Probabilistic Approaches to Optimization of Steel Structures Considering Uncertainty

Do Kim Bach

2023

Contents

List of Tables

List of Figures

Acronyms

ABC	Approximate Bayesian computation
BIC	Bayesian information criterion
BO	Bayesian optimization
CDF	Cumulative distribution function
EI	Expected improvement
EM	Expectation maximization
FE	Finite element
FORM	First-order reliability method
GA	Genetic algorithm
GMM	Gaussian mixture model
GP	Gaussian process
HV	Hypervolume
HVI	Hypervolume indicator
LCB	Lower confidence bound
LSF	Limit-state function
MAP	Maximum a posteriori
MCMC	Markov chain Monte-Carlo simulation
$\mathbf{MCE_R}$	Risk-targeted maximum considered earthquake
MGP	Mixture of Gaussian processes
MCS	Monte-Carlo simulation
MINLP	Mixed-integer nonlinear programming
MOBO	Multi-objective Bayesian optimization
NRHA	Nonlinear response history analysis

NSGA-II Nondominated sorting genetic algorithm II

PDF Probability density function

PEERC The Pacific Earthquake Engineering Research Center

PI Probability of improvement

PSO Particle swarm optimization

RBDO Reliability-based design optimization

RDO Robust design optimization

SA Simulated annealing

SAA Saddlepoint approximation

SDO Sequential deterministic optimization

SOBO Single-objective Bayesian optimization

SQP Sequential quadratic programming

Chapter 1

Introduction and background

1.1 Overview of structural optimization

Structural design naturally involves optimization at its core. From the perspective of structural engineers, the design objective is to find a structure that minimizes the structural cost under constraints on structural performance. The structural cost includes costs of materials and other costs for design, construction, operation, maintenance, and demolition, while the performance is related to the structural soundness at each stage of the structure's life cycle. Sometimes the structural performance is optimized within a limited budget from investors or owners. This research focuses on optimizing the structural cost and performance which is a main concern of structural engineers in the preliminary design phase of steel structures. The reader may consult the monograph by Ohsaki [1] for various technical aspects of design optimization of trusses, building frames, and long-span structures.

The design optimization process facilitating the search for a cost-effective structure without compromising structural performance might be challenging in view of the traditional design process. Although the two processes share the incremental refinement of design to improve the structural performance, as shown in Fig. 1.1, they differ in important features. The optimization design process starts by specifying an objective function for the design, which is cast in the role of the designer, and then systematically updates the design using logical optimization concepts and procedures. In contrast, the traditional design process bypasses the identification of the objective function before updating the design through designer's experience. This not only leads the optimization design process to be more formal than the traditional design process but also can reduce the chance of human error in design [2].

A basic structural design optimization problem is stated as follows:

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & f(\mathbf{s}) \\ \text{subject to} \quad & g_i(\mathbf{s}, \mathbf{r}_0) \leq 0 \quad (i = 1, \ldots, I), \\ & \mathbf{s} \in [\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}], \end{aligned} \tag{1.1}$$

where $\mathbf{s}$ and $\mathbf{r}_0$ are the vectors of design variables and nominal design parameters, respectively; $f(\mathbf{s})$ is the objective function that represents the structural cost; $g_i(\mathbf{s}, \mathbf{r}_0)$ are inequality constraint functions of mechanical properties of the structure; and $\mathbf{s}_\mathrm{l}$ and $\mathbf{s}_\mathrm{u}$ are the pre-specified lower and

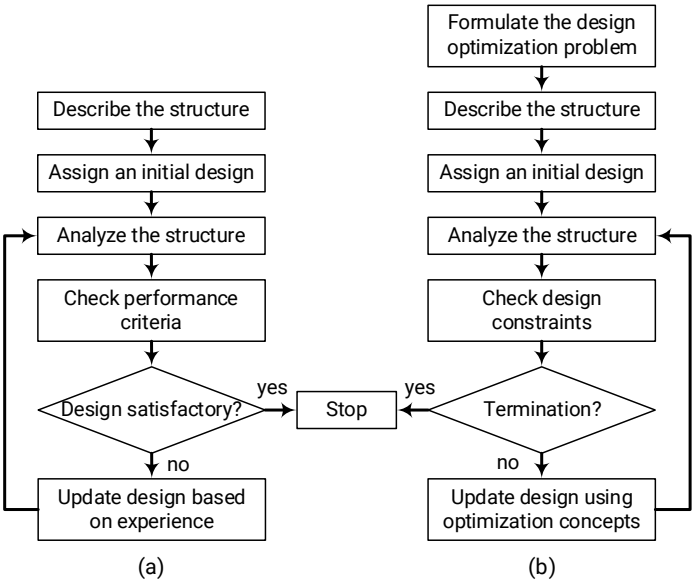

Figure 1.1: Comparison of (a) traditional design process and (b) optimization design process [2].

upper bounds of $\mathbf{s}$, respectively. Although equality constraints are involved in some problems, they are omitted for brevity because we can accommodate each equality constraint by decomposing it into two inequality constraints. Thus, an optimization problem for a steel frame structure may read: minimize the total mass $f(\mathbf{s})$ of the frame subject to design constraints on the stresses $g_1(\mathbf{s}, \mathbf{r}_0)$ in the structural components, long-term displacements $g_2(\mathbf{s}, \mathbf{r}_0)$ of beams, and inter-story drifts $g_3(\mathbf{s}, \mathbf{r}_0)$, where $\mathbf{s}$ consists of the steel sections of beams and columns, and $\mathbf{r}_0$ represents the design loads and mechanical properties of steel materials such as Young's modulus, Poisson's ratio, and yield stress values.

Multi-objective optimization problems are of interest if several objective functions are optimized simultaneously. This arises in structural design because the optimization of a structure is often a trade-off between the structural cost and structural performance, and it is difficult for designers to prioritize these objectives. For example, a bi-objective optimization problem is formulated when the designers try to minimize both the total mass and vertical displacement at the tip of a steel cantilever beam. The strength-ductility trade-off in structural steels is another classical example. In essence, multi-objective optimization deals with a vector of different objective functions for finding a set of design vectors called Pareto-optimal solutions. Such solutions represent the compromise between the conflicting objective functions, and the best solution for the designers is called the most preferred solution or the best compromise solution. The reader may consult the monograph by Ohsaki [1] for a detailed discussion of solutions to structural multi-objective optimization problems.

Discrete optimization problems arise when the design variables take on discrete values. We often encounter this class of problems in structural design, for example, when selecting the sections for the members of a steel structure from a pre-assigned list or catalog of available steel sections, such as the AISC steel shapes database of American wide-flange beams [3]. Different categories of solution techniques for discrete optimization problems are discussed in the monograph by Arora [2].

Structural optimization approaches fall into optimality criteria methods and search methods, each class consists of a broad variety of different optimization algorithms. The class of optimal-

ity criteria methods finds solutions based on optimality criteria that must be satisfied when the objective function is at its minimum under the constraints. The class of search methods numerically finds solutions from an initial design using different approaches for exploration of the design variable space to minimize the objective function in the feasible region. By further transforming a constrained problem into an unconstrained problem through the use of penalty functions, the optimality criteria methods also permit the implementation of search methods that have been developed for unconstrained optimization problems.

Selecting a proper algorithm depends on the problem of interest and is subject to the no free lunch theorem [4]. Accordingly, if an algorithm works better than another algorithm on a class of problems, it may work worse on another class of problems [5]. The following briefly outlines common search methods and their applicability to solving structural optimization problems.

- **Gradient-based methods** rely on the first- or second-order local models constructed from the gradient and/or Hessian of the objective function to incrementally improve a solution until some termination criterion is satisfied. The class of gradient-based methods involves many different optimization algorithms developed based on how the descent direction and the step size are determined using the local model. The methods are suitable for continuous optimization problems where the gradient and/or Hessian of the objective and constraint functions are easy to evaluate. Common descent direction algorithms include gradient descent [6], quasi-Newton [7], and adaptive moment estimation [8] methods for unconstrained problems, and sequential quadratic programming [9] and interior-point methods [10] for constrained problems.

- **Stochastic search methods** rely on random numbers generated during the optimization process by a pseudo-random number generator to help explore the space of design variables. Such an exploration is desirable because randomness can increase the chance of finding a global solution. It is also capable of exploring the design space effectively by assigning a large weight, based on information about past solutions, to regions that are more likely to contain good solutions [5]. However, a large amount of randomness may be ineffective because it can prevent the use of information about the past solutions from guiding the search [5]. The stochastic search methods can also improve the performance of gradient-based method as well as locate approximate global minima for both continuous and discrete mathematical programming problems if a large amount of time can be expended [2]. Stochastic gradient descent [11], simulated annealing (SA) [12], and cross-entropy [13] algorithms are common stochastic search methods.

- **Population-based methods** rely on a population of candidate solutions to explore the space of design variables. These candidate solutions are further diversified after each iteration by a stochastic process. By using the exploration ability of a large number of candidate solutions, the methods are capable of reducing the chance of premature termination or being trapped in a local minimum. Because their search mechanism is only governed by the objective function values, the methods are suitable for both continuous and discrete optimization problems where calculations of the objective and constraint functions do not arise a major computational cost. Genetic algorithm (GA) [14] and particle swarm optimization (PSO) algorithm [15] belong to the family of population-based methods.

Table 1.1: Probabilistic characteristic of geometric properties of wide-flange beam sections [20–22].

Parameter	Nominal value	Statistical property		
		Mean/Nominal	COV	Distribution
Overall depth	d	1.00	0.03	Normal
Flange width	b_{f}	1.00	0.02	Normal
Web thickness	t_{w}	1.00	0.05	Normal
Flange thickness	t_{f}	1.00	0.05	Normal

Table 1.2: Probabilistic characteristics of material and load random parameters [20–22].

Parameter	Nominal value	Statistical property		
		Mean/Nominal	COV	Distribution
Young's modulus	E	1.00	0.04	Normal
Yield stress	$\sigma_{\mathrm{y},0}$	1.10	0.06	Normal
Tensile strength	σ_{u}	1.07	0.08	Normal
Dead load	D_{L}	1.05	0.10	Normal
Live load	L	1.00	0.10	Extreme-value type I
Wind load	W_{L}	0.92	0.37	Extreme-value type I

1.2 Uncertainty in structural engineering

Uncertainty in the field of structural engineering is characterized as either aleatory or epistemic [16, 17] according to how the uncertainty information is described. The former, which is the focus of this study, refers to inherently random variations in material properties and external loads. The information about aleatory uncertainty is encapsulated in a probability density function (PDF), or equivalently, in a cumulative distribution function (CDF). The latter described by uncertainty sets (e.g., interval, ellipsoidal, or polyhedral sets) takes into account incomplete knowledge of designers about simplified models used for structural analysis and design as well as unpredictable factors emerging during the design, construction, and operation processes. A comprehensive treatment for epistemic uncertainty can be found in the monographs by Ben-Tal et al. [18] and by Elishakoff and Ohsaki [19].

The PDFs or CDFs of random resistance and load parameters of a steel structure arise from the data measured from either onsite or laboratory experiments, which are often corrupted by observational noise. More specifically, the measured data of each parameter is mathematically modeled by a best-fit parametric PDF (or CDF) whose underlying parameters are inferred using a statistics tool [20]. By doing so, we are able to identify the probabilistic characteristics of the geometric properties of wide-flange beam sections as well as the randomness of material and load parameters, which are, for example, summarized in Tables 1.1 and 1.2, respectively, where COV is the coefficient of variation. The main properties of common PDFs defer until Chapter 2.

One of the most important tasks of characterizing the PDFs of random parameters in analysis and design of steel structures is the inference of material parameters for structural steels under the effect of noise-corrupted experimental results. This task, however, still limits itself to well-known parameters and is often independent of finite element (FE) models empowered to analyze the structures, even though the uncertainty quantification of structural responses is mostly done through these models. In fact, it is easy to calibrate some material parameters such as Young's modulus,

yield stress, and tensile strength from the experimental results of monotonic tension tests. Yet it is difficult to directly infer some hardening parameters for an FE model that simulates the behavior of structural steels subjected to cyclic loading conditions. This poses a difficulty in quantifying uncertainty in the cyclic response of a steel structure. This difficulty is more emphasized by the substantial computational cost of repeatedly carrying out many nonlinear FE analyses. In this circumstance, the question is, therefore, how to systematically identify and quantify uncertainty in all material parameters of the FE model that is intended for structural analysis and design processes of a steel structure while reducing the number of costly FE simulations as much as possible. This identification may also be desirable for the analysis and design of reinforced concrete or timber structures, but these structures are beyond the scope of this study.

1.3 Why structural optimization considering uncertainty matters?

Although problem (1.1) has been successfully solved to reduce the structural cost and improve the structural performance systematically, its solution, in the presence of uncertainty, may correspond to the structural behavior that is far from the desired performance. This arises from the use of a unique nominal vector $\mathbf{r}_0$ derived from partial factors to account for the randomness of mechanical resistances of structural members and of external loads [23–25]. Since the partial factors have been calibrated using some oversimplified PDFs of random parameters, they may involve uncertainty in structural safety measure. Consequently, the use of $\mathbf{r}_0$ as a shield against uncertainty may lead to a poor optimal solution that often lies on the boundary of the feasible region, and accordingly, is sensitive to even a small fluctuation in the design parameters. The optimal solution obtained from the deterministic approach also fails to address the cost issue of a specific structure because it only manages the trade-off between safety and cost aspects in a general context for similar structures used to derive the partial factors rather than a particular structure of interest [26].

The variance of safety measures may also lead to uncertainty in optimal solutions obtained from deterministic optimization procedures. This arises because there exist different ways a safety margin can be established from structural resistances and external loads [27]. In other words, the nominal vector $\mathbf{r}_0$ in problem (1.1) may not be unique, leading to several "optimal" designs that defy our expectation that only one optimal design should be found for a specified safety level.

1.4 How structural optimization incorporates uncertainty?

The issues listed in Section 1.3 motivate the quest for a unified approach that integrates probabilistic safety measures and optimization algorithms for handling the structural safety and structural cost simultaneously. Such a unified approach not only explicitly considers the effect of random parameters but also should provide a useful way to reduce the resulting computational cost, as a by-product, of uncertainty consideration. Although this unified approach is desirable for designs of any engineering structures [27–29], this study only focuses on the design of steel structures.

The literature provides two main methods for the unified approach, namely robust design optimization (RDO) [18, 19, 30–33] and reliability-based design optimization (RBDO) [34–38]. These

methods differ in how unfavorable effects of uncertainty can be managed. RDO commonly evaluates the sensitivity of uncertain structural responses and uses this information for finding a design that is less sensitive at an acceptable cost, while RBDO seeks a minimum-cost design that, under uncertainty, guarantees an allowable probability of occurring structural failures while ignoring the expected cost due to these failures. Nevertheless, the two methods encounter the same difficulty in finding a good compromise between processing optimization and propagating uncertainty. Before developing a series of approaches to overcoming this difficulty in Chapters 5-8, we briefly describe the optimization formulation of each method in the following.

Let $\mathbf{r}$ denote the vector of random parameters (i.e., mechanical properties of structural steels and external loads) whose uncertainty information is described by the PDF or CDF of each parameter. Let $f_1(\mathbf{s})$ and $f(\mathbf{s},\mathbf{r})$, respectively, denote the structural cost, which is assumed to be independent of $\mathbf{r}$, and an objective function representing the structural performance, for example, the maximum inter-story drift of a steel frame structure. Uncertainty in $\mathbf{r}$ leads to uncertainty in $f(\mathbf{s},\mathbf{r})$, and the mean and variance of $f(\mathbf{s},\mathbf{r})$ can be evaluated and denoted as $f_2(\mathbf{s}) = \mathbb{E}\big[f(\mathbf{s},\mathbf{r})\big]$ and $f_3(\mathbf{s}) = \mathbb{V}\big[f(\mathbf{s},\mathbf{r})\big]$, respectively. Let $g_i(\mathbf{s},\mathbf{r})$ represent a limit-state function (LSF) of a design requirement on the serviceability or strength of the structure so that $g_i(\mathbf{s},\mathbf{r}) > 0$ indicates a failure mode, and $h_j(\mathbf{s})$ denote a deterministic constraint function such as a requirement on the geometry of structural components. $g_i(\mathbf{s},\mathbf{r})$ is commonly defined as the difference of the load effect and the resistance of the structure. Thus, $g_i(\mathbf{s},\mathbf{r})$ takes random values due to the randomness of $\mathbf{s}$ and $\mathbf{r}$. Since the reduction of structural cost often leads to an increase in the inter-story drift, we formulate the following RDO problem as a multi-objective optimization problem for finding a minimum-cost design at which the corresponding structural performance is less sensitive to uncertainty:

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})] \\ \text{subject to} \quad & \mathbb{E}[g_i(\mathbf{s},\mathbf{r})] \le 0 \quad (i = 1,\dots,I), \\ & h_j(\mathbf{s}) \le 0 \quad (j = 1,\dots,J), \\ & \mathbf{s} \in [\mathbf{s}_\text{l}, \mathbf{s}_\text{u}]. \end{aligned} \tag{1.2}$$

Although problem (1.2) is a deterministic optimization problem, it is able to incorporate all realizations of $g_i(\mathbf{s},\mathbf{r})$ into $\mathbb{E}[g_i(\mathbf{s},\mathbf{r})]$ when $\mathbf{r}$ varies on the space of random parameters rather than being fixed at $\mathbf{r}_0$. The problem coincides with the less-variation concept [39] of RDO that simultaneously minimizes the mean and variance of an objective function under set-based uncertainty.

The RBDO problem incorporating probabilistic safety measures into structural optimization reads

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & f_1(\mathbf{s}) \\ \text{subject to} \quad & \mathbb{P}[g_i(\mathbf{s},\mathbf{r}) \le 0] \ge 1 - \epsilon_i \quad (i = 1,\dots,I), \\ & h_j(\mathbf{s}) \le 0 \quad (j = 1,\dots,J), \\ & \mathbf{s} \in [\mathbf{s}_\text{l}, \mathbf{s}_\text{u}], \end{aligned} \tag{1.3}$$

where $\mathbb{P}[\cdot]$ denotes the probability of occurring $[\cdot]$ with respect to all realizations of $\mathbf{r}$, and $\epsilon_i \in (0,1)$ is a prescribed risk level of the i-th probabilistic constraint, for example, $\epsilon_i = 0.1$, 0.05, or 0.01. Problem (1.3) is to minimize the structural cost $f_1(\mathbf{s})$ so that the probability the structure remains intact (i.e., probability of safety) under the effect of $\mathbf{r}$ is greater than or equal to $1 - \epsilon_i$, which is

intimately related to the so-called reliability index for the case of normal distribution of the LSF [27].

By further incorporating the individual and joint probabilistic constraints into problem (1.2), other two RDO problems can be formulated as follows:

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})] \\ \text{subject to} \quad & \mathbb{P}[g_i(\mathbf{s}, \mathbf{r}) \leq 0] \geq 1 - \epsilon_i \quad (i = 1, \ldots, I), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & \mathbf{s} \in [\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}]. \end{aligned} \tag{1.4}$$

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})] \\ \text{subject to} \quad & \mathbb{P}[g_i(\mathbf{s}, \mathbf{r}) \leq 0, \, i = 1, \ldots, I] \geq 1 - \epsilon, \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & \mathbf{s} \in [\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}], \end{aligned} \tag{1.5}$$

where $\epsilon \in (0, 1)$ is a prescribed risk level of the joint probabilistic constraint. Although problems (1.4) and (1.5) can be classified as RDO or RBDO problems, they are termed as RDO problems in this dissertation.

Problems (1.4) and (1.5) simultaneously minimize the structural cost and the sensitivity of structural performance while preventing the structure from two different ways it reaches an unsafe state. Problem (1.4) consists of a finite number of statistically independent constraints of random LSFs so that the structure is unsafe if one of the constraints is violated, while problem (1.5) is formulated with statistically dependent constraints where the LSFs correlate with each other and the probability is taken over the entire system of random LSFs. The two problems also set the basis for further application to a general case where all possible failure modes for a structure can be anticipated.

1.5 Solution approaches to structural optimization considering uncertainty

1.5.1 RDO problem

The most important task in solving problem (1.2) is to propagate uncertainty in random parameters $\mathbf{r}$ for estimations of the mean and variance of the performance function $f(\mathbf{s}, \mathbf{r})$ and the mean of the LSF $g_i(\mathbf{s}, \mathbf{r})$. A simple but powerful approach employs Monte-Carlo simulation (MCS) [40] detailed in Appendix B through an FE model of the structure, thereby providing sample mean and sample variance associated with a particular design point. Although this approach can lead to moderately conservative designs, it may be computationally expensive especially when the RDO problem has many design variables and random parameters. Uncertainty can also be propagated through a surrogate-based approach that attempts to establish the cause-effect relationship between the random parameters and the performance (or LSF) function. After constructing the surrogate model for the performance function, the associated statistical moments can be estimated using either MCS or Taylor series expansion [41] about the mean or the most-probable point of random parameters. A shortcoming of this approach is that a high-fidelity surrogate model itself involves interpolation

uncertainty [42] that becomes considerable as the number of random parameters increases. Another approach relies on the polynomial chaos expansion fitting the performance (or LSF) function through a linear combination of polynomial basis functions [43]. The resulting approximate model is used to derive the statistical moments of interest directly. However, the use of polynomial chaos expansion is subject to the curse of dimensionality that increases the number of multivariate basis functions as a polynomial order of the number of input variables (i.e., the design variables and random parameters).

Studies in the field of structural engineering that focus on solving problem (1.2) have not aimed at statistically understanding the cause-effect relationship between the random parameters and the performance function for solving the problem directly. Instead, MCS has been a common choice for uncertainty propagation. Lagaros and Papadrakakis [30] minimized the total mass and standard deviation of the maximum inter-story drift of a steel structure against seismic motions. MCS was embedded in the nondominated sorting genetic algorithm II (NSGA-II) for estimating the standard deviation of the response required for solving the RDO problem. Liu et al. [32] simultaneously minimized the total mass, mean value of seismic demand, and variation of seismic demand for a steel moment resisting frame against seismic loading with a log-normal distribution. However, the evaluation of mean and variation values of seismic demand was not explicitly given.

It is even more difficult to solve problems (1.4) and (1.5) because checking the feasibility of any candidate solution satisfying the probabilistic constraints intensifies the difficulty in uncertainty propagation for estimating the objective functions of feasible solutions. Current approaches to handling the probabilistic constraints are briefly reviewed in the next section.

1.5.2 RBDO problem

Solving problem (1.3) is challenging because its feasible space defined by the probability constraint functions is generally non-convex with respect to the design variables $\mathbf{s}$. This difficulty is further emphasized by the calculation of probabilistic constraint functions that is unfortunately an intrinsically NP-hard problem [44]. This motivates a rich variety of solution approaches to reformulation of the RBDO problem to a computationally tractable problem, thereby resulting in a reliable solution. Depending on how the probability constraints are handled during the optimization process, the literature classifies the RBDO methods into three main approaches, namely double-loop [45, 46], single-loop [47–50], and decoupling approaches [51, 52].

The double-loop approach consists of an outer optimization loop exploring the space of $\mathbf{s}$ and an inner loop handling the probabilistic constraints in the space of $\mathbf{r}$. The inner loop estimates the failure probabilities associated with all design candidates $\mathbf{s}$ from the outer loop for confirmation of their feasibility using a classical reliability technique, such as the first-order reliability method (FORM) [53] or MCS. The double-loop approach unfortunately is endowed with the poor performance of FORM on highly nonlinear LSFs or the curse of dimensionality of MCS.

The single-loop approach alleviates the shortcoming of the double-loop approach by converting all probabilistic constraints into the corresponding approximate deterministic constraints evaluated at a quantile of the random LSF. The quantile is derived according to the threshold failure probabilities and a sensitivity analysis of the LSF $g_i(\mathbf{s}, \mathbf{r})$. Although this technique is deemed to be more advantageous than the double-loop approach, it may suffer from the strong effect of selecting the starting point for the sensitivity analysis on accuracy of the obtained quantile.

The decoupling approach converts the RBDO problem into a sequence of deterministic optimization problems to avoid the computational cost of the double-loop approach through an inverse probability calculation of a representative of the random parameters. This approach is familiar to practicing engineers as its basis is analogous to the limit state design philosophy stated in a form of problem (1.1) [23].

Recently, the state-of-the-art regression models such as polynomial response surface [54], artificial neural network [55], radial basis function [56], support vector machine [57], and Gaussian process (GP) [58] have been developed as surrogates for the probabilistic constraint functions of the RBDO problem, which are further capable of reducing the inherent high computational cost of structural simulations. This research direction emerges from the fact that FE methods empowered to carry out structural analyses are less advantageous to processing optimization as well as probabilistic computations. Once the surrogate models have been developed, the probabilities of the probabilistic constraints can be obtained using any reliability methods of the double-loop, single-loop, or decoupling approach. For recent reviews on surrogate-based structural reliability analysis and RBDO methods, we refer the reader to Saraygord Afshari et al. [59] and Moustapha and Sudret [60], respectively.

As seen from the above comments, the decoupling approach is advantageous to reduction of the computational cost, and it can be improved if the inverse probability calculation becomes more reliable. The surrogate-based method is promising and likely the most suitable method for solving RBDO problems of large-scale structures. The method, however, limits itself to relatively small training sets that is why most of the previous studies have used adaptive surrogate models starting from a small number of training samples. If the training set becomes larger for well covering the input variable space, the construction of accurate surrogate models over a large range of high-dimensional input variables becomes a challenge.

In the field of mathematical optimization, stochastic programming also finds itself challenging in views of solving a general form of problem (1.3). Existing approaches, including deterministic model [61], scenario [62], sample average approximation [63], and convex approximation [64], are only capable of finding approximate solutions to several special forms of problem (1.3). For example, the deterministic model approach is well suited for solving problems with linear objective and constraint functions, and normal 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 LSFs $g_i(\mathbf{s}, \mathbf{r})$ 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 aforementioned approaches in solving multi-objective problems (e.g., problems (1.4) and (1.5)) remains unknown as they have been developed for single-objective problems only.

1.6 Objectives and contributions

This study aims at developing a variety of useful probabilistic approaches to handling challenging optimization problems of steel structures considering aleatory uncertainty in design parameters such as mechanical properties of structural steels and external loads. The detail of each approach

is presented, followed by the verification of its performance against benchmark problems and its application to solving optimization problems of interest. In particular, we address the following optimization problems:

(1) Single- and multi-objective identification problems of cyclic elastoplastic parameters for structural steels from noise-free and noise-corrupted experimental results. These problems are extensions of problem (1.1).

(2) Discrete multi-objective RDO problem of steel frames, i.e., problem (1.2).

(3) Discrete multi-objective RDO problem of steel frames with individual and joint probability constraints, i.e., problems (1.4) and (1.5).

(4) Discrete and continuous single-objective RBDO problems of steel trusses and steel frames, i.e., problem (1.3).

(5) Discrete multi-objective RBDO problem of seismic-resistant steel frames, which is an extension of problem (1.3).

The main contributions of this study to the structural engineering and optimization communities are twofold:

(1) Effective probabilistic approaches to deterministically and probabilistically identifying material parameters for a costly FE model that simulates the cyclic elastoplastic behavior of structural steels. These approaches, in general, can be extended to identification of parameters for any costly FE models. In addition, a multi-objective identification method is proposed for reducing the chance of finding biased material parameters from different sets of experimental results.

(2) Useful optimization algorithms and procedures based upon probabilistic approaches that help reduce enormous computational costs arising from difficult tasks in optimizing steel structures in the presence of uncertainty, namely uncertainty propagation, handling individual and joint probabilistic constraints, dealing with discrete design variables, and handling multiple objective functions. As mentioned above, there is no preference of using one algorithm over another to solve different optimization problems unless assumptions are made for the probability distribution of objective functions, each algorithm proposed in this study is necessary for the problem of interest.

1.7 Dissertation outline

The remaining chapters of this dissertation and the conceptual dependencies between them are outlined in Fig. 1.2.

Chapter 2 presents a brief review of probability and probabilistic approaches. It covers common continuous PDFs and basic rules of probability as well as the foundations of GP modeling, Bayesian optimization (BO), Gaussian mixture model (GMM), and saddlepoint approximation (SAA) that are main ingredients for the probabilistic approaches presented in later chapters.

Chapters 3 and **4**, respectively, deal with deterministic and probabilistic identification problems of material parameters for a costly FE model that is used to simulate the cyclic elastoplastic behavior of structural steels. In essence, the identification results obtained from Chapter 4 can provide

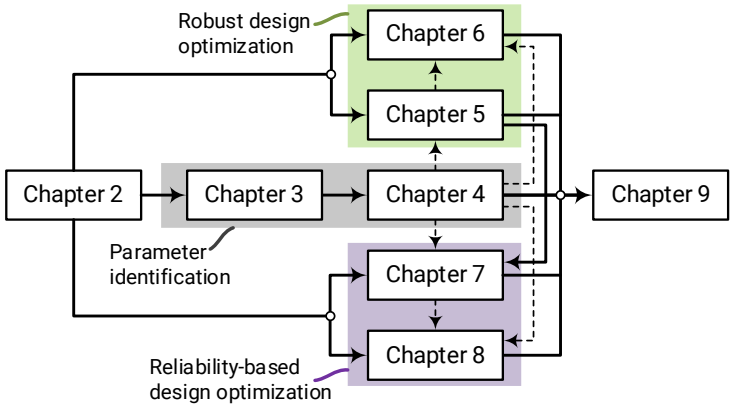

Figure 1.2: Dependencies for the remaining chapters of this study. Dashed arrows indicate weak dependencies.

useful uncertainty information about randomness of material parameters used as input to formulate design optimization problems considering uncertainty, for example, those in Chapters 5−8. Related probabilistic approaches to these chapters consist of GP modeling, single- and multi-objective BO approaches, Bayes's rule, approximate Bayesian computation (ABC), and Markov chain Monte-Carlo simulation (MCMC).

Chapters 5 and **6** show how to handle the discrete multi-objective RDO problems with deterministic and probabilistic constraints, respectively. Related probabilistic approaches to these chapters include GMM, Taylor series expansion for uncertainty propagation, GP modeling, multi-objective BO approach, and SAA.

Chapters 7 and **8** introduce solution approaches to solving single- and multi-objective RBDO problems, respectively. The single-objective RBDO problem in Chapter 7 is formulated for both steel trusses and steel frames, while the multi-objective RBDO problem in Chapter 8 is formulated for moment-resisting steel frames subjected to earthquake excitation. The approaches in these chapters are developed based on GP modeling, GMM, SAA, inverse SAA, and/or MCS.

Chapter 9 summarizes this study and provides possible directions for future research.

1.8 Published works included in the dissertation

This dissertation is the collection of several published papers by the author and co-authors. The detailed information of these papers and how they are organized in this dissertation are summarized as follows:

- **Chapter 3**

  Bach Do and Makoto Ohsaki. "Bayesian optimization for inverse identification of cyclic constitutive law of structural steels from cyclic structural tests". *Structures* 38 (2022), pp. 1079–1097.

  Bach Do and Makoto Ohsaki. "Proximal exploration multi-objective Bayesian optimization for inverse identification of cyclic constitutive law of structural steels". *Structural and Multidisciplinary Optimization* 65(7) (2022), p. 199.

- **Chapter 5**
  Bach Do and Makoto Ohsaki. "Gaussian mixture model for robust design optimization of planar steel frames". *Structural and Multidisciplinary Optimization*, 63(1) (2021), pp. 137–160.
- **Chapter 6**
  Bach Do, Makoto Ohsaki and Makoto Yamakawa. "Bayesian optimization for robust design of steel frames with joint and individual probabilistic constraints". *Engineering Structures* 245 (2021), p. 112859.
- **Chapter 7**
  Bach Do, Makoto Ohsaki and Makoto Yamakawa. "Sequential mixture of Gaussian processes and saddlepoint approximation for reliability-based design optimization of structures". *Structural and Multidisciplinary Optimization* 64(2) (2021), pp. 625–648.
- **Chapter 8**
  Bach Do and Makoto Ohsaki. "Sequential sampling approach to energy-based multiobjective design optimization of steel frames with correlated random parameters". *Earthquake Engineering & Structural Dynamics* 51(3) (2022), pp. 588–611.

Chapter 2

Probability and probabilistic approaches: A brief review

In this chapter, we briefly reviews some basic concepts of probability and probabilistic approaches that are needed in the later chapters. We also briefly introduce the accompanying relationship between each concept and the proposed approaches in the later chapters, although we shall fully cite the related concepts when describing the proposed approaches. For the detail of these concepts, the reader may consult textbooks by Bertsekas and Tsitsiklis [65], Rasmussen and Williams [58], Barber [66], Murphy [67], and Butler [68].

2.1 Basic concepts of continuous random parameters

2.1.1 CDF, PDF, and quantile

Consider a continuous random parameter r. The probability that r is less than or equal to any real value $\bar{r}$, denoted by $\mathbb{P}[r \leq \bar{r}]$, is called the CDF of r. Since the probability preserves the order when $\bar{r}$ increases, $\mathbb{P}[r \leq \bar{r}]$ is a monotonically increasing function. Assume that the CDF of r is differentiable, the PDF of r, denoted as $\pi(r)$, is defined by

$$\pi(r)\mathrm{d}r = \mathbb{P}[\bar{r} \leq r \leq \bar{r} + \mathrm{d}r]. \tag{2.1}$$

Here $\pi(r) \geq 0$ and $\int_{-\infty}^{\infty} \pi(r)\,\mathrm{d}r = 1$ because $\mathbb{P}[r \leq \bar{r}]$ is monotonically increasing and $\mathbb{P}[r \leq \infty] = 1$. If $\pi(r)$ is known, the CDF of r is the area under the PDF from $-\infty$ to $\bar{r}$ illustrated in Fig. 2.1, which is mathematically defined as

$$\mathbb{P}[r \leq \bar{r}] = \int_{-\infty}^{\bar{r}} \pi(r)\,\mathrm{d}r. \tag{2.2}$$

Let $0 < \gamma < 1$. The quantile function associated with γ obtained by the PDF $\pi(r)$ reads

$$Q_\pi(\gamma) = \inf\{\bar{r} | \mathbb{P}[r \leq \bar{r}] = \gamma\}. \tag{2.3}$$

The m-th k-tile ($1 \leq m < k$) is defined as a value of $Q_\pi(\gamma)$ corresponding to $\gamma = m/k$. If $k = 2$,

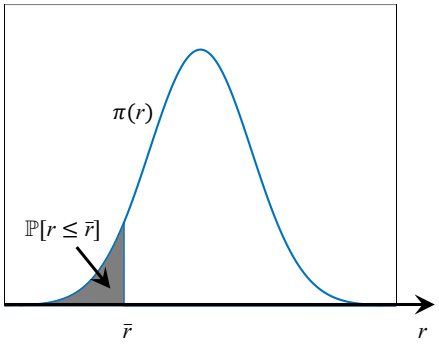

Figure 2.1: Illustration of $\mathbb{P}[r \leq \bar{r}]$.

4, 8, or 100 the quantile quantity is called median, quartile, octile, or percentile, respectively. We will return to the quantile of a dataset in Section 4.4.2 for finding the parameters underlying a distribution generating the dataset. We will also solve RBDO problems through the quantile of a random LSF estimated through an inverse SAA approach in Section 7.3.

2.1.2 Moment, mean, variance, and moment and cumulant generating functions

Let n denote a non-negative integer. The n-th (raw) moment of r is defined as

$$\mu_{r,n} = \int_{-\infty}^{\infty} r^n \pi(r)\, dr. \tag{2.4}$$

The mean (i.e., expected value) of r, denoted as μ_r or $\mathbb{E}[r]$, is the first moment, such that

$$\mu_r = \mu_{r,1} = \int_{-\infty}^{\infty} r\pi(r)\, dr. \tag{2.5}$$

The variance of r, denoted as σ_r^2 or $\mathbb{V}[r]$, measuring "the degree of spread" of its PDF is defined by

$$\sigma_r^2 = \int_{-\infty}^{\infty} (r - \mu_r)^2 \pi(r)\, dr. \tag{2.6}$$

The covariance of two random parameters r_1 and r_1, denoted as $\mathrm{cov}(r_1, r_2)$, reads

$$\mathrm{cov}(r_1, r_2) = \mathbb{E}\left[(r_1 - \mathbb{E}[r_1])\,(r_2 - \mathbb{E}[r_2])\right]. \tag{2.7}$$

Let $h > 0$ denote a real-valued number and ξ denote a variable satisfying $|\xi| < h$. The moment generating function of r, denoted as $M(\xi)$, combines the information of its moments in a single expression. Accordingly, $M(\xi)$ reads

$$M(\xi) = \mathbb{E}[\exp(\xi r)] = \int_{-\infty}^{\infty} \exp(\xi r)\pi(r)\, dr = \mathbb{E}[1 + \xi r + \cdots + \frac{\xi^n r^n}{n!} + \cdots] = \sum_{n=0}^{\infty} \frac{\mu_{r,n}\xi^n}{n!}. \tag{2.8}$$

Thus, the n-th moment $\mu_{r,n}$ is the n-th derivative of $M(\xi)$ evaluated at $\xi = 0$.

The cumulant generating function of r, denoted as $K(\xi)$, is defined by

$$K(\xi) = \log[M(\xi)] = \sum_{n=0}^{\infty} \frac{\kappa_{r,n}\xi^n}{n!}, \tag{2.9}$$

where $\kappa_{r,n} = K^{(n)}(0)$ are cumulants with $\kappa_{r,0} = 0$, which is associated with the zero-th moment $\mu_{r,0} = 1$. According to Eqs. (2.8) and (2.9), the first three cumulants can be expressed in terms of the first three moments, such that

$$\kappa_{r,1} = \mu_{r,1}, \tag{2.10a}$$

$$\kappa_{r,2} = \mu_{r,2} - \mu_{r,1}^2, \tag{2.10b}$$

$$\kappa_{r,3} = \mu_{r,3} - 3\mu_{r,2}\mu_{r,1} + \mu_{r,1}^3. \tag{2.10c}$$

A new truncated function of $K(\xi)$ is proposed in Chapter 7 to facilitate the development of an novel SAA.

2.1.3 Joint probability, conditional probability, and Bayes' rule

Joint probability of two events r_1 and r_2 is the probability that the two events occur at the same time, which is described by the joint PDF $\pi(r_1, r_2)$. Conditional probability is the probability that event r_1 occurs given that event r_2 occurs. When $\pi(r_2) > 0$, the conditional PDF, denoted as $\pi(r_1; r_2)$, is evaluated by

$$\pi(r_1; r_2) = \frac{\pi(r_1, r_2)}{\pi(r_2)}, \tag{2.11}$$

which is the foundation of the GMM approach presented in Section 2.5 and Chapter 5.

Because $\pi(r_1, r_2) = \pi(r_2, r_1)$, the conditional probability in Eq. (2.11) underlies Bayes' rule, which reads

$$\pi(r_1; r_2) = \frac{\pi(r_2; r_1)\pi(r_1)}{\pi(r_2)}. \tag{2.12}$$

This forms the basis of Bayesian inference where $\pi(r_1)$, $\pi(r_2; r_1)$, $\pi(r_2)$, and $\pi(r_1; r_2)$ are termed as prior, likelihood, evidence, and posterior, respectively. By adopting Bayes' rule, the Bayesian inference translates the probabilistic relationship from $\pi(r_2; r_1)$ to $\pi(r_1; r_2)$. This underlies the probabilistic identification approach in Chapter 4.

2.1.4 Correlation and independence

Two random parameters r_1 and r_2 are correlated if their (Pearson) correlation coefficient $\rho(r_1, r_2)$ is not equal to zero, such that

$$\rho(r_1, r_2) = \frac{\mathrm{cov}(r_1, r_2)}{\sigma_{r_1}\sigma_{r_2}} = \frac{\mu_{r_1 r_2} - \mu_{r_1}\mu_{r_2}}{\sigma_{r_1}\sigma_{r_2}} \neq 0, \tag{2.13}$$

where $\mathrm{cov}(r_1, r_2)$ is the covariance between r_1 and r_2 defined in Eq. (2.7); $\mu_{r_1 r_2}$ is the mean of $r_1 r_2$; and σ_{r_1} and σ_{r_2} are the standard deviation values of r_1 and r_2, respectively. The generation of a finite number of samples of correlated random parameters is detailed in Section 8.4.1.

Let $\mathbf{r}$ denote a d_2-dimensional vector of random parameters. Its covariance and correlation

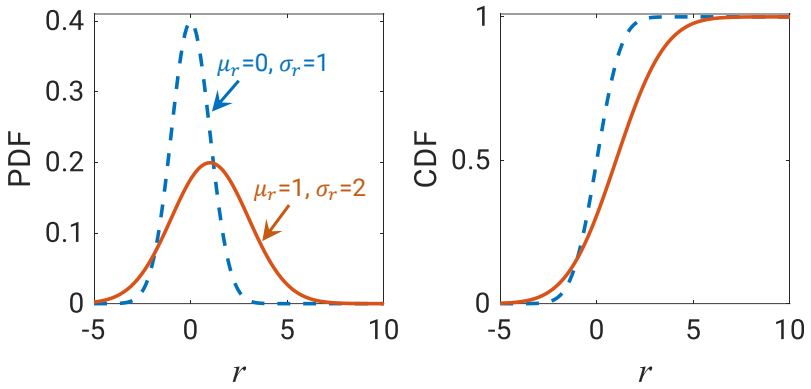

Figure 2.2: Examples of Gaussian PDFs and CDFs.

matrices, respectively, denoted as $\mathbf{\Sigma}_r$ and $\mathbf{C}_r$, are defined as follows:

$$\mathbf{\Sigma}_r = \begin{bmatrix} \sigma_{r_1}^2 & \cdots & \mathrm{cov}(r_1, r_{d_2}) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(r_{d_2}, r_1) & \cdots & \sigma_{r_{d_2}}^2 \end{bmatrix}, \quad \mathbf{C}_r = \begin{bmatrix} \rho(r_1, r_1) & \cdots & \rho(r_1, r_{d_2}) \\ \vdots & \ddots & \vdots \\ \rho(r_{d_2}, r_1) & \cdots & \rho(r_{d_2}, r_{d_2}) \end{bmatrix}. \tag{2.14}$$

Two random parameters r_1 and r_2 are independent if understanding the uncertainty information of r_1 gives no uncertainty information of r_2, and vice versa. This is mathematically expressed by

$$\pi(r_1, r_2) = \pi(r_1)\pi(r_2). \tag{2.15}$$

If r_1 and r_2 are independent, they are also uncorrelated because $\mu_{r_1 r_2} = \mu_{r_1}\mu_{r_2}$, which can be derived from the definition of mean in Eq. (2.5) and the independence condition in Eq. (2.15).

2.2 Common continuous distributions

The probability distributions listed in this section help the reader to understand the probabilistic properties of random parameters given in the subsequent chapters.

2.2.1 Gaussian (normal) distribution

Random parameter r is distributed according to a Gaussian if its PDF has the form of

$$\pi(r; \mu_r, \sigma_r^2) = \mathcal{N}\left(r; \mu_r, \sigma_r^2\right) = \frac{1}{\sigma_r\sqrt{2\pi}} \exp\left[-\frac{(r-\mu_r)^2}{2\sigma_r^2}\right], \tag{2.16}$$

where $\mathcal{N}$ denotes a Gaussian PDF, and mean μ_r and variance σ_r^2 completely characterize the distribution. Figure 2.2 provides examples of PDFs and CDFs for two different univariate Gaussians.

In a general case, the PDF of a d_2-variate Gaussian is defined as

$$\pi(\mathbf{r}; \boldsymbol{\mu}_r, \mathbf{\Sigma}_r) = \mathcal{N}\left(\mathbf{r}; \boldsymbol{\mu}_r, \mathbf{\Sigma}_r\right) = \frac{1}{\det(\mathbf{\Sigma_r})^{1/2}(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_r)^T \mathbf{\Sigma}_r^{-1}(\mathbf{r}-\boldsymbol{\mu}_r)\right], \tag{2.17}$$

where $\mathbf{\Sigma}_r$ is given in Eq. (2.14), and $\det(\mathbf{\Sigma_r})$ denotes the determinant of $\mathbf{\Sigma}_r$, and $(\cdot)^T$ denotes the

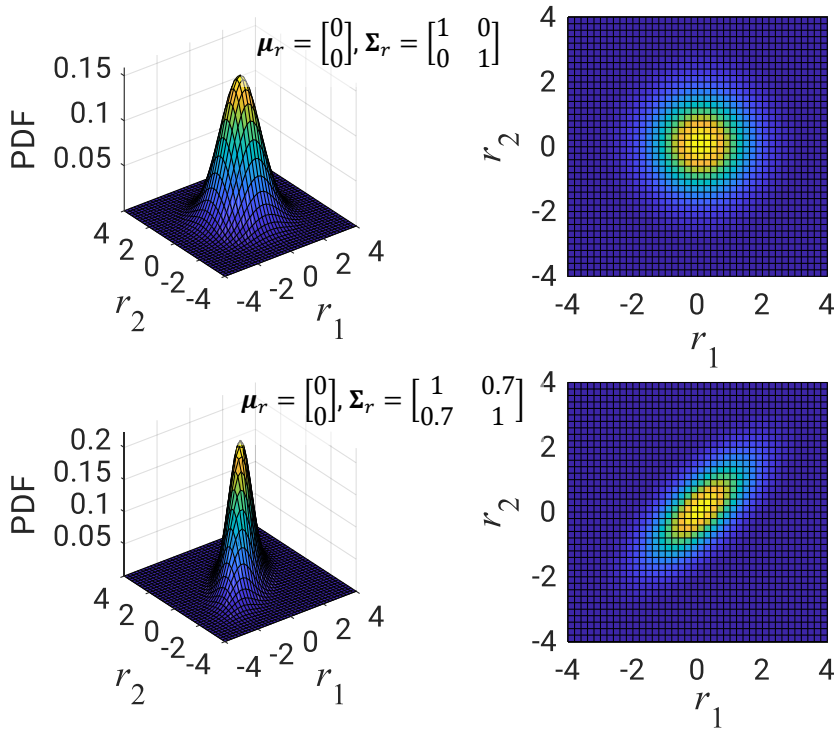

Figure 2.3: Examples of bi-variate Gaussians.

transpose of matrix $(\cdot)$. Figure 2.3 illustrates the PDFs of bi-variate Gaussians for two different sets of $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$.

Now suppose $\mathcal{N}(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$ is defined jointly over two random parameter vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ whose dimensions are not necessarily identical, such that

$$\boldsymbol{\mu}_r = \begin{bmatrix}\boldsymbol{\mu}_{r_1}\\ \boldsymbol{\mu}_{r_2}\end{bmatrix}, \quad \boldsymbol{\Sigma}_r = \begin{bmatrix}\boldsymbol{\Sigma}_{r_1 r_1} & \boldsymbol{\Sigma}_{r_1 r_2}\\ \boldsymbol{\Sigma}_{r_2 r_1} & \boldsymbol{\Sigma}_{r_2 r_2}\end{bmatrix}. \tag{2.18}$$

Here $\boldsymbol{\Sigma}_{r_1 r_2}$ denotes the cross-covariance matrix between random vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ with $\boldsymbol{\Sigma}_{r_1 r_2} = \boldsymbol{\Sigma}_{r_2 r_1}^T$. The marginal PDF $\pi(\mathbf{r}_1)$ and the conditional PDF $\pi(\mathbf{r}_1; \mathbf{r}_2)$ can be evaluated by

$$\pi(\mathbf{r}_1) = \mathcal{N}(\mathbf{r}_1; \boldsymbol{\mu}_{r_1}, \boldsymbol{\Sigma}_{r_1 r_1}), \tag{2.19a}$$

$$\pi(\mathbf{r}_1; \mathbf{r}_2) = \mathcal{N}\left(\mathbf{r}_1; \boldsymbol{\mu}_{r_1} + \boldsymbol{\Sigma}_{r_1 r_2}\boldsymbol{\Sigma}_{r_2 r_2}^{-1}(\mathbf{r}_2 - \boldsymbol{\mu}_{r_2}), \boldsymbol{\Sigma}_{r_1 r_1} - \boldsymbol{\Sigma}_{r_1 r_2}\boldsymbol{\Sigma}_{r_2 r_2}^{-1}\boldsymbol{\Sigma}_{r_2 r_1}\right). \tag{2.19b}$$

2.2.2 Continuous uniform distribution

The continuous random parameter r uniformly distributed over the interval $[a, b]$ has the following PDF:

$$\pi(r; a, b) = U(r; a, b) = \begin{cases}(b-a)^{-1} & \text{if } \ a \le r \le b,\\ 0 & \text{otherwise,}\end{cases} \tag{2.20}$$

where U denotes a uniform PDF, and a and b are two parameters characterizing the distribution. Figure 2.4 shows the PDFs and CDFs for two different continuous uniform distributions.

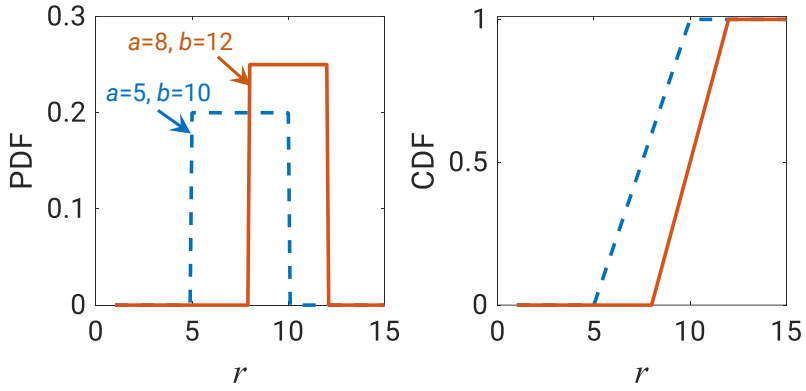

Figure 2.4: Examples of continuous uniform PDFs and CDFs.

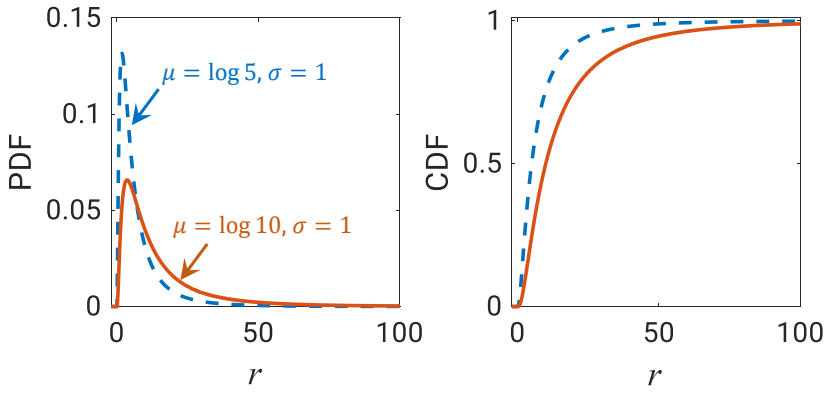

Figure 2.5: Examples of lognormal PDFs and CDFs.

2.2.3 Lognormal distribution

The PDF for the lognormal distribution reads

$$\pi(r;\mu,\sigma) = \frac{1}{r\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log r - \mu)^2}{2\sigma^2}\right] \quad \text{for} \quad r > 0, \tag{2.21}$$

where μ and σ are the mean-of-logarithmic and standard-deviation-of-logarithmic values characterizing the distribution. Figure 2.5 shows the PDFs and CDFs for two different lognormal distributions.

The mean μ_r and variance σ_r^2 of r can be evaluated through μ and σ as

$$\mu_r = \exp(\mu + \sigma^2/2), \quad \sigma_r^2 = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right]. \tag{2.22}$$

2.2.4 Beta distribution

The beta distribution is supported on the interval $[0, 1]$ and its PDF is given as follows:

$$\pi(r;a,b) = \frac{1}{B(a,b)} r^{a-1}(1-r)^{b-1}, \tag{2.23}$$

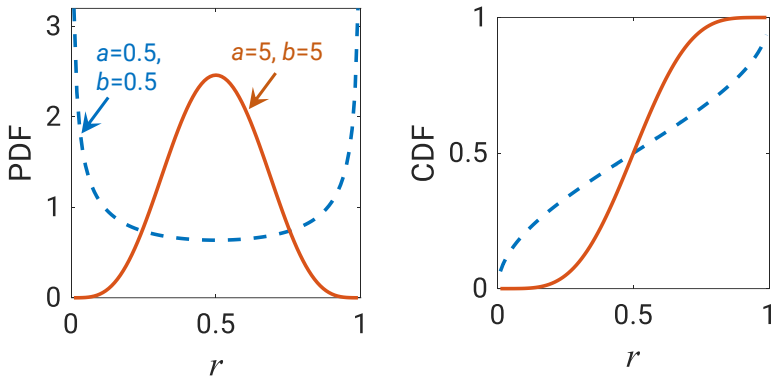

Figure 2.6: Examples of beta PDFs and CDFs.

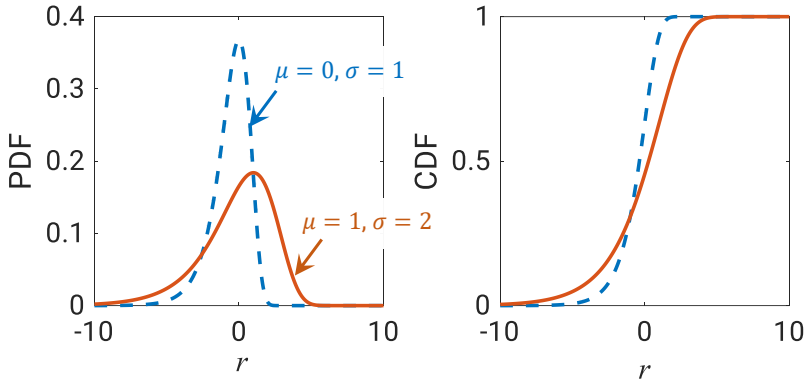

Figure 2.7: Examples of extreme-value type I PDFs and CDFs.

where $B(\cdot,\cdot)$ denotes the beta function, and a and b are the shape parameters of the PDF. Figure 2.6 provides examples of the PDFs and CDFs for two different beta distributions.

2.2.5 Extreme-value type I distribution

The PDF for the extreme-value type I distribution with location parameter μ and scale parameter σ is

$$\pi(r;\mu,\sigma) = \sigma^{-1} \exp\left(\frac{r-\mu}{\sigma}\right) \exp\left[-\exp\left(\frac{r-\mu}{\sigma}\right)\right]. \tag{2.24}$$

Examples of the PDFs and CDFs for two different extreme-value type I distributions are given in Fig. 2.7.

2.2.6 Weibull distribution

The PDF for the Weibull distribution reads

$$\pi(r;a,b) = W(r;a,b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} \exp\left[\left(-\frac{x}{a}\right)^{b}\right], \tag{2.25}$$

where $a > 0$ and $b > 0$ are the scale and shape parameters of the distribution, respectively. Figure 2.8 provides examples of the PDFs and CDFs for two different Weibull distributions.

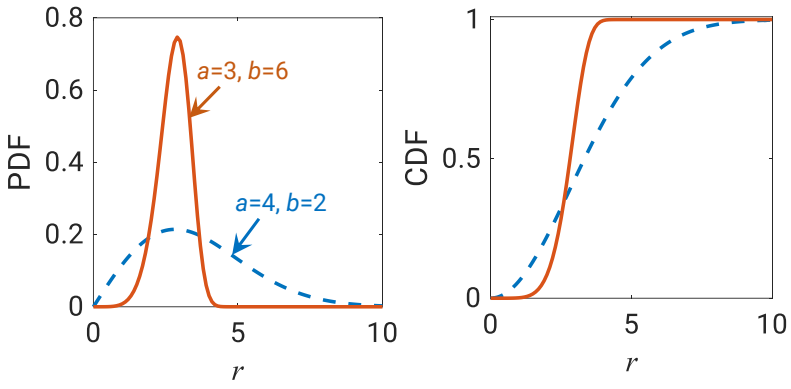

Figure 2.8: Examples of Weibull PDFs and CDFs.

2.3 Gaussian process (GP) modeling

The GP model described in this section is a key ingredient of the BO approaches presented in Section 2.4, Chapters 3, 4, and 6. It also serves as surrogates for the LSFs in Chapters 7 and 8.

The GP modeling aims to construct a probabilistic surrogate model called GP that is capable of quantifying both the prediction of a function and uncertainty in that prediction using a probability distribution over functions [58]. Constructing such a model requires an initial set of input variables and the corresponding output values called training dataset. Given a training dataset $\mathcal{D} = \{\mathbf{x}^i, f^i\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^d$ are d-dimensional vectors of the input variables and $f_i \in \mathbb{R}$ are the corresponding (noise-corrupted) output values. Here $\mathbf{x} = [\mathbf{s}^T, \mathbf{r}^T]^T$ if the design variables $\mathbf{s}$ is distinguished from the random parameters $\mathbf{r}$ or $\mathbf{x} = \mathbf{s}$ if $\mathbf{r}$ does not exist. The goal is to establish a relationship between $\mathbf{x}$ and f using the mapping $f = \hat{y}(\mathbf{x}) + \omega_f : \mathbb{R}^d \to \mathbb{R}$, where $\hat{y}(\mathbf{x})$ is a Gaussian conditioned on $\mathcal{D}$ and $\omega_f \sim \mathcal{N}\left(0, \omega^2\right)$ is additive Gaussian noise.

The GP modeling imposes a prior belief that any finite subset of an infinite set of the output values has a joint Gaussian PDF [58]. Thus, the output values $\{f^1, \ldots, f^N\}$ are distributed according to the following Gaussian:

$$\begin{bmatrix} f^1 \\ \vdots \\ f^N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}^1) \\ \vdots \\ m(\mathbf{x}_N) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & \cdots & k(\mathbf{x}^1, \mathbf{x}^N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}^N, \mathbf{x}^1) & \cdots & k(\mathbf{x}^N, \mathbf{x}^N) \end{bmatrix} + \omega^2 \mathbf{I} \right), \tag{2.26}$$

where $\mathbf{I}$ is the identity matrix, and $m(\mathbf{x}) = \mathbb{E}[\hat{y}(\mathbf{x})]$ and $k(\mathbf{x}, \mathbf{x}')$ represent the mean and covariance kernel functions, respectively. The mean function may be set as $m(\mathbf{x}) = 0$ because the covariance kernel function is flexible enough to handle the role of $m(\mathbf{x})$ [58]. The covariance kernel function is defined for any pair of the input variable vectors $\mathbf{x}$ and $\mathbf{x}'$ to measure the similarity between two corresponding values $\hat{y}(\mathbf{x})$ and $\hat{y}(\mathbf{x}')$, such that

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\big[\big(\hat{y}(\mathbf{x}) - m(\mathbf{x})\big)\big(\hat{y}(\mathbf{x}') - m(\mathbf{x}')\big)\big]. \tag{2.27}$$

This study extensively uses the squared exponential kernel as

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^T (\mathbf{x} - \mathbf{x}')}{2\theta_l^2} \right), \tag{2.28}$$

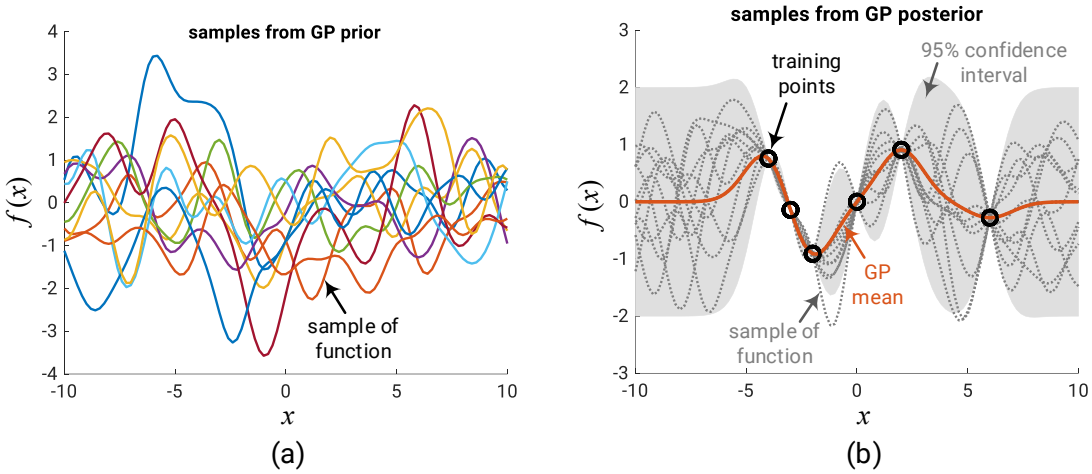

Figure 2.9: Examples of functions sampled from (a) the GP prior and (b) the GP posterior constructed from six training points.

where hyperparameter θ_{l} denotes the characteristic length-scale parameter determined by maximizing the marginal likelihood of $\mathcal{D}$ [58, 67]. ω can be either specified by the user or determined, as a hyperparameter, by maximizing the marginal likelihood of $\mathcal{D}$.

Once θ_{l} has been determined, the information in Eq. (2.26) is utilized for predicting an output value f^* at a new input variable vector $\mathbf{x}^*$, i.e., $f^*|\mathbf{f} = \hat{f}(\mathbf{x}^*)$, where $\mathbf{f} = [f^1, \ldots, f^N]^T$. Because of the GP assumption, the joint PDF of f^* and $\mathbf{f}$ is an $(N+1)$-variate Gaussian, such that

$$\begin{bmatrix} f^* \\ \mathbf{f} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}^*) \\ \mathbf{m}(\mathbf{X}) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}^*, \mathbf{x}^*) & \mathbf{K}(\mathbf{x}^*, \mathbf{X}) \\ \mathbf{K}(\mathbf{x}^*, \mathbf{X})^T & \mathbf{K}(\mathbf{X}, \mathbf{X}) \end{bmatrix} \right), \tag{2.29}$$

where $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^N]^T$, $\mathbf{m}(\mathbf{X}) = [m(\mathbf{x}^1), \ldots, m(\mathbf{x}^N)]^T$, and

$$\mathbf{K}(\mathbf{x}^*, \mathbf{X}) = \big[k(\mathbf{x}^*, \mathbf{x}_1), \ldots, k(\mathbf{x}^*, \mathbf{x}_N)\big], \tag{2.30a}$$

$$\mathbf{K}(\mathbf{X}, \mathbf{X}) = \begin{bmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & \cdots & k(\mathbf{x}^1, \mathbf{x}^N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}^N, \mathbf{x}^1) & \cdots & k(\mathbf{x}^N, \mathbf{x}^N) \end{bmatrix} + \omega^2 \mathbf{I}. \tag{2.30b}$$

The conditional Gaussian $f^*|\mathbf{f} = \hat{f}(\mathbf{x}^*) = \mathcal{N}\left(\mu_{\hat{f}}(\mathbf{x}^*), \sigma^2_{\hat{f}}(\mathbf{x}^*)\right)$ can be derived from Eq. (2.29) using the conditioning rule in Eq. (2.19b). As a result, $\mu_{\hat{f}}(\mathbf{x}^*)$ and $\sigma^2_{\hat{f}}(\mathbf{x}^*)$ read

$$\mu_{\hat{f}}(\mathbf{x}^*) = m(\mathbf{x}^*) + \mathbf{K}(\mathbf{x}^*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\left(\mathbf{f} - \mathbf{m}(\mathbf{X})\right), \tag{2.31a}$$

$$\sigma^2_{\hat{f}}(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\mathbf{x}^*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{x}^*, \mathbf{X})^T. \tag{2.31b}$$

It is worth noting that $\hat{y}(\mathbf{x}^*) = \hat{f}(\mathbf{x}^*)$ if $\omega = 0$, leading to $\hat{y}(\mathbf{x}^*) = \mathcal{N}\left(\mu_{\hat{y}}(\mathbf{x}^*), \sigma^2_{\hat{y}}(\mathbf{x}^*)\right)$, which is called the noise-free GP prediction model.

Figure 2.9 describes the underlying foundation of GP modeling as a Bayesian regression method for prediction of a uni-variate function $f(x)$. GP first defines a prior PDF of functions and then converts this prior into a posterior PDF of functions once some data points have been observed.

Figure 2.9(a) shows 10 functions sampled from the GP prior in Eq. (2.26) with $\omega^2 = 10^{-15}$ and the squared exponential kernel with $\theta_\text{l} = 1$, while Fig. 2.9(b) shows 10 functions sampled from the GP posterior $\mathcal{N}\left(\mu_{\hat{f}}(x^*), \sigma^2_{\hat{f}}(x^*)\right)$ constructed from six training points, the GP posterior mean function in Eq. (2.31a), and 95% confidence interval derived from the GP posterior mean and variance in Eqs. (2.31a) and (2.31b).

2.4 Bayesian optimization (BO)

In this section, we consider the following bound constrained optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X} = [\mathbf{x}_\text{l}, \mathbf{x}_\text{u}], \end{aligned} \tag{2.32}$$

where $\mathcal{X}$ is the feasible region of design variables and the objective function $f(\mathbf{x}) = y(\mathbf{x}) + \mathcal{N}\left(0, \omega^2\right)$ that is either noise-free if $\omega = 0$ (i.e., $f(\mathbf{x}) = y(\mathbf{x})$) or noise-corrupted if $\omega > 0$. $y(\mathbf{x})$ is called the objective-value generator, for example, the output of an FE model of a steel specimen or a steel frame. The BO approaches in this section facilitate the solutions to single-objective inverse identification problems in Chapter 3 and probabilistic identification problems in Chapter 4. They also underlie the multi-objective BO approaches proposed in Chapters 3 and 6 based on a so-called hypervolume measure for solving multi-objective inverse identification problems and multi-objective RDO problems, respectively.

2.4.1 Noise-free objective function

The standard BO is a sequential global optimization method well suited for solving problem (2.32) with an expensive-to-evaluate objective function $f(\mathbf{x})$ [69–71]. It starts by constructing from a (noise-free) training dataset $\mathcal{D} = \{\mathbf{x}^i, f^i\}_{i=1}^N$ and the GP model $\hat{f}(\mathbf{x}) = \mathcal{N}\left(\mu_{\hat{f}}(\mathbf{x}), \sigma^2_{\hat{f}}(\mathbf{x})\right)$ that approximates $f(\mathbf{x})$. By utilizing the information of this GP model, BO formulates an acquisition function $\alpha(\mathbf{x})$ that assists the algorithm in selection of a new, good design point in the next iteration without calling $f(\mathbf{x})$, thereby considerably reducing the number of costly evaluations needed to find a good solution. BO terminates the optimization process and outputs a best-found solution if the number of its iterations reaches a pre-specified upper limit s_max.

Suppose BO has completed the s-th iteration at which a solution has been found from $\mathcal{D}$. By maximizing $\alpha(\mathbf{x})$, BO selects a new design point $\mathbf{x}^{s+1}$ for use in the next iteration, i.e., the $(s+1)$-th iteration, such that

$$\mathbf{x}^{s+1} = \underset{\mathbf{x} \in \mathcal{X}}{\text{argmax}}\ \alpha(\mathbf{x}). \tag{2.33}$$

Once $\mathbf{x}^{s+1}$ has been found, $f(\mathbf{x}^{s+1})$ is evaluated. Subsequently, $\mathbf{x}^{s+1}$ and $f(\mathbf{x}^{s+1})$ are added to the current training dataset for updating the solution as well as the GP model.

Figure 2.10 illustrates the process of two consecutive iterations of BO. In Fig. 2.10(a), the GP model of a uni-variate objective function $f(x)$ is constructed in the first iteration (i.e., $s = 1$) from four training points. The associated acquisition function $\alpha(x)$ is then formulated and maximized. In Fig. 2.10(b), the maximizer of $\alpha(x)$ found in the first iteration is used to update the GP model in the second iteration (i.e., $s = 2$) and the updated GP model formulates a new $\alpha(x)$.

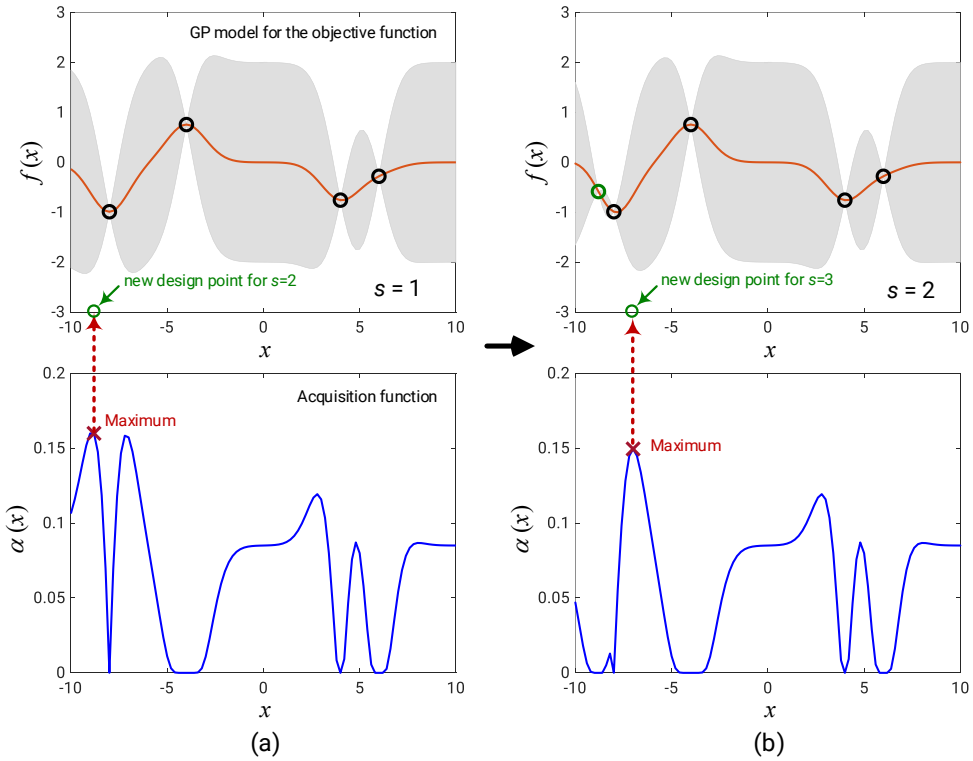

Figure 2.10: Illustration of BO for minimizing an objective function f of a 1-dimensional continuous variable x.

BO formulates $\alpha(\mathbf{x})$ in Eq. (2.33) based on two main criteria: (1) improving the current best-found solution (i.e., exploitation) and (2) reducing the uncertainty in GP predictions (i.e., exploration). The literature has provided many acquisition functions developed based on these criteria [69–71] and three of them are described below.

Let $f_{\min}$ denote the current best value of f associated with the best-found solution, and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the CDF and PDF of the standard Gaussian, respectively. The first acquisition function is the probability of improvement (PI), which is defined as [72]

$$\text{(PI)} \qquad \alpha(\mathbf{x}) = \mathbb{P}\left[f(\mathbf{x}) \leq f_{\min}; \hat{f}(\mathbf{x})\right] = \Phi\left(\frac{f_{\min} - \mu_{\hat{f}}(\mathbf{x})}{\sigma_{\hat{f}}(\mathbf{x})}\right). \tag{2.34}$$

Maximizing PI leads to maximization of the chance that $\mathbf{x}^{s+1}$ is better than the best solution found so far, which corresponds to $f_{\min}$.

The second acquisition function is the expected improvement (EI), which is given by [69]

$$\text{(EI)} \qquad \alpha(\mathbf{x}) = \mathbb{E}\left[\max\left(0, f_{\min} - f(\mathbf{x})\right); \hat{f}(\mathbf{x})\right] = \left(f_{\min} - \mu_{\hat{f}}(\mathbf{x})\right)\Phi\left(\frac{f_{\min} - \mu_{\hat{f}}(\mathbf{x})}{\sigma_{\hat{f}}(\mathbf{x})}\right) + \sigma_{\hat{f}}(\mathbf{x})\phi\left(\frac{f_{\min} - \mu_{\hat{f}}(\mathbf{x})}{\sigma_{\hat{f}}(\mathbf{x})}\right). \tag{2.35}$$

New parameter vector $\mathbf{x}^{s+1}$ obtained by maximizing EI is likely to maximize the expected improvement over the best-found solution $f_{\min}$. Here EI perfectly balances exploitation (i.e., the first term) and exploration (i.e., the second term).

The third acquisition function used for minimizing $f(\mathbf{x})$ is the negative lower confidence bound

(LCB), which is defined by [73]

$$\text{(LCB)} \qquad \alpha(\mathbf{x}) = -\left[\mu_{\hat{f}}(\mathbf{x}) - \beta\sigma_{\hat{f}}(\mathbf{x})\right], \tag{2.36}$$

where β is a tuning parameter defined by the user to control the trade-off between exploitation and exploration. $\beta = 2$ is a common choice. By maximizing LCB, we find $\mathbf{x}^{s+1}$ so that it improves the solution to the lower bound of $f(\mathbf{x})$ as much as possible.

Algorithm 2.1 summarizes the optimization process that implements BO for minimizing noise-free objective functions.

Algorithm 2.1 BO for minimizing noise-free objective functions

Specify $\mathcal{X}$, $\omega = 0$, $s_{\max}$, N, $y(\mathbf{x})$ expensive-to-evaluate function;
Generate N samples of design variable vector $\mathbf{x}^i$ using Latin hypercube sampling [74];
$\mathcal{D} \leftarrow \emptyset$;
for $i = 1 : N$ **do**
    $f^i \leftarrow y(\mathbf{x}^i)$; ▷ Costly step
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^i, f^i\}$;
end for
$f_{\min} \leftarrow \min\{f^i,\ i = 1, \ldots, N\}$;
for $s = 2 : s_{\max}$ **do**
    Construct $\hat{f}(\mathbf{x})$ based on $\mathcal{D}$; see Section 2.3;
    Formulate $\alpha(\mathbf{x})$ (e.g., PI, EI, or LCB);
    Find $\mathbf{x}^{s+1}$ by maximizing $\alpha(\mathbf{x})$;
    $f(\mathbf{x}^{s+1}) \leftarrow y(\mathbf{x}^{s+1})$; ▷ Costly step
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^{s+1}, f(\mathbf{x}^{s+1})\}$;
    $f_{\min} \leftarrow \min\{f^i,\ i = 1, \ldots, N + s\}$;
end for
return $f_{\min}$ and the corresponding vector of design variables.

2.4.2 Noise-corrupted objective function

Minimizing $f(\mathbf{x})$ in the presence of noise, i.e., $\omega > 0$, is more difficult as the value of $f(\mathbf{x})$ is noise-corrupted. From the BO perspective, this difficulty arises from the fact that $f_{\min}$ is not known if $\omega > 0$. Various BO variants [75] have been developed based on the EI acquisition function to address this issue. The main idea is to replace $f_{\min}$ by an evaluable, efficient representative $c_{\min}$ so that BO is still capable of using Eq. (2.35) for guiding the optimization process, which is called the "plug-in" method [75]. Accordingly, the "plug-in" EI reads [76]

$$EI_{\text{p}}(\mathbf{x}) = \left(c_{\min} - \mu_{\hat{f}}(\mathbf{x})\right)\Phi\left(\frac{c_{\min} - \mu_{\hat{f}}(\mathbf{x})}{\sigma_{\hat{f}}(\mathbf{x})}\right) + \sigma_{\hat{f}}(\mathbf{x})\phi\left(\frac{c_{\min} - \mu_{\hat{f}}(\mathbf{x})}{\sigma_{\hat{f}}(\mathbf{x})}\right). \tag{2.37}$$

The value $c_{\min}$ can be assigned as the GP mean evaluated at a so-called effective best solution $\mathbf{x}^*$, hence $c_{\min} = \mu_{\hat{f}}(\mathbf{x}^*)$, where $\mathbf{x}^*$ is selected from $\mathcal{D}$ so that it minimizes $\mu_{\hat{f}}(\mathbf{x}) + \sigma_{\hat{f}}(\mathbf{x})$ [76]. By doing so, $c_{\min}$ is expected to be less sensitive to noise.

To further enhance exploration, a heuristic multiplier is added to $EI_{\text{p}}(\mathbf{x})$, resulting in the fol-

lowing acquisition function [76]:

$$\alpha(\mathbf{x}) = EI_{\mathrm{p}}(\mathbf{x}) \left(1 - \frac{\omega}{\sqrt{\sigma_{\hat{f}}^2(\mathbf{x}) + \omega^2}} \right). \tag{2.38}$$

This acquisition function strengthens exploration by penalizing points with small variance $\sigma_{\hat{f}}^2(\mathbf{x})$ of GP prediction.

To this end, Algorithm 2.2 summarizes the procedure of using BO for minimizing noise-corrupted objective functions.

Algorithm 2.2 BO for minimizing noise-corrupted objective functions

Specify $\mathcal{X}$, $\omega > 0$, $s_{\max}$, N, $y(\mathbf{x})$ expensive-to-evaluate function;
Generate N samples of design variable vector $\mathbf{x}^i$ using Latin hypercube sampling [74];
$\mathcal{D} \leftarrow \emptyset$;
for $i = 1 : N$ **do**
    $f^i \leftarrow y(\mathbf{x}^i) + \mathcal{N}\left(0, \omega^2\right)$; ▷ Costly step
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^i, f^i\}$;
end for
$c_{\min} \leftarrow \min\{\mu_{\hat{f}}(\mathbf{x}^i) + \sigma_{\hat{f}}(\mathbf{x}^i),\, i = 1, \ldots, N\}$;
for $s = 2 : s_{\max}$ **do**
    Construct $\hat{f}(\mathbf{x})$ based on $\mathcal{D}$; see Section 2.3;
    Formulate $\alpha(\mathbf{x})$ as given in Eq. (2.38);
    Find $\mathbf{x}^{s+1}$ by maximizing $\alpha(\mathbf{x})$;
    $f(\mathbf{x}^{s+1}) \leftarrow y(\mathbf{x}^{s+1}) + \mathcal{N}\left(0, \omega^2\right)$; ▷ Costly step
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^{s+1}, f(\mathbf{x}^{s+1})\}$;
    $c_{\min} \leftarrow \min\{\mu_{\hat{f}}(\mathbf{x}^i) + \sigma_{\hat{f}}(\mathbf{x}^i),\, i = 1, \ldots, N + s\}$;
end for
return $c_{\min}$ and the corresponding vector design variables.

2.5 Gaussian mixture model (GMM)

GMM is a probabilistic model that utilizes a weighted sum of Gaussians to approximate the joint PDF of random input parameters and the corresponding output variable [77–79]. Many powerful abilities of the GMM have been demonstrated in the field of statistical learning [78]. For example, it can transform non-Gaussian random input parameters to a weighted sum of Gaussians, thereby enabling an analytical solution for a moment generating function, which has been used to characterize probabilistic properties, and employed in further reliability analyses of structural systems [80]. In Chapter 5, the GMM also finds its success in finding a simple regression function that facilitates the calculations of the mean, variance, and higher-order central moments of an output variable with a modest computational cost for evaluating gradient and Hessian of the output variable. The GMM is characterized by a set of unknown parameters, namely the mixing proportion vector, the number of Gaussians, the mean vectors, and the covariance matrices.

This section describes the mathematical foundation of GMM and how its parameters can be determined via an expectation maximization (EM) algorithm, followed by its application to capturing the probabilistic properties of data points generated from a bi-variate model. The GMM and its parameters underlie the approach presented Chapter 5 to solving the multi-objective RDO problem

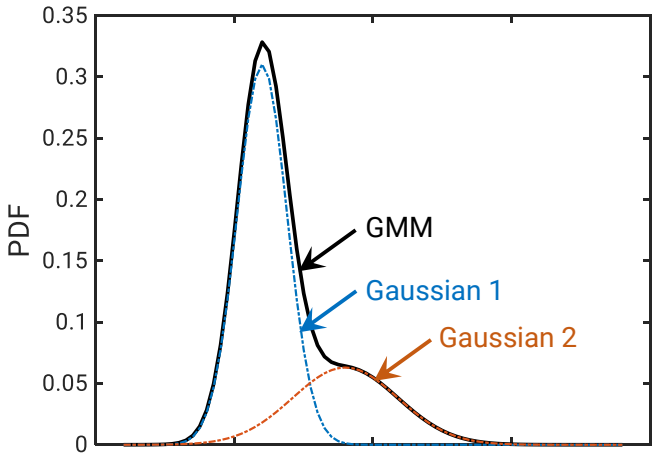

Figure 2.11: Example of a univariate GMM constructed from two Gaussian components.

of steel frames.

2.5.1 Outline of GMM

We again consider the training dataset $\mathcal{D} = \{\mathbf{x}^i, f^i\}_{i=1}^N$. Assume that the training dataset emerges from a joint PDF $\pi(\mathbf{x}, f)$. Our goal is to find a transformation from the space of the random input parameters to that of the random output variable through a conditional PDF $\pi(f; \mathbf{x})$, which is directly derived from $\pi(\mathbf{x}, f)$ and the marginal PDF $\pi(\mathbf{x})$ as follows:

$$\pi(f; \mathbf{x}) = \frac{\pi(\mathbf{x}, f)}{\pi(\mathbf{x})}. \tag{2.39}$$

As its name indicates, GMM describes $\pi(\mathbf{x}, f)$ by a weighted sum of n_g Gaussian components, such that

$$\pi(\mathbf{x}, f) = \sum_{k=1}^{n_\mathrm{g}} w_k \mathcal{N}\left(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \tag{2.40}$$

where

$$\mathbf{y} = \left[\mathbf{x}^T, f\right]^T, \tag{2.41a}$$

$$\sum_{k=1}^{n_\mathrm{g}} w_k = 1, \quad 0 \leq w_k \leq 1, \tag{2.41b}$$

$$\boldsymbol{\mu}_k = \begin{bmatrix} \boldsymbol{\mu}_{x,k} \\ \mu_{f,k} \end{bmatrix}, \quad \boldsymbol{\Sigma}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{xx,k} & \boldsymbol{\Sigma}_{xf,k} \\ \boldsymbol{\Sigma}_{fx,k} & \sigma_{f,k}^2 \end{bmatrix}. \tag{2.41c}$$

Figure 2.11 shows a univariate GMM constructed from two Gaussian components, namely Gaussian 1 and Gaussian 2.

The GMM parameters consist of the mixing proportion w_k, mean vector $\boldsymbol{\mu}_k$, and covariance matrix $\boldsymbol{\Sigma}_k$ of the k-th Gaussian component. From Eq. (2.41c), we can further decompose $\mathcal{N}\left(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$ as

$$\mathcal{N}\left(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \mathcal{N}\left(f(\mathbf{x}); \mu_k(\mathbf{x}), \sigma_k^2(\mathbf{x})\right) \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right), \tag{2.42}$$

which is derived from the conditioning rule in Eqs. (2.19a) and (2.19b), where

$$\mu_k(\mathbf{x}) = \mu_{f,k} + \boldsymbol{\Sigma}_{fx,k}\boldsymbol{\Sigma}_{xx,k}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}_{x,k}\right), \tag{2.43a}$$

$$\sigma_k^2(\mathbf{x}) = \sigma_{f,k}^2 - \boldsymbol{\Sigma}_{fx,k}\boldsymbol{\Sigma}_{xx,k}^{-1}\boldsymbol{\Sigma}_{xf,k}. \tag{2.43b}$$

Substituting Eq. (2.42) into Eq. (2.40) gives

$$\pi(\mathbf{x}, f) = \sum_{k=1}^{n_\mathrm{g}} w_k \mathcal{N}\left(f(\mathbf{x}); \mu_k(\mathbf{x}), \sigma_k^2(\mathbf{x})\right) \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right). \tag{2.44}$$

Thus, the marginal PDF $\pi(\mathbf{x})$ and conditional PDF $\pi(f; \mathbf{x})$ can be derived from the joint PDF $\pi(\mathbf{x}, f)$ as

$$\pi(\mathbf{x}) = \sum_{k=1}^{n_\mathrm{g}} w_k \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right), \tag{2.45a}$$

$$\pi(f; \mathbf{x}) = \sum_{k=1}^{n_\mathrm{g}} W_k(\mathbf{x}) \mathcal{N}\left(f(\mathbf{x}); \mu_k(\mathbf{x}), \sigma_k^2(\mathbf{x})\right), \tag{2.45b}$$

where

$$W_k(\mathbf{x}) = \frac{w_k \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right)}{\sum_{k=1}^{n_\mathrm{g}} w_k \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right)}. \tag{2.46}$$

2.5.2 Determination of GMM parameters

The unknown parameters of GMM, including w_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$, are determined by the EM algorithm that is an iterative solver for maximizing the log-likelihood function of a training dataset [78, 79]. Assume that the data points of the training dataset are independent, the log-likelihood function of GMM measuring how well the GMM model fits the data points is defined as follows:

$$\mathcal{L} = \sum_{i=1}^{N} \log\left[\sum_{k=1}^{n_\mathrm{g}} w_k \mathcal{N}\left(\mathbf{y}^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right]. \tag{2.47}$$

It is difficult to maximize $\mathcal{L}$ because "log" is outside of the inner summation. To overcome this difficulty, the EM algorithm introduces a latent variable $\mathbf{z} = [z^1, \ldots, z^i, \ldots, z^N]$ with $z^i \in \{1, , \ldots, k, \ldots, n_\mathrm{g}\}$ so that if $z^i = k$, the i-th data point emerges from the k-th Gaussian component. Therefore, the joint PDF of the complete set of N data points $\{\mathbf{x}^i, f^i, z^i\}_{i=1}^N$ can be simplified as [78]

$$\pi(\mathbf{x}, f, z) = \prod_{i=1}^{N}\prod_{k=1}^{n_\mathrm{g}} \left[w_k \mathcal{N}\left(\mathbf{y}^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right]^{z^{ik}}, \tag{2.48}$$

where

$$z^{ik} = \begin{cases} 1 & \text{if } z^i = k, \\ 0 & \text{if } z^i \neq k, \end{cases} \tag{2.49}$$

which is the probability that the data point $\mathbf{y}^i$ belongs to the k-th Gaussian component. As a result, the log-likelihood of the complete data reads

$$\mathcal{L}_{\mathrm{c}} = \sum_{i=1}^{N} \sum_{k=1}^{n_{\mathrm{g}}} z^{ik} \log \left[w_k \mathcal{N}\left(\mathbf{y}^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \right]. \tag{2.50}$$

The EM algorithm then maximizes $\mathcal{L}_{\mathrm{c}}$ through E- and M-steps as follows [78]:
E-step: Substitute $\hat{z}^{ik}$ for z^{ik} in Eq. (2.49) that is unknown in advance, where

$$\hat{z}^{ik} = \frac{w_k \mathcal{N}\left(\mathbf{y}^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{k=1}^{n_{\mathrm{g}}} w_k \mathcal{N}\left(\mathbf{y}^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}, \tag{2.51}$$

which is the expectation of z^{ik} in Eq. (2.49).
M-step: Maximize $\mathcal{L}_{\mathrm{c}} = \sum_{i=1}^{N} \sum_{k=1}^{n_{\mathrm{g}}} \hat{z}^{ik} \log \left[w_k \mathcal{N}\left(\mathbf{y}^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \right]$ subject to the constraints in Eq. (2.41b) for updating the mixing proportion, resulting in

$$w_k = \frac{1}{N} \sum_{i=1}^{N} \hat{z}^{ik}. \tag{2.52}$$

The parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are also updated as follows:

$$\boldsymbol{\mu}_k = \sum_{i=1}^{N} \hat{z}^{ik} \mathbf{y}^i \Big/ \sum_{i=1}^{N} \hat{z}^{ik}, \tag{2.53a}$$

$$\boldsymbol{\Sigma}_k = \sum_{i=1}^{N} \hat{z}^{ik} (\mathbf{y}^i - \boldsymbol{\mu}_k)(\mathbf{y}^i - \boldsymbol{\mu}_k)^T \Big/ \sum_{i=1}^{N} \hat{z}^{ik}. \tag{2.53b}$$

Algorithm 2.3 EM algorithm

1: **Specify** $1 \leq n_{\mathrm{g}} \leq n_{\mathrm{c}}$; initial values $w_k^{(0)}$, $\boldsymbol{\mu}_k^{(0)}$, $\boldsymbol{\Sigma}_k^{(0)}$; $tol \leftarrow 1$; $\epsilon \leftarrow 10^{-6}$ a small tolerance value; $s \leftarrow 0$ step number; $\mathcal{L}_{\mathrm{c}}^{(0)} \leftarrow -\infty$;
2: **while** $tol > \epsilon$ **do**
3: $s \leftarrow s + 1$;
4: Compute $\hat{z}^{ik(s)}$ with $w_k^{(s-1)}$, $\boldsymbol{\mu}_k^{(s-1)}$, and $\boldsymbol{\Sigma}_k^{(s-1)}$; see Eq. (2.51);
5: Compute $w_k^{(s)}$ with $\hat{z}^{ik(s)}$; see Eq. (2.52);
6: Compute $\boldsymbol{\mu}_k^{(s)}$ with $\hat{z}^{ik(s)}$; see Eq. (2.53a);
7: Compute $\boldsymbol{\Sigma}_k^{(s)}$ with $\hat{z}^{ik(s)}$ and $\boldsymbol{\mu}_k^{(s)}$; see Eq. (2.53b);
8: Compute $\mathcal{L}_{\mathrm{c}}^{(s)}$; see Eq. (2.50);
9: $tol \leftarrow \left| \mathcal{L}_{\mathrm{c}}^{(s)} - \mathcal{L}_{\mathrm{c}}^{(s-1)} \right| \Big/ \left| \mathcal{L}_{\mathrm{c}}^{(s)} \right|$;
10: **end while**
11: **return** w_k, $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$.

The EM algorithm is summarized in Algorithm 2.3. For a specified value of n_{g}, the initial mixing proportions $w_k^{(0)}$ are uniform, the initial mean vectors $\boldsymbol{\mu}_k^{(0)}$ are n_{g} vectors that are randomly selected from the training set, and the initial variance matrices $\boldsymbol{\Sigma}_k^{(0)}$ for n_{g} Gaussian components

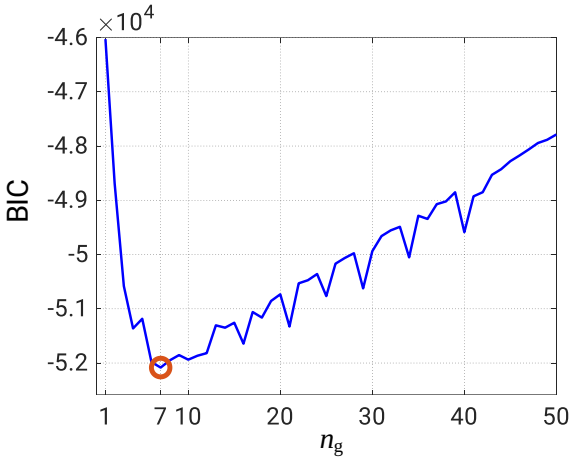

Figure 2.12: BIC versus n_g for selection of the best GMM.

are diagonal, where the i-th diagonal element is the variance of the i-th input variable. To further avoid the ill-conditioning of estimated covariance matrices of the GMM during the implementation of EM algorithm, a small positive regularizer of 0.001 is added to the diagonal components of $\mathbf{\Sigma}_k^{(s)}$. The EM algorithm is assured to terminate after a finite number of iterations since it never reduces the log-likelihood [78].

To determine a proper number of Gaussian components, we increase n_g step by step from 1 to n_c, thereby resulting in a total of n_c GMMs for the model selection. $n_c = 50$ is a common choice to limit the complexity of the GMM [81]. The Bayesian information criterion (BIC) is adopted for the model selection since its effectiveness in determining the number of components for GMMs has been verified [82]. Accordingly, the best GMM from among n_c GMM candidates minimizes BIC that is equivalent to selecting the GMM with the largest posterior probability in the Bayesian inference framework [78].

2.5.3 Test problem: A bi-variate model

Consider the following conditional bi-variate PDF [83]:

$$\pi(f;\mathbf{x}) = \mathcal{N}\left(f; x_1 + x_2, 0.15 + 0.05 x_1 x_2\right), \tag{2.54}$$

where $x_1 \sim U(-0.3, 0.3)$ and $x_2 \sim U(-0.3, 0.3)$.

To verify the performance of GMM, its conditional PDF in Eq. (2.45b) is found to reproduce the PDF in Eq. (2.54). To do so, we generate a total of 4000 samples, in which 2000 samples are randomly selected for constructing GMM and the remaining 2000 samples for testing. Nine pairs of $[x_1, x_2]$, namely $[-0.2, 0.2]$, $[0.0, 0.2]$, $[0.2, 0.2]$, $[-0.2, 0.0]$, $[0.0, 0.0]$, $[0.2, 0.0]$, $[-0.2, -0.2]$, $[0.0, -0.2]$, and $[0.2, -0.2]$, are used for obtaining the corresponding nine true conditional PDFs. By incorporating f values of the test data and each pair of $[x_1, x_2]$ into Eq. (2.45b), the GMM conditional PDF for each pair is obtained and plotted versus the associated true PDF. It is found that the joint PDF $\pi(\mathbf{x}, f)$ in Eq. (2.44) obtained from the trained GMM is a mixture of seven Gaussian components, i.e., $n_g = 7$, which is associated with the minimum BIC shown in Fig. 2.12. The mixing proportion, mean vector, and covariance matrix of each component at $n_g = 7$ are listed

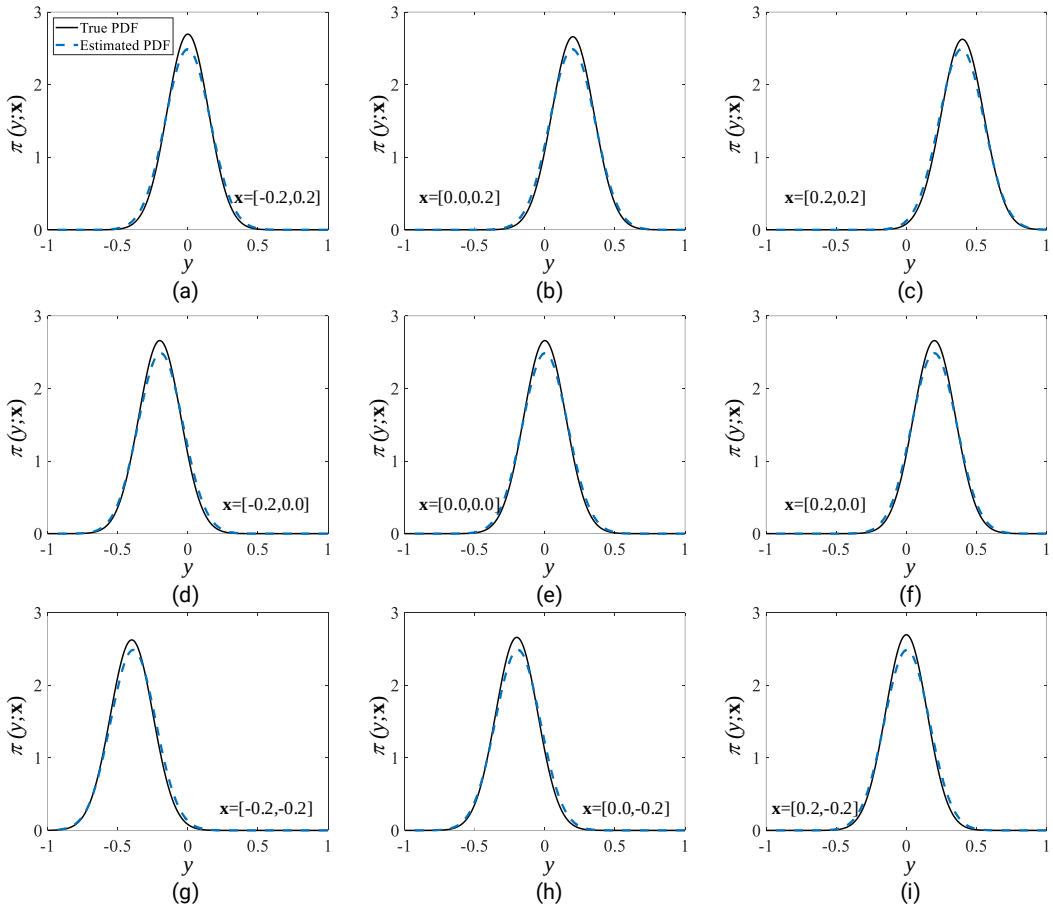

Figure 2.13: Comparison of true and estimated conditional PDFs for the bi-variate model at nine tested input vectors with $n_{\rm g} = 1$.

in Table 2.1. Figures 2.13 and 2.14 show comparisons between the estimated and true conditional PDFs at each pair of $[x_1, x_2]$ for $n_{\rm g} = 1$ and $n_{\rm g} = 7$, respectively. We see that $n_{\rm g} = 7$ shows better estimations of the given conditional PDF at each pair of $[x_1, x_2]$ than $n_{\rm g} = 1$, indicating a reasonable selection of the optimal number of Gaussian components from minimum BIC.

2.6 Saddlepoint approximation (SAA)

Let $K(\xi)$, $K^{(1)}(\xi)$, and $K^{(2)}(\xi)$ denote the cumulant-generating function of a random parameter r, and its first and second derivatives, respectively. The SAA models the PDF of r at a particular value $\bar{r}$ by [68]

$$\pi(\bar{r}) \approx \frac{\exp[K(\xi_{\rm s}) - \xi_{\rm s}\bar{r}]}{\sqrt{2\pi K^{(2)}(\xi_{\rm s})}}, \tag{2.55}$$

which is derived based on the Laplace transform to approximate the integral of $M(\xi)$ in Eq. (2.8) and the Fourier inversion for the approximated integral, where $\xi_{\rm s}$ is called the saddlepoint that is the unique root of the following saddlepoint equation [68]:

$$K^{(1)}(\xi) = \bar{r}. \tag{2.56}$$

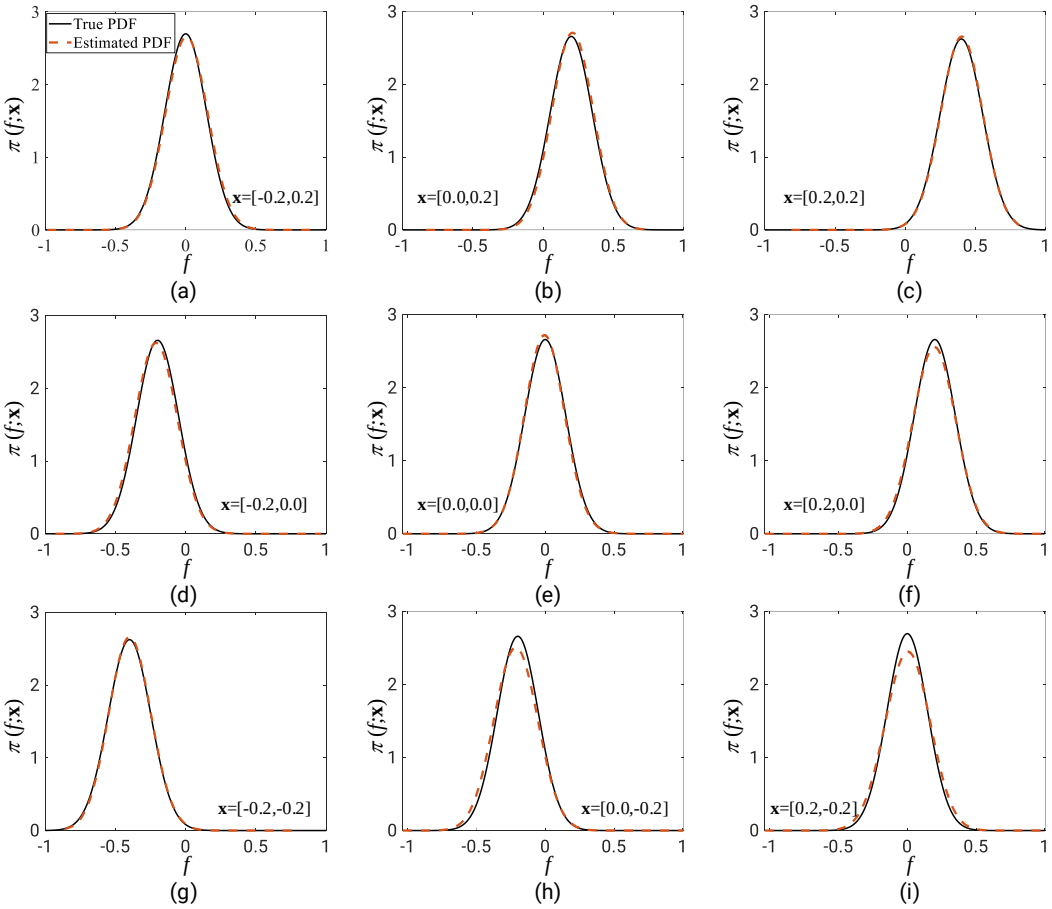

Figure 2.14: Comparison of true and estimated conditional PDFs for the bi-variate model at nine tested input vectors with $n_{\mathrm{g}} = 7$.

Moreover, $\mathbb{P}[r \leq \bar{r}]$ can be approximated by [68]

$$\mathbb{P}[r \leq \bar{r}] \approx \begin{cases} \Phi(q) + \phi(q)\left(q^{-1} - v^{-1}\right) & \text{if} \quad \mu_r \neq \bar{r}, \\ 0.5 + \kappa_{r,3}\left(6\sqrt{2\pi}\kappa_{r,2}^{3/2}\right)^{-1} & \text{if} \quad \mu_r = \bar{r}, \end{cases} \tag{2.57}$$

where $\kappa_{r,2}$ and $\kappa_{r,3}$ are the second and third cumulants of r, respectively; and q and v are given as

$$q = \mathrm{sign}(\xi_{\mathrm{s}})\sqrt{2[\xi_{\mathrm{s}}\bar{r} - K(\xi_{\mathrm{s}})]}, \tag{2.58a}$$

$$v = \xi_{\mathrm{s}}\sqrt{K^{(2)}(\xi_{\mathrm{s}})}, \tag{2.58b}$$

where $\mathrm{sign}(\xi_{\mathrm{s}}) = 1$, -1, or 0 corresponding to $\xi_{\mathrm{s}} > 0$, $\xi_{\mathrm{s}} < 0$, or $\xi_{\mathrm{s}} = 0$, respectively. Thus, an appropriate cumulant-generating function $K(\xi)$ leads to quick estimations of PDF and CDF of r at any $\bar{r}$ through the saddlepoint.

Note that since $K(\xi)$ stated in Eq. (2.9) is a power series expansion of ξ, it should be truncated by keeping a finite number of low order terms for estimating $\mathbb{P}[r \leq \bar{r}]$. Gillespie and Renshaw [84] used the first two or three terms of the power series expansion to approximate the PDF of a random variable. Guo [85] proposed an efficient truncated $K(\xi)$ of a random variable by keeping the

Table 2.1: Mixing proportions, mean vectors, and covariance matrices of seven GMM components for the bi-variate model.

k	w_k	$\boldsymbol{\mu}_k$	$\boldsymbol{\Sigma}_k$
1	0.1902	$[-0.2015, -0.0290, -0.2573]^T$	$\begin{bmatrix} 0.0054 & -0.0001 & 0.0035 \\ -0.0001 & 0.0212 & 0.0177 \\ 0.0035 & 0.0177 & 0.0420 \end{bmatrix}$
2	0.1353	$[0.1421, 0.1909, 0.3632]^T$	$\begin{bmatrix} 0.0106 & 0.0001 & 0.0083 \\ 0.0001 & 0.0063 & 0.0040 \\ 0.0083 & 0.0040 & 0.0323 \end{bmatrix}$
3	0.1669	$[0.1980, -0.0349, 0.1864]^T$	$\begin{bmatrix} 0.0058 & -0.0008 & 0.0029 \\ -0.0008 & 0.0190 & 0.0135 \\ 0.0029 & 0.0135 & 0.0339 \end{bmatrix}$
4	0.0821	$[-0.0014, -0.0130, -0.0167]^T$	$\begin{bmatrix} 0.0159 & -0.0033 & 0.0065 \\ -0.0033 & 0.0153 & 0.0065 \\ 0.0065 & 0.0065 & 0.0269 \end{bmatrix}$
5	0.1826	$[-0.0005, -0.2028, -0.2396]^T$	$\begin{bmatrix} 0.0226 & -0.0001 & 0.0193 \\ -0.0001 & 0.0053 & 0.0032 \\ 0.0193 & 0.0032 & 0.0418 \end{bmatrix}$
6	0.1092	$[-0.0020, -0.0144, -0.0196]^T$	$\begin{bmatrix} 0.0159 & -0.0034 & 0.0064 \\ -0.0034 & 0.0153 & 0.0064 \\ 0.0064 & 0.0064 & 0.0266 \end{bmatrix}$
7	0.1337	$[-0.1074, 0.1921, 0.0993]^T$	$\begin{bmatrix} 0.0133 & -0.0013 & 0.0086 \\ -0.0013 & 0.0063 & 0.0028 \\ 0.0086 & 0.0028 & 0.0286 \end{bmatrix}$

first two terms of the series expansion and introducing an additional logarithmic term to describe the remaining higher-order terms, while the space of the random variables is transformed into a standardized space before performing SAA. However, the condition for the unique root of the saddlepoint equation in Eq. (2.56) and the requirement for the existence of q and v in Eqs. (2.58a) and (2.58b) were not addressed. A new truncated $K(\xi)$ proposed in Chapter 7 is to address the aforementioned issues.

2.7 Summary

This chapter has presented a brief review of probability and probabilistic approaches. It covers common continuous PDFs and basic rules of probability as well as the foundations of GP modeling, BO, GMM, and SAA that are key ingredients for the probabilistic approaches proposed in the subsequent chapters.

Chapter 3

Bayesian optimization approaches to identification of nonlinear hysteresis curve of structural steels

Accurately modeling the cyclic elastoplastic behavior of structural steel plays an important role in establishing reliable analyses of steel structures subjected to earthquake excitation. This chapter is dedicated to inverse problems that infer the cyclic elastoplastic parameters for structural steel based on experimental datasets measured from cyclic tests of a specimen or a structural component. Two inverse problems are formulated, namely single- and multi-objective identification problems. As will be shown in this chapter, the parameters obtained from the single-objective problem are associated with the so-called dataset-specific bias that may lead the parameters identified from a single loading condition to inaccurate predictions of structural responses under other loading conditions, while the parameters obtained from the multi-objective problem can mitigate this dataset-specific bias. Since the traditional identification processes using conventional optimization algorithms still demand a substantial computational cost of repeatedly carrying out many nonlinear analyses, this chapter promotes the use of single-objective Bayesian optimization (SOBO) and multi-objective Bayesian optimization (MOBO) approaches to solving the single- and multi-objective problems, respectively.

3.1 Identification problems

3.1.1 Single-objective identification problem

The bound constrained optimization problem in Eq. (2.32) is formulated as the single-objective inverse problem to identify the vector of cyclic elastoplastic parameters $\mathbf{x} \in [\mathbf{x}_\mathrm{l}, \mathbf{x}_\mathrm{u}]$ from experimental results. Accordingly, we minimize an error function $f(\mathbf{x})$ that measures the discrepancy between the structural responses simulated from an FE model of the experiment, characterized by $\mathbf{x}$, and the corresponding responses measured experimentally. Recall that $f(\mathbf{x}) = y(\mathbf{x}) + \mathcal{N}\left(0, \omega^2\right)$, where $f(\mathbf{x})$ is either noise-free if $\omega = 0$ (i.e., $f(\mathbf{x}) = y(\mathbf{x})$) or noise-corrupted if $\omega > 0$. Thus, $y(\mathbf{x})$ represents the noise-free error function.

Assume that there exist a total of I experimental datasets measured from I different cyclic tests

of a steel specimen (or a structural component) with different loading histories. Let $d^{\mathrm{s}}_{i,t}$ denote the simulated value of the response of interest associated with the t-th time step of the i-th cyclic loading history of N_i discrete steps, and $d^{\mathrm{m}}_{i,t}$ denote the corresponding measured value. Following the root-mean-square deviation, $y(\mathbf{x})$ formulated for the i-th dataset reads [86]

$$y_i(\mathbf{x}) = \sqrt{\frac{1}{N_i}\sum_{t=1}^{N_i}\left(d^{\mathrm{s}}_{i,t}(\mathbf{x}) - d^{\mathrm{m}}_{i,t}\right)^2}. \tag{3.1}$$

The first single-objective identification problem is to find the parameters from each single experimental dataset. Thus, the noise-free error function of this problem reads $y(\mathbf{x}) = y_i(\mathbf{x})$. Meanwhile, the second single-objective identification problem is to incorporate some experimental datasets into an error function as a summation of the associated error functions, which is defined by

$$y(\mathbf{x}) = \sum_{i=1}^{\tilde{I}} y_i(\mathbf{x}), \quad 1 < \tilde{I} \leq I. \tag{3.2}$$

By minimizing this weighted-sum error function, we expect that the parameters identified from the second problem is more reliable than those from the first problem.

It is difficult to solve the above single-objective inverse problems effectively because $y(\mathbf{x})$ is expensive-to-evaluate. This makes common use of population-based optimization algorithms [87, 88] practically inefficient because they require a large number of costly simulations for obtaining a good set of parameters. In addition, the use of gradient-based algorithms [89, 90] is also hindered by the requirement of calculating the gradient and/or Hessian of $y(\mathbf{x})$. Moreover, minimizing noise-corrupted error functions (i.e., $\omega > 0$) has not been fully addressed by the aforementioned optimization algorithms. Since it is desirable to find a good set of parameters while keeping the number of costly simulations as low as possible, SOBO approaches based on the EI acquisition function detailed in Algorithms 2.1 and 2.2 are adopted to solve the two single-objective inverse problems considering noise-free and noise-corrupted error functions, respectively.

3.1.2 Multi-objective identification problem

Another way to find the optimal parameters is to minimize some error functions simultaneously, hence to formulate the following multi-objective identification problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & [f_1(\mathbf{x}), \ldots, f_{\tilde{I}}(\mathbf{x})] \\ \text{subject to} \quad & \mathbf{x} \in [\mathbf{x}_{\mathrm{l}}, \mathbf{x}_{\mathrm{u}}]. \end{aligned} \tag{3.3}$$

Solving this problem provides an approximate Pareto front of parameters that allows designers to select the material parameters for their designs flexibly while the identification is performed only once. The Pareto front is commonly found through a population-based algorithm such as NSGA-II [91]. Such an algorithm directly evaluates the error functions for a population of a large number of candidate solutions, thereby suffering from the computational cost required for calculating $f_i(\mathbf{x})$. To solve the problem effectively, proximal-exploration MOBO is proposed in the next section. To further facilitate the development of MOBO, only noise-free error functions are considered for

problem (3.3). We also drop the subscript of $f_i(\mathbf{x})$ hereafter to simplify the exposition because problem (3.3) treats the error functions equally.

3.2 Identification using proximal-exploration MOBO

3.2.1 Proposed MOBO and identification procedure

The proposed MOBO starts by generating a training dataset $\mathcal{D} = \{\mathbf{x}^k, f^k\}_{k=1}^N$, where $\mathbf{x}^k$ are samples of material parameters and $f^k = y(\mathbf{x}^k)$. The samples $\mathbf{x}^k$ are randomly generated using Latin-hypercube sampling [74]. To evaluate f^k, the FE model of the experiment evaluates $d_{i,t}^s$ in Eq. (3.1) using $\mathbf{x}^k$ as the input of material properties for the i-th loading history. MOBO then finds from the members of $\mathcal{D}$ a set of approximate Pareto-optimal solutions to problem (3.3) using a non-dominated sorting algorithm [91] and sequentially improves this set until it can no longer be improved or until the budgeted computational cost is reached.

Let $\Omega = \{\mathbf{f}_1, ..., \mathbf{f}_M\} \in \mathbb{R}^{\tilde{I}}$ denote the set of M approximate Pareto-optimal solutions sorted from $\mathcal{D}$. The improvement of Ω implements the following five steps:

(1) Construct a total of $\tilde{I}$ GP models as surrogates for the error functions of problem (3.3).

(2) Find the best compromise solution from the members of Ω using a fuzzy-based method.

(3) Formulate an acquisition function based on Ω, the current GP models, and the current best compromise solution.

(4) Maximize the formulated acquisition function for a new parameter vector that updates $\mathcal{D}$.

(5) Sort new Pareto-optimal solutions from the updated $\mathcal{D}$, check the termination conditions, and reiterate from (1) if MOBO is still in process.

In the first step of solution improvement, MOBO utilizes the DACE toolbox [92] to construct a GP model that approximates the error function $f(\mathbf{x}) \in \{f_1(\mathbf{x}), \ldots, f_{\tilde{I}}(\mathbf{x})\}$. Following Eqs. (2.31a) and (2.31b), the GP prediction of the error function f for a particular $\mathbf{x}$ reads

$$\hat{f}(\mathbf{x}) \sim \mathcal{N}\left(\mu_{\hat{f}}(\mathbf{x}), \sigma_{\hat{f}}^2(\mathbf{x})\right). \tag{3.4}$$

where $\mu_{\hat{f}}(\mathbf{x})$ and $\sigma_{\hat{f}}(\mathbf{x})$ are the mean and standard deviation of $\hat{f}(\mathbf{x})$, respectively.

The second step for improving Ω is to find its best compromise member, or equivalently, the best compromise solution of parameters, denoted as $\mathbf{x}_0$. This is done through the fuzzy-based method that computes the following membership function for the i-th error function of each member of the current solution set Ω [93]:

$$m_{i,j} = \begin{cases} 1 & \text{if } \; f_{i,j} = f_i^{\min} \\ \dfrac{f_i^{\max} - f_{i,j}}{f_i^{\max} - f_i^{\min}} & \text{if } \; f_i^{\min} < f_{i,j} < f_i^{\max} \quad (i = 1, \ldots, \tilde{I}; \;\; j = 1, \ldots, M), \\ 0 & \text{if } \; f_{i,j} = f_i^{\max} \end{cases} \tag{3.5}$$

where j indicates the j-th solution among M members of Ω; $f_{i,j}$ is the i-th error function value of the j-th solution; and $f_i^{\min}$ and $f_i^{\max}$ are the minimum and maximum values among M values of

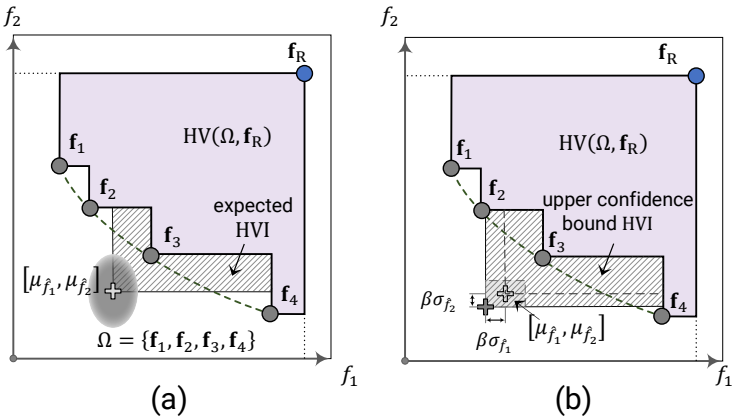

Figure 3.1: Illustrations of the expected and upper confidence bound HVI measures for a bi-objective minimization problem. (a) expected HVI; (b) upper confidence bound HVI.

the i-th error function, respectively. Then, a normalized membership function is defined for $m_{i,j}$, as

$$m_j = \frac{\sum_{i=1}^{\tilde{I}} m_{i,j}}{\sum_{j=1}^{M} \sum_{i=1}^{\tilde{I}} m_{i,j}}. \tag{3.6}$$

The best compromise solution $\mathbf{x}_0$ corresponds to a member of Ω that has the largest value of m_j. If Ω has only two members, either can be selected. Also, the second-best compromise solution, which shall be used in Section 3.3, has the second-largest value of m_j.

The third step for improving Ω is to formulate a hypervolume-based acquisition function [94, 95] for guiding MOBO. Such a hypervolume-based acquisition function is natural because the hypervolume (HV) measure is often used in the field of multi-objective design to assess the quality of different sets of solutions [96, 97]. Let $\mathbf{f}_R \in \mathbb{R}^{\tilde{I}}$ denote a fixed reference point in the error function space so that it is dominated by all members of Ω. Each element of $\mathbf{f}_R$ can be assigned a sufficiently large value of the corresponding error function at which the prediction error is unacceptable. As an example, Fig. 3.1 shows the HV defined by a set of four Pareto-optimal solutions to a bi-objective minimization problem and a reference point dominated by these solutions. Mathematically, the HV defined by Ω and $\mathbf{f}_R$ is a Lebesgue measure of the $\tilde{I}$-dimensional subspace dominated by Ω and bounded above by $\mathbf{f}_R$ [96, 97], such that

$$\mathrm{HV}(\Omega, \mathbf{f}_R) = \Lambda\left(\left\{\mathbf{f} \in \mathbb{R}^{\tilde{I}} \mid \exists \mathbf{f}_j \in \Omega : \mathbf{f}_j \preceq \mathbf{f} \text{ and } \mathbf{f} \preceq \mathbf{f}_R\right\}\right), \tag{3.7}$$

where $\Lambda(\cdot)$ denotes the Lebesgue measure defined for a set on $\mathbb{R}^{\tilde{I}}$ as the $\tilde{I}$-dimensional volume of this set, for example, $\Lambda(\cdot)$ is equivalent to the standard measure of length, area, or volume of set $(\cdot)$ if $\tilde{I} = 1$, 2, or 3, respectively; $\mathbf{f}$ is a point in the error function space; and $\mathbf{f}_j \preceq \mathbf{f}$ indicates $\mathbf{f}_j$ dominates $\mathbf{f}$. The HV in this study is evaluated using an algorithm developed by Fonseca et al. [98].

Now suppose MOBO is processing its s-th iteration and has to specify a new parameter vector $\mathbf{x}^{s+1}$ in the next iteration (i.e., $s+1$) at which costly simulations are carried out to update the current dataset $\mathcal{D}$ as well as the current solution set Ω. As it is desirable to reduce the number

of costly simulations as much as possible, $\mathbf{x}^{s+1}$ should be ideal in the parameter space so that it maximizes the improvement of Ω. Maximizing the improvement of Ω is equivalent to maximizing the difference of the HV defined by the union of $\mathbf{f}(\mathbf{x})$ and Ω, and that defined by Ω, where $\mathbf{f}(\mathbf{x})$ corresponds to an arbitrary vector $\mathbf{x}$ in the parameter space. This difference is represented by the following hypervolume indicator (HVI) [99, 100]:

$$\mathrm{HVI}(\mathbf{f}(\mathbf{x}) \mid \Omega, \mathbf{f}_{\mathrm{R}}) = \mathrm{HV}(\mathbf{f}(\mathbf{x}) \cup \Omega, \mathbf{f}_{\mathrm{R}}) - \mathrm{HV}(\Omega, \mathbf{f}_{\mathrm{R}}), \tag{3.8}$$

where $\mathbf{f}(\mathbf{x}) \cup \Omega$ denotes the union of $\mathbf{f}(\mathbf{x})$ and Ω.

Since the error functions are evaluated through costly simulations, it is inefficient to maximize HVI by direct evaluation of $\mathrm{HV}(\mathbf{f}(\mathbf{x}) \cup \Omega, \mathbf{f}_{\mathrm{R}})$. An expected value of HVI can be estimated by integrating $\mathrm{HV}(\mathbf{f}(\mathbf{x}) \cup \Omega, \mathbf{f}_{\mathrm{R}})$ over a non-dominated region of the error function space defined by the current Ω and the GP prediction models for the error functions (e.g., shading in Fig. 3.1(a)). In this way, the non-dominated region can be decomposed into a set of small disjoint cells over which the integral can be approximated using an analytical form [99]. However, the calculation of the expected HVI using such an integral is not advantageous for quickly finding $\mathbf{x}^{s+1}$. Thus, Chapters 6 and 8 shall replace $\mathrm{HV}(\mathbf{f}(\mathbf{x}) \cup \Omega, \mathbf{f}_{\mathrm{R}})$ with $\mathrm{HV}(\boldsymbol{\mu}_{\hat{f}}(\mathbf{x}) \cup \Omega, \mathbf{f}_{\mathrm{R}})$, leading to an expected measure of HVI as, for example, illustrated by the hatched area in Fig. 3.1(a), where $\boldsymbol{\mu}_{\hat{f}}(\mathbf{x})$ is the Gaussian mean vector of the error functions at $\mathbf{x}$ given in Eqs. (3.4) and (2.31a). This replacement arises because it is often more computationally efficient to perform more optimization iterations than to do an exact calculation of HVI at each iteration. Roussel et al. [101] further enhanced exploration by defining $\mathrm{HV}(\boldsymbol{\mu}_{\hat{f}}(\mathbf{x}) - \beta\boldsymbol{\sigma}_{\hat{f}}(\mathbf{x}) \cup \Omega, \mathbf{f}_{\mathrm{R}})$ as an upper confidence bound HVI, which is depicted by the hatched area in Fig. 3.1(b), where $\boldsymbol{\sigma}_{\hat{f}}(\mathbf{x})$ is the Gaussian standard deviation vector of the error functions at $\mathbf{x}$ given in Eqs. (3.4) and (2.31b), and β is a positive scalar that controls the trade-off between exploration and exploitation of MOBO. A large value of β prioritizes MOBO exploration. In fact, the upper confidence bound HVI is an extension of the negative LCB acquisition function [73] stated in Eq. (2.36).

As we aim to use the best compromise solution $\mathbf{x}_0$ for structural response predictions, we bias MOBO exploration toward the non-dominated region surrounding the best compromise member of Ω and multiply the upper confidence bound HVI by an exponential function of $-\frac{1}{2}\|\mathbf{x}_{\mathrm{n}} - \mathbf{x}_{0\mathrm{n}}\|$, where $\|\cdot\|$ denotes the Euclidean norm, and $\mathbf{x}_{\mathrm{n}}$ and $\mathbf{x}_{0\mathrm{n}}$ are normalized values of $\mathbf{x}$ and $\mathbf{x}_0$, respectively. This is to address the selection of an efficient set of material parameters once the Pareto solutions have been found, and is inspired by two previous exploration schemes. The first exploration scheme by Daulton et al. [102] focused on a trust region (in the error function space) surrounding a member of Ω that has maximum HV contribution. The second exploration scheme by Roussel et al. [101] carried out exploration in the neighborhood of the most recently observed point (in the parameter space) for minimizing the traveling distance by the optimization steps. Accordingly, the following acquisition function is formulated in this study to specify $\mathbf{x}^{s+1}$:

$$\alpha(\mathbf{x}) = \mathrm{HVI}(\boldsymbol{\mu}_{\hat{f}}(\mathbf{x}) - \beta\boldsymbol{\sigma}_{\hat{f}}(\mathbf{x}) | \Omega, \mathbf{f}_{\mathrm{R}}) \exp\left(-\frac{1}{2}\|\mathbf{x}_{\mathrm{n}} - \mathbf{x}_{0\mathrm{n}}\|\right). \tag{3.9}$$

Here $\mathbf{x}_{\mathrm{n}}$ and $\mathbf{x}_{0\mathrm{n}}$ are the normalized values to prevent dependence of the exponential function value on the parameter units. The components of $\mathbf{x}_{\mathrm{n}}$ and $\mathbf{x}_{0\mathrm{n}}$ are derived from the corresponding components

of $\mathbf{x}$, $\mathbf{x}_0$, $\mathbf{x}_\text{l}$, and $\mathbf{x}_\text{u}$ as

$$x_\text{n} = \frac{x - x_\text{l}}{x_\text{u} - x_\text{l}}, \quad x_{0\text{n}} = \frac{x_0 - x_\text{l}}{x_\text{u} - x_\text{l}}. \tag{3.10}$$

Therefore, the next parameter vector $\mathbf{x}^{s+1}$ is found by solving

$$\begin{aligned} \mathbf{x}^{s+1} &= \underset{\mathbf{x}}{\text{argmax}}\ \alpha(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in [\mathbf{x}_\text{l}, \mathbf{x}_\text{u}]. \end{aligned} \tag{3.11}$$

The search for $\mathbf{x}^{s+1}$ is likely to focus on the neighborhood of $\mathbf{x}_0$ that, as expected, can have a major contribution to the improvement in the current HV of Ω. This is referred to as proximal exploration. Note, however, that it is not necessary to restrict $\mathbf{x}^{s+1}$ to the neighborhood of $\mathbf{x}_0$ because a new parameter vector that is far from $\mathbf{x}_0$ can also be selected if it improves the current HV considerably.

The fourth step for improving Ω is to solve problem (3.11) using a proper optimization algorithm. Such an algorithm should avoid using the gradient information of $\alpha(\mathbf{x})$ for finding the search direction because it is difficult to evaluate the gradient of the upper confidence bound HVI in Eq. (3.9). Thus, a population-based method (e.g., GA) or a stochastic method (e.g., SA) can be used.

With $\mathbf{x}^{s+1}$, MOBO terminates if it satisfies one of the following termination conditions: (1) the number of iterations reaches a pre-specified upper limit s_max, and (2) $\mathbf{x}^{s+1}$ is identical to any available sample of parameters in $\mathcal{D}$, i.e., the current HV can no longer be improved. Otherwise, MOBO updates $\mathcal{D}$ and starts a new iteration.

Algorithm 3.1 summaries the procedure of using MOBO for solving the multi-objective identification problem.

Algorithm 3.1 Proposed MOBO for solving multi-objective identification problem

Specify $\mathbf{x}_\text{l}$, $\mathbf{x}_\text{u}$, s_max, N, β, $\mathbf{f}_\text{R}$, (FE) model for evaluation of $\mathbf{y}(\mathbf{x}) = [y_1(\mathbf{x}), \ldots, y_{\tilde{I}}(\mathbf{x})]$;
Generate $\{\mathbf{x}^k,\ k = 1, \ldots, N\}$ using Latin hypercube sampling [74];
$\mathcal{D} \leftarrow \emptyset$;
for $k = 1 : N$ **do**
    $\mathbf{f}^k \leftarrow \mathbf{y}(\mathbf{x}^k)$; ▷ Costly step
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^k, \mathbf{f}^k\}$;
end for
Sort Ω from $\mathcal{D}$;
Find $\mathbf{x}_0$ using Eqs. (3.5) and (3.6);
for $s = 2 : s_\text{max}$ **do**
    Construct $\hat{f}_1(\mathbf{x}), \ldots, \hat{f}_{\tilde{I}}(\mathbf{x})$ based on $\mathcal{D}$; see Section 2.3 of Chapter 2;
    Formulate $\alpha(\mathbf{x})$ as given in Eq. (3.9);
    Find $\mathbf{x}^{s+1}$ by maximizing $\alpha(\mathbf{x})$;
    if $\mathbf{x}^{s+1} \notin \mathcal{D}$ **then**
        $\mathbf{f}(\mathbf{x}^{s+1}) \leftarrow \mathbf{y}(\mathbf{x}^{s+1})$; ▷ Costly step
        $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^{s+1}, \mathbf{f}(\mathbf{x}^{s+1})\}$;
        Sort Ω from $\mathcal{D}$;
        Find $\mathbf{x}_0$ using Eqs. (3.5) and (3.6);
    else
        exit **for**
    end if
end for
return Ω and the corresponding vectors of material parameters.

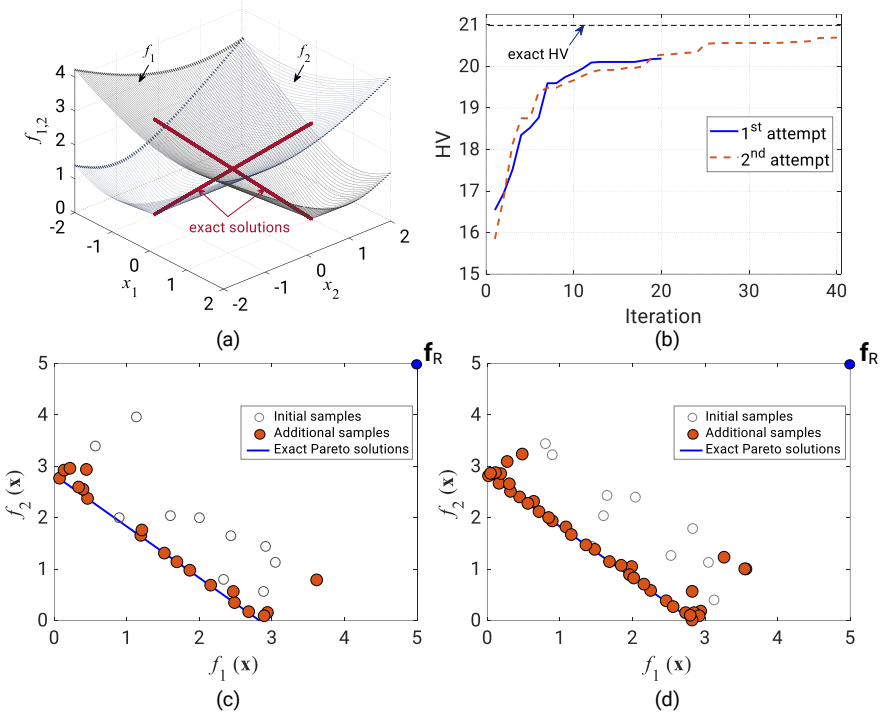

Figure 3.2: Solving the bi-objective minimization problem using MOBO. (a) surfaces of two objective functions and the exact solutions; (b) histories of HVs for two MOBO attempts; (c) and (d) initial and additional design points for the first and second MOBO attempts, respectively.

Table 3.1: Parameters for GA.

Parameter	Value
Population size	200
Maximum number of generations	50
Crossover fraction	65%
Elite transfer	2
Fitness function tolerance	10^{-12}

3.2.2 Test problem

The performance of the proposed MOBO is verified against a bi-objective minimization problem, which is stated as [101]

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & [f_1(\mathbf{x}), f_2(\mathbf{x})] \\ \text{subject to} \quad & x_i \in [-2, 2] \quad (i = 1, 2), \end{aligned} \tag{3.12}$$

where

$$f_1(\mathbf{x}) = \|\mathbf{x} - \mathbf{1}\|, \quad f_2(\mathbf{x}) = \|\mathbf{x} + \mathbf{1}\|, \tag{3.13}$$

and $\mathbf{1}$ denotes the column vector of ones.

Figure 3.2(a) depicts the exact Pareto-optimal solutions to problem (3.12). Their image in the objective function space shown in Fig. 3.2(c) is the line segment connecting $[f_1, f_2] = [0, 2\sqrt{2}]$ and $[2\sqrt{2}, 0]$.

MOBO is performed two times, each starts with a random set of ten sampling points. The numbers of iterations for the first and second MOBO attempts are limited at 20 and 40, respectively.

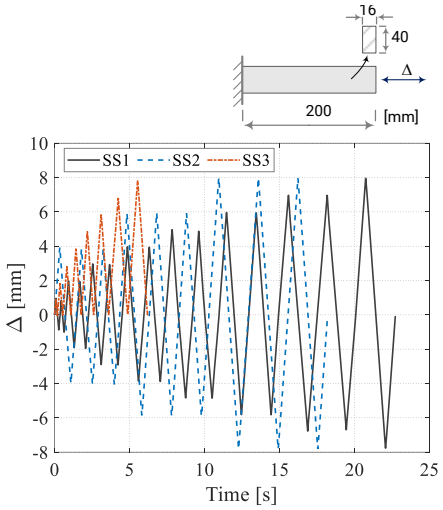

Figure 3.3: Example 1: Steel specimen and three loading histories for cyclic tests [103].

The assigned reference point $\mathbf{f}_{\mathrm{R}} = [5, 5]$ leads to an exact HV of 21 (i.e., $5 \times 5 - 0.5 \times 2\sqrt{2} \times 2\sqrt{2}$). Problem (3.11) is solved in each iteration of MOBO using GA whose parameters are listed in Table 3.1. The parameter β in Eq. (3.9) is assigned as 0.01 since we wish MOBO to focus on exploitation for improving HV. The effect of β on the convergence speed of MOBO may be of interest when fixing the initial dataset; however, this is not our focus here. The HV and additional sampling point after each MOBO iteration are recorded.

Figure 3.2(b) confirms that the HV values during the two MOBO attempts increase as much as possible and tend to converge to the exact HV as the number of MOBO iterations increases. As a result, the additional sampling points by MOBO shown in Figs. 3.2(c) and (d) well capture the exact solutions. These results indicate a good performance of the proposed MOBO in solving problem (3.12).

3.3 Example 1: A steel specimen

This section identifies the elastoplastic parameters to simulate the uniaxial cyclic behavior of a steel specimen in Fig. 3.3. The specimen was tested under three different static cyclic loading histories SS1, SS2, and SS3 by Yamada and Jiao [103]. Test results for the specimen consist of three experimental datasets of true stress σ and true strain ϵ corresponding to the three loading histories.

The true stress σ and true strain ϵ are derived as follows:

$$\sigma = (1 + \epsilon_{\mathrm{e}})\sigma_{\mathrm{e}}, \quad \epsilon = \log(1 + \epsilon_{\mathrm{e}}), \tag{3.14}$$

where the engineering stress σ_{e} is calculated by dividing the measured axial force by the initial cross-sectional area of the specimen, and the engineering strain ϵ_{e} is estimated by dividing the axial deformation of the specimen by its initial length.

Since it was axially loaded during the cyclic tests, the specimen is modeled using one Abaqus linear hexahedral element with reduced integration of type C3D8R [104]. For every time instant of interest, Abaqus directly extracts the true stress and true strain in the direction of loading at the

Table 3.2: Example 1: Material parameter intervals for the specimen.

Parameter	Lower bound	Upper bound
E [GPa]	205.94	–
ν	0.3	–
$\sigma_{y,0}$ [MPa]	250	260
Q_∞ [MPa]	10	100
b	5	25
C_1 [MPa]	2000	8000
γ_1	10	100

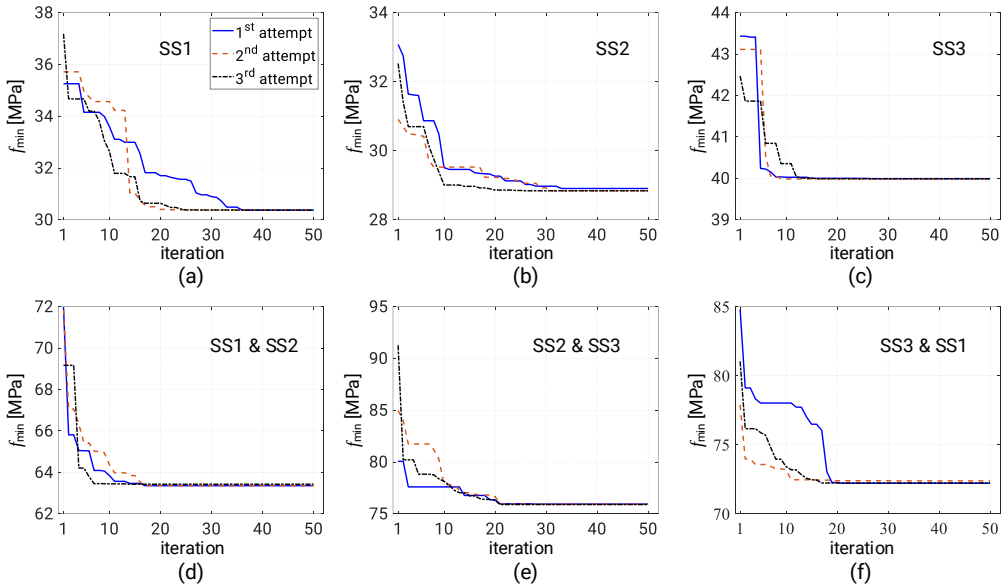

Figure 3.4: Example 1: Histories of SOBO with noise-free error function for different groups of experimental datasets. (a)–(f) groups 1–6, respectively.

integration point of the element. The nonlinear combined isotropic/kinematic hardening, detailed in Appendix A, is used as the material model that is characterized by seven parameters, namely $\mathbf{x} = [E, \nu, \sigma_{y,0}, Q_\infty, b, C_1, \gamma_1]$, where E is Young's modulus, ν is Poisson's ratio, $\sigma_{y,0}$ [MPa] is the initial yield stress, Q_∞ [MPa] and b are two parameters for isotropic hardening, and C_1 [MPa] and γ_1 are two parameters for nonlinear kinematic hardening. For identification, Young's modulus and Poisson's ratio of the specimen are respectively kept constant at $E = 205.94$ GPa and $\nu = 0.3$, where E is directly calibrated from the experiment results [103]. Thus, there are five plastic parameters to be identified for the specimen. The interval for each of these parameters is listed in Table 3.2, where the interval for $\sigma_{y,0}$ is derived from the experiment results [103] and those for Q_∞, b, C_1, and γ_1 are taken from many sources of the literature.

3.3.1 Identification results from SOBO

For noise-free error functions, we assign $\omega = 0$ to the datasets of SS1, SS2, and SS3. For noise-corrupted error functions, we set $\omega = 0.40$, 0.45, and 0.78 MPa for SS1, SS2, and SS3, respectively. Empirical calculations of these values from noise involved in each measurement of stress are detailed in Do and Ohsaki [105].

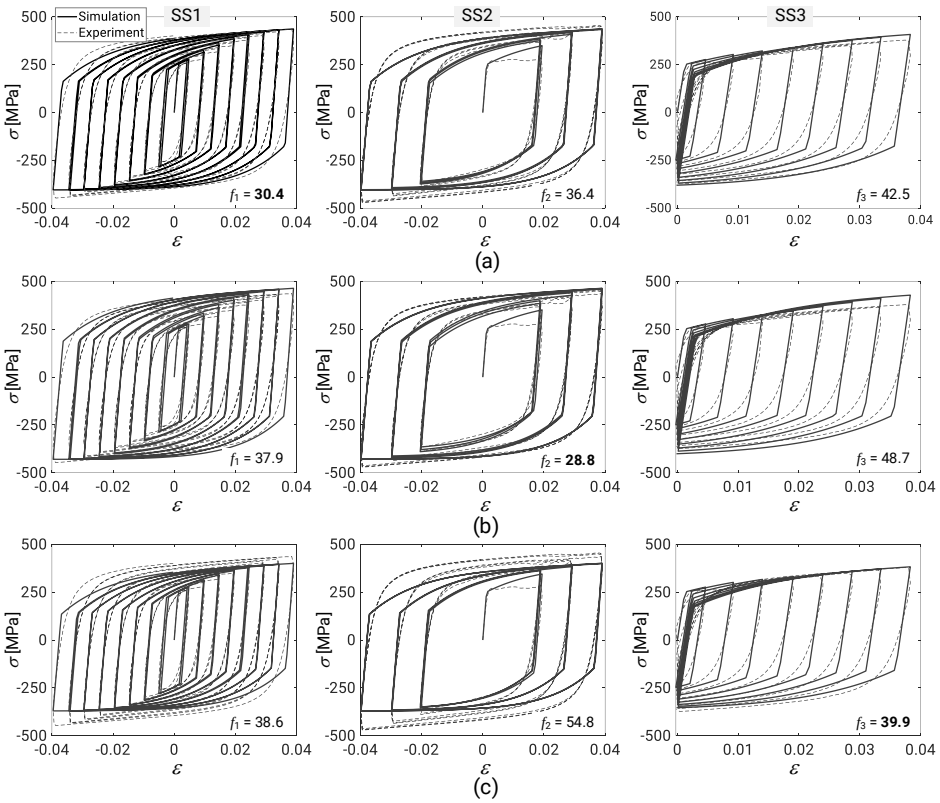

Figure 3.5: Example 1: Comparison of test data and model predictions for parameters identified from SOBO with noise-free error function and different groups of experimental datasets. (a)–(c) parameters from groups 1–3, respectively.

To investigate how experimental datasets used for identification affect the resulting parameters as well as their prediction ability, the three experimental datasets are classified into six groups for formulation of two single-objective identification problems, which are indexed as 1, 2, 3, 4, 5, and 6 corresponding to SS1, SS2, SS3, (SS1 & SS2), (SS2 & SS3), and (SS3 & SS1), respectively.

To further investigate the robustness of SOBO, we randomly generate three different training datasets for each group. Thus, three sets of parameters are found from each group corresponding to the three SOBO attempts. The initial training datasets of SOBO for each group consist of 50 samples of the parameters and the corresponding error functions obtained by performing FE analyses.

SOBO limits the number of its iterations at 50. Thus, the number of costly simulations required for SOBO of group 1, 2, or 3 is 100, and for that of group 4, 5, or 6 is 200. In each iteration, SOBO uses GA with a population size of 4000 for maximizing the acquisition functions. Here, a large population size is to increase the chance of finding the global optimizer of the acquisition function in each SOBO iteration that can mitigate the effect of GA's randomness on the performance of SOBO. Other parameters characterizing GA are given in Table 3.1.

Figure 3.4 shows the histories of three SOBO attempts of each group with noise-free error function, where $f_{\min}$ defined in Eq. (2.35) denotes the best value of the error function f in each iteration. Three SOBO attempts of each group tend to converge to an optimal error function value after 50 iterations, even though they start from different initial training datasets.

Figures 3.5 and 3.6 compare the measured and predicted $\sigma - \epsilon$ curves associated with the best

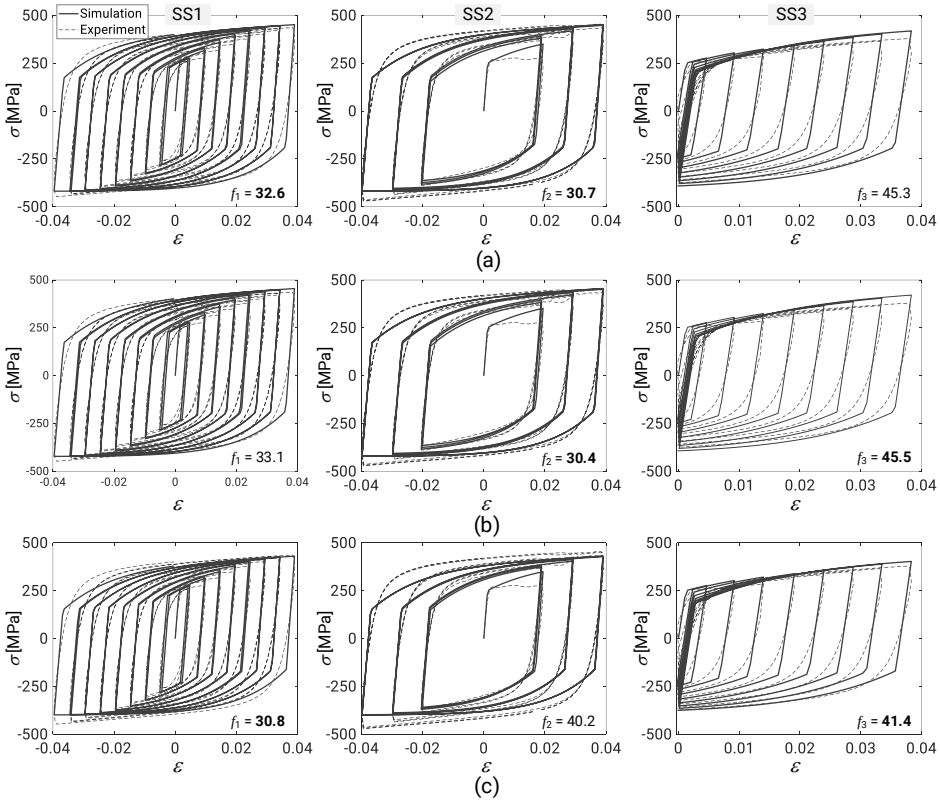

Figure 3.6: Example 1: Comparison of test data and model predictions for parameters identified from SOBO with noise-free error function and different groups of experimental datasets. (a)–(c) parameters from groups 4–6, respectively.

Table 3.3: Example 1: Comparison of the identification results obtained from SOBO with noise-free error function and different groups of experimental datasets.

Group	$\sigma_{y,0}$ [MPa]	Q_∞ [MPa]	b	C_1 [MPa]	γ_1	f_1 [MPa]	f_2 [MPa]	f_3 [MPa]
1	250.004	42.105	5.001	7999.997	69.188	**30.372**	36.370	42.499
2	250.010	70.283	5.015	7999.769	70.308	37.898	**28.840**	48.692
3	250.008	11.939	5.000	8000.000	72.634	38.635	54.834	**39.986**
4	250.011	56.630	5.005	7999.996	69.096	**32.627**	**30.732**	45.261
5	250.056	57.661	5.000	8000.000	67.594	33.092	**30.397**	**45.497**
6	250.002	34.289	5.003	8000.000	67.985	**30.827**	40.150	**41.389**

Table 3.4: Example 1: Comparison of identification results obtained from SOBO, GA, and PSO.

Group	$\sigma_{y,0}$ [MPa]	Q_∞ [MPa]	b	C_1 [MPa]	γ_1	f_1 [MPa]	f_2 [MPa]	f_3 [MPa]	No. simulations
1 (SOBO)	250.004	42.105	5.001	7999.997	69.188	**30.372**	36.370	42.499	100
1 (GA-1)	250.123	29.700	12.616	7397.003	61.045	**34.021**	43.787	43.110	100
1 (GA-2)	256.562	32.546	6.265	7409.248	65.632	**32.253**	44.282	39.747	200
1 (PSO-1)	251.419	41.499	7.598	6940.747	54.853	**33.010**	39.311	43.911	100
1 (PSO-2)	251.540	37.018	6.139	8000.000	64.631	**31.260**	37.088	43.217	200
4 (SOBO)	250.011	56.630	5.005	7999.996	69.096	**32.627**	**30.732**	45.261	200
4 (GA-1)	250.815	63.098	8.169	6440.088	59.086	**35.640**	**32.788**	49.728	200
4 (GA-2)	252.821	51.212	6.885	7536.697	65.882	**33.324**	**32.419**	47.323	400
4 (PSO-1)	250.000	55.367	6.836	7979.444	64.027	**34.491**	**30.918**	47.514	200
4 (PSO-2)	250.000	55.113	5.000	8000.000	72.489	**32.068**	**31.608**	44.957	400

Table 3.5: Example 1: Comparison of the identification results obtained from SOBO with noise-corrupted error function and different groups of experimental datasets.

Group	$\sigma_{y,0}$ [MPa]	Q_∞ [MPa]	b	C_1 [MPa]	γ_1	y_1 [MPa]	y_2 [MPa]	y_3 [MPa]
1	250.000	41.297	5.000	8000.000	70.058	**30.368**	36.913	42.373
2	250.000	69.963	5.000	8000.000	74.521	37.422	**29.094**	48.642
3	250.000	10.000	9.754	8000.000	68.307	38.929	54.451	**40.177**
4	250.000	56.210	5.000	8000.000	69.732	**32.453**	**30.905**	45.161
5	250.000	49.819	5.000	8000.000	69.491	31.045	**33.049**	**43.837**
6	250.000	34.486	5.000	8000.000	68.391	**30.804**	40.112	**41.408**

set of parameters obtained from each group, where f_1, f_2, and f_3 in the figures denote the error function values at the identified parameters corresponding to SS1, SS2, and SS3, respectively, and the boldface value indicates the experimental dataset used for identification. It is observed that the prediction ability of the obtained parameters is strongly affected by the loading history (i.e., dataset) used to formulate the identification problem. The bias toward a specific loading history is exhibited in Fig. 3.5 (see along each column) as the error function corresponding to the loading history used for identification is smaller than those predicted by the parameters identified from other loading histories. The parameters identified from a specific loading history (e.g., SS3) may lead to large errors in prediction of $\sigma - \epsilon$ curves associated with other loading histories (e.g., SS1 and SS2). These errors can be reduced by using the experimental datasets from two loading histories for identification as observed in Fig. 3.6.

Table 3.3 lists the identified parameters from each group with noise-free error function, where the boldface value indicates the error function corresponding to the experimental dataset used for identification. There is no major difference in values of $\sigma_{y,0}$, b, C_1, and γ_1 among the groups. However, Q_∞ is affected by the datasets.

We also compare the identification results by SOBO of groups 1 and 4 with those by minimizing the corresponding error functions using GA and PSO algorithms. For each group (i.e., 1 or 4), GA and PSO are performed two times, namely GA-1, GA-2, PSO-1, and PSO-2. Each algorithm has a population of 20 individuals (or particles) and the numbers of iterations for GA(PSO)-1 and GA(PSO)-2 are limited at five and ten, respectively. Thus, the numbers of simulations required for GA(PSO)-1 and GA(PSO)-2 are 100 and 200 when working on group 1, and are 200 and 400 when working on group 4, respectively. To enable fair comparisons with SOBO, other parameters for GA and PSO follow default settings of the MATLAB R2018a Global Optimization Toolbox [106].

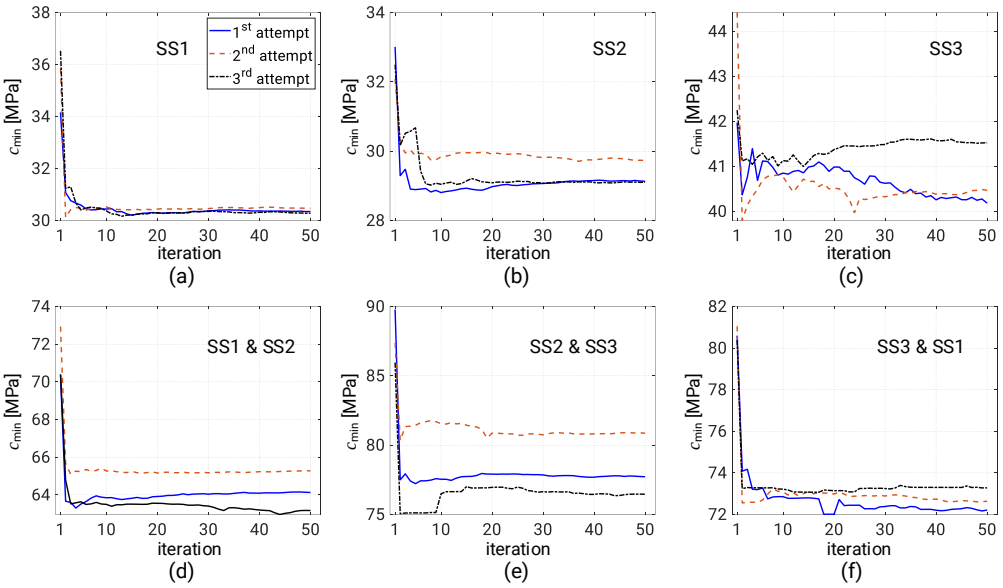

Figure 3.7: Example 1: Histories of SOBO with noise-corrupted error function for different groups of experimental datasets. (a)–(f) groups 1–6, respectively.

Comparison results in Table 3.4 show that SOBO outperforms GA and PSO in terms of the minimized error function value as well as the prediction ability of identified parameters when expending the same number of costly simulations. Interestingly, GA and PSO should double the number of simulations to offer performance approaching that of SOBO.

Figure 3.7 shows the histories of three SOBO attempts of each group with noise-corrupted error function, where $c_{\min}$ defined in Eq. (2.37) is the representative of $f_{\min}$. In the presence of noise, three SOBO attempts of each group result in different error function values after 50 iterations. However, there is still a good agreement between the predicted $\sigma - \epsilon$ curves at the best set of the parameters identified from each group and the experimental ones, as shown in Figs. 3.8 and 3.9, where y_1, y_2, and y_3 are the noise-free error function at the identified parameters corresponding to SS1, SS2, and SS3, respectively. It is also confirmed in Figs. 3.8 and 3.9 that noise does not affect the bias toward a set of experimental measures that is enormous when using a specific loading history for identification and can be reduced when using a pair of loading histories for identification.

Table 3.5 reports the best identified material parameters from each group with noise-corrupted error functions, where the boldface value indicates the noise-free error function corresponding to the experimental dataset used for identification. The values of $\sigma_{\mathrm{y},0}$, b, C_1, and γ_1 among the groups are similar, while Q_∞ again varies across the groups as it plays an important role in mitigating the bias toward a set of experimental results.

3.3.2 Identification results from MOBO

To investigate the effect of experimental datasets used to formulate the multi-objective inverse problem on the resulting parameters, we also classify the three experimental datasets of the specimen into three groups, namely 7, 8, and 9, corresponding to (SS1 & SS2), (SS2 & SS3), and (SS3 & SS1), respectively. Problem (3.3) is formulated for each group with two objectives that are the error functions evaluated using Eq. (3.1) for the measured and simulated $\sigma - \epsilon$ curves of individual tests.

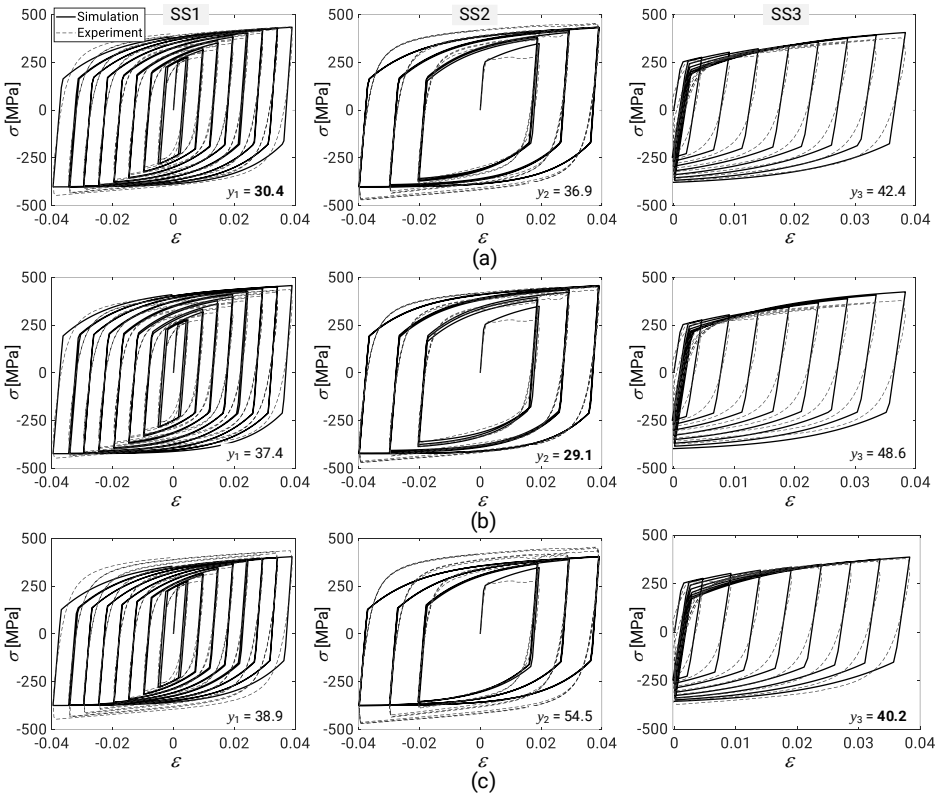

Figure 3.8: Example 1: Comparison of test data and model predictions for parameters identified from SOBO with noise-corrupted error function and different groups of experimental datasets. (a)–(c) parameters from groups 1–3, respectively.

To examine the dependence of MOBO on the training dataset, ten different training datasets are generated for each group. Each of these training datasets has 50 random samples of parameters and the corresponding error function values. Thus, a total of ten Pareto fronts of material parameters are found from each group. The best Pareto front of each group corresponds to the largest HV value among ten values associated with the ten Pareto fronts. To further demonstrate their prediction ability, we use the best and second-best compromise solutions on the best Pareto front of each group as input to the FE model for evaluating the error function associated with the loading history not used for identification.

We limit the number of MOBO iterations at 50. Thus, the maximum number of simulations required for each MOBO attempt is 200 (i.e., 100 for generating the initial training dataset and maximum $50 \times 2 = 100$ for processing MOBO iterations). The reference point $\mathbf{f}_{\mathrm{R}}$ and the parameter β in Eq. (3.9) are assigned as $\mathbf{f}_{\mathrm{R}} = [150, 150]$ MPa and 0.01, respectively. We solve problem (3.11) in each MOBO iteration using GA whose parameters are the same as those for the GA embedded in SOBO.

For comparison, we perform NSGA-II ten times, each treats the variables as real numbers and requires 20 individuals and five generations (i.e., $20 \times 2 \times 5 = 200$ simulation calls) for finding approximate Pareto fronts for each group. The solutions from NSGA-II are baselines to assess the quality of the Pareto fronts by SOBO. Moreover, we carry out ParEGO [107], which is an extension of the single-objective efficient global optimization algorithm to solving multi-objective optimization

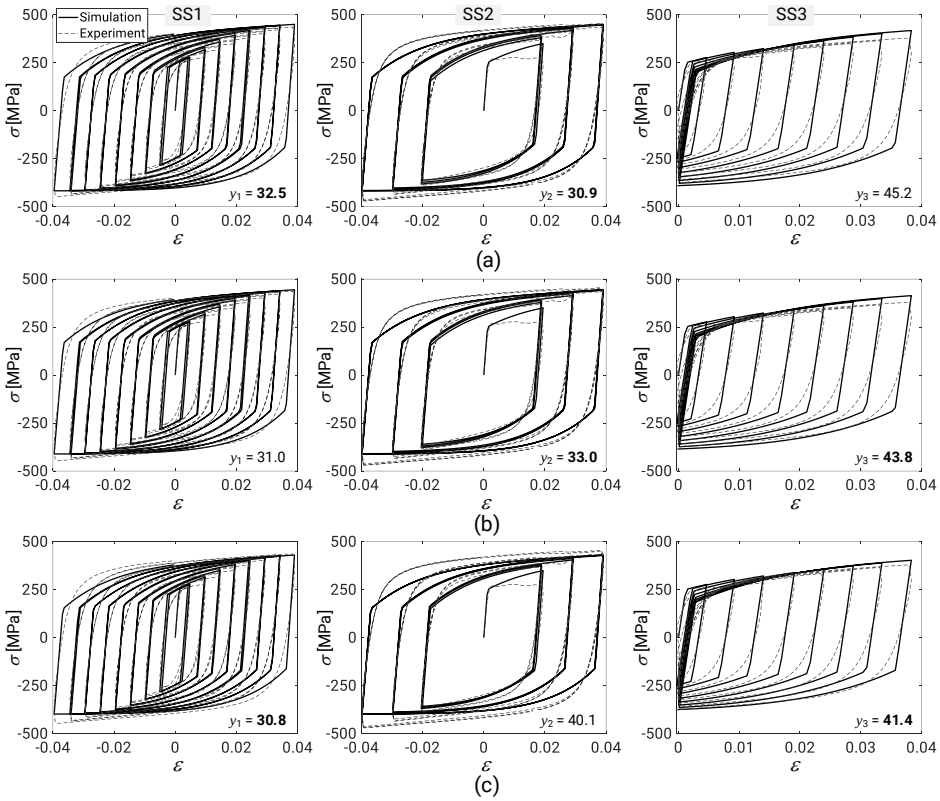

Figure 3.9: Example 1: Comparison of test data and model predictions for parameters identified from SOBO with noise-corrupted error function and different groups of experimental datasets. (a)–(c) parameters from groups 4–6, respectively.

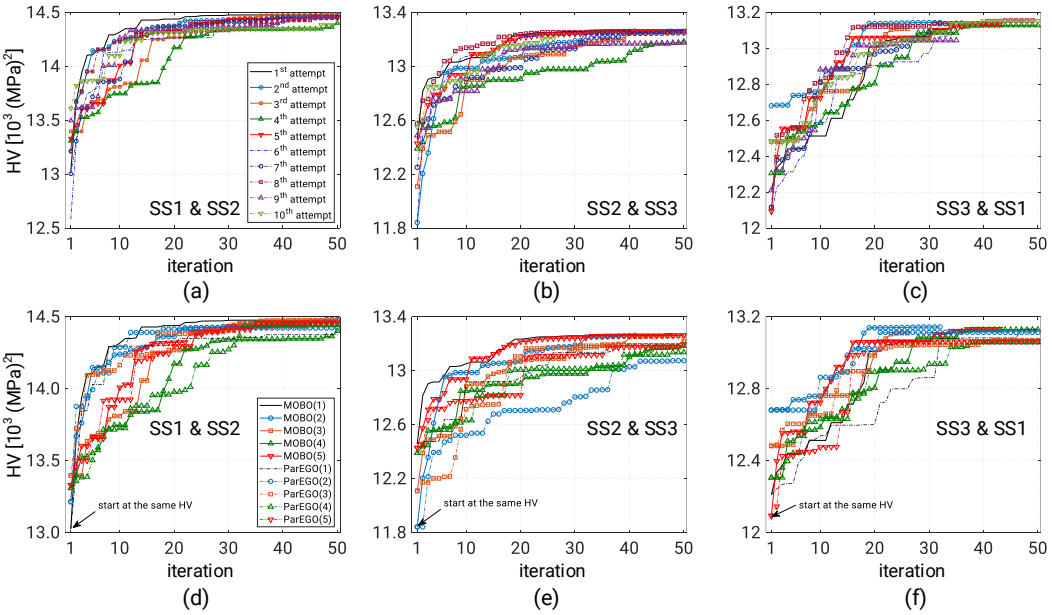

Figure 3.10: Example 1: Histories of HVs for different groups of experimental datasets. (a)–(c) ten MOBO attempts of groups 7–9, respectively; (d)–(f) the first five MOBO attempts and five ParEGO attempts associated with groups 7–9, respectively.

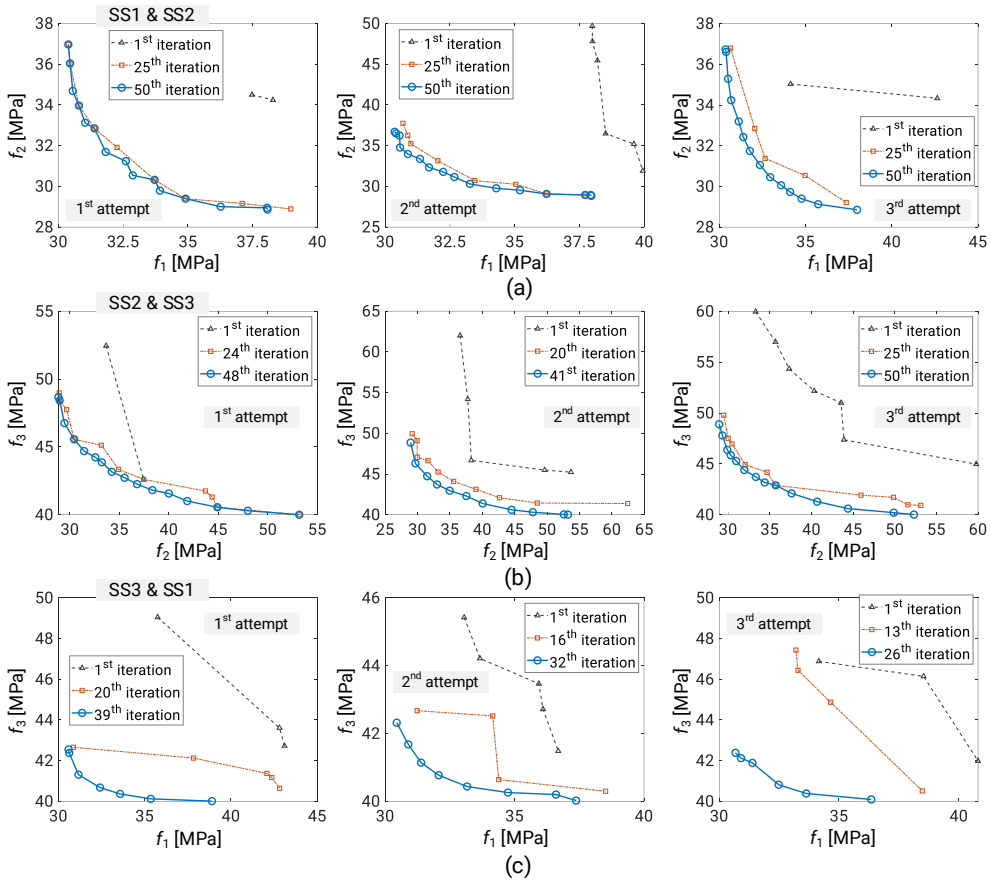

Figure 3.11: Example 1: Evolution of solutions during the first three MOBO attempts for different groups of experimental datasets. (a)–(c) groups 7–9, respectively.

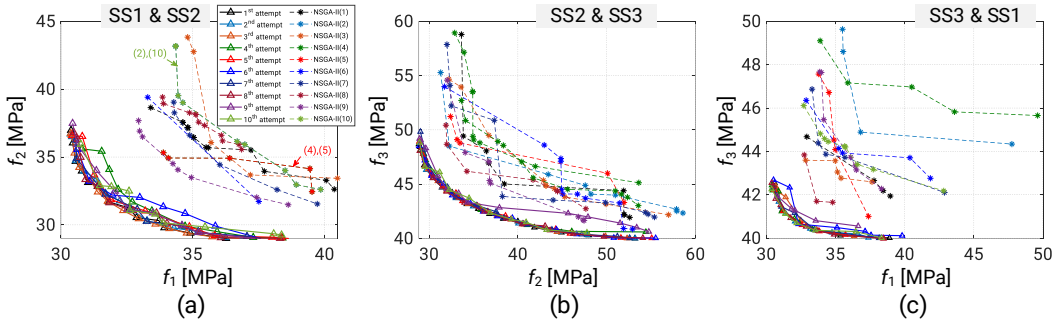

Figure 3.12: Example 1: Comparison of solutions by ten MOBO attempts and by ten NSGA-II attempts for different groups of experimental datasets. (a)–(c) groups 7–9, respectively.

problems, five times, each is characterized by 50 iterations using the same training dataset as that at the beginning of each of the first five MOBO attempts. This is to compare the largest HV values associated with the final Pareto fronts from MOBO and ParEGO when they start at the same training dataset.

Figure 3.10 shows the histories of HV from the ten MOBO attempts of each group as well as the comparison between the HV histories from the first five MOBO attempts and those from the corresponding five ParEGO attempts. Although starting at different initial values, the HVs from

Table 3.6: Example 1: Optimized HVs associated with ten MOBO attempts for different groups of experimental datasets [$\times 10^3$ (MPa)^2].

Attempt	Group 7	Group 8	Group 9
1	14.477	13.275	13.129
2	14.480	13.261	13.143
3	14.479	13.262	13.109
4	14.402	13.183	13.128
5	14.457	13.262	13.132
6	14.444	13.265	13.123
7	14.454	13.248	13.140
8	14.446	13.259	13.154
9	14.454	13.175	13.044
10	14.383	13.206	13.152

Table 3.7: Example 1: Comparison of the identification results obtained from MOBO for different groups of experimental datasets. (1) and (2) indicate the best and second-best compromise solutions, respectively.

Group	$\sigma_{y,0}$ [MPa]	Q_∞ [MPa]	b	C_1 [MPa]	γ_1	f_1 [MPa]	f_2 [MPa]	f_3 [MPa]
7 (1)	250.373	58.208	5.058	7987.986	69.641	**33.271**	**30.298**	45.902
8 (1)	250.004	46.273	5.005	7982.653	65.429	30.861	**34.236**	**43.168**
9 (1)	250.346	39.932	5.013	7999.983	67.686	**30.447**	37.028	**42.319**
7 (2)	251.806	54.956	5.003	7855.806	68.741	**32.669**	**31.147**	45.669
8 (2)	250.005	40.323	5.025	7999.934	65.076	30.599	**36.794**	**42.257**
9 (2)	250.093	35.395	5.186	7969.030	65.229	**30.895**	39.244	**41.680**

each group tend to converge to the similar value as MOBO terminates. Moreover, the final HV values by most of the MOBO attempts are better than those by the ParEGO attempts when they start at the same training dataset. The optimized HV values associated with ten MOBO attempts of each group are reported in Table 3.6.

Figure 3.11 illustrates the evolution of approximate Pareto-optimal solutions during the first three MOBO attempts of each group. As observed, MOBO considerably improves the solution quality. Figure 3.12 shows the similarity in shapes of the Pareto fronts at the last iterations of ten MOBO attempts of each group, regardless of their difference at the very first iterations (see the first iteration along each row of Fig. 3.11 for the first three MOBO attempts of each group). This observation is consistent with the convergence of HV in Fig. 3.10. Figure 3.12 also confirms that the optimization results by all MOBO attempts of each group outperform those by the corresponding ten NSGA-II attempts, even though the number of simulation calls required for MOBO does not exceed that required for NSGA-II.

Table 3.7 provides the best (1) and second-best (2) compromise solutions selected from the best Pareto front of each group. It also provides the error function values associated with SS1, SS2, and SS3, namely f_1, f_2, and f_3, respectively, where the boldface error value corresponds to the experimental dataset used for identification. We see that there is no major difference in the values of $\sigma_{y,0}$, b, C_1, and γ_1 for the best and second-best compromise solutions among those obtained from the three groups. However, Q_∞ is dataset-dependent.

Figure 3.13 compares the measured and simulated $\sigma - \epsilon$ curves from each loading history using the best compromise solution of each group listed in Table 3.7 for the simulation. Although the yield plateau observed in test SS2 cannot be captured because of the nature of the nonlinear combined

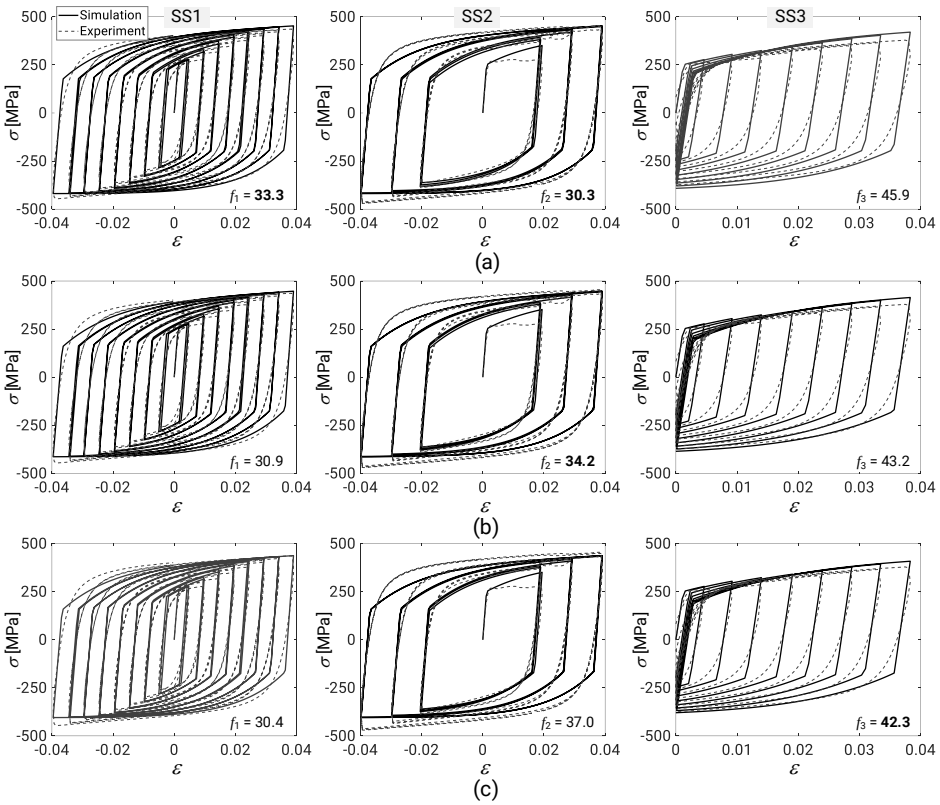

Figure 3.13: Example 1: Comparison of test data and model predictions for the best compromise solutions of parameters identified from MOBO with different groups of experimental datasets. (a)–(c) parameters from groups 7–9, respectively.

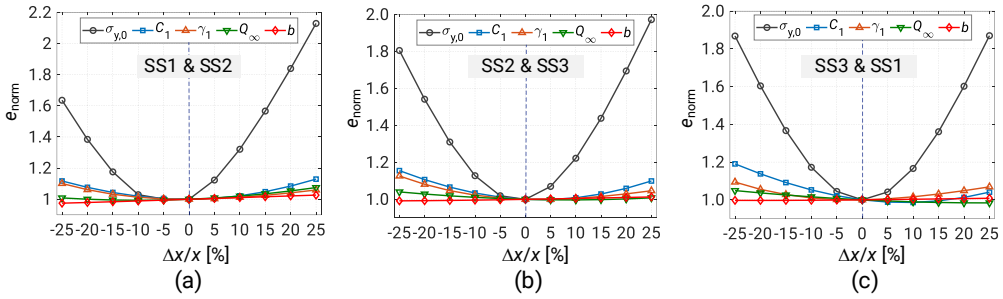

Figure 3.14: Example 1: Average sensitivity results over different groups of experimental datasets. (a)–(c) parameters from groups 7–9, respectively.

isotropic/kinematic hardening model, the identified parameters can reproduce the $\sigma - \epsilon$ curves for the loading histories not used for identification with good accuracy.

To further rank the importance of the identified material parameters, we assess the sensitivity of error functions to the variation of each parameter, indicated by the ratio $\Delta x/x$, in the neighborhood of the best compromise value (in the parameter space) as shown in Fig. 3.14 while keeping other parameters constant. The sensitivity of each error function is represented by e_{norm} that is the ratio of the error function of interest to that associated with the best compromise solution. Thus, the sensitivity result at a particular value of $\Delta x/x$ shown in Fig. 3.14 is the average of three e_{norm} values corresponding to the three error functions f_1, f_2, and f_3. It is found that the error functions are

Table 3.8: Example 1: Dataset-specific indexes for different sets of identification results evaluated using the experimental results from tests SS1, SS2, and SS3 of the specimen [MPa].

Set of results	ψ_1	ψ_2	ψ_3	η
Single-objective (groups $1-3$)	8.263	25.994	8.706	14.321
Weighted-sum (groups $4-6$)	2.265	9.418	4.109	5.264
Multi-objective (1) (groups $7-9$)	2.823	6.730	3.583	4.379
Multi-objective (2) (groups $7-9$)	2.070	8.096	3.990	4.719

most and least sensitive to the variations of $\sigma_{\mathrm{y},0}$ and b, respectively.

3.3.3 A dataset-specific index

Using the identification results in Tables 3.3 and 3.7, we now wish to compare the dataset-specific bias levels for different sets of identification results obtained from different inverse problem formulations. Here we consider three different types of identification results from three formulations, namely single-objective (groups 1–3), weighted-sum (groups 4–6), and multi-objective (groups 7–9) formulations. The multi-objective formulation offers two sets of identification results, i.e., the best compromise (1) and second-best (2) compromise sets, while each of the remaining two formulations provides one set of identification results. To enable a rigorous comparison, we propose in the following a dataset-specific index η for each set of identification results.

Let $P \geq 2$ and $J \geq 2$, respectively, denote the number of parameter sets in each set of identification results and the number of tests in consideration. Let $f_{j,p}$ with $j \in \{1,\ldots,J\}$ and $p \in \{1,\ldots,P\}$ indicate the j-th error corresponding to the p-th parameter set. Thus, each set of identification results in Tables 3.3 and 3.7 reads $J = 3$ and $P = 3$ while the best compromise set in Tables 3.7, for example, reads $f_{1,1} = \mathbf{33.271}$, $f_{1,2} = 30.861$, $f_{1,3} = \mathbf{30.447}$ MPa, and so on. For the j-th test of each set of identification results, we define

$$\psi_j = f_j^{\max} - f_j^{\min} \quad (j = 1,\ldots,J), \tag{3.15}$$

where

$$f_j^{\min} = \min\{f_{j,1},\ldots,f_{j,P}\}, \quad f_j^{\max} = \max\{f_{j,1},\ldots,f_{j,P}\}. \tag{3.16}$$

Here $f_j^{\min}$, due to the dataset-specific bias, is a minimized value of f_j (i.e., a boldface value in Tables 3.3 and 3.7). That means $f_j^{\min}$ is found by minimizing the error function formulated from the experimental results of test j. Meanwhile, $f_j^{\max}$ in most cases corresponds to a parameter set that is found by minimizing the error function formulated from the experimental results of other tests rather than test j. For example, the set of identification results of groups 1–3 in Table 3.3 has $f_1^{\min} = \mathbf{30.372}$, $f_2^{\min} = \mathbf{28.840}$, $f_3^{\min} = \mathbf{39.986}$, $f_1^{\max} = 38.635$, $f_2^{\max} = 54.834$, and $f_3^{\max} = 48.692$ MPa. Thus, $f_j^{\min}$ and $f_j^{\max}$ represent the quality of the solution to the inverse problem and its prediction performance, respectively. The larger the difference of $f_j^{\max}$ and $f_j^{\min}$ (i.e., ψ_j), the higher the dataset-specific bias level observed on the j-th test. Without loss of generality, we further assume that the errors of the tests in consideration have the same unit. Normalization can be used when we have different units of the errors. The following dataset-specific index η is defined for each

set of identification results based on a total of J tests of consideration:

$$\eta = \frac{1}{J}\sum_{j=1}^{J}\psi_j. \tag{3.17}$$

As a result, the set of identification results with the smallest value of η has the lowest level of dataset-specific bias.

Table 3.8 provides the η index value computed for each set of identification results listed in Tables 3.3 and 3.7, where ψ_1, ψ_2, and ψ_3 are evaluated using tests SS1, SS2, and SS3, respectively. The η values associated with the two sets of identification results from the multi-objective formulation are better than those corresponding to the single-objective and weighted-sum formulations. The single-objective formulation shows the largest value of bias level. These results confirm that the best and second-best compromise solutions obtained from the multi-objective formulation of the inverse problem of the specimen can reduce the dataset-specific bias. Thus, the experimental measures from various loading histories should be used for parameter identification simultaneously.

It is worth noting that the index η can also be evaluated for sets of identification results using the error function values of tests not used in parameter identification. In this case, η represents the dispersion of predictions by the associated parameter sets. A small value of η would indicate the prediction ability of the parameters on the tests being considered.

3.4 Example 2: A bi-material cantilever

In this section, the elastoplastic parameters are identified for a steel cantilever in Fig. 3.15 that was tested under three different static cyclic loading histories RH1, RH2, and RH3 by Yamada and Jiao [103]. The cantilever is a built-up wide-flange beam H-244 $\times$ 175 $\times$ 7 $\times$ 11 of two different materials in which the flange and web have the same Young's modulus, but different plastic material parameters. During the cyclic tests, the left end of the cantilever was fixed while forced vertical displacement was applied at the right end. The deflection angle θ of the cantilever is defined as the ratio of the vertical tip displacement Δ mm to the beam length $L = 800$ mm, i.e., $\theta = \Delta/L$. The test result for the each loading history consists of the bending reaction moment M at the cantilever support and the associated deflection angle θ.

For identification, Young's modulus and Poisson's ratio for the web and flange are fixed at $E = 175.05$ GPa, which is directly calibrated from the experiment, and $\nu = 0.3$, respectively. Thus, a total of ten parameters (i.e., five for the web and five for the flange) are identified for the cantilever. The interval for each parameter is provided in Table 3.9.

The cantilever is modeled using Abaqus [104] that generates a fine mesh consisting of 4960 nodes and 3510 linear hexahedral elements of type C3D8 as shown in Fig. 3.15. The maximum increment size for each loading history is set as 0.01 s. The numerical values of M and Δ at the time instants of interest are directly extracted from the Abaqus model.

For noise-free error functions, $\omega = 0$ is assigned to all the datasets. For noise-corrupted error functions, $\omega = 0.1$, 0.2, and 0.1 kNm are assigned to the datasets of RH1, RH2, and RH3, respectively. Empirical calculations of these values from noise involved in each measurement of bending moment are detailed in Do and Ohsaki [105].

The experimental datasets from the three loading histories are also classified into a total of six

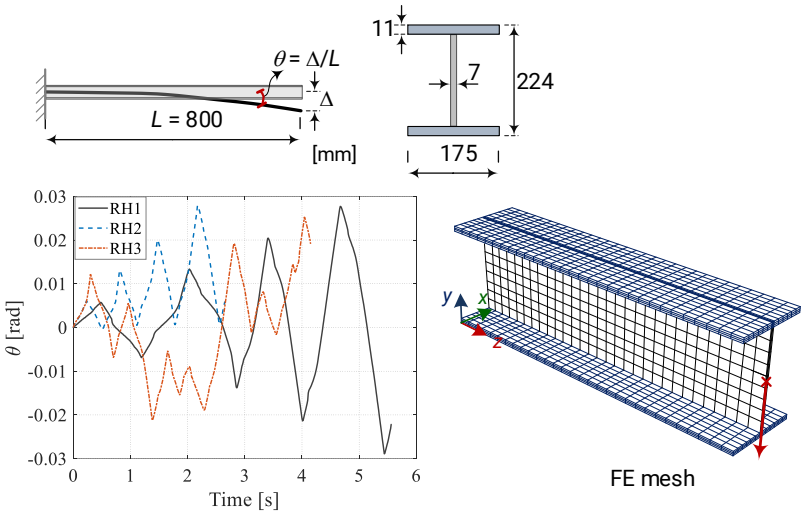

Figure 3.15: Example 2: Bi-material cantilever, its FE mesh, and three loading histories for cyclic tests [103].

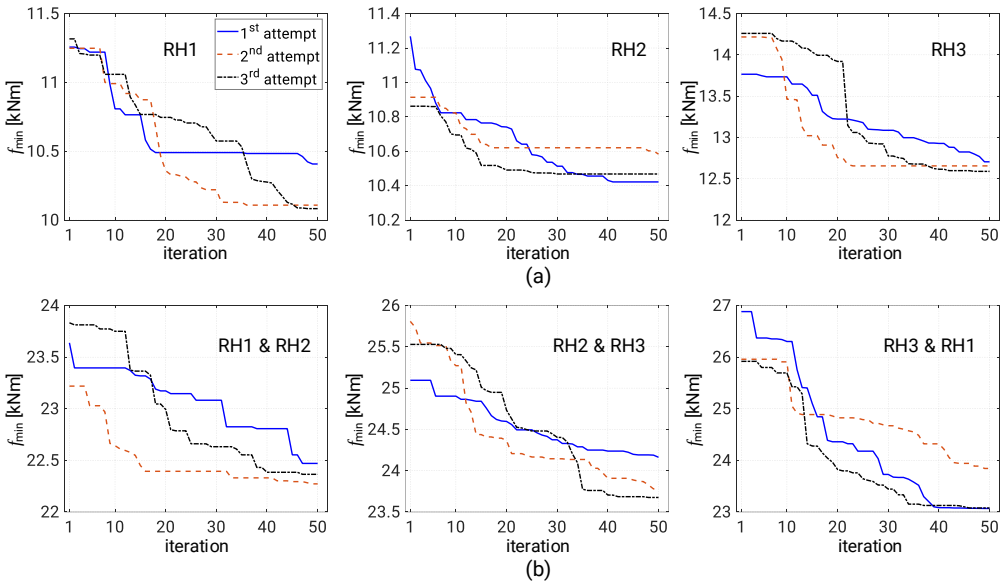

Figure 3.16: Example 2: Histories of SOBO with noise-free error function for different groups of experimental datasets. (a)–(f) groups 1–6, respectively.

Table 3.9: Example 2: Material parameter intervals for the cantilever.

	Parameter	Lower bound	Upper bound
Web	E [GPa]	175.05	–
	ν	0.3	–
	$\sigma_{y,0}$ [MPa]	300	340
	Q_∞ [MPa]	10	100
	b	5	25
	C_1 [MPa]	2000	8000
	γ_1	10	100
Flange	E [GPa]	175.05	–
	ν	0.3	–
	$\sigma_{y,0}$ [MPa]	270	290
	Q_∞ [MPa]	10	100
	b	5	25
	C_1 [MPa]	2000	8000
	γ_1	10	100

Table 3.10: Example 2: Material parameters and error functions obtained from SOBO with noise-free error function and different groups of experimental datasets.

	Parameter	Group					
		1	2	3	4	5	6
Web	$\sigma_{y,0}$ [MPa]	339.957	300.049	339.924	336.672	330.990	339.880
	Q_∞ [MPa]	99.890	99.827	99.923	99.874	99.972	99.974
	b	24.946	24.910	24.952	24.761	24.809	24.869
	C_1 [MPa]	7160.521	7968.280	3596.024	7766.642	7999.973	7682.000
	γ_1	99.622	10.178	99.967	99.999	69.775	99.577
Flange	$\sigma_{y,0}$ [MPa]	271.262	270.004	270.047	270.008	270.001	270.061
	Q_∞ [MPa]	10.065	10.135	10.000	10.140	10.107	10.001
	b	5.008	5.120	5.047	5.534	5.092	5.041
	C_1 [MPa]	7999.840	5170.983	6636.794	6537.440	5140.712	7998.996
	γ_1	66.632	62.092	71.565	60.903	69.258	84.763
Error	f_1 [kNm]	**10.083**	14.166	11.469	**10.781**	12.760	**10.107**
	f_2 [kNm]	12.836	**10.422**	11.377	**11.491**	**10.794**	12.380
	f_3 [kNm]	13.601	14.060	**12.590**	12.959	**12.881**	**12.961**

groups indexed as 1, 2, 3, 4, 5, and 6, corresponding to RH1, RH2, RH3, (RH1 & RH2), (RH2 & RH3), and (RH3 & RH1), respectively. For each value of ω, three different initial training datasets are created for each group to investigate the performance of SOBO. Each training dataset is constructed by randomly generating 100 samples of the parameters and performing FE analysis for each generated sample to evaluate the corresponding error function. The remaining steps of the identification process using the training datasets from each group are identical to those performed in Section 3.3.

3.4.1 Identification results from SOBO

Figure 3.16 shows the histories of three SOBO attempts for three different initial training datasets of each group with noise-free error function. Although SOBO with different initial training datasets cannot provide a unique set of the parameters after 50 iterations, it considerably reduces the error function value as it terminates. This result also poses a question for future research on how to find

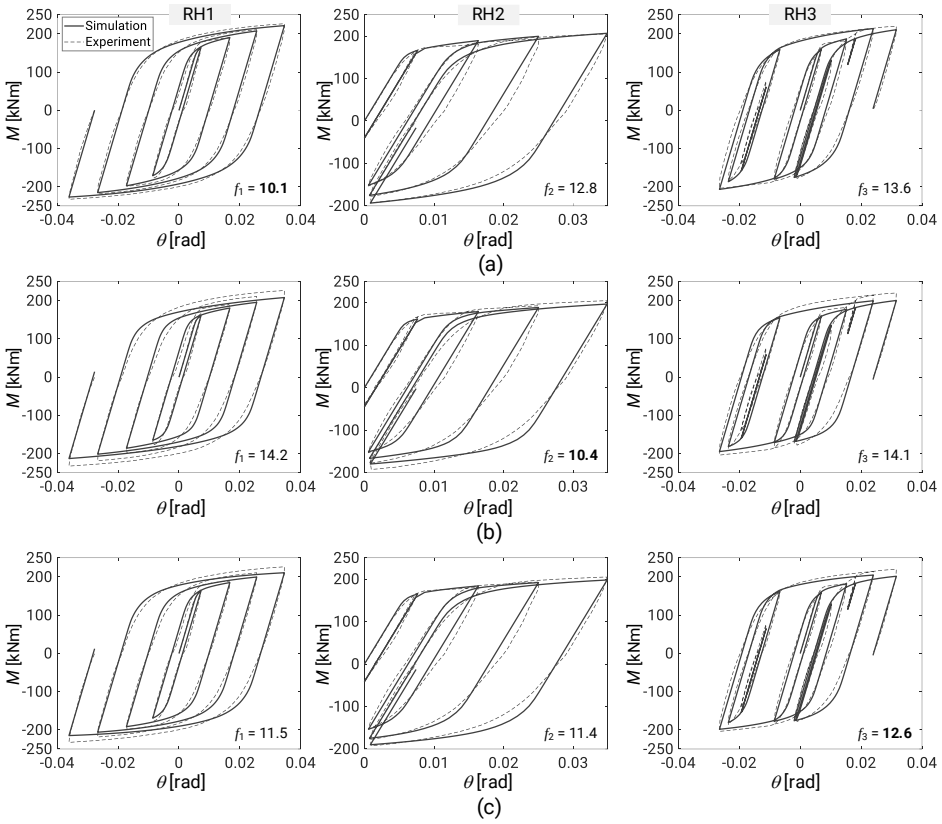

Figure 3.17: Example 2: Comparison of test data and model predictions for parameters identified from SOBO with noise-free error function and different groups of experimental datasets. (a)–(c) parameters from groups 1–3, respectively.

an optimal experimental design of the initial training dataset. In other words, how do we select the initial training dataset to minimize the variation of the inferred parameters as observed?

Table 3.10 provides the best set of identified parameters that gives the smallest error function value among three values obtained from each group. Figures 3.17 and 3.18 compare the predicted $M - \theta$ curves at the parameters identified from each group listed in Table 3.10 and the measured curves, where f_1, f_2, and f_3 are the error function values corresponding to RH1, RH2, and RH3, respectively. As expected, all predicted $M - \theta$ curves associated with the loading histories not used for identification are in good agreement with the measured curves. Results in Table 3.10 and Fig. 3.17 also show that the dataset-specific bias is considerable when the experimental dataset from a single loading history is used for identification. This bias can be mitigated when using the experimental datasets from two loading histories as illustrated in Fig. 3.18.

Figure 3.19 shows the histories of three SOBO attempts of each group with noise-corrupted error function. Under the effect of noise, $c_{\min}$ considerably fluctuates in the very first iterations of SOBO and becomes stable when approaching the 50th iteration. The values of $c_{\min}$ in some of the first iterations are less than those at the last iterations because $c_{\min}$ depends on the variance of the GP model that is updated after new sampling points are added. Regardless of the difference in three SOBO histories from each group, the identified parameters from all groups still arrive at a good agreement between the predicted and measured $M - \theta$ curves, as shown in Figs. 3.20 and

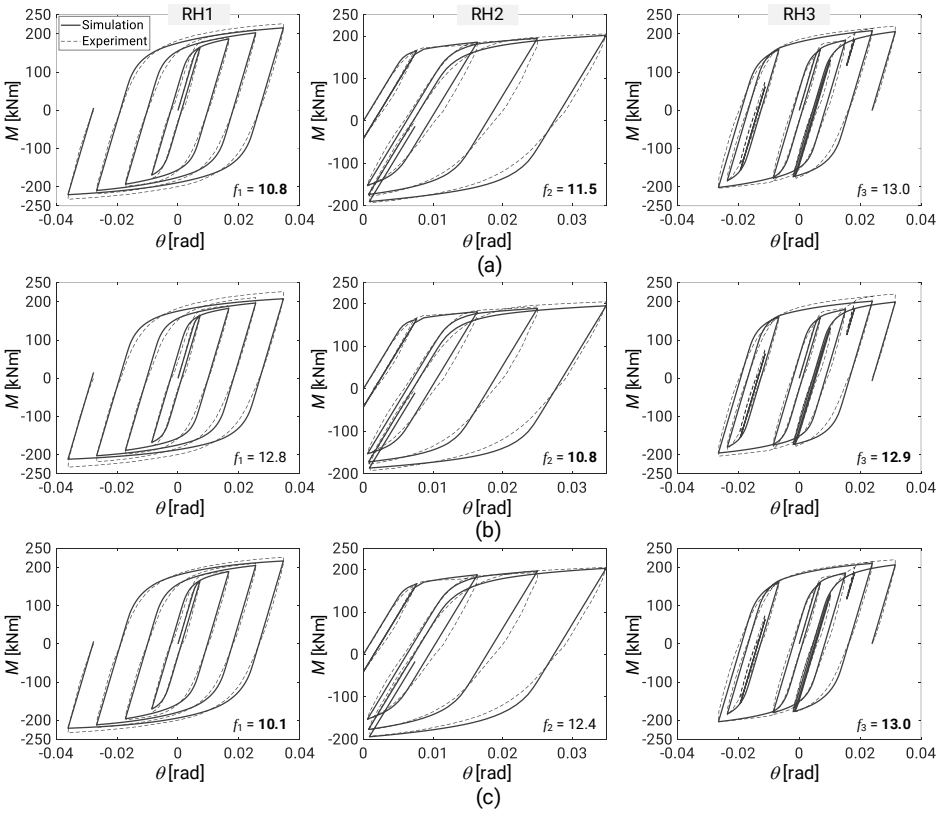

Figure 3.18: Example 2: Comparison of test data and model predictions for parameters identified from SOBO with noise-free error function and different groups of experimental datasets. (a)–(c) parameters from groups 4–6, respectively.

Table 3.11: Example 2: Material parameters and error functions obtained from SOBO with noise-corrupted error function and different groups of experimental datasets.

	Parameter	Group					
		1	2	3	4	5	6
Web	$\sigma_{y,0}$ [MPa]	339.988	339.952	339.986	339.981	339.851	339.998
	Q_∞ [MPa]	99.992	99.95	99.909	99.938	99.916	99.998
	b	24.988	5.034	24.928	24.910	24.983	24.999
	C_1 [MPa]	7999.962	7999.989	2000.104	7999.979	7764.107	7999.998
	γ_1	99.999	10.010	99.968	99.836	10.070	99.998
Flange	$\sigma_{y,0}$ [MPa]	271.038	270.02	270	270.024	270.055	270.001
	Q_∞ [MPa]	10.004	10.044	10.012	10.154	10.019	10.000
	b	5.011	11.312	5.033	8.434	5.067	5.001
	C_1 [MPa]	7999.951	4242.334	7388.138	6432.982	5361.751	7999.998
	γ_1	76.819	59.440	79.299	57.267	90.055	86.975
Error	f_1 [kNm]	**10.006**	14.728	11.242	**10.650**	10.615	**10.115**
	f_2 [kNm]	12.729	**10.592**	11.659	**11.665**	**10.746**	12.382
	f_3 [kNm]	13.311	14.369	**12.582**	13.144	**12.954**	**12.941**

3.21, where y_1, y_2, and y_3 denote the noise-free error function values corresponding to RH1, RH2, and RH3, respectively. Table 3.11 lists the best set of material parameters from each group with noise-corrupted error function. It is observed that Q_∞ and b tend to concentrate to their bounds under the effect of observational noise.

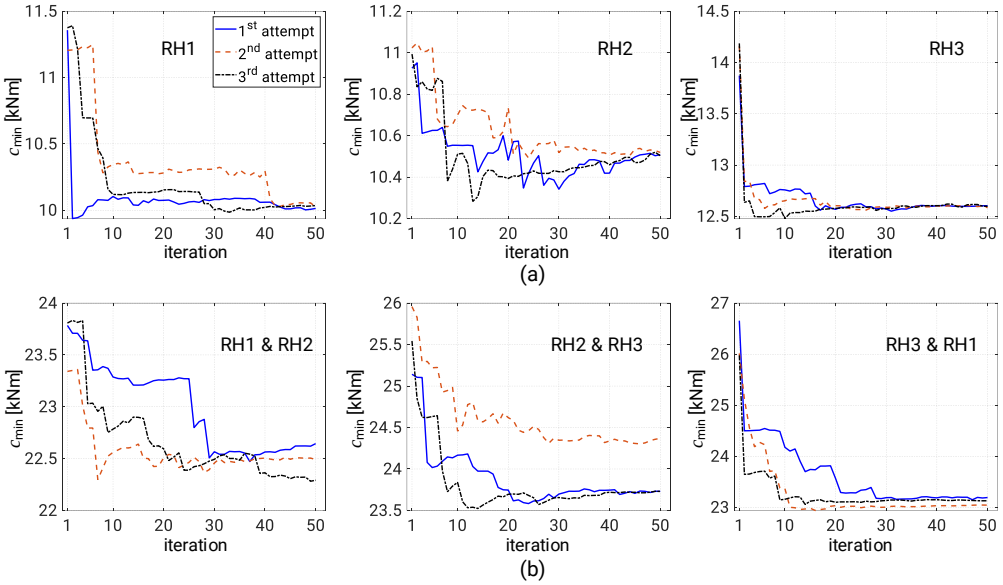

Figure 3.19: Example 2: Histories of SOBO with noise-corrupted error function for different groups of experimental datasets. (a)–(f) groups 1–6, respectively.

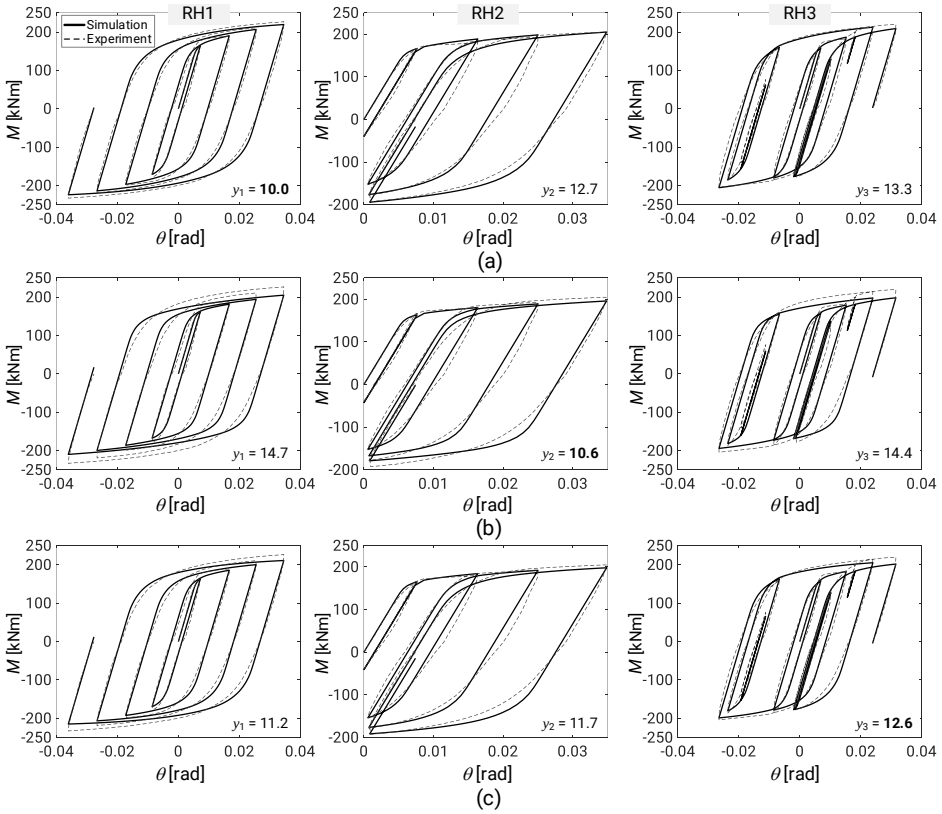

Figure 3.20: Example 2: Comparison of test data and model predictions for parameters identified from SOBO with noise-corrupted error function and different groups of experimental datasets. (a)–(c) parameters from groups 1–3, respectively.

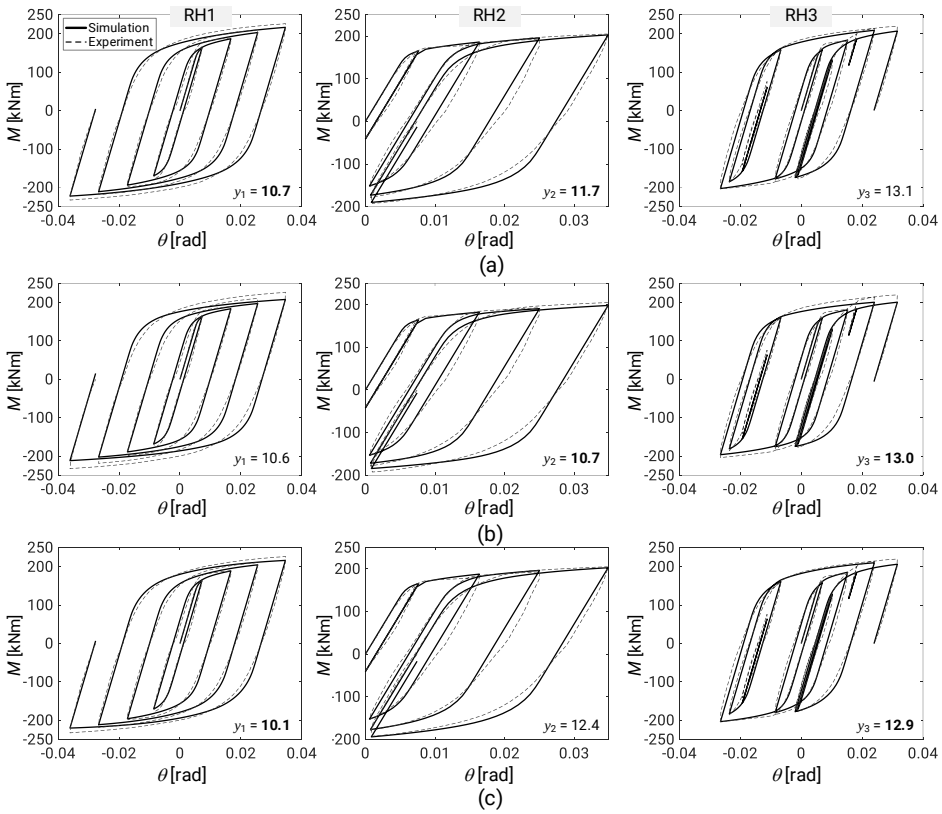

Figure 3.21: Example 2: Comparison of test data and model predictions for parameters identified from SOBO with noise-corrupted error function and different groups of experimental datasets. (a)–(c) parameters from groups 4–6, respectively.

3.4.2 Identification results from MOBO

To investigate how the experimental datasets used for identification affect the resulting parameters, the three experimental datasets of the cantilever are also classified into three different groups indexed as 7, 8, 9, which correspond to (RH1 & RH2), (RH2 & RH3), and (RH3 & RH1), respectively. To examine the dependence of MOBO on the training dataset, each group generates ten different training datasets at the beginning of MOBO. Each of these training datasets has 100 random samples of material parameters and the corresponding error function values.

We carry out MOBO for the ten different training datasets of each group, hence provide ten Pareto fronts of material parameters from each group. The number of MOBO iterations is limited at 50. Thus, the maximum number of simulations required for each MOBO attempt is 300 (i.e., 200 for generation of the initial training dataset and maximum $50 \times 2 = 100$ for MOBO iterations). The reference point and the parameter β are set as $\mathbf{f}_{\mathrm{R}} = [50, 50]$ kNm and 0.01, respectively. Problem (3.11) in each MOBO iteration is solved using GA whose parameters are the same as those for the GA in Section 3.3.2.

To enable a favorable comparison of the identified parameters, we also perform NSGA-II ten times for finding Pareto fronts for each group. Each NSGA-II is characterized by a population of 30 individuals and five generations, thereby requiring a total of $30 \times 2 \times 5 = 300$ simulations, which is the same as the maximum number of simulations required for MOBO.

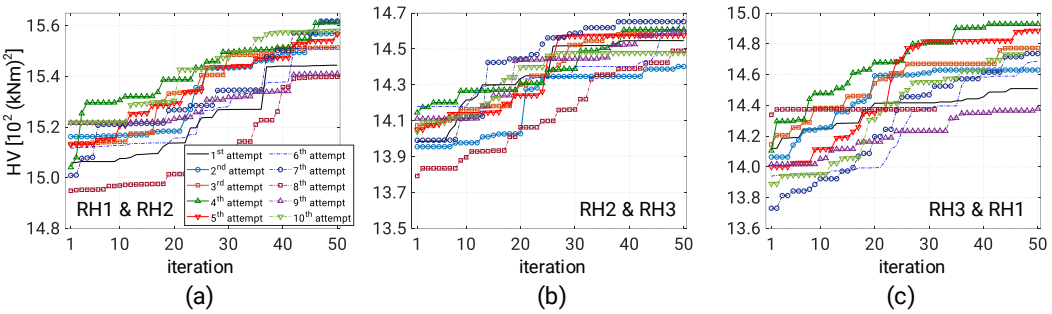

Figure 3.22: Example 2: Histories of HVs associated with ten MOBO attempts of each group. (a)–(c) groups 7–9, respectively.

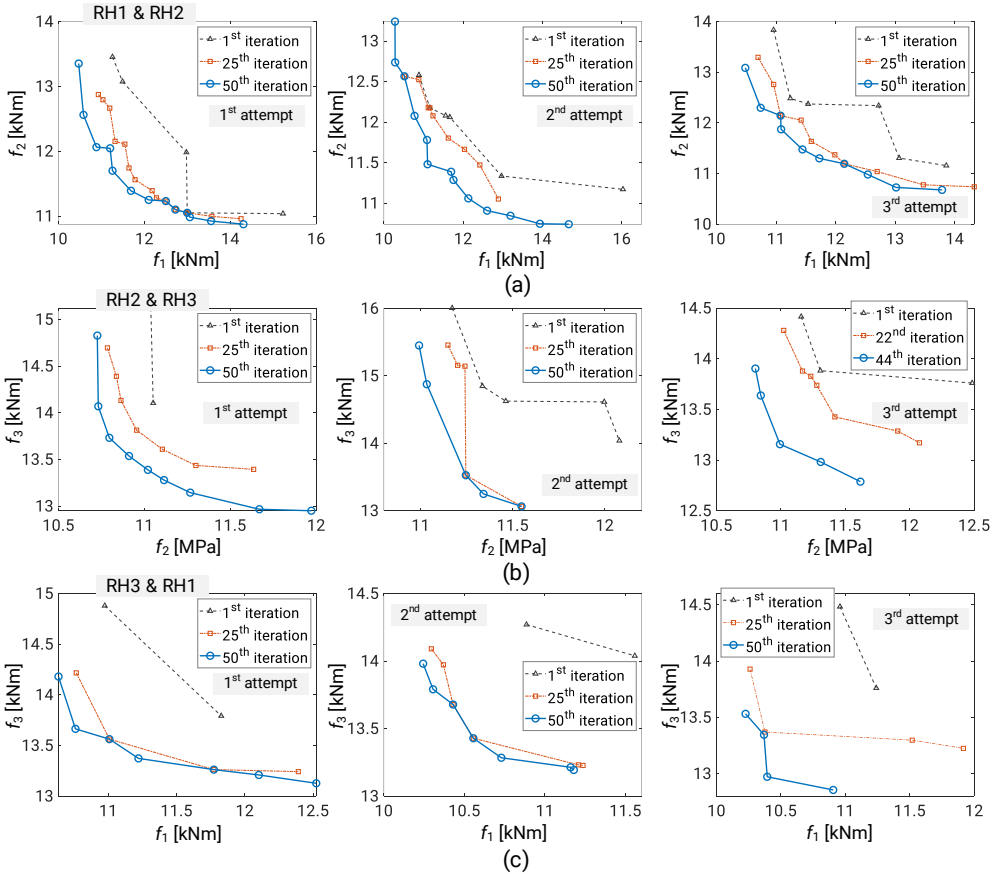

Figure 3.23: Example 2: Evolution of solutions during the first three MOBO attempts for different groups of experimental datasets. (a)–(c) groups 7–9, respectively.

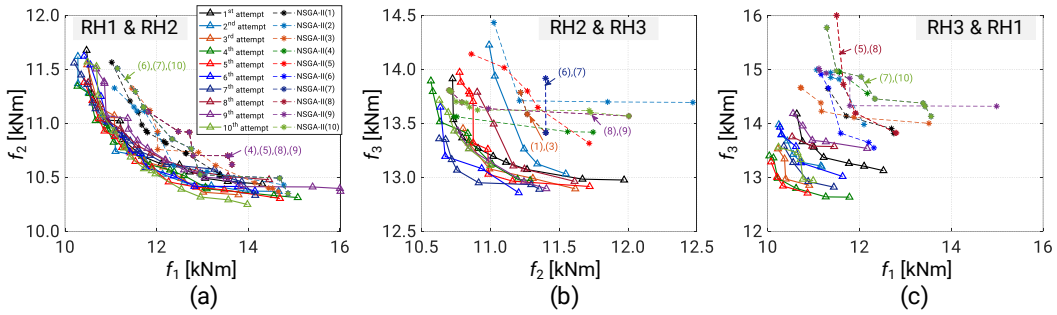

Figure 3.24: Example 2: Comparison of solutions by ten MOBO attempts and by ten NSGA-II attempts for different groups of experimental datasets. (a)–(c) groups 7–9, respectively.

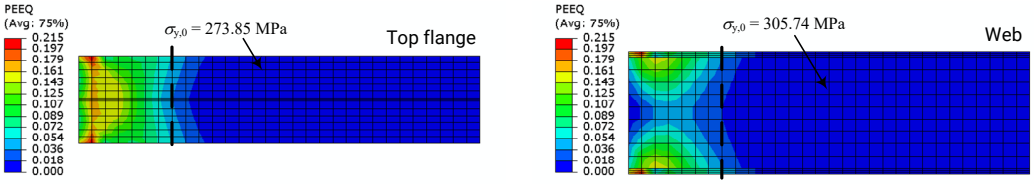

Figure 3.25: Example 2: Equivalent plastic strain distribution over the top flange and web of the cantilever at a time instant of RH1 with the best compromise set of parameters obtained from group 7.

Figure 3.22 shows the histories of HV from the ten MOBO attempts of each group. Although the MOBO attempts starting at different initial training datasets cannot arrive at a unique value of HV after 50 iterations, they considerably improve the solution quality. Since the HVs from most of the MOBO attempts tend to improve in the very last iterations, it is expected that the solutions from each group can be improved if the number of MOBO iterations increases. This result also poses a question on how to characterize the properties of an optimal initial training dataset of MOBO.

Figure 3.23 shows the evolution of approximate Pareto-optimal solutions during the first three MOBO attempts of each group. As is clear, the solutions in the last iteration are much better than those in the very first iteration. Figure 3.24 compares the Pareto fronts from ten MOBO attempts and ten NSGA-II attempts of each group. The shapes of the Pareto fronts by the ten MOBO attempts of the first group are similar (see Fig. 3.24(a)), but those of the other two groups are not (see Figs. 3.24(c) and (d)). The solutions by the MOBO attempts in each group outperform those by the NSGA-II attempts, even though the number of costly simulations required for MOBO does not exceed that required for NSGA-II.

Table 3.12 lists the best (1) and second-best (2) compromise solutions selected from the Pareto front with the highest HV among the ten Pareto fronts found from each group. The error function values corresponding to RH1, RH2, and RH3, denoted as f_1, f_2, and f_3, respectively, are also provided, where the boldface values indicate the experimental datasets used for identification. Identified values of $\sigma_{y,0}$, Q_∞, and b for the flange among those from different groups are similar, while those of C_1 and γ_1 are group-dependent. A change in the dataset (i.e., group) can lead to a major difference in the identified parameters for the web although these parameters are possible to produce the simulated responses consistent with the corresponding experimental measures. This is because the plastic deformation values of web is smaller than those of flange in bending test as illustrated in Fig. 3.25.

Table 3.12: Example 2: Comparison of the identification results obtained from MOBO for different groups of experimental datasets. (1) and (2) indicate the best and second-best compromise solutions, respectively.

	Parameter	Group					
		7 (1)	8 (1)	9 (1)	7 (2)	8 (2)	9 (2)
Web	$\sigma_{y,0}$ [MPa]	305.738	325.737	338.125	311.831	326.628	334.007
	Q_∞ [MPa]	99.938	76.621	99.813	82.085	86.073	99.751
	b	15.077	23.777	24.937	19.234	24.933	24.997
	C_1 [MPa]	6013.083	3405.444	5691.577	5477.573	5946.957	2665.344
	γ_1	66.381	99.536	99.964	35.927	99.909	99.940
Flange	$\sigma_{y,0}$ [MPa]	273.851	270.577	270.007	273.761	271.079	270.233
	Q_∞ [MPa]	12.268	10.315	14.499	10.191	10.517	10.025
	b	6.426	8.090	5.001	6.521	5.783	5.220
	C_1 [MPa]	7502.126	6315.732	7316.046	7410.770	4904.880	7030.846
	γ_1	73.482	58.928	76.793	73.137	44.959	79.902
Error	f_1 [kNm]	**11.607**	12.571	**10.621**	**11.596**	13.360	**11.783**
	f_2 [kNm]	**11.489**	**10.937**	11.846	**11.517**	**10.672**	11.458
	f_3 [kNm]	13.630	**13.163**	**12.799**	13.660	**13.390**	**12.635**

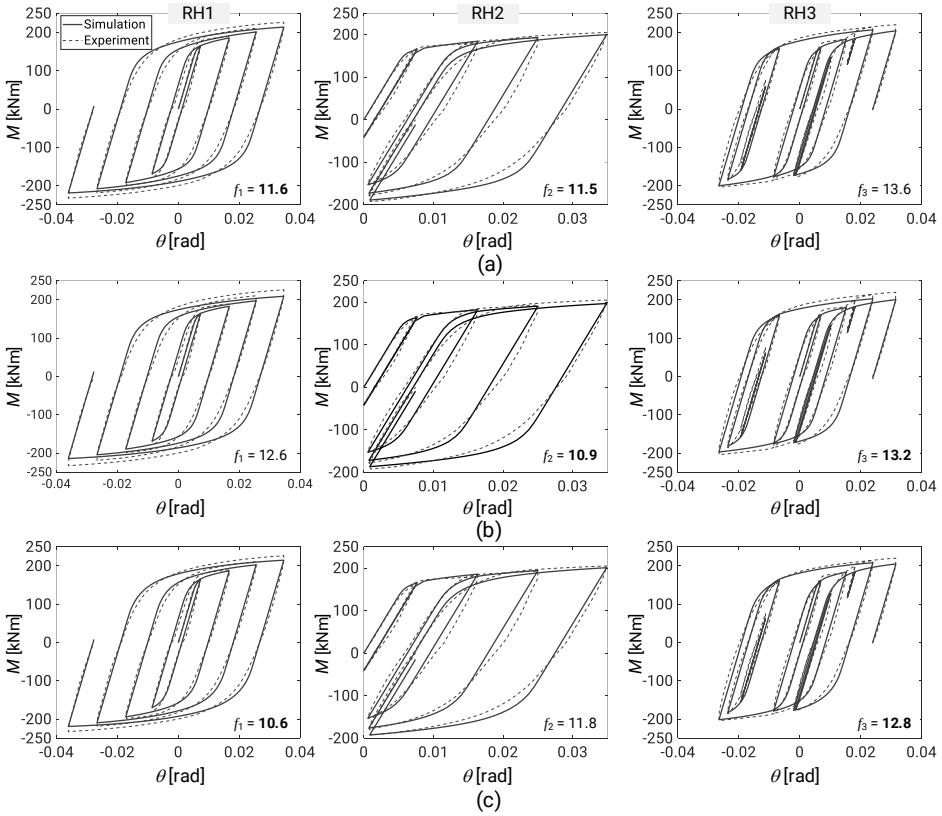

Figure 3.26: Example 2: Comparison of test data and model predictions for the best compromise solutions of parameters identified from MOBO with different groups of experimental datasets. (a)–(c) parameters from groups 7–9, respectively.

Figure 3.26 compares the predicted and measured $M - \theta$ curves from each loading history with use of the best compromise parameters identified from each group for prediction. In general, the identified parameters well reproduce the $M - \theta$ curve corresponding to the loading history that

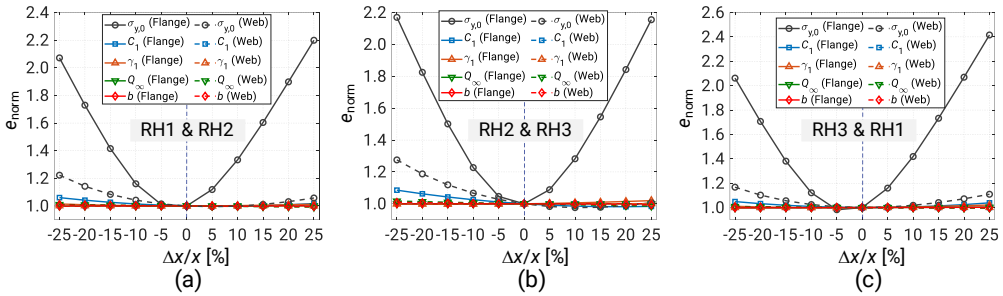

Figure 3.27: Example 2: Average sensitivity results over different groups of experimental datasets. (a)–(c) parameters from groups 7–9, respectively.

Table 3.13: Example 2: Dataset-specific indexes for different sets of identification results evaluated using the experimental results from tests RH1, RH2, and RH3 of the cantilever [kNm].

Set of results	ψ_1	ψ_2	ψ_3	η
Single-objective (groups $1-3$)	4.083	2.414	1.470	2.656
Weighted-sum (groups $4-6$)	2.653	1.586	0.080	1.440
Multi-objective (1) (groups $7-9$)	1.950	0.909	0.831	1.230
Multi-objective (2) (groups $7-9$)	1.765	0.844	1.025	1.211

is not used for identification. Although there exists the bias toward the loading history used for identification, its effect on the prediction performance of identified parameters is not significant as the corresponding minimized error function value is slightly smaller than that predicted from the parameters identified from the other loading histories (see along each column of Fig. 3.26).

From the results in Tables 3.10 and 3.12, η index in Eq. (3.17) is evaluated for the parameter sets obtained from the single-objective (groups 1–3) and weighted-sum (groups 4–6) formulations, and for the best (1) and second-best (2) compromise solutions of the multi-objective formulation. Results in Table 3.13 indicate that η values for the best and second-best compromise sets of identification results from the multi-objective formulation are better than those from the single-objective and weighted-sum formulations. The single-objective formulation again shows the largest value of the dataset-specific bias level.

Figure 3.27 shows the sensitivity results for each best compromise parameters identified from the three experimental dataset groups of the cantilever. The high-sensitivity parameters include $\sigma_{y,0}$ of the flange and web, and C_1 of the flange, in which $\sigma_{y,0}$ of the flange has the greatest influence on the sensitivity of the error functions. Other parameters can be classified as low-sensitivity parameters.

3.5 Conclusions

The cyclic response of a steel structure depends on the constitutive laws of its materials. If this cyclic response can be measured experimentally, single- and multi-objective inverse problems can be formulated for identification of the underlying material parameters. This chapter has successfully applied SOBO to solving the single-objective inverse problem with noise-free and noise-corrupted error function, and proposed a proximal-exploration MOBO approach to solving the multi-objective inverse problem. The main conclusions of this chapter are summarized as follows:

(1) SOBO demonstrates its good ability to identify the parameters for the nonlinear combined

isotropic/kinematic hardening model for structural steel. The material parameters obtained by SOBO with noise-free and noise-corrupted error functions can reliably predict the cyclic behavior of steel subjected to different loading conditions.

(2) Identification results confirm that SOBO outperforms GA and PSO in terms of the prediction performance of identified parameters when the number of simulation calls is small. This promotes the use of SOBO to solving single-objective parameter identification problems with expensive-to-calculate error functions at a small number of simulation calls.

(3) A multi-objective inverse problem is formulated for identifying the material parameters to reduce the dataset-specific bias that may lead the elastoplastic parameters for the cyclic constitutive law identified from a single loading history to inaccurate predictions of structural responses under other loading histories. Accordingly, an efficient proximal-exploration MOBO approach is proposed for solving the formulated multi-objective inverse problem.

(4) The proposed MOBO demonstrates its good performance in solving a simple bi-objective minimization problem. As also confirmed in the identification examples, MOBO outperforms NSGA-II in terms of solution quality when the number of simulation calls is small. This promotes the use of MOBO to solving multi-objective parameter identification problems with expensive-to-calculate error functions at a small number of simulation calls.

(5) Identification results show that the best compromise solution of parameters obtained by MOBO well captures the cyclic behavior of the steel under different loading conditions. Nevertheless, designers are free to select the material parameters for their designs according to the loading conditions.

(6) A dataset-specific index η is proposed for assessing the dataset-specific bias levels for different sets of identification results obtained from different inverse problem formulations. Based on η, we have shown in two identification examples that the best and second-best compromise solutions to the multi-objective inverse problem can reduce the dataset-specific bias, while the solution to the single-objective formulation shows an enormous bias level. Thus, measures from various loading histories should be simultaneously used for parameter identification.

(7) The identification results for the bi-material cantilever also suggest a possibility of identifying material parameters from cyclic tests of a structural component. This is notable because cyclic material tests are difficult and usually not carried out before structural tests.

Chapter 4

Bayesian optimization-assisted approximate Bayesian computation and its application to identifying nonlinear hysteresis curve of structural steels

The single-objective inverse problem in Chapter 3 minimizes an error function for finding a best-fit material parameter vector. Although SOBO enables the minimization of a noise-corrupted error function, the resulting best-fit parameter vector is not very useful in characterizing the effects of observational noise on identification results because many parameter vectors consistent with the measured dataset can be found within the noise, especially for an ill-posed inverse problem. This circumstance motivates the implementation of Bayesian inference for solving the inverse problem that transforms observational noise into uncertainty in identified parameters. However, it is too costly to perform a posterior sampling technique directly for the Bayesian inference in this case because the likelihood function associated with simulating the cyclic elastoplastic behavior of structural steels is computationally intractable. As a continuation of Chapter 3, this chapter introduces BO-assisted approximate Bayesian computation (ABC) to the Bayesian inference of cyclic elastoplastic parameters for structural steels.

4.1 Problem formulation

Consider an inverse problem to infer the vector of material parameters $\mathbf{x}$ incorporated into an FE model for simulating the cyclic elastoplastic behavior of a steel specimen. Let $\mathbf{d}(\mathbf{x}) \in \mathbb{R}^m$ denote the vector of responses of interest simulated from the FE model that inputs $\mathbf{x}$ as material parameters, and $\mathbf{d}_\mathrm{o}$ denote the corresponding vector measured from a cyclic test of the steel specimen. For

formulation of the inverse problem, $\mathbf{d}(\mathbf{x})$ and $\mathbf{d}_\mathrm{o}$ formulate the following discrepancy:

$$\delta = \|\mathbf{d}(\mathbf{x}) - \mathbf{d}_\mathrm{o}\| + e, \tag{4.1}$$

where $\|\cdot\|$ denotes a distance measure, e.g., the Euclidean distance, and e is the discrepancy noise that describes random corruptions of the measured dataset $\mathbf{d}_\mathrm{o}$. Note that e can be evaluated in an average scheme using noise involved in each element of $\mathbf{d}_\mathrm{o}$; see Ref. [105]. If e is modeled by a Gaussian with zero mean and a given standard deviation value, the PDF of δ given $\mathbf{x}$ reads

$$\pi(\delta; \mathbf{x}) = \mathcal{N}\left(y(\mathbf{x}), \tau^2\right), \tag{4.2}$$

where $y(\mathbf{x}) = \|\mathbf{d}(\mathbf{x}) - \mathbf{d}_\mathrm{o}\|$ and τ is a standard deviation value in the space of δ. Note that the uncertainty due to idealization and discretization errors is not considered in this study.

The literature categorizes methods for solving inverse problems into deterministic and probabilistic approaches [108]. The deterministic approach presented in Chapter 3 minimizes $y(\mathbf{x})$ for obtaining a best-fit vector of the parameters. However, it is not very useful to use the best-fit parameter vector for capturing the effects of observational noise e on the identification results because many parameter vectors within the noise can well reproduce $\mathbf{d}_\mathrm{o}$ [109, 110]. In this circumstance, it is preferable to adopt the probabilistic approach that formulates the inverse problem to transform observational noise into uncertainty in the identified parameters. By updating a prior PDF, which represents our belief about the parameters before the dataset is measured, based upon a likelihood of the FE model, which encodes how likely that a given parameter set captures the measured dataset $\mathbf{d}_\mathrm{o}$, the probabilistic approach provides a posterior PDF describing distributions of the parameters when conditioned on $\mathbf{d}_\mathrm{o}$. Then, the samples of parameters are generated from the posterior using a posterior sampling technique, for example, MCMC method [111]. Based on these samples, the sample mean, sample variance, maximum a posteriori (MAP) estimate, or confidence intervals of each uncertain parameter can be found for decision-making or for uncertainty quantification of the FE model. It is worth noting that the deterministic approach is a special case of the probabilistic approach that assigns a constant and the MAP estimate as the prior and the best-fit parameter vector, respectively. This allows a reasonable assessment of the quality of posterior samples if the best-fit parameter vector is available.

The probabilistic approach in this chapter infers $\mathbf{x}$ by the following two steps. The first step expresses our prior belief about $\mathbf{x}$ via the prior PDF $\pi(\mathbf{x})$. The second step updates the prior $\pi(\mathbf{x})$ based upon the likelihood function $p(\mathbf{d}_\mathrm{o}; \mathbf{x})$ of the FE model, resulting in the posterior PDF $\pi(\mathbf{x}; \mathbf{d}_\mathrm{o})$. In essence, if the likelihood function $p(\mathbf{d}_\mathrm{o}; \mathbf{x})$ is available in closed form, $\pi(\mathbf{x}; \mathbf{d}_\mathrm{o})$ is estimated from Bayes' rule in Eq. (2.12), which reads

$$\pi(\mathbf{x}; \mathbf{d}_\mathrm{o}) = \frac{\pi(\mathbf{x}) p(\mathbf{d}_\mathrm{o}; \mathbf{x})}{\pi(\mathbf{d}_\mathrm{o})} = \frac{\pi(\mathbf{x}) p(\mathbf{d}_\mathrm{o}; \mathbf{x})}{\int \pi(\mathbf{x}) p(\mathbf{d}_\mathrm{o}; \mathbf{x}) \, \mathrm{d}\mathbf{x}} \propto \pi(\mathbf{x}) p(\mathbf{d}_\mathrm{o}; \mathbf{x}), \tag{4.3}$$

where the denominator $\pi(\mathbf{d}_\mathrm{o})$ is called evidence that is a normalizing constant to ensure that $\pi(\mathbf{x}; \mathbf{d}_\mathrm{o})$ is a PDF. Thus, $\pi(\mathbf{x}; \mathbf{d}_\mathrm{o})$ is proportional to $\pi(\mathbf{x}) p(\mathbf{d}_\mathrm{o}; \mathbf{x})$.

For model validation or decision making, the standard Bayesian inference generates thousands or millions of parameter samples from Eq. (4.3) using a posterior sampling technique [112], and hence requires a huge number of $p(\mathbf{d}_\mathrm{o}; \mathbf{x})$ evaluations. For our purpose, however, it is difficult

to comply with this requirement because the likelihood function is computationally intractable as modeling the cyclic elastoplastic behavior of structural steels demands a substantial computational cost. Consequently, it is impractical to directly employ the posterior sampling technique using Eq. (4.3).

To deal with this difficulty, recent studies approximated the likelihood using surrogate models and performed MCMC or MCS based upon approximate likelihoods [113, 114]. Thus, it is very important to achieve sufficient accuracy of the surrogate models for having reliable posterior samples. Yet the use of either the goodness-of-fit coefficient evaluated for a test dataset [113] or a space-filling experimental design [114] may not be the best choice for handling the trade-off between the required number of costly training samples and the reliability of MCMC samples. This is due to the fact that neither of these methods incorporates the information from each parameter sample into the improvement of the approximate likelihood accuracy, and it is more efficient if improving the reliability of MCMC samples can totally ignore the parameter region that is unlikely to be consistent with the measured dataset. Thus, the remaining issue is how to distinguish the region the approximate likelihood should be constructed carefully from that should be ignored to reduce the number of costly simulations based on the information from the parameter sample. This is the focus of the subsequent sections.

4.2 Approximate Bayesian computation (ABC)

ABC appears as one of the posterior approximation methods [112, 115–117]. It bypasses the evaluations of $p(\mathbf{d}_\mathrm{o};\mathbf{x})$ using samples generated from the FE model. The main idea is to draw a parameter sample $\mathbf{x}$ from the prior $\pi(\mathbf{x})$, evaluate the numerical dataset $\mathbf{d}(\mathbf{x})$ using the FE model, generate δ from Eq. (4.2), and accept $\mathbf{x}$ as a sample constituting the posterior PDF in Eq. (4.3) if δ is small enough. The underlying foundation of this idea is the generation of sample $(\mathbf{x},\delta)$ from the following joint PDF:

$$\pi_\xi^{\mathrm{ABC}}(\mathbf{x},\delta;\mathbf{d}_\mathrm{o}) \propto \pi(\mathbf{x})\pi(\delta;\mathbf{x})K_\xi(\delta), \tag{4.4}$$

where $K_\xi(\delta)$ is a kernel function indicating whether a parameter sample $\mathbf{x}$ is accepted or not. Most implementations of ABC algorithms use $K_\xi(\delta) = \mathbb{I}\left[\delta \leq \xi\right]$ with the indicator function $\mathbb{I}[\cdot]$ as detailed in Eq. (B.3) and a small threshold value $\xi > 0$. As a result, the ABC posterior can be derived from Eq. (4.4) as

$$\pi_\xi^{\mathrm{ABC}}(\mathbf{x};\mathbf{d}_\mathrm{o}) \propto \pi(\mathbf{x}) \int \pi(\delta;\mathbf{x})K_\xi(\delta)\,\mathrm{d}\delta. \tag{4.5}$$

This can be interpreted as conditioning the joint PDF in Eq. (4.4) on the measured dataset $\mathbf{d}_\mathrm{o}$. The kernel function allows us to assign larger weights for parameter samples with smaller δ.

A high dimensional dataset $\mathbf{d}_\mathrm{o}$ and a small threshold value ξ may result in a small probability of accepting a parameter sample at which the associated vector of simulated responses is close to $\mathbf{d}_\mathrm{o}$ [112]. In this case, a large ξ may be needed for an increase in the acceptance probability. A relatively large ξ, however, can lead to a poor approximation of the posterior because the simulated response vector corresponding to the accepted parameter sample is not actually close to $\mathbf{d}_\mathrm{o}$. To deal with this issue, let $\mathbf{z}(\mathbf{x}) = \eta(\mathbf{d}(\mathbf{x})) : \mathbb{R}^m \to \mathbb{R}^q$ be a mapping (or summary statistics [112]) to reduce the dimension of $\mathbf{d}$, where $q \ll m$. Here $\mathbf{z}(\mathbf{x})$ serves as an informative feature of $\mathbf{d}(\mathbf{x})$. By choosing

an appropriate mapping $\eta(\cdot)$, the ABC posterior in Eq. (4.5) can be approximated by

$$\pi_{\xi}^{\mathrm{ABC}}(\mathbf{x};\mathbf{d}_{\mathrm{o}}) \approx \pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}}) \propto \pi(\mathbf{x}) \int \pi(\Delta;\mathbf{x}) K_{\zeta}(\Delta)\, \mathrm{d}\Delta, \tag{4.6}$$

where $\mathbf{z}_{\mathrm{o}} = \eta(\mathbf{d}_{\mathrm{o}})$, $\pi(\Delta;\mathbf{x}) = \mathcal{N}\left(f(\mathbf{x}), \omega^2\right)$ with $f(\mathbf{x}) = \|\mathbf{z}(\mathbf{x}) - \mathbf{z}_{\mathrm{o}}\|$, ω is a standard deviation value in the space of Δ, and $K_{\zeta}(\Delta) = \mathbb{I}\left[\Delta \leq \zeta\right]$ with a small threshold value $\zeta > 0$.

By further generating a total of n_{Δ} samples $\{\Delta^1, \ldots, \Delta^{n_{\Delta}}\}$ from $\pi(\Delta;\mathbf{x})$ for a given parameter vector $\mathbf{x}$, the ABC posterior can be estimated by replacing the integral in Eq. (4.6) with the Monte-Carlo approximation, such that

$$\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}}) \propto \frac{\pi(\mathbf{x})}{n_{\Delta}} \sum_{t=1}^{n_{\Delta}} \mathbb{I}\left[\Delta^t \leq \zeta\right] \approx \pi(\mathbf{x}) \mathbb{P}\left[\Delta(\mathbf{x}) \leq \zeta\right], \tag{4.7}$$

where $\mathbb{P}\left[\Delta(\mathbf{x}) \leq \zeta\right]$ is the approximate likelihood. Applying Markov's inequality to $\mathbb{P}\left[\Delta(\mathbf{x}) \leq \zeta\right] = 1 - \mathbb{P}\left[\Delta(\mathbf{x}) \geq \zeta\right]$ gives [118]

$$\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}}) \propto\!\!\!\!\!\sim\; \pi(\mathbf{x}) \mathbb{P}\left[\Delta(\mathbf{x}) \leq \zeta\right] \geq \pi(\mathbf{x}) \left[1 - \frac{\mathbb{E}[\Delta(\mathbf{x})]}{\zeta}\right], \tag{4.8}$$

where "$\propto\!\!\!\!\!\sim$" denotes "approximately proportional".

The approximation in Eq. (4.8) implies that a large error due to the estimation of $\mathbb{P}\left[\Delta(\mathbf{x}) \leq \zeta\right]$ can lead to substantial uncertainty in the posterior estimate. This error can be reduced by adopting one of the following refinement approaches developed in the community of statistical modeling. The first approach explores the important region of the parameter space that corresponds to a high probability of accepting a parameter sample $\mathbf{x}$ [118]. This approach is equivalent to maximizing the lower bound of the approximate likelihood or minimizing $\mathbb{E}[\Delta(\mathbf{x})]$ in Eq. (4.8). The second approach sequentially reduces the variance of $\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})$ [119]. Each of these approaches can be formulated as an optimization problem with a costly objective function, and therefore, can be handled by BO.

4.3 Bayesian optimization-based approximate Bayesian computation

At the beginning of the BO-assisted ABC approach, we approximate the mean $f(\mathbf{x}) = \mathbb{E}[\Delta(\mathbf{x})]$ of the discrepancy $\Delta(\mathbf{x})$ using a GP model; see Section 2.3. This GP model is constructed from a training dataset $\mathcal{D} = \{\mathbf{x}^i, f^i\}_{i=1}^N$, where $\mathbf{x}^i$ is a set of parameters, $f^i = \|\mathbf{z}(\mathbf{x}^i) - \mathbf{z}_{\mathrm{o}}\|$, and N is the number of training samples. The sample $\mathbf{x}^i$ is randomly generated from the prior $\pi(\mathbf{x})$. In case $\pi(\mathbf{x})$ is a uniform PDF, it can be better to generate $\mathbf{x}^i$ using Latin-hypercube sampling [74] that provides a good space-filling training dataset. The value f^i is calculated through $\mathbf{d}(\mathbf{x}^i)$, or equivalently $\mathbf{z}(\mathbf{x}^i)$, by performing the FE analysis that inputs $\mathbf{x}^i$ as the vector of material parameters.

The GP model describes the relationship between f and $\mathbf{x}$ using a conditional Gaussian $\hat{f}(\mathbf{x})$. Following Eqs. (2.31a) and (2.31b), the GP model for $f(\mathbf{x})$ reads

$$\hat{f}(\mathbf{x}) \sim \mathcal{N}\left(\mu_{\hat{f}}(\mathbf{x}), \sigma_{\hat{f}}^2(\mathbf{x})\right), \tag{4.9}$$

where $\mu_{\hat{f}}(\mathbf{x})$ and $\sigma_{\hat{f}}(\mathbf{x})$ are the mean and standard deviation of $\hat{f}(\mathbf{x})$, respectively.

The costly simulation of cyclic elastoplastic behavior of structural steels motivates the use of BO to intelligently guide the approximation of ABC posterior. The mathematical foundation BO is detailed in Section 2.4. Suppose BO has completed the s-th iteration at which a solution has been found from a training dataset. By maximizing an acquisition function formulated from the GP model for the objective function, BO strategically selects a new, good design vector (parameter vector) for updating the current solution in the next iteration, i.e., the $(s+1)$-th iteration. In particular, the new parameter vector $\mathbf{x}^{s+1}$ is found by solving

$$\mathbf{x}^{s+1} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmax}} \quad \alpha(\mathbf{x}), \tag{4.10}$$

where $\alpha(\mathbf{x})$ and $\mathcal{X}$ are the acquisition function and the space of parameters, respectively.

After obtaining $\mathbf{x}^{s+1}$, the corresponding objective function is computed by the FE analysis. $\mathbf{x}^{s+1}$ and the corresponding objective function are then added to the current training dataset for updating the solution as well as the GP model.

As discussed in Chapter 2, BO formulates the acquisition function $\alpha(\mathbf{x})$ based on two main criteria: (1) improving the current best-found solution (i.e., exploitation) and (2) reducing the uncertainty in GP predictions (i.e., exploration). In addition to these criteria, this chapter formulates $\alpha(\mathbf{x})$ based on other two approaches that aim to reduce the error in approximating the ABC posterior.

The first approach to reducing the error in the ABC posterior approximation refines $\hat{f}(\mathbf{x})$ in the region that minimizes the discrepancy mean $f(\mathbf{x})$ for an increase in the probability of accepting parameter samples generated from the prior [118]. This is similar to the exploitation of BO when conditioning the objective function on the associated GP model. Three acquisition functions for solving such a problem are provided in Eqs. (2.34), (2.35), and (2.36), namely PI, EI, and LCB, respectively.

The second approach to reducing the error in the ABC posterior estimation refines $\hat{f}(\mathbf{x})$ in the region that reduces the variance of $\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})$ approximation as given in Eq. (4.7) [119]. To formulate an acquisition function for this approach, we first evaluate the mean and variance of $\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})$ when conditioned on the GP model $\hat{f}(\mathbf{x})$, such that

$$\mathbb{E}\left[\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})\right] \propto \pi(\mathbf{x})\Phi\left(\frac{\zeta-\mu_{\hat{f}}(\mathbf{x})}{\sqrt{\omega^2+\sigma_{\hat{f}}^2(\mathbf{x})}}\right), \tag{4.11}$$

$$\mathbb{V}\left[\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})\right] \propto \pi^2(\mathbf{x})\left[\Phi\left(\frac{\zeta-\mu_{\hat{f}}(\mathbf{x})}{\sqrt{\omega^2+\sigma_{\hat{f}}^2(\mathbf{x})}}\right)\Phi\left(\frac{\mu_{\hat{f}}(\mathbf{x})-\zeta}{\sqrt{\omega^2+\sigma_{\hat{f}}^2(\mathbf{x})}}\right) - 2\mathcal{T}\left(\frac{\zeta-\mu_{\hat{f}}(\mathbf{x})}{\sqrt{\omega^2+\sigma_{\hat{f}}^2(\mathbf{x})}}, \frac{\omega}{\sqrt{\omega^2+2\sigma_{\hat{f}}^2(\mathbf{x})}}\right)\right], \tag{4.12}$$

where $\mathcal{T}(\cdot,\cdot)$ denotes the Owen's T-function as

$$\mathcal{T}(x_1,x_2) = \frac{1}{2\pi}\int_0^{x_2}\frac{\mathrm{e}^{-0.5x_1^2(1+x^2)}}{1+x^2}\,\mathrm{d}x. \tag{4.13}$$

Note that Eq. (4.11) is derived from the assumption that $\pi(\Delta;\mathbf{x}) = \mathcal{N}\left(f(\mathbf{x}),\omega^2\right)$, while the derivation of Eq. (4.12) is detailed in Ref. [119].

To reduce the uncertainty in $\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})$ approximation, we wish to generate new parameter

samples in the region that has a large value of $\mathbb{V}\left[\pi_{\zeta}^{\mathrm{ABC}}(\mathbf{x};\mathbf{z}_{\mathrm{o}})\right]$ [119], which is similar to exploration of the standard BO when conditioning the objective function on the associated GP model. Thus, the following acquisition function named VAR is utilized for specifying the new parameter vector:

$$\alpha(\mathbf{x}) = \pi^2(\mathbf{x}) \left[\Phi\left(\frac{\zeta - \mu_{\hat{f}}(\mathbf{x})}{\sqrt{\omega^2 + \sigma_{\hat{f}}^2(\mathbf{x})}} \right) \Phi\left(\frac{\mu_{\hat{f}}(\mathbf{x}) - \zeta}{\sqrt{\omega^2 + \sigma_{\hat{f}}^2(\mathbf{x})}} \right) - 2\mathcal{T}\left(\frac{\zeta - \mu_{\hat{f}}(\mathbf{x})}{\sqrt{\omega^2 + \sigma_{\hat{f}}^2(\mathbf{x})}}, \frac{\omega}{\sqrt{\omega^2 + 2\sigma_{\hat{f}}^2(\mathbf{x})}} \right) \right]. \tag{4.14}$$

Algorithm 4.1 BO-assisted ABC

1: **Specify** $\mathbf{d}_0$, $\pi(\cdot)$, $\eta(\cdot)$, ζ, $s_{\max}$, N, N_{s}, ω (optional);
2: $\mathbf{z}_0 \leftarrow \eta(\mathbf{d}_0)$;
3: $\mathcal{D} \leftarrow \emptyset$;
4: Generate N samples of parameter vector $\mathbf{x}^i \sim \pi(\mathbf{x})$;
5: **for** $i = 1 : N$ **do**
6: Evaluate $\mathbf{d}(\mathbf{x}^i)$ using the FE model; ▷ Costly step
7: $\mathbf{z}(\mathbf{x}^n) \leftarrow \eta(\mathbf{d}(\mathbf{x}^i))$;
8: $f^i \leftarrow \|\mathbf{z}(\mathbf{x}^i) - \mathbf{z}_{\mathrm{o}}\|$;
9: $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^i, f^i\}$;
10: **end for**
11: **for** $s = 0 : s_{\max}$ **do**
12: Construct $\hat{f}(\mathbf{x})$ in Eq. (4.9) from $\mathcal{D}$;
13: Find $\mathbf{x}^{s+1}$ by solving problem (4.10);
14: **if** $\mathbf{x}^{s+1} \notin \mathcal{D}$ **then**
15: Evaluate $\mathbf{d}(\mathbf{x}^{s+1})$ using the FE model; ▷ Costly step
16: $\mathbf{z}(\mathbf{x}^{s+1}) \leftarrow \eta(\mathbf{d}(\mathbf{x}^{s+1}))$;
17: $f^{s+1} \leftarrow \|\mathbf{z}(\mathbf{x}^{s+1}) - \mathbf{z}_{\mathrm{o}}\|$;
18: $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^{s+1}, f^{s+1}\}$;
19: **else**
20: exit **for**
21: **end if**
22: **end for**
23: Generate N_{s} parameter samples from Eq. (4.11) using the adaptive MCMC [120];
24: **return** N_{s} samples of material parameters from the approximate posterior.

To this end, Algorithm 4.1 summaries the identification procedure utilizing the BO-assisted ABC approach, where $s_{\max}$ and N_{s} in step 1 are the pre-specified limit of the number of BO iterations and the number of approximate posterior samples to be generated, respectively. Specifying ω is optional because an optimal ω can be determined in each BO iteration, as a hyperparameter, by maximizing the approximate marginal likelihood of $\pi(\Delta;\mathbf{x})$ derived from the GP model $\hat{f}(\mathbf{x})$; see Section 2.3. The BO nested in the algorithm (i.e., steps 11−22) terminates if either the number of its iterations reaches $s_{\max}$ or $\mathbf{x}^{s+1}$ is identical to any member of $\mathcal{D}$, i.e., the condition in step 14 is not satisfied. Step 13 requires an optimization algorithm for solving problem (4.10). For this task, any conventional optimization algorithm can be used because the form of $\alpha(\mathbf{x})$ is explicitly given. GA with a population size of 4000 is used in this study. Other parameters for GA are listed in Table 3.1.

After the refinement process is completed, the final GP $\hat{f}(\mathbf{x})$ from step 12 and the prior $\pi(\mathbf{x})$ incorporated in Eq. (4.11) are used in step 23 as input to an adaptive MCMC [120] for generating a total of N_{s} parameter samples that are representatives of the posterior samples. The associated

MCMC toolbox is available at `https://github.com/mjlaine/mcmcstat.git` (Accessed on May 10, 2022). Note that the initial parameter vector for starting the MCMC takes the value of new parameter vector obtained from the very last BO iteration of step 13.

4.4 Benchmarks and identification example

In this section, we first verify the performance of Algorithm 4.1 against the benchmark analyses of two datasets generated from a Gaussian model [119] (Section 4.4.1) and a g-and-k distribution [121] (Section 4.4.2). We then carry out the algorithm to identify the approximate posteriors of elastoplastic parameters used for modeling the cyclic elastoplastic behavior of a steel specimen tested under different cyclic loading conditions (Section 4.4.3).

4.4.1 Benchmark 1: Analysis of Gaussian model data

The approximate posterior of $\mathbf{x} = [x_1, x_2]^T \in \mathcal{X} = [0, 8]^2$ is found once the associated dataset $\mathbf{d}_\mathrm{o} = \sum_{i=1}^{5} \mathbf{d}_i/5$ has been observed, where $\mathbf{d}_i \sim p(\mathbf{d}; \mathbf{x}) = \mathcal{N}(\mathbf{d}; \mathbf{x}, \boldsymbol{\Sigma})$ $(i = 1, \ldots, 5)$ and the covariance matrix $\boldsymbol{\Sigma}$ is known. If $\mathbf{x} \sim \pi(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{a}, \mathbf{B})$ truncated to $\mathcal{X}$ for given $\mathbf{a}$ and $\mathbf{B}$, the true posterior can be derived as $\pi(\mathbf{x}; \mathbf{d}_\mathrm{o}) = \mathcal{N}(\mathbf{a}^\star, \mathbf{B}^\star)$, where $\mathbf{a}^\star = \mathbf{B}^\star \left(\mathbf{B}^{-1}\mathbf{a} + 5\boldsymbol{\Sigma}^{-1}\mathbf{d}_\mathrm{o}\right)$ and $\mathbf{B}^\star = \left(\mathbf{B}^{-1} + 5\boldsymbol{\Sigma}^{-1}\right)^{-1}$; see Ref. [119]. The true posterior is the baseline to assess the accuracy of the approximate posterior obtained from Algorithm 4.1.

To generate the dataset $\mathbf{d}_\mathrm{o}$ and specify the prior $\pi(\mathbf{x})$, we set $\mathbf{x} = [2, 2]^T$, $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = 0.5$, $\mathbf{a} = [5, 5]^T$, and $\mathbf{B} = \mathbf{I}$. The summary statistics and the discrepancy are defined by $\mathbf{z}(\mathbf{x}) = \mathbf{d}(\mathbf{x})$ and $f(\mathbf{x}) = \sqrt{(\mathbf{z}(\mathbf{x}) - \mathbf{z}_\mathrm{o})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}(\mathbf{x}) - \mathbf{z}_\mathrm{o})}$, respectively. The parameters for Algorithm 4.1 include $\zeta = \{1, 0.25\}$, $\omega = 0.5$, $N = 10$ samples, $s_\mathrm{max} = 100$ iterations, and $N_\mathrm{s} = 5 \times 10^4$ samples.

In addition to using only one of the four acquisition functions stated in Sections 2.4 and 4.3 (i.e., PI, EI, LCB, and VAR) for each ABC attempt, we employ another strategy named LCB-VAR that switches the acquisition function from LCB to VAR during the ABC attempt. Accordingly, LCB and VAR guide BO in specifying new parameter vectors for the first half and second half of s_max iterations, respectively. To further assess the accuracy of the approximate posterior, we define a distance criterion $d = \sum_{j=1}^{J} \left| \pi_\zeta^\mathrm{ABC}(\mathbf{x}_j; \mathbf{z}_\mathrm{o}) - \pi(\mathbf{x}_j; \mathbf{z}_\mathrm{o}) \right|$, where $\mathbf{x}_j$ are the node coordinate vectors of grid $[0, 0.1, \ldots, 8]^2$ and $J = 81^2$ is the number of nodes.

Figure 4.1 shows the histories of d for different acquisition functions with two levels of discrepancy threshold $\zeta = 1$ and 0.25. For both threshold levels, the refinement of GP model for the discrepancy mean improves the accuracy of ABC posterior approximations considerably. The approximate ABC posterior is better when using the smaller discrepancy threshold as the final values of d associated with $\zeta = 0.25$ are much smaller than those associated with $\zeta = 1$. For $\zeta = 0.25$, similar accuracy levels of the approximate posterior are obtained from LCB and VAR. Switching the acquisition function from LCB to VAR during the ABC attempt can slightly improve the approximate posterior for $\zeta = 0.25$ in Fig. 4.1(b).

Figure 4.2 compares the approximate and true posteriors for different acquisition functions with two discrepancy threshold levels $\zeta = 1$ and 0.25 . It is confirmed that all the approximate posteriors can reproduce the true posterior with an acceptable accuracy. For the same discrepancy threshold level, the posterior approximations brought by LCB, VAR, and LCB-VAR are more accurate than

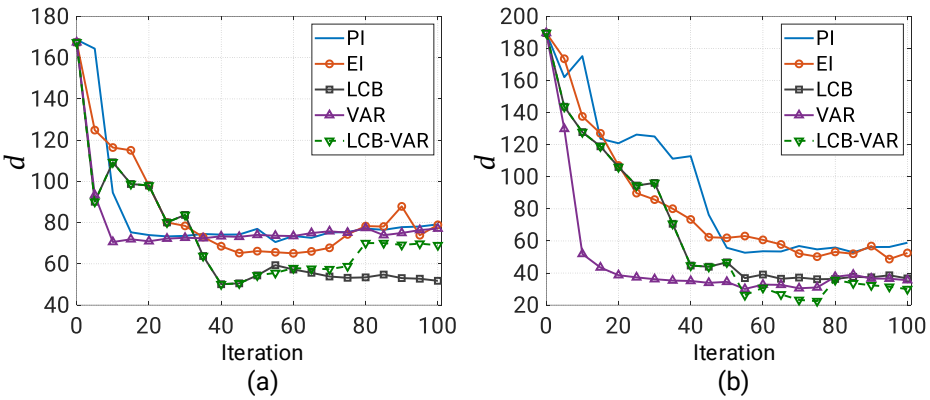

Figure 4.1: Benchmark 1: Histories of the distance criterion d for different acquisition functions with two levels of discrepancy threshold. (a) $\zeta = 1$; (b) $\zeta = 0.25$.

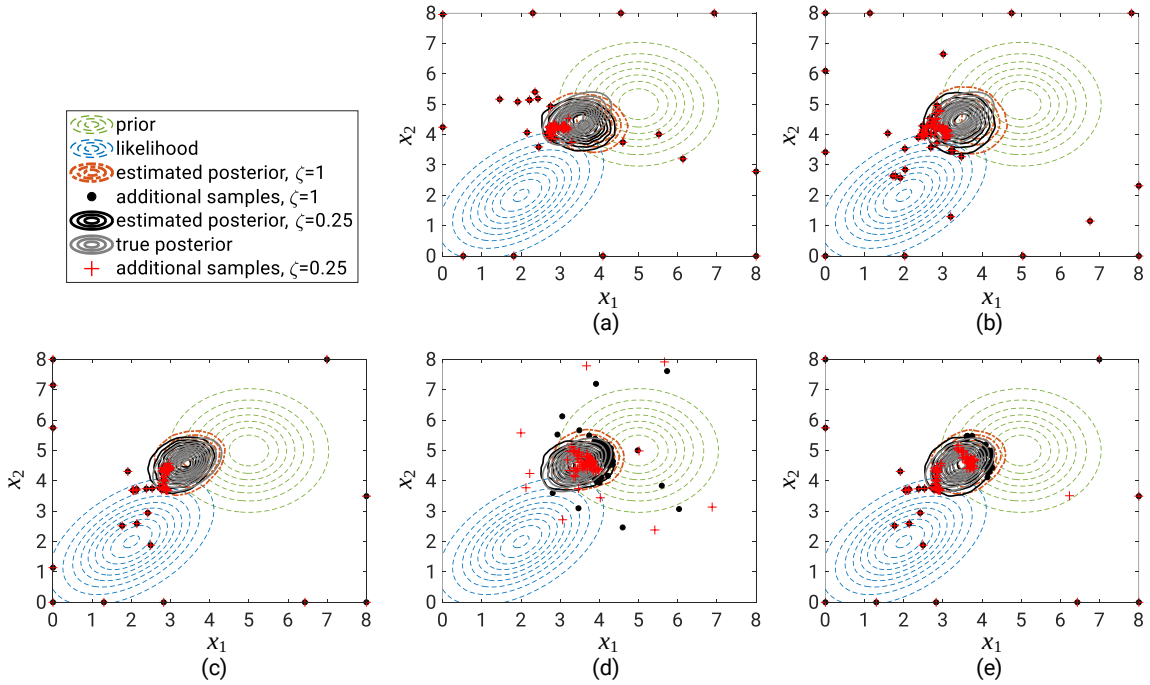

Figure 4.2: Benchmark 1: Comparison of identification results for different acquisition functions with two levels of discrepancy threshold. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

those by PI and EI.

Figure 4.2 also provides insights into how new parameter samples are added to the training dataset when using each acquisition function for a specific threshold level. Accordingly, EI and LCB tend to draw new samples both in the likelihood area where the discrepancy is small, and on the boundary where the uncertainty in GP predictions is large due to the lack of training points. VAR, on the other hand, draws new samples in the prior area only because it incorporates the prior information as stated in Eq. (4.12). We see that the new samples associated with two levels of ζ are identical when using PI, EI, or LCB because of their independence on ζ, but are different when using VAR because of its dependence on ζ. Thus, different levels of discrepancy threshold can share the same GP refinement process if PI, EI, or LCB is the acquisition function of interest.

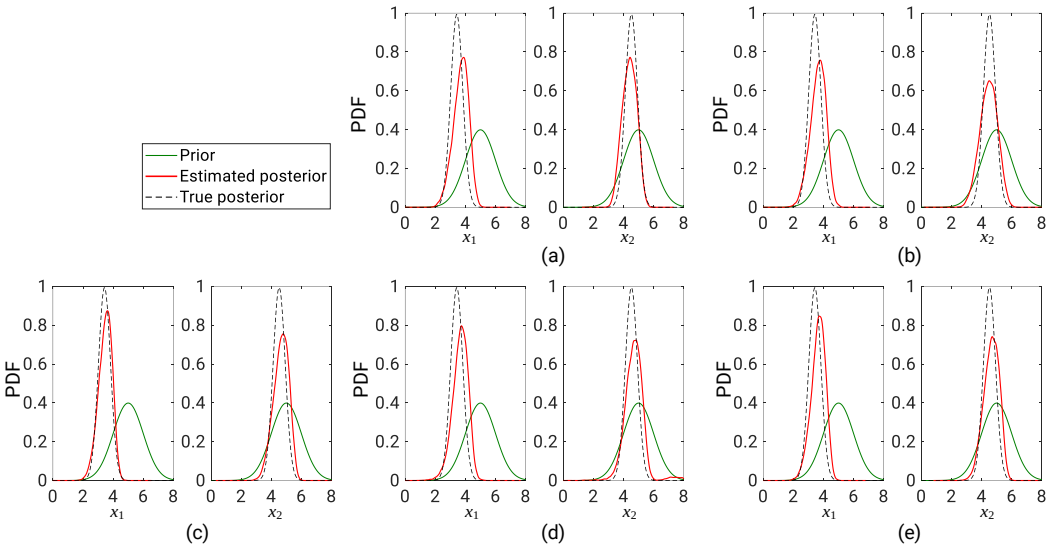

Figure 4.3: Benchmark 1: Comparison of approximate and true marginal posterior PDFs of each parameter for different acquisition functions with $\zeta = 1$. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

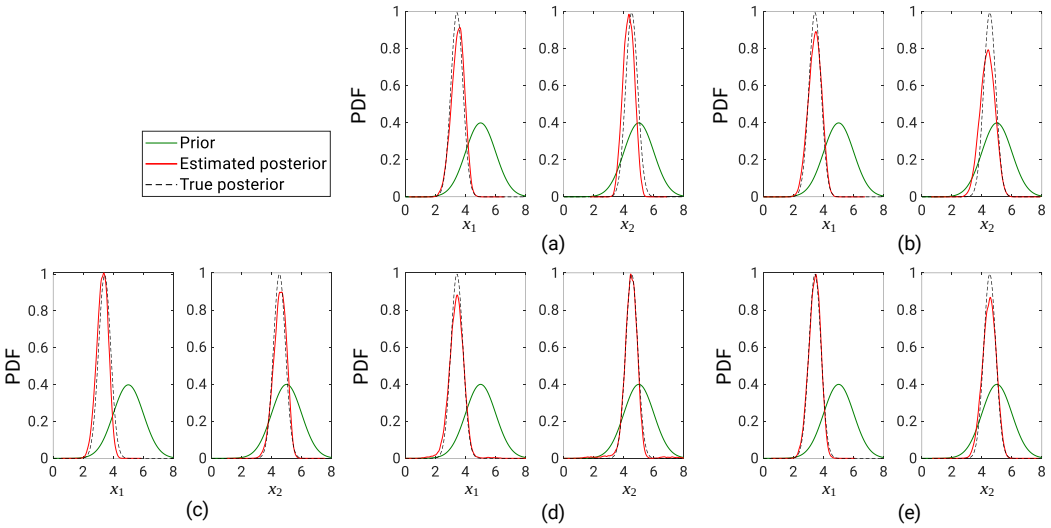

Figure 4.4: Benchmark 1: Comparison of approximate and true marginal posterior PDFs of each parameter for different acquisition functions with $\zeta = 0.25$. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

Figures 4.3 and 4.4 compare the approximate and true marginal posteriors of each parameter for $\zeta = 1$ and 0.25, respectively. These figures again confirm that the improvement of approximate ABC posterior is achieved when using the smaller discrepancy threshold.

4.4.2 Benchmark 2: Analysis of *g*-and-*k* data

This example carries out Algorithm 4.1 to capture the true parameters for a univariate g-and-k distribution, which is a flexible unimodal PDF describing data with large amounts of skewness and kurtosis. Although it has no closed form, the distribution can be defined through a quantile function

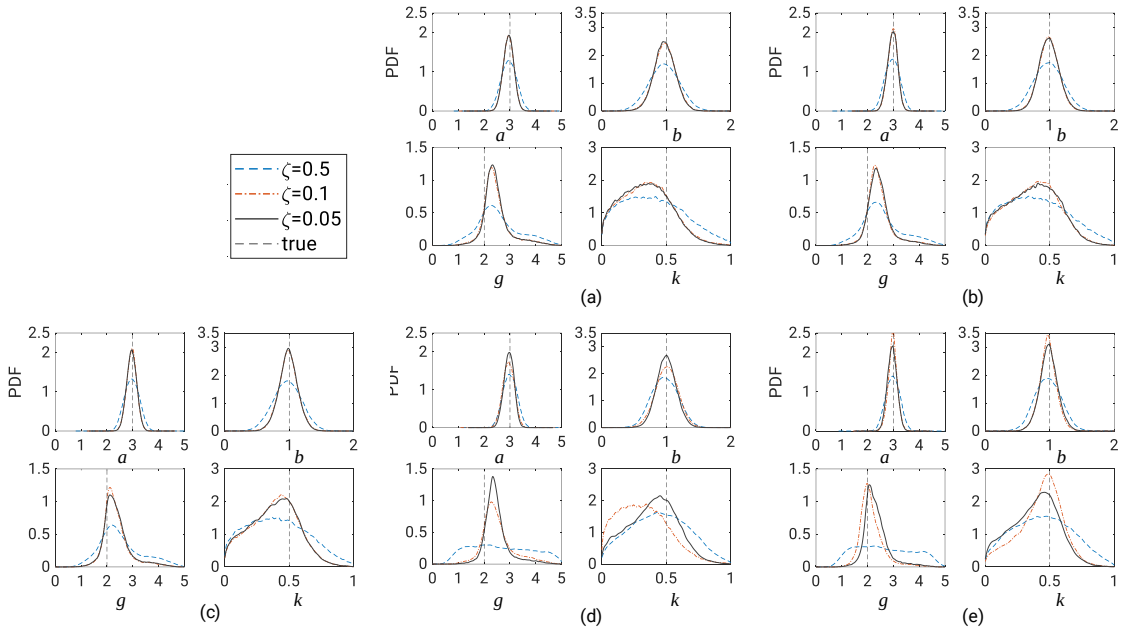

Figure 4.5: Benchmark 2: Approximate posteriors of each parameter for different acquisition functions with three levels of discrepancy threshold. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

(i.e., inverse CDF) as [121]

$$Q_{\mathrm{gk}}\left(\lambda(q);\mathbf{x}\right) = a + b\Big(1 + c\frac{1-\exp(-g\lambda(q))}{1+\exp(-g\lambda(q))}\Big)\left(1+\lambda^2(q)\right)^k \lambda(q), \tag{4.15}$$

where $\lambda(q) = \Phi^{-1}(q)$ is the q-th quantile of the standard Gaussian; $\mathbf{x} = [a, b, g, k]^T$ is the vector of four parameters of the distribution characterizing its location, scale, skewness, and kurtosis, respectively; and c is fixed at 0.8 measuring the overall asymmetry. Our goal is to identify the parameter vector $\mathbf{x} = [a, b, g, k]^T$ that best fits the g-and-k distribution to an observed dataset $\mathbf{d}_{\mathrm{o}}$.

A synthetic dataset $\mathbf{d}_{\mathrm{o}}$ of 10^4 independent samples is generated from Eq. (4.15) with $\mathbf{x} = [3, 1, 2, 0.5]^T$. More specifically, 10^4 independent samples of $\lambda(q)$ are generated from the standard Gaussian, and each sample is then substituted into Eq. (4.15) for obtaining the corresponding sample of $\mathbf{d}_{\mathrm{o}}$. By employing Algorithm 4.1, we wish to find the approximate posterior of $\mathbf{x}$ so that the associated MAP estimate is close to the true parameter vector $\mathbf{x} = [3, 1, 2, 0.5]^T$. For our purpose, let $\pi(\mathbf{x}) = \pi(a)\pi(b)\pi(g)\pi(k) = U(0,5)U(0,2)U(0,5)U(0,1)$ and $\mathbf{z}(\mathbf{x}) = [z_1, z_2, z_3, z_4]^T$ with

$$z_1 = l_2;\ z_2 = l_3 - l_1;\ z_3 = (l_3 + l_1 - 2l_2)/z_2;\ z_4 = (o_7 - o_5 + o_3 - o_1)/z_2. \tag{4.16}$$

Here l_i and o_j denote the i-th quartile and the j-th octile of $\mathbf{d}(\mathbf{x}) \in \mathbb{R}^{10^4}$,respectively. The vector $\mathbf{z}(\mathbf{x})$, following Ref. [121], is called robust summary statistics with z_1, z_2, z_3, and z_4 representing the median, interquartile range, skewness, and kurtosis of the dataset $\mathbf{d}_{\mathrm{o}}$, respectively. If a random seed is used throughout the identification process, each sample of $\mathbf{x}$ drawn from $\pi(\mathbf{x})$ yields a unique vector $\mathbf{z}(\mathbf{x})$ through $\mathbf{d}(\mathbf{x})$ generated from Eq. (4.15).

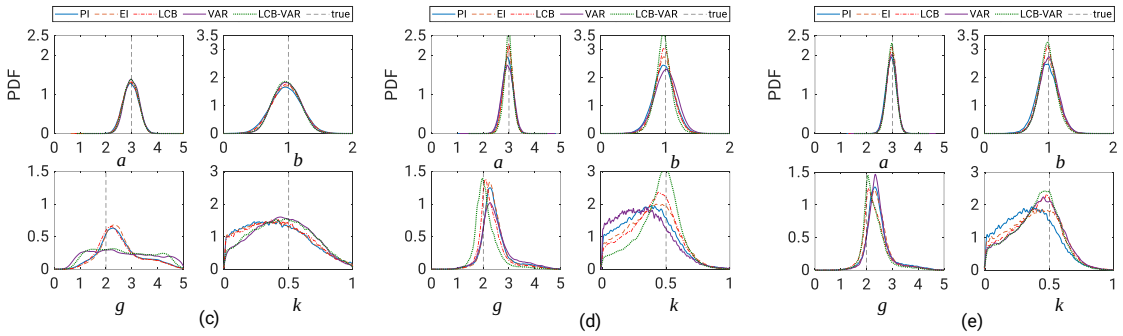

Figure 4.6: Benchmark 2: Comparison of approximate posteriors obtained from different acquisition functions for a specific level of discrepancy threshold. (a) $\zeta = 0.5$; (b) $\zeta = 0.1$; (c) $\zeta = 0.05$.

Table 4.1: Benchmark 2: Comparison of MAP estimates of each parameter for different acquisition functions with $\zeta = 0.5$ and 0.05.

ζ	Parameter	PI	EI	LCB	VAR	LCB-VAR	true
0.5	a	2.939	2.981	2.967	2.963	2.972	3.00
	b	0.956	1.004	0.969	0.952	0.959	1.00
	g	2.893	2.896	2.759	2.873	2.838	2.00
	k	0.415	0.340	0.366	0.436	0.474	0.50
0.05	a	2.962	2.985	2.958	2.973	2.974	3.00
	b	0.951	0.990	0.987	1.002	0.986	1.00
	g	2.743	2.852	2.735	2.863	2.710	2.00
	k	0.361	0.409	0.479	0.451	0.458	0.50

Table 4.2: Benchmark 2: Comparison of standard deviation values of each parameter for different acquisition functions with $\zeta = 0.5$ and 0.05.

ζ	Parameter	PI	EI	LCB	VAR	LCB-VAR
0.5	a	0.297	0.284	0.284	0.270	0.273
	b	0.236	0.226	0.222	0.206	0.205
	g	0.838	0.790	0.816	1.188	1.098
	k	0.237	0.238	0.234	0.223	0.229
0.05	a	0.203	0.191	0.190	0.192	0.181
	b	0.165	0.154	0.144	0.152	0.137
	g	0.524	0.480	0.514	0.469	0.460
	k	0.188	0.195	0.186	0.186	0.178

The discrepancy is defined by the Euclidean norm $f(\mathbf{x}) = \|\mathbf{z}(\mathbf{x}) - \mathbf{z}_o\|_2$. Parameters for Algorithm 4.1 include $\zeta = \{0.5, 0.1, 0.05\}$, $N = 100$ samples, $s_{\max} = 100$ iterations, and $N_s = 10^5$ samples. Because our goal is to find the best-fit parameter vector, ω is automatically optimized in each BO iteration by maximizing the approximate marginal likelihood of $\pi(\Delta; \mathbf{x})$ derived from $\hat{f}(\mathbf{x})$.

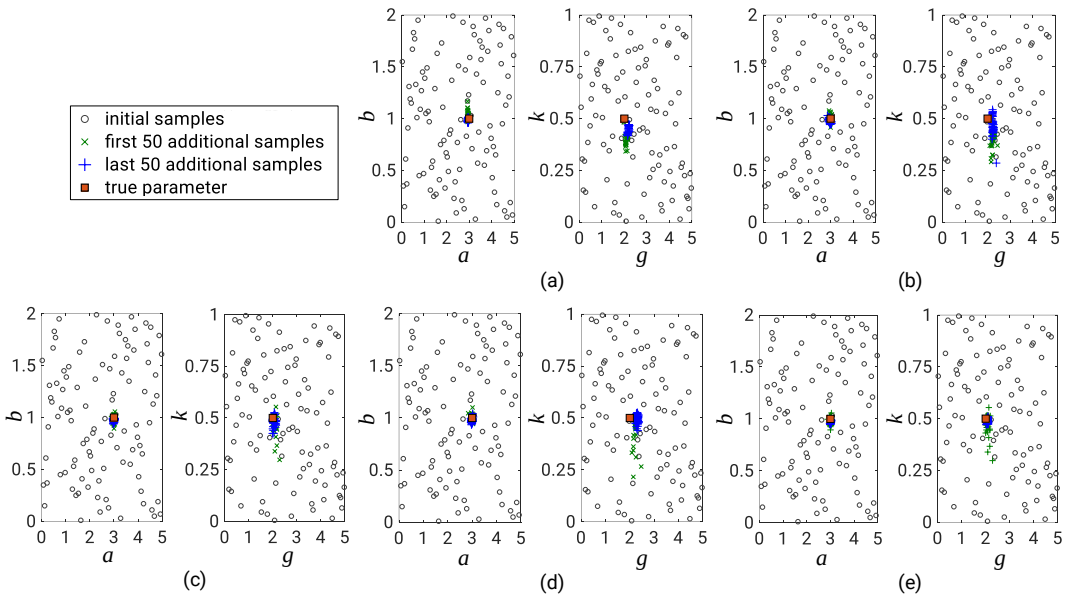

Figure 4.7: Benchmark 2: Additional sampling for different acquisition functions with $\zeta = 0.05$. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

As depicted in Fig. 4.5, Algorithm 4.1 successfully identifies the approximate posterior of each parameter for different acquisition functions with three levels of discrepancy threshold. The MAP estimate of each parameter, as expected, can be found in the neighborhood of the true value. The PDF values at the MAP estimates of all parameters associated with $\zeta = 0.5$ are much smaller than those associated with $\zeta = 0.1$ and 0.05 because there is a large number of accepted samples found from the prior when using a larger ζ, which flattens the approximate posteriors.

Figure 4.6 compares the approximate posteriors of each parameter obtained from different acquisition functions for a specific discrepancy threshold level. For parameters a and b, there is no major difference between the approximate posteriors obtained from different acquisition functions at the same discrepancy threshold level. However, those of parameters g and k are strongly affected by the acquisition function used. LCB provides good approximate posteriors for the three levels of discrepancy threshold.

Tables 4.1 and 4.2 quantitatively compare the MAP estimates and standard deviation values, respectively, associated with the approximate posteriors of each parameter for different acquisition functions. The MAP estimate of each parameter is close to the true parameter as expected. The smaller discrepancy threshold value, i.e., $\zeta = 0.05$, can provide a more accurate MAP estimate as well as a smaller standard deviation value of the approximate posterior. Among the five acquisition function schemes, LCB-VAR with $\zeta = 0.05$ offers the most accurate posterior approximation.

The success of Algorithm 4.1 in identifying the parameters for the g-and-k distribution lies in the special locations of additional samples found by BO. As depicted in Fig 4.7, the additional samples from all the acquisition functions tend to concentrate in the neighborhood of the true parameter vector, where any sample drawn from the prior has a high chance to be accepted as a posterior sample.

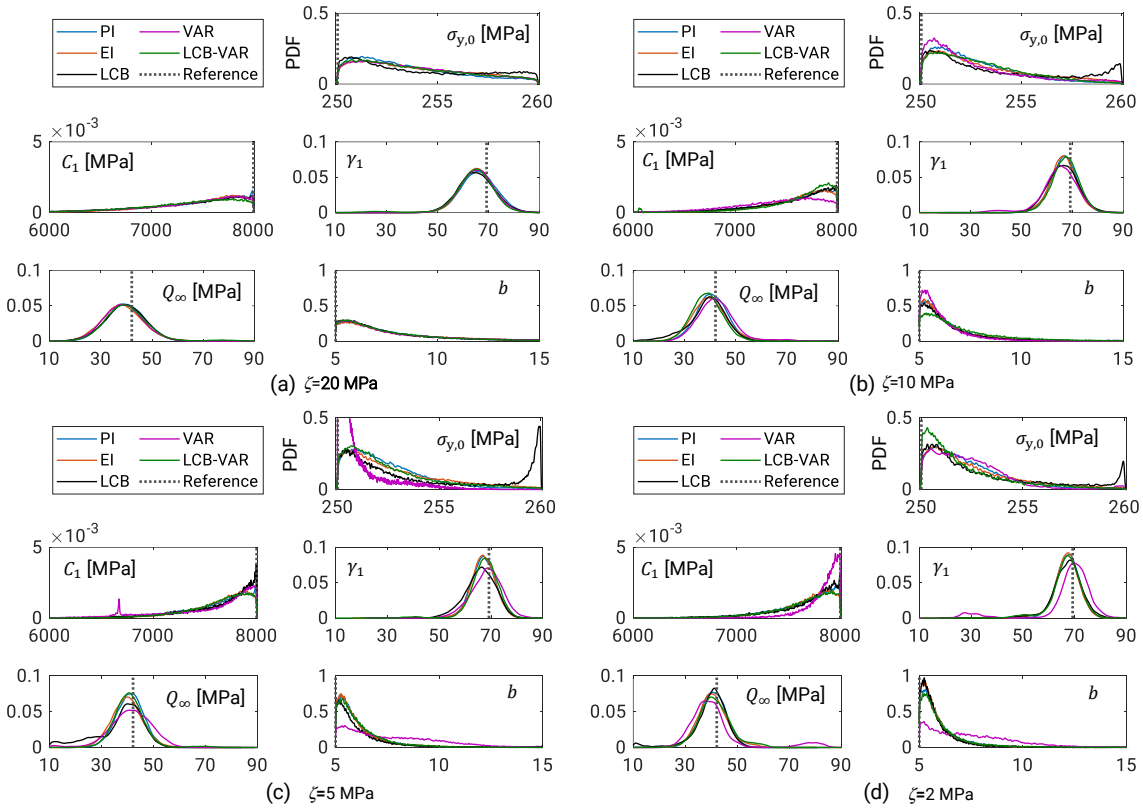

Figure 4.8: Example: Approximate posteriors of each parameter obtained from different acquisition functions, four levels of discrepancy threshold, and dataset SS1.

4.4.3 Example: Parameters for a steel specimen

Algorithm 4.1 is employed to infer the cyclic elastoplastic parameters for the steel specimen in Section 3.3. The specimen was tested under three different loading histories SS1, SS2, and SS3. Three experimental datasets of $\sigma - \epsilon$ curves were obtained corresponding to the three loading histories, where σ and ϵ are the axial true stress and axial true strain, respectively. This section, however, only uses the experimental datasets from SS1 and SS2 for illustration of the identification process.

The material behavior of the specimen is modeled by the nonlinear combined isotropic/kinematic hardening that is characterized by a total of seven parameters. For identification, Young's modulus and Poisson's ratio of the specimen are kept constant as $E = 205.94$ GPa and $\nu = 0.3$, respectively. Thus, there are five material parameters to be identified for the specimen, i.e., $\mathbf{x} = [\sigma_{\mathrm{y},0}, C_1, \gamma_1, Q_\infty, b]$.

Let $\sigma^{\mathrm{s}}_{i,t}(\mathbf{x})$ denote the simulated value of σ at the t-th time step of the i-th cyclic loading history of N_i discrete steps, and $\sigma^{\mathrm{m}}_{i,t}$ denote the corresponding measured value. Therefore, the measured dataset $\mathbf{d}_{\mathrm{o}}$ for the i-th cyclic loading history is a vector of N_i values of σ. Let $\mathbf{z}(\mathbf{x}) = \mathbf{d}(\mathbf{x})$. The discrepancy mean associated with the i-th cyclic loading history for the identification problem of the specimen can be defined by the root-mean-square deviation as

$$f_i(\mathbf{x}) = \sqrt{\frac{1}{N_i}\sum_{t=1}^{N_i}\left(\sigma^{\mathrm{s}}_{i,t}(\mathbf{x}) - \sigma^{\mathrm{m}}_{i,t}\right)^2}. \tag{4.17}$$

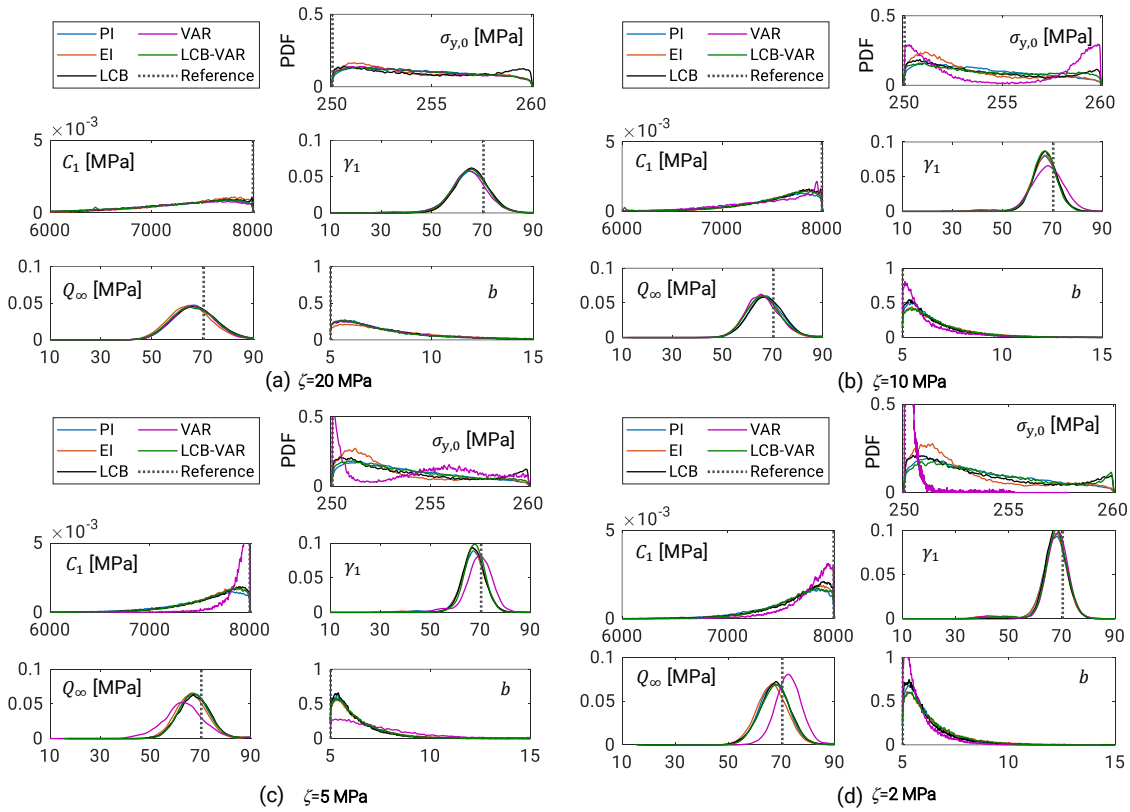

Figure 4.9: Example: Approximate posteriors of each parameter obtained from different acquisition functions, four levels of discrepancy threshold, and dataset SS2.

Using the interval for each parameter in Table 3.2, we define a prior of independent material parameters as $\pi(\mathbf{x}) = \pi(\sigma_{y,0})\pi(C_1)\pi(\gamma_1)\pi(Q_\infty)\pi(b)$, where $\pi(\sigma_{y,0}) = U(250, 260)$ MPa, $\pi(C_1) = U(2000, 8000)$ MPa, $\pi(\gamma_1) = U(10, 100)$, $\pi(Q_\infty) = U(10, 100)$ MPa, and $\pi(b) = U(5, 25)$. For identification, the parameters characterizing Algorithm 4.1 are set as $\zeta = \{20, 10, 5, 2\}$ MPa, $\omega = 3$ MPa, $N = 100$ samples, $s_{\max} = 100$ iterations, and $N_s = 10^5$ samples. Algorithm 4.1 inputs the experimental dataset from SS1 or SS2 for finding the approximate posterior of each parameter. Then, the MAP estimate found from the resulting approximate posterior for each loading history is compared with the corresponding optimal deterministic value (i.e., reference value) given in Table 3.3.

Approximate posteriors and MAP estimates

Figures 4.8 and 4.9 show the approximate posteriors of each parameter, respectively, obtained from the datasets SS1 and SS2 with four discrepancy threshold levels and five acquisition function strategies. The approximate posteriors of each parameter brought by PI, EI, LCB, and LCB-VAR are similar for each threshold level, while the posterior by VAR is quite different from those by the other acquisition functions when the discrepancy threshold level becomes smaller.

Tables 4.3 and 4.4, respectively, show the MAP estimates and standard deviation values of each parameter for different combinations of the acquisition function and measured dataset with $\zeta = 20$ and 2 MPa. The reference value in the last column of Table 4.3 obtained from the deterministic approach in Section 3.3 is the baseline to reasonably assess the accuracy of the corresponding MAP estimate. The results in Table 4.3 show that the MAP estimates of all parameters from different

Table 4.3: Example: Comparison of MAP estimates of each parameter for different combinations of the acquisition function and measured dataset with $\zeta = 20$ and 2 MPa.

Dataset	Parameter	PI	EI	LCB	VAR	LCB-VAR	ref. (Table 3.3)
SS1 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	251.24	251.21	250.65	251.45	251.19	250.00
	C_1 [MPa]	8000.00	7804.99	8000.00	7848.57	7790.70	8000.00
	γ_1	65.50	65.22	64.64	66.51	65.50	69.19
	Q_∞ [MPa]	38.64	38.41	39.64	37.66	38.68	42.11
SS1 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	250.71	250.64	250.56	250.49	250.28	250.00
	C_1 [MPa]	7967.22	7880.43	8000.00	8000.00	7915.22	8000.00
	γ_1	68.03	67.42	68.19	69.61	67.25	69.19
	Q_∞ [MPa]	40.27	39.55	41.15	37.53	39.91	42.11
	b	5.25	5.20	5.22	5.00	5.28	5.00
SS2 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	251.81	251.10	250.75	251.30	251.25	250.01
	C_1 [MPa]	7714.71	7807.73	8000.00	7671.86	7741.45	7999.77
	γ_1	64.83	64.79	65.50	64.11	65.60	70.31
	Q_∞ [MPa]	66.58	64.81	65.92	65.73	64.68	70.28
	b	5.70	5.77	5.65	5.57	5.56	5.01
SS2 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	251.12	251.23	250.67	250.05	251.33	250.01
	C_1 [MPa]	7868.71	7881.34	7906.98	7948.73	7847.83	7999.77
	γ_1	68.08	67.68	67.36	68.80	67.94	70.31
	Q_∞ [MPa]	67.15	66.73	67.92	72.52	68.46	70.28
	b	5.30	5.31	5.29	5.18	5.35	5.01

Table 4.4: Example: Comparison of the standard deviation values of each parameter for different combinations of the acquisition function and measured dataset with $\zeta = 20$ and 2 MPa.

Dataset	Parameter	PI	EI	LCB	VAR	LCB-VAR
SS1 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	2.62	2.72	3.10	2.69	2.75
	C_1 [MPa]	539.87	539.50	558.73	534.93	990.65
	γ_1	7.73	8.31	7.93	7.84	10.67
	Q_∞ [MPa]	3.67	3.94	3.85	3.61	3.56
SS1 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	1.88	2.22	3.92	2.30	1.98
	C_1 [MPa]	341.99	464.31	322.63	957.01	542.55
	γ_1	4.50	4.95	8.65	9.62	5.27
	Q_∞ [MPa]	5.17	5.91	12.60	9.96	5.99
	b	1.37	1.06	5.16	2.04	8.84
SS2 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	2.84	2.96	3.21	2.88	2.86
	C_1 [MPa]	692.79	601.52	593.90	1254.92	600.37
	γ_1	7.53	7.11	6.77	10.20	6.94
	Q_∞ [MPa]	9.12	8.87	9.18	11.16	9.01
	b	3.24	3.79	3.96	3.76	3.49
SS2 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	2.60	2.68	3.06	0.47	2.94
	C_1 [MPa]	504.51	543.85	315.44	525.99	713.21
	γ_1	5.10	5.56	3.97	5.44	6.07
	Q_∞ [MPa]	6.38	7.00	6.02	6.02	8.55
	b	1.12	1.33	1.32	0.75	1.15

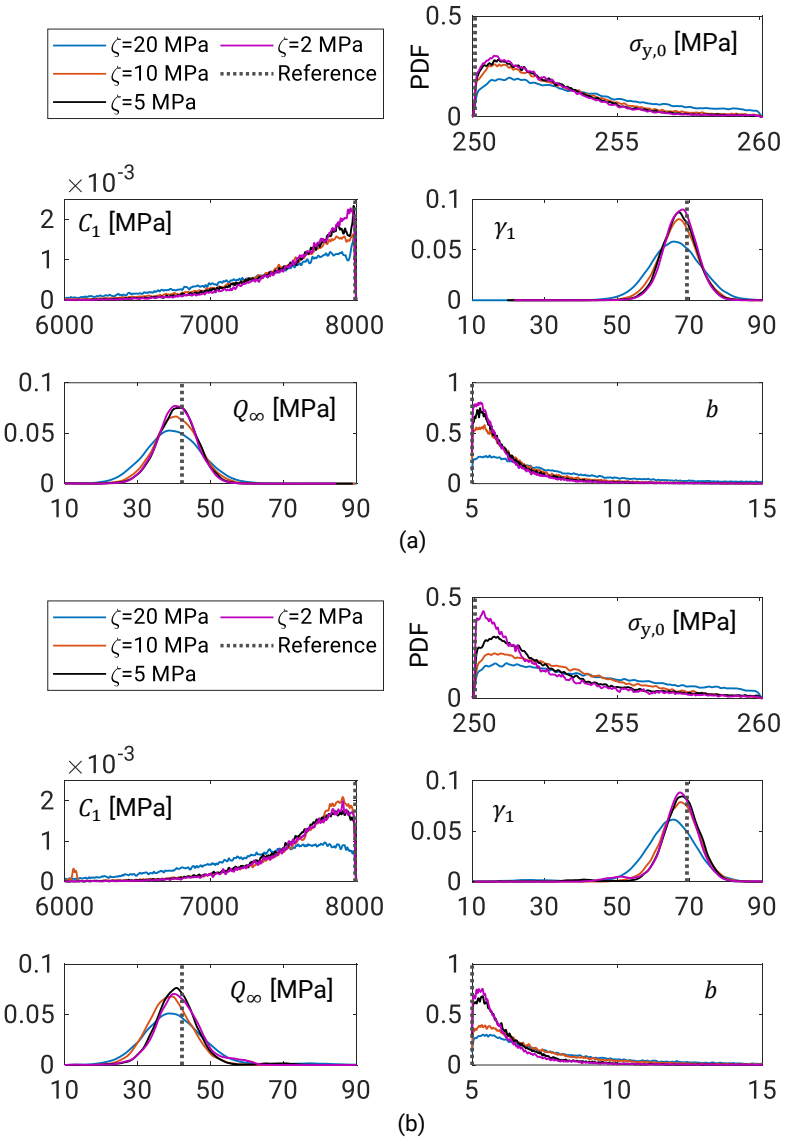

Figure 4.10: Example: Comparison of approximate posteriors of each parameter from four levels of discrepancy threshold and dataset SS1. (a) result from PI; (b) result from LCB-VAR.

acquisition functions are close to the reference values in both cases $\zeta = 20$ and 2 MPa. The results in Table 4.4 confirm that the standard deviation of each parameter tends to be small when using a small discrepancy threshold with PI and EI, indicating an increase in the reliability of posterior approximation. This, however, is not apparent when using LCB, VAR, and LCB-VAR.

Effect of discrepancy threshold

Figures 4.10 and 4.11 compare the approximate posteriors of each parameter, respectively, obtained from datasets SS1 and SS2 with four levels of discrepancy threshold. A large value of the discrepancy threshold flattens the approximate posterior of each parameter because many parameter samples that are not consistent with the measured dataset are accepted as samples of the posterior. The increasing order of PDF values at the MAP estimate of each parameter from both PI and LCB-VAR corresponds to $\zeta = 20$, 10, and (5 & 2) MPa. The difference between the posteriors of each parameter

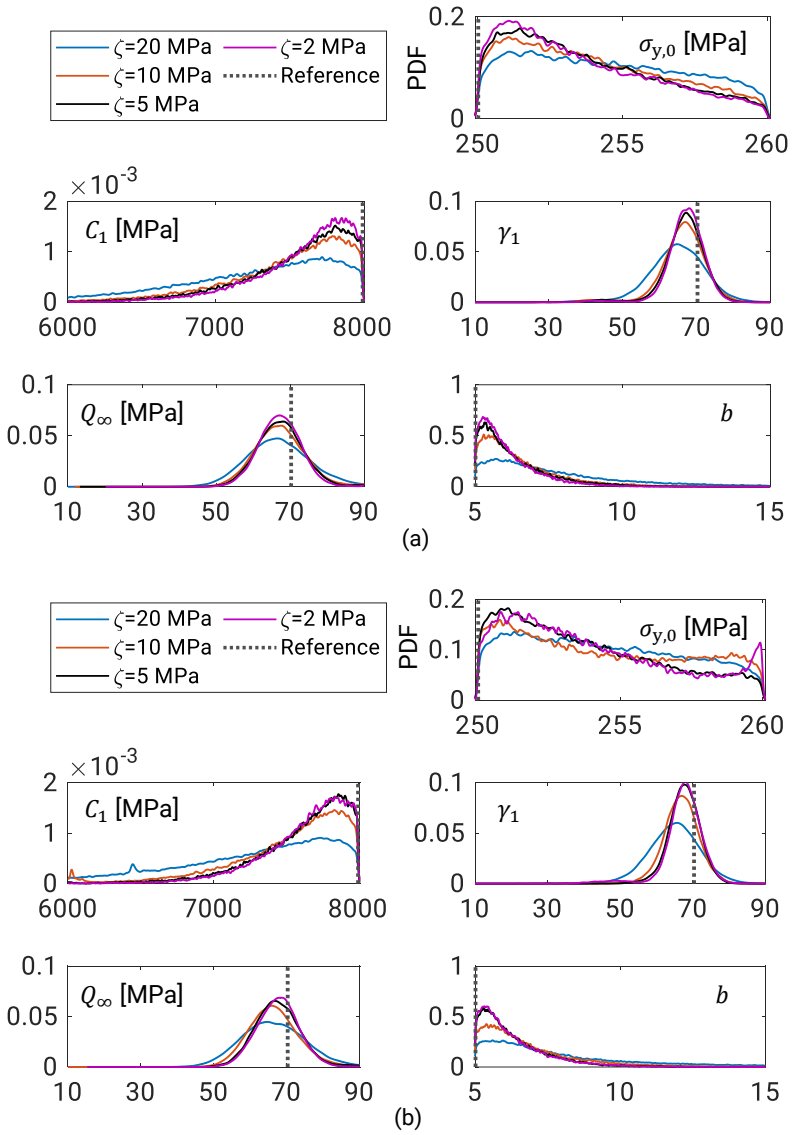

Figure 4.11: Example: Comparison of approximate posteriors of each parameter from four levels of discrepancy threshold and dataset SS2. (a) result from PI; (b) result from LCB-VAR.

by $\zeta = 5$ and 2 MPa in most cases is not pronounced, indicating that it may be impossible to improve the posterior approximation when the discrepancy threshold reaches a small-enough value.

Effect of initial training dataset

Since the GP model for the discrepancy mean function is constructed from the training dataset $\mathcal{D}$, it is desirable to investigate how different initial training datasets affects the resulting approximate posteriors. For this purpose, we generate three different initial training datasets, namely Dataset1, Dataset2, and Dataset3, and then perform ABC based on these datasets. Figures 4.12 and 4.13 compare the posteriors of each parameter associated with the three initial training datasets when ABC is performed using the experimental results from SS1 and SS2, respectively. As observed, there is no major difference in the approximate posteriors obtained from the three initial training datasets when using PI; see Figs. 4.12(a) and 4.13(a). Different initial training datasets, however, can lead

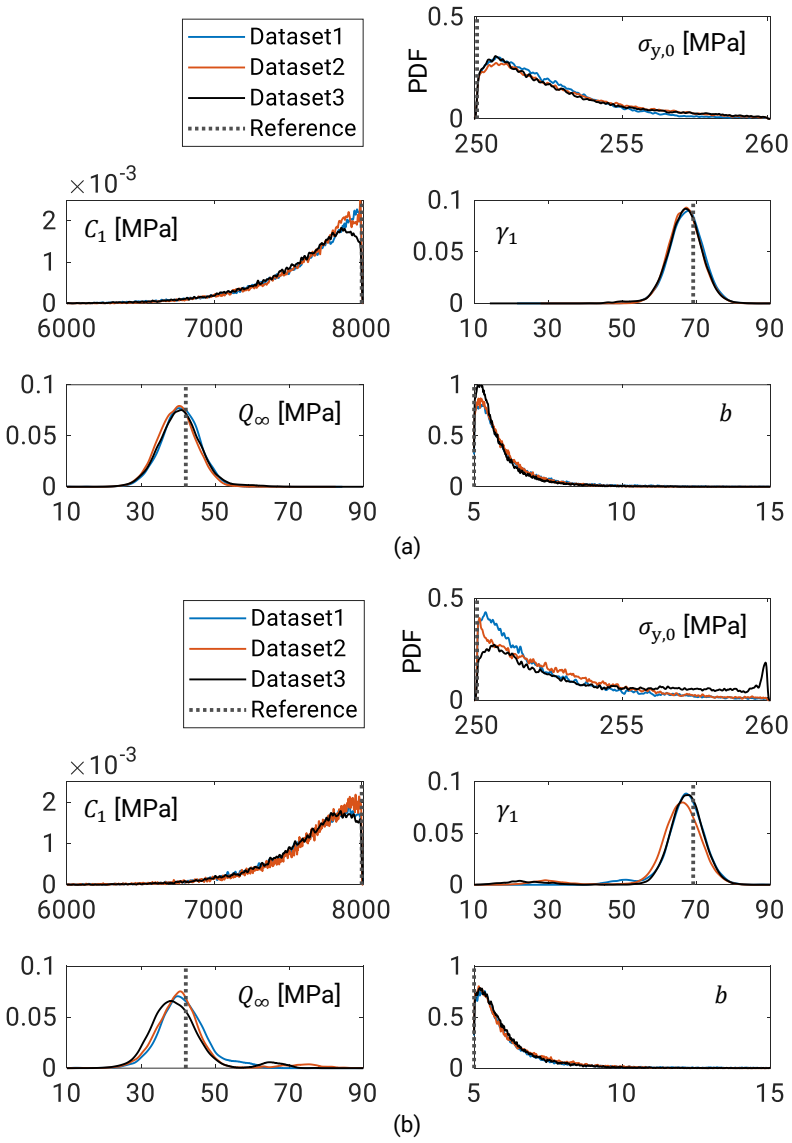

Figure 4.12: Example: Comparison of approximate posteriors of each parameter from dataset SS1, $\zeta = 2$ MPa, and three ABC attempts with different initial training datasets. (a) result from PI; (b) result from LCB-VAR.

to different approximate posteriors when using LCB-VAR; see Figs. 4.12(b) and 4.13(b).

Posterior update

When the experimental datasets from different loading histories are available, we wish to update the approximate posterior of each parameter in an online fashion when each dataset is sequentially observed. For this purpose, we carry out two updating schemes with PI and $\zeta = 2$ MPa, namely SS1-SS2 and SS2-SS1, and the evolution of approximate posteriors during each scheme is shown in Fig. 4.14. More specifically, the SS1-SS2 scheme uses the approximate posterior obtained from SS1 as prior to infer the posterior when the dataset of SS2 is observed. A similar interpretation can be made for the SS2-SS1 scheme. In most cases of Fig. 4.14, adding more experimental datasets can reduce the variance of the posterior of each parameter. An interesting observation is that the

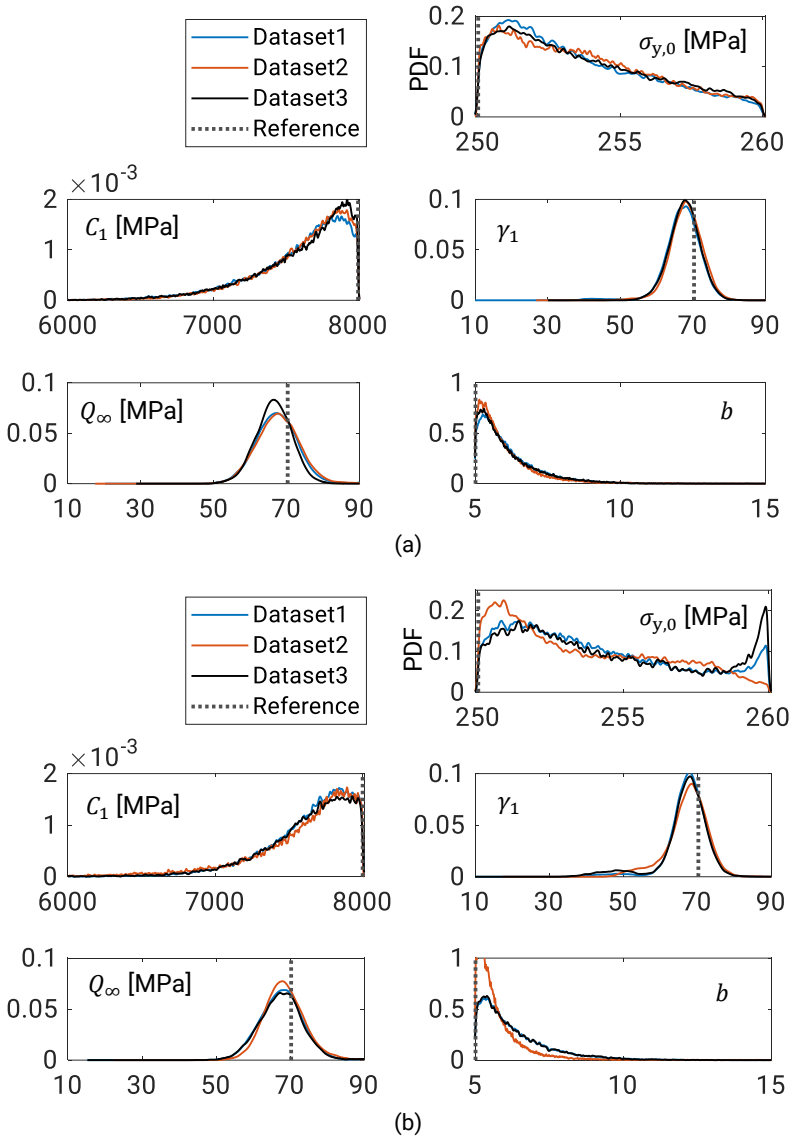

Figure 4.13: Example: Comparison of approximate posteriors of each parameter from dataset SS2, $\zeta = 2$ MPa, and three ABC attempts with different initial training datasets. (a) result from PI; (b) result from LCB-VAR.

posteriors of Q_∞ obtained from the two schemes are capable of representing the trade-off between the reference values found from individual measured datasets.

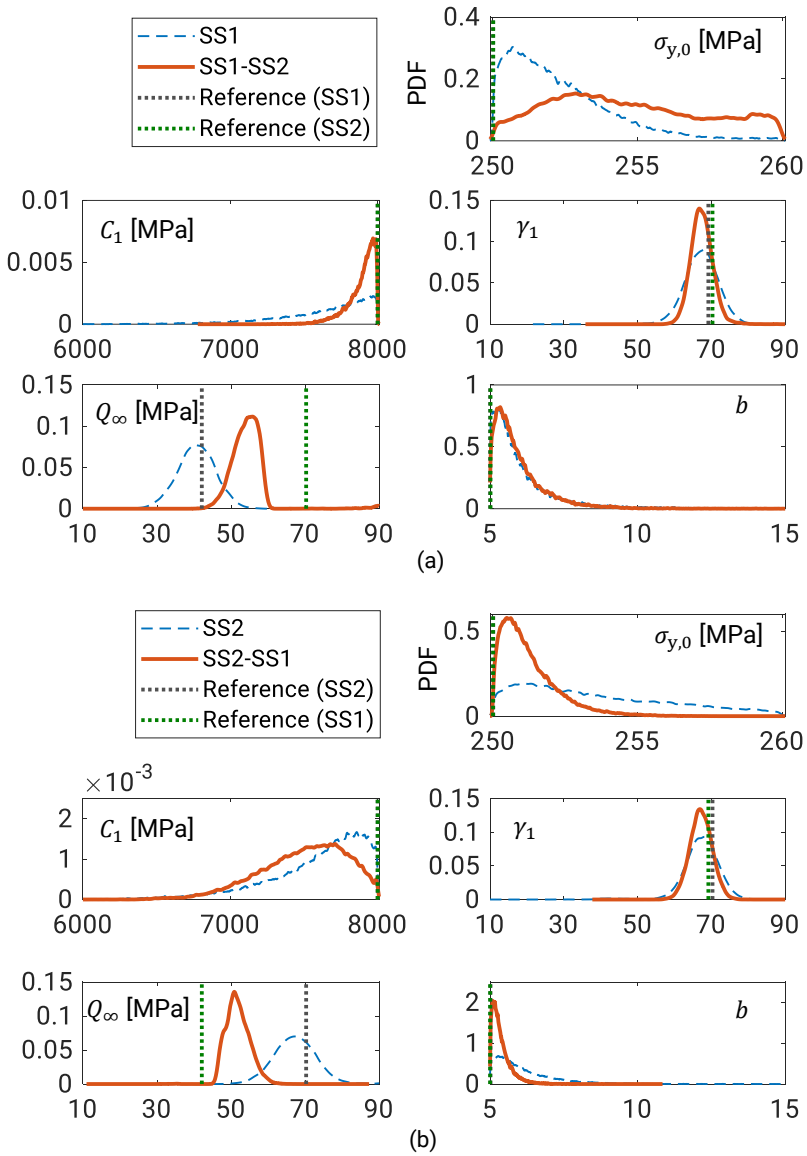

Figure 4.14: Example: Approximate posteriors obtained from two updating schemes with PI and $\zeta = 2$ MPa. (a) SS1-SS2 scheme; (b) SS2-SS1 scheme.

Prediction

To illustrate one of their applications to predictions of the cyclic elastoplastic behavior of the specimen, the approximate posteriors obtained from PI with $\zeta = 2$ MPa are used to reproduce the $\sigma - \epsilon$ curves of SS1 and SS2. For this purpose, we consider three sets of parameter values found from the posterior of each parameter, namely the MAP set as listed in Table 4.3, and the upper and lower sets of the 95% confidence interval of each parameter. Accordingly, the prediction of σ shown in Fig. 4.15 for the $\sigma - \epsilon$ curves of SS1 and SS2 at a specific time instant includes MAP, lower-bound, and upper-bound values. We also see that the MAP, lower-bound, and upper-bound curves, especially in Fig. 4.15(b), are capable of capturing the measured cure from each loading condition.

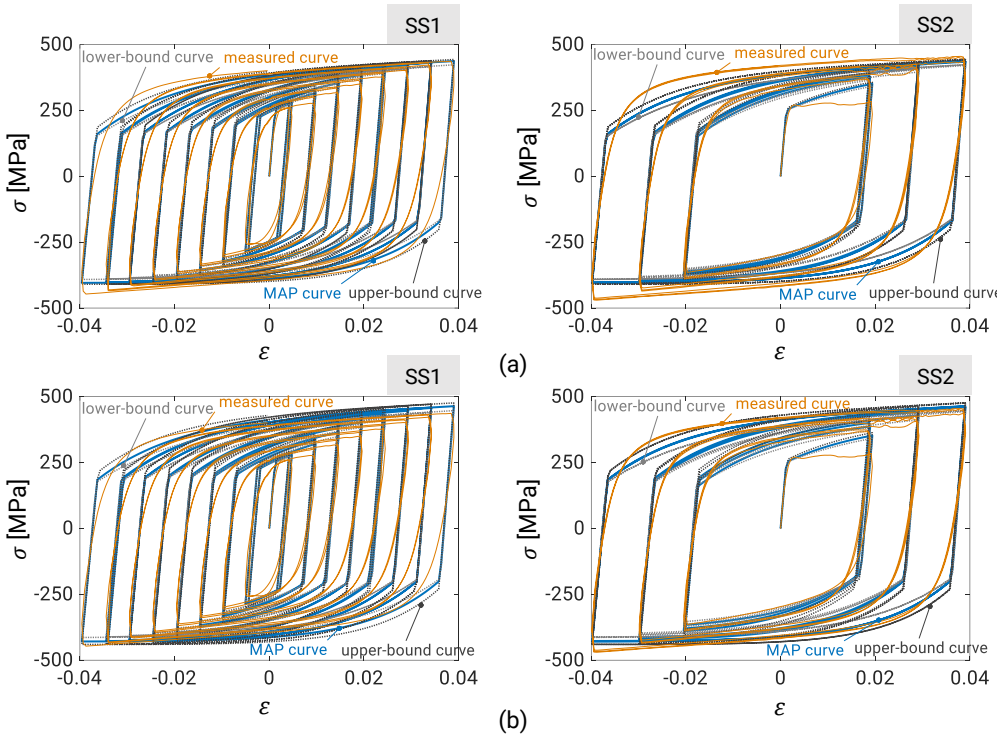

Figure 4.15: Example: Comparison of test data and model predictions for cyclic tests of the specimen with the posteriors obtained from PI. (a) parameters from dataset SS1; (b) parameters from dataset SS2.

4.5 Conclusions

This chapter has presented a BO-assisted ABC approach and its application to inversely identifying the distributions of parameters underlying statistical models as well as the distributions of cyclic elastoplastic parameters for structural steels from noise-corrupted experimental datasets. The approach may be useful for the structural engineering community as it facilitates the Bayesian inference of the parameters for a costly FE model. At the heart of the approach, ABC addresses the computationally intractable likelihood function of the Bayesian inference framework, while BO embedded in ABC tackles the costly simulation for intelligently reducing the uncertainty in the approximation of ABC posterior. The findings of this chapter are summarized as follows:

(1) Results from the illustrative examples show that the BO-assisted ABC approach successfully offers approximate posteriors that not only reproduce the true posterior with acceptable accuracy but also capture the true deterministic parameters of a known statistical model. They also confirm the prediction ability of the approximate posteriors of material parameters in modeling the cyclic elastoplastic behavior of a steel specimen subjected to different cyclic loading conditions. These results are brought by the performance of BO in refining the GP model approximating the discrepancy mean in important regions of the parameter space that improve the accuracy of approximate ABC posterior.

(2) The identification results indicate that the ABC posterior approximation can be better under a smaller discrepancy threshold. When using a sufficiently large number of MCMC samples, a small discrepancy threshold leads to the acceptance of more good samples from the prior as

samples constituting the posterior. However, it may be impossible to improve the posterior approximation when the discrepancy threshold reaches a small-enough value.

(3) It is found that the accuracy of approximate posteriors is affected by the acquisition function used in each ABC attempt. However, it is not clear to point out which acquisition function provides the best approximate posterior from the results of three illustrative examples. In addition, switching the acquisition function from LCB to VAR during an ABC attempt may improve the accuracy of approximate posteriors when using a proper value of discrepancy threshold.

(4) The new samples associated with different levels of discrepancy threshold are identical when BO is guided by PI, EI, or LCB from the same initial training dataset. Thus, the same refinement scheme is used for different levels of discrepancy threshold if PI, EI, or LCB is the acquisition function of interest. This facilitates the investigation of how different levels of discrepancy threshold affect the accuracy of the approximate posteriors.

Chapter 5

Robust design optimization of planar steel frames using Gaussian mixture model

This chapter focuses on the multi-objective RDO problem stated in Eq. (1.2) that is formulated for a planar steel frame under aleatory uncertainty in material properties, external loads, and discrete design variables. The RDO problem simultaneously minimizes the total mass of the frame, and the mean and variance values of its maximum inter-story drift, subject to constraints on design strength and serviceability requirements. The statistical estimates required for solving the problem is estimated through GMMs that are trained based upon a dataset to capture the underlying joint PDF of random input variables and the corresponding random structural responses.

5.1 Uncertainty propagation using Gaussian mixture model

5.1.1 Mean and variance of uncertain structural responses

Consider a planar steel frame structure subjected to uncertainty in random input variables $\mathbf{x} = [\mathbf{s}^T, \mathbf{r}^T]^T$, where $\mathbf{s}$ and $\mathbf{r}$ are vectors of random design variables and random parameters, respectively. Let f denote the structural response of interest. The goal of this section is to establish the cause-effect relationship between $\mathbf{x}$ and f by which the uncertainty information of f can be directly extracted from the uncertainty information of $\mathbf{x}$.

As detailed in Section 2.5, the above cause-effect relationship can be established through a GMM that captures the joint PDF $\pi(f, \mathbf{x})$ of f and $\mathbf{x}$ using a weighted sum of n_g Gaussian components. More specifically, the GMM joint PDF $\pi(f, \mathbf{x})$ is first constructed from a training dataset $\mathcal{D} = \{\mathbf{x}^i, f^i\}_{i=1}^N$, where $\mathbf{x}^i$ is a vector of random input variables and f^i is the corresponding structural response. Then, the cause-effect relationship as a conditional PDF $\pi(f; \mathbf{x})$ is derived from $\pi(f, \mathbf{x})$. Following Eq. (2.45b), $\pi(f; \mathbf{x})$ reads

$$\pi(f; \mathbf{x}) = \sum_{k=1}^{n_\mathrm{g}} W_k(\mathbf{x}) \mathcal{N}\left(f(\mathbf{x}); \mu_k(\mathbf{x}), \sigma_k^2(\mathbf{x})\right), \tag{5.1}$$

where $W_k(\mathbf{x})$, $\mu_k(\mathbf{x})$, and $\sigma_k^2(\mathbf{x})$ stated in Eqs. (2.46), (2.43a), and (2.43b) are the mixing weight, conditional mean, and conditional variance, respectively. They are characterized by a set of GMM parameters including the number of Gaussian components n_g, and the mixing proportion w_k, mean vector $\boldsymbol{\mu}_k$, and covariance matrix $\boldsymbol{\Sigma}_k$ of the k-th Gaussian component. Selecting a proper number of Gaussian components n_g is a task of model selection that can be done through BIC, while the mixing proportion, mean vector, and covariance matrix of each Gaussian component are determined for a particular n_g by the EM algorithm; see Algorithm 2.3.

Once $\pi(f;\mathbf{x})$ has been obtained, the conditional mean function (i.e., regression function) of f, denoted by $\hat{f}(\mathbf{x})$, can be derived as

$$\hat{f}(\mathbf{x}) = \mathbb{E}\left[f;\mathbf{x}\right] = \sum_{k=1}^{n_\mathrm{g}} W_k(\mathbf{x})\mu_k(\mathbf{x}). \tag{5.2}$$

To estimate the mean and variance of $\hat{f}(\mathbf{x})$ under the effect of the randomness in the input variables $\mathbf{x} = [x_1, \ldots, x_d]^T$, we utilize the second-order Taylor expansion [31, 41] with respect to $\mathbf{x}$ at the mean vector $\boldsymbol{\mu}_x = [\mu_{x_1}, \ldots, \mu_{x_d}]^T$, such that

$$\mathbb{E}\left[\hat{f}(\mathbf{x})\right] = \hat{f}(\boldsymbol{\mu}_x) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\left.\frac{\partial^2 \hat{f}}{\partial x_i x_j}\right|_{\boldsymbol{\mu}_x}\sigma_{ij}, \tag{5.3}$$

$$\begin{aligned}\mathbb{V}\left[\hat{f}(\mathbf{x})\right] = &\sum_{i=1}^{d}\sum_{j=1}^{d}\left.\frac{\partial \hat{f}}{\partial x_i}\right|_{\boldsymbol{\mu}_x}\left.\frac{\partial \hat{f}}{\partial x_j}\right|_{\boldsymbol{\mu}_x}\sigma_{ij} + \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\left.\frac{\partial \hat{f}}{\partial x_i}\right|_{\boldsymbol{\mu}_x}\left.\frac{\partial^2 \hat{f}}{\partial x_j x_k}\right|_{\boldsymbol{\mu}_x}\sigma_{ijk} \\ &+ \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d}\left.\frac{\partial^2 \hat{f}}{\partial x_i x_j}\right|_{\boldsymbol{\mu}_x}\left.\frac{\partial^2 \hat{f}}{\partial x_k x_l}\right|_{\boldsymbol{\mu}_x}(\sigma_{ijkl} - \sigma_{ij}\sigma_{kl}),\end{aligned} \tag{5.4}$$

where σ_{ij} is the covariance of x_i and x_j; σ_{ijk} is the third-order co-moment of x_i, x_j, and x_k; and σ_{ijkl} is the fourth-order co-moment of x_i, x_j, x_k, and x_l. The co-moments σ_{ijk} and σ_{ijkl} are defined in a similar way of σ_{ij} in Eq. (2.7). The calculations of the mean and variance of $\hat{f}(\mathbf{x})$ are based on the gradient and Hessian of the GMM of interest, which is the focus of the next section.

5.1.2 Gradient and Hessian of GMMs

Let $\mathbf{y} = [y_1, \ldots, y_M]^T$ denote an M-dimensional vector generated from the M-variate Gaussian $\pi(\mathbf{y}) = \mathcal{N}(\mathbf{y};\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_y)$, which reads

$$\pi(\mathbf{y}) = \frac{1}{\det(\boldsymbol{\Sigma}_y)^{1/2}(2\pi)^{M/2}}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right]. \tag{5.5}$$

Let $\boldsymbol{\Sigma}_y = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ be the singular value decomposition of the covariance matrix $\boldsymbol{\Sigma}_y$, where $\mathbf{U}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a non-singular diagonal matrix with λ_m at the m-th diagonal component. Let $\mathbf{z} = \mathbf{U}^T(\mathbf{y}-\boldsymbol{\mu}_y)$ so that $\partial z_e/\partial y_m = u_{me}$, where $e \in \{1, \ldots, M\}$ and u_{me} denotes

the (m,e)-th element of $\mathbf{U}$. The first derivatives of $\pi(\mathbf{y})$, therefore, read

$$\begin{aligned}\frac{\partial \pi(\mathbf{y})}{\partial y_m} =&\pi(\mathbf{y})\frac{\partial}{\partial y_m}\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right]=\pi(\mathbf{y})\frac{\partial}{\partial y_m}(-\frac{1}{2}\mathbf{z}^T\boldsymbol{\Lambda}^{-1}\mathbf{z})\\ =&\pi(\mathbf{y})\sum_{e=1}^{M}\frac{\partial}{\partial z_e}\left(-\frac{1}{2}\sum_{m=1}^{M}\frac{z_m^2}{\lambda_m}\right)\frac{\partial z_e}{\partial y_m}=-\pi(\mathbf{y})\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{me},\end{aligned} \tag{5.6}$$

which is the m-th element of vector $-\pi(\mathbf{y})\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{z}$. Thus, the gradient of $\pi(\mathbf{y})$ reads

$$\nabla\pi(\mathbf{y})=\pi(\mathbf{y})\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\mu}_y-\mathbf{y}). \tag{5.7}$$

Let $l\in\{1,\ldots,M\}$. Taking the second derivatives of $\pi(\mathbf{y})$ leads to

$$\begin{aligned}\frac{\partial}{\partial y_l}\left(\frac{\partial \pi(\mathbf{y})}{\partial y_m}\right)=&\frac{\partial}{\partial y_l}\left(-\pi(\mathbf{y})\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{me}\right)\\ =&-\pi(\mathbf{y})\frac{\partial}{\partial y_l}\left(\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{me}\right)-\frac{\partial\pi(\mathbf{y})}{\partial y_l}\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{me}\\ =&-\pi(\mathbf{y})\left(\sum_{e=1}^{M}\frac{u_{me}}{\lambda_e}\frac{\partial z_e}{\partial y_l}\right)+\pi(\mathbf{y})\left(\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{le}\right)\left(\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{me}\right)\\ =&-\pi(\mathbf{y})\left(\sum_{e=1}^{M}\frac{u_{me}}{\lambda_e}u_{le}\right)+\pi(\mathbf{y})\left(\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{le}\right)\left(\sum_{e=1}^{M}\frac{z_e}{\lambda_e}u_{me}\right),\end{aligned} \tag{5.8}$$

which is the (l,m)-th element of the matrix $-\pi(\mathbf{y})\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T+[\pi(\mathbf{y})]^{-1}\left[\nabla\pi(\mathbf{y})\right]\left[\nabla\pi(\mathbf{y})\right]^T$; see Ref. [122]. Thus, the Hessian of $\pi(\mathbf{y})$ reads

$$(\nabla\nabla^T)\pi(\mathbf{y})=-\pi(\mathbf{y})\boldsymbol{\Sigma}_y^{-1}+[\pi(\mathbf{y})]^{-1}\left[\nabla\pi(\mathbf{y})\right]\left[\nabla\pi(\mathbf{y})\right]^T=-\pi(\mathbf{y})\boldsymbol{\Sigma}_y^{-1}\left[-\boldsymbol{\Sigma}_y+(\mathbf{y}-\boldsymbol{\mu}_y)(\mathbf{y}-\boldsymbol{\mu}_y)^T\right]\boldsymbol{\Sigma}_y^{-1}. \tag{5.9}$$

Utilizing the results in Eqs. (5.7) and (5.9), the gradient and Hessian of the GMM $\sum_{k=1}^{n_{\mathrm{g}}} w_k\mathcal{N}\left(\mathbf{x};\boldsymbol{\mu}_{x,k},\boldsymbol{\Sigma}_{xx,k}\right)$ in Eq. (2.45a) can be derived as

$$\mathbf{g}=\sum_{k=1}^{n_{\mathrm{g}}} w_k\mathcal{N}\left(\mathbf{x};\boldsymbol{\mu}_{x,k},\boldsymbol{\Sigma}_{xx,k}\right)\boldsymbol{\Sigma}_{xx,k}^{-1}(\boldsymbol{\mu}_{x,k}-\mathbf{x}), \tag{5.10}$$

$$\mathbf{H}=\sum_{k=1}^{n_{\mathrm{g}}} w_k\mathcal{N}\left(\mathbf{x};\boldsymbol{\mu}_{x,k},\boldsymbol{\Sigma}_{xx,k}\right)\boldsymbol{\Sigma}_{xx,k}^{-1}\left[-\boldsymbol{\Sigma}_{xx,k}+(\mathbf{x}-\boldsymbol{\mu}_{x,k})(\mathbf{x}-\boldsymbol{\mu}_{x,k})^T\right]\boldsymbol{\Sigma}_{xx,k}^{-1}. \tag{5.11}$$

By substituting $W_k(\mathbf{x})$ and $\mu_k(\mathbf{x})$, respectively, from Eqs. (2.46) and (2.43a) into Eq. (5.2) and then utilizing the results in Eqs. (5.10) and (5.11), the gradient and Hessian of $\hat{f}(\mathbf{x})$ can be derived easily. Since the mixing weight $W_k(\mathbf{x})$ in Eq. (2.46) is a fraction of two GMMs, i.e., $W_k(\mathbf{x})=$

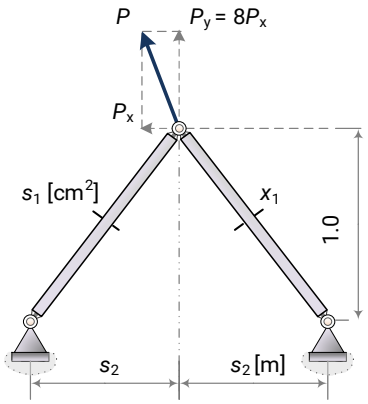

Figure 5.1: Two-bar truss.

$g(\mathbf{x})/h(\mathbf{x})$, the following quotient rules are used to calculate the gradient and Hessian of $\hat{f}(\mathbf{x})$:

$$\frac{\partial W_k}{\partial \mathbf{x}} = \frac{h\partial g/\partial \mathbf{x} - g\partial h/\partial \mathbf{x}}{h^2}, \tag{5.12a}$$

$$\frac{\partial^2 W_k}{\partial \mathbf{x}^2} = \frac{\partial^2 g/\partial \mathbf{x}^2 - 2[\partial W_k/\partial \mathbf{x}][\partial W_k/\partial h]^T - W_k[\partial^2 h/\partial \mathbf{x}^2]}{h}. \tag{5.12b}$$

5.2 Test problems

5.2.1 Two-bar truss

The statistical estimates by GMM are used to solve the RDO problem of a simple two-bar truss in Fig. 5.1, which is taken from Ref. [123]. The cross-sectional area of the truss members s_1 and the horizontal span of the truss s_2 are considered as two design variables $\mathbf{s}$. The density of truss material ρ, the external load P, and the material's yield stress $\sigma_{\mathrm{y},0}$ are three random parameters $\mathbf{r}$. Statistical properties of the design variables and random parameters are given in Table 5.1. The deterministic optimization problem of the truss formulated to minimize its total mass under constraints on the axial stress of the truss members reads

$$\begin{aligned}
\underset{\mathbf{s}}{\text{minimize}} \quad & f(\mathbf{s},\mathbf{r}) = 10^{-4}\rho s_1\sqrt{1+s_2^2} \\
\text{subject to} \quad & g_1(\mathbf{s},\mathbf{r}) = 1 - \frac{5P}{\sqrt{65}s_1\sigma_{\mathrm{y},0}}\sqrt{1+s_2^2}\left(8+\frac{1}{s_2}\right) \geq 0, \\
& g_2(\mathbf{s},\mathbf{r}) = 1 - \frac{5P}{\sqrt{65}s_1\sigma_{\mathrm{y},0}}\sqrt{1+s_2^2}\left(8-\frac{1}{s_2}\right) \geq 0, \\
& s_1 \in [0.2, 20]\,\mathrm{cm}^2,\ s_2 \in [0.1, 1.6]\,\mathrm{m}.
\end{aligned} \tag{5.13}$$

The RDO problem of the truss is formulated with a single function defined as a weighted sum of the mean and standard deviation values of the objective function in problem (5.13). This RDO

Table 5.1: Probabilistic characteristics of random variables of the two-bar truss.

Variable	Description	Mean	COV	Distribution
s_1	Cross-sectional area [cm^2]	μ_{s_1}	0.02	Normal
s_2	Horizontal span [m]	μ_{s_2}	0.02	Normal
ρ	Material density [kg/m^3]	10^4	0.20	Beta*
P	External load [kN]	800	0.25	Extreme-value type I
$\sigma_{\mathrm{y},0}$	Tensile strength [MPa]	1050	0.24	Lognormal

* Shape parameters of beta distribution are $a = b = 5$; see Section 2.2.4

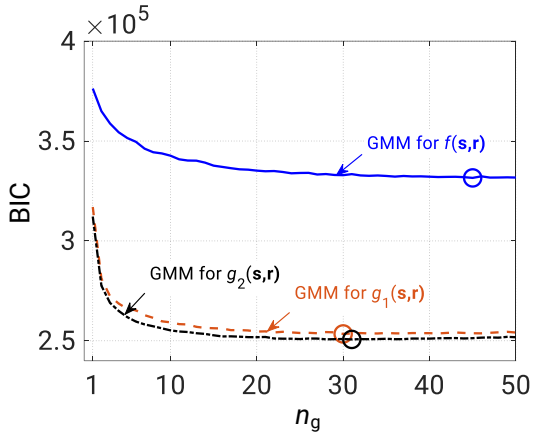

Figure 5.2: Selection of three GMMs for the two-bar truss.

problem reads

$$\begin{aligned} \underset{\mu_{s_1},\mu_{s_2}}{\text{minimize}} \quad & F(\mu_{s_1},\mu_{s_2}) = 0.5\frac{\mu_f}{10} + 0.5\frac{\sigma_f}{2} \\ \text{subject to} \quad & G_1(\mu_{s_1},\mu_{s_2}) = \beta\sigma_{g_1} - \mu_{g_1} \le 0, \\ & G_2(\mu_{s_1},\mu_{s_2}) = \beta\sigma_{g_2} - \mu_{g_2} \le 0, \\ & \mu_{s_1} \in [0.2, 20]\,\mathrm{cm}^2,\ \mu_{s_2} \in [0.1, 1.6]\,\mathrm{m}, \end{aligned} \tag{5.14}$$

where $\mu_{(\cdot)}$ and $\sigma_{(\cdot)}$ denote the mean and standard deviation of $(\cdot)$, respectively; the weight factors associated with the mean and standard deviation of the objective function are 0.5; the scaling factors assigned for the mean and standard deviation of the objective are 10 and 2, respectively; and the risk attitude factor $\beta = 3$.

To develop three GMMs for f, g_1, and g_2, a total of 8000 training samples are randomly generated by MCS. The resulting GMMs for f, g_1, and g_2 shown in Fig. 5.2 are mixtures of 45, 30, and 31 Gaussian components, respectively. The computational times required for training the GMMs for f, g_1, and g_2, using a PC with an Intel(R) i7-7700HQ 2.80 GHz CPU and 8.0 GB memory, are 142, 115, and 132 s, respectively.

Based on the obtained GMMs, Eqs. (5.3) and (5.4) are respectively used to estimate the mean and variance values of the objective and constraint functions of problem (5.14). The RDO solution to this problem is then found by a pattern search.

Table 5.2 compares the resulting RDO solution with those obtained from previous studies [123,

Table 5.2: Comparison of robust optimization results for two-bar truss.

Results	Gaussian quadrature (TPQ) [123]	MCS with 10^5 samples [124]	GMM [this study]
μ_{s_1} [cm^2]	11.5655	11.6795	11.5297
μ_{s_2} [m]	0.3771	0.3771	0.3586
F	1.2393 (1.2428)	1.2516	1.2463 (1.2292)
G_1	0 (0.0096)	(0.0012)	0 (0.0116)
G_2	-0.4978 (-0.4931)	(-0.4973)	-0.4922 (-0.5084)
No. of function calls	3672	$> 10^5$	894

Numbers in parentheses are confirmed values using MCS with 10^5 samples

124], where a tensor product Gaussian quadrature (TPQ) [123] and MCS [124] were used to estimate statistical moments of the objective and constraint functions. All RDO solutions are further confirmed by MCS with 10^5 samples. It is shown that GMM can quickly provide a reasonable RDO solution without using a complex analysis as the Gaussian quadrature method and an enormous computer resource as MCS.

It is desirable to find a more reliable solution to the RDO problem because the current RDO solutions shown in Table 5.2 still violate the constraint G_1. To do so, we replace G_1 in problem (5.14) with a new constraint function as

$$G_1(\mu_{s_1}, \mu_{s_2}) = \beta\sigma_{g_1} - \mu_{g_1} - 0.1. \tag{5.15}$$

With the new G_1, the solution to problem (5.14) is $\mu_{s_1} = 11.4146\,\text{cm}^2$ and $\mu_{s_2} = 0.4063\,\text{m}$. The corresponding objective and constraint values are $F = 1.2540$, $G_1 = 0$, and $G_2 = -0.4650$. This indicates the successful application of GMM to solving problem (5.14) with different bounds of constraints as it equally handles the randomness of design variables and parameters during its construction, and training should be carried out only once before solving the optimization problem with various bounds of constraints.

5.2.2 Steel frame

This section examines the regression and probabilistic modeling performances of GMM through characterizing the uncertain structural response of a four-story three-bay steel frame in Fig. 5.3, which is taken from Ref. [125]. External loads acting on the frame consist of dead loads D_L, short-term live load S, long-term live load L, snow load on the roof S_L, and wind load W_L. They constitute different vertical loads acting on the beam members of the frame, which are indicated in Fig. 5.2 and described in Table 5.3. The structural response of interest is incorporated into a LSF related to the maximum inter-story drift of the frame, which reads

$$g(\mathbf{s}, \mathbf{r}) = \frac{\delta_\text{max}(\mathbf{s}, \mathbf{r})}{\delta_\text{a}} - 1, \tag{5.16}$$

where δ_max is the maximum inter-story drift; $\delta_\text{a} = 11.7$ mm is the allowable upper bound for δ_max; $\mathbf{s}$ is the vector of design variables, including four pairs of the cross-sectional area and moment of inertia $\{(A_1, I_1), (A_2, I_2), (A_3, I_3), (A_4, I_4)\}$ of four groups of structural members indicated in Fig. 5.3; and

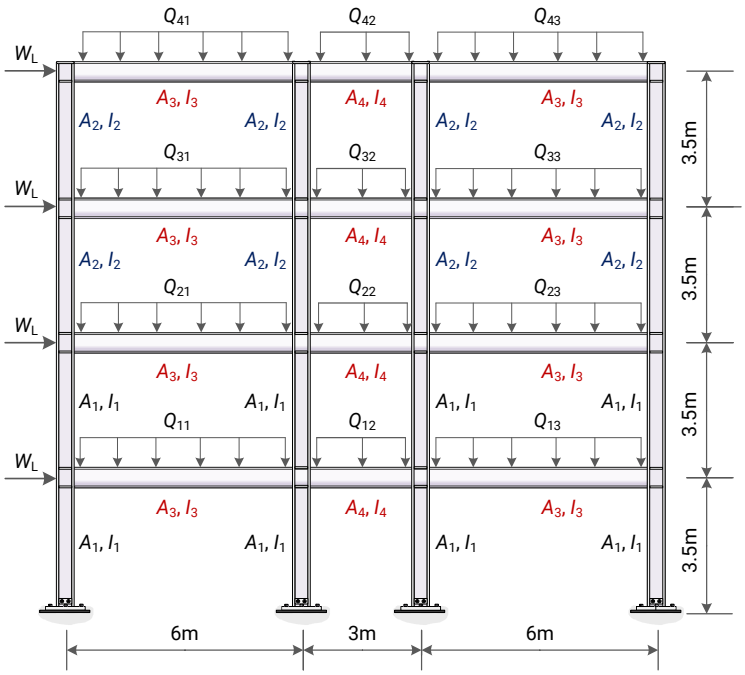

Figure 5.3: Four-story three-bay steel frame.

$\mathbf{r}$ is the vector of random parameters, including Young's modulus E and external loads W_{L}, and $(Q_{11}, Q_{12}, \ldots, Q_{43})$. The uncertainty information of $\mathbf{r}$ is specified in Table 5.4.

A training dataset of 3×10^4 samples of the design variables, random parameters, and g values is created for constructing the GMM. In particular, the samples of the design variables $\mathbf{s}$ are uniformly drawn over eight predefined intervals provided in Table 5.5, while those of the random parameters $\mathbf{r}$ are generated using MCS for the PDFs given in Table 5.4.

Table 5.3: Vertical loads applied to each beam of the four-story three-bay steel frame.

Load	Combination
Q_{11}, Q_{13}, Q_{21}, Q_{23}, Q_{31}, Q_{33}	$D_{\mathrm{L}} + S_1 + L_1$
Q_{12}, Q_{22}, Q_{32}	$D_{\mathrm{L}} + S_2 + L_2$
Q_{41}, Q_{42}, Q_{43}	$D_{\mathrm{L}} + S_{\mathrm{L}}$

Table 5.4: Probabilistic characteristics of random parameters for the four-story three-bay steel frame.

Variable	Description	Mean	COV	Distribution
E	Young's Modulus [GPa]	210	0.08	Normal
D_{L}	Dead load [kN/m]	20	0.10	Normal
S_1	Short term live load 1 [kN/m]	10	0.10	Extreme-value type I
S_2	Short term live load 2 [kN/m]	5	0.30	Extreme-value type I
L_1	Long term live load 1 [kN/m]	10	0.30	Extreme-value type I
L_2	Long term live load 2 [kN/m]	5	0.30	Extreme-value type I
S_{L}	Snow load [kN/m]	5	0.26	Extreme-value type I
W_{L}	Wind load [kN]	8	0.37	Extreme-value type I

Table 5.5: Predefined intervals for the design variables for the four-story three-bay steel frame.

Design variable	Interval [cm^2]	Design variable	Interval [cm^4]
A_1	$[46, 122]$	I_1	$[2118, 12949]$
A_2	$[29, 90]$	I_2	$[735, 6551]$
A_3	$[39, 97]$	I_3	$[4922, 26599]$
A_4	$[24, 62]$	I_4	$[1652, 9610]$

A test dataset of 2×10^4 samples is also created for obtaining a comparison between the performances of GMM and MCS using FE analyses of the frame. The mean value for each design variable is first randomly picked within the corresponding interval. Then, the samples of each design variable are randomly generated based on the assumption that each design variable is normally distributed around the corresponding picked mean with a COV of 0.05 for both the cross-sectional area and the moment of inertia. The samples of the random parameters $\mathbf{r}$ are generated in the same way as creating the training set.

Figure 5.4(a) shows the process of automatically selecting the number of Gaussians n_g using BIC. The optimal number of Gaussians found in this experiment is 14 when BIC gets its minimum value. Results in Figs. 5.4(b), (c), and (d) show that the obtained GMM has an ability to accurately extract the LSF g and characterize its probabilistic property for the frame in the presence of uncertainty in design variables $\mathbf{s}$ and random parameters $\mathbf{r}$. In particular, Fig. 5.4(b) provides a scatter of g values computed by using Eq. (5.2) (estimated) over those obtained from FE analyses (true). A high coefficient of determination $R^2 = 0.996$ indicates a good regression performance of the obtained GMM. Figures 5.4(c) and (d) show a good agreement between the true and estimated PDFs and CDFs, respectively, of the LSF g. Moreover, the mean values for the true and estimated g are 0.2342 and 0.2335, respectively, and the corresponding variance values are 0.0822 and 0.0901.

5.3 Robust design optimization of steel frames

5.3.1 Formulation of robust design optimization (RDO) problem

In this section, we extend the deterministic design optimization problem of steel frames subject to constraints on design strength, serviceability, and constructional requirements specified by the load and resistance design approach [126–128] to incorporate uncertainty and state our multi-objective RDO problem for a planar steel frame of interest. In addition to the total mass of the frame, the mean and variance of the maximum inter-story drift measuring an expected damage level are considered as other two objective functions.

In the RDO problem, the design variable vector, denoted as $\mathbf{s}^0$, represents the set of nominal values of cross-sectional area and moment of inertia of structural members, while the vector of random parameters $\mathbf{r}$ represents Young's modulus, the yield and tensile strengths, and the external loads. Since the sections for structural members are selected from a list of standard steel sections, the components of $\mathbf{s}^0$ are integer and deterministic, and the values of cross-sectional area and moment of inertia as well as the elastic and plastic sectional moduli are linked. The random vector $\mathbf{s}$ is generated in the neighborhood of $\mathbf{s}^0$ based on the specified PDF of geometric parameters such as the overall depth, flange width, web thickness, and flange thickness; see Section 5.3.2.

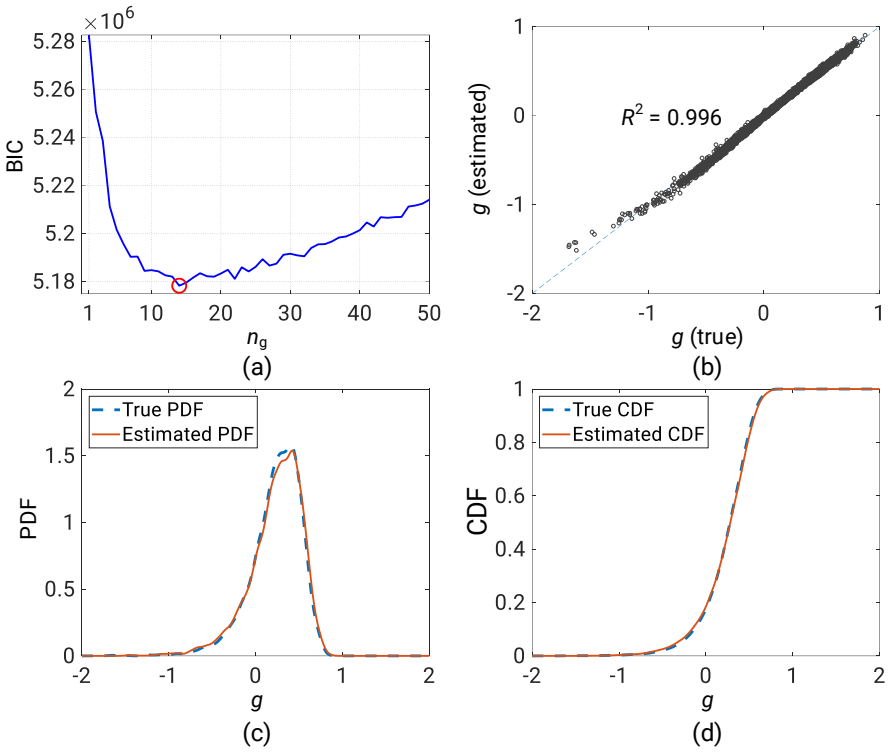

Figure 5.4: Results of GMM performances test for four-story three-bay steel frame. (a) BIC versus n_{g}; (b) Scatter plot; (c) PDF; (d) CDF.

Let n_{e}, W_i, and L_i denote the number of structural members including beams and columns of the frame, the nominal mass [kg/m] of the steel section for the i-th member, and the length of the i-th member, respectively. The first objective function of the RDO problem reads the nominal value of the total structural mass $f_1(\mathbf{s}^0)$ of the frame as

$$f_1(\mathbf{s}^0) = \sum_{i=1}^{n_{\mathrm{e}}} W_i(\mathbf{s}^0) L_i. \tag{5.17}$$

Let δ_j and H_j, respectively, denote the j-th inter-story drift and the height of the j-th story of the frame with n_{s} stories. If the assigned allowable value of the inter-story drift is $H_j/300$, the LSF $f(\mathbf{s}, \mathbf{r})$ related to the maximum inter-story drift of the steel frame is defined as

$$f(\mathbf{s}, \mathbf{r}) = \max\left(\frac{|\delta_1|}{H_1/300}, \ldots, \frac{|\delta_{n_{\mathrm{s}}}|}{H_{n_{\mathrm{s}}}/300}\right) - 1. \tag{5.18}$$

Thus, the mean $f_2(\mathbf{s}^0)$ and variance $f_3(\mathbf{s}^0)$ of $f(\mathbf{s}, \mathbf{r})$ read

$$f_2(\mathbf{s}^0) = \mathbb{E}\left[f(\mathbf{s}, \mathbf{r})\right], \quad f_3(\mathbf{s}^0) = \mathbb{V}\left[f(\mathbf{s}, \mathbf{r})\right]. \tag{5.19}$$

The constraint on the mean performance of the i-th beam-column member is defined based on

the AISC-LRFD interaction formula [24], such that

$$g_{1,i}(\mathbf{s}^0) = \mathbb{E}\left[g_{1,i}(\mathbf{s},\mathbf{r})\right] \leq 0, \quad g_{1,i}(\mathbf{s},\mathbf{r}) = \begin{cases} \frac{P_{\mathrm{r},i}}{P_{\mathrm{c},i}} + \frac{8}{9}\left(\frac{M_{\mathrm{rx},i}}{M_{\mathrm{cx},i}} + \frac{M_{\mathrm{ry},i}}{M_{\mathrm{cy},i}}\right) - 1 & \text{if } \frac{P_{\mathrm{r},i}}{P_{\mathrm{c},i}} \geq 0.2, \\ \frac{P_{\mathrm{r},i}}{2P_{\mathrm{c},i}} + \left(\frac{M_{\mathrm{rx},i}}{M_{\mathrm{cx},i}} + \frac{M_{\mathrm{ry},i}}{M_{\mathrm{cy},i}}\right) - 1 & \text{if } \frac{P_{\mathrm{r},i}}{P_{\mathrm{c},i}} < 0.2, \end{cases} \tag{5.20}$$

where P_{r} and P_{c} denote the required and available axial strengths of the structural member (either tension or compression), respectively; M_{rx} and M_{ry} are the required flexural strengths about the major axis x and the minor axis y, respectively; and M_{cx} and M_{cy} are the available flexural strengths about the major axis x and the minor axis y, respectively. The available axial and flexural strengths are estimated in accordance with chapters E and F of ANSI/AISC 360-16 [24], respectively. They require the evaluation of effective length factors K_{c} for columns of an unbraced frame. In this study, K_{c} is evaluated through the French rule [129] as

$$K_{\mathrm{c}} = \sqrt{\frac{1.6G_{\mathrm{A}}G_{\mathrm{B}} + 4(G_{\mathrm{A}} + G_{\mathrm{B}}) + 7.5}{G_{\mathrm{A}} + G_{\mathrm{B}}}}, \tag{5.21}$$

where

$$G = \frac{\sum I_{\mathrm{c}}/L_{\mathrm{c}}}{\sum I_{\mathrm{b}}/L_{\mathrm{b}}}. \tag{5.22}$$

The subscripts A and B refer to the joints at the lower and upper ends of the column being considered, respectively; and $I_{\mathrm{c(b)}}$ and $L_{\mathrm{c(b)}}$ are the moment of inertia and length of the column (beam) members connected to that joint, respectively. For a column in the first story, $G_{\mathrm{A}} = 10$ if the column is pin-supported, while $G_{\mathrm{A}} = 1$ if it is fixed to a strong foundation.

The constraint on the mean value of the total drift Δ_{t} of the frame is

$$g_2(\mathbf{s}^0) = \mathbb{E}\left[g_2(\mathbf{s},\mathbf{r})\right] \leq 0, \quad g_2(\mathbf{s},\mathbf{r}) = \frac{|\Delta_{\mathrm{t}}|}{H_{\mathrm{t}}/400} - 1, \tag{5.23}$$

where $H_{\mathrm{t}}/400$ is the allowable value of $|\Delta_{\mathrm{t}}|$ with the overall height H_{t} of the frame.

The constraint on the mean value of the long-term deflection $\Delta_{\mathrm{b},k}$ of the k-th beam member is defined by

$$g_{3,k}(\mathbf{s}^0) = \mathbb{E}\left[g_{3,k}(\mathbf{s},\mathbf{r})\right] \leq 0, \quad g_{3,k}(\mathbf{s},\mathbf{r}) = \frac{|\Delta_{\mathrm{b},k}|}{l_{\mathrm{b},k}/360} - 1, \tag{5.24}$$

where $l_{\mathrm{b},k}$ and $l_{\mathrm{b},k}/360$ are the span and the allowable long-term deflection of the k-th beam, respectively.

The strong column-weak beam constraints ensure that the columns are strong enough so that the frame exhibits a desirable yielding mechanism over multiple stories of the frame, rather than an inelastic action concentrated in column hinges at a single story [130]. Accordingly, the strong column-weak beam constraints defined for the frame are stated as

$$g_{4,m}(\mathbf{s}^0) = \mathbb{E}\left[g_{4,m}(\mathbf{s},\mathbf{r})\right] \leq 0, \quad g_{4,m}(\mathbf{s},\mathbf{r}) = 1 - \frac{\sum M_{\mathrm{pc},m}}{\sum M_{\mathrm{pb},m}}, \tag{5.25}$$

where $\sum M_{\mathrm{pc},m}$ is the sum of the flexural strengths of the upper and lower columns that constitute

the m-th joint; and $\sum M_{\mathrm{pb},m}$ is the sum of the flexural strengths of beams connected to the m-th joint. The readers may refer to ANSI/AISC 341-16 [130] for detailed calculations of these strengths.

The geometric constraints that ensure the flange width b_{bf} of beam members connected to a column should not exceed the flange width b_{cf} of the column are defined as

$$h_{1,n}(\mathbf{s}^0) = \frac{b_{\mathrm{bf},n}}{b_{\mathrm{cf},n}} - 1 \leq 0. \tag{5.26}$$

Other geometric constraints are defined to ensure the depth d_{cu} of section for a column on the upper story of a column-column joint should not exceed the depth d_{cl} for the column in the lower story, such that

$$h_{2,v}(\mathbf{s}^0) = \frac{d_{\mathrm{cu},v}}{d_{\mathrm{cl},v}} - 1 \leq 0. \tag{5.27}$$

Let n_{b}, n_{j}, n_{bc}, and n_{cc} denote the number of beams, number of joints between beams and columns, number of beam-column connections, and number of column-column joints, respectively. Then, the multi-objective RDO problem formulated for the frame can be stated as

$$\begin{aligned}
\underset{\mathbf{s}^0}{\text{minimize}} \quad & [f_1(\mathbf{s}^0), f_2(\mathbf{s}^0), f_3(\mathbf{s}^0)] \\
\text{subject to} \quad & g_{1,i}(\mathbf{s}^0) \leq 0 \quad (i = 1, \ldots, n_{\mathrm{e}}), \\
& g_2(\mathbf{s}^0) \leq 0 \\
& g_{3,k}(\mathbf{s}^0) \leq 0 \quad (k = 1, \ldots, n_{\mathrm{b}}), \\
& g_{4,m}(\mathbf{s}^0) \leq 0 \quad (m = 1, \ldots, n_{\mathrm{j}}), \\
& h_{1,n}(\mathbf{s}^0) \leq 0 \quad (n = 1, \ldots, n_{\mathrm{bc}}), \\
& h_{2,v}(\mathbf{s}^0) \leq 0 \quad (v = 1, \ldots, n_{\mathrm{cc}}).
\end{aligned} \tag{5.28}$$

In our design optimization problem, column and beam members of the frame are further classified into b_1 and b_2 groups, respectively. The design variables $\mathbf{s}^0$ represent the nominal values of the cross-sectional area and moment of inertia that are linked. Without compromising the structural integrity, we only construct GMMs corresponding to $f(\mathbf{s}, \mathbf{r})$, $g_1(\mathbf{s}, \mathbf{r})$ values associated with b_1 column groups (each is the maximum value among those for columns in the same group), $g_1(\mathbf{s}, \mathbf{r})$ values associated with b_2 beam groups, $g_2(\mathbf{s}, \mathbf{r})$, and $g_3(\mathbf{s}, \mathbf{r})$ values corresponding to b_2 beam groups. Since it is trivial to calculate g_4, h_1, and h_2 from the list of standard steel sections, they are assigned as inequality constraints to the optimization problem directly. Moreover, the nominal values linked with $\mathbf{s}^0$ are simply used for calculating the total mass $f_1(\mathbf{s}^0)$ because the uncertainty of cross-sectional dimensions are normally distributed around the these values.

5.3.2 Uncertainty characteristic

We limit uncertainty considered in our optimization problem to aleatory type that is associated with the randomness of the random variables $\mathbf{s}$ and $\mathbf{r}$. The randomness properties of geometric properties related to the design variable vector $\mathbf{s}$, including the overall depth d, flange width b_{f}, web thickness t_{w}, and flange thickness t_{f} of standard steel sections, are specified in Table 1.1. Those of the random parameters $\mathbf{r}$, i.e., material properties and external loads are reported in Table 1.2.

Table 5.6: Parameters for NSGA-II.

Parameter	Value
Population size	1000
Probability of mutation	20%
Maximum number of generations	100
Fitness function tolerance	10^{-6}
Constraint tolerance	10^{-6}
Parallel computing	true

5.3.3 Preparation of training dataset

As previously mentioned, the construction of GMM requires a training dataset of samples of random vectors $\mathbf{s}$, random parameters $\mathbf{r}$, and the corresponding structural response of interest, which is obtained through a linear elastic analysis of frame structures.

The training samples of the design variables $\mathbf{s}$, including geometric properties for the column and beam groups, are uniformly generated over associated intervals. MCS is then performed to generate data around each of these samples to characterize the probabilistic property of the sectional geometric dimensions given in Table 1.1. In other words, the sampling technique here consists of two stages: (1) generate uniformly a random set of samples and (2) draw data around these samples. Alternatively, only the first stage may be used because the variance values of the sectional geometric dimensions are very small. Meanwhile, the random samples of Young's modulus, yield strength, tensile strength, and external loads are generated by MCS according to their probabilistic distributions listed in Table 1.2.

To determine a proper value of the number of training samples for a particular problem, a total of five training datasets of design variables and random parameters with different numbers of samples, i.e., 10^4, 2×10^4, 3×10^4, 4×10^4, and 5×10^4, are used for training GMMs. Since problem (5.28) is formulated in a high dimensional space of $\mathbf{s}$ and $\mathbf{r}$, a large number of training samples is required to effectively generalize the design space. After GMMs are obtained, a total of 2×10^4 samples of an independent test dataset are generated and used to compute the coefficient of determination R^2, which evaluates the goodness-of-fit of each GMM and MCS using FE analyses. As a result, the GMMs associated with 5×10^4 samples provide the highest R^2 values for all design examples. Thus, they are used for solving problem (5.28).

5.3.4 Best compromise design

It is desirable to find a compromise design for verifying the structural performance after solving problem (5.28) as the objective functions f_2 and f_3 are totally conflicting. The best compromise solution can be selected from the obtained Pareto-optimal solutions using the fuzzy-based method described in Section 3.2.1. Accordingly, each value of the normalized membership function m_j in Eq. (3.6) represents a level of trade-off between the values of objective functions and the best compromise solution corresponds to the largest value of m_j.

5.3.5 Optimization procedure

To this end, the optimization procedure for solving problem (5.28) using GMMs follows four main steps.

- **Step 1**: Define and characterize the random variable vector $\mathbf{s}$ and random parameter vector $\mathbf{r}$ for the steel frame of interest; see Section 5.3.2. Randomly generate the training data for constructing GMM; see Section 5.3.3.
- **Step 2**: Set the limit n_c for the number of Gaussian components, e.g., $n_c = 50$. Start GMM for the LSF of interest at the number of Gaussian $n_g = 1$ that increases step by step to $n_g = n_c$, thereby producing a total of n_c GMMs. Among them, select the best GMM according to the minimum BIC; see Section 2.5. Generate independent samples of a test dataset and use them for obtaining the discrepancy (or R^2) between the performance of the obtained GMM and MCS using FE analyses. The accuracy of the obtained GMM performance is verified by an value of $R^2 > 0.9$ to decide whether more training data points should be added.
- **Step 3**: Employ NSGA-II for finding approximate Pareto-optimal solutions to problem (5.28). NSGA-II calls Eqs. (5.3) and (5.4) derived from the trained GMMs for estimation of f_2 and f_3, and the mean values of g_1, g_2, and g_3 at a particular candidate solution.
- **Step 4**: Find the best compromise solution among Pareto-optimal solutions based on the fuzzy set theory in Section 3.2.1 and use it for verification of structural performance.

Step 3 utilizes an elitist-strategy NSGA-II [91], available in the MATLAB R2018a Global Optimization Toolbox [106], for finding the approximate Pareto-optimal solutions to problem (5.28). Parameters characterizing NSGA-II are listed in Table 5.6. To further examine its stochastic property, NSGA-II is performed several times for each design example using PC with an Intel(R) Xeon(R) E5-2643V4 3.40 GHz CPU and 64 GB memory.

To carry out NSGA-II for the discrete optimization problem, an array of integers assigned as the initial population is randomly generated at the beginning of the algorithm. A mapping function is defined to transform the integer decision vector $\mathbf{s}^0$ to the linked random vector $\mathbf{s}$, namely the cross-sectional dimensions, cross-sectional area, moment of inertia, elastic and plastic sectional moduli. The crossover and mutation functions are also modified to guarantee that the offsprings of NSGA-II generated after each iteration are integers.

5.4 Design examples

5.4.1 Example 1: Three-story two-bay steel frame

Problem (5.28) is formulated for a three-story two-bay steel frame in Fig. 5.5. Although many deterministic optimal designs have been found for a minimum total mass of this frame using various conventional optimization algorithms [126–128], a robust design is not available for addressing uncertainty in material, geometry, and load parameters simultaneously.

The frame consists of 15 members classified into two groups: group (1) for columns and group (2) for beams, which are indicated Fig. 5.5. The cross-sectional dimensions, external loads, and material properties (i.e., yield strength, tensile strength, and Young's modulus) are sources of uncertainty.

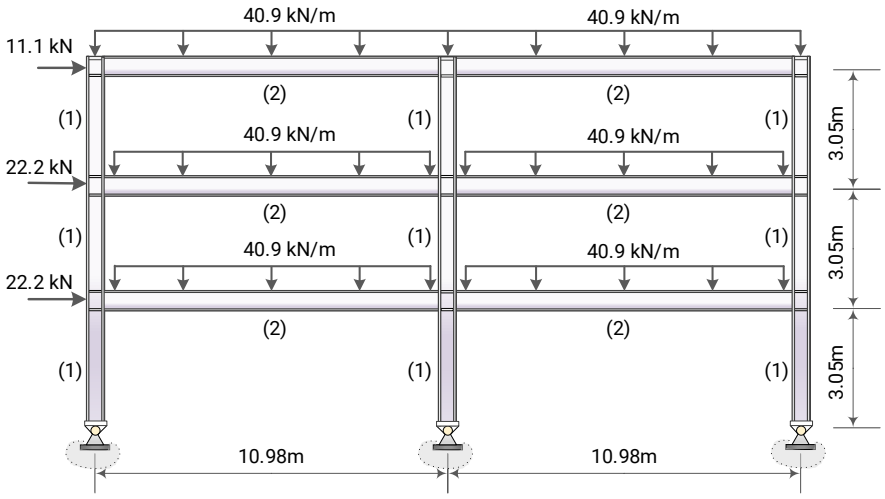

Figure 5.5: Example 1: Three-story two-bay steel frame.

Table 5.7: Example 1: List of column and beam sections for three-story two-bay steel frame.

Column		Beam	
W18 × 86	W12 × 72	W24 × 94	W18 × 65
W18 × 76	W12 × 65	W24 × 84	W18 × 60
W18 × 65	W12 × 58	W24 × 76	W16 × 77
W18 × 60	W12 × 53	W24 × 68	W16 × 67
W18 × 55	W12 × 50	W24 × 62	W16 × 57
W16 × 89	W12 × 45	W24 × 55	W16 × 50
W16 × 77	W12 × 40	W21 × 93	W16 × 45
W16 × 67	W10 × 77	W21 × 83	W16 × 40
W16 × 57	W10 × 68	W21 × 73	W14 × 61
W16 × 50	W10 × 60	W21 × 68	W14 × 53
W14 × 90	W10 × 54	W21 × 62	W14 × 48
W14 × 82	W10 × 49	W21 × 57	W14 × 38
W14 × 74	W10 × 45	W21 × 55	W14 × 34
W14 × 68	W8 × 40	W21 × 50	W14 × 30
W14 × 61	W8 × 35	W21 × 48	W14 × 26
W14 × 53	W8 × 31	W18 × 71	W14 × 22

The nominal values of cross-sectional dimensions of the members in each group are extracted from a list of American wide flange standard steel sections in Table 5.7. The nominal mass [kg/m] of each section can be found in the AISC shapes database [3]. The nominal values of the yield stress, tensile strength, and Young's modulus of the steel material are 250 MPa, 400 MPa, and 200 GPa, respectively. The mean values of external loads are provided in Fig. 5.5. The probabilistic properties of these parameters are described in Table 1.2.

The intervals of the cross-sectional dimensions for the column group are $d \in [150, 500]$ mm, $b_{\mathrm{f}} \in [140, 400]$ mm, $t_{\mathrm{w}} \in [5.5, 16.0]$ mm, and $t_{\mathrm{f}} \in [6, 24]$ mm, and those for the beam group include $d \in [350, 620]$ mm, $b_{\mathrm{f}} \in [160, 250]$ mm, $t_{\mathrm{w}} \in [5.5, 16.0]$ mm, and $t_{\mathrm{f}} \in [9, 24]$ mm. After a training dataset is generated for the frame, a total of five GMMs are constructed for f (GMM-1), $g_{1,1}$ for the column group (GMM-2), $g_{1,2}$ for the beam group (GMM-3), g_2 (GMM-4), and g_3 (GMM-5). Figure 5.6

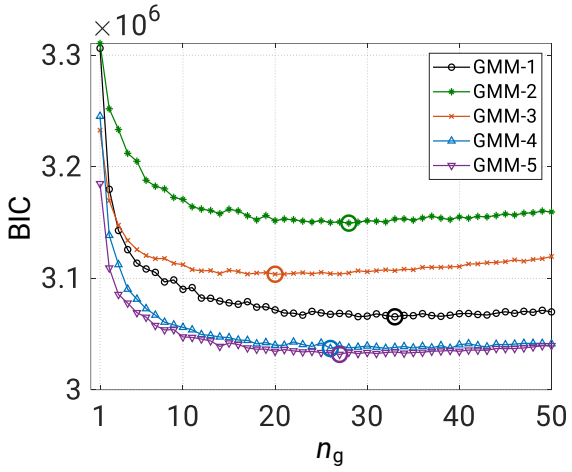

Figure 5.6: Example 1: Selection of five GMMs for three-story two-bay steel frame.

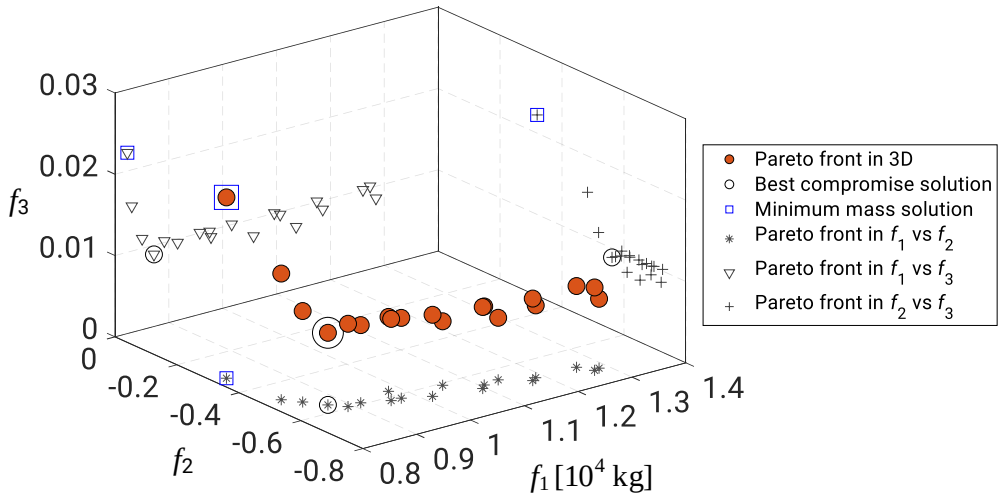

Figure 5.7: Example 1: Pareto front and the best compromise design for three-story two-bay steel frame.

Table 5.8: Example 1: Comparison of optimization results for three-story two-bay steel frame.

Group ID	Deterministic			NSGA-II	
	Ref. [126]	Ref. [127]	Ref. [128]	min. mass	best compromise
Column (1)	W10 × 60	W10 × 49	W10 × 68	W14 × 53	W18 × 65
Beam (2)	W24 × 62	W24 × 62	W21 × 55	W24 × 62	W24 × 62
f_1 [kg]	8504	8069	8174	8229	8724
f_2	–	–	–	−0.319	−0.561
f_3	–	–	–	0.023	0.010
f_2 (MCS)	–	–	–	−0.317	−0.560
f_3 (MCS)	–	–	–	0.028	0.012
$g_{1,1}$ (MCS)	–	–	–	−0.128	−0.303
$g_{1,2}$ (MCS)	–	–	–	−0.108	−0.149
g_2 (MCS)	–	–	–	−0.617	−0.739
g_3 (MCS)	–	–	–	−0.317	−0.395

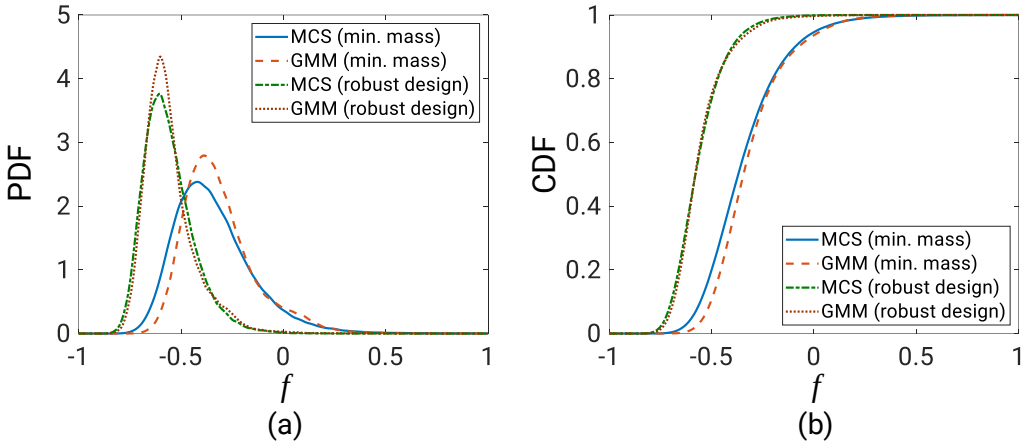

Figure 5.8: Example 1: Statistical properties of the maximum inter-story drift for the min. mass and robust designs of three-story two-bay steel frame. (a) PDFs; (b) CDFs.

shows the selection of the optimal number of Gaussian components for each GMM using BIC. Accordingly, the optimal numbers of Gaussian components are 33, 28, 20, 26, and 27, respectively, for GMM-1, GMM-2, GMM-3, GMM-4, and GMM-5. The variation ranges of the mixing proportions for GMM-1, GMM-2, GMM-3, GMM-4, and GMM-5 are found as $[0.0077, 0.0599]$, $[0.0098, 0.0954]$, $[0.0266, 0.1083]$, $[0.0074, 0.0963]$, and $[0.0103, 0.0596]$, respectively.

After the GMMs are obtained, Eq. (5.3) is used to evaluate f_2 and the mean values of g_1, g_2, and g_3 when the steel section is fixed, while the mean values of the design variables and random parameters are directly determined from the corresponding nominal values. The variance f_3 is computed using Eq. (5.4).

NSGA-II is performed three times. A set of 19 resulting Pareto-optimal solutions, the best compromise solution, and a solution with the minimum value of the structural mass are identically obtained as depicted in Fig. 5.7. Table 5.8 lists the sets of steel sections for the column and beam groups of the minimum mass and best compromise designs. They are also compared with the previous deterministic optimal designs. It is found that the minimum total mass of the frame is 8229 kg that outperforms the design by Pezeshk et al. [126] and is 2% and 0.7% larger than the designs by Toğan [127] and Maheri and Narimani [128], respectively. The best compromise design has the largest total mass of 8724 kg. This design, however, shows a trade-off between the three objective functions, and hence provides a very small variance $f_3 = 0.010$ and reasonable structure design performance with $f_2 = -0.561$, which is far enough from the limit state value $f_2 = 0$.

To confirm their actual robustness and feasibility, the obtained robust optimal designs are used to characterize the probabilistic property of the LSF regarding the maximum inter-story drift f in Eq. (5.18). The mean f_2, variance f_3, and LSFs of $g_{1,1}$, $g_{1,2}$, g_2, and g_3 for the minimum mass and the best compromise designs are evaluated by MCS with 10^5 samples. Results in Table 5.8 confirm the feasibility of the obtained robust optimal designs as the associated constraint values are all negative. The mean f_2 and variance f_3 by MCS, respectively, are -0.560 and 0.012 for the best compromise design, and -0.317 and 0.028 for the minimum structural mass design, which are very close to those obtained from the GMM-1. Figure 5.8 compares the PDFs and CDFs of the maximum inter-story drift limit state for the min. mass and best compromise designs generated by MCS and GMM-1. A good agreement between the PDFs and CDFs generated by GMM-1 and those by MCS

is observed. The robustness levels of the two designs are clearly distinguished.

5.4.2 Example 2: Six-story two-bay steel frame

This design example formulates problem (5.28) for a six-story two-bay steel frame in Fig. 5.9, which is taken from Ref. [131]. The frame includes 30 members classified into six column groups [(1) to (6)] and two beam groups [(7) and (8)]. The cross-sections for the members in beam and column groups are selected from the list in Table 5.9. Uncertainty and its characteristic for all random input variables are described in the same way as those for the three-story two-bay steel frame in Example 1.

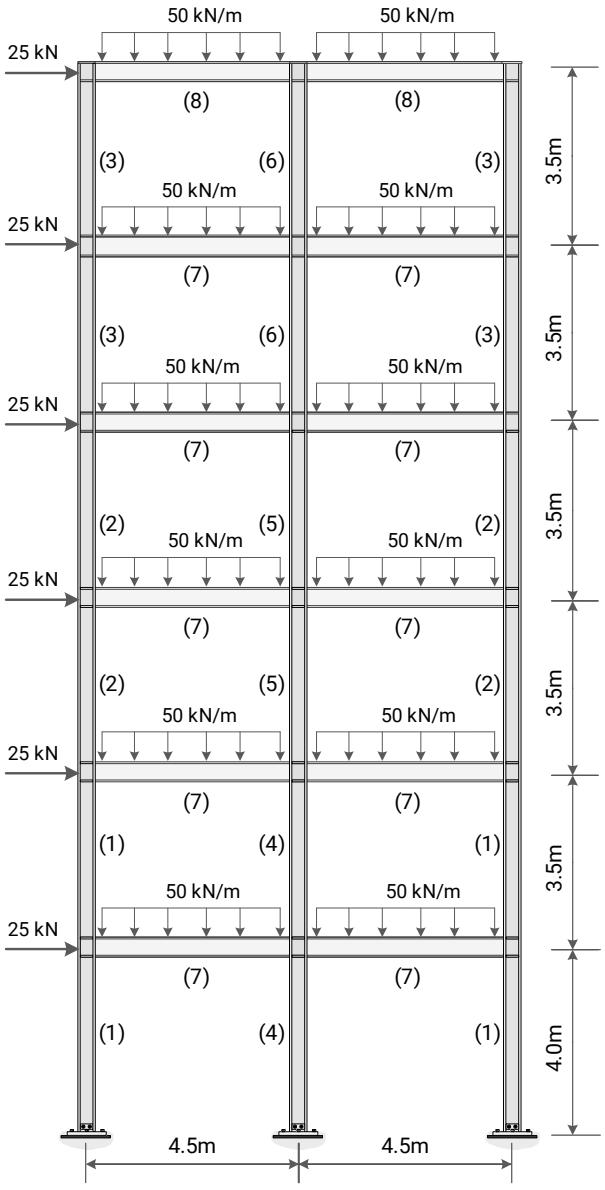

Figure 5.9: Example 2: Six-story two-bay steel frame.

Table 5.9: Example 2: List of column and beam sections for six-story two-bay steel frame.

Column			Beam		
W24 × 84	W16 × 50	W12 × 50	W24 × 62	W16 × 50	W12 × 40
W24 × 76	W16 × 40	W12 × 45	W24 × 55	W16 × 45	W12 × 35
W24 × 68	W14 × 99	W12 × 40	W21 × 62	W16 × 40	W12 × 30
W24 × 62	W14 × 90	W10 × 68	W21 × 55	W16 × 36	W12 × 26
W21 × 93	W14 × 82	W10 × 60	W21 × 48	W16 × 31	W12 × 22
W21 × 83	W14 × 74	W10 × 54	W21 × 57	W16 × 26	W12 × 19
W21 × 73	W14 × 68	W10 × 49	W21 × 50	W14 × 43	W12 × 16
W21 × 68	W14 × 61	W10 × 45	W18 × 71	W14 × 38	W12 × 14
W18 × 86	W14 × 53	W10 × 39	W18 × 65	W14 × 34	W10 × 33
W18 × 76	W14 × 48	W8 × 40	W18 × 60	W14 × 30	W10 × 30
W18 × 65	W12 × 79	W8 × 35	W18 × 55	W14 × 26	W10 × 26
W18 × 60	W12 × 72	W8 × 31	W18 × 50	W14 × 22	W10 × 22
W18 × 55	W12 × 65	W6 × 25	W18 × 46	W12 × 53	W10 × 19
W16 × 77	W12 × 58	W6 × 20	W18 × 40	W12 × 50	W10 × 17
W16 × 67	W12 × 53	W6 × 15	W18 × 35	W12 × 45	W10 × 15
W16 × 57			W16 × 57		

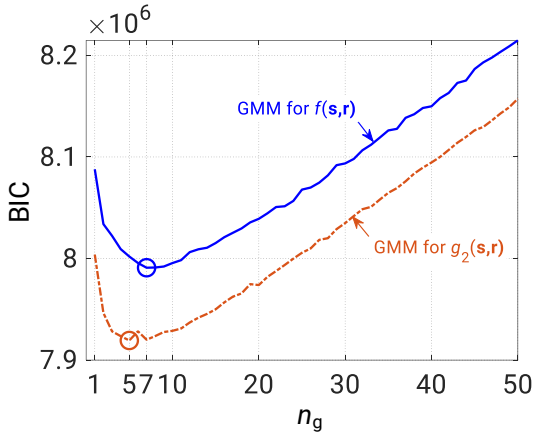

Figure 5.10: Example 2: Selection of GMMs for f and g_2 of six-story two-bay steel frame.

Figure 5.10 shows the selection of the GMMs associated with f and g_2 using BIC. Accordingly, the mixtures of seven and five Gaussian components are constructed for modeling f and g_2. The GMMs associated with the remaining LSFs of the frame are constructed in a similar way.

Since there are many groups of columns and beams in this design example, NSGA-II is performed 25 times to examine the statistical characteristics of the resulting robust optimal solutions. Figure 5.11 shows the variation in the total structural mass for the minimum mass and best compromise solutions. The variation ranges for the total mass values associated with the minimum mass and best compromise solutions are $[7460, 7739]$ and $[8520, 8694]$ kg, respectively. Although there exists a variation in the solution of 25 NSGA-II attempts, the differences in the minimum mass and in the best compromise solutions among the obtained solutions are not much higher than 3.7% and 2.0%, respectively.

Figures 5.12(a), (b), and (c) show a set of 303, 218, and 284 Pareto-optimal solutions obtained from the 1st, 2nd, and 3rd NSGA-II attempts, respectively. Although there are differences in the

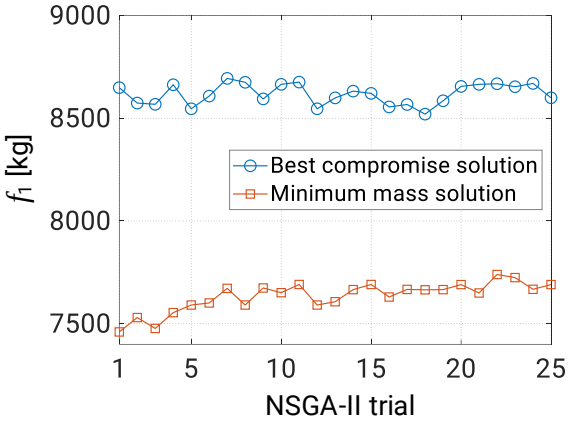

Figure 5.11: Example 2: Variation in the total structural mass due to different NSGA-II trials for six-story two-bay steel frame.

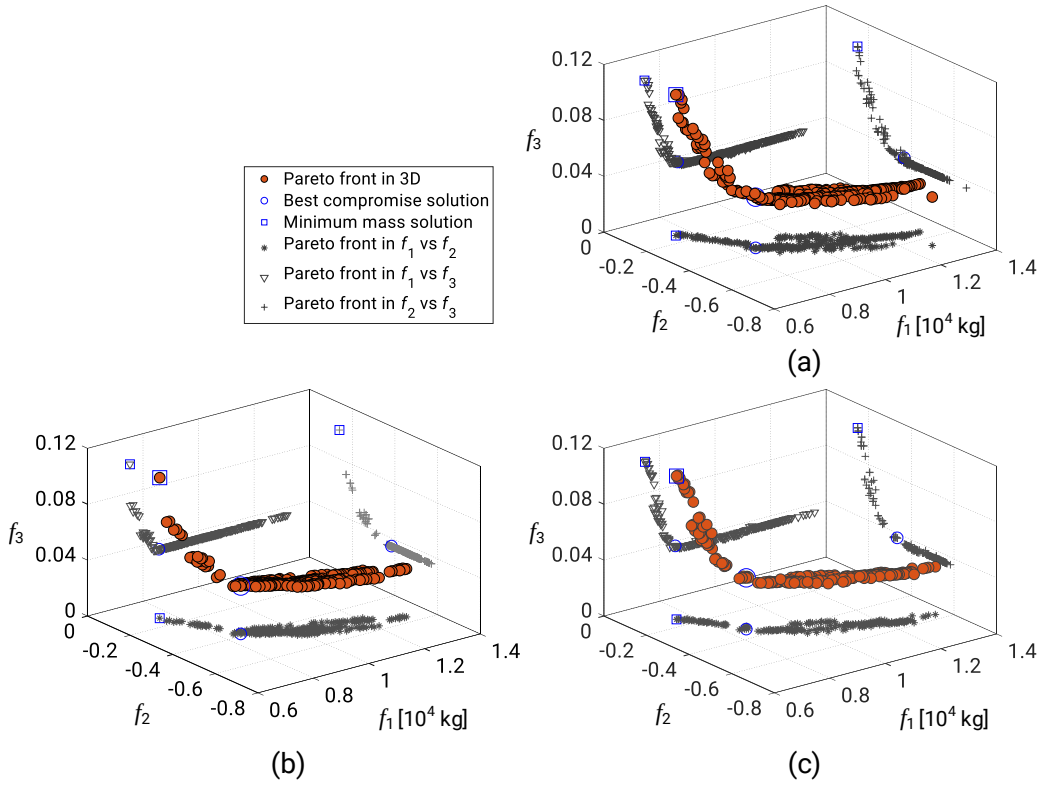

Figure 5.12: Example 2: Pareto front and the best compromise design for six-story two-bay steel frame. (a) 1st NSGA-II; (b) 2nd NSGA-II; (c) 3rd NSGA-II.

Table 5.10: Example 2: Comparison of optimization results for six-story two-bay steel frame.

Group ID	Deterministic		1st NSGA-II		2nd NSGA-II		3rd NSGA-II	
	Ref. [131]	Ref. [131]	min. mass	best compromise	min. mass	best compromise	min. mass	best compromise
Column (1)	W16 × 57	W18 × 55	W18 × 60	W18 × 60	W18 × 60	W24 × 62	W18 × 55	W24 × 62
Column (2)	W16 × 40	W12 × 50	W16 × 40	W16 × 40	W14 × 48	W16 × 40	W16 × 40	W14 × 53
Column (3)	W10 × 39	W8 × 31	W12 × 45	W12 × 50	W12 × 40	W12 × 40	W14 × 48	W12 × 50
Column (4)	W24 × 62	W21 × 73	W18 × 55	W18 × 60	W18 × 55	W18 × 55	W18 × 60	W24 × 62
Column (5)	W24 × 62	W18 × 65	W16 × 40	W16 × 40	W16 × 40	W16 × 50	W16 × 40	W18 × 60
Column (6)	W8 × 40	W12 × 40	W14 × 48	W14 × 53	W14 × 48	W14 × 48	W14 × 48	W18 × 55
Beam (7)	W14 × 30	W16 × 40	W18 × 35	W21 × 50	W18 × 35	W21 × 50	W18 × 35	W18 × 40
Beam (8)	W18 × 65	W14 × 22	W12 × 35	W12 × 35	W12 × 35	W12 × 40	W12 × 35	W14 × 34
f_1 [kg]	7533	7829	7460	8649	7530	8574	7477	8568
f_2	–	–	−0.147	−0.367	−0.140	−0.384	−0.148	−0.333
f_3	–	–	0.101	0.036	0.101	0.034	0.103	0.037
f_2 (MCS)	–	–	−0.172	−0.410	−0.168	−0.391	−0.175	−0.337
f_3 (MCS)	–	–	0.093	0.042	0.095	0.037	0.087	0.041
max. $g_{1,1-6}$ (MCS)	–	–	−0.119	−0.223	−0.118	−0.216	−0.163	−0.197
max. $g_{1,7-8}$ (MCS)	–	–	−0.040	−0.009	−0.040	−0.289	−0.004	−0.079
g_2 (MCS)	–	–	−0.243	−0.457	−0.239	−0.482	−0.245	−0.423
max. $g_{3,1-2}$ (MCS)	–	–	−0.135	−0.177	−0.134	−0.181	−0.162	−0.277

numbers of solutions and in the solutions themselves, the shapes of Pareto fronts obtained from the three attempts are similar.

As listed in Table 5.10, the minimum mass of the frame is 7460 kg (1st NSGA-II) that outperforms the previous deterministic optimal designs obtained using PSO (7533 kg) and harmony search (7829 kg). Meanwhile, the best compromise solution is 8568 kg (3rd NSGA-II) that is 14.9% heavier than that of the minimum mass design. The trade-off between the three objective functions is apparent. The minimum mass design although has a smaller mass, it is both un-safer, with $f_2 = -0.147$, and more sensitive, with $f_3 = 0.101$, than the best compromise design, with $f_2 = -0.333$ and $f_3 = 0.037$.

It is worth noting that a larger number of candidate steel sections assigned to each group, as listed in Table 5.9, leads to a difficulty in finding the best compromise solution because many Pareto-optimal solutions correspond to similar values of normalized membership function defined in Eq. (3.5). In this circumstance, some rules of thumb for selection of both column and beam members may be used to shorten the candidate list.

To further confirm their robustness and feasibility, the minimum mass and the best compromise designs obtained from the first three NSGA-II attempts are used to characterize the probabilistic property of the LSF related to the maximum inter-story drift and those of the constraint functions using MCS with 10^5 samples. Results in Table 5.10 confirm the feasibility of the obtained robust optimal designs as all constraint function values associated with these designs are negative. Furthermore, Fig. 5.13 shows an agreement between the PDF and CDF generated by GMM for the best compromise designs obtained from the first three NSGA-II attempts and those generated by MCS. Note that the GMM curves in Figs. 5.13(b) and (c) show slight disturbances around $f = 0$ because the mean value of a Gaussian component of the GMM is $f = -0.074$ and the associated input variables correspond to large values of mixing weight $W_k(\mathbf{x})$ of that component.

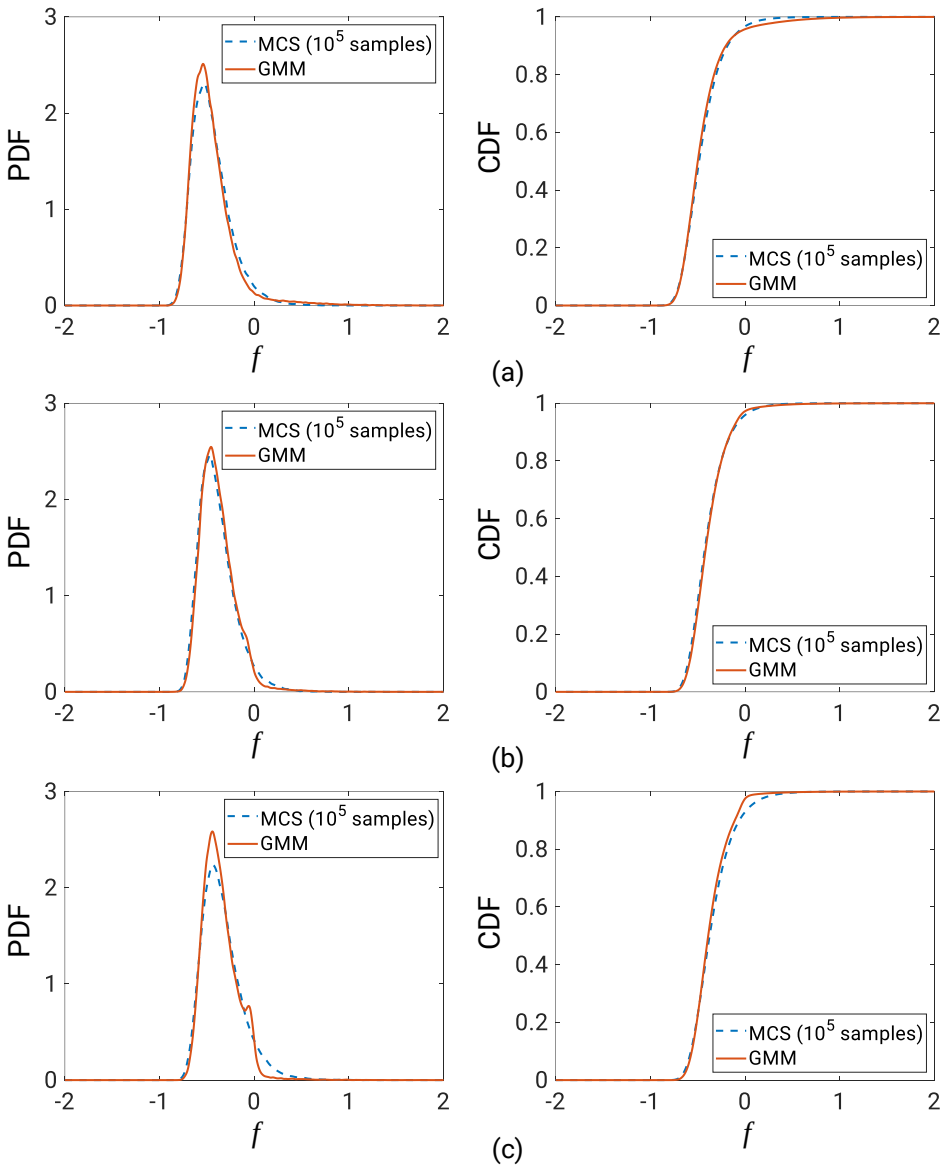

Figure 5.13: Example 2: Statistical property of the maximum inter-story drift for the robust design of six-story two-bay steel frame: (a) PDF and CDF from 1st NSGA-II; (b) PDF and CDF from 2nd NSGA-II; (c) PDF and CDF from 3rd NSGA-II.

5.5 Conclusions

This chapter has presented an application of GMM to solving the multi-objective RDO problem of planar steel frames. The RDO problem is formulated to minimize three conflicting objective functions, namely the total mass, and the mean and variance values of the maximum inter-story drift of the frames, under some constraints on the design strength and serviceability requirements. Uncertainty in the discrete design variables is modeled in the wide range between the smallest and largest values in the catalog of the cross-sectional areas. The main idea of the application is to let random input variables and the corresponding structural response behave according to a joint distribution. GMM aims at statistically understanding this joint distribution based on a sampled

training dataset for both characterizing the probabilistic property of the structural response and providing a simple regression function to facilitate the calculation of the response and its statistical estimates required for solving the RDO problem. Uncertainty in the small ranges of the discrete design variables can also be easily extracted from GMM trained only once before carrying out optimization under various conditions. The main conclusions are summarized as follows:

(1) GMM demonstrates its ability to quickly provide a reasonable RDO design for a two-bar truss and accurately characterize the probabilistic property of a LSF for a steel frame structure before being applied to solving the RDO problem of interest. The latter may be further beneficial for uncertainty quantification of structural responses of a structure that has experienced deterioration or retrofitting in which the GMMs of the responses have been constructed only once.

(2) Approximate Pareto-optimal solutions to the RDO problem are found using NSGA-II with support from the trained GMMs in calculation of the objective and constraint functions. The best trade-off solution is recommended to relieve the conflicting objective functions from the perspective of fuzzy-based compromise programming.

(3) Two numerical design examples have demonstrated the robustness and feasibility of the proposed method. It is confirmed that the GMM-based RDO are capable of finding good optimal solutions that well represent different levels of the trade-off between the conflicting objective functions. Based on these solutions, structural engineers can specify their design according to their preferences and the best compromise solution is one reasonable choice.

Chapter 6

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

This chapter integrates joint and individual probabilistic constraints into the multi-objective RDO problem presented in Chapter 5, and hence formulates two new multi-objective RDO problems. These problems are of interest because they are capable of exploring the space of uncertain constraint functions for confirming the feasibility of a candidate solution, rather than a point estimate as presented in Chapter 5. The basis of these problems is also intimately related to the limit state design philosophy that has adopted the so-called reliability indices to derive the partial factors specified by design codes [23, 25].

Although considering the probabilistic constraints makes the RDO problem more realistic, it considerably increases the complexity of the problem because the calculation of each probabilistic constraint function is NP-hard [44]. In this chapter, a constrained BO approach is proposed to facilitate solutions to the RDO problems with joint and individual probabilistic constraints.

6.1 Robust design optimization of steel frames with probabilistic constraints

Consider a steel frame structure subjected to aleatory uncertainty in design parameters $\mathbf{r}$ that consist of material properties and external loads. Let $\mathbf{s} = [s_1, ..., s_{d_1}]^T \in \mathbb{N}^{d_1}$ denote a d_1-dimensional deterministic vector of discrete design variables of the frame. Each element of $\mathbf{s}$ is selected from a given list of standard steel sections, i.e., $s_k \in \mathcal{S}_k$ $(k = 1, \ldots, d_1)$. Let $f_1(\mathbf{s})$ and $f(\mathbf{s}, \mathbf{r})$ represent the total mass and the LSF associated with the maximum inter-story drift of the frame, respectively. The mean and variance of $f(\mathbf{s}, \mathbf{r})$ are denoted by $f_2(\mathbf{s}) = \mathbb{E}[f(\mathbf{s}, \mathbf{r})]$ and $f_3(\mathbf{s}) = \mathbb{V}[f(\mathbf{s}, \mathbf{r})]$, respectively. Also, let $g_i(\mathbf{s}, \mathbf{r})$ represent the LSFs of certain requirements on the serviceability and strength of

the frame, and $h_j(\mathbf{s})$ indicate deterministic constraints on the geometry of the structural members. By incorporating the joint and individual probabilistic constraints into the multi-objective RDO problem of the frame stated in Eq. (5.28), the following two RDO problems are formulated:

$$\begin{aligned} \text{(Problem J)} \quad \underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})] \\ \text{subject to} \quad & \mathbb{P}[g_i(\mathbf{s}, \mathbf{r}) \leq 0, i = 1, \ldots, I] \geq 1 - \epsilon, \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & s_k \in \mathcal{S}_k \quad (k = 1, \ldots, d_1). \end{aligned} \tag{6.1}$$

$$\begin{aligned} \text{(Problem I)} \quad \underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s}), f_3(\mathbf{s})] \\ \text{subject to} \quad & \mathbb{P}[g_i(\mathbf{s}, \mathbf{r}) \leq 0] \geq 1 - \epsilon_i \quad (i = 1, \ldots, I), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & s_k \in \mathcal{S}_k \quad (k = 1, \ldots, d_1), \end{aligned} \tag{6.2}$$

where $\epsilon \in (0, 1)$ and $\epsilon_i \in (0, 1)$ are the prescribed risk levels of the joint probabilistic constraint and the i-th probabilistic constraint, respectively. Common choices of ϵ and ϵ_i are 0.1 and 0.05. For simplicity, problems (6.1) and (6.2) are termed as problems J and I, respectively.

As mentioned in Section 1.4, problems J and I correspond to two different ways the frame approaches its collapse state. Problem J with a joint probabilistic constraint takes the probability over the entire system of dependent uncertain LSFs, while problem I focuses on the violation of each of statistically independent uncertain LSFs. The two problems also set the basis for further application to a general case where all possible failure modes for a structure can be anticipated, which may consist of both joint and individual failure modes.

Although the two problems are important for finding robust designs of the frame, solving them effectively is hindered by the following four issues. First, it is difficult to check the feasibility of a candidate solution $\mathbf{s}$ because the calculation of probabilistic constraints is intrinsically an NP-hard problem [44]. Second, propagating uncertainty in $\mathbf{r}$ to evaluate the mean and variance of $f(\mathbf{s}, \mathbf{r})$, as discussed in Section 1.5.1, is a challenging task. Third, a search that directly evaluates the objective and constraint functions during the optimization process may be impossible because FE analyses empowered to evaluate these functions are less advantageous to processing optimization as well as probabilistic computations. Finally, it may be impossible to obtain exact optimal solutions to the two problems because their feasible regions defined by the probability functions are non-convex in general. Existing approaches addressing these issues, as reviewed in Section 1.5, are only applicable to special forms of the uncertain objective and probabilistic constraint functions.

In the subsequent sections of this chapter, a constrained BO approach is proposed for solving the two problems as it does not require special forms of the uncertain objective and probabilistic constraint functions. In this way, the non-linear and FE-implicit nature of the problems can be addressed through the mean functions of the GP models that are surrogates for the uncertain objective and probabilistic constraint functions. These surrogates, in turn, facilitate the evaluation of probabilistic constraint functions for a particular candidate solution $\mathbf{s}$ using the SAA presented in Section 2.6 and Chapter 7. Thus, the feasibility of $\mathbf{s}$ can be reasonably confirmed in each BO iteration. Furthermore, the mean and variance of the objective function $f(\mathbf{s}, \mathbf{r})$ for a specific $\mathbf{s}$ 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 evaluated based on the mean function of the corresponding GP surrogate. With reasonable calculations of the objectives and constraints, an optimization strategy is devised to guide BO toward better candidate solutions and to offer good Pareto-optimal solutions to each problem when the optimization process terminates. In particular, an acquisition function is proposed and maximized by solving a mixed-integer nonlinear programming (MINLP) problem using a novel random sampling method coupled with SA.

6.2 Proposed Bayesian optimization

6.2.1 Surrogates for uncertain objective and probabilistic constraint functions

In the proposed BO, the objective function $f(\mathbf{s}, \mathbf{r})$ and the constraint functions $g_i(\mathbf{s}, \mathbf{r})$ are first approximated by GP models. The construction of these GP models from a training dataset $\mathcal{D}$ can be found in Section 2.3. Since it is desirable to obtain highly accurate GP models for $f(\mathbf{s}, \mathbf{r})$ and $g_i(\mathbf{s}, \mathbf{r})$ 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 the non-positive LSF values. After training, the GP models $\hat{f}(\mathbf{s}, \mathbf{r})$ and $\hat{g}_i(\mathbf{s}, \mathbf{r})$ are obtained for $f(\mathbf{s}, \mathbf{r})$ and $g_i(\mathbf{s}, \mathbf{r})$, respectively. $\hat{f}(\mathbf{s}, \mathbf{r})$ is characterized by its mean $\mu_{\hat{f}}(\mathbf{s}, \mathbf{r})$ and its variance $\sigma^2_{\hat{f}}(\mathbf{s}, \mathbf{r})$, while $\hat{g}_i(\mathbf{s}, \mathbf{r})$ is captured by $\mu_{\hat{g}_i}(\mathbf{s}, \mathbf{r})$ and $\sigma^2_{\hat{g}_i}(\mathbf{s}, \mathbf{r})$.

6.2.2 Acquisition functions

The proposed BO guides optimization of each of the two RDO problems through an acquisition function developed based on $\hat{f}(\mathbf{s}, \mathbf{r})$, $\hat{g}_i(\mathbf{s}, \mathbf{r})$, and the MOBO proposed in Section 3.2. Let $\Omega = \{\mathbf{f}_1, ..., \mathbf{f}_M\} \in \mathbb{R}^3$ and $\mathbf{f}_{\mathrm{R}} \in \mathbb{R}^3$, respectively, denote the current set of M Pareto-optimal solutions in the space of three objective functions of problem I (or J) and a fixed reference point that is dominated by all elements of Ω. Let $\mathbf{f}(\mathbf{s})$ denote an arbitrary vector in the objective function space. BO improves the solutions to each RDO problem after each iteration by finding a new sampling point $\mathbf{x}_{\mathrm{n}} = [\mathbf{s}_{\mathrm{n}}^T, \mathbf{r}_{\mathrm{n}}^T]^T$ in the input variable space so that $\mathbf{x}_{\mathrm{n}}$ maximizes $\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_{\mathrm{R}})$ stated in Eq. (3.8).

Although $\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_{\mathrm{R}})$ can be used to assess improvement in the current solutions after BO obtains the new design $\mathbf{s}_{\mathrm{n}}$, it becomes useless if $\mathbf{s}_{\mathrm{n}}$ 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 J, let $\Delta_{\mathrm{J}}(\mathbf{s}, \mathbf{r})$ be a function expressing the probability that $\mathbf{s}$ satisfies the joint probabilistic constraint computed from the GP models. Here $\mathbf{r}$ is referred to as a particular point in the space of random parameters. $\Delta_{\mathrm{J}}(\mathbf{s}, \mathbf{r})$ can be formulated as

$$\Delta_{\mathrm{J}}(\mathbf{s}, \mathbf{r}) = \mathbb{P}\big[\bar{g}(\mathbf{s}, \mathbf{r}) \leq 0\big] - 1 + \epsilon, \tag{6.3}$$

where ϵ is given in problem (6.1), and

$$\bar{g}(\mathbf{s}, \mathbf{r}) = \max\big\{g_1(\mathbf{s}, \mathbf{r}), \ldots, g_I(\mathbf{s}, \mathbf{r})\big\} \approx \max\big\{\mu_{\hat{g}_1}(\mathbf{s}, \mathbf{r}), \ldots, \mu_{\hat{g}_I}(\mathbf{s}, \mathbf{r})\big\}. \tag{6.4}$$

If $\mathbf{x}_{\mathrm{n}}$ maximizes Δ_{J}, it therefore increases the chance that $\mathbf{s}_{\mathrm{n}}$ satisfies the joint probabilistic constraint of problem J.

For problem I, let $\Delta_{\mathrm{I}}(\mathbf{s},\mathbf{r})$ be a function expressing the chance that $\mathbf{s}$ satisfies all individual probabilistic constraints computed from the GP models. 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 variance $\sigma^2_{\hat{g}_i}(\mathbf{s},\mathbf{r})$, and all probabilistic constraints are statistically independent in this problem, $\Delta_{\mathrm{I}}(\mathbf{s},\mathbf{r})$ can be defined as

$$\Delta_{\mathrm{I}}(\mathbf{s},\mathbf{r}) = \prod_{i=1}^{I} \left[\mathbb{P}\left[g_i(\mathbf{s},\mathbf{r}) \leq 0\right] - 1 + \epsilon_i\right] \approx \prod_{i=1}^{I} \left\{ \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{-\mu_{\hat{g}_i}(\mathbf{s},\mathbf{r})}{\sqrt{2}\sigma_{\hat{g}_i}(\mathbf{s},\mathbf{r})}\right)\right] - 1 + \epsilon_i \right\}, \tag{6.5}$$

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

For both problems, the following feasibility indicator function $\mathbb{I}_{\mathrm{FI}}(\mathbf{s})$ is formulated to indicate whether $\mathbf{s}$ satisfies all deterministic constraints $h_j(\mathbf{s}) \leq 0$ or not:

$$\mathbb{I}_{\mathrm{FI}}(\mathbf{s}) = \begin{cases} 1 & \text{if} \quad h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ 0 & \text{otherwise.} \end{cases} \tag{6.6}$$

By incorporating the HIV and feasibility criteria 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 J and I, respectively:

$$\alpha_1(\mathbf{s},\mathbf{r}) = \mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_{\mathrm{R}})\Delta_{\mathrm{J}}(\mathbf{s},\mathbf{r})\mathbb{I}_{\mathrm{FI}}(\mathbf{s}), \tag{6.7a}$$

$$\alpha_2(\mathbf{s},\mathbf{r}) = \mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_{\mathrm{R}})\Delta_{\mathrm{I}}(\mathbf{s},\mathbf{r})\mathbb{I}_{\mathrm{FI}}(\mathbf{s}). \tag{6.7b}$$

Thus, the next sampling point $\mathbf{x}_{\mathrm{n}}$ for problem J or I is the maximizer of $\alpha_1(\mathbf{s},\mathbf{r})$ or $\alpha_2(\mathbf{s},\mathbf{r})$, respectively. Maximizing $\alpha_1(\mathbf{s},\mathbf{r})$ or $\alpha_2(\mathbf{s},\mathbf{r})$ is associated with solving an MINLP problem because the acquisition function is nonlinear, and $\mathbf{s}$ and $\mathbf{r}$ are discrete and continuous vectors, respectively. An optimization strategy is developed in the next section for solving such an MINLP problem.

6.3 Solution approach

6.3.1 Sorting Pareto-optimal solutions

Let Ω_{a} denote a set of already-generated candidate solutions that consists of all design points of the initial training dataset at the first iteration of BO. Based on Ω_{a}, an elitist non-dominated sorting approach [91] is employed to find a set Ω of approximate Pareto-optimal solutions to the RDO problems in each BO iteration. For obtaining a set of feasible solutions, the sorting approach requires efficient evaluations of the mean $f_2(\mathbf{s})$ and variance $f_3(\mathbf{s})$ of the uncertain objective function, and the probabilities in the probabilistic constraints of problems J and I. Since $f(\mathbf{s},\mathbf{r})$ in each BO iteration is approximated by $\hat{f}(\mathbf{s},\mathbf{r})$, Eqs. (5.3) and (5.4), which are derived from the second-order Taylor expansion, use $\mu_{\hat{f}}(\mathbf{s},\mathbf{r})$ as the regression function for evaluation of $f_2(\mathbf{s})$ and $f_3(\mathbf{s})$, respectively. Furthermore, the joint and individual probabilistic constraints in each iteration of the BO can be estimated using the SAA detailed in Section 2.6 of Chapter 2 and Algorithm 7.2 of Chapter 7, which

is developed based upon the GP surrogates $\hat{g}_i(\mathbf{s}, \mathbf{r})$ for the probabilistic constraint functions.

6.3.2 Maximizing the acquisition functions

The MINLP problem of two acquisition functions in Eqs. (6.7a) and (6.7b) reads

$$\begin{aligned} [\mathbf{s}_\mathrm{n}, \mathbf{r}_\mathrm{n}] = & \underset{\mathbf{s},\mathbf{r}}{\operatorname{argmax}} \left[\alpha_{1(2)}(\mathbf{s}, \mathbf{r})\right] \\ \text{subject to} \quad & \mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_\mathrm{R}) > 0, \\ & s_k \in \mathcal{S}_k \quad (k = 1, \ldots, d_1), \\ & \mathbf{r} \in \mathcal{R}, \end{aligned} \tag{6.8}$$

where $\alpha_1(\mathbf{s}, \mathbf{r})$ and $\alpha_2(\mathbf{s}, \mathbf{r})$ correspond to problems J and I, respectively; and the set $\mathcal{R}$ is assigned as the 95% confidence interval of the random parameters $\mathbf{r}$.

Although several techniques are available for solving a convex MINLP problem [132], for example, branch-and-bound method, single-tree method, multi-tree method, cutting plane method, and mixed-integer second-order cone program, they are not applicable for solving problem (6.8) because it is difficult to detect whether $\alpha_1(\mathbf{s}, \mathbf{r})$ and $\alpha_2(\mathbf{s}, \mathbf{r})$ are convex functions. One approach is to replace $\alpha_1(\mathbf{s}, \mathbf{r})$ and $\alpha_2(\mathbf{s}, \mathbf{r})$ with piecewise linear approximations [132] and 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 and black-box functions, modeling their piecewise linear approximations may be impossible. Another approach is to directly employ population-based methods, for example, GA or PSO, 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.

Since $\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_\mathrm{R})$ and $\mathbb{I}_\mathrm{FI}(\mathbf{s})$ are functions of $\mathbf{s}$, and $\mathbf{r}$ only appears in $\Delta_\mathrm{J}(\mathbf{s}, \mathbf{r})$ and $\Delta_\mathrm{I}(\mathbf{s}, \mathbf{r})$, an optimization strategy that couples a random sampling method with SA is developed for solving problem (6.8). This strategy is an extension of a two-stage random search by Do and Ohsaki [133] that includes a stage of determining $\mathbf{r}$, followed by a stage of determining $\mathbf{s}$ for solving upper- and lower-level optimization problems of a single-objective discrete RDO problem with deterministic constraints and unknown-but-bounded uncertainty in the design parameters, respectively. 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 $\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_\mathrm{R})$ and $\mathbb{I}_\mathrm{FI}(\mathbf{s})$ for each member of Ω_s, and retain in Ω_s the members that yield positive $\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_\mathrm{R})$, $\mathbb{I}_\mathrm{FI}(\mathbf{s}) = 1$, and negative values of $\mu_{\hat{g}_i}(\mathbf{s}, \mathbf{r})$, where $\mu_{\hat{g}_i}(\mathbf{s}, \mathbf{r})$ is evaluated using Eq. (5.3).

(3) Formulate problem (6.8) for each retained member of Ω_s and solve it using SA [1] for obtaining the associated random parameters $\mathbf{r}$.

(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 $\mathbf{x}_\mathrm{n} = [\mathbf{s}_\mathrm{n}^T, \mathbf{r}_\mathrm{n}^T]^T$.

Note that any gradient-based or population-based algorithms can be used for solving problem (6.8) formulated in step (3) with a given $\mathbf{s}$. The SA is selected here 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.

Step (1) generates two groups of new candidate solutions $\mathbf{s}$ for forming Ω_{s}. The first group consists of points drawn by random perturbations (in the design variable space) surrounding each of the current Pareto-optimal solutions, which can be regarded as a neighborhood search [133]. 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. It is expected that a significant improvement in the solution quantity can be achieved by performing the neighborhood search in the design variable space of 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 circumstance, the Pareto-optimal solutions in the objective function space can be first divided into a moderately small number of disjoint clusters using the GMM clustering method, which will be discussed in Chapter 7. 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 constituting Ω_{s} is generated by uniformly sampling points from the design domain, which can be regarded as a global search [133]. Once Ω_{s} is created, its members that already appear in the training dataset $\mathcal{D}$ are discarded. To form a new set Ω_{a} of already-generated solutions that is used in the next iteration of the BO, the current set of new candidate solutions Ω_{s} is appended to the current set Ω_{a}.

In summary, the optimization procedure for solving problem J or I using the proposed BO follows six main steps.

- **Step 1**: Generate samples of $\mathbf{s}$ and $\mathbf{r}$ using Latin hypercube sampling [74]. Also generate the training dataset $\mathcal{D}$ by performing FE analyses for the generated samples.
- **Step 2**: Based on $\mathcal{D}$, construct the GP models $\hat{f}(\mathbf{s},\mathbf{r})$ and $\hat{g}_i(\mathbf{s},\mathbf{r})$.
- **Step 3**: Sort the Pareto-optimal solutions from the set Ω_{a} of already-generated candidate solutions Ω_{a}, which consists of members of $\mathcal{D}$ at the first iteration of BO. Perform the neighborhood and global searches to generate the set Ω_{s} used for maximizing the acquisition function.
- **Step 4**: Terminate 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 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} has no feasible solution. Otherwise, proceed to **Step 5**.
- **Step 5**: Maximize the acquisition function corresponding to each RDO problem using the optimization strategy described above to obtain the sampling point $\mathbf{x}_{\mathrm{n}}$ for the next iteration of BO and let $\Omega_{\mathrm{a}} = \Omega_{\mathrm{a}} \cup \Omega_{\mathrm{s}}$.
- **Step 6**: Add $\mathbf{x}_{\mathrm{n}}$ obtained from **Step 5** to the training dataset, determine the associated LSF values, update the GP models, and reiterate from **Step 3**.

Table 6.1: Assigned random parameters for the two-bar truss.

Parameter	Description	Mean	COV	Distribution
ρ	Mass density [kg/m^3]	10^4	0.20	Lognormal
P	External load [kN]	800	0.25	Lognormal
$\sigma_{\mathrm{y},0}$	Yield stress [MPa]	1050	0.24	Normal

6.3.3 Test problem

To demonstrate the feasibility of the proposed BO in solving a simple multi-objective RDO problem with probabilistic constraints, this section investigates the two-bar truss in Fig. 5.1. The truss is subjected to an external load P whose projections onto the horizontal and vertical axes are P_{x} and P_{y}, respectively, which satisfy $P_{\mathrm{y}} = 8P_{\mathrm{x}}$. As stated in Section 5.2.1, two design variables for the truss are the cross-sectional area s_1 of the truss members and the horizontal span s_2, i.e., $\mathbf{s} = [s_1, s_2]^T$. The random parameters include the magnitude of the external load P, and the mass density ρ and yield stress $\sigma_{\mathrm{y},0}$ of the truss material, i.e., $\mathbf{r} = [\rho, P, \sigma_{\mathrm{y},0}]^T$. Probabilistic characteristics of these parameters are given in Table 6.1.

A bi-objective RDO problem is formulated for the truss with two conflicting objective functions, namely the mean and standard deviation of its total mass under individual probabilistic constraints on the axial stresses in the members. Let $\tilde{f}(\mathbf{s}, \mathbf{r})$ denote the total mass of the truss, and $g_1(\mathbf{s}, \mathbf{r})$ and $g_2(\mathbf{s}, \mathbf{r})$ denote the LSFs associated with the axial stresses in its members. The mean and standard deviation of $\tilde{f}(\mathbf{s}, \mathbf{r})$ are $f_1(\mathbf{s}) = \mathbb{E}\left[\tilde{f}(\mathbf{s}, \mathbf{r})\right]$ and $f_2(\mathbf{s}) = \sqrt{\mathbb{V}[\tilde{f}(\mathbf{s}, \mathbf{r})]}$, respectively. Thus, the bi-objective RDO problem formulated for the truss reads

$$
\begin{aligned}
\underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s})] \\
\text{subject to} \quad & \mathbb{P}\big[g_1(\mathbf{s}, \mathbf{r}) \leq 0\big] \geq 1 - \epsilon_1, \\
& \mathbb{P}\big[g_2(\mathbf{s}, \mathbf{r}) \leq 0\big] \geq 1 - \epsilon_2, \\
& s_1 \in \mathcal{S}_1 = \{1.0, 1.5, \ldots, 20.0\}\,\mathrm{cm}^2, \\
& s_2 \in \mathcal{S}_2 = \{0.1, 0.2, \ldots, 2.0\}\,\mathrm{m},
\end{aligned}
\tag{6.9}
$$

which has the form of problem I, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are the lists of possible values of s_1 and s_2, respectively, and

$$\tilde{f}(\mathbf{s}, \mathbf{r}) = 10^{-4} \rho s_1 \sqrt{1 + s_2^2}, \tag{6.10a}$$

$$g_1(\mathbf{s}, \mathbf{r}) = \frac{5P}{\sqrt{65} s_1 \sigma_{\mathrm{y},0}} \sqrt{1 + s_2^2} \left(8 + \frac{1}{s_2}\right) - 1, \tag{6.10b}$$

$$g_2(\mathbf{s}, \mathbf{r}) = \frac{5P}{\sqrt{65} s_1 \sigma_{\mathrm{y},0}} \sqrt{1 + s_2^2} \left(8 - \frac{1}{s_2}\right) - 1. \tag{6.10c}$$

Note that the mean $f_1(\mathbf{s})$ can be derived as a linear function of the standard deviation $f_2(\mathbf{s})$, i.e., $f_1(\mathbf{s}) = \mathbb{E}[\rho]/\sqrt{\mathbb{V}[\rho]} f_2(\mathbf{s}) = 0.2^{-1} f_2(\mathbf{s}) = 5 f_2(\mathbf{s})$, where 0.2 stated in Table 6.1 is the COV of ρ. This derivation relies on the fact that $\tilde{f}(\mathbf{s}, \mathbf{r})$ stated in Eq. (6.10a) is a linear function of the random parameter ρ, and s_1 and s_2 are deterministic. The set of Pareto-optimal solutions to problem (6.9), therefore, has only one solution.

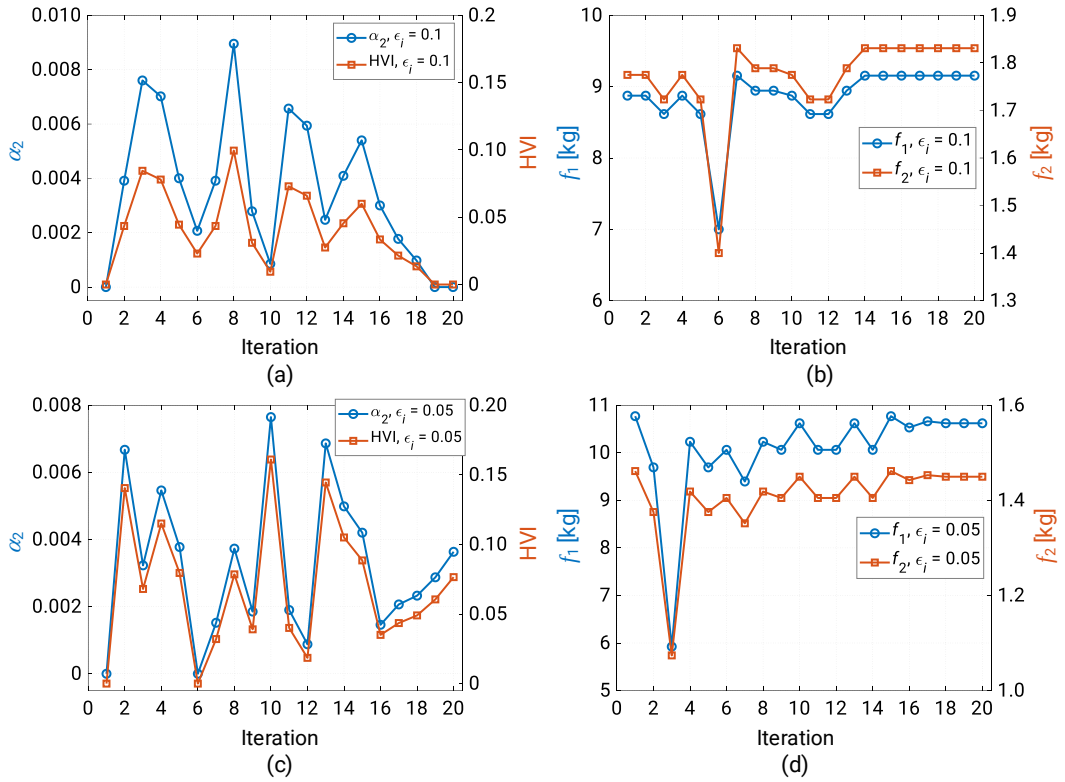

Figure 6.1: Histories of HVI, acquisition function, and objective functions for the two-bar truss with two risk levels. (a)–(b) $\epsilon_i = 0.1$; (c)–(d) $\epsilon_i = 0.05$.

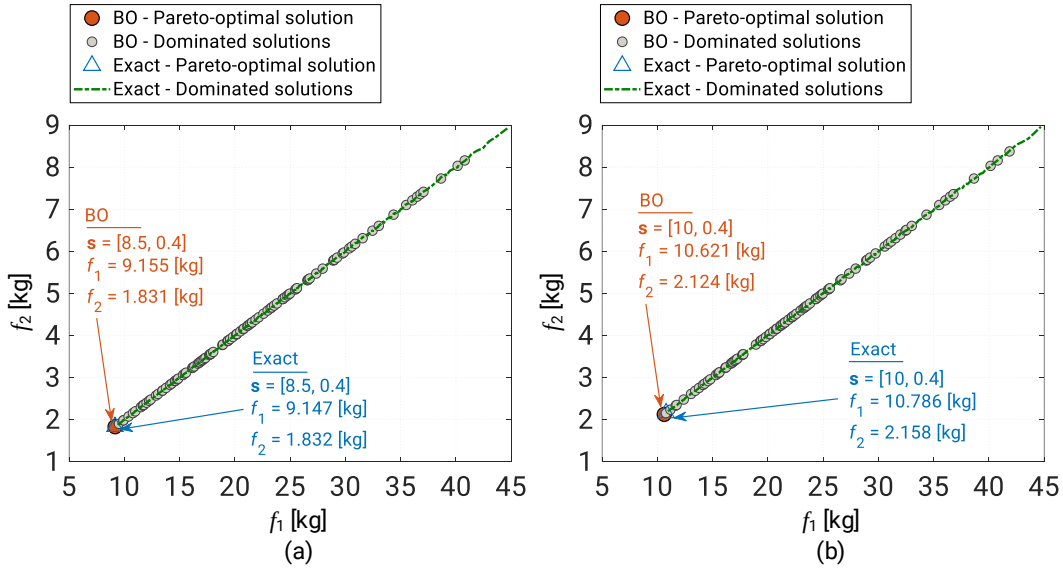

Figure 6.2: Verification of the obtained robust designs for the two-bar truss with two risk levels. (a) $\epsilon_i = 0.1$; (b) $\epsilon_i = 0.05$.

As the first step for solving problem (6.9), a total of 200 samples are generated for the initial training dataset. However, only 140 feasible samples are kept for training because the other 60 samples associated with either positive $g_1(\mathbf{s},\mathbf{r})$ or positive $g_2(\mathbf{s},\mathbf{r})$ are discarded. The DACE toolbox [92] are then used together with a second-degree polynomial mean function for constructing the GP surrogates for $\tilde{f}(\mathbf{s},\mathbf{r})$, $g_1(\mathbf{s},\mathbf{r})$, and $g_2(\mathbf{s},\mathbf{r})$. Based on these surrogates, the proposed BO is employed to solve problem (6.9) for two risk levels, namely $\epsilon_i = 0.1$ and $\epsilon_i = 0.05\,(i = 1, 2)$. The SA incorporated in BO is carried out using MATLAB R2018a Global Optimization Toolbox [106]. 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, and 3000 times the dimension of $\mathbf{r}$, respectively. The maximum number of iterations and the objective function tolerance of SA are assigned as 500 and 10^{-6}, respectively. In each iteration of 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 $\mathbf{f}_{\mathrm{R}} = [50, 10]$ kg and 20, respectively. Here $\mathbf{f}_{\mathrm{R}}$ is chosen from (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 $f_1(\mathbf{s}) = 5f_2(\mathbf{s})$.

Figure 6.1 shows the maximum value of acquisition function α_2, the corresponding HVI, and the objective functions $f_1(\mathbf{s})$ and $f_2(\mathbf{s})$ at each iteration of BO for $\epsilon_i = 0.1$ and $\epsilon_i = 0.05$. The respective robust designs of the truss found at the 20th iteration of the BO are $\mathbf{s} = [8.5, 0.4]$ and $[10, 0.4]$. The computational times required for $\epsilon_i = 0.1$ and $\epsilon_i = 0.05$ are 905 s and 973 s using a PC with an Intel(R) i7-7700HQ 2.80 GHz CPU and 8.0 GB memory.

To further verify the obtained robust designs, the exact robust solution to problem (6.9) is found for each risk level. Since the sets $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, have 39 and 20 elements, a total of $39 \times 20 = 780$ possible designs can be assigned for the truss using a full factorial sampling. Then, 10^5 samples of the random parameters $\mathbf{r}$ are generated by MCS and used to calculate $f_1(\mathbf{s})$, $f_2(\mathbf{s})$, $\mathbb{P}\big[g_1(\mathbf{s},\mathbf{r}) \leq 0\big]$, and $\mathbb{P}\big[g_2(\mathbf{s},\mathbf{r}) \leq 0\big]$ for each possible design. This process is performed three times for each risk level with different sets of MCS samples, which may lead to different solutions. Nevertheless, the solutions by the three trials are identical for each risk level, and therefore they can be regarded as the exact solution. Figure 6.2 confirms that the obtained robust design for each risk level is identical to the exact design, even though the number of function evaluations used by BO (1.30122×10^5 evaluations) is much less than that by the combination of the full factorial sampling and MCS ($3 \times 780 \times 10^5$ evaluations).

6.4 Design examples

6.4.1 LSFs of serviceability and strength requirements for steel frame designs

The designs of two planar steel frames in this section comply with the serviceability and strength requirements stated in ANSI/AISC 360-16 [24]. 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 strengths of column and beam members.

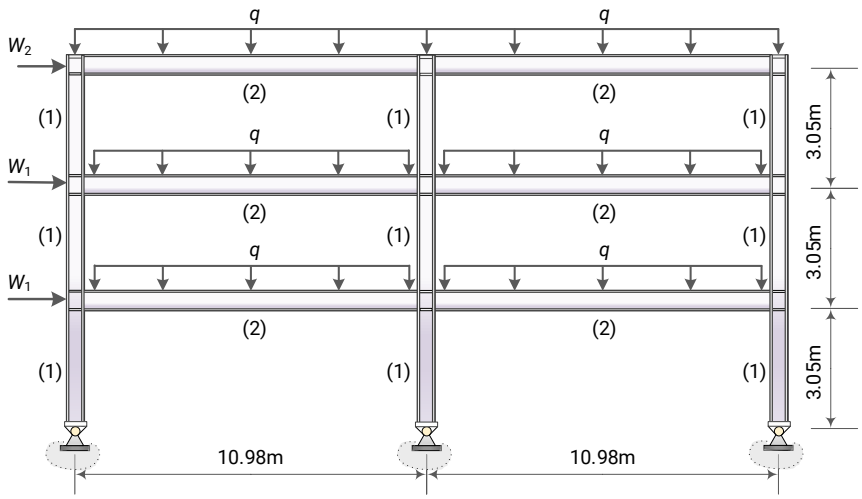

Figure 6.3: Example 1: Three-story two-bay steel frame, groups of members, and external loads.

Table 6.2: Example 1: List of column and beam sections 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

The LSFs corresponding to the maximum inter-story drift $f(\mathbf{s}, \mathbf{r})$ of the frame, the total drift $g_1(\mathbf{s}, \mathbf{r})$ of the frame, and the maximum long-term vertical displacement among beams in the same group are detailed in Section 5.3.1. Let n_b denote the number of beam groups of the frame. The LSFs of long-term vertical displacements among beams in the same group are $g_2(\mathbf{s}, \mathbf{r}), \ldots, g_{1+n_\mathrm{b}}(\mathbf{s}, \mathbf{r})$, i.e., $g_{1+k}(\mathbf{s}, \mathbf{r})$ $(k = 1, \ldots, n_\mathrm{b})$.

Let $q(\mathbf{s}, \mathbf{r})$ denote the strength LSF for a column or a beam, which is normalized using the AISC-LRFD interaction formula in Eq. (5.20), and $q_{\max}(\mathbf{s}, \mathbf{r})$ denote the maximum of $q(\mathbf{s}, \mathbf{r})$ among the

Table 6.3: Example 1: Assigned random 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
$\sigma_{\mathrm{y},0}$	Yield stress [MPa]	250	1.05	0.06	Normal
σ_{u}	Tensile strength [MPa]	400	1.05	0.08	Normal

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

$$g_{1+n_{\mathrm{b}}+j}(\mathbf{s},\mathbf{r}) = q_{\max,j}(\mathbf{s},\mathbf{r}) \quad (j = 1,\ldots,n_{\mathrm{m}}), \tag{6.11}$$

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

6.4.2 Example 1: Three-story two-bay steel frame

Problems J and I are formulated for the three-story two-bay steel frame in Section 5.4.1 with modifying the probabilistic properties of random parameters. The frame, as shown in Fig. 6.3, consisting of 15 members classified into column (1) and beam (2) groups is subjected to static loads q, W_1, and W_2. The steel section for the members in each group is selected from the list of American wide-flange steel sections in Table 6.2. The independent random parameters $\mathbf{r}$, including the vertical load q, lateral loads W_1 and W_2, and material properties (i.e., Young's modulus E, yield stress $\sigma_{\mathrm{y},0}$, and ultimate tensile strength σ_{u}), are described in Table 6.3 with their probabilistic properties. Note that the mass density of steel is not considered as an random parameter because the nominal mass of the standard column or beam section is well controlled by the manufacturers and provided by a deterministic value in Table 6.2.

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, where the total mass $f_1(\mathbf{s})$ reads

$$f_1(\mathbf{s}) = 27.45\rho_1 + 65.88\rho_2. \tag{6.12}$$

Here ρ_1 and ρ_2 provided in Table 6.2 are the nominal mass [kg/m] of the column and beam sections, respectively.

Four probabilistic constraint functions correspond to the total drift $g_1(\mathbf{s},\mathbf{r})$ of the frame, maximum long-term vertical displacement $g_2(\mathbf{s},\mathbf{r})$ of the beams, combined axial-flexural strength $g_3(\mathbf{s},\mathbf{r})$ of the columns, and combined axial-flexural strength $g_4(\mathbf{s},\mathbf{r})$ of the beams. One deterministic geometric constraint $h(\mathbf{s}) \leq 0$ is also 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 $g_1(\mathbf{s},\mathbf{r})$, $g_2(\mathbf{s},\mathbf{r})$, $g_3(\mathbf{s},\mathbf{r})$, and $g_4(\mathbf{s},\mathbf{r})$. The number of new candidate solutions generated in each iteration of BO is 200. A PC with an Intel(R) Xeon(R) E5-2643V4 3.40 GHz CPU and 64 GB memory is used to implement the proposed BO that solves problems J

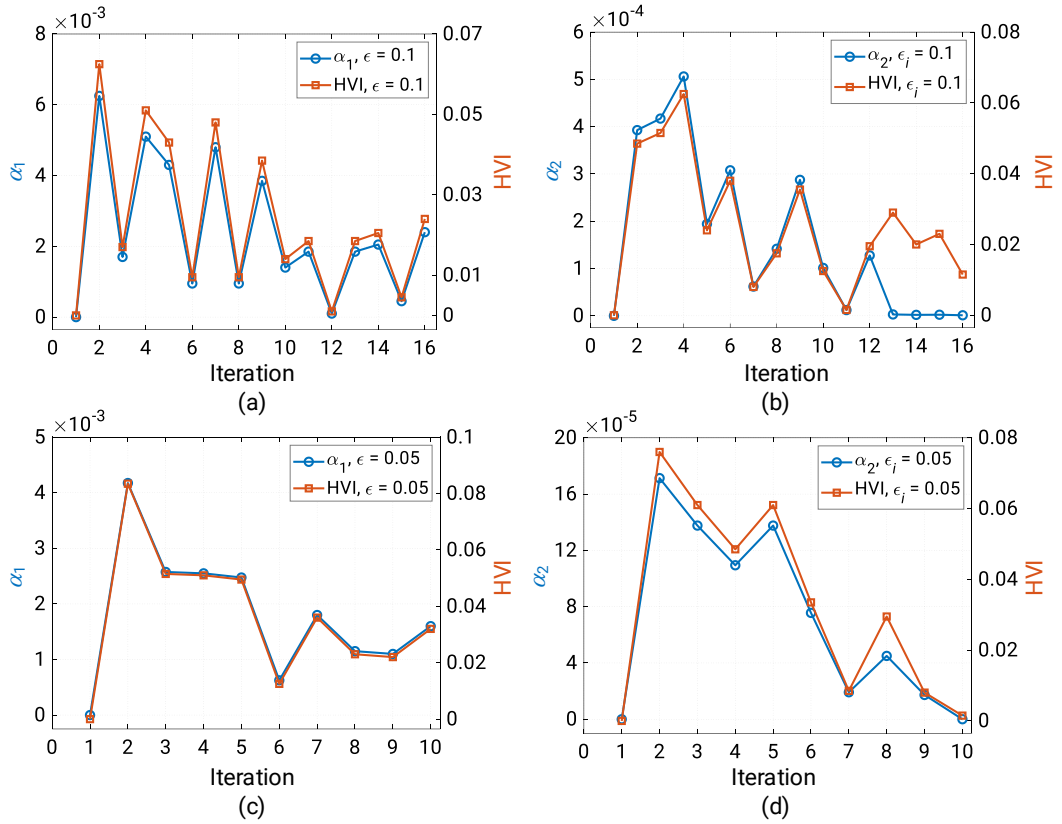

Figure 6.4: Example 1: Histories of HVI and acquisition functions for different RDO problems of the three-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

and I with two risk levels 0.1 and 0.05. For convenient description, we refer problem J formulated for $\epsilon = 0.1$ and $\epsilon = 0.05$ as problems J-0.1 and J-0.05, respectively, and problem I formulated for $\epsilon_i = 0.1$ and $\epsilon_i = 0.05$ $(i = 1, \ldots, 4)$ as problems I-0.1 and I-0.05, respectively. The reference point and the maximum number of iterations of BO are $\mathbf{f}_{\mathrm{R}} = [14000\,\mathrm{kg}, 0.5, 0.5]$ and 20, respectively. As a note to facilitate the selection of $\mathbf{f}_{\mathrm{R}}$, the first element is selected from the maximum value of the frame total mass derived from the list of sections in Table 6.2; the second element associated with the mean of $f(\mathbf{s}, \mathbf{r})$ can be assigned as any positive value so that it bounds from above the expected value of $f(\mathbf{s}, \mathbf{r})$, which should be non-positive to ensure the structural performance; and the last element corresponding to the variance of $f(\mathbf{s}, \mathbf{r})$ should be positive and small enough to ensure the robustness of the frame.

Figure 6.4 shows the histories of HVI and acquisition functions obtained from solving the aforementioned four RDO problems. The graphs in Fig. 6.4(c) 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 BO; see Eq. (6.7a). Since problems I-0.1 and I-0.05 involve the probability of simultaneously meeting all probabilistic constraints, α_2 is much less than α_1 at the same risk level. BO terminates at the 16th iteration when solving problems J-0.1 and I-0.1, and at the 10th iteration when solving problems J-0.05 and I-0.05, 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; see **Step 4** in Sec-

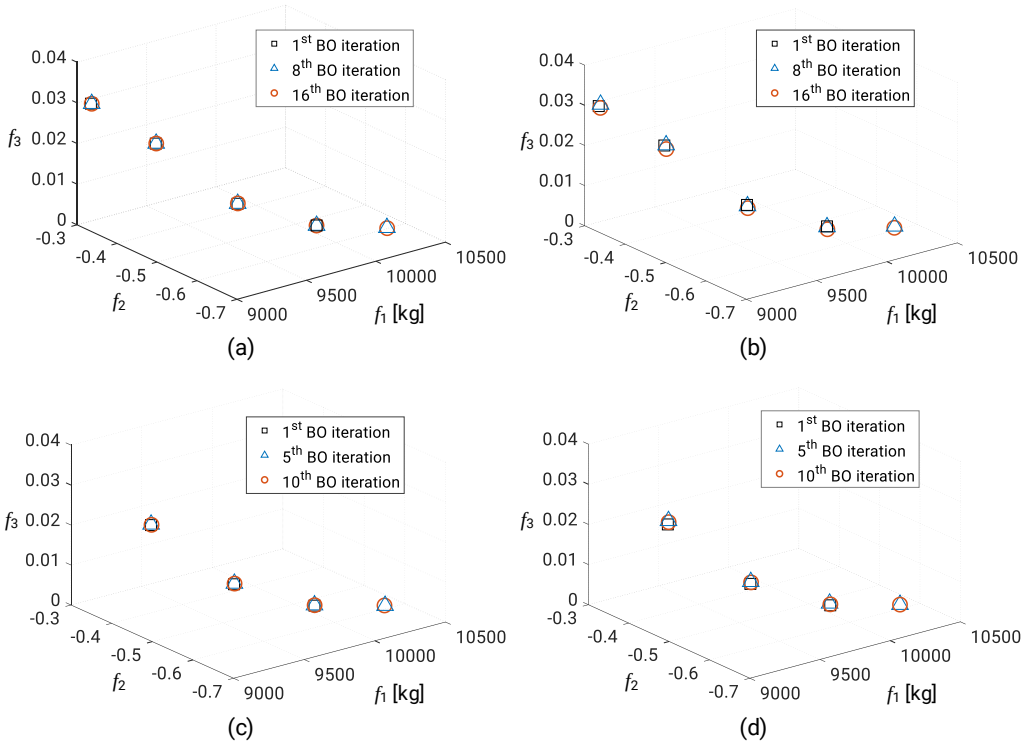

Figure 6.5: Example 1: Pareto-optimal solutions to different RDO problems of the three-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

tion 6.3.2. In addition, the computational times for BO to complete 16 and 10 iterations are 2.68 and 1.81 hours, respectively.

Figure 6.5 provides four sets of Pareto-optimal solutions corresponding to the mentioned four RDO problems. The sets of solutions to problems J and I are identical for each risk level. 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 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 listed in Table 6.2, it is able to generate a total of $20 \times 20 = 400$ possible solutions to each problem. Then, the objective functions $f_2(\mathbf{s})$, $f_3(\mathbf{s})$, and the probabilistic constraint functions are calculated for all possible solutions to each RDO problem using 10^5 samples of the random parameters $\mathbf{r}$ generated by 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 to each RDO problem takes around 10.02 hours to complete, which is much more than that required for 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 $\mathbb{E}\left[g_i(\mathbf{s},\mathbf{r})\right] + \sqrt{\mathbb{V}\left[g_i(\mathbf{s},\mathbf{r})\right]} \leq 0 \;\; (i = 1, \ldots, 4)$ as four constraints. This problem is solved using the GMM-based method presented in Chapter 5 with an expectation that the resulting Pareto-optimal solutions can capture all solutions by BO.

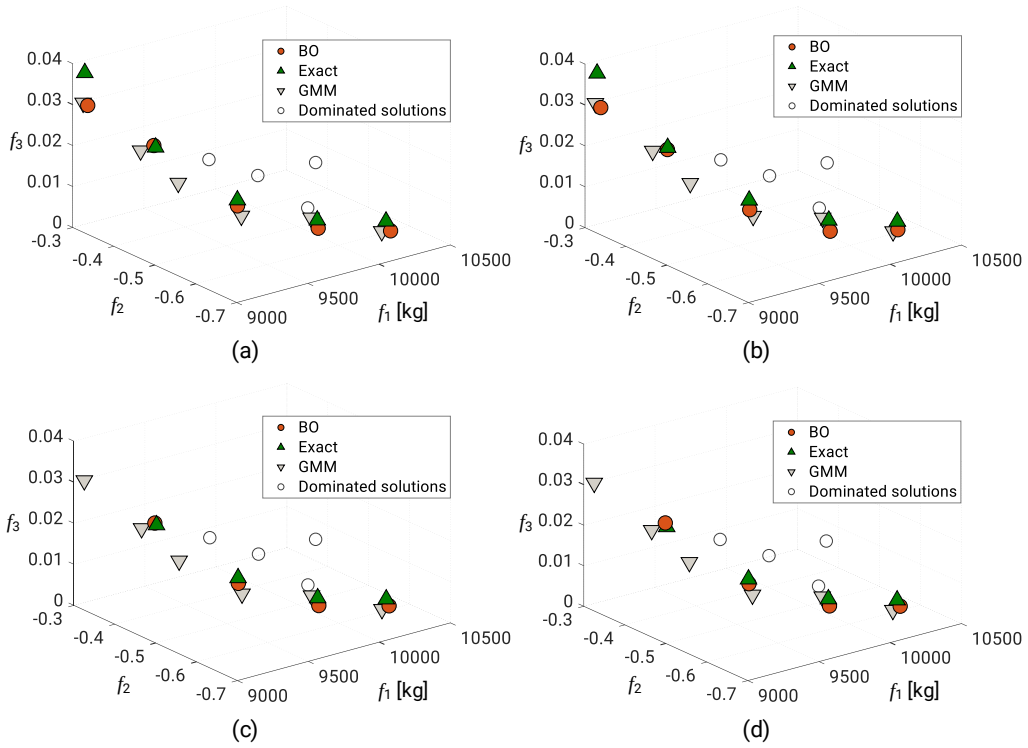

Figure 6.6: Example 1: Verification of the obtained Pareto-optimal solutions to different RDO problems of the three-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

The comparison results in Fig. 6.6 show that the solutions to each RDO problem by 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-based method, as expected, provides a total of six solutions that include all solutions by BO.

Tables 6.4 and 6.5 provide values of the objective and probabilistic constraint functions associated with the solutions to each RDO problem. Figure 6.7 illustrates the CDF and PDF of the uncertain objective function $f(\mathbf{s}, \mathbf{r})$, and the CDFs of four probabilistic constraint functions calculated at solution 1 (i.e., minimum-mass solution) to each RDO problem. Since the design with the larger risk level is less robust than that with the smaller risk level, the PDF in Fig. 6.7(a) is broader than that in Fig. 6.7(c). Moreover, as each solution offers a safety margin in all probabilistic constraints, the obtained Pareto-optimal solutions are feasible.

Table 6.4: Example 1: Pareto-optimal solutions to problems J-0.1 and I-0.1 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–J0.1)	−0.332	−0.445	−0.563	−0.636	−0.670
f_3 (BO–J0.1)	0.031	0.025	0.014	0.008	0.005
f_2 (BO–I0.1)	−0.333	−0.449	−0.564	−0.637	−0.660
f_3 (BO–I0.1)	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
$\mathbb{P}\big[g_{1,\ldots,4}(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.903	0.951	0.967	0.976	0.979
$\mathbb{P}\big[g_1(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.999	1.000	1.000	1.000	1.000
$\mathbb{P}\big[g_2(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.999	1.000	1.000	1.000	1.000
$\mathbb{P}\big[g_3(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.908	0.974	0.998	1.000	1.000
$\mathbb{P}\big[g_4(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.935	0.953	0.967	0.976	0.979

Table 6.5: Example 1: Pareto-optimal solutions to problems J-0.05 and I-0.05 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–J0.05)	−0.445	−0.564	−0.637	−0.666
f_3 (BO–J0.05)	0.025	0.014	0.008	0.006
f_2 (BO–I0.05)	−0.446	−0.564	−0.637	−0.666
f_3 (BO–I0.05)	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
$\mathbb{P}\big[g_{1,\ldots,4}(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.951	0.967	0.976	0.979
$\mathbb{P}\big[g_1(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	1.000	1.000	1.000	1.000
$\mathbb{P}\big[g_2(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	1.000	1.000	1.000	1.000
$\mathbb{P}\big[g_3(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.974	0.998	1.000	1.000
$\mathbb{P}\big[g_4(\mathbf{s},\mathbf{r}) \leq 0\big]$ (MCS)	0.953	0.967	0.976	0.979

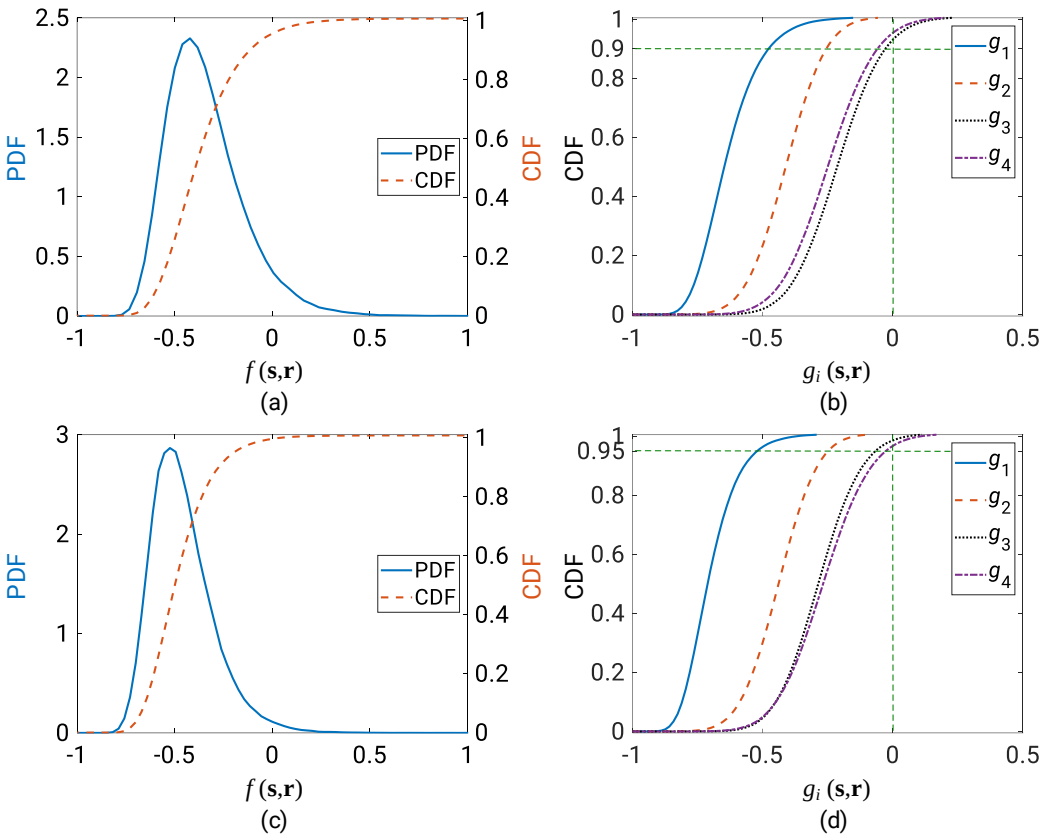

Figure 6.7: Example 1: PDF and CDF of the uncertain objective function $f(\mathbf{s}, \mathbf{r})$, and CDFs of four probabilistic constraint functions at the minimum-mass solutions. (a)–(b) problem J-0.1 (or I-0.1); (c)–(d) problem J-0.05 (or I-0.05).

6.4.3 Example 2: Six-story two-bay steel frame

In this design example, problems J and I are formulated for the six-story two-bay steel frame in Fig. 6.8, which is taken from Section 5.4.2 with modifications of the number of beam groups and the probabilistic properties of random parameters. 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 selected from Table 6.2. The independent random parameters $\mathbf{r}$ for the frame and their probabilistic characteristics are described in Table 6.6, where the vertical load q consists of dead and live loads from the floor.

Table 6.6: Example 2: Assigned random parameters for the six-story two-bay steel frame.

Variable	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
$\sigma_{\mathrm{y},0}$	Yield stress [MPa]	250	1.05	0.06	Normal
σ_{u}	Tensile strength [MPa]	400	1.05	0.08	Normal

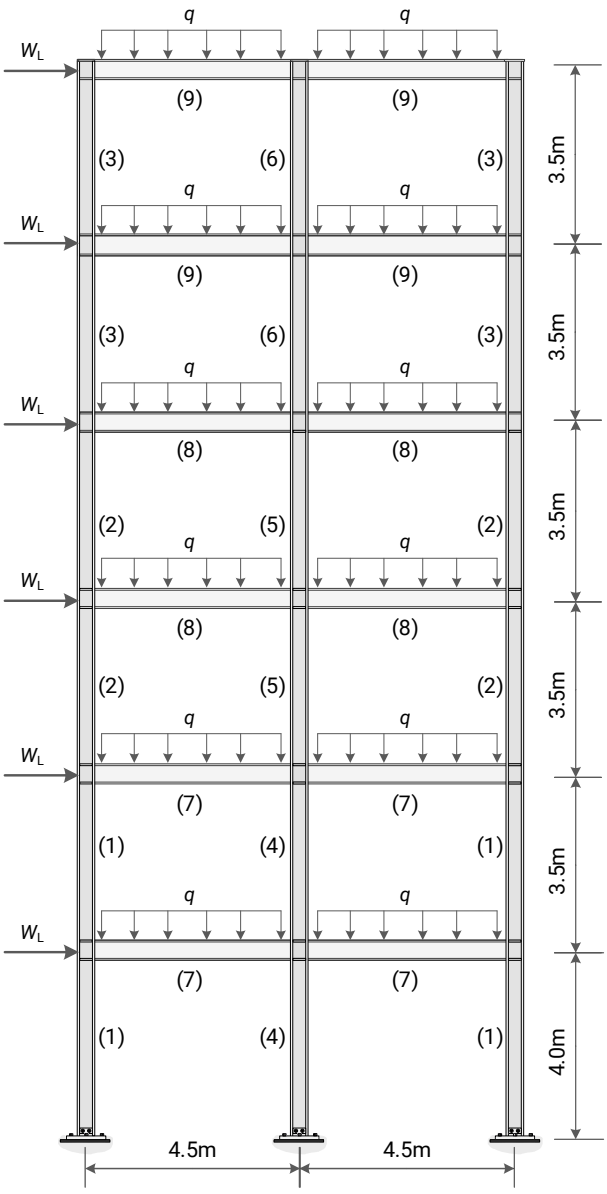

Figure 6.8: Example 2: Six-story two-bay steel frame, groups of members, and external loads.

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 associated with the total drift of the frame, and the maximum vertical long-term 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 $g_{10}(\mathbf{s},\mathbf{r})$, respectively, indicate the LSFs of the combined axial-flexural strength of columns (1), (2), (3), (4), (5), and (6). Also let $g_{11}(\mathbf{s},\mathbf{r})$, $g_{12}(\mathbf{s},\mathbf{r})$, and $g_{13}(\mathbf{s},\mathbf{r})$ represent the LSFs regarding 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., $h_j(\mathbf{s}) \leq 0\,(j = 1,\ldots,10)$, are imposed at beam-column connections and column-column joints of the frame to ensure that (1) the flange width of the 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 J and I of the frame are formulated with 13 probabilistic constraint functions and

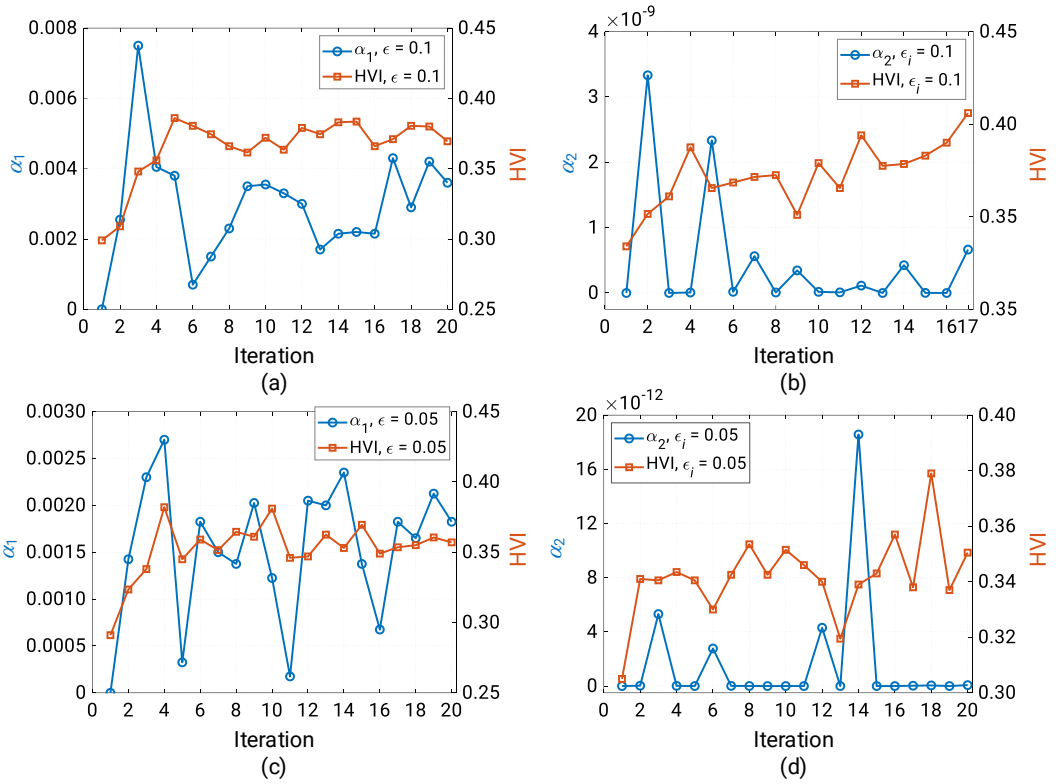

Figure 6.9: Example 2: Histories of HVI and acquisition functions for different RDO problems of the six-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

10 deterministic constraint functions. The total mass $f_1(\mathbf{s})$ of the frame is also given as

$$f_1(\mathbf{s}) = 15\rho_1 + 14(\rho_2 + \rho_3) + 7.5\rho_4 + 7(\rho_5 + \rho_6) + 18(\rho_7 + \rho_8 + \rho_9), \tag{6.13}$$

where ρ_i $(i = 1, \ldots, 9)$ is the nominal mass [kg/m] of the steel section for the i-th group.

To construct a total of 14 GP models for the structural responses of the frame, i.e., one for the uncertain objective function associated with 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 of $\mathbf{s}$ and $\mathbf{r}$ is randomly generated. The proposed BO is then carried out for solving problems J and I with two risk levels 0.1 and 0.05. Similarly to the frame in Example 1, problems J-0.1 and J-0.05, and I-0.1 and I-0.05 refer to problem J formulated with $\epsilon = 0.1$ and $\epsilon = 0.05$, and problem I formulated with $\epsilon_i = 0.1$ and $\epsilon_i = 0.05$ $(i = 1, \ldots, 13)$, respectively. The reference point and the maximum number of BO iterations are $\mathbf{f}_{\mathrm{R}} = [15000\,\mathrm{kg}, 0.5, 0.5]$ and 20, respectively.

Figure 6.9 shows the histories of HVI and acquisition function obtained from solving each RDO problem. BO completes 20 iterations to offer the solutions to problems J-0.1, J-0.05, and I-0.05, while 17 iterations are required for solving problem I-0.1. For the same risk level, the maximum value of the acquisition function α_2 is considerably less than that of α_1 because problems I-0.1 and I-0.05 consider 13 probabilistic constraints independently. The computational times for solving problems J-0.1, J-0.05, I-0.1, and I-0.05 are 11.75, 11.68, 10.19 and 10.56 hours, respectively.

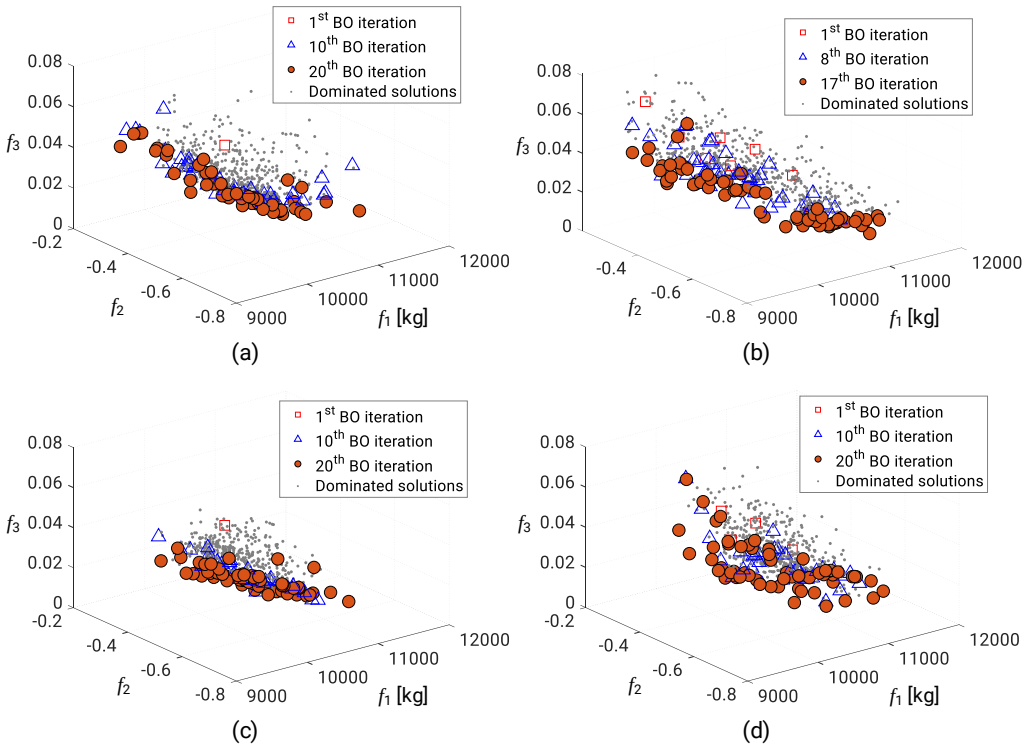

Figure 6.10: Example 2: Pareto-optimal solutions to different RDO problems of the six-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

Figure 6.10 shows the Pareto-optimal solutions to the RDO problems of the frame. The number of solutions to problems J-0.1, I-0.1, J-0.05, and I-0.05, respectively, are 54, 66, 60, and 56, even though they are 1, 7, 1, and 4 after 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 and the solutions themselves may vary when the proposed BO is carried out using different stopping criteria or even when using a unique set of stopping criteria with different random seeds. This can be explained by the following two facts. First, the list of steel sections in Table 6.2 consists of many feasible solutions to each RDO problem of the frame, which are referred to as the dominated solutions in Fig. 6.10. Second, the proposed BO is random in nature because it is governed by a random search method to solve the MINLP problem of the acquisition functions. From the obtained solutions, however, the designer is 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 using MCS.

From the set of obtained Pareto-optimal solutions to each RDO problem, the designer, for example, may wish to select the solution with minimum robustness or maximum robustness as the design for the frame. With each selected design, MCS generates a total of 10^5 samples of the random parameters $\mathbf{r}$ that are further used to compute the associated objective and probabilistic constraint functions. Figure 6.11 compare the PDFs of the uncertain objective function $f(\mathbf{s}, \mathbf{r})$ associated with the minimum- and maximum-robustness designs for each RDO problem of the frame. Figures 6.12 and 6.13 show the CDFs of 13 probabilistic constraint functions, respectively, calculated at the

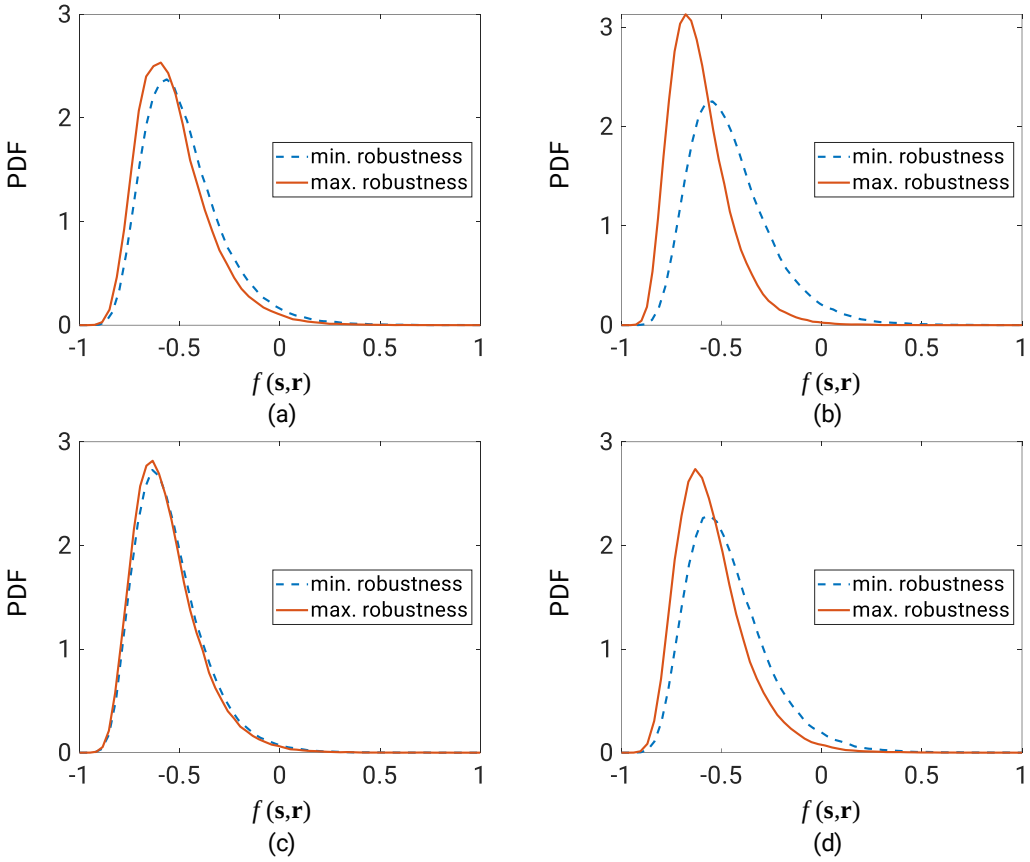

Figure 6.11: Example 2: PDFs the uncertain objective function $f(\mathbf{s}, \mathbf{r})$ at the minimum- and maximum-robustness designs to different RDO problems of the six-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

minimum- and maximum-robustness designs to each RDO problem. Since the design associated with the larger risk level is less robust than that corresponding to the smaller risk level, the PDFs in Figs. 6.11(a) and (b) are broader than those in Figs. 6.11(c) and (d), respectively. In addition, all probabilistic constraints shown in Figs. 6.12 and 6.13 are associated with a safety margin, demonstrating that the minimum-mass solution is a reasonable choice for each RDO problem. The same verification process can be done for any design selected from the obtained Pareto-optimal solutions.

6.5 Conclusions

Although the robust design of steel frames under the effects of random 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 chapter is capable of handling this task with a reasonable computational cost by addressing challenges arising from solving these RDO problems such as the implicit objective and constraint functions, discrete nature of the problems formulated for steel frame structures, and difficulty in evaluation of the probabilistic constraint functions. The feasibility 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 main findings are summarized as follows:

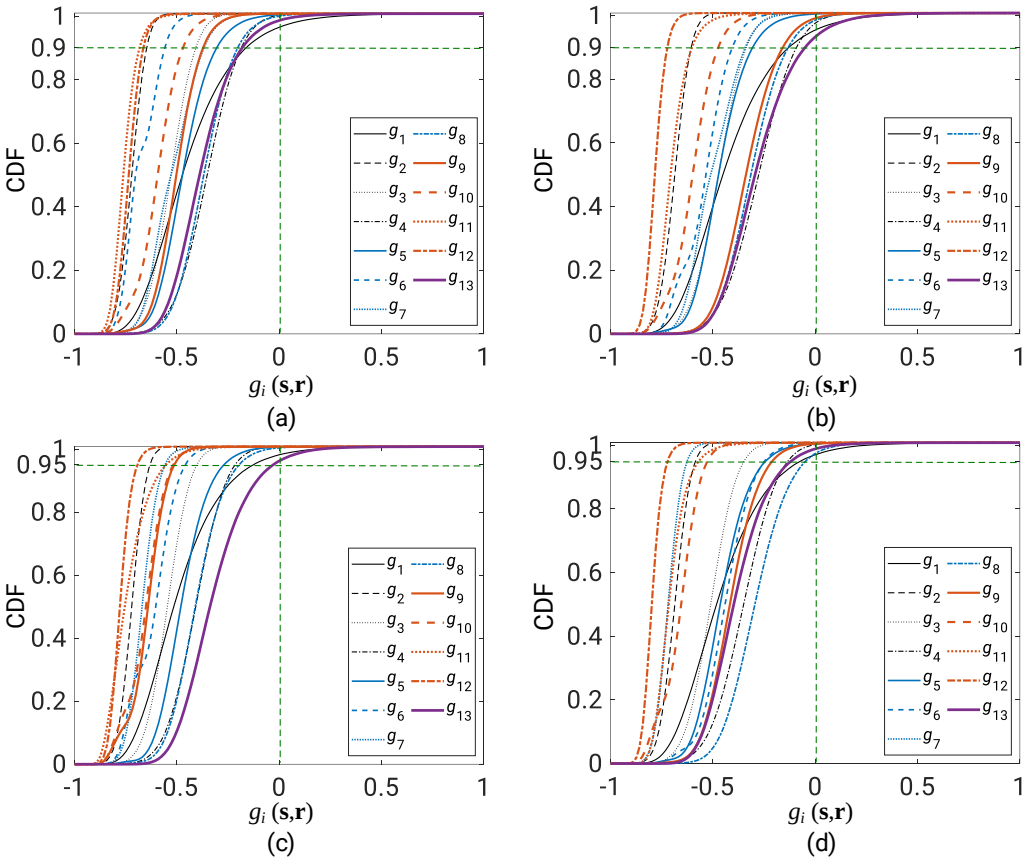

Figure 6.12: Example 2: CDFs of 13 probabilistic constraint functions at the minimum-robustness designs to different RDO problems of the six-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

(1) Two new acquisition functions have been developed for the joint and individual probabilistic constrained RDO problems to guide BO toward better solutions once it completes an iteration. These acquisition functions address the improvement of current solutions and the feasibility of new sampling points simultaneously. This is not only applicable to the RDO problems considered in this chapter but also suggests a basis for the development of constrained BO approaches to solving multi-objective optimization problems with costly objective and/or constraint functions, for example, the approach in Chapter 8.

(2) A new two-stage optimization strategy has been devised for solving the maximization problem of the acquisition functions that decides where the GP surrogates for the structural responses should be refined. Based on this strategy, similar surrogate-based approaches can be proposed for sequentially solving optimization problems that involve uncertain objective and/or constraint functions.

(3) An optimization procedure integrating the proposed BO is introduced to solving the joint and individual probabilistic constrained RDO problems of steel frame structures with discrete design variables.

(4) The optimization results show that the proposed BO can offer exact or good approximate Pareto-optimal solutions to the RDO problems with affordable computational costs.

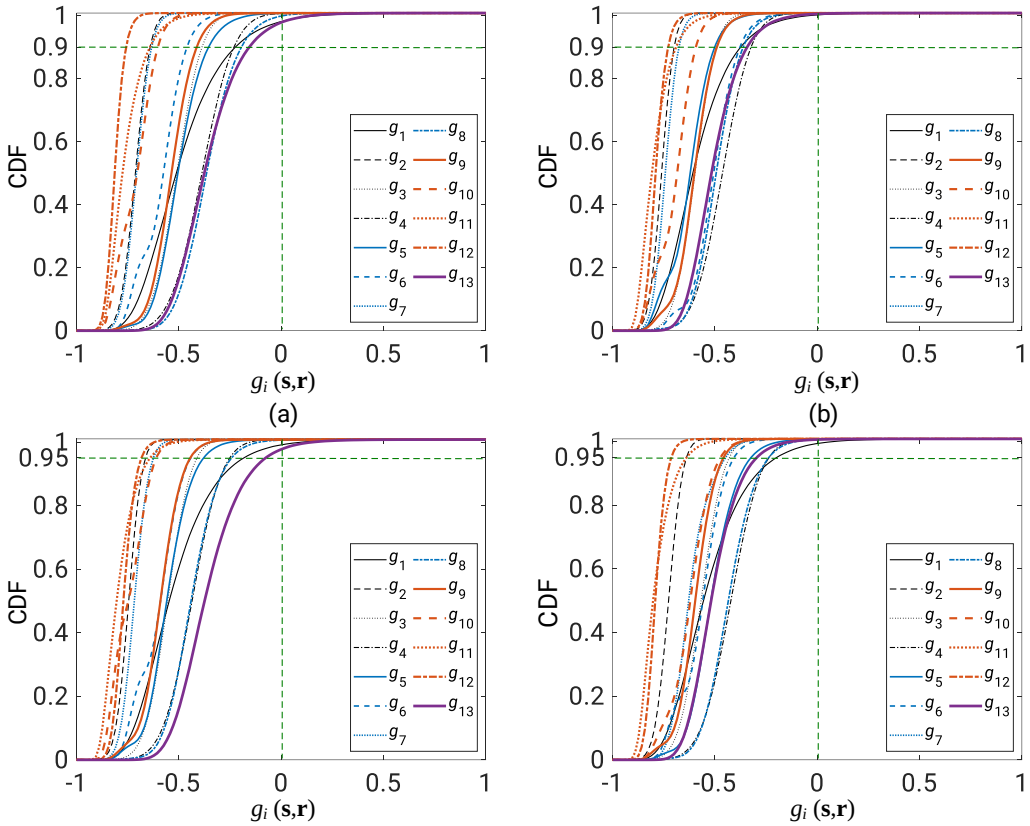

Figure 6.13: Example 2: CDFs of 13 probabilistic constraint functions at the maximum-robustness designs to different RDO problems of the six-story two-bay steel frame. (a) problem J-0.1; (b) problem I-0.1; (c) problem J-0.05; (d) problem I-0.05.

(5) Although the shapes of Pareto fronts to the same RDO problem are similar, different BO trials may lead to different sets of 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. It is, therefore, prudent to select a candidate list of both column and beam sections using design rules of thumb to reduce the discrepancy in the obtained solutions.

Chapter 7

Sequential mixture of Gaussian processes and saddlepoint approximation for reliability-based design optimization of structures

This chapter addresses the RBDO problem stated in Eq. (1.3) using mixture of Gaussian processes (MGP) models and a novel SAA. MGP models serve as surrogates for structural responses that help to reduce the number of structural analysis calls required for processing optimization. MGP is capable of extending the application of GP to large training datasets for well covering the space of input variables, significantly reducing the training time, and improving the overall accuracy of regression models. To handle the probabilistic constraints, the novel SAA is proposed based on the SAA methodology in Section 2.6 and the first three cumulants of LSFs associated with the probabilistic constraints of the RBDO problem. Decoupling the RBDO problem is then processed through a sequential deterministic optimization (SDO) problem in which the MGP models represent the LSFs while the proposed SAA carries out inverse probability evaluations. The SDO problem is strategically solved for exploring a promising region that may contain the optimal solution, refining the MGP models in that region, and offering a reliable optimal solution as it terminates.

7.1 Reliability-based design optimization problem

Recall that problem (1.3) aims to minimize the cost $f(\mathbf{s})$ of a structure so that the probability the structure remains intact $\mathbb{P}[g_i(\mathbf{s},\mathbf{r}) \leq 0]$ (i.e., probability of safety) under the effect of $\mathbf{r}$ is greater than or equal to a value specified by $1-\epsilon_i$, where $\mathbf{s}$ and $\mathbf{r}$ denote vectors of the design variables and random parameters, respectively, ϵ_i is an allowable risk level, and $g_i(\mathbf{s},\mathbf{r})$ denotes the i-th LSF of a design requirement on the serviceability or strength of the structure with $g_i(\mathbf{s},\mathbf{r}) \geq 0$ indicating a failure state.

For a clear description of the reliability method proposed in this chapter, let $l_i(\mathbf{s},\mathbf{r}) = -g_i(\mathbf{s},\mathbf{r})$.

Thus, $l_i(\mathbf{s},\mathbf{r}) \leq 0$ indicates a failure state, and problem (1.3) can be rewritten as

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & f(\mathbf{s}) \\ \text{subject to} \quad & \mathbb{P}[l_i(\mathbf{s},\mathbf{r}) \leq 0] \leq \epsilon_i \quad (i = 1,\ldots,I), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1,\ldots,J), \\ & \mathbf{s} \in [\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}], \end{aligned} \tag{7.1}$$

where $\mathbb{P}[l_i(\mathbf{s},\mathbf{r}) \leq 0]$ is the probability of occurring the i-th failure state and $h_j(\mathbf{s})$ denotes the j-th deterministic constraint function. The design variables $\mathbf{s}$ in this chapter are either continuous or discrete. Since all the probabilistic functions $l_i(\mathbf{s},\mathbf{r})$ are treated equally, we drop the subscript hereafter to simplify the exposition.

Handling the probabilistic constraints is the most important task of solving problem (7.1). As discussed in Section 1.5.2, this can be done through the decoupling approach that is advantageous to reduction of computational cost by the evaluation of inverse probabilities. The decoupling approach can also be combined with a surrogate-based method for solving RBDO problems of large-scale structures [60]. Such a method, however, limits itself to relatively small training datasets. When the training dataset becomes larger for well covering the input variable space (i.e., space of $\mathbf{s}$ and $\mathbf{r}$), the construction of accurate surrogate models over a large range of high-dimensional input variables becomes a challenge. This chapter presents a reliable method for evaluating the inverse probabilities as well as an approach to scaling up the application of the surrogate-based method to large training datasets.

7.2 Mixture of Gaussian processes

7.2.1 Clustering training set using Gaussian mixture model

Consider a training dataset $\mathcal{D} = \{\mathbf{x}^n, l^n\}_{n=1}^N$, where $\mathbf{x}^n = \left[\mathbf{s}^{nT}, \mathbf{r}^{nT}\right]^T$ and $l^n = l(\mathbf{x}^n)$. Since its effectiveness has been verified by a series of successful applications in the previous chapters, the GP model also serves as surrogate for $l(\mathbf{x})$ in this chapter. The GP modeling, however, limits itself to relatively small training sets as it requires $\mathcal{O}(N^3)$ time for training, which is due to the computation of the inversion and determinant of the $N \times N$ kernel matrix [58]. To extend the application of GP to a larger training dataset $\mathcal{D}$, it is desirable to split $\mathcal{D}$ into a finite number of subsets. For this purpose, we isolate relevant samples from other irrelevant samples of $\mathcal{D}$ by grouping them into a subset using the GMM-based clustering method [78]. In this way, $\mathcal{D}$ can be divided into a finite number of independent subsets that have a small number of samples as compared with N, and each subset corresponds to a Gaussian component constituting GMM. From these subsets, the corresponding local GP models are constructed and mixed up to form the MGP.

Clustering a training dataset is to distribute its similar samples to an independent group in which the samples share a general property. Two fundamental steps of clustering a training dataset are measuring the similarity of the samples and selecting a clustering algorithm. Different similarity measures and clustering algorithms can be found in Ref. [134]. This chapter premises that two training samples are similar if they emerge from the same PDF. It is, therefore, convenient to split the joint PDF $\pi(\mathbf{x}, l)$ into different Gaussian components using the GMM presented in Section 2.5, assign

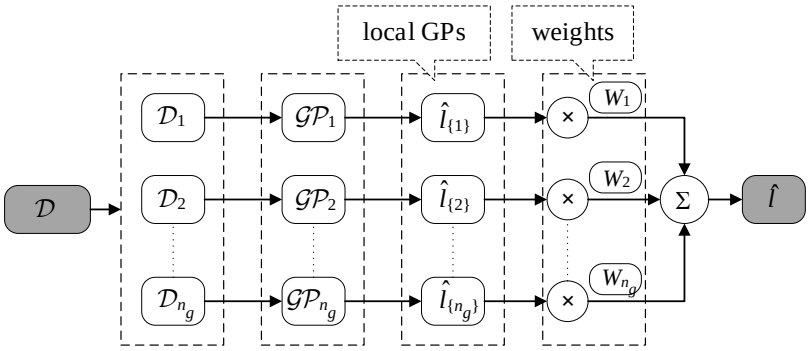

Figure 7.1: Illustration of the construction of an MGP for a given training set.

each component as a subset, and then distribute the training samples into each subset accordingly.

Suppose a GMM has been constructed from $\mathcal{D}$ that is a weighted sum of n_g Gaussian components and characterized by the parameters w_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ ($k = 1, \ldots, n_\mathrm{g}$). Following Eq. (2.51), the probability that an input variable vector $\mathbf{x}$ emerges from the k-th Gaussian component can be determined (without knowing the corresponding $l(\mathbf{x})$) by

$$z^k(\mathbf{x}) = \frac{w_k \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right)}{\sum_{k=1}^{n_\mathrm{g}} w_k \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}\right)}, \tag{7.2}$$

where $\boldsymbol{\mu}_{x,k}$ and $\boldsymbol{\Sigma}_{xx,k}$ are directly extracted from $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ using Eq. (2.41c).

7.2.2 Mixture of Gaussian processes (MGP)

Figure 7.1 illustrates the process of constructing an MGP model for $l(\mathbf{x})$ from $\mathcal{D}$ [135]. After $\mathcal{D}$ is split into n_g independent subsets using the GMM clustering method, i.e., $\mathcal{D} = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_{n_\mathrm{g}}$, we construct the local GP model associated with the k-th subset, denoted by $\hat{l}_{\{k\}}(\mathbf{x})$. The MGP model, denoted by $\hat{l}(\mathbf{x})$, is defined as a weighted average function of the resulting local GP models. By doing so, the overall accuracy of the regression model, as will be shown in Section 7.5, can be improved. The computation time required for training the MGP can also be reduced by carrying out parallel computation for the training process of the independent subsets of $\mathcal{D}$.

To this end, the MGP model is mathematically expressed as

$$\hat{l}(\mathbf{x}) = \sum_{k=1}^{n_\mathrm{g}} W_k(\mathbf{x}) \hat{l}_{\{k\}}(\mathbf{x}), \tag{7.3}$$

where the weight $W_k(\mathbf{x})$ evaluated by Eq. (7.2) denotes the probability that the input variable vector $\mathbf{x}$ belongs to the k-th subset. Algorithm 7.1 details the process of constructing the MGP for $l(\mathbf{x})$ from the training dataset $\mathcal{D}$.

Algorithm 7.1 GMM-based MGP

Input $\mathcal{D} = \{\mathbf{x}^n, l^n\}_{n=1}^N$;
Construct GMM from $\mathcal{D}$ using EM algorithm (Algorithm 2.3) and BIC;
return n_g, w_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$;
Extract $\boldsymbol{\mu}_{x,k}$ and $\boldsymbol{\Sigma}_{xx,k}$ from $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ using Eq. (2.41c);
$\mathcal{D}_k \leftarrow \emptyset \;\; (k = 1, \ldots, n_\mathrm{g})$;
for $n = 1 : N$ **do**
    for $k = 1 : n_\mathrm{g}$ **do**
        Compute $z^k(\mathbf{x}^n)$ for $(\mathbf{x}^n, l^n)$ using Eq. (7.2);
    end for
    $k \leftarrow \max\{z^k(\mathbf{x}^n), k = 1, \ldots, n_\mathrm{g}\}$;
    $\mathcal{D}_k \leftarrow \mathcal{D}_k \cup \{\mathbf{x}^n, l^n\}$; ▷ Distribute each sampling point to suitable subset
end for
for $k = 1 : n_\mathrm{g}$ **do**
    Construct $\hat{l}_{\{k\}}(\mathbf{x})$ from $\mathcal{D}_k$;
end for
return $\hat{l}(\mathbf{x})$ using Eq. (7.3) with $W_k(\mathbf{x})$ given in Eq. (7.2).

7.3 Proposed saddlepoint approximation

7.3.1 Saddlepoint approximation using the first three cumulants of uncertain LSFs

Based on the MGP $\hat{l}(\mathbf{x})$ and the SAA methodology presented in Section 2.6, this section addresses the following forward and inverse probability problems. Forward probability problem: given a threshold value of l, denoted as $\bar{l}$, evaluate $P_\mathrm{f} = \mathbb{P}[l(\mathbf{x}) \leq \bar{l}]$. Inverse probability problem: given P_f, evaluate $\bar{l}$ so that $\mathbb{P}[l(\mathbf{x}) \leq \bar{l}] = P_\mathrm{f}$, which is indeed the calculation of the quantile function value associated with P_f; see Eq. (2.3).

Let μ_l, σ_l^2, and $\sigma_{l,3}$ denote the mean, variance, and third central moment of $l(\mathbf{x})$. The first three cumulants of $l(\mathbf{x})$, under the effects of uncertainty in independent input variables, can be estimated through the first-order Taylor series expansion with respect to $\mathbf{x} = [x_1, \ldots, x_d]^T$ at the mean $\boldsymbol{\mu}_x$, such that

$$\kappa_{l,1} = \mu_l \approx \hat{l}(\boldsymbol{\mu}_x), \tag{7.4a}$$

$$\kappa_{l,2} = \sigma_l^2 \approx \sum_{j=1}^{d} \left(\left. \frac{\partial \hat{l}}{\partial x_j} \right|_{\boldsymbol{\mu}_x} \right)^2 \sigma_{x_j}^2, \tag{7.4b}$$

$$\kappa_{l,3} = \sigma_{l,3} \approx \sum_{j=1}^{d} \left(\left. \frac{\partial \hat{l}}{\partial x_j} \right|_{\boldsymbol{\mu}_x} \right)^3 \sigma_{x_j,3}, \tag{7.4c}$$

where $\sigma_{x_j}^2$ and $\sigma_{x_j,3}$ denote the variance and third central moment of x_j, respectively.

To develop the SAA and its inverse, respectively, for handling the forward and inverse probability problems, we first propose the following truncated cumulant generating function $K_l(\xi)$ based on the

first two terms of the power series expansion and a logarithmic term by Guo [85]:

$$K_l(\xi) = (\kappa_{l,1} - 2ab)\xi + \frac{1}{2}(\kappa_{l,2} - 2ab^2)\xi^2 - a\log\left[(1-b\xi)^2\right], \tag{7.5}$$

where a and b are unknowns to be determined based on (1) the relations between $K_l(\xi)$ and the first three cumulants of of $l(\mathbf{x})$, and (2) the condition for the unique root of the saddlepoint equation in Eq. (2.56).

The first, second, and third derivatives of $K_l(\xi)$ are derived as follows:

$$K_l^{(1)}(\xi) = (\kappa_{l,1} - 2ab) + (\kappa_{l,2} - 2ab^2)\xi + \frac{2ab}{1-b\xi}, \tag{7.6a}$$

$$K_l^{(2)}(\xi) = (\kappa_{l,2} - 2ab^2) + \frac{2ab^2}{(1-b\xi)^2}, \tag{7.6b}$$

$$K_l^{(3)}(\xi) = \frac{4ab^3}{(1-b\xi)^3}. \tag{7.6c}$$

The relations between $K_l(\xi)$ and the first three cumulants of $l(\mathbf{x})$, as described in Section 2.1.2, give

$$K_l^{(1)}(0) = \kappa_{l,1}, \quad K_l^{(2)}(0) = \kappa_{l,2}, \quad K_l^{(3)}(0) = \kappa_{l,3}. \tag{7.7}$$

By letting ξ in Eqs. (7.6a), (7.6b), and (7.6c) be zero, the proposed truncated $K_l(\xi)$ automatically satisfies the first two conditions in Eq. (7.7). The remaining condition leads to

$$4ab^3 = \kappa_{l,3}. \tag{7.8}$$

Substituting $K_l^{(1)}(\xi)$ in Eq. (7.6a) into Eq. (2.56), the saddlepoint equation reads

$$-b(\kappa_{l,2} - 2ab^2)\xi^2 + (-b\kappa_{l,1} + \kappa_{l,2} + \bar{l}b)\xi + \kappa_{l,1} - \bar{l} = 0. \tag{7.9}$$

As mentioned above, the saddlepoint equation should have a unique real root for any value $\bar{l}$. Therefore, the following condition should be satisfied:

$$-b(\kappa_{l,2} - 2ab^2) = 0. \tag{7.10}$$

By solving the system of Eqs. (7.8) and (7.10), a and b can be found as

$$a = \frac{2\kappa_{l,2}^3}{\kappa_{l,3}^2}, \quad b = \frac{\kappa_{l,3}}{2\kappa_{l,2}}. \tag{7.11}$$

If $\kappa_{l,3} = 0$, then $a = b = 0$.

The saddlepoint ξ_s as the root of Eq. (7.9) reads

$$\xi_\mathrm{s} = \frac{\bar{l} - \kappa_{l,1}}{-b\kappa_{l,1} + \kappa_{l,2} + \bar{l}b}. \tag{7.12}$$

By substituting ξ_s, $K_l(\xi_\mathrm{s})$, and $K_l^{(2)}(\xi_\mathrm{s})$ into Eq. (2.57), we obtain $P_\mathrm{f} = \mathbb{P}[l(\mathbf{x}) \leq \bar{l}]$. Note that the terms inside the square root in Eqs. (2.58a) and (2.58b) for evaluation of q and v depend on $\bar{l}$ and must be non-negative to ensure the existence of q and v. Thus, $\bar{l}$ in Eq. (7.12) should be bounded.

In our problems, the proposed $K_l(\xi)$ is valid in the interval $l \in [l_\mathrm{l}, l_\mathrm{u}]$, where $\mathbb{P}[l(\mathbf{x}) \leq l_\mathrm{l}] = 0$ and $\mathbb{P}[l(\mathbf{x}) \leq l_\mathrm{u}] = 1$. The probability of failure is equal to 0 if $l < l_\mathrm{l}$ and 1 if $l > l_\mathrm{u}$.

Algorithms 7.2 and 7.3 are proposed, respectively, for addressing the forward probability problem that calculates P_f when $\bar{l}$ is specified, and for calculating the inverse probability problem that estimates $\bar{l}$ so that $\mathbb{P}[l(\mathbf{x}) \leq \bar{l}] = P_\mathrm{f}$. To implement Algorithm 7.3, we should specify the interval $[l_\mathrm{l}, l_\mathrm{u}]$ over which l varies under the effects of the random parameters. To make the determination of $[l_\mathrm{l}, l_\mathrm{u}]$ easier, we normalize the LSF and add to the normalized LSF an offset value so that if the LSF value is negative at a particular input variable point, a failure occurs at that point. Yet exact values of l_l and l_u are not necessary for calculating the inverse probability. We can simply select a wide range of $[l_\mathrm{l}, l_\mathrm{u}]$ with a small value of l_l and then increase l from l_l step by step for tracking $\bar{l}$. If either q and v does not exist, we set the corresponding probability as zero and continue increasing l until we meet P_f.

Algorithm 7.2 Forward probability using the proposed SAA

Input $\hat{l}(\mathbf{x})$, $\boldsymbol{\mu}_x$, and $\bar{l}$;
Compute $\kappa_{l,1}$, $\kappa_{l,2}$, and $\kappa_{l,3}$ using Eqs. (7.4a), (7.4b), and (7.4c), respectively;
Compute a and b using Eq. (7.11);
Compute ξ_s using Eq. (7.12);
Compute $K_l(\xi_\mathrm{s})$ and $K_l^{(2)}(\xi_\mathrm{s})$ using Eqs. (7.5) and (7.6b), respectively;
Compute q and v using Eqs. (2.58a) and (2.58b), respectively;
return $P_\mathrm{f} = \mathbb{P}[l(\mathbf{x}) \leq \bar{l}]$ using Eq. (2.57).

Algorithm 7.3 Inverse probability using the proposed SAA

Input $\hat{l}(\mathbf{x})$, $\boldsymbol{\mu}_x$, P_f, $[l_\mathrm{l}, l_\mathrm{u}]$, and a small step size $\delta_l = 0.001$;
$l^0 \leftarrow l_\mathrm{l}$, $P_\mathrm{f}^0 \leftarrow 0$, $j \leftarrow 1$;
while $P_\mathrm{f}^{j-1} < P_\mathrm{f}$ **do**
    $l^j \leftarrow l^{j-1} + \delta_l$;
    Compute $\kappa_{l,1}$, $\kappa_{l,2}$, and $\kappa_{l,3}$ using Eqs. (7.4a), (7.4b), and (7.4c), respectively;
    Compute a and b using Eq. (7.11);
    Compute $K_l(\xi_\mathrm{s})$ and $K_l^{(2)}(\xi_\mathrm{s})$ using Eqs. (7.5) and (7.6b), respectively;
    if q and v in Eqs. (2.58a) and (2.58b) exist **then**
        $P_\mathrm{f}^j \leftarrow \mathbb{P}[l(\mathbf{x}) \leq l^j]$ using Eq. (2.57);
    else
        $P_\mathrm{f}^j \leftarrow 0$;
    end if
    $j \leftarrow j + 1$;
end while
return $\bar{l} \leftarrow l^j$.

7.3.2 Application to structural reliability analysis

The proposed SAA is applied to estimating the failure probability of a cantilever beam with a circular hollow section in Fig. 7.2, which is adapted from Ref. [136]. The beam is subjected to the diagonal loads F_1 and F_2, axial load P, and torsion T. The failure mode of the beam is associated with the violation of allowable value of the von Mises stress evaluated at point A located at the upper edge of the fixed support. The random parameters for the beam include the external loads F_1, F_2, P,

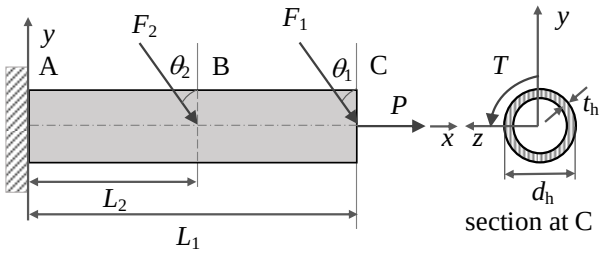

Figure 7.2: Cantilever hollow tube.

Table 7.1: Probabilistic properties of random parameters for the cantilever hollow tube.

Variable	Distribution	Mean (*lower bound)	Standard deviation (*upper bound)	Unit
t_h	Normal	4.0	0.04	mm
d_h	Normal	40.0	0.4	mm
L_1	Uniform	$120^\star$	$125^\star$	mm
L_2	Uniform	$60^\star$	$65^\star$	mm
F_1	Normal	17.0	1.7	kN
F_2	Normal	17.0	1.7	kN
P	Normal	19.0	1.9	kN
T	Normal	90.0	9.0	Nm
$\sigma_{\mathrm{y},0}$	Normal	220	22	MPa
θ_1	Normal	$\pi/36$	$\pi/360$	rad
θ_2	Normal	$\pi/18$	$\pi/180$	rad

and T; the locations of the external loads L_1 and L_2; the load directions θ_1 and θ_2; the diameter d_h and thickness t_h of the hollow tube section; and the yield stress $\sigma_{\mathrm{y},0}$ of the tube material. These random variables are assumed to be independent and their probabilistic properties are described in Table 7.1.

The LSF corresponding to the von Mises stress at A is defined by

$$l = 1 - \frac{\sigma_\mathrm{A}}{\sigma_{\mathrm{y},0}} = 1 - \frac{\sqrt{\sigma_{\mathrm{A,x}}^2 + 3\tau_{\mathrm{A,xz}}^2}}{\sigma_{\mathrm{y},0}}, \tag{7.13}$$

where

$$\sigma_{\mathrm{A,x}} = \frac{P + F_1 \sin\theta_1 + F_2 \sin\theta_2}{\pi\left[d_\mathrm{h}^2 - (d_\mathrm{h} - 2t_\mathrm{h})^2\right]/4} + \frac{(F_1 L_1 \cos\theta_1 + F_2 L_2 \cos\theta_2) d_\mathrm{h}}{\pi\left[d_\mathrm{h}^4 - (d_\mathrm{h} - 2t_\mathrm{h})^4\right]/32}, \tag{7.14a}$$

$$\tau_{\mathrm{A,xz}} = \frac{T d_\mathrm{h}}{\pi\left[d_\mathrm{h}^4 - (d_\mathrm{h} - 2t_\mathrm{h})^4\right]/16}. \tag{7.14b}$$

Algorithm 7.2 is called to generate the CDF of the highly-nonlinear LSF l in Eq. (7.13). The resulting CDF is compared with the CDFs generated by the mean-value first-order SAA [137], MCS with 10^5 samples, and a special case of the SAA by Guo [85] when the unique real root condition of the saddlepoint equation in the standardized space is satisfied. The comparison results in Fig. 7.3 show that the CDFs generated by the proposed SAA and the special case of SAA by Guo [85] are almost identical, which indicates that the transformation of the random parameters from

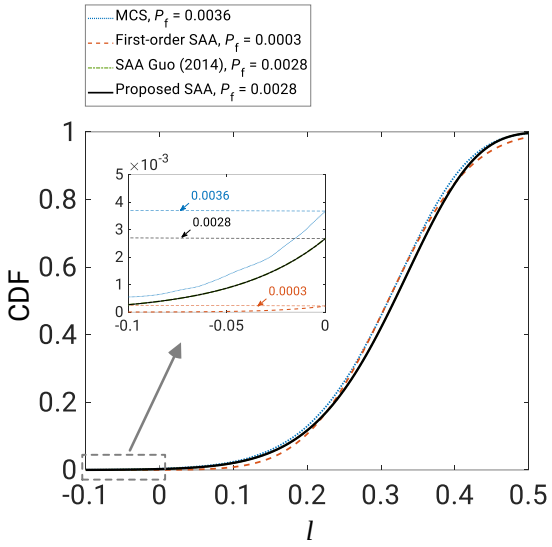

Figure 7.3: Comparison of CDFs of the LSF for the maximum von Mises stress at A.

their physical space into a standardized space [85] is unnecessary if the unique real root condition of the saddlepoint equation is taken into account. The failure probability of the cantilever tube by the mean-value first-order SAA is 0.0003 that underestimates the value 0.0036 by MCS. The failure probability by the proposed SAA is 0.0028, which is 78% of that by MCS. This difference is acceptable since it can represent a trade-off between the error for estimation of the failure probability and the computational cost. This difference, however, should be considered when using a gradient-based algorithm to solve the optimization problem whose probabilistic constraints are estimated by the proposed SAA because the optimal solution in this circumstance is on the boundary of the feasible region, which becomes infeasible due to even a small error in the calculation of the probabilistic constraints.

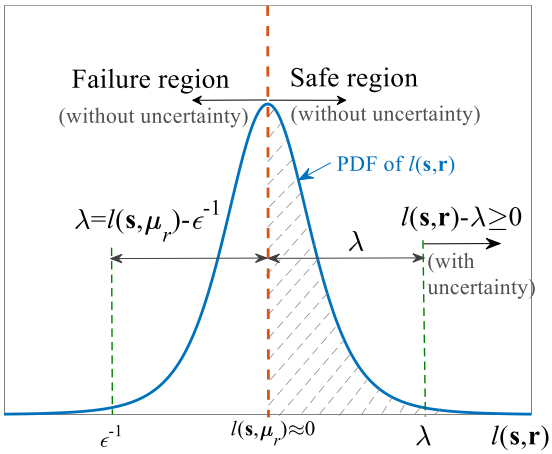

Figure 7.4: SDO problem scheme.

7.4 Design optimization procedure

7.4.1 A sequential deterministic optimization (SDO) for RBDO problem

To explore a promising region that may contain the optimal solution and to improve the accuracy of the MGP models for $l_i(\mathbf{s}, \mathbf{r})$ in that region, we replace problem (7.1) with an SDO problem that is developed by taking the idea of using the partial factors in structural analysis and design, namely, simultaneously increasing the load effects and reducing the structural resistances. In particular, we offset the LSFs at the nominal values of design and random variables to a safe region for considering uncertainty in the load effects and structural resistances. Similar shift approaches to solving RBDO problems can be found in Refs. [51, 138]. Accordingly, the SDO problem associated with problem (7.1) is stated as

$$\begin{aligned} \mathbf{s}^t = &\underset{\mathbf{s}}{\operatorname{argmin}}\ f(\mathbf{s}) \\ \text{subject to}\quad & l_i^t(\mathbf{s}, \boldsymbol{\mu}_r) - \lambda_i^t \geq 0 \quad (i = 1, \ldots, I), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & \mathbf{s} \in [\mathbf{s}_\mathrm{l}, \mathbf{s}_\mathrm{u}], \end{aligned} \tag{7.15}$$

where t denotes the current cycle of the SDO problem, $\boldsymbol{\mu}_r$ is the mean vector of $\mathbf{r}$, $l_i^t(\mathbf{s}, \boldsymbol{\mu}_r)$ is the current nominal value of the i-th LSF that is approximated by the MGP model at the mean vector of input variables, λ_i^t is the current i-th constraint-offset value, and $\mathbf{s}^t$ is the current solution to the SDO problem. The idea is to offset the nominal LSF value evaluated at the mean vector of the input variables, under the effect of random parameters $\mathbf{r}$, to the safe region by at least an amount of λ_i^t, which is depicted in Fig. 7.4. As a result, the most probable point moves discontinuously during the optimization process as the solutions to SDO cycles are generally different.

The update of constraint-offset values is given by

$$\lambda_i^{t+1} = l_i^t(\mathbf{s}^t, \boldsymbol{\mu}_r) - \epsilon_i^{-1} \quad (i = 1, \ldots, I), \tag{7.16}$$

where ϵ_i^{-1} denotes the inverse value of ϵ_i estimated by Algorithm 7.3, which is to ensure that the probability of failure is sequentially approaches the risk level by offsetting the nominal LSFs. At the initial cycle of the sequence of SDO problem, we set l_i^0 as the initial MGPs and $\lambda_i^0 = 0$.

To obtain a new sample for updating the MGP of the current cycle, we perform the FE analysis with the obtained $\mathbf{s}^t$ to determine $l(\mathbf{s}^t, \boldsymbol{\mu}_r)$. The new sample, consisting of $\mathbf{s}^t$, $\boldsymbol{\mu}_r$, and $l(\mathbf{s}^t, \boldsymbol{\mu}_r)$, is added to the cluster that maximizes $z^k(\mathbf{x})$ in Eq. (7.2). Subsequently, only local GP corresponding to such a cluster is reconstructed for updating the current MGP. The updated MGP starts a new cycle of the SDO problem.

The SDO problem terminates when $||\boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}^t|| \leq \epsilon_\mathrm{o}$ or $t > t_\mathrm{max}$, where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_I]^T$ is the constraint-offset vector, ϵ_o is a small positive value, and t_max is the specified limit of the number of SDO cycles.

7.4.2 Summary of the optimization procedure

The optimization procedure for solving the RBDO problem performs the following steps:

- **Step 1**: Generate samples of $\mathbf{s}$ and $\mathbf{r}$ using Latin hypercube sampling [74]. Also generate the training data set $\mathcal{D}$ by performing FE analyses for the generated samples.
- **Step 2**: Based on $\mathcal{D}$, call Algorithm 7.1 for obtaining the initial MGP models for the LSFs $l_i(\mathbf{s}, \mathbf{r})$.
- **Step 3**: Formulate the RBDO problem and its associated SDO problem. Describe the initial cycle of the SDO problem using the MGP models obtained in **Step 2** as surrogates for the LSFs of the probabilistic constraints.
- **Step 4**: Solve the current SDO problem using an appropriate optimization algorithm. Generate new data samples from the resulting optimal solution and the corresponding LSF values, and add them to the training set for updating the current MGPs. Terminate the SDO process when at least one of the stopping conditions is satisfied and output the final solution. Otherwise, formulate the new SDO and solve it with the updated MGP models.

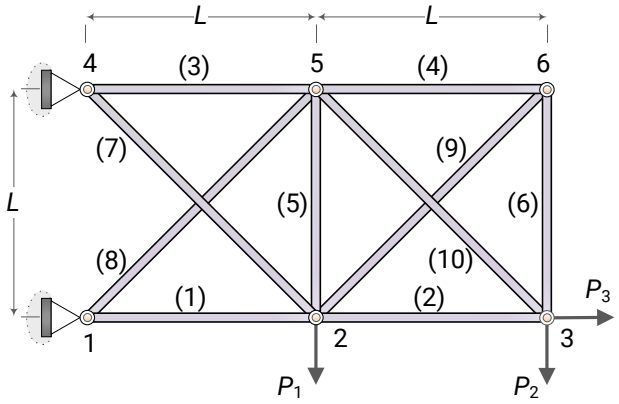

Figure 7.5: Example 1: Ten-bar truss.

7.5 Design examples

7.5.1 Example 1: Ten-bar truss

Problem (7.1) is formulated to minimize the total cross-sectional area of a ten-bar truss in Fig. 7.5. Cross-sectional areas of the truss members are considered as the design variables, namely $\mathbf{s} = [s_1, \ldots, s_{10}]^T$ with $s_i \in [1, 20] \times 10^{-4}\,\mathrm{m}^2$. External loads P_1 and P_2, Young's modulus E of the truss material, and the dimension L constitute vector $\mathbf{r}$ of independent random parameters whose probabilistic properties are given in Table 7.2. The probability that the vertical displacement at node 3, denoted as $\Delta_3(\mathbf{s}, \mathbf{r})$, exceeds an allowable value of 4×10^{-3} m should be less than or equal to $\epsilon = 6.21 \times 10^{-3}$ [139]. Thus, the RBDO problem of the truss is formulated as follows:

Table 7.2: Example 1: Probabilistic properties of random parameters for the ten-bar truss.

Parameter	Distribution	Mean value	COV
P_1 [kN]	Normal	60	0.20
P_2 [kN]	Normal	40	0.20
P_3 [kN]	Normal	10	0.20
E [GPa]	Normal	200	0.10
L [m]	Normal	1	0.05

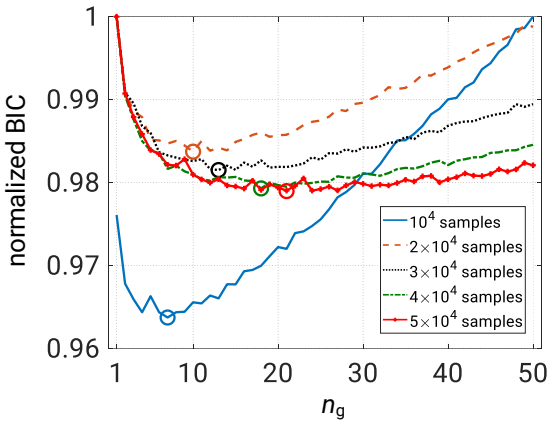

Figure 7.6: Example 1: Selection of GMMs for clustering five training datasets generated for the ten-bar truss.

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & f(\mathbf{s}) = \sum_{i=1}^{10} s_i \\ \text{subject to} \quad & \mathbb{P}[l(\mathbf{s},\mathbf{r}) \le 0] \le \epsilon, \\ & s_i \in [1,20] \times 10^{-4}\,\mathrm{m}^2 \quad (i = 1,\ldots,10), \end{aligned} \tag{7.17}$$

where $l(\mathbf{s},\mathbf{r}) = 4 \times 10^{-3} - \Delta_3(\mathbf{s},\mathbf{r})$.

The MGP for $l(\mathbf{s},\mathbf{r})$ is found for formulation of the SDO problem associated with problem (7.17). To select a proper MGP model, we construct five MGP models from a total of five training datasets with different numbers of samples, namely 10^4, 2×10^4, 3×10^4, 4×10^4, and 5×10^4, using a PC with an Intel(R) Xeon(R) E5-2643V4 3.40 GHz CPU and 64 GB memory. Parallel computation is applied to the construction of MGP models. After training, the prediction performance of the resulting MGP models is assessed through the coefficient of determination R^2 for an independent test set of 2×10^4 random samples.

We also construct a global GP model from each training dataset to obtain a comparison with the corresponding MGP model. This comparison is to examine the computation times required for training the MGP and global GP, and the prediction performance of these models.

Figure 7.6 show the selection of the number of clusters (i.e., components of GMM) for splitting each training dataset generated for the truss. Accordingly, the training datasets of 10^4, 2×10^4, 3×10^4, 4×10^4, and 5×10^4 samples are divided into 7, 10, 13, 18, and 21 clusters, respectively.

Table 7.3: Example 1: Comparison of the training computation time and prediction performance of MGP with those of global GP.

No. sample	Computation time global GP [s]	Computation time MGP [s]	Testing R^2 global GP	Testing R^2 MGP
10^4	4701	1386	0.982	0.991
2×10^4	32364	3208	0.970	0.998
3×10^4	not available	6383	not available	0.995
4×10^4	not available	8502	not available	0.996
5×10^4	not available	12922	not available	0.994

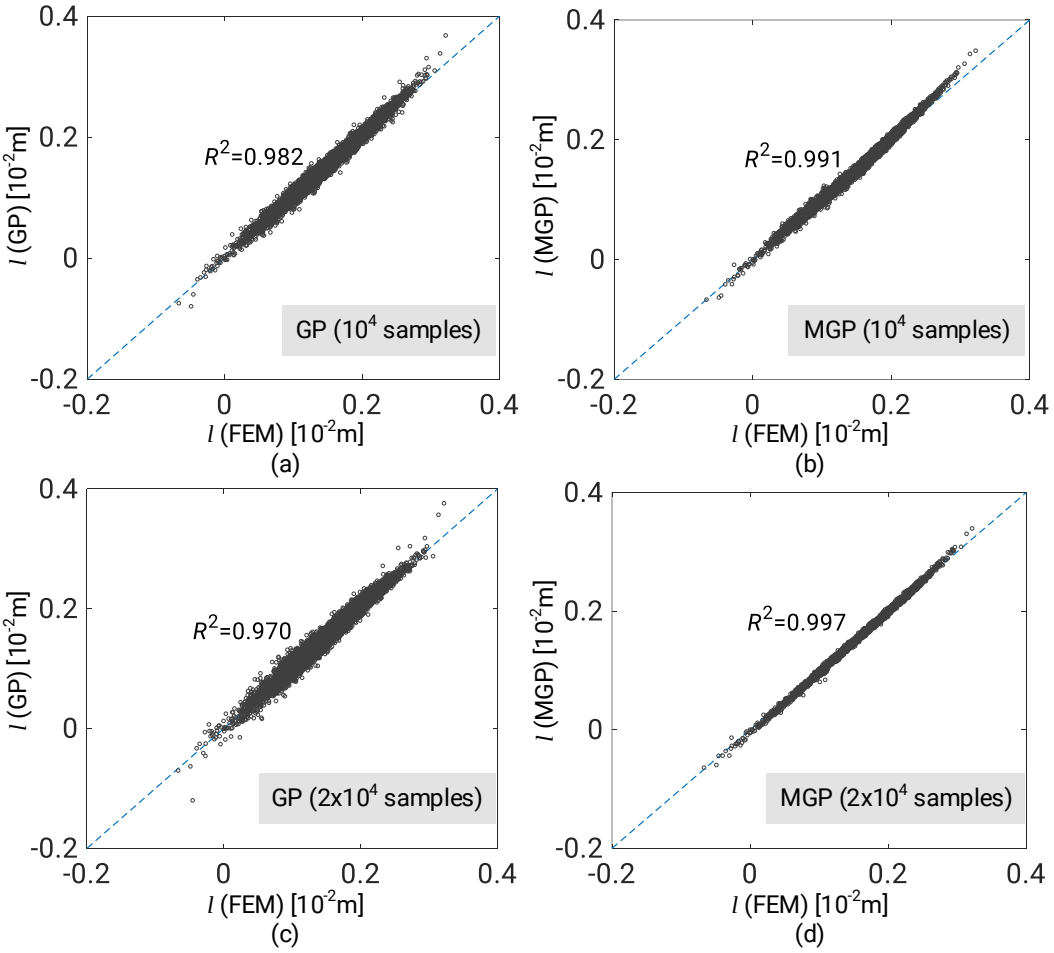

Figure 7.7: Test set prediction performance of the trained global GPs and corresponding MGPs for the ten-bar truss. (a) Global GP trained by 10^4 samples; (b) MGP trained by 10^4 samples; (c) global GP trained by 2×10^4 samples; (d) MGP trained by 2×10^4 samples.

Table 7.3 reports the computation time and R^2 value for MGP and global GP associated with each training dataset. The computation time required for training the MGP (including time for clustering the training dataset) is considerably less than that required for training the corresponding global GP. Although training the global GP is impossible for the mentioned PC specifications when the training dataset becomes larger with 3×10^4, 4×10^4, or 5×10^4 samples, training the associated

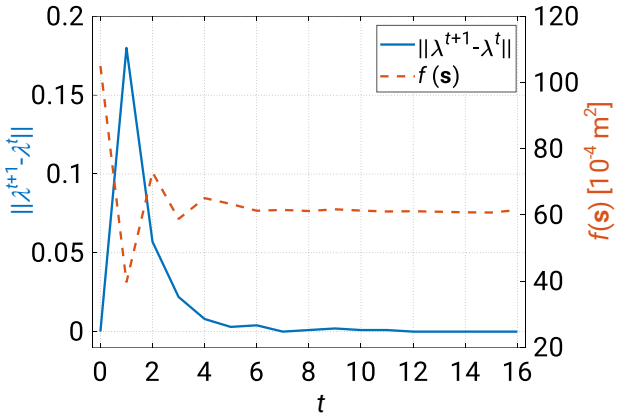

Figure 7.8: Example 1: Histories of the constraint-offset value and objective function with respect to the cycles of the SDO problem of the ten-bar truss.

Table 7.4: Example 1: Comparison of optimization results of the ten-bar truss.

Variable [10^{-4} m^2]	FE and FORM [139]	RSM2, FORM, and MCS [139]	MGP and SA [This study]
s_1	10.493	10.705	10.333
s_2	5.772	5.914	5.371
s_3	14.098	14.424	13.579
s_4	1.000	1.000	1.000
s_5	1.000	1.000	1.000
s_6	1.000	1.000	1.000
s_7	5.460	5.531	6.418
s_8	11.586	11.853	11.273
s_9	1.000	1.000	1.000
s_{10}	10.958	11.223	10.508
$f(\mathbf{s})$ [10^{-4} m^2]	62.367	63.649	61.482
$\mathbb{P}[l(\mathbf{s},\mathbf{r}) \leq 0]$	8.51×10^{-3}	6.19×10^{-3}	4.34×10^{-3}
MCS	4.22×10^{-3}	2.95×10^{-3}	5.64×10^{-3}

MGP is still possible. Moreover, the testing R^2 value by the trained MGP is higher than that by the corresponding global GP, indicating a better prediction performance of the MGP, which is visualized by scatter plots in Fig. 7.7. Since the trained MGP associated with 2×10^4 samples offers the highest R^2 with a reasonable computation time, it is used for formulating the SDO problem.

The SDO problem associated with problem (7.17) is solved by a sequential quadratic programming (SQP) algorithm, available in the function *fmincon* of MATLAB R2018a Optimization Toolbox [140]. The gradient required by SQP is evaluated using the finite difference approximation. Since the solution to each step of the SDO problem obtained using SQP is on the boundary of the feasible region, and the probabilistic constraint is satisfied with equality, the solution is usually infeasible due to a small error tolerance of the inequality constraint. Without loss of generality, we, therefore, use 70% of the risk level ϵ to enforce overestimated λ^t in Eq. (7.16).

Figure 7.8 shows the histories of the constraint-offset value and the objective function during the process of solving the SDO problem. With $\epsilon_o = 10^{-9}$ and $t_{max} = 50$, the optimal solution is found when the SDO problem terminates at its 16th cycle. The constraint value of the SDO

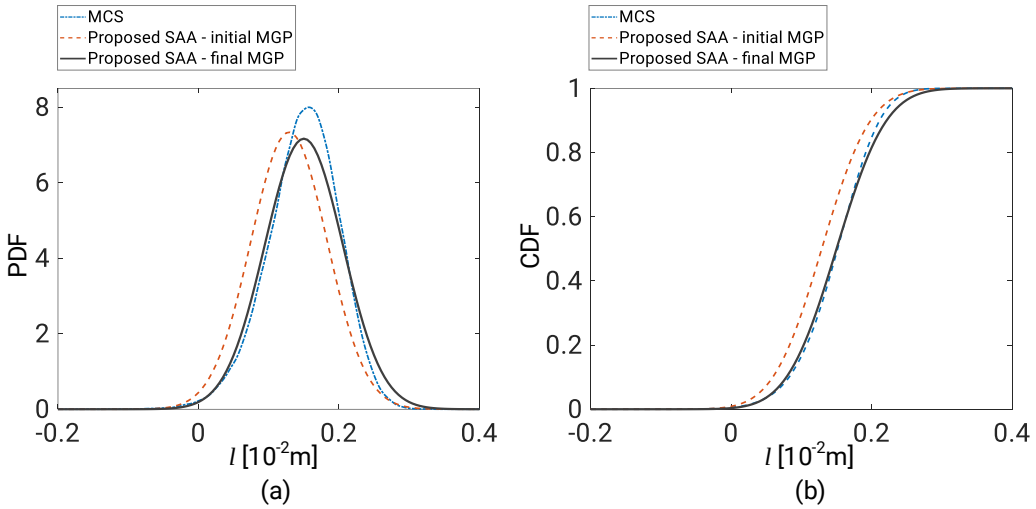

Figure 7.9: Example 1: PDF and CDF of l at the optimal design of the ten-bar truss. (a) PDF; (b) CDF.

problem at the optimal solution is less than a specified constraint tolerance of 10^{-6}, corresponding to $\mathbb{P}[l(\mathbf{s},\mathbf{r}) \leq 0] = 4.34 \times 10^{-3}$ (i.e., 70% of the threshold value ϵ). Table 7.4 shows that the total cross-sectional area by the SDO problem outperforms the results from previous RBDO designs by Zhao and Qiu [139], where FORM and a response surface method named RSM2 were adopted.

To further verify the feasibility of the obtained optimal design, we generate the associated PDF and CDF of l using MCS with 10^5 samples. Those by the combination of the proposed SAA and the initial MGP (at the first cycle of SDO) are also generated to confirm the improvement of accuracy in calculating the PDF and CDF for the optimal design through processing the SDO problem. Reliability analysis results are provided in Table 7.4 and illustrated in Fig. 7.9. It is found that failure probability of the optimal design by MCS is 5.64×10^{-3}, which is very close to the risk level 6.21×10^{-3}. The PDFs and CDFs generated by MCS, and the combination of the proposed SAA and the final MGP over the range $l \leq 0$ are almost identical. This is due to the gradual improvement of accuracy in calculation of the PDF and CDF for the optimal design after each cycle of the SDO problem is processed for obtaining a near-optimal design, as depicted in Fig. 7.9. The MGP contributes to such an improvement by well covering the input variable space so that the LSF l and its gradient at the mean vector of input variables are reasonably evaluated for calculating the first three cumulants of l, and in turn, updating the constraint-offset values.

7.5.2 Example 2: Four-story three-bay steel frame

This section formulates problem (7.1) for the four-story three-bay steel frame in Fig. 7.10, which is taken from Section 5.2.2 with modifications of the beam groups, material properties, and external loads. The frame consists of 28 structural members classified into seven groups of columns and beams. The design variables $\mathbf{s}$ are the sections of column and beam groups, chosen from a list of standard steel sections in Table 7.5, where ρ [kg/m] is the nominal mass of the section, and HBE and IPE represent the standard wide flange and universal beam sections, respectively. External loads and properties of the steel material, including Young's modulus E, yield stress $\sigma_{\mathrm{y},0}$, and ultimate tensile strength σ_{u}, are considered as the independent random parameters $\mathbf{r}$. The probabilistic properties

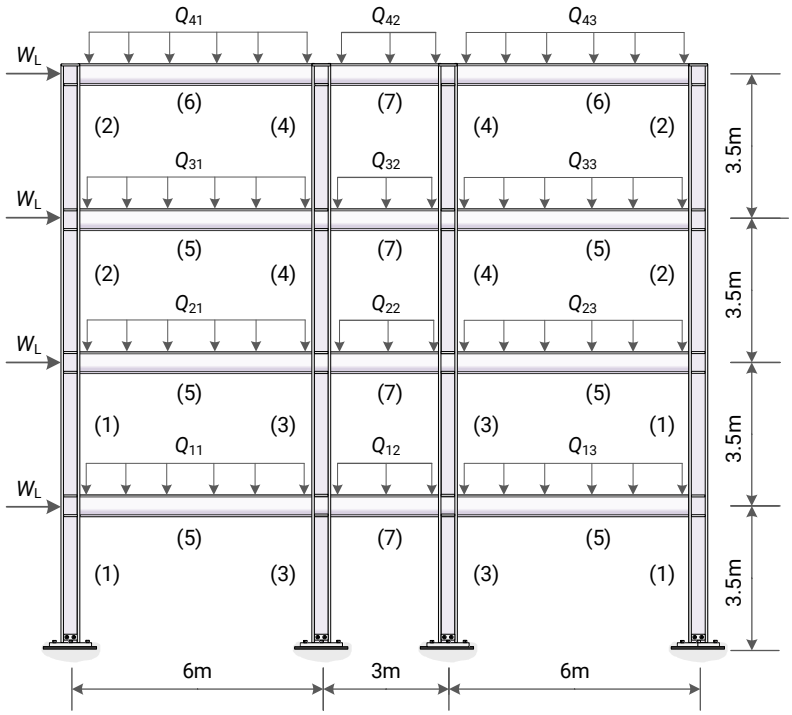

Figure 7.10: Example 2: Four-story three-bay steel frame, groups of structural members, and external loads.

Table 7.5: Example 2: List of column and beam sections for the four-story three-bay steel frame.

ID	Column	ρ [kg/m]	Beam	ρ [kg/m]
1	HEB 140	33.7	IPE 200	22.4
2	HEB 160	42.6	IPE 220	26.2
3	HEB 180	51.2	IPE 240	30.7
4	HEB 200	61.3	IPE 270	36.1
5	HEB 220	71.5	IPE 300	42.2
6	HEB 240	83.2	IPE 330	49.1
7	HEB 260	93.0	IPE 360	57.1
8	HEB 280	103.1	IPE 400	66.3
9	HEB 300	117.0	IPE 450	77.6
10	HEB 320	126.7	IPE 500	90.7
11	HEB 340	134.2	IPE 550	105.5
12	HEB 360	141.8	IPE 600	122.4

of these parameters are described in Table 7.6. The combinations of vertical loads applied to beam members, namely $Q_{11}, \ldots, Q_{43}$, depicted in Fig. 7.10, are detailed in Table 5.3.

The RBDO problem of the frame is to minimize its total mass $f(\mathbf{s})$ under probabilistic and deterministic constraints on the design strength, serviceability, and constructional requirements specified in ANSI/AISC 360-16 [24]. The total mass of the frame reads

$$f(\mathbf{s}) = \sum_{i=1}^{28} \rho_i(s_i) L_i, \tag{7.18}$$

Table 7.6: Example 2: Probabilistic properties of random parameters for the four-story three-bay steel frame.

Parameter	Description	Nominal	Mean/ Nominal	COV	Distribution
D_{L}	Dead load [kN/m]	20	1.00	0.10	Normal
S_1	Short term live load 1 [kN/m]	10	1.00	0.30	Lognormal
S_2	Short term live load 2 [kN/m]	5	1.00	0.30	Lognormal
L_1	Long term live load 1 [kN/m]	10	1.00	0.30	Lognormal
L_2	Long term live load 2 [kN/m]	5	1.00	0.30	Lognormal
S_{L}	Snow load [kN/m]	5	1.00	0.30	Lognormal
W_{L}	Wind load [kN]	8	1.00	0.30	Lognormal
E	Young's Modulus [GPa]	210	1.00	0.04	Normal
$\sigma_{\mathrm{y},0}$	Yield stress [MPa]	235	1.10	0.06	Normal
σ_{u}	Tensile strength [MPa]	360	1.07	0.08	Normal

where L_i and ρ_i denote the length of the i-th member and the nominal mass of the section for that member, respectively. s_i and the corresponding ρ_i are selected from the list of steel sections in Table 7.5.

A total of 12 uncertain LSFs are formulated for the strength and serviceability of the frame, which are summarized as follows: $l_1(\mathbf{s},\mathbf{r})$ to $l_7(\mathbf{s},\mathbf{r})$, respectively, correspond to the maximum strength performance of column or beam member among members of groups 1 to 7; $l_8(\mathbf{s},\mathbf{r})$ and $l_9(\mathbf{s},\mathbf{r})$ are associated with the maximum inter-story drift and total drift of the frame, respectively; $l_{10}(\mathbf{s},\mathbf{r})$, $l_{11}(\mathbf{s},\mathbf{r})$, and $l_{12}(\mathbf{s},\mathbf{r})$, respectively, correspond to the maximum long-term displacements (set as 1.5 times the corresponding elastic displacements) among beams in groups 5, 6, and 7. The detail of each uncertain LSFs for the strength and serviceability of a planar steel frame can be found in Chapters 5 and 6.

In addition to the probabilistic constraints, other 16 deterministic constraints, i.e., $h_j(\mathbf{s}) \leq 0$ $(j = 1,\ldots,16)$, are imposed at beam-column connections and column-column joints to guarantee that (1) the flange width of a beam connected to a column should be less than or equal to the flange width of the column and (2) the depth of the column section on the upper story at a column-column joint should not exceed the depth of the column section in the lower story. The detail of such deterministic constraints can also be found in Chapters 5 and 6.

Assume that the frame is a part of a residential building. According to design codes [23, 25], the risk levels regarding the ultimate and irreversible serviceability limit state failure modes within a 50 years reference period can be assigned as 1.35×10^{-3} and 6.68×10^{-2}, respectively. Thus, the RBDO problem of the frame can be stated as follows:

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & f(\mathbf{s}) \\ \text{subject to} \quad & \mathbb{P}[l_i(\mathbf{s},\mathbf{r}) \leq 0] \leq 1.35 \times 10^{-3} \quad (i = 1,\ldots,7), \\ & \mathbb{P}[l_i(\mathbf{s},\mathbf{r}) \leq 0] \leq 6.68 \times 10^{-2} \quad (i = 8,\ldots,12), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1,\ldots,16). \end{aligned} \tag{7.19}$$

We generate a training set of 2×10^4 samples to construct the initial MGPs for 12 LSFs in the probabilistic constraints of problem (7.19). The random parameters $\mathbf{r}$ in Table 7.6 and the prediction intervals in Table 7.7 for the dimensions of steel sections associated with the design

Table 7.7: Example 2: Empirical relations between b_f and d, and between t_f and t_w for beam and column sections.

Group	d [mm]	t_w [mm]	b_f [mm]	t_f [mm]
Column (HBE)	$[140, 360]$	$[6, 13]$	d	$2.0910t_\mathrm{w} - 3.3595$
Beam (IPE)	$[200, 600]$	$[5, 12]$	$-0.0004d^2 + 0.6384d - 6.3582$	$1.6522t_\mathrm{w} - 0.8304$

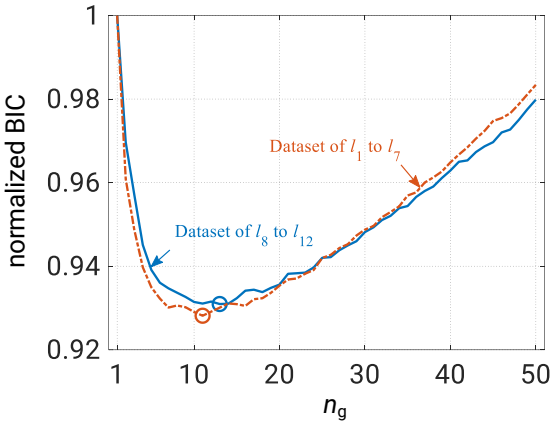

Figure 7.11: Example 2: Selecting GMMs for clustering the datasets of (l_1 to l_7) and (l_8 to l_{12}).

variables $\mathbf{s}$ together define the space of the input variables for generating the training set. Note that the lower and upper bounds of the depth d and the web thickness t_w of the column or beam section in Table 7.7 correspond to the smallest and largest values in the list of sections in Table 7.5. The flange width b_f and the flange thickness t_f of each section are empirically determined according to their relations to d and to t_w, respectively. These relations are established by employing polynomial regressions on the available database of HBE (for columns) and IPE (for beams) sections.

Figure 7.11 shows the clustering process of two training datasets associated with the strength and serviceability LSFs. The first dataset consists of the input variables and the LSFs from l_1 to l_7, while the second dataset consists of the input variables and the LSFs from l_8 to l_{12}. Accordingly, the MGP models for l_1 to l_7 are constructed from 11 clusters and those for l_8 to l_{12} from 13 clusters. The histories of normalized BIC for the two training datasets are similar because they share the same set of input variable samples.

Table 7.8: Example 2: Optimization results for the four-story three-bay steel frame.

Group ID	1st GA	2nd GA	3rd GA
Column (1)	HEB 220	HEB 220	HEB 220
Column (2)	HEB 180	HEB 200	HEB 180
Column (3)	HEB 220	HEB 220	HEB 240
Column (4)	HEB 220	HEB 200	HEB 200
Beam (5)	IPE 400	IPE 400	IPE 400
Beam (6)	IPE 360	IPE 330	IPE 360
Beam (7)	IPE 300	IPE 300	IPE 220
$f(\mathbf{s})$ [kg]	7298	7201	7172

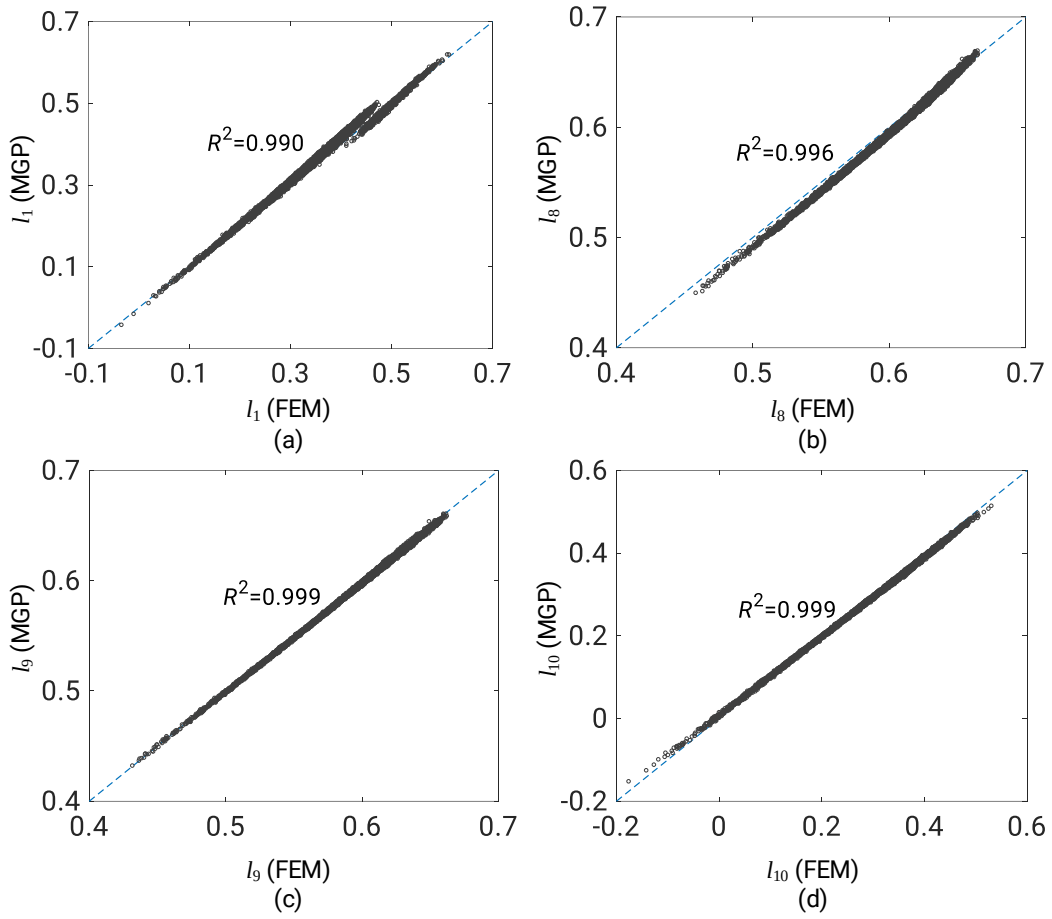

Figure 7.12: Example 2: Test set prediction performance of the MGPs for the steel frame. (a) l_1; (b) l_8; (c) l_9; (d) l_{10}.

Once the MGPs are obtained, we randomly generate a total of 2×10^4 samples of a test set to assess their prediction performance. We use the MGPs to predict the LSFs l_1, l_8, l_9, and l_{10} for the input values in the test set. The prediction results are compared with those by FE analyses. A good agreement is observed in Fig. 7.12 between the MGP predictions and FE analysis results. Hence, the obtained MGPs can be used to formulate the SDO problem associated with problem (7.19).

Since the design variables $\mathbf{s}$ are discrete, we use GA, available in the MATLAB R2018a Global Optimization Toolbox [106], for solving the SDO problem. We set $\epsilon_{\mathrm{o}} = 10^{-3}$ and $t_{\max} = 50$ as the stopping criteria for the optimization procedure. The SDO problem also terminates if its current solution is the same as the solution to its previous cycle. GA is performed three times to investigate the effect of its stochastic property on the resulting solutions.

Figure 7.13 shows the histories of the constraint-offset values corresponding to 12 probabilistic constraints and the norm of the constraint-offset vector during the process of solving the SDO problem with three GA attempts. Although there exists a difference in the histories of the SDO problem by the three GA implementations, the SDO problem can quickly terminate as the norm of the constraint-offset vector meets the threshold value.

Table 7.8 reports the optimal solutions obtained from three GA attempts. The difference in the total mass of the frame among the three solutions is insignificant. The minimum mass of the frame

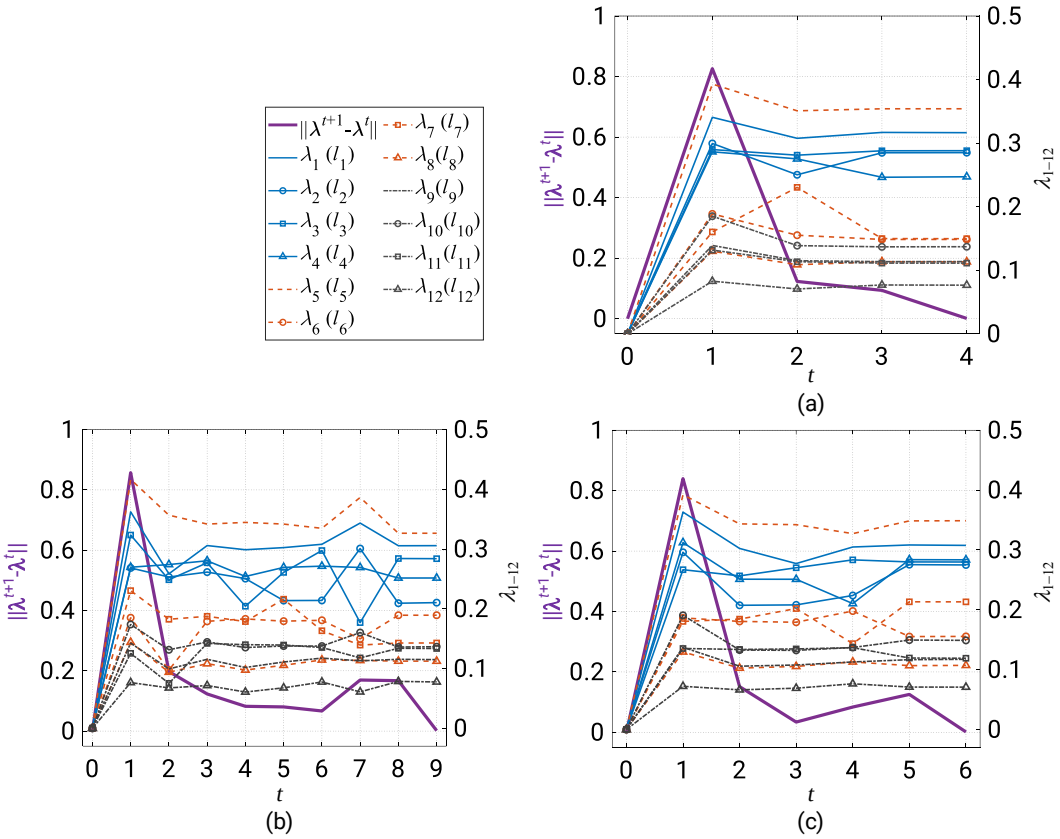

Figure 7.13: Example 2: Histories of the constraint-offset values with respect to the cycles of the SDO problem for the steel frame. (a) 1st GA; (b) 2nd GA; (c) 3rd GA.

is 7172 kg corresponding to the 3rd GA.

To further confirm the feasibility of the obtained optimal designs, we carry out a reliability analysis for each of them. Both the proposed SAA in Algorithm 7.2 and MCS with 10^5 samples are used for calculation of the failure probabilities regarding the LSFs associated with 12 probabilistic constraints of problem (7.19). Results of the reliability analyses are provided in Table 7.9, where $P_{\mathrm{f},1}$ to $P_{\mathrm{f},12}$ correspond to the LSFs l_1 to l_{12}, respectively. It is found that all failure probability values by the proposed SAA and MCS are less than the corresponding risk levels, indicating the feasibility of the obtained optimal designs. The boldface values in Table 7.9 show that the maximum failure probabilities associated with the ultimate and serviceability limit state failure modes refer to the failure probability of column (1) or (3) and the violation of long-term displacement of beam (5) or (6). Figure 7.14 compares the CDFs of the LSFs corresponding to these boldface values generated by the proposed SAA and MCS. A similarity is observed between the CDFs by the proposed SAA and those by MCS over the whole range of the LSFs, which demonstrates the efficiency of the proposed SAA in combination with the MGP to estimate the failure probability of the frame structure.

7.6 Conclusions

This chapter has presented an efficient optimization procedure for solving the RBDO problem of truss and frame structures under aleatory uncertainty in material properties and external loads. The

Table 7.9: Example 2: Probability of failures associated with each probabilistic constraint of the three-story two-bay steel frame.

P_{f}	Risk level	1st GA		2nd GA		3rd GA	
	ϵ	SAA	MCS	SAA	MCS	SAA	MCS
$P_{\mathrm{f},1}$	1.35×10^{-3}	0.29×10^{-3}	0.34×10^{-3}	0.19×10^{-3}	0.19×10^{-3}	$\mathbf{0.37 \times 10^{-3}}$	$\mathbf{0.34 \times 10^{-3}}$
$P_{\mathrm{f},2}$	1.35×10^{-3}	0.09×10^{-3}	0.06×10^{-3}	0.02×10^{-3}	0	0.34×10^{-3}	0.20×10^{-3}
$P_{\mathrm{f},3}$	1.35×10^{-3}	$\mathbf{1.26 \times 10^{-3}}$	$\mathbf{1.32 \times 10^{-3}}$	$\mathbf{0.89 \times 10^{-3}}$	$\mathbf{0.96 \times 10^{-3}}$	0.05×10^{-3}	0.01×10^{-3}
$P_{\mathrm{f},4}$	1.35×10^{-3}	0	0	0	0	0.06×10^{-3}	0.10×10^{-3}
$P_{\mathrm{f},5}$	1.35×10^{-3}	0.06×10^{-3}	0.02×10^{-3}	0	0.06×10^{-3}	0.04×10^{-3}	0.03×10^{-3}
$P_{\mathrm{f},6}$	1.35×10^{-3}	0	0	0	0	0	0
$P_{\mathrm{f},7}$	1.35×10^{-3}	0	0	0	0	0.10×10^{-3}	0.19×10^{-3}
$P_{\mathrm{f},8}$	6.68×10^{-2}	0	0	0	0	0	0
$P_{\mathrm{f},9}$	6.68×10^{-2}	0	0	0	0	0	0
$P_{\mathrm{f},10}$	6.68×10^{-2}	$\mathbf{0.43 \times 10^{-2}}$	$\mathbf{0.53 \times 10^{-2}}$	0.28×10^{-2}	0.20×10^{-2}	$\mathbf{1.30 \times 10^{-2}}$	$\mathbf{1.55 \times 10^{-2}}$
$P_{\mathrm{f},11}$	6.68×10^{-2}	0.02×10^{-2}	0.06×10^{-2}	$\mathbf{3.33 \times 10^{-2}}$	$\mathbf{3.21 \times 10^{-2}}$	0.32×10^{-2}	0.34×10^{-2}
$P_{\mathrm{f},12}$	6.68×10^{-2}	0	0	0	0	0	0

Boldface values are maximum failure probabilities corresponding to the ultimate and irreversible serviceability limit state failure modes.

main findings are summarized as follows:

(1) To reduce the number of structural analysis calls during the optimization process and overcome the prominent weakness of the GP in applying to relatively small training datasets, the MGP models are constructed for prediction of structural responses. A large training dataset of input variables and corresponding LSF values is generated and divided into independent subsets using the GMM clustering method. The GP corresponding to each subset is developed to produce a set of independent GP models. These GP models then together define the MGP as their weighted average function. It is demonstrated in the examples that the MGPs can significantly reduce the computation time required for the training process and provide the predictions with high accuracy.

(2) To handle the probabilistic constraints of the RBDO problem effectively, a novel SAA is proposed based on the first three cumulants of the uncertain LSFs so that the unknown coefficients of their cumulant generating functions are directly derived in the physical space of the LSFs, and the condition for the unique root of the saddlepoint equation and the requirement for the existence of the solution can be addressed. The efficiency of the proposed SAA in calculation of failure probabilities has been verified in comparison with existing SAA formulations and MCS. Thus, the proposed SAA can be a useful technique for a quick implementation of structural reliability analyses or a the determination of quantile of a random LSF. In addition, the combination of the proposed SAA and the MGP has demonstrated itself as a viable choice for solving the RBDO problem.

(3) To improve the optimization strategy, the SDO problem is developed to replace the original RBDO problem. In the SDO problem, the MGP models serve as surrogates for the LSFs in the probabilistic constraints of the RBDO problem. The MGP models are then locally refined after each cycle of solving the SDO problem. The SDO problem is strategically solved, with support from the proposed SAA in calculation of the inverse failure probabilities, for exploring the region that is deemed to contain the optimal solution and improving the accuracy of the

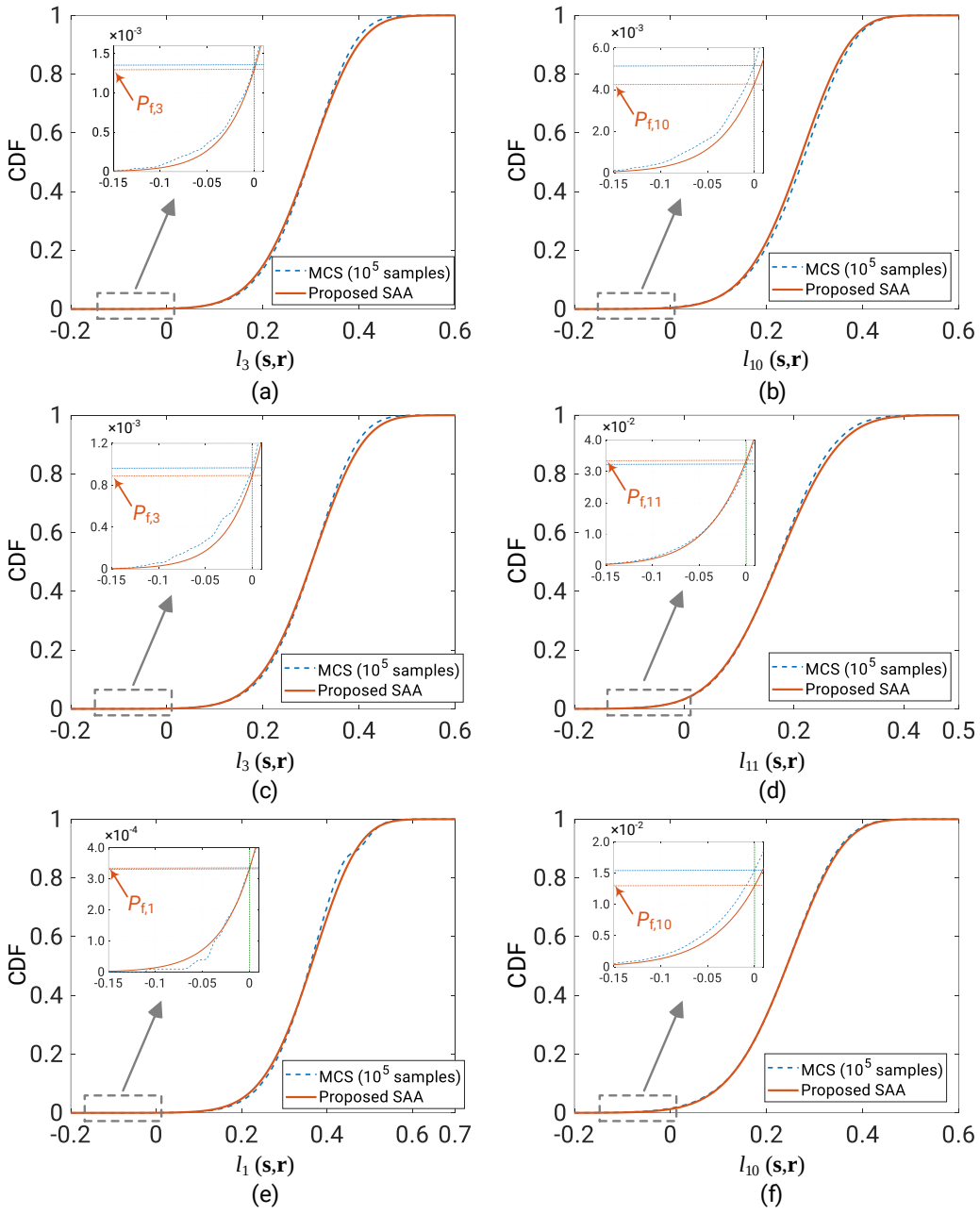

Figure 7.14: Example 2: CDFs of some LSFs at the optimal design of the steel frame. (a) 1st GA, performance of column (3)–l_3; (b) 1st GA, displacement of beam (5)–l_{10}; (c) 2nd GA, performance of column (3)–l_3; (d) 2nd GA, displacement of beam (6)–l_{11};(e) 3rd GA, performance of column (1)–l_1; (f) 3rd GA, displacement of beam (5)–l_{10}.

MGP models in that region. Optimization results for a ten-bar truss and a four-story three-bay steel frame show the effectiveness of the proposed SDO problem in quickly providing a reliable solution to the original RBDO problem.

(4) Since the proposed MGP aims to scale up the application of GP to large training sets for increasing the accuracy of GP predictions, Algorithm 7.1 should be applied to a problem with a large training set that has, for example, $10^4 - 5 \times 10^4$ samples. For problems associated with

a small or medium number of training samples, say, less than 1000, the global GP may be an alternative. Yet the update of the GP, especially for a medium-size training set, becomes computationally expensive since the size of the training set increases after each SDO cycle. Choosing either the MGP or GP for the training purpose would depend on the nature of available dataset.

Chapter 8

Sequential sampling approach to multi-objective reliability-based design optimization of seismic-resistant steel frames

8.1 Introduction

This chapter, as an extension of Chapter 7, formulates a multi-objective RBDO for a moment-resisting steel frame subjected to earthquake excitation. The optimization problem is formulated with two objective functions, namely the total mass and the plastic energy dissipated by beam members of the frame, and subject to uncorrelated probabilistic constraints on dynamic responses under the effect of correlated random parameters of floor masses, external loads, and material properties. The dynamic responses of the frame are evaluated using nonlinear response history analysis (NRHA), hence enables an accurate simulation of the plastic mechanism of the frame. NRHA, however, increases the complexity of the formulated RBDO problem considerably as it demands a substantial computational cost for estimating the uncertain energy dissipation of beams and probabilistic constraints on dynamic structural responses. Therefore, it is desirable to devise a new optimization strategy to keep the advantages of NRHA while reducing the computational cost due to incorporating uncertainty propagation into optimization.

A sequential batch sampling method is proposed in this chapter for solving the formulated multi-objective RBDO problem. Main features of the proposed method are as follows:

1. The dynamic responses for a small number of designs are evaluated using NRHA. The corresponding uncertain LSFs are approximated by GP models for carrying out the optimization process. These GP models facilitate the use of MCS that evaluates the expected value of the uncertain dissipation energy of beam members as well as the probabilistic constraints of the RBDO problem.
2. Approximate solutions to the RBDO problem sorted from the existing candidate solutions

are strategically updated after each optimization iteration. The updated candidate solutions consist of the existing candidates and new candidates generated by performing discrete random local and global searches.

3. The GP models for the uncertain LSFs are refined after each optimization iteration by specifying a batch of new sampling points that lie on the Pareto front of a bi-objective deterministic maximization problem formulated for addressing the improvement in the current solutions and the feasibility of the new sampling points simultaneously. This refinement scheme differs from those of other sequential optimization methods such as BO and SDO, respectively, in Chapters 6 and 7, where only one new sampling point is specified after each optimization iteration. As will be demonstrated in a test problem, the new sampling points tend to distribute in the neighborhood of exact solutions, leading to the robustness of the refinement scheme as well as a quick termination of the optimization process.

8.2 Energy-based RBDO of moment-resisting steel frame

Consider a moment-resisting steel frame subjected to vertical and earthquake loads. The frame is designed according to the capacity-design principle by which the structural components are classified into dissipative and non-dissipative members. The dissipative members, through their inelastic deformations, are primarily responsible for dissipating seismic energy. The failure of these members must occur prior to that of the non-dissipative members to prevent brittle collapse of the whole structure. As beam members of the frame serve as the dissipative members [141], they are expected to experience large inelastic deformations during the earthquake.

This section first describes the evaluation of energy dissipated by the beams using their internal force-deformation histories obtained from NRHA. It then formulates the bi-objective RBDO problem for the frame based on such dissipation energy.

8.2.1 Dissipation energy of beam members

The energy balance equation of the multi-degree-of-freedom system for the frame during the earthquake excitation reads [142]

$$E_{\mathrm{k}}^t + E_{\mathrm{d}}^t + E_{\mathrm{p}}^t = E_{\mathrm{i}}^t, \tag{8.1}$$

where E_{k}^t, E_{d}^t, E_{p}^t, and E_{i}^t represent the kinetic, damping, dissipation, and input energies at a time instant t, respectively.

The dissipation energy E_{p}^t including recoverable-elastic strain and irrecoverable-plastic energies of all structural members can be evaluated based on the force-deformation histories of the members. Note that E_{p}^t excludes viscous dissipation. In case the elastoplastic behavior of each member is simulated using an elastic beam-column element in the middle and two semi-rigid rotational springs with hysteretic properties at the member ends [143], E_{p}^t can be estimated from moment-rotation histories of these springs.

It is reasonable that the cyclic behavior of the rotational springs follows a bilinear hysteretic response incorporated in the modified Ibarra Krawinkler (mIK) deterioration model depicted in Figure 8.1 [144]. This model specifies strength bounds for the spring using a monotonic curve and a cyclic damage rule that captures the deterioration of the bounds as the cyclic excursion accumulates.

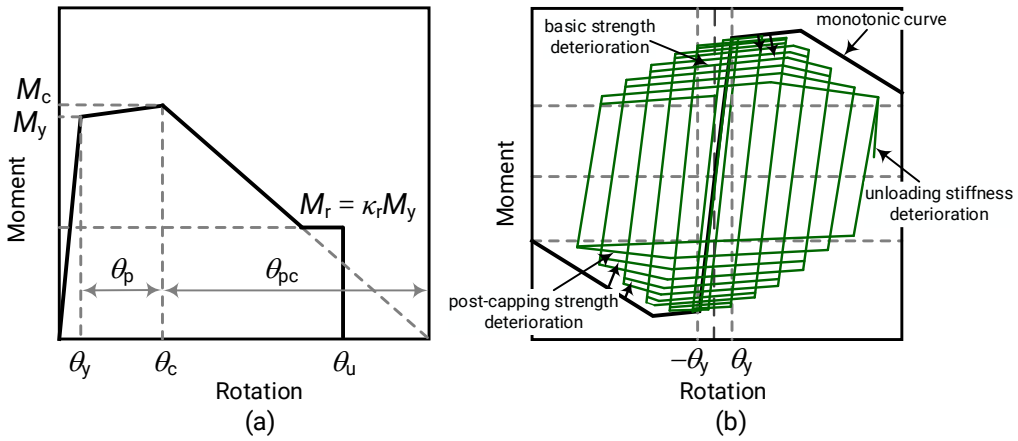

Figure 8.1: Modified IK deterioration model for semi-rigid rotational springs. (a) monotonic curve; (b) basic modes for cyclic deterioration (adapted from Ref. [144]).

The monotonic curve is characterized by three strength parameters and four deformation parameters. Three strength parameters include (1) the effective yield moment M_{y}, equal to 1.1 times the fully plastic moment of the section obtained from plastic section modulus and the material yield stress $\sigma_{\mathrm{y},0}$ [144]; (2) the capping moment strength M_{c}, described by a post-yield strength ratio $M_{\mathrm{c}}/M_{\mathrm{y}}$; and (3) the residual moment M_{r}, defined by a residual strength ratio κ_{r}, i.e., $M_{\mathrm{r}} = \kappa_{\mathrm{r}} M_{\mathrm{y}}$. Four deformation parameters include (1) the yield rotation θ_{y}; (2) the pre-capping plastic rotation for monotonic loading θ_{p}; (3) the post-capping plastic rotation θ_{pc}; and (4) the ultimate rotation capacity θ_{u}, commonly assigned as 0.06 rad [144, 145].

Let $E_{\mathrm{r}} = \lambda_{\mathrm{r}} M_{\mathrm{y}}$ denote a reference hysteretic energy dissipation capacity of each rotational spring, where λ_{r} is the so-called reference cumulative rotation capacity. The rate of cyclic deterioration of the spring in the current excursion can be expressed as a function of E_{r}, the hysteretic energy dissipated in the current excursion, and the total energy dissipated in past excursions. Detailed expression for such a cyclic deterioration rate can be found in the seminal work by Lignos and Krawinkler [144].

Since the energy dissipated by the mentioned beam-column element is primarily due to flexure, the dissipation energy of the i-th beam and that of the i-th column at time t, respectively, denoted as $E^t_{\mathrm{pb},i}$ and $E^t_{\mathrm{pc},i}$, can be evaluated using the following discrete expressions:

$$E^t_{\mathrm{pb},i} = E^{t-\Delta t}_{\mathrm{pb},i} + \frac{M^t_{\mathrm{b1},i} + M^{t-\Delta t}_{\mathrm{b1},i}}{2}\Delta\theta_{\mathrm{b1},i} + \frac{M^t_{\mathrm{b2},i} + M^{t-\Delta t}_{\mathrm{b2},i}}{2}\Delta\theta_{\mathrm{b2},i}, \tag{8.2a}$$

$$E^t_{\mathrm{pc},i} = E^{t-\Delta t}_{\mathrm{pc},i} + \frac{M^t_{\mathrm{c1},i} + M^{t-\Delta t}_{\mathrm{c1},i}}{2}\Delta\theta_{\mathrm{c1},i} + \frac{M^t_{\mathrm{c2},i} + M^{t-\Delta t}_{\mathrm{c2},i}}{2}\Delta\theta_{\mathrm{c2},i}, \tag{8.2b}$$

where $t-\Delta t$ and t represent two consecutive time instants; $M_{\mathrm{b(c)1}}$ and $M_{\mathrm{b(c)2}}$, respectively, stand for internal moments of the first (1) and second (2) semi-rigid rotational springs of the beam (column) element; and $\Delta\theta_{\mathrm{b(c)1(2)},i} = \theta^t_{\mathrm{b(c)1(2)},i} - \theta^{t-\Delta t}_{\mathrm{b(c)1(2)},i}$ with $\theta_{\mathrm{b(c)1}}$ and $\theta_{\mathrm{b(c)2}}$ represent rotation angles of the first (1) and second (2) rotational springs of the beam (column) element, respectively.

Let E_{pb} and E_{pc} represent the energies dissipated by n_{b} beams and n_{c} columns of the frame at

the end of the earthquake, respectively. The dissipation energy ratio of the beams reads

$$\beta = \frac{E_{\mathrm{pb}}}{E_{\mathrm{pb}} + E_{\mathrm{pc}}} = \frac{\sum_{i=1}^{n_{\mathrm{b}}} E_{\mathrm{pb},i}}{\sum_{i=1}^{n_{\mathrm{b}}} E_{\mathrm{pb},i} + \sum_{j=1}^{n_{\mathrm{c}}} E_{\mathrm{pc},j}}. \tag{8.3}$$

8.2.2 Formulation of the RBDO problem

Let $\mathbf{s} = [s_1, ..., s_{d_1}]^T \in \mathbb{N}^{d_1}$ denote a d_1-dimensional vector of discrete design variables of the frame and $\mathbf{r} = [r_1, ..., r_{d_2}]^T \in \mathbb{R}^{d_2}$ denote a d_2-dimensional vector of continuous random parameters of floor masses, vertical loads, and material properties. Each element of $\mathbf{s}$ corresponds to a section number in a list of American wide-flange steel sections, hence $s_k \in \mathcal{S}_k \, (k = 1, \ldots, d_1)$, while probabilistic characteristics of $\mathbf{r}$ are described by the marginal PDFs, or equivalently, CDFs of its elements. The elements of $\mathbf{r}$ correlate with each other according to a given correlation matrix $\mathbf{C}_r$; see Section 2.1.4.

The RBDO problem formulated for the frame is to optimize the steel section of its members considering its total mass and the dissipation energy of beam members. The first objective function associated with the normalized total mass of the frame reads

$$f_1(\mathbf{s}) = \frac{\sum_{i=1}^{n_{\mathrm{e}}} \rho_i(s_i) L_i}{m_{\max}}, \tag{8.4}$$

where ρ_i, L_i, and n_{e} denote the nominal mass [kg/m] of the steel section for the i-th member, the length of that member, and the number of members, respectively; and $m_{\max}$ is the maximum value of the total mass. Thus, $f_1(\mathbf{s})$ ranges from 0 to 1.

The second objective function corresponds to the expected dissipation energy ratio of all beam members as given in Eq. (8.3). For a minimization problem, this objective function is formulated with the minus sign as

$$f_2(\mathbf{s}) = -\mathbb{E}[\beta(\mathbf{s}, \mathbf{r})]. \tag{8.5}$$

Since $0 < \beta(\mathbf{s}, \mathbf{r}) < 1$, $f_2(\mathbf{s})$ ranges from -1 to 0. When using a total of n_{m} ground motions for the design, which are assumed to have the same duration, $\beta(\mathbf{s}, \mathbf{r})$ can be defined as the mean value of the dissipation energy ratios for these motions.

To ensure the frame remains intact during the earthquake, the maximum inter-story drift and maximum plastic rotation of the member ends are limited by some thresholds. Conventionally, plastic deformations are not allowed for the columns. However, we allow the columns to have minor plastifications because minimizing their dissipation energy is consistent with maximizing the dissipation energy of the beams. Let $g_1(\mathbf{s}, \mathbf{r})$, $g_2(\mathbf{s}, \mathbf{r})$, and $g_3(\mathbf{s}, \mathbf{r})$ represent uncertain LSFs corresponding to the absolute maximum of the inter-story drift ratios, absolute maximum of the beam-end plastic rotations, and absolute maximum of the column-end plastic rotations during the earthquake, respectively; and δ_{a}, φ_{a}, and ω_{a} represent the respective allowable thresholds. The LSFs corresponding to

the inter-story drift and member-end plastic rotations read

$$g_1(\mathbf{s},\mathbf{r}) = \max\left(|\delta_1|,\ldots,|\delta_{n_\mathrm{s}}|\right)/\delta_\mathrm{a} - 1, \tag{8.6a}$$

$$g_2(\mathbf{s},\mathbf{r}) = \max\left(|\varphi_1|,\ldots,|\varphi_{2n_\mathrm{b}}|\right)/\varphi_\mathrm{a} - 1, \tag{8.6b}$$

$$g_3(\mathbf{s},\mathbf{r}) = \max\left(|\omega_1|,\ldots,|\omega_{2n_\mathrm{c}}|\right)/\omega_\mathrm{a} - 1, \tag{8.6c}$$

where δ_i is the inter-story drift ratio of the i-th story, defined as the ratio of the story drift to the corresponding story height; and n_s represents the number of stories. The inter-story drift ratio, beam-end rotation, and column-end rotation at a particular time instant are evaluated as the mean value of the corresponding responses for n_m ground motions.

To ensure a column or a beam member can sustain the plastic moment without exhibiting local buckling, the width-thickness ratio of all plates composing the steel section of that member should be constrained. In accordance with Chapter B of ANSI/AISC 360-16 [24], the following two constraints are applied, respectively, to the web and flange plates of the section:

$$\tilde{g}_1(\mathbf{s}) = \frac{d-2t_\mathrm{f}}{t_\mathrm{w}} - 3.76\sqrt{\frac{\mathbb{E}[E]}{\mathbb{E}[\sigma_{\mathrm{y},0}]}} \le 0, \tag{8.7a}$$

$$\tilde{g}_2(\mathbf{s}) = \frac{b_\mathrm{f}}{2t_\mathrm{f}} - 0.38\sqrt{\frac{\mathbb{E}[E]}{\mathbb{E}[\sigma_{\mathrm{y},0}]}} \le 0, \tag{8.7b}$$

where d, b_f, t_f, and t_w are the height, flange width, flange thickness, and web thickness, respectively; and E and $\sigma_{\mathrm{y},0}$ are Young's modulus and the yield stress of the steel material, respectively.

Moreover, a total of J deterministic constructional constraints $h_j(\mathbf{s}) \le 0$ $(j = 1,\ldots,J)$ are imposed at beam-column connections and column-column joints. They ensure (1) the flange width of a beam connected to a column is less than or equal to the flange width of the column and (2) the depth of the column section in the upper story should not exceed that in the lower story.

Using the above objective and constraint functions, the bi-objective RBDO problem formulated for the frame to optimize its total mass and the dissipation energy of its beams can be stated as follows:

$$\begin{aligned}
\underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s})] \\
\text{subject to} \quad & \mathbb{P}\left[g_l(\mathbf{s},\mathbf{r}) \le 0\right] \ge 1-\epsilon_l \quad (l=1,2,3), \\
& \tilde{g}_i(\mathbf{s}) \le 0 \quad (i=1,\ldots,2d_1), \\
& h_j(\mathbf{s}) \le 0 \quad (j=1,\ldots,J), \\
& s_k \in \mathcal{S}_k \quad (k=1,\ldots,d_1),
\end{aligned} \tag{8.8}$$

where the prescribed risk levels ϵ_l $(l = 1,2,3)$ are small positive values. As a connection to current design codes, these risk levels can be derived from corresponding target reliability values specified in each design code.

Moving the left-side terms of the probabilistic constraints in problem (8.8) to the right side and let

$$g_l(\mathbf{s}) = 1 - \epsilon_l - \mathbb{P}\left[g_l(\mathbf{s},\mathbf{r}) \le 0\right] \quad (l=1,2,3). \tag{8.9}$$

Problem (8.8) can be rewritten as

$$\begin{aligned} \underset{\mathbf{s}}{\text{minimize}} \quad & [f_1(\mathbf{s}), f_2(\mathbf{s})] \\ \text{subject to} \quad & g_l(\mathbf{s}) \leq 0 \quad (l = 1, 2, 3), \\ & \tilde{g}_i(\mathbf{s}) \leq 0 \quad (i = 1, \ldots, 2d_1), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & s_k \in \mathcal{S}_k \quad (k = 1, \ldots, d_1). \end{aligned} \tag{8.10}$$

Exact Pareto-optimal solutions to problem (8.10) are difficult to obtain because $f_2(\mathbf{s})$ and $g_l(\mathbf{s})$ $(l = 1, 2, 3)$ are generally nonlinear, nonconvex, and implicit within the NRHA program. A new optimization strategy is proposed for finding approximate Pareto-optimal solutions the problem.

8.3 Design response spectrum and scaling recorded ground motions

According to ASCE 7-16 [23], the design spectral response acceleration S_a can be evaluated using the risk-targeted maximum considered earthquake ($\mathrm{MCE_R}$). The $\mathrm{MCE_R}$ is constructed based on the uniform-hazard (2% in 50-year) ground motions, which underline the ASCE 7-16 $\mathrm{MCE_R}$ ground motion maps. Let $S_{(\mathrm{S})}$ and $S_{(1)}$, respectively, denote the mapped $\mathrm{MCE_R}$, 5%-damped, spectral response acceleration parameters at short periods and at a period of 1 s. Design values of $S_{(\mathrm{S})}$ and $S_{(1)}$, denoted as S_DS and S_D1, are determined as follows [23]:

$$S_\mathrm{DS} = \frac{2}{3} F_\mathrm{a} S_{(\mathrm{S})}, \quad S_\mathrm{D1} = \frac{2}{3} F_\mathrm{v} S_{(1)}, \tag{8.11}$$

where F_a and F_v are two coefficients considering the site soil properties.

Once S_DS and S_D1 have been obtained, S_a can be evaluated as [23]

$$S_\mathrm{a} = \begin{cases} S_\mathrm{DS}\,(0.4 + 0.6T/T_0) & \text{if } \ T \leq T_0, \\ S_\mathrm{DS} & \text{if } \ T_0 < T \leq T_\mathrm{S}, \\ S_\mathrm{D1}/T & \text{if } \ T_\mathrm{S} < T \leq T_\mathrm{L}, \\ S_\mathrm{D1} T_\mathrm{L}/T^2 & \text{if } \ T > T_\mathrm{L}, \end{cases} \tag{8.12}$$

where T denotes the fundamental natural period of the structure, $T_0 = 0.2 S_\mathrm{D1}/S_\mathrm{DS}$, $T_\mathrm{S} = S_\mathrm{D1}/S_\mathrm{DS}$, and T_L is the long-period transition period [23].

When different recorded earthquake ground motions are used for the design, they should be scaled such that the mean of 5%-damped response spectra for the scaled motions is not less than the design $\mathrm{MCE_R}$ spectrum over the period range of $0.2T - 1.5T$ [23]. For designing the frame in this study, a total of six recorded earthquake ground motions are selected from the Pacific Earthquake Engineering Research Center (PEERC) database [146] as listed in Table 8.1. The acceleration spectra of the selected ground motions are scaled to simulate the target $\mathrm{MCE_R}$ acceleration spectrum using a scaling procedure by Reyes and Kalkan [147]. The design acceleration time histories of the selected motions are then evaluated by multiplying the recorded acceleration time histories by the

Table 8.1: Six ground motions selected from PEERC database [146].

ID	Event	Year	Station	Magnitude	Fault normal component		
					PGA [g]	PGV [cm/s]	PGD [cm]
1	Imperial Valley-06	1979	Delta	6.53	0.6	63.6	30.8
2	Loma Prieta	1989	Gilroy Array $\neq 4$	6.93	1.0	100.8	32.5
3	Northridge-01	1994	Canoga Park	6.69	0.8	132.2	56.7
4	Kobe Japan	1995	Kakogawa	6.90	0.5	39.2	12.8
5	Kobe Japan	1995	Shin-Osaka	6.90	0.6	89.9	26.3
6	Chi-Chi Taiwan	1991	CHY036	7.62	0.9	101.0	47.8

PGA = peak ground acceleration, PGV = peak ground velocity, PGD = peak ground displacement

corresponding scale factors.

8.4 Proposed sequential batch sampling approach

8.4.1 Generating correlated random parameters

As previously mentioned, the vector of random parameters $\mathbf{r} = [r_1, \ldots, r_{d_2}]^T$ is described by the marginal PDFs or CDFs of its elements, namely $\pi(r_i)$ or $F(r_i)$, respectively. The correlations between these elements are characterized by the following correlation matrix:

$$\mathbf{C}_r = \begin{bmatrix} 1 & \rho(r_1, r_2) & \cdots & \rho(r_1, r_{d_2}) \\ \rho(r_2, r_1) & 1 & \cdots & \rho(r_2, r_{d_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(r_{d_2}, r_1) & \rho_(r_{d_2}, r_2) & \cdots & 1 \end{bmatrix}, \tag{8.13}$$

where $\rho(r_i, r_j)$ $(i, j = 1, \ldots, d_2)$ is the correlation coefficient between r_i and r_j so that $\rho(r_i, r_i) = 1$ and $\rho(r_i, r_j) = \rho(r_j, r_i)$. The goal is to generate a finite number of $\mathbf{r}$ samples for processing the optimization using $\pi(r_i)$, $F(r_i)$, and $\mathbf{C}_r$.

Let $\mathbf{c} = [c_1, \ldots, c_{d_2}]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_r)$ denote a d_2-variate Gaussian vector in the standardized space. The samples of $\mathbf{c}$ can be generated using the built-in MATLAB function *normrnd* [148]. It is also trivial to evaluate the CDF for each sample of the element c_i $(i = 1, \ldots, d_2)$, denoted as $F(c_i)$, because $c_i \sim \mathcal{N}(0, 1)$.

Suppose there exists an iso-probabilistic mapping that transforms the physical space of $\mathbf{r}$ into the standardized space of $\mathbf{c}$. This mapping preserves the CDFs at two corresponding points r_i and c_i, i.e., $F(r_i) = F(c_i)$. Thus, the random samples of r_i can be generated by

$$r_i = F^{-1}\left(F(c_i)\right), \tag{8.14}$$

where $F^{-1}(\cdot)$ denotes the inverse CDF function with respect to r_i. It is worth noting that Eq. (8.14) is applicable to any distributions of $\mathbf{r}$ and can also be used for generating samples of uncorrelated random parameters.

To further demonstrate its performance in generating samples of random parameters, Eq. (8.14) is used to draw a total of 5×10^4 samples of two Weibull random parameters $r_1 \sim W(r; a = 4, b = 2)$

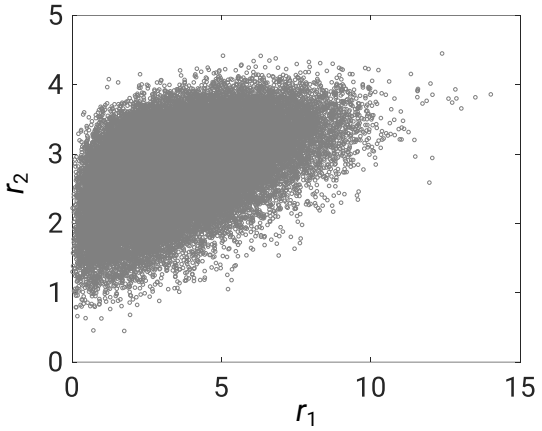

Figure 8.2: Samples of r_1 and r_2 generated from Eq. (8.14) for Weibull distribution.

Table 8.2: Comparison of sampled mean, variance, and correlation matrix estimates of r_1 and r_2 with exact ones.

	Estimated using Eq. (8.14)	Exact
$\mathbb{E}[r_1]$	3.5449	3.5449
$\mathbb{V}[r_1]$	3.4101	3.4336
$\mathbb{E}[r_2]$	2.7822	2.7832
$\mathbb{V}[r_2]$	0.2908	0.2908
$\mathbf{C}_r$	$\begin{bmatrix} 1 & 0.4924 \\ 0.4924 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

and $r_2 \sim W(r; a = 3, b = 6)$ with $\mathbf{C}_r = [1, 0.5; 0.5, 1]$.

Figure 8.2 shows the samples of r_1 and r_2 generated using Eq. (8.14). Table 8.2 shows a good performance of Eq. (8.14) as the sample mean, sample variance, and sample correlation matrix of r_1 and r_2 well agree with the exact ones.

8.4.2 Approximate uncertain objective and probabilistic constraint functions

As the first step for solving problem (8.10), $\beta(\mathbf{s}, \mathbf{r})$ and $g_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$ are approximated by the corresponding GP models trained based upon a finite number of sampling points. To do so, a training dataset $\mathcal{D} = \{\mathbf{x}^i, y^i\}_{i=1}^N$ is generated, where $\mathbf{x}^i = [{\mathbf{s}^i}^T, {\mathbf{r}^i}^T]^T \in \mathbb{R}^d$ $(d = d_1 + d_2)$ are d-dimensional vectors of uncertain input variables, and $y^i \in \mathbb{R}$ are the corresponding outputs (i.e., β^i or g_l^i). The number of initial training samples N depends on the number of input variables d. Samples of $\mathbf{s}$ and $\mathbf{r}$ are generated using Latin-hypercube sampling [74] and Eq. (8.14), respectively. Integer samples of $\mathbf{s}$ are obtained by rounding the corresponding real samples by Latin-hypercube sampling to the nearest integers. Each sample $\mathbf{x}^i$ then serves as an input to NRHA for evaluation of the corresponding dynamic responses. The sample that provides non-positive values of the LSFs $g_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$ is retained in $\mathcal{D}$ as a feasible training sample, which is used for constructing the GP models.

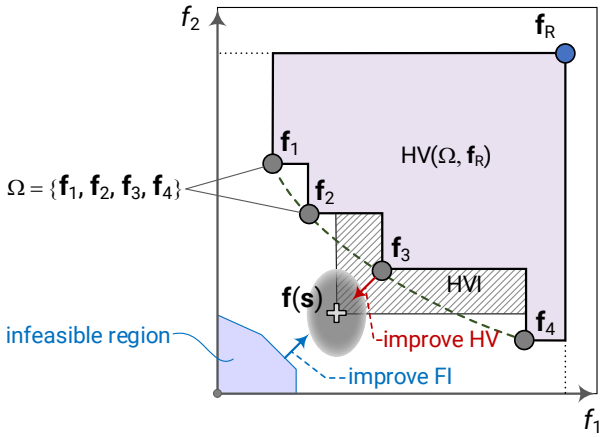

Figure 8.3: Illustration of the trade-off between maximizing HVI and maximizing FI.

Once $\mathcal{D}$ has been created, the GP models for $\beta(\mathbf{s}, \mathbf{r})$ and $g_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$ can be constructed; see Section 2.3, and the resulting GP models at each realization of the input variable vector are, respectively, the Gaussians $\hat{\beta}(\mathbf{s}, \mathbf{r})$ and $\hat{g}_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$. The mean and variance characterizing each of these GP models, for example, $\mu_{\hat{g}_l}(\mathbf{s}, \mathbf{r})$ and $\sigma^2_{\hat{g}_l}(\mathbf{s}, \mathbf{r})$ of $\hat{g}_l(\mathbf{s}, \mathbf{r})$, follow Eqs. (2.31a) and (2.31b), respectively.

8.4.3 Sorting approximate Pareto-optimal solutions

Let Ω_a denote the existing candidate solutions that are the samples of $\mathcal{D}$ at the beginning of the optimization process and enriched after each optimization iteration, which is discussed later. Among the members of Ω_a, a non-dominated sorting approach [91] seeks a set of approximate Pareto-optimal solutions to problem (8.10) once $\hat{\beta}(\mathbf{s}, \mathbf{r})$ and $\hat{g}_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$ have been obtained. In this way, $f_2(\mathbf{s})$ and $g_l(\mathbf{s})$ for each member of Ω_a can be evaluated through MCS using the mean functions of the corresponding GP models, such that

$$f_2(\mathbf{s}) = -\mathbb{E}\left[\beta(\mathbf{s}, \mathbf{r})\right] \approx -\frac{1}{n_\mathrm{r}} \sum_{i=1}^{n_\mathrm{r}} \mu_{\hat{\beta}}(\mathbf{s}, \mathbf{r}_i), \tag{8.15a}$$

$$g_l(\mathbf{s}) = 1 - \epsilon_l - \mathbb{P}\left[g_l(\mathbf{s}, \mathbf{r}) \leq 0\right] \approx 1 - \epsilon_l - \frac{1}{n_\mathrm{r}} \sum_{i=1}^{n_\mathrm{r}} \mathbb{I}[\mu_{\hat{g}_l}(\mathbf{s}, \mathbf{r}_i) \leq 0] \quad (l = 1, 2, 3), \tag{8.15b}$$

where n_r is the number of $\mathbf{r}$ realizations and $\mathbb{I}[\cdot] = 1$ if $[\cdot]$ is true and $\mathbb{I}[\cdot] = 0$ otherwise. Parallel computation is also carried out to speed up the solution-sorting process.

8.4.4 Next sampling points of discrete design variables

Since the current approximate solutions to problem (8.10) are found based on the GP models $\hat{\beta}(\mathbf{s}, \mathbf{r})$ and $\hat{g}_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$, it is desirable to update these solutions by sequentially specifying a promising region of the input variables in which the current solutions and the accuracy of the current GP models are deemed to be improved. It follows that the input variable vectors belonging to this promising region should have the following three properties: (1) they improve the current solutions considerably; (2) they have a high chance for being feasible solutions to problem (8.10); and (3) they

do not appear in the current training dataset. As the input variable space consists of the spaces of the discrete design variables $\mathbf{s}$ and continuous random parameters $\mathbf{r}$, it is natural to divide the exploration of the promising region into two phases as presented in Chapter 6. The first phase, discussed in this section, specifies new sampling vectors of $\mathbf{s}$, denoted as $\mathbf{s}_{\mathrm{n}}$, which are responsible for the three aforementioned properties. The second phase, discussed in the next section, finds new sampling vectors of $\mathbf{r}$, denoted as $\mathbf{r}_{\mathrm{n}}$, for addressing the third property.

The hypervolume-based approach presented Sections 3.2.1 and 6.2.2 is adopted for finding the new sampling points $\mathbf{s}_{\mathrm{n}}$ that improve the current solutions. Let $\Omega = \{\mathbf{f}_1, ..., \mathbf{f}_M\} \in \mathbb{R}^2$ and $\mathbf{f}_{\mathrm{R}} \in \mathbb{R}^2$ denote the current set of M approximate Pareto-optimal solutions to problem (8.10) and a fixed reference point dominated by all elements of Ω, respectively. Ω and $\mathbf{f}_{\mathrm{R}}$ together define the HV measure for the confined space surrounded by them.For two arbitrary sets of solutions, the set with the larger HV is better than the other one. Therefore, if each of the new sampling points $\mathbf{s}_{\mathrm{n}}$ improves the current approximate solutions to problem (8.10), the union of the corresponding objective function vector $\mathbf{f}(\mathbf{s})$ and Ω should form a new HV greater than that of the current Ω. This improvement is further represented by $\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_{\mathrm{R}})$ detailed in Eq. (3.8). As a result, HVI should be maximized for finding the new sampling points $\mathbf{s}_{\mathrm{n}}$ that correspond to a major improvement in the current HV.

Another important requirement for the vectors $\mathbf{s}_{\mathrm{n}}$ is that they should have a high chance to become feasible solutions to the problem (8.10). Accordingly, $\mathbf{s}_{\mathrm{n}}$ must satisfy the deterministic constraints of problem (8.10) and belong to a region in which $g_l(\mathbf{s})$ $(l = 1, 2, 3)$ are minimized simultaneously. In other words, $\mathbb{P}\left[g_l(\mathbf{s}, \mathbf{r}) \leq 0\right]$ $(l = 1, 2, 3)$ in Eq. (8.9) should be simultaneously maximized, which is further transformed to maximizing the following feasibility indicator:

$$\mathrm{FI}(\mathbf{s}) = \prod_{l=1}^{3} \mathbb{P}\left[g_l(\mathbf{s}, \mathbf{r}) \leq 0\right]. \tag{8.16}$$

Here FI is formulated from the fact that $g_l(\mathbf{s})$ $(l = 1, 2, 3)$ are independent, and $\mathbb{P}\left[g_l(\mathbf{s}, \mathbf{r}) \leq 0\right]$ always take positive values. It is apparent that FI conflicts with HVI because its maximizer tends to minimize the HV; see Fig. 8.3. Thus, it is rational to formulate a bi-objective deterministic maximization problem for managing these conflicting criteria.

The last property of the promising region requires that the vectors $\mathbf{s}_{\mathrm{n}}$ do not belong to the current training dataset $\mathcal{D}$. Thus, $\mathbf{s}_{\mathrm{n}}$ can be found after each optimization iteration by solving the following bi-objective deterministic maximization problem:

$$\begin{aligned} \mathbf{s}_{\mathrm{n}} = \underset{\mathbf{s} \notin \mathcal{D}}{\operatorname{argmax}} \quad & \left[\mathrm{HVI}(\mathbf{f}(\mathbf{s}) \mid \Omega, \mathbf{f}_{\mathrm{R}}), \mathrm{FI}(\mathbf{s})\right] \\ \text{subject to} \quad & \tilde{g}_i(\mathbf{s}) \leq 0 \quad (i = 1, \ldots, 2d_1), \\ & h_j(\mathbf{s}) \leq 0 \quad (j = 1, \ldots, J), \\ & s_k \in \mathcal{S}_k \quad (k = 1, \ldots, d_1). \end{aligned} \tag{8.17}$$

Problem (8.17) is solved using NSGA-II [91] whose parameters including the population size, maximum number of generations, crossover fraction, tolerance for the objective and constraint functions are, respectively, assigned as 2000, 100, 80%, and 10^{-6}. Since the exact values of FI at $\mathbf{s}_{\mathrm{n}}$ are not important for solving problem (8.10), problem (8.17) can be quickly solved by using the SAA in

Chapter 7 for reasonable estimations of the probabilities incorporated in the FI rather than using MCS. Solutions to problem (8.17) are then added to the current training dataset $\mathcal{D}$ and the set Ω_{a} for updating the current GP models and for sorting the solutions in the next optimization iteration, respectively.

8.4.5 Next sampling points of correlated random parameters

Once the sampling points $\mathbf{s}_{\mathrm{n}}$ have been found, the new sampling points $\mathbf{r}_{\mathrm{n}}$ of the correlated random parameters are specified accordingly. Since the vectors $\mathbf{r}_{\mathrm{n}}$ address the last property of the promising region, they are randomly generated using Eq. (8.14) so that they do not belong to the current training dataset $\mathcal{D}$. The number of $\mathbf{r}_{\mathrm{n}}$ equals that of $\mathbf{s}_{\mathrm{n}}$.

8.4.6 Enrichment of the existing candidate solutions

The set Ω_{a} of the existing candidate solutions is enriched before starting a new optimization iteration. The enriched Ω_{a} consists of three different groups of the discrete candidate solutions $\mathbf{s}$. The first group includes the samples of the updated training dataset $\mathcal{D}$ because the new sampling points $\mathbf{s}_{\mathrm{n}}$, which improve the solution quality, are expected to become the solutions to problem (8.10). The second group consists of the new candidates generated by performing a total of k_1 random perturbations (in the design variable space) surrounding each of the current approximate Pareto-optimal solutions of the set Ω, which can be regarded as performing discrete local searches. Each perturbation randomly increases or decreases each integer element of every approximate Pareto-optimal solution by an integer value, such as 1, 2, 3, or 4. Thus, the random perturbations can be expected to improve the solutions through the discrete local searches in the design variable space of the current solutions even though the neighborhood in this space differs from that in the objective function space. With the same expectation for the solution improvement, the third group consists of a total of k_2 new candidates generated uniformly over the design domain, which can be regarded as performing discrete global searches.

8.4.7 Optimization procedure

In summary, the proposed optimization procedure for solving problem (8.10) sequentially executes the following six steps:

- **Step 1**: Generate samples of $\mathbf{s}$ and $\mathbf{r}$. Then, create the training dataset $\mathcal{D}$ by performing NRHA for each sample.
- **Step 2**: Construct GP models for $\beta(\mathbf{s}, \mathbf{r})$ and $g_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$.
- **Step 3**: Sort the approximate Pareto-optimal solutions among the existing candidate solutions of the set Ω_{a}.
- **Step 4**: Terminate the optimization process and output the approximate Pareto-optimal solutions if one of the following stopping criteria is satisfied: (1) the number of optimization iterations reaches an upper limit $t_{\max}$ specified the user, and (2) the relative difference of the HVs at the current and previous iterations is less than or equal to a small positive value ϵ_{HV}. Otherwise, proceed to **Step 5**.

- **Step 5**: Find new sampling points $\mathbf{s}_\mathrm{n}$ and $\mathbf{r}_\mathrm{n}$. If no new sampling point is found, i.e., problem (8.17) has no solution, terminate the optimization process. Otherwise, enrich Ω_a and go to **Step 6**.
- **Step 6** Evaluate the LSFs of interest for $\mathbf{s}_\mathrm{n}$ and $\mathbf{r}_\mathrm{n}$ found in **Step 5** using NRHA; update $\mathcal{D}$ and the current GP models for $\beta(\mathbf{s},\mathbf{r})$ and $g_l(\mathbf{s},\mathbf{r})$ $(l=1,2,3)$; and reiterate from **Step 3**.

8.5 Test problem: Two-bar truss

To carefully assess its performance, the proposed optimization method is used for optimizing the two-bar truss shown in Fig. 5.1, which is taken from Sections 5.2.1 and 6.3.3. The vector $\mathbf{s}=[s_1,s_2]^T$ of design variables consists of the cross-sectional area s_1 of the truss members and the horizontal span s_2. The random parameters $\mathbf{r}$ include the magnitude of the external load P, the mass density ρ and yield stress $\sigma_{\mathrm{y},0}$ of the truss material. They are assumed to be uncorrelated as $\mathbf{C}_r=\mathbf{I}$. Probabilistic characteristics of $\mathbf{r}$ are described in Table 6.1.

A bi-objective RBDO problem whose formulation is similar to that of problem (8.10) is formulated for the truss. The mean and standard deviation of the total mass of the truss are considered as two objective functions, while the probabilistic constraints are associated with the axial stress in the truss members. Let $f_1(\mathbf{s},\mathbf{r})$ indicate the total mass of the truss, $f_{11}(\mathbf{s})=\mathbb{E}\left[f_1(\mathbf{s},\mathbf{r})\right]$ and $f_{12}(\mathbf{s})=\sqrt{\mathbb{V}\left[f_1(\mathbf{s},\mathbf{r})\right]}$ represent the mean and standard deviation of $f_1(\mathbf{s},\mathbf{r})$, respectively. Also, let $g_1(\mathbf{s},\mathbf{r})$ and $g_2(\mathbf{s},\mathbf{r})$ indicate the LSFs corresponding to the axial stress in the truss members. For simplification, $g_1(\mathbf{s},\mathbf{r})$ and $g_2(\mathbf{s},\mathbf{r})$ do not account for the self-weight of the truss. The bi-objective RBDO problem of the truss is stated as

$$\begin{aligned}
\underset{\mathbf{s}}{\text{minimize}} \quad & [f_{11}(\mathbf{s}),f_{12}(\mathbf{s})] \\
\text{subject to} \quad & \mathbb{P}\big[g_1(\mathbf{s},\mathbf{r})\le 0\big]\ge 1-\epsilon_1, \\
& \mathbb{P}\big[g_2(\mathbf{s},\mathbf{r})\le 0\big]\ge 1-\epsilon_2, \\
& s_1\in\mathcal{S}_1=\{1.0,1.5,\ldots,20.0\}\,\mathrm{cm}^2, \\
& s_2\in\mathcal{S}_2=\{0.1,0.15,\ldots,2.0\}\,\mathrm{m}.
\end{aligned} \tag{8.18}$$

where $\mathcal{S}_1$ and $\mathcal{S}_2$ are candidate lists for selecting s_1 and s_2, respectively; and $f_1(\mathbf{s},\mathbf{r})$, $g_1(\mathbf{s},\mathbf{r})$, and $g_2(\mathbf{s},\mathbf{r})$, respectively, derived from Eqs. (6.10b), (6.10c), and (6.10a) read

$$f_1(\mathbf{s},\mathbf{r})=\frac{10^{-4}\rho s_1\sqrt{1+s_2^2}}{m_{\max}}, \tag{8.19a}$$

$$g_1(\mathbf{s},\mathbf{r})=\frac{5P}{\sqrt{65}s_1\sigma_{\mathrm{y},0}}\sqrt{1+s_2^2}\left(8+\frac{1}{s_2}\right)-1, \tag{8.19b}$$

$$g_2(\mathbf{s},\mathbf{r})=\frac{5P}{\sqrt{65}s_1\sigma_{\mathrm{y},0}}\sqrt{1+s_2^2}\left(8-\frac{1}{s_2}\right)-1. \tag{8.19c}$$

Here $m_{\max}=45$ kg is the maximum nominal mass of the truss. Since $f_1(\mathbf{s},\mathbf{r})$ is a linear function of ρ, $f_{11}(\mathbf{s})$ is derived as $f_{11}(\mathbf{s})=\left(\mathbb{E}[\rho]/\sqrt{\mathbb{V}[\rho]}\right)f_{12}(\mathbf{s})=5f_{12}(\mathbf{s})$. Therefore, the set of Pareto-optimal solutions to problem (8.18) has only one solution.

To examine the robustness of the proposed method, three different training datasets are generated

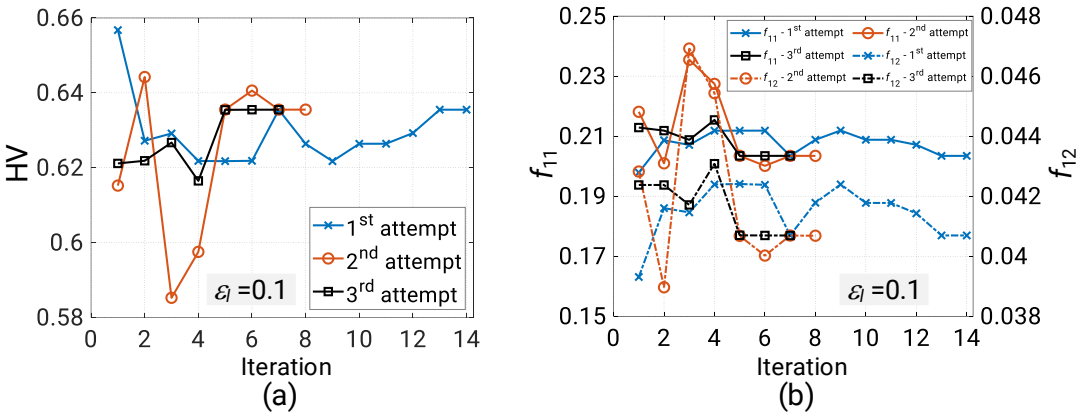

Figure 8.4: Histories of the optimization process for solving the two-bar truss with $\epsilon_l = 0.1$. (a) HV; (b) objective functions.

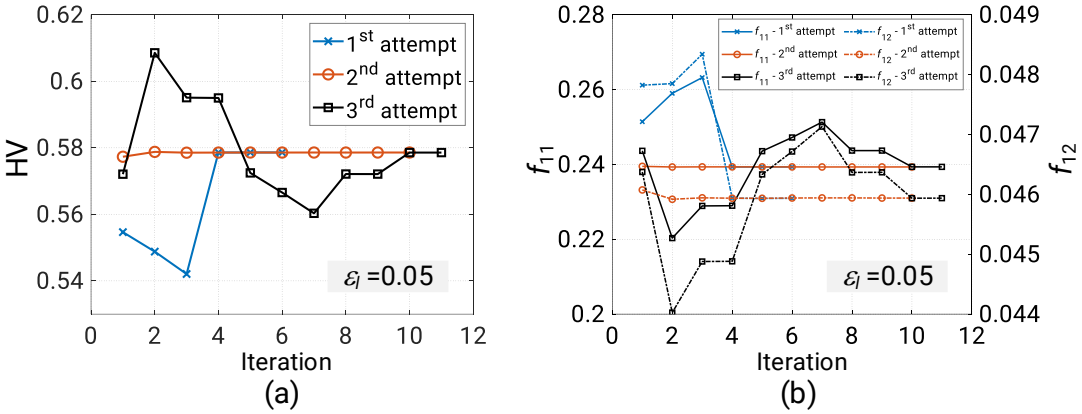

Figure 8.5: Histories of the optimization process for solving the two-bar truss with $\epsilon_l = 0.05$. (a) HV; (b) objective functions.

for processing optimization of the truss. The first, second, and third training datasets having 50, 100, and 200 samples of the input variables, respectively, are used for calculating $f_1(\mathbf{s}, \mathbf{r})$, $g_1(\mathbf{s}, \mathbf{r})$, and $g_2(\mathbf{s}, \mathbf{r})$. However, only 34, 69, and 143 samples offering non-positive values of $f_1(\mathbf{s}, \mathbf{r})$, $g_1(\mathbf{s}, \mathbf{r})$, and $g_2(\mathbf{s}, \mathbf{r})$ serve as the feasible training samples. To develop the GP models for $f_1(\mathbf{s}, \mathbf{r})$, $g_1(\mathbf{s}, \mathbf{r})$, and $g_2(\mathbf{s}, \mathbf{r})$, the DACE toolbox [92] is used together with a first-degree-polynomial mean function. It is desirable to validate the accuracy of the GP models against a test dataset if only one set of the models is used throughout the optimization process. Yet it is not necessary to do so because the GP models are updated sequentially during the optimization process. In the proposed optimization method, the accuracy of the GP models is refined intelligently in the promising region of the input variable space, and therefore, the improvement in the solutions indicates the improvement in the GP models in such a region.

Problem (8.18) is solved for two risk levels $\epsilon_l = 0.1$ and 0.05 ($l = 1, 2$). For each risk level, the optimization process is carried out three times corresponding to the three training datasets. In each optimization iteration, MCS uses $n_\mathrm{r} = 4 \times 10^4$ samples of $\mathbf{r}$ for evaluating the objective and probabilistic constraint functions. To update Ω_a, we set $k_1 = 100$ and $k_2 = 500$. The reference point is $\mathbf{f}_\mathrm{R} = [1, 0.2]$. The stopping criteria include $t_\mathrm{max} = 20$ iterations and $\epsilon_\mathrm{HV} = 10^{-9}$.

Figures 8.4 and 8.5 show the histories of the optimization process for $\epsilon_l = 0.1$ and 0.05, respec-

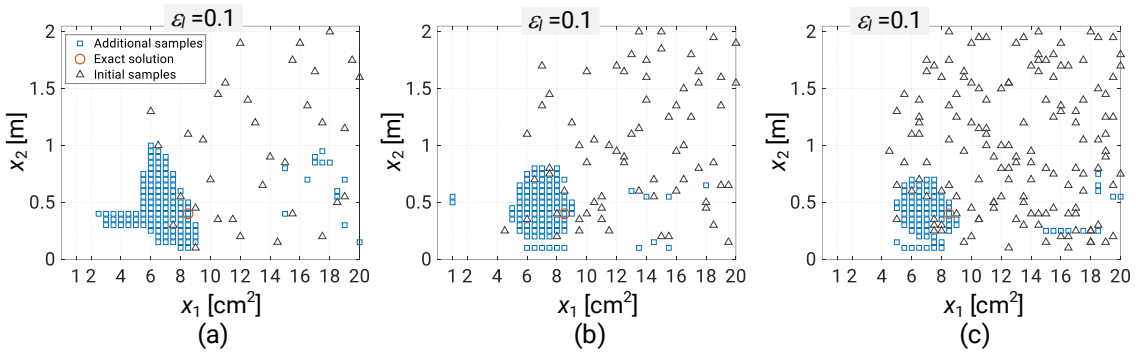

Figure 8.6: Histories of specifying new sampling points of the design variables for the two-bar truss with $\epsilon_l = 0.1$. (a) 1st attempt; (b) 2nd attempt; (c) 3rd attempt.

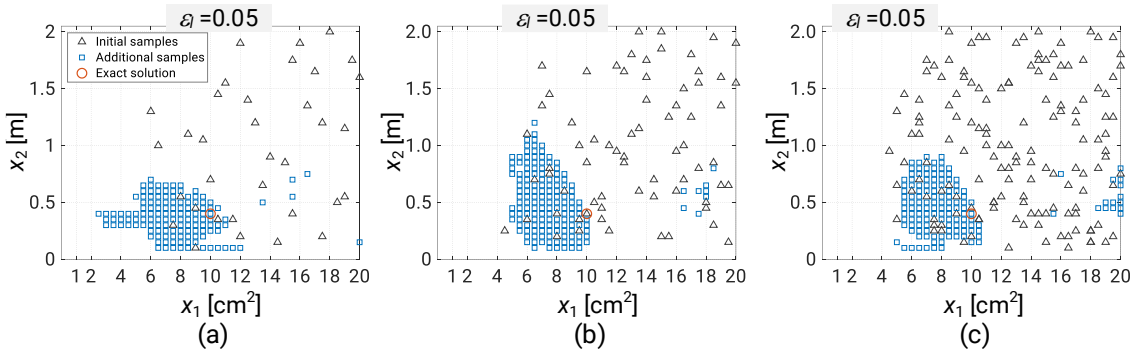

Figure 8.7: Histories of specifying new sampling points of the design variables for the two-bar truss with $\epsilon_l = 0.05$. (a) 1st attempt; (b) 2nd attempt; (c) 3rd attempt.

tively. Although the evolution of HV and that of the objective functions for the three optimization trials of each risk level follow different patterns due to the use of different training datasets, the optimization process is still able to quickly terminate and provides a unique solution. The designs of the truss for $\epsilon_l = 0.1$ and 0.05 are $\mathbf{s} = [8.5, 0.4]$ and $[10.0, 0.4]$, respectively. The maximum computational times required for $\epsilon_l = 0.1$ and 0.05 are 5257 and 4058 s using an Intel(R) i7-1165G7 2.80 GHz CPU and 8.0 GB memory, respectively.

Figures 8.6 and 8.7 show the histories of specifying the new sampling points $\mathbf{s}_\mathrm{n}$ during the optimization process for $\epsilon_l = 0.1$ and 0.05, respectively. As is clear, the new sampling points $\mathbf{s}_\mathrm{n}$ distribute in the neighborhood of the exact solution to problem (8.18), leading to quick termination of the optimization process as observed.

To further verify the obtained designs, the exact solution to problem (8.18) is found for each risk level. Since both $\mathcal{S}_1$ and $\mathcal{S}_1$ have 39 elements, a total of $39 \times 39 = 1521$ possible designs can be assigned for the truss. A total of 10^5 samples of $\mathbf{r}$ are then generated for evaluating $f_{11}(\mathbf{s})$, $f_{12}(\mathbf{s})$, $\mathbb{P}\left[g_1(\mathbf{s}, \mathbf{r}) \leq 0\right]$, and $\mathbb{P}\left[g_2(\mathbf{s}, \mathbf{r}) \leq 0\right]$ associated with each design. In this way, the exact solution can be sorted for each risk level. Table 8.3 indicates a good agreement between the designs by the proposed method and the exact ones to which the associated PDFs of f_1, and CDFs of g_1 and g_2 are illustrated in Fig. 8.8. As the CDFs of g_1 at the exact solutions reach the specified risk levels, the probabilistic constraint associated with g_1 is the active constraint of problem (8.18).

The performance of the proposed method is compared with that of the BO method presented in Chapter 6 because the bi-objective problem turned out to have a single optimal solution. Fig-

Table 8.3: Comparison of optimization results for the two-bar truss.

Variable/	$\epsilon_1 = \epsilon_2 = 0.1$				$\epsilon_1 = \epsilon_2 = 0.05$			
Objective	1st	2nd	3rd	Exact	1st	2nd	3rd	Exact
s_1 [cm^2]	8.5	8.5	8.5	8.5	10.0	10.0	10.0	10.0
s_2 [m]	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
f_{11}	0.2034	0.2001	0.2034	0.2033	0.2393	0.2393	0.2394	0.2397
f_{12}	0.0407	0.0400	0.0407	0.0407	0.0479	0.0479	0.0479	0.0480

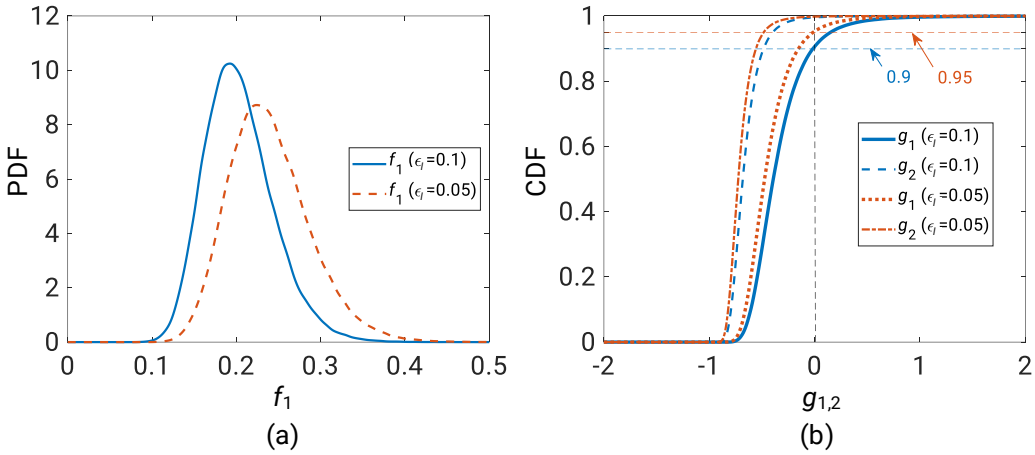

Figure 8.8: PDFs of the objective LSF and CDFs of constraint LSFs corresponding to exact optimal solutions of the two-bar truss. (a) CDF of f_1; (b) PDFs of g_1 and g_2.

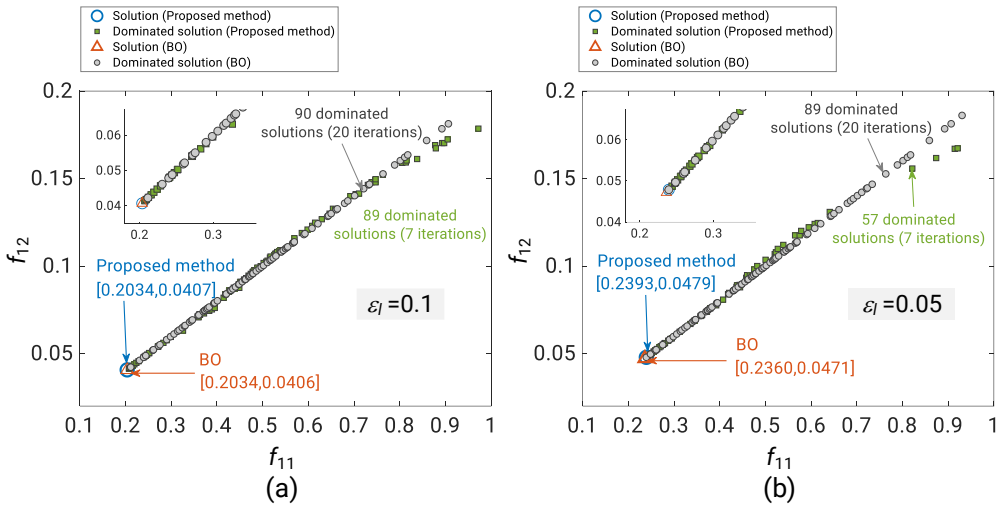

Figure 8.9: Comparison of optimization results for the two-bar truss by the proposed method and by Bayesian optimization. (a) $\epsilon_l = 0.1$; (b) $\epsilon_l = 0.05$.

ure 8.9 shows that the proposed method requires less number of optimization iterations than BO. Furthermore, it generates dominated solutions with high density in the neighborhood of the exact solutions.

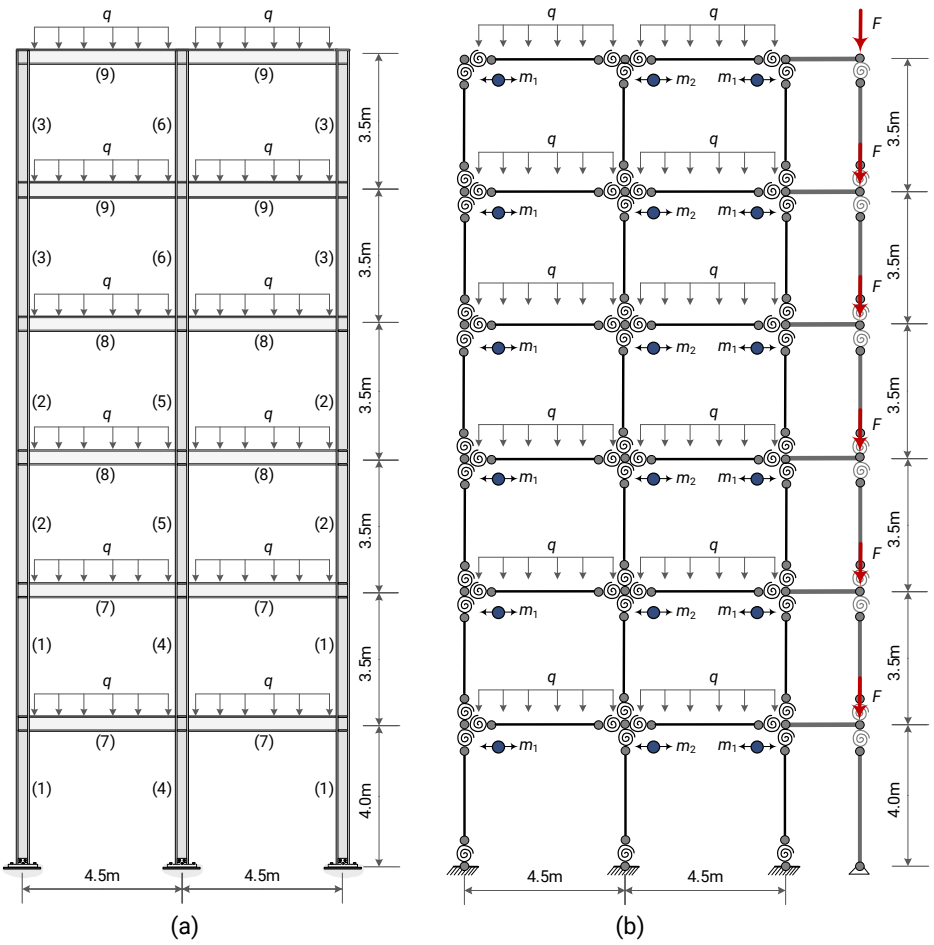

Figure 8.10: (a) Six-story two-bay frame and (b) its finite element model.

8.6 Design example: Six-story two-bay frame

This section investigates a six-story two-bay frame in Fig. 8.10(a), which is taken from Sections 5.4.2 and 6.4.3. The frame is considered as a lateral load resisting system for a residential building. The site soil is assumed to be stiff as the property of site class D according to ASCE 7-16 [23]. Parameters for the mapped MCE_R, 5%-damped, spectral response acceleration involve $T_\text{S} = 1.250\text{g}$ and $S_{(1)} = 0.4\text{g}$. The long-period transition period is $T_\text{L} = 10$ s.

The design ground motions are evaluated using an upper bound value of the fundamental natural period of the frame $T = 1.2$ s, which is derived from equation (12.8-7) of ASCE 7-16 [23]. Once the preliminary design of the frame is specified, the nominal value of its fundamental natural period can be evaluated, and that should not exceed 1.2 s.

The scale factor for each of the selected ground motions in Table 8.1 and the associated scaled acceleration spectrum are given in Fig. 8.11. The design acceleration history of each ground motion is determined by multiplying the recorded acceleration history by the corresponding scale factor.

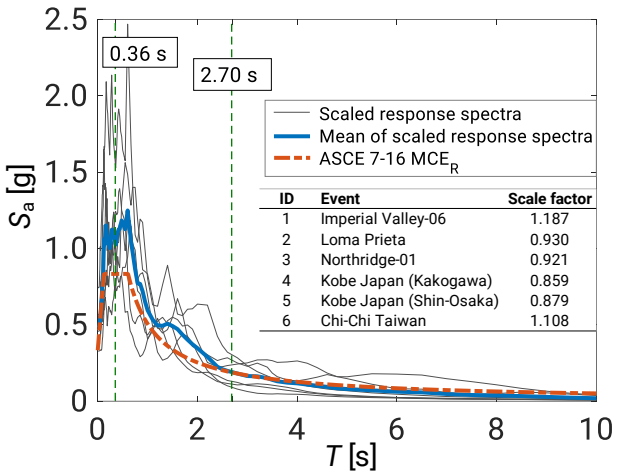

Figure 8.11: Comparison of the mean, 5%-damped response spectrum of scaled ground motions with ASCE 7-MCE$_R$ spectrum.

8.6.1 Numerical model

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). Possible steel sections for the columns and beams in each group are given in Table 8.4.

Table 8.4: List of sections for the columns and beams of the frame.

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

The OpenSees [149] is used to develop a numerical model for the frame as shown in Fig. 8.10(b). In this model, a leaning column with gravity load F is linked to the frame at each floor by a rigid truss element to account for P-Δ effects, where F vertically acts on a generic floor of the building for

Table 8.5: Assigned random parameters for the frame.

Parameter	Description	Mean	COV	Distribution
q	Distributed mass [t/m]	3.06	0.20	Normal
F	Gravity load [kN]	500	0.10	Normal
E	Young's modulus [GPa]	200	0.04	Normal
$\sigma_{y,0}$	Yield stress [MPa]	262.50	0.06	Normal
θ_p	Pre-capping plastic rotation [rad]	0.022	0.27	Lognormal*
θ_{pc}	Post-capping plastic rotation [rad]	0.17	0.35	Lognormal*
θ_u	Ultimate rotation capacity [rad]	0.06	–	–
λ_r	Reference cumulative rotation capacity [rad]	1.10	0.44	Lognormal*
M_c/M_y	Post-yield strength ratio	1.11	0.05	Normal
κ_r	Residual strength ratio	0.40	0.10	Normal

* Mean and standard deviation of logarithmic value

producing overturning action and secondary internal forces to the frame members through the inter-story drift. The leaning column is modeled using a rigid elastic beam-column element connected by two rotational springs with very small rotational stiffness. Geometric nonlinearity is also considered in structural analysis.

The distributed mass q of the floor acting on each beam member is divided into two equal parts assigned to the end nodes of the element; see Fig. 8.10(b). The mass of each structural member is also assigned to the end nodes of the corresponding element. Thus, the lumped masses m_1 and m_2 in the earthquake direction, as depicted in Figure 8.10(b), are derived from the masses of the connecting structural members, the distributed mass q, and the mass associated with F. The probabilistic characteristics of q and F are assumed as provided in Table 8.5. Furthermore, the Rayleigh damping matrix is formulated based on a linear combination of the mass and stiffness matrices. The damping coefficients are calculated by using 5% damping ratio for the first two modes of the frame. The stiffness damping coefficient due to the use of the elastic beam-column element with rotational springs at both ends is also modified according to equation (9) of the work by Zareian and Medina [143]. The time increment for NRHA is 0.01 s.

8.6.2 Random parameters for deterioration model

As described in Section 8.2.1, the material properties for the mIK deterioration model include the Young's modulus E; the yield stress $\sigma_{y,0}$; the rotation capacities θ_p, θ_{pc}, and θ_u ; the reference cumulative rotation capacity λ_r; and the strength ratios M_c/M_y and κ_r. The yield rotation θ_y is directly evaluated from E, $\sigma_{y,0}$, and the section modulus. The probabilistic characteristics of these material parameters and their correlations taken from the works by Lignos and Krawinkler [144] and Liu et al. [145] are listed in Tables 8.5 and 8.6, respectively.

8.6.3 Optimization results

Problem (8.10) is formulated for the frame with two risk levels $\epsilon_l = 0.1$ and 0.05 ($l = 1, 2, 3$). The limit of the inter-story drift ratios is $\delta_a = 2\%$. The allowable plastic rotation angles for the columns and beams are assigned as $\varphi_a = 0.002$ rad and $\omega_a = 0.015$ rad, respectively. The objective function

Table 8.6: Correlation coefficients for the random parameters [145].

	E	$\sigma_{\mathrm{y},0}$	θ_{p}	θ_{pc}	λ_{r}
E	1.00	0.71	0	0	0
$\sigma_{\mathrm{y},0}$	0.71	1.00	0	0	0
θ_{p}	0	0	1.00	0.69	0.44
θ_{pc}	0	0	0.69	1.00	0.67
λ_{r}	0	0	0.44	0.67	1.00

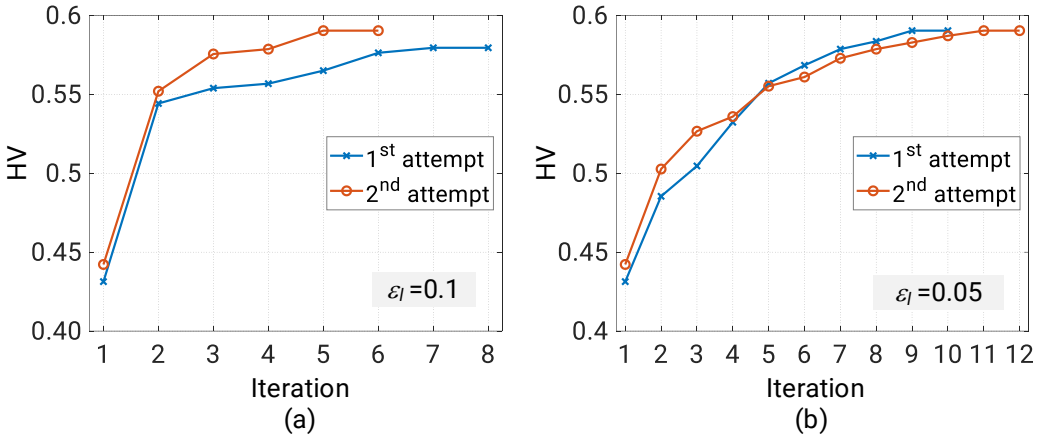

Figure 8.12: Histories of the HVs by different optimization trials for the frame with two risk levels. (a) $\epsilon_l = 0.1$; (b) $\epsilon_l = 0.05$.

f_1 for the frame is

$$f_1(\mathbf{s}) = \frac{15\rho_1 + 14(\rho_2 + \rho_3) + 7.5\rho_4 + 7(\rho_5 + \rho_6) + 18(\rho_7 + \rho_8 + \rho_9)}{m_{\max}}, \tag{8.20}$$

where $\rho_i\,(i = 1, \ldots, 9)$ and $m_{\max} = 13107$ kg are selected and derived from the list of sections in Table 8.4, respectively.

Two different training datasets are generated for performing the optimization process. The first and second datasets with 500 and 1000 feasible sampling points of the input variables, respectively, are used as inputs to NRHA for evaluating $\beta(\mathbf{s}, \mathbf{r})$ and $g_l(\mathbf{s}, \mathbf{r})$ $(l = 1, 2, 3)$.

For each risk level, the optimization process is performed two times corresponding to the two training datasets using a PC with an Intel(R) Xeon(R) E5-2643V4 3.40 GHz CPU and 64 GB memory. $f_2(\mathbf{s})$ and $g_l(\mathbf{s})$ $(l = 1, 2, 3)$ are evaluated for each candidate solution using a total of $n_{\mathrm{r}} = 4 \times 10^4$ samples of $\mathbf{r}$. To update Ω_{a}, k_1 and k_2 are set as 200 and 1000, respectively. The reference point and stopping criteria for the optimization process are, respectively, $\mathbf{f}_{\mathrm{R}} = [1, 0]$, and $t_{\max} = 20$ iterations and $\epsilon_{\mathrm{HV}} = 10^{-9}$. The obtained solutions are further compared with reference solutions that are found by performing NSGA-II without sequential framework and with the GP models for $\beta(\mathbf{s}, \mathbf{r})$ and $g_l(\mathbf{s}, \mathbf{r})$ constructed based upon a total of 1000 training samples and the probabilistic constraints evaluated using SAA in Chapter 7.

Figures 8.12(a) and (b) show the HV histories by the two optimization trials for $\epsilon_l = 0.1$ and 0.05, respectively. The HVs are considerably improved in the very first optimization iterations and gradually increased in the later iterations. For each risk level, the final HV corresponding to the

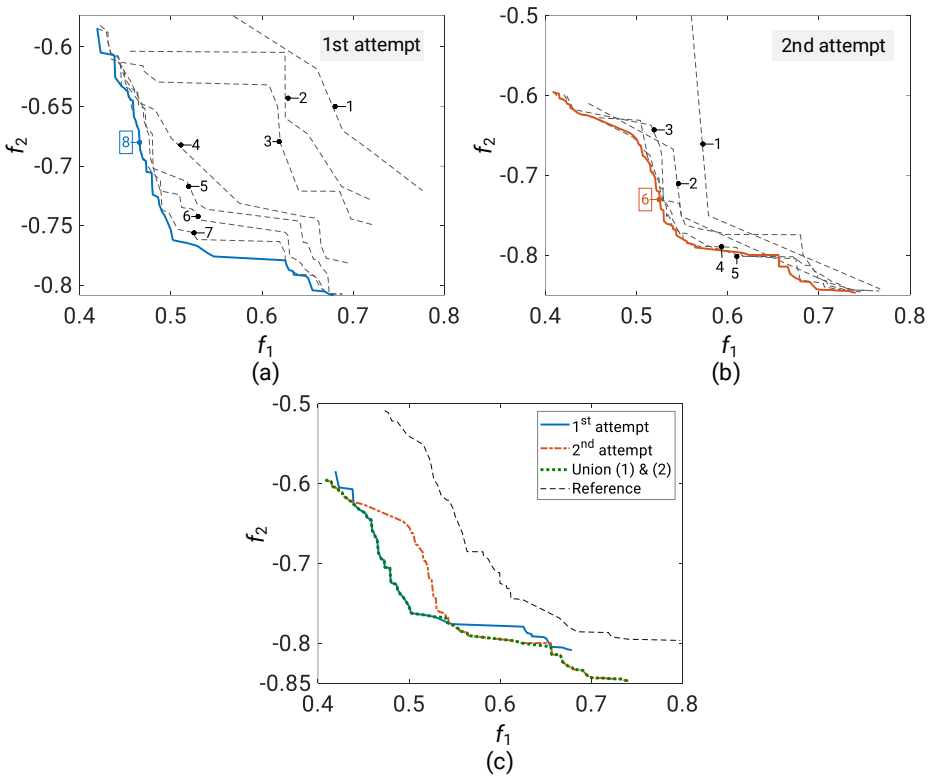

Figure 8.13: Histories of approximate Pareto-optimal solutions for the frame with $\epsilon_l = 0.1$. (a) 1st attempt; (b) 2nd attempt; (c) comparison of solutions by the two attempts and the reference ones.

second training dataset is slightly greater than that associated with the first one. Notably, the approximate Pareto-optimal solutions are always found for each risk level as the corresponding HV converges.

Figures 8.13 and 8.14 compare the obtained solutions by the two optimization trials for $\epsilon_l = 0.1$ and 0.05, respectively. It is clear that the solution quality is guaranteed to improve after each optimization iteration. The Pareto fronts for each risk level by the two optimization trials are different due to the use of different initial training datasets and the randomness of the proposed optimization method. The Pareto front can be further improved by taking the union of solutions by both trials, as shown in Figures 8.13(c) and 8.14(c). More interestingly, the obtained solutions for each risk level completely dominate the corresponding reference solutions, highlighting the importance of the proposed refinement scheme. For $\epsilon_l = 0.1$, the numbers of solutions by the first and second optimization trials are 45 and 69, respectively, and those for $\epsilon_l = 0.05$ are 43 and 82, respectively. The computational times required for the first and second trials with $\epsilon_l = 0.1$ are 8.22 and 12.03 hours, respectively, and those with $\epsilon_l = 0.05$ are 9.88 and 16.52 hours, respectively.

As it is impossible to find the global solutions to problem (8.10), it is not known whether the solutions obtained by the proposed method are the best optimal ones or not. Here the proposed method can only provide the best solutions for a particular setting of the initial training dataset and its underlying parameters (i.e., k_1, k_2, $t_{\max}$, and ϵ_{HV}) because the solutions are always found when their quality can no longer be improved. This is analogous with gradient-based algorithms that are always guaranteed to converge to a local solution. Thus, it is desirable to perform optimization

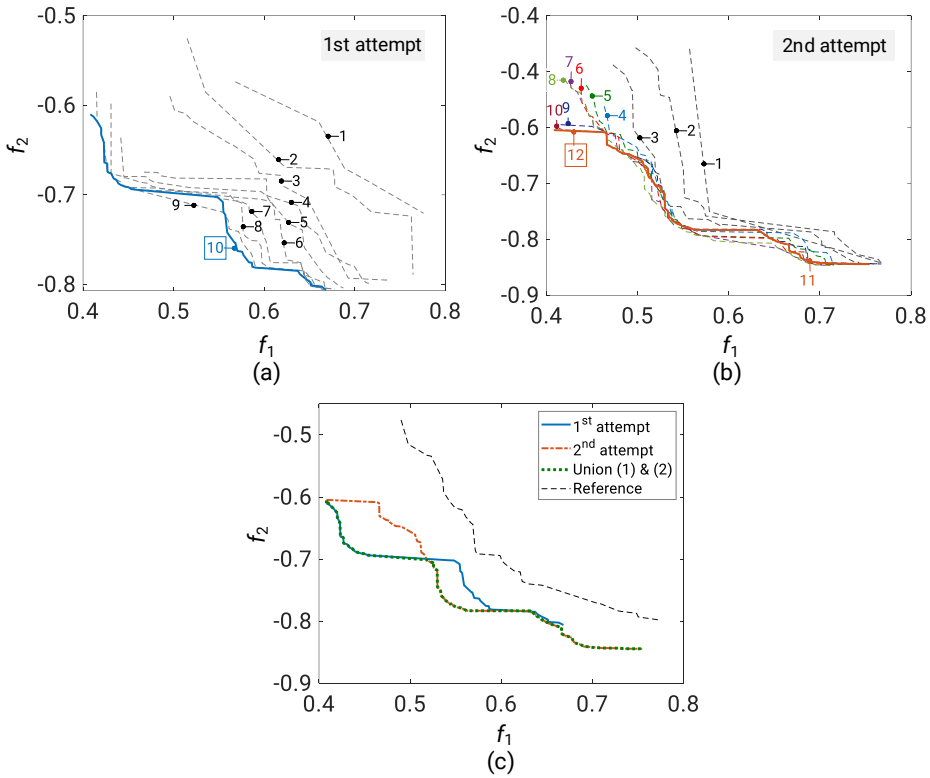

Figure 8.14: Histories of approximate Pareto-optimal solutions for the frame with $\epsilon_l = 0.05$. (a) 1st attempt; (b) 2nd attempt; (c) comparison of solutions by the two attempts and the reference ones.

several times with different settings of the algorithm so that the best solutions among those from the attempts can be found.

8.6.4 Selection of preliminary design

From the Pareto front obtained for each risk level, the solution with maximum energy dissipation ratio of the beams may be assigned as the preliminary design of the frame. In this way, the designs corresponding to the first and second optimization trials for $\epsilon_l = 0.1$ are, respectively, $\mathbf{s}_1 = [2, 6, 6, 1, 2, 4, 18, 8, 18]$ and $\mathbf{s}_2 = [2, 2, 3, 3, 5, 12, 15, 17, 18]$, and those for $\epsilon_l = 0.05$ are $\mathbf{s}_1 = [2, 9, 9, 3, 6, 6, 18, 8, 18]$ and $\mathbf{s}_2 = [2, 2, 3, 4, 4, 11, 15, 18, 18]$, respectively.

To verify the feasibility of these designs, the nominal fundamental natural period, mean of the energy dissipation of the beams, and uncertain LSFs corresponding to each design are evaluated using NRHA with 1000 samples of $\mathbf{r}$. As a result, the nominal fundamental natural period values corresponding to $\mathbf{s}_1$ and $\mathbf{s}_2$ for $\epsilon_l = 0.1$ are 1.04 and 1.08 s, respectively, and those for $\epsilon_l = 0.05$ are 1.09 and 1.09 s, respectively, which are all less than 1.2 s. The expected energy dissipation ratio of the beams and uncertain constraint functions associated with each design by the proposed method agree with those evaluated by NRHA, as shown in Table 8.7. The preliminary designs are feasible as the corresponding probabilistic constraints provide safety margins, which is illustrated by the CDFs of g_l $(l = 1, 2, 3)$ corresponding to each design in Fig. 8.15.

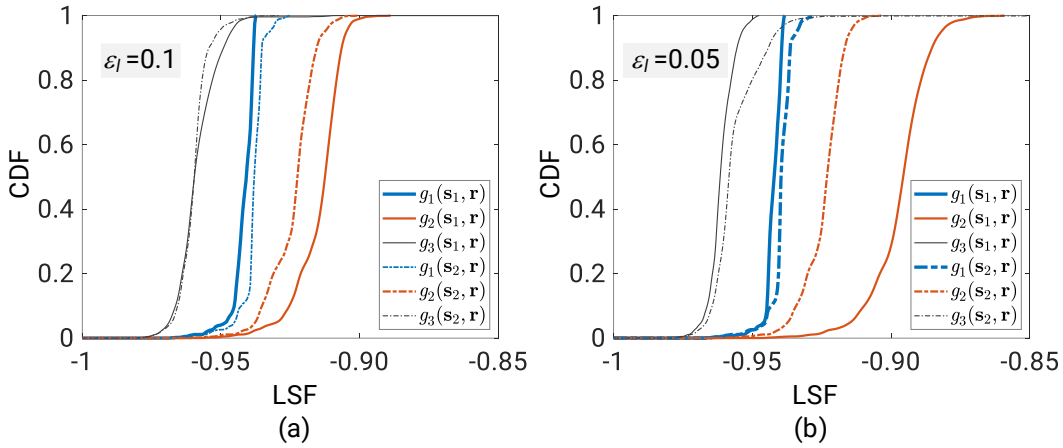

Figure 8.15: CDFs of probabilistic constraint functions at the selected designs for each risk level. (a) $\epsilon_l = 0.1$; (b) $\epsilon_l = 0.05$.

Table 8.7: Comparisons of the energy dissipation ratios of beam members and the uncertain constraints between different selected designs of the frame.

Design	Objective/ constraint	$\epsilon_l = 0.1\,(l = 1,2,3)$		$\epsilon_l = 0.05$	
		Proposed method	NRHA	Proposed method	NRHA
$\mathbf{s}_1$	$f_2(\mathbf{s}_1)$	0.808	0.819	0.806	0.808
	$g_l(\mathbf{s}_1)$ $(l = 1,2,3)$	-0.1	-0.1	-0.05	-0.05
$\mathbf{s}_2$	$f_2(\mathbf{s}_2)$	0.847	0.854	0.844	0.859
	$g_l(\mathbf{s}_2)$	-0.1	-0.1	-0.05	-0.05

8.7 Conclusions

This chapter has introduced a novel sequential batch sampling approach to solving a discrete bi-objective RBDO problem of moment-resisting steel frames subjected to earthquake excitation. The problem is formulated to optimize the total mass of the frame and energy dissipation of the beam members under unfavorable effects of correlated random parameters of floor masses, external loads, and material properties. The probabilities of exceeding allowable values of both the maximum inter-story drift and the rotational angles of the structural members are constrained, while the compactness of the steel sections is required for fully sustaining plastic deformations. The main conclusions of this chapter are summarized as follows:

(1) Approximations of the dynamic responses using the corresponding GP models facilitate solving the bi-objective RBDO problem of the frame, which may be computationally intractable if NRHA is directly used for uncertainty propagation.

(2) A quick termination and the robustness of the proposed method arise from the fact that the new sampling points of the design variables to refine the accuracy of the GP models tend to distribute in the neighborhood of the exact solutions to the RBDO problem. In other words, the bi-objective deterministic maximization problem formulated for specifying the new sampling points of the proposed refinement scheme is suitable for sequentially solving the RBDO problem.

The proposed refinement scheme may further accelerate the optimization process if one can carry out parallel computation for NRHA associated with the new sampling points.

(3) The approximate solutions to the RBDO problem are always found when their quality can no longer be improved, regardless of using a very small number of 10^{-9} as the termination condition on the change in the solution quality and considering the maximum number of optimization iterations as another stopping criterion for the proposed method.

Chapter 9

Conclusion and future direction

9.1 Summary

This dissertation has introduced various probabilistic approaches to inverse identification of material parameters for costly FE models that simulate the cyclic elastoplastic behavior of structural steels and to solving RDO and RBDO problems formulated for steel frame structures under aleatory uncertainty in design parameters. The ultimate goal is to reduce enormous computational costs arising from repeatedly carrying out many costly FE simulations during solving the parameter identification problems with and without considering observational noise as well as from carrying out difficult tasks in optimization of steel structures in the presence of uncertainty, namely uncertainty propagation, handling individual and joint probabilistic constraints, dealing with discrete design variables, and handling multiple objective functions. The background of this study is described in Chapter 1, followed by a selective review of basic concepts of probability theory and some important probabilistic approaches in Chapter 2. Chapters 3–8 detail the approaches to solving parameter identification, RDO, and RBDO problems. The findings and contributions of this dissertation are summarized as follows.

9.1.1 For identification problems of cyclic elastoplastic parameters for structural steels

Single- and multi-objective inverse problems are formulated and, respectively, solved by SOBO and MOBO approaches in Chapter 3 for identifying the elastoplastic parameters used to simulate the cyclic behavior of structural steels subjected to different cyclic loading conditions. Reliable cyclic elastoplastic parameters can be found by SOBO with considering noise-free and noise-corrupted error functions. When expending the same number of simulation calls, SOBO outperforms GA and PSO in terms of the prediction performance of identified parameters. Nevertheless, the parameters identified from SOBO exhibit the dataset-specific bias that may lead the elastoplastic parameters for the cyclic constitutive law identified from a single loading history to inaccurate predictions of structural responses under other loading histories. The best and second-best compromise solutions of parameters on the Pareto front obtained from MOBO can mitigate this bias. A dataset-specific index has also been proposed for a rigorous assessment of the dataset-specific bias levels for different

sets of identification results obtained from different inverse problem formulations.

A BO-assisted ABC approach introduced in Chapter 4 can facilitate the Bayesian inference of cyclic elastoplastic parameters for structural steels from noise-corrupted experimental datasets. ABC addresses the computationally intractable likelihood function of the Bayesian inference framework, while BO nested in ABC handles the costly simulation for intelligently reducing the uncertainty in the approximation of ABC posterior. As demonstrated in the illustrative examples, the BO-assisted ABC approach is capable of providing approximate posteriors that not only reproduce the true posterior with acceptable accuracy and capture the true deterministic parameters of a known statistical model but also reliably predict the cyclic elastoplastic behavior of a steel specimen subjected to different cyclic loading conditions. These results are brought by the performance of BO in intelligently refining the approximate likelihood function in important regions of the parameter space that substantially contributes to the improvement of the accuracy of approximate ABC posterior. When using a sufficiently large number of MCMC samples, a small discrepancy threshold leads to the acceptance of more good samples from the prior as samples constituting the posterior. However, it may be impossible to improve the posterior approximation when the discrepancy threshold reaches a small-enough value. Although their potential in probabilistically modeling the cyclic elastoplastic parameters has been demonstrated, the accuracy of approximate posteriors by the BO-assisted ABC approach is strongly affected by the acquisition function used for guiding BO.

9.1.2 For RDO problems

An important task for solving the less-variance RDO problems is to understand the cause-effect relationship between the random input variables and the performance function that, in turn, facilitates the propagation of uncertainty for estimating the mean and variance of the performance function for solving the problems. By letting the random input variables and the corresponding structural response behave according to a joint PDF under the presence of uncertainty, GMM presented in Chapter 5 is capable of approximating such a joint PDF based on a training dataset, and hence characterizing the probabilistic property of the structural response and providing a simple regression function for calculation of this response as well as its statistical estimates. The method also incorporates uncertainty in the small ranges of discrete design variables before carrying out optimization under various conditions. Results in Chapter 5 show the GMM ability not only to accurately characterize the probabilistic property of a LSF for a steel frame structure but also to find good optimal solutions that well represent different levels of the trade-off between the conflicting objective functions of the multi-objective RDO problem formulated for planar steel frames.

Although considering the joint and individual probabilistic constraints makes the multi-objective RDO problem more realistic, it increases the complexity of the resulting two multi-objective RDO problems, namely joint and individual probabilistic constrained multi-objective RDO problems. The proposed BO approach in Chapter 6 is able to address challenges arising from solving these RDO problems such as the implicit objective and constraint functions, discrete nature of the problems formulated for steel frame structures, and difficulty in evaluation of the probabilistic constraints with a reasonable computational cost. Guided by two new acquisition functions developed for the two RDO problems, the proposed BO approach can quickly offer exact or good approximate Pareto-optimal solutions to the RDO problems of steel frame structures. Nevertheless, different BO trials may lead to different approximate Pareto-optimal solutions because the proposed method is random

in nature. The effect of this randomness is pronounced when the feasible region of the formulated RDO problems is large. Thus, it is important to select a candidate list of both column and beam sections using some design rules of thumb for reduction of the discrepancy in the obtained solutions.

9.1.3 For RBDO problems

An efficient optimization procedure combining MGP modeling, decoupling approach, and SAA is proposed in Chapter 7 for solving continuous and discrete RBDO problems of truss and frame structures under aleatory uncertainty in material properties and external loads. With support from GMM in splitting a training dataset into independent subsets, MGP modeling scales up the application of GP to a large training dataset using a weighted average function of local GP models constructed from the subsets. By doing so, MGP models can well cover the input variable space for improving the overall accuracy of the regression models and significantly reduce the computation time required for the training process. This further enables the development of a novel SAA and its inverse based upon the first three cumulants of the uncertain LSFs to handle the probabilistic constraints of the RBDO problem effectively. The efficiency of the proposed SAA in calculation of failure probabilities has also been verified in comparison with MCS. An important contribution of Chapter 7 is the replacement of the original RBDO problem with a SDO problem that uses the MGP models as surrogates for the LSFs in the probabilistic constraints. The SDO problem is strategically solved, with support from the proposed SAA in calculation of the inverse failure probabilities, for exploring the region that is deemed to contain the optimal solution and improving the accuracy of the MGPs in that region. As shown in the optimization results for a ten-bar truss and a four-story three-bay steel frame, the proposed SDO problem can quickly provide a reliable solution to the original RBDO problem.

As an extension of Chapter 7, Chapter 8 introduces a novel sequential batch sampling approach to solving a discrete bi-objective RBDO problem of moment-resisting steel frames subjected to earthquake excitation. The RBDO problem is formulated to optimize the total mass of the frame and the energy dissipation of its beam members under unfavorable effects of correlated random parameters of floor masses, external loads, and material properties. The dynamic responses are approximated by the corresponding GP models to facilitate solving the RBDO problem, which may be computationally intractable if NRHA is directly incorporated into MCS for uncertainty propagation. A bi-objective deterministic maximization problem is then proposed for specifying the new sampling points that refine the accuracy of the GP models. This bi-objective deterministic maximization problem is formulated based on two conflicting criteria, namely the improvement of the current solutions and the feasibility of new sampling points. To this end, the approximate solutions to the bi-objective RBDO problem are always found when their quality can no longer be improved. This is due to the fact that the new sampling points of the design variables to refine the accuracy of the GP models tend to distribute in the neighborhood of the exact solutions to the RBDO problem. In other words, the bi-objective deterministic maximization problem formulated for specifying the new sampling points of the proposed refinement scheme is suitable for sequentially solving the RBDO problem.

9.2 Possible directions for future research

Aside from the proposed probabilistic approaches, this dissertation initiates some problems related to the applications of these approaches to real-world optimization problems that require further research.

Since its feasibility has only been verified against simple problems of a steel specimen and a cantilever beam, it is desirable to apply the SOBO approach in Chapter 3 to complicated parameter identification problems in the field of system identification. All we need before the implementation of the SOBO approach is to formulate an error function as a function of the parameters to be identified and specify the noise involved in each measurement of the response of interest.

The multi-objective inverse problem in Chapter 3 is formulated based on only three sets of experimental results. It is, therefore, desirable to scale up the application of the multi-objective formulation to the case where many sets of experimental results are available. Either of the following two strategies may be applicable. First, the experimental sets can be split into disjoint clusters based on the similarity or the correlation between the test results. In this way, improving the objective value of a set may not worsen the objective values of other sets in the same cluster and therefore, each cluster can be represented by a single objective function, which is formulated from a representative set or from individual objectives of the sets in that cluster using the weighted-sum approach. Second, if it is not clear to determine the similarity or correlation between the experimental sets, the multi-objective inverse problem can also be formulated for every combination of three sets, and the remaining sets are used for validation. Three sets for each combination are recommended because they allow us to visualize how the solutions are distributed in the objective function space. Then, the dataset-specific bias index values for the solutions to each problem can be evaluated using the associated validation sets. Different sets of dataset-specific bias index values can be found based on the validation sets for rational decision-making.

Since the BO-assisted ABC in Chapter 4 is still in the first stage of the development of a robust posterior approximation method facilitating the Bayesian inference of costly FE models, it is outside the realm of the following important aspects.

- It is desirable to develop a scalable method for BO over high-dimensional parameter spaces because BO often limits itself to problems with less than 20 parameters [71].
- It is also important to characterize robust priors to ensure the consistency of the approximate posterior. For this, hierarchical or conjugate priors may be viable choices.
- Post-processing of parameters characterizing the approximate posterior is needed to further consider the discrepancy between simulated and measured summary statistics. This process adjusts the obtained approximate posteriors based on a regression model, such as local linear or nonlinear model [116].
- Modeling errors can be incorporated by defining the standard deviation of the discrepancy as a function of material parameters, observational noise, and modeling errors.
- The problem of selecting an appropriate constitutive law may be of interest because the ABC approach has been widely used for statistical model selection [115] while many cyclic elasto-plastic constitutive laws are available.

The GMM in Chapter 5 can be further developed with multi-objective functions so that the

performances of all structural members can be fully taken into account instead of using maximum response values for particular member groups. That means, $f(\mathbf{x})$ in Eq. (2.42) becomes a vector of structural responses, rather than a scalar. In addition, it is possible to start the construction of GMM based on a small number of training samples that can be further enriched to refine the GMM accuracy intelligently by utilizing the variance information of the GMM model.

The proposed BO approach in Chapter 6 may be computationally expensive if it is developed based on a large training dataset for increasing the accuracy of the optimization results because the GP surrogates for the structural responses are extremely sensitive to the number of samples [58]. The GMM clustering approach presented in Chapter 7 is a viable choice to increase the scalability of the proposed BO approach.

The RDO problem in Chapter 6 is limited to the strength requirements for individual structural members and does not consider the local and global collapse mechanisms of the frames. It is possible to formulate a new RDO problem by providing additional probabilistic constraints on the local and global collapse mechanisms. Such a new optimization problem may include both correlated and uncorrelated LSFs of different collapse mechanisms. Thus, a new optimization method addressing both the joint and individual probabilistic constrained RDO problems may be of interest.

As an extension of Chapter 7, future studies may focus on using more flexible clustering methods, e.g., Dirichlet process mixture model [150], to replace the GMM clustering method if the size of the training data tends to grow for a real-world problem. Also to further enhance the accuracy of the proposed SAA method, it is possible to use either higher-order derivatives of the LSFs, e.g., the Hessian, for calculating the first three cumulants of the random variables or the higher-order cumulants of the random input variables.

A nontrivial extension to the RBDO problem in Chapter 8 is to incorporate uncertainty in the earthquake ground motions and damping ratio into the design optimization process. The serviceability, repairability, and ultimate limit state verifications should also be carried out for the frame under different seismic intensities. The proposed optimization method in Chapter 8 can be extended to solving either continuous multi-objective or discrete/continuous single-objective RBDO problems. For continuous multi-objective RBDO problems, Gaussian local searches [151] may be employed for enrichment of the existing candidate solutions instead of the random perturbations as presented in Chapter 8. For discrete/continuous single-objective RBDO problems, the acquisition function for SOBO may be a viable choice to replace the HVI. Besides, the use of precise numerical models for predictions of dynamic responses or the combination of low fidelity models and high fidelity models may be another opportunity for future research.

Appendix A

Nonlinear combined isotropic/kinematic hardening model

Let $\boldsymbol{\sigma}$, $\boldsymbol{\alpha}$, and σ_{y} denote the stress tensor at a point of the steel specimen, the back-stress tensor, and the yield stress, respectively. Assume that our focus is on a small strain problem. The relationship between strain and stress states of structural steel is described by its current status that is either elastic or plastic, and can be detected using the following von Mises yield criterion:

$$F = \sqrt{\frac{3}{2}}(\boldsymbol{\xi} : \boldsymbol{\xi})^{1/2} - \sigma_{\mathrm{y}} = 0, \tag{A.1}$$

where $\boldsymbol{\xi} = \mathrm{dev}[\boldsymbol{\sigma}] - \mathrm{dev}[\boldsymbol{\alpha}]$ is the shifted-stress tensor, $\mathrm{dev}[\cdot]$ denotes the deviatoric part of $[\cdot]$, and the symbol ':' is the double dot product of two second-order tensors.

We commonly use isotropic and kinematic hardening models to describe the strain hardening process of structural steels subjected to cyclic loading. The isotropic hardening model only increases the size of the yield surface $F = 0$ during the evolution of plastic deformations. As a result, $\boldsymbol{\alpha}$ does not appear in Eq. (A.1), leading the yield surface to be an isotropic function of the stress that cannot capture the Bauschinger effect [152]. As structural steels exhibit a saturation point of the stress at large deformation, the isotropic hardening model can describe the increment of the size of the yield surface using the following Voce hardening law [153]:

$$\sigma_{\mathrm{y}} = \sigma_{\mathrm{y},0} + Q_{\infty}[1 - \exp(-b\epsilon^{\mathrm{p}}_{\mathrm{eq}})], \tag{A.2}$$

where $\sigma_{\mathrm{y},0}$ denotes the initial yield stress, Q_{∞} is the difference of the stress saturation and $\sigma_{\mathrm{y},0}$, b is the isotropic saturation rate, and $\epsilon^{\mathrm{p}}_{\mathrm{eq}}$ is the current equivalent plastic strain determined based on its previous state and the rate $\dot{\epsilon}^{\mathrm{p}}_{\mathrm{eq}}$.

The kinematic hardening model does not change the size and shape of the yield surface. Instead, it updates the center of the yield surface using a rigid translation in the evolution direction of the plastic strain. This enables the kinematic hardening model to capture the Bauschinger effect. The back-stress tensor $\boldsymbol{\alpha}$ can be further decomposed into n_{k} back-stress components for a better

approximation as [154]

$$\boldsymbol{\alpha} = \sum_{k=1}^{n_{\mathrm{k}}} \boldsymbol{\alpha}_k, \tag{A.3}$$

where the rate of $\boldsymbol{\alpha}_k$ can be described by a the following nonlinear kinematic hardening rule [155]:

$$\dot{\boldsymbol{\alpha}}_k = \sqrt{\frac{2}{3}} C_k \dot{\epsilon}^{\mathrm{p}}_{\mathrm{eq}} \mathbf{n} - \gamma_k \dot{\epsilon}^{\mathrm{p}}_{\mathrm{eq}} \boldsymbol{\alpha}_k, \tag{A.4}$$

where $\mathbf{n} = \boldsymbol{\xi}/\|\boldsymbol{\xi}\|$ denotes the unit normal vector of the yield surface; and C_k and γ_k are the translation and relaxation rates of $\boldsymbol{\alpha}_k$, respectively.

The nonlinear combined isotropic/kinematic hardening model was developed using both Eqs. (A.2) and (A.4). Let $\mathbf{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d$ denote the vector of n material parameters for this combined hardening model. Thus, $\mathbf{x} = [E, \nu, Q_\infty, b, \sigma_{\mathrm{y},0}, C_1, \gamma_1]$ if we use one back-stress component in Eq. (A.3), where E and ν are Young's modulus and Poisson's ratio of the material, respectively.

Appendix B

Monte-Carlo approximations

In general, it is difficult to directly use the integrals in Eqs. (2.5) and (2.6) for calculation of the mean and variance of r. One simple but powerful alternative is to carry out Monte-Carlo integration, which is described as follows. We first randomly generate a total n_{r} independent and identical distributed samples of r, namely $r^1, \ldots, r^{n_{\mathrm{r}}}$. Then, we can approximate the distribution of $\pi(r)$ by using the empirical distribution of $\{\pi(r^i)\}_{i=1}^{n_{\mathrm{r}}}$. As a result, the sample mean μ_r and empirical variance σ_r^2 of r are given by

$$\mu_r \approx \bar{\mu}_r = \frac{1}{n_{\mathrm{r}}} \sum_{i=1}^{n_{\mathrm{r}}} r^i, \tag{B.1a}$$

$$\sigma_r^2 \approx \frac{1}{n_{\mathrm{r}}} \sum_{i=1}^{n_{\mathrm{r}}} (r^i - \bar{\mu}_r)^2. \tag{B.1b}$$

The integral in Eq. (2.2) for calculation of $\mathbb{P}[r \leq \bar{r}]$ can also be approximated by

$$\mathbb{P}[r \leq \bar{r}] \approx \frac{1}{n_{\mathrm{r}}} \sum_{i=1}^{n_{\mathrm{r}}} \mathbb{I}\left[r^i - \bar{r}\right], \tag{B.2}$$

where indicator function $\mathbb{I}[r^i - \bar{r}]$ reads

$$\mathbb{I}\big[r^i - \bar{r}\big] = \begin{cases} 1 & \text{if } \ r^i - \bar{r} \leq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{B.3}$$

Bibliography

[1] Makoto Ohsaki. *Optimization of finite dimensional structures*. 1st. CRC Press, 2010.

[2] Arora Jasbir S. *Introduction to optimum design*. 4th. CRC Press, 2016.

[3] AISC. *AISC Shapes Database v15.0*. 2017.

[4] David H. Wolpert and William G. Macready. "No free lunch theorems for optimization". *IEEE Transactions on Evolutionary Computation* 1.1 (1997), pp. 67–82.

[5] Mykel J. Kochenderfer and Tim A. Wheeler. *Algorithms for optimization*. Cambridge, Massachusetts: The MIT Press, 2019.

[6] Richard Courant. *Variational methods for the solution of problems of equilibrium and vibrations*. 1943.

[7] John E. Dennis Jr and Jorge J. Moré. "Quasi-Newton methods, motivation and theory". *SIAM review* 19.1 (1977), pp. 46–89.

[8] Diederik P. Kingma and Jimmy Ba. *Adam: A method for stochastic optimization*. 2014.

[9] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. New York, USA: Springer, 2006.

[10] Arkadi S. Nemirovski and Michael J. Todd. "Interior-point methods for optimization". *Acta Numerica* 17 (2008), pp. 191–234.

[11] Léon Bottou. "On-line learning and stochastic approximations". In: ed. by David Saad. On-Line Learning in Neural Networks. Cambridge: Cambridge University Press, 1999, pp. 9–42.

[12] Scott Kirkpatrick, C. Daniel Gelatt, and Mario P. Vecchi. "Optimization by simulated annealing". *Science* 220.4598 (1983), pp. 671–680.

[13] Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. "A tutorial on the cross-entropy method". *Annals of Operations Research* 134.1 (2005), pp. 19–67.

[14] Melanie Mitchell. *An introduction to genetic algorithms*. Cambridge, Massachusetts: The MIT Press, 1998.

[15] James Kennedy and Russell Eberhart. "Particle swarm optimization". In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. Vol. 4. 1995, pp. 1942–1948.

[16] Armen Der Kiureghian and Ove Ditlevsen. "Aleatory or epistemic? Does it matter?" *Structural Safety* 31.2 (2009), pp. 105–112.

[17] Christopher J. Roy and William L. Oberkampf. "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing". *Computer Methods in Applied Mechanics and Engineering* 200.25 (2011), pp. 2131–2144.

[18] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.

[19] Isaac Elishakoff and Makoto Ohsaki. *Optimization and anti-optimization of structures under uncertainty*. London: Imperial College Press, 2010, p. 424.

[20] Paul E. Hess, Daniel Bruchman, Ibrahim A. Assakkaf, and Bilal M. Ayyub. "Uncertainties in material and geometric strength and load variables". *Naval Engineers Journal* 114.2 (Apr. 2002), pp. 139–166.

[21] F. Michael Bartlett, Robert J. Dexter, Mark D. Graeser, Jason J. Jelinek, Bradley J. Schmidt, and Theodore V. Galambos. "Updating standard shape material properties database for design and reliability". *Engineering Journal-American Institute of Steel Construction* 40.1 (2003), pp. 2–14.

[22] Bach Do and Makoto Ohsaki. "Gaussian mixture model for robust design optimization of planar steel frames". *Structural and Multidisciplinary Optimization* 63 (2021), pp. 137–160.

[23] ASCE/SEI 7-16. *Minimum design loads and associated criteria for buildings and other structures*. Reston, Virginia, USA, 2017.

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

[25] EN 1990. *Eurocode - Basis of structural design*. Brussels, Belgium, 2002.

[26] Bach Do and Makoto Ohsaki. "Sequential sampling approach to energy-based multi-objective design optimization of steel frames with correlated random parameters". *Earthquake Engineering and Structural Dynamics* 51.3 (2022), pp. 588–611.

[27] Robert E Melchers and André T Beck. *Structural reliability analysis and prediction*. 3th. New York: John Wiley & Sons, 2018.

[28] Bulleit William M. "Uncertainty in structural engineering". *Practice Periodical on Structural Design and Construction* 13.1 (2008), pp. 24–30.

[29] Ang Alfredo H-S. and Tang Wilson H. *Probability concepts in engineering: Emphasis on applications to civil and environmental engineering*. 2nd. New Jersey, USA: Wiley, 2007.

[30] Nikos D. Lagaros and Manolis Papadrakakis. "Robust seismic design optimization of steel structures". *Structural and Multidisciplinary Optimization* 33.6 (2007), pp. 457–469.

[31] Hans-Georg Beyer and Bernhard Sendhoff. "Robust optimization – A comprehensive survey". *Computer Methods in Applied Mechanics and Engineering* 196.33 (2007), pp. 3190–3218.

[32] Zhifeng Liu, Sez Atamturktur, and C. Hsein Juang. "Performance based robust design optimization of steel moment resisting frames". *Journal of Constructional Steel Research* 89 (2013), pp. 165–174.

[33] Makoto Ohsaki, Makoto Yamakawa, Wenliang Fan, and Zhengliang Li. "An order statistics approach to multiobjective structural optimization considering robustness and confidence of responses". *Mechanics Research Communications* 97 (2019), pp. 33–38.

[34] James L. Beck, Eduardo Chan, Ayhan Irfanoglu, and Costas Papadimitriou. "Multi-criteria optimal structural design under uncertainty". *Earthquake Engineering & Structural Dynamics* 28.7 (1999), pp. 741–761.

[35] Dan M Frangopol and Kurt Maute. "Life-cycle reliability-based optimization of civil and aerospace structures". *Computers & Structures* 81.7 (2003), pp. 397–410.

[36] Byeng D Youn and Kyung K Choi. "A new response surface methodology for reliability-based design optimization". *Computers & Structures* 82.2 (2004), pp. 241–256.

[37] Manolis Papadrakakis, Nikos D Lagaros, and Vagelis Plevris. "Design optimization of steel structures considering uncertainties". *Engineering Structures* 27.9 (2005), pp. 1408–1418.

[38] Marcos A Valdebenito and Gerhart I Schuëller. "A survey on approaches for reliability-based optimization". *Structural and Multidisciplinary Optimization* 42.5 (2010), pp. 645–663.

[39] Yoshihiro Kanno. "On three concepts in robust design optimization: absolute robustness, relative robustness, and less variance". *Structural and Multidisciplinary Optimization* 62.2 (2020), pp. 979–1000.

[40] Russel E. Caflisch. "Monte Carlo and quasi-Monte Carlo methods". *Acta Numerica* 7 (1998), pp. 1–49.

[41] Travis V. Anderson and Christopher A. Mattson. "Propagating skewness and kurtosis through engineering models for low-cost, meaningful, nondeterministic design". *Journal of Mechanical Design* 134.10 (2012).

[42] Paul D. Arendt, Daniel W. Apley, and Wei Chen. "Objective-oriented sequential sampling for simulation based robust design considering multiple sources of uncertainty". *Journal of Mechanical Design* 135.5 (2013).

[43] Thierry Crestaux, Olivier Le Maıˆtre, and Jean-Marc Martinez. "Polynomial chaos expansion for sensitivity analysis". *Reliability Engineering & System Safety* 94.7 (2009), pp. 1161–1172.

[44] Xinbo Geng and Le Xie. "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization". *Annual Reviews in Control* 47 (2019), pp. 341–363.

[45] Younes Aoues and Alaa Chateauneuf. "Benchmark study of numerical methods for reliability-based design optimization". *Structural and Multidisciplinary Optimization* 41.2 (2010), pp. 277–294.

[46] David Lehký, Ondřej Slowik, and Drahomír Novák. "Reliability-based design: Artificial neural networks and double-loop reliability-based optimization approaches". *Advances in Engineering Software* 117 (2018), pp. 123–135.

[47] Harish Agarwal, Chandan K. Mozumder, John E. Renaud, and Layne T. Watson. "An inverse-measure-based unilevel architecture for reliability-based design optimization". *Structural and Multidisciplinary Optimization* 33.3 (2007), pp. 217–227.

[48] Jinghong Liang, Zissimos P. Mourelatos, and Efstratios Nikolaidis. "A single-loop approach for system reliability-based design optimization". *Journal of Mechanical Design* 129.12 (2007), pp. 1215–1224.

[49] Songqing Shan and G. Gary Wang. "Reliable design space and complete single-loop reliability-based design optimization". *Reliability Engineering & System Safety* 93.8 (2008), pp. 1218–1230.

[50] Chen Jiang, Haobo Qiu, Liang Gao, Xiwen Cai, and Peigen Li. "An adaptive hybrid single-loop method for reliability-based design optimization using iterative control strategy". *Structural and Multidisciplinary Optimization* 56.6 (2017), pp. 1271–1286.

[51] Xiaoping Du and Wei Chen. "Sequential optimization and reliability assessment method for efficient probabilistic design". *Journal of Mechanical Design* 126.2 (2004), pp. 225–233.

[52] Gengdong Cheng, Lin Xu, and Lei Jiang. "A sequential approximate programming strategy for reliability-based structural optimization". *Computers & Structures* 84.21 (2006), pp. 1353–1367.

[53] Yan-Gang Zhao and Tetsuro Ono. "A general procedure for first/second-order reliabilitymethod (FORM/SORM)". *Structural Safety* 21.2 (1999), pp. 95–112.

[54] Rodrigo C. Soares, Alaa Mohamed, Wilson S. Venturini, and Maurice Lemaire. "Reliability analysis of non-linear reinforced concrete frames using the response surface method". *Reliability Engineering & System Safety* 75.1 (2002), pp. 1–16.

[55] Jian Deng, Desheng Gu, Xibing Li, and Zhong Qi Yue. "Structural reliability analysis for implicit performance functions using artificial neural network". *Structural Safety* 27.1 (2005), pp. 25–48.

[56] Jian Deng. "Structural reliability analysis for implicit performance function using radial basis function network". *International Journal of Solids and Structures* 43.11 (2006), pp. 3255–3291.

[57] Jean-Marc Bourinet, François Deheeger, and Maurice Lemaire. "Assessing small failure probabilities by combined subset simulation and Support Vector Machines". *Structural Safety* 33.6 (2011), pp. 343–353.

[58] Carl Edward Rasmussen and Christopher K I Williams. *Gaussian processes for machine learning*. Cambridge, Massachusetts: The MIT Press, 2006.

[59] Sajad Saraygord Afshari, Fatemeh Enayatollahi, Xiangyang Xu, and Xihui Liang. "Machine learning-based methods in structural reliability analysis: A review". *Reliability Engineering & System Safety* 219 (2022), p. 108223.

[60] Maliki Moustapha and Bruno Sudret. "Surrogate-assisted reliability-based design optimization: a survey and a unified modular framework". *Structural and Multidisciplinary Optimization* 60.5 (2019), pp. 2157–2176.

[61] Juan Alvarez Lopez, Kumaraswamy Ponnambalam, and Víctor H. Quintana. "Generation and transmission expansion under risk using stochastic programming". *IEEE Transactions on Power Systems* 22.3 (2007), pp. 1369–1378.

[62] Giuseppe Carlo Calafiore and Marco C. Campi. "The scenario approach to robust control design". *IEEE Transactions on Automatic Control* 51.5 (2006), pp. 742–753.

[63] James Luedtke and Shabbir Ahmed. "A sample approximation approach for optimization with probabilistic constraints". *SIAM Journal on Optimization* 19.2 (2008), pp. 674–699.

[64] Arkadi Nemirovski and Alexander Shapiro. "Convex approximations of chance constrained programs". *SIAM Journal on Optimization* 17.4 (2007), pp. 969–996.

[65] Dimitri Bertsekas and John N. Tsitsiklis. *Introduction to probability*. 2nd. New Hampshire: Athena Scientific, 2008.

[66] David Barber. *Bayesian reasoning and machine learning*. New York City, New York: Cambridge University Press, 2012.

[67] Kevin P. Murphy. *Machine learning: A probabilistic perspective*. Cambridge, Massachusetts: The MIT Press, 2012.

[68] Ronald W. Butler. *Saddlepoint approximations with applications*. Cambridge: Cambridge University Press, 2007.

[69] Donald R. Jones, Matthias Schonlau, and William J. Welch. "Efficient global optimization of expensive black-box functions". *Journal of Global Optimization* 13.4 (1998), pp. 455–492.

[70] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. "Taking the human out of the loop: A review of Bayesian optimization". *Proceedings of the IEEE* 104.1 (2016), pp. 148–175.

[71] Peter I. Frazier. "A tutorial on Bayesian optimization". *arXiv preprint* (2018).

[72] Harold J. Kushner. "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise". *Journal Basic Engineering* 86.1 (1964), pp. 97–106.

[73] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. "Gaussian process optimization in the bandit setting: No regret and experimental design". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. 2010, pp. 1015–1022.

[74] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The design and analysis of computer experiments*. 2nd ed. New York, NY: Springer, 2018.

[75] Victor Picheny, Tobias Wagner, and David Ginsbourger. "A benchmark of Kriging-based infill criteria for noisy optimization". *Structural and Multidisciplinary Optimization* 48.3 (2013), pp. 607–626.

[76] Deng Huang, Theodore T. Allen, William I. Notz, and Ning Zeng. "Global optimization of stochastic black-box systems via sequential Kriging meta-models". *Journal of Global Optimization* 34.3 (2006), pp. 441–466.

[77] Hsi Guang Sung. "Gaussian mixture regression and classification". PhD thesis. Rice University, 2004.

[78] Trevor Hastie, Robert Tibshirani, Jerome H. Friedman, and Jerome H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.

[79] Geoffrey J. McLachlan, Sharon X. Lee, and Suren I. Rathnayake. "Finite mixture models". *Annual Review of Statistics and Its Application* 6.1 (2019), pp. 355–378.

[80] Dimitrios I. Papadimitriou, Zissimos P. Mourelatos, and Zhen Hu. "Reliability analysis using second-order saddlepoint approximation and mixture distributions". *Journal of Mechanical Design* 141.2 (2018).

[81] Faicel Chamroukhi. "Unsupervised learning of regression mixture models with unknown number of components". *Journal of Statistical Computation and Simulation* 86.12 (2016), pp. 2308–2334.

[82] Geoffrey J. McLachlan and Suren Rathnayake. "On the number of components in a Gaussian mixture model". *WIREs Data Mining Knowl Discov* 4.5 (2014), pp. 341–355.

[83] Gang Fu, Frank Y. Shih, and Haimin Wang. "A kernel-based parametric method for conditional density estimation". *Pattern Recognition* 44.2 (2011), pp. 284–294.

[84] Colin S. Gillespie and Eric Renshaw. "An improved saddlepoint approximation". *Mathematical Biosciences* 208.2 (2007), pp. 359–374.

[85] Shuxiang Guo. "An efficient third-moment saddlepoint approximation for probabilistic uncertainty analysis and reliability evaluation of structures". *Applied Mathematical Modelling* 38.1 (2014), pp. 221–232.

[86] Makoto Ohsaki, Miyamura Miyamura, and Jingyao Zhang. "A piecewise linear isotropic-kinematic hardening model with semi-implicit rules for cyclic loading and its parameter identification". *Computer Modeling in Engineering & Sciences* 111.4 (2016), pp. 303–333.

[87] Amir-Hossein Mahmoudi, S. M. Pezeshki-Najafabadi, and Hojjat Badnava. "Parameter determination of Chaboche kinematic hardening model using a multi objective genetic algorithm". *Computational Materials Science* 50.3 (2011), pp. 1114–1122.

[88] Atri Nath, Kalyan Kumar Ray, and Sudhirkumar V. Barai. "Evaluation of ratcheting behaviour in cyclically stable steels through use of a combined kinematic-isotropic hardening rule and a genetic algorithm optimization technique". *International Journal of Mechanical Sciences* 152 (2019), pp. 138–150.

[89] A. Andrade-Campos, S. Thuillier, P. Pilvin, and F. Teixeira-Dias. "On the determination of material parameters for internal variable thermoelastic-viscoplastic constitutive models". *International Journal of Plasticity* 23.8 (2007), pp. 1349–1379.

[90] Bruno M. Chaparro, Sandrine Thuillier, Luís Filipe Menezes, Pierre Yves Manach, and Jose Valdemar Fernandes. "Material parameters identification: Gradient-based, genetic and hybrid optimization algorithms". *Computational Materials Science* 44.2 (2008), pp. 339–346.

[91] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. "A fast and elitist multiobjective genetic algorithm: NSGA-II". *IEEE transactions on evolutionary computation* 6.2 (2002), pp. 182–197.

[92] Søren Nymand Lophaven, Hans Bruun Nielsen, and Jacob Søndergaard. *DACE-A Matlab Kriging toolbox*. Tech. rep. Lyngby, Denmark: Informatics and Mathematical Modelling, Technical University of Denmark, 2002, p. 34.

[93] Mohamed A. Abido. "A novel multiobjective evolutionary algorithm for environmental/economic power dispatch". *Electric Power Systems Research* 65.1 (2003), pp. 71–81.

[94] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and V. G. da Fonseca. "Performance assessment of multiobjective optimizers: an analysis and review". *IEEE Transactions on Evolutionary Computation* 7.2 (2003), pp. 117–132.

[95] Michael T. M. Emmerich, Kyriakos C. Giannakoglou, and Boris Naujoks. "Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels". *IEEE Transactions on Evolutionary Computation* 10.4 (2006), pp. 421–439.

[96] Eckart Zitzler and Lothar Thiele. "Multiobjective optimization using evolutionary algorithms – A comparative case study". In: *Parallel Problem Solving from Nature – PPSN V*. Ed. by Agoston E. Eiben, Thomas Bäck, Marc Schoenauer, and Hans-Paul Schwefel. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 292–301.

[97] Nicola Beume, Boris Naujoks, and Michael Emmerich. "SMS-EMOA: Multiobjective selection based on dominated hypervolume". *European Journal of Operational Research* 181.3 (2007), pp. 1653–1669.

[98] Carlos M. Fonseca, Luís Paquete, and Manuel Lopez-Ibanez. "An improved dimension-sweep algorithm for the hypervolume indicator". In: *2006 IEEE International Conference on Evolutionary Computation*. 2006, pp. 1157–1163.

[99] Ivo Couckuyt, Dirk Deschrijver, and Tom Dhaene. "Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization". *Journal of Global Optimization* 60.3 (2014), pp. 575–594.

[100] Alexandre Mathern, Olof Skogby Steinholtz, Anders Sjöberg, Magnus Önnheim, Kristine Ek, Rasmus Rempling, Emil Gustavsson, and Mats Jirstrand. "Multi-objective constrained Bayesian optimization for structural design". *Structural and Multidisciplinary Optimization* 63.2 (2021), pp. 689–701.

[101] Ryan Roussel, Adi Hanuka, and Auralee Edelen. "Multiobjective Bayesian optimization for online accelerator tuning". *Physical Review Accelerators and Beams* 24.6 (2021), p. 062801.

[102] Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. "Multi-objective Bayesian optimization over high-dimensional search spaces". *arXiv preprint* (2021).

[103] Satoshi Yamada and Yu Jiao. "A concise hysteretic model of structural steel considering the Bauschinger effect". *International Journal of Steel Structures* 16.3 (2016), pp. 671–683.

[104] Dassault Systèmes. *Abaqus User's Manual Version 2017*. 2017.

[105] Bach Do and Makoto Ohsaki. "Bayesian optimization for inverse identification of cyclic constitutive law of structural steels from cyclic structural tests". *Structures* 38 (2022), pp. 1079–1097.

[106] MathWorks. *Global optimization toolbox (R2018a)*. Berkeley, California, USA, 2018.

[107] Joshua Knowles. "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems". *IEEE Transactions on Evolutionary Computation* 10.1 (2006), pp. 50–66.

[108] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

[109] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. "Solving inverse problems using data-driven models". *Acta Numerica* 28 (2019), pp. 1–174.

[110] Omar Ghattas and Karen Willcox. "Learning physics-based models from data: perspectives from inverse problems and model reduction". *Acta Numerica* 30 (2021), pp. 445–554.

[111] Dani Gamerman and Hedibert F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC press, 2006.

[112] Scott A. Sisson, Yanan Fan, and Mark Beaumont. *Handbook of approximate Bayesian computation*. CRC Press, 2018.

[113] Hu Wang, Yang Zeng, Xiancheng Yu, Guangyao Li, and Enying Li. "Surrogate-assisted Bayesian inference inverse material identification method and application to advanced high strength steel". *Inverse Problems in Science and Engineering* 24.7 (2016), pp. 1133–1161.

[114] Patxi Fernandez-Zelaia, V. Roshan Joseph, Surya R. Kalidindi, and Shreyes N. Melkote. "Estimating mechanical properties from spherical indentation using Bayesian approaches". *Materials & Design* 147 (2018), pp. 92–105.

[115] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P. H. Stumpf. "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems". *Journal of The Royal Society Interface* 6.31 (2009), pp. 187–202.

[116] Jean-Michel Marin, Pierre Pudlo, Christian P. Robert, and Robin J. Ryder. "Approximate Bayesian computational methods". *Statistics and Computing* 22.6 (2012), pp. 1167–1180.

[117] Brandon M. Turner and Trisha Van Zandt. "A tutorial on approximate Bayesian computation". *Journal of Mathematical Psychology* 56.2 (2012), pp. 69–85.

[118] Michael U. Gutmann and Jukka Corander. "Bayesian optimization for likelihood-free inference of simulator-based statistical models". *Journal of Machine Learning Research* 17.125 (2016), pp. 1–47.

[119] Marko Järvenpää, Michael U. Gutmann, Arijus Pleska, Aki Vehtari, and Pekka Marttinen. "Efficient acquisition rules for model-based approximate Bayesian computation". *Bayesian Analysis* 14.2 (2019), pp. 595–622.

[120] Heikki Haario, Marko Laine, Antonietta Mira, and Eero Saksman. "DRAM: Efficient adaptive MCMC". *Statistics and Computing* 16.4 (2006), pp. 339–354.

[121] Christopher C. Drovandi and Anthony N. Pettitt. "Likelihood-free Bayesian estimation of multivariate quantile distributions". *Computational Statistics & Data Analysis* 55.9 (2011), pp. 2541–2556.

[122] Miguel A. Carreira-Perpinan. "Mode-finding for mixtures of Gaussian distributions". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (2000), pp. 1318–1323.

[123] Sang Hoon Lee, Wei Chen, and Byung Man Kwak. "Robust design with arbitrary distributions using Gauss-type quadrature formula". *Structural and Multidisciplinary Optimization* 39.3 (2009), pp. 227–243.

[124] Tanmoy Chatterjee, Souvik Chakraborty, and Rajib Chowdhury. "A critical review of surrogate assisted robust design optimization". *Archives of Computational Methods in Engineering* 26.1 (2019), pp. 245–274.

[125] Xufang Zhang and Mahesh D. Pandey. "Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method". *Structural Safety* 43 (2013), pp. 28–40.

[126] Shahram Pezeshk, V. Camp Charles, and D. Chen. "Design of nonlinear framed structures using genetic optimization". *Journal of Structural Engineering* 126.3 (2000), pp. 382–388.

[127] Vedat Toğan. "Design of planar steel frames using teaching-learning based optimization". *Engineering Structures* 34 (2012), pp. 225–232.

[128] Mahmoud R. Maheri and M. M. Narimani. "An enhanced harmony search algorithm for optimum design of side sway steel frames". *Computers & Structures* 136 (2014), pp. 78–89.

[129] Pierre Dumonteil. "Simple equations for effective length factors". *Engineering Journal-American Institute of Steel Construction* 29.3 (1992), pp. 111–115.

[130] ANSI/AISC 341-16. *Seismic provisions for structural steel buildings*. Chicago, Illinois, USA, 2016.

[131] Ercan Doğan and Mehmet Polat Saka. "Optimum design of unbraced steel frames to LRFD-AISC using particle swarm optimization". *Advances in Engineering Software* 46.1 (2012), pp. 27–34.

[132] Pietro Belotti, Christian Kirches, Sven Leyffer, Jeff Linderoth, James Luedtke, and Ashutosh Mahajan. "Mixed-integer nonlinear optimization". *Acta Numerica* 22 (2013), pp. 1–131.

[133] Bach Do and Makoto Ohsaki. "A random search for discrete robust design optimization of linear-elastic steel frames under interval parametric uncertainty". *Computers & Structures* 249 (2021), p. 106506.

[134] Rui Xu and Donald Wunsch. "Survey of clustering algorithms". *IEEE Transactions on Neural Networks* 16.3 (2005), pp. 645–678.

[135] H. Liu, Y. S. Ong, X. Shen, and J. Cai. "When Gaussian process meets big data: A review of scalable GPs". *IEEE Transactions on Neural Networks and Learning Systems* 31.11 (2020), pp. 4405–4423.

[136] Chao Jiang, G. Y. Lu, Xu Han, and L. X. Liu. "A new reliability analysis method for uncertain structures with random and interval variables". *International Journal of Mechanics and Materials in Design* 8.2 (2012), pp. 169–182.

[137] Beiqing Huang and Xiaoping Du. "Probabilistic uncertainty analysis by mean-value first order saddlepoint approximation". *Reliability Engineering & System Safety* 93.2 (2008), pp. 325–336.

[138] Xu Li, Chunlin Gong, Liangxian Gu, Zhao Jing, Hai Fang, and Ruichao Gao. "A reliability-based optimization method using sequential surrogate model and Monte Carlo simulation". *Structural and Multidisciplinary Optimization* 59.2 (2019), pp. 439–460.

[139] Weitao Zhao and Zhiping Qiu. "An efficient response surface method and its application to structural reliability and reliability-basedoptimization". *Finite Elements in Analysis and Design* 67 (2013), pp. 34–42.

[140] MathWorks. *Optimization toolbox (R2018a)*. Berkeley, California, USA, 2018.

[141] Yanglin Gong, Yusong Xue, Lei Xu, and Donald E. Grierson. "Energy-based design optimization of steel building frameworks using nonlinear response history analysis". *Journal of Constructional Steel Research* 68.1 (2012), pp. 43–50.

[142] Anil K. Chopra. *Dynamics of structures: Theory and applications to earthquake engineering*. 4th. Hoboken, New Jersey: Prentice Hall, 2013.

[143] Farzin Zareian and Ricardo A. Medina. "A practical method for proper modeling of structural damping in inelastic plane structural systems". *Computers & Structures* 88.1 (2010), pp. 45–53.

[144] Dimitrios G. Lignos and Helmut Krawinkler. "Deterioration modeling of steel components in support of collapse prediction of steel moment frames under earthquake loading". *Journal of Structural Engineering* 137.11 (2011), pp. 1291–1302.

[145] Zhifeng Liu, Sez Atamturktur, and C Hsein Juang. "Reliability based multi-objective robust design optimization of steel moment resisting frame considering spatial variability of connection parameters". *Engineering Structures* 76 (2014), pp. 393–403.

[146] PEERC. *PEER strong ground motion databases*. "`https://peer.berkeley.edu/peer-strong-ground-motion-databases`". [Accessed February 5, 2021]. Berkeley, California, USA, Feb. 2021.

[147] Juan C Reyes and Erol Kalkan. "How many records should be used in an ASCE/SEI-7 ground motion scaling procedure?" *Earthquake Spectra* 28.3 (2012), pp. 1223–1242.

[148] MathWorks. *Statistics and machine learning toolbox user's guide (R2019b)*. [Accessed October 20, 2019]. Berkeley, California, USA, Oct. 2019.

[149] PEERC. *Open System for Earthquake Engineering Simulation (OpenSees)*. "`https://opensees.berkeley.edu/`". [Accessed December 10, 2020]. Berkeley, California, USA, 2020.

[150] Charles E. Antoniak. "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems". *The Annals of Statistics* 2.6 (1974), pp. 1152–1174.

[151] Rommel G. Regis and Christine A. Shoemaker. "A stochastic radial basis function method for the global optimization of expensive functions". *INFORMS Journal on Computing* 19.4 (2007), pp. 497–509.

[152] Jean Lemaitre and Jean-Louis Chaboche. *Mechanics of solid materials*. Cambridge University Press, 1994.

[153] E. Voce. "The relationship between stress and strain for homogeneous deformation". *Journal of the Institute of Metals* 74 (1948), pp. 537–562.

[154] Jean-Louis Chaboche and Gilles Rousselier. "On the plastic and viscoplastic constitutive equations–Part I: Rules developed with internal variable concept". *Journal of Pressure Vessel Technology* 105.2 (1983), pp. 153–158.

[155] Peter J. Armstrong and Charles O. Frederick. *A mathematical representation of the multiaxial Bauschinger effect*. Report RD/B/N731. Berkeley, UK: CEGB, Central Electricity Generating Board, 1966.

Publications

Peer-reviewed articles associated with this dissertation

[1] Bach Do and Makoto Ohsaki. "Proximal exploration multi-objective Bayesian optimization for inverse identification of cyclic constitutive law of structural steels". *Structural and Multidisciplinary Optimization* 65(7) (2022), p. 199.

[2] Bach Do and Makoto Ohsaki. "Bayesian optimization for inverse identification of cyclic constitutive law of structural steels from cyclic structural tests". *Structures* 38 (2022), pp. 1079–1097.

[3] Bach Do and Makoto Ohsaki. "Sequential sampling approach to energy-based multiobjective design optimization of steel frames with correlated random parameters". *Earthquake Engineering & Structural Dynamics* 51(3) (2022), pp. 588–611.

[4] Bach Do, Makoto Ohsaki and Makoto Yamakawa. "Bayesian optimization for robust design of steel frames with joint and individual probabilistic constraints". *Engineering Structures* 245 (2021), p. 112859.

[5] Bach Do, Makoto Ohsaki and Makoto Yamakawa. "Sequential mixture of Gaussian processes and saddlepoint approximation for reliability-based design optimization of structures". *Structural and Multidisciplinary Optimization* 64(2) (2021), pp. 625–648.

[6] Bach Do and Makoto Ohsaki. "Gaussian mixture model for robust design optimization of planar steel frames". *Structural and Multidisciplinary Optimization* 63(1) (2021), pp. 137–160.

Oral presentations associated with this dissertation

[1] Bach Do and Makoto Ohsaki. "Bayesian optimization for inverse identification of cyclic constitute law of structural steels". In *the 15th World Congress on Computational Mechanics & 8th Asian Pacific Congress on Computational Mechanics (WCCM-APCOM 2022)*, Yokohama, Japan (2022), No. 557.

[2] Bach Do and Makoto Ohsaki. "Inverse identification of cyclic constitutive law of structural steels using multi-objective Bayesian optimization". In *the 17th East Asia-Pacific Conference on Structural Engineering and Construction (EASEC-17)*, Singapore (2022), ID 17.

[3] Bach Do and Makoto Ohsaki. "Sequential batch sampling approach to discrete multi-objective reliability-based design optimization". In *Asian Congress of Structural and Multidisciplinary Optimization 2022 (ACSMO2022)*, Matsue, Japan (2022), pp.158, No. 2F1−3.

[4] Bach Do and Makoto Ohsaki. "A novel saddlepoint approximation and its application to structural reliability-based design optimization". In *the 14th World Congress of Structural and Multidisciplinary Optimization (WCSMO14)*, Boulder, Colorado, USA (2021), ID 58.

[5] Bach Do and Makoto Ohsaki. "A Gaussian mixture model for multi-objective robust design optimization of steel frames under aleatory uncertainty". In *the 3rd International Conference on Computational Engineering and Science for Safety and Environmental Problems (COMPSAFE2020)*, Kobe, Japan (2020), pp. 394, No. MS13−3−03.

[6] Bach Do and Makoto Ohsaki. "Multi-objective robust design optimization of planar steel frames using a Gaussian mixture model". In *Summaries of technical papers of annual meeting (Structure I)*. Architectural Institute of Japan (AIJ), Kanto, Japan (2020), pp.283−284, No. 20142.

Other peer-reviewed articles

[1] Bach Do and Makoto Ohsaki. "A random search for discrete robust design optimization of linear-elastic steel frames under interval parametric uncertainty". *Computers & Structures* 249 (2021), p. 106506.

[2] Makoto Ohsaki, Bach Do, Jun Fujiwara, Toshiaki Kimura and Takuzo Yamashita. "Two-step parameter identification of multi-axial cyclic constitutive law of structural steels from cyclic structural responses". *Structures* 46 (2022), pp. 2014–2030.

[3] Bach Do and Akhrawat Lenwari. "Optimization of fiber-reinforced polymer patches for repairing fatigue cracks in steel plates using a genetic algorithm". *Journal of Composites for Construction (ASCE)* 24(2) (2020), p. 04020006.

[4] Thu Van Huynh, Sawekchai Tangaramvonga, Bach Do, Wei Gao and Suchart Limkatanyu. "Sequential most probable point update combining Gaussian process and comprehensive learning PSO for structural reliability-based design optimization". *Reliability Engineering & System Safety* (2022) – in revision.

Acknowledgments

This dissertation has been done when I am a research student and then a doctoral student at Ohsaki & Zhang Laboratory, the Department of Architecture and Architectural Engineering, Kyoto University. I gratefully acknowledge the financial support over the years from the Japan International Cooperation Agency (JICA) project for the ASEAN University Network/Southeast Asia Engineering Education Development Network (AUN/SEED-Net).

Above all, I would like to express my deepest gratitude to my academic advisor, Professor Makoto Ohsaki, for his guidance, support, and encouragement. His expertise in the field of structural optimization and his supporting advising style not only encourage me in obtaining more knowledge skills to shape my understanding of any problem, give me the freedom in finding my own inclinations but also teach me how to carefully prepare for an academic career starting from constructing a solid foundation of knowledge.

I would like to thank the committee members, Professor Yoshiki Ikeda and Associate Professor Kohei Fujita, for spending their precious time reading, discussing, and providing constructive feedback on the contents of my dissertation and my presentation.

I am grateful to Professor Makoto Yamakawa at Tokyo University of Science for his fruitful comments that have contributed to the formation of approaches in Chapters 5–7. I would like to thank Associate Professor Jingyao Zhang for his discussions of important aspects related to steel frame designs. I would also like to send many thanks to Dr. Kazuki Hayashi for his assistance in every aspect of my "laboratory life". I appreciate Professor Satoshi Yamada at The University of Tokyo for his generosity in providing the experimental results I used in Chapters 3 and 4. I am grateful to my Master's advisor, Professor Akhrawat Lenwari at Chulalongkorn University, for guiding me through the very first steps of how to conduct a research project properly.

I want to thank the former and current members of Ohsaki & Zhang Laboratory, especially, to Dr. Wei Shen for his thorough discussion of the approaches I proposed in this dissertation as we are both interested in structural optimization under uncertainty, to Kentaro Hayakawa for his support in setting up and configuring my PC and for his diligence to keep everything in room 315 organized, to Chi-tathon Kupwiwat for sending me some presentation materials of doctor seminars in the last year of my doctoral study, and to Kazuma Sakaguchi and Kosuke Kogiso for taking me around the Katsura Campus when I first came there. I thank Safumi Saiki for her administrative assistance that has helped me to totally focus on my research over the years.

My gratitude also goes to Kento Hirai from JICA for getting everything well-prepared when I arrived in Japan in fall 2019, to Rika Aoki from JICA for her countless support in the first two years of my study and for the towel as a gift she gave me when she came to the Katsura Campus for our monitoring meeting in winter 2021, and to Takuya Ohara from JICA for scheduling my trip back.

Three and half years of my life in Kyoto are full of great experiences from day trips, thanks to Manoj Pradhan, Indra Bahadur Chhetr, and Thu Dao. I would also like to thank Dr. Duy Vo for taking me around Tokyo on the very first days of the year 2020.

Finally, my heartfelt thanks go to my parents, my younger brothers, my nephews, my grandma, and my dear girlfriend Tuyet for their wholehearted love, their understanding, and their calls every day over the years just to have a virtual dinner together. I dedicate this dissertation to them.

Kyoto, January 20, 2023