbf{t})$ has been established, the non-continuous covariance function $\kappa_{\mathrm{t}}\left(\mathbf{t}, \mathbf{t}'|\boldsymbol{\phi}_t\right)$ is defined as

$$\kappa_{\mathrm{t}}\left(\mathbf{t}, \mathbf{t}'|\boldsymbol{\phi}_{\mathrm{t}}\right) = \kappa_{\mathrm{z}}\left(\mathbf{z}(\mathbf{t}), \mathbf{z}(\mathbf{t}')|\boldsymbol{\phi}_{\mathrm{z}}\right), \tag{35}$$

where $\kappa_{\mathrm{z}}\left(\mathbf{z}(\mathbf{t}), \mathbf{z}(\mathbf{t}')|\boldsymbol{\phi}_{\mathrm{z}}\right)$ is a standard covariance function defined in the continuous space of $\mathbf{z}$.

To further select an appropriate non-continuous covariance function for a specific problem, [195] tested and compared the modeling performances of the hypersphere decomposition, compound symmetry, and latent mapping covariance functions. The main characteristics of these covariance functions were also compared to point out their advantages and disadvantages.

Instead of relying on a point-wise product to define κ_{g} as in Eq. (31), [198] used a sum of covariance functions to define κ_{g} when maximizing an objective function $f(\mathbf{x}) = g(t, \mathbf{x}) - \delta_t(\mathbf{x})$. Here $t \in \mathcal{T} = \{t_1, \ldots, t_T\}$ is a categorical variable indicating the noisy information source, and therefore can be seen a fidelity variable. $g(t, \mathbf{x})$ is the noisy value of $f(\mathbf{x})$ when observing t at $\mathbf{x}$, thus $g(t_T, \mathbf{x}) = f(\mathbf{x})$ is the noise-free observation. $\delta_t(\mathbf{x})$ is the discrepancy function defined as a GP independent of f, such that $\delta_t(\cdot) \sim \mathcal{GP}(m_{\delta,t}(\cdot), \kappa_{\delta,t}\left(\cdot, \cdot|\boldsymbol{\phi}_{\delta,t}\right))$. The objective function is $f(\cdot) \sim \mathcal{GP}(m_{\mathrm{f}}(\cdot), \kappa_{\mathrm{x}}\left(\cdot, \cdot|\boldsymbol{\phi}_{\mathrm{x}}\right))$. Accordingly, $\kappa_{\mathrm{g}}\left((t, \mathbf{x}), (t', \mathbf{x}')|\boldsymbol{\phi}_{\mathrm{g}}\right)$ can be derived from $g(t, \mathbf{x}) = f(\mathbf{x}) + \delta_t(\mathbf{x})$, as

$$\kappa_{\mathrm{g}}\left((t, \mathbf{x}), (t', \mathbf{x}')|\boldsymbol{\phi}_{\mathrm{g}}\right) = \mathrm{cov}[g(t, \mathbf{x}), g(t', \mathbf{x}')] = \kappa_{\mathrm{x}}\left(\mathbf{x}, \mathbf{x}'|\boldsymbol{\phi}_{\mathrm{x}}\right) + \mathbb{1}_{t,t'}\kappa_{\delta,t}\left(\mathbf{x}, \mathbf{x}'|\boldsymbol{\phi}_{\delta,t}\right)), \tag{36}$$

where $\mathbb{1}_{t,t'}$ with $t, t' \in \mathcal{T}$ denotes the Kronecker delta, and κ_{x} and $\kappa_{\delta,t}$ are standard parameterized covariance functions.

V. Acquisition functions

An acquisition function formulated in each iteration of BO maps one of our preferences for the next design point to each point in the design variable space. Here, our preferences under incomplete knowledge about the objective function include the improvement in the objective function value, the gain in information on the true minimizer, and the gain in information on the true minimum. Maximizing the acquisition function, therefore, guides BO toward a new design point with the highest score of solution improvement. In this section, we first review notable popular acquisition functions of the generic BO (Section V.A). We then describe several ways of modifying these acquisition functions for use of MF BO (Section V.B), followed by a brief discussion on the portfolio of different design points when maximizing a set of multiple acquisition functions (Section V.C). We finally discuss several optimization algorithms to maximize the acquisition functions (Section V.D) and recommend open-source software to implement the generic BO (Section V.E).

A. Acquisition functions for Bayesian optimization

Two classes of acquisition functions of the generic BO include one-step look-ahead (myopic) and multi-step look-ahead (or non-myopic). *One-step look-ahead acquisition functions*, which are predominantly used in BO literature, select a new design point using a utility measure, without forecasting the potential impact of all future selections beyond the immediate next selection. They can be further classified improvement-based (Section V.A.1), optimistic (Section V.A.2), and information-based (Section V.A.3) acquisition functions. In contrast, *multi-step look-ahead acquisition functions* (Section V.A.4) consider the impact of future selections on the decision of next design points by adding a new term to one-step look-ahead acquisition functions. This aims to mitigate a limitation of one-step look-ahead acquisition functions that they tend to prefer exploitation over exploration [199].

Assume that we are at iteration k of BO and wish to select a new design point $\mathbf{x}^k$ by maximizing an acquisition function $\alpha(\mathbf{x})$. What we currently know to formulate $\alpha(\mathbf{x})$ in Line 9 of Algorithm 1 are the GP posterior $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ constructed from the available (HF) data $\mathcal{D}^{k-1}$ and, under noise-free observations, the best solution $\{\mathbf{x}_{\min}, f_{\min}\}$ we found in $\mathcal{D}^{k-1}$. Let $\mathcal{D}^k$ be the data after iteration k has completed, i.e., $\mathcal{D}^k = \mathcal{D}^{k-1} \cup \{\mathbf{x}^k, f_{\mathrm{H}}(\mathbf{x}^k)\}$.

1. Improvement-based acquisition functions

Improvement-based acquisition functions arise from a thought experiment in which we expect that the objective function at the new design point is better (i.e., smaller) than the best-observed objective function value $f_{\min}$. To describe this, we define the following solution improvement measure at iteration k of BO:

$$I(\mathbf{x}) = \max\{f_{\min} - f(\mathbf{x}), 0\}. \tag{37}$$

Probability of improvement (PI) [200], one of the earliest improvement-based acquisition functions, measures the

probability that $I(\mathbf{x})$ is greater than a non-negative target value τ. This is equivalent to the chance of having a solution improvement. Conditioning it on the current GP posterior, PI can be written in the following analytical form:

$$\alpha(\mathbf{x}) = \mathbb{P}\left[I(\mathbf{x}) > \tau | \hat{f}_{\mathrm{H}}^{k}(\mathbf{x})\right] = \Phi\left(\frac{f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x}) - \tau}{\sigma_{\mathrm{f}}^{k}(\mathbf{x})}\right), \tag{38}$$

where $\tau \geq 0$ is the improvement target, $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF), and $\mu_{\mathrm{f}}^{k}(\mathbf{x})$ and $\sigma_{\mathrm{f}}^{k}(\mathbf{x})$ are the predictive mean and standard deviation of $\hat{f}_{\mathrm{H}}^{k}(\mathbf{x})$ given in Eqs. (A.6) and (A.7), respectively.

While maximizing PI tends to reduce the solution over time, it does not necessarily result in a substantial improvement in the objective function. This is attributed to the fact that PI is not a direct quantitative measure of an improvement in the objective function [2]. Additionally, we should select a value of the improvement target carefully to obtain a desired solution improvement because we favor exploitation for a small τ or exploration for a large τ. A data-driven approach to choosing τ values can be found in [182].

Expected improvement (EI) [164, 165] is an acquisition function that measures the solution improvement quantitatively. As its name suggests, EI calculates the expected value of $I(\mathbf{x})$, given the current GP posterior. Mathematically, EI reads

$$\alpha(\mathbf{x}) = \mathbb{E}\left[I(\mathbf{x}) | \hat{f}_{\mathrm{H}}^{k}(\mathbf{x})\right] = \left(f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x})\right) \Phi\left(\frac{f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x})}{\sigma_{\mathrm{f}}^{k}(\mathbf{x})}\right) + \sigma_{\mathrm{f}}^{k}(\mathbf{x}) \phi\left(\frac{f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x})}{\sigma_{\mathrm{f}}^{k}(\mathbf{x})}\right), \tag{39}$$

where $\phi(\cdot)$ denotes the standard normal PDF. This analytical form is derived using integration by parts [2, 165].

EI provides a way to conceptualize the balance between exploitation and exploration in optimization. In its expression, the first term of EI embodies exploitation, guiding the search toward a new design point with a high probability of improvement. Its second term embodies exploration, directing the search to regions where there is considerable uncertainty in the prediction of objective function. However, EI does not allow a direct control over the exploitation-exploration balance when necessary. For example, it is preferable to bias exploitation if the objective function tends to be unimodal. Conversely, exploration can work well if the objective function is extremely multimodal [201].

To gain control over the exploitation-exploration balance, [201] introduced the *weighted expected improvement* (WEI) as a weighted sum of the two terms of EI, such that

$$\alpha(\mathbf{x}) = w\left(f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x})\right) \Phi\left(\frac{f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x})}{\sigma_{\mathrm{f}}^{k}(\mathbf{x})}\right) + (1-w)\sigma_{\mathrm{f}}^{k}(\mathbf{x}) \phi\left(\frac{f_{\min} - \mu_{\mathrm{f}}^{k}(\mathbf{x})}{\sigma_{\mathrm{f}}^{k}(\mathbf{x})}\right), \tag{40}$$

where the weighting factor $w \in [0, 1]$.

If the new design point is selected by maximizing WEI in some set of candidate solutions, then this only results in selecting $\mathbf{x}^k$ in the maximal non-dominated set for a relatively small range of w, regardless the fact that w can vary over $[0, 1]$ [202]. By carefully examining this phenomenon, [202] recommended restricting the values of w to the interval

$[0.185, 0.5]$, which guarantees the selection of a new design point that lies on the Pareto frontier of available solutions.

Knowledge gradient (KG) [203] is an improvement-based acquisition function that is closely related to EI. By introducing a departure from one of the key assumptions underlying EI that the observations are noise-free, KG operates independently of the best-observed objective function value. This is justified by considering that the best solution might possess some level of uncertainty.

Let $\mu_{\mathrm{f}}^{k+1}(\cdot)$ be the unknown predictive posterior mean obtained from $\mathcal{D}^k = \mathcal{D}^{k-1} \cup \{\mathbf{x}^k, f_{\mathrm{H}}(\mathbf{x}^k)\}$. Let $\mathbf{x}_\star^k$ denote the minimizer of the predictive posterior mean $\mu_{\mathrm{f}}^k(\cdot)$. The KG acquisition function for finding $\mathbf{x}^k$ is defined as [203]

$$\alpha(\mathbf{x}) = \mathbb{E}\left[\mu_{\mathrm{f}}^k(\mathbf{x}_\star^k) - \mu_{\mathrm{f}}^{k+1}(\mathbf{x})\right], \tag{41}$$

which integrates over all possible points $\{\mathbf{x}^k, f_{\mathrm{H}}(\mathbf{x}^k)\}$ for a given $\mathbf{x}^k$ under the posterior predictive PDF of $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$. Thus, the simplest way to estimate KG at a specific value $\mathbf{x}$ is via the Monte-Carlo integration.

Rather than balancing the exploitation and exploration, KG balances the so-called influence of alternative $\mathbf{x}$ and variance, where the benefit of variance is the same as that of exploration [203].

2. Optimistic acquisition function

While the solution to the exploitation-exploration trade-off in multi-armed bandit problems is almost intractable, an intuitive approach is to use confidence bounds to handle this trade-off [204, 205]. This is underpinned by the principle of optimism in the face of uncertainty, which takes greedy actions based on optimistic estimates of their rewards [25].

In the context of GP-based optimization, [183] proposed the *GP upper confidence bound* (GP-UCB) to maximize the sum of rewards while handling the exploitation-exploration trade-off. Based on GP-UCB, we can define the following *GP negative lower confidence bound* (GP-NLCB) as an optimistic acquisition function for minimization problems:

$$\alpha(\mathbf{x}) = -\left[\mu_{\mathrm{f}}^k(\mathbf{x}) - \sqrt{\beta^k}\sigma_{\mathrm{f}}^k(\mathbf{x})\right], \tag{42}$$

where $\sqrt{\beta^k} \geq 0$ is the tuning parameter to control the exploitation-exploration trade-off.

The convergence proof by [183] is based on the scheduled values of $\sqrt{\beta^k}$. In particular, $\sqrt{\beta^k}$ increases as logarithm of the number of past evaluations of the objective function. This means, the search biases toward exploration after each iteration of BO.

3. Information-based acquisition functions

Information-based acquisition functions focus on gaining the solution information from the posterior PDF of unknown minimizer $\mathbf{x}_\star$ or the posterior PDF of unknown minimum $f(\mathbf{x}_\star)$. Two policies in this class include Thompson sampling and entropy search.

Thompson sampling (TS), also known as posterior sampling or randomized probability matching, selects the next action in a multi-armed bandit problem by maximizing the reward with respect to a reward function randomly drawn from the associated posterior [see e.g., 206–208]. More specifically, TS randomly selects each action in proportion to the posterior PDF of the optimal action [209].

In the context of BO, instead of maximizing an explicit acquisition function, TS selects the new design point using a randomized strategy, which simulates a function $q_{\mathrm{H}}^k(\mathbf{x})$ from the current GP posterior $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ and then finds the minimizer of this function. This two-stage implementation is indeed the random generation of the next design point from the posterior PDF of unknown minimizer $\mathbf{x}_\star$. To sample $q_{\mathrm{H}}^k(\mathbf{x})$ from a GP posterior with a stationary covariance function, an approach is via GP spectral sampling that approximates the GP prior using a covariance function by a Bayesian linear model using random features [210]. Then, the Bayesian linear model that corresponds to the posterior covariance function is considered a sample of the GP posterior $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$. In another approach, the GP posterior $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ is defined as the sum of a weight-space prior and a function-space update [211]. The weight-space prior and function-space update can be separately discretized using Fourier basis functions and canonical basis functions, respectively. A pseudo-code for TS is given in Algorithm 2.

Algorithm 2 Sequential Thompson sampling.

1: **Input:** $\mathcal{X}$, K, N;
2: Generate N samples of $\mathbf{x}$;
3: **for** $i = 1 : N$ **do**
4: $\quad f^i \leftarrow f_{\mathrm{H}}(\mathbf{x}^i)$; ▷ Costly step
5: **end for**
6: $\mathcal{D}^0 \leftarrow \{\mathbf{x}^i, f^i\}_{i=1}^N$;
7: $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow \min\{f^i,\, i = 1, \ldots, N\}$;
8: **for** $k = 1 : K$ **do**
9: $\quad$ Construct $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ from $\mathcal{D}^{k-1}$;
10: $\quad$ Sample $q_{\mathrm{H}}^k(\mathbf{x})$ from $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$;
11: $\quad \mathbf{x}^k \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\ q_{\mathrm{H}}^k(\mathbf{x})$ s.t. $\mathbf{x} \in \mathcal{X}$;
12: $\quad f^k \leftarrow f_{\mathrm{H}}(\mathbf{x}^k)$; ▷ Costly step
13: $\quad \mathcal{D}^k \leftarrow \mathcal{D}^{k-1} \cup \{\mathbf{x}^k, f^k\}$;
14: $\quad \{\mathbf{x}_{\min}, f_{\min}\} \leftarrow \min\{f_{\min}, f^k\}$;
15: **end for**
16: **return** $\{\mathbf{x}_{\min}, f_{\min}\}$.

Algorithm 2 implies that if the posterior $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ is highly uncertain, TS tends to sample many different functions $q_{\mathrm{H}}^k(\mathbf{x})$ in Line 10, which turns out that TS favors exploration. As the uncertainty reduces, it starts exploiting the knowledge about the true objective function. Thus, TS handles the exploitation-exploration trade-off naturally, even though it is not known to maximize any specific acquisition function. In fact, the exploitation ability of TS is inferior because of its inherent randomness [209].

Some theoretical results allow for establishing the connection between TS and GP-UCB via the translation from

regret bounds developed for GP-UCB into Bayesian regret bounds developed for TS [212]. It is also possible to sample multiple functions $q_{\mathrm{H}}^{k}(\mathbf{x})$ (Line 10 of Algorithm 2) in parallel to select a batch of new design points [213]. Nevertheless, TS may address the following four problem features inadequately: problems that do not require exploration, problems that do not require exploitation, problems that are time-sensitive, and problems that require careful assessment of information gain [214].

Information-based acquisition functions with entropy search (ES) policy select the next design points to gain information in the true minimizer $\mathbf{x}_\star$ or in the true minimum $f_{\mathrm{H}}(\mathbf{x}_\star)$. The former includes information approach optimization algorithm, entropy search, and predictive entropy search. The latter includes maximum-value entropy search.

Information approach optimization algorithm (IAGO) [215] and *entropy search* (ES) [184] define the acquisition function based on the following location-information loss:

$$\lambda(\mathcal{D}^k) = \mathbb{H}(\mathbf{x}_\star|\mathcal{D}^k). \tag{43}$$

Here $\mathcal{D}^k$ denotes the training data obtained after adding $\{\mathbf{x}^k, f_{\mathrm{H}}(\mathbf{x}^k)\}$ to $\mathcal{D}^{k-1}$. $\mathbb{H}(\cdot)$ is the entropy of the random variable inside the parentheses, measuring the spread (uncertainty) of its PDF. Low entropy corresponds to high information gain. Thus, if a new design point $\mathbf{x}^k$ is selected by minimizing $\lambda(\mathcal{D}^k)$, then it minimizes the entropy of $\mathbf{x}_\star$ given the data in the next iteration.

Let $\mathbf{U}$ denote a vector of continuous random variables distributed according to $p(\mathbf{u})$, and $\mathbf{V}$ denote a vector of discrete random variables, which take values in a discrete set $\mathcal{V}$. The mathematical descriptions of the information entropy values of $\mathbf{U}$ and $\mathbf{V}$ are

$$\mathbb{H}(\mathbf{U}) = -\int p(\mathbf{u}) \log p(\mathbf{u}) \mathrm{d}\mathbf{u}, \tag{44a}$$

$$\mathbb{H}(\mathbf{V}) = -\sum_{\mathbf{v}\in\mathcal{V}} p(\mathbf{V}=\mathbf{v}) \log p(\mathbf{V}=\mathbf{v}), \tag{44b}$$

where $\mathbf{u}$ and $\mathbf{v}$ are realizations of $\mathbf{U}$ and $\mathbf{V}$, respectively.

Since $\lambda(\mathcal{D}^k)$ depends on the unknown design point $\mathbf{x}^k$, we define IAGO and ES using the expected value of $\lambda(\mathcal{D}^k)$, such that

$$\alpha(\mathbf{x}) = -\mathbb{E}_{f_{\mathrm{H}}}\left[\lambda(\mathcal{D}^k)\right] = -\int \mathbb{H}(\mathbf{x}_\star|\mathcal{D}^k) p(f_{\mathrm{H}}|\mathbf{x}, \mathcal{D}^{k-1}) \mathrm{d}f_{\mathrm{H}} = -\mathbb{E}_{f_{\mathrm{H}}}\left[\mathbb{H}(\mathbf{x}_\star|\mathbf{x}, f_{\mathrm{H}}, \mathcal{D}^{k-1})\right], \tag{45}$$

where " $-$ " reformulates the minimization of entropy to the maximization of acquisition function, $p(f_{\mathrm{H}}|\mathbf{x}, \mathcal{D}^{k-1})$ denotes the PDF of f_{H} conditioned on $\mathbf{x}$ and $\mathcal{D}^{k-1}$, and $\mathbb{E}_{f_{\mathrm{H}}}\left[\mathbb{H}(\mathbf{x}_\star|\mathbf{x}, f_{\mathrm{H}}, \mathcal{D}^{k-1})\right]$ is the expected entropy of $\mathbf{x}_\star$ given

that the new design point and the corresponding value of the objective function has been added to $\mathcal{D}^{k-1}$.

Since computing the entropy for a continuous $\mathbf{x}_\star$ is analytical intractable, discretizing Eq. (45) makes the computation tractable. While IAGO and ES share the same acquisition function form, they differ in their ways of discretizing Eq. (45).

Predictive entropy search (PES) [210] is another information-based acquisition function derived from $\lambda(\mathcal{D}^k)$. PES arises from an important observation that the maximizer of $-\mathbb{E}_{f_\mathrm{H}}\left[\mathbb{H}(\mathbf{x}_\star|\mathbf{x}, f_\mathrm{H}, \mathcal{D}^{k-1})\right]$ is identical to that of $\mathbb{H}(\mathbf{x}_\star|\mathcal{D}^{k-1}) - \mathbb{E}_{f_\mathrm{H}}\left[\mathbb{H}(\mathbf{x}_\star|\mathbf{x}, f_\mathrm{H}, \mathcal{D}^{k-1})\right]$. This is due to the fact that $\mathbb{H}(\mathbf{x}_\star|\mathcal{D}^{k-1})$ does not depend on $\mathbf{x}^k$. By further leveraging the mutual information $\mathcal{I}(\mathbf{x}_\star; f_\mathrm{H})$ between $\mathbf{x}_\star$ and f_H, we obtain the following relation [216]:

$$\mathcal{I}(\mathbf{x}_\star; f_\mathrm{H}) = \mathbb{H}(\mathbf{x}_\star|\mathcal{D}^{k-1}) - \mathbb{E}_{f_\mathrm{H}}\left[\mathbb{H}(\mathbf{x}_\star|\mathbf{x}, f_\mathrm{H}, \mathcal{D}^{k-1})\right] = \mathbb{H}(f_\mathrm{H}|\mathbf{x}, \mathcal{D}^{k-1}) - \mathbb{E}_{\mathbf{x}_\star}\left[\mathbb{H}(f_\mathrm{H}|\mathbf{x}_\star, \mathbf{x}, \mathcal{D}^{k-1})\right], \tag{46}$$

where $\mathbb{E}_{\mathbf{x}_\star}\left[\mathbb{H}(f_\mathrm{H}|\mathbf{x}_\star, \mathbf{x}, \mathcal{D}^{k-1})\right]$ is the expected entropy of f_H given $\mathbf{x}$, $\mathbf{x}_\star$, and $\mathcal{D}^{k-1}$. Based on this, PES makes use of the following acquisition function:

$$\alpha(\mathbf{x}) = \mathbb{H}(f_\mathrm{H}|\mathbf{x}, \mathcal{D}^{k-1}) - \mathbb{E}_{\mathbf{x}_\star}\left[\mathbb{H}(f_\mathrm{H}|\mathbf{x}_\star, \mathbf{x}, \mathcal{D}^{k-1})\right]. \tag{47}$$

The calculation of PES is much easier than that of ES because the first and second terms in Eq. (47) only require evaluating the entropy of univariate Gaussians. Nevertheless, the calculation of the second term is still intricate as it necessitates marginalization over the multivariate posterior $p(\mathbf{x}_\star|\mathcal{D}^{k-1})$.

Rather than using the entropy of unknown minimizer $\mathbf{x}_\star$, *maximum-value entropy search* (MES) [217] formulates its acquisition function using the entropy of unknown minimum $f_\mathrm{H}(\mathbf{x}_\star)$. In fact, MES modifies PES in Eq. (47) by replacing the minimizer $\mathbf{x}_\star$ of the second term with the minimum $f_\star = f_\mathrm{H}(\mathbf{x}_\star)$. Accordingly, MES reads

$$\alpha(\mathbf{x}) = \mathbb{H}(f_\mathrm{H}|\mathbf{x}, \mathcal{D}^{k-1}) - \mathbb{E}_{f_\star}\left[\mathbb{H}(f_\mathrm{H}|f_\star, \mathbf{x}, \mathcal{D}^{k-1})\right]. \tag{48}$$

MES has advantages in implementation over PES as its second term only requires marginalization over the univariate posterior $p(f_\star|\mathcal{D}^{k-1})$.

4. Multi-step look-ahead acquisition functions

Although they are simple and computationally efficient, one-step look-ahead acquisition functions tend to favor exploitation [199]. To address this, multi-step look-ahead acquisition functions consider the influence of future selections on the decision of next design points [see e.g., 218, 219]. However, it still remains challenging to handle an exact multi-step look-ahead problem because it requires marginalizing uncertain objective function value and design point location in each step [220]. What we can expect is to use either two-step look-ahead acquisition functions via the Monte-Carlo integration [221] or multi-step look-ahead acquisition functions via approximation techniques, such as

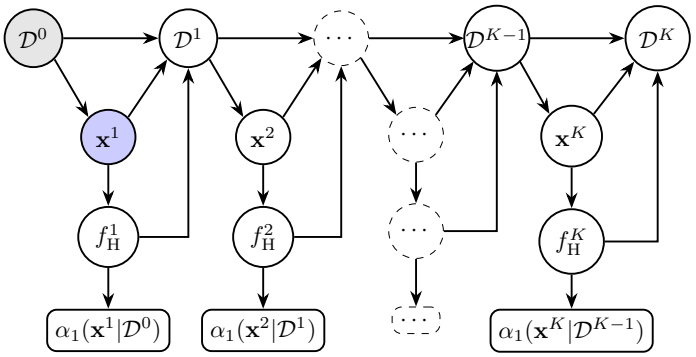

Fig. 8 Illustration of a K-step look-ahead problem. Node $\mathcal{D}^0$ represents the data we observed so far, and node $\mathbf{x}^1$ is the design point we wish to find. Finding $\mathbf{x}^1$ is influenced by future selections of $\mathbf{x}^2, \ldots, \mathbf{x}^K$.

rollout [222, 223], GLASSES [220], and multi-step [224]. In the following, we briefly describe some multi-step look-ahead acquisition functions. For the detailed implementation, the reader is encouraged to refer to the corresponding references.

Figure 8 describes BO as a multi-stage decision problem for finding $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^K$. Our objective is to select $\mathbf{x}^1$ based on the information available in the observed data $\mathcal{D}^0$. However, this selection is influenced by future selections of $\mathbf{x}^2, \ldots, \mathbf{x}^K$.

For any data $\mathcal{D}^k$, $k = 0, \ldots, K$, we define

$$u(\mathcal{D}^k) = \min_{(\mathbf{x}, f_{\mathrm{H}}) \in \mathcal{D}^k} f_{\mathrm{H}}(\mathbf{x}), \tag{49}$$

which returns the best solution found among the data points of $\mathcal{D}^k$. Based on $u(\cdot)$, we can define the one-step look-ahead acquisition function $\alpha_1(\mathbf{x}^1|\mathcal{D}^0)$ for selecting $\mathbf{x}^1$ as

$$\alpha_1(\mathbf{x}^1|\mathcal{D}^0) = \mathbb{E}_{f_{\mathrm{H}}^1}\left[\max\left(u(\mathcal{D}^0) - u(\mathcal{D}^1), 0\right)|\mathbf{x}^1, \mathcal{D}^0\right], \tag{50}$$

where $f_{\mathrm{H}}^1 = f_{\mathrm{H}}(\mathbf{x}^1)$ and $\alpha_1(\mathbf{x}^1|\mathcal{D}^0)$ is EI given in Eq. (39).

The Bellman's principle of optimality allows the computation of a *K-step look-ahead acquisition function* recursively, such that [225]

$$\alpha_K(\mathbf{x}^1|\mathcal{D}^0) = \alpha_1(\mathbf{x}^1|\mathcal{D}^0) + \mathbb{E}_{f_{\mathrm{H}}^1}\left[\max_{\mathbf{x}^2} \alpha_{K-1}(\mathbf{x}^2|\mathcal{D}^1)\right], \tag{51}$$

where $\alpha_K(\cdot)$ and $\alpha_{K-1}(\cdot)$ represent the K- and $(K-1)$-step look-ahead acquisition functions, respectively. Since $\alpha_1(\mathbf{x}^1|\mathcal{D}^0)$ favors exploitation, we can view the first and second terms of Eq. (51) as exploitation and exploration terms, respectively.

[224] used the following K-step look-ahead acquisition function for finding $\mathbf{x}^1$:

$$\alpha_K(\mathbf{x}^1|\mathcal{D}^0) = \alpha_1(\mathbf{x}^1|\mathcal{D}^0) + \mathbb{E}_{f_{\mathrm{H}}^1}\left[\max_{\mathbf{x}^2}\left(\alpha_1(\mathbf{x}^2|\mathcal{D}^1) + \mathbb{E}_{f_{\mathrm{H}}^2}\left[\max_{\mathbf{x}^3}\alpha_1(\mathbf{x}^3|\mathcal{D}^2) + \dots\right]\right.\right., \tag{52}$$

where $f_{\mathrm{H}}^k = f_{\mathrm{H}}(\mathbf{x}^k)$, $k = 1, \dots, K$.

[221] proposed *two-step look-ahead acquisition function* $\alpha_2(\mathbf{x}^1|\mathcal{D}^0)$ that, via the Monte-Carlo integration, can be computationally efficient. Accordingly, $\alpha_2(\mathbf{x}^1|\mathcal{D}^0)$ reads

$$\alpha_2(\mathbf{x}^1|\mathcal{D}^0) = \alpha_1(\mathbf{x}^1|\mathcal{D}^0) + \mathbb{E}_{f_{\mathrm{H}}^1}\left[\max_{\mathbf{x}^2}\alpha_1(\mathbf{x}^2|\mathcal{D}^1)\right]. \tag{53}$$

[220] introduced *GLASSES* acquisition function by assuming a joint PDF of the future selections from which a batch of design points can be generated in each iteration. GLASSES reads

$$\alpha_K(\mathbf{x}^1|\mathcal{D}^0) \approx \alpha_1(\mathbf{x}^1|\mathcal{D}^0) + \mathbb{E}_{f_{\mathrm{H}}^1}\left[\Lambda_{K-1}(\mathbf{X}_{\mathrm{x}}|\mathcal{D}^1)\right], \tag{54}$$

where $\mathbf{X}_{\mathrm{x}} \in \mathbb{R}^{(K-1)\times d}$ is a matrix whose rows represent the future design points $\mathbf{x}^2, \dots, \mathbf{x}^K$. $\Lambda_{K-1}(\mathbf{X}_{\mathrm{x}}|\mathcal{D}^1)$ is a batch value function for the matrix $\mathbf{X}_{\mathrm{x}}$ conditioned on the unknown data $\mathcal{D}^1$, such that

$$\Lambda_{K-1}(\mathbf{X}_{\mathrm{x}}|\mathcal{D}^1) = \mathbb{E}_{f_{\mathrm{H}}^2,\dots,f_{\mathrm{H}}^K}\left[\max\left(u(\mathcal{D}^1) - u(\mathcal{D}^K), 0\right)|\mathbf{X}_{\mathrm{x}}, \mathcal{D}^1\right]. \tag{55}$$

Alternatively, *rollout strategies* [222, 223] formulate the selection of new design points as a Markov decision process. Thus, the multi-step look-ahead acquisition function is the expected total reward of this Markov decision process. Then, the maximum of such reward can be found by approximate dynamic programming [226].

Algorithm 3 MF BO, no-fidelity consideration.

Input: $\mathcal{X}$, K, $\mathcal{D}_t$ $(t = 1, \dots, T)$;
$\mathcal{D}^0 \leftarrow \mathcal{D}_1 \cup \dots \cup \mathcal{D}_T$;
$\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow u(\mathcal{D}^0)$;
for $k = 1 : K$ **do**
  Construct MF surrogate $\hat{f}_{\mathrm{H}}^k(\mathbf{x})$ from $\mathcal{D}^{k-1}$;
  Formulate MF acquisition function $\alpha(\mathbf{x})$;
  $\mathbf{x}^k \leftarrow \underset{\mathbf{x}}{\operatorname{argmax}}\ \alpha(\mathbf{x})$ s.t. $\mathbf{x} \in \mathcal{X}$;
  $\mathcal{D}_t \leftarrow \mathcal{D}_t \cup \{\mathbf{x}^k, f_t(\mathbf{x}^k)\}$, $t = 1, \dots, T$;
  $\mathcal{D}^k \leftarrow \mathcal{D}_1 \cup \dots \cup \mathcal{D}_T$;
  $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow u(\mathcal{D}^k)$;
end for
return $\{\mathbf{x}_{\min}, f_{\min}\}$.

B. Acquisition functions considering fidelities

In MF BO, we replace $\hat{f}_{\mathrm{H}}^{k}(\mathbf{x})$ in Line 9 of Algorithm 1 with one of the MF surrogates described in Section IV. This raises a further question of how to incorporate the information about fidelities, i.e., $t \in \{1, \ldots, T\}$, into the acquisition function so that we can select a new design point and an appropriate computational model for estimating the corresponding objective function value.

The use of different MF surrogates and/or different acquisition functions of the generic BO results in a variety of MF acquisition functions. Nevertheless, we can categorize the MF acquisition functions into the following three groups of approaches:

- No-fidelity consideration (Section V.B.1).
- Heuristic approach (Section V.B.2).
- Sequential selection (Section V.B.3).

1. No-fidelity consideration

The acquisition functions of this approach only depend on design variables $\mathbf{x}$. Once the new design point $\mathbf{x}^k$ has been found, it is fed to T computational models associated with T fidelities for predictions of the corresponding objective function value. These predictions are used for updating the current solution and MF surrogate. Algorithm 3 shows a pseudo-code for MF BO without considering fidelities. $u(\mathcal{D}^k)$, $k = 0, \ldots, K$, in Lines 3 and 10 is defined in Eq. (49). The MF acquisition function in Line 6 can be any acquisition function of the generic BO. For example, [16] used EI with $f_{\min}$ selected from the highest fidelity, i.e., $\mathcal{D}^k$ in Lines 3 and 10 was fixed at $\mathcal{D}^K$. [128] formulated EI, but with $f_{\min}$ defined as the best objective value among those from all fidelities.

Algorithm 4 MF BO using heuristic approach.

1: **Input:** $\mathcal{X}$, K, $\mathcal{D}_t$ $(t = 1, \ldots, T)$;
2: $\mathcal{D}^0 \leftarrow \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_T$;
3: $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow u(\mathcal{D}^0)$;
4: **for** $k = 1 : K$ **do**
5: Construct MF surrogate $\hat{f}_{\mathrm{H}}^{k}(\mathbf{x})$ from $\mathcal{D}^{k-1}$;
6: Formulate MF acquisition function $\alpha(\mathbf{x}, t)$;
7: **for** $t = 1 : T$ **do**
8: $\{\mathbf{x}_t^k, t, \alpha_t^k\} \leftarrow \max \alpha(\mathbf{x}, t)$ s.t. $\mathbf{x} \in \mathcal{X}$;
9: **end for**
10: $\{\mathbf{x}^k, t\} \leftarrow \max\{\alpha_t^k,\ t = 1, \ldots, T\}$
11: $\mathcal{D}_t \leftarrow \mathcal{D}_t \cup \{\mathbf{x}^k, f_t(\mathbf{x}^k)\}$;
12: $\mathcal{D}^k \leftarrow \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_T$;
13: $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow u(\mathcal{D}^k)$;
14: **end for**
15: **return** $\{\mathbf{x}_{\min}, f_{\min}\}$.

2. Heuristic approach

The acquisition functions of this approach depend on both $\mathbf{x}$ and t. They are often derived by using auxiliary functions to modify one of the one-step look-ahead acquisition functions of the generic BO. These auxiliary functions consider the computational cost of each fidelity and/or how the selection of each fidelity for handling the new design point affects the accuracy improvement of MF surrogate. Algorithm 4 shows a pseudo-code for MF BO using the heuristic approach. In Lines 7–10, a new design point $\mathbf{x}^k$ is found for each enumerated value of t, then the pair $\{\mathbf{x}^k, t\}$ that provides the best acquisition function value $\alpha_t^k = \alpha(\mathbf{x}^k, t)$ is selected for the next iteration. Due to its simplicity, the heuristic approach has been widely used in engineering design optimization [see e.g., 15, 134, 141, 153, 198].

Let $\mu_{f_\mathrm{H}}^k(\mathbf{x})$ and $\sigma_{f_\mathrm{H}}^{2,k}(\mathbf{x})$ denote the predictive mean and predictive variance of the MF surrogate $\hat{f}_\mathrm{H}^k(\mathbf{x})$. Let $c(t)$, $t = 1, \ldots, T$, denote the computational cost associated with the model of fidelity t.

One of the first heuristic MF acquisition functions by [15] was defined as the product of the so-called augmented EI (AEI) [227], developed for the generic BO with noisy objective functions, and two auxiliary functions $\alpha_1(\mathbf{x}, t)$ and $\alpha_2(t)$ for considering the influence of fidelities. This heuristic MF acquisition function reads

$$\alpha(\mathbf{x}, t) = \mathrm{AEI}(\mathbf{x})\alpha_1(\mathbf{x}, t)\alpha_2(t). \tag{56}$$

Here $\mathrm{AEI}(\mathbf{x})$ becomes $\mathrm{EI}(\mathbf{x})$ when the objective function is noise-free, and $\alpha_1(\mathbf{x}, t)$ is the correlation between the posterior PDF of MF surrogate t and that of MF surrogate T. Since $\alpha_1 = 1$ when $t = T$, α_1 tends to promote high fidelities for maximizing $\alpha(\mathbf{x}, t)$. Meanwhile, $\alpha_2(t) = c(T)/c(t)$ is the ratio between the computational cost per model on the highest fidelity T and that on the fidelity t. In contrast to α_1, α_2 favors low fidelities for maximizing $\alpha(\mathbf{x}, t)$.

Attempts to modify Eq. (56) are problem-dependent. One of them is to replace the standard deviation constituting $\mathrm{AEI}(\mathbf{x})$, which depends on $\mathbf{x}$ only, with a new standard deviation that is a function of both $\mathbf{x}$ and t, given that only two fidelities are considered [70, 139]. Some attempts aim to use a different form of $\alpha_1(\mathbf{x}, t)$. For example, [146] used $\alpha_1(\mathbf{x}, t) = \max\left(0, 1 - \sigma_{f_\mathrm{H}}^{2,k+1}(\mathbf{x}|t)/\sigma_{f_\mathrm{H}}^{2,k}(\mathbf{x})\right)$, where $\sigma_{f_\mathrm{H}}^{2,k+1}(\mathbf{x}|t)$ denotes the predictive variance of the updated MF surrogate, which is formed by adding the new design point generated by fidelity t to the current training dataset. Other attempts define the MF acquisition function as the ratio between an acquisition function of the generic BO and the computational cost per model on the fidelity [151, 157]. Beyond the constraint of most one-step look-ahead heuristic MF acquisition functions, the empirical performance of two-step look-ahead EIs has recently been assessed [134, 154].

3. Sequential selection

The sequential selection approach consists of two steps: (i) select the new design point $\mathbf{x}^k$ and (ii) select the fidelity t with $\mathbf{x}^k$ found in step (i). The main difference between this approach and the heuristic approach is that the selection of $\mathbf{x}^k$ is independent of t. This means, the solution improvement is separated from the consideration of computational

Algorithm 5 MF BO using sequential selection.

1: **Input:** $\mathcal{X}$, K, $\mathcal{D}_t$ $(t = 1, \ldots, T)$;
2: $\mathcal{D}^0 \leftarrow \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_T$;
3: $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow u(\mathcal{D}^0)$;
4: **for** $k = 1 : K$ **do**
5: Construct MF surrogate $\hat{f}^k_{\mathrm{H}}(\mathbf{x})$ from $\mathcal{D}^{k-1}$;
6: Formulate MF acquisition function $\alpha(\mathbf{x})$;
7: $\mathbf{x}^k \leftarrow \underset{\mathbf{x}}{\operatorname{argmax}}\ \alpha(\mathbf{x})$ s.t. $\mathbf{x} \in \mathcal{X}$;
8: Formulate fidelity-query function $\gamma(\mathbf{x}^k, t)$;
9: $t \leftarrow \underset{t}{\operatorname{argmax}}\ \gamma(\mathbf{x}^k, t)$ s.t. $t \in \{1, \ldots, T\}$;
10: $\mathcal{D}_t \leftarrow \mathcal{D}_t \cup \{\mathbf{x}^k, f_t(\mathbf{x}^k)\}$;
11: $\mathcal{D}^k \leftarrow \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_T$;
12: $\{\mathbf{x}_{\min}, f_{\min}\} \leftarrow u(\mathcal{D}^k)$;
13: **end for**
14: **return** $\{\mathbf{x}_{\min}, f_{\min}\}$.

cost and the accuracy improvement of the MF surrogate. Algorithm 5 shows the pseudo-code for MF BO using the sequential selection approach.

One of the first methods of the sequential approach by [129] used the so-called preposterior analysis to formulate a fidelity-query function $\gamma(\mathbf{x}^k, t)$ (Line 8 of Algorithm 5) after maximizing EI in Line 7 for $\mathbf{x}^k$. The preposterior analysis was to examine how the standard deviation of the MF surrogate at $\mathbf{x}^k$ reduces when fictitious simulation data for each fidelity are added to the currently real training data. Because they used auto-regressive approach, [129] could generate the samples constituting the fictitious simulation data for fidelity t at $\mathbf{x}^k$ from a Gaussian characterized by the predictive mean and predictive variance at $\mathbf{x}^k$. In this way, a reduction of the standard deviation associated with fidelity t was defined as $\sigma^k_{f_{\mathrm{H}}}(\mathbf{x}^k) - \bar{\sigma}(\mathbf{x}^k, t)$, where $\bar{\sigma}(\mathbf{x}^k, t)$ denotes the expected standard deviation from the updated GP posterior after the fictitious simulation data for fidelity t are added to the currently real training data several times. By further considering the computational cost per model on the fidelity, the fidelity-query function was defined as

$$\gamma(\mathbf{x}^k, t) = \frac{\sigma^k_{f_{\mathrm{H}}}(\mathbf{x}^k) - \bar{\sigma}(\mathbf{x}^k, t)}{\sigma^k_{f_{\mathrm{H}}}(\mathbf{x}^k) - \bar{\sigma}(\mathbf{x}^k, T)} \frac{c(T)}{c(t)}. \tag{57}$$

We see that, a large value of $\gamma(\mathbf{x}^k, t)$ indicates a large reduction in the prediction uncertainty per unit of computational cost at $(\mathbf{x}^k, t)$. A similar fidelity-query function can be found in [28, 138].

Alternatively, [116] defined $\gamma(\mathbf{x}^k, t)$ as the negative value of the computational cost per model on the fidelity. They also imposed two constraints on the maximization of $\gamma(\mathbf{x}^k, t)$. The first constraint, aiming at promoting exploration, required that the posterior variance associated with $(\mathbf{x}^k, t)$ should be larger than a pre-specified threshold value. The second constraint required that the maximizer of $\gamma(\mathbf{x}^k, t)$ should be found in a neighborhood of T that can shrink over time. As a result, the constraints in Line 9 of Algorithm 5 should consist of the aforementioned two constraints and a bound constraint on t as t is considered because a continuous variable [116]; see Section IV.F.1.

In another method of the sequential approach, [27] formulated $\gamma(\mathbf{x}^k, t)$ based on the idea that the use of LF data favors exploration, and that of HF data favors exploitation. Thus, multiple fidelities can be used simultaneously for updating MF surrogates. More specifically, $\gamma(\mathbf{x}^k, t)$ can be defined as the ratio between the total uncertainty reduction when adding the new data generated from models whose fidelities are not greater than t and the total computational cost of these models. As a result, the models whose fidelities are not greater than t are invoked in Line 10 of Algorithm 5, which differs from the above-mentioned methods of the sequential approach. Note that this method is a type of no-fidelity consideration approach when $t = T$.

C. Portfolio of acquisition functions

No acquisition function works well on all classes of problems because the preferred search strategy may vary during different phases of a sequential optimization process [25]. To address this, a promising solution involves employing a *portfolio of multiple acquisition functions* [228, 229]. The idea is to leverage the interaction between different acquisition functions to safeguard against the potential failure of any single search strategy. This collaborative interaction is quantified via a unified portfolio metric, which can take the form of a meta-criterion [228] or an entropy search metric [229]. The approach then requires two steps. First, it finds a collection of new design points by maximizing individual acquisition functions within the portfolio. Second, from the collection of design points, it selects the actual design point that maximizes the portfolio metric.

D. Maximization of acquisition functions

Because it is difficult to examine the convexity of most acquisition functions, we wish to maximize $\alpha(\mathbf{x})$ or $\alpha(\mathbf{x}, t)$ using a global optimization algorithm [230]. Nevertheless, robust derivative-based algorithms can be used if an acquisition function has an analytical form. When maximizing an information-based acquisition function, it may be useful to discretize the design variable space or to use a sampling method. Table 5 lists several optimization algorithms and techniques that are found in the literature to maximize the acquisition functions (and their variants) described in Sections V.A and V.B.

E. Software implementations

There exist dozens of open-source BO libraries and most of them implement one-step look-ahead acquisition functions. We recommend the Emukit package that provides a fully-featured sublibrary for BO and supports PI, EI, GP-NLCB, ES, and MES [235]. Other packages include BoTorch [236] and SMAC3 [237].

A few BO libraries implement multi-step look-ahead acquisition functions. We recommend a repository for two-step look-ahead EI [221] and a repository for rollout dynamic programming [223].

Table 5 Summary of optimization algorithms for maximizing acquisition functions.

Acquisition	Optimization algorithm	Reference
EI*	Branch-and-bound algorithm	[165]
	Nelder-Mead simplex method (NM)	[15, 227]
	Broyden-Fletcher-Goldfarb-Shanno	[201]
	Limited-memory BFGS (L-BFGS)	[26, 133]
	Discretizing design variable space	[134, 153]
	Genetic algorithm (GA)	[16, 70, 129, 135, 231]
	Particle swarm optimization	[80, 232]
	Evolution w/ covariance matrix adaptation	[28, 138]
GP-UCB*	Direct optimization algorithm	[116, 233]
PI*	GA	[139]
KG	Multi-start stochastic gradient	[234]
IAGO	Discretizing design variable space	[215]
ES	Sampling + L-BFGS	[184]
PES	Sampling + (local search or NM)	[210]
MES	Sampling + (local search or NM)	[217]

* And its variants.

VI. Further research topics

Recent advances in MFO and especially BO have focused on solving intricate yet important design optimization problems, which stem from the nature of the problems or from the inherent limitations of the generic BO [31]. In this section, we briefly describe sophisticated BO and several MFO approaches to addressing some of these problems, including constrained optimization (Section VI.A), high-dimensional optimization (Section VI.B), optimization under uncertainty (Section VI.C), and multi-objective optimization (Section VI.D). By doing so, we expect to shed light on potential opportunities for future research in MF BO.

A. Constrained optimization

Constrained design optimization is formulated as

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X}} \ f_{\mathrm{H}}(\mathbf{x}) \\ &\text{s.t.} \ \ g_{\mathrm{H},i}(\mathbf{x}) \le 0, \ \ i = 1, \ldots, I, \end{aligned} \tag{58}$$

where the constraint functions are conditionally independent given a vector of design variables and expensive to compute, therefore can be estimated via HF evaluations $g_{\mathrm{H},i}(\mathbf{x})$.

There are three approaches to solving problem (58) via BO. The first approach formulates constrained acquisition functions $\alpha_{\mathrm{c}}(\mathbf{x})$ that consider the influence of constraints on the solution improvement. The second approach reformulates problem (58) as an unconstrained problem which can be addressed by the generic BO. The third approach employs TS,

as described in Algorithm 2, via the realizations of GP posteriors for the objective and constraint functions.

An attempt of the first approach by [238] formulated a *constrained expected improvement* (CEI) acquisition function as the product of EI and the probability that all inequality constraints are satisfied (i.e., probability of feasibility). Accordingly, CEI reads

$$\alpha_{\mathrm{c}}(\mathbf{x}) = \mathrm{EI}(\mathbf{x})\mathbb{P}[g_{\mathrm{H},1}(\mathbf{x}) \leq 0, \ldots, g_{\mathrm{H},I}(\mathbf{x}) \leq 0]. \tag{59}$$

Here the probability of feasibility $\mathbb{P}[g_{\mathrm{H},1}(\mathbf{x}) \leq 0, \ldots, g_{\mathrm{H},I}(\mathbf{x}) \leq 0]$ tightens the search space for the new design point by penalizing unfeasible regions of an approximate problem of problem (58). This approach, however, may put more weight on the feasible regions away from the boundary of feasible space, thereby overlooking the boundary where the optimal solution has a high chance of being found.

When conditioning $g_{\mathrm{H},i}(\mathbf{x})$ on the GP posteriors $\hat{g}_{\mathrm{H},i}(\mathbf{x})$, CEI becomes

$$\alpha_{\mathrm{c}}(\mathbf{x}) = \mathrm{EI}(\mathbf{x}) \prod_{i=1}^{I} \Phi\left(\frac{-\mu_{\mathrm{g},i}(\mathbf{x})}{\sigma_{\mathrm{g},i}(\mathbf{x})}\right), \tag{60}$$

where $\mu_{\mathrm{g},i}(\mathbf{x})$ and $\sigma_{\mathrm{g},i}(\mathbf{x})$ are the predictive mean and standard deviation of $\hat{g}_{\mathrm{H},i}(\mathbf{x})$, respectively. The performance of CEI has been verified by [232, 239, 240].

Note that it is not necessary to formulate $\alpha_{\mathrm{c}}(\mathbf{x})$ if the constraint functions are easy to compute. In this case, we can simply maximize any acquisition function of the generic BO under a set of easy-to-check constraints for finding the new design point.

MF BO has recently incorporated the probability of feasibility or its variants into heuristic MF acquisition functions to address several design optimization problems [see e.g., 80, 134, 139]. However, decoupling the selection of a new design point and the selection of fidelity for constrained optimization is still an open issue.

[241] introduced the concept of *integrated expected conditional improvement* (IECI) for solving problem (58) when it involves inexpensive constraint functions. This concept was then extended to solve problems featuring expensive constraint functions. IECI reads

$$\alpha_{\mathrm{c}}(\mathbf{x}) = \int_{\mathcal{X}} [\mathrm{EI}(\mathbf{x}') - \mathrm{EI}(\mathbf{x}'|\mathbf{x})]\, p(\mathbf{x}')\mathrm{d}\mathbf{x}', \tag{61}$$

where $\mathrm{EI}(\mathbf{x}')$ is EI at a reference point $\mathbf{x}'$ that is distributed according to $p(\mathbf{x}')$, and $\mathrm{EI}(\mathbf{x}'|\mathbf{x})$ is EI at $\mathbf{x}'$ given that $\{\mathbf{x}, f_{\mathrm{H}}(\mathbf{x})\}$ is added to the current data.

We see that, IECI handles the constraints via $p(\mathbf{x}')$, which may be uniform in a bounded feasible region $\mathbf{x}' \in \mathcal{X}$ and zero otherwise. When the constraint functions are costly, $p(\mathbf{x}')$ is approximated by performing the Monte-Carlo simulation based on their surrogates [241]. An interesting property of IECI is that it, via flexible choices of $p(\mathbf{x}')$, allows the selection of infeasible design points at cost of gaining useful information about the objective function.

The following considerations should be taken into account if one wishes to use IECI for MF BO: (i) how to select

$p(\mathbf{x}')$ when considering the fidelities and (ii) how to reason about the selection of an appropriate fidelity if an infeasible design point is found after maximizing $\alpha_{\mathrm{c}}(\mathbf{x})$.

Expected volume minimization (EVM) by [242] is another constrained acquisition function for solving problem (58) via BO. By integrating the product of the probability of improvement and the probability of feasibility, EVM is defined as

$$\begin{aligned}\alpha_{\mathrm{c}}(\mathbf{x}) = &\int_{\mathcal{X}} \mathbb{P}\left[f_{\mathrm{H}}(\mathbf{x}) \leq \min\{f_{\mathrm{H}}(\mathbf{x}), f_{\min}\}\right] \mathbb{P}[g_{\mathrm{H},1}(\mathbf{x}) \leq 0, \ldots, g_{\mathrm{H},I}(\mathbf{x}) \leq 0] \mathrm{d}\mathbf{x} \\ &+ \int_{\mathcal{X}} \mathbb{P}\left[f_{\mathrm{H}}(\mathbf{x}) \leq f_{\min}\right] \left(1 - \mathbb{P}[g_{\mathrm{H},1}(\mathbf{x}) \leq 0, \ldots, g_{\mathrm{H},I}(\mathbf{x}) \leq 0]\right) \mathrm{d}\mathbf{x}.\end{aligned} \tag{62}$$

The first and second integrals correspond to the probability of improvement considering the feasibility and unfeasibility of the new design point, respectively. If the new design point is feasible, the best value of the objective found so far is $\min\{f_{\mathrm{H}}(\mathbf{x}), f_{\min}\}$. Otherwise, the best value of the objective found so far is still $f_{\min}$. Although it is an attractive concept, maximizing EVM is costly because it requires numerical integration over $\mathcal{X}$.

In an attempt of the reformulation approach, [243] adopted BO to minimize the *augmented Lagrangian* of problem (58), which reads

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\lambda}, \rho_0) = f_{\mathrm{H}}(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i g_{\mathrm{H},i}(\mathbf{x}) + \frac{1}{2\rho_0} \sum_{i=1}^{I} \max\{0, g_{\mathrm{H},i}(\mathbf{x})\}^2, \tag{63}$$

where $\rho_0 > 0$ is a penalty parameter and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_I]^\intercal$ a vector of non-negative Lagrange multipliers, i.e., $\lambda_1 \geq 0, \ldots, \lambda_I \geq 0$.

The augmented Lagrangian method starts by assigning initial values of ρ_0 and $\boldsymbol{\lambda}$. A sufficiently large initial value of ρ_0 is required to enforce feasibility. It then sequentially finds the minimizer of the augmented Lagrangian corresponding to the given values of $\boldsymbol{\lambda}$ and ρ_0, and updates them for the next iteration with the resulting minimizer. Algorithm 6 implements the augmented Lagrangian method (not in BO), where η_1 and η_2 in Line 1 are two small thresholds for checking termination conditions in Line 6.

In the context of BO, the idea is to construct in Line 3 of Algorithm 6 a GP for $\mathcal{L}(\mathbf{x}|\boldsymbol{\lambda}^{k-1}, \rho_0^{k-1})$ using the observations for several samples of design variable vectors and the corresponding values of the objective and constraint functions. Based on this GP, the next design point $\mathbf{x}^k$ in Line 4 is found by maximizing EI, and the new parameters ρ_0 and $\boldsymbol{\lambda}$ are updated accordingly. The advantage of this approach is that it, via BO, can increase the chance of finding a global solution as compared with the traditional augmented Lagrangian method, which is a local optimization algorithm.

The reformulation approach enables a straightforward application of MF BO in solving constrained design optimization problems. Once the augmented Lagrangian has been formulated, we can use any established unconstrained MF BO algorithms, which are more mature than the constrained counterparts.

The third approach to solving problem (58) is via TS [244]. Let $q_{\mathrm{H},i}^k(\mathbf{x})$ and $h_{\mathrm{H},i}^k(\mathbf{x})$ be the realizations of $\hat{f}_{\mathrm{H}}(\mathbf{x})$ and

Algorithm 6 Augmented Lagrangian method.

Require: ρ_0^0, $\boldsymbol{\lambda}^0$, bounded region $\mathcal{X}$, $\eta_1 > 0$, $\eta_2 > 0$;
for $k = 1, 2, \ldots$ **do**
  Formulate $\mathcal{L}(\mathbf{x}|\boldsymbol{\lambda}^{k-1}, \rho_0^{k-1})$;
  $\mathbf{x}^k \leftarrow \underset{\mathbf{x}}{\text{argmin}}\ \mathcal{L}(\mathbf{x}|\boldsymbol{\lambda}^{k-1}, \rho_0^{k-1})$ s.t. $\mathbf{x} \in \mathcal{X}$;
  $\lambda_i^k \leftarrow \max\{0, g_{\text{H},i}(\mathbf{x}^k)/\rho_0^{k-1}\}$;
  if $\|\max\{0, g_{\text{H},1}(\mathbf{x}^k), \ldots, g_{\text{H},I}(\mathbf{x}^k)\}\| \leq \eta_1$ or $\|\nabla f_{\text{H}}(\mathbf{x}^k) + \sum_{i=1}^I \lambda_i^k \nabla g_{\text{H},i}(\mathbf{x}^k)\| \leq \eta_2$ **then**
    $\mathbf{x}_{\min} \leftarrow \mathbf{x}^k$
  else
    if $\mathbf{g}_{\text{H}}(\mathbf{x}^k) \leq \mathbf{0}$ **then**
      $\rho_0^k \leftarrow \rho_0^{k-1}$;
    else
      $\rho_0^k \leftarrow 0.5\rho_0^{k-1}$;
    end if
  end if
end for
return $\mathbf{x}_{\min}$.

$\hat{g}_{\text{H},i}(\mathbf{x})$ at iteration k of TS, respectively. Assume that we perform Line 10 of Algorithm 2 in a batch setting and obtain a total of q new candidates, denoted by $\mathbf{x}_1^k, \ldots, \mathbf{x}_q^k$. Let $\mathcal{F}^k = \{\mathbf{x}_l^k | \, h_{\text{H},i}^k(\mathbf{x}) \leq 0, \forall l \in \{1, \ldots q\}, \forall i \in \{1, \ldots I\}\}$ be a set of points whose realizations are feasible. If $\mathcal{F}^k \neq \emptyset$, we select one of its elements that minimizes $q_{\text{H},i}^k(\mathbf{x})$. Otherwise, we select a candidate with minimum total violation $\sum_{i=1}^I \max\{0, h_{\text{H},i}(\mathbf{x})\}$.

B. High-dimensional optimization

Based on the size of optimization problems, we can classify them into the following three classes:

- Small-dimensional problems, which have five or fewer design variables and constraints.
- Moderate-dimensional problems, which have from five to a thousand design variables and constraints.
- High-dimensional problems, which have thousands or even millions of variables and constraints.

While this classification is not rigid, it reflects the fundamental differences in solution approach associated with varying problem sizes [245].

Extending the BO framework to high-dimensional optimization problems faces two computational challenges. First, a conventional GP model scales cubically with the number of training data points [13], which often increases as the number of design variables increases. Second, maximizing high-dimensional acquisition functions is nontrivial because they are mostly flat at high dimension [246]. These two challenges make high-dimensional BO related to, but distinct from scalable GP modeling approaches, which only focus on the first challenge [42].

Recent advances in high-dimensional BO have followed three main approaches [247]:

- Exploiting low, active/effective dimensional subspace of the design variables (Section VI.B.1).
- Exploiting the additive structures of the objective and/or constraint functions (Section VI.B.2).
- Trust-region BO (Section VI.B.3).

1. Subspace-based approach

This approach constructs surrogates for the objective/constraint functions in an unknown low-dimensional subspace of the design variables, and then performs BO in this subspace. Methods of this approach differ in their ways of defining the low-dimensional subspace.

Sensitivity analysis [248] is a classical method of the subspace-based approach that selects the most influential design variables and ignores the less influential ones. A drawback of this method is that it cannot capture the models varying most prominently along the directions that are not aligned with the original coordinate system of the design variable space [249], leading to the quest for active/effective subspace.

[249] constructed an active subspace for applications of GPs using the eigenvalue decomposition of a covariance matrix associated with the gradient of the function of interest. Then, the first d_e eigenvectors were selected to form a reduced-order basis. This is justified because an active subspace represents the directions of the largest variability of a function. However, the exact calculation of the gradient covariance matrix is impossible, leading to use of the Monte-Carlo integration that requires a large number of HF data points.

A random linear embedding [250] is one of the important methods of the subspace-based approach. It states that for any design variable vector $\mathbf{x} \in \mathbb{R}^d$ with an unknown active subspace dimension of d_e, there exists, with probability 1, a vector $\mathbf{y} \in \mathbb{R}^{d_\mathrm{s}}$ ($d_\mathrm{s} \geq d_\mathrm{e}$) such that $f_\mathrm{H}(\mathbf{x}) = f_\mathrm{H}(\mathbf{A}\mathbf{y})$, where $\mathbf{A} \in \mathbb{R}^{d \times d_\mathrm{s}}$ is a random projection matrix whose independent elements are sampled from $\mathcal{N}(0, 1)$. This enables performing BO in the low-dimensional space of $\mathbf{y}$ once $\mathbf{A}$ and the domain of $\mathbf{y}$ have been determined. The choice of covariance function to construct GPs in the space of $\mathbf{y}$ is another important aspect that affects the performance of the random linear embedding. Theoretically, [251] showed that any GP-based BO algorithm runs on the embedded low-dimensional space as it would if it was run on an unknown active subspace. [252] further listed several crucial issues and misconceptions about the use of random linear embedding for high-dimensional BO.

Apart from random linear embedding methods, nonlinear embedding techniques have been used to explore more effective low-dimensional spaces. For example, [253] learned a low-dimensional space using variational autoencoders. [254] performed high-dimensional BO via a nonlinear feature mapping for dimensionality reduction and a reconstruction mapping for evaluating the objective function. A drawback of the nonlinear embedding techniques is that they require a large number of HF data points for embedding learning.

To adopt the subspace-based approach for MF BO, in addition to formulating an MF surrogate and an MF acquisition function in an approximate low-dimensional subspace, addressing the following questions is of interest:

- Is it possible to facilitate learning the approximate low-dimensional subspace using the information from multiple fidelities?
- How can we select a new design point and a fidelity level to gain the information used for updating the approximate subspace as much as possible after each iteration?

2. Additive structure approach

This approach makes a strong assumption that the objective function can be written in an additive form, such that [255]

$$f_{\mathrm{H}}(\mathbf{x}) = \sum_{m=1}^{M} f_m(\mathbf{x}_m), \tag{64}$$

where $\mathbf{x}_m \in \mathcal{X}_m$ are disjoint lower dimensional components of a high-dimensional vector of design variables.

Using the assumption in Eq. (64), we can construct GP models for $f_m(\mathbf{x}_m)$ and maximize the acquisition functions formulated from these models to progress BO, regardless of our knowledge on the structure of $f_m(\mathbf{x}_m)$. If $\mathbf{x}_m^k$ is the new point from maximizing the acquisition function m, then the new design point is $\mathbf{x}^k = \bigcup_{m=1}^{M} \mathbf{x}_m^k$. Alternatively, $\mathbf{x}^k$ can be found by maximizing an acquisition function derived from an overall GP for $f_{\mathrm{H}}(\mathbf{x})$ whose mean (or covariance) function is assigned as the sum of mean (or covariance) functions of the GPs for $f_m(\mathbf{x}_m)$.

Unfortunately, it is hard to reason about a good, unknown decomposition structure, especially when the objective function is non-additive. Recent attempts have tried to find possible model decompositions via Markov chain Monte-Carlo (MCMC) algorithms. For example, a large number of possible additive structures that well explain the data in each BO iteration can be sampled from a model posterior via Metropolis–Hastings algorithm [256]. Gibbs sampling can be performed to construct local GPs for use of the so-called ensemble BO after the input space is partitioned via a Mondrian process [257].

Different combinations of HF and LF data may completely change the additive form in Eq. (64) because learning possible additive structures is data-driven. Thus, examining the sensitivity of decomposition structures with respect to the data used may be the first step if one wishes to use the additive structure approach for MF BO.

3. Trust-region BO (TuRBO) approach

This approach simultaneously uses independent local BO runs for global optimization [258]. Each local BO relies on a local GP constructed from a trust region that is defined as a hyperrectangle centered at the best solution found from a set of data points. The size of such a trust region can be expanded when better solutions are found after several consecutive iterations, or be contracted when no better solution is found after several consecutive iterations.

Assume that TuRBO maintains m trust regions, i.e., $\mathrm{TR}_1, \ldots, \mathrm{TR}_m$, at its iteration k. The local GPs constructed from these trust regions are denoted by $\mathcal{GP}_1^k, \ldots, \mathcal{GP}_m^k$. In each iteration, TuRBO selects a batch of q new design points, i.e., $\{\mathbf{x}_1^k, \ldots, \mathbf{x}_q^k\}$, which are drawn from the union of the trust regions, and then updates the local GPs associated with the trust regions from which the new design points are drawn [258]. This can be done via performing TS described in Algorithm 2.

To select a new design point i in iteration k of TuRBO, i.e., $\mathbf{x}_i^k$, TS randomly draws m functions $q_{\mathrm{H},1}^{i,k}(\mathbf{x}), \ldots, q_{\mathrm{H},m}^{i,k}(\mathbf{x})$ from $\mathcal{GP}_1^k, \ldots, \mathcal{GP}_m^k$, respectively. Then, it selects $\mathbf{x}_i^k$, $i = 1, \ldots, q$, as the point that minimizes the function value

across all m sampled functions [258], such that

$$\mathbf{x}_i^k = \underset{l\in\{1,\ldots,m\}}{\operatorname{argmin}} \ \underset{\mathbf{x}\in \mathrm{TR}_l}{\operatorname{argmin}} \ q_{\mathrm{H},l}^{i,k}(\mathbf{x}), \quad i = 1,\ldots,q. \tag{65}$$

By solving problem (65), we obtain the information about $\mathbf{x}_i^k$ as well as the trust region from which $\mathbf{x}_i^k$ is drawn. This information informs the updates of the size of the trust regions and the local GPs. Note that while problem (65) is still a high-dimensional minimization problem, solving it in an efficient manner is not the focus of TuRBO.

The following are two possible ways to develop a trust-region MF BO algorithm based on the TuRBO approach:

- Use local GP-based MF surrogates for local BO runs and then adopt the no-fidelity consideration approach (Section V.B.1) to update these surrogates after $\mathbf{x}_i^k$ is obtained by solving problem (65).
- Use local GP-based MF surrogates for local BO runs and then adopt the sequential selection approach (Section V.B.3) to select fidelity variables after $\mathbf{x}_i^k$ is obtained by solving problem (65). In this way, the fidelity-query function $\gamma(\mathbf{x}_i^k, t)$ described in Eq. (57) can be formulated for the trust region from which $\mathbf{x}_i^k$ is drawn.

C. Optimization under uncertainty

Managing uncertainty is one of the important tasks in scientist computing [259, 260] and engineering design [261]. To perform this task, we can follow two key steps.

- First, determine and capture the sources of uncertainty for the problem of interest. These sources may include model parameters, the mathematical model itself, measurement noise, and the lack of knowledge of the modelers or engineers. Each source can be described by a PDF or an interval-valued quantity if it is classified into aleatory or epistemic uncertainty, respectively.
- Second, select a method to propagate the uncertainty through the problem for evaluating quantified uncertainty in the quantities of interest. The choice of the propagation method depends on the quality of information we have in the first step. This choice is also problem-dependent.

Optimization under uncertainty is crucial because the optimal solution is sensitive to even a small change in the problem. This may be attributed to, for example, the fact that optimal solutions are often on the boundary of the feasible space.

Although there is a rich literature on techniques for optimization under uncertainty, processing this class of optimization in engineering design is still a challenging task. In this section, we briefly review recent applications of MF modeling methods and BO in two branches of optimization under uncertainty: robust optimization (RO) (Section VI.C.1) and reliability-based optimization (RBO) (Section VI.C.2). We also attempt to provide possible directions for future research on MF BO to address these problems.

Let $\mathbf{s}$ denote a vector of random parameters that encapsulates uncertainty in our problem. If a design variable is

random, then it can be defined as a sum of a nominal variable, which is considered a design variable, and a random parameter before the optimization problem is formulated. We describe uncertainty in $\mathbf{s}$ using an interval $[\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}]$ for set-based uncertainty and a PDF $p(\mathbf{s})$ for probabilistic uncertainty. The latter is the focal point of our discussion below.

1. Robust optimization

An RO problem is often formulated using one of the following three concepts: absolute robustness, relative robustness, and less variance [262]. The absolute robustness formulates the problem based on the worst values of the objective and constraint functions under set-based uncertainty with a fixed set $[\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}]$. Particularly, we minimize the worst value of the objective function under the constraints on the worst values of constraint functions, which is also called the worst-case scenario approach (or minimax approach) [see e.g., 263, 264]. The relative robustness formulates the problem under set-based uncertainty to maximize the gap between the nominal value and the worst-case value of the objective function. The less variance concept formulates the problem based on probabilistic uncertainty.

Note that there exists another robustness concept that formulates the problem using the information-gap decision theory [265]. In this concept, the uncertainty set, often defined as a closed Euclidean ball of radius ε_s centered at a nominal vector $\mathbf{s}_0$, can vary via adjusting ε_s. A design is considered robust if it remains feasible for large uncertainty sets.

In the less variance concept, a design that is less sensitive to uncertainty is considered more robust. Then, we may wish to minimize the mean and variance of the objective function simultaneously, thereby formulating a bi-objective RO problem, as discussed in Section VI.D. Nevertheless, we can reformulate this bi-objective RO problem using the following weighted-sum formulation:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X}} \; w\mathbb{E}_\mathbf{s}\left[f_\mathrm{H}(\mathbf{x},\mathbf{s})\right] + (1-w)\sqrt{\mathbb{V}_\mathbf{s}\left[f_\mathrm{H}(\mathbf{x},\mathbf{s})\right]} \\ &\text{s.t.} \; \mathbb{E}_\mathbf{s}\left[g_{\mathrm{H},i}(\mathbf{x},\mathbf{s})\right] + \beta_i\sqrt{\mathbb{V}_\mathbf{s}\left[g_{\mathrm{H},i}(\mathbf{x},\mathbf{s})\right]} \le 0, \;\; i = 1,\ldots,I, \end{aligned} \tag{66}$$

where $\mathbb{V}_\mathbf{s}[\cdot]$ denotes the variance of the quantity inside the brackets under uncertainty in $\mathbf{s}$, $w \in (0,1)$ is a weight value, and β_i is a risk attitude factor for constraint i.

Difficulties in solving problem (66) arise from two aspects: (i) uncertainty propagation, i.e., evaluating statistical estimates of the objective and constraint functions for each candidate solution, and (ii) search strategy that involves an inner loop of uncertainty propagation.

For the first aspect, the Monte-Carlo integration [266] and polynomial chaos expansion [267] are traditional methods, but they come at cost of the curse of dimensionality. The Taylor series approximation [268] and Bayes-Hermite quadrature [269] may serve as alternatives if the uncertain functions are differentiable and the parameters $\mathbf{s}$ are normally distributed. For the second aspect, derivative-free methods [270] are often used because they do not require derivative

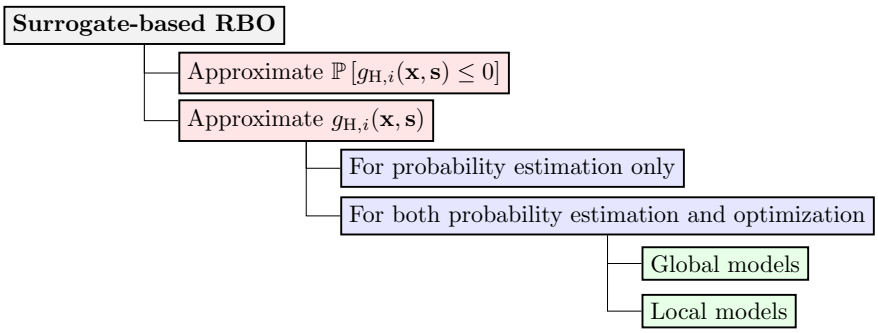

Fig. 9 Different schemes of surrogate-based RBO [278].

information of the statistical estimates, which are too expensive to extract. Meanwhile, only a few works in the literature have solved RO problems via BO [271, 272].

As an early attempt to solve problem (66) via MFO, [273] estimated the mean and variance at a candidate solution via the Monte-Carlo integration while adjusting solutions using a derivative-free method. Given two computational models with different infidelities and under a fixed computational budget, the MF mean was evaluated from the HF and LF mean values, which required n calls of the HF model and m calls of the LF model, for $m > n$. The MF variance was evaluated in a similar manner.

Other attempts to solve problem (66) via MFO differ in their ways of using MF surrogates, uncertainty propagation methods, and optimization solvers [see e.g., 274–277]. For example, [274], under the model-then-optimize approach, employed a composition method via polynomial chaos for the mean and variance estimates in each iteration of the sequential least squares quadratic programming. [277] incorporated the hierarchical Kriging model and the Monte-Carlo integration into a GA solver.

While MFO approaches can enhance the feasibility of addressing problem (66), it does so at the expense of supplementary uncertainty induced by use of LF models. This uncertainty has the potential to misguide the optimization process or render the RO solutions infeasible if not considered meticulously. Hence, it is of great interest to devise efficient methods that can account for this uncertainty if one wishes to solve problem (66) via an MFO approach.

2. Reliability-based optimization

Under probabilistic uncertainty in $\mathbf{s}$, RBO, or chance-constrained optimization [279], minimizes the mean of the objective function under probabilistic constraints. The general formulation of RBO is

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X}} \ \mathbb{E}_{\mathbf{s}}\left[f_{\mathrm{H}}(\mathbf{x},\mathbf{s})\right] \\ &\text{s.t.} \ \ \mathbb{P}_{\mathbf{s}}\left[g_{\mathrm{H},i}(\mathbf{x},\mathbf{s}) \leq 0\right] \geq 1-\epsilon_i, \ \ i=1,\ldots,I, \end{aligned} \tag{67}$$

where $\epsilon_i \in (0, 1)$ is a prescribed risk level associated with probabilistic constraint i, e.g., $\epsilon_i = 0.1$, 0.05, or 0.01. Large values of ϵ_i sacrifice the reliability of the solution, while small values lead to conservative designs. We also assume that the constraint functions are conditionally independent given a vector of design variables. Hence, problem (67) is an RBO with individual probabilistic constraints, which differs from another class of RBO problems that involve joint probabilistic constraints [280].

Although problem (67) is very important for optimization under uncertainty, it is too difficult to solve. The difficulty arises from three aspects. First, checking the feasibility of a candidate solution is nontrivial because the computation of probabilistic constraints is generally an NP-hard problem [281]. Second, finding an exact optimal solution may be impossible because the feasible region defined by the probabilistic constraints is generally non-convex [282]. Third, estimating the mean of the objective is computationally demanding, which is discussed in Section VI.C.1.

The Monte-Carlo integration is a traditional method for estimating the probabilistic constraint functions. While this method is simple to implement and able to provide unbiased statistical estimates, its rate of convergence and the convergence stability often depend on the quality of random generators [283], leading to the difficulty in capturing the sensitivity of constraint functions. This method also introduces a considerable computational burden when evaluating the feasibility of a candidate solution, which violates a reasonable expectation of achieving a greater computational efficiency through an increased number of optimization iterations, rather than dedicating excessive time to each iteration.

In engineering design optimization, problem (67) is often solved approximately. Classical RBO approximation techniques fall into three distinct categories: top-level, mono-level, and decoupled approaches. A benchmark study and a comprehensive review of methods in each of these approaches can be found in [284] and [285], respectively. Other approximation techniques solve problem (67) sequentially by leveraging surrogates for $\mathbb{P}_{\mathbf{s}}\left[g_{\mathrm{H},i}(\mathbf{x}, \mathbf{s}) \leq 0\right]$ or $g_{\mathrm{H},i}(\mathbf{x}, \mathbf{s})$. Figure 9 shows different schemes of the surrogate-based RBO, which was extensively discussed by [278].

Existing approaches in the field of mathematical optimization typically rely on several special forms of problem (67). For example, the scenario approach [286] is formulated for linear objective functions and continuous, convex constraint functions $g_{\mathrm{H},i}(\mathbf{x}, \mathbf{s})$ with respect to $\mathbf{x}$. The convex approximation requires that $g_{\mathrm{H},i}(\mathbf{x}, \mathbf{s})$ is convex in $\mathbf{x}$ for every instance of $\mathbf{s}$ and the set defined by the deterministic constraints is convex [282].

In BO literature, [287] estimated failure probabilities using a modification of EGO that relies on the GPs for constraint functions and an expected feasibility acquisition function. This acquisition function aims to increase the accuracy of the GPs in the vicinity of the limit state, i.e., the boundary of the feasible region. Once the GPs have been refined by EGO, they can be used to estimate the probabilities via multimodal adaptive importance sampling. [288] provided a class of likelihood-weighted acquisition functions to uncertainty quantification of the quantities of interest. [289] extended the approach by [287] and solved problem (67) by embedding EGO in a sequential strategy of the decoupled RBO approach. [290] developed an RBO approach that couples reusing samples for a posteriori biasing density nearby designs for probability estimation with randomly exploiting in the neighborhood of the current design for

optimization.

Although many methods, inspired by [287], have adopted MF BO to facilitate reliability analyses (see e.g., [291–294]), the use of MF BO for addressing RBO problems is still an open issue. This may be attributed to the conflict between accuracy and efficiency properties of a good optimization algorithm. While enhancing optimization accuracy by using sophisticated MF reliability analysis methods, it unintentionally requires excessive computer time or storage, thereby reducing optimization efficiency.

The above observations present potential opportunities for MF BO to address problem (67). In addition to improving reliability analyses, considerations should include other important aspects such as improving the solution quality, evaluating the feasibility, selecting fidelities, and addressing uncertainty due to use of LF models. We believe that the sequential selection approach (Section V.B) would be suitable for such serious considerations because it can decouple the selections of design variables $\mathbf{x}$, random parameters $\mathbf{s}$, and fidelity variables for different purposes. After these considerations, MF BO may look on a far horizon to explore its possibility of solving RBO problems with joint probabilistic constraints, which are also very important for decision-making under uncertainty, but generally more difficult to solve than problem (67).

D. Multi-objective optimization

We formulate a multi-objective (or vector-valued) optimization problem when we wish to optimize several competing objectives but do not know how to prioritize them. A general m-objective optimization problem reads

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X}} \quad \left[f_{\mathrm{H},1}(\mathbf{x}),\ldots,f_{\mathrm{H},m}(\mathbf{x})\right] \\ &\text{s.t.} \quad g_{\mathrm{H},i}(\mathbf{x}) \le 0, \quad i=1,\ldots,I. \end{aligned} \tag{68}$$

Solving problem (68) yields a set of potentially optimal solutions lying on the so-called Pareto frontier. There are typically two approaches to finding the Pareto frontier: decomposition and population-based.

Decomposition approaches convert the vector-valued objective function to a series of single-valued objective functions using constraint-based or weight-based methods. In comparison, *population-based approaches* involve the adaptation of standard population-based algorithms to iteratively improve both the accuracy and diversity of approximate Pareto frontiers. The adaptation may combine several techniques such as population partitioning, nondomination ranking, Pareto filtering, and niche methods [2]. One of the popular population-based multi-objective algorithms is the non-dominated sorting genetic algorithm (NSGA-II) [295]. There also exist several algorithms, e.g., MOEA/D [296], that decompose the problem into scalar-valued optimization subproblems which are solved simultaneously by evolving a population of candidate solutions.

Multi-objective BO (MOBO) algorithms devised for finding approximate Pareto frontiers of problem (68) can also

be classified into decomposition and population-based approaches. The former, for example, ParEGO [297], converts m objective functions to a single objective function via an m-dimensional weight vector, thereby enabling the generic BO to identify a new design point in each iteration. By sweeping over the space of possible weight vectors and employing one value in each iteration, an approximate Pareto frontier can be incrementally constructed. The limitation of this approach is that it only works well on small-dimensional problems and often fails in capturing noncovex parts of the Pareto frontier when using the weighted sum of objectives [298].

The latter views the initial set of samples generated for MOBO as a population that evolves to improve approximate Pareto frontiers. This evolution strategy often receives support from a non-dominated sorting algorithm and an acquisition function. While the sorting algorithm is to find non-dominated solutions constituting the approximate Pareto frontier in each iteration, the acquisition function is tailored to improve the quality of the approximate Pareto frontier incrementally. Some examples of MOBO's acquisition functions include the expected hypervolume improvement [299], expected value of Euclidean distance improvement [300], $\mathcal{S}$-metric [301], and Pareto-frontier entropy search [302].

Recent works have exploited MFO and MF BO approaches to multi-objective optimization. [125], via the model-then-optimize approach, constructed recursive MF surrogates for the objective functions to be optimized by a population-based multi-objective algorithm. By limiting the total computing budget allocation, [145] combined a model composition and the expected hypervolume improvement. Notably, [303] proposed a one-step look-ahead acquisition function based on a hypervolume KG that is capable of conditioning on MF evaluations and decoupling the objective evaluations. This allows independent evaluations of the objective functions at different fidelities, which can be efficient when the computational cost varies substantially between the objectives.

Future research on MF MOBO may consider the following questions:

- Can the computational cost of multi-objective acquisition functions be effectively reduced through the application of MF approaches? This is prompted by the inherent expense associated with maximizing multi-objective acquisition functions, which often involves integration over the (feasible) space of objective functions.
- What are possible impacts of varying fidelities within a fixed computational budget on the accuracy and diversity of approximate Pareto frontiers? Addressing this aspect could unveil insights into the development of more efficient MF multi-objective acquisition functions for MOBO.
- How to reason about multi-step look-ahead multi-objective acquisition functions when considering fidelities and do we really need them to further balance exploitation and exploration?

VII. Concluding remarks

MF BO is capable of facilitating costly design optimization problems by making efficient use of HF data while extracting all information we have about the mathematical models and/or the structures of the physical processes underlying these problems. As shown from the literature surveyed in this work, MF BO has achieved an important impact

on engineering design optimization, specifically when it involves running HF computational models or conducting time-consuming physical experiments. Ever since the seminal works by Huang et al. [15] and Forrester et al. [16], MF BO has continually evolved in three directions: (i) improving MF modeling methods to enhance the information transfer between HF and MF models, (ii) devising novel acquisition functions for judicious selections of both new design points and model fidelities, and (iii) applying MF BO methods to various domains of engineering design.

Specific to this work, we highlight two essential ingredients of MF BO that exist in the literature: GP-based MF surrogates and acquisition functions. We first classify the existing MF modeling methods and MFO strategies to determine where MF BO is located in the rich literature on SbO and MFO algorithms. We then survey important GP-based MF surrogates, followed by a comprehensive review of various acquisition functions of BO and how to modify them for use of MF BO. While describing the techniques from each ingredient, we attempt to exploit their common properties. For example, we show that the GMGP model is more general than the KOH model while being a type of the LMC model. We focus on the exploitation-exploration trade-off to see how it can be addressed by some popular acquisition functions, including EI, WEI, GP-NLCB, and TS, and to provide a new perspective that multi-step look-ahead acquisition functions indeed aim at improving exploration by considering the impact of future selections of new design points. Our ultimate goal is to provide a structured understanding of MF BO.

Additionally, we provide recent advances in MFO and BO to address intricate yet important design optimization problems, including constrained optimization, high-dimensional optimization, optimization under uncertainty, and multi-objective optimization. For each of these problems, we attempt to reveal important aspects that require further research for applications of MF BO. We expect to open up new avenues for MF BO, bringing it closer to being a comprehensive optimization technique.

Although they are not detailed in this work, other potential opportunities for future research on MF BO include

- MF BO for combinatorial optimization. This is a natural extension of several advanced BO algorithms that have recently been devised for solving combinatorial optimization problems. These algorithms have focused on developing covariance functions of non-continuous design variables [see e.g., 253, 304–306]. Another aspect that should be considered when using MF BO for combinatorial optimization is to find an efficient optimizer for maximizing acquisition functions in the combinatorial space of the design variables and fidelities.
- MF BO with gradients. It is shown that exploiting gradients to construct a GP for the objective function can help BO provide good solutions with fewer objective function evaluations [307].

While it is beyond the scope of this work, there is notable value in the development of open-source software that offers full-featured sublibraries for MFO, BO, and MF BO. Such a platform would facilitate in-depth comparative investigations on the performance of various MF BO techniques reviewed in this work across a wide spectrum of engineering design.

Appendix A: Univariate Gaussian process

Consider a training dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}^k, y^k\}_{k=1}^N$, where $\mathbf{x}^k \in \mathbb{R}^d$ are d-dimensional vectors of design variables and $y^k \in \mathbb{R}$ the corresponding observations of the function of interest (e.g., objective or constraint function). Based on $\mathcal{D}$, we wish to find the mapping $y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon_\mathrm{y} : \mathbb{R}^d \mapsto \mathbb{R}$, where $f(\mathbf{x})$ is a regression function conditioned on $\mathcal{D}$, and $\varepsilon_\mathrm{y} \sim \mathcal{N}(0, \sigma_\mathrm{y}^2)$ is an additive zero-mean Gaussian noise with variance σ_y^2. Note that the noise is assumed to be independent and identically distributed.

GP assumes that any finite subset of an infinite set of the values of $f(\mathbf{x})$ has a joint Gaussian distribution [13]. This assumption is encoded in the following GP prior:

$$f(\cdot) \sim \mathcal{GP}\left(m(\cdot|\boldsymbol{\beta}_\mathrm{m}), \kappa(\cdot,\cdot|\boldsymbol{\phi}_\mathrm{x})\right), \tag{A.1}$$

where $m(\cdot|\boldsymbol{\beta}_\mathrm{m})$: $\mathbb{R}^d \mapsto \mathbb{R}$ is a mean function characterized by the hyperparameter $\boldsymbol{\beta}_\mathrm{m}$ and $\kappa(\mathbf{x}, \mathbf{x}'|\boldsymbol{\phi}_\mathrm{x}) = \mathrm{cov}[f(\mathbf{x}), f(\mathbf{x}')] : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is a positive semi-definite covariance function parameterized by the hyperparameter vector $\boldsymbol{\phi}_\mathrm{x}$.

Thus, for the set of N parameter vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, the vector of the corresponding error function values $\mathbf{F}(\mathbf{X}) = [f(\mathbf{x}^1), \ldots, f(\mathbf{x}^N)]^\intercal \in \mathbb{R}^N$ is distributed according to an N-variate Gaussian with parameters determined by the mean and covariance functions, such that

$$\mathbf{F}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{M}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X})\right), \tag{A.2}$$

where $\mathbf{M}(\mathbf{X}) = [m(\mathbf{x}^1|\boldsymbol{\beta}_\mathrm{m}), \ldots, m(\mathbf{x}^N|\boldsymbol{\beta}_\mathrm{m})]^\intercal \in \mathbb{R}^N$ is the mean vector, $\mathbf{K}(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{N\times N}$ is the covariance matrix with the (i, j)th element $(\mathbf{K}(\mathbf{X}, \mathbf{X}))_{i,j} = \kappa(\mathbf{x}^i, \mathbf{x}^j|\boldsymbol{\phi}_\mathrm{x})$, $i, j = 1, \ldots, N$.

Furthermore, the covariance of the observations is

$$\mathrm{cov}[y(\mathbf{x}^i), y(\mathbf{x}^j)] = \mathrm{cov}[f(\mathbf{x}^i), f(\mathbf{x}^j)] + \mathrm{cov}[\varepsilon_\mathrm{y}^i, \varepsilon_\mathrm{y}^j] = \kappa(\mathbf{x}^i, \mathbf{x}^j|\boldsymbol{\phi}_\mathrm{x}) + \mathbb{1}_{i,j}, \tag{A.3}$$

where $\mathbb{1}_{i,j}$ denotes the Kronecker delta. Eqs. (A.2) and (A.3) lead to

$$\mathbf{Y}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{M}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\mathrm{y}^2\mathbf{I}_N\right), \tag{A.4}$$

where $\mathbf{Y}(\mathbf{X}) = [y(\mathbf{x}^1), \ldots, y(\mathbf{x}^N)]^\intercal \in \mathbb{R}^N$ is the vector of observations.

Let $\boldsymbol{\phi} = [\boldsymbol{\beta}_\mathrm{m}^\intercal, \sigma_\mathrm{y}^2, \boldsymbol{\phi}_\mathrm{x}^\intercal]^\intercal$ denote the vector of hyperparameters of the GP model. We can find the optimal value $\boldsymbol{\phi}^\star$ of $\boldsymbol{\phi}$ by maximizing the likelihood of $\mathbf{Y}(\mathbf{X})$ derived from Eq. (A.4).

Once $\boldsymbol{\phi}^\star$ has been found, we wish to predict $f(\mathbf{x}^\star)$, where $\mathbf{x}^\star$ is an unseen variable vector. The posterior predictive

PDF at $\mathbf{x}^\star$ reads

$$p\left(f(\mathbf{x}^\star)|\mathbf{x}^\star, \mathcal{D}, \boldsymbol{\phi}\right) = \mathcal{N}\left(\mu_{\mathrm{f}}(\mathbf{x}^\star), \sigma_{\mathrm{f}}^2(\mathbf{x}^\star)\right). \tag{A.5}$$

The mean and variance of this PDF are

$$\mu_{\mathrm{f}}(\mathbf{x}^\star) = m(\mathbf{x}^\star|\boldsymbol{\beta}_{\mathrm{m}}) + \mathbf{K}(\mathbf{x}^\star, \mathbf{X})^\intercal \left[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_{\mathrm{y}}^2\mathbf{I}\right]^{-1} (\mathbf{Y}(\mathbf{X}) - \mathbf{M}(\mathbf{X})), \tag{A.6}$$

$$\sigma_{\mathrm{f}}^2(\mathbf{x}^\star) = \kappa(\mathbf{x}^\star, \mathbf{x}^\star|\boldsymbol{\phi}_{\mathrm{x}}) - \mathbf{K}(\mathbf{x}^\star, \mathbf{X})^\intercal \left[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_{\mathrm{y}}^2\mathbf{I}\right]^{-1} \mathbf{K}(\mathbf{x}^\star, \mathbf{X}), \tag{A.7}$$

where

$$\mathbf{K}(\mathbf{x}^\star, \mathbf{X}) = \left[\kappa(\mathbf{x}^\star, \mathbf{x}^1|\boldsymbol{\phi}_{\mathrm{x}}), \ldots, \kappa(\mathbf{x}^\star, \mathbf{x}^N|\boldsymbol{\phi}_{\mathrm{x}})\right]^\intercal. \tag{A.8}$$

Appendix B: Multiivariate Gaussian process

We find the mapping $\mathbf{y}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}_{\mathrm{y}} : \mathbb{R}^d \mapsto \mathbb{R}^T$ to explain the training dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}^k, \mathbf{y}^k\}_{k=1}^N$, where $\mathbf{x}^k \in \mathbb{R}^d$, $\mathbf{y}^k \in \mathbb{R}^T$, $\mathbf{f}(\mathbf{x})$ is a regression function, and $\boldsymbol{\varepsilon}_{\mathrm{y}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{y}})$ is a zero-mean Gaussian noise with covariance matrix $\boldsymbol{\Sigma}_{\mathrm{y}} \in \mathbb{R}^{T\times T}$.

Multivariate GP for vector-valued functions is a natural extension of the univariate GP. The multivariate GP prior reads

$$\mathbf{f}(\cdot) \sim \mathcal{GP}\left(\mathbf{m}(\cdot|\boldsymbol{\beta}_{\mathrm{m}}), \mathbf{S}(\cdot, \cdot|\boldsymbol{\phi}_{\mathrm{x}})\right), \tag{B.1}$$

where $\mathbf{m}(\cdot|\boldsymbol{\beta}_{\mathrm{m}})$: $\mathbb{R}^d \mapsto \mathbb{R}^T$ is the vector-valued mean function characterized by the hyperparameter vector $\boldsymbol{\beta}_{\mathrm{m}}$ and $\mathbf{S}(\mathbf{x}, \mathbf{x}'|\boldsymbol{\phi}_{\mathrm{x}}) = \mathrm{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')|\boldsymbol{\phi}_{\mathrm{x}}] \in \mathbb{R}^{T\times T}$ denotes the inter-group covariance matrix.

Let $\mathbf{F}(\mathbf{X}) = [\mathbf{f}^\intercal(\mathbf{x}^1), \ldots, \mathbf{f}^\intercal(\mathbf{x}^N)]^\intercal \in \mathbb{R}^{TN}$ be a vector that concatenates the vectors $\mathbf{f}(\mathbf{x}^i)$, $i = 1, \ldots, N$. Let $\mathbf{M}(\mathbf{X}) = [\mathbf{m}^\intercal(\mathbf{x}^1|\boldsymbol{\beta}_{\mathrm{m}}), \ldots, \mathbf{m}^\intercal(\mathbf{x}^N|\boldsymbol{\beta}_{\mathrm{m}})]^\intercal \in \mathbb{R}^{TN}$ be a vector that concatenates the mean vectors $\mathbf{m}(\mathbf{x}^i)$. According to the GP prior in Eq. (B.1), the distribution of $\mathbf{F}(\mathbf{X})$ reads

$$\mathbf{F}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{M}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X})\right), \tag{B.2}$$

where the covariance matrix $\mathbf{K}(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{TN\times TN}$ is the block partitioned matrix with $(\mathbf{K}(\mathbf{X}, \mathbf{X}))_{i,j} = \mathbf{S}(\mathbf{x}^i, \mathbf{x}^j|\boldsymbol{\phi}_{\mathrm{x}})$, $i, j = 1, \ldots, N$.

Let $\mathbf{Y}(\mathbf{X}) = [\mathbf{y}^\intercal(\mathbf{x}^1), \ldots, \mathbf{y}^\intercal(\mathbf{x}^N)]^\intercal \in \mathbb{R}^{TN}$ be a vector of output observations. $\mathbf{Y}(\mathbf{X})$ is distributed according to

$$\mathbf{Y}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{M}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\mathrm{y}} \otimes \mathbf{I}_N\right), \tag{B.3}$$

where $\otimes$ denotes the Kronecker product between matrices.

By maximizing the likelihood of $\mathbf{Y}(\mathbf{X})$, we obtain the optimal set of hyperparameters $\boldsymbol{\phi} = \{\boldsymbol{\phi}_{\mathrm{x}}, \boldsymbol{\Sigma}_{\mathrm{y}}, \boldsymbol{\beta}_{\mathrm{m}}\}$. Then, the posterior predictive PDF at $\mathbf{x}^{\star}$ can be derived as

$$p\left(\mathbf{f}(\mathbf{x}^{\star})|\mathbf{x}^{\star}, \mathcal{D}, \boldsymbol{\phi}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{f}}(\mathbf{x}^{\star}), \boldsymbol{\Sigma}_{\mathrm{f}}(\mathbf{x}^{\star})\right). \tag{B.4}$$

The mean and variance of this PDF are

$$\boldsymbol{\mu}_{\mathrm{f}}(\mathbf{x}^{\star}) = \mathbf{m}(\mathbf{x}^{\star}|\boldsymbol{\beta}_{\mathrm{m}}) + \mathbf{K}(\mathbf{x}^{\star}, \mathbf{X})^{\intercal}\left[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\mathrm{y}} \otimes \mathbf{I}\right]^{-1}\left(\mathbf{Y}(\mathbf{X}) - \mathbf{M}(\mathbf{X})\right), \tag{B.5}$$

$$\boldsymbol{\Sigma}_{\mathrm{f}}(\mathbf{x}^{\star}) = \mathbf{S}(\mathbf{x}^{\star}, \mathbf{x}^{\star}|\boldsymbol{\phi}_{\mathrm{x}}) - \mathbf{K}(\mathbf{x}^{\star}, \mathbf{X})^{\intercal}\left[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\mathrm{y}} \otimes \mathbf{I}\right]^{-1}\mathbf{K}(\mathbf{x}^{\star}, \mathbf{X}), \tag{B.6}$$

where $\mathbf{K}(\mathbf{x}^{\star}, \mathbf{X}) \in \mathbb{R}^{TN \times T}$, which concatenates the following matrices:

$$\left(\mathbf{K}(\mathbf{x}^{\star}, \mathbf{X})\right)_i = \mathbf{S}(\mathbf{x}^{\star}, \mathbf{x}^{i}|\boldsymbol{\phi}_{\mathrm{x}}), \quad i = 1, \ldots, N. \tag{B.7}$$

Acknowledgments

The authors thank the University of Houston for providing startup fund to support this research.

References

[1] Arora, J. S., *Introduction to optimum design*, 4th ed., Academic Press, Massachusetts, USA, 2016.

[2] Kochenderfer, M. J., and Wheeler, T. A., *Algorithms for optimization*, MIT Press, Massachusetts, USA, 2019.

[3] Martins, J. R. R. A., and Ning, A., *Engineering Design Optimization*, Cambridge University Press, Cambridge, 2021. https://doi.org/10.1017/9781108980647.

[4] Bathe, K.-J., *Finite element procedures*, Prentice Hall, New Jersey, USA, 2006.

[5] Peherstorfer, B., Willcox, K., and Gunzburger, M., "Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization," *SIAM Review*, Vol. 60, No. 3, 2018, pp. 550–591. https://doi.org/10.1137/16M1082469.

[6] Nocedal, J., and Wright, S. J., *Numerical optimization*, 2nd ed., Springer, New York, USA, 2006. https://doi.org/10.1007/978-0-387-40065-5.

[7] Wang, G. G., and Shan, S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, 2006, pp. 370–380. https://doi.org/10.1115/1.2429697.

[8] Ohsaki, M., Miyamura, T., Kohiyama, M., Hori, M., Noguchi, H., Akiba, H., Kajiwara, K., and Ine, T., "High-precision finite element analysis of elastoplastic dynamic responses of super-high-rise steel frames," *Earthquake Engineering and Structural Dynamics*, Vol. 38, No. 5, 2009, pp. 635–654. https://doi.org/10.1002/eqe.900.

[9] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Kevin Tucker, P., "Surrogate-based analysis and optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, 2005, pp. 1–28. https://doi.org/10.1016/j.paerosci.2005.02.001.

[10] Forrester, A. I. J., Sóbester, A., and Keane, A., *Engineering design via surrogate modelling: a practical guide*, John Wiley & Sons, West Sussex, UK, 2008.

[11] Simpson, T., Toropov, V., Balabanov, V., and Viana, F., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - Or Not," *The 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, No. 5802 in Multidisciplinary Analysis Optimization Conferences, American Institute of Aeronautics and Astronautics, 2008. https://doi.org/10.2514/6.2008-5802.

[12] Forrester, A. I. J., and Keane, A. J., "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, Vol. 45, No. 1, 2009, pp. 50–79. https://doi.org/10.1016/j.paerosci.2008.11.001.

[13] Rasmussen, C. E., and Williams, C. K. I., *Gaussian processes for machine learning*, The MIT Press, Massachusetts, USA, 2006. https://doi.org/10.7551/mitpress/3206.001.0001.

[14] Alexandrov, N. M., Dennis, J. E., Lewis, R. M., and Torczon, V., "A trust-region framework for managing the use of approximation models in optimization," *Structural optimization*, Vol. 15, No. 1, 1998, pp. 16–23. https://doi.org/10.1007/BF01197433.

[15] Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A., "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, Vol. 32, No. 5, 2006, pp. 369–382. https://doi.org/10.1007/s00158-005-0587-0.

[16] Forrester, A. I. J., Sóbester, A., and Keane, A. J., "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 463, No. 2088, 2007, pp. 3251–3269. https://doi.org/10.1098/rspa.2007.1900.

[17] Viana, F. A. C., Simpson, T. W., Balabanov, V., and Toropov, V., "Special Section on Multidisciplinary Design Optimization: Metamodeling in Multidisciplinary Design Optimization: How Far Have We Really Come?" *AIAA Journal*, Vol. 52, No. 4, 2014, pp. 670–690. https://doi.org/10.2514/1.J052375.

[18] Kennedy, M. C., and O'Hagan, A., "Predicting the Output from a Complex Computer Code When Fast Approximations Are Available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13. https://doi.org/10.1093/biomet/87.1.1.

[19] Han, Z.-H., and Görtz, S., "Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 9, 2012, pp. 1885–1896. https://doi.org/10.2514/1.J051354.

[20] Gratiet, L. L., and Garnier, J., "Recursive cokriging model for design ofcomputer experiments with multiple levels of fidelity," *International Journal for Uncertainty Quantification*, Vol. 4, No. 5, 2014, pp. 365–386. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2014006914.

[21] Bandler, J. W., Biernacki, R. M., Chen, S. H., Grobelny, P. A., and Hemmers, R. H., "Space mapping technique for electromagnetic optimization," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 42, No. 12, 1994, pp. 2536–2544. https://doi.org/10.1109/22.339794.

[22] Bandler, J. W., Cheng, Q. S., Nikolova, N. K., and Ismail, M. A., "Implicit space mapping optimization exploiting preassigned parameters," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 52, No. 1, 2004, pp. 378–385. https://doi.org/10.1109/TMTT.2003.820892.

[23] Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., and Karniadakis, G. E., "Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 473, No. 2198, 2017, p. 20160751. https://doi.org/10.1098/rspa.2016.0751.

[24] Snoek, J., Larochelle, H., and Adams, R. P., "Practical Bayesian Optimization of Machine Learning Algorithms," *Advances in Neural Information Processing Systems*, Vol. 25, NIPS 2012, 2012, pp. 2951–2959.

[25] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N., "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, Vol. 104, No. 1, 2016, pp. 148–175. https://doi.org/10.1109/JPROC.2015.2494218.

[26] Frazier, P. I., "Bayesian Optimization," *Recent Advances in Optimization and Modeling of Contemporary Problems*, INFORMS TutORials in Operations Research, INFORMS, 2018, Chap. 11, pp. 255–278. https://doi.org/10.1287/educ.2018.0188.

[27] Meliani, M., Bartoli, N., Lefebvre, T., Bouhlel, M.-A., Martins, J. R. R. A., and Morlier, J., "Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design," *AIAA Aviation 2019 Forum*, No. 3236 in AIAA AVIATION Forum, American Institute of Aeronautics and Astronautics, 2019. https://doi.org/10.2514/6.2019-3236.

[28] Tran, A., Wildey, T., and McCann, S., "sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization framework for design applications," *Journal of Computing and Information Science in Engineering*, Vol. 20, No. 3, 2020. https://doi.org/10.1115/1.4046697.

[29] Hebbal, A., Brevault, L., Balesdent, M., Talbi, E.-G., and Melab, N., "Bayesian optimization using deep Gaussian processes with applications to aerospace system design," *Optimization and Engineering*, Vol. 22, No. 1, 2021, pp. 321–361. https://doi.org/10.1007/s11081-020-09517-8.

[30] Khatamsaz, D., Molkeri, A., Couperthwaite, R., James, J., Arróyave, R., Srivastava, A., and Allaire, D., "Adaptive active subspace-based efficient multifidelity materials design," *Materials & Design*, Vol. 209, 2021, p. 110001. https://doi.org/10.1016/j.matdes.2021.110001.

[31] Wang, X., Jin, Y., Schmitt, S., and Olhofer, M., "Recent advances in Bayesian optimization," *ACM Computing Surveys*, Vol. 55, No. 13s, 2023. https://doi.org/10.1145/3582078.

[32] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–423. https://doi.org/10.1214/ss/1177012413.

[33] Barthelemy, J. F. M., and Haftka, R. T., "Approximation concepts for optimum structural design – a review," *Structural optimization*, Vol. 5, No. 3, 1993, pp. 129–144. https://doi.org/10.1007/BF01743349.

[34] Box, G. E. P., and Wilson, K. B., "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 13, No. 1, 1951, pp. 1–45.

[35] Sobieszczanski-Sobieski, J., and Haftka, R. T., "Multidisciplinary aerospace design optimization: survey of recent developments," *Structural optimization*, Vol. 14, No. 1, 1997, pp. 1–23. https://doi.org/10.1007/BF01197554.

[36] Hussain, M. F., Barton, R. R., and Joshi, S. B., "Metamodeling: Radial basis functions, versus polynomials," *European Journal of Operational Research*, Vol. 138, No. 1, 2002, pp. 142–154. https://doi.org/10.1016/S0377-2217(01)00076-5.

[37] Girosi, F., "An Equivalence Between Sparse Approximation and Support Vector Machines," *Neural Computation*, Vol. 10, No. 6, 1998, pp. 1455–1480. https://doi.org/10.1162/089976698300017269.

[38] Cressie, N., "The origins of kriging," *Mathematical Geology*, Vol. 22, No. 3, 1990, pp. 239–252. https://doi.org/10.1007/BF00889887.

[39] Kleijnen, J. P. C., "Kriging metamodeling in simulation: A review," *European Journal of Operational Research*, Vol. 192, No. 3, 2009, pp. 707–716. https://doi.org/10.1016/j.ejor.2007.10.013.

[40] Antoulas, A. C., *Approximation of Large-Scale Dynamical Systems*, Society for Industrial and Applied Mathematics, 2005. https://doi.org/10.1137/1.9780898718713.

[41] Goel, T., Haftka, R. T., Shyy, W., and Queipo, N. V., "Ensemble of surrogates," *Structural and Multidisciplinary Optimization*, Vol. 33, No. 3, 2007, pp. 199–216. https://doi.org/10.1007/s00158-006-0051-9.

[42] Liu, H., Ong, Y. S., Shen, X., and Cai, J., "When Gaussian process meets big data: A review of scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 11, 2020, pp. 4405–4423. https://doi.org/10.1109/TNNLS.2019.2957109.

[43] Haftka, R. T., "Combining global and local approximations," *AIAA Journal*, Vol. 29, No. 9, 1991, pp. 1523–1525. https://doi.org/10.2514/3.10768.

[44] Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., "Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models," *Journal of Aircraft*, Vol. 38, No. 6, 2001, pp. 1093–1101. https://doi.org/10.2514/2.2877.

[45] Rodríguez, J. F., Pérez, V. M., Padmanabhan, D., and Renaud, J. E., "Sequential approximate optimization using variable fidelity response surface approximations," *Structural and Multidisciplinary Optimization*, Vol. 22, No. 1, 2001, pp. 24–34. https://doi.org/10.1007/s001580100122.

[46] Knill, D. L., Giunta, A. A., Baker, C. A., Grossman, B., Mason, W. H., Haftka, R. T., and Watson, L. T., "Response Surface Models Combining Linear and Euler Aerodynamics for Supersonic Transport Design," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 75–86. https://doi.org/10.2514/2.2415.

[47] Gano, S. E., Renaud, J. E., and Sanders, B., "Hybrid Variable Fidelity Optimization by Using a Kriging-Based Scaling Function," *AIAA Journal*, Vol. 43, No. 11, 2005, pp. 2422–2433. https://doi.org/10.2514/1.12466.

[48] Fernández-Godino, M. G., Park, C., Kim, N.-H., and Haftka, R. T., "Review of multi-fidelity models," *arXiv preprint arXiv.1609.07196*, 2016. No published.

[49] Lewis, R., and Nash, S., "A multigrid approach to the optimization of systems governed by differential equations," *The 8th Symposium on Multidisciplinary Analysis and Optimization*, No. 4890 in Multidisciplinary Analysis Optimization Conferences, American Institute of Aeronautics and Astronautics, 2000. https://doi.org/10.2514/6.2000-4890.

[50] Gano, S. E., Renaud, J. E., Martin, J. D., and Simpson, T. W., "Update strategies for kriging models used in variable fidelity optimization," *Structural and Multidisciplinary Optimization*, Vol. 32, No. 4, 2006, pp. 287–298. https://doi.org/10.1007/s00158-006-0025-y.

[51] Viana, F. A. C., Steffen, V., Butkewitsch, S., and de Freitas Leal, M., "Optimization of aircraft structural components by using nature-inspired algorithms and multi-fidelity approximations," *Journal of Global Optimization*, Vol. 45, No. 3, 2009, pp. 427–449. https://doi.org/10.1007/s10898-008-9383-x.

[52] Palar, P. S., Tsuchiya, T., and Parks, G. T., "Multi-fidelity non-intrusive polynomial chaos based on regression," *Computer Methods in Applied Mechanics and Engineering*, Vol. 305, 2016, pp. 579–606. https://doi.org/10.1016/j.cma.2016.03.022.

[53] Zhang, Y., Kim, N. H., Park, C., and Haftka, R. T., "Multifidelity Surrogate Based on Single Linear Regression," *AIAA Journal*, Vol. 56, No. 12, 2018, pp. 4944–4952. https://doi.org/10.2514/1.J057299.

[54] Fernández-Godino, M. G., Dubreuil, S., Bartoli, N., Gogu, C., Balachandar, S., and Haftka, R. T., "Linear regression-based multifidelity surrogate for disturbance amplification in multiphase explosion," *Structural and Multidisciplinary Optimization*, Vol. 60, No. 6, 2019, pp. 2205–2220. https://doi.org/10.1007/s00158-019-02387-4.

[55] Song, X., Lv, L., Sun, W., and Zhang, J., "A radial basis function-based multi-fidelity surrogate model: exploring correlation between high-fidelity and low-fidelity models," *Structural and Multidisciplinary Optimization*, Vol. 60, No. 3, 2019, pp. 965–981. https://doi.org/10.1007/s00158-019-02248-0.

[56] Kou, J., and Zhang, W., "Multi-fidelity modeling framework for nonlinear unsteady aerodynamics of airfoils," *Applied Mathematical Modelling*, Vol. 76, 2019, pp. 832–855. https://doi.org/10.1016/j.apm.2019.06.034.

[57] Meng, X., and Karniadakis, G. E., "A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems," *Journal of Computational Physics*, Vol. 401, 2020, p. 109020. https://doi.org/10.1016/j.jcp.2019.109020.

[58] Durantin, C., Rouxel, J., Désidéri, J.-A., and Glière, A., "Multifidelity surrogate modeling based on radial basis functions," *Structural and Multidisciplinary Optimization*, Vol. 56, No. 5, 2017, pp. 1061–1075. https://doi.org/10.1007/s00158-017-1703-7.

[59] Teichert, G. H., and Garikipati, K., "Machine learning materials physics: Surrogate optimization and multi-fidelity algorithms predict precipitate morphology in an alternative to phase field dynamics," *Computer Methods in Applied Mechanics and Engineering*, Vol. 344, 2019, pp. 666–693. https://doi.org/10.1016/j.cma.2018.10.025.

[60] Yan, L., and Zhou, T., "Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems," *Journal of Computational Physics*, Vol. 381, 2019, pp. 110–128. https://doi.org/10.1016/j.jcp.2018.12.025.

[61] Leary, S. J., Bhaskar, A., and Keane, A. J., "A Knowledge-Based Approach To Response Surface Modelling in Multifidelity Optimization," *Journal of Global Optimization*, Vol. 26, No. 3, 2003, pp. 297–319. https://doi.org/10.1023/A:1023283917997.

[62] Xiong, Y., Chen, W., and Tsui, K.-L., "A New Variable-Fidelity Optimization Framework Based on Model Fusion and Objective-Oriented Sequential Sampling," *Journal of Mechanical Design*, Vol. 130, No. 11, 2008. https://doi.org/10.1115/1.2976449.

[63] Kuya, Y., Takeda, K., Zhang, X., and Forrester, A. I. J., "Multifidelity Surrogate Modeling of Experimental and Computational Aerodynamic Data Sets," *AIAA Journal*, Vol. 49, No. 2, 2011, pp. 289–298. https://doi.org/10.2514/1.J050384.

[64] Toal, D. J. J., and Keane, A. J., "Efficient Multipoint Aerodynamic Design Optimization Via Cokriging," *Journal of Aircraft*, Vol. 48, No. 5, 2011, pp. 1685–1695. https://doi.org/10.2514/1.C031342.

[65] Keane, A. J., "Cokriging for Robust Design Optimization," *AIAA Journal*, Vol. 50, No. 11, 2012, pp. 2351–2364. https://doi.org/10.2514/1.J051391.

[66] Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., and Rutter, E., "Prediction and Computer Model Calibration Using Outputs From Multifidelity Simulators," *Technometrics*, Vol. 55, No. 4, 2013, pp. 501–512. https://doi.org/10.1080/00401706.2013.838910.

[67] Zheng, J., Shao, X., Gao, L., Jiang, P., and Li, Z., "A hybrid variable-fidelity global approximation modelling method combining tuned radial basis function base and kriging correction," *Journal of Engineering Design*, Vol. 24, No. 8, 2013, pp. 604–622. https://doi.org/10.1080/09544828.2013.788135.

[68] de Baar, J., Roberts, S., Dwight, R., and Mallol, B., "Uncertainty quantification for a sailing yacht hull, using multi-fidelity kriging," *Computers & Fluids*, Vol. 123, 2015, pp. 185–201. https://doi.org/10.1016/j.compfluid.2015.10.004.

[69] Park, C., Haftka, R. T., and Kim, N. H., "Remarks on multi-fidelity surrogates," *Structural and Multidisciplinary Optimization*, Vol. 55, No. 3, 2017, pp. 1029–1050. https://doi.org/10.1007/s00158-016-1550-y.

[70] Zhang, Y., Han, Z.-H., and Zhang, K.-S., "Variable-fidelity expected improvement method for efficient global optimization of expensive functions," *Structural and Multidisciplinary Optimization*, Vol. 58, No. 4, 2018, pp. 1431–1451. https://doi.org/10.1007/s00158-018-1971-x.

[71] Rokita, T., and Friedmann, P. P., "Multifidelity coKriging for High-Dimensional Output Functions with Application to Hypersonic Airloads Computation," *AIAA Journal*, Vol. 56, No. 8, 2018, pp. 3060–3070. https://doi.org/10.2514/1.J056620.

[72] Xiao, M., Zhang, G., Breitkopf, P., Villon, P., and Zhang, W., "Extended Co-Kriging interpolation method based on multi-fidelity data," *Applied Mathematics and Computation*, Vol. 323, 2018, pp. 120–131. https://doi.org/10.1016/j.amc.2017.10.055.

[73] Jiang, P., Cheng, J., Zhou, Q., Shu, L., and Hu, J., "Variable-Fidelity Lower Confidence Bounding Approach for Engineering Optimization Problems with Expensive Simulations," *AIAA Journal*, Vol. 57, No. 12, 2019, pp. 5416–5430. https://doi.org/10.2514/1.J058283.

[74] Serani, A., Pellegrini, R., Wackers, J., Jeanson, C.-E., Queutey, P., Visonneau, M., and Diez, M., "Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels," *International Journal of Computational Fluid Dynamics*, Vol. 33, No. 6-7, 2019, pp. 237–255. https://doi.org/10.1080/10618562.2019.1683164.

[75] Zhou, Q., Wu, Y., Guo, Z., Hu, J., and Jin, P., "A generalized hierarchical co-Kriging model for multi-fidelity data fusion," *Structural and Multidisciplinary Optimization*, Vol. 62, No. 4, 2020, pp. 1885–1904. https://doi.org/10.1007/s00158-020-02583-7.

[76] Shu, L., Jiang, P., and Wang, Y., "A multi-fidelity Bayesian optimization approach based on the expected further improvement," *Structural and Multidisciplinary Optimization*, Vol. 63, No. 4, 2021, pp. 1709–1719. https://doi.org/10.1007/s00158-020-02772-4.

[77] Kaps, A., Czech, C., and Duddeck, F., "A hierarchical kriging approach for multi-fidelity optimization of automotive crashworthiness problems," *Structural and Multidisciplinary Optimization*, Vol. 65, No. 4, 2022, p. 114. https://doi.org/10.1007/s00158-022-03211-2.

[78] Toal, D. J. J., "Applications of multi-fidelity multi-output Kriging to engineering design optimization," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 6, 2023, p. 125. https://doi.org/10.1007/s00158-023-03567-z.

[79] Xu, K., Shu, L., Zhong, L., Jiang, P., and Zhou, Q., "A bi-fidelity Bayesian optimization method for multi-objective optimization with a novel acquisition function," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 3, 2023, p. 53. https://doi.org/10.1007/s00158-023-03509-9.

[80] Ribeiro, L. G., Parente, E., and de Melo, A. M. C., "Alternative variable-fidelity acquisition functions for efficient global optimization of black-box functions," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 7, 2023, p. 147. https://doi.org/10.1007/s00158-023-03607-8.

[81] Peng, X., Kou, J., and Zhang, W., "Multi-fidelity nonlinear unsteady aerodynamic modeling and uncertainty estimation based on Herarchical Kriging," *Applied Mathematical Modelling*, Vol. 122, 2023, pp. 1–21. https://doi.org/10.1016/j.apm.2023.05.031.

[82] Wiangkham, A., Ariyarit, A., Timtong, A., and Aengchuan, P., "Multi-fidelity model using GRNN and ANFIS algorithms-based fracture criterion for predicting mixed-mode I-II of sugarcane leaves/epoxy composite," *Theoretical and Applied Fracture Mechanics*, Vol. 125, 2023, p. 103892. https://doi.org/10.1016/j.tafmec.2023.103892.

[83] Chang, K. J., Haftka, R. T., Giles, G. L., and Kao, P.-J., "Sensitivity-based scaling for approximating structural response," *Journal of Aircraft*, Vol. 30, No. 2, 1993, pp. 283–288. https://doi.org/10.2514/3.48278.

[84] Goldfeld, Y., Vervenne, K., Arbocz, J., and van Keulen, F., "Multi-fidelity optimization of laminated conical shells for buckling," *Structural and Multidisciplinary Optimization*, Vol. 30, No. 2, 2005, pp. 128–141. https://doi.org/10.1007/s00158-004-0506-9.

[85] Hino, R., Yoshida, F., and Toropov, V. V., "Optimum blank design for sheet metal forming based on the interaction of high- and low-fidelity FE models," *Archive of Applied Mechanics*, Vol. 75, No. 10, 2006, pp. 679–691. https://doi.org/10.1007/s00419-006-0047-3.

[86] Sun, G., Li, G., Stone, M., and Li, Q., "A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials," *Computational Materials Science*, Vol. 49, No. 3, 2010, pp. 500–511. https://doi.org/10.1016/j.commatsci.2010.05.041.

[87] Gano, S. E., Renaud, J. E., Agarwal, H., and Tovar, A., "Reliability-based design using variable-fidelity optimization," *Structure and Infrastructure Engineering*, Vol. 2, No. 3-4, 2006, pp. 247–260. https://doi.org/10.1080/15732470600590408.

[88] Han, Z.-H., Görtz, S., and Zimmermann, R., "Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function," *Aerospace Science and Technology*, Vol. 25, No. 1, 2013, pp. 177–189. https://doi.org/10.1016/j.ast.2012.01.006.

[89] Tyan, M., Nguyen, N. V., and Lee, J.-W., "Improving variable-fidelity modelling by exploring global design space and radial basis function networks for aerofoil design," *Engineering Optimization*, Vol. 47, No. 7, 2015, pp. 885–908. https://doi.org/10.1080/0305215X.2014.941290.

[90] Nguyen, N. V., Tyan, M., and Lee, J.-W., "A modified variable complexity modeling for efficient multidisciplinary aircraft conceptual design," *Optimization and Engineering*, Vol. 16, No. 2, 2015, pp. 483–505. https://doi.org/10.1007/s11081-014-9273-7.

[91] Hu, D., Mao, J., Wang, R., Jia, Z., and Song, J., "Optimization Strategy for a Shrouded Turbine Blade Using Variable-Complexity Modeling Methodology," *AIAA Journal*, Vol. 54, No. 9, 2016, pp. 2808–2818. https://doi.org/10.2514/1.J054742.

[92] Absi, G. N., and Mahadevan, S., "Multi-fidelity approach to dynamics model calibration," *Mechanical Systems and Signal Processing*, Vol. 68-69, 2016, pp. 189–206. https://doi.org/10.1016/j.ymssp.2015.07.019.

[93] Rumpfkeil, M. P., and Beran, P. S., "Construction of Dynamic Multifidelity Locally Optimized Surrogate Models," *AIAA Journal*, Vol. 55, No. 9, 2017, pp. 3169–3179. https://doi.org/10.2514/1.J055834.

[94] Bryson, D. E., and Rumpfkeil, M. P., "All-at-once approach to multifidelity polynomial chaos expansion surrogate modeling," *Aerospace Science and Technology*, Vol. 70, 2017, pp. 121–136. https://doi.org/10.1016/j.ast.2017.07.043.

[95] Rumpfkeil, M. P., and Beran, P. S., "Multifidelity Sparse Polynomial Chaos Surrogate Models Applied to Flutter Databases," *AIAA Journal*, Vol. 58, No. 3, 2019, pp. 1292–1303. https://doi.org/10.2514/1.J058452.

[96] Wang, S., Liu, Y., Zhou, Q., Yuan, Y., Lv, L., and Song, X., "A multi-fidelity surrogate model based on moving least squares: fusing different fidelity data for engineering design," *Structural and Multidisciplinary Optimization*, Vol. 64, No. 6, 2021, pp. 3637–3652. https://doi.org/10.1007/s00158-021-03044-5.

[97] Qian, P. Z. G., and Wu, C. F. J., "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments," *Technometrics*, Vol. 50, No. 2, 2008, pp. 192–204. https://doi.org/10.1198/004017008000000082.

[98] Gratiet, L. L., "Bayesian Analysis of Hierarchical Multifidelity Codes," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 1, No. 1, 2013, pp. 244–269. https://doi.org/10.1137/120884122, submitted 2012-07-10; accepted 2013-06-10; online 2013-08-22.

[99] Parussini, L., Venturi, D., Perdikaris, P., and Karniadakis, G. E., "Multi-fidelity Gaussian process regression for prediction of random fields," *Journal of Computational Physics*, Vol. 336, 2017, pp. 36–50. https://doi.org/10.1016/j.jcp.2017.01.047.

[100] Hao, P., Feng, S., Li, Y., Wang, B., and Chen, H., "Adaptive infill sampling criterion for multi-fidelity gradient-enhanced kriging model," *Structural and Multidisciplinary Optimization*, Vol. 62, No. 1, 2020, pp. 353–373. https://doi.org/10.1007/s00158-020-02493-8.

[101] Ji, Y., Mak, S., Soeder, D., Paquet, J. F., and Bass, S. A., "A graphical multi-fidelity Gaussian process model, with application to emulation of heavy-ion collisions," *Technometrics*, Vol. 0, No. ja, 2023, pp. 1–25. https://doi.org/10.1080/00401706.2023.2281940.

[102] Cheng, M., Jiang, P., Hu, J., Shu, L., and Zhou, Q., "A multi-fidelity surrogate modeling method based on variance-weighted sum for the fusion of multiple non-hierarchical low-fidelity data," *Structural and Multidisciplinary Optimization*, Vol. 64, No. 6, 2021, pp. 3797–3818. https://doi.org/10.1007/s00158-021-03055-2.

[103] Liu, B., Koziel, S., and Zhang, Q., "A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems," *Journal of Computational Science*, Vol. 12, 2016, pp. 28–37. https://doi.org/10.1016/j.jocs.2015.11.004.

[104] Allaire, D., and Willcox, K., "A Mathematical and Computational Framework for Multifidelity Design and Analysis with Computer Models," *International Journal for Uncertainty Quantification*, Vol. 4, No. 1, 2014, pp. 1–20. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2013004121.

[105] Zhou, Q., Wang, Y., Choi, S.-K., Jiang, P., Shao, X., and Hu, J., "A sequential multi-fidelity metamodeling approach for data regression," *Knowledge-Based Systems*, Vol. 134, 2017, pp. 199–212. https://doi.org/10.1016/j.knosys.2017.07.033.

[106] Koziel, S., Bandler, J. W., and Madsen, K., "A space-mapping framework for engineering optimization–Theory and implementation," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 54, No. 10, 2006, pp. 3721–3730. https://doi.org/10.1109/TMTT.2006.882894.

[107] Robinson, T. D., Eldred, M. S., Willcox, K., and Haimes, R., "Surrogate-Based Optimization Using Multifidelity Models with Variable Parameterization and Corrected Space Mapping," *AIAA Journal*, Vol. 46, No. 11, 2008, pp. 2814–2822. https://doi.org/10.2514/1.36043.

[108] Tao, S., Apley, D. W., Chen, W., Garbo, A., Pate, D. J., and German, B. J., "Input Mapping for Model Calibration with Application to Wing Aerodynamics," *AIAA Journal*, Vol. 57, No. 7, 2019, pp. 2734–2745. https://doi.org/10.2514/1.J057711.

[109] Zheng, J., Shao, X., Gao, L., Jiang, P., and Qiu, H., "A prior-knowledge input LSSVR metamodeling method with tuning based on cellular particle swarm optimization for engineering design," *Expert Systems with Applications*, Vol. 41, No. 5, 2014, pp. 2111–2125. https://doi.org/10.1016/j.eswa.2013.09.010.

[110] Zhou, Q., Jiang, P., Shao, X., Hu, J., Cao, L., and Wan, L., "A variable fidelity information fusion method based on radial basis function," *Advanced Engineering Informatics*, Vol. 32, 2017, pp. 26–39. https://doi.org/10.1016/j.aei.2016.12.005.

[111] Jiang, P., Xie, T., Zhou, Q., Shao, X., Hu, J., and Cao, L., "A space mapping method based on Gaussian process model for variable fidelity metamodeling," *Simulation Modelling Practice and Theory*, Vol. 81, 2018, pp. 64–84. https://doi.org/10.1016/j.simpat.2017.11.010.

[112] Cutajar, K., Pullin, M., Damianou, A., Lawrence, N., and González, J., "Deep Gaussian processes for multi-fidelity modeling," *Third workshop on Bayesian Deep Learning (NeurIPS 2018)*, Montréal, Canada, 2019.

[113] Hebbal, A., Brevault, L., Balesdent, M., Talbi, E.-G., and Melab, N., "Multi-fidelity modeling with different input domain definitions using deep Gaussian processes," *Structural and Multidisciplinary Optimization*, Vol. 63, No. 5, 2021, pp. 2267–2288. https://doi.org/10.1007/s00158-020-02802-1.

[114] Li, K., Li, Q., Lv, L., Song, X., Ma, Y., and Lee, I., "A nonlinearity integrated bi-fidelity surrogate model based on nonlinear mapping," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 9, 2023, p. 196. https://doi.org/10.1007/s00158-023-03633-6.

[115] Brevault, L., Balesdent, M., and Hebbal, A., "Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems," *Aerospace Science and Technology*, Vol. 107, 2020, p. 106339. https://doi.org/10.1016/j.ast.2020.106339.

[116] Kandasamy, K., Dasarathy, G., Schneider, J., and Póczos, B., "Multi-fidelity Bayesian Optimisation with Continuous Approximations," *Proceedings of the 34th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 70, edited by D. Precup and Y. W. Teh, PMLR, 2017, pp. 1799–1808.

[117] Alexandrov, N., Nielsen, E., Lewis, R., and Anderson, W., "First-order model management with variable-fidelity physics applied to multi-element airfoil optimization," *The 8th Symposium on Multidisciplinary Analysis and Optimization*, No. 4886 in Multidisciplinary Analysis Optimization Conferences, American Institute of Aeronautics and Astronautics, 2000. https://doi.org/10.2514/6.2000-4886.

[118] March, A., Willcox, K., and Wang, Q., "Gradient-based multifidelity optimisation for aircraft design using Bayesian model calibration," *The Aeronautical Journal*, Vol. 115, No. 1174, 2011, pp. 729–738. https://doi.org/10.1017/S0001924000006473.

[119] Elham, A., "Adjoint quasi-three-dimensional aerodynamic solver for multi-fidelity wing aerodynamic shape optimization," *Aerospace Science and Technology*, Vol. 41, 2015, pp. 241–249. https://doi.org/10.1016/j.ast.2014.12.024.

[120] Bryson, D. E., and Rumpfkeil, M. P., "Multifidelity Quasi-Newton Method for Design Optimization," *AIAA Journal*, Vol. 56, No. 10, 2018, pp. 4074–4086. https://doi.org/10.2514/1.J056840.

[121] De, S., Maute, K., and Doostan, A., "Bi-fidelity stochastic gradient descent for structural optimization under uncertainty," *Computational Mechanics*, Vol. 66, No. 4, 2020, pp. 745–771. https://doi.org/10.1007/s00466-020-01870-w.

[122] Wu, N., Mader, C. A., and Martins, J. R. R. A., "A Gradient-based Sequential Multifidelity Approach to Multidisciplinary Design Optimization," *Structural and Multidisciplinary Optimization*, Vol. 65, No. 4, 2022, p. 131. https://doi.org/10.1007/s00158-022-03204-1.

[123] Zhang, T., Barakos, G. N., Furqan, and Foster, M., "Multi-fidelity aerodynamic design and analysis of propellers for a heavy-lift eVTOL," *Aerospace Science and Technology*, Vol. 135, 2023, p. 108185. https://doi.org/10.1016/j.ast.2023.108185.

[124] Leusink, D., Alfano, D., and Cinnella, P., "Multi-fidelity optimization strategy for the industrial aerodynamic design of helicopter rotor blades," *Aerospace Science and Technology*, Vol. 42, 2015, pp. 136–147. https://doi.org/10.1016/j.ast.2015.01.005.

[125] Singh, P., Couckuyt, I., Elsayed, K., Deschrijver, D., and Dhaene, T., "Multi-objective Geometry Optimization of a Gas Cyclone Using Triple-Fidelity Co-Kriging Surrogate Models," *Journal of Optimization Theory and Applications*, Vol. 175, No. 1, 2017, pp. 172–193. https://doi.org/10.1007/s10957-017-1114-3.

[126] Yang, Y., Gao, Z., and Cao, L., "Identifying optimal process parameters in deep penetration laser welding by adopting Hierarchical-Kriging model," *Infrared Physics & Technology*, Vol. 92, 2018, pp. 443–453. https://doi.org/10.1016/j.infrared.2018.07.006.

[127] Sóbester, A., Leary, S. J., and Keane, A. J., "A parallel updating scheme for approximating and optimizing high fidelity computer simulations," *Structural and Multidisciplinary Optimization*, Vol. 27, No. 5, 2004, pp. 371–383. https://doi.org/10.1007/s00158-004-0397-9.

[128] Perdikaris, P., and Karniadakis, G. E., "Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond," *Journal of The Royal Society Interface*, Vol. 13, No. 118, 2016, p. 20151107. https://doi.org/10.1098/rsif.2015.1107.

[129] Chen, S., Jiang, Z., Yang, S., and Chen, W., "Multimodel Fusion Based Sequential Optimization," *AIAA Journal*, Vol. 55, No. 1, 2016, pp. 241–254. https://doi.org/10.2514/1.J054729.

[130] Pang, G., Perdikaris, P., Cai, W., and Karniadakis, G. E., "Discovering variable fractional orders of advection-dispersion equations from field data using multi-fidelity Bayesian optimization," *Journal of Computational Physics*, Vol. 348, 2017, pp. 694–714. https://doi.org/10.1016/j.jcp.2017.07.052.

[131] Amrit, A., Leifsson, L., and Koziel, S., "Multi-fidelity aerodynamic design trade-off exploration using point-by-point Pareto set identification," *Aerospace Science and Technology*, Vol. 79, 2018, pp. 399–412. https://doi.org/10.1016/j.ast.2018.05.023.

[132] Bonfiglio, L., Perdikaris, P., del Águila, J., and Karniadakis, G. E., "A probabilistic framework for multidisciplinary design: Application to the hydrostructural optimization of supercavitating hydrofoils," *International Journal for Numerical Methods in Engineering*, Vol. 116, No. 4, 2018, pp. 246–269. https://doi.org/10.1002/nme.5923.

[133] Bonfiglio, L., Perdikaris, P., Brizzolara, S., and Karniadakis, G. E., "Multi-fidelity optimization of super-cavitating hydrofoils," *Computer Methods in Applied Mechanics and Engineering*, Vol. 332, 2018, pp. 63–85. https://doi.org/10.1016/j.cma.2017.12.009.

[134] Ghoreishi, S. F., and Allaire, D., "Multi-information source constrained Bayesian optimization," *Structural and Multidisciplinary Optimization*, Vol. 59, No. 3, 2019, pp. 977–991. https://doi.org/10.1007/s00158-018-2115-z.

[135] Bailly, J., and Bailly, D., "Multifidelity Aerodynamic Optimization of a Helicopter Rotor Blade," *AIAA Journal*, Vol. 57, No. 8, 2019, pp. 3132–3144. https://doi.org/10.2514/1.J056513.

[136] Shi, R., Liu, L., Long, T., Wu, Y., and Gary Wang, G., "Multi-Fidelity Modeling and Adaptive Co-Kriging-Based Optimization for All-Electric Geostationary Orbit Satellite Systems," *Journal of Mechanical Design*, Vol. 142, No. 2, 2020. https://doi.org/10.1115/1.4044321.

[137] Kontogiannis, S. G., Demange, J., Savill, A. M., and Kipouros, T., "A comparison study of two multifidelity methods for aerodynamic optimization," *Aerospace Science and Technology*, Vol. 97, 2020, p. 105592. https://doi.org/10.1016/j.ast.2019.105592.

[138] Tran, A., Tranchida, J., Wildey, T., and Thompson, A. P., "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys," *The Journal of Chemical Physics*, Vol. 153, No. 7, 2020, p. 074705. https://doi.org/10.1063/5.0015672.

[139] Ruan, X., Jiang, P., Zhou, Q., Hu, J., and Shu, L., "Variable-fidelity probability of improvement method for efficient global optimization of expensive black-box problems," *Structural and Multidisciplinary Optimization*, Vol. 62, No. 6, 2020, pp. 3021–3052. https://doi.org/10.1007/s00158-020-02646-9.

[140] Nachar, S., Boucard, P.-A., Néron, D., and Rey, C., "Multi-fidelity Bayesian optimization using model-order reduction for viscoplastic structures," *Finite Elements in Analysis and Design*, Vol. 176, 2020, p. 103400. https://doi.org/10.1016/j.finel.2020.103400.

[141] Fiore, D. F., Maggiore, P., and Mainini, L., "Multifidelity domain-aware learning for the design of re-entry vehicles," *Structural and Multidisciplinary Optimization*, Vol. 64, No. 5, 2021, pp. 3017–3035. https://doi.org/10.1007/s00158-021-03037-4.

[142] Wu, Y., Lin, Q., Zhou, Q., Hu, J., Wang, S., and Peng, Y., "An adaptive space preselection method for the multi-fidelity global optimization," *Aerospace Science and Technology*, Vol. 113, 2021, p. 106728. https://doi.org/10.1016/j.ast.2021.106728.

[143] He, Y., Sun, J., Song, P., and Wang, X., "Variable-fidelity expected improvement based efficient global optimization of expensive problems in presence of simulation failures and its parallelization," *Aerospace Science and Technology*, Vol. 111, 2021, p. 106572. https://doi.org/10.1016/j.ast.2021.106572.

[144] Zhang, X., Xie, F., Ji, T., Zhu, Z., and Zheng, Y., "Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization," *Computer Methods in Applied Mechanics and Engineering*, Vol. 373, 2021, p. 113485. https://doi.org/10.1016/j.cma.2020.113485.

[145] Khatamsaz, D., Peddareddygari, L., Friedman, S., and Allaire, D., "Bayesian Optimization of Multiobjective Functions Using Multiple Information Sources," *AIAA Journal*, Vol. 59, No. 6, 2021, pp. 1964–1974. https://doi.org/10.2514/1.J059803.

[146] Sacher, M., Maître, O. L., Duvigneau, R., Hauville, F., Durand, M., and Lothodé, C., "A Non-Nested Infilling Strategy for Multi-Fidelity based Efficient Global Optimization," *International Journal for Uncertainty Quantification*, Vol. 11, No. 1, 2021, pp. 1–30. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020032982.

[147] Wu, W., Bonneville, C., and Earls, C., "A principled approach to design using high fidelity fluid-structure interaction simulations," *Finite Elements in Analysis and Design*, Vol. 194, 2021, p. 103562. https://doi.org/10.1016/j.finel.2021.103562.

[148] Renganathan, S. A., Maulik, R., and Ahuja, J., "Enhanced data efficiency using deep neural networks and Gaussian processes for aerodynamic design optimization," *Aerospace Science and Technology*, Vol. 111, 2021, p. 106522. https://doi.org/10.1016/j.ast.2021.106522.

[149] Kishi, Y., Kanazaki, M., and Makino, Y., "Supersonic Forward-Swept Wing Design Using Multifidelity Efficient Global Optimization," *Journal of Aircraft*, Vol. 59, No. 4, 2022, pp. 1027–1040. https://doi.org/10.2514/1.C036422.

[150] Cheng, J., Lin, Q., and Yi, J., "An enhanced variable-fidelity optimization approach for constrained optimization problems and its parallelization," *Structural and Multidisciplinary Optimization*, Vol. 65, No. 7, 2022, p. 188. https://doi.org/10.1007/s00158-022-03283-0.

[151] Foumani, Z. Z., Shishehbor, M., Yousefpour, A., and Bostanabad, R., "Multi-fidelity cost-aware Bayesian optimization," *Computer Methods in Applied Mechanics and Engineering*, Vol. 407, 2023, p. 115937. https://doi.org/10.1016/j.cma.2023.115937.

[152] Huang, H., Liu, Z., Zheng, H., Xu, X., and Duan, Y., "A proportional expected improvement criterion-based multi-fidelity sequential optimization method," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 2, 2023, p. 30. https://doi.org/10.1007/s00158-022-03484-7.

[153] Grassi, F., Manganini, G., Garraffa, M., and Mainini, L., "RAAL: Resource aware active learning for multifidelity efficient optimization," *AIAA Journal*, Vol. 61, No. 6, 2023, pp. 2744–2753. https://doi.org/10.2514/1.J061383.

[154] Fiore, D. F., and Mainini, L., "NM-MF: Non-Myopic Multifidelity Framework for Constrained Multi-Regime Aerodynamic Optimization," *AIAA Journal*, Vol. 61, No. 3, 2023, pp. 1270–1280. https://doi.org/10.2514/1.J062219.

[155] Shintani, K., Sugai, T., and Yamada, T., "A set-based approach for hierarchical optimization problem using Bayesian active learning," *International Journal for Numerical Methods in Engineering*, Vol. 124, No. 10, 2023, pp. 2196–2214. https://doi.org/10.1002/nme.7206.

[156] Lin, Q., Zheng, A., Hu, J., Shu, L., and Zhou, Q., "A multi-objective Bayesian optimization approach based on variable-fidelity multi-output metamodeling," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 5, 2023, p. 100. https://doi.org/10.1007/s00158-023-03536-6.

[157] Winter, J. M., Abaidi, R., Kaiser, J. W. J., Adami, S., and Adams, N. A., "Multi-fidelity Bayesian optimization to solve the inverse Stefan problem," *Computer Methods in Applied Mechanics and Engineering*, Vol. 410, 2023, p. 115946. https://doi.org/10.1016/j.cma.2023.115946.

[158] Conn, A. R., Scheinberg, K., and Vicente, L. N., "Global Convergence of General Derivative-Free Trust-Region Algorithms to First- and Second-Order Critical Points," *SIAM Journal on Optimization*, Vol. 20, No. 1, 2009, pp. 387–415. https://doi.org/10.1137/060673424.

[159] Wild, S. M., and Shoemaker, C., "Global Convergence of Radial Basis Function Trust Region Derivative-Free Algorithms," *SIAM Journal on Optimization*, Vol. 21, No. 3, 2011, pp. 761–781. https://doi.org/10.1137/09074927X.

[160] March, A., and Willcox, K., "Constrained multifidelity optimization using model calibration," *Structural and Multidisciplinary Optimization*, Vol. 46, No. 1, 2012, pp. 93–109. https://doi.org/10.1007/s00158-011-0749-1.

[161] March, A., and Willcox, K., "Provably Convergent Multifidelity Optimization Algorithm Not Requiring High-Fidelity Derivatives," *AIAA Journal*, Vol. 50, No. 5, 2012, pp. 1079–1089. https://doi.org/10.2514/1.J051125.

[162] Zhang, F., Song, J., Bowden, J. C., Ladd, A., Yue, Y., Desautels, T., and Chen, Y., "Learning Regions of Interest for Bayesian Optimization with Adaptive Level-Set Estimation," *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 202, PMLR, 2023, pp. 41579–41595.

[163] Bossek, J., Doerr, C., and Kerschke, P., "Initial Design Strategies and Their Effects on Sequential Model-Based Optimization: An Exploratory Case Study Based on BBOB," *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, Association for Computing Machinery, New York, NY, USA, 2020, p. 778–786. https://doi.org/10.1145/3377930.3390155.

[164] Močkus, J., "On Bayesian methods for seeking the extremum," *Optimization Techniques IFIP Technical Conference: Novosibirsk, July 1-7, 1974*, Springer, 1975, pp. 400–404. https://doi.org/10.1007/3-540-07165-2_55.

[165] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492. https://doi.org/10.1023/A:1008306431147.

[166] Chiles, J.-P., and Delfiner, P., *Geostatistics: modeling spatial uncertainty*, John Wiley & Sons, New York, USA, 1999. https://doi.org/10.1002/9781118136188.

[167] Bull, A. D., "Convergence rates of efficient global optimization algorithms." *Journal of Machine Learning Research*, Vol. 12, No. 10, 2011, pp. 2879–2904.

[168] Christianson, R. B., Pollyea, R. M., and Gramacy, R. B., "Traditional kriging versus modern Gaussian processes for large-scale mining data," *Statistical Analysis and Data Mining*, Vol. 16, No. 5, 2023, pp. 488–506. https://doi.org/10.1002/sam.11635.

[169] Erickson, C. B., Ankenman, B. E., and Sanchez, S. M., "Comparison of Gaussian process modeling software," *European Journal of Operational Research*, Vol. 266, No. 1, 2018, pp. 179–192. https://doi.org/10.1016/j.ejor.2017.10.002.

[170] Garnett, R., *Bayesian Optimization*, Cambridge University Press, Cambridge, 2023. https://doi.org/10.1017/9781108348973.

[171] Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B., "Algorithms for hyper-parameter optimization," *Advances in Neural Information Processing Systems*, Vol. 24, NIPS 2011, 2011, pp. 2546–2554.

[172] Priem, R., Bartoli, N., Diouane, Y., and Sgueglia, A., "Upper trust bound feasibility criterion for mixed constrained Bayesian optimization with application to aircraft design," *Aerospace Science and Technology*, Vol. 105, 2020, p. 105980. https://doi.org/10.1016/j.ast.2020.105980.

[173] Tran, A., Tran, M., and Wang, Y., "Constrained mixed-integer Gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials," *Structural and Multidisciplinary Optimization*, Vol. 59, No. 6, 2019, pp. 2131–2154. https://doi.org/10.1007/s00158-018-2182-1.

[174] Greenhill, S., Rana, S., Gupta, S., Vellanki, P., and Venkatesh, S., "Bayesian Optimization for Adaptive Experimental Design: A Review," *IEEE Access*, Vol. 8, 2020, pp. 13937–13948. https://doi.org/10.1109/ACCESS.2020.2966228.

[175] Ueno, T., Rhone, T. D., Hou, Z., Mizoguchi, T., and Tsuda, K., "COMBO: An efficient Bayesian optimization library for materials science," *Materials Discovery*, Vol. 4, 2016, pp. 18–21. https://doi.org/10.1016/j.md.2016.04.001.

[176] Mathern, A., Steinholtz, O. S., Sjöberg, A., Önnheim, M., Ek, K., Rempling, R., Gustavsson, E., and Jirstrand, M., "Multi-objective constrained Bayesian optimization for structural design," *Structural and Multidisciplinary Optimization*, Vol. 63, No. 2, 2021, pp. 689–701. https://doi.org/10.1007/s00158-020-02720-2.

[177] Shi, R., Xu, X., Li, J., and Li, Y., "Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization," *Applied Soft Computing*, Vol. 109, 2021, p. 107538. https://doi.org/10.1016/j.asoc.2021.107538.

[178] Park, S., Na, J., Kim, M., and Lee, J. M., "Multi-objective Bayesian optimization of chemical reactor design using computational fluid dynamics," *Computers & Chemical Engineering*, Vol. 119, 2018, pp. 25–37. https://doi.org/10.1016/j.compchemeng.2018.08.005.

[179] Torun, H. M., Swaminathan, M., Davis, A. K., and Bellaredj, M. L. F., "A Global Bayesian Optimization Algorithm and Its Application to Integrated System Design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 26, No. 4, 2018, pp. 792–802. https://doi.org/10.1109/TVLSI.2017.2784783.

[180] Manheim, D. C., and Detwiler, R. L., "Accurate and reliable estimation of kinetic parameters for environmental engineering applications: A global, multi objective, Bayesian optimization approach," *Methods X*, Vol. 6, 2019, pp. 1398–1414. https://doi.org/10.1016/j.mex.2019.05.035.

[181] Roussel, R., Hanuka, A., and Edelen, A., "Multiobjective Bayesian optimization for online accelerator tuning," *Physical Review Accelerators and Beams*, Vol. 24, 2021, p. 062801. https://doi.org/10.1103/PhysRevAccelBeams.24.062801.

[182] Jones, D. R., "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *Journal of Global Optimization*, Vol. 21, No. 4, 2001, pp. 345–383. https://doi.org/10.1023/A:1012771025575.

[183] Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M., "Gaussian process optimization in the bandit setting: No regret and experimental design," *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 1015–1022.

[184] Hennig, P., and Schuler, C. J., "Entropy Search for Information-Efficient Global Optimization." *Journal of Machine Learning Research*, Vol. 13, No. 6, 2012, pp. 1809–1837.

[185] Blanchard, A., and Sapsis, T., "Bayesian optimization with output-weighted optimal sampling," *Journal of Computational Physics*, Vol. 425, 2021, p. 109901. https://doi.org/10.1016/j.jcp.2020.109901.

[186] Brochu, E., Cora, V. M., and De Freitas, N., "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010. https://doi.org/10.48550/arXiv.1012.2599.

[187] Fricker, T. E., Oakley, J. E., and Urban, N. M., "Multivariate Gaussian Process Emulators With Nonseparable Covariance Structures," *Technometrics*, Vol. 55, No. 1, 2013, pp. 47–56. https://doi.org/10.1080/00401706.2012.715835.

[188] Palar, P. S., Parussini, L., Bregant, L., Shimoyama, K., and Zuhal, L. R., "On kernel functions for bi-fidelity Gaussian process regressions," *Structural and Multidisciplinary Optimization*, Vol. 66, No. 2, 2023, p. 37. https://doi.org/10.1007/s00158-023-03487-y.

[189] Garland, N., Le Riche, R., Richet, Y., Durrande, N., Brevault, L., Balesdent, M., and Morio, J., "Multi-fidelity for MDO using Gaussian processes," *Aerospace System Analysis and Optimization in Uncertainty*, Springer International Publishing, Cham, 2020, pp. 295–320. https://doi.org/10.1007/978-3-030-39126-3_8.

[190] Verweij, B., Ahmed, S., Kleywegt, A. J., Nemhauser, G., and Shapiro, A., "The Sample Average Approximation Method Applied to Stochastic Routing Problems: A Computational Study," *Computational Optimization and Applications*, Vol. 24, No. 2, 2003, pp. 289–333. https://doi.org/10.1023/A:1021814225969.

[191] Damianou, A., and Lawrence, N. D., "Deep Gaussian processes," *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, Vol. 31, PMLR, 2013, pp. 207–215.

[192] Salimbeni, H., and Deisenroth, M., "Doubly stochastic variational inference for deep Gaussian processes," *Advances in Neural Information Processing Systems*, Vol. 30, NIPS 2017, 2017.

[193] Titsias, M., "Variational Learning of Inducing Variables in Sparse Gaussian Processes," *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, Vol. 5, edited by D. van Dyk and M. Welling, PMLR, 2009, pp. 567–574.

[194] Zhou, Q., Qian, P. Z. G., and Zhou, S., "A Simple Approach to Emulation for Computer Models With Qualitative and Quantitative Factors," *Technometrics*, Vol. 53, No. 3, 2011, pp. 266–273. https://doi.org/10.1198/TECH.2011.10025.

[195] Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., "Mixed Variable Gaussian Process-Based Surrogate Modeling Techniques: Application to Aerospace Design," *Journal of Aerospace Information Systems*, Vol. 18, No. 11, 2021, pp. 813–837. https://doi.org/10.2514/1.I010965.

[196] Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H., "Group Kernels for Gaussian Process Metamodels with Categorical Inputs," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 8, No. 2, 2020, pp. 775–806. https://doi.org/10.1137/18M1209386.

[197] Oune, N., and Bostanabad, R., "Latent map Gaussian processes for mixed variable metamodeling," *Computer Methods in Applied Mechanics and Engineering*, Vol. 387, 2021, p. 114128. https://doi.org/10.1016/j.cma.2021.114128.

[198] Poloczek, M., Wang, J., and Frazier, P. I., "Multi-Information Source Optimization," *Advances in Neural Information Processing Systems*, Vol. 30, NIPS 2017, 2017.

[199] Hennig, P., Osborne, M. A., and Kersting, H. P., *Probabilistic numerics: Computation as machine learning*, Cambridge University Press, 2022.

[200] Kushner, H. J., "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Journal Basic Engineering*, Vol. 86, No. 1, 1964, pp. 97–106. https://doi.org/10.1115/1.3653121.

[201] Sóbester, A., Leary, S. J., and Keane, A. J., "On the Design of Optimization Strategies Based on Global Response Surface Approximation Models," *Journal of Global Optimization*, Vol. 33, No. 1, 2005, pp. 31–59. https://doi.org/10.1007/s10898-004-6733-1.

[202] De Ath, G., Everson, R. M., Rahat, A. A. M., and Fieldsend, J. E., "Greed Is Good: Exploration and Exploitation Trade-Offs in Bayesian Optimisation," *ACM Transactions on Evolutionary Learning and Optimization*, Vol. 1, No. 1, 2021. https://doi.org/10.1145/3425501.

[203] Frazier, P. I., Powell, W. B., and Dayanik, S., "A Knowledge-Gradient Policy for Sequential Information Collection," *SIAM Journal on Control and Optimization*, Vol. 47, No. 5, 2008, pp. 2410–2439. https://doi.org/10.1137/070693424.

[204] Lai, T. L., and Robbins, H., "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, Vol. 6, No. 1, 1985, pp. 4–22. https://doi.org/10.1016/0196-8858(85)90002-8.

[205] Auer, P., "Using confidence bounds for exploitation-exploration trade-offs," *Journal of Machine Learning Research*, Vol. 3, No. Nov, 2002, pp. 397–422.

[206] Chapelle, O., and Li, L., "An empirical evaluation of Thompson sampling," *Advances in Neural Information Processing Systems*, Vol. 24, NIPS 2011, 2011, pp. 2249–2257.

[207] Agrawal, S., and Goyal, N., "Analysis of Thompson sampling for the multi-armed bandit problem," *25th Annual Conference on Learning Theory*, Vol. 23, JMLR Workshop and Conference Proceedings, 2012, pp. 39.1–39.26.

[208] Bijl, H., Schön, T. B., van Wingerden, J.-W., and Verhaegen, M., "A sequential Monte Carlo approach to Thompson sampling for Bayesian optimization," *arXiv preprint arXiv:1604.00169*, 2016. https://doi.org/10.48550/arXiv.1604.00169.

[209] Scott, S. L., "A modern Bayesian look at the multi-armed bandit," *Applied Stochastic Models in Business and Industry*, Vol. 26, No. 6, 2010, pp. 639–658. https://doi.org/10.1002/asmb.874.

[210] Hernández-Lobato, J. M., Hoffman, M. W., and Ghahramani, Z., "Predictive entropy search for efficient global optimization of black-box functions," *Advances in Neural Information Processing Systems*, Vol. 27, NIPS 2014, 2014, pp. 918–926.

[211] Wilson, J. T., Borovitskiy, V., Terenin, A., Mostowsky, P., and Deisenroth, M. P., "Efficiently Sampling Functions from Gaussian Process Posteriors," *Proceedings of the 37th International Conference on Machine Learning*, PMLR, 2020, pp. 10292–10302.

[212] Russo, D. J., and Van Roy, B., "Learning to Optimize via Posterior Sampling," *Mathematics of Operations Research*, Vol. 39, No. 4, 2014, pp. 1221–1243. https://doi.org/10.1287/moor.2014.0650.

[213] Kandasamy, K., Krishnamurthy, A., Schneider, J., and Poczos, B., "Parallelised Bayesian Optimisation via Thompson Sampling," *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 84, PMLR, 2018, pp. 133–142.

[214] Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z., "A Tutorial on Thompson Sampling," *Foundations and Trends in Machine Learning*, Vol. 11, No. 1, 2018, p. 1–96. https://doi.org/10.1561/2200000070.

[215] Villemonteix, J., Vazquez, E., and Walter, E., "An informational approach to the global optimization of expensive-to-evaluate functions," *Journal of Global Optimization*, Vol. 44, No. 4, 2009, pp. 509–534. https://doi.org/10.1007/s10898-008-9354-2.

[216] Houlsby, N., Huszar, F., Ghahramani, Z., and Hernández-lobato, J., "Collaborative Gaussian Processes for Preference Learning," *Advances in Neural Information Processing Systems*, Vol. 25, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Curran Associates, Inc., 2012.

[217] Wang, Z., and Jegelka, S., "Max-value entropy search for efficient Bayesian optimization," *Proceedings of the 34th International Conference on Machine Learning*, PMLR, 2017, pp. 3627–3635.

[218] Streltsov, S., and Vakili, P., "A Non-myopic Utility Function for Statistical Global Optimization Algorithms," *Journal of Global Optimization*, Vol. 14, No. 3, 1999, pp. 283–298. https://doi.org/10.1023/A:1008284229931.

[219] Ginsbourger, D., and Le Riche, R., "Towards Gaussian Process-based Optimization with Finite Time Horizon," *mODa 9 - Advances in Model-Oriented Design and Analysis*, edited by A. Giovagnoli, A. C. Atkinson, B. Torsney, and C. May, Physica-Verlag HD, Heidelberg, 2010, pp. 89–96.

[220] González, J., Osborne, M., and Lawrence, N., "GLASSES: Relieving the myopia of Bayesian optimisation," *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, PMLR, 2016, pp. 790–799.

[221] Wu, J., and Frazier, P. I., "Practical Two-Step Lookahead Bayesian Optimization," *Advances in Neural Information Processing Systems*, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, NeurIPS 2019, 2019.

[222] Lam, R., Willcox, K., and Wolpert, D. H., "Bayesian Optimization with a Finite Budget: An Approximate Dynamic Programming Approach," *Advances in Neural Information Processing Systems*, Vol. 29, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, NIPS 2016, 2016.

[223] Lee, E., Eriksson, D., Bindel, D., Cheng, B., and Mccourt, M., "Efficient rollout strategies for Bayesian optimization," *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, PMLR, 2020, pp. 260–269.

[224] Jiang, S., Jiang, D. R., Balandat, M., Karrer, B., Gardner, J. R., and Garnett, R., "Efficient Nonmyopic Bayesian Optimization via One-Shot Multi-Step Trees," *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY, USA, 2020.

[225] Bellman, R., "On the Theory of Dynamic Programming," *Proceedings of the National Academy of Sciences*, Vol. 38, No. 8, 1952, pp. 716–719. https://doi.org/10.1073/pnas.38.8.716.

[226] Powell, W. B., *Approximate Dynamic Programming: Solving the curses of dimensionality*, John Wiley & Sons, 2011.

[227] Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models," *Journal of Global Optimization*, Vol. 34, No. 3, 2006, pp. 441–466. https://doi.org/10.1007/s10898-005-2454-3.

[228] Hoffman, M., Shahriari, B., and Freitas, N., "On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning," *Proceedings of the 17th International Conference on Artificial Intelligence and Statistic*, PMLR, 2014, pp. 365–374.

[229] Shahriari, B., Wang, Z., Hoffman, M. W., Bouchard-Côté, A., and de Freitas, N., "An entropy search portfolio for Bayesian optimization," *NIPS Workshop on Bayesian Optimization*, 2014. BayesOpt 2014.

[230] Neumaier, A., "Complete search in continuous global optimization and constraint satisfaction," *Acta Numerica*, Vol. 13, 2004, pp. 271–369. https://doi.org/10.1017/S0962492904000194.

[231] Do, B., and Ohsaki, M., "Proximal-exploration multi-objective Bayesian optimization for inverse identification of cyclic constitutive law of structural steels," *Structural and Multidisciplinary Optimization*, Vol. 65, No. 7, 2022, p. 199. https://doi.org/10.1007/s00158-022-03297-8.

[232] Kontogiannis, S. G., and Savill, M. A., "A generalized methodology for multidisciplinary design optimization using surrogate modelling and multifidelity analysis," *Optimization and Engineering*, Vol. 21, No. 3, 2020, pp. 723–759. https://doi.org/10.1007/s11081-020-09504-z.

[233] Kandasamy, K., Dasarathy, G., Oliva, J. B., Schneider, J., and Poczos, B., "Gaussian Process Bandit Optimisation with Multi-fidelity Evaluations," *Advances in Neural Information Processing Systems*, Vol. 29, NIPS 2016, 2016, pp. 992–1000.

[234] Wu, J., and Frazier, P. I., "The parallel knowledge gradient method for batch Bayesian optimization," *Proceedings of the 30th Conference on Neural Information Processing Systems*, NIPS 2016, 2016.

[235] Paleyes, A., Pullin, M., Mahsereci, M., McCollum, C., Lawrence, N. D., and González, J., "Emulation of physical processes with Emukit," *Second workshop on machine learning and the physical sciences (NeurIPS 2019)*, Vancouver, Canada, 2019.

[236] Balandat, M., Karrer, B., Jiang, D., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E., "BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization," *Advances in Neural Information Processing Systems*, Vol. 33, NeurIPS 2020, 2020, pp. 21524–21538.

[237] Lindauer, M., Eggensperger, K., Feurer, M., Biedenkapp, A., Deng, D., Benjamins, C., Ruhkopf, T., Sass, R., and Hutter, F., "SMAC3: A Versatile Bayesian Optimization Package for Hyperparameter Optimization," *Journal of Machine Learning Research*, Vol. 23, No. 1, 2022.

[238] Schonlau, M., Welch, W. J., and Jones, D. R., "Global versus Local Search in Constrained Optimization of Computer Models," *New developments and applications in experimental design*, IMS Lecture Notes-Monograph Series, Vol. 34, edited by N. Flournoy, W. F. Rosenberger, and W. K. Wong, Institute of Mathematical Statistics, 1998, pp. 11–25. https://doi.org/10.1214/lnms/1215456182.

[239] Gardner, J. R., Kusner, M. J., Xu, Z. E., Weinberger, K. Q., and Cunningham, J. P., "Bayesian optimization with inequality constraints," *Proceedings of the 31st International Conference on Machine Learning*, Vol. 2014, JMLR, 2014, pp. 937–945.

[240] Sóbester, A., Forrester, A. I. J., Toal, D. J. J., Tresidder, E., and Tucker, S., "Engineering design applications of surrogate-assisted optimization techniques," *Optimization and Engineering*, Vol. 15, No. 1, 2014, pp. 243–265. https://doi.org/10.1007/s11081-012-9199-x.

[241] Gramacy, R. B., and Lee, H. K. H., "Optimization Under Unknown Constraints," *Bayesian Statistics 9*, edited by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, Oxford University Press, 2011, pp. 229–256. https://doi.org/10.1093/acprof:oso/9780199694587.003.0008.

[242] Picheny, V., "A stepwise uncertainty reduction approach to constrained global optimization," *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, PMLR, 2014, pp. 787–795.

[243] Gramacy, R. B., Gray, G. A., Le Digabel, S., Lee, H. K. H., Ranjan, P., Wells, G., and Wild, S. M., "Modeling an Augmented Lagrangian for Blackbox Constrained Optimization," *Technometrics*, Vol. 58, No. 1, 2016, pp. 1–11. https://doi.org/10.1080/00401706.2015.1014065.

[244] Eriksson, D., and Poloczek, M., "Scalable constrained Bayesian optimization," *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, Vol. 130, PMLR, 2021, pp. 730–738.

[245] Luenberger, D. G., and Ye, Y., *Linear and nonlinear programming*, Springer, California, USA, 2008. https://doi.org/10.1007/978-0-387-74503-9.

[246] Rana, S., Li, C., Gupta, S., Nguyen, V., and Venkatesh, S., "High Dimensional Bayesian Optimization with Elastic Gaussian Process," *Proceedings of the 34th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 70, edited by D. Precup and Y. W. Teh, PMLR, 2017, pp. 2883–2891.

[247] Daulton, S., Eriksson, D., Balandat, M., and Bakshy, E., "Multi-objective Bayesian optimization over high-dimensional search spaces," *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, Vol. 180, PMLR, 2022, pp. 507–517.

[248] Spagnol, A., Riche, R. L., and Veiga, S. D., "Global Sensitivity Analysis for Optimization with Variable Selection," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 7, No. 2, 2019, pp. 417–443. https://doi.org/10.1137/18M1167978.

[249] Constantine, P. G., Dow, E., and Wang, Q., "Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces," *SIAM Journal on Scientific Computing*, Vol. 36, No. 4, 2014, pp. A1500–A1524. https://doi.org/10.1137/130916138.

[250] Wang, Z., Hutter, F., Zoghi, M., Matheson, D., and De Freitas, N., "Bayesian Optimization in a Billion Dimensions via Random Embeddings," *Journal of Artificial Intelligence Research*, Vol. 55, No. 1, 2016, pp. 361–387.

[251] Nayebi, A., Munteanu, A., and Poloczek, M., "A Framework for Bayesian Optimization in Embedded Subspaces," *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 97, PMLR, 2019, pp. 4752–4761.

[252] Letham, B., Calandra, R., Rai, A., and Bakshy, E., "Re-examining linear embeddings for high-dimensional Bayesian optimization," *Advances in Neural Information Processing Systems*, Vol. 33, NeurIPS 2020, 2020, pp. 1546–1558.

[253] Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A., "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules," *ACS Central Science*, Vol. 4, No. 2, 2018, pp. 268–276. https://doi.org/10.1021/acscentsci.7b00572.

[254] Moriconi, R., Deisenroth, M. P., and Sesh Kumar, K. S., "High-dimensional Bayesian optimization using low-dimensional feature spaces," *Machine Learning*, Vol. 109, No. 9, 2020, pp. 1925–1943. https://doi.org/10.1007/s10994-020-05899-z.

[255] Kandasamy, K., Schneider, J., and Póczos, B., "High Dimensional Bayesian Optimisation and Bandits via Additive Models," *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Vol. 37, JMLR, Lille, France, 2015, pp. 295–304.

[256] Gardner, J., Guo, C., Weinberger, K., Garnett, R., and Grosse, R., "Discovering and exploiting additive structure for Bayesian optimization," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 54, edited by A. Singh and J. Zhu, PMLR, 2017, pp. 1311–1319.

[257] Wang, Z., Gehring, C., Kohli, P., and Jegelka, S., "Batched large-scale Bayesian optimization in high-dimensional spaces," *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, Vol. 84, PMLR, 2018, pp. 745–754.

[258] Eriksson, D., Pearce, M., Gardner, J., Turner, R. D., and Poloczek, M., "Scalable global optimization via local Bayesian optimization," *Advances in Neural Information Processing Systems*, Vol. 32, NIPS 2019, 2019, pp. 5496–5507.

[259] Oberkampf, W. L., and Roy, C. J., *Verification and validation in scientific computing*, Cambridge University Press, Cambridge, UK, 2010.

[260] Roy, C. J., and Oberkampf, W. L., "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," *Computer Methods in Applied Mechanics and Engineering*, Vol. 200, No. 25, 2011, pp. 2131–2144. https://doi.org/10.1016/j.cma.2011.03.016.

[261] Beyer, H.-G., and Sendhoff, B., "Robust optimization - A comprehensive survey," *Computer Methods in Applied Mechanics and Engineering*, Vol. 196, No. 33, 2007, pp. 3190–3218. https://doi.org/10.1016/j.cma.2007.03.003.

[262] Kanno, Y., "On three concepts in robust design optimization: absolute robustness, relative robustness, and less variance," *Structural and Multidisciplinary Optimization*, Vol. 62, No. 2, 2020, pp. 979–1000. https://doi.org/10.1007/s00158-020-02503-9.

[263] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A., *Robust optimization*, Princeton University Press, New Jersey, USA, 2009. https://doi.org/10.1515/9781400831050.

[264] Elishakoff, I., and Ohsaki, M., *Optimization and anti-optimization of structures under uncertainty*, Imperial College Press, London, 2010.

[265] Hemez, F. M., and Ben-Haim, Y., "Info-gap robustness for the correlation of tests and simulations of a non-linear transient," *Mechanical Systems and Signal Processing*, Vol. 18, No. 6, 2004, pp. 1443–1467. https://doi.org/10.1016/j.ymssp.2004.03.001.

[266] Caflisch, R. E., "Monte Carlo and quasi-Monte Carlo methods," *Acta Numerica*, Vol. 7, 1998, pp. 1–49. https://doi.org/10.1017/S0962492900002804.

[267] Crestaux, T., Le Maı^tre, O., and Martinez, J.-M., "Polynomial chaos expansion for sensitivity analysis," *Reliability Engineering & System Safety*, Vol. 94, No. 7, 2009, pp. 1161–1172. https://doi.org/10.1016/j.ress.2008.10.008.

[268] Anderson, T. V., and Mattson, C. A., "Propagating Skewness and Kurtosis Through Engineering Models for Low-Cost, Meaningful, Nondeterministic Design," *Journal of Mechanical Design*, Vol. 134, No. 10, 2012. https://doi.org/10.1115/1.4007389.

[269] O'Hagan, A., "Bayes-Hermite quadrature," *Journal of Statistical Planning and Inference*, Vol. 29, No. 3, 1991, pp. 245–260. https://doi.org/10.1016/0378-3758(91)90002-V.

[270] Larson, J., Menickelly, M., and Wild, S. M., "Derivative-free optimization methods," *Acta Numerica*, Vol. 28, 2019, pp. 287–404. https://doi.org/10.1017/S0962492919000060.

[271] Do, B., Ohsaki, M., and Yamakawa, M., "Bayesian optimization for robust design of steel frames with joint and individual probabilistic constraints," *Engineering Structures*, Vol. 245, 2021, p. 112859. https://doi.org/10.1016/j.engstruct.2021.112859.

[272] Daulton, S., Cakmak, S., Balandat, M., Osborne, M. A., Zhou, E., and Bakshy, E., "Robust multi-objective Bayesian optimization under input noise," *Proceedings of the 39th International Conference on Machine Learning*, PMLR, 2022, pp. 4831–4866.

[273] Ng, L. W. T., and Willcox, K., "Multifidelity approaches for optimization under uncertainty," *International Journal for Numerical Methods in Engineering*, Vol. 100, No. 10, 2014, pp. 746–772. https://doi.org/10.1002/nme.4761.

[274] Shah, H., Hosder, S., Koziel, S., Tesfahunegn, Y. A., and Leifsson, L., "Multi-fidelity robust aerodynamic design optimization under mixed uncertainty," *Aerospace Science and Technology*, Vol. 45, 2015, pp. 17–29. https://doi.org/10.1016/j.ast.2015.04.011.

[275] Fusi, F., Guardone, A., Quaranta, G., and Congedo, P. M., "Multifidelity Physics-Based Method for Robust Optimization Applied to a Hovering Rotor Airfoil," *AIAA Journal*, Vol. 53, No. 11, 2015, pp. 3448–3465. https://doi.org/10.2514/1.J053952.

[276] Chakraborty, S., Chatterjee, T., Chowdhury, R., and Adhikari, S., "A surrogate based multi-fidelity approach for robust design optimization," *Applied Mathematical Modelling*, Vol. 47, 2017, pp. 726–744. https://doi.org/10.1016/j.apm.2017.03.040.

[277] Zhou, Q., Wang, Y., Choi, S.-K., Jiang, P., Shao, X., Hu, J., and Shu, L., "A robust optimization approach based on multi-fidelity metamodel," *Structural and Multidisciplinary Optimization*, Vol. 57, No. 2, 2018, pp. 775–797. https://doi.org/10.1007/s00158-017-1783-4.

[278] Moustapha, M., and Sudret, B., "Surrogate-assisted reliability-based design optimization: a survey and a unified modular framework," *Structural and Multidisciplinary Optimization*, Vol. 60, No. 5, 2019, pp. 2157–2176. https://doi.org/10.1007/s00158-019-02290-y.

[279] Campi, M. C., and Garatti, S., "A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality," *Journal of Optimization Theory and Applications*, Vol. 148, No. 2, 2011, pp. 257–280. https://doi.org/10.1007/s10957-010-9754-6.

[280] Xie, W., and Ahmed, S., "On Deterministic Reformulations of Distributionally Robust Joint Chance Constrained Optimization Problems," *SIAM Journal on Optimization*, Vol. 28, No. 2, 2018, pp. 1151–1182. https://doi.org/10.1137/16M1094725.

[281] Geng, X., and Xie, L., "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization," *Annual Reviews in Control*, Vol. 47, 2019, pp. 341–363. https://doi.org/10.1016/j.arcontrol.2019.05.005.

[282] Nemirovski, A., "On safe tractable approximations of chance constraints," *European Journal of Operational Research*, Vol. 219, No. 3, 2012, pp. 707–718. https://doi.org/10.1016/j.ejor.2011.11.006.

[283] Melchers, R. E., and Beck, A. T., *Structural reliability analysis and prediction*, 3rd ed., John Wiley & Sons, New York, 2018.

[284] Aoues, Y., and Chateauneuf, A., "Benchmark study of numerical methods for reliability-based design optimization," *Structural and Multidisciplinary Optimization*, Vol. 41, No. 2, 2010, pp. 277–294. https://doi.org/10.1007/s00158-009-0412-2.

[285] Valdebenito, M. A., and Schuëller, G. I., "A survey on approaches for reliability-based optimization," *Structural and Multidisciplinary Optimization*, Vol. 42, No. 5, 2010, pp. 645–663. https://doi.org/10.1007/s00158-010-0518-6.

[286] Calafiore, G. C., and Campi, M. C., "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, Vol. 51, No. 5, 2006, pp. 742–753. https://doi.org/10.1109/TAC.2006.875041.

[287] Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M., "Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions," *AIAA Journal*, Vol. 46, No. 10, 2008, pp. 2459–2468. https://doi.org/10.2514/1.34321.

[288] Blanchard, A., and Sapsis, T., "Output-Weighted Optimal Sampling for Bayesian Experimental Design and Uncertainty Quantification," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 9, No. 2, 2021, pp. 564–592. https://doi.org/10.1137/20M1347486.

[289] Huynh, V. T., Tangaramvong, S., Do, B., Gao, W., and Limkatanyu, S., "Sequential most probable point update combining Gaussian process and comprehensive learning PSO for structural reliability-based design optimization," *Reliability Engineering & System Safety*, Vol. 235, 2023, p. 109164. https://doi.org/10.1016/j.ress.2023.109164.

[290] Chaudhuri, A., Kramer, B., and Willcox, K. E., "Information Reuse for Importance Sampling in Reliability-Based Design Optimization," *Reliability Engineering & System Safety*, Vol. 201, 2020, p. 106853. https://doi.org/10.1016/j.ress.2020.106853.

[291] Chaudhuri, A., Marques, A. N., and Willcox, K., "mfEGRA: Multifidelity efficient global reliability analysis through active learning for failure boundary location," *Structural and Multidisciplinary Optimization*, Vol. 64, No. 2, 2021, pp. 797–811. https://doi.org/10.1007/s00158-021-02892-5.

[292] Patsialis, D., and Taflanidis, A. A., "Multi-fidelity Monte Carlo for seismic risk assessment applications," *Structural Safety*, Vol. 93, 2021, p. 102129. https://doi.org/10.1016/j.strusafe.2021.102129.

[293] Zhang, C., Song, C., and Shafieezadeh, A., "Adaptive reliability analysis for multi-fidelity models using a collective learning strategy," *Structural Safety*, Vol. 94, 2022, p. 102141. https://doi.org/10.1016/j.strusafe.2021.102141.

[294] Ashwin Renganathan, S., Rao, V., and Navon, I. M., "CAMERA: A method for cost-aware, adaptive, multifidelity, efficient reliability analysis," *Journal of Computational Physics*, Vol. 472, 2023, p. 111698. https://doi.org/10.1016/j.jcp.2022.111698.

[295] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, 2002, pp. 182–197. https://doi.org/10.1109/4235.996017.

[296] Zhang, Q., and Li, H., "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, Vol. 11, No. 6, 2007, pp. 712–731. https://doi.org/10.1109/TEVC.2007.892759.

[297] Knowles, J., "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, Vol. 10, No. 1, 2006, pp. 50–66. https://doi.org/10.1109/TEVC.2005.851274.

[298] Das, I., and Dennis, J. E., "A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems," *Structural optimization*, Vol. 14, No. 1, 1997, pp. 63–69. https://doi.org/10.1007/BF01197559.

[299] Emmerich, M. T. M., Giannakoglou, K. C., and Naujoks, B., "Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels," *IEEE Transactions on Evolutionary Computation*, Vol. 10, No. 4, 2006, pp. 421–439. https://doi.org/10.1109/TEVC.2005.859463.

[300] Keane, A. J., "Statistical Improvement Criteria for Use in Multiobjective Design Optimization," *AIAA Journal*, Vol. 44, No. 4, 2006, pp. 879–891. https://doi.org/10.2514/1.16875.

[301] Ponweiser, W., Wagner, T., Biermann, D., and Vincze, M., "Multiobjective Optimization on a Limited Budget of Evaluations Using Model-Assisted $\mathcal{S}$-Metric Selection," *Parallel Problem Solving from Nature - PPSN X*, edited by G. Rudolph, T. Jansen, N. Beume, S. Lucas, and C. Poloni, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 784–794.

[302] Suzuki, S., Takeno, S., Tamura, T., Shitara, K., and Karasuyama, M., "Multi-objective Bayesian Optimization using Pareto-frontier Entropy," *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 119, PMLR, 2020, pp. 9279–9288.

[303] Daulton, S., Balandat, M., and Bakshy, E., "Hypervolume knowledge gradient: A lookahead approach for multi-objective Bayesian optimization with partial information," *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 202, PMLR, 2023, pp. 7167–7204.

[304] Garrido-Merchán, E. C., and Hernández-Lobato, D., "Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes," *Neurocomputing*, Vol. 380, 2020, pp. 20–35. https://doi.org/10.1016/j.neucom.2019.11.004.

[305] Zhang, Y., Apley, D. W., and Chen, W., "Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables," *Scientific Reports*, Vol. 10, No. 1, 2020, p. 4924. https://doi.org/10.1038/s41598-020-60652-9.

[306] Vangelatos, Z., Sheikh, H. M., Marcus, P. S., Grigoropoulos, C. P., Lopez, V. Z., Flamourakis, G., and Farsari, M., "Strength through defects: A novel Bayesian approach for the optimization of architected materials," *Science Advances*, Vol. 7, No. 41, 2021, p. eabk2218. https://doi.org/10.1126/sciadv.abk2218.

[307] Wu, J., Poloczek, M., Wilson, A. G., and Frazier, P., "Bayesian Optimization with Gradients," *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Curran Associates, Inc., 2017, pp. 5267–5278.