Quantification of Uncertainty in Flow Simulations Using Probabilistic Methods

Gianluca Iaccarino
Mechanical Engineering
Institute for Computational Mathematical Engineering
Stanford University

Non-equilibrium gas dynamics
From physical models to hypersonic flights

VKI Lecture Series - Sept. 8-12, 2008

Contents

1 Introduction and motivation

In the last three decades, computer simulation tools have achieved wide spread use in the design and analysis of engineering devices. This has shortened the overall product design cycle and it has also provided better understanding of the operating behavior of the systems of interest. As a consequence numerical simulations have lead to a reduction of physical prototyping and to lower costs.

In spite of this considerable success, it remains difficult to provide objective confidence levels in quantitative information obtained from numerical predictions. The complexity arises from the amount of uncertainties related to the inputs of any computation attempting to represent a physical system. As a result, especially in the area of reliability and safety, physical testing remains the dominant mechanism of certification of new devices. Rigorous quantification of the errors and uncertainties[1] introduced in numerical simulations is required to establish objectively their predictive capabilities.

Procedures to establish the quality of numerical simulations have been organized within the framework of Verification and Validation (V&V) activities [34]. Verification is a mathematical process that aims at answering the question: "are we solving the equations correctly?". The objective is to quantify the errors associated to the algorithms employed to obtain the solution of the governing equations. Validation, on the other hand, aims at answering the question "are we solving the correct equations?". The goal is to identify the appropriateness of the selected mathematical/physical formulation to represent the device to be analyzed. Validation always involves comparisons of the numerical predictions to reality, whereas verification *only* involves numerical analysis and tests.

There is a growing recognition of the fact that validation cannot be carried out without explicitly accounting for the uncertainties present in both the measurements and the computations. Experimentalists are typically required to report *uncertainty bars* to clearly identify the repeatability and the errors associated to the measurements. Validation must be carried out acknowledging the nature of the experimental uncertainties and by providing a similar indication of the computational *error bars*. One of the objective of uncertainty quantification methods is to construct a framework to estimate the error bars associated to given predictions. Another objective is to evaluate the likelihood of a certain outcome; this obviously leads to better understanding of risks and improves the decision making process.

[1] The difference between errors and uncertainties will be given later

2 Definitions and basic concepts

The uncertainty quantification community has introduced precise definitions to characterize various types of uncertainties.

2.1 Errors vs. uncertainties

The American Institute of Aeronautics and Astronautics (AIAA) "Guide for the Verification and Validation of CFD Simulations" [35] defines *errors* as recognisable deficiencies of the models or the algorithms employed and *uncertainties* as a potential deficiency that is due to lack of knowledge.

This definition is not completely satisfactory because does not precisely distinguish between the mathematics and the physics. It is more useful to define *errors* as associated to the *translation* of a mathematical formulation into a numerical algorithm (and a computational code).

Errors are typically also further classified in two categories: acknowledged errors are known to be present but their effect on the results is deemed negligible. Examples are round-off errors and limited convergence of certain iterative algorithms. On the other end, unacknowledged errors are not recognizable[2] but might be present; implementation mistakes (bugs) or usage errors can only be characterized by comprehensive verification tests and procedures.

Using the present definition of errors, the *uncertainties* are naturally associated to the choice of the physical models and to the specification of the input parameters required for performing the analysis. As an example, numerical simulations require the precise specification of boundary conditions and typically only limited information are available from corresponding experiments and observations. Therefore variability, vagueness, ambiguity and confusion are all factors that introduce uncertainties in the simulations. A more precise characterization is based on the distinction in aleatory and epistemic uncertainties.

2.2 Aleatory uncertainty

Aleatory uncertainty[3] is the physical variability present in the system being analysed or its environment. It is not strictly due to a lack of knowledge

[2] In principle, using the AIAA definition, unacknowledged errors could be considered uncertainties because they are associated to lack of knowledge

[3] Aleatory uncertainty is also referred to as variability, stochastic uncertainty or irreducible uncertainty.

and cannot be reduced. The determination of material properties or operating conditions of a physical system typically leads to aleatory uncertainties; additional experimental characterization might provide more conclusive evidence of the variability but cannot eliminate it completely. Aleatory uncertainty is normally characterized using probabilistic approaches.

2.3 Epistemic uncertainty

Epistemic uncertainty[4]. is what is indicated in the AIAA Guide (AIAA 1998) as "uncertainty"[5], i.e. a potential deficiency that is due to a lack of knowledge. It can arise from assumptions introduced in the derivation of the mathematical model used or simplifications related to the correlation or dependence between physical processes. It is obviously possible to *reduce* the epistemic uncertainty by using, for example, a combination of calibration, inference from experimental observations and improvement of the physical models. Epistemic uncertainty is not well characterized by probabilistic approaches because it might be difficult to infer any statistical information due to the *nominal* lack of knowledge. A variety of approaches have been introduced to provide a more suitable framework for these analysis. Typical examples of sources of epistemic uncertainties are turbulence modeling assumptions and surrogate chemical kinetics models.

2.4 Sensitivity vs. uncertainty analysis

Sensitivity analysis (SA) investigates the connection between inputs and outputs of a (computational) model; more specifically, it allows to identify how the variability in an output quantity of interest is connected to an input in the model and which input sources will dominate the response of the system. On the other hand, uncertainty analysis aims at identifying the overall output uncertainty in a given system. The main difference is that sensitivity analysis does not require input data uncertainty characterization from a real device; it can be conducted purely based on the mathematical form of the model. As a conclusion large output sensitivities (identified using SA) do not necessarily translate in important uncertainties because the input uncertainty might be very small in a device of interest.

SA is often based on the concept of sensitivity derivatives, the gradient of the output of interest with respect to input variables. The overall sensitivity

[4]Epistemic uncertainty is also called reducible uncertainty or incertitude

[5]Aleatory uncertainty is not mentioned in the AIAA Guidelines

is then evaluated using a Taylor series expansion, which, to first order, would be equivalent to a linear relationship between inputs and outputs.

3 Predictions under uncertainty

Computer simulations of an engineering device are performed following a sequence of steps.

Initially the system of interest and desired performance measures are defined. The geometrical characterization of the device, its operating conditions, the physical processes involved are identified and their relative importance must be quantified. It is worthwhile to point out that the definition of the system response of interest is a fundamental aspect of this phase. The next step is the formulation of a mathematical representation of the system. The governing equations and the phenomenological models required to capture the relevant physical processes need to be defined. In addition, the precise geometrical definition of the device is introduced. This step introduces simplification with respect to the real system; for example small geometrical components are eliminated, or artificial boundaries are introduced to reduce the scope of the analysis. With a well defined mathematical representation of the system, the next step if to formulate a discretized representation. Numerical methods are devised to convert the continuous form of the governing equations into an algorithm that produces the solution. This step typically requires, for example, the generation computational grid, which is a tessellation of the physical domain. Finally the analysis can be carried out.

The introduction of uncertainty in numerical simulations does not alter this process but introduces considerable difficulties in each phase. It is useful to distinguish three steps: data assimilation, uncertainty propagation and certification.

3.1 Data assimilation

Data assimilation consists of a study of the system of interest that aims at identifying the properties, physical processes and other factors required to fully characterize it. The analysis is typically focused on the specific inputs required by the mathematical framework that will be applied in the simulations. As an example, the boundary conditions required in numerical simulations should be *inferred* from observation of the device of interest or specific experiments. Given the limited degree of reproducibility of experimental measurements and the errors associated to the measurement

techniques, these quantity are known with a certain degree of uncertainty (typically specified as an interval, $x \pm y\%$). Probabilistic approaches treat these quantities, that overall characterize the aleatory uncertainty, as *random* variables assuming values within specified intervals. In mathematical terms this is equivalent to define random variables with a specified probability distribution functions (PDF) [13]. The obvious choice is to use random variables defined using analytical distributions (Gaussian, uniform, etc.). It is difficult to justify this choice [5] solely from experimental evidence because of the limited amount of data typically available[6]; in many situation the only data available are obtained from *expert opinions* and can lead to ambiguous or conflicting estimates. Alternative approaches have been devised to provide a more flexible framework to handle this situation, evidence theory is one such approach [37, 38].

In the context of probabilistic approaches, the objective of data assimilation is to define PDFs of each of the input quantity used in the computational tool.

3.2 Uncertainty propagation

Once probability distributions are available for all the input quantities in the computational algorithm, the objective is to compute the PDFs of the output quantities of interest. This step is usually the most complex and computationally intensive for realistic engineering simulations. A variety of methods are available in the literature, from sampling based approaches (e.g. Monte Carlo) to more sophysticated stochastic spectral Galerkin approaches. In the next section these methods are described in detail.

3.3 Certification

Once the statistics of the quantity of interest have been computed, various metrics can be used to characterize the system output, depending on the specific application. The most common use of such statistical information is a reliability assessments, where the likelihood of a certain outcome is estimated and compared to operating margins. In a validation context, the PDF (or more typically the cumulative distribution function) is compared to experimental observation to extract a *measure* of the confidence in the numerical results. The characterization of these measures, so-called validation

[6]Note that the specification of an interval is not equivalent to a specification of a uniform probability distribution. An interval formally does not contain probabilistic information!

metrics, is an active area of research [33].

4 Probabilistic uncertainty propagation

Within a probabilistic framework, the problem of uncertainty propagation consists of the generation of PDFs of the outcomes given (known) distribution of all the input parameters. Several classes of methods have been developed to solve this problems; in this section three popular approaches are described.

Consider the vector $\mathbf{x} = (x_1, ...x_D)$ containing the input quantities to the computational model; assume that $y = g(\mathbf{x})$ is the output of interest; g is possibly the result of a complex fluid dynamic simulation.

In probabilistic uncertainty quantification approaches the stochastic, input quantities $\mathbf{x}$ are represented as independent continuous random variables $x_i(\omega_i)$[7] mapping the sample space Ω_i to real numbers, $x_i : \Omega_i \rightarrow \mathbb{R}$. This assumption in practical terms increases the dimensionality of the problem: the original *deterministic* outcome $y = f(x_1, ..., x_i, ...x_D)$ becomes a stochastic quantity $y = f(x_1, ..., x_i, ...x_D : \omega_1, ..., \omega_i, ..., \omega_D)$. The objective is to compute the PDF of y, f_y, in order to evaluate the likelihood of a certain outcome, or, in general, statistics of y. The expected value $\mathbb{E}[y]$ and the variance $Var[y]$ are defined as:

$$\mathbb{E}[y] = \int_{\infty}^{\infty} z f_y(z) dz \tag{1}$$

$$Var[y] = \int_{\infty}^{\infty} (z - \mathbb{E}[y])^2 f_y(z) dz = \mathbb{E}[y^2] - (\mathbb{E}[y])^2 . \tag{2}$$

Note that y is a stochastic variable while the expected value and the variance are deterministic quantities.

4.1 Sampling techniques

Sampling-based techniques are the simplest approaches to propagate uncertainty in numerical simulations: they involve repeated simulations (also called realizations) with a proper selection of the input values. All the results are then collected to generate a statistical characterization of the outcome. In the following the Monte Carlo approach and the Latin Hypercube sampling strategy are described.

[7] The assumption of independence is mathematically stated as $f_{x_1,...,x_D} = f_{x_1} \cdots f_{x_D}$, where $f_{x_1,...,x_D}$ the joint PDF is the product of the marginals

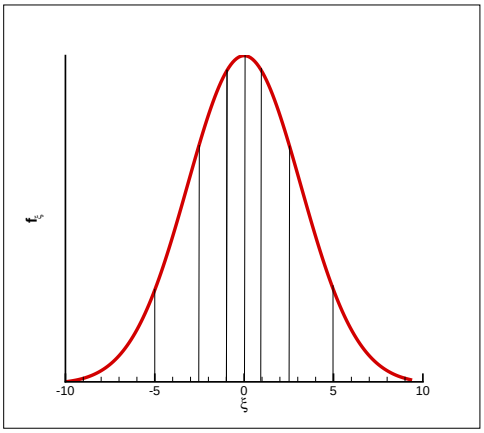

Figure 1: Equally probability intervals defined on the input parameter space for a Gaussian PDF. Each of the interval is sampled in the Latin hypercube method.

4.1.1 Monte Carlo method

The Monte Carlo method [31] is the oldest and most popular sampling approach. It involves random sampling from the space of the random variables x_i according to the given PDFs. The outcome is typically organized as a histogram and the statistics are readily computed from Eq. (2) by replacing the integrals with sums over the number of samples. The method has the advantage that it is simple, universally applicable and does not require any modification to the available (deterministic) computational tools. It is important to note that while the method converges to the exact stochastic solution as the number of samples goes to infinity, the convergence of the mean error estimate is slow. Hence thousands or millions of data samples may be required to obtained accurate estimations. However, the convergence does not directly depend on the number of random variables in the problem. In this form the Monte Carlo methods always give the correct answer, but a prohibitively large number of realizations may be required to accurately estimate responses that have a small probability of occurrence. On the other hand, the convergence of the low order statistics (expected value and variance) require much smaller number of samples.

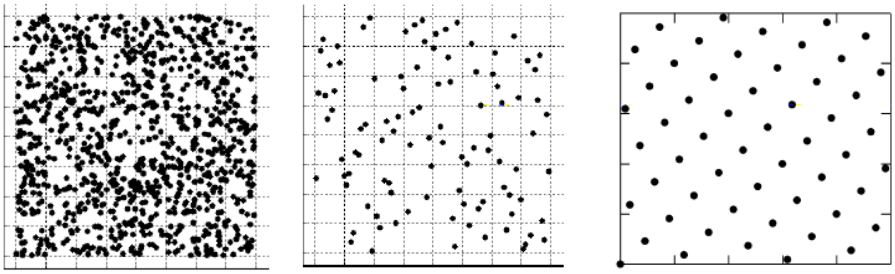

Figure 2: Sampling based techniques with two uncertain inputs. (left) Monte Carlo; (center) Latin Hypercube; (right) Lattice based

4.1.2 Latin Hypecube approach

Several methods have been developed to accelerate the Monte Carlo approach. One of the most successful is the Latin Hypercube sampling (LHS) [29] approach. The range of each input random variable x_i is divided in M intervals with equal probability. In Fig. 1 this is illustrated for a single input Gaussian random variable. M random samples are drawn, one from each of the intervals and the convergence is faster than Monte Carlo because the occurrence of low probability samples is reduced. The construction is especially effective in multiple dimensions; in Fig. 2 an example of samples generated using Monte Carlo and Latin Hypercube for two uniform random variables is reported. The LHS samples projected on each of the axis provide *optimal* coverage of the parameter space [30]. On the other hand, the Monte Carlo samples show clusters and holes that eventually lead to the slower convergence. In Fig. 2 another strategy is illustrated: Lattice-based sampling. In this approach, the parameter space is *discretized* using regularly spaced points (lattice); once the solution is evaluated at those locations a random shift is applied to the lattice and another set of solutions are computed[31]. The choice of the lattice is related to the distribution of the input variables similarly to the LHS approach and the objective is again to reduce the occurrence of clusters and holes in the samples.

4.2 Quadrature methods

One of the objective of UQ propagation methods is the computation of the statistics of an outcome of interest, such as its expectation and the variance. As shown earlier, these require the evaluation of integrals (over

the parameter space) and it is therefore natural to employ *conventional* numerical integration techniques[8].

Let's consider a problem with one uncertain parameter ξ; the objective is to compute integrals of $y(\xi)$. A class of quadrature rules are based on interpolating basis functions that are easy to integrate, typically polynomials. The integral is expressed as a weighted sum of the integrand y evaluated in a finite number of locations on the ξ-axis (abscissas). The choice of the polynomial basis defines the weights and the corresponding abscissas.

The simplest example is the *midpoint rule* in which a sequence of M equally spaced values $\xi^{(i)}$ in the interval $[a, b]$ ($x_1 = a$ and $x_M = b$) is considered. The function y is evaluated at $M-1$ abscissas as $y((\xi^{(i)} + \xi^{(i+1)})/2) = y_{i+1/2}$. The integral is simply obtained as the sum of the $M-1$ values $y_{i+1/2}$ times the size of the interval $(b-a)$. In this case the basis functions are piecewise constant and the weights are all equal to $(b-a)/M$. Quadratures based on equally spaced abscissas include the commonly used trapezoidal and Simpson rules and are, in general, referred to as Newton-Cotes formulas.

4.2.1 Stochastic collocation

Stochastic collocation refers to quadrature methods used to compute integrands of random variables, thus over a stochastic domain. Although Newton-Cotes formulas are applicable in this context, it is usually preferable to consider more general approaches, in which the abscissas are not equally spaced. Gaussian quadratures are popular in the field of uncertainty analysis because of their high accuracy[9]; the general format is:

$$\int_a^b y(\xi) f(\xi) d\xi = \sum_{k=0}^{M} w(\xi_k) y(\xi_k) + R_M(y) \tag{3}$$

where the abscissas ξ_k are associated with zeros of *orthogonal* polynomials and w_k are weights and $f(\xi)$ is a weighting function; in the present context, f represents a PDF. R_M is the remainder term which determines the order of accuracy of the integration. In particular, the formula (3) is said to have polynomial degree of exactness p if $R_M(y) = 0$ when $y(\xi)$ is a polynomial of degree $\leq p$. Newton-Cotes formulae have $p = M - 1$ whereas Gaussian

[8] Numerical integration and quadrature are typically synonym for one dimensional functions. In higher dimensions the term *cubature* is often used instead of quadrature

[9] High order polynomial interpolating functions constructed on equally spaced grids also suffer from the Runge's phenomenon.

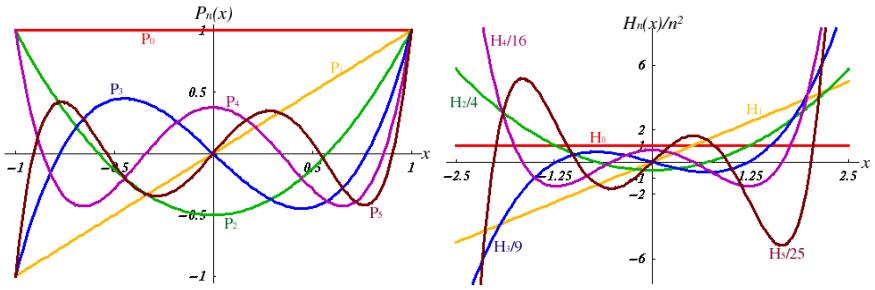

Figure 3: (Legendre (left) and Hermite (right) polynomials.

quadratures achieve higher p with the same number of function evaluations $(M+1)$, specifically $p = 2M+1$. This is the main advantage of Guassian quadrature and is accomplished by virtue of the orthogonality condition of the interpolating basis functions[24] [4], which is in general stated as:

$$\langle p_m, p_n \rangle = \int_a^b p_m(\xi) p_n(\xi) f(\xi) d\xi = \delta_{nm} \tag{4}$$

where δ_{nm} is the Kroneker symbol.

The most commonly used form of Guassian quadrature is the Gauss-Legendre integration formula which is based on Legendre polynomials $\ell_k(\xi)$. In Fig. 3 the first few polynomials are reported. These are constructed using the three term recurrence relation (with $\ell_0(\xi) = 1$ and $\ell_1(\xi) = \xi$):

$$(n+1)\ell_{n+1}(\xi) = (2n+1)\xi\ell_n(\xi) - n\ell_{n-1}(\xi)$$

The weight function in Eq. (3) is simply $f(\xi) = 1$; the integration weights are computed as

$$w_k = \int_a^b \ell_k(\xi) d\xi$$

and the abscissas ξ_k for an M-point quadrature rule are the zeros of the ℓ_{M+1} polynomial. As an example the computation of expectation of an output in the presence of an uncertain input defined in a interval $[a, b]$ (as a uniform random variable) can be computed as:

$$\mathbb{E}(y) = \int_a^b y(\xi) f_y(\xi) d\xi = \sum_{k=0}^{M} y(\xi_k) w_k \tag{5}$$

For uncertain parameters defined on unbounded domains such as random variables described by Gaussian probability distributions, the Gauss-Hermite quadrature is used [19]. In this case the three term recurrence relation is ($h_0(\xi) = 1$ and $h_1(\xi) = \xi$):

$$h_{n+1}(\xi) = 2\xi h_n(\xi) - 2n h_{n-1}(\xi)$$

and the weight function is $f(\xi) = e^{-\xi^2}$. In this case the integration weights w_k are expressed as:

$$w_k = \int_{-\infty}^{\infty} h_k(\xi) e^{-\xi^2} d\xi$$

and the expectation can be computed as before just evaluating the outcome at $M + 1$ abscissas that correspond to the zeros of the Hermite polynomial h_{M+1}.

There is an obvious connection between the choice of the polynomial and the probability distribution of the input variables: the orthogonality condition (Eq. 4). In the present context, in order to achieve the polynomial exactness of Gaussian quadratures for computing the expectation, the polynomials have to be orthogonal with respect to the weight function that corresponds to the PDF of the input variable.

The two Gauss quadrature rules described above require the computation of a new set of abscissas (and corresponding function evaluations) for each desired order M; alternative approaches have the property that the abscissas are nested, e.g. different accuracy formulas share abscissas. A popular nested formulation rule is the Clenshaw-Curtis quadrature[17].

In practical terms, collocation methods for uncertainty propagation require the evaluation of zeros and weights for a family of orthogonal basis functions; these can be computed and stored in advance. A set of $M + 1$ independent computations are performed and the results are combined to obtain the statistics of the output of interest. Collocation can therefore be interpreted as a sampling technique; it retains the main advantage of the Monte Carlo method because it does not require modifications to the existing computational tool.

The evaluation of the PDF of the output quantities is somewhat more complicated for stochastic collocation methods than the computations of the output statistics. The first step is the construction of the polynomial interpolant of the solution $y(\xi)$ in the parameter space which involves the use of the $M+1$ solutions $y(\xi^{(i)})$. At this point, the interpolant is used as a

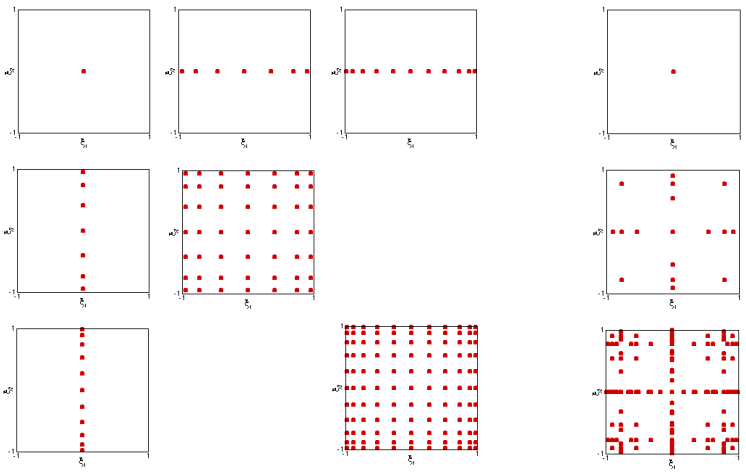

Figure 4: Tensor (left) and sparse (right) grid quadrature rules for two-dimensional functions.

replacement for the original function[10] and Monte Carlo sampling is used.

4.2.2 Extension to multiple dimensions

The natural extension of interpolation techniques to multiple dimensions is a tensor product of one-dimensional interpolants. Let $G(\xi_1, \ldots, \xi_D)$ be the outcome of interest, expressed as a function of D random variables. The set of points for the multi-dimensional interpolant is the tensor product of the abscissas for the one-dimensional interpolants; in the case that the polynomial order is the same in every direction, the total number of abscissas is $(M+1)D$. Although the tensor product extension of the one-dimensional quadrature rules is simple and accurate, it clearly suffers from the exponential increase in the number of abscissas - and consequently required function evaluations - as the number of dimensions increases [20]. To combat this, some have proposed using an alternative extension based on the sparse grid construction, attributed to Smolyak [15]. In this case the multi-dimensional product is constructed as a linear combination of tensor product interpolants, each with relatively few points in its respective point set. This greatly reduces the total number of abscissas - particularly as the dimension D increases - while retaining the accuracy and convergence

[10]The approximated function is typically called a surrogate response surface

properties of one-dimensional interpolants.

Fig. 4 illustrates the extension of Gauss-Legendre quadrature in two-dimensions using tensor products and sparse grid rules. Different set of abscissas (M is varied from 1 to 11) are reported to illustrate how, for sparse grids, the number of points and corresponding function evaluations remains relatively low even for high order polynomial interpolants.

Sparse grid constructions reduce the *curse of dimensionality* and converge very rapidly if the function of interest is smooth. Precise convergence results and comparisons to Monte Carlo methods are reported in [9]. Extensions of this approach to construct anisotropic grids are currently an active area of research; the objective in this case is to increase the order of the interpolating polynomials only in selected directions to further reduce the computational cost.

Other integration methods in high dimensions are also available in the literature, for example Stroud formulae [10] although have not yet been explored in the area of uncertainty quantification.

4.3 Spectral methods

The third class of methods for uncertainty propagation is based on spectral methods in which the solution is expressed using a suitable series expansion. These approaches are *intrusive* in the sense that the mathematical formulation requires modification of the existing deterministic codes - in contrast to the sampling approaches and the stochastic collocation methods described before. For this reason, spectral methods have to be described with reference to a given mathematical problem; in the following the Burgers equation is used.

4.3.1 Stochastic Galerkin approach

In this approach the unknown quantities are expresses as an infinite series of orthogonal polynomials defined on the space of he random input variable[11]. This representation has its roots in the work of Wiener [23] who expressed a Gaussian process as an infinite series of Hermite polynomials. Ghanem and Spanos [22] truncated Wiener's representation and used the resulting finite series as a key ingredient in a stochastic finite element method.

As an example, consider the 1D Burgers equations

$$u_t + uu_x = 0$$

[11] Stochastic Galerkin methods based on polynomial expansions are often referred to as *Polynomial Choas* approaches [22]

and assume that the initial condition is uncertain and characterized by one parameter, ξ a Guassian random variable.

$$u(x, t = 0) = \xi * sin(x)$$

In stochastic Galerkin approaches the (unknown) solution $u = u(x, t, \xi)$ is expressed as a spectral expansion as

$$u(x, t, \xi) = \sum_{i=0}^{\infty} u_i(x, t)\Psi_i(\xi) \tag{6}$$

where the coefficients $u_i(x, t)$ are independent of the random variable and, therefore, deterministic (they are still function of the physical dimensions). The $\Psi_i(\xi)$ are Hermite polynomials and form a complete orthogonal set of basis functions.

The expansion (6) is truncated to a certain order M and the approximated expression for u is inserted in the governing equation:

$$\sum_{i=0}^{M} \frac{\partial u_i}{\partial t}\Psi_i(\xi) + \left(\sum_{j=0}^{M} u_j\Psi_j(\xi)\right)\left(\sum_{i=0}^{M} \frac{\partial u_i}{\partial x}\Psi_i(\xi)\right) = 0$$

Multiplying by $\Psi_k(\xi)$ and integrating over the probability space - Galerkin projection - we obtain a system of M coupled & deterministic equations for the coefficients u_i:

$$\sum_{i=0}^{M} \frac{\partial u_i}{\partial t}\langle\Psi_i\Psi_k\rangle + \sum_{i=0}^{M}\sum_{j=0}^{M} u_i\frac{\partial u_j}{\partial x}\langle\Psi_i\Psi_j\Psi_k\rangle = 0 \qquad \text{for } k = 0, 1, ..., M. \tag{7}$$

Note that in Eq. (7) the double and triple products of the Hermite polynomials are simply numbers [24] (cfr. Eq. 4):

$$\langle\Psi_i\Psi_j\rangle = \delta_{ij} i!$$

$$\langle\Psi_i\Psi_j\Psi_k\rangle = 0 \quad \text{if } i + j + k \text{ is odd or } \max(i, j, k) > s$$

$$\langle\Psi_i\Psi_j\Psi_k\rangle = \frac{i!j!k!}{(s - i)!(s - j)!(s - k)!} \quad \text{otherwise}$$

and $s = (i + j + k)/2$.

The statistics of the solutions can be readily computed from the expansions coefficients. In particular the expectation is

$$\mathbb{E}[u] = \mathbb{E}\left[\sum_{k=0}^{M} u_k \Psi_k\right] = u_0 \mathbb{E}[\Psi_0] + \sum_{k=1}^{M} u_k \mathbb{E}[u_k] = u_0$$

for the orthogonality of the polynomial basis. The variance is computed as:

$$Var[u] = \mathbb{E}\left[(u - \mathbb{E}[u])^2\right] = \mathbb{E}\left[\left(\left(\sum_{k=0}^{M} u_k \Psi_k\right) - u_0\right)^2\right] =$$

$$\mathbb{E}\left[\left(\sum_{k=1}^{M} u_k \Psi_k\right)^2\right] = \sum_{k=1}^{M} u_k^2 \mathbb{E}\left[\Psi_k^2\right] = \sum_{k=1}^{M} u_k^2 (k!)^2$$

As an example, the polynomial chaos formulations for the Burgers equations can be written explicitly. For $M = 2$ the expansion (6) is $u(x,t,\xi) = u_0 + \xi u_1$ and the governing system of equations is:

$$\begin{aligned} \frac{\partial u_0}{\partial t} + u_0 \frac{\partial u_0}{\partial x} + u_1 \frac{\partial u_1}{\partial x} &= 0 \\ \frac{\partial u_1}{\partial t} + u_1 \frac{\partial u_0}{\partial x} + u_0 \frac{\partial u_1}{\partial x} &= 0 \end{aligned}$$

For $M = 3$ the expansion (6) is $u(x,t,\xi) = u_0 + \xi u_1 + 2(\xi^2 - 1)u_2$ and the governing system of equations is:

$$\begin{aligned} \frac{\partial u_0}{\partial t} + u_0 \frac{\partial u_0}{\partial x} + u_1 \frac{\partial u_1}{\partial x} + 2u_2 \frac{\partial u_2}{\partial x} &= 0 \\ \frac{\partial u_1}{\partial t} + u_1 \frac{\partial u_0}{\partial x} + (u_0 + 2u_2) \frac{\partial u_1}{\partial x} + 2u_1 \frac{\partial u_2}{\partial x} &= 0 \\ \frac{\partial u_2}{\partial t} + u_2 \frac{\partial u_0}{\partial x} + u_1 \frac{\partial u_1}{\partial x} + (u_0 + 4u_2) \frac{\partial u_2}{\partial x} &= 0 \end{aligned}$$

It is worth to note that the choice of the polynomial basis is again connected to the distribution of the input random variables. As for stochastic collocation approaches Hermite and Legendre polynomials are used for Gaussian and uniform random variables respectively. A general formulation based on the Askey polynomials is presented in [21].

While the stochastic Galerkin technique has been shown to produce highly accurate statistics, it does requires extensive modifications to existing computational methods to solve for the coefficients of the expansion.

4.3.2 Extension to multiple dimensions

The presence of multiple uncertain parameters increases considerably the computational complexity of the stochastic Galerkin approach although it does not change the overall formulation.

It is formally possible to formulate the basis polynomials in terms of a full tensor product of one-dimensional polynomials similarly to what considered for the stochastic collocation approach. Ghanem and Spanos [22] introduced a different expansion. For example for a problem with two uncertain parameters expressed in terms of Gaussian random variables the first few bi-variate Hermite polynomials are:

$$\begin{aligned} \Phi_0(\xi_1,\xi_2) &= 1 \\ \Phi_1(\xi_1,\xi_2) &= \xi_1 \\ \Phi_2(\xi_1,\xi_2) &= \xi_2 \\ \Phi_3(\xi_1,\xi_2) &= \xi_1^2 - 1 \\ \Phi_4(\xi_1,\xi_2) &= \xi_1\xi_2 \\ \Phi_5(\xi_1,\xi_2) &= \xi_2^2 - 1 \end{aligned} \tag{8}$$

In multiple dimensions the number of coefficients M is determined by the number of dimensions D and the highest polynomial order P:

$$M = \frac{(D+P)!}{D!P!} \tag{9}$$

As in stochastic collocation approach, the stochastic Galerkin method suffers from the curse of dimensionality. Smolyak-type constructions have not been attempted for the polynomial basis used in the spectral expansions but offer a possible remedy.

5 Examples

The examples reported in the following offer indications of the relative computational cost of the uncertainty propagation approaches illustrated before. In these test cases the uncertainty is assumed as part of the problem specification.

5.1 Steady Burgers equation

The first test problem is based on the $1D$ viscous Burgers equations [36]:

$$\frac{1}{2}(1-u)\frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial u^2}$$

where the convective flux is modified so that an exact solution can be obtained (manufactured solution [34]):

$$u(x) = \frac{1}{2}\left[1 + tanh\left(\frac{x}{4\mu}\right)\right]$$

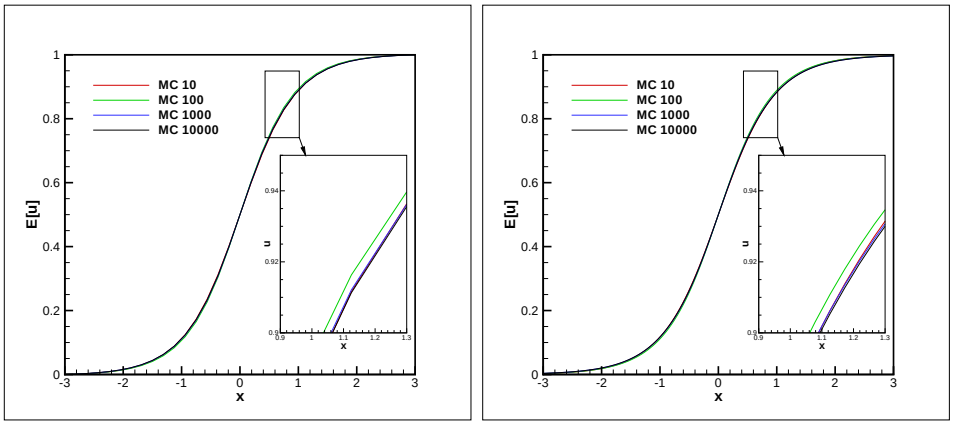

Figure 5: Expectation of the solution for the Burgers steady state problem with uncertain viscosity. Monte Carlo sampling with different number of realization. A numerical solution computed using 32 cells (left) is compared to the exact solution (right).

In this case it is assumed that the viscosity is uncertain, for example as a consequence to an unknown fluid temperature. A Gaussian random variable is considered with mean and variance given by $\mathbb{E}[\mu] = 0.25$ and $var[\mu] = 0.0025$[12]

Calculations have been performed using a finite volume discretization on the domain [-3:3]; boundary conditions are derived from the exact solution. Monte Carlo sampling is initially used as the uncertainty propagation technique. The expectation of the velocity is reported in Fig. 5.1 and has been computed using a different number of samples ranging from 10 to 10,000. The results shows only small dependency on the number of realization; in the same figure the statistical convergence is also evaluated using the exact solution of the deterministic problem. In Fig. 5.1 the solution variance is reported. In this case it is clear that 10 samples are not sufficient and *converged* results are only obtained for 1,000 and 10,000 samples. Note that the variance is zero at $x = 0$ to satisfy the symmetry constraint.

As a second step the stochastic Galerkin method has been implemented and tested. In this case the expansion is based on Hermite polynomials. Computations have been performed using the same discretization used in the deterministic code [36]. Fig. 5.1 shows results obtained using different order of expansions, from 1 to 3; the comparisons to Monte Carlo show that for this simple problem - characterized by small variability - even the low order expansion provide a reasonable representation

[12]It should be noted that a Gaussian distribution for the viscosity implies that there is a non-zero probability of having very low - or even negative - values of the viscosity. In this case the high mean value and the small variance reduce the occurrence of negative viscosity to virtually zero.

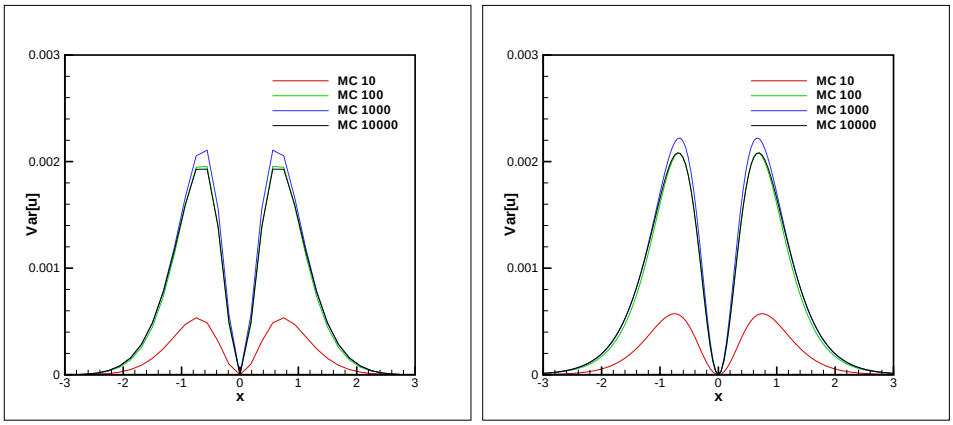

Figure 6: Variance of the solution for the Burgers steady state problem with uncertain viscosity. Monte Carlo sampling with different number of realization. A numerical solution computed using 32 cells (left) is compared to the exct solution (right).

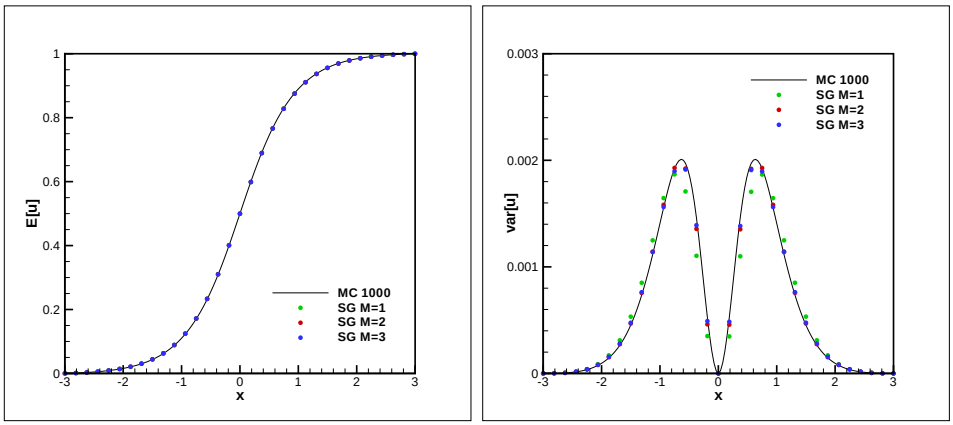

Figure 7: Expectation (left) and variance (right) of the solution for the Burgers steady state problem with uncertain viscosity. The Stochastic Galerkin method is used and the results are compared to Monte Carlo sampling.

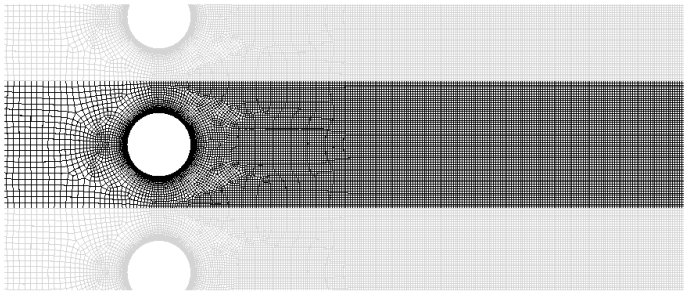

Figure 8: Computational grid for the flow around an array of cylinders with uncertain surface heat flux

of the variance of the solution.

5.2 Heat transfer computations under uncertainty

In this problem the goal is the computation of the wall temperature for an array of cylinders in crossflow under uncertain free stream conditions and surface heat flux. The mathematical framework is based on the two-dimensional Reynolds-averaged Navier-Stokes equations and the computational domain is a channel with a specified inflow, periodic boundary conditions on the top and bottom boundaries, and an outflow. An unstructured grid is used, see Fig. 5.2.

We assume that the sources of uncertainty are the specification of the velocity boundary condition on the incoming flow and the definition of the thermal condition on the surface of the cylinder. The inlet velocity profile is constructed as a linear combination of two cosine functions:

$$U_{inlet}(y) = 1 + 0.25(\xi_1 \cos(2\pi y) + \xi_2 \cos(10\pi y)). \tag{10}$$

where ξ_1 and ξ_2 are two uniform random variables in the interval [-1:1]; this inflow specification ensures that the amplitude of the random fluctuations did not cause the velocity to be negative. The wave numbers 2 and 10 were chosen to give one low and one high frequency fluctuation while maintaining symmetry in the problem. The heat flux is specified as an exponential function of ξ_3 over the cylinder, namely

$$\frac{\partial T}{\partial n}_{cylinder}(x) = e^{-(0.1+0.05\xi_3)(x-0.5)} \tag{11}$$

where n is the normal to the cylinder surface. The heat flux is slightly greater at the left side of the cylinder at the stagnation point; the value of ξ_3 determines precisely how much greater. Again ξ_3 is a uniform random variable in [-1:1].

We employed Monte Carlo sampling and stochastic collocation to solve this problem. In particular, we used both tensor product and sparse grid formalism to

l	# tensor grid points	# sparse grid points
2	27	25
3	125	111
4	729	351
5	2187	1026

Table 1: Number of function evaluation required to achieve a certain degree of polynomial exactness using tensor product and sparse grid formulations. The parameter l is called the *level* and is directly connected to the accuracy of the interpolant on tensor grids; on sparse grids the relationship is more complex.

construct quadrature rules in the three-dimensional parameter space (ξ_1, ξ_2, ξ_3). In Table 1 the number of function evaluations (abscissas) required to achieve a certain order of polynomial exactness are reported.

The expectation and variance of the wall temperature are reported in Fig. 5.2 and 5.2, respectively.

In this application, it is clear that accurate temperature statistics can be obtained efficiently using stochastic collocation with respect to Monte Carlo sampling. The difference between tensor product and sparse grid formulations is very small.

5.3 Shock dominated problems

Uncertainty propagation algorithms based on polynomial basis are particularly effective when the output of interest are smoothly varying with respect to the input uncertainties. The presence of discontinuities in the response surface poses a challenge; in physical terms, these discontinuities represent sharp system transitions such as the occurrence of flow separation or shock interactions.

Consider the problem governed by the non-homogenous Burgers equations [27]:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(\frac{u^2}{2}) = \frac{\partial}{\partial x}(\frac{\sin^2 x}{2}), \qquad 0 \leq x \leq \pi, \quad t > 0$$

with the initial condition $u(x,0) = \beta \sin x$, and boundary conditions $u(0,t) = u(\pi,t) = 0$. β represents the uncertainty in the initial condition and is defined as

$$\beta = \frac{-1 + \sqrt{1 + 4\sigma^2 \xi}}{2\sigma^2 \xi}$$

where ξ is a Gaussian random variable and σ is used to control the amount of uncertainty.

The exact steady solution contains a shock located at $x_s = f(\beta)$ and the desired output is the PDF of the shock location. The response surface u as a function of β is reported in Fig. 12.

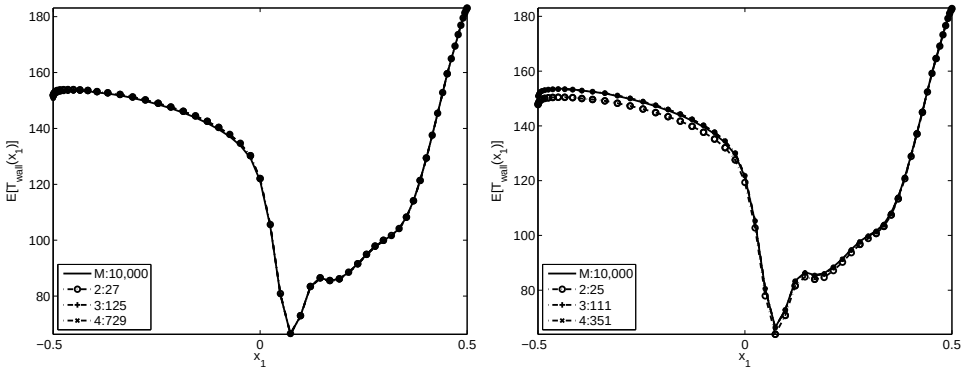

Figure 9: Expectation of the wall temperature. Tensor product (left) versus sparse grids (right). Monte Carlo (MC) statistics are based on 10,000 samples; multiple quadrature formulas are reported for different levels (see Table 1).

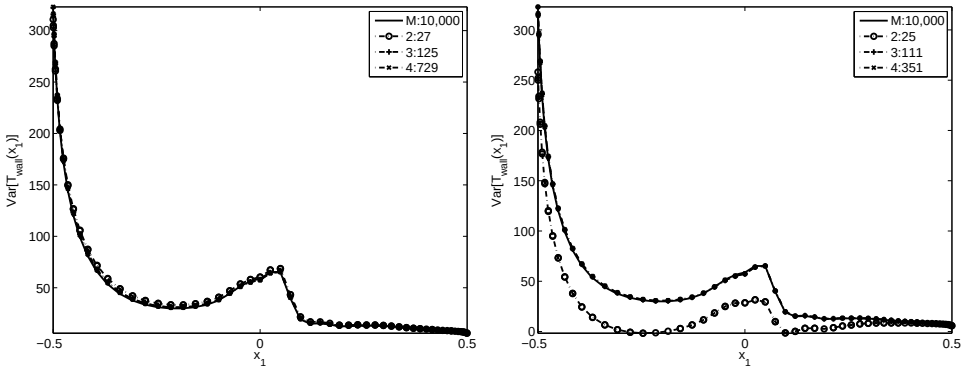

Figure 10: Variance of the wall temperature. Tensor product (left) versus sparse grids (right). Monte Carlo (MC) statistics are based on 10,000 samples; multiple quadrature formulas are reported for different levels (see Table 1).

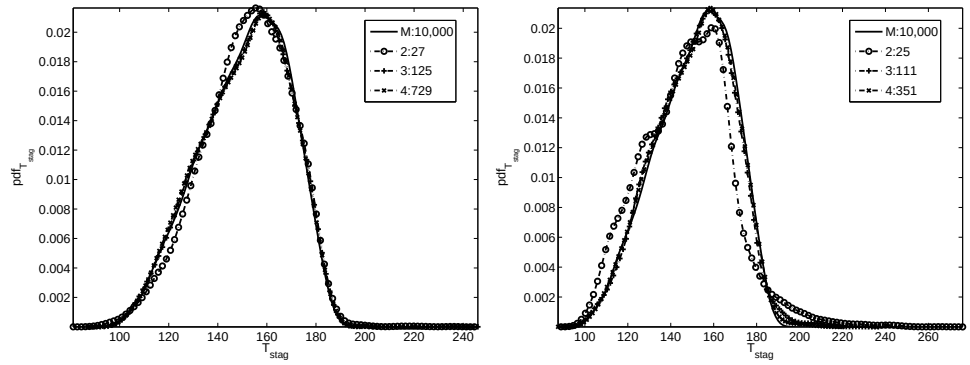

Figure 11: Probability distribution function of the wall temperature at the stagnation point. Tensor product (left) versus sparse grids (right). Monte Carlo (MC) statistics are based on 10,000 samples; multiple quadrature formulas are reported for different levels (see Table 1).

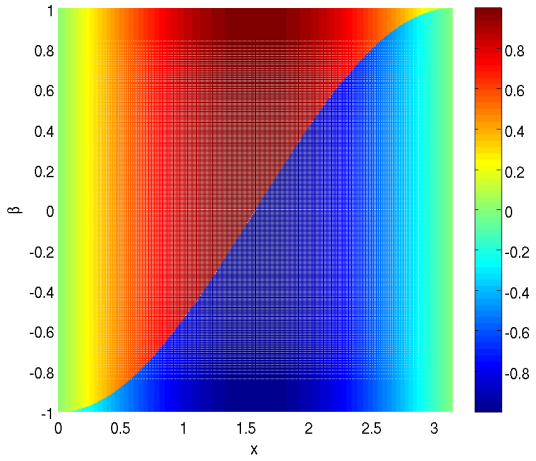

Figure 12: Solution of the in-homogeneous Burgers equations with uncertain initial condition. The sharp transition corresponds to a shock whose location is a function of the input uncertainty.

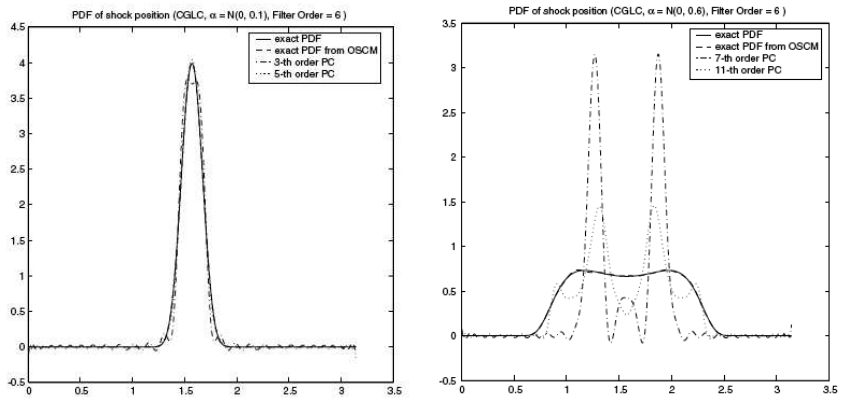

Figure 13: Probability distribution of the shock location for the inhomogeneous Burgers equations with uncertain initial condition. Low , $\sigma = 0.1$ (left) and high input variability, $\sigma = 0.6$ (right). Results obtained using high-order stochastic Galerkin expansions [28]

Hestevan et al. [25] have applied the stochastic Galerkin approach to study this problem; high order expansions in Hermite polynomials were used to compute the PDF of the shock location. The availability of exact PDFs allows to identify the limitation of the numerical approach. In Fig. 5.3 the solution is reported for low and high degree of input variability (σ set to 0.1 and 0.6 respectively). For the low variability case the computed results are in reasonable agreement with the exact PDF, but some oscillations are present. For the high variability case, the oscillations are the dominant feature and the shape of the PDF is completely misrepresented. In this case, the problem is that the solution is discontinuous and the polynomial basis (underlying the stochastic expansion) is highly oscillatory - Gibbs phenomenon.

An alternative response reconstruction has been developed to handle strong non-linear or discontinuous surfaces. The method is based on a Pade-Legendre interpolatory formula [32]; it is a stochastic collocation approach that eliminates the Gibbs phenomenon. Assume N solution evaluations are available; given the integers M and L, the pair of Legendre polynomials P and Q of order less or equal than M and L, respectively, are said to be the solution of the (N, M, L) Pade-Legendre interpolation problem [26] of u if

$$\langle P - Qu, \phi \rangle = 0 \tag{12}$$

and the rational function $R(u) := P/Q$ is defined as the approximation of u. The details of the construction of R are reported in Chantrasmi et al. [32]; here it is sufficient to observe that if u is discontinuous, the polynomial Q can be regarded

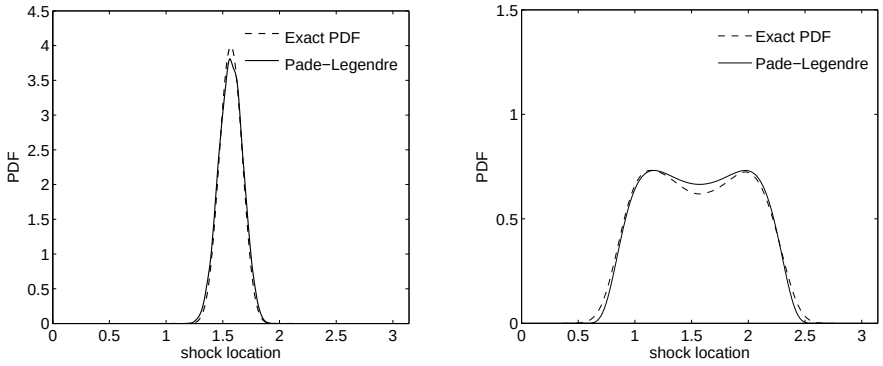

Figure 14: Probability distribution of the shock location for the inhomogeneous Burgers equations with uncertain initial condition. Low , $\sigma = 0.1$ (left) and high input variability, $\sigma = 0.6$ (right). Results obtained using the Pade-Legendre stochastic collocation approach [32]

as a preconditioner (or a filter) such that $u \cdot Q$ is smooth and can be efficiently and accurately interpolated by the polynomial P.

Results obtained using the Pade-Legendre stochastic collocation approach are reported in Fig. 5.3 in terms of the PDF of the shock locations. In this case the accuracy is preserved even for high input variability.

6 Conclusions and outlook

The growing interest in uncertainty quantification has motivated the development of a variety of mathematical approaches. In particular, probabilistic uncertainty propagation methods have received considerable attention.

The choice of the appropriate method to use for a specific application is not obvious. For typical fluid mechanics simulations some common considerations are:

- *expensive function evaluation*: sampling based methods are typically not appropriate because they might require several thousand full computations to build the statistics of the outputs;
- *large number of uncertainties*: boundary conditions, material properties, geometry specification, etc. introduce many independent input parameters that have to be characterize. Methods that suffer from curse of dimensionality quickly become unfeasible;
- *non-linear system responses*: transitions and bifurcations are typical of fluid mechanics, especially for compressible flows. Methods that strictly require a smooth dependency between inputs and outputs can be ineffective.

In spite of these difficulties, probabilistic UQ is naturally amenable to mathematical and numerical analysis and, therefore, is expected to achieve a high level of maturity in the near future.

In addition to the methods presented here several other methods have been applied especially in the field of structural mechanics. It is also worth mentioning that alternative approaches not based on probabilistic reasoning have been proposed and used with some success. It is not generally clear when probabilistic methods fail or are insufficient; the treatment of epistemic uncertainty remains difficult and possibly the greatest challenge in uncertainty quantification.

Acknowledgments

This work is supported by the Department of Energy within the Advanced Simulation and Computing Program. The notes are based on work carried out at Stanford by Paul Constantine and Tonkid Chantrasmi.

References

[1] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids, Adv. Comput. Math. 12 (2000) 273-288.

[2] Bowman, A. W., A. Azzalini, Applied Smoothing Techniques for Data Analysis, Oxford University Press, 1997.

[3] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, J. Comput. Phys. (2007), doi:10.1017/j.jcp.2006.12.014

[4] W. Gautschi, Algorithm 726: ORTHOPOL-A package of routines for generating orthogonal polynomials and Gauss-type quadrature rules, ACM Trans. Math. Software20, (1994) 2162.

[5] R. Ghanem and A. Doostan, On the construction and analysis of stochastic models: characterization and propagation of the errors associated with limited data, J. of Comput. Phys. 217 (2006) 6381.

[6] E. Isaacson, H.B. Keller, Analysis of Numerical Methods, Dover, 1994.

[7] A. S. Kronrod, Nodes and Weights of Quadrature Formulas. Consultants Bureau, New York, 1965

[8] M. Loeve, Probability Theory, 4th ed., Springer-Verlag, New York, 1997.

[9] F. Nobile, R. Tempone and C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. Report number: TRITA-NA 2007:07. To appear SINUM.

[10] E. Novak, K. Ritter, Simple cubature formulas with high polynomial exactness, Constructive Approx., 15 (1999) 499-522.

[11] E. Novak, K. Ritter, High dimensional integration of smooth functions over cubes, Numer. Math. 75 (1996) 79-97.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipes in C. Cambridge University Press, 2nd ed., 1992.

[13] J.A. Rice, Mathematical Statistics and Data Analysis, 2nd ed., Duxbury Press, 1995.

[14] B.D. Ripley, Stochastic Simulation, 1st ed., Wiley, 1987.

[15] S.A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, Soviet math. Dokl., 4 (1963) 240-243.

[16] D.M. Tartakovsky, D. Xiu, Stochastic analysis of transport in tubes with rough walls, J. Comput. Phys. 217 (2006) 248-259.

[17] L.N. Trefethen, Is Gauss quandrature better than Clenshaw-Curtis? SIAM Review, to appear, 2007.

[18] J.H. VanLint, R.M. Wilson, A Course in Combinatorics, 2nd ed., Cambridge University Press, 2001.

[19] D. Xiu, J.S. Hesthaven, High order collocation methods for the differential equations with random inputs, SIAM J. Sci. Comput. 27 (2005) 1118-1139.

[20] G.W. Wasilkowski and H. Wozniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, J. of Complexity, 11:1-56 1995.

[21] D. Xiu and G. E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.

[22] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.

[23] N. Wiener, *The homogeneous chaos*, Amer. J. Math., 60 (1938), pp. 897–936.

[24] Szegő Orthogonal Polynomials, 4th ed. Providence, RI: Amer. Math. Soc., 1975.

[25] J. S. Hesthaven , S. M. Kaber , L. Lurati, Pad-Legendre Interpolants for Gibbs Reconstruction, Journal of Scientific Computing, v.28 n.2-3, pp. 337-359, September 2006

[26] G. A. Baker, Pade Approximants, Cambridge Univ Pr.

[27] M.D. Salas, S. Abarbanel and D. Gottlieb, Multiple steady states for characteristic initial value problems, Applied Numerical Mathematics 2 (1986), pp. 193-210.

[28] Q.-Y. Chen, D. Gottlieb, J. S. Hesthaven, Uncertainty analysis for the steady-state flows in a dual throat nozzle, Journal of Computational Physics, v.204 (1), pp. 378-398, March 2005

[29] J.C. Helton and F.J. Davis, Sampling-based Methods, in Sensitivity Analysis, A. Saltelli, C. Chan, and E.M. Scott, Eds. New York: Wiley, 2000, pp. 101-153.

[30] J.C. Helton and F.J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliability Engineering and System Safety, vol. 81, pp. 23-69, 2003.

[31] Z. Sandor and P. Andras, "Alternative Sampling Methods for Estimating Multivariate Normal Probabilities", Econometric Institute Report EI 2003,05

[32] T. Chantrasmi, A. Doostan and G. Iaccarino, Pade-Legendre Approximants for Uncertainty Analysis with Discontinuous Response Surfaces, submitted to JCP, 2008

[33] W.L. Oberkampf and M. F. Barone, Measures of agreement between computation and experiment: validation metrics. Journal of Computational Physics, V. 217, 2006, pp. 5–36.

[34] Knupp, P., and Salari, K. (2003), Verification of Computer Codes in Computational Science and Engineering, Chapman and Hall/CRC, Boca Raton, FL.

[35] AIAA (1998), "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations," American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA.

[36] R. W. Walters and L. Huyse, Uncertainty Analysis for Fluid Mechanics with Applications, NASA/CR 2002-211449.

[37] S. Fersin, R. B. Nelsen, et al. Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis, SAND Report 2004-3072.

[38] Oberkampf, W.L. and J. C. Helton. (2002). Investigation of evidence theory for engineering applications. AIAA Non-Deterministic Approaches Forum, April 2002, Denver, Colorado, paper 2002-1569.