SAND2014-19420PE

Uncertainty Quantification in Computational Models

Habib N. Najm

hnnajm@sandia.gov

Sandia National Laboratories
Livermore, CA, USA

Auburn University
Auburn, MA
Sept. 19, 2014

Acknowledgement

B.J. Debusschere, R.D. Berry, K. Sargsyan, C. Safta,
K. Chowdhary, M. Khalil — Sandia National Laboratories, CA

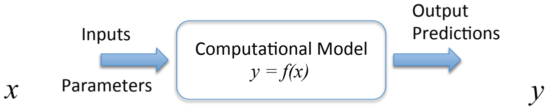
R.G. Ghanem — U. South. California, Los Angeles, CA
O.M. Knio — Duke Univ., Durham, NC
O.P. Le Maître — CNRS, Paris, France
Y.M. Marzouk — Mass. Inst. of Tech., Cambridge, MA

This work was supported by:

- DOE Office of Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC)
- DOE ASCR Applied Mathematics program.
- DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.

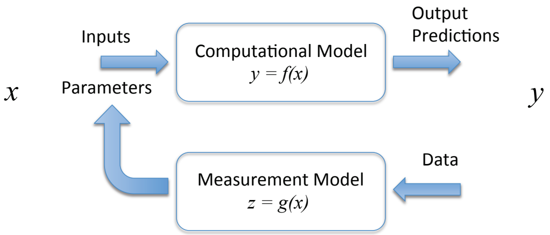
Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94-AL85000.

Uncertainty Quantification and Computational Science



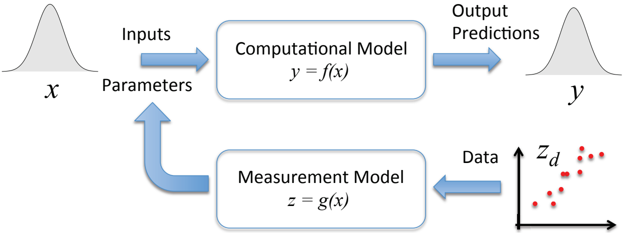
Forward problem

Uncertainty Quantification and Computational Science



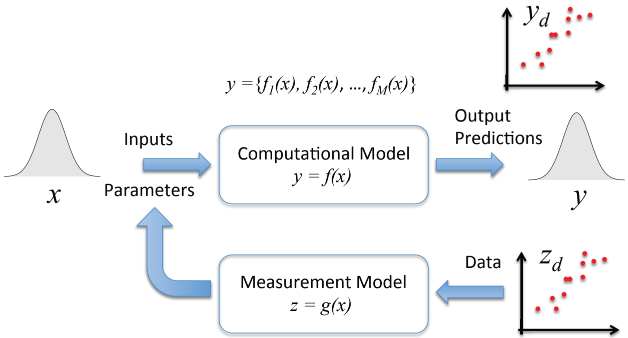
Inverse & Forward problems

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Model validation & comparison, Hypothesis testing

Outline

Forward propagation of parametric uncertainty

Forward model: $y = f(x)$

- Local sensitivity analysis (SA) and error propagation

$$\Delta y = \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} \Delta x$$

  This is ok for:

  – small uncertainty
  – low degree of non-linearity in $f(x)$

- Non-probabilistic methods
  - Fuzzy logic
  - Evidence theory – Dempster-Shafer theory
  - Interval math

- Probabilistic methods – this is our focus

Probabilistic Forward UQ – $y = f(x)$

Represent uncertain quantities using probability theory

- Random sampling, MC, QMC
  - Generate random samples $\{x^i\}_{i=1}^N$ from the PDF of x, $p(x)$
  - Bin the corresponding $\{y^i\}$ to construct $p(y)$
  - Not feasible for computationally expensive $f(x)$
    - slow convergence of MC/QMC methods
      $\Rightarrow$ very large N required for reliable estimates
- Build a cheap surrogate for $f(x)$, then use MC
  - Collocation – interpolants
  - Regression – fitting
- Galerkin methods
  - Polynomial Chaos (PC)
  - Intrusive and non-intrusive PC methods

Probabilistic Forward UQ & Polynomial Chaos Representation of Random Variables

With $y = f(x)$, x a random variable, estimate the RV y

- Can describe a RV in terms of its
  - density, moments, characteristic function, or
  - as a function on a probability space
- Constraining the analysis to RVs with finite variance
  - $\Rightarrow$ Represent RV as a spectral expansion in terms of orthogonal functions of standard RVs
    - Polynomial Chaos Expansion
- Enables the use of available functional analysis methods for forward UQ

Polynomial Chaos Expansion (PCE)

- Model uncertain quantities as random variables (RVs)
- Given a *germ* $\boldsymbol{\xi}(\omega) = \{\xi_1, \cdots, \xi_n\}$ – a set of *i.i.d.* RVs
  - where $p(\boldsymbol{\xi})$ is uniquely determined by its moments

Any RV in $L^2(\Omega, \mathfrak{S}(\boldsymbol{\xi}), P)$ can be written as a PCE:

$$u(\boldsymbol{x}, t, \omega) = f(\boldsymbol{x}, t, \boldsymbol{\xi}) \simeq \sum_{k=0}^{P} u_k(\boldsymbol{x}, t)\Psi_k(\boldsymbol{\xi}(\omega))$$

- $u_k(\boldsymbol{x}, t)$ are mode strengths
- $\Psi_k()$ are multivariate functions orthogonal w.r.t. $p(\boldsymbol{\xi})$

With dimension n and order p: $$P + 1 = \frac{(n+p)!}{n!p!}$$

Orthogonality

By construction, the functions $\Psi_k()$ are orthogonal with respect to the density of ξ

$$u_k(\boldsymbol{x}, t) = \frac{\langle u\Psi_k \rangle}{\langle \Psi_k^2 \rangle} = \frac{1}{\langle \Psi_k^2 \rangle} \int u(\boldsymbol{x}, t; \lambda(\boldsymbol{\xi}))\Psi_k(\boldsymbol{\xi})\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

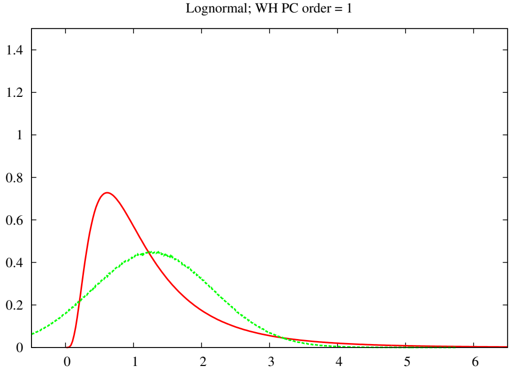
Examples:

- Hermite polynomials with Gaussian basis
- Legendre polynomials with Uniform basis, ...
- Global versus Local PC methods
  - Adaptive domain decomposition of the support of ξ

PC Illustration: WH PCE for a Lognormal RV

- Wiener-Hermite PCE constructed for a Lognormal RV
- PCE-sampled PDF superposed on true PDF
- Order = 1

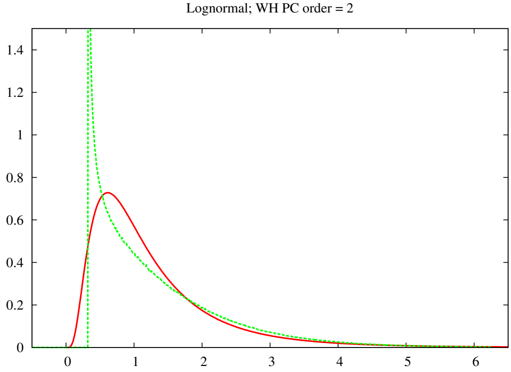
$$u = \sum_{k=0}^{P} u_k \Psi_k(\xi) = u_0 + u_1 \xi$$



PC Illustration: WH PCE for a Lognormal RV

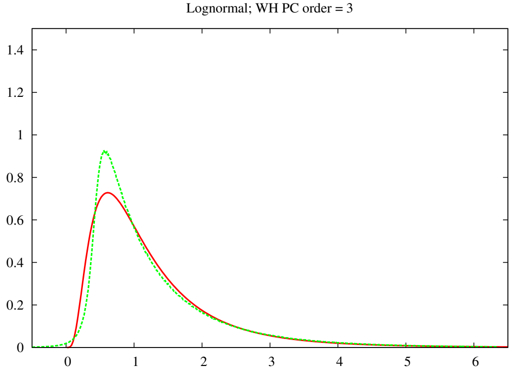
- Wiener-Hermite PCE constructed for a Lognormal RV
- PCE-sampled PDF superposed on true PDF
- Order = 2

$$
\begin{aligned}
u &= \sum_{k=0}^{P} u_k \Psi_k(\xi) \\
&= u_0 + u_1 \xi + u_2(\xi^2 - 1)
\end{aligned}
$$



PC Illustration: WH PCE for a Lognormal RV

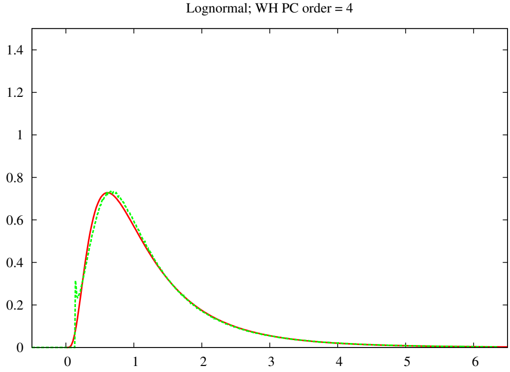
- Wiener-Hermite PCE constructed for a Lognormal RV
- PCE-sampled PDF superposed on true PDF
- Order = 3



$$
\begin{aligned}
u &= \sum_{k=0}^{P} u_k \Psi_k(\xi) \\
&= u_0 + u_1 \xi + u_2(\xi^2 - 1) + u_3(\xi^3 - 3\xi)
\end{aligned}
$$

PC Illustration: WH PCE for a Lognormal RV

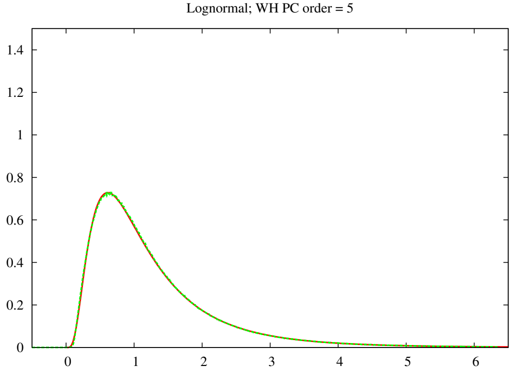
- Wiener-Hermite PCE constructed for a Lognormal RV
- PCE-sampled PDF superposed on true PDF
- Order = 4



$$
\begin{aligned}
u &= \sum_{k=0}^{P} u_k \Psi_k(\xi) \\
&= u_0 + u_1\xi + u_2(\xi^2 - 1) + u_3(\xi^3 - 3\xi) + u_4(\xi^4 - 6\xi^2 + 3)
\end{aligned}
$$

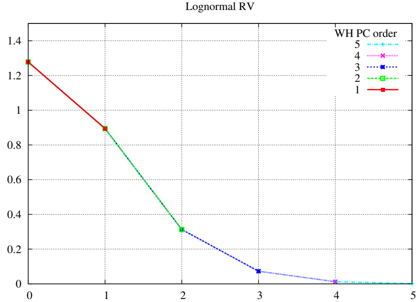
PC Illustration: WH PCE for a Lognormal RV

- Wiener-Hermite PCE constructed for a Lognormal RV
- PCE-sampled PDF superposed on true PDF
- Order = 5



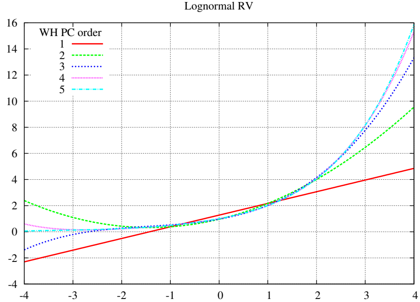
$$
\begin{aligned}
u &= \sum_{k=0}^{P} u_k \Psi_k(\xi) \\
&= u_0 + u_1\xi + u_2(\xi^2 - 1) + u_3(\xi^3 - 3\xi) + u_4(\xi^4 - 6\xi^2 + 3) \\
&\quad + u_5(\xi^5 - 10\xi^3 + 15\xi)
\end{aligned}
$$

PC Illustration: WH PCE for a Lognormal RV



PC mode amplitudes

u_0–u_5



PC function $u(\xi)$

Order 1–5

- Fifth-order Wiener-Hermite PCE represents the given Lognormal well
- Higher order terms are negligible

Essential Use of PC in UQ

Strategy:

- Represent model parameters/solution as random variables
- Construct PCEs for uncertain parameters
- Evaluate PCEs for model outputs

Advantages:

- Computational efficiency
- Utility
  - Moments: $\mathrm{E}(u) = u_0, \ \mathrm{var}(u) = \sum_{k=1}^{P} u_k^2 \langle \Psi_k^2 \rangle, \ \ldots$
  - Global Sensitivities – fractional variances, Sobol' indices
  - Surrogate for forward model

Requirement:

- RVs in L^2, *i.e.* with finite variance, on $(\Omega, \mathfrak{S}(\xi), P)$

Intrusive PC UQ: A direct *non-sampling* method

- Given model equations: $\mathcal{M}(u(\boldsymbol{x},t);\lambda) = 0$
- Express uncertain parameters/variables using PCEs

$$u = \sum_{k=0}^{P} u_k \Psi_k; \quad \lambda = \sum_{k=0}^{P} \lambda_k \Psi_k$$

- Substitute in model equations; apply Galerkin projection
- New set of equations: $\mathcal{G}(U(\boldsymbol{x},t),\Lambda) = 0$
  - with $U = [u_0, \ldots, u_P]^T$, $\Lambda = [\lambda_0, \ldots, \lambda_P]^T$
- Solving this deterministic system once provides the full specification of uncertain model ouputs

Intrusive Galerkin PC ODE System

$$\frac{du}{dt} = f(u;\lambda)$$

$$\lambda = \sum_{i=0}^{P} \lambda_i \Psi_i \qquad u(t) = \sum_{i=0}^{P} u_i(t)\Psi_i$$

$$\frac{du_i}{dt} = \frac{\langle f(u;\lambda)\Psi_i\rangle}{\langle \Psi_i^2\rangle} \qquad i = 0,\ldots,P$$

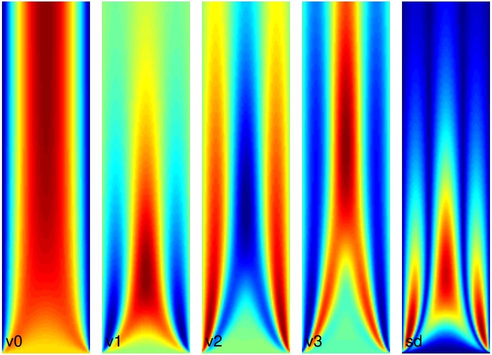
Say $f(u;\lambda) = \lambda u$, then

$$\frac{du_i}{dt} = \sum_{p=0}^{P}\sum_{q=0}^{P} \lambda_p u_q C_{pqi}, \quad i = 0,\cdots,P$$

where the tensor $C_{pqi} = \langle \Psi_p \Psi_q \Psi_i\rangle / \langle \Psi_i^2\rangle$ is readily evaluated

Laminar 2D Channel Flow with Uncertain Viscosity

- Incompressible flow
- Viscosity PCE
  - $\nu = \nu_0 + \nu_1 \xi$
- Streamwise velocity
  - $\mathrm{v} = \sum_{i=0}^{P} \mathrm{v}_i \Psi_i$
  - v_0: mean
  - v_i: i-th order mode
  - $\sigma^2 = \sum_{i=1}^{P} \mathrm{v}_i^2 \langle \Psi_i^2 \rangle$



v_0 v_1 v_2 v_3 σ

(Le Maître *et al.*, J. Comput. Phys., 2001)

Intrusive PC UQ Pros/Cons

Cons:

- Reformulation of governing equations
- New discretizations
- New numerical solution method
  - Consistency, Convergence, Stability
  - Global vs. multi-element local PC constructions
- New solvers and model codes
  - Opportunities for automated code transformation
- New preconditioners

Pros:

- Tailored solvers <u>can</u> deliver superior performance

Stability of Intrusive Galerkin ODEs

- Setting:
  - Nonlinear ODEs with uncertain initial conditions/parameters
- Intrusive Galerkin ODE system can be unstable in general
  - Spurious positive eigenvalues
  - Fast growth of PC coefficients
- Equation structure similar to that of semidiscretized spectral Galerkin/Fourier equations for conservation laws
  - Generally unstable for nonlinear conservation laws
  - Filtering is useful to ensure stability
- Stability achievable with local multi-element PC methods
- Need to develop stable global PC methods – filtering
- Application setting:
  - Uncertain chemical kinetics model – homogeneous ignition

A chemically-relevant model ODE system

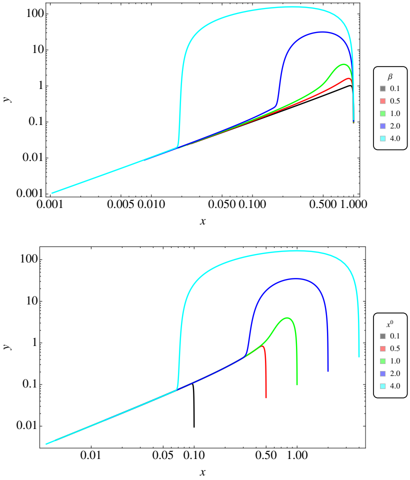
State vector $(x(t), y(t))$

$$\begin{aligned}\frac{dx}{dt} &= -x(1+y)\\ \frac{dy}{dt} &= \frac{1}{\epsilon}(x - \gamma y + \beta x y)\\ x(0) &= x^0\\ y(0) &= y^0\end{aligned}$$

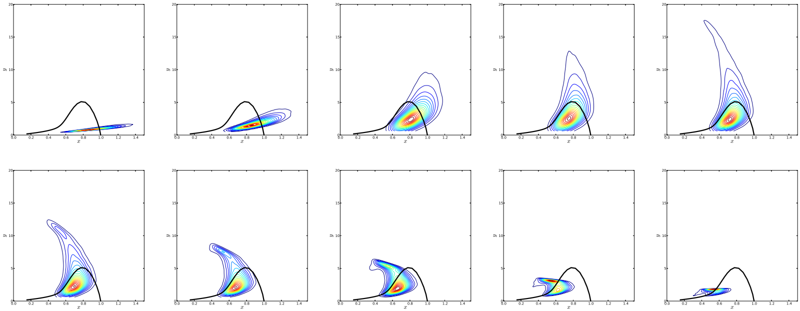
$\epsilon, \gamma > 0$, $\beta \geq 0$
$x(t), y(t) \geq 0$

$\epsilon << 1 \Rightarrow$ stiff

- x: slow variable (stable species)
- y: fast variable (radical species)



Uncertain model chemical system



$y^0 = 0, \gamma = 1, \epsilon = 0.01$
x^0, β : Uncertain, independent, lognormal

$\mathrm{E}[\beta] = 1.0, \quad \mathrm{Stdv}[\beta] = 0.202$
$\mathrm{E}[x^0] = 1.0, \quad \mathrm{Stdv}[x^0] = 0.202, 0.343, 0.372$ (Case: 1,2,3)

Intrusive Galerkin global PC ODE system

$\boldsymbol{\xi} \equiv (\xi_1, \xi_2)$ *i.i.d.* N(0,1)

$$x^0(\boldsymbol{\xi}) = \sum_k x_k^0 \Psi_k(\boldsymbol{\xi}) = x^0(\xi_1)$$
$$\beta(\boldsymbol{\xi}) = \sum_k \beta_k \Psi_k(\boldsymbol{\xi}) = \beta(\xi_2)$$

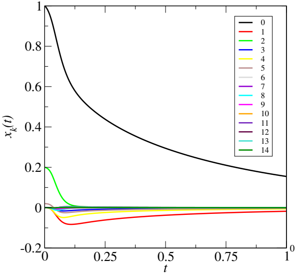
$$x(t, \boldsymbol{\xi}) = \sum_k x_k(t) \Psi_k(\boldsymbol{\xi})$$
$$y(t, \boldsymbol{\xi}) = \sum_k y_k(t) \Psi_k(\boldsymbol{\xi}).$$

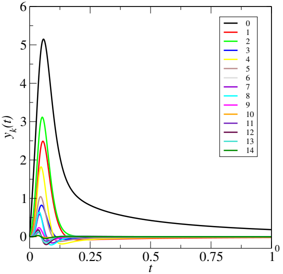
Galerkin ODE system: $k = 0, \dots, P$,

$$\dot{x}_k = -x_k - \sum_{i,j} x_i y_j \; C_{ijk}$$
$$\dot{y}_k = \epsilon^{-1}(x_k - \gamma y_k + \sum_{i,j,\ell} x_i y_j \beta_\ell \; C_{ij\ell k})$$

where

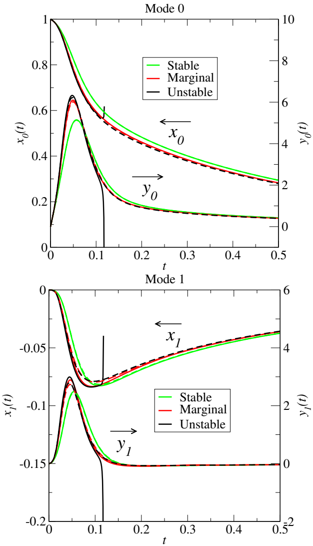
$$C_{ijk} \equiv \langle \Psi_i \Psi_j \Psi_k \rangle / \langle \Psi_k^2 \rangle$$
$$C_{ij\ell k} \equiv \langle \Psi_i \Psi_j \Psi_\ell \Psi_k \rangle / \langle \Psi_k^2 \rangle$$





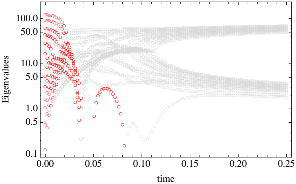
Galerkin system instability for high σ_{x^0}

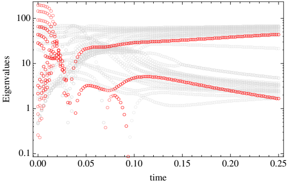
- Fourth order PC solution
- Implicit stiff time integrator
  - **DVODE**
- Consequences of increased uncertainty in x^0
  - Case 1 – stable
  - Case 2 – marginal
  - Case 3 – unstable
- Increasing PC order makes the problem worse
  - Unstable for lower degrees of uncertainty

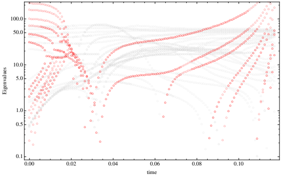


Spurious persistent growth of positive eigenvalues

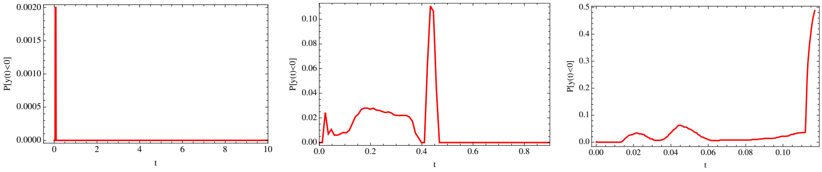
- Physically meaningful positive eigenvalues during brief phases of system dynamics
  - Initial fast time evolution of the system
  - Brief phase of explosive growth typical of ignition
  - Growth of initial uncertainty
- Marginal case exhibits persistent positive eigenvalues
- Unstable case exhibits fast growth of positive eigenvalues







Instability consequences on $p(x, y)$



Case 1 Case 2 Case 3

- Probability density function $p(x(t), y(t))$ based on sampling PCEs of $(x(t), y(t))$
- Instability leads to large/increasing $P(Y < 0)$
  - Unphysical and inconsistent with sampled solution PDF

- Employ filtering that minimizes $\max(P_{X<0}, P_{Y<0})$

Filter Design

Given a PCE $\tilde{u}(\xi) = \sum_{k=0}^{P} \tilde{u}_k \Psi_k(\xi)$.

Then, with $U \equiv (u_0, u_1, \ldots, u_P)^T$, and $u = \sum_k u_k \Psi_k(\xi) = u(U, \xi)$,

seek filtered PCE $\hat{u}(\xi) = \sum_{k=0}^{P} \hat{u}_k \Psi_k(\xi)$, where

$$\hat{U} = \underset{U}{\text{argmin}}\{\|u - \tilde{u}\|_2^2 + w\ \Phi_N(U)\}$$

with

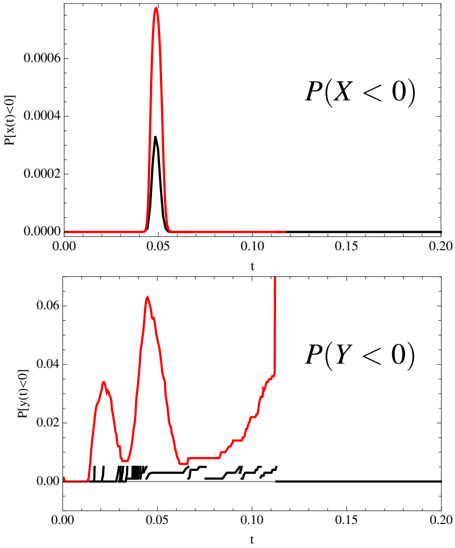
$$\Phi_N(U) \equiv \frac{1}{N} \sum_{i=1}^{N} (1 - H(u(U, \xi^i))) \approx P(u < 0)$$

where w is a chosen weight factor

- Filter applied at any time step where $\Phi_N > \Phi_{\text{thr}}$
- Procedure finds $\hat{u}(\xi)$ that is both
  - near to $\tilde{u}(\xi)$, and has minimal Φ_N

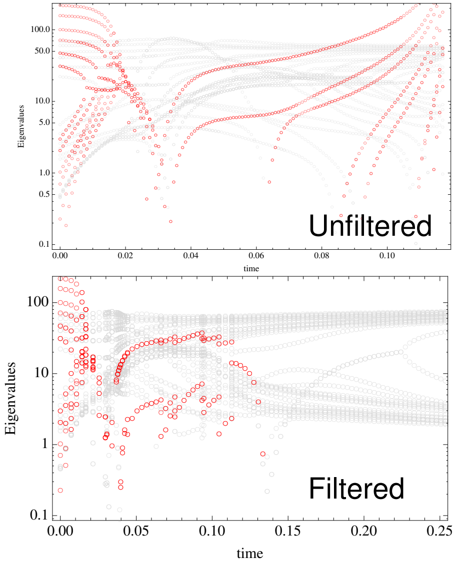
Filtered system reduction in $P_{\text{neg}} = \max[P(X < 0), P(Y < 0)]$

- Growth in P_{neg} is halted
- Filtered time integration maintains $P_{\text{neg}} < \Phi_{\text{thr}}$



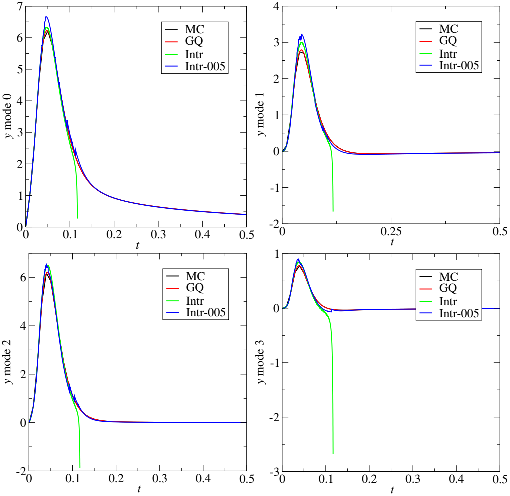
Filter eliminates spurious positive eigenvalues

- Filtered system has positive eigenvalues over a limited time range
- Persistent and growing spurious positive eigenvalues eliminated



Filtered Galerkin ODE system is Stable

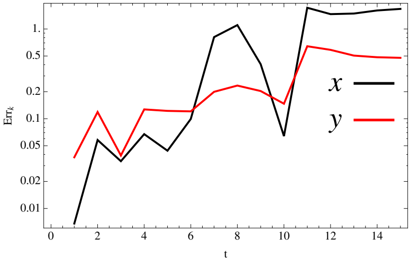
- Stable time evolution of PC mode coefficients
- Filtered system intrusive solution consistent with that based on non-intrusive Monte Carlo and Gauss Quadrature results



Filtered system PC mode errors

- PC mode errors are higher in the higher-order terms
- $\mathcal{O}(10\%)$ normalized RMS errors in low-order modes
- Other challenges:

  Filtered system stability impacted by
  - degree of uncertainty
  - PC order
  - P_{neg} threshold ϕ_{thr}



Error between PC mode coefficients from the intrusive Galerkin solution and the non-intrusive Gauss-Quadrature solution

Non-intrusive PC UQ

- *Sampling*-based
- Relies on black-box utilization of the computational model
- Evaluate projection integrals *numerically*
- For any quantity of interest $\phi(\boldsymbol{x}, t; \lambda) = \sum_{k=0}^{P} \phi_k(\boldsymbol{x}, t)\Psi_k(\boldsymbol{\xi})$

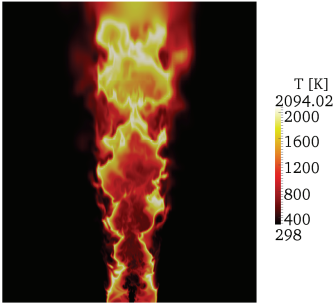
$$\phi_k(\boldsymbol{x}, t) = \frac{1}{\langle \Psi_k^2 \rangle} \int \phi(\boldsymbol{x}, t; \lambda(\boldsymbol{\xi}))\, \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \quad k = 0, \ldots, P$$

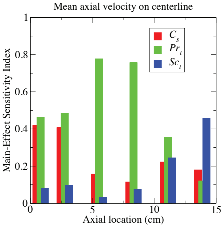
- Integrals can be evaluated using
  - A variety of (Quasi) Monte Carlo methods
    - Slow convergence; $\sim$ indep. of dimensionality
  - Quadrature/Sparse-Quadrature methods
    - Fast convergence; depends on dimensionality

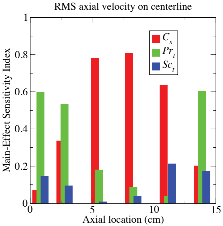
UQ in LES computations: turbulent bluff-body flame

with M. Khalil, G. Lacaze, & J. Oefelein, Sandia Nat. Labs

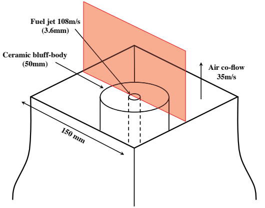
- CH_4-H_2 jet, air coflow, 3D flow
- Re=9500, LES subgrid modeling
- 12×10^6 mesh cells, 1024 cores
- 3 days run time, 2×10^5 time steps
- 3 uncertain parameters (C_s, Pr_t, Sc_t)
- 2^{nd}-order PC, 25 sparse-quad. pts







Main-Effect Sensitivity Indices



J. Oefelein & G. Lacaze, SNL

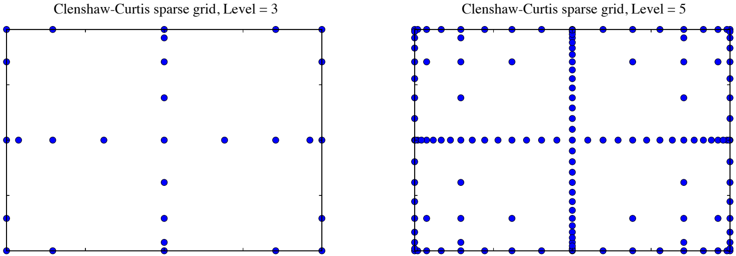
PC and High-Dimensionality

Dimensionality n of the PC basis: $\boldsymbol{\xi} = \{\xi_1, \dots, \xi_n\}$

- $n \approx$ number of uncertain parameters
- $P + 1 = (n+p)!/n!p!$ grows fast with n

Impacts:

- Size of intrusive PC system
- Hi-D projection integrals $\Rightarrow$ large # non-intrusive samples
  - Sparse quadrature methods



PC Sparse Quadrature in hiD – Climate land model

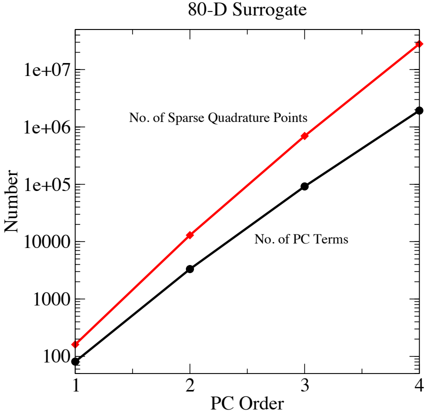
Full quadrature: $N = (N_{1D})^n$

Sparse Quadrature

- Wide range of methods
- Nested & hierarchical
- Clenshaw-Curtis: $N = \mathcal{O}(n^p)$
- Adaptive – greedy algorithms

Number of points can still be excessive in hi-D

- Large no. of terms
- Reduction/sparsity



Other non-intrusive methods

- Response surface employing PC or other functional basis
- Collocation: Fit interpolant to samples
  - Oscillation concern in multi-D
- Regression: Estimate best-fit response surface
  - Least-squares
    - Sparsity via ℓ_1 constraints; compressive sensing
  - Bayesian inference
    - Sparsity via Laplace priors; Bayesian compressive sensing
  - Useful when quadrature methods are infeasible, e.g.:
    - Samples given *a priori*
    - Can't choose sample locations
    - Can't take enough samples
    - Forward model is noisy

Challenges in PC UQ – Non-Linearity

- Bifurcative response at critical parameter values
  - Rayleigh-Bénard convection
  - Transition to turbulence
  - Chemical ignition
- Discontinuous $u(\lambda(\boldsymbol{\xi}))$
  - Failure of global PCEs in terms of smooth $\Psi_k()$
  - $\Leftrightarrow$ failure of Fourier series in representing a step function
- Local PC methods
  - Subdivide support of $\lambda(\boldsymbol{\xi})$ into regions of smooth $u \circ \lambda(\boldsymbol{\xi})$
  - Employ PC with compact support basis on each region
  - A spectral-element vs. spectral construction
  - Domain mapping

Challenges in PC UQ – Time Dynamics

- Systems with limit-cycle or chaotic dynamics
- Large amplification of phase errors over long time horizon
- PC order needs to be increased in time to retain accuracy
- Time shifting/scaling remedies

- Futile to attempt representation of detailed turbulent velocity field $\boldsymbol{v}(\boldsymbol{x}, t; \lambda(\boldsymbol{\xi}))$ as a PCE
  - Fast loss of correlation due to energy cascade
  - Problem studied in 60's and 70's
- Focus on flow statistics, *e.g.* Mean/RMS quantities
  - Well behaved
  - Argues for non-intrusive methods with DNS/LES of turbulent flow

Inverse UQ – Estimation of Uncertain Parameters

Forward UQ requires specification of uncertain inputs

Probabilistic setting

- Require joint PDF on input space
- Statistical inference – an inverse problem

Bayesian setting

- Given Data: PDF on uncertain inputs can be estimated using Bayes formula
  - Bayesian Inference
- Given Constraints: PDF on uncertain inputs can be estimated using the Maximum Entropy principle
  - MaxEnt Methods

Bayes formula for Parameter Inference

- Data Model (fit model + noise model): $y = f(\lambda) * g(\epsilon)$
- Bayes Formula:

$$p(\lambda, y) = p(\lambda|y)p(y) = p(y|\lambda)p(\lambda)$$

$$\underbrace{p(\lambda|y)}_{\text{Posterior}} = \frac{\overbrace{p(y|\lambda)}^{\text{Likelihood}}\ \overbrace{p(\lambda)}^{\text{Prior}}}{\underbrace{p(y)}_{\text{Evidence}}}$$

- Prior: knowledge of λ prior to data
- Likelihood: forward model and measurement noise
- Posterior: combines information from prior and data
- Evidence: normalizing constant for present context

The Prior

- Prior $p(\lambda)$ comes from
  - Physical constraints
  - Prior data
  - Prior knowledge
- The prior can be **uninformative**
- It can be chosen to impose **regularization**
- Unknown aspects of the prior can be added to the rest of the parameters as hyperparameters
- The choice of prior can be crucial when there is little information in the data relative to the number of degrees of freedom in the inference problem
- When there is sufficient information in the data, the data can overrule the prior

Construction of the Likelihood $p(y|\lambda)$

- Where does probability enter the mapping $\lambda \to y$ in $p(y|\lambda)$?
- Through a presumed error model:
- Example:
  - Model:

$$y_m = g(\lambda)$$

  - Data: y
  - Error between data and model prediction: ϵ

$$y = g(\lambda) + \epsilon$$

- Model this error as a random variable
- Example
  - Error is due to instrument measurement noise
  - Instrument has Gaussian errors, with no bias

$$\epsilon \sim N(0, \sigma^2)$$

Construction of the Likelihood $p(y|\lambda)$ – cont'd

For any given λ, this implies

$$y|\lambda,\sigma \sim N(g(\lambda),\sigma^2)$$

or

$$p(y|\lambda,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(y-g(\lambda))^2}{2\sigma^2}\right)$$

Given N measurements $(y_1,\ldots,y_N)$, and presuming independent identically distributed (*iid*) noise

$$\begin{aligned} y_i &= g(\lambda)+\epsilon_i \\ \epsilon_i &\sim N(0,\sigma^2) \\ L(\lambda)=p(y_1,\ldots,y_N|\lambda,\sigma) &= \prod_{i=1}^{N} p(y_i|\lambda,\sigma) \end{aligned}$$

Likelihood Modeling

- This is frequently the *core* modeling challenge
  - Error model: a statistical model for the discrepancy between the forward model and the data
  - composition of the error model with the forward model
- Error model composed of discrepancy between
  - data and the truth – (data error)
  - model prediction and the truth – (model error)
- Mean bias and correlated/uncorrelated noise structure
- Hierarchical Bayes modeling, and dependence trees

$$p(\phi, \theta | D) = p(\phi | \theta, D) p(\theta | D)$$

- Choice of observable – constraint on Quantity of Interest?

Posterior

$$p(\lambda|y) \propto p(y|\lambda)p(\lambda)$$

Continuing the above *iid* Gaussian likelihood example, consider also an *iid* Gaussian prior on λ with

$$\lambda \sim N(m, s^2)$$

$$p(\lambda) = \frac{1}{\sqrt{2\pi}\, s} \exp\left(-\frac{(\lambda - m)^2}{2s^2}\right)$$

Posterior cont'd

Then the posterior is

$$p(\lambda|y) \propto \; e^{-||y-g(\lambda)||} \; e^{-||\lambda-m||}$$

and the log posterior is

$$\ln p(\lambda|y) \quad = \quad -||y-g(\lambda)|| - ||\lambda - m|| + C_\lambda$$

Thus, the maximum a-posteriori (MAP) estimate of λ is equivalent to the solution of the regularized least-squares problem

$$\underset{\lambda}{\operatorname{argmin}}(||y-g(\lambda)|| + ||\lambda - m||)$$

The prior plays the role of a regularizer

Exploring the Posterior

- Given any sample λ, the un-normalized posterior probability can be easily computed

$$p(\lambda|y) \propto p(y|\lambda)p(\lambda)$$

- Explore posterior w/ Markov Chain Monte Carlo (MCMC)
  - Metropolis-Hastings algorithm:
    - Random walk with proposal PDF & rejection rules
  - Computationally intensive, $\mathcal{O}(10^5)$ samples
  - Each sample: evaluation of the forward model
    - Surrogate models
- Evaluate moments/marginals from the MCMC statistics

Surrogate Models for Bayesian Inference

- Need an inexpensive response surface for
  - Observables of interest y
  - as functions of parameters of interest x
- Gaussian Process (GP) surrogate
  - GP goes through all data points with probability 1.0
  - Uncertainty between the points
- Fit a convenient polynomial to $y = f(x)$
  - over the range of uncertainty in x

  - Employ a number of samples (x_i, y_i)
  - Fit with interpolants, regression, ... global/local
  - With uncertain x :
    - Construct Polynomial Chaos response surface

(Marzouk *et al.* JCP 2007; Marzouk & Najm JCP 2009)

Bayesian inference – High Dimensionality Challenge

- Judgement on local/global posterior peaks is difficult
  - Multiple chains
  - Tempering
- Choosing a good starting point is very important
  - An initial optimization strategy is useful, albeit not trivial
- Choosing good MCMC proposals, and attaining good mixing, is a signficant challenge
  - Likelihood-informed proposals
  - Adaptive learning of proposal based on available samples
  - Hessian informs best local multivariate normal approximation of posterior
  - Adaptive, Geometric, and Langevin MCMC methods

Bayesian inference – Model Error Challenge

- Quantifying model error, as distinct from data noise, is important for assessing confidence in model validity
- Available statistical methods for accounting for model error have shortcomings when applied to physical models
- New methods are needed/under-development for assessing how best to model model error such that
  - physical constraints are satisfied
  - disambiguation of model error and data noise is feasible
  - calibrated model error terms adequately impact all model outputs of interest
  - uncertainties in predictions from calibrated model reflect the range of discrepancy from the truth

UQ Challenges - Characterization of Uncertain Inputs

- Computational Model $\mathcal{M}(u, \lambda) = 0$
  - Uncertain input parameter λ
  - Experimental Measurement $F(y, \lambda) = 0$
- Uncertain model inputs can be estimated from data on y
  - Regression
  - Bayesian inference
- Quite frequently, we have partial data/information
  - Partial missing data, *e.g.* failed measurements
  - Full data loss – No data, but have summary information, *e.g.* moments and/or quantiles on
    - data – processed data products
    - fitted parameters

Parameter Estimation in the Absence of Data

- Frequently:
  - we know summary statistics about data or parameters from previous work
  - the raw data used to arrive at these statistics is not available
- How can we construct a joint PDF on the parameters?
- In the absence of data, the structure of the fit model, combined with the summary statistics, implicitly inform the joint PDF on the parameters
- Goal: Make available information *explicit* in the joint PDF

Data Free Inference (DFI)

- Discover a consensus joint PDF on the parameters consistent with given information in the absence of data
  - MaxEnt and Approximate Bayesian Computation

Berry JCP 2012, Najm IJUQ 2014

DFI Algorithm Structure

Basic idea:

- Explore the space of hypothetical data sets
  - MCMC chain on the data
  - Each state defines a data set
- For each data set:
  - MCMC chain on the parameters
  - Evaluate statistics on resulting posterior
  - Accept data set if posterior is consistent with given information
- Evaluate pooled posterior from all acceptable posteriors
  Logarithmic pooling:

$$p(\lambda|y) = \left[\prod_{i=1}^{K} p(\lambda|y_i)\right]^{1/K}$$

Closure

- Probabilistic UQ framework
  - Polynomial Chaos representation of random variables
- Forward UQ
  - Intrusive and non-intrusive forward PC UQ methods
- Inverse UQ
  - Parameter estimation via Bayesian inference
  - Missing data and MaxEnt
- Challenges
  - High dimensionality
  - Intrusive Galerkin stability
  - Nonlinearity
  - Time dynamics
  - Model error