Structural and Multidisciplinary Optimization (2021) 63:2425–2457
https://doi.org/10.1007/s00158-020-02820-z

RESEARCH PAPER

Robust design optimization under dependent random variables by a generalized polynomial chaos expansion

Dongjin Lee[1] · **Sharif Rahman**[1]

Received: 5 August 2020 / Revised: 10 November 2020 / Accepted: 9 December 2020 / Published online: 1 March 2021

Abstract

New computational methods are proposed for robust design optimization (RDO) of complex engineering systems subject to input random variables with arbitrary, dependent probability distributions. The methods are built on a generalized polynomial chaos expansion (GPCE) for determining the second-moment statistics of a general output function of dependent input random variables, an innovative coupling between GPCE and score functions for calculating the second-moment sensitivities with respect to the design variables, and a standard gradient-based optimization algorithm, establishing direct GPCE, single-step GPCE, and multi-point single-step GPCE design processes. New analytical formulae are unveiled for design sensitivity analysis that is synchronously performed with statistical moment analysis. Numerical results confirm that the proposed methods yield not only accurate but also computationally efficient optimal solutions of several mathematical and simple RDO problems. Finally, the success of conducting stochastic shape optimization of a steering knuckle demonstrates the power of the multi-point single-step GPCE method in solving industrial-scale engineering problems.

Keywords RDO · Second-moment analysis · GPCE · Design sensitivity analysis · Score functions · Stochastic optimization

1 Introduction

Robust design optimization, commonly referred to as RDO, is a prime exemplar for engineering design in the presence of uncertainty (Taguchi 1993). Unlike a conservative design optimization using heuristically derived safety factors, RDO manages risks explicitly by propagating input uncertainties to the objective and constraint functions, eventually leading to insensitive designs. The success of RDO is well documented in many real-world applications, such as those found in the design of aerospace, automotive, civil, and electronic structures, systems, or devices (Chen et al. 1996; Du and Chen 2000; Mourelatos and Liang 2006; Zaman et al. 2011; Park et al. 2006; Yao et al. 2011; Ren and Rahman 2013; Chatterjee et al. 2019).

In practical applications, the objective or constraint functions are often described algorithmically via finite-element analysis (FEA), which is generally expensive. Therefore, numerous surrogate methods for RDO, encompassing Taylor series or perturbation expansion (Sundaresan et al. 1995), the point estimate method (Huang and Du 2007), polynomial chaos expansion (PCE) (Shen and Braatz 2016), the tensor-product quadrature rule (Lee et al. 2009), dimension-reduction methods (Lee et al. 2009), the polynomial dimensional decomposition (PDD) method (Ren and Rahman 2013), Kriging (Jin et al. 2003), and artificial neural network (Chatterjee et al. 2019) have appeared. Unfortunately, the foregoing methods, including many others not listed here for brevity, are largely predicated on the independence assumption of input random variables. In reality, there may exist significant correlation or dependence among input variables, hindering or invalidating most existing RDO methods available today. Indeed, ignoring these correlations or dependencies, whether emanating from loads, material properties, or manufacturing variables, may produce inaccurate or inadequate designs (Noh et al. 2009). The authors rule out Rosenblatt transformation (Rosenblatt 1952) or Nataf transformation (Nataf 1962), commonly used for mapping dependent to independent variables, as they may induce overly large nonlinearity to a stochastic

Responsible Editor: Byeng D Youn

✉ Dongjin Lee
dongjin-lee@uiowa.edu

[1] Department of Mechanical Engineering, The University of Iowa, Iowa City, IA 52242, USA

response, potentially degrading the convergence properties of probabilistic solutions (Rahman 2009a). Therefore, the existing methods must be generalized or new methods should be developed from scratch for uncertainty quantification (UQ) analysis and subsequent design optimization when dealing with dependent or correlated input random variables.

A few additional studies on design optimization under uncertainty using a copula-based approach for dependent variables have been reported (Noh et al. 2009; Lee et al. 2011). In particular, Noh et al. (2009) selected the Gaussian copula to describe the dependence between random variables. While such copula and other available variants facilitate a practical way to deal with correlated variables, finding the right copula when the joint distribution of random variables is arbitrary but unknown is highly nontrivial. More often than not, a gradient-based method, such as sequential quadratic programming, is used to solve the underlying optimization problem. This is mainly because it provides fast convergence and an efficient way to integrate optimization and stochastic simulation. However, there are multiple choices for optimization, such as evolutionary algorithms (Cramer et al. 2008) and swarm algorithms (Ono et al. 2009), to name a few.

Recently, the authors introduced a practical version of the generalized polynomial chaos expansion (GPCE) for UQ analysis under arbitrary, dependent probability distribution of input random variables (Lee and Rahman 2020).[1] A remarkable feature of this work, in contrast to the prequel (Rahman 2018), is that the multivariate orthonormal polynomial basis functions consistent with any non-product-type probability measure of input random variables can be generated without the need for a Rodrigues-type formula. However, the aforementioned GPCE is limited to forward UQ analysis only. As a result, there remain three important challenges for GPCE to address RDO problems: (1) how to simultaneously determine design sensitivities with statistical moments of output functions for a given design with no added computational cost, (2) how to sidestep repetitive calculations of statistical moments and design sensitivities to the extent possible during design iterations, and (3) how to markedly reduce the number of function evaluations or FEA in conjunction with standard gradient-based optimization algorithms for problems with large design spaces. Only by tackling these challenges successfully will the GPCE method be further strengthened to effectively solve RDO problems subject to dependent input variables.

The overarching goal of this work is to build a solid theoretical foundation, accompanied by robust numerical algorithms, for UQ analysis and design optimization of complex systems subject to random input following an arbitrary dependent probability measure. In the context of RDO, three new design methods, aimed at solving simple to complex problems, are proposed to meet the goal. They are premised on (1) a GPCE for determining the second-moment statistics of a general output function of dependent input random variables; (2) an innovative coupling between GPCE and score functions for calculating the second-moment sensitivities with respect to the design variables; and (3) a standard gradient-based optimization algorithm, encompassing direct GPCE, single-step GPCE, and multi-point single-step GPCE design processes.

The paper is organized as follows. In Section 2, a general RDO problem is formally defined with associated mathematical statements. In Section 3, a brief exposition of GPCE is provided, including a three-step algorithm to construct arbitrary measure-consistent multivariate orthonormal polynomial basis and two regression techniques to estimate the expansion coefficients. In Section 4, score functions are defined and new closed-form formulae for design sensitivities of statistical moments are disclosed. In Section 5, three new design methods, integrating the stochastic and sensitivity analyses and using a standard gradient-based optimization, are introduced. In Section 6, four numerical examples, ranging from simple mathematical functions to an industrial-scale engineering problem, are dealt with to investigate the accuracy, convergence properties, and computational efforts of all three methods. In Section 7, the novelty of this work and the efficiency and relevance of proposed design methods are discussed. Finally, in Section 8, the conclusions are drawn.

2 Robust design optimization

Let $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$, and $\mathbb{R}_0^+$ be the sets of positive integer, non-negative integer, real number, and non-negative real number, respectively. For a positive integer $N \in \mathbb{N}$, denote by $\mathbb{R}^N$ the N-dimensional real vector space. Finally, denote by $\mathbb{A}^N \subseteq \mathbb{R}^N$ and $\bar{\mathbb{A}}^N \subseteq \mathbb{R}^N$ two bounded or unbounded domains.

Consider a measurable space $(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}})$, where $\Omega_{\mathbf{d}}$ is a sample space and $\mathcal{F}_{\mathbf{d}}$ is a σ-field on $\Omega_{\mathbf{d}}$. Defined over $(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}})$, let $\{\mathbb{P}_{\mathbf{d}} : \mathcal{F}_{\mathbf{d}} \to [0,1]\}$ be a family of probability measures where, for $M \in \mathbb{N}$ and $N \in \mathbb{N}$, $\mathbf{d} = (d_1, \cdots, d_M)^T \in \mathcal{D}$ is an M-dimensional design vector with non-empty closed set $\mathcal{D} \subset \mathbb{R}^M$. Here, $\mathbf{X} := (X_1, \cdots, X_N)^T : (\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}}) \to (\mathbb{A}^N, \mathcal{B}^N)$ is an $\mathbb{A}^N$-valued input random vector with $\mathcal{B}^N$ representing the Borel σ-field on $\mathbb{A}^N$, describing the statistical uncertainties in loads, material properties, and geometry of a complex mechanical system. The probability law of $\mathbf{X}$ is completely defined by a family of the joint probability density functions

[1] In contrast to the existing GPCE (Xiu and Karniadakis 2002), which accounts for only independent random variables, the authors' GPCE can handle dependent as well as independent random variables with arbitrary probability distributions.

(PDF) $\left\{ f_{\mathbf{X}}(\mathbf{x};\mathbf{d}) : \mathbf{x} \in \mathbb{R}^N,\ \mathbf{d} \in \mathcal{D} \right\}$ that are associated with probability measures $\{\mathbb{P}_{\mathbf{d}} : \mathbf{d} \in \mathcal{D}\}$, so that the probability triple $(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}}, \mathbb{P}_{\mathbf{d}})$ of $\mathbf{X}$ depends on $\mathbf{d}$. In theory, a design variable d_k can be any distribution parameter or a statistic; however, here, d_k is limited to the mean of random variable X_k. Indeed, the design parameters as mean values are commonly used in almost all engineering problems.

2.1 Problem definition

Let $y_l(\mathbf{X}) := y_l(X_1, \ldots, X_N)$, $l = 0, 1, \ldots, K$, represent a collection of $K + 1$ real-valued, square-integrable, measurable transformations on $(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}})$, describing output functions of a complex system. They are commonly referred to as response or performance functions in applications. It is assumed that $y_l : (\mathbb{A}^N, \mathcal{B}^N) \to (\mathbb{R}, \mathcal{B})$ is not an explicit function of $\mathbf{d}$, although y_l implicitly depends on $\mathbf{d}$ via the probability law of $\mathbf{X}$. This is not a major limitation, as most, if not all, RDO problems involve means of random variables as design variables. In addition, let $\mathcal{D} = \times_{k=1}^{M} [d_{k,L}, d_{k,R}]$ be a closed rectangular subdomain of $\mathbb{R}^M$. From a fundamental standpoint, RDO is performed by minimizing the mean and standard deviation of the performance individually. It generally leads to a bi-objective optimization problem, demanding one to

$$\begin{aligned} \min_{\mathbf{d} \in \mathcal{D} \subseteq \mathbb{R}^M} \quad & \{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})],\ \sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}\}, \\ \text{subject to} \quad & \alpha_l \sqrt{\mathrm{var}_{\mathbf{d}}[y_l(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X})] \leqslant 0, \\ & l = 1, \ldots, K,\ 1 \leqslant K < \infty \\ & d_{k,L} \leqslant d_k \leqslant d_{k,U},\ k = 1, \ldots, M, \end{aligned}$$

where

$$\mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X})] := \int_{\mathbb{A}^N} y_l(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x};\mathbf{d}) d\mathbf{x}$$

is the mean of $y_l(\mathbf{X})$ and

$$\mathrm{var}_{\mathbf{d}}[y_l(\mathbf{x})] := \mathbb{E}_{\mathbf{d}}\left[y_l(\mathbf{X}) - \mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X})] \right]^2$$

is the variance of $y_l(\mathbf{X})$. Here, $\mathbb{E}_{\mathbf{d}}$ and $\mathrm{var}_{\mathbf{d}}$ are the expectation and variance operators, respectively, with respect to the probability measure $\mathbb{P}_{\mathbf{d}}$ or $f_{\mathbf{X}}(\mathbf{x};\mathbf{d})d\mathbf{x}$; $\alpha_l \in \mathbb{R}_0^+$, $l = 1, \ldots, K$, are non-negative, real-valued constants associated with the probabilities of constraint satisfaction; and $d_{k,L}$ and $d_{k,U}$ are the lower and upper bounds of the kth design variable d_k.

In many realistic cases, the bi-objective optimization problem may require to make optimal decisions in the presence of trade-offs between two conflicting objective functions $\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]$ and $\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}$. In that case, there exist an infinite number of optimal solutions, typically called Pareto optimal solutions, where none of the objective function values can be amended without deteriorating the other. To find either multiple Pareto optimal solutions or a single solution that meets the preferences of a decision-maker, the commonly used scalarization approach transforms the bi-objective optimization problem into a single-objective optimization problem. Representative scalarization approaches include weighted-sum approach (Marler and Arora 2010), ϵ-constraint approach (Bashiri et al. 2020), weighted-Tchebycheff approach (Chen et al. 1999; Shin and Cho 2008), goal programming (Nha et al. 2013), and lexicographic approach (Bhushan et al. 2008), and others (Miettinen 2012). In this study, any choice of scalarization approaches is applicable to solve the bi-objective optimization problem.

2.2 Proposed formulations

Two mathematical formulations of RDO—one expressed with respect to the original input random variables and the other described with respect to transformed input random variables—are presented in the remainder of this section. The formulations are equivalent, that is, they yield identical solutions to a design optimization problem. However, the latter is more beneficial than the former in light of GPCE approximations, as will be discussed in forthcoming sections.

2.2.1 Original formulation

The mathematical formulation for RDO in most engineering applications involving an objective function $c_0 : \mathcal{D} \to \mathbb{R}$ and constraint functions $c_l : \mathcal{D} \to \mathbb{R}$, where $l = 1, \ldots, K$ and $1 \leqslant K < \infty$, requires one to (Chen et al. 1996; Du and Chen 2000; Ren and Rahman 2013)

$$\begin{aligned} \min_{\mathbf{d} \in \mathcal{D} \subseteq \mathbb{R}^M} \quad & c_0(\mathbf{d}) := G\left(\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})],\ \sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]} \right), \\ \text{subject to} \quad & c_l(\mathbf{d}) := \alpha_l \sqrt{\mathrm{var}_{\mathbf{d}}[y_l(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X})] \leqslant 0, \\ & l = 1, \ldots, K, \\ & d_{k,L} \leqslant d_k \leqslant d_{k,U}, k = 1, \ldots, M, \end{aligned} \tag{1}$$

where $G(\cdot,\cdot)$ is an arbitrary function determined by the choice of scalarization. Two commonly used variants of the scalarized objective function are illustrated as follows.

For the first example, the weighted sum approach presents a linear aggregation of the objectives, yielding

$$\begin{aligned} & G\left(\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})],\ \sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]} \right) \\ & := w_1 \frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]}{\mu_0^*} + w_2 \frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}}{\sigma_0^*}, \end{aligned}$$

where $w_1 \in \mathbb{R}_0^+$ and $w_2 \in \mathbb{R}_0^+$ are two non-negative, real-valued weights such that $w_1 + w_2 = 1$; $\mu_0^* \in \mathbb{R}\backslash\{0\}$ and $\sigma_0^* \in \mathbb{R}_0^+\backslash\{0\}$ are two non-zero, real-valued scaling factors.

For the second example, the weighted Tchebycheff approach requires a reference point

$$(\mu_f^* := \mathbb{E}_{\mathbf{d}_1^*}[y_0(\mathbf{X})],\ \sigma_f^* := \sqrt{\mathrm{var}_{\mathbf{d}_2^*}[y_0(\mathbf{X})]})$$

where $\mathbf{d}_1^* = \arg\min_{\mathbf{d}} \mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]$ subject to $c_l(\mathbf{d}) \leqslant 0$ from (1) and $\mathbf{d}_2^* = \arg\min_{\mathbf{d}} \sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}$ subject to $c_l(\mathbf{d}) \leqslant 0$ from (1). Then

$$G\left(\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})],\ \sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}\right) := \max\left(w_1 \frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})] - \mu_f^*}{\mu_0^*},\ w_2 \frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]} - \sigma_f^*}{\sigma_0^*}\right).$$

In both scalarization choices, equal weights are usually chosen, but they can be distinct and biased, depending on the objective set forth by a designer. By contrast, the scaling factors are relatively arbitrary and chosen to better condition, such as normalize, the objective function.

2.2.2 Alternative formulation

Since the design variables are the means of some or all input random variables, a linear transformation, such as shifting or scaling of random variables, provides an alternative formulation of RDO. To do so, let $(X_{i_1}, \ldots, X_{i_M})^\intercal$ be an M-dimensional sub-vector of $\mathbf{X} := (X_1, \ldots, X_N)^\intercal$, $1 \leqslant i_1 \leqslant \cdots \leqslant i_M \leqslant N$, $M \leqslant N$, such that the means of its components are M design variables. In other words, $\mathbb{E}_{\mathbf{d}}[X_{i_k}] = d_k, k = 1, \ldots, M$.

Shifting Let $\mathbf{Z} := (Z_1, \ldots, Z_N)^\intercal$ be an N-dimensional vector of new random variables obtained by shifting $\mathbf{X}$ as

$$\mathbf{Z} = \mathbf{X} + \mathbf{r}, \tag{2}$$

where $\mathbf{r} := (r_1, \ldots, r_N)^\intercal$ is an N-dimensional vector of deterministic variables. Define $g_i := \mathbb{E}_{\mathbf{d}}[Z_i]$ as the mean of the ith component of $\mathbf{Z}$. Denote by $(Z_{i_1}, \ldots, Z_{i_M})^\intercal$ a subvector of $\mathbf{Z}$, where the i_kth new random variable Z_{i_k} corresponds to the i_kth original random variable X_{i_k}. From the shifting transformation, the mean of Z_{i_k} is

$$\mathbb{E}_{\mathbf{d}}[Z_{i_k}] = d_k + r_{i_k} = g_k$$

and the PDF of $\mathbf{Z}$ is

$$f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g}) = |\mathbf{J}|\, f_{\mathbf{X}}(\mathbf{x}; \mathbf{d}) = f_{\mathbf{X}}(\mathbf{x}; \mathbf{d}) = f_{\mathbf{X}}(\mathbf{z} - \mathbf{r}; \mathbf{d}),$$

supported on $\bar{\mathbb{A}}^N \subseteq \mathbb{R}^N$ (say). Here, the absolute value of the determinant of the Jacobian matrix is $|\mathbf{J}| = |\det[\partial \mathbf{x}/\partial \mathbf{z}]| = 1$ and the M-dimensional vector $\mathbf{g} := (g_1, \ldots, g_M)^\intercal$ has its kth component $g_k = \mathbb{E}_{\mathbf{d}}[Z_{i_k}], k = 1, \ldots, M$.

Scaling Let $\mathbf{Z} := (Z_1, \ldots, Z_N)^\intercal$ be an N-dimensional vector of new random variables obtained by scaling $\mathbf{X}$ as

$$\mathbf{Z} = \mathrm{diag}[r_1, \ldots, r_N]\mathbf{X}, \tag{3}$$

where $\mathbf{r} := (r_1, \ldots, r_N)^\intercal$ is an N-dimensional vector of deterministic variables. Define $g_i := \mathbb{E}_{\mathbf{d}}[Z_i]$ as the mean of the ith component of $\mathbf{Z}$. Denote by $(Z_{i_1}, \ldots, Z_{i_M})^\intercal$ a subvector of $\mathbf{Z}$, where the i_kth new random variable Z_{i_k} corresponds to the i_kth original random variable X_{i_k}. From the scaling transformation, the mean of Z_{i_k} is

$$\mathbb{E}_{\mathbf{d}}[Z_{i_k}] = d_k r_{i_k} = g_k$$

and the PDF of $\mathbf{Z}$ is

$$f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g}) = |\mathbf{J}|\, f_{\mathbf{X}}(\mathbf{x}; \mathbf{d}) = \left|\frac{1}{r_1 \ldots r_N}\right| f_{\mathbf{X}}(\mathbf{x}; \mathbf{d}) = \left|\frac{1}{r_1 \ldots r_N}\right| f_{\mathbf{X}}(\mathrm{diag}[1/r_1, \ldots, 1/r_N]\mathbf{z}; \mathbf{d}),$$

supported on $\bar{\mathbb{A}}^N \subseteq \mathbb{R}^N$ (say). Here, the absolute value of the determinant of the Jacobian matrix is $|\mathbf{J}| = |\det[\partial \mathbf{x}/\partial \mathbf{z}]| = |1/(r_1 \ldots r_N)|$ and the M-dimensional vector $\mathbf{g} := (g_1, \ldots, g_M)^\intercal$ has its kth component $g_k = \mathbb{E}_{\mathbf{d}}[Z_{i_k}], k = 1, \ldots, M$.

For each $l = 1, 2, \ldots, K$, define $h_l(\mathbf{Z}; \mathbf{r}) := y_l(\mathbf{X})$ to be the generic output function of the new random variables $\mathbf{Z}$, where the relation between $\mathbf{Z}$ and $\mathbf{X}$ is obtained by either shifting transformation in (2) or scaling transformation in (3). In both cases, the RDO formulation requires one to

$$\begin{aligned} \min_{\mathbf{d} \in \mathcal{D} \subseteq \mathbb{R}^M}\ & c_0(\mathbf{d}) := G\left(\mathbb{E}_{\mathbf{g}(\mathbf{d})}[h_0(\mathbf{Z}; \mathbf{r})],\ \sqrt{\mathrm{var}_{\mathbf{g}(\mathbf{d})}[h_0(\mathbf{Z}; \mathbf{r})]}\right), \\ \text{subject to}\ & c_l(\mathbf{d}) := \alpha_l \sqrt{\mathrm{var}_{\mathbf{g}(\mathbf{d})}[h_l(\mathbf{Z}; \mathbf{r})]} - \mathbb{E}_{\mathbf{g}(\mathbf{d})}[h_l(\mathbf{Z}; \mathbf{r})] \leqslant 0, \\ & l = 1, \ldots, K, \\ & d_{k,L} \leqslant d_k \leqslant d_{k,U},\ k = 1, \ldots, M, \end{aligned} \tag{4}$$

where

$$\mathbb{E}_{\mathbf{g}(\mathbf{d})}[h_l(\mathbf{Z}; \mathbf{r})] := \int_{\bar{\mathbb{A}}^N} h_l(\mathbf{z}; \mathbf{r}) f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g}) d\mathbf{z}$$

is the mean of $h_l(\mathbf{Z}; \mathbf{r})$ and

$$\mathrm{var}_{\mathbf{g}(\mathbf{d})}[h_l(\mathbf{Z}; \mathbf{r})] := \mathbb{E}_{\mathbf{g}(\mathbf{d})}\left[h_l(\mathbf{Z}; \mathbf{r}) - \mathbb{E}_{\mathbf{g}(\mathbf{d})}[h_l(\mathbf{Z}; \mathbf{r})]\right]^2$$

is the variance of $h_l(\mathbf{Z}; \mathbf{r})$. Here, $\mathbb{E}_{\mathbf{g}(\mathbf{d})}$ and $\mathrm{var}_{\mathbf{g}(\mathbf{d})}$ are the expectation and variance operators, respectively, with respect to the probability measure $f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})d\mathbf{z}$, which depends on $\mathbf{d}$. For brevity, the subscript "$\mathbf{g}(\mathbf{d})$" of the expectation operator will be denoted by "$\mathbf{g}$" in the rest of the paper.

The alternative formulation in (4) is simply a rephrasing of (1), but it is now expressed in terms of the transformed

input random variables $\mathbf{Z}$. In doing so, the probability measure of $\mathbf{Z}$ is fixed during design iterations, thus avoiding the need to recalculate measure-associated quantities. For the rest of the paper, the solution of an RDO problem will be described with respect to the alternative formulation. In addition, $\mathbf{X}$ or $\mathbf{Z}$ and y_l or h_l will be referred to, interchangeably, as input random vector and output function, respectively.

2.2.3 Construction of sub-problems

A gradient-based solution to the RDO problem in (4) mandates adequate smoothness in objective and constraint functions. Therefore, both functions are assumed to be differentiable with respect to design variables. Moreover, as these functions are generally nonlinear, iterative approximations of (4), resulting in a sequence of RDO sub-problems, are required.

Let $q = 1, 2, \ldots, Q$, $Q \in \mathbb{N}$, be a design iteration count describing the qth RDO sub-problem for (4). Given q, denote by $\mathbf{d}^{\{q\}}$, $\mathbf{g}^{\{q\}}$, and $\mathbf{r}^{\{q\}}$ the qth iterative versions of $\mathbf{d}$, $\mathbf{g}$, and $\mathbf{r}$, respectively. Then, the qth RDO sub-problem asks to

$$\begin{aligned}
\min_{\mathbf{d}^{\{q\}} \in \mathcal{D} \subseteq \mathbb{R}^M} \; & c_0^{\{q\}}(\mathbf{d}^{\{q\}}) := T\left[G\left(\mathbb{E}_{\mathbf{g}^{\{q\}}}[h_0(\mathbf{Z}; \mathbf{r}^{\{q\}})], \right.\right. \\
& \left.\left. \sqrt{\mathrm{var}_{\mathbf{g}^{\{q\}}}[h_0(\mathbf{Z}; \mathbf{r}^{\{q\}})]} \right) \right], \\
\text{subject to } & c_l^{\{q\}}(\mathbf{d}^{\{q\}}) := T\left[\alpha_l \sqrt{\mathrm{var}_{\mathbf{g}^{\{q\}}}[h_l(\mathbf{Z}; \mathbf{r}^{\{q\}})]} \right. \\
& \left. - \mathbb{E}_{\mathbf{g}^{\{q\}}}[h_l(\mathbf{Z}; \mathbf{r}^{\{q\}})] \right] \leqslant 0, \\
& l = 1, \ldots, K, \\
& d_{k,L} \leqslant d_k^{\{q\}} \leqslant d_{k,U}, \; k = 1, \ldots, M,
\end{aligned} \tag{5}$$

where $c_0^{\{q\}}$ and $c_l^{\{q\}}$ are the qth objective and the qth constraint functions, respectively. They are obtained iteratively from first- or higher-order Taylor series expansions T of c_0 and c_l at $\mathbf{d}_0^{\{q\}} = (d_{1,0}^{\{q\}}, \ldots, d_{M,0}^{\{q\}})^\intercal$. The solution of (5), denoted by $\mathbf{d}_*^{\{q\}} = (d_{1,*}^{\{q\}}, \ldots, d_{M,*}^{\{q\}})$, is traditionally produced using a suitable programming method, such as the well-known sequential linear and quadratic programming methods. Then, the qth RDO sub-problem solution $\mathbf{d}_*^{\{q\}}$ is used as the initial design for the $(q+1)$th RDO sub-problem by setting $\mathbf{d}_0^{\{q+1\}} = \mathbf{d}_*^{\{q\}}$. This process is repeated from a chosen initial design $\mathbf{d}_0 = \mathbf{d}_0^{\{1\}}$ to the final optimal design $\mathbf{d}^* = \mathbf{d}_*^{\{Q\}}$ during all $Q \in \mathbb{N}$ iterations to reach convergence. In this paper, the iterations with respect to q are referred to as design iterations.

3 Statistical moment analysis

Given an input random vector $\mathbf{X} := (X_1, \ldots, X_N)^\intercal$ or its transformed version $\mathbf{Z} := (Z_1, \ldots, Z_N)^\intercal$ with known PDF $f_{\mathbf{X}}(\mathbf{x}; \mathbf{d})$ or $f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})$, let $h(\mathbf{Z}; \mathbf{r})$ represent any one of the random output functions $h_l(\mathbf{Z}; \mathbf{r})$, $l = 1, \ldots, K$, introduced in Section 2. Here, $h(\mathbf{Z}; \mathbf{r})$ is assumed to belong to a reasonably large class of random variables, such as the Hilbert space

$$L^2(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}}, \mathbb{P}_{\mathbf{d}}) := \left\{ h : \Omega_{\mathbf{d}} \to \mathbb{R} : \int_{\Omega_{\mathbf{d}}} h^2(\mathbf{Z}; \mathbf{r}) d\mathbb{P}_{\mathbf{d}} < \infty \right\}.$$

This is tantamount to saying that the real-valued function $h(\mathbf{z}; \mathbf{r})$ lives in the equivalent Hilbert space

$$\left\{ h : \bar{\mathbb{A}}^N \to \mathbb{R} : \int_{\bar{\mathbb{A}}^N} h^2(\mathbf{z}; \mathbf{r}) f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g}) d\mathbf{z} < \infty \right\}.$$

The assumption guarantees existence of the first two moments of $h(\mathbf{Z}; \mathbf{r})$, facilitating a solution of the RDO problem in (4).

3.1 Measure-consistent multivariate orthonormal polynomials

When $\mathbf{Z} = (Z_1, \ldots, Z_N)^\intercal$ comprises statistically dependent random variables, the resultant probability measure, in general, is not a product-type, meaning that the joint distribution of $\mathbf{Z}$ cannot be obtained strictly from its marginal distributions. Consequently, measure-consistent multivariate orthonormal polynomials in $\mathbf{z} = (z_1, \ldots, z_N)^\intercal$ cannot be built from an N-dimensional tensor product of measure-consistent univariate orthonormal polynomials. In this section, a three-step algorithm founded on a whitening transformation of the monomial basis is briefly summarized to generate multivariate orthonormal polynomials that are consistent with an arbitrary, non-product-type probability measure $f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})d\mathbf{z}$ of $\mathbf{Z}$. Readers interested in additional details should review the prior work of the authors (Lee and Rahman 2020).

Let $\mathbf{j} := (j_1, \ldots, j_N) \in \mathbb{N}_0^N$ be an N-dimensional multi-index. For $\mathbf{z} = (z_1, \ldots, z_N)^\intercal \in \bar{\mathbb{A}}^N \subseteq \mathbb{R}^N$, a monomial in the real variables $z_1, \ldots, z_N$ is the product $\mathbf{z}^{\mathbf{j}} = z_1^{j_1} \ldots z_N^{j_N}$ and has a total degree $|\mathbf{j}| = j_1 + \cdots + j_N$. A linear combination of $\mathbf{z}^{\mathbf{j}}$, where $|\mathbf{j}| = l, l \in \mathbb{N}_0$, is a homogeneous polynomial of degree l. Consider for each $m \in \mathbb{N}_0$ the elements of the multi-index set $\{\mathbf{j} \in \mathbb{N}_0^N : |\mathbf{j}| \leqslant m\}$, which is arranged as $\mathbf{j}^{(1)}, \ldots, \mathbf{j}^{(L_{N,m})}$, $\mathbf{j}^{(1)} = \mathbf{0}$, according to a monomial order of choice. The set has cardinality

$$L_{N,m} := \sum_{l=0}^{m} \binom{N+l-1}{l} = \binom{N+m}{m}.$$

Denote by

$$\Psi_m(\mathbf{z};\mathbf{g}) = (\Psi_1(\mathbf{z};\mathbf{g}),\ldots,\Psi_{L_{N,m}}(\mathbf{z};\mathbf{g}))^\intercal,$$

an $L_{N,m}$-dimensional vector of multivariate orthonormal polynomials that is consistent with the probability measure $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})d\mathbf{z}$ of $\mathbf{Z}$. It is determined as follows.

(1) Given $m \in \mathbb{N}_0$, create an $L_{N,m}$-dimensional column vector

$$\mathbf{P}_m(\mathbf{z}) = (\mathbf{z}^{\mathbf{j}^{(1)}},\ldots,\mathbf{z}^{\mathbf{j}^{(L_{N,m})}})^\intercal,$$

of monomials whose elements are the monomials $\mathbf{z}^{\mathbf{j}}$ for $|\mathbf{j}| \leqslant m$ arranged in the aforementioned order. It is referred to as the monomial vector in $\mathbf{z} = (z_1,\ldots,z_N)^\intercal$ of degree at most m.

(2) Construct an $L_{N,m} \times L_{N,m}$ monomial moment matrix of $\mathbf{P}_m(\mathbf{Z})$, defined as

$$\begin{aligned}\mathbf{G}_m &:= \mathbb{E}_{\mathbf{g}}[\mathbf{P}_m(\mathbf{Z})\mathbf{P}_m^\intercal(\mathbf{Z})]\\ &:= \int_{\bar{\mathbb{A}}^N} \mathbf{P}_m(\mathbf{z})\mathbf{P}_m^\intercal(\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})d\mathbf{z}.\end{aligned}$$

For an arbitrary PDF $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})$, $\mathbf{G}_m$ cannot be determined exactly, but it can be estimated with good accuracy by numerical integration and/or sampling methods (Lee and Rahman 2020).

(3) Select the $L_{N,m} \times L_{N,m}$ whitening matrix $\mathbf{W}_m$ from the Cholesky decomposition of the monomial moment matrix $\mathbf{G}_m$ such that

$$\mathbf{W}_m^\intercal\mathbf{W}_m = \mathbf{G}_m^{-1} \text{ or } \mathbf{W}_m^{-1}\mathbf{W}_m^{-\intercal} = \mathbf{G}_m.$$

Then, employ the whitening transformation to generate multivariate orthonormal polynomials from

$$\Psi_m(\mathbf{z};\mathbf{g}) = \mathbf{W}_m\mathbf{P}_m(\mathbf{z}).$$

The effectiveness of the three-step algorithm is dependent on reliable construction of a well-conditioned monomial moment matrix, facilitating Cholesky factorization by standard techniques of linear algebra. From past experience, the authors obtained good estimates of $\mathbf{G}_m$ if m is not overly large (Lee and Rahman 2020).

For an ith element $\Psi_i(\mathbf{Z};\mathbf{g})$ of the polynomial vector $\Psi_m(\mathbf{Z};\mathbf{g}) = (\Psi_1(\mathbf{Z};\mathbf{g}),\ldots,\Psi_{L_{N,m}}(\mathbf{Z};\mathbf{g}))^\intercal$, the first- and second-order moments are (Lee and Rahman 2020)

$$\mathbb{E}_{\mathbf{g}}\left[\Psi_i(\mathbf{Z};\mathbf{g})\right] = \begin{cases}1, & \text{if } i = 1,\\ 0, & \text{if } i \neq 1,\end{cases} \tag{6}$$

and

$$\mathbb{E}_{\mathbf{g}}\left[\Psi_i(\mathbf{Z};\mathbf{g})\Psi_j(\mathbf{Z};\mathbf{g})\right] = \begin{cases}1, & i = j,\\ 0, & i \neq j,\end{cases} \tag{7}$$

respectively. These properties are essential to GPCE, to be invoked in a forthcoming section.

Note that the above three-step algorithm is described in terms of orthonormal polynomials in $\mathbf{z}$, not $\mathbf{x}$. This is mainly because $\mathbf{g}$ and hence $\Psi_m(\mathbf{z};\mathbf{g})$ are desired to be invariant when updating the design vector $\mathbf{d}$ during design iterations. To explain this further, consider the qth RDO sub-problem in (5), where the shifting and scaling transformations for the kth initial design variable yield

$$\mathbb{E}_{\mathbf{d}^{\{q\}}}[Z_{i_k}] = g_k^{\{q\}} = \begin{cases}d_k^{\{q\}} + r_{i_k}^{\{q\}}, & \text{shifting},\\ d_k^{\{q\}} r_{i_k}^{\{q\}}, & \text{scaling}.\end{cases} \tag{8}$$

Here, one is free to choose the value of $g_k^{\{q\}}$ with respect to $d_{k,0}^{\{q\}}$ in (8). For instance, when setting $d_k^{\{q\}}$ to $d_{k,0}^{\{q\}}$ at initial design, update $g_k^{\{q\}}$ to be *zero* and *one* in shifting and scaling transformations, respectively. Then, $r_{i_k}^{\{q\}}$ is determined to be $-d_{k,0}^{\{q\}}$ and $1/d_{k,0}^{\{q\}}$, respectively, from (8). Solving the qth RDO sub-problem with the initial design $\mathbf{d}_0^{\{q\}}$ yields $\mathbf{d}_*^{\{q\}}$; thereby $g_k^{\{q\}}$ becomes $d_{k,*}^{\{q\}} - d_{k,0}^{\{q\}}$ and $d_{k,*}^{\{q\}}/d_{k,0}^{\{q\}}$ in shifting or scaling transformations, respectively. In fact, it doesn't matter what value of $g_k^{\{q\}}$ is set with respect to $d_{k,0}^{\{q\}}$ to solve the RDO problem. However, in the updating process from qth to $(q+1)$th design iterations, choosing the same values of $g_k^{\{q\}}$ for $q = 1,2,\ldots,Q$ contributes to only one sequence of calculation of the measure-consistent orthonormal polynomials $\Psi_m(\mathbf{z};\mathbf{g})$ throughout all design iterations.

3.2 Generalized polynomial chaos expansion

According to (6) and (7), any two distinct elements $\Psi_i(\mathbf{z};\mathbf{g})$ and $\Psi_j(\mathbf{z};\mathbf{g})$, $i,j = 1,\ldots,L_{N,m}$, of the polynomial vector $\Psi_m(\mathbf{z};\mathbf{g})$ are mutually orthonormal with respect to the probability measure of $\mathbf{Z}$. Therefore, the set $\{\Psi_i(\mathbf{z};\mathbf{g}), 1 \leqslant i \leqslant L_{N,m}\}$, constructed from the elements of $\Psi_m(\mathbf{z};\mathbf{g})$, is linearly independent. Moreover, the set has cardinality $L_{N,m}$, which matches the dimension of the polynomial space of degree at most m. As $m \to \infty$, $L_{N,m} \to \infty$ as well. In this case, the resulting set $\{\Psi_i(\mathbf{z};\mathbf{g}), 1 \leqslant i < \infty\}$ comprises an infinite number of basis functions. If the PDF of random input $\mathbf{Z}$ is compactly supported or is exponentially integrable (Rahman 2018), as assumed here, then the set of random orthonormal polynomials $\{\Psi_i(\mathbf{Z};\mathbf{g}), 1 \leqslant i < \infty\}$ forms an orthonormal basis of $L^2(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}}, \mathbb{P}_{\mathbf{d}})$. Consequently, any random variable $h(\mathbf{Z};\mathbf{r}) \in L^2(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}}, \mathbb{P}_{\mathbf{d}})$ can be expanded as a Fourier series comprising multivariate orthonormal polynomials in $\mathbf{Z}$, referred to as the GPCE of[2]

$$h(\mathbf{Z};\mathbf{r}) \sim \sum_{i=1}^{\infty} C_i(\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g}), \tag{9}$$

[2]Here, the symbol $\sim$ represents equality in a weaker sense, such as equality in mean-square, but not necessarily pointwise, nor almost everywhere.

where the expansion coefficients $C_i \in \mathbb{R}$, $i = 1, \ldots, \infty$, are defined as

$$C_i(\mathbf{r}) := \mathbb{E}_{\mathbf{g}}\left[h(\mathbf{Z};\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g})\right] := \int_{\bar{\mathbb{A}}^N} h(\mathbf{z};\mathbf{r})\Psi_i(\mathbf{z};\mathbf{g}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z}. \tag{10}$$

According to Lee and Rahman (2020), the GPCE of $h(\mathbf{Z};\mathbf{r}) \in L^2(\Omega_{\mathbf{d}}, \mathcal{F}_{\mathbf{d}}, \mathbb{P}_{\mathbf{d}})$ converges in mean-square, in probability, and in distribution.

The GPCE contains an infinite number of orthonormal polynomials or coefficients. In a practical setting, the number must be finite, meaning that the GPCE must be truncated. However, there are multiple ways to perform a truncation, such as those involving tensor-product, total-degree, and hyperbolic-cross index sets. In this work, the truncation stemming from the total-degree index set is adopted, which entails retaining polynomial expansion orders less than or equal to $m \in \mathbb{N}_0$. The result is an mth-order GPCE approximation

$$h_m(\mathbf{Z};\mathbf{r}) = \sum_{i=1}^{L_{N,m}} C_i(\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g}) \tag{11}$$

of $h(\mathbf{Z};\mathbf{r})$, which contains $L_{N,m}$ expansion coefficients defined by (10).

The GPCE in (9) and (10) should not be mixed up with that of Xiu and Karniadakis (2002). The GPCE presented here is meant for an arbitrary dependent probability distribution of random input. In contrast, the existing PCE, whether classical (Wiener 1938) or generalized (Xiu and Karniadakis 2002), still needs independence of random input.

3.3 Statistical moments

The mth-order GPCE approximation $h_m(\mathbf{Z};\mathbf{r})$ can be viewed as an inexpensive surrogate of an expensive-to-calculate function $h(\mathbf{Z};\mathbf{r})$. Therefore, relevant statistical properties of the latter, such as its first two moments, can be estimated from those of the former.

Applying the expectation operator on $h_m(\mathbf{Z};\mathbf{r})$ in (11) and recognizing (6), its mean

$$\mathbb{E}_{\mathbf{g}}[h_m(\mathbf{Z};\mathbf{r})] = C_1(\mathbf{r}) = \mathbb{E}_{\mathbf{g}}\left[h(\mathbf{Z};\mathbf{r})\right]$$

matches the exact mean of $h(\mathbf{Z};\mathbf{r})$ for any $m \in \mathbb{N}_0$. Enforcing the expectation operator again, this time on $(h_m(\mathbf{Z};\mathbf{r}) - \mathbb{E}_{\mathbf{g}(\mathbf{d})}[h_m(\mathbf{Z};\mathbf{r})])^2$, and using (7) results in the variance

$$\begin{aligned}\mathrm{var}_{\mathbf{g}}[h_m(\mathbf{Z};\mathbf{r})] &= \sum_{i=1}^{L_{N,m}} C_i^2(\mathbf{r}) - C_1^2(\mathbf{r}) \\ &= \sum_{i=2}^{L_{N,m}} C_i^2(\mathbf{r}) \leqslant \mathrm{var}_{\mathbf{g}}\left[h(\mathbf{Z};\mathbf{r})\right]\end{aligned}$$

of $h_m(\mathbf{Z};\mathbf{r})$, where the equality before the last term operates when $m \to \infty$. Therefore, the second-moment statistics of a GPCE approximation are solely determined by an appropriately truncated set of expansion coefficients.

3.4 Expansion coefficients

The expansion coefficients of an mth-order GPCE approximation $h_m(\mathbf{Z};\mathbf{r})$ involve various N-dimensional integrations. For an arbitrary function h and an arbitrary probability distribution of random input $\mathbf{Z}$, their exact evaluations from the definition alone are impossible. Numerical integration involving a multivariate, tensor-product Gauss-type quadrature rule is computationally prohibitive for high-dimensional ($N \geqslant 10$, say) UQ/RDO problems. To surmount this hurdle, two regression methods, namely, standard least-squares (SLS) and diffeomorphic modulation under observable response preserving homotopy (D-MORPH), were employed to obtain associated estimates of the coefficients. Here, only a brief summary of SLS and D-MORPH regression is given for the paper to be self-contained. For additional details, readers are advised to consult related works of Li and Rabitz (2010) and Lee and Rahman (2020).

3.4.1 Standard least-squares regression

From the known distribution of random input $\mathbf{Z}$ and an output function $h : \bar{\mathbb{A}}^N \to \mathbb{R}$, consider an input-output data set $\{\mathbf{z}^{(l)}, h(\mathbf{z}^{(l)};\mathbf{r})\}_{l=1}^{L}$ of size $L \in \mathbb{N}$, where $\mathbf{r}$ is decided from the knowledge of $\mathbf{d}$ and $\mathbf{g}$, as discussed earlier. The data set, often referred to as the experimental design, is generated by calculating the function h at each input data $\mathbf{z}^{(l)}$. Various sampling methods, namely, standard Monte Carlo simulation (MCS), quasi MCS (QMCS), and Latin hypercube sampling, can be used to build the experimental design. Using the experimental design, the approximate GPCE coefficients $\tilde{C}_i(\mathbf{d})$, $i = 1, \ldots, L_{N,m}$, satisfy the linear system

$$\mathbf{A}\mathbf{c} = \mathbf{b},$$

where

$$\mathbf{A} := \begin{bmatrix} \tilde{\Psi}_1(\mathbf{z}^{(1)};\mathbf{g}) & \cdots & \tilde{\Psi}_{L_{N,m}}(\mathbf{z}^{(1)};\mathbf{g}) \\ \vdots & \ddots & \vdots \\ \tilde{\Psi}_1(\mathbf{z}^{(L)};\mathbf{g}) & \cdots & \tilde{\Psi}_{L_{N,m}}(\mathbf{z}^{(L)};\mathbf{g}) \end{bmatrix},$$
$$\mathbf{b} := (h(\mathbf{z}^{(1)};\mathbf{r}), \ldots, h(\mathbf{z}^{(L)};\mathbf{r}))^{\mathsf{T}}, \text{ and}$$
$$\mathbf{c} := (\tilde{C}_1(\mathbf{r}), \ldots, \tilde{C}_{L_{N,m}}(\mathbf{r}))^{\mathsf{T}}.$$

Here, $\tilde{\Psi}_i(\mathbf{z}^{(l)};\mathbf{g})$ represents an estimate of $\Psi_i(\mathbf{z}^{(l)};\mathbf{r})$ due to approximations involved in constructing the monomial

matrix (Lee and Rahman 2020). According to SLS, the expansion coefficients of GPCE are estimated by minimizing the residual

$$\hat{e}_m := \frac{1}{L}\sum_{l=1}^{L}\left[h(\mathbf{z}^{(l)};\mathbf{r}) - \sum_{i=1}^{L_{N,m}} \tilde{C}_i \tilde{\Psi}_i(\mathbf{z}^{(l)};\mathbf{g})\right]^2.$$

As such, the SLS solution $\hat{C}_i,\ i = 1,\ldots,L_{N,m}$, is obtained from

$$\mathbf{A}^\intercal\mathbf{A}\hat{\mathbf{c}} = \mathbf{A}^\intercal\mathbf{b},$$

where $\hat{\mathbf{c}} := (\hat{C}_1(\mathbf{r}),\ldots,\hat{C}_{L_{N,m}}(\mathbf{r}))^\intercal$ and the $L_{N,m}\times L_{N,m}$ matrix $\mathbf{A}^\intercal\mathbf{A}$ is referred to as the information or data matrix. The inversion of the data matrix, if it is positive-definite, yields the best estimate

$$\hat{\mathbf{c}} = (\mathbf{A}^\intercal\mathbf{A})^{-1}\mathbf{A}^\intercal\mathbf{b}$$

of the approximate GPCE coefficients. A necessary condition for the inversion is $L > L_{N,m}$, often referred to as an overdetermined system. Even when the condition is satisfied, the experimental design must be wisely selected, so that the matrix $\mathbf{A}^\intercal\mathbf{A}$ is well-conditioned.

3.4.2 Partitioned D-MORPH regression

In an overdetermined system ($L > L_{N,m}$), if L is not sufficiently larger than $L_{N,m}$, then there may not be enough information, rendering SLS regression inaccurate for estimating the coefficients. Moreover, in an underdetermined system ($L < L_{N,m}$), SLS becomes invalid because the data matrix is no longer invertible. In either case, an alternative regression method, such as the partitioned D-MORPH, was employed to obtain reliable and efficient estimates of the GPCE coefficients. A more detailed theory of the partitioned D-MORPH is available in the prior work (Lee and Rahman 2020). Here, it is summarized, including the final solutions, and evaluated later in Example 3.

For an overdetermined ($L > L_{N,m}$) or underdetermined ($L < L_{N,m}$) system, consider dividing the GPCE basis functions into two groups: (1) a primary group consisting of $L_p \leqslant L_{N,m}$ basis functions and (2) a secondary group comprising the remaining $L_{N,m} - L_p$ basis functions. In many real-life problems, the low-order basis functions of GPCE contribute to a function value more significantly than the high-order basis functions of GPCE. In such a case, the low-order basis functions form the primary group, while the rest are lumped into the secondary group. One can then utilize specific criteria, introduced by Lee and Rahman (2020), to group the primary and secondary basis functions.

According to Lee and Rahman (2020), two types of the partitioned D-MORPH, namely, the direct approach and extended approach, are available. The direct approach, which entails a straightforward version of the partitioned D-MORPH, is explained in Appendix 1. The extended approach describes an iterated variant of the partitioned D-MORPH, revising the approximate expansion coefficients estimated by the direct approach. In this section, the final solution of the extended approach is concisely presented. Readers interested in additional details should consult the original work (Lee and Rahman 2020).

In the extended approach, the best estimate of the expansion coefficients is obtained from the following two principal steps: (1) the GPCE coefficients using the direct approach of the partitioned D-MORPH are calculated, obtaining $\check{\mathbf{c}} \in \mathbb{R}^{L_{N,m}}$ from (A1.2) in Appendix 1; (2) using the coefficients from the direct approach, a revised initial solution $\mathbf{c}'_0 = (\check{C}_1(\mathbf{d}),\ldots,\check{C}_{L_p}(\mathbf{d}), C'_{0,L_p+1}(\mathbf{d}),\ldots,C'_{0,L_{N,m}}(\mathbf{d}))^\intercal \in \mathbb{R}^{L_{N,m}}$ of the GPCE coefficients is defined, where $|C'_{0,i}| \leqslant |\check{C}_{L_p}|$ for $i = L_{p+1},\ldots,L_{N,m}$ by one of weighting methods presented by Lee and Rahman (2020). Then, the final solution of the GPCE coefficients by the extended approach, denoted by $\mathbf{c}' = (C'_1(\mathbf{d}),\ldots,C'_{L_{N,m}}(\mathbf{d}))^\intercal$, is

$$\begin{aligned}\mathbf{c}' = {}& \bar{\mathbf{F}}'_{L_{N,m}-r}(\bar{\mathbf{E}}'^\intercal_{L_{N,m}-r}\bar{\mathbf{F}}'_{L_{N,m}-r})^{-1}\bar{\mathbf{E}}'^\intercal_{L_{N,m}-r}\,\bar{\mathbf{A}}^+\bar{\mathbf{b}}\\ &+\bar{\mathbf{F}}'_r(\bar{\mathbf{E}}'^\intercal_r\bar{\mathbf{F}}'_r)^{-1}\bar{\mathbf{T}}'^{-1}_r\bar{\mathbf{E}}'^\intercal_r\,\bar{\Phi}\mathbf{c}'_0,\end{aligned}$$

where $\bar{\mathbf{E}}'_r$ and $\bar{\mathbf{E}}'_{L_{N,m}-r}$, $\bar{\mathbf{F}}'_r$ and $\bar{\mathbf{F}}'_{L_{N,m}-r}$ are constructed from the first r and the last $L_{N,m}-r$ columns of matrices $\bar{\mathbf{E}}'$ and $\bar{\mathbf{F}}'$, respectively; they are generated from the singular value decomposition of $\bar{\Phi}$ as

$$\bar{\Phi} = \bar{\mathbf{E}}'\begin{bmatrix}\bar{\mathbf{T}}'_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\bar{\mathbf{F}}'^\intercal,$$

where the $L_{N,m}\times L_{N,m}$ matrix $\bar{\Phi}$ is presented in (A1.3) of Appendix 1.

The extended version of the partitioned D-MORPH regression will be employed for estimating the GPCE coefficients in Example 4.

4 Proposed methods for design sensitivity analysis

When solving an RDO problem with a typical gradient-based optimization algorithm, such as linear and sequential quadratic programming, at least the first-order derivatives of the first- and second-order moments of $h_l(\mathbf{Z};\mathbf{r})$, $l = 0,1,\ldots,K$, with respect to each design variable d_k, $k = 1,\ldots,M$, are necessary. In this section, an analytical formulation for design sensitivity analysis is unveiled by combining GPCE coefficients with score functions for dependent input random variables. For such sensitivity analysis, the following regularity conditions are required:

1. The probability density function $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})$ of $\mathbf{Z}$ is continuous. In addition, the partial derivative $\partial f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})/\partial g_k$, $k = 1,\ldots,M$, exists and is finite for all possible values of $\mathbf{z}$ and g_k. Furthermore, the statistical moments of $h(\mathbf{Z};\mathbf{r})$ are differentiable functions of $\mathbf{g}$.
2. There exists a Lebesgue integrable dominating function $t(\mathbf{z})$ such that

$$\left| h^r(\mathbf{z};\mathbf{r}) \frac{\partial f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})}{\partial d_k} \right| \leqslant t(\mathbf{z}),\ r = 1,2;\ k = 1,\ldots,M.$$

The proposed formulation is novel when compared with the existing sensitivity analysis restricted to independent random variables only (Ren and Rahman 2013; Rahman and Ren 2014).

4.1 Score functions

Suppose the first-order derivatives of the first two moments, $\mathbb{E}_{\mathbf{g}}[h^r(\mathbf{Z};\mathbf{r})]$, $r = 1, 2$, of a generic output function $h(\mathbf{Z};\mathbf{r})$ with respect to a design variable d_k are wanted to solve the qth RDO sub-problem in (5). During the sub-iteration process of the qth design iteration, $\mathbf{g}^{\{q\}}$ changes, but $\mathbf{r}^{\{q\}}$ remains constant locally. For brevity, the iteration count q is omitted from $\mathbf{d}^{\{q\}}$, $\mathbf{g}^{\{q\}}$, and $\mathbf{r}^{\{q\}}$ in the remainder of this section.

Applying the partial derivative of these moments with respect to d_k and then invoking the chain rule and Lebesgue dominated convergence theorem (Browder 1996), which permits the differential and integral operators to be interchanged, yields the sensitivities

$$\begin{aligned}\frac{\partial \mathbb{E}_{\mathbf{g}}\left[h^r(\mathbf{Z};\mathbf{r})\right]}{\partial d_k} &= \frac{\partial}{\partial d_k}\int_{\bar{\mathbb{A}}^N} h^r(\mathbf{z};\mathbf{r}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z} \\ &= \frac{\partial g_k}{\partial d_k}\frac{\partial}{\partial g_k}\int_{\bar{\mathbb{A}}^N} h^r(\mathbf{z};\mathbf{r}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z} \\ &= \frac{\partial g_k}{\partial d_k}\int_{\bar{\mathbb{A}}^N} h^r(\mathbf{z};\mathbf{r}) \frac{\partial \ln f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})}{\partial g_k} f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z}, \\ & \qquad r = 1,2;\ k = 1,\ldots,M, \end{aligned} \tag{12}$$

provided that $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) > 0$ on $\bar{\mathbb{A}}^N$. Here, $\partial g_k/\partial d_k$ is 1 or r_{i_k} for shifting or scaling transformations, respectively. Define by

$$s_k(\mathbf{Z};\mathbf{g}) := \frac{\partial \ln f_{\mathbf{Z}}(\mathbf{Z};\mathbf{g})}{\partial g_k} \tag{13}$$

the first-order score function (Rubinstein and Shapiro 1993; Rahman 2009b) for the variable g_k. In many cases, the score functions can be determined numerically or analytically—for instance, when $\mathbf{Z}$ follows classical probability distributions, such as those obtained for multivariate Gaussian and lognormal density functions in Table 1. Thereafter, the sensitivities in (12) can also be expressed by

$$\begin{aligned}\frac{\partial \mathbb{E}_{\mathbf{g}}[h^r(\mathbf{Z};\mathbf{r})]}{\partial g_k} &= \frac{\partial g_k}{\partial d_k}\int_{\bar{\mathbb{A}}^N} h^r(\mathbf{z};\mathbf{r}) s_k(\mathbf{z};\mathbf{g}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z} \\ &= \frac{\partial g_k}{\partial d_k}\mathbb{E}_{\mathbf{g}}\left[h^r(\mathbf{Z};\mathbf{r}) s_k(\mathbf{Z};\mathbf{g})\right]. \end{aligned} \tag{14}$$

According to (14), the moments and their sensitivities have both been formulated as expectations of stochastic quantities with respect to the same probability measure, making their concurrent evaluations possible in a single stochastic simulation or analysis.

4.2 Exact sensitivities

Given the input random vector $\mathbf{Z}$ with PDF $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})$, consider the full GPCE of the kth score function

$$s_k(\mathbf{Z};\mathbf{g}) = \sum_{i=2}^{\infty} D_{k,i}(\mathbf{g})\Psi_i(\mathbf{Z};\mathbf{g}), \tag{15}$$

with its own GPCE coefficients

$$D_{k,i}(\mathbf{g}) = \int_{\bar{\mathbb{A}}^N} s_k(\mathbf{z};\mathbf{g})\Psi_i(\mathbf{z};\mathbf{g}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z},\ i = 2,3,\ldots,\infty.$$

Note that the lowest orthonormal polynomial function or coefficient of the GPCE in (15) starts from $i = 2$, not $i = 1$. This is because

$$\begin{aligned} D_{k,1}(\mathbf{g}) &= \int_{\bar{\mathbb{A}}^N} s_k(\mathbf{z};\mathbf{g}) f_{\mathbf{Z}}(\mathbf{z};\mathbf{g}) d\mathbf{z} \\ &=: \mathbb{E}_{\mathbf{g}}\left[s_k(\mathbf{Z};\mathbf{g})\right] = 0, \end{aligned}$$

Table 1 Derivatives of log-density functions for two types of multivariate distributions

Type	Score function for design variables $\mathbf{g} = (g_1,\ldots,g_M)$.
1	Gaussian density on $(-\infty,\infty)^N$ $s_k(\mathbf{z};\mathbf{g})^{(a)} = \sum_{j=1}^N p_{i_k,j}(z_j - \mu_j)$, $\mu_{i_k} = g_k, k = 1,\ldots,M, 1 \leqslant i_1 < \cdots < i_M \leqslant N$, $[p_{i,j}] = \Sigma_{\mathbf{Z}}^{-1} \in \mathbb{R}^{N\times N}$, $\Sigma_{\mathbf{Z}} = [\rho_{ij}\sigma_i\sigma_j]$, $0 < \sigma_i < \infty, -1 < \rho_{ij} < 1, i, j = 1,\ldots,N$.
2	Lognormal density on $[0,\infty)^N$ $s_k(\mathbf{z};\mathbf{g})^{(a)} = l_k \sum_{j=1}^N \tilde{p}_{i_k,j}(\ln z_j - \tilde{\mu}_j)^{(b)}$, $\tilde{\mu}_i = \ln[\mu_i^2/\sigma_i] - 1/2, \tilde{\sigma}_i = \sqrt{\ln\left[\sigma_i^2/\mu_i^2\right] + 1}$, $\mu_{i_k} = g_k, k = 1,\ldots,M, 1 \leqslant i_1 < \cdots < i_M \leqslant N$, $[\tilde{p}_{i,j}] = \tilde{\Sigma}_{\ln\mathbf{Z}}^{-1} \in \mathbb{R}^{N\times N}$, $\tilde{\Sigma}_{\ln\mathbf{Z}} = [\tilde{\rho}_{ij}\tilde{\sigma}_i\tilde{\sigma}_j]$, $\tilde{\rho}_{ij} = \ln\left[\rho_{ij}/(\mu_i\mu_j) + 1\right]/(\tilde{\sigma}_i\tilde{\sigma}_j)$, $0 < \sigma_i < \infty, -1 < \rho_{ij} < 1, i, j = 1,\ldots,N$.

[a] $s_k(\mathbf{z};\mathbf{g}) = \partial \ln f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})/\partial g_k$

[b] $l_k = [1 + 2(\sigma_{i_k}/g_k)^2]/[g_k\{1 + (\sigma_{i_k}/g_k)^2\}]$

according to Appendix 2. Then, combining (9) and (15), (14) produces the sensitivity of rth-order moment of $h(\mathbf{Z};\mathbf{r})$ with respect to kth design variable d_k as

$$\frac{\partial \mathbb{E}_{\mathbf{g}}[h^r(\mathbf{Z};\mathbf{r})]}{\partial d_k} = \frac{\partial g_k}{\partial d_k}\mathbb{E}_{\mathbf{g}}\left[\left(\sum_{i=1}^{\infty} C_i(\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g})\right)^r \times\left(\sum_{j=2}^{\infty} D_{k,j}(\mathbf{g})\Psi_j(\mathbf{Z};\mathbf{g})\right)\right]. \tag{16}$$

On the right side of (16), the expectation operator contains expansions of response and score functions with respect to the same multivariate orthonormal polynomial basis, consistent with the same probability measure $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})d\mathbf{z}$. Thereafter, using the second-moment properties in (6) and (7), the sensitivities of the first-order ($r = 1$) and second-order ($r = 2$) moments with respect to d_k are finally derived as

$$\frac{\partial \mathbb{E}_{\mathbf{g}}[h(\mathbf{Z};\mathbf{r})]}{\partial d_k} = \frac{\partial g_k}{\partial d_k}\sum_{i=2}^{\infty} C_i(\mathbf{r})D_{k,i}(\mathbf{g}) \tag{17}$$

and

$$\frac{\partial \mathbb{E}_{\mathbf{g}}[h^2(\mathbf{Z};\mathbf{r})]}{\partial d_k} = \frac{\partial g_k}{\partial d_k}\sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty}\sum_{i_3=2}^{\infty} C_{i_1}(\mathbf{r})C_{i_2}(\mathbf{r})D_{k,i_3}(\mathbf{g}) \times\mathbb{E}_{\mathbf{g}}\left[\prod_{p=1}^{3}\Psi_{i_p}(\mathbf{Z};\mathbf{g})\right], \tag{18}$$

respectively.

The closed-form expressions of the moment sensitivities in (17) and (18) mainly consist of GPCE coefficients for $h(\mathbf{Z};\mathbf{r})$ and $s_k(\mathbf{Z};\mathbf{g})$. Therefore, these sensitivity equations are exact because GPCE is mean-square convergent for any square-integrable function.

4.3 Approximate sensitivities

The full GPCE solution for the sensitivities of response moments comprises an infinite number of basis functions or coefficients. Therefore, in practice, the solution must be truncated. Two options are suggested as follows.

Option 1. Given two non-negative integers $m \in \mathbb{N}_0$ and $m' \in \mathbb{N}_0$, consider replacing $h(\mathbf{Z};\mathbf{r})$ and $s_k(\mathbf{Z};\mathbf{g})$ by their mth-order and m'th-order truncations or approximations, respectively. The resultant first- and second-moment sensitivities then become

$$\frac{\partial \mathbb{E}_{\mathbf{g}}[h_m(\mathbf{Z};\mathbf{r})]}{\partial d_k} = \frac{\partial g_k}{\partial d_k}\sum_{i=2}^{L_{\min}} C_i(\mathbf{r})D_{k,i}(\mathbf{g}), \tag{19}$$

and

$$\frac{\partial \mathbb{E}_{\mathbf{g}}[h_m^2(\mathbf{Z};\mathbf{r})]}{\partial d_k} = \frac{\partial g_k}{\partial d_k}\sum_{i_1=1}^{L_{N,m}}\sum_{i_2=1}^{L_{N,m}}\sum_{i_3=2}^{L_{N,m'}} C_{i_1}(\mathbf{r})C_{i_2}(\mathbf{r}) \times D_{k,i_3}(\mathbf{g})\mathbb{E}_{\mathbf{g}}\left[\prod_{p=1}^{3}\Psi_{i_p}(\mathbf{Z};\mathbf{g})\right], \tag{20}$$

respectively, where $L_{\min} := \min(L_{N,m}, L_{N,m'})$. The approximate sensitivities in (19) and (20) converge to $\partial\mathbb{E}_{\mathbf{g}}[h(\mathbf{Z};\mathbf{r})]\,/\,\partial d_k$ and $\partial\mathbb{E}_{\mathbf{g}}[h^2(\mathbf{Z};\mathbf{r})]\,/\,\partial d_k$, respectively, when $m \to \infty$ and $m' \to \infty$.

Since the score function is solely described by the PDF of input random variables, the order m' for its GPCE approximation is generally different than the order m employed for GPCE approximations of output functions. Moreover, the size L' (say) of the input-output data set is also different when estimating the coefficients $D_{k,i}$, $i = 2, 3, \ldots, L_{N,m'}$.

In (20), the expectations of products of three distinct multivariate orthonormal polynomials need to be calculated $L_{N,m}\times L_{N,m}\times(L_{N,m'}-1)$ times. For an arbitrary dependent random vector $\mathbf{Z}$, such expectations or integrals cannot be calculated exactly. This is in contrast to independent variables where exact solutions exist for a few classical distributions (Busbridge 1948; Rahman and Ren 2014). Therefore, for dependent variables, they must be estimated, say, by numerical integration or sampling methods. If the dimension is too high, then the sampling methods, such as MCS, QMCS, or Latin hypercube sampling, can be used to estimate these integrals in two steps:

1. Consistent with the probability measure $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})d\mathbf{z}$, generate an input data set $\{\mathbf{z}^{(l)}\}_{l=1}^{L''}$ of size $L'' \in \mathbb{N}$ by a sampling method of choice.
2. Estimate the expectation of triple product as an arithmetic mean, producing

$$\mathbb{E}_{\mathbf{g}}\left[\prod_{p=1}^{3}\Psi_{i_p}(\mathbf{Z};\mathbf{g})\right] \approx \begin{cases} \frac{1}{L''}\sum_{l=1}^{L''}\prod_{p=1}^{3}\Psi_{i_p}(\mathbf{z}^{(l)};\mathbf{g}), \\ \quad \text{if } i_1 \neq 1, i_2 \neq 1, \\ 1, \\ \quad \text{if } i_1 = 1, i_2 \neq 1, i_2 = i_3; \\ \quad i_2 = 1, i_1 \neq 1, i_1 = i_3, \\ 0, \\ \quad \text{if } i_1 = i_2 = 1; \\ \quad i_1 = 1, i_2 \neq 1, i_2 \neq i_3; \\ \quad i_2 = 1, i_1 \neq 1, i_1 \neq i_3. \end{cases} \tag{21}$$

Note that some of these expectations are either *one* or *zero*, depending on the polynomial indices. Also, the various

triple products of orthonormal polynomials inside the expectation operator in (21) are repetitive, meaning that the total number of the expectations can be reduced. Furthermore, in a design process, the expectations of products of these polynomials need not be recalculated since the orthonormal polynomials are preserved during design iterations.

Since the score function and measure-consistent orthonormal polynomials are not associated with any output function, the response moments and their design sensitivities are estimated from GPCE coefficients simultaneously in a single stochastic analysis. Therefore, a significant cost savings is anticipated when FEA-generated output functions are involved in practical applications.

Option 2. For high-dimensional RDO problems, the three-dimensional sums in (20) mandate numerous expectations of triple products of orthonormal polynomials. As a result, the computational expense of Option 1, even when the expectations are calculated only once, can be high. This is despite the fact that no output functions, which are generally expensive to evaluate, are involved. In such a case, Option 2 provides a more economical route to the sensitivity analysis by performing an $\bar{m}$th-order (say) GPCE approximation of the product term inside the expectation of (14), yielding

$$h^r(\mathbf{Z};\mathbf{r})s_k(\mathbf{Z};\mathbf{g}) \approx \sum_{i=1}^{L_{N,\bar{m}}} H_{k,i}^{(r)}(\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g}), \tag{22}$$

where

$$\begin{aligned} H_{k,i}^{(r)}(\mathbf{r}) &= \int_{\bar{\mathbb{A}}^N} h^r(\mathbf{z};\mathbf{r})s_k(\mathbf{z};\mathbf{g})\Psi_i(\mathbf{z};\mathbf{g})f_{\mathbf{Z}}(\mathbf{z};\mathbf{r})d\mathbf{z}, \\ i &= 1,\ldots,L_{N,\bar{m}}, \end{aligned} \tag{23}$$

represent the affiliated expansion coefficients. Hereafter, combining (14) and (22) and invoking the property in (6), the sensitivities of first- and second-order moments are approximated by

$$\begin{aligned} \frac{\partial \mathbb{E}_{\mathbf{g}}[h_m(\mathbf{Z};\mathbf{r})]}{\partial d_k} &\approx \frac{\partial g_k}{\partial d_k}\mathbb{E}_{\mathbf{g}}\left[\sum_{i=1}^{L_{N,\bar{m}}} H_{k,i}^{(1)}(\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g})\right], \\ &= \frac{\partial g_k}{\partial d_k}H_{k,1}^{(1)}(\mathbf{r}) \end{aligned} \tag{24}$$

and

$$\begin{aligned} \frac{\partial \mathbb{E}_{\mathbf{g}}[h_m^2(\mathbf{Z};\mathbf{r})]}{\partial d_k} &\approx \frac{\partial g_k}{\partial d_k}\mathbb{E}_{\mathbf{g}}\left[\sum_{i=1}^{L_{N,\bar{m}}} H_{k,i}^{(2)}(\mathbf{r})\Psi_i(\mathbf{Z};\mathbf{g})\right], \\ &= \frac{\partial g_k}{\partial d_k}H_{k,1}^{(2)}(\mathbf{r}). \end{aligned} \tag{25}$$

As (24) and (25) sidestep the need for calculating the expectations of products of three orthonormal polynomials, a hefty computational savings is anticipated in Option 2 when the expectations are expensive to evaluate. Having said this, as the product term in (22) becomes a non-polynomial function, higher-order GPCE approximations with larger order ($\bar{m}$) may be required to warrant a satisfactory approximation quality. Additionally, even when $\bar{m} = m$, a data set of larger size ($\bar{L} > L$) may be needed for estimating the respective coefficients. Nonetheless, Option 2 is worth trying and will be featured in relevant numerical examples where output functions are inexpensive to evaluate.

5 Proposed methods for robust design optimization

The GPCE approximations described in the foregoing sections are intended to evaluate the objective and constraint functions, including their design sensitivities, from a single stochastic analysis. An integration of statistical moment analysis, design sensitivity analysis, and a suitable optimization algorithm is expected to deliver a convergent solution of a generic RDO problem defined in (4). However, such an integration depends on the complexity of the RDO problem at hand, pointing to a need for multiple design methods. The following sections describe three distinct approaches to the integration, resulting in three design optimization methods: (1) the direct GPCE method, (2) the single-step GPCE method, and (3) the multi-point single-step GPCE method.

5.1 Direct GPCE

The direct GPCE method entails a plain vanilla incorporation of the GPCE-based stochastic and design sensitivity analyses with a chosen gradient-based optimization algorithm. Given a design vector at the current iteration and the corresponding values of the objective and constraint functions and their sensitivities, the design vector at the next iteration is calculated from the optimization algorithm. However, new statistical moment analysis and new sensitivity analysis, requiring recalculations of the GPCE expansion coefficients from additional output function evaluations, are required at every design iteration.

For a more elaborate explanation, consider a change of design variables from an old design $\mathbf{d}$ to a new design $\mathbf{d}'$ during the design iteration process. Then, the probability measure of $\mathbf{X}$ varies from $f_{\mathbf{X}}(\mathbf{x};\mathbf{d})d\mathbf{x}$ to $f_{\mathbf{X}}(\mathbf{x};\mathbf{d}')d\mathbf{x}$, corresponding to old and new designs, respectively. Analogously, the old deterministic vector $\mathbf{r}$ evolves to its newer version $\mathbf{r}'$ under shifting or scaling transformations of $\mathbf{X}$, as described in Section 2. However, the probability measure $f_{\mathbf{Z}}(\mathbf{z};\mathbf{g})d\mathbf{z}$ of $\mathbf{Z}$ remains unaltered as $\mathbf{g}$ remains constant at all design iterations from either transformation. Nonetheless, a new set of output data is called for when the design changes.

Let the input-output data sets generated independently for the old and new designs be $\{\mathbf{z}^{(l)}, h(\mathbf{z}^{(l)};\mathbf{r})\}_{l=1}^{L}$ and $\{\mathbf{z}^{(l)}, h(\mathbf{z}^{(l)};\mathbf{r}')\}_{l=1}^{L}$, respectively. In these two sets, the input data are the same, but the output data are different as $\mathbf{r}$ and $\mathbf{r}'$ are different. Denote by $C_i(\mathbf{r})$ and $C_i(\mathbf{r}')$ the expansion coefficients for the old and new designs, respectively. Then, the coefficients for both designs are obtained by minimizing the associated residuals

$$\hat{e}_m := \frac{1}{L}\sum_{l=1}^{L}\left[h(\mathbf{z}^{(l)};\mathbf{r}) - \sum_{i=1}^{L_{N,m}} C_i(\mathbf{r})\Psi_i(\mathbf{z}^{(l)};\mathbf{g})\right]^2 \tag{26}$$

and

$$\hat{e}'_m := \frac{1}{L}\sum_{l=1}^{L}\left[h(\mathbf{z}^{(l)};\mathbf{r}') - \sum_{i=1}^{L_{N,m}} C_i(\mathbf{r}')\Psi_i(\mathbf{z}^{(l)};\mathbf{g})\right]^2, \tag{27}$$

respectively, using either SLS or D-MORPH regression explained in Section 3. According to (26) and (27), there is no need to regenerate the input data and recalculate the multivariate orthonormal polynomials, but, still, new output data sets are mandated at all design iterations. In consequence, the direct GPCE method can be expensive, depending on the cost of evaluating the objective and constraint functions and the requisite number of design iterations to attain convergence.

5.2 Single-step GPCE

The single-step GPCE method is intended to solve the entire RDO problem from a single stochastic analysis by circumventing the need to recalculate the GPCE expansion coefficients from new input-output data sets at every design iteration. However, it is predicated on two important assumptions: (1) an mth-order GPCE approximation $h_m(\mathbf{Z};\mathbf{r})$ of $h(\mathbf{Z};\mathbf{r})$ at the initial design is adequate for all possible designs and (2) the GPCE coefficients for a new design, derived by recycling those generated for an old design, are accurate.

Under the above two assumptions, consider again the vectors $\mathbf{r}$ and $\mathbf{r}'$ associated with the old and new designs, respectively. Assume that the GPCE coefficients $C_i(\mathbf{r})$, $i = 1,\ldots,L_{N,m}$, for the old design have been calculated from the input-output data $\{\mathbf{z}^{(l)}, h(\mathbf{z}^{(l)};\mathbf{r})\}_{l=1}^{L}$ already. Then, the GPCE coefficients $C_i(\mathbf{r}')$, $i = 1,\ldots,L_{N,m}$, for the new design are estimated by modifying the old input data $\{\mathbf{z}^{(l)}\}_{l=1}^{L}$ to the new input data $\{\mathbf{z}'^{(l)}\}_{l=1}^{L}$, depending on the scaling or shifting transformations, as follows.

$$\mathbf{z}'^{(l)} = \begin{cases} \mathbf{z}^{(l)} - \mathbf{r}' + \mathbf{r} & \text{shifting,} \\ \operatorname{diag}\left(\frac{r_1}{r'_1},\ldots,\frac{r_N}{r'_N}\right)\mathbf{z}^{(l)}, & \text{scaling.} \end{cases}$$

To explain these modifications, first consider the shifting transformation. In this case, the lth sample of the output function is

$$\begin{aligned} h(\mathbf{z}^{(l)};\mathbf{r}') &:= y(\mathbf{z}^{(l)} - \mathbf{r}') = y(\mathbf{z}^{(l)} - \mathbf{r}' + \mathbf{r} - \mathbf{r}) \\ &= y(\mathbf{z}'^{(l)} - \mathbf{r}) =: h(\mathbf{z}'^{(l)};\mathbf{r}), \end{aligned}$$

where $\mathbf{z}'^{(l)} := \mathbf{z}^{(l)} - \mathbf{r}' + \mathbf{r}$ is the modified lth input sample. Second, for the scaling transformation, the lth sample of the output function is

$$\begin{aligned} h(\mathbf{z}^{(l)};\mathbf{r}') &:= y\left(\operatorname{diag}\left[\frac{1}{r'_1},\ldots,\frac{1}{r'_N}\right]\mathbf{z}^{(l)}\right) \\ &= y\left(\operatorname{diag}\left[\frac{1}{r_1},\ldots,\frac{1}{r_N}\right]\right. \\ &\qquad \left.\times \operatorname{diag}\left[\frac{r_1}{r'_1},\ldots,\frac{r_N}{r'_N}\right]\mathbf{z}^{(l)}\right) \\ &= y\left(\operatorname{diag}\left[\frac{1}{r_1},\ldots,\frac{1}{r_N}\right]\mathbf{z}'^{(l)}\right) =: h(\mathbf{z}'^{(l)};\mathbf{r}), \end{aligned}$$

where $\mathbf{z}'^{(l)} := \operatorname{diag}[r_1/r'_1,\ldots,r_N/r'_N]\mathbf{z}^{(l)}$ is the modified lth input sample. These modifications are intended to evaluate the output function at the new design to be approximated by the output function at the old design, that is,

$$h(\mathbf{z}^{(l)};\mathbf{r}') = h(\mathbf{z}'^{(l)};\mathbf{r}) \approx \sum_{i=1}^{L_{N,m}} C_i(\mathbf{r})\Psi_i(\mathbf{z}'^{(l)};\mathbf{g}), \tag{28}$$

where the last term reflects mth-order GPCE approximation. Applying (28) to (27) yields yet another residual

$$\hat{e}''_m := \frac{1}{L}\sum_{l=1}^{L}\left[\sum_{i=1}^{L_{N,m}} C_i(\mathbf{r})\Psi_i(\mathbf{z}'^{(l)};\mathbf{g}) - \sum_{i=1}^{L_{N,m}} C_i(\mathbf{r}')\Psi_i(\mathbf{z}^{(l)};\mathbf{g})\right]^2, \tag{29}$$

the minimization of which by SLS or D-MORPH regression produces GPCE coefficients for the new design. Compared with the minimization of $\hat{e}'_m$ in (27), no new output data obtained from the original function, that is, $h(\mathbf{z}^{(l)};\mathbf{r}')$, are required. Instead, the output data involved in (29) are generated reusing the old coefficients and invoking the GPCE approximation. Subsequently, new statistical moment and design sensitivity analyses, all employing an mth-order GPCE approximation at the initial design, are conducted with little extra cost during all design iterations. Therefore, the single-step GPCE method holds the potential to substantially curtail the computational effort in solving an RDO problem.

5.3 Multi-point single-step GPCE

The direct and single-step methods described in the former sections are grounded on GPCE approximations of stochastic responses, supplying surrogates of objective and

constraint functions for the entire design space. Therefore, these methods are globally formulated and may not be practical when the order of GPCE approximation is required to be overly large to capture highly nonlinear response characteristics. Furthermore, a global method using a truncated GPCE, obtained by retaining only low-order terms, may not even find a true optimal solution. An appealing substitute, referred to as the multi-point single-step GPCE method, asks for local implementations of the GPCE approximation that are built on subregions of the entire design space. According to this latter method, the original RDO problem is swapped for a series of local RDO problems, where the objective and constraint functions in each local RDO problem represent their multi-point approximations (Toropov et al. 1993). The design solution of an individual local RDO problem, obtained by the single-step GPCE method, constitutes the initial design for the next local RDO problem. Then, the move limits are updated, and the optimization is repeated iteratively until the optimal solution is acquired. Due to the local approach, the multi-point single-step GPCE method is anticipated to solve practical engineering problems using low-order GPCE approximations.

For the rectangular design space

$$\mathcal{D} = \times_{k=1}^{k=M}[d_{k,L}, d_{k,U}] \subseteq \mathbb{R}^M$$

of the RDO problem described in (4), denote by $q' = 1, 2, \ldots, Q'$, $Q' \in \mathbb{N}$, an index representing the q'th subregion of $\mathcal{D}$ with the initial design vector $\mathbf{d}_0^{(q')} = (d_{1,0}^{(q')}, \ldots, d_{M,0}^{(q')})^\intercal$. Given a sizing factor $0 < \beta_k^{(q')} \leqslant 1$, the domain of the q'th subregion is expressed by

$$\mathcal{D}^{(q')} = \bigtimes_{k=1}^{k=M}\left[d_{k,0}^{(q')} - \beta_k^{(q')}\frac{(d_{k,U} - d_{k,L})}{2}, d_{k,0}^{(q')} + \beta_k^{(q')}\frac{(d_{k,U} - d_{k,L})}{2}\right] \subseteq \mathcal{D} \subseteq \mathbb{R}^M, \quad q' = 1, \ldots, Q'.$$

According to the multi-point single-step GPCE method, the RDO problem in (4) is converted to a succession of local RDO problems defined for Q' subregions. For the q'th subregion, the local RDO problem requires one to

$$\begin{aligned}
\min_{\mathbf{d}\in\mathcal{D}^{(q')}\subseteq\mathbb{R}^M} \tilde{c}_{0,m}^{(q')}(\mathbf{d}) &:= G\Big(\mathbb{E}_{\mathbf{g}}[\tilde{h}_{0,m}^{(q')}(\mathbf{Z};\mathbf{r})], \sqrt{\mathrm{var}_{\mathbf{g}}[\tilde{h}_{0,m}^{(q')}(\mathbf{Z};\mathbf{r})]}\Big),\\
\text{subject to } \tilde{c}_{l,m}^{(q')}(\mathbf{d}) &:= \alpha_l\sqrt{\mathrm{var}_{\mathbf{g}}[\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r})]} - \mathbb{E}_{\mathbf{g}}[\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r})] \leqslant 0,\\
d_k &\in \big[d_{k,0}^{(q')} - \beta_k^{(q')}(d_{k,U} - d_{k,L})/2,\ d_{k,0}^{(q')} + \beta_k^{(q')}(d_{k,U} - d_{k,L})/2\big],\\
& l = 1, \ldots, K,\ k = 1, \ldots, M,
\end{aligned} \tag{30}$$

where

$$\begin{aligned}
\mathbb{E}_{\mathbf{g}}\big[\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r})\big] &:= \int_{\tilde{\mathbb{A}}^N} \tilde{h}_{l,m}^{(q')}(\mathbf{z};\mathbf{r}) f_{\mathbf{Z}}\big(\mathbf{z};\mathbf{g}(\mathbf{d})\big)d\mathbf{z},\\
\mathrm{var}_{\mathbf{g}}\big[\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r})\big] &:= \mathbb{E}_{\mathbf{g}}\big[\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r}) - \mathbb{E}_{\mathbf{g}}[\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r})]\big]^2,
\end{aligned}$$

and $\tilde{c}_{l,m}^{(q')}(\mathbf{d})$, $\tilde{y}_{l,m}^{(q')}(\mathbf{X})$, and $\tilde{h}_{l,m}^{(q')}(\mathbf{Z};\mathbf{r})$, $l = 0, 1, \ldots, K$, are mth-order GPCE approximations of $c_l(\mathbf{d})$, $y_l(\mathbf{X})$, and $h_l(\mathbf{Z};\mathbf{r})$, respectively, for the q'th subregion. Furthermore, $d_{k,0}^{(q')} - \beta_k^{(q')}(d_{k,U} - d_{k,L})/2$ and $d_{k,0}^{(q')} + \beta_k^{(q')}(d_{k,U} - d_{k,L})/2$, also known as the move limits, are the lower and upper bounds, respectively, of the subregion $\mathcal{D}^{(q')}$. Here, the iterations with respect to q' are associated with solving local RDO problems and should not be confused with q describing the iteration count for design iterations.

The multi-point single-step GPCE method is schematically depicted in Fig. 1. Here, $\mathbf{d}_*^{(q')}$ is the optimal design

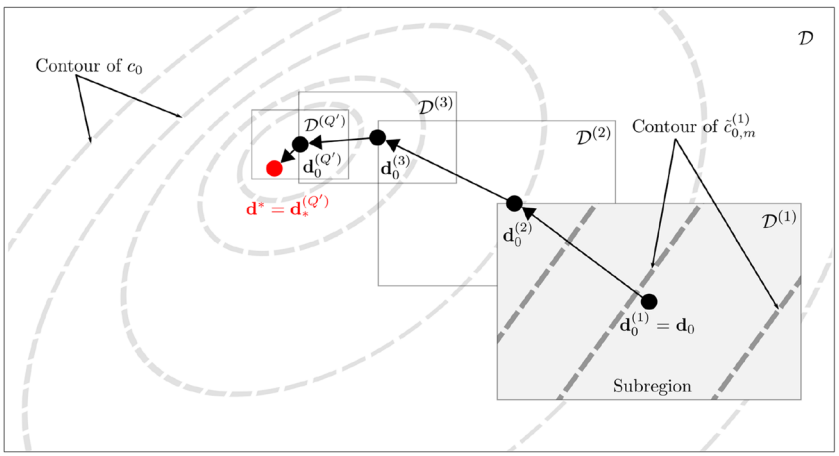

Fig. 1 A schematic description of the multi-point single-step design process during Q' iterations to get the final optimum $\mathbf{d}^*$

solution obtained using the single-step GPCE method for the q'th local RDO problem in (30). By setting the initial design $\mathbf{d}_0^{(q'+1)}$ equal to $\mathbf{d}_*^{(q')}$ at the next local RDO problem on $\mathcal{D}^{(q'+1)}$, the process is repeated until attaining a final, convergent solution $\mathbf{d}^*$. The flow chart of the method is presented in Figs. 2 and 3 with supplementary explanations of each step as follows.

1. Initialize all parameters and tolerances as follows: set termination criteria $0 < \epsilon_1, \epsilon_2 << 1$; set tolerances for sizing sub-regions $0 < \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6, \epsilon_7 < 1$; set size parameters $0 < \beta_k^{(q')} \leqslant 1$, $k = 1, \ldots, M$, of $\mathcal{D}^{(q')}$; set an initial design vector $\mathbf{d}_0^{(q')} = (d_{1,0}^{(q')}, \ldots, d_{M,0}^{(q')})$. The initial design can be either feasible or infeasible to the constraint conditions.
2. Transform the random input vector $\mathbf{X}$ to a new random vector $\mathbf{Z}$ such that $\mathbb{E}_{\mathbf{d}}[Z_{i_k}] = g_k = 0$ or 1, $k = 1, \ldots, M$, via shifting or scaling transformations, respectively.
3. Choose the orders m and m' of GPCE approximations for generic responses and score functions, respectively. Construct an $L_{N,m}$- and $L_{N,m'}$-dimensional vectors of measure consistent orthonormal polynomials $\Psi_m(\mathbf{Z}; \mathbf{g})$ and $\Psi_{m'}(\mathbf{Z}; \mathbf{g})$ via the three-step algorithm.
4. Update the current design vector $\mathbf{d}$ as follows. If $q' = 1$, create input samples $\mathbf{z}^{(l)}$, $l = 1, \ldots, L''$, where $L'' >> L$ and $L'' >> L'$, via the QMCS method. Use the samples to generate input-output data sets $\{\mathbf{z}^{(l)}, h(\mathbf{z}^{(l)}; \mathbf{r})\}_{l=1}^{L}$ of the sample size $L > L_{N,m}$ (say, $L/L_{N,m} \geqslant 3$) and $\{\mathbf{z}^{(l)}, s_k(\mathbf{z}^{(l)}; \mathbf{g})\}_{l=1}^{L'}$ of the sample size $L' > L_{N,m}$ (say, $L'/L_{N,m} \geqslant 3$). If $q' > 1$, reuse the input samples to generate new input-output data sets $\{\mathbf{z}^{(l)}, h(\mathbf{z}^{(l)}; \mathbf{r}')\}_{l=1}^{L}$. In each iteration, use SLS to estimate GPCE coefficients with respect to $\Psi_m(\mathbf{z}; \mathbf{g})$ for a generic response, but for score functions, use SLS to estimate GPCE coefficients with respect to $\Psi_{m'}(\mathbf{z}; \mathbf{g})$ only in the initial iteration ($q' = 1$). If $q' > 1$, reuse the expansion coefficients of score functions. In each iteration, use the second-moment properties to approximate the objective function $\tilde{c}_{0,m}$ and constraint function $\tilde{c}_{l,m}$, $l = 1, \ldots, K$. Also, if $q' = 1$, calculate expectations of the multiple-product of three orthonormal polynomials and preserve it to reuse the values for the next iterations.
5. If $q' = 1$, use the default values of size parameters $0 < \beta_k^{(q')} \leqslant 1$, $k = 1, \ldots, M$, in Step 1. If $q' > 1$ and $s = 1$, determine the size parameters from three conditions: (1) the accuracy of GPCE to approximate the objective and constraint functions, (2) active or inactive conditions of design to boundaries of the subregion, and (3) converging condition to the final optimality. The details of these three conditions are described in the following steps of Fig. 3. Otherwise, skip Step 5.

5-1. (First condition) For all $l = 0, 1, \ldots, K$, if $||\tilde{c}_{l,m}^{(q')}(\mathbf{d}_0^{(q')}) - \tilde{c}_{l,m}^{(q'-1)}(\mathbf{d}_0^{(q')})|| \leqslant \epsilon_3$, then increase all $\beta_k^{(q')}$, $k = 1, \ldots, M$. Otherwise, go to Step 5-2.

5-2. (First condition) For any $l = 0, 1, \ldots, K$, if $||\tilde{c}_{l,m}^{(q')}(\mathbf{d}_0^{(q')}) - \tilde{c}_{l,m}^{(q'-1)}(\mathbf{d}_0^{(q')})|| > \epsilon_4$, then decrease all $\beta_k^{(q')}$, $k = 1, \ldots, M$. Otherwise, go to Step 5-3.

5-3. (Second condition) If $||d_{k,0}^{(q')} - d_{k,L}^{(q'-1)}|| \leqslant \epsilon_5$ or $||d_{k,0}^{(q')} - d_{k,U}^{(q'-1)}|| \leqslant \epsilon_5$, increase $\beta_k^{(q')}$. Otherwise, go to Step 5-4.

5-4. (Third condition) If $||d_{k,0}^{(q')} - d_{k,0}^{(q'-1)}|| \leqslant \epsilon_6$, decrease $\beta_k^{(q')}$. Otherwise, go to Step 5-5.

5-5. (Move limit) If $\beta_k^{(q')}(d_{k,U} - d_{k,L}) < \epsilon_7$, set $\beta_k^{(q')} = \epsilon_7/(d_{k,U} - d_{k,L})$. Otherwise, increase k and repeat the process until satisfying the loop condition $k \leqslant M$.

6. If the current design is infeasible to constraint conditions, go to Step 7. Otherwise, set the current feasible design $\mathbf{d}_f^{(q')} = \mathbf{d}$, then go to Step 8.
7. Interpolate between the current infeasible design $\mathbf{d}$ and the previous feasible design $\mathbf{d}_f^{(q'-1)}$. If an initial design (at $q' = 1$) is infeasible, interpolate with upper or lower bounds of the design space depending on problems at hand. One can follow the golden ratio, about 1.618, to prevent excessive withdrawal of a solution during interpolation.
8. If a termination condition is satisfied, such that $\|\mathbf{d}_f^{(q')} - \mathbf{d}_f^{(q'-1)}\| \leqslant \epsilon_1$ or $\|\tilde{c}_{0,m}^{(q')}(\mathbf{d}_f^{(q')}) - \tilde{c}_{0,m}^{(q')}(\mathbf{d}_f^{(q'-1)})\| \leqslant \epsilon_2$, set $\mathbf{d}_f^{(q')}$ to the final optimal design $\mathbf{d}^*$ and terminate the optimization process. Otherwise, go to Step 9.
9. Solve the q'th local RDO problem with the single-step GPCE using a gradient-based algorithm (e.g., sequential quadratic programming), obtaining the local optimal solution $\mathbf{d}_*^{(q')}$. Increase the sub-region or iteration count q'. Set $\mathbf{d}_0^{(q')} = \mathbf{d}_*^{(q'-1)}$ and go to Step 4.

6 Numerical examples

Five numerical examples are presented to illustrate the proposed RDO methods as follows: the direct GPCE

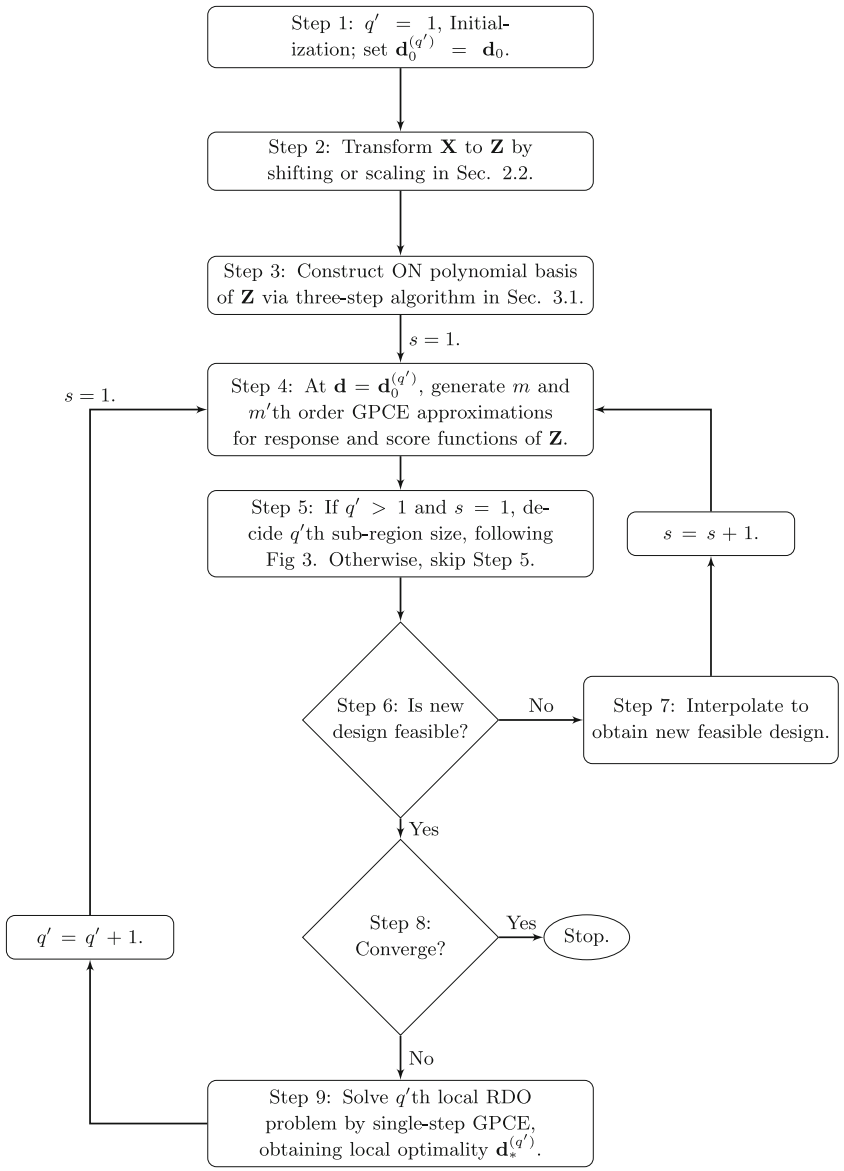

Fig. 2 A flow chart of the multi-point single-step GPCE

method in Examples 1, 3, and 4; the single-step GPCE method in Examples 1 and 2; and the multi-point single-step GPCE method in Examples 3–5. To illustrate multiple choices for scalarization, the weighted sum and Tchebycheff approaches were employed in Example 2. All other examples used a single objective function or the weighted sum approach. The objective and constraint functions in these examples are elementary mathematical functions or derived from practical applications, from simple truss problems to an industrial-scale steering knuckle problem from the automotive industry. Both size and shape design problems, in the context of RDO, were studied. In all examples, the design variables are the statistical means of some or all input random variables.

In all examples, each component of $\mathbf{g}$ is either *zero* or *one*, depending on whether the shifting or scaling transformation is employed. The multivariate orthonormal polynomials consistent with the probability measure of $\mathbf{Z}$ were constructed using the three-step algorithm. The monomial moment matrix was estimated using a 13-point Gauss quadrature in Examples 1 and 2 and QMCS with a sample size of $L'' = 5 \times 10^6$ in conjunction with the Sobol sequence (Lee and Rahman 2020) in Examples 3–5. The GPCE orders $(m, m', \bar{m})$ and sample sizes $(L, L', L'', \bar{L})$ vary from example to example, depending on the objective or constraint functions and score functions at hand. Table 2 lists the specific values used in all five examples. For estimating the GPCE coefficients, SLS regression were used

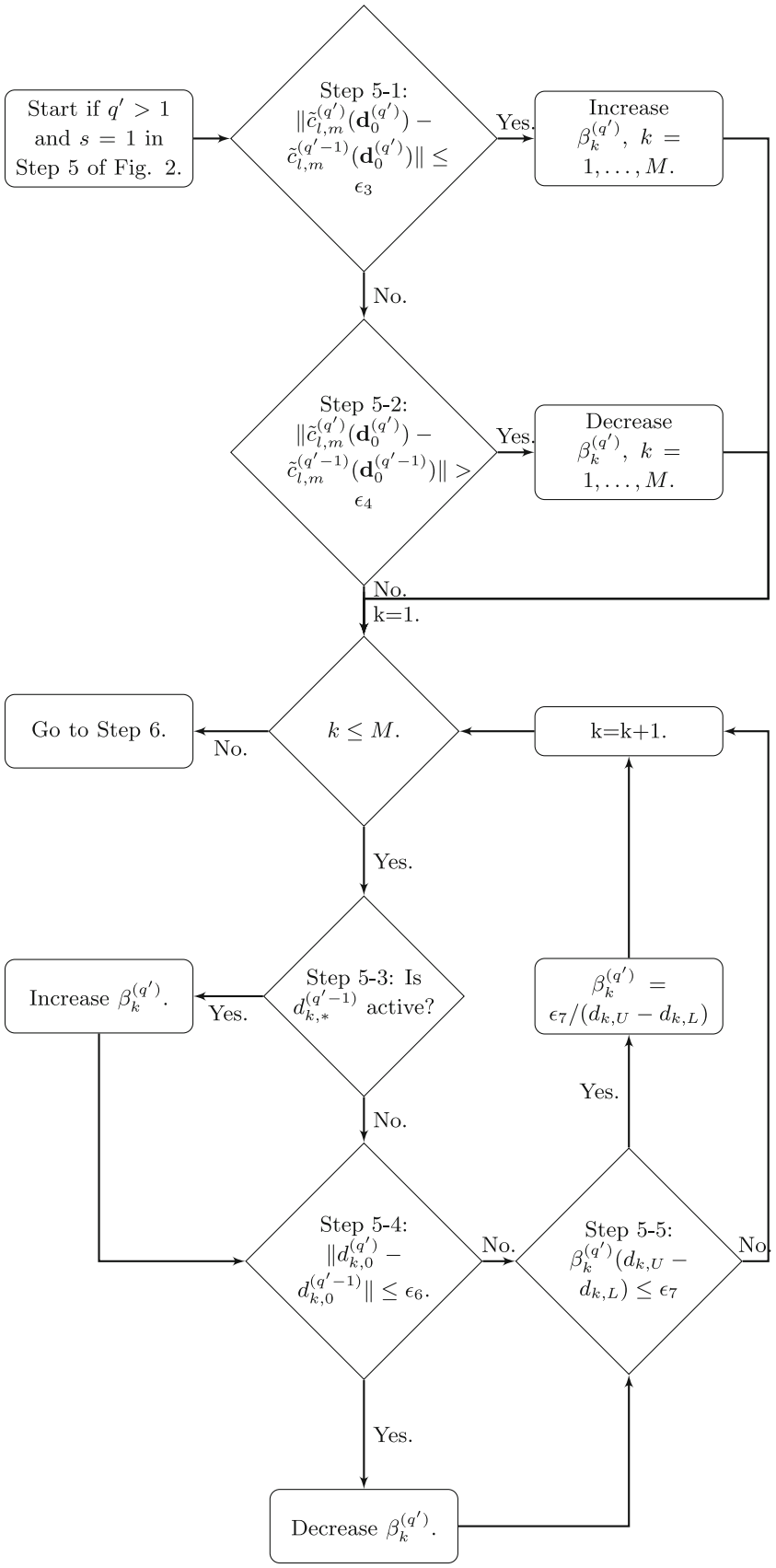

Fig. 3 A flow chart of sizing the q'th sub-region in the multi-point single-step GPCE

Table 2 The list of parameters (Examples 1–5): GPCE orders (m, m', $\bar{m}$) and sample sizes (L, L', L'', $\bar{L}$)

Examples 1 and 2

	$m^{(a)}$		m' (b)	$L^{(c)}$		L' (d)	L'' (e)
Methods	y_0	y_1	s_k	y_0	y_1	s_k	–
Direct GPCE	4	1	1	45	9	30	1×10^6
Single-step GPCE	4	1	1	45	9	30	1×10^6

Example 3

		$m^{(a)}$	m' (b)	$\bar{m}$ (f)	$L^{(c)}$	L' (d)	L'' (e)	$\bar{L}$ (g)
Methods		y_0, y_1, y_2	s_k	y_0, y_1, y_2	y_0, y_1, y_2	s_k	–	–
Direct GPCE								
	Option 1	3	1	–	168	60	2×10^6	–
		4	1	–	378	60	2×10^6	–
	Option 2	–	–	3	–	–	–	1800
		–	–	4	–	–	–	1980
Multi-point single-step GPCE		1	1	–	18	60	2×10^6	–

Example 4

		$m^{(a)}$	m' (b)	$L^{(c)}$	L' (d)	L'' (e)
Methods		$y_l, l = 0 - 11$	s_k	$y_l, l = 0 - 11$	s_k	–
Direct GPCE						
	SLS	2	3	198	2860	2×10^6
		3	3	858	2860	2×10^6
	Partitioned D-MORPH	2	3	100	2860	2×10^6
		3	3	200	2860	2×10^6
Multi-point single-step GPCE		1	3	33	2860	2×10^6

Example 5

	$m^{(a)}$		m' (b)	$L^{(c)}$	L' (d)	L'' (e)
Method	y_0	y_1, y_2	s_k	y_0 y_1, y_2	s_k	–
Multi-point single-step GPCE	1	1	3	21 33	840	2×10^6

[a]The degree of GPCE approximation for an output response $y_l,\ l = 0, \ldots, K,\ 0 \leqslant K < \infty$

[b]The degree of GPCE approximation for a score function s_k

[c]The sample size of input-output data set for expansion coefficients of an output response $y_l,\ l = 0, \ldots, K,\ 0 \leqslant K < \infty$

[d]The sample size of input-output data set for expansion coefficients of a score function s_k

[e]The sample size for expectation of multiple product of three orthonormal polynomials

[f]The degree of GPCE approximation for integrand in (24) and (25)

[g]The sample size of input-output data set for expansion coefficients of an integrand in (24) and (25)

in all five examples, whereas the partitioned D-MORPH was employed only in Example 4. The input data in the experimental design were generated using QMCS.

For the gradient-based optimization, the sequential quadratic programming was chosen in all examples. In the multi-point single-step GPCE, the tolerances and initial scale parameter are as follows: $\epsilon_1 = 1 \times 10^{-6}$, $\epsilon_2 = 1 \times 10^{-6}$, $\epsilon_3 = 0.01$, $\epsilon_4 = 0.07$, $\epsilon_5 = 0.01$, $\epsilon 6 = 0.5$, $\epsilon_7 = 0.05$, and $\beta_k^{(1)} = 0.3,\ k = 1, \ldots, M$, in Examples 3–5.

All numerical results were generated using MATLAB (version 2019b) (MATLAB 2019), CREO parametric (version 4.0) (CREO 2016), and ABAQUS (version 2019) (ABAQUS 2019) on an Intel Core i7-7700K 4.20 GHz processor with 64 GB of RAM.

6.1 Example 1: Optimization of a mathematical function

Consider a mathematical problem involving a two-dimensional Gaussian random vector $\mathbf{X} = (X_1, X_2)^\intercal$ with dependent components, which have means $\mathbb{E}_{\mathbf{d}}[X_1] = d_1$ and $\mathbb{E}_{\mathbf{d}}[X_2] = d_2$. Given the design vector $\mathbf{d} = (d_1, d_2)^\intercal$, the objective of this example is to

$$\min_{\mathbf{d}\in\mathcal{D}} \; c_0(\mathbf{d}) := \frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}}{\sigma_0^*},$$
$$\text{subject to } c_1(\mathbf{d}) := 3\sqrt{\mathrm{var}_{\mathbf{d}}[y_1(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})] \leqslant 0,$$
$$0 \leqslant d_1 \leqslant 10, \; 0 \leqslant d_2 \leqslant 10,$$

where

$$y_0(\mathbf{X}) = (X_1 - 4)^3 + (X_1 - 3)^4 + (X_2 - 5)^2 + 10, \quad (31)$$

and

$$y_1(\mathbf{X}) = X_1 + X_2 - 6.45 \quad (32)$$

are two random output functions of $\mathbf{X}$. The initial design vector $\mathbf{d}_0 = (5, 5)^\intercal$. The approximate optimal solution is denoted by $\tilde{\mathbf{d}}^* = (\tilde{d}_1^*, \tilde{d}_2^*)^\intercal$.

Two distinct cases of dependent variables, demonstrating the respective needs of the shifting (Case 1) and scaling (Case 2) transformations, were examined:

1. The standard deviations of X_1 and X_2 are the same as 0.4. The correlation coefficient between X_1 and X_2 is 0.4. The normalizing factor $\sigma_0^* = 17$.
2. The standard deviations of X_1 and X_2 are $0.15d_1$ and $0.15d_2$, respectively. The correlation coefficient between X_1 and X_2 is -0.5. The normalizing factor $\sigma_0^* = 45$.

Formerly studied by Lee et al. (2009) and Ren and Rahman (2013) for independent Gaussian variables, this example was slightly modified by defining two cases of correlated Gaussian variables.

Table 3 presents the means and variances of $y_0(\mathbf{X})$ and $y_1(\mathbf{X})$, including their first-order design sensitivities, by GPCE approximations, at the initial design $\mathbf{d}_0 = (5, 5)^\intercal$. For the sensitivity analysis by GPCE and score functions, Option 1 was employed. The shifting and scaling transformations were applied in Cases 1 and 2, respectively. When compared with the respective exact solutions, which

Table 3 The results of second moment properties and sensitivities of y_0 and y_1 at $\mathbf{d}_0 = (5, 5)$ (Example 1)

	Case 1 (shifting)		Case 2 (scaling)	
(1) Response function: $y_0(X_1, X_2) = (X_1 - 4)^3 + (X_1 - 3)^4 + (X_2 - 5)^2 + 10$				
Results	GPCE approx.[a]	Exact[b]	GPCE approx.[a]	Exact[b]
$\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]$	31.5568	31.5568	43.6992	43.6992
$\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]$[c]	289.9119	289.9119	2099.8191	2099.8191
$\partial\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]/\partial d_1$	39.32	39.32	50.1875	50.1875
$\partial\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]/\partial d_2$	1.8754×10^{-13}	0	-1.1660×10^{-9}	0
$\partial\mathbb{E}_{\mathbf{d}}[y_0^2(\mathbf{X})]/\partial d_1$	3264.3591	3264.7502	8939.0636	8939.7114
$\partial\mathbb{E}_{\mathbf{d}}[y_0^2(\mathbf{X})]/\partial d_2$	10.6860	10.0659	−54.0485	−56.4609
No. of y_0 evaluations	45	–	45	–
(2) Response function: $y_1(X_1, X_2) = X_1 + X_2 - 6.45$				
Results	GPCE approx.[d]	Exact[b]	GPCE approx.[d]	Exact[b]
$\mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})]$	3.5500	3.5500	3.5500	3.5500
$\mathrm{var}_{\mathbf{d}}[y_1(\mathbf{X})]$[c]	0.4480	0.4480	0.5625	0.5625
$\partial\mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})]/\partial d_1$	1.0000	1	1.0000	1
$\partial\mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})]/\partial d_2$	1.0000	1	1.0000	1
$\partial\mathbb{E}_{\mathbf{d}}[y_1^2(\mathbf{X})]/\partial d_1$	7.1004	7.1	7.1000	7.1
$\partial\mathbb{E}_{\mathbf{d}}[y_1^2(\mathbf{X})]/\partial d_2$	7.1000	7.1	7.0997	7.1
No. of y_0 evaluations	9	–	9	–

[a]The order (m) of GPCE is four

[b]The exact closed forms of sensitivities are used

[c]$\mathrm{var}_{\mathbf{d}}[y_l(\mathbf{x})] := \mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X}) - \mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X})]]^2, l = 0, 1$

[d]The order (m) of GPCE is one

exist for these two functions and are also reported in Table 3, the GPCE estimates of response moments and their design sensitivities are excellent.

Table 4 summarizes the approximate optimal solutions for Cases 1 and 2, including the requisite numbers of design iterations and function evaluations, by the direct GPCE and single-step GPCE methods. For comparison, the exact solutions, obtained employing the exact expressions of objective and constraint functions and their design sensitivities, are also included. According to Table 4, both design methods yield identical optimal solutions in seven to nine iterations. This is possible as the selected orders of GPCE approximations in both methods are the same. More importantly, both design methods deliver optimization results remarkably close to the exact optimal solutions. Hence, each method can be used to solve this optimization problem. However, the numbers of function evaluations required to reach optimal solutions reduce dramatically when the single-step GPCE is employed. This is because the chosen GPCE approximation at the initial design is adequate for the entire design space, facilitating accurate calculations of the GPCE coefficients by exploiting (29) for any design. In this case, the coefficients need to be calculated only once during all design iterations.

Lastly, Table 4 also incorporates the optimization results for both cases when the input variables are statistically independent. The results are quite different than those when the input variables are dependent, especially when Case 2 is considered. Therefore, dependence or correlation in input random variables, if it exists, should be accounted for in design optimization under uncertainty.

6.2 Example 2: Bi-objective optimization of a mathematical function

The second example involves the bi-objective version of the first example, requiring one to

$$\begin{aligned} \min_{\mathbf{d}\in\mathcal{D}} \quad & \{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})],\ \sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}\}, \\ \text{subject to} \quad & 3\sqrt{\mathrm{var}_{\mathbf{d}}[y_1(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})] \leqslant 0, \\ & 0 \leqslant d_1 \leqslant 10,\ 0 \leqslant d_2 \leqslant 10, \end{aligned} \tag{33}$$

Table 4 Optimization results of mathematical formulations (Example 1)

Results	Direct GPCE	Single-step GPCE	Exact [c]	Exact[c]
Case 1 (shifting)	Dependent[a]			Independent[b]
$\tilde{d}_1{}^*$	3.3908	3.3908	3.3906	3.3577
$\tilde{d}_2{}^*$	5.0672	5.0672	5.0673	5.0000
$c_0(\tilde{\mathbf{d}}^*)$	0.0682	0.0682	0.0682	0.0667
$c_1(\tilde{\mathbf{d}}^*)$	7.7539×10^{-8}	7.7539×10^{-8}	-6.6613×10^{-15}	-0.2107
$\sqrt{\mathrm{var}_{\tilde{\mathbf{d}}^*}[y_0(\mathbf{X})]}$	1.1592	1.1592	1.1592	1.1338
No. of iterations	7	7	7	7
No. of y_0 evaluations	585	45	–	–
No. of y_1 evaluations	117	9	–	–
Case 2 (scaling)	Dependent[d]			Independent[b]
$\tilde{d}_1{}^*$	3.1964	3.1964	3.1964	3.4239
$\tilde{d}_2{}^*$	5.3978	5.3978	5.3976	6.1912
$c_0(\tilde{\mathbf{d}}^*)$	0.0377	0.0377	0.0377	0.0721
$c_1(\tilde{\mathbf{d}}^*)$	-0.0287	-0.0287	-0.0286	0.0186
$\sqrt{\mathrm{var}_{\tilde{\mathbf{d}}^*}[y_0(\mathbf{X})]}$	1.6987	1.6987	1.6987	3.2449
No. of iterations	9	9	9	7
No. of y_0 evaluations	1890	45	–	–
No. of y_1 evaluations	378	9	–	–

[a] X_1 and X_2 are mutually dependent with the correlation coefficient of 0.4

[b] X_1 and X_2 are independent

[c] Exact closed forms of objective, constraint, and their gradient functions are used

[d] X_1 and X_2 are mutually dependent with the correlation coefficient of -0.5

where $\mathbf{X} = (X_1, X_2)^T$ is a dependent Gaussian vector as before and y_0 and y_1 are defined in (31) and (32), respectively. The initial design vector $\mathbf{d}_0 = (5, 5)^\intercal$. The standard deviations of X_1 and X_2 are the same as 0.4; and the correlation coefficient between X_1 and X_2 is 0.4.

The objective of this example is to evaluate the single-step GPCE method for obtaining Pareto solutions in two distinct scalarization approaches, comprising the weighted sum approach and the weighted Tchebycheff approach.

6.2.1 The weighted sum approach

The bi-objective functions in (33) are linearly aggregated by weights $w_1 \in \mathbb{R}_0^+$ and $w_2 \in \mathbb{R}_0^+$, such that $w_1 + w_2 = 1$. Then, the weighted sum problem demands one to

$$\begin{aligned} \min_{\mathbf{d}\in\mathcal{D}} \; c_0(\mathbf{d}) := w_1 \left(\frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]}{31.5568}\right) \\ + w_2 \left(\frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}}{17.0268}\right), \\ \text{subject to } c_1(\mathbf{d}) := 3\sqrt{\mathrm{var}_{\mathbf{d}}[y_1(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})] \leqslant 0. \end{aligned} \tag{34}$$

6.2.2 The weighted Tchebycheff approach

To define the weighted Tchebycheff problem, a reference point was obtained as $(\mu_f^* \simeq 4.4307,\ \sigma_f^* \simeq 1.1592)$, where $\mu_f^* = \mathbb{E}_{\tilde{\mathbf{d}}_1^*}[y_0(\mathbf{X})]$ at $\tilde{\mathbf{d}}_1^* = (1.5713, 6.8867)$ and $\sigma_f^* = \sqrt{\mathrm{var}_{\tilde{\mathbf{d}}_2^*}[y_0(\mathbf{X})]}$ at $\tilde{\mathbf{d}}_2^* = (3.3934, 5.0646)$. Then, the scalarized objective function is defined as

$$\begin{aligned} c_0(\mathbf{d}) := \max \Bigg[w_1 \left(\frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})] - 4.4307}{31.5568}\right), \\ w_2 \left(\frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]} - 1.1592}{17.0268}\right) \Bigg]. \end{aligned} \tag{35}$$

The so-called min-max problem in (35) was tackled by treating the bi-objective functions as additional side constraints whose values are bounded by a real-valued scalar variable $\lambda \in \mathbb{R}_{\mathbf{0}}^+$ and asking to

$$\begin{aligned} \min_{\substack{\mathbf{d}\in\mathcal{D} \\ \lambda\in\mathbb{R}_0^+}} \quad & \lambda, \\ \text{subject to} \quad & w_1 \frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})] - 4.4307}{31.5568} - \lambda \leqslant 0, \\ & w_2 \frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]} - 1.1592}{17.0268} - \lambda \leqslant 0, \\ & 3\sqrt{\mathrm{var}_{\mathbf{d}}[y_1(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})] \leqslant 0. \end{aligned} \tag{36}$$

To solve (34) and (36), the single-step GPCE was applied to obtain Pareto optimal solutions for nine cases of evenly distributed combinations of weights w_1 and w_2.

Table 5 summarizes the results of the weighted Tchebycheff approach and the weighted sum approach for the aforementioned nine cases of w_1 and w_2. Comparing Table 5a and b, the Pareto solutions $\tilde{d}_1^*$ and $\tilde{d}_2^*$ between two scalarization approaches are noticeably different in each case of the combination of weights. The corresponding values of the mean and the standard deviation generated from the two scalarization approaches, listed in the fifth and sixth columns, respectively, of Table 5a and b, are also remarkably different. However, whether using the weighted Techebycheff approach or the weighted sum approach, the single-step GPCE needs only a single stochastic analysis during the design process to obtain all Pareto solutions. Indeed, to complete all optimizations of nine cases of w_1 and w_2, the requisite numbers of function evaluations for y_0 and y_1 are still 45 and 9. Thus, the single-step GPCE method is computationally expedient in estimating Pareto solutions of RDO problems, regardless of the function $G(\cdot, \cdot)$ selected.

Figure 4a and b present the plots of the mean $\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$ versus the standard deviation $\sqrt{\mathrm{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$, obtained using the weighted Tchebycheff and weighted sum approaches, respectively. According to Fig. 4a, the weighted Tchebycheff approach yields an evenly distributed trade-off relationship between the mean and standard deviation. On the other hand, as displayed in Fig. 4b, the weighted sum approach explores mostly two corners of the Pareto front. This is because the weighted sum approach has a limit to capture the Pareto solutions on non-convex portions of the Pareto optimal curve. In that case, the weighted Tchebycheff approach is clearly a better choice, and the single-step GPCE performs effectively, requiring the same computational efforts in terms of function evaluations for both scalarization approaches.

6.3 Example 3: Optimal sizing design of a two-bar truss structure

The third example was created by recasting a two-bar truss problem previously investigated by Ramakrishnan and Rao (1996) and Ren and Rahman (2013). Depicted in Fig. 5, there are seven random variables in this problem: cross-sectional areas X_1 and X_2; half-span lengths X_3 and X_4; mass density X_5; yield strength X_6; and load magnitude X_7. The second-moment characteristics and marginal probability distribution of each random variable are described in Table 6.

Additionally, two random variable pairs (X_1, X_2) and (X_3, X_4) are statistically dependent with correlation coefficients of 0.4 and -0.4, respectively. There are four design variables, as follows: $d_1 = \mathbb{E}_{\mathbf{d}}[X_1]$, $d_2 = \mathbb{E}_{\mathbf{d}}[X_2]$, $d_3 = \mathbb{E}_{\mathbf{d}}[X_3]$, and $d_4 = \mathbb{E}_{\mathbf{d}}[X_4]$. The objective is to minimize the second-moment properties of the mass of the structure subject to two constraints, limiting axial stresses of both

Table 5 Bi-objective optimization results of a mathematical function (Example 2)

Weights		Pareto solutions		Mean	Standard deviation	No. of function evaluations	
(a) The weighted Tchebycheff approach							
w_1	w_2	$\tilde{d}_1^*$	$\tilde{d}_2^*$	$\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$	$\sqrt{\mathrm{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$	y_0	y_1
0.1	0.9	3.1092	5.3488	9.2355	1.4473	45	9
0.2	0.8	2.9545	5.5035	8.8476	1.7550	—[a]	—[a]
0.3	0.7	2.8076	5.6504	8.4288	2.0837	—[a]	—[a]
0.4	0.6	2.6583	5.7997	7.9427	2.4225	—[a]	—[a]
0.5	0.5	2.5040	5.9540	7.3773	2.7491	—[a]	—[a]
0.6	0.4	2.3449	6.1131	6.7436	3.0311	—[a]	—[a]
0.7	0.3	2.1829	6.2751	6.0772	3.2321	—[a]	—[a]
0.8	0.2	1.4281	7.0299	4.5877	2.8772	—[a]	—[a]
0.9	0.1	1.4278	7.0302	4.5884	2.8772	—[a]	—[a]
(b) The weighted sum approach							
w_1	w_2	$\tilde{d}_1^*$	$\tilde{d}_2^*$	$\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$	$\sqrt{\mathrm{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$	y_0	y_1
0.1	0.9	3.3748	5.0832	9.8539	1.1603	45	9
0.2	0.8	3.3546	5.1034	9.8053	1.1650	—[a]	—[a]
0.3	0.7	3.3278	5.1302	9.7420	1.1766	—[a]	—[a]
0.4	0.6	3.2895	5.1685	9.6530	1.2029	—[a]	—[a]
0.5	0.5	3.2262	5.2318	9.5075	1.2688	—[a]	—[a]
0.6	0.4	2.9975	5.4605	8.9600	1.6639	—[a]	—[a]
0.7	0.3	1.5214	6.9366	4.4488	2.9264	—[a]	—[a]
0.8	0.2	1.5386	6.9194	4.4384	2.9432	—[a]	—[a]
0.9	0.1	1.5555	6.9025	4.4325	2.9612	—[a]	—[a]

[a]After obtaining GPCE coefficients for the first case ($w_1 = 0.1, w_2 = 0.9$), the single-step GPCE recycled the coefficients for the remaining eight cases, thus requiring no additional function evaluations

members at or below the yield strength of the material with 99.875% probability if y_l, $l = 1, 2$, are standard Gaussian. Here, the RDO problem is devised to

$$\begin{aligned}
\min_{\mathbf{d}\in\mathcal{D}\subseteq\mathbb{R}^M} \; & c_0(\mathbf{d}) := 0.5\frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]}{56.5744} + 0.5\frac{\sqrt{\mathrm{var}_{\mathbf{d}}[y_0(\mathbf{X})]}}{17.0059}, \\
\text{subject to } & c_1(\mathbf{d}) := 3\sqrt{\mathrm{var}_{\mathbf{d}}[y_1(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})] \leqslant 0, \\
& c_2(\mathbf{d}) := 3\sqrt{\mathrm{var}_{\mathbf{d}}[y_2(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_2(\mathbf{X})] \leqslant 0, \\
& 2\ \mathrm{cm}^2 \leqslant d_1, d_2 \leqslant 25\ \mathrm{cm}^2, \\
& 0.3\ \mathrm{m} \leqslant d_3, d_4 \leqslant 1.4\ \mathrm{m},
\end{aligned}$$

where

$$y_0(\mathbf{X}) = X_5\left(X_1\sqrt{1+X_3^2} + X_2\sqrt{1+X_4^2}\right),$$

and

$$y_1(\mathbf{X}) = 1 - \frac{10\sqrt{1+X_3^2}(1+8X_4)X_7}{\sqrt{65}X_1(X_3+X_4)X_6},$$

and

$$y_2(\mathbf{X}) = 1 - \frac{10\sqrt{1+X_4^2}(-1+8X_3)X_7}{\sqrt{65}X_2(X_3+X_4)X_6}$$

are three random response functions of $\mathbf{X}$. They represent the total mass of two bars, the axial stress on the left bar, and the axial stress on the right bar, respectively. The design vector $\mathbf{d} = (d_1, d_2, d_3, d_4)^\intercal$ with the initial value $\mathbf{d}_0 = (20\ \mathrm{cm}^2, 20\ \mathrm{cm}^2, 1\ \mathrm{m}, 1\ \mathrm{m})^T$. The corresponding mean and standard deviation of $y_0(\mathbf{X})$ at the initial design, calculated by the QMCS method with 5×10^5 samples, are 56.5744 kg and 17.0059 kg, respectively. The approximate optimal solution is denoted by $\tilde{\mathbf{d}}^* = (\tilde{d}_1^*, \tilde{d}_2^*, \tilde{d}_3^*, \tilde{d}_4^*)^T$.

Three design methods comprising direct GPCE, multi-point single-step GPCE, and QMCS were applied to solve this RDO problem. In the direct GPCE method, both Options 1 and 2 were used for obtaining design sensitivities. Several orders of GPCE approximations were employed in the GPCE-based methods. The QMCS method was conducted with the motivation of providing a benchmark solution. The sample size of QMCS is 5×10^5 for stochastic

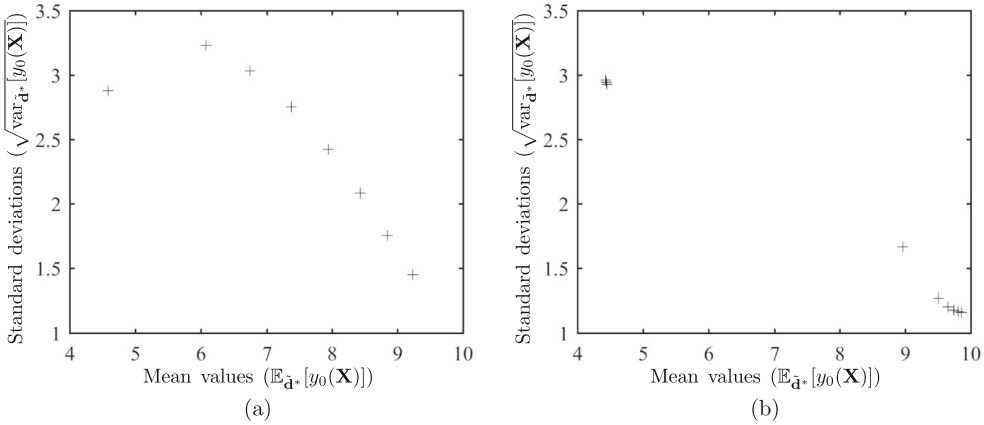

Fig. 4 Pareto optimal set of mean and standard deviation of y_0 for RDO problem (Example 2): **a** the weighted Tchebycheff approach; **b** the weighted sum approach

analysis, where design sensitivities were calculated from the central finite-difference approximation. The results of all three methods are outlined in Table 7.

According to Table 7, the optimal solutions by both variants of the GPCE-based method and QMCS are very close to each other, all indicating that the constraints are active ($c_1 \simeq 0$, $c_2 \simeq 0$). The individual results of the direct GPCE method generated for the third-order and fourth-order GPCE approximations are practically convergent, regardless of which option is selected for calculating design sensitivities. However, when comparing the numbers of function evaluations, Option 2 requires more function evaluations than Option 1. This is primarily because, although $\bar{m} = m$ was used in (22), a larger set of experimental design was required to estimate the coefficients $H_{k,i}^{(1)}(\mathbf{r})$ and $H_{k,i}^{(2)}(\mathbf{r})$. Conversely, though, the CPU times in Option 2 (11–23 s) are substantially lower than those in Option 1 (3558–19,935 s), as the computational expense in calculating numerous expectations of the triple product dominates the time and effort needed to evaluate the functions. Therefore, Option 2 is more economical than Option 1 when the objective and constraint functions are simple enough for rapid evaluations, as is the case in this problem. However, for more realistic RDO problems mandating FEA-generated functions, to be illustrated in the next two examples, the computational advantage of Option 2 may disappear, rendering Option 1 to be more effective than the other.

Since this problem was also tackled by the multi-point single-step GPCE method, a few additional comments on this method are in order. When compared with the results of the higher-order ($m = 4$) direct GPCE method and QMCS, the first-order ($m = 1$) multi-point single-step GPCE method produces good estimates of the optimal solution as well. Although there is some discrepancy in the values of $\tilde{d}_1^*$ and $\tilde{d}_2^*$ calculated by the latter method, any differences

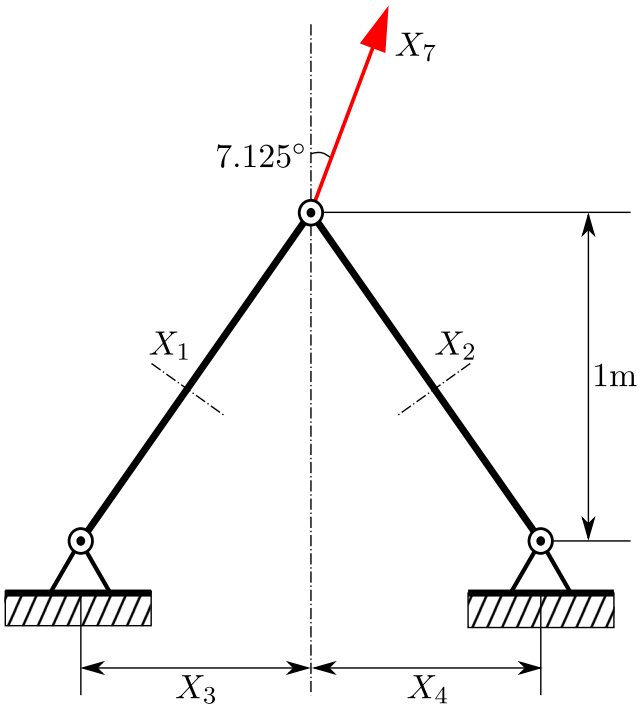

Fig. 5 A two-bar truss structure

Table 6 Statistical properties of random variables of two-bar truss (Example 3)

Random variables	Mean	Standard deviation	Probability distribution
Cross-sectional area (X_1), cm$^{2(a)}$	d_1	$0.02d_1$	Gaussian
Cross-sectional area (X_2), cm$^{2(a)}$	d_2	$0.02d_2$	Gaussian
Half-horizontal span (X_3), m$^{(b)}$	d_3	$0.02d_3$	Gaussian
Half-horizontal span (X_4), m$^{(b)}$	d_4	$0.02d_4$	Gaussian
Mass density (X_5), kg/m^3	10,000	3,000	Weibull
Yield strength (X_6), MPa	2,050	488	Gumbel
Load magnitude (X_7), kN	200	50	Gumbel

[a] X_1 and X_2 are mutually dependent with the correlation coefficient of 0.4

[b] X_3 and X_4 are mutually dependent with the correlation coefficient of -0.4

Table 7 Optimization results of a two-bar truss (Example 3)

	Direct GPCE				Multi-point	
	Option 1[(a)]		Option 2[(b)]		single-step GPCE[(c)]	
Results	$m = 3$	$m = 4$	$m = 3$	$m = 4$	$m = 1$	Benchmark[(d)]
$\tilde{d}_1{}^*$, cm^2	15.3760	15.2294	15.2056	15.1852	15.8173	15.1853
$\tilde{d}_2{}^*$, cm^2	6.3489	6.2758	6.2700	6.2623	6.4966	6.2628
$\tilde{d}_3{}^*$, m	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000
$\tilde{d}_4{}^*$, m	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000
$c_0(\tilde{\mathbf{d}}^*)$	0.4009	0.3969	0.3963	0.3958	0.4111	0.3958
$c_1(\tilde{\mathbf{d}}^*)$	1.3874×10^{-7}	4.5317×10^{-8}	-2.0120×10^{-10}	4.9160×10^{-12}	-1.7285×10^{-5}	8.8818×10^{-16}
$c_2(\tilde{\mathbf{d}}^*)$	-2.4614×10^{-7}	-3.8889×10^{-7}	-6.2194×10^{-6}	-1.0281×10^{-6}	5.5536×10^{-7}	-2.8155×10^{-13}
$\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$, kg	22.6817	22.4522	22.4214	22.3921	23.3319	22.3940
$\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$, kg	6.8176	6.7487	6.7394	6.7295	6.9706	6.7312
No. of iterations	21	19	25	17	193	18
No. of y_0 evaluations	14,760	36,630	88,200	81,180	792	1.665×10^8
No. of y_1 evaluations	14,760	36,630	88,200	81,180	792	1.665×10^8
No. of y_2 evaluations	14,760	36,630	88,200	81,180	792	1.665×10^8

[a]Option 1 is the approximate sensitivity method of the first two moments in (19) and (20)

[b]Option 2 is the approximate sensitivity method of the first two moments in (24) and (25)

[c]Option 1 of approximate sensitivities is used

[d]QMCS method of 5×10^5 samples is employed for stochastic and sensitivity analysis based the central finite-difference method

in the objective functions at respective optima by all three methods are negligibly small. More notably, the multi-point single-step GPCE method delivers an acceptable design solution, asking for a fraction of the function evaluations required by any other methods.

6.4 Example 4: Optimal sizing design of a ten-bar truss structure

In the fourth example, a linear-elastic ten-bar truss, studied by Elishakoff et al. (1994) and Lee and Rahman (2020), was redesigned to evaluate the direct GPCE and multi-point single-step GPCE methods. As shown in Fig. 6, the truss is simply supported at nodes 1 and 4 and is subjected to two vertically downward concentrated forces of 100,000 lb at nodes 2 and 3 and a horizontal concentrated force of 400,000 lb at node 3. The material is aluminium alloy, which has Young's modulus of 10^7 psi and a mass density of 0.1 lb/in^3. There are ten ($N = 10$) random variables $\mathbf{X} = (X_1, \ldots, X_{10})^\intercal$, representing random cross-sectional areas of all ten bars. Modeled as correlated lognormal random variables, they have means $\mathbb{E}_{\mathbf{d}}[X_i]$, $i = 1, \ldots, 10$; standard deviations equal to $0.05\mathbb{E}_{\mathbf{d}}[X_i]$, $i = 1, \ldots, 10$; and correlation coefficients $\rho_{ij} = 0.3997$, $i, j = 1, \ldots, 10$, $i \neq j$. There are ten design variables, as follows: $d_k = \mathbb{E}_{\mathbf{d}}[X_k]$, $k = 1, \ldots, 10$. The objective is to minimize the second-moment properties of the mass of the entire truss structure, constrained by specifying the upper limits of the vertical displacement (v_2) at node 2 and axial stresses σ_i, $i = 1, \ldots, 10$, at all ten bars, such that the limits are satisfied with 99.865% probability if the distribution of each response $y_l(\mathbf{X})$, $l = 1, \ldots, 11$, is standard Gaussian. More specifically, the RDO problem is defined to

$$\min_{\mathbf{d} \in \mathcal{D} \subseteq \mathbb{R}^M} c_0(\mathbf{d}) := 0.5\frac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]}{12,589} + 0.5\frac{\sqrt{\text{var}_{\mathbf{d}}[y_0(\mathbf{X})]}}{334},$$

$$\text{subject to } c_l(\mathbf{d}) := 3\sqrt{\text{var}_{\mathbf{d}}[y_l(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_l(\mathbf{X})] \leqslant 0, \quad l = 1, \ldots, 11,$$

$$0.1 \text{ in}^2 \leqslant d_k \leqslant 30 \text{ in}^2, \ k = 1, \ldots, 10,$$

where

$$y_0(\mathbf{X}) = 0.1 \sum_{i=1}^{10} l_i X_i$$

is the random mass of the truss with l_i, $i = 1, \ldots, 10$, representing bar lengths and

$$y_l(\mathbf{X}) = \begin{cases} 25,000 - \sigma_l(\mathbf{X}), & \text{if } l = 1, \ldots, 9, \\ 75,000 - \sigma_l(\mathbf{X}), & \text{if } l = 10, \\ 5 - v_2(\mathbf{X}), & \text{if } l = 11, \end{cases}$$

are 11 stochastic performance functions. The initial design is $\mathbf{d}_0 = (15, \ldots, 15)^\intercal$ in^2. The approximate optimal solution is denoted by $\tilde{\mathbf{d}}^* = (\tilde{d}_1^*, \ldots, \tilde{d}_{10}^*)^\intercal$.

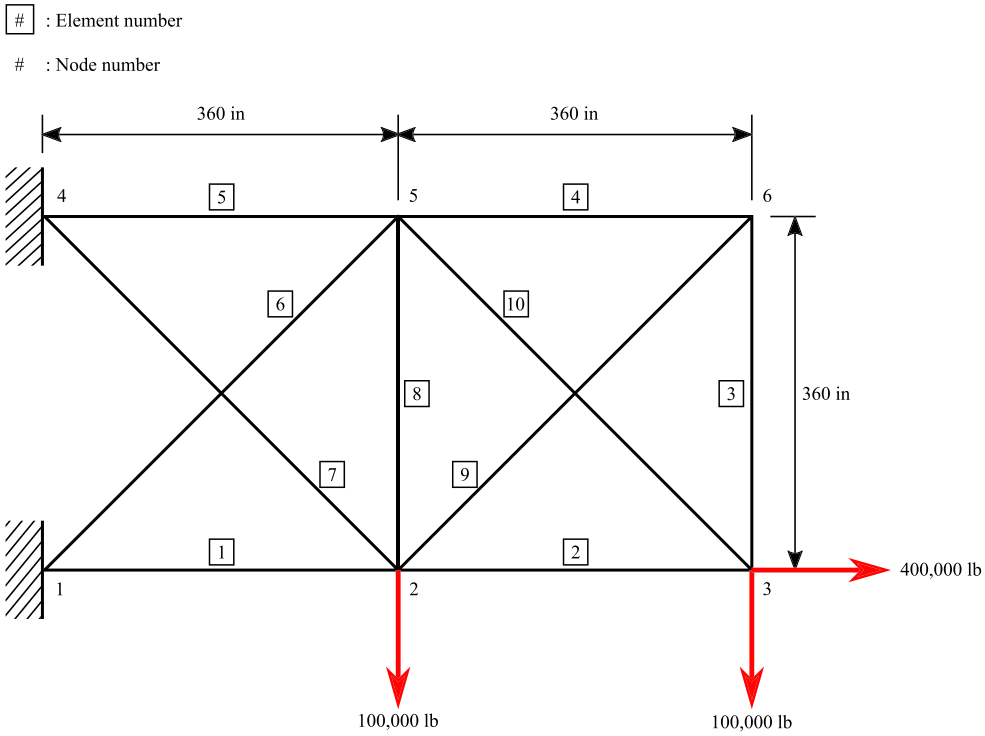

Fig. 6 A ten-bar truss structure

Table 8 summarizes the assorted results of design optimization by the direct GPCE and multi-point single-step GPCE methods. In the second and third columns of Table 8, the direct GPCE solutions obtained from the second-order ($m = 2$) and third-order ($m = 3$) approximations with expansion coefficients estimated by the SLS regression are almost the same. Therefore, the SLS-produced second-order direct GPCE solution is adequate and credible. However, for the third-order GPCE approximation, the requisite number of FEA escalates significantly, making the SLS-produced solution impractical. This is because the SLS regression, depending on the number of GPCE coefficients, requires a large data set; for instance, when $m = 3$, the data size is $L = 858$, incurring over one hundred thousand FEA to complete all design iterations. To mitigate such high computational demand, the extended version of the partitioned D-MORPH regression was utilized to estimate the GPCE coefficients for two different data sizes ($L = 100, 200$), and the same RDO problem was re-solved. The solutions are tabulated in the fourth and fifth columns for $m = 2$ and $m = 3$, respectively. The direct GPCE method, obtained using the D-MORPH regression, yields nearly matching optimal solutions of SLS regression, especially when the data size is larger. However, the associated computational expense is now significantly lower than before. For example, when $m = 3$, a data size of $L = 200$ is satisfactory for the D-MORPH regression, shrinking the number of FEA by more than a factor of four.

While the direct GPCE method, in conjunction with the D-MORPH regression, has greatly reduced the computational cost, the absolute numbers of FEA are still very high. To alleviate the cost even further, the first-order multi-point single-step GPCE method with SLS-estimated coefficients was tested, the results of which are shown in the last column of Table 8. Indeed, a low-order multi-point single-step GPCE method also produces optimal solutions very close to those obtained by higher-order direct GPCE methods, but sustaining less than 1,000 FEA. Therefore, the multi-point single-step GPCE method provides a computationally efficient solution to RDO problems, at least in this example.

6.5 Example 5: Shape optimization of a steering knuckle

The last example establishes the efficacy of the multi-point single-step GPCE method in designing an industrial-scale mechanical component, known as steering knuckle. As illustrated in Fig. 7, the steering knuckle of an automotive suspension system is located in the spindle of a tire-wheel assembly, attaching to the suspension and steering components. In Fig. 7a, the knuckle connects to a strut at the top and lower control arms through the ball joint at the bottom to stabilize the driving motion. A steering arm of the knuckle is connected to the steering mechanism to turn the tire-wheel assembly. The knuckles are commonly designed to maintain

Table 8 Optimization results of a ten-bar truss (Example 4)

	Direct GPCE				Multi-point
	SLS[a]		Partitioned D-MORPH[b]		single-step GPCE[c]
Results	$m = 2$ $L = 198$[d]	$m = 3$ $L = 858$[d]	$m = 2$ $L = 100$[d]	$m = 3$ $L = 200$[d]	$m = 1$ $L = 33$[d]
$\tilde{d}_1^*$, in^2	4.6435	4.6434	4.6435	4.6430	4.6295
$\tilde{d}_2^*$, in^2	13.9252	13.9280	14.2013	13.9236	13.9185
$\tilde{d}_3^*$, in^2	0.1000	0.1000	0.3935	0.1000	0.1007
$\tilde{d}_4^*$, in^2	0.1000	0.1000	0.3931	0.1000	0.1000
$\tilde{d}_5^*$, in^2	4.6427	4.6433	4.6425	4.6424	4.6419
$\tilde{d}_6^*$, in^2	0.1000	0.1000	0.1000	0.1000	0.1000
$\tilde{d}_7^*$, in^2	12.9792	12.9824	12.9771	12.9782	12.9777
$\tilde{d}_8^*$, in^2	4.4737	4.4740	4.1823	4.4739	4.4820
$\tilde{d}_9^*$, in^2	0.1385	0.1385	0.5455	0.1385	0.1387
$\tilde{d}_{10}^*$, in^2	3.2952	3.2955	3.0717	3.2947	3.3161
$c_0(\tilde{\mathbf{d}}^*)$	1.5909×10^{-1}	1.5912×10^{-1}	1.6110×10^{-1}	1.5908×10^{-1}	1.5913×10^{-1}
$c_1(\tilde{\mathbf{d}}^*)$	-1.1999×10^{-1}	-1.2006×10^{-1}	-1.2032×10^{-1}	-1.2003×10^{-1}	-1.2244×10^{-1}
$c_2(\tilde{\mathbf{d}}^*)$	1.8118×10^{-8}	-1.1053×10^{-9}	-1.1850×10^{-6}	-6.5266×10^{-7}	-1.0283×10^{-5}
$c_3(\tilde{\mathbf{d}}^*)$	-7.9680×10^{-8}	6.6478×10^{-9}	-3.9597×10^{-8}	-1.3862×10^{-7}	-4.0113×10^{-6}
$c_4(\tilde{\mathbf{d}}^*)$	2.0464×10^{-7}	1.3008×10^{-8}	9.7856×10^{-6}	3.7328×10^{-6}	-3.2441×10^{-6}
$c_5(\tilde{\mathbf{d}}^*)$	-1.6853×10^{-7}	2.0698×10^{-8}	2.0882×10^{-6}	-1.6241×10^{-6}	-4.2739×10^{-6}
$c_6(\tilde{\mathbf{d}}^*)$	-1.2132×10^{-7}	1.3089×10^{-8}	-2.1549×10^{-7}	-1.4687×10^{-6}	-6.4120×10^{-6}
$c_7(\tilde{\mathbf{d}}^*)$	-4.1197×10^{-1}	-4.1203×10^{-1}	-4.1018×10^{-1}	-4.1197×10^{-1}	-4.1351×10^{-1}
$c_8(\tilde{\mathbf{d}}^*)$	1.3304×10^{-8}	1.4487×10^{-8}	-8.4815×10^{-7}	5.1011×10^{-7}	-6.1153×10^{-6}
$c_9(\tilde{\mathbf{d}}^*)$	4.8502×10^{-7}	-1.7907×10^{-8}	-3.1965×10^{-6}	6.7594×10^{-6}	5.6970×10^{-7}
$c_{10}(\tilde{\mathbf{d}}^*)$	-1.0894×10^{-8}	-7.9334×10^{-10}	3.0665×10^{-6}	3.7827×10^{-7}	-5.1394×10^{-6}
$c_{11}(\tilde{\mathbf{d}}^*)$	-3.5549×10^{-1}	-3.5548×10^{-1}	-3.5502×10^{-1}	-3.5556×10^{-1}	-3.5571×10^{-1}
$\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$, lb	1844.5687	1844.8778	1874.3617	1844.4106	1845.1160
$\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$, lb	57.4072	57.4179	57.9574	57.4022	57.4148
No. of iterations	55	53	57	55	428
No. of FEA	25,938	101,244	13,300	23,600	924

[a]The standard least squares (SLS), used for estimating GPCE coefficients

[b]The extended version of the partitioned D-MORPH, used for estimating GPCE coefficients

[c]The multi-point single-step GPCE is in conjunction with the SLS

[d]The sample size (L) of input-output data set

adequate fatigue durability under various driving conditions, such as forward/backward accelerations, cornering, and braking, and must sustain satisfactory performances on uneven road conditions. However, any uncertainties in material properties and/or manufacturing variables, if they exist, result in the randomness of fatigue life. A traditional deterministic design optimization incorporating large safety factors may lead to increased weight of a vehicle, causing a loss of fuel efficiency. Therefore, incorporating uncertainty in fatigue life, as done in RDO, under various driving environments is essential to creating lightweight knuckle design.

Ten input random variables were identified by modeling the randomness in manufacturing tolerances of the knuckle geometry and in fatigue parameters of the knuckle material. Figure 7b depicts a computer-aided design (CAD) model of a steering knuckle with six random manufacturing variables X_1 through X_6, which are marked in the side and front views. The variables X_1 and X_2 represent the widths of two arms, while the variables X_3 through X_6 describe the heights and widths of two voids introduced in two arms to bring down the mass of the knuckle as much as possible. The knuckle material is made of cast steel and has the following random fatigue parameters: fatigue

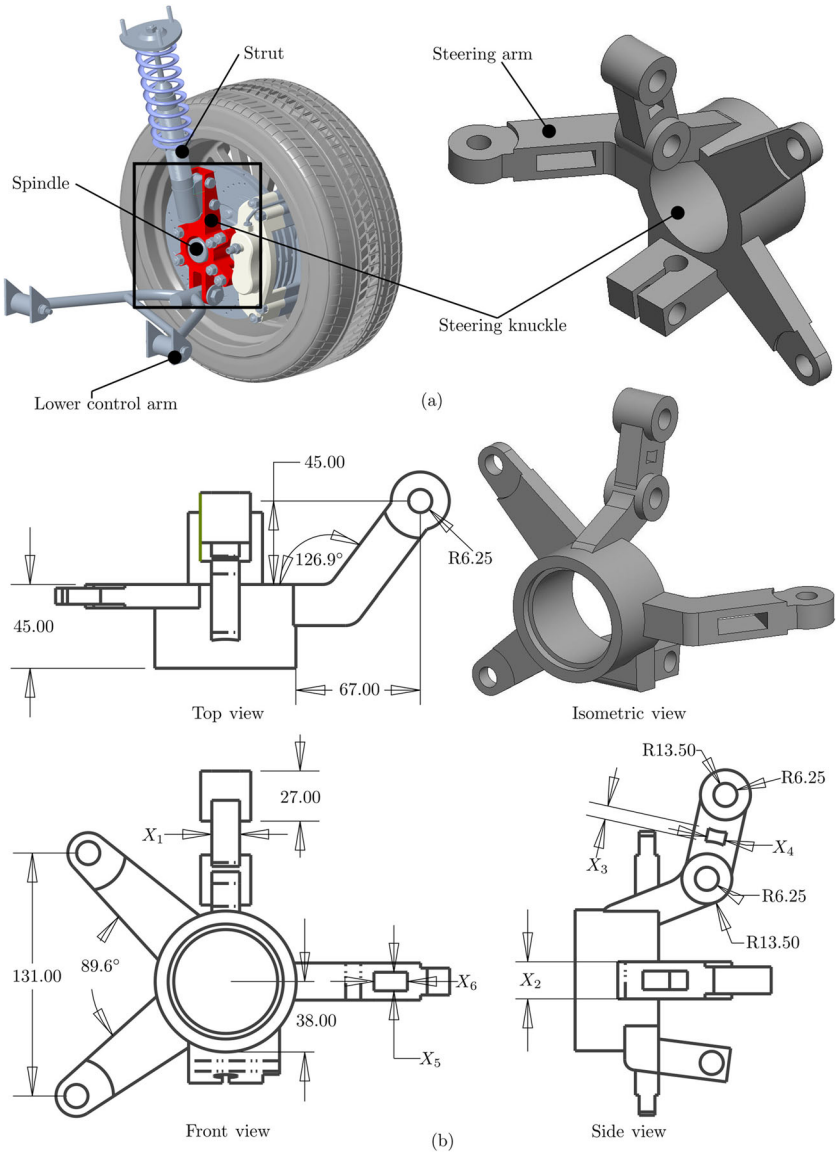

Fig. 7 A steering knuckle (Example 5): **a** a photo of the knuckle/suspension assembly; **b** a CAD model of the steering knuckle (unit: mm)

strength coefficient X_7, fatigue ductility coefficient X_8, fatigue strength exponent X_9, fatigue ductility exponent X_{10}. The statistical dependence among/between random variables is as follows: (1) $(X_1, \ldots, X_6)^\intercal$ is a hexavariate lognormal vector; (2) $(X_7, X_8)^\intercal$ is a bivariate lognormal vector; and (3) $(X_9, X_{10})^\intercal$ is a bivariate Gaussian vector. The probabilistic characteristics of all ten random variables are fully described in Table 9.

The deterministic material properties are as follows: mass density $\rho = 7800\ \text{kg/m}^3$, Young's modulus $E = 203$ GPa, and Poisson's ratio $\nu = 0.3$. The design variables are the means of the first six manufacturing variables, that is, $d_k = \mathbb{E}_{\mathbf{d}}[X_k], k = 1, \ldots, 6$.

The stochastic performance of the knuckle was determined by fatigue durability analysis under two distinct loading conditions: a road impact condition comprising two separate applied loads F_1 and F_2, as shown in Fig. 8a; and a cornering condition entailing a single applied load F_3, as shown in Fig. 8b. In both loading conditions, the knuckle experiences constant-amplitude cyclic loads with the maximum and minimum values of these three loads, as follows: $1.16 \leqslant F_1 \leqslant 2.24$ kN (Y-axis); $1.16 \leqslant F_2 \leqslant 2.24$ kN (Y-axis); $-1.28 \leqslant F_3 \leqslant 1.28$ kN (Z-axis). The essential boundary condition involves fixing the center of the spindle of the knuckle in all three directions. The fatigue durability analysis involved (1) calculating maximum principal strain and mean stress at a critical point; and (2) calculating the fatigue crack-initiation life at that point from the well-known Coffin-Manson-Morrow equation (Stephens et al. 2000). The critical point is the location where the von

Table 9 Statistical properties of random variables of the steering knuckle (Example 5)

(1) Means and standard deviations of random variables

Random variables	Mean	Standard deviation	Probability distribution
Dimension (X_1), mm	d_1	$0.05d_1$	Lognormal
Dimension (X_2), mm	d_2	$0.05d_2$	Lognormal
Dimension (X_3), mm	d_3	$0.05d_3$	Lognormal
Dimension (X_4), mm	d_4	$0.05d_4$	Lognormal
Dimension (X_5), mm	d_5	$0.05d_5$	Lognormal
Dimension (X_6), mm	d_6	$0.05d_6$	Lognormal
σ_f' (X_7), MPa[a]	1332	66.6	Lognormal
b (X_8)[b]	-0.1085	5.425×10^{-4}	Lognormal
ϵ_f' (X_9)[c]	0.375	1.875×10^{-2}	Gaussian
c (X_{10})[d]	-0.6354	3.177×10^{-2}	Gaussian

(2) Covariance matrix of input random variables

$X_i,\ i = 1, \ldots, 6$

$$\begin{bmatrix} 0.0025d_1^2 & 0.0012d_1d_2 & 0.0012d_1d_3 & 0.0012d_1d_4 & 0.0012d_1d_5 & 0.0012d_1d_6 \\ 0.0012d_1d_2 & 0.0025d_2^2 & 0.0012d_2d_3 & 0.0012d_2d_4 & 0.0012d_2d_5 & 0.0012d_2d_6 \\ 0.0012d_1d_3 & 0.0012d_2d_3 & 0.0025d_3^2 & 0.0012d_3d_4 & 0.0012d_3d_5 & 0.0012d_3d_6 \\ 0.0012d_1d_4 & 0.0012d_2d_4 & 0.0012d_3d_4 & 0.0025d_4^2 & 0.0012d_4d_5 & 0.0012d_4d_6 \\ 0.0012d_1d_5 & 0.0012d_2d_5 & 0.0012d_3d_5 & 0.0012d_4d_5 & 0.0025d_5^2 & 0.0012d_5d_6 \\ 0.0012d_1d_6 & 0.0012d_2d_6 & 0.0012d_3d_6 & 0.0012d_4d_6 & 0.0012d_5d_6 & 0.0025d_6^2 \end{bmatrix}$$

X_7, X_8

$$\begin{bmatrix} 4.4222 \times 10^3 & 0.2520 \\ 0.2520 & 2.9431 \times 10^{-5} \end{bmatrix}$$

X_9, X_{10}

$$\begin{bmatrix} 3.5156 \times 10^{-4} & 5.3612 \times 10^{-4} \\ 5.3612 \times 10^{-4} & 0.0010 \end{bmatrix}$$

[a] σ_f' represents the fatigue strength coefficient

[b] b represents the fatigue strength exponent

[c] ϵ_f' represents the fatigue ductility coefficient

[d] c represents the fatigue ductility exponent

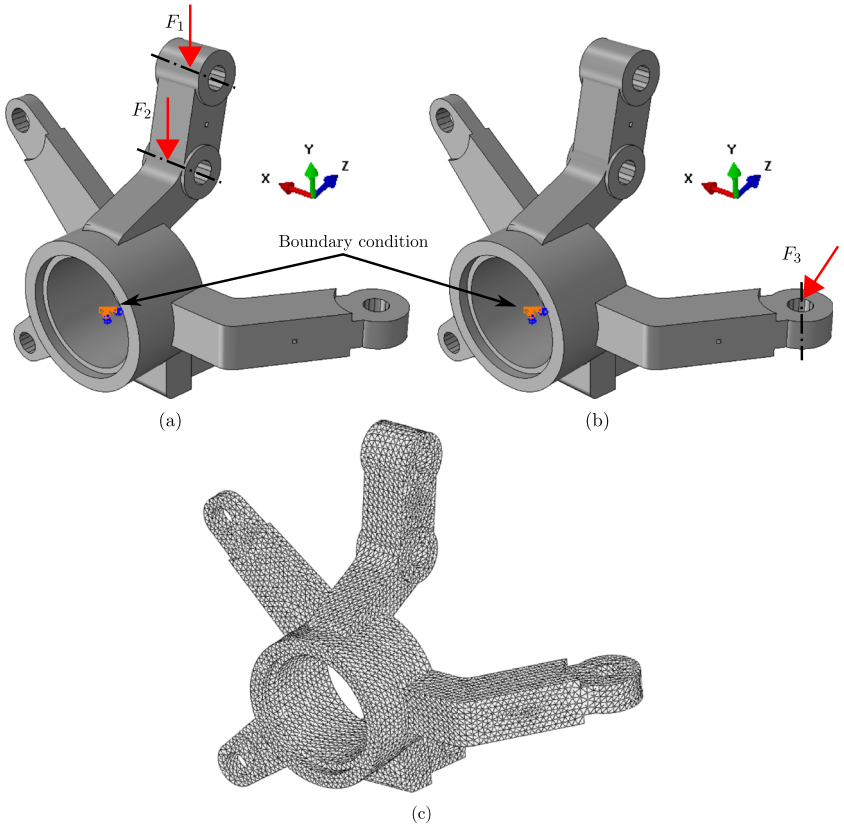

Fig. 8 An FEA of the steering knuckle (Example 5): **a** road impact load and boundary condition; **b** cornering load and boundary condition; **c** a tetrahedral mesh comprising 75,271 elements

Mises stress is the largest, identified from an FEA where all input random variables are assigned their mean values. The objective is to minimize the mass of the knuckle by changing the shape of the geometry such that the fatigue crack-initiation lives $N_1(\mathbf{X})$ and $N_2(\mathbf{X})$ under the two aforementioned loading conditions at the critical point exceed a design threshold of a million loading cycles with 98.21% probability. Mathematically, the RDO for this problem is defined to

$$\begin{aligned}
\min_{\mathbf{d}\in\mathcal{D}} \; & c_0(\mathbf{d}) := \tfrac{\mathbb{E}_{\mathbf{d}}[y_0(\mathbf{X})]}{2.2964}, \\
\text{subject to } & c_1(\mathbf{d}) := 2.1\sqrt{\text{var}_{\mathbf{d}}[y_1(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_1(\mathbf{X})] \leqslant 0, \\
& c_l(\mathbf{d}) := 2.1\sqrt{\text{var}_{\mathbf{d}}[y_2(\mathbf{X})]} - \mathbb{E}_{\mathbf{d}}[y_2(\mathbf{X})] \leqslant 0, \\
& 10 \text{ mm} \leqslant d_1 \leqslant 25 \text{ mm}, \\
& 15 \text{ mm} \leqslant d_2 \leqslant 25 \text{ mm}, \\
& 2 \text{ mm} \leqslant d_3 \leqslant 16 \text{ mm}, \\
& 2 \text{ mm} \leqslant d_4 \leqslant 20 \text{ mm}, \\
& 2 \text{ mm} \leqslant d_5 \leqslant 10 \text{ mm}, \\
& 2 \text{ mm} \leqslant d_6 \leqslant 30 \text{ mm},
\end{aligned}$$

where

$$y_0(\mathbf{X}) = \rho \int_{\mathcal{D}'} d\mathcal{D}'$$

is the random mass of the knuckle, and

$$y_1(\mathbf{X}) = \frac{N_1(\mathbf{X})}{1 \times 10^6} - 1,$$

and

$$y_2(\mathbf{X}) = \frac{N_2(\mathbf{X})}{1 \times 10^6} - 1,$$

are two stochastic performance functions based on normalized fatigue crack-initiation lives. The initial design $\mathbf{d}_0 = (23, 23, 2, 2, 2, 2)^\intercal$ mm. Figure 8c presents an FEA mesh for the initial knuckle design, which comprises 75,271 tetrahedral elements. The approximate optimal solution is denoted by $\tilde{\mathbf{d}}^* = (\tilde{d}_1^*, \ldots, \tilde{d}_6^*)^\intercal$.

The multi-point single-step GPCE method was applied to solve this steering knuckle design problem employing first-order ($m = 1$) GPCE approximations for second-moment analyses of $y_1(\mathbf{X})$ and $y_2(\mathbf{X})$. As a result, the optimal design is at $\tilde{\mathbf{d}}^* = (10.0, 15.0, 16.0, 20.0, 10.0, 30.0)^\intercal$ mm. At the optima, the design variables $\tilde{d}_1^*$ and $\tilde{d}_2^*$ reached the lower limits and $\tilde{d}_3^*$, $\tilde{d}_4^*$, $\tilde{d}_5^*$ and $\tilde{d}_6^*$ the upper limits, satisfying inactive constraints ($c_1 \simeq -0.58$, $c_2 \simeq -8.68$).

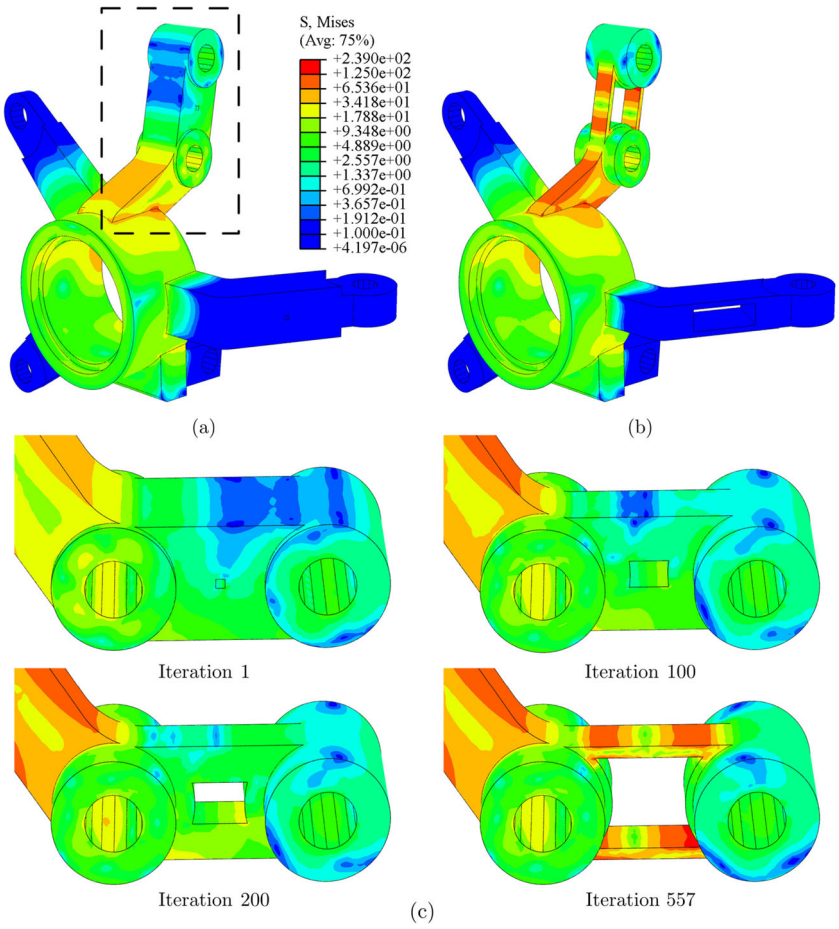

Fig. 9 Contours of von Mises stress under the mean road impact load condition at mean shapes of the steering knuckle optimized by the multi-point single-step GPCE ($m = 1$) method: **a** initial design; **b** final optimal design at iteration 557; **c** design history at iterations 1, 100, 200, and 557 for the component inside the dot-lined box in Fig. 9a

The mean optimal mass of the knuckle is 1.95 kg, which represents about a 15% reduction from the initial mass of 2.30 kg. Correspondingly, the standard deviation of the mass declines from 2.60×10^{-2} to 1.00×10^{-2} kg—a drop of about 61%. To complete the design process, the requisite number of FEA is 2310.

Figures 9 and 10 show the contour plots of the von Mises stress at the initial (Figs. 9a and 10a) and optimal (Figs. 9b and 10b) designs under road impact and cornering conditions, respectively. They both illustrate deterministic stresses obtained at mean shapes of these designs when all random variables take on their respective mean values. These two figures also include close-up views of the stress contours at several design iterations (Figs. 9c and 10c). At the optimal solution, there is, indeed, a considerable reduction of the overall volume of the knuckle, but without exceeding stresses responsible for violating the required fatigue lives.

Figure 11 presents the iteration histories of the objective function c_0 and six design variables during the RDO process attained by the first-order multi-point single-step GPCE method. In Fig. 11a, the monotonic convergence is achieved such that the value of objective function c_0 is reduced from 1.00 at the initial design to nearly 0.85 at the optimal design, about a 15% change. According to Fig. 11b, all design variables have undergone significant changes from their initial values, influencing substantial modifications of the knuckle arms and their hollow shapes and sizes. This concluding example indicates that the RDO methods developed—in particular, the first-order multi-point single-step GPCE method—are capable of solving industrial-scale engineering design problems using only a few thousand FEA.

Finally, Table 10 outlines the percentage changes in the mean and standard deviation of y_0 from initial to optimal designs in all four examples. The second-moment statistics at optimal designs are averages of all GPCE solutions described earlier. The largest reduction of the mean is 85.3%, whereas the standard deviation diminishes by at most 96.3%. Clearly, robustness in design optimization has played an important role in reducing the statistical moments of objective functions.

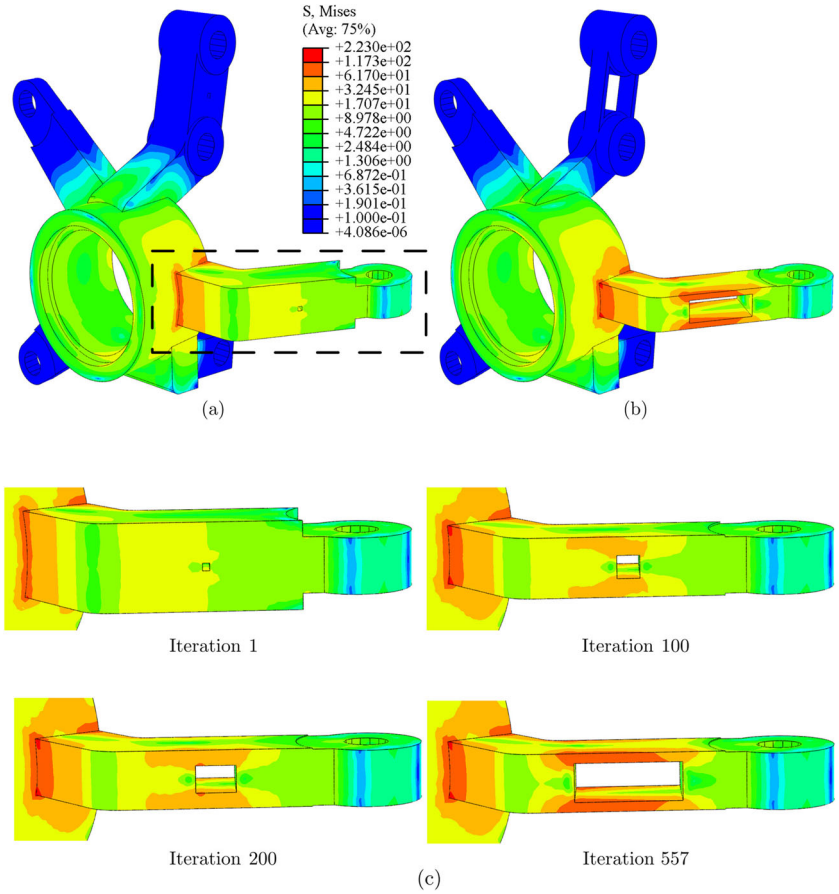

Fig. 10 Contours of von Mises stress under the mean cornering load condition at mean shapes of the steering knuckle optimized by the multi-point single-step GPCE ($m = 1$) method: **a** initial design; **b** final optimal design at iteration 557; **c** design history at iterations 1, 100, 200, and 557 for the component inside of the dot-lined box in Fig. 10a

7 Discussion

It is important to differentiate this work from related past studies on UQ and design optimization (Ren and Rahman paper, Lee and Rahman paper) conducted in the authors' group. First, Lee and Rahman (2020) solved only forward UQ problems for dependent random variables without considering stochastic sensitivity analysis and design optimization. Second, Ren and Rahman (2013) tackled RDO problems, including formulating sensitivity analysis and design optimization algorithms, but their work was limited strictly to independent random variables. In contrast, the current study is new in the sense that a more general class of RDO problems, grappling with dependent input

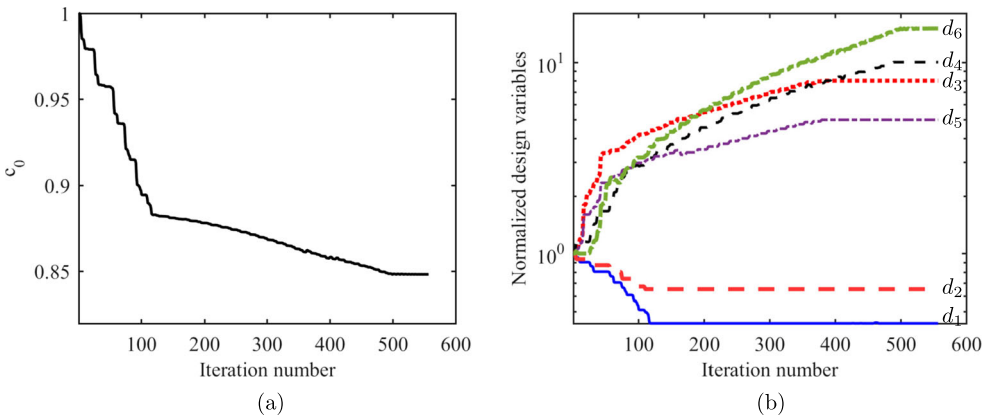

Fig. 11 RDO iteration histories for the knuckle: **a** objective function at $m = 1$; **b** normalized design variables d_1, d_2, d_3, d_4, d_5, and d_6 at $m = 1$

Table 10 Reductions in the mean and standard deviation of y_0 from initial to optimal designs

Example	$\frac{\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]-\mathbb{E}_{\mathbf{d}_0}[y_0(\mathbf{X})]}{\mathbb{E}_{\mathbf{d}_0}[y_0(\mathbf{X})]}$	$\frac{\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}-\sqrt{\text{var}_{\mathbf{d}_0}[y_0(\mathbf{X})]}}{\sqrt{\text{var}_{\mathbf{d}_0}[y_0(\mathbf{X})]}}$
1-1[a]	− 68.7%	− 93.2%
1-2[b]	− 77.2%	− 96.3%
3[c]	− 60.0%	− 60.0%
4[d]	− 85.3%	− 82.8%
5	− 15.2%	− 61.4%

[a]The value of $\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$ and $\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$ is the average of all corresponding GPCE results in Table 4-(1)

[b]The value of $\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$ and $\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$ is the average of all corresponding GPCE results in Table 4-(2)

[c]The value of $\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$ and $\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$ is the average of all corresponding GPCE results in Table 7

[d]The value of $\mathbb{E}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]$ and $\sqrt{\text{var}_{\tilde{\mathbf{d}}*}[y_0(\mathbf{X})]}$ is the average of all corresponding GPCE results in Table 8

random variables head-on, is addressed. There are at least two key elements of novelty in this paper: (1) development of new design sensitivity equations for input random variables following an arbitrary, dependent probability measure; (2) development of three new design optimization algorithms tailored to solve diverse RDO problems subject to dependent input random variables.

As multiple design methods have been proffered, providing a brief overview of their efficiency and relevance is in order. Table 11 presents such an overview, including a few comments supported by the numerical examples of the preceding section. While all three methods have their own applications, the multi-point single-step GPCE is especially suited to solving practical RDO problems.

8 Conclusion

Three new computational methods were created for robust design optimization of complex engineering systems subject to input random variables with arbitrary, dependent probability distributions. The methods are built on a GPCE for determining the second-moment statistics of a general output function of dependent input random variables, an innovative coupling between GPCE and score functions for calculating the second-moment sensitivities with respect to the design variables, and a standard gradient-based optimization algorithm, establishing direct GPCE, single-step GPCE, and multi-point single-step GPCE design processes. When blended with score functions, all three GPCE methods subsume analytical formulae for calculating the design sensitivities. More importantly, the statistical moments and their respective design sensitivities are both determined concurrently from a single stochastic analysis or simulation.

Among the three methods, the direct GPCE method is most straightforward, but it mandates re-computations of the GPCE coefficients at each design iteration. Consequently, it can be expensive, depending on the cost of evaluating the objective and constraint functions and the required number of design iterations. The single-step GPCE method avoids the necessity of recomputing the coefficients by reprocessing the old coefficients, potentially bringing down the computational cost substantially. However, it relies heavily on the GPCE's approximation quality and the accuracy of the estimated coefficients during all design iterations. However, both of these methods are globally formulated on the entire design space, raising a concern about the GPCE's approximation quality when the design space is too large and/or the objective and constraint functions

Table 11 Efficiency and relevance of three GPCE methods

Method	Efficiency	Relevance	Comments
Direct GPCE	Low	Inexpensive-to-evaluate, lowly or highly nonlinear functions	Costly due to recalculation of the expansion coefficients from new experimental design. Not practical for complex system designs.
Single-step GPCE	High	Lowly nonlinear functions with small or large design spaces	Highly efficient due to recycling of old expansion coefficients, but may produce premature or inaccurate solutions for complex system designs.
Multi-point single-step GPCE	Medium	Highly nonlinear functions with large design spaces	Suitable for solving complex, practical design problems using low-order GPCE approximations.

are overly nonlinear. The multi-point single-step GPCE method comes to the rescue in this problem by adopting a local enforcement of GPCE approximation, where the original RDO problem is exchanged with a series of local RDO problems defined on a subregion of the entire design space. This is precisely why the latter method is capable of solving practical engineering problems using only low-order approximations, as demonstrated by shape design optimization of an industrial-scale steering knuckle.

Appendix 1. Direct approach of partitioned D-MORPH regression

The partitioned D-MORPH regression involves primary and secondary groups of the GPCE basis functions. According to Lee and Rahman (2020), the number of primary basis functions L_p is defined as

$$L_p := \min\left(\left\lceil \frac{L_{N,m}}{K_1} \right\rceil, \left\lceil \frac{L}{K_2} \right\rceil\right),$$

where $K_1, K_2 \geqslant 1$ are two real-valued factors such that $L_p < L_{N,m}$ and $L_p < L$. Here, $\lceil\cdot\rceil$ is a symbol for the ceiling function. The proper values of K_1, K_2 are problem dependent, meaning that they are determined from trial and error. Nonetheless, there are two cases for how L_p is determined: (1) $L_p = \lceil L_{N,m}/K_1 \rceil$ and (2) $L_p = \lceil L/K_2 \rceil$.

In the direct approach, partition the $L \times L_{N,m}$ data matrix $\mathbf{A} = [\mathbf{A}_p, \mathbf{A}_s]$ into two parts, where $\mathbf{A}_p$ is the $L \times L_p$ matrix representing the first L_p columns of $\mathbf{A}$ and $\mathbf{A}_s$ is the $L \times (L_{N,m} - L_p)$ matrix describing the last $L_{N,m} - L_p$ columns of $\mathbf{A}$. Then, the approximation solution of expansion coefficients is obtained from

$$\bar{\mathbf{A}}\hat{\mathbf{c}} = \bar{\mathbf{b}}, \tag{A1.1}$$

where $\bar{\mathbf{A}} := \mathbf{A}_p^\intercal[\mathbf{A}_p, \mathbf{A}_s]$ is the reduced $L_p \times L_{N,m}$ data matrix and $\bar{\mathbf{b}} := \mathbf{A}_p^\intercal \mathbf{b}$ is the reduced L_p-dimensional vector. As (A1.1) is an underdetermined system, the direct approach produces partitioned D-MORPH solution $\check{\mathbf{c}} = (\check{C}_1(\mathbf{d}), \ldots, \check{C}_{L_{N,m}}(\mathbf{d}))^\intercal \in \mathbb{R}^{L_{N,m}}$ of expansion coefficients as

$$\check{\mathbf{c}} = \bar{\mathbf{F}}_{L_{N,m}-r}(\bar{\mathbf{E}}_{L_{N,m}-r}^\intercal \bar{\mathbf{F}}_{L_{N,m}-r})^{-1}\bar{\mathbf{E}}_{L_{N,m}-r}^\intercal \bar{\mathbf{A}}^+\bar{\mathbf{b}}, \tag{A1.2}$$

where $\bar{\mathbf{A}}^+$ is the Moore-Penrose inverse of $\bar{\mathbf{A}}$. Here, $\bar{\mathbf{E}}_{L_{N,m}-r}$ and $\bar{\mathbf{F}}_{L_{N,m}-r}$ are constructed from the last $L_{N,m}-r$ columns of $\bar{\mathbf{E}}$ and $\bar{\mathbf{F}}$, generated from the singular value decomposition of

$$\bar{\Phi}\bar{\mathbf{D}} = \bar{\mathbf{E}}\begin{bmatrix} \bar{\mathbf{T}}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\bar{\mathbf{F}}^\intercal,$$

with

$$\bar{\Phi} := (\mathbf{I}_{L_{N,m}} - \bar{\mathbf{A}}^+\bar{\mathbf{A}}) \in \mathbb{R}^{L_{N,m}\times L_{N,m}}, \tag{A1.3}$$
$$\bar{\mathbf{D}} = \mathrm{diag}[\underbrace{0, \cdots, 0}_{L_p}, \underbrace{1, \cdots, 1}_{L_{N,m}-L_p}].$$

Appendix 2. Expectation of score function

Consider the kth first-order score function $s_k(\mathbf{Z}; \mathbf{g})$ in (13). The expectation of $s_k(\mathbf{Z}; \mathbf{g})$ with respect to the probability measure $f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})d\mathbf{z}$ is reformulated by replacing the score function with the ratio given by the derivative of the natural logarithm as follows:

$$\begin{aligned}\mathbb{E}_{\mathbf{g}}[s_k(\mathbf{Z}; \mathbf{g})] &= \int_{\bar{\mathbb{A}}^N} \frac{\partial \ln f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})}{\partial g_k} f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})d\mathbf{z} \\ &= \int_{\bar{\mathbb{A}}^N} \frac{\partial f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})}{\partial g_k} \frac{1}{f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})} f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})d\mathbf{z} \\ &= \int_{\bar{\mathbb{A}}^N} \frac{\partial f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})}{\partial g_k} d\mathbf{z}. \end{aligned} \tag{A2.1}$$

Then, under the regularity conditions in Section 4, interchanging the order of differentiation and integration in the final equality of (A2.1) yields

$$\int_{\bar{\mathbb{A}}^N} \frac{\partial f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})}{\partial g_k} d\mathbf{z} = \frac{\partial}{\partial g_k}\int_{\bar{\mathbb{A}}^N} f_{\mathbf{Z}}(\mathbf{z}; \mathbf{g})d\mathbf{z} = \frac{\partial}{\partial g_k} 1 = 0.$$

Therefore, the mean value of the score function is always zero for any probability measure of $\mathbf{Z}$.

Funding This work received financial support from the U.S. National Science Foundation under Grant Nos. CMMI-1462385 and CMMI-1933114.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Replication of results The MATLAB code files of the Examples 1–4 are provided in the following Github repository: https://github.com/icdsigma-lee/Robust-Design-Optimization-under-Dependent-Random-Variables-by-a-Generalized-Polynomial-Chaos-Expans.git. Unfortunately, the FEM or other source codes of Example 5 cannot be shared due to the external code license. Readers further interested in the codes are encouraged to contact the authors by e-mail.

References

ABAQUS (2019) version 2019. Dassault Systèmes Simulia Corp

Bashiri M, Moslemi A, Akhavan Niaki ST (2020) Robust multi-response surface optimization: a posterior preference approach. Int Trans Oper Res 27(3):1751–1770

Bhushan M, Narasimhan S, Rengaswamy R (2008) Robust sensor network design for fault diagnosis. Comput Chem Eng 32(4-5):1067–1084

Browder A (1996) Mathematical analysis: an introduction. Undergraduate texts in mathematics. Springer, Berlin
Busbridge I (1948) Some integrals involving hermite polynomials. J Lond Math Soc 23:135–141
Chatterjee T, Chakraborty S, Chowdhury R (2019) A critical review of surrogate assisted robust design optimization. Archiv Comput Methods Eng 26(1):245–274
Chen W, Allen J, Tsui K, Mistree F (1996) Procedure for robust design: minimizing variations caused by noise factors and control factors. J Mechan Design Trans ASME 118(4):478–485
Chen W, Wiecek MM, Zhang J (1999) Quality utility—a compromise programming approach to robust design. J Mech Des 121(2):179–187
Cramer AM, Sudhoff SD, Zivi EL (2008) Evolutionary algorithms for minimax problems in robust design. IEEE Trans Evol Comput 13(2):444–453
CREO (2016) version 4.0. PTC
Du X, Chen W (2000) Towards a better understanding of modeling feasibility robustness in engineering design. J Mechan Design 122(4):385–394
Elishakoff I, Haftka R, Fang J (1994) Structural design under bounded uncertainty—optimization with anti-optimization. Comput Struct 53(6):1401–1405
Huang B, Du X (2007) Analytical robustness assessment for robust design. Struct Multidiscip Optim 34(2):123–137
Jin R, Du X, Chen W (2003) The use of metamodeling techniques for optimization under uncertainty. Struct Multidiscip Optim 25(2):99–116
Lee D, Rahman S (2020) Practical uncertainty quantification analysis involving statistically dependent random variables. Appl Math Model 84:324–356
Lee I, Choi KK, Noh Y, Zhao L, Gorsich D (2011) Sampling-based stochastic sensitivity analysis using score functions for rbdo problems with correlated random variables. J Mech Des 133(2)
Lee S, Chen W, Kwak B (2009) Robust design with arbitrary distributions using gauss-type quadrature formula. Struct Multidiscip Optim 39(3):227–243
Li G, Rabitz H (2010) D-MORPH regression: application to modeling with unknown parameters more than observation data. J Math Chem 48:1010–1035
Marler RT, Arora JS (2010) The weighted sum method for multi-objective optimization: new insights. Struct Multidiscip Opt 41(6):853–862
MATLAB (2019) Version 9.7.0 (R2019b). The MathWorks Inc., Natick, Massachusetts
Miettinen K (2012) Nonlinear multiobjective optimization, vol 12. Springer Science & Business Media, Berlin
Mourelatos Z, Liang J (2006) A methodology for trading-off performance and robustness under uncertainty. J Mechan Design 128(4):856–863
Nataf A (1962) Détermination des distributions de probabilités dont les marges sont données. C R Acad Sci, Paris 255:42–43
Nha VT, Shin S, Jeong SH (2013) Lexicographical dynamic goal programming approach to a robust design optimization within the pharmaceutical environment. Eur J Oper Res 229(2):505–517
Noh Y, Choi K, Du L (2009) Reliability-based design optimization of problems with correlated input variables using a gaussian copula. Struct Multidiscip Opt 38(1):1–16
Ono S, Yoshitake Y, Nakayama S (2009) Robust optimization using multi-objective particle swarm optimization. Artificial Life Robot 14(2):174
Park G, Lee T, Kwon H, Hwang K (2006) Robust design: an overview. AIAA J 44(1):181–191
Rahman S (2009) Extended polynomial dimensional decomposition for arbitrary probability distributions. J Eng Mechan-ASCE 135(12):1439–1451
Rahman S (2009) Stochastic sensitivity analysis by dimensional decomposition and score functions. Probab Eng Mechan 24(3):278–287
Rahman S (2018) A polynomial chaos expansion in dependent random variables. J Math Anal Appl 464(1):749–775
Rahman S, Ren X (2014) Novel computational methods for high-dimensional stochastic sensitivity analysis. Int J Numer Methods Eng 98(12):881–916
Ramakrishnan B, Rao S (1996) A general loss function based optimization procedure for robust design. Eng Optim 25(4):255–276
Ren X, Rahman S (2013) Robust design optimization by polynomial dimensional decomposition. Struct Multidiscip Optim 48(1):127–148
Rosenblatt M (1952) Remarks on a multivariate transformation. Ann Math Statist 23:470–472
Rubinstein R, Shapiro A (1993) Discrete event systems: sensitivity analysis and stochastic optimization by the score function method. Wiley series in probability and mathematical statistics. Wiley, New York
Shen DE, Braatz RD (2016) Polynomial chaos-based robust design of systems with probabilistic uncertainties. AIChE J 62(9):3310–3318
Shin S, Cho BR (2008) Development of a sequential optimization procedure for robust design and tolerance design within a bi-objective paradigm. Eng Optim 40(11):989–1009
Stephens R, Fatemi A, Stephens RR, Fuchs H (2000) Metal fatigue in engineering. Wiley-Interscience
Sundaresan S, Ishii K, Houser DR (1995) A robust optimization procedure with variations on design variables and constraints. Eng Opt+ A35 24(2):101–117
Taguchi G (1993) Taguchi on robust technology development: bringing quality engineering upstream. ASME Press series on international advances in design productivity, ASME Press
Toropov V, Filatov A, Polynkin A (1993) Multiparameter structural optimization using FEM and multipoint explicit approximations. Struct Multidiscip Optim 6(1):7–14
Wiener N (1938) The homogeneous chaos. Am J Math 60(4):897–936
Xiu D, Karniadakis GE (2002) The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 24:619–644
Yao W, Chen X, Luo W, Van Tooren M, Guo J (2011) Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. Prog Aerosp Sci 47(6):450–479
Zaman K, McDonald M, Mahadevan S, Green L (2011) Robustness-based design optimization under data uncertainty. Struct Multidiscip Optim 44(2):183–197