Journal of Computational Physics 250 (2013) 477–510

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

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

A probabilistic graphical model approach to stochastic multiscale partial differential equations

Jiang Wan [a], Nicholas Zabaras [a,b,*]

[a] Materials Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, Cornell University, 101 Frank H.T. Rhodes Hall, Ithaca, NY 14853-3801, USA
[b] Center for Applied Mathematics, Cornell University, 657 Frank H.T. Rhodes Hall, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 4 December 2012
Received in revised form 30 March 2013
Accepted 7 May 2013
Available online 24 May 2013

Keywords:
Bayesian
Probabilistic graphical models
Uncertainty quantification
Multiscale modeling
Sequential Monte Carlo
Stochastic partial differential equations
Belief propagation

ABSTRACT

We develop a probabilistic graphical model based methodology to efficiently perform uncertainty quantification in the presence of both stochastic input and multiple scales. Both the stochastic input and model responses are treated as random variables in this framework. Their relationships are modeled by graphical models which give explicit factorization of a high-dimensional joint probability distribution. The hyperparameters in the probabilistic model are learned using sequential Monte Carlo (SMC) method, which is superior to standard Markov chain Monte Carlo (MCMC) methods for multi-modal distributions. Finally, we make predictions from the probabilistic graphical model using the belief propagation algorithm. Numerical examples are presented to show the accuracy and efficiency of the predictive capability of the developed graphical model.

1. Introduction

Physical systems generally have inherent randomness which could result from uncertainties in boundary or initial conditions, material heterogeneities and so on. It is therefore necessary to include parameters characterizing these uncertainties into the model system. In order to study the propagation of uncertainties from the parameter space to the response space, stochastic differential equations are constructed such that both input parameters and model responses are modeled as random variables. The scope of this work focuses on stochastic partial differential equations (SPDEs).

Over the past few decades, many methods and algorithms have been developed to solve SPDEs. The most celebrated one is the Monte Carlo (MC) method. As a sampling method, the deterministic solver is called for each realization of the stochastic input for one to obtain the statistics of the solution. The convergence rate does not depend on the dimension of the parameter space, but is of order $\mathcal{O}(n^{-1/2})$ with n realizations. To accelerate convergence, quasi Monte Carlo methods and several efficient sampling techniques have been developed as alternatives. Another group of methods refers to nonsampling approaches, typically perturbation algorithms based on a series representation of the stochastic solution. These methods are limited to small fluctuations and low-order statistics of the solution. Recently, many research efforts have been devoted to the study of schemes based on spectral representation of the stochastic solution [1]. For example, the well-established stochastic Galerkin method approximates the solution in a multivariate polynomial space or in anisotropic tensor product polynomial spaces [2]. Stochastic collocation methods using sparse grids based on the Smolyak algorithm

* Corresponding author at: Materials Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, Cornell University, 101 Frank H.T. Rhodes Hall, Ithaca, NY 14853-3801, USA. Tel.: +1 607 255 9104; fax: +1 607 255 9410.
E-mail addresses: nzabaras@gmail.com, zabaras@cornell.edu (N. Zabaras).
URL: http://mpdc.mae.cornell.edu/ (N. Zabaras).

0021-9991/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcp.2013.05.016

[3] have a weaker dependence on the dimensionality of the problem and recently have been applied extensively to various uncertainty quantification problems [4–8].

In spite of the tremendous progress in solving SPDEs, the curse of dimensionality remains a long-standing challenge. For conventional stochastic Galerkin methods, the computational cost depends on the number of expansion terms which grows exponentially as a function of the dimensionality of stochastic input space [9]. Stochastic collocation methods can achieve fast convergence rate by taking advantage of multidimensional polynomial interpolation. However, the number of collocation points required to achieve sufficient accuracy increases exponentially for high-dimensional problems [10]. Therefore, many efforts have been devoted to stochastic methods that deal with high-dimensions. In [7], an adaptive sparse grid collocation (ASGC) method is proposed such that the collocation points are selected automatically based on the smoothness of the stochastic domain as detected by the magnitude of the hierarchical surpluses. The ASGC can successfully solve stochastic elliptic problems up to 100 dimensions when not all stochastic dimensions are equally important [8]. However, the convergence rate deteriorates even for problems with moderate input dimensionality when all stochastic dimensions are equally weighted. Another direction is to decompose the original problem into sub-problems with low-dimensional input. A typical example is the high-dimensional model representation (HDMR) techniques that capture the high-dimensional relationships between input and output model variables and generate a collection of low-dimensional sub-problems in stochastic space [11]. Recent progress in HDMR can be found in [8,12]. In [9], a low-rank separated representation of the solution to SPDEs with high-dimensional inputs is obtained using an alternating least-squares approach.

More challenges arise when multiscale phenomena are taken into account in high-dimensional stochastic problems. In such cases, information across scales contains a certain level of uncertainties but assessment of uncertainty propagation often leads to large computational cost. Let us take, for example, fluid flow through porous media occurring from large geological scales down to microscopic scales. Full-scale spatial and temporal resolution simulations may require significant computational resources. Since the sample-based stochastic methods mentioned above, such as MC, ASGC, and HDMR, call the deterministic solvers repetitively, efficient solvers for multiscale partial differential equations are of great importance in reducing overall computational cost. For this purpose, computational techniques, such as the multiscale finite element (MsFEM) method [13–15], variational multiscale (VMS) method [16,17], the heterogeneous multiscale (HMM) method as well as their variants [10,18,19] and multigrid methods [20,21], have been developed to solve a coarse-scale problem that captures fine-scale effects without resolving all the fine-scale features.

So far, many efforts for solving multiscale stochastic problems focus on decoupling multiscale deterministic solvers from stochastic approaches. In this work, we propose a new scheme to quantify the uncertainties propagated in multiscale systems based on a probabilistic model of SPDE solutions. Inference problems can be solved directly on this probabilistic model without sampling-based methods or calling expensive deterministic solvers. The stochastic input and model responses are all treated as random variables. However, conventional regression models are inefficient or even impractical to represent their relationships when the stochastic input is in a high-dimensional space. This curse of dimensionality can be overcome by utilizing probabilistic graphical models which have been intensively studied and widely used in machine learning and Bayesian statistics for multivariate statistical modeling [22,23]. A probabilistic graphical model consists of a collection of nodes and edges that connect pairs of nodes. Each node represents a random variable or a group of random variables and the edges express probabilistic relationships between variables, e.g. conditional independence properties. By combining both graph theory and probability theory, the complicated relationships between all variables can be modeled explicitly and the resulting graph expresses a decomposition of a joint distribution as a product of functions of subsets of variables. Given the graph-based probabilistic model of model responses, efficient algorithms for inference on graphical models can be directly applied. If we treat the stochastic input as observed variables, the probabilistic model becomes a conditional distribution of model responses on input variables, which leads to a surrogate model. The predictions of model responses can be evaluated by inference algorithms on the graphical model associated with this surrogate model. If the stochastic input field also has an explicit graphical representation, we can directly estimate the marginal distributions of unobserved variables in the graphical model by integrating out all the other variables with an efficient algorithm.

This paper is organized as follows. In Section 2, we introduce the classical flow in random porous media benchmark multiscale problem to set forward the presentation of the methodology. In Section 3, we construct the framework of probabilistic modeling of coarse-scale model responses. Then we introduce a sequential Monte Carlo method to learn the hyperparameters in the graphical model from training data in Section 4. The belief propagation algorithm presented in Section 5 is utilized to predict model responses based on the probabilistic graphical model. Case studies on the stochastic porous media flow problem are provided in Section 6.

2. Problem definition

To model uncertainties in a physical system, we define a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where Ω is a sample space, $\mathcal{F}$ a σ-algebra of subsects of Ω and $\mathcal{P} : \mathcal{F} \to [0, 1]$ a probability measure. Let $D \subset \mathbb{R}^d$ be a fixed d-dimensional bounded domain with boundary ∂D. A general stochastic partial differential equation (SPDE) is formulated as

$$\mathcal{L}(\boldsymbol{x}, \omega; \boldsymbol{y}) = 0, \quad \forall \boldsymbol{x} \in D, \tag{1}$$

with boundary conditions