Submitted to Engineering Structures

Bayesian optimization for robust design of steel frames with joint and individual probabilistic constraints

Bach Do[1*], Makoto Ohsaki[1] and Makoto Yamakawa[2]

[1]Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto-Daigaku Katsura, Nishikyo, Kyoto 615-8540, Japan.

[2] Department of Architecture, Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo, 125-8585, Japan.

(*Corresponding author email: se.do@archi.kyoto-u.ac.jp)

Abstract

This work proposes a Bayesian optimization (BO) method for solving multi-objective robust design optimization (RDO) problems of steel frames under aleatory uncertainty in external loads and material properties. Joint and individual probabilistic constrained RDO problems are formulated to consider two different ways the frame reaches its collapse state. Each problem involves three conflicting objective functions, namely, the total mass of the frame, the mean and variance of the maximum inter-story drift. Since the uncertain objective and probabilistic constraint functions of both problems are implicit within a finite element analysis program and the computation of the probabilistic constraints is an NP-hard problem, BO is used to guide the optimization process toward better solutions after it completes an iteration and offers a set of near Pareto-optimal solutions when it terminates. Specifically, Bayesian regression models called Gaussian processes (GPs) serve as surrogates for the structural responses. Two acquisition functions are then developed for the two RDO problems and a maximization problem of these functions is formulated as a mixed-integer nonlinear programming (MINLP) problem. A new random search coupled with simulated annealing is devised to solve the MINLP problem, thereby locating the most promising point in the input variable space at which the current solutions maximize their chance to be improved and the GP models are refined before the BO starts a new iteration. A test problem and two design examples show that exact or good Pareto-optimal solutions to the RDO problems can be found by the proposed method with 20 iterations.

Keywords: Bayesian optimization; Robust design optimization; Probabilistic constraints; Steel frames

1. Introduction

In designing an engineering structure, design parameters such as structural dimensions, material properties, and external loads are uncertain in nature [1,2]. Designers often accommodate this issue by seeking a robust design for the structure according to specified design objectives and specifications. With the aim of minimizing the worst value of the objective function, the robust design can be found using a minimax approach that takes into account the worst effects of the uncertain parameters [3–6]. If the minimum gap between the worst and nominal values of the objective function is of interest, the designers can use a gap minimization approach for their design [7]. Meanwhile, when the goal is to minimize the sensitivity of the uncertain objective function, a variance minimization approach [8,9] can be adopted by which the robustness of the structure can be measured by variation of the uncertain objective function. This study focuses on the latter approach.

In the variance minimization approach, the uncertain parameters are described by their probability density functions (PDFs), or equivalently, their cumulative distribution functions (CDFs). Then, the robust design optimization (RDO) problem is commonly formulated as a multi-objective optimization problem to manage the trade-off between the expected and variance values of the uncertain objective function [10–15]. Since the uncertain parameters also affect the constraints of the RDO problem, it is natural to formulate the problem using a set of probabilistic constraints (i.e., chance constraints [4]), which are related to the probabilities that certain design requirements are satisfied under the effects of uncertainty. Particularly, in this study, two multi-objective RDO problems of a steel frame are investigated under aleatory uncertainty in material properties and external loads. These problems correspond to two different ways the frame approaches its collapse state. When the limit state functions (LSFs) correlate with each other, the first RDO problem is formulated with a joint probabilistic constraint, where the probability is taken over the entire system of random LSFs. On the other hand, the second RDO problem is formulated with a finite number of individual probabilistic constraints on the uncertain LSFs, which are statistically independent. The joint and individual constrained RDO

problems also set the basis for further application to a general case where all possible failure modes for a structure can be identified.

Let $\mathbf{s} = [s_1, \dots, s_{d_1}]^T \subset \mathbb{N}^{d_1}$ denote a d_1-dimensional vector of discrete design variables of the steel frame. Each element of $\mathbf{s}$ is selected from a given list of standard steel sections, i.e., $s_i \in \mathcal{S}_i$ $(i = 1, \dots, d_1)$. Let $\mathbf{r}$ denote a d_2-dimensional vector of continuous uncertain parameters, i.e., material properties and external loads. The marginal PDFs of $\mathbf{r}$ elements are given and supported on a set $\mathcal{R} \subset \mathbb{R}^{d_2}$. Let $f_1(\mathbf{s}): \mathbb{N}^{d_1} \to \mathbb{R}$ and $f(\mathbf{s}, \mathbf{r}): \mathbb{N}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ represent the total mass and the LSF corresponding to the maximum inter-story drift of the frame, respectively. The mean and variance of $f(\mathbf{s}, \mathbf{r})$ are denoted as $f_2(\mathbf{s}) = \mathbb{E}[f(\mathbf{s}, \mathbf{r})]$ and $f_3(\mathbf{s}) = \mathrm{var}[f(\mathbf{s}, \mathbf{r})]$, respectively. Also, let $g_i(\mathbf{s}, \mathbf{r})$: $\mathbb{N}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ represent the LSFs of certain requirements on the serviceability and strength of the frame, and $h_j(\mathbf{s}): \mathbb{N}^{d_1} \to \mathbb{R}$ indicate deterministic constraints on the geometry of the structural elements. By extending the authors' work [15] to include the joint and individual probabilistic constraints in the optimization problem, this study formulates the following two multi-objective RDO problems:

$$\min_{\mathbf{s}} [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})]$$

subject to

$$\mathbb{P}[g_i(\mathbf{s}, \mathbf{r}) \le 0,\ i = 1, \dots, I\ ] \ge 1 - \varepsilon$$

$$h_j(\mathbf{s}) \le 0\ ,\ j = 1, \dots, J$$

$$s_i \in \mathcal{S}_i,\ i = 1, \dots, d_1 \qquad (1)$$

where $\mathbb{P}[\cdot]$ denotes the probability of occurring $[\cdot]$ with respect to all instances of $\mathbf{r}$ and $\varepsilon \in (0,1)$ is a prescribed risk level of the joint probabilistic constraint.

$$\min_{\mathbf{s}} [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})]$$

subject to

$$\mathbb{P}[g_i(\mathbf{s}, \mathbf{r}) \le 0\ ] \ge 1 - \varepsilon_i,\ i = 1, \dots, I$$

$$h_j(\mathbf{s}) \le 0\ ,\ j = 1, \dots, J \qquad (2)$$

$s_i \in \mathcal{S}_i,\ i = 1, \ldots, d_1$

where $\varepsilon_i \in (0,1)$ is a prescribed risk level of the ith probabilistic constraint.

Although problems (1) and (2) are important for finding a robust design of a frame, they are very difficult to solve due to the following four issues. First, a search that directly calculates the objective and constraint functions during the optimization process may be impossible because uncertain objective and probabilistic constraint functions of these problems are non-linear and implicit within a finite element analysis program. Second, checking the feasibility of a specified candidate solution to both problems is challenging as the computation of the probabilistic constraints is an NP-hard problem [16]. For this purpose, sampling methods, e.g., the Monte-Carlo simulation (MCS), can be used, but they are computationally demanding due to the curse of dimensionality. Third, evaluation of the mean and variance values of the uncertain objective function is a challenging task. Although the Monte-Carlo integration [17] and polynomial chaos expansion [18] are two reliable and simple methods to handle this task, they are also subject to the curse of dimensionality. The Taylor series approximation [19] and Bayes-Hermite quadrature [20] may be used with respective restrictions that the uncertain objective function is differentiable and the uncertain parameters should be normally distributed. Finally, it may be impossible to obtain exact optimal solutions to the two problems because their feasible regions, defined by the probability function $\mathbb{P}[\cdot]$, are non-convex in general. Existing approaches in the literature, e.g., deterministic model approach [21], scenario approach [22], sample average approximation approach [23], and convex approximation approach [24], have attempted to find approximate solutions to several special cases of problems (1) and (2). For example, the deterministic model approach solves the problems with linear objective and constraint functions, and normally distributed random parameters. The scenario approach requires that the objective and probabilistic constraint functions should be convex with respect to the design variables $\mathbf{s}$. The sample average approximation approach works well when the objective function is linear and explicitly given, while the convex approximation approach relies on the assumption that the constraint functions $g_i(\mathbf{s}, \mathbf{r})$ are convex with respect to $\mathbf{s}$ for every instance of $\mathbf{r}$. In addition to the scope of their applications, the performance of the existing approaches in solving multi-objective problems remains unknown as they

have been only applied to single-objective problems. Obviously, it is inconvenient to use the aforementioned approaches to solve the RDO problems of the steel frames in this study.

Bayesian optimization (BO) [25,26], also well-known as the Efficient Global Optimization (EGO) [27], is a powerful sequential method to solve optimization problems with unknown objective and/or constraint functions as problems (1) and (2). It consists of two primary ingredients, i.e., a Bayesian regression model called Gaussian process (GP) and an acquisition function. The GP approximates the uncertain objective and probabilistic constraint functions, while the acquisition function guides the algorithm toward better solutions by deciding where the GP approximations should be refined. Different acquisition functions for solving single-objective problems are available in a recent tutorial [28]. The goal is then to find a new point in the joint space of $\mathbf{s}$ and $\mathbf{r}$ (or input variable space) that maximizes the acquisition function, and to use this point for improving accuracy of the approximate objective and/or constraint functions. The BO, therefore, iterates through constructing GP models, maximizing the acquisition function to locate the new sampling point, and refining the GP models at that point. It outputs the solution when the number of iterations reaches an upper bound. Recently, the BO has demonstrated its ability in solving combinatorial optimization problems [29], designing materials with mixed quantitative and qualitative variables [30], and solving nominal multi-objective optimization problems of structures [31,32].

This study proposes using the BO for solving problems (1) and (2) as it does not require special forms of the uncertain objective and probabilistic constraint functions, contrary to the existing approaches. In this way, the non-linear and implicit nature of problems (1) and (2) can be addressed by using the mean functions of the GP models as surrogates for the uncertain objective and probabilistic constraint functions. These surrogates, in turn, facilitate the evaluation of the probabilistic constraints for a particular candidate solution $\mathbf{s}$ using a saddlepoint approximation developed by the authors [33]. Thus, the feasibility of $\mathbf{s}$ can be reasonably confirmed in each iteration of the BO. Furthermore, the mean and variance of the objective function $f(\mathbf{s},\mathbf{r})$ for a given $\mathbf{s}$ can be estimated using the second-order Taylor series approximation, in which the gradient and Hessian of $f(\mathbf{s},\mathbf{r})$ with respect to $\mathbf{r}$ are

calculated based on the mean function of its GP surrogate. With reasonable calculations of the objectives and constraints of problems (1) and (2), an optimization strategy is devised to guide the BO toward better candidate solutions and to offer good Pareto-optimal solutions to each problem when it terminates. In particular, the maximization problem of an acquisition function is formulated in each iteration of the BO as a mixed-integer nonlinear programming (MINLP) problem that is solved using a novel random sampling method coupled with simulated annealing. Finally, the proposed BO is applied to finding the robust designs of a two-bar truss and two planar steel frames. The obtained robust designs are then compared with exact robust designs or verified by the MCS.

The remainder of this paper is structured as follows. Section 2 describes two key ingredients of the BO in which two acquisition functions for problems (1) and (2) are developed. Section 3 introduces a new optimization strategy for solving the MINLP problem of the acquisition functions. The optimization procedure with the proposed BO is also introduced in this section. In Section 4, the performance of the proposed BO is demonstrated through a simple bi-objective RDO problem of a two-bar truss. Section 5 provides two design examples of planar steel frames. Finally, Section 6 draws main conclusions of this paper. Detailed mathematical foundations of the GP model and the saddlepoint approximation are provided in Appendices A and B, respectively.

2. Proposed Bayesian optimization

2.1. GP models as surrogates for uncertain objective and probabilistic constraint functions

Consider a training dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$, where $\mathbf{x}_i = \left[\mathbf{s}_i^T, \mathbf{r}_i^T\right]^T \in \mathbb{R}^d$ $(d = d_1 + d_2)$ are d-dimensional vectors of uncertain input variables and $y_i \in \mathbb{R}$ are the corresponding output variables (or uncertain LSFs). $\mathcal{D}$ is created by randomly generating a set of N feasible samples of the discrete design variables $\mathbf{s}$ and the uncertain parameters $\mathbf{r}$ using Latin-hypercube sampling (LHS). The number of initial sample points N depends on the number of design variables and uncertain parameters d, for example, $N \geq 15d$ [34]. As $\mathbf{s}$ for a frame design is integer, the values of its samples are determined by rounding the corresponding real values generated by LHS to the nearest integers. The LSFs y_i are

calculated using a finite element analysis. Since it is desirable to obtain highly accurate GP models for the uncertain objective function and the constraint functions over the region of non-positive LSFs $f(\mathbf{s},\mathbf{r}) \leq 0$ and $g_i(\mathbf{s},\mathbf{r}) \leq 0$, the feasible samples for creating $\mathcal{D}$ are defined as those associated with these non-positive LSFs.

To describe the relationship between the uncertain input and output variables, the mapping $y = \hat{g}(\mathbf{x})\colon \mathbb{R}^d \to \mathbb{R}$ is established, where $\hat{g}(\mathbf{x})$ at a particular input variable vector $\mathbf{x}$ is a probabilistic regression model that is characterized by a mean and a variance, for which the detailed derivations are provided in Appendix A. More specifically, the posterior distribution in Eq. (A.7) is used in each iteration of the BO as $\hat{g}(\mathbf{x})$, which can be either the uncertain objective function or the probabilistic constraint functions of problems (1) and (2). Starting with a Gaussian prior over $\hat{g}(\mathbf{x})$ and a likelihood in Eq. (A.1), the posterior distribution is derived by conditioning on the training dataset $\mathcal{D}$. This posterior distribution contains information about the mean and variance of the prediction of an output variable for a new set of the input variables, as provided in Eqs. (A.8) and (A.9), respectively.

2.2. Acquisition functions

In the BO context, the acquisition function maps our belief about an improvement in the current solution to a measure of how promising each point in the input variable space is if it is evaluated in the next iteration. Thus, the most promising point should maximize the acquisition function. This point is then used to refine the GP surrogates for the uncertain objective and probabilistic constraint functions before the BO starts a new iteration.

Let $\Omega=\{\mathbf{f}_1, \dots, \mathbf{f}_M\} \in \mathbb{R}^k$ and $\mathbf{f}_\mathrm{R} \in \mathbb{R}^k$ denote the current set of M Pareto-optimal solutions in a space of k objective functions and a fixed reference point that is dominated by all elements of Ω, respectively. Here, $\mathbf{f}_\mathrm{R}$ does not necessarily correspond to a particular vector of the input variables. As a result, Ω and $\mathbf{f}_\mathrm{R}$ together define a so-called hypervolume (HV) indicator measure [35] that is used to assess the quality of different sets of Pareto-optimal solutions to a multi-objective optimization problem. In other words, if two different sets of Pareto-optimal solutions to the multi-objective optimization problem are in consideration, the set with larger HV is better than the other one. For a better

understanding, Fig. 1(a) shows an example of the HV for a set of four Pareto-optimal solutions to a bi-objective minimization problem. In this case, the HV is defined by an area measure. In a general case, however, it is a Lebesgue measure of the k-dimensional subspace dominated by Ω, and is bounded above by $\mathbf{f}_{\mathrm{R}}$. Mathematically, the HV is given as follow:

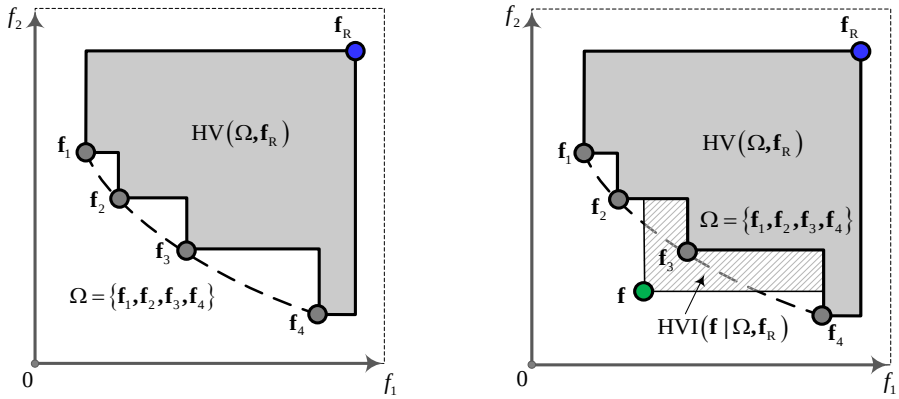
$$\mathrm{HV}(\Omega, \mathbf{f}_{\mathrm{R}}) = \Lambda\left(\left\{\mathbf{f} \in \mathbb{R}^k \middle| \exists \mathbf{f}_m \in \Omega : \mathbf{f}_m \preccurlyeq \mathbf{f} \text{ and } \mathbf{f} \preccurlyeq \mathbf{f}_{\mathrm{R}}\right\}\right) \tag{3}$$

where $\Lambda(\cdot)$ denotes the Lebesgue measure, $\mathbf{f}$ is a point in the k-dimensional space of the objective functions, and $\mathbf{f}_m \preccurlyeq \mathbf{f}$ implies $\mathbf{f}_m$ dominates $\mathbf{f}$. The HV in this study is estimated using a sampling-based technique incorporated in the MATLAB function *hypervolume* [36].

To improve the solutions after each iteration of the BO, a new sampling point $\mathbf{x}_{\mathrm{n}} = [\mathbf{s}_{\mathrm{n}}^T, \mathbf{r}_{\mathrm{n}}^T]^T$ in the input variable space should be determined so that the union of its objective function vector and Ω forms a new measure of HV better than that of the previous iteration. To indicate this improvement, a hypervolume improvement indicator [37], denoted as HVI, is formulated as

$$\mathrm{HVI}(\mathbf{f}|\Omega, \mathbf{f}_{\mathrm{R}}) = \mathrm{HV}(\mathbf{f} \cup \Omega, \mathbf{f}_{\mathrm{R}}) - \mathrm{HV}(\Omega, \mathbf{f}_{\mathrm{R}}) \tag{4}$$

where $\mathbf{f} \cup \Omega$ denotes the union of $\mathbf{f}$ and Ω. It is desirable that $\mathrm{HVI} > 0$ for an improvement in the HV. For a significant improvement, the HVI should be maximized. Fig. 1(b) shows an example of the HVI (i.e., hatched area) for an objective function vector $\mathbf{f}$ and a set of four Pareto-optimal solutions to a bi-objective minimization problem.



(a) (b)

Fig. 1. Examples of the HV and HVI with two objective functions: (a) HV (gray region); (b) HVI (hatched region).

Although the HVI in Eq. (4) can be used to assess whether an improvement in the current solutions is achieved after the BO specifies the new design $\mathbf{s}_n$ for the next iteration, it becomes useless if $\mathbf{s}_n$ is infeasible. This requirement, in the presence of uncertainty, can be transformed into maximizing the chance that a candidate solution $\mathbf{s}$ satisfies both the probabilistic and deterministic constraints of each RDO problem.

For problem (1), let $\Delta_{1,1}(\mathbf{s},\mathbf{r})$ be a function to express the expectation that $\mathbf{s}$ satisfies the joint probabilistic constraint. Here, $\mathbf{r}$ is referred to as a particular point in the random parameter space. $\Delta_{1,1}(\mathbf{s},\mathbf{r})$, therefore, can be formulated as follows:

$$\bar{g}(\mathbf{s},\mathbf{r}) = \max\{g_i(\mathbf{s},\mathbf{r}),\dots,g_I(\mathbf{s},\mathbf{r})\} \approx \max\{\hat{g}_i(\mathbf{s},\mathbf{r}),\dots,\hat{g}_I(\mathbf{s},\mathbf{r})\} \tag{5}$$

$$\Delta_{1,1}(\mathbf{s},\mathbf{r}) = \mathbb{P}[\bar{g}(\mathbf{s},\mathbf{r}) \le 0\,] - 1 + \varepsilon \tag{6}$$

where $\hat{g}_i(\mathbf{s},\mathbf{r})$, as described by Eq. (A.7), is the GP model for the probabilistic constraint function $g_i(\mathbf{s},\mathbf{r})$. Thus, if $\mathbf{x} = [\mathbf{s}^T,\mathbf{r}^T]^T$ maximizes $\Delta_{1,1}$, it increases the chance for $\mathbf{s}$ to satisfy the joint probabilistic constraint of problem (1).

For problem (2), let $\Delta_{1,2}(\mathbf{s},\mathbf{r})$ be a function expressing the expectation that $\mathbf{s}$ satisfies all individual probabilistic constraints. Since $g_i(\mathbf{s},\mathbf{r})$ is represented by the GP model $\hat{g}_i(\mathbf{s},\mathbf{r})$ with mean $\mu_{\hat{g}_i}(\mathbf{s},\mathbf{r})$ and standard deviation $\sigma_{\hat{g}_i}(\mathbf{s},\mathbf{r})$, and all probabilistic constraints are statistically independent in this problem, $\Delta_{1,2}(\mathbf{s},\mathbf{r})$ can be defined as follows:

$$\mathbb{P}[g_i(\mathbf{s},\mathbf{r}) \le 0\,] \approx \mathbb{P}[\hat{g}_i(\mathbf{s},\mathbf{r}) \le 0\,] = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{-\mu_{\hat{g}_i}(\mathbf{s},\mathbf{r})}{\sqrt{2}\sigma_{\hat{g}_i}(\mathbf{s},\mathbf{r})}\right)\right] \tag{7}$$

$$\begin{aligned}\Delta_{1,2}(\mathbf{s},\mathbf{r}) &= \prod_{i=1}^{I}[\mathbb{P}[g_i(\mathbf{s},\mathbf{r}) \le 0\,] - 1 + \varepsilon_i] \\ &\approx \prod_{i=1}^{I}\left\{\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{-\mu_{\hat{g}_i}(\mathbf{s},\mathbf{r})}{\sqrt{2}\sigma_{\hat{g}_i}(\mathbf{s},\mathbf{r})}\right)\right] - 1 + \varepsilon_i\right\}\end{aligned} \tag{8}$$

where erf(·) indicates the Gauss error function to evaluate the CDF of the Gaussian variable $\hat{g}_i(\mathbf{s},\mathbf{r})$. If $\mathbf{x}$ maximizes $\Delta_{1,2}$, it also increases the chance for $\mathbf{s}$ to satisfy all individual probabilistic constraints of problem (2).

For both problems, the following feasibility indicator function $\Delta_2(\mathbf{s})$ is used to indicate whether $\mathbf{s}$ satisfies all deterministic constraints $h_j(\mathbf{s}) \leq 0$

$$\Delta_2(\mathbf{s}) = \begin{cases} 1, & \text{if } h_j(\mathbf{s}) \leq 0,\ j = 1,\dots,J \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Finally, to incorporate different improvement criteria in the objective functions and in the constraint functions into a single improvement criterion, the following acquisition functions $\alpha_1(\mathbf{s},\mathbf{r})$ and $\alpha_2(\mathbf{s},\mathbf{r})$ are formulated for problems (1) and (2), respectively.

$$\alpha_1(\mathbf{s},\mathbf{r}) = \mathrm{HVI}(\mathbf{f}|\Omega,\mathbf{f}_\mathrm{R})\Delta_{1,1}(\mathbf{s},\mathbf{r})\Delta_2(\mathbf{s}) \tag{10}$$

$$\alpha_2(\mathbf{s},\mathbf{r}) = \mathrm{HVI}(\mathbf{f}|\Omega,\mathbf{f}_\mathrm{R})\Delta_{1,2}(\mathbf{s},\mathbf{r})\Delta_2(\mathbf{s}) \tag{11}$$

For each RDO problem, therefore, the next sampling point, denoted as $\mathbf{x}_\mathrm{n} = [\mathbf{s}_\mathrm{n}^T, \mathbf{r}_\mathrm{n}^T]^T$, is the maximizer of the corresponding acquisition function. Maximizing $\alpha_1(\mathbf{s},\mathbf{r})$ or $\alpha_2(\mathbf{s},\mathbf{r})$ is associated with solving an MINLP problem because each acquisition function is nonlinear, and $\mathbf{s}$ and $\mathbf{r}$ are discrete and continuous vectors, respectively. To handle this MINLP problem, the next section develops an optimization strategy.

3. Solution approach

3.1. Sorting Pareto-optimal solutions

Let Ω_a denote a set of already-generated candidate solutions $\mathbf{s}$ that consists of all design points of the initial training dataset at the first iteration of the BO as well as all candidate solutions generated so far at each of the other iterations. Based on Ω_a, an elitist non-dominated sorting approach [38] is employed to find approximate Pareto-optimal solutions to the RDO problems in each iteration of the BO. This approach requires efficient evaluations of the mean and variance of the uncertain objective function, i.e., $f_2(\mathbf{s})$ and $f_3(\mathbf{s})$, and the probabilities in the probabilistic constraints of problems (1) and (2) for

obtaining a set of feasible solutions, denoted as Ω_f. Since $f(\mathbf{s},\mathbf{r})$ in each BO iteration is approximated by $\hat{f}(\mathbf{s},\mathbf{r})$, Eqs. (B.1) and (B.2), which are derived from the second-order Taylor series expansion, are used to evaluate the mean $f_2(\mathbf{s})$ and the variance $f_3(\mathbf{s})$, respectively. Furthermore, the joint and individual probabilistic constraints in each iteration of the BO can be estimated using a saddlepoint approximation [33], which is derived based on the GP surrogates for the probabilistic constraint functions; cf. Appendix B.

3.2. Maximizing the acquisition functions

The MINLP problem of two acquisition functions in Eqs. (10) and (11) can be stated as follows:

$$[\mathbf{s}_n,\mathbf{r}_n]=\underset{\mathbf{s},\mathbf{r}}{\operatorname{argmax}}\left[\alpha_{1(2)}(\mathbf{s},\mathbf{r})\right]$$

subject to

$$\mathrm{HVI}(\mathbf{f}|\Omega,\mathbf{f}_R)>0$$

$$s_i\in\mathcal{S}_i,\ \ i=1,\dots,d_1$$

$$\mathbf{r}\in\mathcal{R} \tag{12}$$

where $\alpha_1(\mathbf{s},\mathbf{r})$ and $\alpha_2(\mathbf{s},\mathbf{r})$ correspond to problems (1) and (2), respectively; and the set $\mathcal{R}$ is independent of the risk levels and assigned as the 95% confidence interval of the uncertain parameters $\mathbf{r}$ to solve problem (12).

Although several techniques are available for solving a convex MINLP problem such as branch-and-bound method, single-tree method, multi-tree method, cutting plane method, and mixed-integer second-order cone program, they are not applicable to problem (12) because $\alpha_1(\mathbf{s},\mathbf{r})$ and $\alpha_2(\mathbf{s},\mathbf{r})$ are non-convex functions. One approach is first to replace $\alpha_1(\mathbf{s},\mathbf{r})$ and $\alpha_2(\mathbf{s},\mathbf{r})$ with piecewise linear approximations [39] and then to solve the resulting approximate problem using a mixed-integer linear programming algorithm. However, as both $\alpha_1(\mathbf{s},\mathbf{r})$ and $\alpha_2(\mathbf{s},\mathbf{r})$ are multivariate-implicit functions, modeling their piecewise linear approximations may be impossible. Another approach is to use population-based heuristic methods, e.g., genetic algorithm, differential evolution, or particle swarm optimization, but they would significantly increase the computational burden for the task of maximizing the acquisition function, especially when either $\mathbf{s}$ or $\mathbf{r}$ is a high-dimensional vector.

As $\mathrm{HVI}(\mathbf{f}|\Omega, \mathbf{f}_{\mathrm{R}})$ and $\Delta_2(\mathbf{s})$ are functions of $\mathbf{s}$, and $\mathbf{r}$ only appears in $\Delta_{1,1}(\mathbf{s}, \mathbf{r})$ and $\Delta_{1,2}(\mathbf{s}, \mathbf{r})$, an optimization strategy that couples a random sampling method with simulated annealing is developed for solving problem (12). This strategy is an extension of a two-stage random search proposed by the authors [40], which includes a stage of determining $\mathbf{r}$, followed by a stage of determining $\mathbf{s}$ for solving upper- and lower-level optimization of a single-objective discrete RDO problem with deterministic constraints and unknown-but-bounded uncertainty in the design parameters. In this study, these two stages are further divided into the following four steps: (1) generate a set Ω_{s} of a finite number of new candidate solutions $\mathbf{s}$; (2) calculate $\mathrm{HVI}(\mathbf{f}|\Omega, \mathbf{f}_{\mathrm{R}})$ and $\Delta_2(\mathbf{s})$ for each member of Ω_{s}, and retain in Ω_{s} the members that yield positive $\mathrm{HVI}(\mathbf{f}|\Omega, \mathbf{f}_{\mathrm{R}})$, $\Delta_2(\mathbf{s}) = 1$, and negative values of the approximate constraint functions; (3) formulate problem (12) for each retained member of Ω_{s} and solve it using simulated annealing [41] to find the associated uncertain parameters $\mathbf{r}$; and (4) select the set of $\mathbf{s}$ and $\mathbf{r}$ that maximizes $\alpha_1(\mathbf{s}, \mathbf{r})$ or $\alpha_2(\mathbf{s}, \mathbf{r})$ and assign it as the next sampling point $\mathbf{x}_{\mathrm{n}} = [\mathbf{s}_{\mathrm{n}}^T, \mathbf{r}_{\mathrm{n}}^T]^T$. It is worth noting that any gradient-based or population-based algorithms can be used for solving problem (12) formulated in step (3) with a given $\mathbf{s}$. However, the simulated annealing is selected in this study because it is able to overcome the drawbacks of both the gradient-based and population-based algorithms that are premature termination and computationally demanding, respectively.

To create Ω_{s} in the first step, two groups of new candidate solutions $\mathbf{s}$ are generated. The first group consists of the points generated by performing random perturbations (in the design variable space) surrounding each of the current Pareto-optimal solutions, which can be regarded as a neighborhood search. In fact, each integer element of every Pareto-optimal solution is randomly increased or decreased by an integer value such as 1, 2, 3, or 4. Here, it is expected that a significant improvement in HV can be achieved by performing the neighborhood search in the design variable space for the current Pareto-optimal solutions even though the neighborhood in this space differs from that in the objective function space. Note that the generation of the first group may slow the proposed BO if the current Pareto-optimal solutions involve many members, say, more than 20. In this case, the Pareto-optimal solutions in the objective function space can be first divided into a moderately small number of disjoint clusters using the Gaussian mixture model clustering method [42]. In this way, the solutions

from the same cluster can be referred to as the samples generated from a Gaussian of which the mean vector is defined as the center of the cluster. Then, a representative member for each cluster can be selected as it is nearest to the center of the cluster. Hence, the random perturbations, in the design variable space, can be performed surrounding each of the representative members. The second group of Ω_s is generated by uniformly sampling points from the design domain, which can be regarded as a global search. Once Ω_s is created, its members that already appear in the training dataset $\mathcal{D}$ are deleted.

To enrich the feasible set Ω_f used for sorting Pareto-optimal solutions (cf. Section 3.1), the current set of new candidate solutions Ω_s is added to the current set of already-generated solutions Ω_a to construct a new set Ω_a that is used in the next iteration of the BO.

3.3. Optimization procedure

Fig. 2 summarizes the optimization procedure for solving problems (1) and (2) using the proposed BO. The following six steps are executed sequentially.

Step 1: Generate samples of the design variables **s** and the uncertain parameters **r** using LHS. Also generate the training dataset $\mathcal{D}$ by performing the finite element analysis for these samples; cf. Section 2.1.

Step 2: Based on the generated training dataset, construct GP models to approximate LSFs of the uncertain objective and probabilistic constraint functions; cf. Section 2.1 and Appendix A.

Step 3: Sort the Pareto-optimal solutions from the set of already-generated candidate solutions Ω_a; cf. Section 3.1.

Step 4: Terminate the BO and output the Pareto-optimal solutions if one of the following criteria is satisfied: (1) the number of BO iterations reaches an upper limit, which is specified by the user to manage the trade-off between the solution quality and the computational cost for carrying out the BO; (2) the difference of the current HV and that of the previous iteration is less than a small positive value, e.g., 10^{-9}; and (3) the set Ω_s used for maximizing the acquisition function has no feasible solution; cf. Section 3.2. Otherwise, go to Step 5.

Step 5: Maximize the acquisition function for each RDO problem using the optimization strategy as described in Section 3.2 to obtain the sampling point $\mathbf{x}_n$ for the next iteration of the BO.

Step 6: Add $\mathbf{x}_n$ obtained from Step 5 to the training dataset, determine the associated structural responses, update the GP models for LSFs of the uncertain objective and probabilistic constraint functions, and reiterate from Step 3.

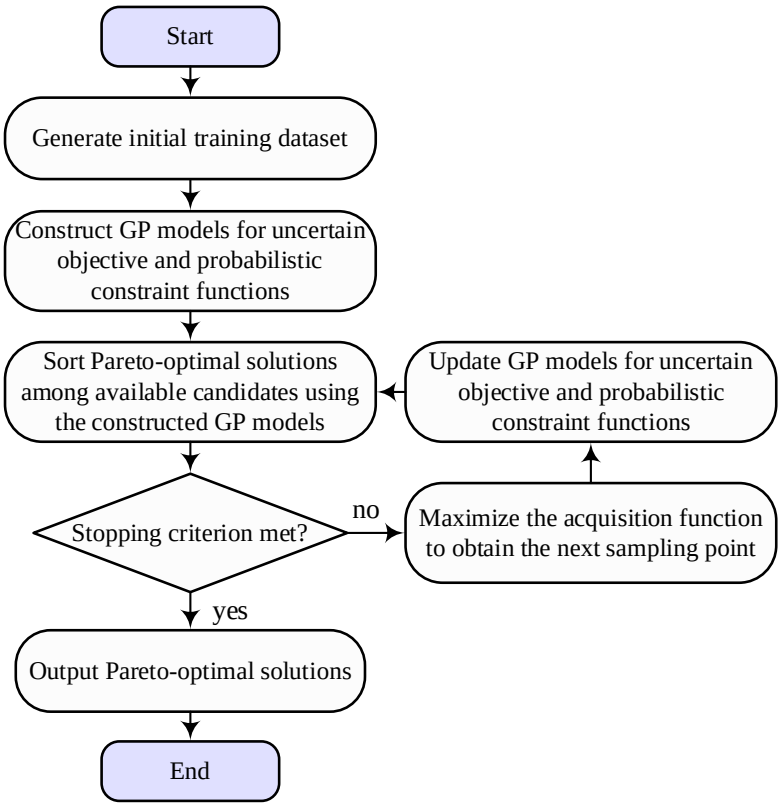


Fig. 2. Proposed BO procedure.

4. Test problem

To demonstrate the feasibility of the proposed BO in solving a simple multi-objective RDO problem with probabilistic constraints, this section investigates a two-bar truss, as shown in Fig. 3, which is taken from Ref. [15]. The truss is subjected to an external load P whose projections onto the horizontal and vertical axes, namely, P_x and P_y, satisfy $P_y = 8P_x$. As stated in the original problem [15], two design variables of the truss are the cross-sectional area s_1 of its members and the horizontal span s_2, i.e., $\mathbf{s} = [s_1, s_2]^T$. The uncertain parameters include the magnitude of the external load P, and the mass

density ρ and yield stress σ_{y} of the truss material, i.e., $\mathbf{r} = \left[\rho, P, \sigma_{\mathrm{y}}\right]^T$. Probabilistic characteristics of these parameters are taken from Ref. [15] as given in Table 1, where COV is the coefficient of variation computed by dividing the standard deviation by the mean value.

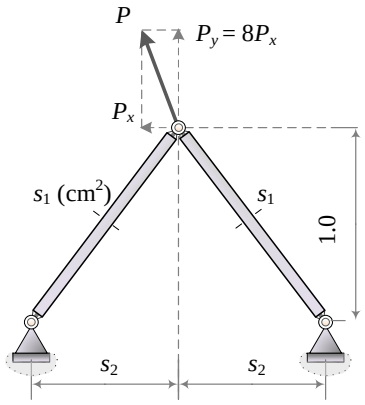


Fig. 3. Two-bar truss.

Table 1 Uncertain parameters for the two-bar truss [15].

Parameter	Description	Mean	COV	Distribution
ρ	Mass density [kg/m^3]	10^4	0.20	Lognormal
P	External load [kN]	800	0.25	Lognormal*
σ_{y}	Yield stress [MPa]	1050	0.24	Normal

* assumed

As it is desirable to formulate a multi-objective RDO problem with conflicting objective functions as problems (1) and (2) for assessing the performance of the proposed BO, the truss is designed to minimize both the mean and standard deviation of its total mass under individual probabilistic constraints on the axial stress in its members. Let $f^*(\mathbf{s}, \mathbf{r})$ denote the total mass of the truss, and $g_1(\mathbf{s}, \mathbf{r})$ and $g_2(\mathbf{s}, \mathbf{r})$ indicate the LSFs of the axial stress in its members. The axial stress due to the self-weight of the truss elements is not considered to keep the problem as simple as the original problem [15]. Let $f_1(\mathbf{s}) = \mathbb{E}[f^*(\mathbf{s}, \mathbf{r})]$ and $f_2(\mathbf{s}) = \sqrt{\mathrm{var}[f^*(\mathbf{s}, \mathbf{r})]}$ represent the mean and standard deviation of $f^*(\mathbf{s}, \mathbf{r})$, respectively. The RDO problem of the truss, therefore, is formulated as follows:

$$\min_{\mathbf{s}}.[f_1(\mathbf{s}), f_2(\mathbf{s})]$$

subject to

$$\mathbb{P}[g_1(\mathbf{s},\mathbf{r}) \le 0\ ] \ge 1-\varepsilon_1$$

$$\mathbb{P}[g_2(\mathbf{s},\mathbf{r}) \le 0\ ] \ge 1-\varepsilon_2$$

$$s_1 \in \mathcal{S}_1 = \{1.0, 1.5, \dots, 20.0\}\ \text{cm}^2$$

$$s_2 \in \mathcal{S}_2 = \{0.1, 0.2, \dots, 2.0\}\ \text{m} \tag{13}$$

where $\mathcal{S}_1$ and $\mathcal{S}_2$ are respectively the list of possible values of s_1 and s_2, and

$$f^*(\mathbf{s},\mathbf{r}) = 10^{-4}\rho s_1\sqrt{1+s_2^2} \tag{14}$$

$$g_1(\mathbf{s},\mathbf{r}) = \frac{5P}{\sqrt{65}s_1\sigma_y}\sqrt{1+s_2^2}\left(8+\frac{1}{s_2}\right)-1 \tag{15}$$

$$g_2(\mathbf{s},\mathbf{r}) = \frac{5P}{\sqrt{65}s_1\sigma_y}\sqrt{1+s_2^2}\left(8-\frac{1}{s_2}\right)-1 \tag{16}$$

Before applying the proposed BO to solving problem (13), it should be noted that the mean $f_1(\mathbf{s})$ can be derived as a linear function of the standard deviation $f_2(\mathbf{s})$, i.e., $f_1(\mathbf{s}) = \left(\mathbb{E}[\rho]/\sqrt{\text{var}[\rho]}\right)f_2(\mathbf{s}) = 5f_2(\mathbf{s})$. This derivation relies on the fact that $f^*(\mathbf{s},\mathbf{r})$ in Eq. (14) is a linear function of the uncertain parameter ρ. The set of Pareto-optimal solutions to problem (13), therefore, includes only one solution.

As the first step in solving problem (13), a total of 200 samples are generated for the initial training dataset. Only 140 feasible samples are kept for the training purpose as the other 60 samples associated with either positive $g_1(\mathbf{s},\mathbf{r})$ or positive $g_2(\mathbf{s},\mathbf{r})$ are discarded. Then, the DACE toolbox [43] are used together with a second-degree polynomial mean function (cf. Appendix A) for constructing three GP surrogates for $f^*(\mathbf{s},\mathbf{r})$, $g_1(\mathbf{s},\mathbf{r})$, and $g_2(\mathbf{s},\mathbf{r})$ in Eqs. (14), (15), and (16), respectively. Based on these surrogates, the proposed BO is employed to solve problem (13) with two risk levels, namely, $\varepsilon_i = 0.1$ and $\varepsilon_i = 0.05$ $(i = 1,2)$. The simulated annealing incorporated in the BO is carried out using MATLAB R2018a Global Optimization Toolbox. Default settings for the acceptance probability function, annealing schedule, initial annealing temperature, and maximum number of evaluations of the

objective function are 'acceptancesa', 'annealingfast', 100, 3000 times the dimension of **r**, respectively. The maximum number of iterations and the objective function tolerance of the simulated annealing are assigned as 500 and 10^{-6}, respectively. In each iteration of the BO, a total of 200 new candidate solutions are generated for the set Ω_s. Furthermore, the reference point and the maximum iteration of the BO are assigned as $\mathbf{f}_R = [50, 10]^T$ and 20, respectively. Here, $\mathbf{f}_R$ is chosen based on (1) the maximum value of the truss total mass, computed from possible combinations of $\mathcal{S}_1$ and $\mathcal{S}_2$ elements, and (2) the relation $f_1(\mathbf{s}) = 5f_2(\mathbf{s})$.

Fig. 4 shows the maximum acquisition function α_2, the corresponding HVI, and the objective functions $f_1(\mathbf{s})$ and $f_2(\mathbf{s})$ at each iteration of the BO for the two risk levels $\varepsilon_i = 0.1$ and $\varepsilon_i = 0.05$. The respective robust designs of the truss found at the 20th iteration of the BO are $\mathbf{s} = [8.5, 0.4]^T$ and $\mathbf{s} = [10, 0.4]^T$. The computational times required for $\varepsilon_i = 0.1$ and $\varepsilon_i = 0.05$ are 905 and 973 s using a PC with an Intel i7-7700HQ 2.80 GHz CPU and 8.0 GB memory.

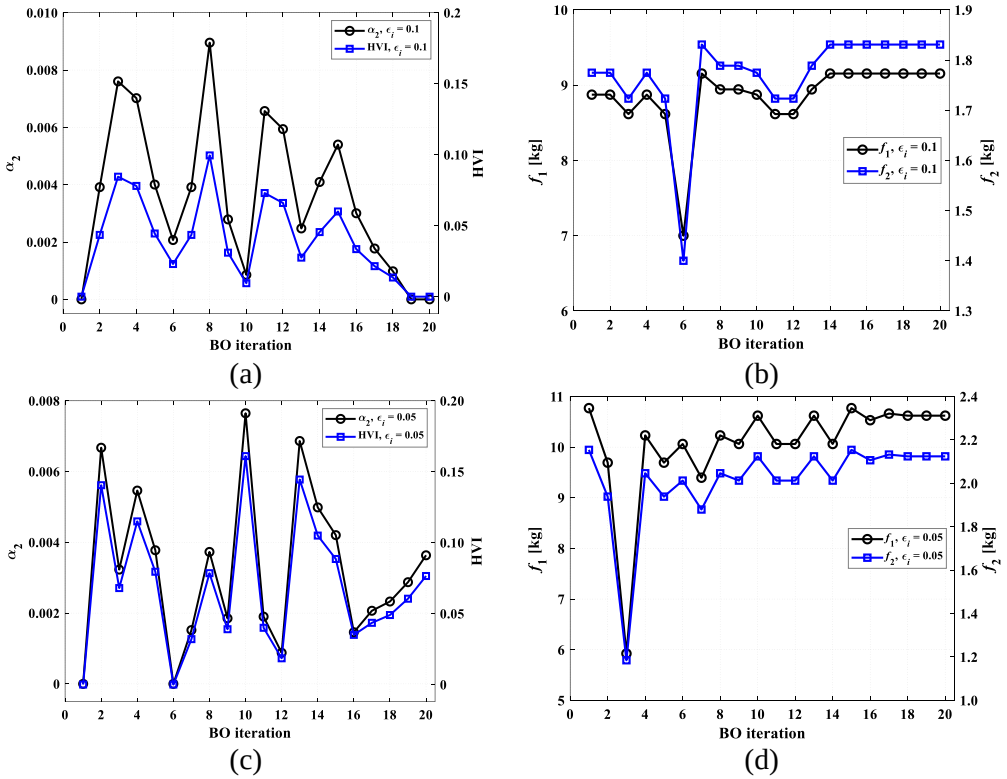


Fig. 4. Histories of the BO for the two-bar truss with two risk levels: (a) and (b) $\varepsilon_i = 0.1$; (c) and (d) $\varepsilon_i = 0.05$.

To verify the obtained robust designs, the exact robust solution to problem (13) is found for each risk level. As the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ have 39 and 20 elements, respectively, a total of 39×20 = 780 possible designs can be generated for the truss using a full factorial sampling. Then, 10^5 samples of the uncertain parameters $\mathbf{r}$ are generated by the MCS, and used to calculate $f_1(\mathbf{s})$, $f_2(\mathbf{s})$, $\mathbb{P}[g_1(\mathbf{s},\mathbf{r}) \leq 0\,]$, and $\mathbb{P}[g_2(\mathbf{s},\mathbf{r}) \leq 0\,]$ for each possible design. This process is performed three times for each risk level with different sets of the MCS samples, which may lead to different solutions. As a result, the solutions by the three trials are identical for each risk level, and therefore they can be regarded as the exact solution. Fig. 5 shows that the robust design by the BO for each risk level is identical to the exact design even though the number of function evaluations used by the BO (1.30122×10^5 evaluations) is much less than that by the combination of the full factorial sampling and MCS ($3\times780\times10^5$ evaluations).

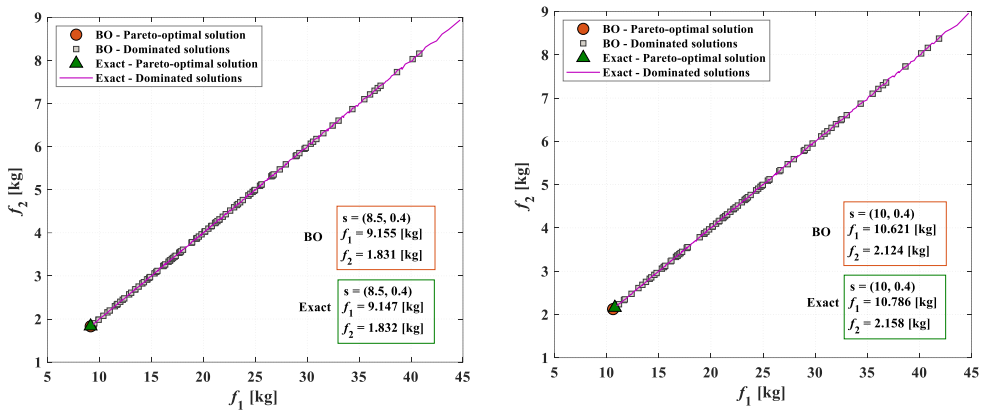


Fig. 5. Verification of the obtained robust designs for the two-bar truss with two risk levels: (a) ε_i = 0.1; (b) ε_i = 0.05.

5. Design examples

5.1. LSFs of serviceability and strength requirements for steel frame designs

The design of two planar steel frames in this section complies with certain serviceability and strength requirements in ANSI/AISC 360-16 [44]. The serviceability requirements are concerned with the inter-story drifts, total drift of the frame, and long-term vertical displacements of the beam members, while the strength requirements are associated with the strength of the column and beam members.

Let δ_s, Δ_{t}, and Δ_k denote the inter-story drift of the sth story of a frame having a total of n_{s} stories, the total drift of the frame, and the maximum long-term vertical displacement among beams in the kth beam group, respectively. Note that the long-term vertical displacement of each beam member is evaluated as 1.5 times the associated elastic displacement [33]. Also, let h_{s}, H, and L_k denote the height of the sth story, the overall height of the frame, and the span of the beams in the kth beam group, respectively. As specified in ANSI/AISC 360-16 [44], the inter-story drift of the story, the total drift of the frame, and the long-term vertical displacement of the kth beam should not exceed $h_{\mathrm{s}}/300$, $H/400$, and $L_k/360$, respectively. Therefore, the LSFs corresponding to the maximum inter-story drift of the frame, i.e., $f(\mathbf{s},\mathbf{r})$, the total drift of the frame, i.e., $g_1(\mathbf{s},\mathbf{r})$, and the maximum long-term vertical displacement among beams in the same group, i.e., $g_{1+k}(\mathbf{s},\mathbf{r})$, can be formulated as follows:

$$f(\mathbf{s},\mathbf{r}) = \max\left\{\frac{|\delta_1|}{h_1/300},\dots,\frac{|\delta_{n_{\mathrm{s}}}|}{h_{n_{\mathrm{s}}}/300}\right\} - 1 \tag{17}$$

$$g_1(\mathbf{s},\mathbf{r}) = \frac{|\Delta_{\mathrm{t}}|}{H/400} - 1 \tag{18}$$

$$g_{1+k}(\mathbf{s},\mathbf{r}) = \frac{\Delta_k}{L_k/360} - 1, \quad k = 1,\dots,n_{\mathrm{b}} \tag{19}$$

where n_{b} is the number of beam groups.

In the same manner, the strength LSF for a column or a beam can be normalized using the following AISC-LRFD interaction formula [44]:

$$q(\mathbf{s},\mathbf{r}) = \begin{cases} \dfrac{P_{\mathrm{r}}}{P_{\mathrm{c}}} + \dfrac{8}{9}\left(\dfrac{M_{\mathrm{r}x}}{M_{cx}} + \dfrac{M_{\mathrm{r}y}}{M_{cy}}\right), & \text{if } \dfrac{P_{\mathrm{r}}}{P_{\mathrm{c}}} \geq 0.2 \\ \dfrac{P_{\mathrm{r}}}{2P_{\mathrm{c}}} + \left(\dfrac{M_{\mathrm{r}x}}{M_{cx}} + \dfrac{M_{\mathrm{r}y}}{M_{cy}}\right), & \text{if } \dfrac{P_{\mathrm{r}}}{P_{\mathrm{c}}} < 0.2 \end{cases} \tag{20}$$

where P_{r} and P_{c} are the required and available axial strengths of the member, respectively; $M_{\mathrm{r}x}$ and $M_{\mathrm{r}y}$ are respectively the required flexural strengths about the major axis x, and the minor axis y; and M_{cx} and M_{cy} denote the available flexural strengths about the major axis x, and the minor axis y, respectively. P_{r}, $M_{\mathrm{r}x}$, and $M_{\mathrm{r}y}$ are obtained from a linear elastic frame analysis under static loads. P_{c}, M_{cx}, and M_{cy} can be evaluated based on specifications in chapters E, F, and H of ANSI/AISC 360-16

[44], which are described as functions of the cross-sectional properties of the steel section, Young's modulus E, the yield stress σ_{y} and the ultimate tensile strength σ_{u} of the steel material, and the effective length of the structural members.

Let $q_{\max}(\mathbf{s},\mathbf{r})$ denote the maximum of $q(\mathbf{s},\mathbf{r})$ among the values for the columns or beams in the same group. Thus, the LSFs corresponding to the combined axial-flexural strength of column and beam members of the frame can be defined as

$$g_{1+n_{\mathrm{b}}+j}(\mathbf{s},\mathbf{r}) = q_{\max,j} - 1, \;\; j = 1,\dots,n_{\mathrm{m}} \tag{21}$$

where n_{m} is the number of column and beam groups.

5.2. Three-story two-bay steel frame

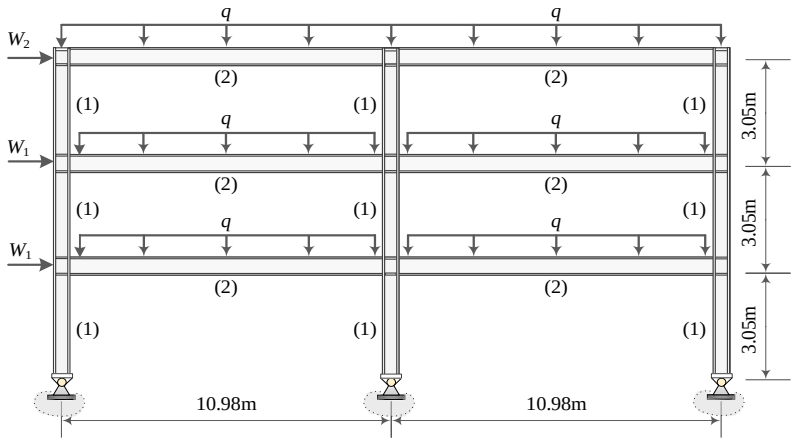


Fig. 6. Three-story two-bay steel frame.

Consider a three-story two-bay steel frame subjected to static loads q, W_1, and W_2, as shown in Fig. 6. The frame consists of 15 members classified into column (1) and beam (2) groups. Steel section for the members in each group is selected from the list of American wide-flange steel sections in Table 2. The independent uncertain parameters $\mathbf{r}$ include the vertical load q (i.e., dead and live loads from the floor), lateral loads W_1 and W_2, and material properties, i.e., Young's modulus E, yield stress σ_{y}, and ultimate tensile strength σ_{u}. Their probabilistic properties, as described in Table 3, are taken from ASCE/SEI 7-16 [45] and a previous survey on uncertainty in the material strength [46]. Note that the mass density of steel is not considered as an uncertain parameter because the nominal mass of the standard column

or beam sections is well controlled by the manufacturers and is provided as a deterministic value as shown in Table 2.

To formulate problems (1) and (2) for the frame, the total mass $f_1(\mathbf{s})$, and the mean $f_2(\mathbf{s})$ and variance $f_3(\mathbf{s})$ of the LSF regarding the maximum inter-story drift $f(\mathbf{s},\mathbf{r})$ are considered as three objective functions, in which $f_1(\mathbf{s})$ is explicitly given as

$$f_1(\mathbf{s}) = 27.45\rho_1 + 65.88\rho_2 \tag{22}$$

where ρ_1 and ρ_2, as provided in Table 2, are the nominal mass [kg/m] of the column and beam sections, respectively.

Table 2 List of sections of columns and beams for the three-story two-bay steel frame.

ID	Column $\mathcal{S}_1$	ρ_1 [kg/m]	Beam $\mathcal{S}_2$	ρ_2 [kg/m]
1	W18×86	128.0	W24×68	101.0
2	W18×76	113.0	W24×62	92.0
3	W16×77	114.0	W24×55	82.0
4	W16×67	100.0	W21×57	85.0
5	W14×82	122.0	W21×55	82.0
6	W14×74	110.0	W21×50	74.0
7	W14×68	101.0	W18×65	97.0
8	W14×61	91.0	W18×60	89.0
9	W14×53	79.0	W18×40	60.0
10	W14×48	72.0	W18×35	52.0
11	W12×58	86.0	W16×57	85.0
12	W12×53	79.0	W16×50	75.0
13	W12×50	74.0	W16×45	67.0
14	W12×45	67.0	W16×40	60.0
15	W10×54	80.0	W14×61	91.0
16	W10×49	73.0	W14×53	79.0
17	W10×45	67.0	W14×48	72.0
18	W8×40	59.0	W14×38	57.8
19	W8×35	52.0	W14×34	51.0
20	W8×31	46.1	W14×30	44.0

Table 3 Uncertain parameters for the three-story two-bay steel frame.

Parameter	Description	Nominal	Mean/ Nominal	COV	Distribution
q	Vertical load [kN/m]	40.9	1.00	0.20	Normal
W_1	Lateral load [kN]	22.2	1.00	0.37	Lognormal
W_2	Lateral load [kN]	11.1	1.00	0.37	Lognormal
E	Young's Modulus [GPa]	200	1.00	0.04	Normal
σ_{y}	Yield stress [MPa]	250	1.05	0.06	Normal
σ_{u}	Tensile strength [MPa]	400	1.05	0.08	Normal

Four probabilistic constraint functions include the total drift of the frame $g_1(\mathbf{s},\mathbf{r})$, maximum long-term vertical displacement of the beams $g_2(\mathbf{s},\mathbf{r})$, combined axial-flexural strength of the columns $g_3(\mathbf{s},\mathbf{r})$, and combined axial-flexural strength of the beams $g_4(\mathbf{s},\mathbf{r})$. Also, one deterministic geometric constraint $h(\mathbf{s}) \leq 0$ is applied to beam-column connections to guarantee that the flange width of the beam connected to a column should not exceed the flange width of the column.

A total of 150 feasible samples are generated to construct five GP models for $f(\mathbf{s},\mathbf{r})$ and four probabilistic constraint functions. The number of new candidate solutions generated in each iteration of the BO is assigned as 200. A PC with an Intel Xeon E5-2643V4 3.40 GHz CPU and 63.9 GB memory are used to implement the proposed BO that solves problems (1) and (2) with two risk levels 0.1 and 0.05. For convenient description, let problems (1-1) and (1-2) refer to problem (1) formulated for $\varepsilon = 0.1$ and $\varepsilon = 0.05$, respectively, and problems (2-1) and (2-2) denote problem (2) formulated with $\varepsilon_i = 0.1$ and $\varepsilon_i = 0.05$ $(i = 1, \ldots, 4)$, respectively. The reference point and the maximum number of iterations of the BO are set as $\mathbf{f}_{\mathrm{R}} = [14000, 0.5, 0.5]^T$ and 20, respectively. As a note to facilitate the selection of $\mathbf{f}_{\mathrm{R}}$, its first element can be selected based on the maximum value of the frame total mass, which is derived from the list of sections in Table 2; the second element of $\mathbf{f}_{\mathrm{R}}$ associated with the mean of $f(\mathbf{s},\mathbf{r})$ in Eq. (17) can be assigned as any positive value so that it bounds above the expected value of $f(\mathbf{s},\mathbf{r})$, which should be non-positive to ensure the structural performance; and the last element of $\mathbf{f}_{\mathrm{R}}$ associated with the variance of $f(\mathbf{s},\mathbf{r})$ should be positive and small enough to ensure the robustness of the frame.

Fig. 7 shows histories of the BO for solving the aforementioned four RDO problems of the frame. As depicted in Fig. 7(c), the graphs associated with the maximum acquisition function and the HVI are similar with respect to an appropriate scale because they are proportional to each other at each iteration of the BO; cf. Eq. (10). Since problems (2-1) and (2-2) involve the probability of simultaneously meeting all probabilistic constraints, α_2 is much less than α_1 at the same risk level. The BO terminates at the 16th iteration when solving problems (1-1) and (2-1), and at the 10th iteration when solving problems (1-2) and (2-2), even though its maximum number of iterations is assigned as 20. These early terminations arise from the fact that the set Ω_s used to maximize the acquisition function in each problem has no feasible solution; cf. Step 4 in Section 3.3. Furthermore, the computational times for the BO to complete 16 and 10 iterations are 2.68 and 1.81 hours, respectively.

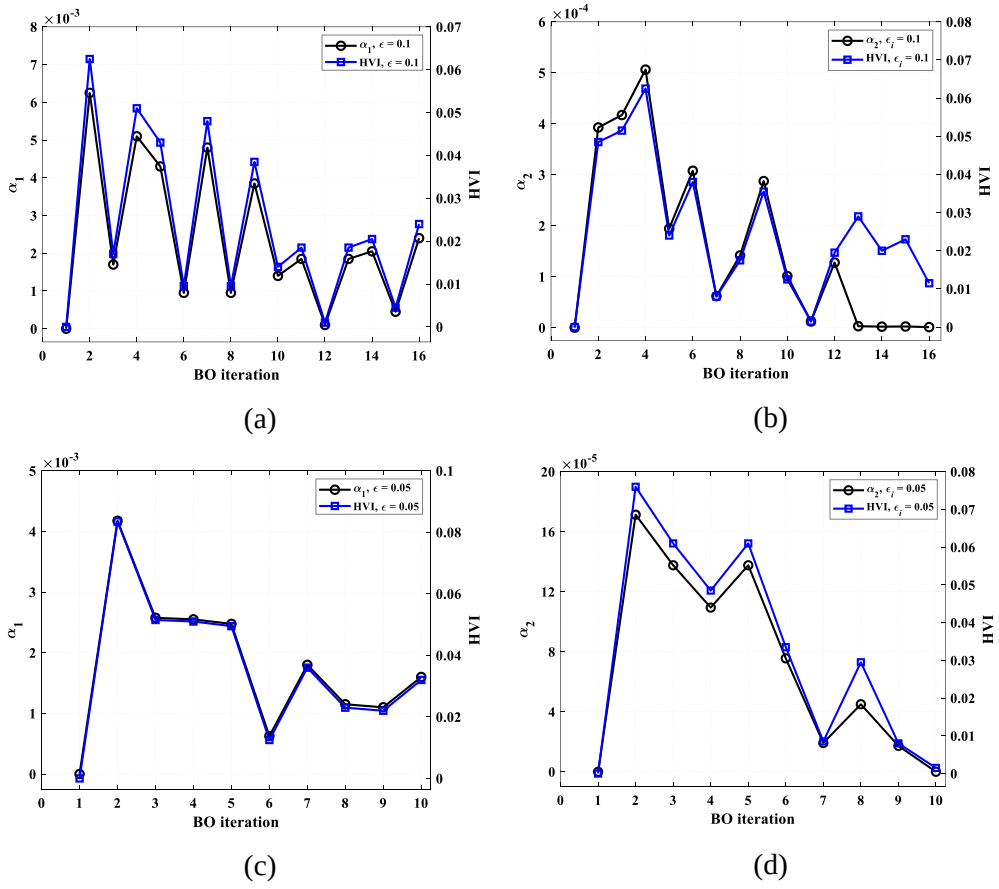


Fig. 7. Histories of the BO for different RDO problems of the three-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).

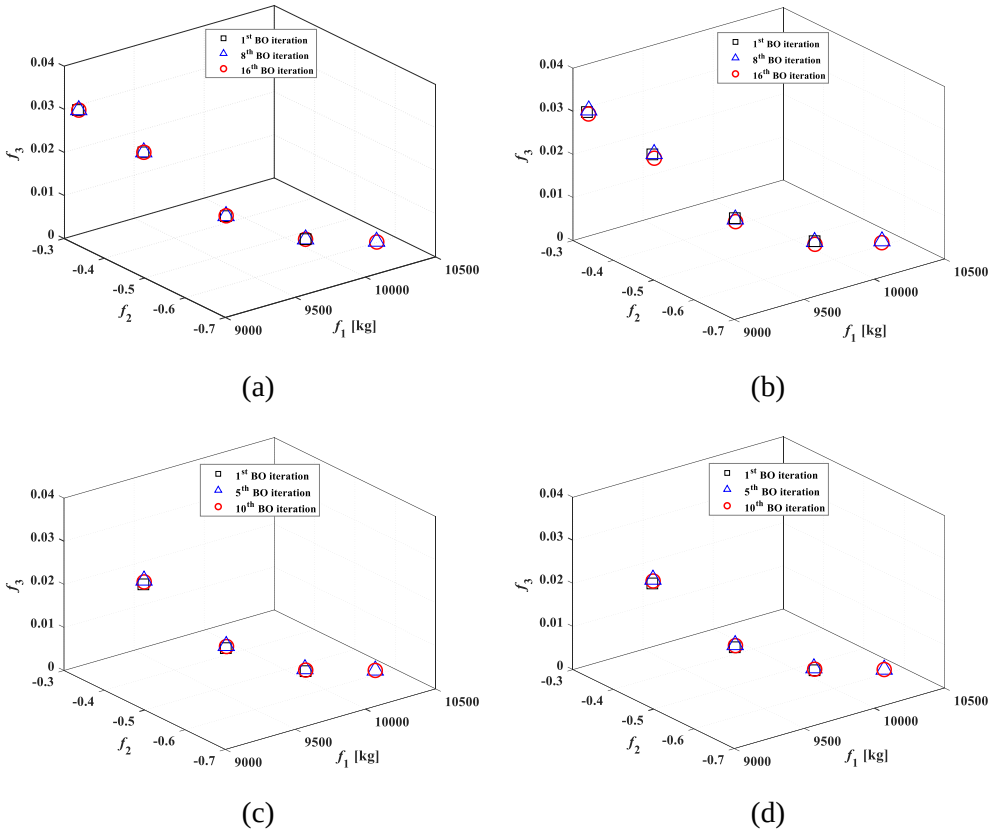


Fig. 8. Pareto-optimal solutions to different RDO problems of the three-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).

Fig. 8 provides four sets of Pareto-optimal solutions corresponding to the mentioned four RDO problems. The solutions to the joint and individual probability constrained RDO problems are identical for each risk level. More interestingly, four Pareto-optimal solutions associated with the risk level 0.05 can be obtained by removing the solution with minimum total mass $f_1(\mathbf{s})$ from the set of five Pareto-solutions for the risk level 0.1. Moreover, as the objective function $f_2(\mathbf{s})$ for all obtained solutions is negative, the mean value of the maximum inter-story drift of each robust design, as expected, is less than the allowable value.

Exact robust solutions to each RDO problem of the frame are found for verifying the obtained solutions. Because the RDO problems of the frame have only two design variables and each variable has 20 possible values as provided in Table 2, it is able to generate a total of 20×20 = 400 possible solutions to each problem. Then, the objective functions $f_2(\mathbf{s})$, $f_3(\mathbf{s})$, and the probabilistic constraints are calculated for all possible solutions to each RDO problem using a total of 10^5 samples of the

uncertain parameters **r** generated by the MCS. These values enable the sorting of the exact Pareto-optimal solutions to each RDO problem. Note that the process for finding the exact solutions takes 10.02 hours to complete, which is much more than that required for the BO. For further comparison, a multi-objective RDO problem for the frame is formulated with considering $f_1(\mathbf{s})$, $f_2(\mathbf{s})$, and $f_3(\mathbf{s})$ as three objective functions, and $\mathbb{E}[g_i(\mathbf{s},\mathbf{r})] + \sqrt{\mathrm{var}[g_i(\mathbf{s},\mathbf{r})]} \leq 0$, $(i = 1,\dots,4)$ as four deterministic constraints. This problem is solved using a Gaussian mixture model (GMM) method [15] with an expectation that the resulting Pareto-optimal solutions can capture all solutions by the BO. As a result, Fig. 9 shows that the solutions to each RDO problem by the BO agree with the exact solutions even though there exists a slight difference in the objective functions $f_2(\mathbf{s})$ and $f_3(\mathbf{s})$. The GMM, as expected, provides a total of six solutions that include all solutions by the BO. Moreover, Tables 4 and 5 provide detailed values of the objective functions and the probabilistic constraints associated with the solutions to each RDO problem. As each solution offers a safety margin in all probabilistic constraints, the obtained Pareto-optimal solutions are feasible.

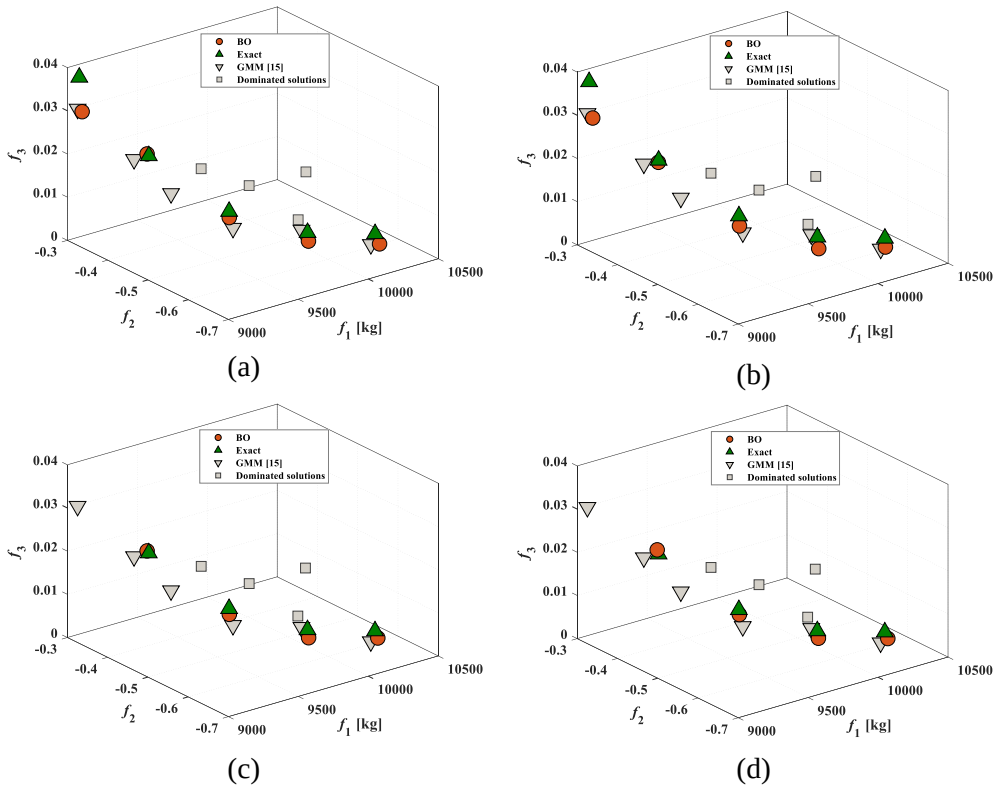


Fig. 9. Verification of the obtained Pareto-optimal solutions to different RDO problems of the three-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).

Table 4 Pareto-optimal solutions to problems (1-1) and (2-1) of the three-story two-bay steel frame and corresponding values of the objective and probabilistic constraint functions.

Variable, objective and constraint functions	Solution 1	Solution 2	Solution 3	Solution 4	Solution 5
Column (1)	W12×58	W14×61	W16×67	W18×76	W18×86
Beam (2)	W24×68	W24×68	W24×68	W24×68	W24×68
f_1 [kg]	9015	9152	9399	9756	10167
f_2 (BO- problem (1-1))	-0.332	-0.445	-0.563	-0.636	-0.670
f_3 (BO - problem (1-1))	0.031	0.025	0.014	0.008	0.005
f_2 (BO- problem (2-1))	-0.333	-0.449	-0.564	-0.637	-0.660
f_3 (BO - problem (2-1))	0.034	0.025	0.013	0.008	0.005
f_2 (GMM)	-0.320	-0.413	-0.573	-0.619	-0.648
f_3 (GMM)	0.031	0.022	0.012	0.010	0.004
f_2 (MCS)	-0.325	-0.444	-0.563	-0.634	-0.659
f_3 (MCS)	0.039	0.025	0.015	0.010	0.007
$\mathbb{P}\left[g_{1,..,4}(\mathbf{s},\mathbf{r}) \leq 0\right]$ (MCS)	0.903	0.951	0.967	0.976	0.979
$\mathbb{P}[g_1(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	0.999	1.000	1.000	1.000	1.000
$\mathbb{P}[g_2(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	0.999	1.000	1.000	1.000	1.000
$\mathbb{P}[g_3(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	0.908	0.974	0.998	1.000	1.000
$\mathbb{P}[g_4(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	0.935	0.953	0.967	0.976	0.979

Table 5 Pareto-optimal solutions to problems (1-2) and (2-2) of the three-story two-bay steel frame and corresponding values of the objective and probabilistic constraint functions.

Variable, objective and constraint functions	Solution 1	Solution 2	Solution 3	Solution 4
Column (1)	W14×61	W16×67	W18×76	W18×86
Beam (2)	W24×68	W24×68	W24×68	W24×68
f_1 [kg]	9152	9399	9756	10167
f_2 (BO- problem (1-2))	-0.445	-0.564	-0.637	-0.666
f_3 (BO - problem (1-2))	0.025	0.014	0.008	0.006
f_2 (BO- problem (2-2))	-0.446	-0.564	-0.637	-0.666
f_3 (BO - problem (2-2))	0.026	0.014	0.008	0.006
f_2 (GMM)	-0.413	-0.573	-0.619	-0.648
f_3 (GMM)	0.022	0.012	0.010	0.004
f_2 (MCS)	-0.444	-0.482	-0.634	-0.659
f_3 (MCS)	0.025	0.022	0.010	0.007
$\mathbb{P}\left[g_{1,..,4}(\mathbf{s},\mathbf{r}) \leq 0\right]$ (MCS)	0.951	0.967	0.976	0.979
$\mathbb{P}[g_1(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	1.000	1.000	1.000	1.000
$\mathbb{P}[g_2(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	1.000	1.000	1.000	1.000
$\mathbb{P}[g_3(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	0.974	0.998	1.000	1.000
$\mathbb{P}[g_4(\mathbf{s},\mathbf{r}) \leq 0]$ (MCS)	0.953	0.967	0.976	0.979

5.3. Six-story two-bay steel frame

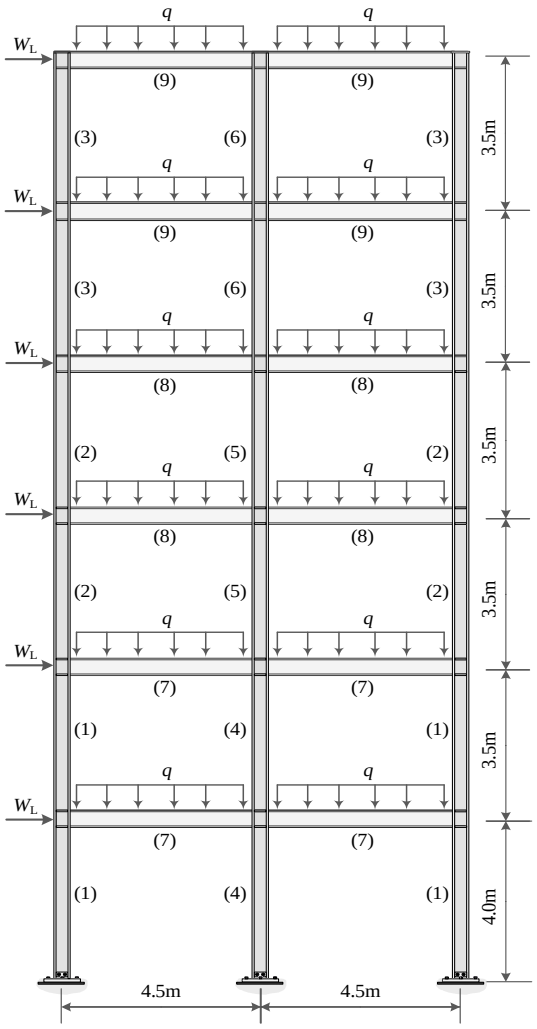


Fig. 10. Six-story two-bay steel frame.

In this design example, problems (1) and (2) are formulated for a six-story two-bay steel frame, as shown Fig. 10, which is taken from Do and Ohsaki [15]. The frame has 30 members classified into six column groups, i.e., groups (1) to (6), and three beam groups, i.e., groups (7), (8), and (9). Steel section for members in each group is also selected from Table 2. The independent uncertain parameters **r** for the frame and their probabilistic characteristics are described in Table 6, where the vertical load q consists of dead and live loads from the floor.

Table 6 Uncertain parameters for the six-story two-bay steel frame.

Parameter	Description	Nominal	Mean/ Nominal	COV	Distribution
q	Vertical load [kN/m]	50	1.00	0.20	Normal
W_{L}	Wind load [kN]	25	1.00	0.37	Lognormal
E	Young's Modulus [GPa]	200	1.00	0.04	Normal
σ_{y}	Yield stress [MPa]	250	1.05	0.06	Normal
σ_{u}	Tensile strength [MPa]	400	1.05	0.08	Normal

Let $g_1(\mathbf{s},\mathbf{r})$, $g_2(\mathbf{s},\mathbf{r})$, $g_3(\mathbf{s},\mathbf{r})$, and $g_4(\mathbf{s},\mathbf{r})$ denote the LSFs of the total drift of the frame, and the maximum long-term vertical displacements of beams (7), (8), and (9), respectively. Let $g_5(\mathbf{s},\mathbf{r})$, $g_6(\mathbf{s},\mathbf{r})$, $g_7(\mathbf{s},\mathbf{r})$, $g_8(\mathbf{s},\mathbf{r})$, $g_9(\mathbf{s},\mathbf{r})$, and $g_{10}(\mathbf{s},\mathbf{r})$ indicate the LSFs of the combined axial-flexural strength of columns (1), (2), (3), (4), (5), and (6), respectively. $g_{11}(\mathbf{s},\mathbf{r})$, $g_{12}(\mathbf{s},\mathbf{r})$, and $g_{13}(\mathbf{s},\mathbf{r})$ represent the LSFs of the combined axial-flexural strength of beams (7), (8), and (9), respectively. In addition to the constraints on the structural responses, 10 deterministic constraints, i.e., $h_j(\mathbf{s}) \leq 0$ $(j = 1,\dots,10)$, are imposed at beam-column connections and column-column joints of the frame to ensure that (1) the flange width of a beam connected to a column should not exceed the flange width of the column and (2) the depth of the column section in the upper story should be less than or equal to the depth of the column section in the lower story. Thus, problems (1) and (2) of the frame are formulated using 13 probabilistic constraint functions and 10 deterministic constraint functions. The total mass $f_1(\mathbf{s})$ of the frame is also given as

$$f_1(\mathbf{s}) = 15\rho_1 + 14(\rho_2 + \rho_3) + 7.5\rho_4 + 7(\rho_5 + \rho_6) + 18(\rho_7 + \rho_8 + \rho_9) \tag{23}$$

where ρ_i $(l = 1,\dots,9)$ is the nominal mass [kg/m] of the steel section for the ith group.

To construct a total of 14 GP models for the structural responses of the frame, one for the uncertain objective function, i.e., the LSF of the maximum inter-story drift $f(\mathbf{s},\mathbf{r})$, and 13 others for $g_1(\mathbf{s},\mathbf{r})$ to $g_{13}(\mathbf{s},\mathbf{r})$, a set of 1000 feasible samples (cf. Section 2.1) of $\mathbf{s}$ and $\mathbf{r}$ is randomly generated. The generation of these samples is detailed in Section 2.1. The proposed BO is then carried out for solving problems (1) and (2) with two risk levels 0.1 and 0.05. Similarly as the frame in the first design example, problems (1-1), (1-2), and (2-1), (2-2) are used to refer to problem (1) formulated with $\varepsilon =$

0.1 and $\varepsilon = 0.05$, and problem (2) formulated with $\varepsilon_i = 0.1$ and $\varepsilon_i = 0.05$ $(i = 1,\ldots,13)$, respectively. The reference point and the maximum number of BO iterations are $\mathbf{f}_{\mathrm{R}} = [15000, 0.5, 0.5]^T$ and 20, respectively.

Fig. 11 shows the history of the BO for each RDO problem of the frame. The BO completes 20 iterations to offer the solutions to problems (1-1), (1-2), and (2-2) while a total of 17 iterations are required for solving problem (2-1). For the same risk level, the maximum value of the acquisition function α_2 is considerably less than that of α_1 because problems (2-1) and (2-2) consider 13 probabilistic constraints independently. The computational times for solving problems (1-1), (1-2), (2-1), and (2-2) are 11.75, 11.68, 10.19 and 10.56 hours, respectively.

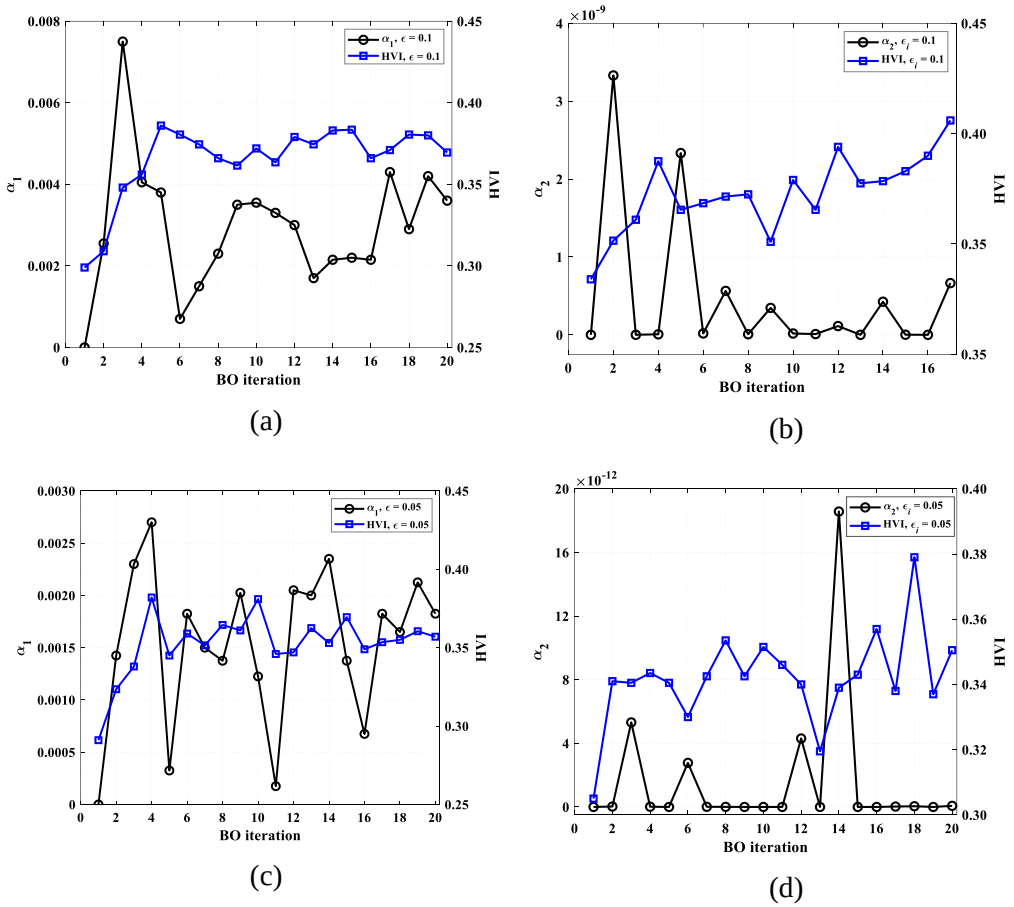


Fig. 11. Histories of the BO for different RDO problems of the six-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).

Fig. 12 shows the Pareto-optimal solutions to the RDO problems of the frame. The number of solutions to problems (1-1), (2-1), (1-2), and (2-2) respectively are 54, 66, 60, and 56 even though they

are 1, 7, 1, and 4 after the BO completes its first iteration. The shapes of Pareto fronts to the same RDO problem are similar, regardless of the difference in the solution distributions due to different risk levels. It is worth noting that the number of solutions, as well as the solutions themselves, may vary when the proposed BO is carried out using different stopping criteria or even when using unique stopping criteria, but it is performed from different random seeds. This can be explained by the following two facts. First, the list of steel sections in Table 2 consists of many feasible solutions to each RDO problem of the frame, which are referred to as the dominated solutions in Fig. 12. Second, the proposed BO is random in nature because it performs a random search method to solve the MINLP problem of the acquisition functions. From the obtained solutions, however, designers are still able to specify a robust design for the frame by handling the trade-off between the three objective functions. Once a robust design has been specified, the remaining task is simply to verify that it satisfies all probabilistic constraints as expected using the MCS.

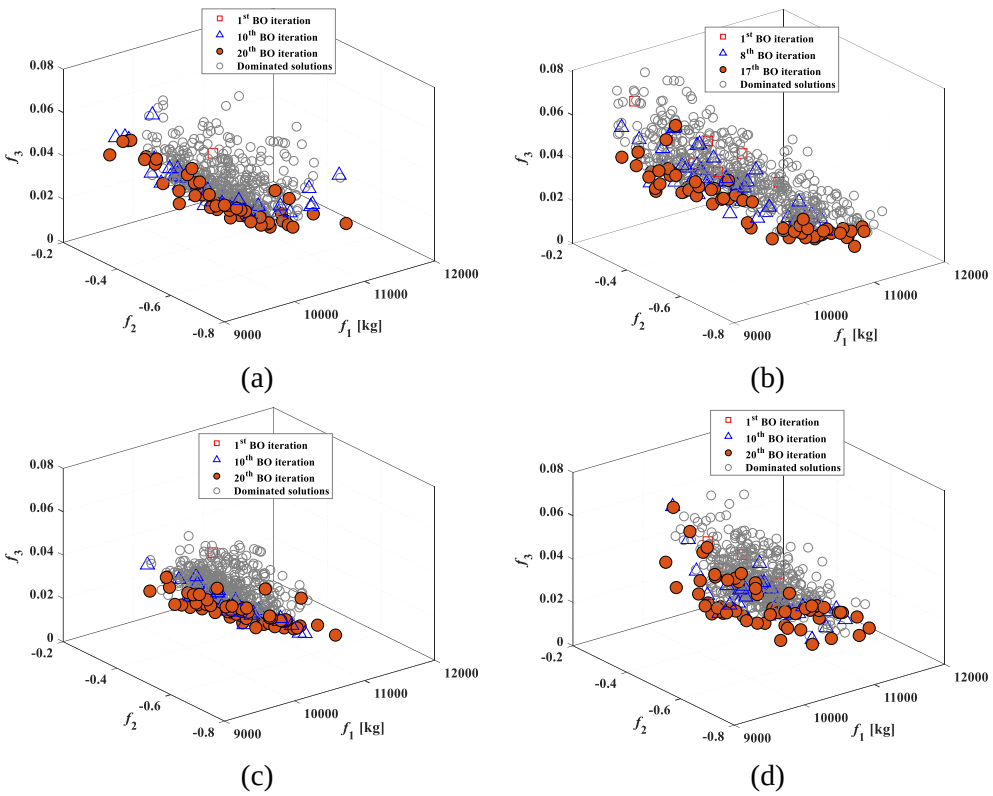


Fig. 12. Pareto-optimal solutions to different RDO problems of the six-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).

For example, among the three conflicting objectives, the designer may assume that the total mass of the frame is the most important objective. Thus, from the set of obtained Pareto-optimal solutions to each RDO problem, the solution with minimum total mass may be selected as the robust design for the frame. With the selected design, the MCS can be performed to generate a total of 10^5 samples of the uncertain parameters **r** that are further used to compute the objective and probabilistic constraint functions for the selected design. Figs. 13 and 14 show the PDF of the uncertain objective function $f(\mathbf{s},\mathbf{r})$ and the CDFs of 13 probabilistic constraint functions calculated at the minimum-mass solution to each RDO problem of the frame, respectively. Since the design associated with the larger risk level is less robust than that corresponding to the smaller risk level, the PDFs in Figs. 13(a) and (b) are broader than those in Figs. 13(c) and (d), respectively. Furthermore, all probabilistic constraints associated with a safety margin, as shown in Fig. 14, demonstrate that the minimum-mass solution is a reasonable choice for each RDO problem.

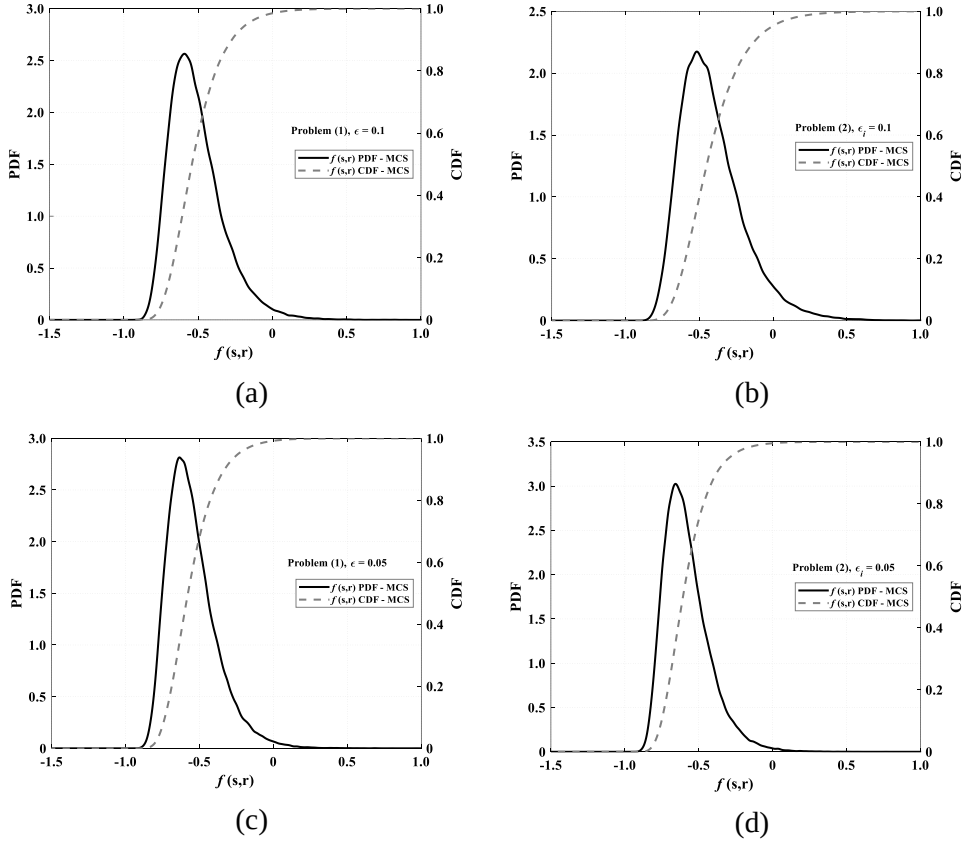
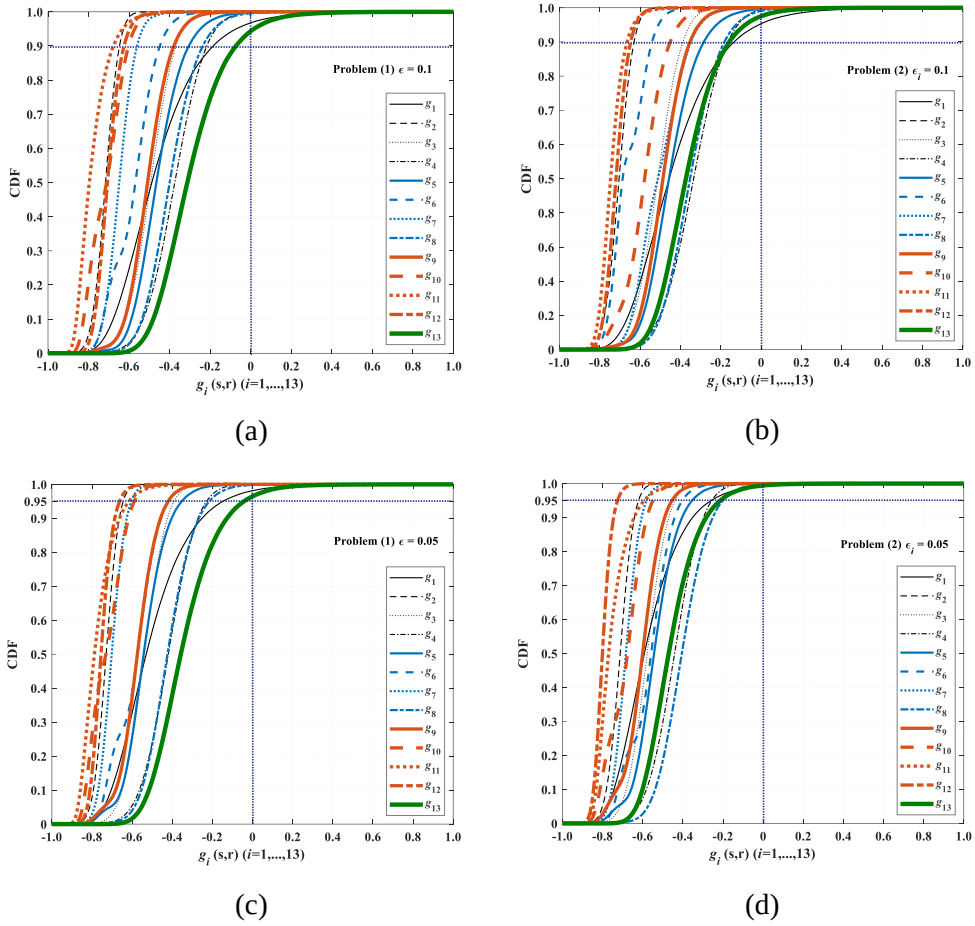


Fig. 13. PDF and CDF of the uncertain objective function $f(\mathbf{s},\mathbf{r})$ at the minimum-mass solutions to different RDO problems of the six-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).



(a) (b)

(c) (d)

Fig. 14. CDFs of 13 probabilistic constraint functions at the minimum-mass solutions to different RDO problems of the six-story two-bay steel frame: (a) problem (1-1); (b) problem (2-1); (c) problem (1-2); (d) problem (2-2).

6. Conclusion

Although the robust design of a steel frame under the effects of uncertain design parameters has been investigated extensively, solving the joint and individual probabilistic constrained RDO problems is still a challenging task. The proposed BO in this work is able to handle this task by addressing challenges arising from solving these RDO problems such as the implicit objective and constraint

functions, discrete nature of the problems formulated for frame structures, and difficulty in evaluation of the probabilistic constraints with a reasonable computational cost. Two new acquisition functions have been developed for the two RDO problems to guide the BO toward better solutions once it completes an iteration. A new optimization strategy is then devised for solving the maximization problem of the acquisition functions that decides where the GP surrogates for the structural responses should be refined. Finally, an optimization procedure integrating the proposed BO is introduced for solving probabilistic constrained multi-objective RDO problems of structures.

Effectiveness of the proposed BO has been demonstrated through a test problem of a two-bar truss and two design examples of two planar steel frames. The optimization results show that the BO can offer exact or good approximate Pareto-optimal solutions to the RDO problems after 20 iterations. Although the shapes of Pareto fronts to the same RDO problem are similar, different BO trials may lead to different Pareto-optimal solutions because the proposed method is random in nature. The effect of this randomness is pronounced when the feasible region of the RDO problems is large. Thus, it is prudent to select a decision list of both column and beam members using design rules of thumb to reduce the discrepancy in the obtained solutions.

The proposed method may be computationally expensive if it is developed based on a large training set to increase the accuracy of the optimization results because the GP surrogates for the structural responses are extremely sensitive to the number of training samples [47]. To scale up the application of the method to a complex problem having a larger training set, the training set may be divided into different independent subsets using a clustering method [42], construct GP models for each subset, and then combine these GP models to form a new mixture probabilistic model for the structural responses. The proposed method may be also extended by first replacing the GP models with deterministic regression models, e.g., polynomial, radial basis function, or support vector regression models, and then developing the acquisition functions using additive Gaussian errors.

This study is limited to the strength requirements for individual structural members and does not consider the local and global collapse mechanisms of the frames. However, the proposed method can

be extended to the same RDO problems with additional probabilistic constraints on these collapse mechanisms in future works. In case the optimization includes both correlated and uncorrelated LSFs of different collapse mechanisms, it is desirable to develop a new optimization method addressing both the joint and individual probabilistic constrained RDO problems in this study.

Acknowledgments

This work is supported by the JICA AUN/SEED-Net project and by JSPS KAKENHI under Grant Number JP19H02286.

Appendix A: Gaussian process

Based on the training dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$, the relationship between the input and output variables can be established using the mapping $y = \hat{g}(\mathbf{x})\colon \mathbb{R}^d \to \mathbb{R}$, where $\hat{g}(\mathbf{x})$ is a probabilistic regression model conditioned on $\mathcal{D}$.

A GP is an infinite set of the output random variables, any finite subset of which has a joint Gaussian distribution [47]. Thus, for the set of the input variables $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, the associated output variables $\{y_1, \dots, y_N\}$ are distributed according to

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N}_N \left(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \right) \tag{A.1}$$

where $\mathcal{N}_N$ denotes an N-variate Gaussian, and $m(\mathbf{x}) = \mathbb{E}[\hat{g}(\mathbf{x})]$ and $k(\mathbf{x}, \mathbf{x}')$ are the mean and covariance kernel functions, respectively. Here, the covariance kernel function defined for any pair of input variable vectors $\mathbf{x}$ and $\mathbf{x}'$ measures the similarity between two corresponding random output values $\hat{g}(\mathbf{x})$ and $\hat{g}(\mathbf{x}')$, such that

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\big[\big(\hat{g}(\mathbf{x}) - m(\mathbf{x})\big)\big(\hat{g}(\mathbf{x}') - m(\mathbf{x}')\big)\big] \tag{A.2}$$

The following Gaussian kernel is used in this study.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^T(\mathbf{x} - \mathbf{x}')}{2l^2}\right] \tag{A.3}$$

where l is the characteristic length-scale parameter determined using the maximum likelihood estimation of $\mathcal{D}$ [47].

One now wishes to use the information from $\mathcal{D}$ for predicting the output value y^* at a new input vector $\mathbf{x}^*$. Let $\mathbf{m}(\mathbf{X}) = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_N)]^T$, the joint PDF of y^* and $\mathbf{y}$ is also a Gaussian because the PDF of any finite subset of the output random variables is a Gaussian, such that

$$\begin{bmatrix} y^* \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}_{N+1}\left(\begin{bmatrix} m(\mathbf{x}^*) \\ \mathbf{m}(\mathbf{X}) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}^*, \mathbf{x}^*) & \mathbf{K}(\mathbf{x}^*, \mathbf{X}) \\ \mathbf{K}(\mathbf{x}^*, \mathbf{X})^T & \mathbf{K}(\mathbf{X}, \mathbf{X}) \end{bmatrix}\right) \tag{A.4}$$

where

$$\mathbf{K}(\mathbf{x}^*, \mathbf{X}) = [k(\mathbf{x}^*, \mathbf{x}_1), \dots, k(\mathbf{x}^*, \mathbf{x}_N)] \tag{A.5}$$

$$\mathbf{K}(\mathbf{X}, \mathbf{X}) = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \tag{A.6}$$

Using the standard conditioning rule for the conditional Gaussian random variable $y^*|\mathbf{y} = \hat{g}(\mathbf{x}^*)$, the associated PDF can be directly derived from Eq. (A.4) as

$$y^*|\mathbf{y} = \hat{g}(\mathbf{x}^*) \sim \mathcal{N}\left(\mu_{\hat{g}}(\mathbf{x}^*), \sigma_{\hat{g}}^2(\mathbf{x}^*)\right) \tag{A.7}$$

where

$$\mu_{\hat{g}}(\mathbf{x}^*) = m(\mathbf{x}^*) + \mathbf{K}(\mathbf{x}^*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\left(\mathbf{y} - \mathbf{m}(\mathbf{X})\right) \tag{A.8}$$

$$\sigma_{\hat{g}}^2(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\mathbf{x}^*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{x}^*, \mathbf{X})^T \tag{A.9}$$

Appendix B: Saddlepoint approximation

Let $\boldsymbol{\mu}_{\mathbf{x}} = \left[\mu_{x_1}, \dots, \mu_{x_d}\right]^T$ be the mean vector of the random input variables $\mathbf{x} = [\mathbf{s}^T, \mathbf{r}^T]^T = [x_1, \dots, x_d]^T$. Let μ_y, $\mu_{2,y}$, and $\mu_{3,y}$ denote the first three cumulants of the random variable $y = \mu_{\hat{g}}(\mathbf{x})$ in Eq. (A.8). These values are estimated based on the second-order Taylor series expansion with respect to $\mathbf{x}$ at the mean $\boldsymbol{\mu}_{\mathbf{x}}$, as follows [19]:

$$\mu_y \approx \mu_{\hat{g}}(\boldsymbol{\mu}_\mathbf{x}) + \frac{1}{2}\sum_{i=1}^{d} \left.\frac{\partial^2 \mu_{\hat{g}}}{\partial x_i^2}\right|_{\boldsymbol{\mu}_\mathbf{x}} \mu_{2,x_i} \tag{B.1}$$

$$\begin{aligned}\mu_{2,y} \approx & \sum_{i=1}^{d} \left(\left.\frac{\partial \mu_{\hat{g}}}{\partial x_i}\right|_{\boldsymbol{\mu}_\mathbf{x}}\right)^2 \mu_{2,x_i} + \sum_{i=1}^{d} \left.\frac{\partial \mu_{\hat{g}}}{\partial x_i}\right|_{\boldsymbol{\mu}_\mathbf{x}} \left.\frac{\partial^2 \mu_{\hat{g}}}{\partial x_i^2}\right|_{\boldsymbol{\mu}_\mathbf{x}} \mu_{3,x_i} \\ & + \frac{1}{4}\sum_{i=1}^{d} \left(\left.\frac{\partial^2 \mu_{\hat{g}}}{\partial x_i^2}\right|_{\boldsymbol{\mu}_\mathbf{x}}\right)^2 \left(\mu_{4,x_i} - \mu_{2,x_i}^2\right)\end{aligned} \tag{B.2}$$

$$\begin{aligned}\mu_{3,y} \approx & \sum_{i=1}^{d} \left(\left.\frac{\partial \mu_{\hat{g}}}{\partial x_i}\right|_{\boldsymbol{\mu}_\mathbf{x}}\right)^3 \mu_{3,x_i} \\ & + \frac{3}{2}\sum_{i=1}^{d} \left(\left.\frac{\partial \mu_{\hat{g}}}{\partial x_i}\right|_{\boldsymbol{\mu}_\mathbf{x}}\right)^2 \left.\frac{\partial^2 \mu_{\hat{g}}}{\partial x_i^2}\right|_{\boldsymbol{\mu}_\mathbf{x}} \left(\mu_{4,x_i} - \mu_{2,x_i}^2\right) \\ & + \sum_{i=1}^{d} \left.\frac{\partial \mu_{\hat{g}}}{\partial x_i}\right|_{\boldsymbol{\mu}_\mathbf{x}} \left(\left.\frac{\partial^2 \mu_{\hat{g}}}{\partial x_i^2}\right|_{\boldsymbol{\mu}_\mathbf{x}}\right)^2 \left(\frac{3}{4}\mu_{5,x_i} - \frac{3}{2}\mu_{2,x_i}\mu_{3,x_i}\right) \\ & + \sum_{i=1}^{d} \left(\left.\frac{\partial^2 \mu_{\hat{g}}}{\partial x_i^2}\right|_{\boldsymbol{\mu}_\mathbf{x}}\right)^3 \left(\frac{1}{4}\mu_{2,x_i}^3 - \frac{3}{8}\mu_{2,x_i}\mu_{4,x_i} + \frac{1}{8}\mu_{6,x_i}\right)\end{aligned} \tag{B.3}$$

where μ_{k,x_i} $(k = 2, \ldots, 6)$ is the kth central moment of x_i, determined based on a total 2×10^6 sampled values of x_i.

Let K_y, $K_y^{(1)}$, $K_y^{(2)}$, and $K_y^{(3)}$ denote the cumulant generating function of y, and its first, second, and third derivatives, respectively. K_y is taken from Ref. [33], as

$$K_y(\xi) = \left(\mu_y - 2ab\right)\xi + \frac{1}{2}\left(\mu_{2,y} - 2ab^2\right)\xi^2 - a\log[(1 - b\xi)^2] \tag{B.4}$$

where ξ represents a real value, and a and b are unknowns.

The unknowns a and b are determined based on the relationship between K_y and the first three cumulants of y, i.e., $K_y^{(1)}(0) = \mu_y$, $K_y^{(2)}(0) = \mu_{2,y}$, and $K_y^{(3)}(0) = \mu_{3,y}$, and the unique root condition for the saddlepoint equation in Eq. (B.7), such that

$$a = \frac{2\mu_{2,y}^3}{\mu_{3,y}^2}, \quad b = \frac{\mu_{3,y}}{2\mu_{2,y}} \tag{B.5}$$

If $\mu_{3,y} = 0$, then $a = 0$ and $b = 0$. The second cumulant $\mu_{2,y}$ is always positive.

In the saddlepoint approximation context, the PDF of the random variable $y = \mu_{\hat{g}}(\mathbf{x})$, denoted as $p_y(y)$, can be estimated by [48]

$$p_y(y) \approx \frac{\exp\left[K_y(\xi_s) - \xi_s y\right]}{\sqrt{2\pi K_y^{(2)}(\xi_s)}} \tag{B.6}$$

where ξ_s is called the saddlepoint that is the unique root to the following saddlepoint equation

$$K_y^{(1)}(\xi) = y \tag{B.7}$$

Also, the probability that $\mu_{\hat{g}}(\mathbf{x}) \leq y$ is evaluated by [48]

$$\mathbb{P}\left[\mu_{\hat{g}}(\mathbf{x}) \leq y\right] \approx \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{v}\right) \tag{B.8}$$

where Φ and ϕ are the standard normal CDF and PDF, respectively; r and v are given as

$$r = \text{sign}(\xi_s)\sqrt{2\left[\xi_s y - K_y(\xi_s)\right]} \tag{B.9}$$

$$v = \xi_s\sqrt{K_y^{(2)}(\xi_s)} \tag{B.10}$$

where $\text{sign}(\xi_s)$= 1,–1, or 0 corresponding to $\xi_s > 0$, $\xi_s < 0$, or $\xi_s = 0$, respectively.

When either r or v does not exist, $\mathbb{P}\left[\mu_{\hat{g}}(\mathbf{x}) \leq y\right] = 0$ if $y \leq \mu_y$ and $\mathbb{P}\left[\mu_{\hat{g}}(\mathbf{x}) \leq y\right] = 1$ if $y > \mu_y$.

References

[1] Roy CJ, Oberkampf WL. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Comput Methods Appl Mech Eng 2011;200:2131–44. https://doi.org/10.1016/j.cma.2011.03.016.

[2] Melchers RE, Beck AT. Structural reliability analysis and prediction. 3rd ed. New York: John Wiley & Sons; 2018. https://doi.org/10.1002/9781119266105.

[3] Schuëller GI, Jensen HA. Computational methods in optimization considering uncertainties – An overview. Comput Methods Appl Mech Eng 2008;198:2–13. https://doi.org/10.1016/j.cma.2008.05.004.

[4] Ben-Tal A, El Ghaoui L, Nemirovski A. Robust optimization. Princeton University Press; 2009. https://doi.org/10.1515/9781400831050.

[5] Elishakoff I, Ohsaki M. Optimization and anti-optimization of structures under uncertainty. Imperial College Press; 2010. https://doi.org/10.1142/p678.

[6] Bertsimas D, Brown DB, Caramanis C. Theory and applications of robust optimization. SIAM Rev 2011;53:464–501. https://doi.org/10.1137/080734510.

[7] Kanno Y. On three concepts in robust design optimization: absolute robustness, relative robustness, and less variance. Struct Multidiscip Optim 2020;62:979–1000. https://doi.org/10.1007/s00158-020-02503-9.

[8] Doltsinis I, Kang Z. Robust design of structures using optimization methods. Comput Methods Appl Mech Eng 2004;193:2221–37. https://doi.org/10.1016/j.cma.2003.12.055.

[9] Beyer H-G, Sendhoff B. Robust optimization – A comprehensive survey. Comput Methods Appl Mech Eng 2007;196:3190–218. https://doi.org/10.1016/j.cma.2007.03.003.

[10] Du X, Chen W. Towards a better understanding of modeling feasibility robustness in engineering design. J Mech Des 2000;122:385–94. https://doi.org/10.1115/1.1290247.

[11] Papadrakakis M, Lagaros ND, Plevris V. Design optimization of steel structures considering uncertainties. Eng Struct 2005;27:1408–18. https://doi.org/10.1016/j.engstruct.2005.04.002.

[12] Lee I, Choi KK, Du L, Gorsich D. Dimension reduction method for reliability-based robust design optimization. Comput Struct 2008;86:1550–62. https://doi.org/10.1016/j.compstruc.2007.05.020.

[13] Richardson JN, Filomeno Coelho R, Adriaenssens S. Robust topology optimization of truss structures with random loading and material properties: A multiobjective perspective. Comput Struct 2015;154:41–7. https://doi.org/10.1016/j.compstruc.2015.03.011.

[14] Kriegesmann B, Lüdeker JK. Robust compliance topology optimization using the first-order second-moment method. Struct Multidiscip Optim 2019;60:269–86. https://doi.org/10.1007/s00158-019-02216-8.

[15] Do B, Ohsaki M. Gaussian mixture model for robust design optimization of planar steel frames. Struct Multidiscip Optim 2021;63:137–60. https://doi.org/10.1007/s00158-020-02676-3.

[16] Geng X, Xie L. Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization. Annu Rev Control 2019;47:341–63. https://doi.org/10.1016/j.arcontrol.2019.05.005.

[17] Caflisch RE. Monte Carlo and quasi-Monte Carlo methods. Acta Numer 1998;7:1–49. https://doi.org/10.1017/S0962492900002804.

[18] Crestaux T, Le Maıˆtre O, Martinez J-M. Polynomial chaos expansion for sensitivity analysis. Reliab Eng Syst Saf 2009;94:1161–72. https://doi.org/10.1016/j.ress.2008.10.008.

[19] Anderson T V, Mattson CA. Propagating skewness and kurtosis through engineering models for low-cost, meaningful, nondeterministic design. J Mech Des 2012;134:100911. https://doi.org/10.1115/1.4007389.

[20] O'Hagan A. Bayes–Hermite quadrature. J Stat Plan Inference 1991;29:245–60. https://doi.org/10.1016/0378-3758(91)90002-V.

[21] Lopez JA, Ponnambalam K, Quintana VH. Generation and transmission expansion under risk using stochastic programming. IEEE Trans Power Syst 2007;22:1369–78. https://doi.org/10.1109/TPWRS.2007.901741.

[22] Calafiore GC, Campi MC. The scenario approach to robust control design. IEEE Trans Automat Contr 2006;51:742–53. https://doi.org/10.1109/TAC.2006.875041.

[23] Luedtke J, Ahmed S. A sample approximation approach for optimization with probabilistic constraints. SIAM J Optim 2008;19:674–99. https://doi.org/10.1137/070702928

[24] Nemirovski A, Shapiro A. Convex approximations of chance constrained programs. SIAM J Optim 2006;17:969–96. https://doi.org/10.1137/050622328.

[25] Shahriari B, Swersky K, Wang Z, Adams RP, Freitas N de. Taking the human out of the loop: A review of Bayesian optimization. Proc IEEE 2016;104:148–75. https://doi.org/10.1109/JPROC.2015.2494218.

[26] Feliot P, Bect J, Vazquez E. A Bayesian approach to constrained single- and multi-objective optimization. J Glob Optim 2017;67:97–133. https://doi.org/10.1007/s10898-016-0427-3.

[27] Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. J Glob Optim 1998;13:455–92. https://doi.org/10.1023/A:1008306431147.

[28] Frazier PI. A tutorial on Bayesian optimization. arXiv preprint 2018. https://arxiv.org/abs/1807.02811.

[29] Baptista R, Poloczek M. Bayesian optimization of combinatorial structures. Proceedings of the 35th international conference on machine learning, Stockholmsmässan, Stockholm Sweden: PMLR vol. 80; 2018, p. 462–71. http://proceedings.mlr.press/v80/baptista18a.html.

[30] Zhang Y, Apley DW, Chen W. Bayesian optimization for materials design with mixed quantitative and qualitative variables. Sci Rep 2020;10:4924. https://doi.org/10.1038/s41598-020-60652-9.

[31] Mathern A, Steinholtz OS, Sjöberg A, Önnheim M, Ek K, Rempling R, et al. Multi-objective constrained Bayesian optimization for structural design. Struct Multidiscip Optim 2020. https://doi.org/10.1007/s00158-020-02720-2.

[32] Shu L, Jiang P, Shao X, Wang Y. A new multi-objective Bayesian optimization formulation with the acquisition function for convergence and diversity. J Mech Des 2020;142. https://doi.org/10.1115/1.4046508.

[33] Do B, Ohsaki M, Yamakawa M. Sequential mixture of Gaussian processes and saddlepoint approximation for reliability-based design optimization of structures. Struct Multidiscip Optim 2021. https://doi.org/10.1007/s00158-021-02855-w.

[34] Afzal A, Kim K-Y, Seo J. Effects of Latin hypercube sampling on surrogate modeling and optimization. Int J Fluid Mach Syst 2017;10:240–53. https://doi.org/10.5293/IJFMS.2017.10.3.240.

[35] Couckuyt I, Deschrijver D, Dhaene T. Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. J Glob Optim 2014;60:575–94. https://doi.org/10.1007/s10898-013-0118-2.

[36] Cao Y. Hypervolume indicator. MATLAB Central File Exchange 2008. https://www.mathworks.com/matlabcentral/fileexchange/19651-hypervolume-indicator (accessed October 10, 2020).

[37] Emmerich MTM, Giannakoglou KC, Naujoks B. Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. IEEE Trans Evol Comput 2006;10:421–39. https://doi.org/10.1109/TEVC.2005.859463.

[38] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 2002;6:182–97. https://doi.org/10.1109/4235.996017.

[39] Belotti P, Kirches C, Leyffer S, Linderoth J, Luedtke J, Mahajan A. Mixed-integer nonlinear optimization. Acta Numer 2013;22:1–131. https://doi.org/10.1017/S0962492913000032.

[40] Do B, Ohsaki M. A random search for discrete robust design optimization of linear-elastic steel frames under interval parametric uncertainty. Comput Struct 2021;249:106506. https://doi.org/10.1016/j.compstruc.2021.106506.

[41] Ohsaki M. Optimization of finite dimensional structures. 1st ed. Boca Raton: CRC Press; 2010. https://doi.org/10.1201/EBK1439820032.

[42] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media; 2009. https://doi.org/10.1007/978-0-387-84858-7.

[43] Lophaven SN, Nielsen HB, Søndergaard J. DACE-A Matlab Kriging toolbox, version 2.0. Lyngby, Denmark: Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark; 2002.

[44] ANSI/AISC 360-16. Specification for structural steel buildings. Chicago, Illinois, USA; 2016.

[45] ASCE/SEI 7-16. Minimum design loads and associated criteria for buildings and other structures. Reston, Virginia, USA; 2017. https://doi.org/10.1061/9780784414248.

[46] Hess PE, Bruchman D, Assakkaf IA, Ayyub BM. Uncertainties in material and geometric strength and load variables. Nav Eng J 2002;114:139–66. https://doi.org/10.1111/j.1559-3584.2002.tb00128.x.

[47] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Cambridge,

Massachusetts: The MIT Press; 2006. https://doi.org/10.7551/mitpress/3206.001.0001.

[48] Butler RW. Saddlepoint approximations with applications. Cambridge: Cambridge University Press; 2007. https://doi.org/10.1017/CBO9780511619083.