IDENTIFICATION OF PROBABILISTIC MODELS FROM DATA

ROGER GHANEM
USC, LOS ANGELES

There has been a recent surge of interest within computational science and engineering, in the numerical resolution of physical problems described by probabilistic models. This generates a pressing need for the construction of probabilistic models that reflect, with controllable fidelity, the weight of evidence present in the experimental data.

In these two talks I will describe the mathematical structure of this problem and its associated challenges. I will also describe recent efforts aimed at constructing models that simultaneously lend themselves to numerical resolution using function approximation techniques (a-la Polynomial Chaos), and to model validation.

STOCHASTIC PROJECTION METHODS FOR UNCERTAIN FLOW MODELS

OLIVIER LE MÂITRE
(LIMSI, PARIS)

In this lecture, I will review the Stochastic projection methods for uncertainty propagation in numerical models involving stochastic coefficients.

I will first introduce the stochastic approximation bases (Polynomial chaos and generalizations) before proceeding with the Galerkin projection of the model equations. The derivation of the Galerkin problem will be detailed for the Eulerian formulation of the incompressible Navier-Stokes equations. Efficient solution techniques decoupling the stochastic mode resolution will then be discussed.

In the second part of the lecture, the extension of the Galerkin projection method to low Mach number flows and Lagrangian formulations (particle methods) will be shortly reviewed.

REDUCED BASIS APPROXIMATIONS FOR UNCERTAINTY PROPAGATION

OLIVIER LE MÂITRE
(LIMSI, PARIS)

In the second lecture, I will discuss the so-called Generalized Spectral Decomposition (GSD) method for the approximation of solutions to models involving stochastic coefficients.

Contrary to the Galerkin projection method where the stochastic basis is selected a priori, GSD aims at constructing iteratively low dimensional approximation spaces yielding a reduced solution that minimizes (in some sense) the stochastic equation residual.

Different algorithms for the construction of the reduced approximation space will be discussed. Applications to both linear and nonlinear models will be detailed and the efficiency of the GSD algorithms will be contrasted to support a discussion on the advantages and limits of the method.

SPARSE ADAPTIVE TENSOR FEM FOR OPERATOR EQUATIONS WITH STOCHASTIC DATA

CHRIS SCHWAB
(ETH ZÜRICH)

Let $A: V \to V'$ be a linear, strongly elliptic operator on a d-dimensional manifold D (polyhedra or boundaries of polyhedra are also allowed).

An operator equation $Au = f$ with stochastic data f is considered. The goal of the computation is the mean field and k-point correlation functions $\mathcal{M}^1 u \in V$, $\mathcal{M}^2 u \in V \otimes V$, $\dots$, $\mathcal{M}^k u \in V \otimes \cdots \otimes V$ of the random solution u.

We discretize the mean field problem using an h-Version FEM in D with hierarchical basis and N degrees of freedom. We analyze a Monte-Carlo (MC) algorithm and a deterministic algorithm for the approximation of the moment $\mathcal{M}^k u$ for $k \geq 1$.

Both algorithms are based on a "sparse tensor product" of multilevel Finite Element spaces in D with N degrees of freedom. They allow the approximation of the kth moment $\mathcal{M}^k u$ with $O\left(N(\log N)^{k-1}\right)$ degrees of freedom, instead of N^k degrees of freedom for the full tensor product FE space.

Algorithm 1, a sparse tensor MC-FEM with M samples (i.e., deterministic solves) is proved to yield approximations to $\mathcal{M}^k u$ with work of $O(MN(\log N)^{k-1})$ operations. The solutions are shown to converge with the optimal rates with respect to the number of Finite Element degrees of freedom N and the number M of samples.

Algorithm 2, the deterministic FEM, is based on deterministic, hypoelliptic equations for the kth moment $\mathcal{M}^k u$ of the random solution u in $D^k \subset \mathbb{R}^{kd}$. We analyze their discretization using sparse tensor products of the hierarchic FE spaces in D.

For nonlocal operators A wavelet compression is employed in both algorithms. The linear systems are solved iteratively with multilevel preconditioning. We prove that this yields an approximation for $\mathcal{M}^k u$ with at most $O\left(N(\log N)^{k+1}\right)$ operations.

For nonlinear problems, we employ a linearization to derive deterministic equations for the statistics of the random solution in terms of the statistics of the input data. An example for elliptic problems in domains with small amplitude stochastic boundary perturbation based on a second order shape calculus and a boundary integral reformulation of the hypoelliptic equation for the second moment illustrates this approach.

Joint work with Alexej Chernov (ETH Zurich), Helmut Harbrecht (Bonn), R. Schneider (TU Berlin), Rob Stevenson (Amsterdam) and T. von Petersdorff (College Park and ETH Zurich).

References

[TS06] Ch. Schwab and R.A. Todor, Karhúnen-Loève Approximation of Random Fields by Generalized Fast Multipole Methods. Journal of Computational Physics **217** (2006), 100-122.

[TS] R.A. Todor and Ch. Schwab, Convergence Rates of Sparse Chaos Approximations of Elliptic Problems with stochastic coefficients. IMA Journ. Numer. Anal. (2007).

[BS] M. Bieri and Ch. Schwab, Sparse high order FEM for elliptic sPDEs. Comp. Meth. Appl. Mech. Engg. **198** (2009) 1149-1170

[ABS] M. Bieri, R. Andreev and Ch. Schwab, Sparse Tensor Discretization of Elliptic sPDEs (to appear in SISC 2010).

CONVERGENCE RATES OF STOCHASTIC GALERKIN FEM FOR ELLIPTIC SPDES

CHRIS SCHWAB
(ETH ZÜRICH)

We consider the Finite Element Solution of second order elliptic problems in a physical domain $D \subset \mathbb{R}^d$ with spatially inhomogeneous random coefficients.

We present convergence rates and complexity estimates for sparse Galerkin semidiscretization in the probability domain of the random solution. It is parametric in the first M Karhúnen-Loève (KL) variables of the input data [TS06].

Two cases are distinguished:

(i) Exponential decay of the input's KL expansion based on [TS]

and

(ii) algebraic decay of the input's KL expansion.

In (i), a "polynomial chaos" type Galerkin discretization is shown to yield spectral convergence rates *in terms of* N_Ω, the number of deterministic elliptic problems to be solved. In (ii), first approximation rates in terms of N_Ω are available.

Finally, in

(iii) ongoing work [BS, B, CDS09, G] on the total complexity vs. accuracy of (adaptive) tensor Galerkin discretizations in both, stochastic as well as in the deterministic domain D will be addressed.

Sufficient conditions on the joint pdf's of the random field input to ensure better complexity than with (Quasi) Monte Carlo in the probability domain and with Galerkin discretization in D will be identified and implementational issues will be addressed in each case.

References

[CDS09] A. Cohen, R. DeVore and Ch. Schwab, Convergence Rates of best N-term approximations for a class of elliptic sPDEs (in review 2009).

[TS06] Ch. Schwab and R.A. Todor, Karhúnen-Loève Approximation of Random Fields by Generalized Fast Multipole Methods. Journal of Computational Physics **217** (2006), 100-122.

[TS] R.A. Todor and Ch. Schwab, Convergence Rates of Sparse Chaos Approximations of Elliptic Problems with stochastic coefficients. IMA Journ. Numer. Anal. (2007).

[BS] M. Bieri and Ch. Schwab, Sparse high order FEM for elliptic sPDEs. Comp. Meth. Appl. Mech. Engg. **198** (2009) 1149-1170

[B] M. Bieri, A sparse, composite collocation method for elliptic stochastic PDEs, Report 2009-08, Seminar for Applied Mathematics, ETH Zürich.

[G] C.J. Gittelson, (to appear in M3AS 2010).

[ABS] M. Bieri, R. Andreev and Ch. Schwab, Sparse Tensor Discretization of Elliptic sPDEs (to appear in SISC 2010).

STOCHASTIC COLLOCATION FOR PDES WITH RANDOM INPUT DATA

RAÚL TEMPONE
(KAUST, SAUDI ARABIA)

We consider the problem of numerically approximating statistical moments of the solution of a linear elliptic or parabolic partial differential equation (PDE), whose coefficients and/or forcing terms are spatially correlated random fields. The stochastic coefficients of the PDE are approximated for instance by truncated Karhunen-Loève expansions driven by a finite number of uncorrelated random variables. After approximating the stochastic coefficients the original stochastic PDE turns into a new deterministic parametric PDE of the same type, the dimension of the parameter set being equal to the number of random variables introduced.

Space discretization is achieved via finite elements. Thanks to the fact that the solution to the PDE typically features analytic regularity with respect to the parameters, we consider global polynomial approximations in the probability space.

In this talk, we review Collocation type polynomial approximations of the stochastic PDE in the probability space based either on full or sparse tensor product polynomial spaces. In particular, the Collocation technique requires to solve a parametric problem for input parameter values in a sparse grids of Gauss points and thus it naturally leads to the solution of uncoupled deterministic problems as in the Monte Carlo approach.

We will focus on anisotropic polynomial approximations, which are particularly attractive in the case of input data obtained as truncated expansions of random fields, since the anisotropy can be tuned on the decay properties of the expansion. We will present a priori and a posteriori procedures to choose the anisotropy of the polynomial space that are extremely effective in some situations. Convergence studies show that for certain problems the anisotropic sparse collocation approximation actually breaks the curse of dimensionality.

Numerical examples illustrate the theoretical results and are used to compare the different approaches with the more traditional Monte Carlo technique.

References

[1] I. Babuška, F. Nobile, and R. Tempone, "A stochastic collocation method for elliptic partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 45, no. 3, pp. 1005–1034, 2007.

[2] F. Nobile, R. Tempone, and C. Webster, "A sparse grid stochastic collocation method for partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 46, no. 5, pp. 2309–2345, 2008.

[3] F. Nobile, R. Tempone, and C. Webster, "An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 46, no. 5, pp. 2411–2442, 2008.

[4] F. Nobile and R. Tempone, "Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients," *International Journal for Numerical Methods in Engineering*, vol. 80/6-7, pp. 979–1006, 2009.

[5] J. Bäck, F. Nobile, L. Tamellini and R. Tempone, "Stochastic Spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison". Accepted for publication in ICOSAHOM proceedings, 2010.

UNCERTAINTY ANALYSIS FOR COMPLEX SYSTEMS: ALGORITHMS AND DATA

DONGBIN XIU
(PURDUE UNIVERSITY)

The field of uncertainty quantification has received increasing amount of attention recently. Extensive research efforts have been devoted to it and many novel numerical techniques have been developed. These techniques aim to conduct stochastic simulations for large-scale complex systems.

In this talk we will review one of the most widely approaches – generalized polynomial chaos (gPC). The gPC based methods employ orthogonal polynomials in random space and take advantage of the solution smoothness (whenever possible).

The features of various gPC numerical schemes will be reviewed. Furthermore, we will discuss how real observational data can be utilized and combined with stochastic simulations. The resulting data-driven uncertainty analysis can provide much more insight to the true physics and produce predictions with high fidelity.

A DIMENSION-REDUCTION METHOD FOR STOCHASTIC PDES

NICHOLAS ZABARAS
CORNELL UNIVERSITY

We utilize High Dimensional Model Representation (HDMR) technique in the stochastic space to represent the model output as a finite hierarchical correlated function expansion in terms of the stochastic inputs starting from lower-order to higher-order component functions. HDMR is efficient at capturing the high-dimensional input-output relationship such that the behavior for many physical systems can be modeled only by the first few lower-order terms.

An adaptive version of HDMR is developed to automatically detect the important stochastic dimensions and construct higher-order terms only as a function of the important dimensions. In this work, we integrate the adaptive sparse grid collocation (ASGC) method with HDMR to solve the resulting sub-problems. This results in a low-dimensional stochastic reduced-order model of the high-dimensional stochastic problem.

A number of examples will be shown including stochastic multiscale modeling of flow through random heterogeneous media. The cases examined show that the method provides accurate results for stochastic dimensionality as high as 500 even with large input variability.

DATA-DRIVEN MODEL REDUCTION OF STOCHASTIC INPUT MODELS WITH APPLICATIONS TO MULTISCALE MATERIALS MODELING

NICHOLAS ZABARAS
CORNELL UNIVERSITY

We will discuss model reduction techniques for data-driven stochastic input models. They include multidimensional scaling, kernel PCA and a bio-orthogonal KLE expansion for capturing variability of microstructure topology in the continuum.

Particular developments will be shown in deriving reduced-order stochastic input models to represent the uncertainty in heterogeneous media, polycrystalline materials and geological media. These models are essential for developing computationally efficient multiscale models of physical processes in random media (thermal and hydrodynamic transport, deformation, etc.).

Several examples will be discussed to demonstrate the importance of such developments in the analysis and design in the presence of uncertainty of multiscale engineering systems and processes.

REPRESENTATION OF GAUSSIAN FIELDS IN SERIES WITH INDEPENDENT COEFFICIENTS

CLAUDE J. GITTELSON
ETH ZÜRICH

The numerical discretization of problems with stochastic data or stochastic parameters generally involves the introduction of coordinates that describe the stochastic behavior, such as coefficients in a series expansion or values at discrete points. The series expansion of a Gaussian field with respect to any orthonormal basis of its Cameron–Martin space has independent standard normal coefficients.

A standard choice for numerical simulations is the Karhunen–Loève series, which is based on eigenfunctions of the covariance operator. We suggest an alternative basis that can be constructed directly from the covariance kernel. The resulting basis functions are often well localized, and the convergence of the series expansion seems to be comparable to that of the Karhunen–Loève series.

We provide explicit formulas for particular cases, and general numerical methods for computing exact representations or approximations of the basis functions. Finally, we relate our approach to numerical discretizations based on replacing a random field by its values on a finite set.

This research is supported in part by the Swiss National Science Foundation under grant No. 200021-120290/1.

INTEGRATION OF FICTITIOUS DOMAIN AND CHAOS COLLOCATION METHODS FOR THE ANALYSIS OF GEOMETRIC UNCERTAINTIES IN FLUID DYNAMICS

LUCIA PARUSSINI
UNIVERSITY OF TRIESTE

In this talk an approach is presented for the analysis of the effects of geo- metric tolerances in fluid dynamic behaviour of manufactured components. In the engineering design phase of a component tolerance specifications are provided and, since geometric tolerances can influence the performance of the component, an analysis on the way this affects its behavior should be per- formed. Moreover the sensitivity of the performances respect to the geometric uncertainties should be investigated. Therefore there is a great interest in developing a methodology to face differential problems where the geometrical domain is treated as a stochastic phenomenon.

Here the Tensorial-expanded Chaos Collocation method coupled to Fictitious Domain Method is used to solve Fluid Dynamic problems with geometric uncertainties.

The main advantage of the Tensorial-expanded Chaos Collocation method is its non-intrusive formulation, so existing deterministic solvers can be used. The Least-Squares Spectral Element Method has been employed for the analysis of the deterministic differential problems obtained by Tensorial-expanded Chaos Collocation. This algorithm exploits a Fictitious Domain approach, so it is particularly suitable to solve differential problems defined on stochastic domains. The capabilities of the Tensorial-expanded Chaos Collocation method com- bined to the Fictitious Domain-Least-Squares Spectral Element Method are demonstrated by a numerical experiment.

COMPUTER COMMUTATIVE ALGEBRA AND POLYNOMIALS IN NORMAL VARIABLES

GIOVANNI PISTONE
POLITECNICO DI TORINO

Computational methods based on polynomial algebra software such as CoCoA, http://cocoa.dima.unige.it/, have been used in Statistics for Design of Experiments and for various problems in statistical modeling. A recent overview of this new field, termed Algebraic Statistics, is in Gibilisco, Riccomagno, Rogantin, Wynn (eds), Algebraic and Geometric Methods in Statistics, CUP 2010.

In Design of Experiments a set of trial points is described as the solution of a system of polynomial equations and the identification of various classes of models is computed by the use of special bases of the ideal generated. Here we present the first results of a research in progress in which we explore the applicability of these ideas when the defining equations are derived from Hermite polynomials, e.g. the system is $H_3(x) = 0$, $H_3(y) = 0$,$H_2(x)-H_2(y) = 0$.

A NUMERICAL COMPARISON BETWEEN STOCHASTIC GALERKIN AND COLLOCATION TECHNIQUES FOR ELLIPTIC EQUATIONS WITH UNIFORM AND LOGNORMAL RANDOM VARIABLES

LORENZO TAMELLINI
POLITECNICO DI MILANO

The topic of this talk is a comparison between Stochastic Galerkin and Stochatic Collocation methods for the solution of linear elliptic equations with random coefficients. We will recall some results from our previous work, where the random coefficients were modeled as uniform variables, and extend these results to the case of lognormal variables.

In particular, we will show how to set the Stochastic Galerkin and Collocation methods in the same polynomial space, and how to choose these spaces depending on the underlying random variables to have convergence in a probability space.

The next step for a fair comparison between the two methods is a computational cost definition in terms of the number of calls to the deterministic solver, so that the comparison is independent of the actual implementation of the methods.

In this setting we are able to compare the two methods in terms of accuracy of the computed solution versus the computational cost. We will first recall the results for the case with uniform variables and then move to a case with lognormal variables, introducing a suitable multivariate polynomial space that relies on the convergence properties of the Hermite polynomial expansion. We will highlight the different performances of each method between the test with uniform random variables and the test with lognormal random variables. The focus will also be pointed to some "numerical" issues, like sparsity pattern and preconditiong of matrices for Stohastic Galerkin.

A GALERKIN METHOD FOR UNCERTAIN HYPERBOLIC SYSTEMS: ROE SOLVER AND ENTROPY CORRECTOR

JULIE TRYOEN
LIMSI PARIS

We consider hyperbolic systems with uncertainties in input quantities (e.g. model parameters or initial and boundary conditions) which are parametrized by independent random variables with known distribution functions. Because of uncertainty, the solution can present discontinuities in the spatial as well as in the stochastic domains. To prevent the emergence of Gibbs phenomena and aliasing errors associated with spectral representations using smooth stochastic functionals, we rely on piecewise continuous approximations at the stochastic level. The stochastic solution is then sought by means of a stochastic Galerkin projection procedure.

A finite volume scheme with a Roe solver is used for the discretization in physical space and time of the resulting Galerkin system. This is motivated by the mathematical analysis of the Galerkin system. Approximate eigenvalues and vectors for the Galerkin system Jacobian are proposed and used to derive upwind matrices. Finally, an entropy corrector is presented to avoid entropy-violating shocks in the vicinity of sonic points.

The effectiveness of the proposed method is assessed on nonlinear hyperbolic problems (Burger and Euler equations) with uncertain initial conditions and physical properties. Simulation results are contrasted with reference solutions computed by means of Monte Carlo sampling.

STOCHASTIC BIFURCATION ANALYSIS OF RAYLEIGH-BÉNARD CONVECTION

DANIELE VANTURI
UNIVERSITY OF BOLOGNA

Stochastic bifurcations and stability of natural convection within two-dimensional square enclosures are investigated by different stochastic modelling approaches.

Deterministic stability analysis is carried out first to obtain steady state solutions and primary bifurcations. It is found that multiple stable steady states coexist, in agreement with recent results, within specific ranges of Rayleigh number.

Stochastic simulations are then conducted around bifurcation points and transitional regimes. The influence of random initial flow states on the development of supercritical convection patterns is also investigated. It is found that a multi-element polynomial chaos method captures accurately the onset of convective instability as well as multiple convection patterns corresponding to random initial flow states.

GEOSTATISTICAL INVERSION OF MOMENT EQUATIONS OF GROUNDWATER FLOW

MONICA RIVA
POLITECNICO DI MILANO

Predictions of groundwater flow in porous media are typically affected by different sources of uncertainty: (i) conceptual uncertainties (including model uncertainties and incomplete knowledge of governing equations); (ii) measurement uncertainties and (iii) parameter uncertainties. These, together with uncertainties in forcing terms, are conveniently tackled upon casting the governing equations in a stochastic framework. In this context, different methods have been developed to condition hydrogeological models not only on direct measurements of parameters but also on measurements of state variables. Linearized stochastic inverse solutions based on cokriging were developed for steady state flow. These methods yield reliable parameter estimates for moderate variability but relatively poor estimates and unduly small estimation variances when variability or non-linearity is pronounced. An alternative methodology is offered by Monte Carlo. These require the generation of a (potentially) large set of random inverse solutions that honor measurements. This may consume a large amount of computer time. Applying them to only a few realizations, as has been the practice to date, may yield plausible representations of reality which however are random and therefore nonunique.

Recently, we have formulated a nonlinear methodology for the inversion of steady-state and transient (ensemble) Moment Equations of groundwater flow. Log-conductivity, Y, is parameterized geostatistically from measured values at discrete locations and unknown values at discrete "pilot points". Whereas prior values of Y at pilot points are obtained by a variant of kriging, posterior estimates at pilot points are obtained through a maximum likelihood fit of computed to measured heads. Optionally, the maximum likelihood function may include a regularization term reflecting prior information on Y, i.e., prior measurements/estimates of parameters and associated errors/uncertainties. The approach provides predictions of hydraulic head and flux through their conditional first moments. Variances of hydraulic head (and flux) are then calculated a posteriori upon solving the corresponding equations. We have compared the relative performance of this Moment Equations-based inverse method and several types of Monte Carlo and semi-analytical inverse methodologies. We have explored, on the basis of a synthetic example, the influence of (i) the order of approximation (zero- or second-order) of the governing mean flow equation and (ii) the number of pilot points on our ability to properly reconstruct the log-conductivity field and to identify the statistical parameters of the underlying variogram of Y, the plausibility weight and the uncertainty associated with the available measurements. Estimation of statistical parameters characterizing the geostatistical model defining the spatial variability of the system is performed on the basis of formal model information/discrimination criteria. Finally, the methodology has been applied on a field case located in the Neckar River Valley near Tbingen, Germany.