Formulation and Stochastic Galerkin Methods for Stochastic Partial Differential Equations II

Hermann G. Matthies
with Andreas Keese

Institute of Scientific Computing
Technische Universität Braunschweig, Germany

wire@tu-bs.de
http://www.wire.tu-bs.de

Repetition of First Summary

- Motivation, Probabiliy, aleatoric and epistemic Uncertainty
- Formulation as a well-posed problem
- RVs, Stochastic Processes and Random Fields
- Spectral Expansion, Karhunen-Loève Expansion
- Still open:
  - How to discretise RVs ?
  - How to actually compute $u(\omega)$?
  - How to perform integration ?

Overview II

1. Approximating Random Variables

2. Computational Approaches

3. Direct Integration and Collocation Methods

4. Stochastic Galerkin Methods

5. Stability Issues

6. Convergence

Remember Karhunen-Loève Expansion (KLE)

Karhunen-Loève Eigenproblem gives spectrum $\{\kappa_\jmath^2\}$ and orthogonal KLE eigenfunctions $g_\jmath(x) \Rightarrow$ Representation of C_κ and κ:

$$\mathsf{C}_\kappa(x,y) = \sum_{\jmath=1}^{\infty} \varkappa_\jmath^2 \, g_\jmath(x) g_\jmath(y)$$

and

$$\kappa(x,\omega) = \bar{\kappa}(x) + \sum_{\jmath=1}^{\infty} \varkappa_\jmath \, g_\jmath(x)\xi_\jmath(\omega) =: \sum_{\jmath=0}^{\infty} \varkappa_\jmath \, g_\jmath(x)\xi_\jmath(\omega)$$

with centred, uncorrelated random variables $\xi_\jmath(\omega)$.
i.e. $\mathbb{E}\left(\xi_\imath\right) = \langle \xi_\imath \rangle = 0$ and $\mathrm{cov}(\xi_\imath, \xi_\jmath) = \langle \xi_\imath \xi_\jmath \rangle = \langle \xi_\imath, \xi_\jmath \rangle_{L_2(\Omega)} = \delta_{\imath\jmath}$.

Truncation $\Rightarrow$ optimal—in variance—expansion in m variables.

Approximating RVs

The solution $u(x, \omega)$ will be a random field through $\xi_\jmath(\omega)$,
i.e. $u(x, \omega) = u(x, \xi_\jmath(\omega))$.

How to deal with RVs $\xi_\jmath(\omega)$?

- Use $\xi_\jmath(\omega)$ directly. Assume $\xi_\jmath(\omega)$ to be independent, only a finite number M. Transform measure $\mathbb{P}$ to $Y = \mathbb{R}^M$ with image measure from $\{\xi_\jmath(\omega)\}_{\jmath=1,\ldots,M}$. Ansatz for solution $u(x, \omega)$ in (doubly orthogonal) polynomials in $\boldsymbol{y} = (y_1, \ldots, y_M) \in Y$ w.r.t. image measures.

- Represent $\xi_\jmath(\omega)$ as functions of other—simpler—RVs.

Functions of Simpler RVs

What kind of simpler RVs ?
What kind of functions? — Usually polynomials or other algebras.

- Gaussian RVs —classical Wiener Chaos
- Poissonian RVs —discrete Poisson Chaos
- other RVs, e.g. uniform, exponential, Gamma, Beta, etc.
  This is called generalised Polynomial Chaos (gPC).

Best is to use orthogonal polynomials w.r.t. relevant measure, i.e. Hermite polynomials for Gaussian RVs, Charlier polynomials for Poisson RVs, Legendre polynomials for uniform RVs, Laguerre polynomials for exponential RVs, etc. $\Rightarrow$ Askey scheme.

Why White Noise Analysis?

Comes from directly constructing Ω as (a subset of) $\mathcal{S}'(\mathcal{G})$ (tempered distributions) with a Gaussian or Poissonian measure $\mathbb{P}$ $\Rightarrow$ Gaussian or Poissonian white noise.

Elements from $\mathcal{S}(\mathcal{G})$ (rapidly falling test functions) are then naturally Gaussian or Poissonian RVs.

Let $\mathfrak{F} = \mathfrak{F}(\{\xi_\jmath(\omega)\}_{\jmath=1,\dots,\infty})$ be the σ-algebra generated by $\xi_\jmath(\omega)$. Want to approximate $L_2(\Omega, \mathfrak{F}, \mathbb{P}) \subseteq L_2(\Omega, \mathbb{P})$.

Density results: Polynomial algebra, algebra of exponentials, and algebra of trigonometric polynomials of Gaussian RVs is dense in $L_2(\Omega, \mathfrak{F}, \mathbb{P})$, polynomial algebra of Poissonian RVs is dense in $L_2(\Omega, \mathfrak{F}, \mathbb{P})$.

Polynomial Chaos Expansion in Gaussians (PCE)

Each $\xi_\jmath(\omega) = \sum_\alpha \xi_\jmath^{(\alpha)} H_\alpha(\boldsymbol{\theta}(\omega))$ from KLE may be expanded in polynomial chaos expansion (PCE), with orthogonal polynomials of independent Gaussian RVs $\{\theta_m(\omega)\}_{m=1}^\infty =: \boldsymbol{\theta}(\omega)$:

$$H_\alpha(\boldsymbol{\theta}(\omega)) = \prod_{\jmath=1}^{\infty} h_{\alpha_\jmath}(\theta_\jmath(\omega)),$$

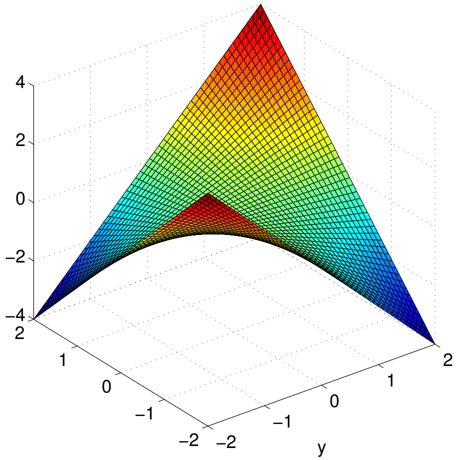
where $h_\ell(\vartheta)$ are the usual Hermite polynomials, and

$$\mathcal{J} := \{\alpha \,|\, \alpha = (\alpha_1, \dots, \alpha_\jmath, \dots), \; \alpha_\jmath \in \mathbb{N}_0, \; |\alpha| := \sum_{\jmath=1}^{\infty} \alpha_\jmath < \infty\}$$

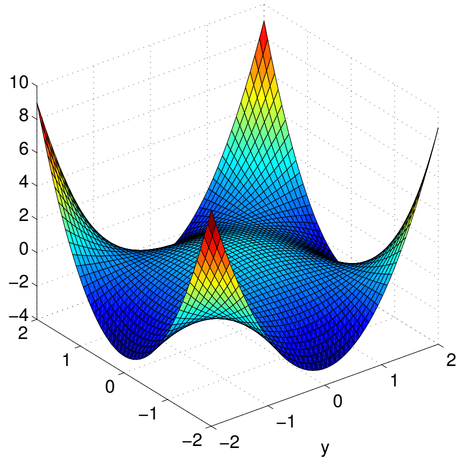
are multi-indices, where only finitely many of the $\alpha_\jmath$ are non-zero.

Here $\langle H_\alpha, H_\beta \rangle_{L_2(\Omega)} = \mathbb{E}\,(H_\alpha H_\beta) = \alpha!\, \delta_{\alpha\beta}$, where $\alpha! := \prod_{\jmath=1}^{\infty}(\alpha_\jmath!)$.

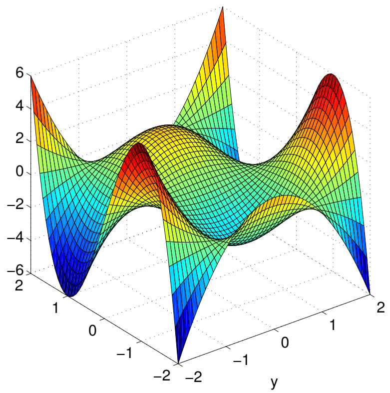
Polynomial Chaos



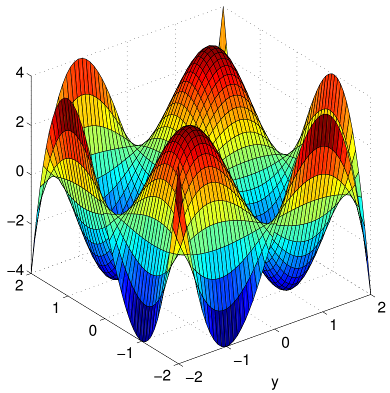
Hermite$_{(1,1)}$



Hermite$_{(2,2)}$



Hermite$_{(2,3)}$



Hermite$_{(3,3)}$

Hermite Algebra

Hermite polynomials $H_\alpha(\boldsymbol{\theta})$ are considered on $\Theta = \mathbb{R}^{\mathbb{N}}$ with image product measure $\Gamma = \bigotimes_m \Gamma_m$ from Gaussian RVs $\{\theta_m(\omega)\}_{m=1}^{\infty} =: \boldsymbol{\theta}(\omega)$.
Remember that polynomials are an algebra:

$$h_k(\vartheta)h_\ell(\vartheta) = \sum_{m=0}^{k+\ell} c_{k\ell}^{(m)} h_m(\vartheta)$$

The coefficients $c_{k\ell}^{(m)}$ are explicitly known—structure constants of the algebra. Similarly for multi-polynomials H_α:

$$H_\alpha(\boldsymbol{\theta})H_\beta(\boldsymbol{\theta}) = \sum_\gamma C_{\alpha\beta}^{(\gamma)} H_\gamma(\boldsymbol{\theta})$$

Structure constants $C_{\alpha\beta}^{(\gamma)}$ are explicitly known in terms of $c_{k\ell}^{(m)}$.

Stochastic or White Noise Hilbert Spaces

Start with formal PCE: $R(\theta) = \sum_{\alpha \in \mathcal{J}} R^{(\alpha)} H_\alpha(\theta)$, where $R^{(\alpha)} \in \mathcal{V}$, and $\mathcal{V}$ some other Hilbert space. Define for $|\rho| \leq 1$ and $p \geq 0$ inner product and corresponding norm (with $(2\mathbb{N})^\beta := \prod_{\jmath \in \mathbb{N}} (2\jmath)^{\beta_\jmath}$):

$$\langle R_1, R_2 \rangle_{\rho,p} = \sum_\alpha \langle R_1^{(\alpha)}, R_2^{(\alpha)} \rangle_{\mathcal{V}} \, (\alpha!)^{1+\rho} (2\mathbb{N})^{p\alpha}.$$

Define for $1 \geq \rho \geq 0, p \geq 0$ (with $\|R\|^2_{\rho,p} = \langle R, R \rangle_{\rho,p}$):

$$(\mathcal{S})^{\rho,p} = \{ R(\theta) = \sum_{\alpha \in \mathcal{J}} R^{(\alpha)} H_\alpha(\theta) : \quad \|R\|_{\rho,p} < \infty \}.$$

These are Hilbert spaces, the duals are denoted by $(\mathcal{S})^{-\rho,-p}$, and $L_2(\Omega) = (\mathcal{S})^{0,0}$. One has Gelfand triplets $(\mathcal{S})^{\rho,p} \subset (\mathcal{S})^{0,0} \subset (\mathcal{S})^{-\rho,-p}$.

White Noise Hilbert Spaces

The scale of Hilbert spaces $\{(\mathcal{S})^{\rho,p}\}$ allows definitions of various stochastic distribution spaces via toplogical limits.

PCE allows definition of stochastic Sobolev spaces via inner products

$$\langle R_1, R_2 \rangle_k = \sum_{n=0}^{\infty} (n+1)^k \sum_{|\alpha|=n} \langle R_1^{(\alpha)}, R_2^{(\alpha)} \rangle_{\mathcal{V}}.$$

Define for $k \in \mathbb{N}_0$ (with $\|R\|_{k,2}^2 = \langle R, R \rangle_k$):

$$\mathcal{D}_2^k = \{R(\theta) = \sum_{\alpha \in \mathcal{J}} R^{(\alpha)} H_\alpha(\theta) : \quad \|R\|_{k,2} < \infty\}.$$

Knowing that a random variable R is in one of these spaces gives regularity results (differentiability, smoothness of distribution function).

Computational Approaches

The principal computational approaches are:

Perturbation Assume that stochastics is a small perturbation around mean value, do Taylor expansion and truncate.

Direct Integration (e.g. Monte Carlo) Directly compute statistic by quadrature: $\boldsymbol{\Psi}_u = \mathbb{E}\left(\Psi(u(\omega), \omega)\right) = \int_\Theta \Psi(u(\boldsymbol{\theta}), \boldsymbol{\theta})\, \Gamma(d\boldsymbol{\theta})$ by numerical integration. Needs solution $u(\theta_z)$.

Direct Response Surface Try to find a functional fit $u(\boldsymbol{\theta}) \approx v(\boldsymbol{\theta})$, then compute with $v(\boldsymbol{\theta})$. Needs solution $u(\theta_z)$. Integrand is now cheap. One possibility is PCE.

Stochastic Galerkin This is one possible way to compute PCE.

Stability Issues

For direct methods expansions (both KLE and PCE) pose stability problems: Both only converge in L_2, not in L_∞ (uniformly) as required $\Rightarrow$ spatially discrete problems to compute $u(\theta_z)$ for a specific realisation θ_z (like Monte Carlo) may not be well posed.

Convergence of KLE may be uniform if covariance $\mathsf{C}_\kappa(x_1, x_2)$ smooth enough, but e.g. not possible for $\mathsf{C}_\kappa(x_1, x_2) = \exp(-a|x_1 - x_2|)$

Truncation of PCE gives a polynomial, as soon as one $\alpha_\jmath$ is odd, there are regions where κ is negative—compare approximating $\exp(\xi)$ with a truncated Taylor poplynomial at odd power.

This can not be repaired. Like negative Jacobian in normal FEM.

Method $\kappa(x, \omega) = \phi(x, \gamma(x, \omega))$ possible with KLE of Gaussian $\gamma(x, \omega)$.

Stochastic Galerkin I

Variational formulation discretised in space, e.g. via finite element ansatz $u(x,\omega) = \sum_{\ell=1}^{n} u_\ell(\boldsymbol{\theta}) N_\ell(x) = [N_1(x), \dots, N_n(x)][u_1(\boldsymbol{\theta}), \dots, u_n(\boldsymbol{\theta})]^T = \boldsymbol{N}(x)^T \boldsymbol{u}(\boldsymbol{\theta})$:

$$\boldsymbol{K}(\boldsymbol{\theta})[\boldsymbol{u}(\boldsymbol{\theta})] = \boldsymbol{f}(\boldsymbol{\theta}).$$

Recipe: Stochastic ansatz and projection in stochastic dimensions

$$\boldsymbol{u}(\boldsymbol{\theta}) = \sum_{\beta} \boldsymbol{u}^{(\beta)} H_\beta(\boldsymbol{\theta}) = [\dots, \boldsymbol{u}^{(\beta)}, \dots][\dots, H_\beta(\boldsymbol{\theta}), \dots]^T = \mathbf{u}\boldsymbol{H}(\boldsymbol{\theta})$$

Goal: Compute coefficients $\boldsymbol{u}^{(\beta)}$ through stochastic Galerkin Methods,

$$\forall \gamma : \quad \mathbb{E}\left((\boldsymbol{f}(\boldsymbol{\theta}) - \boldsymbol{K}(\boldsymbol{\theta})[\mathbf{u}\boldsymbol{H}(\boldsymbol{\theta})]) H_\gamma(\boldsymbol{\theta}) \right) = 0,$$

requires solution of one huge system, only integrals of residuals.

Stochastic Galerkin II

Of course we can not use all $\alpha \in \mathcal{J}$, but take only a finite subset

$$\mathcal{J}_{k,m} = \{\alpha \in \mathcal{J} | \quad |\alpha| \leq k,\ \imath > m\ \alpha_\imath = 0\} \subset \mathcal{J}.$$

Let $\mathscr{S}_{k,m} = \mathrm{span}\{H_\alpha : \alpha \in \mathcal{J}_{k,m}\}$,

then $\dim \mathscr{S}_{k,m} = \begin{pmatrix} m+p+1 \\ p+1 \end{pmatrix}$

better to use other subsets—
best is adaptive choice.

m	k	dim $\mathscr{S}_{k,m}$
3	3	35
	5	84
5	3	126
	5	462
10	3	1001
	5	8008
20	3	10626
	5	230230
	10	$\approx 8.5 \cdot 10^7$
100	3	$\approx 4.6 \cdot 10^6$
	5	$\approx 1.7 \cdot 10^9$
	10	$\approx 4.7 \cdot 10^{14}$

Galerkin-Methods for the General Linear Case

$\forall \gamma \in \mathcal{J}_{k,m} = \{\alpha \in \mathcal{J} | \quad |\alpha| \le k,\ \imath > m \ \Rightarrow\ \alpha_\imath = 0\}$ satisfy:

$$\sum_\beta \left[\int_{\mathcal{G}} \nabla \boldsymbol{N}(x)\, \mathbb{E}\left(\kappa(x,\boldsymbol{\theta}) H_\beta(\boldsymbol{\theta}) H_\gamma(\boldsymbol{\theta})\right) \nabla \boldsymbol{N}(x)^T \, dx \right] \boldsymbol{u}^{(\beta)} =$$

$$\underbrace{\mathbb{E}\left(\boldsymbol{f}(\boldsymbol{\theta}) H_\gamma(\boldsymbol{\theta})\right)}_{=:\gamma!\boldsymbol{f}^{(\gamma)}}$$

More efficient representation through direct expansion of κ in KLE and PCE and analytic computation of expectations.

$$\kappa(x,\boldsymbol{\theta}) = \sum_{\jmath=0}^{\infty} \varkappa_\jmath \, \xi_\jmath(\boldsymbol{\theta}) g_\jmath(x) \approx \sum_{\jmath=0}^{r} \sum_{\alpha \in \mathcal{J}_{2k,m}} \varkappa_\jmath \, \xi_\jmath^{(\alpha)} H_\alpha(\boldsymbol{\theta}) g_\jmath(x).$$

Resulting Equations

Insertion of expansion of κ $\#\text{dof}_{\text{space}} \cdot \#\text{dof}_{\text{stoch}}$ linear equations.

$$\sum_\alpha \sum_\beta \sum_\jmath \xi_\jmath^{(\alpha)} \underbrace{\mathbb{E}\left(H_\alpha H_\beta H_\gamma\right)}_{=:\Delta^{(\alpha)}_{\beta,\gamma}} \underbrace{\int \nabla \boldsymbol{N}(x) \varkappa_\jmath\, g_\jmath(x) \nabla \boldsymbol{N}(x)^T \, dx}_{\boldsymbol{K}_\jmath} \; \boldsymbol{u}^{(\beta)} = \boldsymbol{f}^{(\gamma)}$$

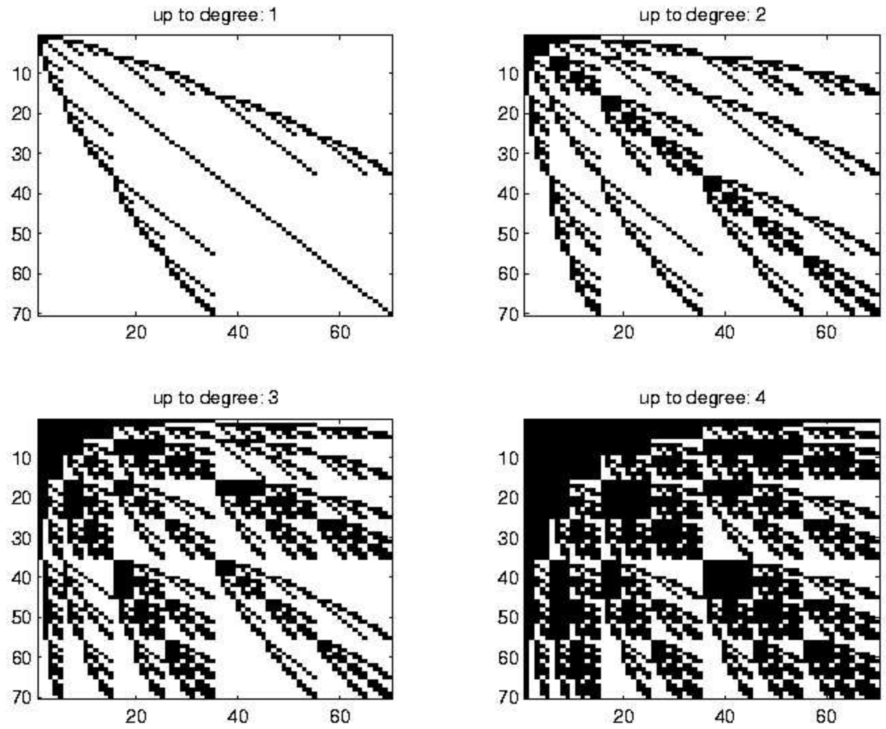
- $\boldsymbol{K}_\jmath$ is stiffness matrix of a FEM discretisation for the material $\varkappa_\jmath\, g_\jmath(x)$.
- $\Rightarrow$ Use deterministic FEM program in black-box-fashion.
- Equations have structure of a tensor product (storage and use).

$$\mathbf{K}\,\mathbf{u} = \sum_\jmath \sum_\alpha \xi_\jmath^{(\alpha)} \Delta^{(\alpha)} \otimes \boldsymbol{K}_\jmath\, \mathbf{u} = \mathbf{f}$$

- $\mathbb{E}\left(H_\alpha H_\beta H_\gamma\right) = \mathbb{E}\left(H_\alpha \sum_\varepsilon C^{(\varepsilon)}_{\beta\gamma} H_\varepsilon\right) = \sum_\varepsilon C^{(\varepsilon)}_{\beta\gamma} \langle H_\alpha, H_\varepsilon \rangle = C^{(\alpha)}_{\beta\gamma}\, \alpha!$

Sparsity Structure

Non-zero blocks of $\Delta^{(\alpha)}$ for increasing degree of H_α



Properties of Global Equations

$$\mathbf{K}\,\mathbf{u} = \sum_{\jmath} \sum_{\alpha} \xi_{\jmath}^{(\alpha)} \Delta^{(\alpha)} \otimes \boldsymbol{K}_{\jmath}\,\mathbf{u} = \mathbf{f}$$

- Each $\boldsymbol{K}_{\jmath}$ is symmetric, and each $\Delta^{(\alpha)} \Rightarrow$ Block-matrix $\mathbf{K}$ is symmetric.
- Appropriate expansion of $\kappa \Rightarrow \mathbf{K}$ is uniformly positive definite.
- Never assemble block-matrix explicitly.
- $\Delta^{(\alpha)}$ are known analytically. No need to store explicitly.
- Use $\mathbf{K}$ only as multiplication.
- Use Krylov method (here CG) with pre-conditioner.

Block-Diagonal Pre-Conditioner

Let $\overline{\boldsymbol{K}} = \boldsymbol{K}_0 =$ stiffness-matrix for average material $\overline{\kappa}(x)$.

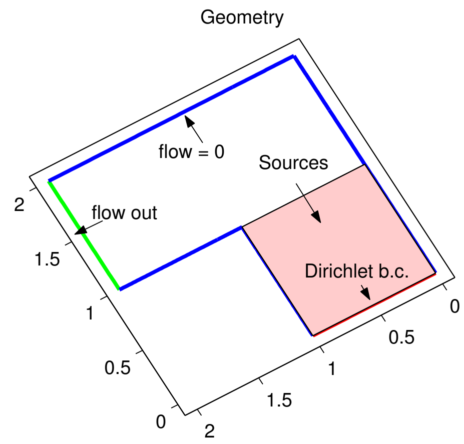
Use deterministic solver as pre-conditioner:

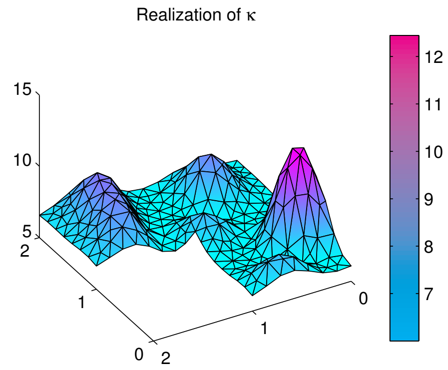
$$\mathbf{P} = \begin{pmatrix} \overline{\boldsymbol{K}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \overline{\boldsymbol{K}} \end{pmatrix} = \boldsymbol{I} \otimes \overline{\boldsymbol{K}}$$

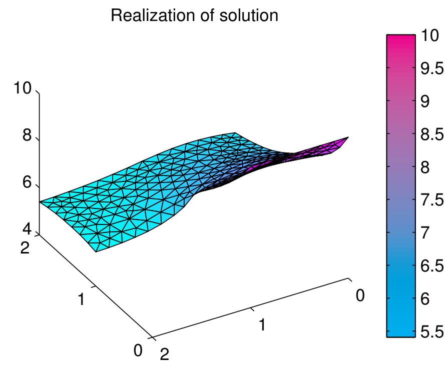
Good pre-conditioner, when variance of κ not too large.
Otherwise use $\mathbf{P} =$ block-diag$(\mathbf{K})$.
This may again be done with existing deterministic solver.

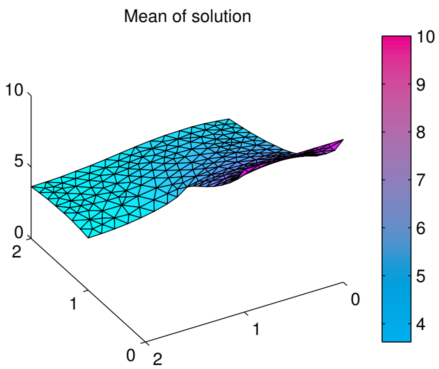
Block-diagonal $\mathbf{P}$ is well suited for parallelisation.

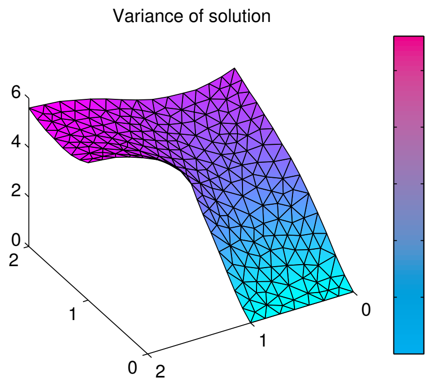
Example Solution

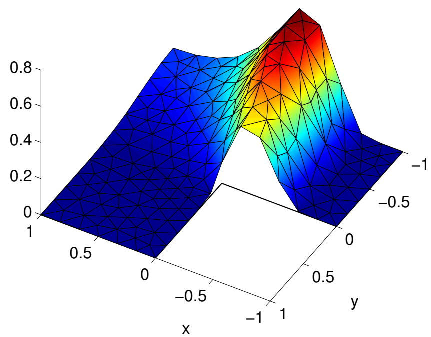








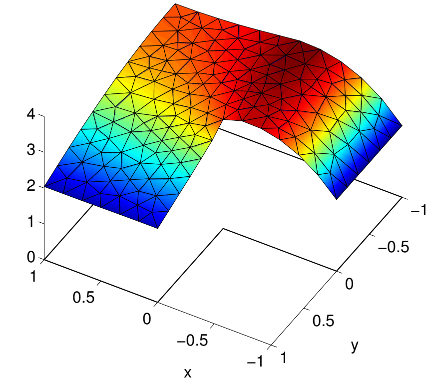




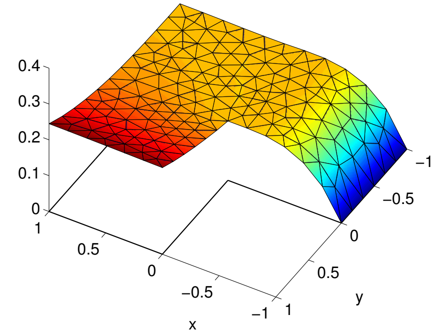
$\Pr\{u(x) > 8\}$

Results of Galerkin Method

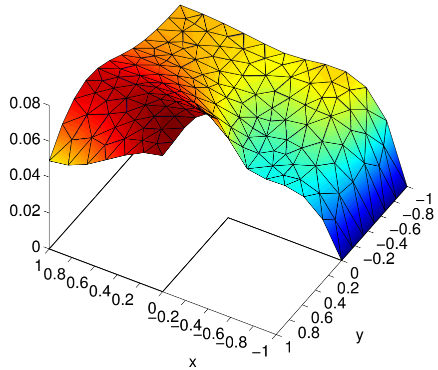
err. $\cdot 10^4$ in mean $m = 6$ $k = 2$



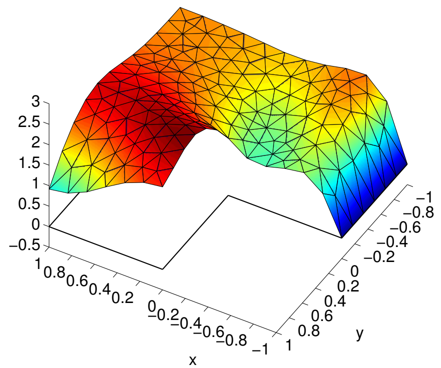
err. $\cdot 10^4$ in std dev $m = 6$ $k = 2$



$u^{(\alpha)}$ for $\alpha = (0, 0, 0, 1, 0$



Error $\cdot 10^4$ in $u^{(\alpha)}$ Galerkin sche-me.



Parallelising the Matrix-Vector Product

$$\forall \gamma : \qquad (\boldsymbol{K}\,\mathbf{u})^{(\gamma)} = \sum_{\jmath}^{J} \sum_{\beta}^{N} \sum_{\alpha}^{n_\alpha} \xi_{\jmath}^{(\alpha)}\, \Delta_{\beta,\gamma}^{(\alpha)} \cdot \boldsymbol{K}_{\jmath}\, \boldsymbol{u}_{\beta}$$

- $\boldsymbol{K}_{\jmath} =$ deterministic solver.
  This may be a (lower-level) parallel program to do $\boldsymbol{K}_{\jmath}\, \boldsymbol{u}_{\beta}$.
- Parallelise operator-sum in $\jmath$
  $\Rightarrow$ several instances of deterministic solver in parallel.
- Distribute $\mathbf{u}$ and $\mathbf{f}$ $\Rightarrow$ Parallelise sum in β.
- Sum in α may also be done in parallel, but usually not essential.

Parallelisation

$$\forall \gamma : \qquad (\boldsymbol{K}\,\mathbf{u})^{(\gamma)} = \sum_{\jmath} \sum_{\beta} \sum_{\alpha} \xi_{\jmath}^{(\alpha)} \, \Delta_{\beta,\gamma}^{(\alpha)} \cdot \boldsymbol{K}_{\jmath} \, \boldsymbol{u}_{\beta}$$

- Obviously Parallel in γ.
- Block-vectors $\mathbf{u}$ and $\mathbf{f}$ distributed. May be replicated, in order to reduce communication.
- Matrices $\boldsymbol{K}_{\jmath}$ distributed over processors. May be replicated, in order to reduce parallel communication, and use more processors than number of $\boldsymbol{K}_{\jmath}$.

Several processor-groups, where each uses a subset of the $\boldsymbol{K}_{\jmath}$ and stores a subset of $\mathbf{u}$ and $\mathbf{f}$. On Cray T3E with 128 proc. we have solved systems with more than 5×10^7 equations with high parallel efficiency.

Approximation Theory

- Stability of discrete approximation under truncated KLE and PCE. Matrix stays uniformly positive definite.
- Convergence follows from Céa's Lemma.
- Convergence rates under stochastic regularity in stochastic Hilbert spaces—stochastic regularitry theory?.
- Error estimation via dual weighted residuals possible.

Theorem: Let $p > 0$, $r > 1$ and let $|\rho| \leq 1$. Then for any $R \in (\mathcal{S})^{\rho,p}$:

$$\|R - P_{k,m}(R)\|^2_{\rho,-p} \leq \|R\|^2_{\rho,-p+r}\, c(m,k,r)^2,$$

where $c(m,k,r)^2 = c_1(r)m^{1-r} + c_2(r)2^{-kr}$. But $\dim \mathscr{S}_{k,m}$ grows too quickly with k and m. Sparser spaces and error estimates needed.

Second Summary

Stochstic Galerkin methods work.

Galerkin procedure is numerically stable $\Rightarrow$ convergence.

Convergence rates seemingly only with regularity.

Stochastic calculations produce huge amounts of data, which is expensive to operate on and to store.

Results a priori live in very high dimensional spaces.

They have a natural tensor product structure.