Probabilistic Engineering Mechanics 24 (2009) 382–396

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

Stochastic optimization using a sparse grid collocation scheme

Sethuraman Sankaran *

Department of Mechanical and Aerospace Engineering, EBU II 467, University of California San Diego, La Jolla, CA 92093-0411, USA

ARTICLE INFO

Article history:
Received 3 October 2007
Received in revised form
5 November 2008
Accepted 14 November 2008
Available online 27 November 2008

Keywords:
Stochastic optimization
Sparse grid collocation
Smolyak algorithm
Inverse problems
Stochastic sensitivities
Robust design

ABSTRACT

In computational sciences, optimization problems are frequently encountered in solving inverse problems for computing system parameters based on data measurements at specific sensor locations, or to perform design of system parameters. This task becomes increasingly complicated in the presence of uncertainties in boundary conditions or material properties. The task of computing the optimal probability density function (PDF) of parameters based on measurements of physical fields of interest in the form of a PDF, is posed as a stochastic optimization problem. This stochastic optimization problem is solved by dividing it into two problems—an auxiliary optimization problem to construct stochastic space representations from the PDF of measurement data, and a stochastic optimization problem to compute the PDF of problem parameters. The auxiliary optimization problem is solved using a downhill simplex method, whilst a gradient based approach is employed for solving the stochastic optimization problem. The gradients required for stochastic optimization are defined, using appropriate stochastic sensitivity problems. A computationally efficient sparse grid collocation scheme is utilized to compute the solution of these stochastic sensitivity problems. The implementation discussed, requires minimum intrusion into existing deterministic solvers, and it is thus applicable to a variety of problems. Numerical examples involving stochastic inverse heat conduction problems, contamination source identification problems and large deformation robust design problems are discussed.

1. Introduction

Stochastic optimization is the process of computing certain parameters of a system when either these parameters are random, some known properties of the system are random, and/or when measurement data are given in the form of a probability distribution function (PDF). Such stochastic problems are rife in the context of estimating boundary heat flux and heat transfer coefficients from temperature measurements, or computing initial concentration (contamination source) profiles from concentration measurements at a later time. The problem of stochastic optimization is also encountered in robust design problems (i.e. design problems wherein either known or unknown parameters of the system are stochastic).

The main goal of this work is to build an efficient computational model that can solve stochastic optimization problems. We are motivated by the following aspects, which have not been answered satisfactorily in literature—(i) How to account for data (in inverse problems) when the experiment/measurement technique is not repeatable or the data is random, (ii) How to perform estimation when the parameters are stochastic and (iii) How to design control parameters of a system when there is uncertainty in either the known or unknown parameters of a system. For example, consider the problem where we want to design the shape of the preform (raw workpiece) so that we minimize flash and underfill in a forged product. The forging velocity will be uncertain (since this is not exactly controllable) in addition to ambient temperatures, pressures etc. It is also possible that the shape of the preform cannot be exactly manufactured. We want the process to perform optimally in spite of such uncertainties. The same is the case when we try to infer probabilistic heat flux based on random temperature measurements at specific sensor locations. Essentially, we are interested in higher order statistics which have not been considered in previous research.

The transition from deterministic optimization problems to stochastic optimization problems has its share of problems and pitfalls – How will you find descent directions (in a gradient-based framework) at different iterative stages in the algorithm (which will be stochastic)? How will the random dimensions be resolved – using Monte-Carlo or more advanced techniques? The answers to these questions are quite pertinent with respect to computational efficiency, as well as to the amount of coding required to overhaul existing deterministic solvers. To simplify the discussion in the paper, we will be referring frequently to the Stochastic Inverse Heat Conduction Problem (SIHCP), though the concept is by no means restricted to this particular problem (this can be applied to a broad class of robust design problems as well). Also, we use the

* Tel.: +1 607 2274154.
E-mail address: sesankar@ucsd.edu.

0266-8920/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.probengmech.2008.11.002

terms stochastic optimization, stochastic inverse and robust design problem interchangeably.

Deterministic techniques for solving inverse problems involving heat conduction applications are detailed in [1]. Numerical techniques for inferring heat fluxes from temperature measurements using iterative regularization schemes are discussed in [2,3]. In [4], an adjoint based approach is employed to compute PDFs of heat fluxes based on PDFs of temperatures. A polynomial chaos approach is employed to resolve the stochastic dimensions present in the problem. In [5], a similar problem is resolved using a Bayesian approach.

The simplest and most intuitive means to deal with randomness in any system is the use of Monte-Carlo techniques—compute optimal parameters for many deterministic realizations of the system and thereby, compute the optimal PDF of parameters. However, the convergence of Monte-Carlo schemes is extremely slow, and in systems where it is computationally expensive to obtain solutions for deterministic algorithms, the task of stochastic optimization becomes increasingly burdensome. The idea used in this paper is based on stochastic collocation, where one carefully chooses realizations of the system where computations will be made such that the convergence with respect to the number of stochastic dimensions is significantly better than Monte-Carlo schemes. The mathematical formalism for doing the same follows from the sparse-grid stochastic collocation scheme [6].

We develop a generic mathematical framework incorporating the sparse-grid collocation framework for solving high-dimensional SIHCPs. In this aspect, we highlight the drawbacks of the stochastic adjoint approach in [4] wherein a similar problem has been dealt with—(a) inability to perform well under increasingly higher-order of stochastic dimensions (b) assumption that the data is provided directly in the stochastic space rather than constructing this stochastic space and (c) its inability to be extensible to non-linear problems wherein adjoint operators cannot be derived (refer to Problem 4 in this paper). In [4], evaluations of the temperatures done using the Generalized Polynomial Chaos approach (GPCE) involves a set of coupled equations. To overcome this disadvantage, a sparse grid stochastic collocation technique is utilized in [7] wherein the direct problem is solved at specific collocation points in the stochastic space. Each direct problem is decoupled and hence, computational complexity is significantly reduced. In addition, we develop a framework for computing stochastic sensitivities from a series of deterministic sensitivity problems. This is needed for computing stochastic gradients in the steepest gradient descent scheme. In order to compute these, we derive a scheme where we only compute deterministic sensitivities at specific collocation points. Another distinct advantage to using such an approach is the minimum coding effort to overhaul existing deterministic solvers and its applicability for systems with non-linear governing equations.

The paper is divided as follows: In Section 2, we provide some mathematical background necessary for the mathematical formalism discussed in this paper. In Section 3, we provide a working definition of stochastic inverse problems. Stochastic sensitivities are dealt with in Section 4. We discuss the mathematical procedure for performing the task of stochastic optimization in Section 5 and follow it up with some numerical examples for inverse heat conduction, inverse concentration reconstruction and robust design problems.

2. Mathematical background

This paper requires some background into probability theory, such as the definition of probability spaces, probability measures, random variables and space–time stochastic processes. There are several texts available in the literature. [8] is an excellent and comprehensive primer on probability theory. In the framework employed here, we define a stochastic space $\boldsymbol{\xi} = [\xi^1, \xi^2, \ldots, \xi^N]$ where ξ^i may represent either uniform or normally distributed random variables. Any construct on the stochastic space has a unique PDF associated with it, and we frequently work with stochastic spaces as opposed to PDFs for the analysis of random fields.

In practice, data (such as temperature measurements) is available only as a PDF. Since we work in the stochastic space for mathematical convenience, the following algorithm was developed to convert a PDF, say $p_{\text{given}}(f)$, into its representation in the stochastic space. In this paper, we restrict $\xi's$ to have either normal or uniform distribution, though the algorithm is generic in nature. Potentially, CDFs derived from the PDF can be utilized to construct the stochastic space. However, this means that an independent random variable is chosen for each sensor location, which is both impractical as well as computationally inefficient. Hence, the following algorithm is carried out for constructing the stochastic space from $p_{\text{given}}(f)$ using two choices-ξ has normal distribution and ξ has uniform distribution:

(1) Set $k = 1$
(2) Set $N = k$ and $P = 0$.
   (a) The representation in stochastic space is given by $f(\boldsymbol{\xi}) = \sum_{i=0}^{P} f_i \psi_i(\boldsymbol{\xi})$ where $\boldsymbol{\xi} = [\xi^1, \xi^2, \ldots, \xi^N]$
   (b) Compute the set of parameters f_i by optimizing: $\int (pdf(f(\boldsymbol{\xi})) - p_{\text{given}}(f))^2 \mathrm{d}f$ using the Nelder–Mead algorithm.
   (c) If the optimal objective function is greater than tolerance, set $P = P + 1$. If $P \leq 6$, continue with Step 2. If $P > 6$, go to Step 3. If the optimal objective function is less than tolerance, terminate.
(3) Set $k = k + 1$ and go to Step 2.

Nelder–Mead is a commonly used nonlinear optimization algorithm for extremely complex search spaces which is the situation herein. The technique employs the simplex technique at successive iterations, and gradually approaches the optimal solution. For further details, the interested reader is referred to [9] and *MATLAB™* is employed for the solution of this problem. This does not affect the time involved in the main algorithm, since this is used as a preprocessing step.

As part of the solution of the stochastic inverse problems, one needs to utilize computational tools for the solution of direct SPDEs. The governing SPDEs are represented in general as $\mathcal{L}u = 0$. The boundary conditions are represented as $\mathcal{B}u = 0$. The uncertainty in the input variables is commonly represented through a Karhunen–Loève (KL) expansion. This expansion can be used only if the covariance function is known a priori and hence, is only suitable for input random fields. The idea behind PCE is to perform the spectral expansion of a random process in terms of polynomial functions. The reader is referred to [10–12] for the solution to SPDEs computed using the GPCE technique. The main disadvantage of this method is that the number of terms grows combinatorially with the number of stochastic dimensions, and the equations for computing the unknown coefficients are coupled. Since the inverse problem requires repeated solutions to such direct problems, we utilize a recently developed technique, the sparse grid stochastic collocation method.

In the stochastic collocation method, we compute solutions at certain fixed locations in the multidimensional stochastic space and use interpolating functions to represent the solutions at other stochastic points [6,13,14]. The Smolyak's algorithm is a way to reduce the number of collocation points necessary for the interpolation in the multi-dimensional random space while simultaneously ensuring that the error does not increase significantly. This has been explained in [7] wherein the sparse grid interpolant is employed. Further details of the algorithm described

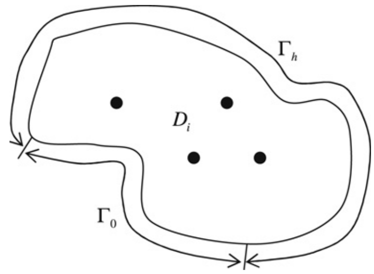

Fig. 1. The figure shows a schematic of the SIHCP problem. Γ_0 represents the boundaries where the heat flux is to be computed, Γ_h represents the domain with known heat flux and D_i represents specific points where PDF of temperature is provided. For robust design problems, only stages 2 and 3 are employed.

in this paper are given in [15–17]. Algorithms for integrations based on sparse grids are provided in [18,19]. Such a sparse grid algorithm was first used for stochastic applications recently in [7] for tackling natural convection problems and in [20] for tackling diffusion problems in random heterogeneous media.

3. The stochastic inverse heat conduction problem (SIHCP)

Let $\mathcal{D}$ be a bounded region in $\mathbb{R}^d$, $d = 1, 2, 3$ with boundary Γ. Let the thermal conductivity $k(\mathbf{x}, \theta)$ and heat capacity $C(\mathbf{x}, \theta)$ be random fields. Let the boundary Γ be divided into Γ_h and Γ_0 with $\Gamma_h \cap \Gamma_0 = \emptyset$, where Γ_h is the part of the boundary Γ with known thermal boundary conditions (here, heat flux). There is no loss of generality in this assumption since problems where there is a boundary condition of the form $T = T_0$ can also be dealt with using the methodology given below. The PDF of heat flux on the boundary Γ_0 is considered unknown. We have to compute the PDF of the unknown stochastic flux on the boundary Γ_0 that yields the PDF of the measured stochastic temperature $Y(\mathbf{x}(D_i), t, \theta)$ at specific points D_i where $i = 1, 2, \ldots, s$, s represents the number of sensors where data is measured (ref. Fig. 1).

The stochastic partial differential equations involved in the direct heat conduction problem are summarized below:

$$C\frac{\partial T}{\partial t} = \nabla \cdot (k\nabla T), \quad (\mathbf{x}, t, \theta) \in (\mathcal{D}, \mathcal{T}, \Omega),$$
$$T(\mathbf{x}, 0, \theta) = \hat{T}(\mathbf{x}, \theta), \quad (\mathbf{x}, t, \theta) \in (\mathcal{D}, \Omega),$$
$$k\frac{\partial T}{\partial n}(\mathbf{x}, t, \theta) = q(\mathbf{x}, t, \theta), \quad (\mathbf{x}, t, \theta) \in (\Gamma_0, \mathcal{T}, \Omega),$$
$$k\frac{\partial T}{\partial n}(\mathbf{x}, t, \theta) = \hat{f}(\mathbf{x}, t, \theta), \quad (\mathbf{x}, t, \theta) \in (\Gamma_h, \mathcal{T}, \Omega). \tag{1}$$

The heat flux q on the boundary Γ_0 is used here as a parameter. It is apparent that for any given q, one can compute the solution $T(\mathbf{x}, t, \theta; q)$.

In the inverse problem, we are seeking a heat flux that minimizes the L_2-error norm between the measured and actual temperatures as computed at the sensor locations. In particular, one looks for a flux $\bar{q}(\mathbf{x}, t, \theta) \in L_2(\Gamma_0 \times \mathcal{T} \times \Omega)$ such that:

$$\mathcal{J}(\bar{q}) \le \mathcal{J}(q), \quad \forall\, q \in L_2(\Gamma_0 \times \mathcal{T} \times \Omega), \tag{2}$$

where, $L_2(\Gamma_0 \times \mathcal{T} \times \Omega)$ is the space of all mean square integrable stochastic processes defined over the spatial and temporal domain Γ_0 and $\mathcal{T}$.

$$\mathcal{J}(q) = \frac{1}{2}\|T(\mathbf{x}, t, \theta; q) - Y(\mathbf{x}, t, \theta)\|^2_{L_2(D_i \times \mathcal{T} \times \Omega)}$$
$$= \frac{1}{2}\int_{\mathcal{T}}\int_{\Omega}\sum_{i=1}^{s}\{T(\mathbf{x}(D_i), t, \theta; q) - Y(\mathbf{x}(D_i), t, \theta)\}^2 \mathrm{d}P\mathrm{d}t, \tag{3}$$

where $T(\mathbf{x}, t, \theta; q) \equiv T(\mathbf{x}, t, \theta; q(\Gamma_0, t, \theta))$ is the solution of the parametric direct stochastic heat conduction problem and $\int_\Omega \bullet\, \mathrm{d}P$ denotes an integral with respect to the probability measure on $(\Omega, \mathcal{F}, \mathcal{P})$.

The task of computing the solution to the inverse problem is divided into the following tasks (see also Fig. 2):

(1) *Pre-processing*—Obtaining measurement of temperature in a discrete form and computing its probability mass function.
(2) *Stochastic optimization*—
   (a) Auxiliary optimization problem defined to represent the input data in stochastic space.
   (b) Solution of direct SPDEs and sensitivity SPDEs.
   (c) Solution to the stochastic optimization method using stochastic gradient based algorithms.
(3) *Post-processing*—Conversion of discrete problem parameters to their respective PDFs.

4. Stochastic sensitivities

4.1. Definition

The main difficulty in solving the optimization problem defined in Eq. (3) is the calculation of the gradient $\mathcal{J}'(q)$ of the cost functional in the function space $L_2(\Gamma_0 \times \mathcal{T} \times \Omega)$. Stochastic sensitivities are interpreted as the change in PDF of the temperature at the sensor locations when the PDF of the heat flux q is perturbed. The sensitivity temperature field (directional derivative) $\Theta(\mathbf{x}, t, \theta; q, \Delta q) \equiv D_{\Delta q}T(\mathbf{x}, t, \theta; q)$ is defined as the linear part in Δq in the Taylor expansion of the process $T(\mathbf{x}, t, \theta; q + \Delta q)$ i.e.

$$T(\mathbf{x}, t, \theta; q + \Delta q) = T(\mathbf{x}, t, \theta; q) + D_{\Delta q}T(\mathbf{x}, t, \theta; q)$$
$$+ \mathcal{O}(\|\Delta q\|^2_{L_2(\Gamma_0 \times \mathcal{T} \times \Omega)}) \tag{4}$$

where $\Delta q \equiv \Delta q(\mathbf{x}, t, \theta)$.

4.2. Governing equations

The stochastic sensitivity problem is defined by linearization of the system of Eq. (1). The governing equations for computing the sensitivity of the temperature with respect to the heat flux are summarized below (refer [4]).

$$C\frac{\partial \Theta}{\partial t} = \nabla \cdot (k\nabla \Theta), \quad (\mathbf{x}, t, \theta) \in (\mathcal{D}, \mathcal{T}, \Omega),$$
$$\Theta(\mathbf{x}, 0, \theta; q, \Delta q) = 0, \quad (\mathbf{x}, \theta) \in (\mathcal{D}, \Omega),$$
$$k\frac{\partial \Theta}{\partial n}(\mathbf{x}, t, \theta; q, \Delta q) = \Delta q(\mathbf{x}, t, \theta), \quad (\mathbf{x}, t, \theta) \in (\Gamma_0, \mathcal{T}, \Omega),$$
$$k\frac{\partial \Theta}{\partial n}(\mathbf{x}, t, \theta; q, \Delta q) = 0, \quad (\mathbf{x}, t, \theta) \in (\Gamma_h, \mathcal{T}.\Omega). \tag{5}$$

It is to be noted that $\Delta q(\mathbf{x}, t, \theta)$ drives the sensitivity problem and hence, it is important to define it numerically.

4.3. Numerical definition of perturbations

There are a number of issues regarding the numerical implementation of stochastic sensitivities which include—(i) What is Δq?, (ii) How do you ensure that $q + \Delta q$ remains normalized and (iii) Which technique is used to solve the set of Eq. (5). Stochastic sensitivities are essential to solve stochastic optimization problems using gradient-based approaches. Since Δq is a stochastic field, we explore if this perturbation can be defined by either perturbing specific coefficients in its spectral expansion or values of q at collocation points.

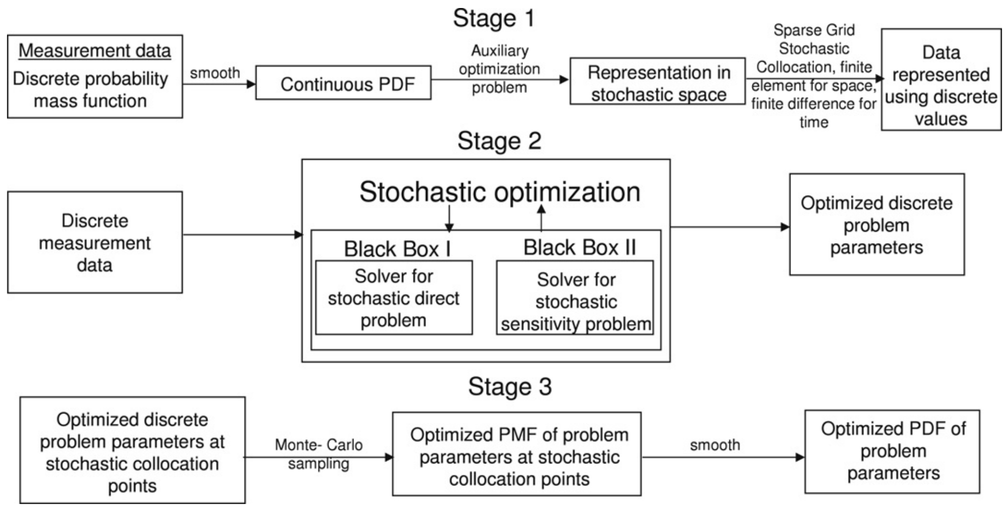

Fig. 2. The figure shows a schematic of the technique used to solve the SIHCP problem.

4.3.1. Method I: Perturbation of spectral coefficients

The unknown random parameters are represented as:

$$q(\mathbf{x}, t, \theta) = \sum_{i=0}^{N} q_i(\mathbf{x}, t)\psi_i(\boldsymbol{\xi}). \tag{6}$$

The perturbation Δq can be defined using perturbations to q_i. This is because the resultant PDF will still be normalized, as justified below.

Normalization: Let $pdf(.)$ be denoted as $h(.)$ and let $cdf(.)$ be denoted as $H(.)$. We have:

$$h(q) = \frac{\partial H}{\partial q}, \tag{7}$$

where $H(\hat{q}) = Prob(\sum_{i=0}^{N} q_i\psi_i(\boldsymbol{\xi}) \leq \hat{q})$. Hence,

$$\int_{-\infty}^{\infty} h(q)\mathrm{d}q = \int_{-\infty}^{\infty} \frac{\partial H}{\partial q}\mathrm{d}q = H(\infty) - H(-\infty) = 1. \tag{8}$$

Note that the GPCE should not be interpreted as a direct representation of the PDF and hence, normalization is not an explicit constraint while using GPCE. The spectral coefficients themselves do not lend any constraints to make the PDF normalized.

4.3.2. Method II: Perturbation of heat flux at collocation points

Perturbation in a collocation coefficient of a stochastic variable implies a valid perturbation of the PDF of the stochastic variable by ensuring its normality. Let us consider interpolation of the heat flux q with Lagrange polynomials using the Smolyak quadrature rule:

$$q(\mathbf{x}, t, \boldsymbol{\xi}) = \sum q_i(\mathbf{x}, t)\mathcal{L}_i(\boldsymbol{\xi}). \tag{9}$$

The proof directly follows from the relations given in the previous part. The only dependence on $\boldsymbol{\xi}$ lies in the interpolating function $\mathcal{L}_i(\boldsymbol{\xi})$. Hence, each coefficient $q_i(\mathbf{x}, t)$ may be perturbed independently and arbitrarily.

We can associate a specific $\Delta q(\mathbf{x}, t, \theta)$ with either perturbations to its GPCE coefficients or perturbations at specific collocation points as shown in Fig. 3. We also show how a perturbation in the collocation space can be converted into a corresponding perturbation in the GPCE coefficients. If $\Delta q(\mathbf{x}, t, \theta) = \Delta q_s(\mathbf{x}, t)\psi_s(\boldsymbol{\xi})$ for some s, then $\Delta q_s(\mathbf{x}, t)\psi_s(\boldsymbol{\xi}) = \sum_i \Delta q(\mathbf{x}, t, \boldsymbol{\xi}_i)L_i(\boldsymbol{\xi})$ where the $\boldsymbol{\xi}_i$'s are the cubature points and $\boldsymbol{\xi}$ is any arbitrary point in the random support space. From the last equation we can then derive that the perturbation at each collocation point is given as:

$$\Delta q(\mathbf{x}, t, \boldsymbol{\xi}_i) = \Delta q_s(\mathbf{x}, t)\psi_s(\boldsymbol{\xi}_i). \tag{10}$$

4.3.3. Implementation in spatial and temporal coordinates

In both the schemes dealt with in the previous section, the perturbation of the coefficients are in the form: $\Delta q_i(\mathbf{x}, t)$. We show how this perturbation will be implemented numerically. $q(\mathbf{x}, t)$ will be represented using its value at specific points in space and time $q(\mathbf{x}_j, t_k)$. The continuous field $q(\mathbf{x}, t)$ is extracted from $q(\mathbf{x}_j, t_k)$ by using linear interpolants in space and time. Finite element and finite-difference schemes are utilized for space and time, respectively. We sequentially choose $\Delta q_i(\mathbf{x}, t) \equiv \Delta q_i(\mathbf{x}_j, t_k)$ for each i. Hence, the number of sensitivity problems computed for each i is the product of the number of spatial and temporal degrees of freedom (DOF). Also, $\Delta q_i(\mathbf{x}_j, t_k) = \delta$, if $\mathbf{x} = \mathbf{x}_j$ and $t = t_k$ and 0 otherwise. δ is chosen to be 0.001 in all our computations.

4.4. Solution to the sensitivity equations

Using the stochastic collocation technique, the sensitivity field as defined in Eq. (5) is here represented as $\Theta(\mathbf{x}, t, \boldsymbol{\xi}; q, \Delta q) = \sum_i \Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)L_i(\boldsymbol{\xi})$ where each of the $\Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)$ is defined from the solution of a deterministic sensitivity problem as given in Eq. (11).

$$\begin{aligned}
&C\frac{\partial \Theta(x, t, \boldsymbol{\xi}_i; q, \Delta q)}{\partial t} = \nabla \cdot (k(\boldsymbol{\xi}_i)\nabla\Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)),\\
&\Theta(\mathbf{x}, 0, \boldsymbol{\xi}_i; q, \Delta q) = 0 \quad (\mathbf{x}, \theta) \in (\mathcal{D}, \Omega),\\
&k(\boldsymbol{\xi}_i)\frac{\partial \Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)}{\partial n} = \Delta q(\mathbf{x}, t, \boldsymbol{\xi}_i) \quad \mathbf{x} \in \Gamma_0\\
&k(\boldsymbol{\xi}_i)\frac{\partial \Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)}{\partial n} = 0 \quad \mathbf{x} \in \Gamma_h.
\end{aligned} \tag{11}$$

Similarly, using perturbations to the GPCE coefficients, the sensitivity field can be represented as $\Theta(\mathbf{x}, t, \boldsymbol{\xi}; q, \Delta q) = \sum_i \Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)\psi_i(\boldsymbol{\xi})$ where each of the $\Theta(\mathbf{x}, t, \boldsymbol{\xi}_i; q, \Delta q)$ is defined from the solution of a deterministic sensitivity problem as given in Eq. (12).

$$\begin{aligned}
&C\frac{\partial \Theta_k(\mathbf{x}, t, \theta; q, \Delta q)}{\partial t}\\
&\quad = \sum_j \nabla \cdot (\langle k(\mathbf{x}, \theta)\psi_j(\theta)\psi_k(\theta)\rangle \nabla\Theta_j(\mathbf{x}, t, \theta; q, \Delta q)),\\
&\Theta_k(\mathbf{x}, 0, \theta; q, \Delta q) = 0, \quad (\mathbf{x}, \theta) \in (\mathcal{D}, \Omega),\\
&\langle k(\mathbf{x}, t, \theta)\psi_j(\theta)\psi_i(\theta)\rangle \frac{\partial \Theta_j(\mathbf{x}, t, \theta; q, \Delta q)}{\partial n} = \Delta q_i(\mathbf{x}, t), \quad \mathbf{x} \in \Gamma_0\\
&\langle k(\mathbf{x}, t, \theta)\psi_i(\theta)\psi_j(\theta)\rangle \frac{\partial \Theta_i(\mathbf{x}, t, \theta; q, \Delta q)}{\partial n} = 0, \quad \mathbf{x} \in \Gamma_h.
\end{aligned} \tag{12}$$

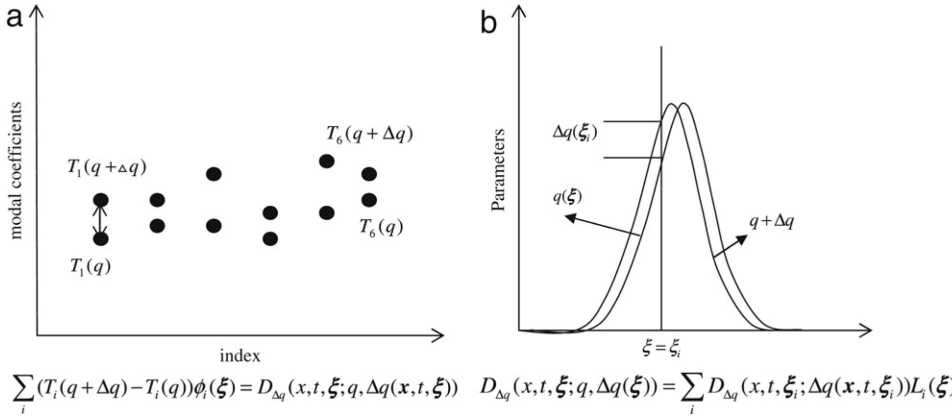

Fig. 3. The figure shows how a specific Δq can be associated with (a) on the left, the perturbation of GPCE coefficients of the heat flux and (b) on the right, the perturbations of heat flux at specific points in the stochastic space.

Note that the equations for Θ are uncoupled in Eq. (11) while they are coupled in Eq. (12). While the stochastic collocation technique was employed for solution of stochastic sensitivity PDE in Eq. (11), the GPCE technique was employed in Eq. (12).

4.5. Discrete optimization problem

It is apparent that the stochastic optimization problem explained in the previous section has to be posed discretely so that numerical schemes can be used to compute the solution. We have shown how each DOF – the spatial, temporal as well as stochastic dimensions – will be discretized. Hence, we are interested in finding the tuple, $q_v = [q_0 q_1 \dots q_{N-1} q_N]$ so as to minimize the objective function, $\mathcal{J}(q)$ (q is trivially constructed from q_v by using the corresponding interpolant).

As a result, we reduce the problem to the following form in the stochastic space (q_i here refers to the values of q at the cubature points):

$$q_v^* = \underset{q_i}{\operatorname{argmin}}\, \mathcal{J}\left(\mathbf{x}, t, \boldsymbol{\xi}; \sum_i q_i L_i(\boldsymbol{\xi})\right). \tag{13}$$

A similar problem can also be defined by using the spectral expansion on q (q_i here refers to coefficients of q in its GPCE expansion).

$$q_v^* = \underset{q_i}{\operatorname{argmin}}\, \mathcal{J}\left(\mathbf{x}, t, \boldsymbol{\xi}; \sum_i q_i . \psi_i(\boldsymbol{\xi})\right). \tag{14}$$

In this work, we will concentrate on the problem of Eq. (13). For a gradient optimization approach to this problem, we will need to utilize sensitivities of the form: $\mathcal{S}_{jki}(\mathbf{x}, t) = \frac{\Theta(\mathbf{x}, t, \xi_i; q, \Delta q(\mathbf{x}_j, t_k, \xi_i))}{\Delta q(\mathbf{x}_j, t_k, \xi_i)}$, where the continuum sensitivity Θ is here defined as the solution of Eq. (11) evaluated for the specific Δq as defined.

5. Optimization scheme

As we had discussed earlier, the aim is to compute a flux $\bar{q}(\mathbf{x}, t, \theta) \in L_2(\Gamma_0 \times \mathcal{T} \times \Omega)$ such that:

$$\mathcal{J}(\bar{q}) \le \mathcal{J}(q), \quad \forall\, q \in L_2(\Gamma_0 \times \mathcal{T} \times \Omega). \tag{15}$$

The strategy employed here is built on (a) the ability to solve stochastic optimization problems with minimal overhaul of existing deterministic codes and (b) be versatile enough to work for a wide range of PDEs. To satiate this need, we employ only direct and sensitivity PDE's for computing the optimal solution. It is to be noted that the efficiency of the optimization algorithm can be improved for specific problems by solving an auxiliary set of equations. For instance, the use of adjoint equations may be utilized in a SIHCP setting wherein conjugate gradient algorithms could be employed. However, the derivation of adjoint operators may not be feasible in complex fluid flow problems, and it is in this spirit that we stick to employing steepest descent schemes in this work. This ensures that the algorithm is generic in nature, while more sophisticated algorithms can be derived for specific problems.

The objective function can be written as:

$$\begin{aligned}\mathcal{J}(\mathbf{q}(\mathbf{x}, t, \boldsymbol{\xi})) = \frac{1}{2}\sum_{k=1}^{s}\int \sum_i \sum_j (T(\mathbf{x}(D_k), t, \boldsymbol{\xi}_i) \\ - Y(\mathbf{x}(D_k), t, \boldsymbol{\xi}_i))(T(\mathbf{x}(D_k), t, \boldsymbol{\xi}_j) \\ - Y(\mathbf{x}(D_k), t, \boldsymbol{\xi}_j))\mathrm{d}t \int L_i(\boldsymbol{\xi}) L_j(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi},\end{aligned} \tag{16}$$

where $\int .\mathrm{d}\boldsymbol{\xi}$ implies that the integration is done as $\int .pdf(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$. The integrals of the form $\int L_i(\boldsymbol{\xi})L_j(\boldsymbol{\xi})pdf(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$ are computed using Monte-Carlo schemes.

The steps to be followed in performing the task of stochastic optimization are summarized below. (The measure $\mathcal{S}_{jki}$ drives the optimization procedure which indicates the variation of physical fields when parameters at specific points in space, time and stochastic space are perturbed.)

(1) Initialize values for $q(\mathbf{x}_j, t_k, \boldsymbol{\xi}_i)$, $q^0(\mathbf{x}_j, t_k, \boldsymbol{\xi}_i) = 0$. Set $k = 0$. The heat flux is $q^0(\mathbf{x}_j, t_k, \boldsymbol{\xi}) = \sum_i q^0(\mathbf{x}_j, t_k, \boldsymbol{\xi}_i)L_i(\boldsymbol{\xi})$.
(2) Solve the direct problem to compute the objective function $\mathcal{J}(q^k(\mathbf{x}, t, \boldsymbol{\xi}))$. Terminate if $k > 0$ and $\mathcal{J}(q^{k+1}) - \mathcal{J}(q^k) < tol$.
(3) Solve a set of $X \times M$ sensitivity problems where X represents the number of spatial and temporal discretizations and M denotes the stochastic discretizations of q. Compute $\mathcal{S}_{jki}(\mathbf{x}, t)$ (defined earlier) and $\mathbf{d}_{ijk} = \frac{\partial \mathcal{J}}{\partial q(\boldsymbol{\xi}_i, \bar{\mathbf{x}}_j, \bar{t}_k)} = \sum_{m=1}^{s}\sum_n \int (T(\mathbf{x}(D_m), t, \boldsymbol{\xi}_n; q) - Y(\mathbf{x}(D_m), t, \boldsymbol{\xi}_n))\mathcal{S}_{jki}(\mathbf{x}(D_m), t)\mathrm{d}t \int L_i(\boldsymbol{\xi})L_n(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$ where $\mathbf{d}_{ijk}$ is written in the form of a vector, say $\mathbf{d}_s$ where s traverses the whole range of ijk.
(4) Update $k = k + 1$. $q_i^k(\mathbf{x}_j, t_k) = q_i^{k-1}(\mathbf{x}_j, t_k) + \alpha \mathbf{d}_{ijk}$ where $\alpha = -\frac{\mathbf{d}^T\mathbf{d}}{\mathbf{d}^T\mathbf{R}\mathbf{d}}$. If $s = ijk$ and $t = lmn$, then $R_{st} = \sum_{m=1}^{s}\int \mathcal{S}_{ijk}(\mathbf{x}(D_m), t)\mathcal{S}_{lmn}(\mathbf{x}(D_m), t)\mathrm{d}t$. Go to step 2.

NOTE: It is to be noted that a similar technique can also be established by utilizing the GPCE scheme for representing the unknown heat fluxes.

For the solution of eigenvalue problems while using Karhunen–Loeve expansion, the SLEPC parallel eigenvalue solver was utilized.

In addition, all computations utilized the PETSC library and were parallelized using MPI. The computing clusters available in Cornell Theory Center (CTC) were utilized for performing the computations.

6. Numerical examples

6.1. Example 1

A Gaussian triangular heat flux (see Eqs. (17) *and* (18)*) is applied at the left end of a one-dimensional heat conducting rod of length* $L = 1$ *units while the right end is insulated (motivated from* [4]*). The temperature is measured at a specific sensor location at* $x^+ = 0.3$ *in the time interval* $[0, 1]$ *and it is desired to reproduce the flux based on these temperature measurements.*

$$q^+_{\text{tri}} = \begin{cases} 2.5t^+, & 0 \le t^+ \le 0.4 \\ 2.0 - 2.5t^+, & 0.4 \le t^+ \le 0.8 \\ 0, & t^+ > 0.8 \end{cases} \tag{17}$$

$$q^+ = \mathcal{N}(q^+_{\text{tri}}, 0.1q^+_{\text{tri}}). \tag{18}$$

The system of direct governing equations are given in Eq. (1) and the system of sensitivity governing equations are given in Eq. (5).

The temperature measurements are taken at a point, $x^+ = d^+ = 0.3$ for all times $t^+ \in \mathcal{T} = [0, 1]$. Deterministic values of $k^+ = 1$ and $C^+ = 1$ were used. An explicit finite difference technique (central-difference in space and forward difference in time) with space discretization of $\Delta x^+ = 0.0025$ and time discretization $\Delta t^+ = 0.025$ was used along with Monte-Carlo technique for obtaining the random temperature 'measurements' (100 temperature measurements at each time). The auxiliary optimization scheme was run to convert the temperature measurements to its representation in the stochastic space. Since the randomness comes due to a Gaussian variation of the heat flux (Eq. (18)), a value of $N = 1$ was sufficient to represent the input randomness. The PDF of the temperature at the sensor location at a specific time and its representation in the stochastic space are shown in Figs. 4 and 5, respectively. It is to be noted that in Fig. 5, the stochastic space construct for measurement data was done using a linear polynomial where ξ is a normal variable. For visualization purposes, since the support is not finite, it was transformed into the uniform interval from 0 to 1. This is done by using the transformation $\hat{\xi} = \frac{erf(\xi)+1}{2}$ where $erf(.)$ is the error function.

The stochastic optimization scheme chalked out in this paper was utilized to compute the optimal PDF of heat flux based on the PDF of temperatures. For the solution of direct and sensitivity problems that are required during the optimization procedure, a space discretization, $\Delta x^+ = 0.025$ and a time discretization, $\Delta t^+ = 0.025$ were used in a finite element framework with linear two-noded elements. Results using the SC and GPCE based optimization algorithms are shown in Figs. 6 and 7 respectively. For performing stochastic optimization using the sparse grid collocation scheme, we used a depth of interpolation 8, and for solving using the polynomial chaos scheme, a third-order GPCE expansion was used.

In Figs. 6 and 7, it is clear that the first four moments of the heat flux is captured sufficiently. It is to be noted that there is a difference noticed at the peak value owing to a step change in the derivative. The same trend is also noticed in [1] and some problems in [4] as shown in Fig. 8 (where the solution is exactly the same everywhere except the peak values where a small error is noticed). It is clear that the stochastic optimization framework developed is accurate enough to capture randomness in the heat flux.

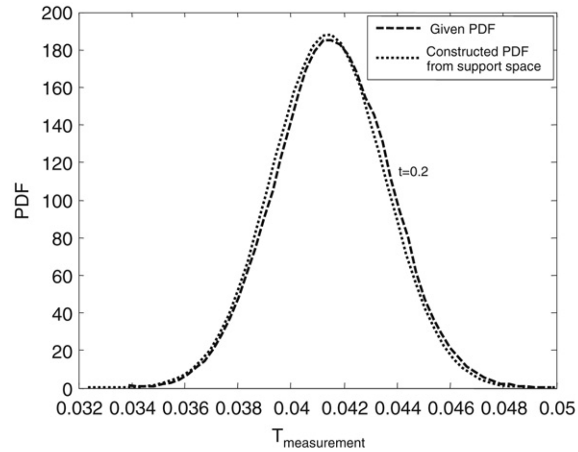

Fig. 4. The figure shows comparison of the measurement PDF with the PDF constructed by solving the auxiliary optimization problem at time $t^+ = 0.2$.

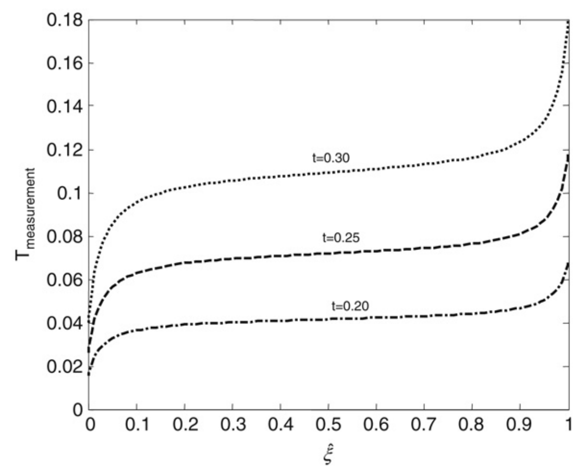

Fig. 5. The figure shows representation of measurement temperature in the support space (the corresponding PDF for $t^+ = 0.2$ is shown in Fig. 4).

NOTE: We show numerically that both the optimal heat flux as well as sensitivities computed at different stages in the spectral and collocation algorithms are identical. We construct a set of numerical tests wherein different Δq's are constructed by varying different coefficients in the GPCE expansion. Table 1 uses the measure $\mathcal{S}_{jki}(\mathbf{x}, t) = \frac{\Theta(\mathbf{x}, t, \xi_i; q, \Delta q(\mathbf{x}_j, t_k, \xi_i))}{\Delta q(\mathbf{x}_j, t_k, \xi_i)}$. In Table 1, we use $\mathcal{S}_{ijk}(0.3, 0.025)$ where the arguments mean that sensitivities are computed at the point $\mathbf{x} = 0.3$ and time $t = 0.025$. Also, the indices in the subscript denote how the heat flux was perturbed, as explained below:

(1) $\mathcal{S}_{jki}(., .)$ indicates a heat flux which is perturbed at $x = x_j$, $t = t_k$ and $\boldsymbol{\xi} = \boldsymbol{\xi}_i$. Since this is a one-dimensional problem, $x_j = 0$ is the only point where we have a boundary heat flux and we specify the perturbation to be at $t_k = 0$.
(2) Each row in Table 1 indicates perturbation of a specific GPCE coefficient. d is used to indicate the dimensionality of the problem and δp indicates which term is perturbed (the perturbation has a magnitude of 0.001). For instance, a value of $\delta p = 3$ in the table means that the third GPCE term was perturbed, which is $0.5 * (3\xi^2 - 1)$ for Legendre polynomials. Using Eq. (10), the same was converted into perturbations at different stochastic collocation points (chosen as Chebychev points with a depth of interpolation, 8).
(3) The measure $\mathcal{S}_{jki}(., .)$ is computed from $\Theta(.)$ which is computed using both the GPCE as well as collocation

Fig. 6. The figure shows comparison of the optimal flux computed using the SC scheme with that of the actual heat flux (the first four moments). Note that nth moment means $\mathbf{E}(.)^n$ where $\mathbf{E}$ denotes expectation operator.

techniques, as defined in Eqs. (12) and (11) respectively. For stochastic collocation, statistics of $\mathcal{S}_{jki}(.,.)$ are constructed and shown in Table 1 (the statistics are taken over the index i since i represents $\boldsymbol{\xi}$). For GPCE, the values shown are statistics of $\frac{\Theta(\mathbf{x}_j, t_k, \boldsymbol{\xi}; q, \Delta q)}{0.001}$. These computed statistics are denoted by $\langle x \rangle$ (mean), $\langle x^2 \rangle$ (second moment) and so on in Table 1.

6.2. Problem 2: Two-dimensional SIHCP on a rolling body

In this section, a practical application of the SIHCP is discussed. Here, we consider rotating bodies that are subject to a boundary heat flux (Fig. 9). It is desired to reconstruct the heat flux based on certain temperature sensors within the rolling body.

Problem definition: A two-dimensional rolling body (Fig. 9(a)) *is subject to a unknown heat flux in one-quadrant on the outer boundary* ($-3\pi/4 \leq \theta \leq -\pi/4$ *where* θ *is measured from positive x-axis). The inner-boundary is insulated while the rest of the outer boundary is subject to a constant temperature of* $T = 25$ °C. *The problem is to recompute the PDF of heat flux given temperature measurements at four points as shown in* Fig. 9(b). *The inner radius is* 1.5 m *while the outer radius is* 3.0 m. *The temperature measurements are made at* $r = 2.8$ m *and* $\theta = -135°, -45°, 45°, 135°$. *The thermal conductivity is taken to be random with a given exponential correlation function. The physical problem is treated as quasi-static.*

The following are the problem parameters used for this problem: Number of sensor nodes = 4, number of stochastic dimensions for thermal conductivity, which are known = 4 (obtained using eigenvalue decomposition and extracting eigenmodes with 99.5% energy), number of unknowns in the equivalent deterministic optimization problem = 35360, number of spatial nodes where the heat flux is unknown = 32.

We assume a quasi-steady state problem. Hence $\frac{\partial T}{\partial t} = 0$ and we do not consider boundary conditions containing time as a parameter. Naturally, the stochastic optimization algorithms detailed, are reduced to a simpler form owing to the absence of a temporal dimension.

Such problems occur frequently in processes such as rolling, where it is desired to compute heat fluxes at the roll contact regions. During the rolling process for manufacturing components, many practical problems such as the wear of the rolls, amount of coolant required at the contact zone, and stresses induced in the workpiece, as well as the rolls, is determined by the amount of heat generated at the contact zone. However, it is impractical to directly place thermo-couples in the contact zone, since they can easily wear away. Hence, a practical solution to this problem is to embed thermocouples within the rolls and recompute the heat flux on the boundary using temperature measurements within the body.

We use a random thermal conductivity defined by an exponential correlation of correlation length, 10, $f(r) = \exp(-r/10)$. The thermal conductivity, k, has the following expansion:

$$k(r, \theta, \boldsymbol{\xi}) = k_0(\theta) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i(\theta) k_i(\mathbf{x}). \tag{19}$$

The resultant eigenvalue problem was solved and the first four modes of k were used in the final analysis, based on the eigenvalue decomposition. Some specific samples of conductivity are shown in Fig. 10.