Bayesian Inference for Inverse Problems Using Surrogate Model

Dongbin Xiu

Department of Mathematics, Purdue University

Supported by AFOSR, DOE, NNSA, NSF

Overview

- **Uncertainty quantification (UQ) and stochastic modeling**

- **Forward problem: uncertainty propagation**
  - Brief introduction of generalized polynomial chaos (gPC)

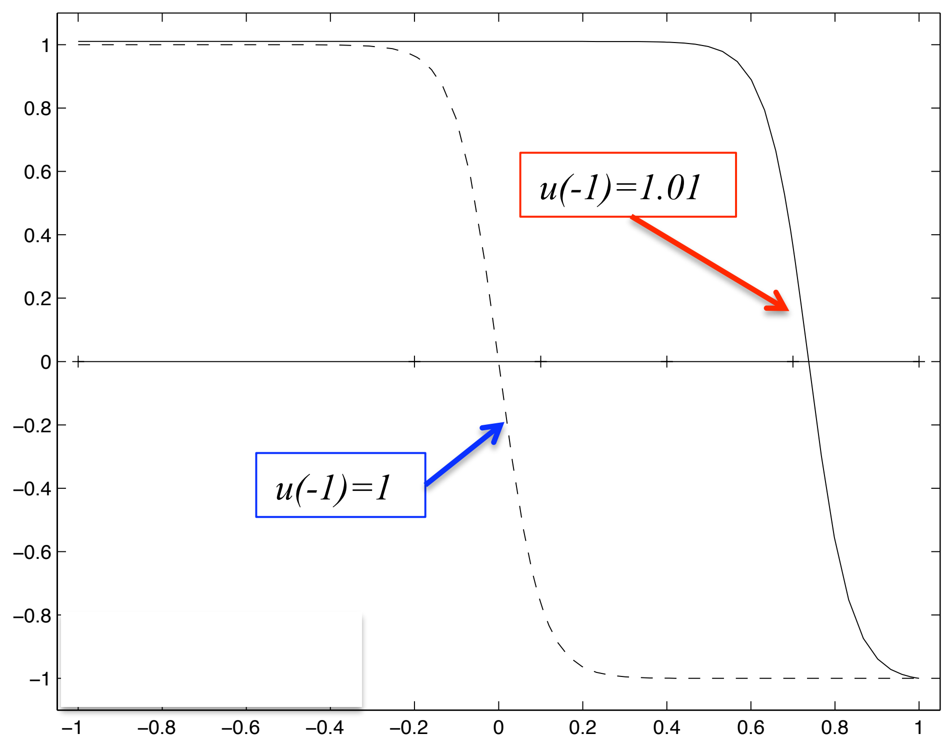
- **Inverse problem: Bayesian inference**
  - Use generalized polynomial chaos (gPC)

- **Back to the forward problem**
  - Issues and challenges

- **Key Issues:**
  - **Efficiency**
  - **Curse-of-dimensionality**

Illustrative Example: Burgers Equation

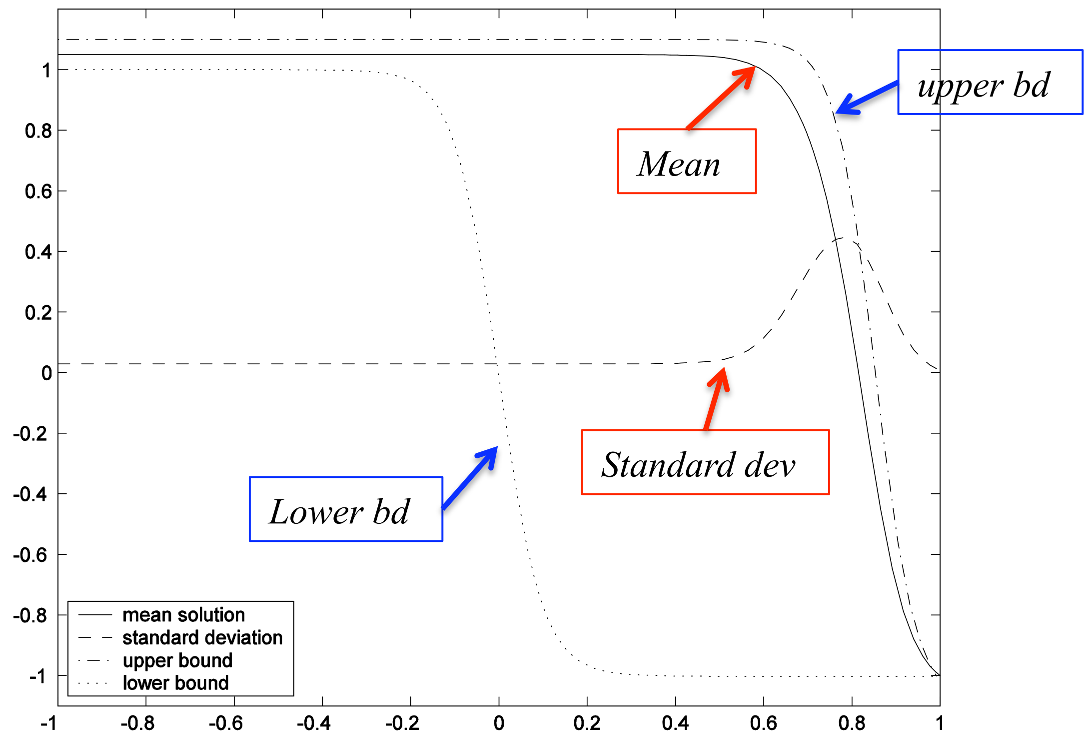
- Burgers' equation : $$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \quad u(-1) = 1, \quad u(1) = -1$$
- Steady state solution: $$u(x) = -A\tanh\left[\frac{A}{2\nu}\left(x - z\right)\right], \quad \text{where } u(z) = 0,\ A = -\left.\frac{\partial u}{\partial x}\right|_{x=z}$$



ν=0.05

Effects of Uncertainty – "Supersensitivity"

Method	(Mean, STD)	Cost (unit)
Interval analysis	N/A	<2
1st-order perturbation	**(0.823, 0.349)**	~2
4th-order perturbation	**(0.824, 0.328)**	~5
4th-order gPC Galerkin	**(0.814, 0.414)**	~5
Monte Carlo simulation	**(0.814, 0.414)**	~10,000



(Xiu & Karniadakis, Int. J. Numer. Eng., vol. 61, 2004)

(Re-)Formulation of PDE: Input Parameterization

$$\frac{\partial u}{\partial t}(t,x) = \mathcal{L}(u) \quad + \text{ boundary/initial conditions}$$

- **Goal:** To characterize the random inputs by a set of random variables
  - Finite number
  - Mutual independence

- **If inputs == parameters**
  - Identify the (smallest) independent set
  - Prescribe probability distribution

- **Else if inputs == fields/processes**
  - Approximate the field by a function of finite number of RVs
  - Well-studied for Gaussian processes
  - Under-developed for non-Gaussian processes
  - Examples: Karhunen-Loeve expansion, spectral decomposition, etc.

$$a(x,\omega) \approx \mu_a(x) + \sum_{i=1}^{d} \tilde{a}_i(x) Z_i(\omega)$$

The Reformulation

- **Stochastic PDE:**

$$\frac{\partial u}{\partial t}(t,x,Z) = \mathcal{L}(u) \quad + \text{ boundary/initial conditions}$$

- **Solution:** $u(t,x,Z):[0,T]\times \bar{D}\times \mathbb{R}^{n_z} \rightarrow \mathbb{R}$

- Uncertain inputs are characterized by n_z random variables Z

- Probability distribution of Z is prescribed

$$F_Z(s) = \Pr(Z \le s), \quad s \in \mathbb{R}^{n_z}$$

Non-trivial task

Generalized Polynomial Chaos (gPC)

$$\frac{\partial u}{\partial t}(t,x,Z) = \mathcal{L}(u) \quad + \text{ boundary/initial conditions}$$

- **Focus on dependence on *Z*:** $u(\bullet,Z):\mathbb{R}^{n_Z} \to \mathbb{R}$

- ***N*th-order gPC expansion:**

$$u_N(Z) \in \mathcal{P}_N = \{\text{space of } n_Z\text{-variate polynomials of degree up to } N\}$$

- **Orthogonal basis:** $\mathbf{i} = (i_1,\cdots,i_{n_Z}), \quad |\mathbf{i}| = i_1 + \cdots + i_{n_Z}$

$$\mathbb{E}\left[\Phi_{\mathbf{i}}(Z)\Phi_{\mathbf{j}}(Z)\right] = \int \Phi_{\mathbf{i}}(z)\Phi_{\mathbf{j}}(z)\rho(z)\,dz = \delta_{\mathbf{ij}}$$

$$u_N(t,x,Z) \triangleq \sum_{|\mathbf{k}|=0}^{N} \hat{u}_{\mathbf{k}}(t,x)\Phi_{\mathbf{k}}(Z) \qquad \dim \mathcal{P}_N = \begin{pmatrix} n_z + N \\ N \end{pmatrix}$$

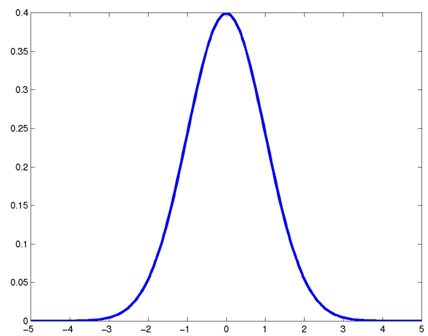
gPC: Basis

- **Expectation:**

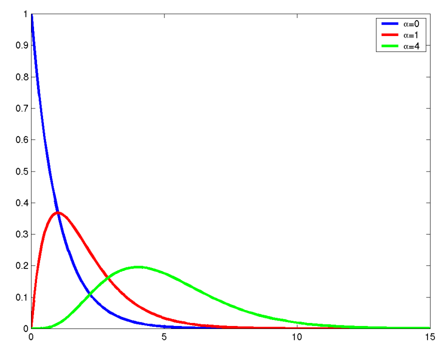
$$\mathbb{E}\big(g(Z)\big) = \int_{\mathbb{R}} g(z)\rho(z)\,dz$$

- **Orthogonality:**

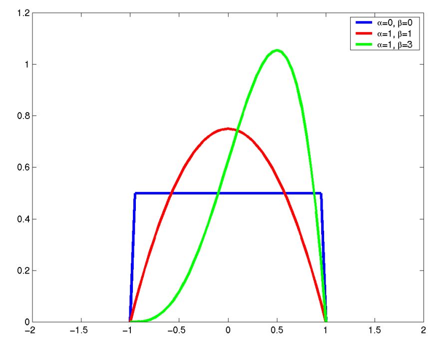
$$\int \Phi_{\mathrm{i}}(z)\Phi_{\mathrm{j}}(z)\rho(z)\,dz = \mathbb{E}\big[\Phi_{\mathrm{i}}(Z)\Phi_{\mathrm{j}}(Z)\big] = \delta_{\mathrm{ij}}$$



Gaussian distribution



Gamma distribution



Beta distribution

$$\int_{-\infty}^{\infty} \Phi_{\mathrm{i}}(z)\Phi_{\mathrm{j}}(z)e^{-z^2}\,dz = \delta_{\mathrm{ij}}$$

$$\int_{0}^{\infty} \Phi_{\mathrm{i}}(z)\Phi_{\mathrm{j}}(z)e^{-z}\,dz = \delta_{\mathrm{ij}}$$

$$\int_{-1}^{1} \Phi_{\mathrm{i}}(z)\Phi_{\mathrm{j}}(z)\,dz = \delta_{\mathrm{ij}}$$

Hermite polynomial

Laguerre polynomial

Legendre polynomial

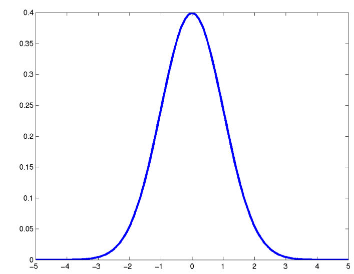
(Xiu & Karniadakis, SISC, 2002)

gPC Basis: the Choices

- **Orthogonality:** $$\int \Phi_{\mathbf{i}}(z)\Phi_{\mathbf{j}}(z)\rho(z)\,dz = \mathbb{E}\left[\Phi_{\mathbf{i}}(Z)\Phi_{\mathbf{j}}(Z)\right] = \delta_{\mathbf{ij}}$$

- **Example: Hermite polynomial**

$$\int_{-\infty}^{\infty} \Phi_{\mathbf{i}}(z)\Phi_{\mathbf{j}}(z)e^{-z^2}\,dz = \delta_{\mathbf{ij}}$$

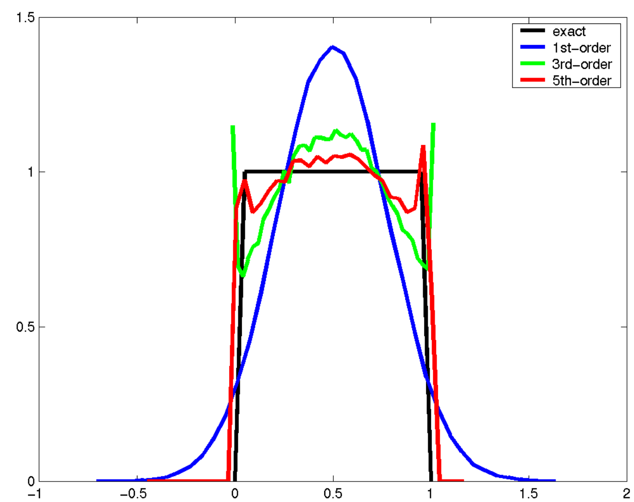


- **The polynomials:** *Z~N(0,1)*

$$\Phi_0 = 1, \quad \Phi_1 = Z, \quad \Phi_2 = Z^2 - 1, \quad \Phi_3 = Z^3 - 3Z, \quad \cdots$$

- **Approximation of arbitrary random variable:** Requires L^2 integrability

- **Example:** Uniform random variable
  - Convergence
  - Non-optimal
  - First-order Legendre is exact



Stochastic Galerkin

$$\frac{\partial u}{\partial t}(t,x,Z) = \mathcal{L}(u) \quad + \text{ boundary/initial conditions}$$

- **Galerkin method:** Seek

$$u_N(t,x,Z) \triangleq \sum_{|\mathbf{k}|=0}^{N} \hat{u}_{\mathbf{k}}(t,x)\Phi_{\mathbf{k}}(Z)$$

Such that

$$\mathbb{E}\left[\frac{\partial u_N}{\partial t}(t,x,Z)\mathbf{\Phi}_{\mathbf{m}}(Z)\right] = \mathbb{E}\left[\mathcal{L}(u_N)\mathbf{\Phi}_{\mathbf{m}}(Z)\right], \quad \forall |\mathbf{m}| \le N$$

- **The result:**
  - Residue is orthogonal to the gPC space
  - A set of deterministic equations for the coefficients
  - The equations are usually coupled – requires new solver

Stochastic Collocation

$$\frac{\partial u}{\partial t}(t, x, Z) = \mathcal{L}(u) \quad + \text{ boundary/initial conditions}$$

- **Collocation:** To satisfy governing equations at selected nodes
  - Allow one to use existing deterministic codes repetitively

Let $Z^1, Z^2, \ldots Z^Q \in \mathbb{R}^{n_z}$, be a set of nodes/samples, then solve

$$\frac{\partial u}{\partial t}(t, x, Z^j) = \mathcal{L}(u), \quad j = 1, \ldots, Q.$$

- **Sampling:** (solution statistics only)
  - Random (Monte Carlo)
  - Deterministic (lattice rule, tensor grid, cubature)

- **Stochastic collocation:** To construct **polynomial approximations**
  - Node selection is critical to efficiency and accuracy
  - More than sampling

Stochastic Collocation: Interpolation

$$\frac{\partial u}{\partial t}(t,x,Z) = \mathcal{L}(u) \quad + \text{ boundary/initial conditions}$$

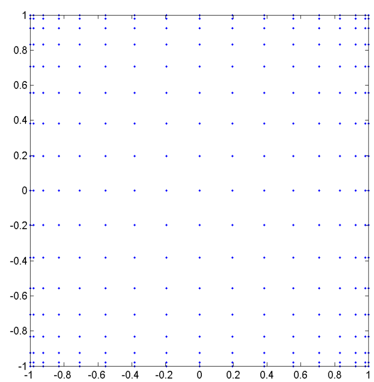
- **Definition:** Given a set of nodes and solution ensemble, find u_N in a proper polynomial space, such that $u_N \approx u$ in a proper sense.

- **Interpolation Approaches:**

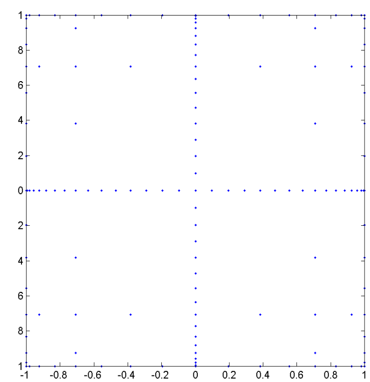
$$u^Q(Z) \triangleq \sum_{j=1}^{Q} u(Z^j) L_j(Z)$$

$$L_i(Z^j) = \delta_{ij}, \quad 1 \le i,j \le Q$$

- *Optimal nodal distribution in high dimensions…*



Tensor grids: inefficient



Sparse grids: more efficient

(Xiu & Hesthaven, SIAM J. Sci. Comput., 05)

Stochastic Collocation: Discrete Projection

- **Orthogonal projection:**

$$u_N(t,x,Z) \triangleq \mathbb{P}_N u = \sum_{|\mathbf{k}|=0}^{N} \hat{u}_{\mathbf{k}}(t,x)\Phi_{\mathbf{k}}(Z)$$

$$\hat{u}_{\mathbf{k}} = \mathbb{E}[u(Z)\Phi_{\mathbf{k}}(Z)] = \int u(z)\Phi_{\mathbf{k}}(z)\rho(z)\,dz$$

- **Discrete projection:**

$$w_N(t,x,Z) = \sum_{|\mathbf{k}|=0}^{N} \hat{w}_{\mathbf{k}}(t,x)\Phi_{\mathbf{k}}(Z)$$

$$\hat{w}_{\mathbf{k}} = \sum_{j=1}^{Q} u(t,x,Z^j)\Phi_{\mathbf{k}}(Z^j)\alpha^j \approx \int u(z)\Phi_{\mathbf{k}}(z)\rho(z)\,dz$$

$$\hat{w}_{\mathbf{k}}(t,x) \to \hat{u}_{\mathbf{k}}(t,x), \quad Q \to \infty$$

- **Aliasing Error:** $\varepsilon_Q \triangleq \left\| u_N - w_N \right\|_{L^2_\rho}$

Inverse Parameter Estimation

- **Solution of a stochastic system:**

$$u(t,x,Z):[0,T]\times\overline{D}\times\mathbb{R}^{n_z}\mapsto\mathbb{R}^{n_u}$$

- **Need: Probability distribution of the parameters *Z***

  - Prior distribution: $\pi_Z(z)=\prod_{i=1}^{n_z}\pi_i(z_i)$

- **Information of the prior distribution is critical**
  - Requires direct measurements of the parameters
  - No/not enough direct measurements? (Use experience/intuition …)
  - How to take advantage of measurements of other variables?

Bayesian Inference for Parameter Estimation

- **Setup:**
  - Prior distribution of the parameters *Z*: $\pi(z)$
  - Forward problem simulation:

    $$y=G(Z)$$
  - Measurement/data: d

- **Error:** $d = G(Z) + e, \quad e \in \mathbb{R}^{n_d}$ is i.i.d.

- **Goal:** To estimation the distribution of *Z* --- posterior distribution

- **Posterior distribution:** $\pi(Z \mid d) \propto \pi(d \mid Z)\pi(Z)$

- **Likelihood function:** $\pi(d \mid Z) = \prod_{i=1}^{n_d} \pi_{e_i}(d_i - G_i(Z))$

- **Notes:**
  - Difficult to manipulate
  - Classical sampling approaches can be time consuming (MCMC, etc)

Surrogate-based Bayesian Estimation

- **Surrogate:** $$y_N = G_N(Z)$$

- **Approximate Bayes rule:**

$$\pi_N(Z \mid d) \propto \pi_N(d \mid Z)\pi(Z)$$

where $$\pi_N(d \mid Z) = \prod_{i=1}^{n_d} \pi_{e_i}(d_i - G_{N,i}(Z))$$

- **Properties:**
  - Allows direct sampling with arbitrarily large samples
  - No additional simulations – forward problem solver only

Convergence Analysis

- **Kullback-Leibler divergence:** $D(\pi_1 \| \pi_2) \triangleq \int \pi_1(z) \log \frac{\pi_1(z)}{\pi_2(z)} dz$

- **Basic assumption:** observation error is i.i.d. Gaussian

Theorem. If the gPC expansion G_N converges to G in $L^2_{\pi_Z}$, then the posterior density π_N^d converges to π^d in the sense

$$\left| D\left(\pi_N^d \| \pi^d\right) \right| \to 0, \quad N \to \infty.$$

Moreover, if

$$\left\| G_i(Z) - G_{N,i}(Z) \right\|_{L^2_{\pi_z}} \leq CN^{-\alpha}, \quad 1 \leq i \leq n_d, \alpha > 0, \ C \text{ independent of } N,$$

then for sufficiently large N,

$$\left| D\left(\pi_N^d \| \pi^d\right) \right| \lesssim N^{-\alpha}.$$

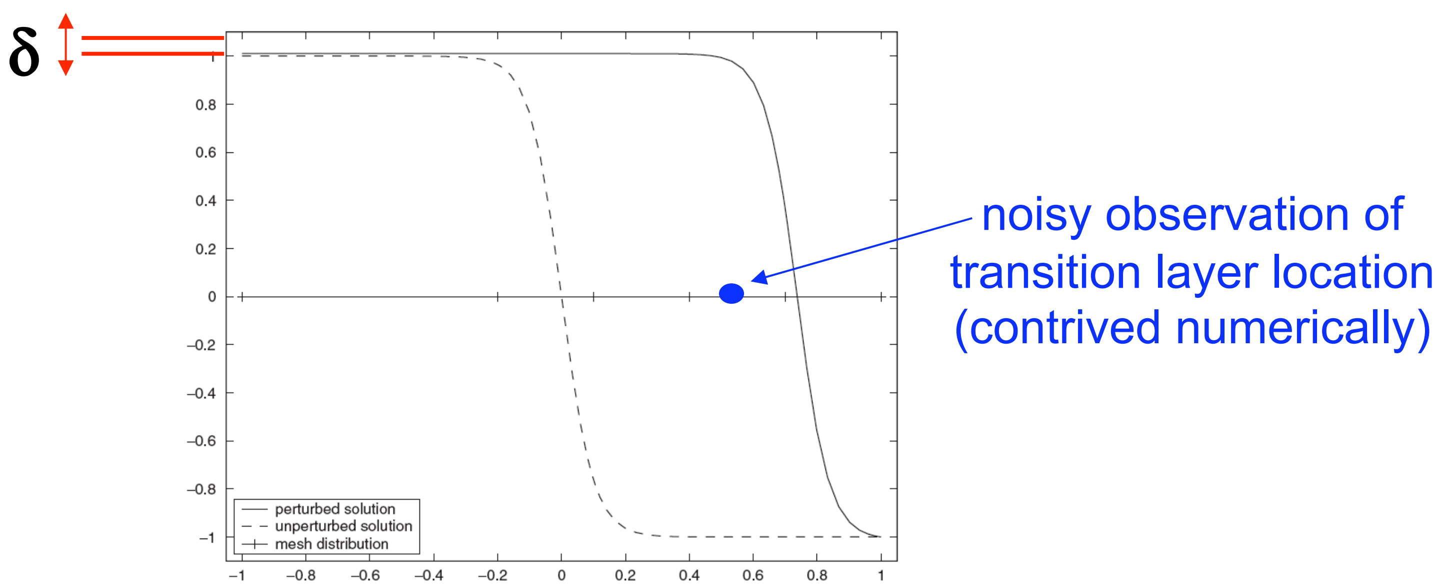
- **Notes:**
  - Fast (exponential) convergence rate is retained
  - So is the slow convergence

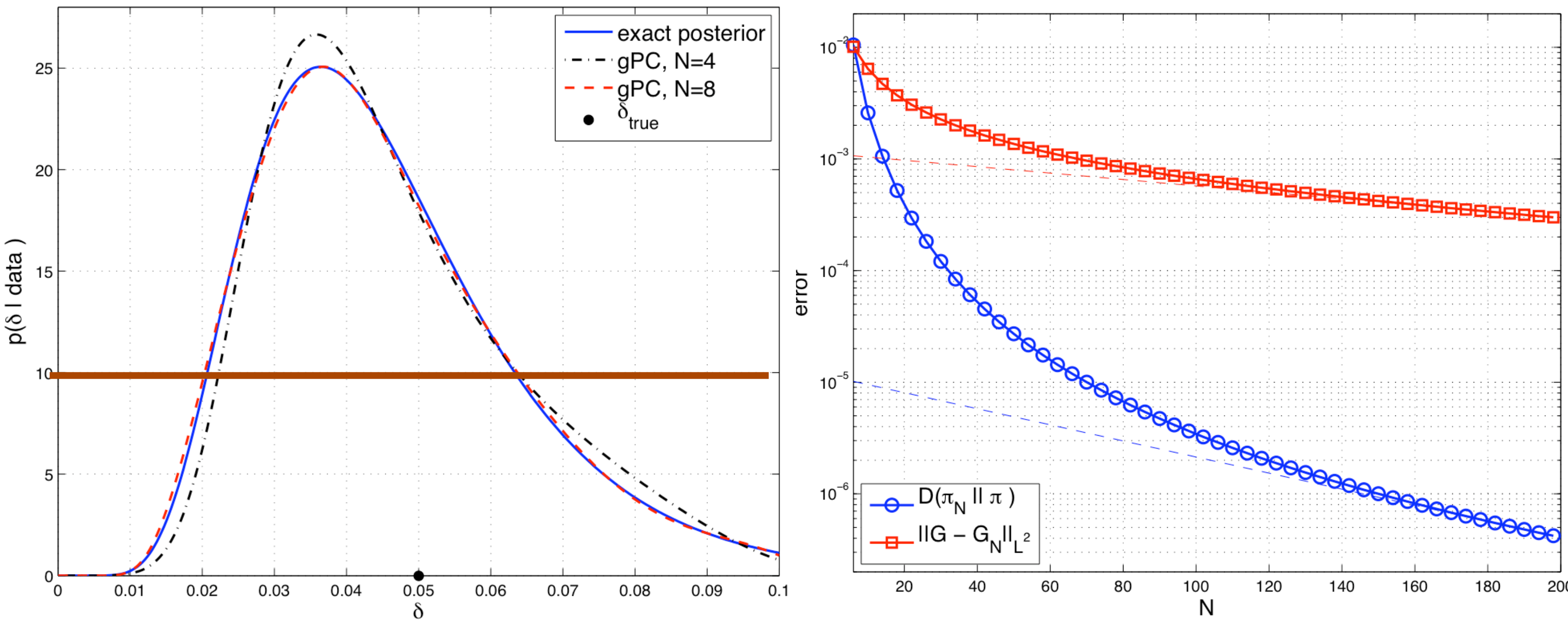
(Marzouk & Xiu, Comm. Comput. Phys.,vol. 6, 08)

Parameter Estimation: Supersensitivity Example

- Burgers' equation : $$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}, \quad x \in [-1,1]$$

- Boundary conditions : $$u(-1) = 1 + \delta(Z); \quad u(1) = -1; \quad 0 < \delta \ll 1$$



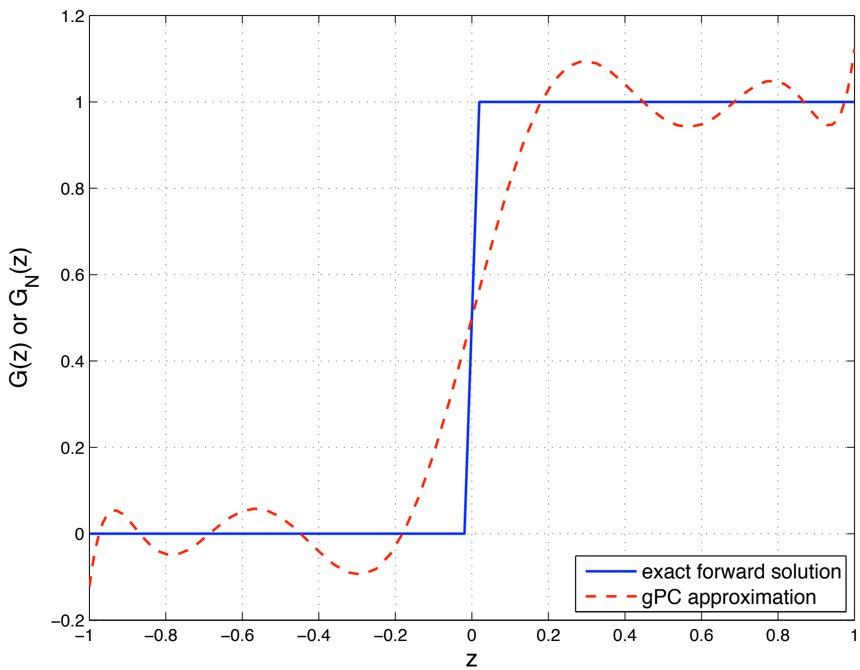


Prior distribution is uniform *Z~(0,0.1)*

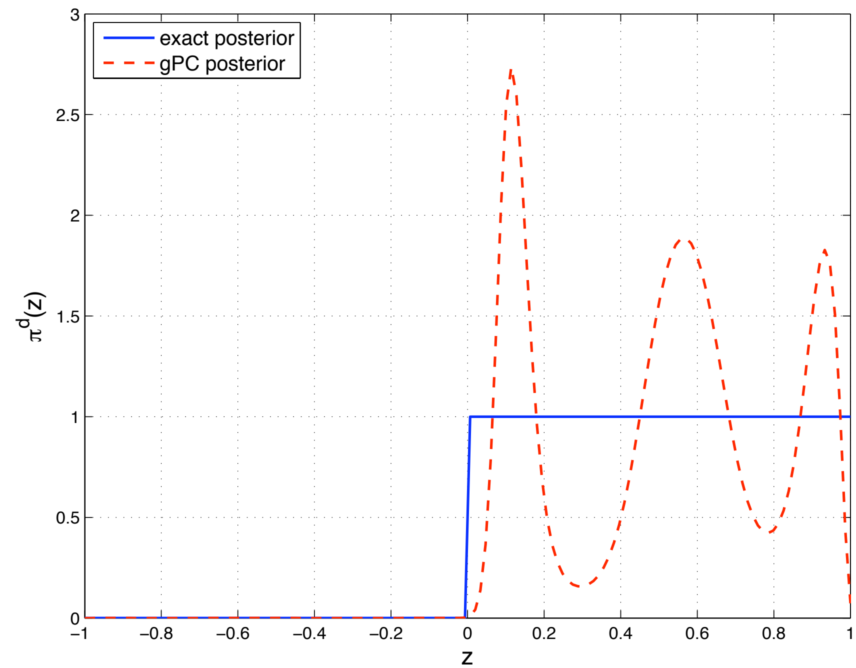
Measurement noise: *e~N(0,0.05²)*

Parameter Estimation: Step Function

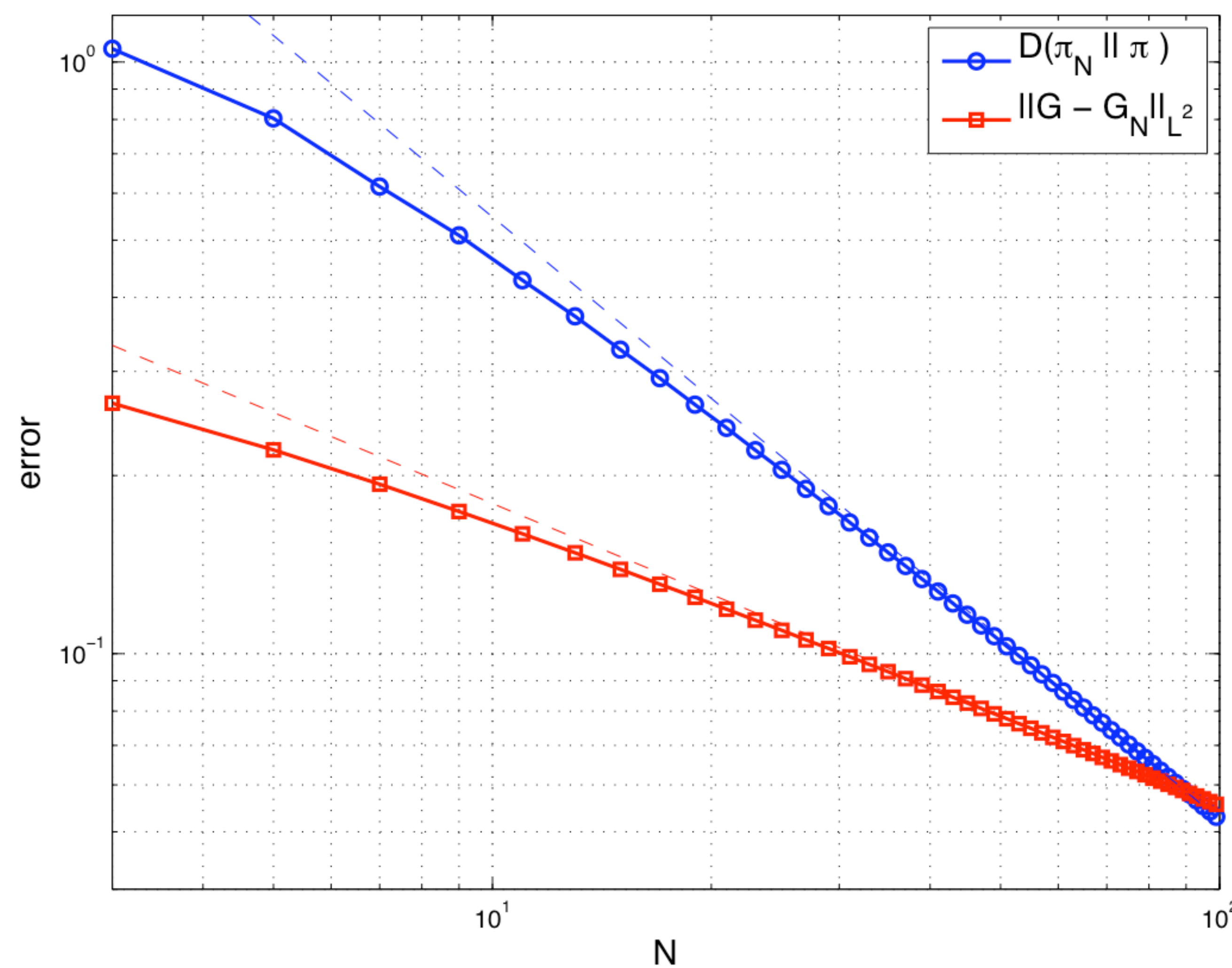
- Assume the forward model is a step function
- Posterior distribution is discontinuous
- Gibb’s oscillations exist
- Slow convergence with global gPC basis functions



Forward model and its approximation



Posterior distribution and its approximation



D(π_N || π)
||G − G_N||_{L^2}
error
N
10^0
10^{−1}
10^1
10^2

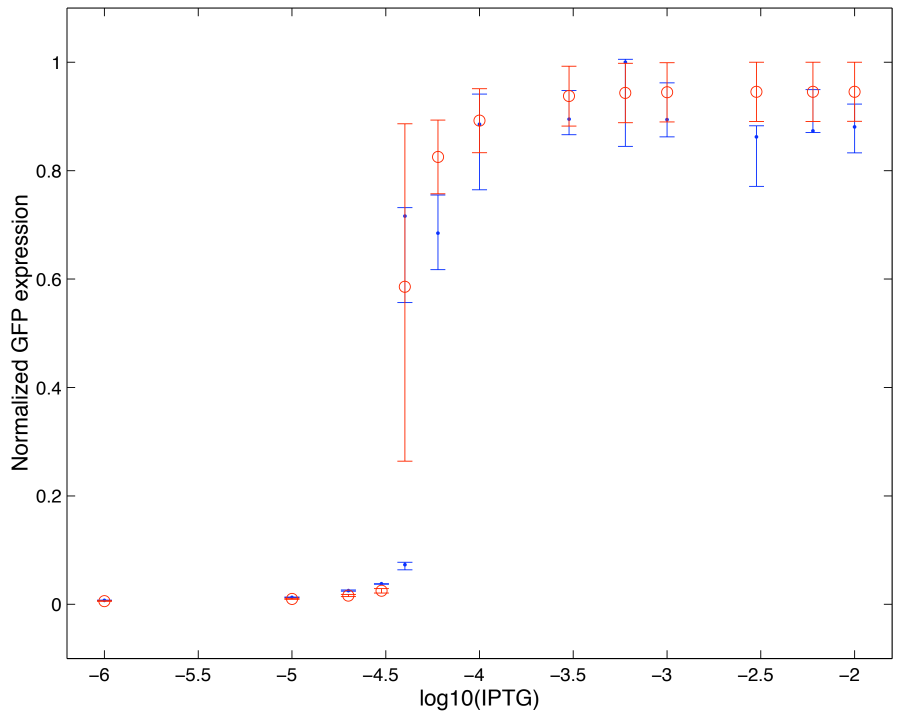
Biological Application

- **Example:** estimate kinetic parameters in a *genetic toggle switch*
  - Differential-algebraic equations model from [Gardner *et al*, Nature, 2000]
  - **Real experimental data**: steady-state expression levels of one gene (v)

$$\frac{du}{dt} = \frac{\alpha_1}{1+v^{\beta}} - u$$

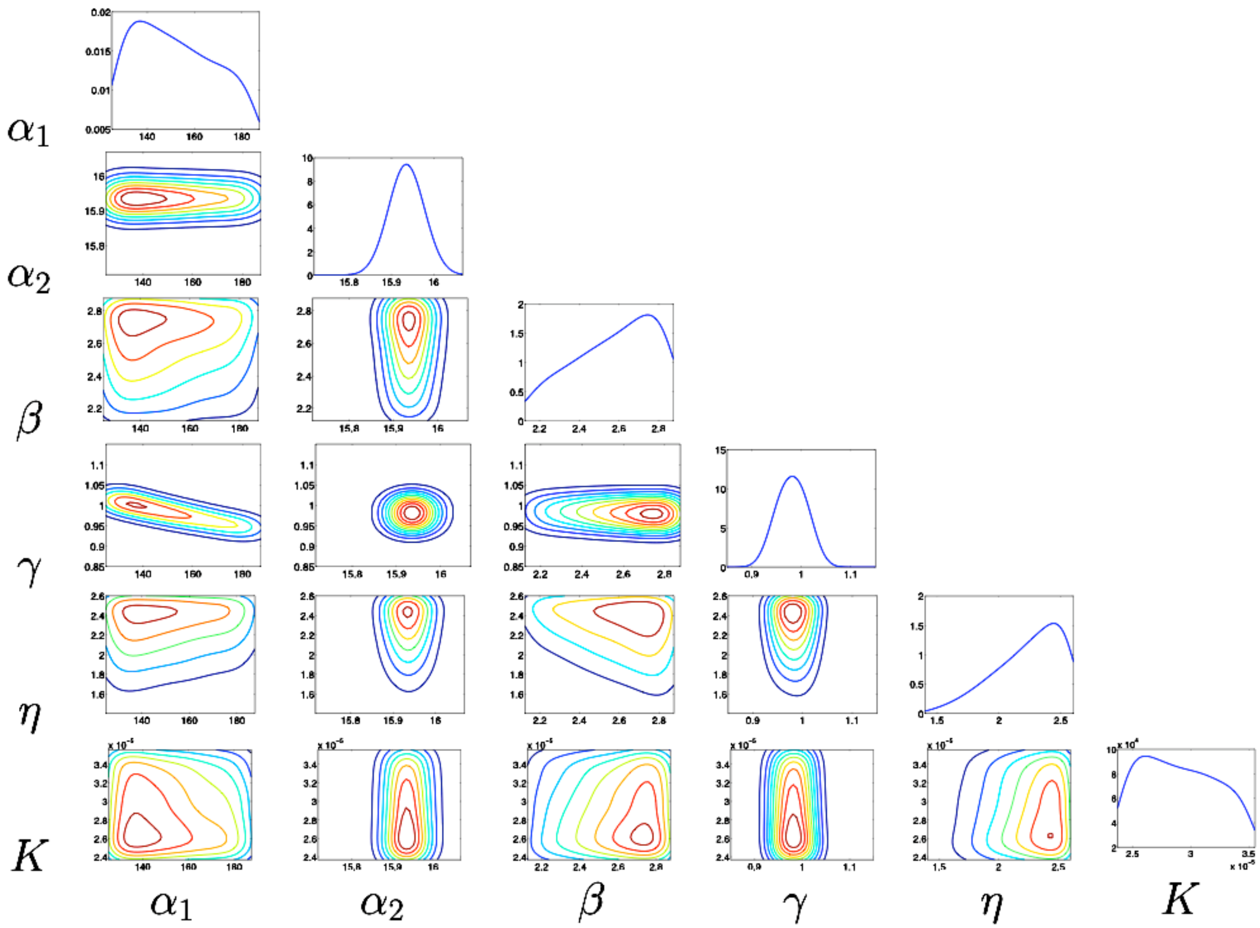
$$\frac{dv}{dt} = \frac{\alpha_2}{1+w^{\gamma}} - v$$

$$w = \frac{u}{(1+[IPTG]/K)^{\eta}}$$



- 6 uncertain parameters
- Assumed to be independent and uniform in the forward problem
- Good agreement with measurement --- but let's now use the data *again*

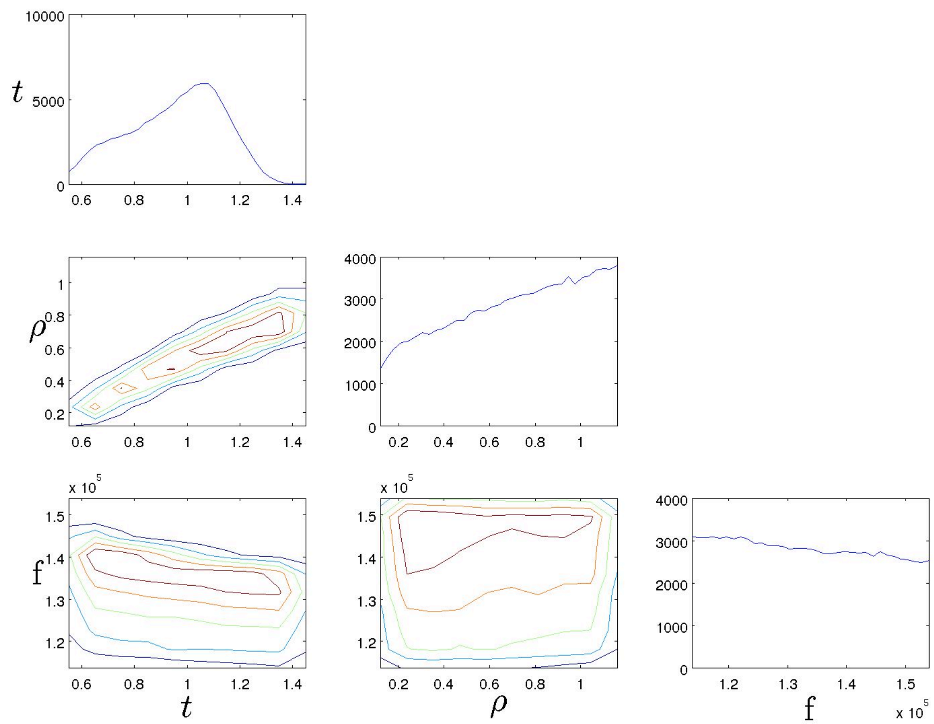
Using the Data Again – Parameter estimation

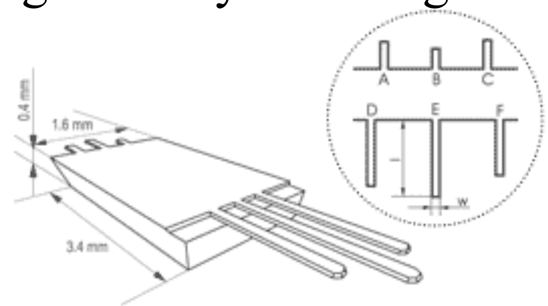


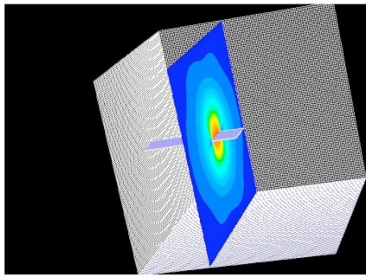
1-parameter and 2-parameter marginal posterior densities
(*dim* = 6, *N* = 4, 6-level **sparse grid** forward problem solver)

An (almost) Industrial Application: Free-Cantilever Damping

Gas damping of freely-vibrating cantilever

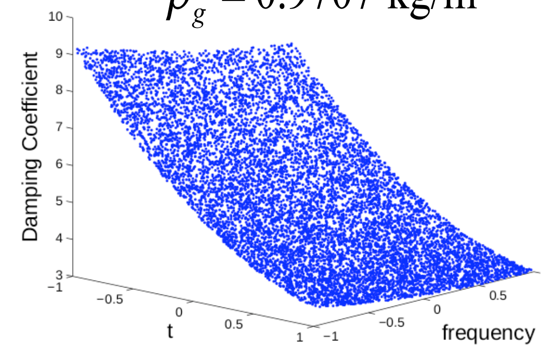






MEMOSA for damping simulation

$\rho_g = 0.9707 \text{ kg/m}^3$



3-parameter marginal posterior densities
Level 2 sparse grid in thickness, gas density and frequency

Summary

- Uncertainty Analysis: To provide improved prediction
  - Input characterization
  - Uncertainty propagation
  - Post processing
- Generalized polynomial chaos (gPC)
  - Multivariate approximation theory
  - An highly efficient uncertainty propagation strategy
  - Makes many UQ tasks "post processing"
- **Data, *any* data, can help**
  - UQ simulation needs to be dynamically data driven.

- Support:
  - AFOSR: FA9550-08-1-0353
  - DOE:
    - DE-FC52-08NA28617
    - DE-SC0005713
  - NSF:
    - DMS-0645035
    - IIS-0914447
    - IIS-1028291