Stochastic galerkin and collocation methods for quantifying uncertainty in differential equations: a review

J. D. Jakeman¹  S. G. Roberts²

(Received 14 August 2008; revised 21 January 2009)

Abstract

The article reviews the mathematical theory of stochastic Galerkin and stochastic collocation methods, focusing on their strengths and limitations. The aim is to construct a first stop, widely accessible document that directs a reader to more detailed descriptions of stochastic Galerkin and stochastic collocation methods that are suitable for their application of interest. References point to rigorous convergence proofs and accuracy estimates, computational considerations and numerical examples. A supplementary document gives a quick look-up guide to the strengths and weaknesses of stochastic Galerkin and collocation methods.

Contents

1 Introduction
1 Introduction

Differential equations are used to model a wide range of systems and processes in engineering, physics, biology, chemistry and the environmental sciences. These systems are subject to a wide range of uncertainty in initial and boundary conditions, model coefficients, forcing terms and geometry. The effects of such uncertainty should be traced through the system thoroughly enough to allow one to evaluate their effects on prediction of model outputs. In addition to comparing model predictions to observed data, detailed investigation of model behaviour, independent of observational data, should also be used to help ascertain confidence levels in model predictions [14].

Associated computational methods and software tools are needed to facilitate analysis of model sensitivity and uncertainty. Traditional Monte Carlo techniques are often infeasible due to the large CPU time needed to run the model in question. Polynomial chaos (PC) expansions [29] arose as an efficient means of representing stochastic processes. PC expansions are based on a probabilistic framework and represent stochastic quantities as spectral
1 Introduction

expansions of orthogonal polynomials. Section 2 presents the mathematical framework of polynomial chaos.

Stochastic Galerkin (SG) methods employ PC expansions to represent the solution and inputs to stochastic differential equations [2, 11, 33]. A Galerkin projection minimises the error of the truncated expansion and the resulting set of coupled equations solved to obtain the expansion coefficients. SG methods are highly suited to dealing with ordinary and partial differential equations and have the ability to deal with steep non-linear dependence of the solution on random model data [13]. Provided sufficient smoothness conditions are met, PC estimates of uncertainty converge exponentially with the order of the expansion and, for low dimensions, come with small computational cost. Section 3 describes the general procedure of stochastic Galerkin methods and common variants.

SG necessitates the solution of a system of coupled equations that require efficient and robust solvers and the modification of existing deterministic code. Often the forms of the governing equations and/or the deterministic code used to solve the equations are complicated and make implementing PC difficult or even impossible. A non-intrusive method, referred to as stochastic collocation (SC) [20, 26, 32], addresses this limitation. SC methods utilise interpolation methods and project a set of deterministic simulations, evaluated using carefully chosen sampled parameter sets, onto a polynomial basis. This approach is very useful when endeavouring to quantify uncertainty in models implemented with complex deterministic code which are not easily modified. Similar to SG methods, SC methods achieve fast convergence when the solutions possess sufficient smoothness in random space. Section 4 discusses SC methods.

This review should be read in conjunction with that of Xiu [31] who provides illustrative examples of some of the points made here.
2 Problem definition

To quantify the uncertainty in a system of differential equations we adopt a probabilistic approach and define a complete probability space \((\Omega, \mathcal{F}, \mathcal{P})\). This space consists of an event space \(\Omega\), comprising of possible outcomes \(\omega\), a \(\sigma\)-algebra \(\mathcal{F}\) and a probability measure \(\mathcal{P}\). Utilising this framework, the uncertainty in a model is introduced by representing the model input data as random fields.

2.1 Governing equations

Consider the general differential equation defined on a \(d\)-dimensional bounded domain \(D \in \mathbb{R}^d\), \(d = 1, 2, 3\),

\[
\mathcal{L}(x, t, \omega, p(\omega); u) = f(x, t, \omega, p(\omega)), \quad \text{for all } x \in D, \ t \in (t_0, T],
\]

where \(x = (x_1, \ldots, x_d) \in \mathbb{R}^d\), \(d \geq 1\), are the coordinates in \(\mathbb{R}^d\), \(\mathcal{L}\) is a linear or non-linear differential operator, \(u(\omega) = (u_1(\omega), \ldots, u_i(\omega)) \in \mathbb{R}^i\), \(i \geq 1\), are the unknown solution quantities, and \(p(\omega) = (p_1(\omega), \ldots, p_j(\omega)) \in \mathbb{R}^j\), \(j \geq 1\), are the input data, either parameters or stochastic processes, characterising the governing equations. Note that we omit equations for the boundary and initial conditions for convenience.

We are interested in finding the stochastic solution \(u : \Omega \times D \rightarrow \mathbb{R}\) such that for \(\mathcal{P}\)-almost everywhere \(\omega \in \Omega\), equation (1) holds.

2.2 Representing random input data

To solve equation (1) numerically we must invoke the ‘finite noise assumption’ \([1]\) to reduce the infinite dimensional probability space \((\Omega, \mathcal{F}, \mathcal{P})\) to a finite dimensional space. Employing any truncated spectral expansion of the stochastic process in the probability space we characterise the random inputs
by a set of $N$ random variables $\xi = (\xi_1(\omega), \ldots, \xi_N(\omega))$. By the Doob–Dynkin Lemma [24], the solution can then be described by the same set of random variables, that is $u(x, t, \omega) = u(x, t, \xi)$.

The parametric uncertainty of (1) can be found by modelling the parameters $p = (p_1(\omega), \ldots, p_j(\omega)), j = N$, as an $N$-variate random vector $\xi$. If the input data $p = (p_1(x, t, \omega), \ldots, p_j(x, t, \omega)), j \geq 1$, of equation (1) are spatially and/or temporally dependent random fields, an alternative decomposition of the stochastic processes must be used. A popular choice used to represent a stochastic field is the finite term, Karhunen–Loève (KL) expansion [17]. This decomposition is based on the spectral expansion of the covariance function of the process $p_j(x, t, \omega)$:

$$p_j(x, t, \omega) = E(p_j)(x, t) + \sum_{i=1}^{N} \sqrt{\lambda_i} \phi_i(x, t) \xi_i(\omega), \quad (2)$$

where $E(p_j)(x, t)$ is the expected value of the process $p_j$, $\lambda_i$ and $\phi_i(x, t)$ are the eigenvalues and eigenfunctions, also referred to as Empirical Orthogonal Functions (EOFs), of the covariance function, and $\xi_i$ are uncorrelated random variables dependent on the form of the covariance kernel. Mathematical analysis and numerical studies of the convergence of KL expansions were performed by Sewhab [25] and Huang [12], respectively.

### 2.3 Polynomial chaos

Polynomial chaos expansions provide a means of representing the second order stochastic processes in equation (1). Specifically we represent any second order stochastic process $X(\omega)$ by

$$X(x, t, \omega) = \sum_{i=0}^{\infty} x_i(x, t) \Phi_i(\xi(\omega))$$
and this approximation converges in an $L_2$-sense to any continuous functional in $L_2$ \cite{4}. Here $\{\Phi_i(\xi)\}$ forms a complete orthogonal basis, so that

$$\langle \Phi_i, \Phi_j \rangle = \langle \Phi_i^2 \rangle \delta_{ij},$$

where $\delta_{ij}$ is the Kroneker delta and $\langle \cdot, \cdot \rangle$ is the inner product in the Hilbert space $L_2(\Omega, \mathcal{F}, \mathcal{P})$ determined by the support $\Gamma$ of the random variable

$$\langle f(\xi), g(\xi) \rangle = \int_{\Gamma} f(\xi)g(\xi)\rho(\xi) \, d\xi,$$

with $\rho(\xi)$ denoting the weighting function of the type of random variable used. The Hermite chaos expansion (PC expansion with Hermite polynomial basis) has been used effectively to solve stochastic differential equations with Gaussian inputs \cite{33, 34}. But according to the theorem of Cameron and Martin \cite{4}, the Hermite chaos expansion will converge for arbitrary second order random processes. For example, Ghana \cite{11} employed Hermite chaos to model log-normal processes. However, the use of Gaussian processes results in optimal exponential convergence \cite{19}. In some cases, the performance of Hermite chaos has been shown to decrease substantially when representing random processes with non-Gaussian inputs \cite{34}. To achieve faster convergence rates for non-Gaussian inputs, Xiu and Karniadakis \cite{33} developed a generalisation of Wiener’s chaos known as Wiener–Askey polynomial chaos or generalized polynomial chaos (gPC). This method chooses the orthogonal basis $\{\Phi\}$ so that the weighting function in the orthogonality relation (3) has the same form as the probability distribution of the random variables $\xi$. Xiu and Karniadakis \cite{33} showed computationally that the Wiener–Askey scheme achieves exponential convergence, with respect to the order of the approximation, when the appropriate random distribution and orthogonal basis are chosen. A description of the numerical challenges of employing PC expansions of stochastic fields was given by Debusschere \cite{6}.
According to the Cameron and Martin Theorem we can approximate the solution \( u \) of (1) as a truncated generalised polynomial chaos expansion involving \( P + 1 \) expansion terms:

\[
\begin{align*}
    u(x, t; \xi) = \sum_{i=0}^{P} u_i(x, t) \Phi_i(\xi)
\end{align*}
\]

where \( P \) depends on the number of dimensions \( N \) of \( \xi \) and the order \( M \) of the polynomial \( \{\Phi\} \) is set according to the required accuracy \( P + 1 = (N + M)!/(N!M!) \). Wiener–Askey polynomials result in the best rate of convergence when used with their associated random distribution [33]. Orthogonal polynomials can also be constructed numerically to deal with arbitrary probability measure [33]. Once an appropriate basis has been chosen the truncated gPC expansions of the solution and input data are substituted into the governing equations. If the input data are random fields, then the Karhunen–Loève expansion (2) must first be projected onto a gPC basis before being substituted. A Galerkin projection then minimises the error in the gPC expansion. This is achieved by finding the solution \( \mathbf{u}(x, t, \xi) \in V_\Gamma \subset L_2(\Omega, \mathcal{F}, \mathcal{P}) \) to the weak form of (1):

\[
\begin{align*}
    \int_{\Gamma} \mathcal{L}(x, t, \xi; \mathbf{u}) v(\xi) \rho(\xi) \, d\xi = \int_{\Gamma} f(x, t, \xi) v(\xi) \rho(\xi) \, d\xi,
\end{align*}
\]

for all \( v(\xi) \in V_\Gamma \) and \( x \in D \). Utilizing the orthogonality condition equation (3) yields a set of \( P + 1 \) coupled deterministic equations that can then be solved using appropriate numerical methods. The form of an SG method is dependent on the construction of the subspace \( V_\Gamma \) used in (5). The space spanned by the gPC Wiener–Askey polynomials are one such construction. Other choices include tensor products of one dimensional, fixed order, polynomial spaces [2, 5] and sparse tensor product spaces [27] (not to be confused with sparse grid interpolation).
Once the gPC approximation of the solution is obtained, various statistical measures are easily calculated. For instance, the mean and variance of the gPC approximation are \( \bar{u}(x, t) = u_0(x, t) \) and \( \sigma^2(x, t) = \sum_{i=1}^{P} u_i^2(x, t) \langle \Phi_i^2 \rangle \), respectively. The gPC expansion provides a complete functional representation of the stochastic solution to (1). One can easily evaluate a particular solution corresponding to a specific realisation of the random variables \( \xi \), provided the data lies within the support of \( \xi \). Knio [13] overviews SG methods in the context of computational fluid dynamics.

### 3.1 Alternative formulations

The SG method has four main limitations.

1. Computational complexity increases rapidly with the number of random inputs and the order of the expansion.

2. A high expansion order is required when the dependence of the solution on the random input data varies rapidly or if a singularity exists in the random space [28].

3. Errors in the gPC approximation of transient solutions may become unacceptably large even after only a short time [18].

4. SG necessitates the solution of a system of coupled equations requiring efficient, robust solvers and modification of deterministic code.

Several techniques have been proposed to overcome the problems associated with long term integration and steep dependence in the random space. Wan and Karniadakis [28], Le Maître et al. [15] and Babuska et al. [2] developed \( k-p \) gPC approximations which combine local \( (-k) \) refinement in the stochastic space with increases in the polynomial degree \( (-p) \). Wan [28] advocated the use of Multi-Element generalised polynomial chaos (ME-gPC) which adaptively decomposes the random input space into sub-domains or elements. This technique takes advantage of the observation that SG is more
efficient for random variables with standard deviation which is relatively small in comparison to its mean. Although utilising the fast (exponential) convergence of gPC approximation, the complexity of this method reduces the observed rate of convergence. Le Maître [15] advocates the use of an orthogonal projection of uncertain data and solution variables onto a wavelet basis, consisting of compact, piecewise smooth, polynomial functions. The use of a localised polynomial basis results in a low order rate of convergence. Both these localised expansions produce more robust schemes than the global gPC expansion and the adaptive decomposition of random space efficiently controls the error of long term integration. However, these methods are based upon tensor product spaces in higher dimensions and suffer the curse of dimensionality.

Adaptive-p-type methods were developed by Li [16] and Lucor [18]. These methods are based on the observation that some of the terms in the gPC representation of the solution to (1) do not contribute significantly to its value. Only the terms which have the greatest contribution to the solution are kept.

## 4 Stochastic collocation methods

Recently a new technique referred to as stochastic collocation (SC) [1, 20, 32] has arisen, to address some limitations of SG. SC endeavours to combine the strengths of non-intrusive sampling and SG. As with Monte Carlo methods, SC requires only the solution of a set of decoupled equations, allowing the model to be treated as a black box and run with existing deterministic solvers. Provided the solutions possess sufficient smoothness in the random space, SC methods maintain the fast convergence of SG [32].

In 1D SC seeks to approximate the solution to (1) using Lagrange interpo-
4 Stochastic collocation methods

\[ u(x, t, \xi) = \sum_{k=0}^{Q} u_k(x, t, \xi^{(k)}) L_k(\xi), \] (6)

where \( u_k(x, t, \xi^{(k)}) \) is the solution at the collocation points \( \{\xi^{(k)}\}_{k=0}^{Q} \) and \( L_i \) is the Lagrange polynomial of order \( Q + 1 \) that satisfies the usual relation \( L_i(\xi^{(j)}) = \delta_{ij} \).

Similar to the SG approach we wish to minimise the error in the SC approximation using the Galerkin method to find the solution \( u(x, t, \xi) \in V_r \) to the weak form (5) of the governing equations (1). The orthogonality property of the Lagrange interpolating polynomials decouples the resulting set of equations so that the SC method only requires solving \( Q + 1 \) uncoupled deterministic problems at each collocation point:

\[ L(x, t, \xi^{(k)}; u) = f(x, t, \xi^{(k)}). \]

Furthermore, we use Gaussian quadrature to recover the gPC polynomial expansion of the solution:

\[ \hat{u}_k = \frac{\langle u(x, t; \xi) \Phi_k(\xi) \rangle}{\langle \Phi_k^2(\xi) \rangle} \simeq \frac{1}{\langle \Phi_k^2(\xi) \rangle} \sum_{k=0}^{n} u(\xi^{(k)}) \Phi_k(\xi^{(k)}) w^{(k)}, \quad k = 0, \ldots, P, \]

where \( w^{(k)} \) is the Gaussian quadrature weights associated with \( \xi^{(k)} \), and \( \langle \Phi_k^2(\xi) \rangle \) is evaluated analytically. To best utilise the interpolation and Gaussian quadrature framework, abscissa should correspond to the collocation points of the \( Q + 1 \) Gaussian quadrature rule associated with the weighting function of the random variable \( \xi \). For example, if \( \xi \) is uniformly distributed, the abscissa are the Gauss–Legendre collocation points.

Xiu [30] showed that the overall error of SC can be decomposed into three parts: the truncation error of the finite term gPC expansion; the aliasing error of the \( Q + 1 \) point integration rule; and the error introduced by discretising the deterministic problem through the use of a finite element scheme such as the Runge–Kutta solver.
4 Stochastic collocation methods

4.1 Multi-dimensional problems

Although quadrature is well developed for the univariate case, less is known about the multivariate case. But univariate quadrature can easily be extended to multiple dimensions using tensor products of one dimensional approximations. Mathelin [20] and Babuska [1] presented tensor product approximation schemes. Consider the interpolant of the solution $u$ to (6) defined over the support of $\xi$. For every dimension we select $Q_i + 1$ nodal points $(\xi_i^{(0)}, \ldots, \xi_i^{(Q_i)})$ in the support of $\xi_i$ and subsequently construct the one dimensional quadrature approximation. The multivariate ($N > 1$) quadrature rule is then simply the tensor product of the one dimensional approximations based on the $N$-dimensional nodal set and associated weights $\{\xi^{(k)}, w^{(k)}\}_{k=0}^Q$:

$$(I_1 \otimes \cdots \otimes I_N)[u] = \sum_{j_1=0}^{Q} \cdots \sum_{j_N=0}^{Q} u(\xi_1^{(j_1)}, \ldots, \xi_N^{(j_N)}) \cdot (w_1^{(j_1)} \cdot w_2^{(j_2)} \cdots w_N^{(j_N)})$$

where $\xi^{(k)} = (\xi_1^{(k)}, \ldots, \xi_N^{(k)})$, $w^{(k)} = (w_1^{(k)}, \ldots, w_N^{(k)})$ and $I_i$ is the 1D quadrature rule in the $i$th dimension. If an equal number of abscissa are chosen in each dimension, then the above formula requires $(Q + 1)^N$ collocation points. The full tensor product scheme shown above becomes infeasible for $N$ large enough. Consequently the need arises for more sophisticated sampling techniques. This leads directly to the use of sparse grid tensor product approximations.

4.2 Sparse Grid spaces

Sparse grid tensor product approximation has emerged as the most useful tool to address the curse of dimensionality experienced by SC [23, 32]. Sparse grids have been extensively used for high dimensional interpolation and quadrature [3, 10]. The approach can yield several orders of magnitude reduction in the number of collocation points required to achieve the same level of accuracy as the full tensor grid approach.
Standard sparse grids are isotropic, treating all dimensions equally. Although an advance on full tensor product spaces, such approximations can still be improved. Many problems vary rapidly in only some dimensions, remaining much smoother in other dimensions. Consequently, it is advantageous to increase the level of accuracy only in certain non-smooth dimensions, resulting in so-called adaptive or anisotropic grids. In some cases the important dimensions can be determined a priori, but in most cases the collocation points must be chosen during the computational procedure. Examples of dimension adaptive, sparse grid, collocation schemes were given by Gana [9] and Nobile [22].

Foo et al. [8] proposed a multi-element probabilistic collocation method (ME-PCM) which adaptively discretises the random space into elements and implements a local isotropic sparse grid collocation problem on each element. To avoid an exponential growth in the number of elements, only a small subset of the important dimensions are refined.

Xiu [32] employed Stroud cubature points as an alternative to sparse grids in higher dimensions. The method has relatively low order accuracy but employs the minimal number of points for its corresponding algebraic accuracy [21]. To address the low order accuracy of Stroud SC, Ding et al. [7] proposed an adaptive Stroud cubature method which subdivides the leading random dimensions and applies Stroud SC in these subdivided elements.

References


References


References


Author addresses

1. **J. D. Jakeman**, Department of Mathematics, The Australian National University, Canberra, AUSTRALIA. 
   mailto:john.jakeman@anu.edu.au

2. **S. G. Roberts**, Department of Mathematics, The Australian National University, Canberra, AUSTRALIA.