

# Solving Stochastic Inverse Problems: A Sparse Grid Collocation Approach

**N. Zabararas**

**Cornell University**

## **0.1 Introduction**

In recent years there has been significant progress in quantifying and modeling the effect of input uncertainties in the response of PDEs using non-statistical methods. The presence of uncertainties is incorporated by transforming the PDEs representing the system into a set of stochastic PDEs (SPDEs). The spectral representation of the stochastic space resulted in the development of the Generalized Polynomial Chaos Expansion (GPCE) methods (Ghanem 1991; Ghanem and Spanos 1991; Ghanem and Sharkar 2002; Xiu and Karniadakis 2002, 2003). This approach requires extensive revamping of a deterministic simulator to convert it into its stochastic counterpart and has issues with scalability when applied to large scale problems. To solve large-scale problems involving high-dimensional stochastic spaces (in a scalable way) and to allow non-smooth variations of the solution in the random space, there have been recent efforts to couple the fast convergence of the Galerkin methods with the decoupled nature of Monte-Carlo sampling (Velamuri Asokan and Zabararas 2005; Babuska et al. 2005a,b). The Smolyak algorithm has been used recently to build sparse grid interpolants in high-dimensional space (Nobile et al. 2008; Xiu

*Computational Methods for Large-Scale Inverse Problems and Quantification of Uncertainty*. Edited by People on Earth

© 2001 John Wiley & Sons, Ltd

---

*This is a Book Title* **Name of the Author/Editor**

© XXXX John Wiley & Sons, Ltd

and Hesthaven 2005; Xiu 2007). Using this method, interpolation schemes for the stochastic solution can be constructed with orders of magnitude reduction in the number of sampled points. The sparse grid collocation strategy provides a seamless way to scalably incorporate the effects of multiple sources of uncertainty in an embarrassingly parallel way. It only utilizes solutions of deterministic problems to construct the stochastic solution. This allows the use of deterministic legacy codes in a stochastic setting. Furthermore, reduction in computational effort can result from adaptively constructing the representation based on the local behavior of the stochastic solution (Ganapathysubramanian and Zabaras 2007; Ma and Zabaras 2009). The mathematical framework to solve SPDEs is a thriving, mature field with a few more issues that have to be resolved (Ma and Zabaras 2009; Wan and Karniadakis 2005).

These developments in direct stochastic analysis raise the possibility of solving the corresponding stochastic inverse problem: the problem of designing/estimating the evolution of a system in the presence of multiple sources of uncertainty given limited information. Specifically, *the development of a non-statistical framework that only utilizes deterministic simulators for the inverse analysis and/or design of complex systems in the presence of multiple sources of uncertainties*. We are interested in two classes of stochastic inverse problems in one mathematical setting.

The first class of problems is *stochastic inverse/estimation problems*. Consider a complex system that is inherently stochastic due to uncertainties in boundary conditions or property distribution. The problem is to reconstruct these *stochastic* boundary conditions, source terms, and/or property variations given statistical sensor data (in the form of moments or PDFs) of the dependent variables. Posing the problem in terms of available statistical sensor data encompasses most traditional scenarios: (i) when the sensor data is from a single experiment, where the stochasticity creeps in due to measurement errors/uncertainties, and (ii) when the sensor data is from statistics of multiple experiments, where the stochasticity creeps in due to uncertainties in both measurement and parameters (operating conditions, property variability).

The second class of problems is *stochastic design problems*. The motivation for posing these problems is to design operating conditions which ensure that a system exhibits a specific desired response even when the property variation in the system is uncertain. The goal is to construct the optimal stochastic input (since the input also cannot be imposed with infinite certainty) such that a system response is achieved in a probabilistic sense in the presence of other sources of uncertainty. Based on our recent work in constructing realistic input models of topological and material/microstructure variability using limited and gappy data (Ganapathysubramanian and Zabaras 2008, 2007), the proposed framework has wide ranging applicability. Fig. 1 shows two applications where the underlying geometric, topological and material uncertainties have to be accounted for during the design of the optimal operating conditions.

Deterministic inverse techniques (based on exact matching or least-squares optimization) lead to point estimates of unknowns without rigorously considering system uncertainties and without providing quantification of the uncertainty in the inverse

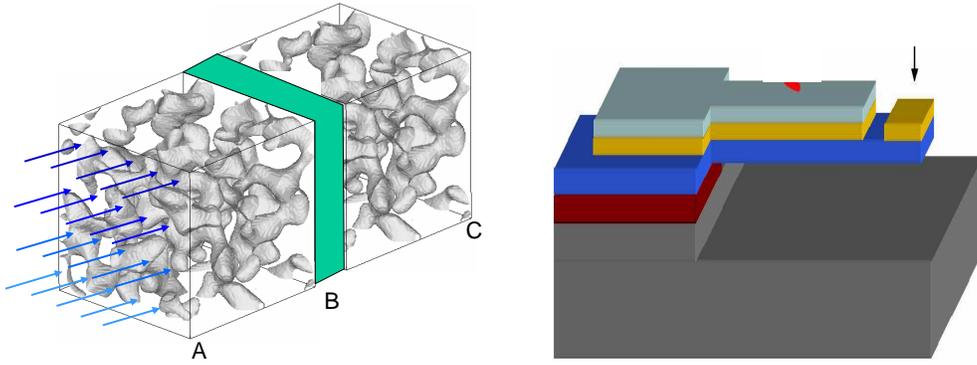


Figure 1 *Two proof-of-concept design applications with topological, material and operating uncertainties. Left: A micro-scale heat-sink. The exact material distribution in the device is unknown (random heterogeneous material). However, we want to maintain a specific temperature profile in region ‘B’ where also the heat flux is zero. The stochastic design problem is to find the optimal heat flux at ‘A’ such that this condition is satisfied in the presence of material uncertainty. Right: The operating conditions (potential difference) across the MEMS cantilever-switch has to be designed in the presence of geometric and material uncertainties.*

or design solution (Sampath and Zabarar 2001). It is clearly necessary to incorporate the effects of uncertainty in property variation and operating conditions for the design of components and devices that perform satisfactorily in a variety of conditions. Beyond the obvious technological significance of the proposed framework, there are several mathematical challenges involved. New mathematical metrics and strategies have to be developed to pose and solve stochastic inverse and design problems. The high stochastic dimensionality of the direct stochastic problems necessitates the development of computationally tractable yet accurate optimization strategies with low storage overhead. One also needs to investigate the possibility of stochastic model reduction to alleviate these issues.

The framework discussed here is a non-statistical framework for stochastic design and estimation. This is in contrast to the Bayesian inference approach that provides a complete probabilistic description of the unknowns and uncertainties given measurement data (Kaipio and Somersalo 2005; Somersalo and Calvetti 2007; Wang and Zabarar 2005a, 2004, 2005b, 2006). The Bayesian approach incorporates the known information regarding the unknown parameters and system uncertainties into a prior distribution model that is then combined with the likelihood to formulate the posterior probability density function (PPDF). There are a few issues with Bayesian inference that have motivated the development of alternate means of solving stochastic inverse/design problems. The higher-order statistics of the unknown are highly-dependent on the form of the prior distribution chosen. Furthermore, the inputs to this framework are a finite number of deterministic sensor measurements, i.e. this

methodology provides no means of incorporating statistics and/or PDF of the measured quantities. Furthermore, this framework has not been applied to design problems with multiple sources of uncertainty. In addition, a non-statistical framework allows one the flexibility of working with different representation of uncertain fields and naturally incorporating the effects of correlation statistics. Furthermore, the proposed mathematical framework provides quantitative measures of convergence and the ability to selectively refine the stochastic solution.

With the direct stochastic solution of large-scale complex systems having reached a level of maturity, this chapter turns to the problem of utilizing these strategies in the control and design of such systems in the presence of uncertainties. Table 1 below summarizes the main features of this work. We will highlight a stochastic optimization framework where the uncertainty representation is based on a sparse grid collocation approach. Utilizing a sparse grid collocation strategy guarantees scalability, the ability to seamlessly incorporate multiple sources of uncertainty and more importantly relies entirely on *multiple calls to deterministic simulators* (Ma and Zabarar 2009). Issues with regularization and the mathematics of posing and solving the sensitivity equations will be considered. Using a sparse grid representation of the design variable, we propose to convert the stochastic optimization problem into a deterministic optimization problem in a higher dimensional space (Zabarar and Ganapathysubramanian 2008). This deterministic optimization problem will subsequently be solved using gradient based optimization. A stochastic sensitivity computation method to compute the sensitivity of the dependent stochastic variables with respect to the estimated/design stochastic variables has to be formulated. This formulation arises naturally from posing the problem in the sparse grid framework and involves multiple calls to deterministic sensitivity problems. Stochastic model reduction strategies and efficient stochastic Hessian calculation strategies will accelerate the optimization framework while keeping the computational and storage overhead manageable. We will examine a number of proof-of-concept applications that have been studied widely within the Bayesian framework as well as stochastic design problems for which the present methodology maybe more appropriate.

*Table 1. Tasks for the solution of stochastic inverse problems*

- Develop a stochastic inverse/design framework based on recent advances in representing and solving SPDEs using non-statistical, non-intrusive methods.
- Develop a mathematical framework that seamlessly utilizes deterministic legacy simulators (including off-the-shelf optimization algorithms) and incorporates multiple sources of uncertainty.
- Develop highly scalable algorithms for computing stochastic sensitivities and stochastic Hessian.
- Utilizing stochastic model reduction strategies to accelerate the optimization framework.

## 0.2 Mathematical developments

We herein introduce various mathematical and algorithmic aspects of posing and solving stochastic inverse problems. The critical questions to be addressed include: (a) representation of the stochastic estimated/design variables, (b) modeling the effect of multiple sources of uncertainties in the inverse solution and addressing the curse of dimensionality (c) the definition of appropriate stochastic metrics to compare stochastic solutions, (d) the definition of the stochastic sensitivity variables and equations and (e) the need for a mathematical framework for model reduction to accelerate the optimization problem for large-scale problems. Each of the above proposed developments are described in some detail next.

### 0.2.1 The stochastic inverse problem: Mathematical problem definition

Consider a PDE-defined system over a domain,  $\mathcal{D}$ . Assume that we are interested in the behavior of the system in a time interval  $\mathcal{T}$ . This system is affected by multiple sources of uncertainty. This is because the input conditions (initial conditions, boundary conditions, forces, material properties) cannot be known/imposed with infinite certainty. Denote the set of input conditions as  $\{\mathbf{q}, \alpha\}$ , e.g.  $q$  can represent a boundary forcing term and  $\alpha$  a distributed material property. The dependent variable  $\mathbf{u}$  is a function of space, time and the input parameters  $\{\mathbf{q}, \alpha\}$ . The evolution of the system is described by a set of coupled SPDEs as

$$\mathcal{B}(\mathbf{u}, \mathbf{x}, t : \{\mathbf{q}, \alpha\}) = 0, \quad (1)$$

with the appropriate boundary and initial conditions given by

$$\mathcal{L}(\mathbf{u}, \mathbf{x}, t : \{\mathbf{q}, \alpha\}) = 0. \quad (2)$$

To numerically solve Eqs. (1–2), it is necessary to have a finite dimensional representation of the input uncertainty Deb et al. (2001). Denote this finite dimensional representation as

$$\mathbf{q} = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_q), \quad \mathbf{Y}_q \in \Gamma_q \quad (a), \quad \alpha = \mathcal{G}_2(\mathbf{x}, t, \mathbf{Y}_\alpha), \quad \mathbf{Y}_\alpha \in \Gamma_\alpha \quad (b), \quad (3)$$

where  $\mathcal{G}_1 : \Gamma_q \rightarrow \Omega_q$ , where  $\Omega_q$  is the space of all possible realizations of  $\mathbf{q}$ . Similar definitions hold for  $\mathcal{G}_2 : \Gamma_\alpha \rightarrow \Omega_\alpha$ . The set of inputs belong to a tensor product space, i.e. they have spatial and temporal dependence ( $\mathbf{q}, \alpha \in \mathcal{D} \otimes \mathcal{T}$ ) as well as depend on the input uncertainties ( $\mathbf{q} \in \Gamma_q, \alpha \in \Gamma_\alpha$ ). By the Doob-Dynkin lemma, the dependent variable  $\mathbf{u}$  also belongs to the corresponding tensor-product space, albeit one that is transformed by the differential operator,  $\mathcal{B} (\mathbf{u} \in \mathcal{D} \otimes \mathcal{T} \otimes \Gamma_\alpha \otimes \Gamma_q)$ .

Given a numerical representation of the input quantities (i.e  $\mathcal{G}_1$  and  $\mathcal{G}_2$ ), we utilize sparse grid collocation strategies to solve this direct stochastic problem. The

basic idea is to represent the input stochastic quantities in terms of a finite number of realizations and a set of multi-dimensional interpolating polynomials:

$$\mathbf{q} = \sum_{i=1}^{n_q} \mathbf{q}_i L_i^q(\mathbf{q}), \quad \alpha = \sum_{i=1}^{n_\alpha} \alpha_i L_i^\alpha(\alpha). \quad (4)$$

Eqs. (1–2) are solved for the finite set  $(\{\mathbf{q}_i\}, \{\alpha_i\})$  of realizations of the input stochastic quantities and the stochastic solution is represented as

$$\mathbf{u}(\dots, \mathbf{q}, \alpha) = \sum_{i=1}^{n_q} \sum_{j=1}^{n_\alpha} \mathbf{u}_{i,j}(\dots, \mathbf{q}_i, \alpha_j) L_i^q(\mathbf{q}) L_j^\alpha(\alpha). \quad (5)$$

Hence given an abstract representation of the input stochastic quantities in terms of a finite number of random variables  $\mathbf{Y}_q$  and  $\mathbf{Y}_\alpha$ , it is straightforward to solve for the stochastic dependent variable.

The inverse/design problems of interest can take several forms. In all cases, part of the boundary conditions or a source term or a material property need to be computed using some additional data provided either as sensor measurement of the dependent variable  $u$  or as desired system response. *Without loss of generality, we assume that the stochastic representation of one set of input conditions  $\alpha$  is specified, while the representation of the other set of stochastic input conditions  $\mathbf{q}$  has to be determined.* Mathematically, this is stated as:

Given  $\alpha = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_\alpha)$  find  $\mathbf{q}$  such that  $\mathbf{u}$  satisfies,

$$\mathcal{B}(\mathbf{u}, \mathbf{x}, t : \{\mathbf{q}, \alpha\}) = 0, \quad (6)$$

$$\mathcal{L}(\mathbf{u}, \mathbf{x}, t : \{\mathbf{q}, \alpha\}) = 0, \quad (7)$$

$$\text{and } \mathbf{u} = \mathcal{H}, \quad (8)$$

where  $\mathcal{H}$  is the desired stochastic evolution or measured stochastic system response in some locations in space and time.  $\alpha$  can represent the known material property variability and  $q$  an unknown boundary condition.

A naive formulation would be to construct  $\mathbf{q} = \mathcal{G}_2(\mathbf{Y}_q)$  given the input data and measurement. But there are inherent difficulties with this: specifically, (a) the dependent variable belongs to the tensor product space  $\Gamma_u \equiv \Gamma_q \otimes \Gamma_\alpha$  and there is no simple way to decompose it back into its constituent spaces. Furthermore, it is important to emphasize that the stochastic representation,  $\mathbf{q} = \mathcal{G}_2(\mathbf{Y}_q)$ , of any random process is a purely abstract scheme. From a physically realizable/observable perspective, only *operations* on the stochastic representation,  $\mathbf{q} = \mathcal{G}_2(\mathbf{Y}_q)$ , make sense (for instance, realizations of events, moments or the PDF). These operations are *reduced* (or averaged) representations of the abstract random process. *This observability/measurability argument strongly points to the necessity of posing the inverse stochastic problem in a way that the measured or desired dependent variables are given in the form of moments or PDF.* Moreover, it is appropriate to define the expected/measured/designed quantity  $\mathcal{L}(\mathbf{u})$  in terms of realizations, moments or

PDFs. In addition, recent work (Petersdorff and Schwab 2006; Schwab and Todur 2003, 2001) suggests that the direct stochastic equations posed in the form of moments or PDF have rigorous existence and uniqueness arguments (Zabaras and Ganapathysubramanian 2008).

In this work, the solution  $q$  of the inverse problem is computed in the form of moments or PDF. For simplicity, the domain  $(\mathbf{x}, t) \in (\mathcal{D}_q, \mathcal{T})$  of  $q$  is not shown explicitly in the following equations. That is, given  $\alpha = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_\alpha)$

Find (a)  $\langle \mathbf{q}^k \rangle$  or, (b)  $PDF(\mathbf{q})$  such that  $\mathbf{u}$  satisfies

$$\mathcal{B}(\mathbf{u}, \mathbf{x}, t; \{\mathbf{q}, \alpha\}) = 0, \quad (9)$$

and the desired behavior given by, (a)  $\langle \mathbf{u}^k \rangle = \mathcal{H}$  or, (b)  $PDF(\mathbf{u}) = \mathcal{H}$ . (10)

## 0.2.2 The stochastic metrics and representation of the inverse stochastic solution $q$

As discussed in Section 0.2.1, the measurement values or the desired system response  $\mathbf{u}$  is given at some locations within the domain in two forms, e.g. for  $(\mathbf{x}, t) \in (\mathcal{D}_s, \mathcal{T})$ : Case I: The mean and higher order moments of the system response are given,  $\langle \mathbf{u}^k(\mathbf{x}, t) \rangle = \mathcal{H}_k(\mathbf{x}, t)$ , for  $k \geq 1$  and Case II: The PDF of the system response is given  $PDF(\mathbf{u}^k(\mathbf{x}, t)) = \mathcal{H}(\mathbf{x}, t)$ . Corresponding to these physically meaningful input data statistics, the inverse problem solution will construct the statistics of the designed field as follows:

### Case I:

*Given the known stochastic input parameters,  $\alpha = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_\alpha)$ , find the moments of the stochastic input,  $\langle \mathbf{q}^k(\mathbf{x}, t, \mathbf{Y}_q) \rangle$ ,  $k = 1, \dots, p$ , such that the stochastic system defined by Eq. (1), results in statistics of the dependent variable corresponding to  $\langle \mathbf{u}^k(\mathbf{x}, t, \cdot) \rangle = \mathcal{H}_k(\mathbf{x}, t)$ ,  $k = 1, \dots, p$ ,  $(\mathbf{x}, t) \in (\mathcal{D}_s, \mathcal{T})$ .*

### Case II:

*Given the known stochastic input parameters,  $\alpha = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_\alpha)$ , find the PDF of the stochastic input,  $PDF[\mathbf{q}(\mathbf{x}, t, \cdot)]$ , such that the stochastic system defined by Eq. (1), results in statistics of the dependent variable corresponding to  $PDF[\mathbf{u}(\mathbf{x}, t, \cdot)] = \mathcal{H}(\mathbf{x}, t)$ ,  $(\mathbf{x}, t) \in (\mathcal{D}_s, \mathcal{T})$ .*

In design applications, the given moments or PDFs should be interpreted as ‘desired’ variability or performance robustness. In the context of inverse problems driven by data, this variability maybe induced either by sensor noise (in the case of a single experiment) or e.g. because of the variability of the random topology (repeated experimentation with random realizations of the medium will lead to variability in measurements even without measurement noise). All of these problems can be stated in the forms proposed above.

In this chapter, we are interested in designing for higher order statistics of the input fields. However posing and solving for the statistics directly will have the additional bottleneck of having to construct appropriate closure arguments. This will make the framework application specific. We propose to construct the complete stochastic representation instead. That is, we propose to construct  $\mathbf{q}(\mathbf{x}, t, \mathbf{Y}_q)$  such that its statistics satisfy the two cases defined above. This presents an apparent paradox because the construction of  $\mathbf{q}$  requires an a priori knowledge of the support of  $\mathbf{q}$ , i.e. knowledge of  $\mathbf{Y}_q \in \Gamma_q$ . However, since we are only interested in the statistics of  $\mathbf{q}$  and these statistics are essentially integrals over the support, *they are independent of the choice of the support,  $\Gamma_q$*  (see Fig. 2 for a verification using some of our preliminary work in (Zabaras and Ganapathysubramanian 2008)). Utilizing this rational, we represent the designed input stochastic field using a sparse grid representation *for an arbitrarily chosen support*. Our preliminary investigations have revealed that this representation is in fact valid and is a key step to effectively decoupling the stochastic inverse/design problem.

Without loss of generality, we can assume that  $\mathbf{Y}_q = \{Y_q^1, \dots, Y_q^n\}$  are uncorrelated uniform random variables defined in  $\Gamma_q \equiv [0, 1]^n$  as

$$\mathbf{q}(\mathbf{x}, t, \mathbf{Y}_q) = \sum_{i=1}^{n_q} q(\mathbf{x}, t, \mathbf{Y}_q^i) L_i(\mathbf{Y}_q). \quad (11)$$

With this assumption, the transform the problem of computing a stochastic function,  $q(\mathbf{x}, t, \mathbf{Y}_q)$ , into a problem of computing *a finite set of deterministic functions*,  $q(\mathbf{x}, t, \mathbf{Y}_q^i)$ ,  $i = 1, \dots, n_q$ .

The redefined stochastic inverse/design problems can now be stated as follows:

**Case I:**

*Given the known stochastic input parameters,  $\alpha = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_\alpha)$ , find the  $n_q$  deterministic functions,  $q(\mathbf{x}, t, \mathbf{Y}_q^i)$ ,  $i = 1, \dots, n_q$ , such that the stochastic system defined by Eq. (1), results in statistics of the dependent variable corresponding to  $\langle \mathbf{u}^k(\mathbf{x}, t, \cdot) \rangle = \mathcal{H}_k(\mathbf{x}, t)$ ,  $k = 1, \dots, p$ ,  $(\mathbf{x}, t) \in (\mathcal{D}_s, \mathcal{T})$ .*

**Case II:**

*Given the known stochastic input parameters,  $\alpha = \mathcal{G}_1(\mathbf{x}, t, \mathbf{Y}_\alpha)$ , find the  $n_q$  deterministic functions,  $q(\mathbf{x}, t, \mathbf{Y}_q^i)$ ,  $i = 1, \dots, n_q$ , such that the stochastic system defined by Eq. (1), results in statistics of the dependent variable corresponding to  $PDF[\mathbf{u}(\mathbf{x}, t, \cdot)] = \mathcal{H}(\mathbf{x}, t)$ ,  $(\mathbf{x}, t) \in (\mathcal{D}_s, \mathcal{T})$ .*

Eq. (9) together with the known initial conditions,  $\alpha$  and the sensor/desired conditions define an *ill-posed* problem that can be solved for the stochastic field  $\mathbf{q}$ . In the proposed work, we assume that a solution to the inverse problem exists in the sense of Tikhonov. That is, we look for a solution  $\{q_i^*\}_{i=1}^{n_q}$  such that:

$$\mathcal{CF}[\{q_i^*\}_{i=1}^{n_q}] \leq \mathcal{CF}[\{q_i\}_{i=1}^{n_q}], \quad \forall q(\mathbf{x}, t, \mathbf{Y}_q) = \sum_{i=1}^{n_q} q_i L_i(\mathbf{Y}_q). \quad (12)$$

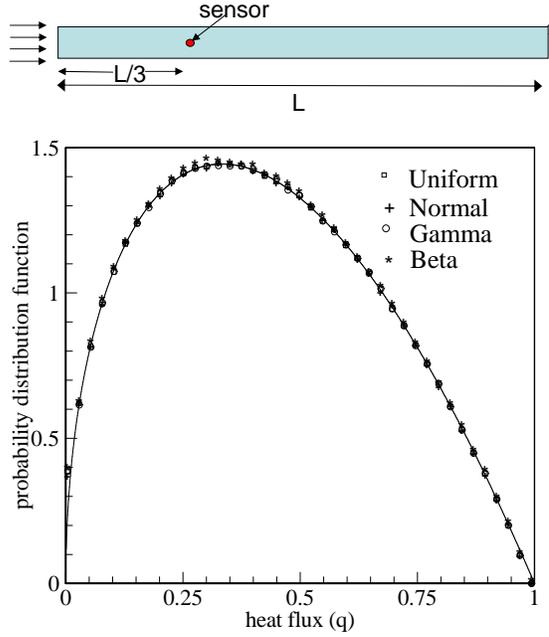


Figure 2 Preliminary results for a simple 1D stochastic inverse heat conduction illustrating the support independent formalism (Zabaras and Ganapathysubramanian 2008). A 1D domain of length  $L$  is considered (top figure). A specific temperature is maintained at one end. An (unknown) stochastic heat flux,  $q$ , is applied at the other end. Furthermore, the thermal conductivity is a known stochastic field. From the PDF of temperature taken at the sensor, the applied stochastic heat flux is reconstructed. The unknown heat flux is represented using four different support spaces. The PDF (and other moments) of the optimal solution is independent of the choice of the support as can be seen in the bottom figure.

Here,  $\mathcal{CF}[\{q_i\}_{i=1}^{n_q}]$  is a cost functional that quantifies how well the given data are matched from the simulated solution for a guessed random function  $q$ . For a given  $\{q_i\}_{i=1}^{n_q}$ , the cost functional is computed using the dependent stochastic variable,  $\mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1}^{n_q})$ . This is obtained from the solution of the direct stochastic problem in Eq. (9) with  $q$  as a (known) stochastic function, that is, using  $\{q_i\}_{i=1}^{n_q}$  as the guessed stochastic input conditions along with the given conditions  $\alpha$ . We will utilize quadratic cost functionals to allow use of the extensive (deterministic) quadratic optimization tools available.

For Case I, the cost functional measures how well the stochastic solution satisfies the given moments:

$$\mathcal{CF}[\{\mathbf{q}_i\}_{i=1}^{n_q}] = \frac{1}{2} \int_{\mathcal{D}_s} \int_{t=0}^{t_{max}} \sum_{k=1}^P [\langle \mathbf{u}(\mathbf{x}, t, \cdot)^k \rangle - \mathcal{H}_k(\mathbf{x}, t)]^2 d\mathbf{x} dt, \quad (13)$$

where  $\langle \mathbf{u}(\mathbf{x}, t, \cdot)^k \rangle$  is the  $k$ -th moment of the dependent variable and  $\mathcal{D}_s$  is the region over which sensor measurements are available. In Case II, the cost functional measures how well the stochastic solution satisfies the measured/desired PDF. Since the PDF, the CDF and the inverse CDF represent the same distribution in different ways, for the sake of computational effectiveness, we propose to define the cost functional in terms of the inverse CDF. This choice is motivated by the fact that the inverse CDF has a fixed, known support  $[0, 1]$  and this greatly simplifies defining the cost functional. Denote  $\Upsilon^{-1}[f, \mathbf{Y}]$  as the inverse cumulative distribution of the random variable  $f$ .  $\mathbf{Y}$  is the spanning variable,  $\mathbf{Y} \in [0, 1]$ . The inverse CDF is utilized because it is computationally easy to represent it using the collocation based polynomial interpolation representation. We define the cost functional in this case as:

$$\mathcal{CF}[\{\mathbf{q}_i\}_{i=1}^{n_q}] = \frac{1}{2} \int_{\mathcal{D}_s} \int_{t=0}^{t_{max}} (\Upsilon^{-1}[\mathbf{u}(\mathbf{x}, t), \mathbf{Y}] - \Upsilon^{-1}[\mathcal{H}(\mathbf{x}, t), \mathbf{Y}])^2 d\mathbf{x} dt. \quad (14)$$

*The stochastic inverse problem has now been converted into a deterministic optimization problem in a larger dimensional space – the space where the collocated values of  $q$  at each  $(\mathbf{x}, t) \in (\mathcal{D}_q, \mathcal{T})$  lie. Without spatial or temporal regularization (see also Section 0.2.5), we treat the nodal values of  $q$  at each finite element point  $(\mathbf{x}_j, t_k)$  in the discretization of  $(\mathcal{D}_q, \mathcal{T})$  as independent uncorrelated random variables. Their collocated values  $\{\mathbf{q}_i(\mathbf{x}_j, t_k)\}_{i=1}^{n_q}$  (on the same support space  $\Gamma_q$ ) define the unknowns of the overall optimization problem.*

### 0.2.3 Solving the direct stochastic problem: Adaptivity sparse grid collocation

As part of the inverse analysis one needs to be able to efficiently construct the solution  $\mathbf{u}$  to the underlying SPDEs. We utilize a sparse grid collocation strategy. This framework represents the stochastic solution as a polynomial approximation. This interpolant is constructed via independent function calls to the deterministic PDE with different realizations of the stochastic input. The sparse grid collocation strategy follows naturally from a full-tensor product representation of a multi-variate function in terms of univariate interpolation formulae. Consider a univariate interpolation operator  $\mathcal{U}$  to represent a one-dimensional function  $f$  in terms of realizations of  $f$  at some sampling points. That is,  $\mathcal{U}^i = \sum_{x^i \in X^i} a_{x^i} \cdot f(x^i)$  with the set of support nodes  $X^i = \{x_1^i, \dots, x_{m_i}^i\}$ ,  $x_k \in [0, 1]$ ,  $1 \leq k \leq m_i$  and the polynomial basis functions  $a_{x^i}$ . A multi-dimensional function  $f(x_1, \dots, x_N)$  can be represented as product of one-dimensional functions as  $\mathcal{U}^N f = (\mathcal{U}^{i_1} \otimes \dots \otimes \mathcal{U}^{i_N})(u)$ .

The tensor product representation quickly suffers the ‘curse-of-dimensionality’ problem – where if one utilized  $\mathcal{O}(k)$  points in one dimension, one would require  $\mathcal{O}(k^N)$  points in  $N$ -dimensions. *The sparse grid collocation strategy selectively chooses points in this uniform  $N$ -dimensional sampling* to significantly reduce the number of sampling points. It has been shown the the number of points reduces from  $\mathcal{O}(k^N)$  to  $\mathcal{O}((\log(k))^N)$  (Gerstner and Griebel 1998). Because of this advantage, the sparse grid collocation strategy has emerged as a very attractive alternative to the spectral stochastic paradigm. Nevertheless, two issues still remain to be satisfactorily resolved: (1) The conventional sparse grid strategies are isotropic. The efficiency can be significantly improved by incorporating adaptivity. However, the state-of-art adaptive sparse grid techniques developed still utilize global basis (Nobile et al. 2007). (2) It is necessary to develop an adaptive framework that scales linearly ( $\mathcal{O}(N)$ ) with dimensionality instead of the  $\mathcal{O}(2^N)$  scaling of current adaptive stochastic methods (e.g. Me-GPCE).

*We utilize a hierarchical basis based approach towards adaptivity that resolves the issues of locality and curse-of-dimensionality.* This borrows ideas from wavelet-based representation of functions. One of the key motivations towards using hierarchical basis functions is their linear scaling with dimensionality, in contrast to the  $n - d$  tree ( $2^d$ ) scaling of other h-type adaptive frameworks (for instance, the Me-GPCE framework (Wan and Karniadakis 2005)). This naturally results in rigorous convergence estimates for the adaptive sparse grid methods developed. This includes construction of appropriate hierarchical surpluses (described later in this section) that relate the interpolant to the local variance error. Such construction offers a natural means of adaptively refining the stochastic solution. We briefly describe the proposed developments below.

**Hierarchical-basis based adaptive sparse collocation:** The key to incorporating scalable adaptivity is defining incremental or hierarchical linear interpolants (Ma and Zabarar 2009). Define the incremental interpolant (Ganapathysubramanian and Zabarar 2007; Klimke 2006) as  $\Delta^i = \mathcal{U}^i - \mathcal{U}^{i-1}$ , where  $i$  is the level of interpolation used in the interpolation operator. Denote the number of interpolation points in multiple dimensions by the index  $\mathbf{i} = (i_1, \dots, i_N)$  with  $|\mathbf{i}| = i_1 + \dots + i_N$ . The conventional Smolyak algorithm is given by  $\mathcal{A}(q, N) = \sum_{|\mathbf{i}| \leq q} (\Delta^{i_1} \otimes \dots \otimes \Delta^{i_N})$ . To compute  $\mathcal{A}(q, N)$ , one needs to compute the function values at the sparse grid points given by  $\mathcal{H}(q, N) = \bigcup_{q-N+1 \leq |\mathbf{i}| \leq q} (X^{i_1} \times \dots \times X^{i_N})$  where  $X^i = \{y_1^i, \dots, y_{m_i}^i\}$  are the set of points used by  $\mathcal{U}^i$ .

One should select the set  $X^i$  in a nested fashion such that  $X^i \subset X^{i+1}$  to obtain many recurring points with increasing  $q$ . The basic idea towards adaptivity here is to use hierarchical surplus as an error indicator and only refine the grid points whose hierarchical surplus is larger than a pre-defined threshold. The hierarchical surplus at any point is just the value of the incremental interpolant. That is

$$\Delta^i(f) = \sum_{j=1}^{m_i^\Delta} a_j^i \cdot \underbrace{(f(x_j^i) - \mathcal{U}^{i-1}(f)(x_j^i))}_{w_j^i}, \quad (15)$$

where  $w_j^i$  are the hierarchical surpluses. We can apply the above equation to obtain the sparse grid interpolation formula for the multivariate case in a hierarchical form as  $\mathcal{A}_{q,N}(f) = \mathcal{A}_{q-1,N}(f) + \Delta\mathcal{A}_{q,N}(f)$  and

$$\mathcal{A}_{q-1,N}(f) = \sum_{|\mathbf{i}| \leq q-1} (\Delta^{i_1} \otimes \dots \otimes \Delta^{i_N}), \quad (16)$$

$$\Delta\mathcal{A}_{q,N}(f) = \sum_{|\mathbf{i}|=q-1} \sum_{\mathbf{j}} (a_{j_1}^{i_1} \otimes \dots \otimes a_{j_N}^{i_N}) \cdot w_{\mathbf{j}}^{\mathbf{i}}, \quad (17)$$

$$w_{\mathbf{j}}^{\mathbf{i}} = f(x_{j_1}^{i_1}, \dots, x_{j_N}^{i_N}) - \mathcal{A}_{q-1,N}(f)(x_{j_1}^{i_1}, \dots, x_{j_N}^{i_N}). \quad (18)$$

Defining the interpolation strategy in terms of the hierarchical surpluses provides a rigorous framework for adaptively interpolating multivariate functions. For continuous functions, the hierarchical surpluses tend to zero as the interpolation level tends to infinity. Furthermore, for non-smooth functions, details about the singularities are indicated by the magnitude of the hierarchical surplus (Ma and Zabarar 2009).

**Choice of nested points:** Our preliminary work on these aspects (Ma and Zabarar 2009) has revealed that the choice of the univariate sampling points is absolutely critical in determining the scalability of the framework. With an aim to make this framework scale efficiently to high-dimensions, one can utilize three types of nested grids: the Clenshaw-Curtis-type grid, the Chebyshev type grid and the maximum norm grid. For these grids, the number of points required as the dimensionality increases scales relatively slowly.

**Data structure for dimensional scalability:** As the dimensionality of the problem increases, the memory requirements for storing and constructing the interpolant increases substantially. One needs to utilize a tree-based data structure to efficiently store and retrieve the hierarchical surpluses. Given some user defined tolerance,  $\epsilon$ , the generalized sparse grid strategy is utilized to construct interpolants. The interpolant utilizes points *until the hierarchical surpluses associated with all most recently added points becomes less than the tolerance*. The choice of the new points to be sampled is governed by the tree structure (the daughter nodes of a node whose surplus  $> \epsilon$  are sampled in the next stage of the adaptive procedure).

**Error control through variance estimation:** The hierarchical surpluses can further be utilized to provide estimates on the convergence of the stochastic solution. Similar idea is used in the Me-GPCE (Wan and Karniadakis 2005). By developing explicit relations between the hierarchical surpluses to extract information about the local variance of the solution, one can adaptively sample only those regions whose variance is larger than a prescribed tolerance. This approach seems very promising because it circumvents the concept of discretization of the stochastic space into elements (which results in the N-dimensional tree structure that scales as  $2^N$ ) with a linear scaling model (since *only*  $2N$  points are required in this framework) (Mathelin and Le Maitre 2007).

Our preliminary investigations into the adaptive framework (Ma and Zabarar 2009) have shown the feasibility of this approach. We were able to solve using

adaptive hierarchical interpolants SPDEs driven by random inputs that lie in a 100-dimensional space– this being an order of magnitude increase over conventional stochastic solution strategies.

## 0.2.4 Stochastic sensitivity equations and gradient-based optimization framework

We utilize a gradient based optimization strategy to design the optimal stochastic input  $\{q_i\}$ . The first step is to compute the gradient of the cost functional with respect to the design variables. The directional derivative,  $D_{\Delta q_v} \mathcal{C}\mathcal{F}[\{q_i\}_{i=1}^{n_q}]$  of the cost functionals have to be computed:

<p><b>Gradient of the cost functional- Case I:</b></p> $D_{\Delta q_v} \mathcal{C}\mathcal{F}[\{q_i\}_{i=1}^{n_q}] = \int_{\mathcal{D}_s} \int_{t=0}^{t_{max}} \sum_{k=1}^p [\langle \mathbf{u}(\mathbf{x}, t, \cdot)^k \rangle - \mathcal{H}_k(\mathbf{x}, t)] D_{\Delta q_v}(\mathbf{u}(\mathbf{x}, t, \cdot)) d\mathbf{x} dt.$
<p><b>Gradient of the cost functional- Case II:</b></p> $D_{\Delta q_v} \mathcal{C}\mathcal{F}[\{q_i\}_{i=1}^{n_q}] = \int_{\mathcal{D}_s} \int_{t=0}^{t_{max}} (\Upsilon^{-1}[\mathbf{u}(\mathbf{x}, t), \mathbf{Y}] - \Upsilon^{-1}[\mathcal{H}(\mathbf{x}, t), \mathbf{Y}]) D_{\Delta q_v}(\mathbf{u}(\mathbf{x}, t, \cdot)) d\mathbf{x} dt. \quad (19)$

Recall that the stochastic variable  $\mathbf{u}$  is represented in terms of the stochastic interpolants as

$$\mathbf{u} = \sum_{i=1}^{n_q} \sum_{j=1}^{n_\alpha} \mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q^i, \mathbf{Y}_\alpha^j) L_q^i(\mathbf{Y}_q) L_\alpha^j(\mathbf{Y}_\alpha). \quad (20)$$

The calculation of the stochastic gradient requires the calculation of the directional derivative of  $\mathbf{u}$  w.r.t.  $q_v$ , i.e. of  $D_{\Delta q_v} \mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q^r, \mathbf{Y}_\alpha^e)$ . Denote the directional derivative of the stochastic variable with respect to each design variable as  $\hat{\mathbf{u}}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1}^{n_q}, \Delta q_v) \equiv D_{\Delta q_v} \mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1}^{n_q})$ . This defines the *sensitivity field of the dependent variable  $\mathbf{u}$  as the linear in  $\Delta q_v$  part of  $\mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1}^{n_q}, q_v + \Delta q_v)$ , where  $\Delta q_v$  is a perturbation to one of the unknown (collocated values of  $q$ ) variables.*

$$\begin{aligned} & \mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1, i \neq v}^{n_q}, q_v + \Delta q_v) = \\ & \mathbf{u}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1}^{n_q}) + \hat{\mathbf{u}}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_i\}_{i=1}^{n_q}, \Delta q_v) + h.o.t. \end{aligned} \quad (21)$$

The stochastic sensitivity equations are simply obtained by taking the directional derivative of the equations that define the parametric direct problem used to compute  $\mathbf{u}$  for each input  $q$ , i.e. the direct equations and the boundary conditions are linearized w.r.t. the design variables  $\{q_i\}$ . We denote directional derivatives of operators and

fields with  $\hat{\cdot}$ . These directional derivatives are taken with respect to each of the  $n_q$  collocation points used to represent  $q$ . The sensitivity equations take the following general form:

**Continuum stochastic sensitivity equations:**

$$\hat{\mathcal{B}}(\hat{\mathbf{u}}, \mathbf{u}, \mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha, \{q_i\}_{i=1}^{n_q}, \Delta q_v) = 0, \quad (22)$$

$$\hat{\mathcal{L}}(\hat{\mathbf{u}}, \mathbf{u}, \mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha, \{q_i\}_{i=1}^{n_q}, \Delta q_v) = 0. \quad (23)$$

The sensitivity of the stochastic variable computed at the  $\mathbf{Y}_q^v$  collocation point depends only on perturbations to the  $q_v$ -th design variable. Thus the sensitivity equations for perturbation  $\Delta q_v$  to each design variable  $q_v(x_j, t_k)$  are SPDEs defined in an  $M$  dimensional support space where  $M = n_q \times n_\alpha$ . These equations are solved with the sparse grid collocation method highlighted earlier for the direct analysis. Thus the complete stochastic sensitivity field w.r.t.  $q_v$  can be re-constructed using only deterministic sensitivity problems as

$$\hat{\mathbf{u}}(\mathbf{x}, t, \mathbf{Y}_q, \mathbf{Y}_\alpha : \{q_k\}_{k=1}^{n_q}, \Delta q_v) = \sum_{r=1}^{n_q} \sum_{e=1}^{n_\alpha} \hat{\mathbf{u}}(\mathbf{x}, t, \mathbf{Y}_q^r, \mathbf{Y}_\alpha^e : \{q_k\}_{k=1}^{n_q}, \Delta q_v) L_r(\mathbf{Y}_q) L_e(\mathbf{Y}_\alpha). \quad (24)$$

Simplified sensitivity calculations for linear problems are given in (Zabaras and Ganapathysubramanian 2008). Let us now denote the gradient of the cost functional with respect to the design variables  $\{q_i\}_{i=1}^{n_q}$  as  $\mathbf{d} = \{d_i\}^T$ . The gradient of the cost functional can be written in terms of the directional derivative (the directional derivative is just the gradient computed in a specific direction). Since we have a scheme to compute the directional derivative by solving the continuum stochastic sensitivity equations, the gradient of the cost-function is then simply given as  $\mathbf{d} = \{d_v\}^T = \{D_{\Delta q_v} \mathcal{C} \mathcal{F}[\{q_i\}_{i=1}^{n_q}] / \Delta q_v\}^T$ . Preliminary implementation of these ideas within a steepest descent approach for a stochastic diffusion problem are reported in our work in (Zabaras and Ganapathysubramanian 2008).

There are two aspects of these developments that play a crucial part towards making the framework scalable. The first involves developing mathematical strategies to make the solution of the direct stochastic problem more efficient— through the incorporation of adaptivity. This part of the proposed work is discussed in a previous section. The second involves developing mathematical frameworks that make the stochastic optimization problem efficient - through fast computation of the stochastic Hessian and accelerated convergence.

**Efficient optimization - Incorporating stochastic gradient information:** There is significant work that needs to be considered in the optimization to reduce the computational overhead including using high fidelity optimization schemes utilizing the Hessian information. The Hessian matrix is defined as:  $H_{ij} = \frac{\partial^2 \mathcal{C} \mathcal{F}}{\partial \mathbf{q}_i \partial \mathbf{q}_j}$ . Once could investigate the use of stochastic quasi-Newtonian schemes for computing the Hessian, specifically BFGS like schemes (Bashir et al. 2007; Dennis and More

1977). The key advantage is that the computational complexity is  $\mathcal{O}(n_q M)$  while the convergence rate is improved from linear (e.g. for the steepest descent method) to quadratic (for quasi-Newton schemes). The stochastic Hessian is computed as a set of decoupled deterministic Hessians at the  $M$  collocated points. For the stochastic optimization problem, at each search iteration  $k$ , the Hessian at a specific collocation point is updated as

$$H_{ij}^{k+1} = H_{ij}^k + \frac{r_i^k r_j^k}{\sum_p r_p^k s_p^k} - \frac{H_{ip}^k s_p^k H_{jq}^k s_q^k}{\sum_p \sum_q H_{pq}^k s_p^k s_q^k}, \quad (25)$$

where  $r^k = \nabla \mathcal{C}\mathcal{F}(\mathbf{q}^{k+1}) - \nabla \mathcal{C}\mathcal{F}(\mathbf{q}^k)$  and  $\mathbf{q}^{k+1} = \mathbf{q}^k + \gamma^k s^k$  where  $\gamma^k$  is computed by performing a line search to minimize  $\mathcal{C}\mathcal{F}$  along  $s^k$ . The search direction is computed from  $H^k s^k = -\nabla \mathcal{C}\mathcal{F}(\mathbf{q}^k)$ . Note that each step above is performed independently for each collocation point. This ensures that these computations are easily scalable.

**Efficient optimization - Hierarchical optimization strategies:** The stochastic design field  $\mathbf{q}$  is typically represented as a set of  $n_q$  deterministic fields. The computational complexity of the optimization scales with increasing  $n_q$  as does the accuracy of representation. Our preliminary numerical investigations in (Zabaras and Ganapathysubramanian 2008) revealed the following trends (a) The optimization problem with coarser representation (i.e. smaller  $n_q$ ) of the design variable converges faster initially, but becomes very slowly converging later and (2) The optimization problem with coarser representation is numerically faster to solve. These two observations naturally lead to the possibility of formulating a *hierarchical stochastic optimization* problem, where a coarser problem is solved to quickly compute a coarse solution that is in turn used as an initial guess to solve a finer problem. This idea (similar to the accelerated convergence using multi-grid methods) seems to offer great promise in terms of rapid solutions to inverse problems through solution of a hierarchy of coarser optimization problems. We propose to utilize this strategy to accelerate the optimization framework. One key issue that has to be satisfactorily resolved is when to stop the current optimization and use the solution as the initial guess for a refined optimization problem. This can be decided based on a trade-off between accuracy of representation and computational effort (Fig. 3).

**Analytical gradient calculations - The continuum stochastic adjoint equations:** For some classes of problems the gradient of the cost functional can be analytically computed using appropriate adjoint variables (Sampath and Zabaras 2001). In our recent work (Velamur Asokan and Zabaras 2004), we have extended this to formulating the stochastic adjoint equations to analytically compute the stochastic gradient. The continuum stochastic adjoint equations provide a direct means of computing the gradient while the continuum stochastic sensitivity equations provide the step size. One can utilize sparse grid collocation to solve the stochastic adjoint equations to explicitly compute the stochastic gradient. This strategy does not require the explicit computation of the stochastic Hessian. While our earlier implementation of these techniques for a stochastic inverse heat conduction problem using a GPCE approach revealed slow convergence in higher-order GPCE terms, one can revisit

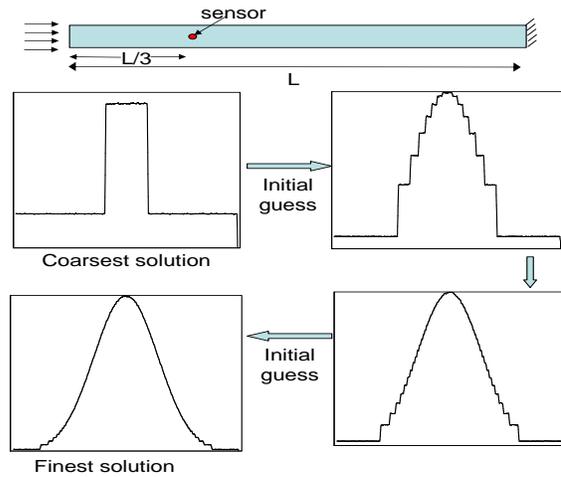


Figure 3 A simple 1D stochastic inverse heat conduction illustrating the hierarchical optimization strategy (Zabaras and Ganapathysubramanian 2008). A 1D domain of length  $L$  is considered (top figure). A specific temperature is maintained on the right end. An (unknown) stochastic heat flux,  $q$ , is applied at the left end. Furthermore, the thermal conductivity is a known stochastic field. From the PDF of temperature taken at the sensor, the applied stochastic heat flux is reconstructed. The optimization is accelerated by first solving for a coarse representation of the stochastic variable and utilizing this solution as an initial guess to a refined optimization problem. This resulted in 2 orders of magnitude reduction in iterations.

these techniques using the adaptive hierarchical sparse grid collocation discussed in Section 0.2.3.

## 0.2.5 Incorporating correlation statistics and investigating regularization

In classical inverse problems (Alifanov 1994), the ill-posedness requires some form of regularization or smoothing assumptions for computing the solution. But in the context of stochastic inverse problems, we are searching for a solution *in a distribution sense* in a much larger space (tensor product space of spatio-temporal and stochastic variations). This expanded search space essentially guarantees a stochastic solution (Kaipio and Somersalo 2005). Nevertheless, regularization arguments play a very significant role in reducing the computational effort by imposing some smoothness criteria on the designed stochastic solution. One needs to investigate

mathematical approaches to stochastic regularization as a purely computational strategy. A potential approach is two-pronged: (a) utilize spatial (temporal) regularization via parametrization of the spatial (temporal) representation of the unknown stochastic fields, and (b) utilize stochastic regularization via parametrization of the stochastic field.

**Spatial (temporal) parametrization of the unknown stochastic fields:** We here extend deterministic parametric representation of fields in a straightforward manner for the spatial representation of stochastic fields. There are many methodologies to parametrically represent the spatial variation of a stochastic field  $q(\mathbf{x}, \mathbf{Y})$  (wavelets, polynomials, Bézier surfaces among others). Consider a Bézier representation of  $q(\mathbf{x}, \mathbf{Y})$ - a stochastic field in two spatial dimensions. A Bézier curve of order  $(m_{x1}, m_{x2})$  is defined by a set of  $(m_{x1} + 1) \times (m_{x2} + 1)$  control points  $P_{i,j}$  as

$$q(x1, x2, \mathbf{Y}) = \sum_{i=0}^{m_{x1}} \sum_{j=0}^{m_{x2}} P_{i,j}(\mathbf{Y}) B_i^{m_{x1}}(x1) B_j^{m_{x2}}(x2), \quad (26)$$

$$\text{where, } B_i^{m_x}(x) = \frac{m_x!}{i!(m_x - i)!} x^i (1 - x)^{m_x - i}. \quad (27)$$

The design variables are now the stochastic control values  $P_{i,j}(\mathbf{Y})$ . These design variables can be subsequently represented in terms of  $n_q$  collocated values (similar to Eq. 11) as

$$P_{i,j}(\mathbf{Y}) = \sum_{r=1}^{n_q} P_{i,j}^r(\mathbf{Y}_r) L_r(\mathbf{Y}). \quad (28)$$

The sensitivity equations are now defined in terms of these  $n_q$  deterministic parameters. Since the choice of the parametrization controls the spatial smoothness of the solution, it is necessary to construct convergence estimates as the spatial parametrization is varied. Some preliminary work on convergence of this approximation is given in (Zabaras and Ganapathysubramanian 2008). Similar arguments can be utilized to parameterize the temporal variation of the stochastic field.

**Stochastic regularization of the unknown stochastic fields:** An alternative strategy is to consider some stochastic regularization. This can be imposed by enforcing some correlation structure on the unknown stochastic field. The correlation kernel  $C(\mathbf{x})$  determines the stochastic field and imposes some smoothness on its spatial variability. One can convert the problem of designing the stochastic field  $q$  to the problem of designing its correlation structure  $C(\mathbf{x})$ . The unknown correlation function is represented in terms of its spectral expansion

$$C(r) = \sum_{i=1}^K a_i \sin(ix/L), \quad (29)$$

where  $L$  is the characteristic length of the system and  $\{a_i\}$  is the set of  $K$  Fourier modes of the correlation function (and we have assumed that the field is isotropic,

without loss of generality). By representing the correlation this way, the stochastic optimization problem is converted to computing the modal coefficients  $\{a_i\}$ . The design parameters are now the set of  $K$  scalars  $\{a_i\}$ . Given any correlation kernel, the Karhunen-Loève expansion can be utilized to explicitly represent the stochastic field as

$$q(\mathbf{x}, \omega) = E(q(\mathbf{x}, \omega)) + \sum_{i=0}^N \sqrt{\lambda_i} f_i(\mathbf{x}) Y_i(\omega), \quad (30)$$

where  $(\lambda_i, f_i(\mathbf{x}))$  are the eigen-pairs of the correlation function. The stochastic sensitivity equations are formulated in terms of these  $K$  deterministic scalars. The choice of the number of modes  $K$  used in the representation of the unknown correlation provides a measure of the regularization. Including higher frequencies (larger  $K$ ) allows  $C(r)$  to be very steep representing a highly uncorrelated field. We can utilize an iterative solution strategy to sequentially increase the number of Fourier modes in the expansion to guarantee the accuracy of the stochastic solution. Defining and constructing the gradient of the cost functional is important along with developing a strategy for estimating the number of modes required based on the statistics of the measurements.

Designing/estimating the correlation statistics also has significant physical significance. Spatial correlation statistics provides some notion of how easy or difficult it is to physically realize the designed process. A very large or small correlation signifies expensive manufacturing, fabrication or operating conditions. This introduces the possibility of computing the designed fields with some constraints to obtain physically accessible fields. The above discussed strategies naturally take into account these issues.

## 0.2.6 Stochastic low-dimensional modeling

In the stochastic inverse/design problem, we need to solve the direct problem and the sensitivity problem for all of the collocation points at each iteration. For large stochastic dimensionality, each iteration in the optimization procedure will take substantial amount of computational time. The solution procedure can be accelerated significantly if a reduced order model of the system can be constructed (Holmes et al. 1996). *This reduced representation can either be created off-line or adaptively modified on-the-fly* (Ito and Ravindran 2001; Ravindran 2002a,b).

There has been recent work on extending the classical POD method to stochastic evolution equations (Venturi et al. 2008). A compact expansion of the stochastic dependent variable has been proposed into stochastic spatial modes. We consider a random field  $\mathbf{u}(\mathbf{x}, t; \mathbf{Y})$  in a space-time domain and we look for biorthogonal representations as

$$\mathbf{u}(\mathbf{x}, t; \mathbf{Y}) = \sum_{i=1}^K b_i^{(h)}(t) \mathbf{a}_i^{(h)}(\mathbf{x}; \mathbf{Y}). \quad (31)$$

The spatial-stochastic modes  $\mathbf{a}_i^{(h)}(\mathbf{x}; \mathbf{Y})$  are constructed from snapshots of the stochastic field (either from off-line calculations, available numerical data or experimental

data) by solving an eigenvalue problem. This eigen-value problem arises from the Euler-Lagrange equations resulting from minimizing the residual of the projection of the stochastic field onto the basis functions (Holmes et al. 1996; Venturi et al. 2008). The residual is calculated based on some norm  $\{\cdot, \cdot\}_h$ . Recent work (Venturi et al. 2008) has suggested that this stochastic counterpart of the POD framework will produce low-order ODEs (in terms of the temporal coefficients  $b_i^{(h)}(t)$ ) that are independent of the stochastic dimensionality of the system. It is important that care be taken in choosing the appropriate norm over which to construct the spatial stochastic basis. As a preliminary step, we propose to utilize inner products based on the mean, variance and standard deviation (Venturi et al. 2008): We propose to construct these spatial-stochastic modes  $\mathbf{a}_i$  using sparse grid collocation. That is, the stochastic modes are represented as

$$\mathbf{a}_k^{(h)}(\mathbf{x}; \mathbf{Y}) = \sum_{i=1}^M \hat{\mathbf{a}}_{ki}^{(h)}(\mathbf{x}, \mathbf{Y}_i) L_i(\mathbf{Y}). \quad (32)$$

An arbitrary stochastic field is represented in terms of these stochastic modes as

$$\mathbf{u}(\mathbf{x}, t; \mathbf{Y}) = \sum_{i=1}^K b_i^{(h)}(t) \sum_{j=1}^M \mathbf{a}_{ij}^{(h)}(\mathbf{x}; \mathbf{Y}_j) L_j(\mathbf{Y}). \quad (33)$$

The unknowns are the deterministic coefficients  $\{b_i\}$ . Inserting this representation, Eq. (33) into Eqs. (1)-(2) and using the orthogonality properties of the modes results in a set of ODEs for the modal coefficients,  $\{b_i\}$ .

$$\left( \mathcal{B} \left( \sum_{i=1}^K b_i^{(h)}(t) \mathbf{a}_i^{(h)} : \{\mathbf{q}, \alpha\} \right), b_j^{(h)} \right) = 0, \quad j = 1, \dots, K, \quad (34)$$

which can be written in compact form as

$$\frac{\partial b_j^h}{\partial t} = \mathcal{F}_j(b_j^h), \quad j = 1, \dots, K. \quad (35)$$

The cost functional, its gradient and the stochastic optimization problem can now be posed in terms of this set of ODEs instead of the set of SPDEs.

$$\begin{aligned} & \text{Minimize } \mathcal{CF}(\{b_i\}, \{q_i\}), \\ & \text{subject to } \frac{\partial b_i^h}{\partial t} = \mathcal{F}_i(b_i^h), \quad j = 1, \dots, K, \\ & b_i(0) = b_i^0. \end{aligned} \quad (36)$$

Once can utilize this strategy to compute stochastic modes (Eq. ??) for the control/optimization of SPDEs.

### 0.3 Numerical Examples

Let us consider a two-dimensional domain  $\mathcal{D} \equiv [-0.5, 0.5] \times [-0.5, 0.5]$  with the temperature at the right boundary specified as  $\theta = 0.5$  (Fig. 4). At  $t = 0$ , the temperature in the domain is  $\theta = 0$ . The spatial variation of the thermal diffusivity follows an exponential correlation with correlation length  $b = 10$  and mean value of  $\alpha = 10$ . We consider  $s = 41$  equally spaced sensors inside the domain at a distance  $d = 0.1$  from the left boundary with each sensor collecting data over the time interval  $[0, 0.5]$ . This data is given in terms of PDF of the temperature at each of these sensor location at 50 equally spaced time intervals in time range  $[0, 0.5]$ . The inverse problem of interest is posed as follows: *Identify the PDF of the heat flux on the left boundary in the time range  $[0, 0.5]$  such that the experimental measurements are reconstructed.*

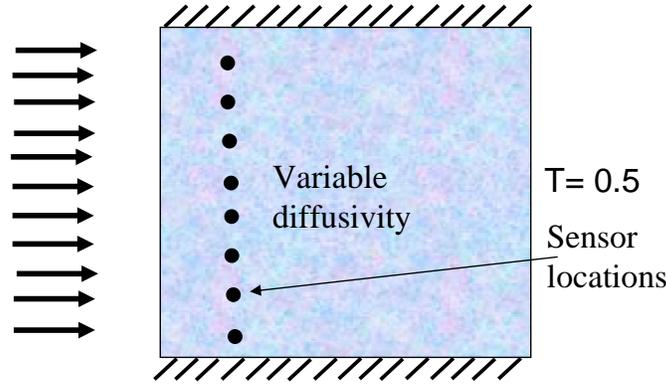


Figure 4 Schematic of the problem of interest: The temperature is given at the sensor locations shown by the dark circles. The spatial variation of the thermal diffusivity is defined by a known correlation kernel. We are interested to compute the unknown stochastic heat flux on the left boundary.

Let us expand the thermal diffusivity using a Karhunen-Loève expansion:

$$\alpha(x, y, \omega_\alpha) = \alpha_{mean}(x, y) + \sum_{i=1}^m \sqrt{\lambda_i} f_i(x, y) \xi_i, \quad (37)$$

where  $\lambda_i$  and  $f_i$  are the eigenvalues and eigenvectors of the correlation kernel. Fig. 5 plots the first few eigenvalues of the correlation matrix. The first three eigenvalues represent about 96% of the variation. Correspondingly, the thermal diffusivity is represented using three random variables  $\xi_i$ ,  $i = 1, 2, 3$ . Fig. 6 shows the eigenmodes corresponding to these random variables.

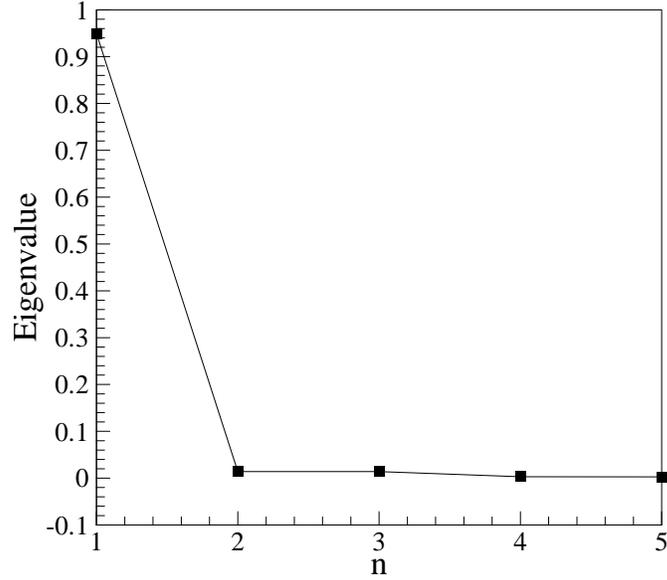


Figure 5 The eigenvalues of the correlation kernel that represents the thermal diffusivity variation.

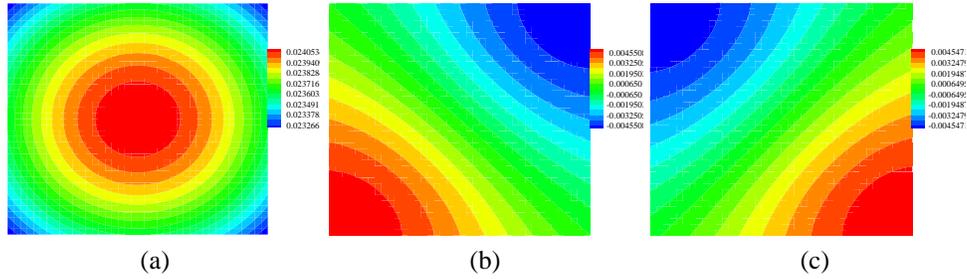


Figure 6 The first three modes of the Karhunen-Loève expansion of the thermal diffusivity.

### Computing the 'experimental' PDF

A direct problem is solved using an assumed stochastic variation for the heat flux applied on the left boundary. The PDFs of the temperature variation at the sensor locations are computed. This serves as the 'experimental' data that is used to run the inverse problem. The spatial variation of the heat flux is assumed to have an exponential correlation  $C_1(y_1, y_2) = \exp(-|y_1 - y_2|/b_1)$ , with  $b_1 = 0.5$ . This stochastic

heat flux is expanded using a Karhunen-Loève expansion and the eigenvalues are plotted in Fig. 7. The first three eigenvalues represent about  $\sim 94\%$  of the variation. The applied heat flux is thus represented using three random variables.

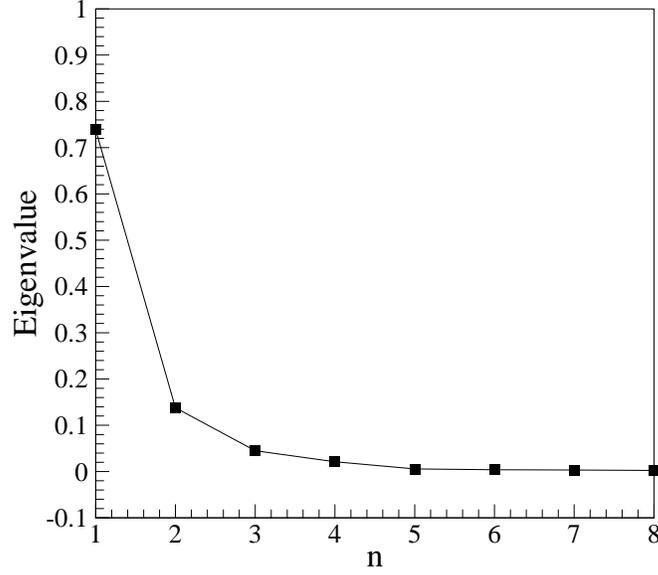


Figure 7 The eigenvalues of the correlation kernel that represents the ‘experimental’ heat flux.

The heat flux applied has a mean value of 20 and is given a time-dependent damping term  $exp(-\beta t)$  with  $\beta = 2.0$ . The input heat flux is thus of the form:

$$q_{exp}(y, t, \omega_q) = e^{-\beta t} \left[ 20.0 + 5.0 \sum_{i=1}^3 \sqrt{\lambda_i} g_i(y) \xi_i \right]. \quad (38)$$

The direct problem to obtain the ‘experimental’ statistics is solved using a  $80 \times 80$  quad element discretization of the spatial domain using  $\Delta t = 0.0025$ . In this problem, the temperature variation resides in a six-dimensional stochastic space (3D uncertainty in the thermal diffusivity and 3D uncertainty in the stochastic flux). A sparse grid collocation strategy is used to compute the temperature variability at the sensor locations. The thermal diffusivity variation is represented using a level 6 depth of interpolation (1073 points in 3D space). Similarly, 1073 points are used for the heat flux variation. The stochastic direct problem involves the solution of  $1073 \times 1073 = 1.151 \times 10^6$  direct deterministic problems.

The PDF of the applied ‘experimental’ heat flux at different times at  $(x, y) = (-0.5, 0.0)$  is given in Fig. 8. Notice that the initially diffuse PDF peaks and shifts towards a value of zero with increasing time because of the damping effect. This is clearly seen in Fig. 9 that plots the time variation of the mean flux applied.

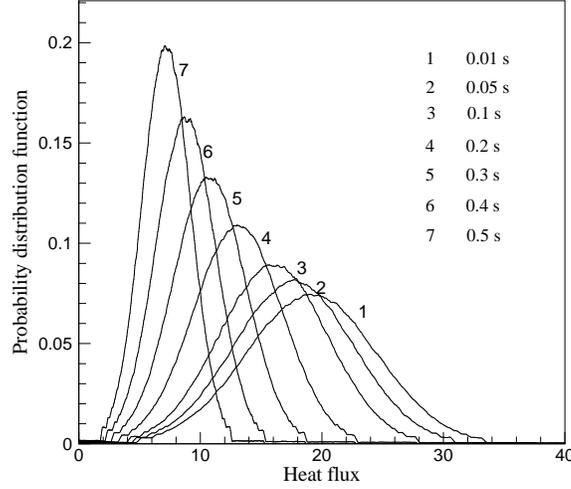


Figure 8 The ‘experimental’ PDF of the heat flux at the location  $(x, y) = (-0.5, 0.0)$  at different times.

The PDF of the resultant ‘experimental’ temperature at location  $(x, y) = (-0.4, 0.0)$  is given in Fig. 10. The initially diffuse PDF slowly peaks and shifts towards zero with increasing time because of the damping effect. The effect of the uncertain thermal diffusivity is also seen in the bimodal structure of the resultant PDF.

### The optimization problem: computational details and results

A  $40 \times 40$  quad element discretization of the domain is utilized to solve the inverse problem. The time domain is discretized into  $n_t = 50$  equal time steps. The total number of nodal points on the left vertical boundary is  $n_y = 41$  and the heat flux at each of these nodal points is assumed to be an independent random variable. The variation in the heat flux at each nodal point at each time instant has to be estimated. The total number of random variables that have to be estimated is equal to  $n_y \times n_t = 2050$ . The thermal diffusivity variation is represented using a level 5 depth of interpolation corresponding to using  $n_\alpha = 441$  realizations of the thermal diffusivity field. Each of the random fluxes is represented as

$$q(x, t, \xi) = \sum_{i=1}^{n_q} q(x, t, \xi_i) L_i(\xi). \quad (39)$$

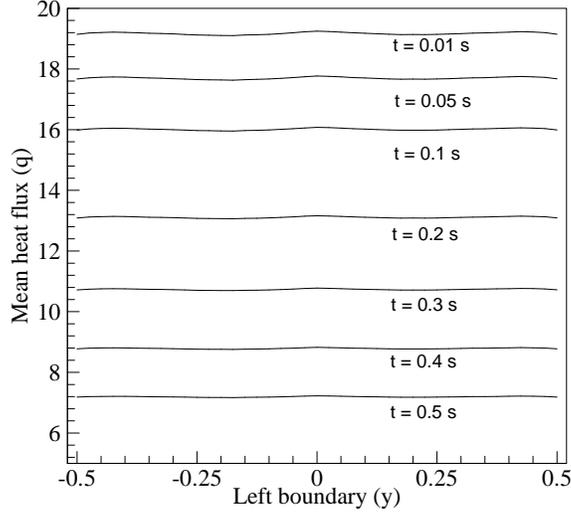


Figure 9 The time variation of the mean ‘experimental’ heat flux applied.

Without loss of generality, we assume that the heat flux can be represented using one uniform random variable. That is,  $\xi = U[0, 1]$ . A level 6 depth of interpolation is used to represent each random variable. This corresponds to  $n_q = 65$ . The total number of design variables is consequently  $n_q \times n_y \times n_t = 133250$ .

The optimization problem requires the estimation of the sensitivity of the temperature at the sensor locations to perturbations to each of the design variables. A set of decoupled deterministic sensitivity problems were run to construct the stochastic temperature sensitivity. The number of such deterministic sensitivity problems run is  $n_\alpha \times n_y \times n_t = 0.9 \times 10^6$ . Each iteration of the optimization problem requires the solution of the stochastic forward problem. The stochastic forward problem is solved as a set of decoupled direct deterministic problems. The total number of such direct deterministic problems was  $N_{run} = n_\alpha \times n_q = 28665$ . Each deterministic problem requires the solution on a  $40 \times 40$  quad grid for  $n_t$  time steps. The number of degrees of freedom (DOF) in each deterministic problem is  $N_{det} = (41)^2 \times n_t = 84050$ . The total number of DOF in *each* direct stochastic solve of the optimization algorithm is  $N_{stochastic} = N_{det} \times N_{run} = 2.41 \times 10^9$ . Thus, more than a *billion* DOF are solved at each iteration of the stochastic optimization problem.

The reduction in the cost functional with the number of iterations of the stochastic optimization is shown in Fig. 11. The optimization problem was run using our in-house Linux super computing cluster. Forty nodes, corresponding to 160 processors were utilized for the current problem. Each optimization iteration of the problem took 74 minutes to complete.

Fig. 12 plots the time evolution of the PDF of the heat flux at one location on

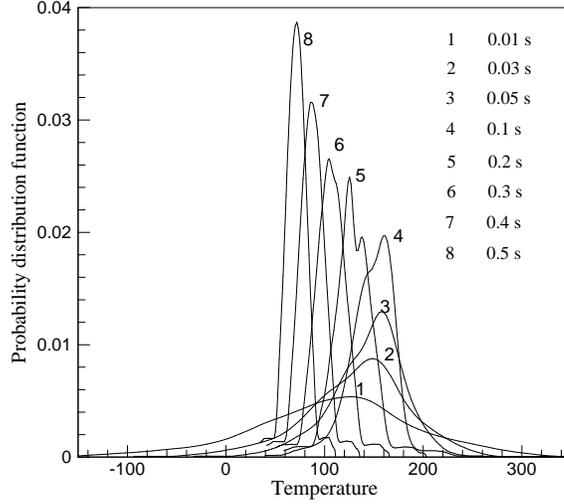


Figure 10 The ‘experimental’ PDF of temperature at the sensor location  $(x, y) = (-0.4, 0.0)$  at different times.

the boundary. The stochastic heat fluxes are reconstructed very well though there is some pixelation near the tails of the PDF.

To see if this pixelation disappears when a finer representation of the stochastic flux is used, another optimization was run using a depth of interpolation 8 of the stochastic heat flux. This corresponds to  $n_q = 257$  design variables to represent each unknown. We utilize the hierarchical stochastic optimization method to solve this larger optimization problem. That is, the solution of the previous optimization problem is used as the initial guess of this finer stochastic optimization problem. Fig. 13 plots the time evolution of the PDF of the heat flux at the same location on the boundary. The pixelation near the tails of the PDF in Fig. 12 is smoothed out by using a higher-depth of interpolation of the stochastic space.

The difference between the exact and the reconstructed stochastic heat flux is defined in terms of the error,  $e = \sum_{i=1}^{n_y} (\Upsilon^{-1}[q_{ex}(\mathbf{x}_i), u] - \Upsilon^{-1}[q_{rc}(\mathbf{x}_i), u])^2$ , where  $\Upsilon^{-1}(\cdot, u)$  is the corresponding inverse CDF of the heat flux at each nodal point on the left boundary and  $q_{ex}$  denotes the actual heat flux while  $q_{rc}$  denotes the reconstructed solution. Fig. 14 plots the error for the two optimization problems.

## 0.4 Summary

A scalable methodology was introduced that provides the ability to perform design and estimation in the presence of multiple sources of uncertainties while relying

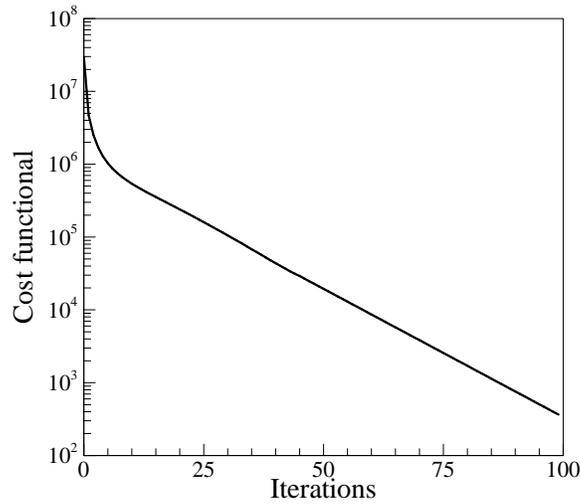


Figure 11 Reduction in the cost functional.

purely on the use of deterministic simulators. This is accomplished through constructing the solution of the stochastic direct problem using sparse grid interpolation strategies. A physically motivated framework based on arguments of measurability/observability of random processes is used to pose the stochastic inverse problem. The stochastic sensitivity solution is constructed via repeated calls to deterministic sensitivity/direct problems. By using the sparse grid interpolation formulation, the stochastic optimization problem is converted to a deterministic optimization problem in a higher-dimensional space. This naturally allows us to utilize mature deterministic optimization algorithms to solve the stochastic optimization. We also briefly discussed a hierarchical stochastic optimization algorithm which seems to provide significant computational gains. We are currently investigating extensions/improvements of the proposed stochastic optimization framework (coupling with more advanced optimization tools, using model reduction for the stochastic direct and sensitivity problems, etc.) along with applications to multiscale estimation in geological applications.

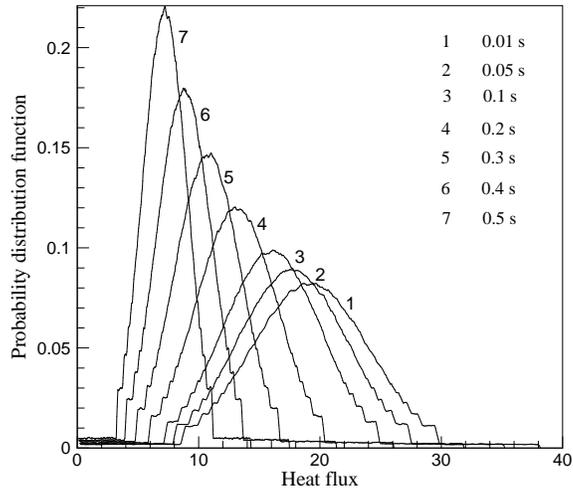


Figure 12 The reconstructed PDF of heat flux at the boundary location  $(x, y) = (-0.5, 0.0)$  using a depth of interpolation 6).

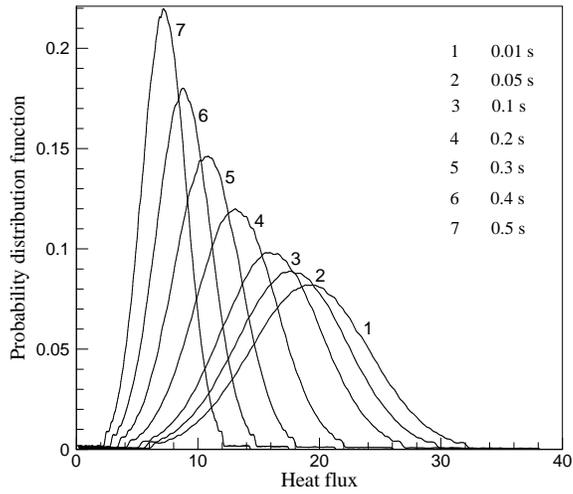


Figure 13 The reconstructed PDF of heat flux at the boundary location  $(x, y) = (-0.5, 0.0)$  at different times (with depth of interpolation 8).

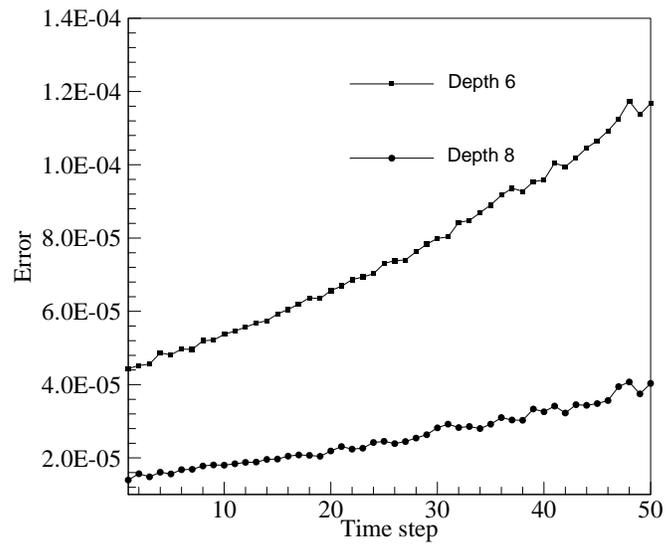


Figure 14 Error in the heat flux between the actual solution and the reconstructed solution.

# Bibliography

- Alifanov O M 1994 *Inverse Heat Transfer Problems*, Springer-Verlag, Berlin.
- Velamur Asokan B and Zabarar N 2004 Stochastic inverse heat conduction using a spectral approach. *Int. J. Numer. Meth. Eng.* **60**, 1569–1593.
- Velamur Asokan B and Zabarar N 2005 Using stochastic analysis to capture unstable equilibrium in natural convection. *J. Comp. Physics* **209** 134-153.
- Babuska I, Tempone R and Zouraris GE 2005 Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.* **194** 1251-1294.
- Babuska I, Nobile F and Tempone R 2005 A stochastic collocation method for elliptic partial differential equations with random input data. *ICES Report* 05-47.
- Bashir O, Ghattas O, Hill J, van Bloemen Waanders B and Willcox K 2007 *Hessian-Based Model Reduction for Large-Scale Data Assimilation Problems, Lecture Notes in Computer Science*, Springer Berlin/Heidelberg.
- Deb M K, Babuska I K and Oden J T 2001 Solution of stochastic partial differential equations using the Galerkin finite element techniques. *Comput. Methods Appl. Mech. Engrg.* **190** 6359–6372.
- Dennis J E and More J J 1997 Quasi Newton methods, motivation and theory. *SIAM Review* **19/1** 46–89.
- Desceliers C, Ghanem R and Soize C 2007 Maximum likelihood representation of stochastic chaos representations from experimental data. *Int. J. Num. Met. Engr.* **66** 978-1001.
- Ganapathysubramanian B and Zabarar N 2007 Sparse grid collocation schemes for stochastic natural convection problems. *J. Comp. Physics* **225** 652-685.
- Ganapathysubramanian B, Zabarar N 2008 A non-linear dimension reduction methodology for generating data-driven stochastic input models. *J. Comp. Physics* **227** 6612–6637.
- Ganapathysubramanian B and Zabarar N 2007 Modelling diffusion in random heterogeneous media: Data-driven models, stochastic collocation and the variational multi-scale method. *J. Comp. Physics* **226** 326–353.
- Gerstner T and Griebel M 1998 Numerical integration using sparse grids. *Numerical Algorithms* **18** 209-232.
- Ghanem R 1991 Probabilistic characterization of transport in heterogeneous porous media. *Comput. Methods Appl. Mech. Engrg.* **158** 199–220.
- Ghanem RG and Spanos PD 1991 *Stochastic Finite Elements: A Spectral Approach* Dover publications.
- Ghanem R and Sarkar A 2002, Mid-frequency structural dynamics with parameter uncertainty. *Comput. Methods Appl. Mech. Engrg.* **191** 5499–5513.

- Holmes P, Lumley J L and Berkooz G 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* Cambridge Univ press.
- Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems*, Springer, New York.
- Klimke A 2006 *Uncertainty modeling using fuzzy arithmetic and sparse grids*, Ph.D. Thesis, Universitt Stuttgart, Shaker Verlag, Aachen.
- Ito K and Ravindran S S 2001 Reduced Basis Method for Optimal Control of Unsteady Viscous Flows. *International Journal of Computational Fluid Dynamics* **15** 97–113.
- Ma X and Zabaras N 2009 An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comp. Physics* **228**, 3084–3113.
- Mathelin L and Le Maitre O 2007 Dual-Based a posteriori Error Estimate for Stochastic Finite Element Methods. *Communications in Applied Mathematics and Computational Sciences* **2**, 83–115.
- Nobile F, Tempone R and Webster C 2008 A sparse grid stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* **46** 2309-2345.
- Nobile F, Tempone R and Webster C 2007 An anisotropic sparse grid collocation method for elliptic partial differential equations with random input data, MOX, Dipartimento di Matematica, Politecnico di Milano, Italy.
- Petersdorff T and Schwab Ch 2006 Sparse finite element methods for operator equations with stochastic data. *Applications of Mathematics* **51** 145–180.
- Ravindran S S 2002 Adaptive Reduced-Order Controllers for a Thermal Flow System. *SIAM Journal on Scientific Computing* **23** 1925–1943.
- Ravindran S S 2002 Control of flow separation over a forward-facing step by Model reduction. *Comput. Methods Appl. Mech. Engrg.* **191** , 4599–4617.
- Sampath R and Zabaras N 2001 A functional optimization approach to an inverse magneto-convection problem. *Comput. Methods Appl. Mech. Engrg.* **190** 2063–2097.
- Schwab Ch and Todur R A 2003 Sparse finite elements for stochastic elliptic problems higher moments. *Computing* **71** 43-63.
- Schwab Ch and Todur R A 2001 Sparse finite elements for elliptic problems with stochastic loading. *Numerische Mathematik* **95**, 707–734.
- Somersalo E and Calvetti D 2007 *Subjective Computing: Introduction to Bayesian Scientific Computing*, Springer Verlag.
- Venturi D, Wan X and Karniadakis G E 2008 Stochastic low-dimensional modeling of random laminar wake past a circular cylinder. *J. Fluid Mechanics* **606**, 339-367.
- Wan X and Karniadakis G E 2005 An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comp. Physics* **209**, 617–642.
- Wang J and Zabaras N 2005 Hierarchical Bayesian models for inverse problems in heat conduction. *Inverse Problems* **21**, 183–206.
- Wang J and Zabaras N 2004 A Bayesian inference approach to the stochastic inverse heat conduction problem. *Int. J. Heat Mass Transfer* **47**, 3927-3941.
- Wang J and Zabaras N 2005 Using Bayesian statistics in the estimation of heat source in radiation. *Int. J. Heat Mass Transfer* **48**, 15-29.
- Wang J and Zabaras N 2006 A Markov random field model to contamination source identification in porous media flow. *Int. J. Heat Mass Transfer* **49**, 939–950.
- Xiu D and Hesthaven JS 2005 High-order collocation methods for the differential equation with random inputs. *SIAM J. Sci. Comput.* **27**, 1118-1139.

- Xiu D and Karniadakis GE 2002 Modeling uncertainty in steady-state diffusion problems via generalized polynomial chaos. *Comput. Methods Appl. Mech. Engrg.* **191**, 4927–4948.
- Xiu D and Karniadakis GE 2003 Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comp. Physics* **187**, 137-167.
- Xiu D 2007 Efficient collocation approach for parametric uncertainty analysis. *Comm. Computational Physics* **2**, 293-309.
- Zabaras N and Ganapathysubramanian B 2008 A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach. *J. Comp. Physics* **227**, 4697–4735.