Discrete least squares polynomial approximations for high dimensional uncertainty propagation

Fabio Nobile

CSQI-MATHICSE, EPFL, Switzerland

Acknowlegments: G. Migliorati (EPFL), R. Tempone (KAUST), A. Cohen, A. Chkifa (UPMC - Paris VI), E. von Schwerin (KTH),

MascotNum Workshop on Computer Experiments and Meta-models for Uncertainty Quantification, April 23-25, 2014, ETH Zürich,

Italian project FIRB-IDEAS ('09) *Advanced Numerical Techniques for Uncertainty Quantification in Engineering and Life Science Problems*

Center for Advanced Modeling and Science

Outline

Outline

UQ for deterministic PDE models

- Consider the parametric equation (typically a PDE)

$$\text{find } u : \quad \mathcal{F}(\mathbf{y}, u) = 0 \tag{1}$$

  where $\mathbf{y}$ is a vector of N parameters: $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{R}^N$ ($N = \infty$ when dealing with distributed fields).

- Often in applications the parameters $\mathbf{y}$ are not perfectly known or are intrinsically variable. Examples are:
  - subsurface modeling: porous media flows; seismic waves; basin evolutions; ...
  - modeling of living tissues: mechanical response; growth models;
  - material science: properties of composite materials
- Probabilistic approach: $\mathbf{y}$ is a random vector with probability density function $\rho : \Gamma \subset \mathbb{R}^N \to \mathbb{R}_+$.

UQ for deterministic PDE models

- Consider the parametric equation (typically a PDE)

$$\text{find } u : \quad \mathcal{F}(\mathbf{y}, u) = 0 \tag{1}$$

  where $\mathbf{y}$ is a vector of N parameters: $\mathbf{y} = (y_1, \dots, y_N) \in \mathbb{R}^N$ ($N = \infty$ when dealing with distributed fields).
- Often in applications the parameters $\mathbf{y}$ are not perfectly known or are intrinsically variable. Examples are:
  - subsurface modeling: porous media flows; seismic waves; basin evolutions; ...
  - modeling of living tissues: mechanical response; growth models;
  - material science: properties of composite materials
- Probabilistic approach: $\mathbf{y}$ is a random vector with probability density function $\rho : \Gamma \subset \mathbb{R}^N \to \mathbb{R}_+$.

UQ for deterministic PDE models

- Consider the parametric equation (typically a PDE)

$$\text{find } u : \quad \mathcal{F}(\mathbf{y}, u) = 0 \tag{1}$$

where $\mathbf{y}$ is a vector of N parameters: $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{R}^N$ ($N = \infty$ when dealing with distributed fields).

- Often in applications the parameters $\mathbf{y}$ are not perfectly known or are intrinsically variable. Examples are:
  - subsurface modeling: porous media flows; seismic waves; basin evolutions; ...
  - modeling of living tissues: mechanical response; growth models;
  - material science: properties of composite materials
- Probabilistic approach: $\mathbf{y}$ is a random vector with probability density function $\rho : \Gamma \subset \mathbb{R}^N \to \mathbb{R}_+$.

UQ for deterministic PDE models

Assumption: $\forall \mathbf{y} \in \Gamma$ the problem admits a unique solution $u \in V$ in a suitable finite or infinite dimensional Hilbert space V. Moreover,

$$\forall \mathbf{y} \in \Gamma, \quad \exists C(y) > 0; \qquad \|u(\mathbf{y})\|_V \leq C(\mathbf{y})$$

- Then, equation (1) induces a map $u = u(\mathbf{y}) : \Gamma \to V$.
- if $\int_\Gamma C(\mathbf{y})^p \rho(\mathbf{y}) d\mathbf{y} < \infty$, then $u \in L^p_\rho(\Gamma, V)$.

Goals:

- Construct a reduced model $u_\Lambda(\mathbf{y}) \approx u(\mathbf{y})$
- Compute statistics of the solution

Expected value: $\bar{u} \approx \mathbb{E}[u_\Lambda]$

Variance: $Var[u] \approx \mathbb{E}[u_\Lambda^2] - \mathbb{E}[u_\Lambda]^2$

two points corr. (if u is a distributed field)

$$Cov_u(x_1, x_2) \approx \mathbb{E}[u_\Lambda(x_1)u_\Lambda(x_2)] - \mathbb{E}[u_\Lambda(x_1)]\mathbb{E}[u_\Lambda(x_2)]$$

UQ for deterministic PDE models

Assumption: $\forall \mathbf{y} \in \Gamma$ the problem admits a unique solution $u \in V$ in a suitable finite or infinite dimensional Hilbert space V. Moreover,

$$\forall \mathbf{y} \in \Gamma, \;\; \exists C(y) > 0; \qquad \|u(\mathbf{y})\|_V \leq C(\mathbf{y})$$

- Then, equation (1) induces a map $u = u(\mathbf{y}) : \Gamma \to V$.
- if $\int_\Gamma C(\mathbf{y})^p \rho(\mathbf{y}) d\mathbf{y} < \infty$, then $u \in L^p_\rho(\Gamma, V)$.

Goals:

- Construct a reduced model $u_\Lambda(\mathbf{y}) \approx u(\mathbf{y})$
- Compute statistics of the solution

  Expected value: $\bar{u} \approx \mathbb{E}[u_\Lambda]$
  Variance: $Var[u] \approx \mathbb{E}[u_\Lambda^2] - \mathbb{E}[u_\Lambda]^2$
  two points corr. (if u is a distributed field)

$$\mathrm{Cov}_u(x_1, x_2) \approx \mathbb{E}[u_\Lambda(x_1)u_\Lambda(x_2)] - \mathbb{E}[u_\Lambda(x_1)]\mathbb{E}[u_\Lambda(x_2)]$$

UQ for deterministic PDE models

Assumption: $\forall \mathbf{y} \in \Gamma$ the problem admits a unique solution $u \in V$ in a suitable finite or infinite dimensional Hilbert space V. Moreover,

$$\forall \mathbf{y} \in \Gamma, \quad \exists C(y) > 0; \qquad \|u(\mathbf{y})\|_V \leq C(\mathbf{y})$$

- Then, equation (1) induces a map $u = u(\mathbf{y}) : \Gamma \to V$.
- if $\int_\Gamma C(\mathbf{y})^p \rho(\mathbf{y}) d\mathbf{y} < \infty$, then $u \in L^p_\rho(\Gamma, V)$.

Goals:

- Construct a reduced model $u_\Lambda(\mathbf{y}) \approx u(\mathbf{y})$
- Compute statistics of the solution

  Expected value: $\bar{u} \approx \mathbb{E}[u_\Lambda]$
  Variance: $Var[u] \approx \mathbb{E}[u_\Lambda^2] - \mathbb{E}[u_\Lambda]^2$
  two points corr. (if u is a distributed field)

$$\mathrm{Cov}_u(x_1, x_2) \approx \mathbb{E}[u_\Lambda(x_1)u_\Lambda(x_2)] - \mathbb{E}[u_\Lambda(x_1)]\mathbb{E}[u_\Lambda(x_2)]$$

Example: Elliptic PDE with random coefficients

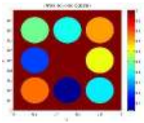
$$\begin{cases} -\operatorname{div}(a(\mathbf{y},x)\nabla u(\mathbf{y},x)) = f(x) & x \in D, \quad \mathbf{y} \in \Gamma, \\ u(\mathbf{y},x) = 0 & x \in \partial D, \quad \mathbf{y} \in \Gamma \end{cases}$$

with $a_{min}(\mathbf{y}) = \inf_{x \in D} a(\mathbf{y},x) > 0$ for all $\mathbf{y} \in \Gamma$ and $f \in L^2(D)$. Then

$$\forall \mathbf{y} \in \Gamma, \quad u(\mathbf{y}) \in V \equiv H_0^1(D), \quad \text{and} \quad \|u(\mathbf{y})\|_V \leq \frac{C_P}{a_{min}(\mathbf{y})} \|f\|_{L^2(D)}.$$

Inclusions problem

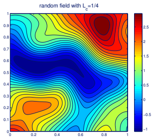
$\mathbf{y}$ describes the conductivitiy in each inclusion



$$a(\mathbf{y},x) = a_0 + \sum_{n=N}^{N} y_n \mathbb{1}_{D_n}(x)$$

Random fields problem

$a(\mathbf{y},x)$ is a random field, e.g. lognormal: $a(\mathbf{y},x) = e^{\gamma(\mathbf{y},x)}$ with γ expanded e.g. in Karhunen-Loève series



$$\gamma(\mathbf{y},x) = \sum_{n=1}^{\infty} \sqrt{\lambda_n} y_n b_n(x), \quad y_n \sim N(0,1) \ i.i.d.$$

Outline

1. Introduction – parametric / stochastic equations
2. Stochastic polynomial approximation
3. Discrete least squares approx. using random evaluations
   - Convergence analysis
   - Numerical results
4. Adaptive algorithms
5. Conclusions

Stochastic multivariate polynomial approximation

- The parameter-to-solution map $u(\mathbf{y}) : \Gamma \to V$ is often smooth (even analytic for the elliptic diffusion model). It is therefore sound to approximate it by global multivariate polynomials.
- Let $\Lambda \subset \mathbb{N}^N$ be an index set of cardinality $|\Lambda| = M$, and consider the multivariate polynomial space

$$\mathbb{P}_\Lambda(\Gamma) = span\left\{\prod_{n=1}^N y_n^{p_n}, \quad \text{with } \mathbf{p} = (p_1, \ldots, p_N) \in \Lambda\right\}$$

  We seek an approximation $P_\Lambda u \in \mathbb{P}_\Lambda(\Gamma) \otimes V$.
- The optimal choice of Λ depends heavily on the problem at hand and the "structure" of the map $u(\mathbf{y})$.

Definition. An index set Λ is downward closed (or *lower set*) if

$$\mathbf{p} \in \Lambda \text{ and } \mathbf{q} \leq \mathbf{p} \quad \Longrightarrow \quad \mathbf{q} \in \Lambda$$

Stochastic multivariate polynomial approximation

- The parameter-to-solution map $u(\mathbf{y}) : \Gamma \to V$ is often smooth (even analytic for the elliptic diffusion model). It is therefore sound to approximate it by global multivariate polynomials.
- Let $\Lambda \subset \mathbb{N}^N$ be an index set of cardinality $|\Lambda| = M$, and consider the multivariate polynomial space

$$\mathbb{P}_\Lambda(\Gamma) = span\left\{ \prod_{n=1}^{N} y_n^{p_n}, \quad \text{with } \mathbf{p} = (p_1, \ldots, p_N) \in \Lambda \right\}$$

We seek an approximation $P_\Lambda u \in \mathbb{P}_\Lambda(\Gamma) \otimes V$.

- The optimal choice of Λ depends heavily on the problem at hand and the "structure" of the map $u(\mathbf{y})$.

Definition. An index set Λ is downward closed (or *lower set*) if

$$\mathbf{p} \in \Lambda \text{ and } \mathbf{q} \leq \mathbf{p} \implies \mathbf{q} \in \Lambda$$

Stochastic multivariate polynomial approximation

- The parameter-to-solution map $u(\mathbf{y}) : \Gamma \to V$ is often smooth (even analytic for the elliptic diffusion model). It is therefore sound to approximate it by global multivariate polynomials.
- Let $\Lambda \subset \mathbb{N}^N$ be an index set of cardinality $|\Lambda| = M$, and consider the multivariate polynomial space

$$\mathbb{P}_\Lambda(\Gamma) = span\left\{\prod_{n=1}^N y_n^{p_n}, \quad \text{with } \mathbf{p} = (p_1, \dots, p_N) \in \Lambda\right\}$$

We seek an approximation $P_\Lambda u \in \mathbb{P}_\Lambda(\Gamma) \otimes V$.

- The optimal choice of Λ depends heavily on the problem at hand and the “structure” of the map $u(\mathbf{y})$.

Definition. An index set Λ is downward closed (or *lower set*) if

$$\mathbf{p} \in \Lambda \text{ and } \mathbf{q} \le \mathbf{p} \implies \mathbf{q} \in \Lambda$$

Stochastic multivariate polynomial approximation

- The parameter-to-solution map $u(\mathbf{y}) : \Gamma \to V$ is often smooth (even analytic for the elliptic diffusion model). It is therefore sound to approximate it by global multivariate polynomials.
- Let $\Lambda \subset \mathbb{N}^N$ be an index set of cardinality $|\Lambda| = M$, and consider the multivariate polynomial space

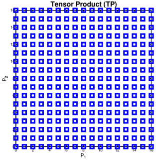
$$\mathbb{P}_\Lambda(\Gamma) = span\left\{\prod_{n=1}^N y_n^{p_n}, \quad \text{with } \mathbf{p} = (p_1, \dots, p_N) \in \Lambda\right\}$$

  We seek an approximation $P_\Lambda u \in \mathbb{P}_\Lambda(\Gamma) \otimes V$.
- The optimal choice of Λ depends heavily on the problem at hand and the "structure" of the map $u(\mathbf{y})$.

Definition. An index set Λ is downward closed (or *lower set*) if

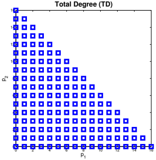
$$\mathbf{p} \in \Lambda \text{ and } \mathbf{q} \leq \mathbf{p} \implies \mathbf{q} \in \Lambda$$

Common choices of polynomial spaces



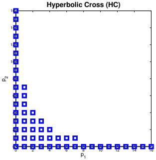
tensor product (TP)

$\Lambda(w) = \{\mathbf{p} : \max_n p_n \leq w\}$



total degree (TD)

$\Lambda(w) = \{\mathbf{p} : \sum_{n=1}^{N} p_n \leq w\}$

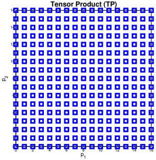


hyperbolic cross (HC)

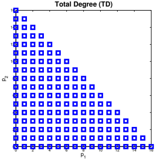
$\Lambda(w) = \{\mathbf{p} : \prod_{n=1}^{N}(p_n + 1) \leq w + 1\}$

- Anisotropic versions are also possible.
- All these index sets are all downward closed.

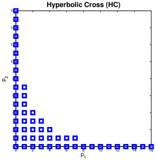
Common choices of polynomial spaces



tensor product (TP)
$\Lambda(w) = \{\mathbf{p} : \max_n p_n \leq w\}$



total degree (TD)
$\Lambda(w) = \{\mathbf{p} : \sum_{n=1}^{N} p_n \leq w\}$



hyperbolic cross (HC)
$\Lambda(w) = \{\mathbf{p} : \prod_{n=1}^{N}(p_n + 1) \leq w + 1\}$

- Anisotropic versions are also possible.
- All these index sets are all downward closed.

Outline

Approximation using random evaluations

Goal: construct a polynomial approximation using random evaluations (Monte Carlo sampling):

1. Generate M random i.i.d. samples $\mathbf{y}^{(k)} \sim \rho(\mathbf{y})d\mathbf{y}$, $k = 1, \ldots, M$
2. Compute the corresponding solutions $u^{(k)} = u(\mathbf{y}^{(k)})$
3. Construct a suitable approximation $P_\Lambda^{M,\omega} u \in \mathbb{P}_\Lambda(\Gamma) \otimes V$

Notation

given two functions $u, v \in L^2_\rho(\Gamma; V)$

- Continuous inner product: $\mathbb{E}[(u,v)_V] = \int_\Gamma (u(\mathbf{y}), v(\mathbf{y}))_V \rho(\mathbf{y}) d\mathbf{y}$
- Continuous norm: $\|v\|^2_{L^2_\rho(\Gamma;V)} = \mathbb{E}[(v,v)_V]$.
- Discrete inner product: $\mathbb{E}_M[(u,v)_V] = \frac{1}{M}\sum_{i=1}^M (u(\mathbf{y}^{(i)}), v(\mathbf{y}^{(i)}))_V$
- Discrete norm: $\|v\|^2_{M,V} = \mathbb{E}_M[(v,v)_V]$

Let $\{\psi_\mathbf{p}\}_{\mathbf{p}\in\Lambda}$ be an orthonormal basis of $\mathbb{P}_\Lambda$ w.r.t the weight ρ.
Then, the best approximation of u in $\mathbb{P}_\Lambda(\Gamma)\otimes V$ (exact L^2 projection) is

$$P_\Lambda u = \underset{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V}{\operatorname{argmin}} \mathbb{E}[\|u-v\|^2_V] = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}[u\psi_\mathbf{p}]\psi_\mathbf{p}$$

How to compute an approx. projection using the random sample?
Replace the exact expectation $\mathbb{E}[\cdot]$ with the sample average $E_M[\cdot]$.

Notation

given two functions $u, v \in L^2_\rho(\Gamma; V)$

- Continuous inner product: $\mathbb{E}[(u,v)_V] = \int_\Gamma (u(\mathbf{y}), v(\mathbf{y}))_V \rho(\mathbf{y}) d\mathbf{y}$
- Continuous norm: $\|v\|^2_{L^2_\rho(\Gamma;V)} = \mathbb{E}[(v,v)_V]$.
- Discrete inner product: $\mathbb{E}_M[(u,v)_V] = \frac{1}{M}\sum_{i=1}^M (u(\mathbf{y}^{(i)}), v(\mathbf{y}^{(i)}))_V$
- Discrete norm: $\|v\|^2_{M,V} = \mathbb{E}_M[(v,v)_V]$

Let $\{\psi_\mathbf{p}\}_{\mathbf{p}\in\Lambda}$ be an orthonormal basis of $\mathbb{P}_\Lambda$ w.r.t the weight ρ.
Then, the best approximation of u in $\mathbb{P}_\Lambda(\Gamma) \otimes V$ (exact L^2 projection) is

$$P_\Lambda u = \underset{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V}{\operatorname{argmin}} \mathbb{E}[\|u-v\|^2_V] = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}[u\psi_\mathbf{p}]\psi_\mathbf{p}$$

How to compute an approx. projection using the random sample?
Replace the exact expectation $\mathbb{E}[\cdot]$ with the sample average $E_M[\cdot]$.

Notation

given two functions $u, v \in L^2_\rho(\Gamma; V)$

- Continuous inner product: $\mathbb{E}[(u,v)_V] = \int_\Gamma (u(\mathbf{y}), v(\mathbf{y}))_V \rho(\mathbf{y}) d\mathbf{y}$
- Continuous norm: $\|v\|^2_{L^2_\rho(\Gamma;V)} = \mathbb{E}[(v,v)_V]$.
- Discrete inner product: $\mathbb{E}_M[(u,v)_V] = \frac{1}{M}\sum_{i=1}^M (u(\mathbf{y}^{(i)}), v(\mathbf{y}^{(i)}))_V$
- Discrete norm: $\|v\|^2_{M,V} = \mathbb{E}_M[(v,v)_V]$

Let $\{\psi_\mathbf{p}\}_{\mathbf{p}\in\Lambda}$ be an orthonormal basis of $\mathbb{P}_\Lambda$ w.r.t the weight ρ.
Then, the best approximation of u in $\mathbb{P}_\Lambda(\Gamma) \otimes V$ (exact L^2 projection) is

$$P_\Lambda u = \underset{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V}{\operatorname{argmin}} \mathbb{E}[\|u-v\|^2_V] = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}[u\psi_\mathbf{p}]\psi_\mathbf{p}$$

How to compute an approx. projection using the random sample?
Replace the exact expectation $\mathbb{E}[\cdot]$ with the sample average $E_M[\cdot]$.

First idea (bad): discrete projection

We construct an approximation as

$$P_\Lambda^{M,\omega} u = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}_M[u\psi_\mathbf{p}]\psi_\mathbf{p} = \sum_{\mathbf{p}\in\Lambda} \left(\frac{1}{M} \sum_{i=1}^{M} u(\mathbf{y}^{(i)})\psi_\mathbf{p}(\mathbf{y}^{(i)}) \right) \psi_\mathbf{p}$$

Error analysis:

$$\|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} = \underbrace{\sum_{\mathbf{q}\notin\Lambda} \|\mathbb{E}[u\psi_\mathbf{q}]\|^2_V}_{L^2 \text{ projection error}} + \sum_{\mathbf{p}\in\Lambda} \underbrace{\|\mathbb{E}[u\psi_\mathbf{p}] - \mathbb{E}_M[u\psi_\mathbf{p}]\|^2_V}_{\text{Monte Carlo error}\sim O(M^{-1})}$$

from which, setting $K(\Lambda) = \sup_{\mathbf{y}\in\Gamma} \left(\sum_{\mathbf{p}\in\Lambda} |\psi_\mathbf{p}(\mathbf{y})|^2 \right)$, one can deduce

$$\mathbb{E}^\omega \|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} \leq \inf_{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|^2_{L^2_\rho(\Gamma;V)} + \frac{K(\Lambda)}{M} \|u\|^2_{L^2_\rho(\Gamma;V)}$$

Even for smooth functions the convergence is $O(\sqrt{M})$!

First idea (bad): discrete projection

We construct an approximation as

$$P_\Lambda^{M,\omega} u = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}_M[u\psi_\mathbf{p}]\psi_\mathbf{p} = \sum_{\mathbf{p}\in\Lambda} \left(\frac{1}{M} \sum_{i=1}^{M} u(\mathbf{y}^{(i)})\psi_\mathbf{p}(\mathbf{y}^{(i)}) \right) \psi_\mathbf{p}$$

Error analysis:

$$\|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} = \underbrace{\sum_{\mathbf{q}\notin\Lambda} \|\mathbb{E}[u\psi_\mathbf{q}]\|^2_V}_{L^2 \text{ projection error}} + \sum_{\mathbf{p}\in\Lambda} \underbrace{\|\mathbb{E}[u\psi_\mathbf{p}] - \mathbb{E}_M[u\psi_\mathbf{p}]\|^2_V}_{\text{Monte Carlo error} \sim O(M^{-1})}$$

from which, setting $K(\Lambda) = \sup_{\mathbf{y}\in\Gamma} \left(\sum_{\mathbf{p}\in\Lambda} |\psi_\mathbf{p}(\mathbf{y})|^2 \right)$, one can deduce

$$\mathbb{E}^\omega \|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} \leq \inf_{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|^2_{L^2_\rho(\Gamma;V)} + \frac{K(\Lambda)}{M} \|u\|^2_{L^2_\rho(\Gamma;V)}$$

Even for smooth functions the convergence is $O(\sqrt{M})$!

First idea (bad): discrete projection

We construct an approximation as

$$P_\Lambda^{M,\omega} u = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}_M[u\psi_\mathbf{p}]\psi_\mathbf{p} = \sum_{\mathbf{p}\in\Lambda} \left(\frac{1}{M} \sum_{i=1}^{M} u(\mathbf{y}^{(i)})\psi_\mathbf{p}(\mathbf{y}^{(i)}) \right) \psi_\mathbf{p}$$

Error analysis:

$$\|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} = \underbrace{\sum_{\mathbf{q}\notin\Lambda} \|\mathbb{E}[u\psi_\mathbf{q}]\|^2_V}_{L^2 \text{ projection error}} + \sum_{\mathbf{p}\in\Lambda} \underbrace{\|\mathbb{E}[u\psi_\mathbf{p}] - \mathbb{E}_M[u\psi_\mathbf{p}]\|^2_V}_{\text{Monte Carlo error} \sim O(M^{-1})}$$

from which, setting $K(\Lambda) = \sup_{\mathbf{y}\in\Gamma} \left(\sum_{\mathbf{p}\in\Lambda} |\psi_\mathbf{p}(\mathbf{y})|^2 \right)$, one can deduce

$$\mathbb{E}^\omega \|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} \leq \inf_{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|^2_{L^2_\rho(\Gamma;V)} + \frac{K(\Lambda)}{M} \|u\|^2_{L^2_\rho(\Gamma;V)}$$

Even for smooth functions the convergence is $O(\sqrt{M})$!

First idea (bad): discrete projection

We construct an approximation as

$$P_\Lambda^{M,\omega} u = \sum_{\mathbf{p}\in\Lambda} \mathbb{E}_M[u\psi_\mathbf{p}]\psi_\mathbf{p} = \sum_{\mathbf{p}\in\Lambda} \left(\frac{1}{M} \sum_{i=1}^{M} u(\mathbf{y}^{(i)})\psi_\mathbf{p}(\mathbf{y}^{(i)}) \right) \psi_\mathbf{p}$$

Error analysis:

$$\|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} = \underbrace{\sum_{\mathbf{q}\notin\Lambda} \|\mathbb{E}[u\psi_\mathbf{q}]\|^2_V}_{L^2 \text{ projection error}} + \sum_{\mathbf{p}\in\Lambda} \underbrace{\|\mathbb{E}[u\psi_\mathbf{p}] - \mathbb{E}_M[u\psi_\mathbf{p}]\|^2_V}_{\text{Monte Carlo error}\sim O(M^{-1})}$$

from which, setting $K(\Lambda) = \sup_{\mathbf{y}\in\Gamma} \left(\sum_{\mathbf{p}\in\Lambda} |\psi_\mathbf{p}(\mathbf{y})|^2 \right)$, one can deduce

$$\mathbb{E}^\omega \|u - P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)} \leq \inf_{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|^2_{L^2_\rho(\Gamma;V)} + \frac{K(\Lambda)}{M} \|u\|^2_{L^2_\rho(\Gamma;V)}$$

Even for smooth functions the convergence is $O(\sqrt{M})$!

Second idea (good): Discrete least squares approximation

(see e.g. [Hosder-Walters et al. 2010, Blatman-Sudret 2008, Burkardt-Eldred 2009, Eldred 2011, Yan-Guo-Xiu 2012, Cohen-Davenport-Leviatan 2013, Migliorati etal 2011-2014])

$$P_{\Lambda}^{M,\omega} u = \underset{v \in \mathbb{P}_{\Lambda}(\Gamma) \otimes V}{\operatorname{argmin}} \frac{1}{M} \sum_{k=1}^{M} \| u^{(k)} - v(\mathbf{y}^{(k)}) \|_V^2$$

Two relevant questions

- What is the accuracy of the random discrete least square approximation?
- For a given set Λ, how many samples should one use?

Second idea (good): Discrete least squares approximation

(see e.g. [Hosder-Walters et al. 2010, Blatman-Sudret 2008, Burkardt-Eldred 2009, Eldred 2011, Yan-Guo-Xiu 2012, Cohen-Davenport-Leviatan 2013, Migliorati etal 2011-2014])

$$P_\Lambda^{M,\omega} u = \operatorname*{argmin}_{v \in \mathbb{P}_\Lambda(\Gamma) \otimes V} \frac{1}{M} \sum_{k=1}^{M} \| u^{(k)} - v(\mathbf{y}^{(k)}) \|_V^2$$

Two relevant questions

- What is the accuracy of the random discrete least square approximation?
- For a given set Λ, how many samples should one use?

Algebraic formulation

- Let $V_h \subset V$ be a finite dimensional subspace (e.g. finite elements) and $\{\phi_j\}_{j=1}^{N_h}$ a basis; mass matrix $\mathcal{M}_{ij} = (\phi_j, \phi_i)_V$.
- Let $\{\psi_{\mathbf{p}}\}_{\mathbf{p}\in\Lambda}$ be an orthonormal basis of $\mathbb{P}_\Lambda$ w.r.t the weight ρ; define the design matrix $D_{i\mathbf{p}} = \frac{1}{\sqrt{M}}\psi_{\mathbf{p}}(\mathbf{y}^{(i)})$.

Then $P_\Lambda^{M,\omega} u(x,\mathbf{y}) = \sum_{\mathbf{p}\in\Lambda}\sum_{j=1}^{N_h} c_{\mathbf{p}j}\phi_j(x)\psi_{\mathbf{p}}(\mathbf{y})$ and the tensor $C = \{c_{\mathbf{p}j}\}$ satisfies the normal equations

$$(D^T D \otimes \mathcal{M})C = (D^T \otimes \mathcal{M})U$$

with $u(\mathbf{y}^{(i)}) = \sum_{j=1}^{N_h} u_j(\mathbf{y}^{(i)})\phi_j$ and $U_{ij} = \frac{1}{\sqrt{M}} u_j(\mathbf{y}^{(i)})$.

Since the matrix $\mathcal{M}$ is invertible, the previous problem decouples in N_h least square problems, one for each spatial dof.

$$D^T D C_{:,j} = D^T U_{:,j}$$

Algebraic formulation

- Let $V_h \subset V$ be a finite dimensional subspace (e.g. finite elements) and $\{\phi_j\}_{j=1}^{N_h}$ a basis; mass matrix $\mathcal{M}_{ij} = (\phi_j, \phi_i)_V$.
- Let $\{\psi_{\mathbf{p}}\}_{\mathbf{p}\in\Lambda}$ be an orthonormal basis of $\mathbb{P}_\Lambda$ w.r.t the weight ρ; define the design matrix $D_{i\mathbf{p}} = \frac{1}{\sqrt{M}}\psi_{\mathbf{p}}(\mathbf{y}^{(i)})$.

Then $P_\Lambda^{M,\omega} u(x,\mathbf{y}) = \sum_{\mathbf{p}\in\Lambda}\sum_{j=1}^{N_h} c_{\mathbf{p}j}\phi_j(x)\psi_{\mathbf{p}}(\mathbf{y})$ and the tensor $C = \{c_{\mathbf{p}j}\}$ satisfies the normal equations

$$(D^T D \otimes \mathcal{M})C = (D^T \otimes \mathcal{M})U$$

with $u(\mathbf{y}^{(i)}) = \sum_{j=1}^{N_h} u_j(\mathbf{y}^{(i)})\phi_j$ and $U_{ij} = \frac{1}{\sqrt{M}} u_j(\mathbf{y}^{(i)})$.

Since the matrix $\mathcal{M}$ is invertible, the previous problem decouples in N_h least square problems, one for each spatial dof.

$$D^T D C_{:,j} = D^T U_{:,j}$$

Algebraic formulation

- Let $V_h \subset V$ be a finite dimensional subspace (e.g. finite elements) and $\{\phi_j\}_{j=1}^{N_h}$ a basis; mass matrix $\mathcal{M}_{ij} = (\phi_j, \phi_i)_V$.
- Let $\{\psi_{\mathbf{p}}\}_{\mathbf{p}\in\Lambda}$ be an orthonormal basis of $\mathbb{P}_\Lambda$ w.r.t the weight ρ; define the design matrix $D_{i\mathbf{p}} = \frac{1}{\sqrt{M}}\psi_{\mathbf{p}}(\mathbf{y}^{(i)})$.

Then $P_\Lambda^{M,\omega} u(x,\mathbf{y}) = \sum_{\mathbf{p}\in\Lambda}\sum_{j=1}^{N_h} c_{\mathbf{p}j}\phi_j(x)\psi_{\mathbf{p}}(\mathbf{y})$ and the tensor $C = \{c_{\mathbf{p}j}\}$ satisfies the normal equations

$$(D^T D \otimes \mathcal{M})C = (D^T \otimes \mathcal{M})U$$

with $u(\mathbf{y}^{(i)}) = \sum_{j=1}^{N_h} u_j(\mathbf{y}^{(i)})\phi_j$ and $U_{ij} = \frac{1}{\sqrt{M}} u_j(\mathbf{y}^{(i)})$.

Since the matrix $\mathcal{M}$ is invertible, the previous problem decouples in N_h least square problems, one for each spatial dof.

$$D^T D C_{:,j} = D^T U_{:,j}$$

Error analysis

The error analysis goes through the *equivalence of norms* on the polynomial space. Remind:

- continuous norm: $\|v\|^2_{L^2_\rho(\Gamma;V)} = \int_\Gamma \|v(\mathbf{y})\|^2_V \rho(\mathbf{y}) d\mathbf{y}$
- discrete norm: $\|v\|^2_{M,V} = \frac{1}{M}\sum_{i=1}^M \|v(\mathbf{y}^{(i)})\|^2_V$

Define the random variable

$$\delta := \sup_{v\in\mathbb{P}_\Lambda(\Gamma)\otimes V} \left| \frac{\|v\|^2_{M,V}}{\|v\|^2_{L^2_\rho(\Gamma,V)}} - 1 \right|.$$

Whenever $\delta < 1$, we have norm equivalence

$$(1-\delta)\|v\|^2_{L^2_\rho(\Gamma,V)} \leq \|v\|^2_{M,V} \leq (1+\delta)\|v\|^2_{L^2_\rho(\Gamma,V)}, \quad \forall v \in \mathbb{P}_\Lambda(\Gamma)\otimes V$$

(analogous to RIP in compressed sensing, see [Candès-Tao 2006, Rahout-Ward 2012, ...])

Error analysis

The error analysis goes through the *equivalence of norms* on the polynomial space. Remind:

- continuous norm: $\|v\|^2_{L^2_\rho(\Gamma;V)} = \int_\Gamma \|v(\mathbf{y})\|^2_V \rho(\mathbf{y}) d\mathbf{y}$
- discrete norm: $\|v\|^2_{M,V} = \frac{1}{M}\sum_{i=1}^M \|v(\mathbf{y}^{(i)})\|^2_V$

Define the random variable

$$\delta := \sup_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \left| \frac{\|v\|^2_{M,V}}{\|v\|^2_{L^2_\rho(\Gamma,V)}} - 1 \right|.$$

Whenever $\delta < 1$, we have norm equivalence

$$(1-\delta)\|v\|^2_{L^2_\rho(\Gamma,V)} \le \|v\|^2_{M,V} \le (1+\delta)\|v\|^2_{L^2_\rho(\Gamma,V)}, \quad \forall v \in \mathbb{P}_\Lambda(\Gamma)\otimes V$$

(analogous to RIP in compressed sensing, see [Candès-Tao 2006, Rahout-Ward 2012, ...])

Theorem [Migliorati-Nobile-von Schwerin-Tempone '11]

1. $\delta \to 0$ almost surely when $M \to \infty$
2. $\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \le \left(1 + \sqrt{\frac{1}{1-\delta}}\right) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$

Proof: for any $v \in \mathbb{P}_\Lambda \otimes V$:

$$
\begin{aligned}
\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} &\le \|u - v\|_{L^2_\rho(\Gamma,V)} + \|v - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \\
&\le \|u - v\|_{L^2_\rho \otimes V} + \sqrt{\frac{1}{1-\delta}} \|v - P_\Lambda^{M,\omega} u\|_{M,V} \\
&\le \|u - v\|_{L^2_\rho(\Gamma,V)} + \sqrt{\frac{1}{1-\delta}} \|u - v\|_{M,V}
\end{aligned}
$$

A first bound on the constant $C_\delta = 1/\sqrt{1-\delta}$ has been given in [Migliorati-Nobile-von Schwerin-Tempone '11] in the case of 1 uniform random variable $y \sim \mathcal{U}([-1,1])$ using order statistics: $\forall \alpha \in (0,1)$, if $M \propto \#\Lambda^2$, then $C_\delta \le 3 \log \frac{M+1}{\alpha}$ with probability larger than $1-\alpha$.

Theorem [Migliorati-Nobile-von Schwerin-Tempone '11]

1. $\delta \to 0$ almost surely when $M \to \infty$
2. $\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \leq (1 + \sqrt{\frac{1}{1-\delta}}) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$

Proof: for any $v \in \mathbb{P}_\Lambda \otimes V$:

$$\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \|v - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)}$$

$$\leq \|u - v\|_{L^2_\rho \otimes V} + \sqrt{\frac{1}{1-\delta}} \|v - P_\Lambda^{M,\omega} u\|_{M,V}$$

$$\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \sqrt{\frac{1}{1-\delta}} \|u - v\|_{M,V}$$

A first bound on the constant $C_\delta = 1/\sqrt{1-\delta}$ has been given in

[Migliorati-Nobile-von Schwerin-Tempone '11] in the case of 1 uniform random variable $y \sim \mathcal{U}([-1,1])$ using order statistics: $\forall \alpha \in (0,1)$, if $M \propto \#\Lambda^2$, then $C_\delta \leq 3 \log \frac{M+1}{\alpha}$ with probability larger than $1 - \alpha$.

Theorem [Migliorati-Nobile-von Schwerin-Tempone '11]

1. $\delta \to 0$ almost surely when $M \to \infty$
2. $\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \leq (1 + \sqrt{\frac{1}{1-\delta}}) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$

Proof: for any $v \in \mathbb{P}_\Lambda \otimes V$:

$$\begin{aligned}\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \|v - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \\ &\leq \|u - v\|_{L^2_\rho \otimes V} + \sqrt{\frac{1}{1-\delta}} \|v - P_\Lambda^{M,\omega} u\|_{M,V} \\ &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \sqrt{\frac{1}{1-\delta}} \|u - v\|_{M,V}\end{aligned}$$

A first bound on the constant $C_\delta = 1/\sqrt{1-\delta}$ has been given in [Migliorati-Nobile-von Schwerin-Tempone '11] in the case of 1 uniform random variable $y \sim \mathcal{U}([-1,1])$ using order statistics: $\forall \alpha \in (0,1)$, if $M \propto \#\Lambda^2$, then $C_\delta \leq 3 \log \frac{M+1}{\alpha}$ with probability larger than $1 - \alpha$.

Theorem [Migliorati-Nobile-von Schwerin-Tempone '11]

1. $\delta \to 0$ almost surely when $M \to \infty$
2. $\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \leq \left(1 + \sqrt{\frac{1}{1-\delta}}\right) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$

Proof: for any $v \in \mathbb{P}_\Lambda \otimes V$:

$$\begin{aligned}\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \|v - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \\ &\leq \|u - v\|_{L^2_\rho \otimes V} + \sqrt{\frac{1}{1-\delta}} \|v - P_\Lambda^{M,\omega} u\|_{M,V} \\ &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \sqrt{\frac{1}{1-\delta}} \|u - v\|_{M,V}\end{aligned}$$

A first bound on the constant $C_\delta = 1/\sqrt{1-\delta}$ has been given in

[Migliorati-Nobile-von Schwerin-Tempone '11] in the case of 1 uniform random variable $y \sim \mathcal{U}([-1,1])$ using order statistics: $\forall \alpha \in (0,1)$, if $M \propto \#\Lambda^2$, then $C_\delta \leq 3 \log \frac{M+1}{\alpha}$ with probability larger than $1-\alpha$.

Theorem [Migliorati-Nobile-von Schwerin-Tempone '11]

1. $\delta \to 0$ almost surely when $M \to \infty$
2. $\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \leq (1 + \sqrt{\frac{1}{1-\delta}}) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$

Proof: for any $v \in \mathbb{P}_\Lambda \otimes V$:

$$\begin{aligned}\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \|v - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \\ &\leq \|u - v\|_{L^2_\rho \otimes V} + \sqrt{\frac{1}{1-\delta}} \|v - P_\Lambda^{M,\omega} u\|_{M,V} \\ &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \sqrt{\frac{1}{1-\delta}} \|u - v\|_{M,V}\end{aligned}$$

A first bound on the constant $C_\delta = 1/\sqrt{1-\delta}$ has been given in

[Migliorati-Nobile-von Schwerin-Tempone '11] in the case of 1 uniform random variable $y \sim \mathcal{U}([-1,1])$ using order statistics: $\forall \alpha \in (0,1)$, if $M \propto \#\Lambda^2$, then $C_\delta \leq 3 \log \frac{M+1}{\alpha}$ with probability larger than $1-\alpha$.

Theorem [Migliorati-Nobile-von Schwerin-Tempone '11]

1. $\delta \to 0$ almost surely when $M \to \infty$
2. $\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \leq \left(1 + \sqrt{\frac{1}{1-\delta}}\right) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$

Proof: for any $v \in \mathbb{P}_\Lambda \otimes V$:

$$\begin{aligned}
\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} &\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \|v - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma,V)} \\
&\leq \|u - v\|_{L^2_\rho \otimes V} + \sqrt{\frac{1}{1-\delta}} \|v - P_\Lambda^{M,\omega} u\|_{M,V} \\
&\leq \|u - v\|_{L^2_\rho(\Gamma,V)} + \sqrt{\frac{1}{1-\delta}} \|u - v\|_{M,V}
\end{aligned}$$

A first bound on the constant $C_\delta = 1/\sqrt{1-\delta}$ has been given in

[Migliorati-Nobile-von Schwerin-Tempone '11] in the case of 1 uniform random variable $y \sim \mathcal{U}([-1,1])$ using order statistics: $\forall \alpha \in (0,1)$, if $M \propto \#\Lambda^2$, then $C_\delta \leq 3 \log \frac{M+1}{\alpha}$ with probability larger than $1 - \alpha$.

A general result

[Cohen-Davenport-Leviatan '12], [Chkifa-Cohen-Migliorati-Nobile-Tempone '13]
Let $\{\psi_{\mathbf{p}}\}$ be any orthonormal basis of $L^2_\rho(\Gamma)$ and define

$$K(\Lambda) := \sup_{\mathbf{y}\in\Gamma}\left(\sum_{\mathbf{p}\in\Lambda}|\psi_{\mathbf{p}}(\mathbf{y})|^2\right) = \sup_{v\in\mathbb{P}_\Lambda}\frac{\|v\|^2_{L^\infty(\Gamma)}}{\|v\|^2_{L^2_\rho(\Gamma)}}$$

Theorem [Cohen-Davenport-Leviatan '13]

For any $\gamma > 0$, and $0 < \delta < 1$, and $\beta_\delta = \delta + (1-\delta)\log(1-\delta)$, if

$$\frac{M}{\log M} \geq \frac{1+\gamma}{\beta_\delta}K(\Lambda), \tag{2}$$

Then $P\left((1-\delta)\|v\|^2_{L^2_\rho(\Gamma;V)} \leq \|v\|^2_{M,V} \leq (1+\delta)\|v\|^2_{L^2_\rho(\Gamma;V)}\right) \geq 1 - 2M^{-\gamma}$

The result is based on properties of random matrices.

Implications

Convergence in probability: with probability greater than $1 - 2M^{-\gamma}$

$$\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma;V)} \leq \left(1 + \sqrt{\frac{1}{1-\delta}}\right) \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$$

Convergence in expectation: assume $\|u\|_{L^\infty(\Gamma,V)} \leq \tau$ and define the truncation operator

$$T_\tau : V \to V, \qquad T_\tau(v) = \begin{cases} v & \text{if } \|v\|_V \leq \tau \\ \frac{\tau}{\|v\|_V} v, & \text{if } \|v\|_V > \tau \end{cases}$$

Then $\quad \mathbb{E}^\omega(\|u - T_\tau P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)}) \leq C\|u - P_\Lambda u\|^2_{L^2_\rho(\Gamma,V)} + 8\tau^2 M^{-\gamma}$

Stability of discrete least squares: $cond(D^T D) \leq \frac{1+\delta}{1-\delta}$.

Implications

Convergence in probability: with probability greater than $1 - 2M^{-\gamma}$

$$\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma;V)} \leq \left(1 + \sqrt{\frac{1}{1-\delta}}\right) \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$$

Convergence in expectation: assume $\|u\|_{L^\infty(\Gamma,V)} \leq \tau$ and define the truncation operator

$$T_\tau : V \to V, \qquad T_\tau(v) = \begin{cases} v & \text{if } \|v\|_V \leq \tau \\ \frac{\tau}{\|v\|_V} v, & \text{if } \|v\|_V > \tau \end{cases}$$

Then $\quad \mathbb{E}^\omega(\|u - T_\tau P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)}) \leq C\|u - P_\Lambda u\|^2_{L^2_\rho(\Gamma,V)} + 8\tau^2 M^{-\gamma}$

Stability of discrete least squares: $cond(D^T D) \leq \frac{1+\delta}{1-\delta}$.

Implications

Convergence in probability: with probability greater than $1 - 2M^{-\gamma}$

$$\|u - P_\Lambda^{M,\omega} u\|_{L^2_\rho(\Gamma;V)} \leq \left(1 + \sqrt{\frac{1}{1-\delta}}\right) \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{L^\infty(\Gamma,V)}$$

Convergence in expectation: assume $\|u\|_{L^\infty(\Gamma,V)} \leq \tau$ and define the truncation operator

$$T_\tau : V \to V, \qquad T_\tau(v) = \begin{cases} v & \text{if } \|v\|_V \leq \tau \\ \frac{\tau}{\|v\|_V} v, & \text{if } \|v\|_V > \tau \end{cases}$$

Then $\mathbb{E}^\omega(\|u - T_\tau P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma;V)}) \leq C\|u - P_\Lambda u\|^2_{L^2_\rho(\Gamma,V)} + 8\tau^2 M^{-\gamma}$

Stability of discrete least squares: $cond(D^T D) \leq \frac{1+\delta}{1-\delta}$.

Uniform random vector in $[-1,1]^N$

Let $y_1, \dots, y_N$ be i.i.d. uniform random variables in $[-1,1]$.

Theorem [Chkifa-Cohen-Migliorati-Nobile-Tempone '13]

For any N and any downward closed set $\Lambda \subset \mathbb{N}^N$ it holds

$$K(\Lambda) \leq (\#\Lambda)^2.$$

Therefore, the discrete L^2 projection is stable and optimally convergent under the condition

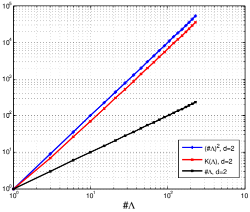
$$\frac{M}{\log M} \geq \frac{1+\gamma}{\beta_\delta}(\#\Lambda)^2$$

The result uses expansion on Legendre polynomials for which

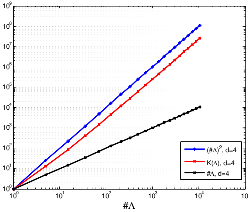
$$|\psi_{\mathbf{p}}(\mathbf{y})| \leq \prod_{n=1}^{N} \sqrt{2p_n+1}, \quad \forall \mathbf{y} \in [-1,1]^N.$$

Uniform random vector in $[-1,1]^N$

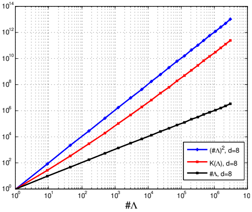
- For specific sets Λ the condition can be improved.
- For instance for the Total Degree polynomial space of degree w the bound $K(\Lambda) \leq (\#\Lambda)^2$ is very conservative



dimension $N = 2$



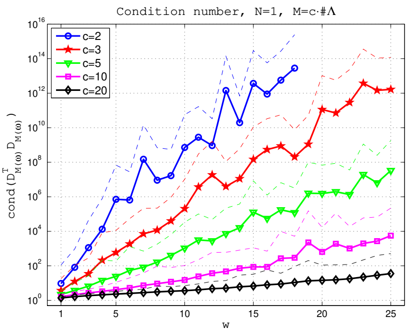
dimension $N = 4$



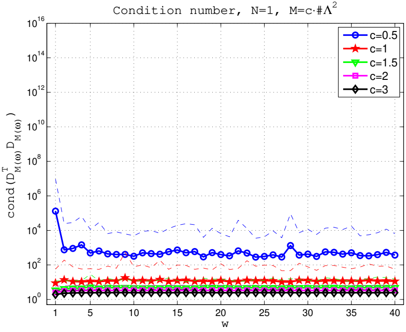
dimension $N = 8$

Some numerical examples – 1D function

Condition number of $D^T D$



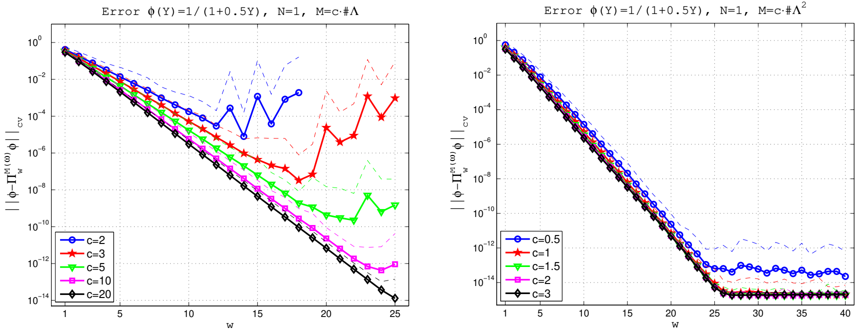
$M = c \cdot \#\Lambda$



$M = c \cdot (\#\Lambda)^2$

Some numerical examples – 1D function

Approximation of the meromorphic function $\phi(y) = \frac{1}{1+0.5y}$

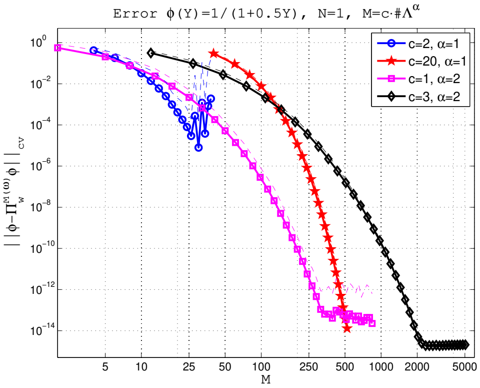


$M = c \cdot \#\Lambda$ $\qquad$ $M = c \cdot (\#\Lambda)^2$

error with respect to polynomial degree.

Some numerical examples – 1D function

Approximation of the meromorphic function $\phi(y) = \frac{1}{1+0.5y}$

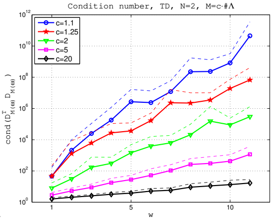


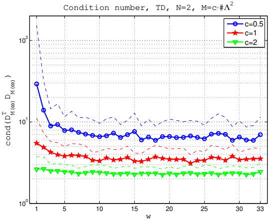
error with respect to total number of sampling points.

Some numerical examples

Condition number of $D^T D$ – multiD – Total Degree poly. space

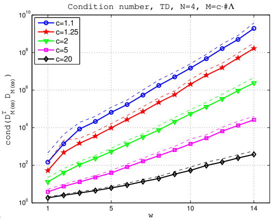
$M = c \cdot \#\Lambda$

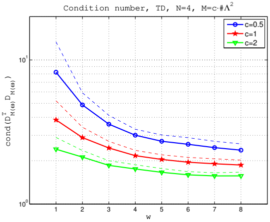




$M = c \cdot (\#\Lambda)^2$

$M = c \cdot \#\Lambda$





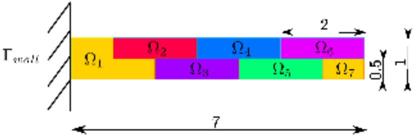
$M = c \cdot (\#\Lambda)^2$

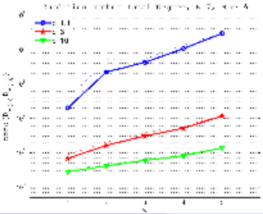
Cantilever beam

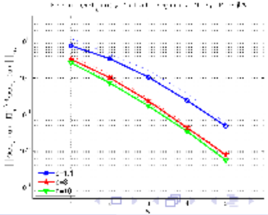
- linear elasticity equations
- Young modulus uncertain in each brick:

$$E_i = e^{7+Y_i}, \quad \text{in } \Omega_i,$$
$$Y_i \sim \mathcal{U}([-1,1]), \text{ iid}$$

- Uncertainty analysis on maximum vertical displacement.



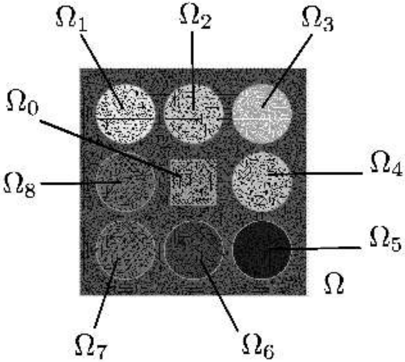


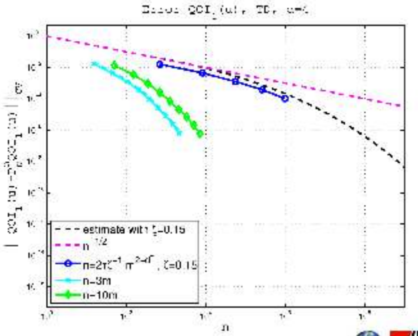


Elliptic PDE with random inclusions

The following bound has been derived in
[Chkifa-Cohen-Migliorati-Nobile-Tempone 13]

$$\mathbb{E}(\|u - T_\tau \circ P_\Lambda^{M,\omega} u\|^2_{L^2_\rho(\Gamma, H^1_0(D))}) \leq c_1 e^{-c_2 N M^{\frac{1}{1+2N}}}$$





4 random variables

Improvements on the quadratic relation

Improvements can be obtained by sampling from a different distribution $\hat{\rho}$.
Let us consider the weighted least squares approx.

$$P_\Lambda^M u = \underset{v \in \mathbb{P}_\Lambda(\Gamma) \otimes V}{\operatorname{argmin}} \frac{1}{M} \sum_{k=1}^{M} \frac{\rho(\mathbf{y}^{(k)})}{\hat{\rho}(\mathbf{y}^{(k)})} \| u^{(k)} - v(\mathbf{y}^{(k)}) \|_V^2$$

where the sample $\{\mathbf{y}^{(k)}\}_k$ is drawn from the distribution $\hat{\rho}(\mathbf{y})d\mathbf{y}$.

- $\rho(\mathbf{y}) = \hat{\rho}(\mathbf{y}) =$ Chebyshev distribution in $[-1,1]^N$, then the relation $M \propto \min\{2^N \# \Lambda, (\# \Lambda)^{\frac{\log(3)}{\log(2)}}\}$ is enough to guarantee optimal convergence [Chkifa-Cohen-Migliorati-N.-Tempone '13]
- $\rho(\mathbf{y})$=uniform and $\hat{\rho}(\mathbf{y})$=Chebyshev distribution in $[-1,1]^N$, then, the relation $M \propto 2^N \# \Lambda$ guarantees optimal convergence [Rauhut-Ward '12]. However, the constant depends on N [Yan-Guo-Xiu '12].
- $\rho(\mathbf{y})$=Gaussian: still unclear. Numerically, the situation seems to be worse. Improvements suggestd in [Tang-Zhou '14].

Improvements on the quadratic relation

Improvements can be obtained by sampling from a different distribution $\hat{\rho}$. Let us consider the weighted least squares approx.

$$P_\Lambda^M u = \underset{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V}{\operatorname{argmin}} \frac{1}{M} \sum_{k=1}^{M} \frac{\rho(\mathbf{y}^{(k)})}{\hat{\rho}(\mathbf{y}^{(k)})} \| u^{(k)} - v(\mathbf{y}^{(k)}) \|_V^2$$

where the sample $\{\mathbf{y}^{(k)}\}_k$ is drawn from the distribution $\hat{\rho}(\mathbf{y})d\mathbf{y}$.

- $\rho(\mathbf{y}) = \hat{\rho}(\mathbf{y}) =$ Chebyshev distribution in $[-1,1]^N$, then the relation $M \propto \min\{2^N \#\Lambda, (\#\Lambda)^{\frac{\log(3)}{\log(2)}}\}$ is enough to guarantee optimal convergence [Chkifa-Cohen-Migliorati-N.-Tempone '13]
- $\rho(\mathbf{y})$=uniform and $\hat{\rho}(\mathbf{y})$=Chebyshev distribution in $[-1,1]^N$, then, the relation $M \propto 2^N \#\Lambda$ guarantees optimal convergence [Rauhut-Ward '12]. However, the constant depends on N [Yan-Guo-Xiu '12].
- $\rho(\mathbf{y})$=Gaussian: still unclear. Numerically, the situation seems to be worse. Improvements suggestd in [Tang-Zhou '14].

Improvements on the quadratic relation

Improvements can be obtained by sampling from a different distribution $\hat{\rho}$. Let us consider the weighted least squares approx.

$$P_\Lambda^M u = \underset{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V}{\operatorname{argmin}} \frac{1}{M}\sum_{k=1}^{M} \frac{\rho(\mathbf{y}^{(k)})}{\hat{\rho}(\mathbf{y}^{(k)})} \|u^{(k)} - v(\mathbf{y}^{(k)})\|_V^2$$

where the sample $\{\mathbf{y}^{(k)}\}_k$ is drawn from the distribution $\hat{\rho}(\mathbf{y})d\mathbf{y}$.

- $\rho(\mathbf{y}) = \hat{\rho}(\mathbf{y}) =$ Chebyshev distribution in $[-1,1]^N$, then the relation $M \propto \min\{2^N \#\Lambda, (\#\Lambda)^{\frac{\log(3)}{\log(2)}}\}$ is enough to guarantee optimal convergence [Chkifa-Cohen-Migliorati-N.-Tempone '13]
- $\rho(\mathbf{y})$=uniform and $\hat{\rho}(\mathbf{y})$=Chebyshev distribution in $[-1,1]^N$, then, the relation $M \propto 2^N \#\Lambda$ guarantees optimal convergence [Rauhut-Ward '12]. However, the constant depends on N [Yan-Guo-Xiu '12].
- $\rho(\mathbf{y})$=Gaussian: still unclear. Numerically, the situation seems to be worse. Improvements suggestd in [Tang-Zhou '14].

Improvements on the quadratic relation

Improvements can be obtained by sampling from a different distribution $\hat{\rho}$. Let us consider the weighted least squares approx.

$$P_\Lambda^M u = \underset{v \in \mathbb{P}_\Lambda(\Gamma)\otimes V}{\operatorname{argmin}} \frac{1}{M} \sum_{k=1}^{M} \frac{\rho(\mathbf{y}^{(k)})}{\hat{\rho}(\mathbf{y}^{(k)})} \|u^{(k)} - v(\mathbf{y}^{(k)})\|_V^2$$

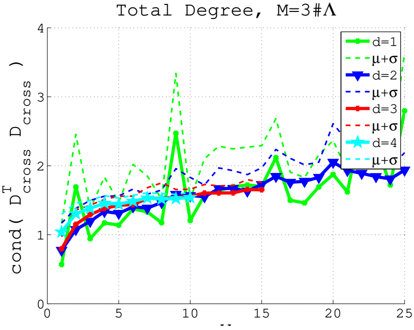
where the sample $\{\mathbf{y}^{(k)}\}_k$ is drawn from the distribution $\hat{\rho}(\mathbf{y})d\mathbf{y}$.

- $\rho(\mathbf{y}) = \hat{\rho}(\mathbf{y}) =$ Chebyshev distribution in $[-1,1]^N$, then the relation $M \propto \min\{2^N \#\Lambda, (\#\Lambda)^{\frac{\log(3)}{\log(2)}}\}$ is enough to guarantee optimal convergence [Chkifa-Cohen-Migliorati-N.-Tempone '13]
- $\rho(\mathbf{y})$=uniform and $\hat{\rho}(\mathbf{y})$=Chebyshev distribution in $[-1,1]^N$, then, the relation $M \propto 2^N \#\Lambda$ guarantees optimal convergence [Rauhut-Ward '12]. However, the constant depends on N [Yan-Guo-Xiu '12].
- $\rho(\mathbf{y})$=Gaussian: still unclear. Numerically, the situation seems to be worse. Improvements suggestd in [Tang-Zhou '14]

Numerical example with Chebyshev preconditioning

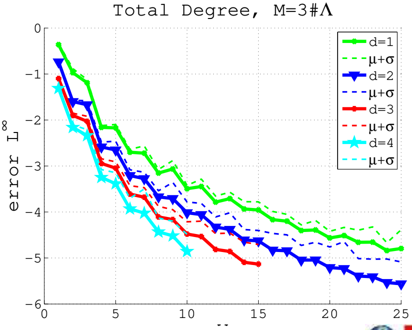
Expansion in Legendre polynomials ($\rho(\mathbf{y})$=uniform) and samples from Chebyshev distribution ($\hat{\rho}(\mathbf{y})$=Chebyshev)

condition number $cond(D^T D)$



$M = 3 \cdot \#\Lambda$

error for $u(y) = \left(1 + \frac{0.7}{2N}\sum_{n=1}^{N} y_n\right)^{-1}$



$M = 3 \cdot \#\Lambda$

Outline

Adaptive construction of polynomial spaces

$\{\Lambda_k\}_{k\geq 0}$ sequence of downward closed multi-index sets, with $\Lambda_0 = \{\mathbf{0}\}$.
The sequence is adaptively computed by means of greedy algorithms based on the random discrete L^2 projection.

Definitions:

- Margin $\mathcal{M}(\Lambda)$ associated to a multi-index set Λ:

$$\mathcal{M}(\Lambda) = \{\mathbf{p} \ : \ \mathbf{p} \notin \Lambda \text{ and } \exists j > 0 \ : \ \mathbf{p} - \mathbf{e}_j \in \Lambda\}$$

- Reduced margin $\mathcal{R}(\Lambda)$ associated to a multi-index set Λ:

$$\mathcal{R}(\Lambda) = \{\mathbf{p} \ : \ \mathbf{p} \notin \Lambda \text{ and } \forall j = 1, \dots, d \ : \ p_j \neq 0 \Rightarrow \mathbf{p} - \mathbf{e}_j \in \Lambda\}$$

set Λ and its Margin   set Λ and its Reduced margin

Adaptive construction of polynomial spaces

$\{\Lambda_k\}_{k\geq 0}$ sequence of downward closed multi-index sets, with $\Lambda_0 = \{\mathbf{0}\}$.
The sequence is adaptively computed by means of greedy algorithms based on the random discrete L^2 projection.

Definitions:

- Margin $\mathcal{M}(\Lambda)$ associated to a multi-index set Λ:

$$\mathcal{M}(\Lambda) = \{\mathbf{p} \ : \ \mathbf{p} \notin \Lambda \text{ and } \exists j > 0 \ : \ \mathbf{p} - \mathbf{e}_j \in \Lambda\}$$

- Reduced margin $\mathcal{R}(\Lambda)$ associated to a multi-index set Λ:

$$\mathcal{R}(\Lambda) = \{\mathbf{p} \ : \ \mathbf{p} \notin \Lambda \text{ and } \forall j = 1, \ldots, d \ : \ p_j \neq 0 \Rightarrow \mathbf{p} - \mathbf{e}_j \in \Lambda\}$$

set Λ and its Margin set Λ and its Reduced margin

Adaptive construction of polynomial spaces

$\{\Lambda_k\}_{k\geq 0}$ sequence of downward closed multi-index sets, with $\Lambda_0 = \{\mathbf{0}\}$.
The sequence is adaptively computed by means of greedy algorithms based on the random discrete L^2 projection.

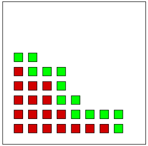
Definitions:

- Margin $\mathcal{M}(\Lambda)$ associated to a multi-index set Λ:

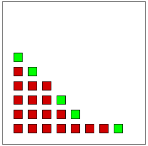
$$\mathcal{M}(\Lambda) = \{\mathbf{p} \ : \ \mathbf{p} \notin \Lambda \text{ and } \exists j > 0 \ : \ \mathbf{p} - \mathbf{e}_j \in \Lambda\}$$

- Reduced margin $\mathcal{R}(\Lambda)$ associated to a multi-index set Λ:

$$\mathcal{R}(\Lambda) = \{\mathbf{p} \ : \ \mathbf{p} \notin \Lambda \text{ and } \forall j = 1, \ldots, d \ : \ p_j \neq 0 \Rightarrow \mathbf{p} - \mathbf{e}_j \in \Lambda\}$$



set Λ and its Margin



set Λ and its Reduced margin

The Dörfler marking

Idea proposed by W. Dörfler in 1996 for Adaptive Finite Elements.

Given a multi-index set Λ, a subset $R \subseteq \mathcal{R}(\Lambda)$, a (continuous) function $e : R \to \mathbb{R}$ and a parameter $\theta \in (0, 1]$, we define a procedure

$$\text{Dörfler} = \text{Dörfler}(R, e, \theta)$$

that computes a set $F \subseteq R \subseteq \mathcal{R}(\Lambda)$ of minimal cardinality such that

$$\sum_{\mathbf{p} \in F} e(\mathbf{p})^2 \geq \theta \sum_{\mathbf{p} \in R} e(\mathbf{p})^2.$$

In practice, for any $\mathbf{p} \in R$, the error indicator $e(\mathbf{p})$ will be either an estimator of the coefficient $c_{\mathbf{p}}$ of the function u expanded over the Legendre basis or the projected residual on the $\mathbf{p}$-th Legendre basis function.

This corresponds to choose a fraction θ of the energy associated with the (estimates of the) coefficients in the set R.

The Dörfler marking

Idea proposed by W. Dörfler in 1996 for Adaptive Finite Elements.

Given a multi-index set Λ, a subset $R \subseteq \mathcal{R}(\Lambda)$, a (continuous) function $e : R \to \mathbb{R}$ and a parameter $\theta \in (0, 1]$, we define a procedure

$$\mathrm{Dörfler} = \mathrm{Dörfler}(R, e, \theta)$$

that computes a set $F \subseteq R \subseteq \mathcal{R}(\Lambda)$ of minimal cardinality such that

$$\sum_{\mathbf{p} \in F} e(\mathbf{p})^2 \geq \theta \sum_{\mathbf{p} \in R} e(\mathbf{p})^2.$$

In practice, for any $\mathbf{p} \in R$, the error indicator $e(\mathbf{p})$ will be either an estimator of the coefficient $c_{\mathbf{p}}$ of the function u expanded over the Legendre basis or the projected residual on the $\mathbf{p}$-th Legendre basis function.

This corresponds to choose a fraction θ of the energy associated with the (estimates of the) coefficients in the set R.

Orthogonal Matching Pursuit with Dörfler marking

Algorithm 1 Orthogonal Matching Pursuit with Dörfler marking

Set $r_0 = u(\mathbf{y})$, $u_0 \equiv 0$ and $\Lambda_0 = \{\mathbf{0}\}$,
for $k = 1, \dots, k_{max}$ **do**
$F_1 = \text{Dörfler}(\mathcal{R}(\Lambda_{k-1}), \{|(r_{k-1}, \psi_{\mathbf{p}})_{M,V}|\}_{\mathbf{p}}, \theta_1)$
$\widetilde{\Lambda}_k = \Lambda_{k-1} \cup F_1$
$u_k = \text{argmin}_{v \in \mathbb{P}_{\widetilde{\Lambda}_k}} \|u - v\|_{M,V}, \qquad u_k = \sum_{\mathbf{p} \in \tilde{\Lambda}_k} c_{\mathbf{p}}^{(k)} \psi_{\mathbf{p}}$
$F_2 = \text{Dörfler}(F_1, \{c_{\mathbf{p}}^{(k)}\}_{\mathbf{p}}, \theta_2)$
$\Lambda_k = \Lambda_{k-1} \cup F_2$
$r_k = u - u_k|_{\Lambda_k}$
end for

$\theta_1 \in (0,1)$ and $\theta_2 = 1$: Dörfler marking only with the correlations.

$\theta_1 = 1$ and $\theta_2 \in (0,1)$: Dörfler marking only with the random discrete L^2 projection on $\Lambda_{k-1} \cup \mathcal{R}(\Lambda_k)$.

Some remarks and open issues

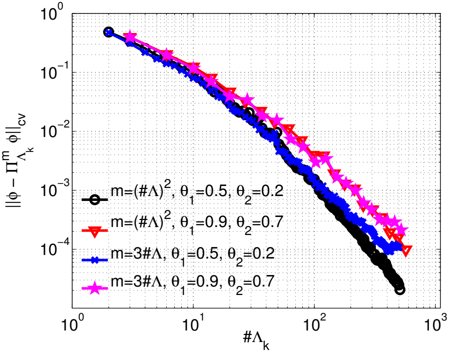
- The first Dörfler marking performs a screening of the reduced margin, to avoid an L^2 discrete minimization over a too large polynomial space.
- At each step the correlations $\{|(r_{k-1}, \psi_{\mathbf{p}})_{M,V}| \ : \ \mathbf{p} \in \mathcal{R}(\Lambda_k)\}$ are mutually uncoupled and cheap to compute, but might provide only a rough estimate of the coefficients (depending on the choice of M_k).
- The second Dörfler marking performs a selection based on the more accurate estimates of the coefficients coming from the L^2 projection.
- At each step the adaptive algorithm remains stable and accurate by choosing $M_k \propto (\#\Lambda_k)^2$ (consequence of the theory in the first part).
- The adaptive algorithm generates a sequence $\{\Lambda_k\}_{k\geq 0}$ of only quasi best N-term sets.
- Rate of convergence? Choice of θ_1, θ_2? What if $M_k \propto \#\Lambda_k$?

A numerical test

Approximation of a meromorphic function (16-variables)

$$\phi(\mathbf{y}) = \frac{1}{1 - \boldsymbol{\gamma} \cdot \mathbf{y}}, \quad \mathbf{y} \sim \mathcal{U}([-1,1]^{16})$$

$$\gamma = 0.3 * (1, 5 \cdot 10^{-1}, 10^{-1}, 5 \cdot 10^{-2}, \ldots, 5 \cdot 10^{-8})$$



Outline

1. Introduction – parametric / stochastic equations
2. Stochastic polynomial approximation
3. Discrete least squares approx. using random evaluations
   - Convergence analysis
   - Numerical results
4. Adaptive algorithms
5. Conclusions

Conclusions

- We have derived conditions under which the random discrete least squares approximation is stable and optimally convergent.
- The condition $M \geq C(\#\Lambda)^2$ for uniform random variables holds in any dimension and for any "shape" of the polynomial space, opening the possibility for adaptive algorithms.
- The condition $M \sim (\#\Lambda)^2$ seems to be too stringent in high dimension and a linear scaling is often enough, making this technique more attractive for high dimensional problems.
- Still open questions on preconditioned least squares or unbounded random variables.
- We have proposed an adaptive algorithm based on a double Dörfler marking that performs very well. The analysis is still ongoing.

Conclusions

- We have derived conditions under which the random discrete least squares approximation is stable and optimally convergent.
- The condition $M \geq C(\#\Lambda)^2$ for uniform random variables holds in any dimension and for any "shape" of the polynomial space, opening the possibility for adaptive algorithms.
- The condition $M \sim (\#\Lambda)^2$ seems to be too stringent in high dimension and a linear scaling is often enough, making this technique more attractive for high dimensional problems.
- Still open questions on preconditioned least squares or unbounded random variables.
- We have proposed an adaptive algorithm based on a double Dörfler marking that performs very well. The analysis is still ongoing.

Conclusions

- We have derived conditions under which the random discrete least squares approximation is stable and optimally convergent.
- The condition $M \geq C(\#\Lambda)^2$ for uniform random variables holds in any dimension and for any "shape" of the polynomial space, opening the possibility for adaptive algorithms.
- The condition $M \sim (\#\Lambda)^2$ seems to be too stringent in high dimension and a linear scaling is often enough, making this technique more attractive for high dimensional problems.
- Still open questions on preconditioned least squares or unbounded random variables.
- We have proposed an adaptive algorithm based on a double Dörfler marking that performs very well. The analysis is still ongoing.

Conclusions

- We have derived conditions under which the random discrete least squares approximation is stable and optimally convergent.
- The condition $M \geq C(\#\Lambda)^2$ for uniform random variables holds in any dimension and for any "shape" of the polynomial space, opening the possibility for adaptive algorithms.
- The condition $M \sim (\#\Lambda)^2$ seems to be too stringent in high dimension and a linear scaling is often enough, making this technique more attractive for high dimensional problems.
- Still open questions on preconditioned least squares or unbounded random variables.
- We have proposed an adaptive algorithm based on a double Dörfler marking that performs very well. The analysis is still ongoing.

Conclusions

- We have derived conditions under which the random discrete least squares approximation is stable and optimally convergent.
- The condition $M \geq C(\#\Lambda)^2$ for uniform random variables holds in any dimension and for any "shape" of the polynomial space, opening the possibility for adaptive algorithms.
- The condition $M \sim (\#\Lambda)^2$ seems to be too stringent in high dimension and a linear scaling is often enough, making this technique more attractive for high dimensional problems.
- Still open questions on preconditioned least squares or unbounded random variables.
- We have proposed an adaptive algorithm based on a double Dörfler marking that performs very well. The analysis is still ongoing.

Thank you for your attention!

References

A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone
Discrete least squares polynomial approximation with random evaluations application to parametric and stochastic elliptic PDEs, MATHICSE Tech. Rep. 35.2013. Submitted.

G. Migliorati,
Adaptive polynomial approximation by means of random discrete least squares, to appear in ENUMATH 2013 Proceedings, LNCSE Springer.

G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone
Analysis of the discrete L^2 projection on polynomial spaces with random evaluations, Found. Comp. Math., 2014. available online.

G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone,
Approximation of quantities of interest in stochastic PDEs by the random discrete L2 projection on polynomial spaces, SISC 35(3), 2013

J. Beck, F. Nobile, L. Tamellini, and R. Tempone.
Convergence of quasi-optimal stochastic Galerkin methods for a class of PDEs with random coefficients, Comput. & Math. with Appl., 2013