Comput. Methods Appl. Mech. Engrg. 245–246 (2012) 36–46

Contents lists available at SciVerse ScienceDirect

Comput. Methods Appl. Mech. Engrg.

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

Stochastic collocation with kernel density estimation ☆

Howard C. Elman [a,*], Christopher W. Miller [b]

[a] Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, United States
[b] Department of Applied Mathematics and Scientific Computation, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 7 September 2011
Received in revised form 25 June 2012
Accepted 26 June 2012
Available online 16 July 2012

Keywords:
Stochastic partial differential equation
Stochastic collocation
Kernel density estimation
Adaptive

ABSTRACT

The stochastic collocation method has recently received much attention for solving partial differential equations posed with uncertainty, i.e., where coefficients in the differential operator, boundary terms or right-hand sides are random fields. Recent work has led to the formulation of an adaptive collocation method that is capable of accurately approximating functions with discontinuities and steep gradients. These methods, however, usually depend on an assumption that the random variables involved in expressing the uncertainty are independent with marginal probability distributions that are known explicitly. In this work we combine the adaptive collocation technique with kernel density estimation to approximate the statistics of the solution when the joint distribution of the random variables is unknown.

1. Problem statement

Let (Ω, Σ, P) be a complete probability space with sample space Ω, σ-algebra $\Sigma \subset 2^{\Omega}$ and probability measure $P : \Sigma \to [0,1]$. Let $D \subset \mathbb{R}^d$ be a d-dimensional bounded domain with boundary ∂D. We investigate partial differential equations (PDEs) of the form

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \omega; u) &= f(\mathbf{x}), \quad \forall \mathbf{x} \in D, \quad \omega \in \Omega \\ \mathcal{B}(\mathbf{x}, \omega; u) &= g(\mathbf{x}), \quad \forall \mathbf{x} \in \partial D, \quad \omega \in \Omega. \end{aligned} \tag{1.1}$$

Here $\mathcal{L}$ is a partial differential operator with boundary operator $\mathcal{B}$, both of which can depend on the random parameter ω. As a consequence of the Doob–Dynkin lemma, it follows that u is also a random field, dependent on both the spatial location $\mathbf{x}$ and the event ω. In order to work numerically with the expressions in (1.1), we must first represent the operators in terms of a finite number of random variables $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_M]^T$. This is often accomplished using a truncated Karhunen–Loève (KL) expansion [17]. If we denote $\Gamma = \mathit{Image}(\boldsymbol{\xi})$, then we can write (1.1) as

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\xi}; u) &= f(\mathbf{x}), \quad \forall \mathbf{x} \in D, \quad \boldsymbol{\xi} \in \Gamma \\ \mathcal{B}(\mathbf{x}, \boldsymbol{\xi}; u) &= g(\mathbf{x}), \quad \forall \mathbf{x} \in \partial D, \quad \boldsymbol{\xi} \in \Gamma. \end{aligned} \tag{1.2}$$

For a given realization of the random vector $\boldsymbol{\xi}$, the system (1.2) is a deterministic partial differential equation that can be solved using a deterministic solver. Throughout this paper we assume that $D, \mathcal{L}, \mathcal{B}, f$, and g are defined so that the above problem (1.2) is well posed for all values of $\boldsymbol{\xi} \in \Gamma$. In this paper we will explore several different sampling methods for solving the system (1.2).

One is typically interested in methods that allow statistical properties of u to be computed. If $\rho(\boldsymbol{\xi})$ denotes the joint probability density function of the random vector $\boldsymbol{\xi}$, then the kth moment of the solution u is defined as

$$\mathbb{E}(u^k) = \int_{\Gamma} u^k \rho(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{1.3}$$

One may also be interested in computing probability distributions associated with u, for example $P(u(\mathbf{x}, \boldsymbol{\xi}) \geqslant c)$.

Several methods have been developed for computing approximations to the random field u and the associated statistical quantities. The most widely known is the Monte–Carlo method, where the desired statistics are obtained by repeatedly sampling the distribution of $\boldsymbol{\xi}$, solving each of the resulting deterministic PDEs, and then estimating the desired quantities by averaging. Recently, much attention has been paid to alternative approaches such as the stochastic Galerkin and stochastic sparse grid collocation methods [2,9,12,22,21,27]. These methods typically approximate the solution u as a high-degree multivariate polynomial in $\boldsymbol{\xi}$. If this approximation is denoted $u_p(\mathbf{x}, \boldsymbol{\xi})$, then the error $u - u_p$ can be measured in terms of an augmented Sobolev norm

$$\|\cdot\|_{L^2_P;V} = \left(\int_{\Omega} \|\cdot\|_V^2 dP \right)^{\frac{1}{2}}. \tag{1.4}$$

Here V is an appropriate Sobolev space that depends on the spatial component of the problem and $\|\cdot\|_V$ is the norm over this space. It can be shown that as the total degree of the polynomial approximation is increased, the error in the above norm, $\|u - u_p\|_{L^2_P;V}$, decays very rapidly provided that the solution u is sufficiently smooth in

☆ This work was supported in part by the US Department of Energy under Grant DEFG0204ER25619 and the US National Science Foundation under Grants CCF0726017 and DMS1115317.

* Corresponding author.
E-mail addresses: elman@cs.umd.edu (H.C. Elman), cmiller@math.umd.edu (C.W. Miller).

0045-7825/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cma.2012.06.020

ξ [22]. If u is not sufficiently smooth then the convergence of these methods can stall or they may not converge at all [18]. Several methods have been proposed for treating problems that are discontinuous in the stochastic space. One approach partitions the stochastic space into elements and approximates the solution locally within elements by polynomials, continuous on the domain [3,25]. Another approach is to use a hierarchical basis method developed in [16], which approximates u using a hierarchical basis of piecewise linear functions defined on a sparse grid. This idea was used with stochastic collocation in [18] where the sparse grid is refined adaptively using an a posteriori error estimator.

If the truncated Karhunen–Loève expansion is used to express $\mathcal{L}$ and $\mathcal{B}$, then the random variables $\xi_1, \xi_2, \ldots, \xi_M$ have zero mean and are uncorrelated [17]. It is frequently assumed that the random variables are independent and that their marginal density functions $\rho_i(\xi_i)$ are known explicitly. In this case the joint density function is simply the product of the marginal densities $\rho(\xi) = \Pi_{k=1}^{M} \rho_i(\xi_i)$. This assumption simplifies the evaluation of the moments of the solution since the multidimensional integral in (1.3) can be written as the product of one-dimensional integrals. It is not the case, however, that uncorrelated random variables are necessarily independent, and in the worst case the support of the product of the marginal densities may contain points that are not in the support of the true joint density. Thus, it may not be appropriate to define the joint density function as the product of the marginal density functions. See [13] for further discussion of this point. In this paper we explore a method for approximating the statistics of the solution u when an explicit form of the joint distribution is not available and we only have access to a finite number of samples of the random vector ξ. In particular, we are able to treat the case where information on the parameters of the problem is only available in the form of experimental data. The method works by constructing an approximation $\hat{\rho}(\xi)$ to the joint probability distribution $\rho(\xi)$ using kernel density estimations [23]. This construction is then combined with an adaptive collocation strategy similar to the one derived in [18] to compute an approximation to the random field u. Moments can then be efficiently evaluated by integrating this approximation with respect to the approximate probability measure $\hat{\rho}(\xi)$.

The remainder of this paper proceeds as follows. Section 2 discusses the adaptive collocation method in [18]. Section 3 presents an overview of the kernel density estimation technique used for approximating the unknown distribution of ξ. Section 4 presents the method developed in this paper for approximating solutions to problems of the form (1.2). An error bound for the method is given in Section 4.1, and Section 4.2 presents techniques for extracting solution statistics. Section 5 presents the results of numerical experiments showing the performance of the new method and comparing this performance with that of the Monte Carlo method. Finally in Section 6 we draw some conclusions.

2. The adaptive collocation method

Collocation methods work by solving the Eq. (1.2) for a finite number of pre-determined parameters $\{\xi^{(1)}, \ldots, \xi^{(N_c)}\}$ using a suitable deterministic solver. The solutions at each sample point are then used to construct an interpolant to the solution for arbitrary choices of the random vector ξ. We denote such an approximation generally as $\mathcal{A}(u)(\xi)$. Collocation methods were first used for solving PDEs with random coefficients in [2]. The interpolant was formed using a Lagrangian polynomial basis defined on tensor product grids. The cardinality of these grids is exponential in the dimension of the random vector so that this method is not viable for problems with high-dimensional random inputs. Sparse grid collocation methods were developed in [27] and an error analysis of the method was presented in [22]. These methods use the Smolyak interpolation formula [24] to construct a high-order polynomial interpolant using many fewer points than the full tensor grid. A refinement of this method for problems where the solution depends on the parameters in an anisotropic manner was presented in [21]. For all of these methods, the solution random field is expressed globally as a polynomial in the random vector ξ. These methods are therefore only useful when the random field u is sufficiently regular in ξ.

An adaptive collocation method was developed in [18]. This method is designed to compute approximations of random fields that possess discontinuities or strong gradients, and for which the image set Γ is bounded.[1] In the following, we present an overview of this method and our proposed modifications. To simplify the presentation we describe the case of a function u defined by a single random parameter whose image is a subset of $[0,1]$. This can be generalized in a straightforward manner to a function defined by M parameters with image contained in any M-dimensional hypercube. Define

$$m_i = \begin{cases} 1 & \text{if } i = 1, \\ 2^{i-1} + 1 & \text{if } i > 1, \end{cases} \tag{2.1}$$

$$\xi_j^i = \begin{cases} \frac{j-1}{m_i - 1} & \text{for } j = 1, \ldots, m_i, \quad \text{if } m_i > 1, \\ 0.5 & \text{for } j = 1, \quad \text{if } m_i = 1. \end{cases} \tag{2.2}$$

For $i = 1, 2, \ldots$, we have that $\theta^i = \{\xi_j^i\}_{j=1}^{m_i}$ consists of m_i distinct equally spaced points on $[0,1]$. We also have that $\theta^i \subset \theta^{i+1}$. Since these points are equidistant, the use of global polynomial interpolation as in [27] is not appropriate due to the Runge phenomenon. We make no assumptions on the smoothness of u; for example, it may contain singularities that global polynomial approximations will not resolve. To address these issues, a hierarchical basis of piecewise linear functions is used to construct the interpolant. Define $\theta^0 = \emptyset$ and $\Delta\theta^i = \theta^i \setminus \theta^{i-1}$. Note that $|\Delta\theta^i| = m_i - m_{i-1}$. Let the members of $\Delta\theta^i$ be denoted $\{\xi_j^{\Delta i}\}_{j=0}^{|\Delta\theta^i|-1}$. The hierarchical basis is defined on the interval $[0,1]$ as

$$a_0^1(\xi) = 1 \tag{2.3}$$

$$a_j^i(\xi) = \begin{cases} 1 - (m_i - 1)|\xi - \xi_j^{\Delta i}| & \text{if } |\xi - \xi_j^{\Delta i}| < 1/(m_i - 1), \\ 0 & \text{otherwise}, \end{cases} \tag{2.4}$$

for $i > 1$ and $j = 0, \ldots, |\Delta\theta^i| - 1$; see Fig. 2.1. These functions are piecewise linear and have the property that $a_j^i(\xi_k^{\Delta i}) = \delta_{jk}$, and $a_j^i(\xi_k^s) = 0$ for all $s < i$. Note that there is a binary tree structure on the nodes in θ^i. That is, we can define the set of children of a point $\xi_j^{\Delta i}$ as

$$child(\xi_j^{\Delta i}) = \begin{cases} \{\xi_j^{\Delta i+1}\} & \text{if } i = 2 \\ \{\xi_{2j}^{\Delta i+1}, \xi_{2j+1}^{\Delta i+1}\} & \text{otherwise}. \end{cases} \tag{2.5}$$

We also denote the parent of a point in this tree as $par(\xi_j^{\Delta i})$.

Algorithm 1 defines an interpolation scheme using the hierarchical basis functions.

Algorithm 1. Interpolation with hierarchical basis functions

Define $\mathcal{A}_0(u)(\xi) = 0$.
Define $k = 1$
repeat
  Construct $\Delta\theta^k$
  Evaluate $u(\xi_j^{\Delta k})\ \forall \xi_j^{\Delta k} \in \Delta\theta^k$
  $w_j^k = u(\xi_j^{\Delta k}) - \mathcal{A}_{k-1}(u)(\xi_j^{\Delta k})\ \forall \xi_j^{\Delta k} \in \Delta\theta^k$
  Define $\mathcal{A}_k(u)(\xi) = \sum_{i=1}^{k} \sum_{j=0}^{|\Delta\theta^i|-1} w_j^i a_j^i(\xi)$.
  $k = k + 1$
until $max(|w_j^{k-1}|) < \tau$

[1] For unbounded Γ, interpolation is carried out on a bounded subset of Γ, see e.g. [26].