Stochastic collocation and MLMC methods for elliptic PDEs with random coefficients

Fabio Nobile

CSQI-MATHICSE, EPFL, Switzerland

Acknowlegments: L. Tamellini, F. Tesei (EPFL),
R. Tempone, (KAUST), J. Beck (UCL)

Workshop "Partial Differential Equations with Random Coefficients"
WIAS, Berlin, 13 15 November, 2013

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

Center for Advanced Modeling and Science

Outline

1. Model problem: elliptic PDEs with random coefficients
2. Polynomial approximation by sparse grid collocation
   - Quasi optimal sparse grid construction
   - Convergence result for elliptic PDEs with random inclusions
   - Numerical results
3. Combined MLMC / Sparse grid method
   - MLMC with Sparse Grid Control Variate
   - Variance analysis and algorithm tuning
   - Numerical results
4. Conclusions

Outline

1. Model problem: elliptic PDEs with random coefficients
2. Polynomial approximation by sparse grid collocation
   - Quasi optimal sparse grid construction
   - Convergence result for elliptic PDEs with random inclusions
   - Numerical results
3. Combined MLMC / Sparse grid method
   - MLMC with Sparse Grid Control Variate
   - Variance analysis and algorithm tuning
   - Numerical results
4. Conclusions

Model problem: elliptic PDE with random coeffs

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $D \subset \mathbb{R}^d$ and open bounded domain.

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

with $f \in L^2(D)$ and $a(\omega, x) : \Omega \times D \to \mathbb{R}$ an almost surely bounded random field.

Coercivity assumption: $a_{min}(\omega) = \operatorname{essinf}_{x \in D} a(\omega, x) > 0$ almost surely and $\mathbb{E}[a_{min}^{-\bar{p}}] < \infty$ for some $\bar{p} \geq 2$.

Then $u \in V = H_0^1(D)$ almost surely and

$$\|u(\omega, \cdot)\|_V \leq \frac{C_P}{a_{min}(\omega)} \|f\|_{L^2(D)}, \quad \text{a.s. in } \Omega$$

Therefore, $u \in L_P^p(\Omega, V)$ for all $p \leq \bar{p}$. In particular, $u \in L_P^2(\Omega, V)$.

Model problem: elliptic PDE with random coeffs

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $D \subset \mathbb{R}^d$ and open bounded domain.

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

with $f \in L^2(D)$ and $a(\omega,x) : \Omega \times D \to \mathbb{R}$ an almost surely bounded random field.

Coercivity assumption: $a_{min}(\omega) = \operatorname{essinf}_{x \in D} a(\omega,x) > 0$ almost surely and $\mathbb{E}[a_{min}^{-\bar{p}}] < \infty$ for some $\bar{p} \geq 2$.

Then $u \in V = H_0^1(D)$ almost surely and

$$\|u(\omega,\cdot)\|_V \leq \frac{C_P}{a_{min}(\omega)} \|f\|_{L^2(D)}, \quad \text{a.s. in } \Omega$$

Therefore, $u \in L_P^p(\Omega, V)$ for all $p \leq \bar{p}$. In particular, $u \in L_P^2(\Omega, V)$.

Model problem: elliptic PDE with random coeffs

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $D \subset \mathbb{R}^d$ and open bounded domain.

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

with $f \in L^2(D)$ and $a(\omega, x) : \Omega \times D \to \mathbb{R}$ an almost surely bounded random field.

Coercivity assumption: $a_{min}(\omega) = \operatorname{essinf}_{x \in D} a(\omega, x) > 0$ almost surely and $\mathbb{E}[a_{min}^{-\bar{p}}] < \infty$ for some $\bar{p} \geq 2$.

Then $u \in V = H_0^1(D)$ almost surely and

$$\|u(\omega, \cdot)\|_V \leq \frac{C_P}{a_{min}(\omega)} \|f\|_{L^2(D)}, \quad \text{a.s. in } \Omega$$

Therefore, $u \in L_P^p(\Omega, V)$ for all $p \leq \bar{p}$. In particular, $u \in L_P^2(\Omega, V)$.

Random field parametrization

In view of polynomial approximation we parametrize the random field $a(\omega, x)$ by a finite or countable sequence of random variables $\mathbf{y}(\omega) = (y_1(\omega), \ldots, y_N(\omega))$ with range $\Gamma = \mathbf{y}(\Omega) \subset \mathbb{R}^N$ and probability density function $\rho : \Gamma \to \mathbb{R}_+$:

$$a(\omega, x) = a(\mathbf{y}(\omega), x)$$

Then the stochastic solution u depends on ω only through the vector $\mathbf{y}(\omega)$: $u(\omega, x) = u(\mathbf{y}(\omega), x)$

parameter-to-solution map: $u(\mathbf{y}) : \Gamma \to V, \quad u \in L^2_\rho(\Gamma, V).$

Random field parametrization

In view of polynomial approximation we parametrize the random field $a(\omega, x)$ by a finite or countable sequence of random variables $\mathbf{y}(\omega) = (y_1(\omega), \ldots, y_N(\omega))$ with range $\Gamma = \mathbf{y}(\Omega) \subset \mathbb{R}^N$ and probability density function $\rho : \Gamma \to \mathbb{R}_+$:

$$a(\omega, x) = a(\mathbf{y}(\omega), x)$$

Then the stochastic solution u depends on ω only through the vector $\mathbf{y}(\omega)$: $u(\omega, x) = u(\mathbf{y}(\omega), x)$

parameter-to-solution map: $u(\mathbf{y}) : \Gamma \to V, \quad u \in L^2_\rho(\Gamma, V).$

Random field parametrization

In view of polynomial approximation we parametrize the random field $a(\omega, x)$ by a finite or countable sequence of random variables $\mathbf{y}(\omega) = (y_1(\omega), \ldots, y_N(\omega))$ with range $\Gamma = \mathbf{y}(\Omega) \subset \mathbb{R}^N$ and probability density function $\rho : \Gamma \to \mathbb{R}_+$:

$$a(\omega, x) = a(\mathbf{y}(\omega), x)$$

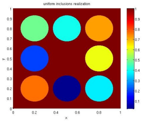
Then the stochastic solution u depends on ω only through the vector $\mathbf{y}(\omega)$: $u(\omega, x) = u(\mathbf{y}(\omega), x)$

parameter-to-solution map: $\quad u(\mathbf{y}) : \Gamma \to V, \quad u \in L^2_\rho(\Gamma, V).$

Examples of random diffusion coefficients

Inclusions problem

y describes the conductivity in each inclusion



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

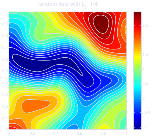
with $y_n \sim \mathcal{U}([y_{min}, y_{max}])$ and $y_{min} > -a_0$.

Therefore $a_{min}(\mathbf{y}) \in L^p_\rho(\Gamma)$ for any $1 \leq p \leq \infty$.

$$\Longrightarrow \quad u \in L^p_\rho(\Gamma, H^1_0(D)),\ \forall 1 \leq p \leq \infty$$

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



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

If $Cov[\gamma]$ is Holder continuous, then $a_{min} \in L^p_\rho(\Gamma)$ for any $1 \leq p < \infty$ (see e.g. [Charrier, 2011])

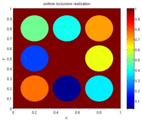
$$\Longrightarrow \quad u \in L^p_\rho(\Gamma, H^1_0(D)),\ \forall 1 \leq p < \infty$$

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Examples of random diffusion coefficients

Inclusions problem

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



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

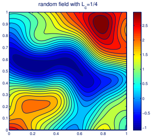
with $y_n \sim \mathcal{U}([y_{min}, y_{max}])$ and $y_{min} > -a_0$.

Therefore $a_{min}(\mathbf{y}) \in L^p_\rho(\Gamma)$ for any $1 \leq p \leq \infty$.

$$\implies \quad u \in L^p_\rho(\Gamma, H^1_0(D)), \ \forall 1 \leq p \leq \infty$$

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



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

If $Cov[\gamma]$ is Holder continuous, then $a_{min} \in L^p_\rho(\Gamma)$ for any $1 \leq p < \infty$ (see e.g. [Charrier, 2011])

$$\implies \quad u \in L^p_\rho(\Gamma, H^1_0(D)), \ \forall 1 \leq p < \infty$$

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

1 Model problem: elliptic PDEs with random coefficients

2 Polynomial approximation by sparse grid collocation
- Quasi optimal sparse grid construction
- Convergence result for elliptic PDEs with random inclusions
- Numerical results

3 Combined MLMC / Sparse grid method
- MLMC with Sparse Grid Control Variate
- Variance analysis and algorithm tuning
- Numerical results

4 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

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

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

Collocation approaches

Construct a polynomial approximation of $u(\mathbf{y}) : \Gamma \to V$ using only point evaluations $u_i = u(\mathbf{y}_i)$ where $\{\mathbf{y}_i\}_{i=1}^{\tilde{M}}$ is a set of suitable collocation points, with $\tilde{M} \geq M$.

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

Collocation approaches

Construct a polynomial approximation of $u(\mathbf{y}) : \Gamma \to V$ using only point evaluations $u_i = u(\mathbf{y}_i)$ where $\{\mathbf{y}_i\}_{i=1}^{\widetilde{M}}$ is a set of suitable collocation points, with $\widetilde{M} \geq M$.

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

Collocation approaches

Construct a polynomial approximation of $u(\mathbf{y}) : \Gamma \to V$ using only point evaluations $u_i = u(\mathbf{y}_i)$ where $\{\mathbf{y}_i\}_{i=1}^{\tilde{M}}$ is a set of suitable collocation points, with $\widetilde{M} \geq M$.

Collocation on a (generalized) Sparse Grid

Let $\mathbf{i} = [i_1, \dots, i_N] \in \mathbb{N}_+^N$ and $m(i) : \mathbb{N}_+ \to \mathbb{N}_+$ an increasing function

1. 1D polynomial interpolant operators: $\mathscr{U}_n^{m(i_n)}$ on $m(i_n)$ abscissas. We use either
   - Clenshaw-Curtis (extrema on Chebyshev polynomials)
   - Gauss points w.r.t. the weight ρ_n, assuming that the probability density factorizes as $\rho(\mathbf{y}) = \prod_{n=1}^N \rho_n(y_n)$
2. Detail operator: $\Delta_n^{m(i_n)} = \mathscr{U}_n^{m(i_n)} - \mathscr{U}_n^{m(i_n-1)}$, $\mathscr{U}_n^{m(0)} = 0$.
3. Hierarchical surplus: $\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^N \Delta_n^{m(i_n)}$.
4. Sparse grid approximation: on an index set $\mathcal{I} \subset \mathbb{N}^N$

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} \Delta^{m(\mathbf{i})}[u]$$

Assumption: The set $\mathcal{I}$ is downward closed:
$\mathbf{i} \in \mathcal{I} \;\Rightarrow\; \mathbf{i} - \mathbf{e}_n \in \mathcal{I}$, $n = 1, \dots, N$.

Collocation on a (generalized) Sparse Grid

Let $\mathbf{i} = [i_1, \ldots, i_N] \in \mathbb{N}_+^N$ and $m(i) : \mathbb{N}_+ \to \mathbb{N}_+$ an increasing function

1. 1D polynomial interpolant operators: $\mathscr{U}_n^{m(i_n)}$ on $m(i_n)$ abscissas. We use either
   - Clenshaw-Curtis (extrema on Chebyshev polynomials)
   - Gauss points w.r.t. the weight ρ_n, assuming that the probability density factorizes as $\rho(\mathbf{y}) = \prod_{n=1}^N \rho_n(y_n)$
2. Detail operator: $\Delta_n^{m(i_n)} = \mathscr{U}_n^{m(i_n)} - \mathscr{U}_n^{m(i_n-1)}$, $\mathscr{U}_n^{m(0)} = 0$.
3. Hierarchical surplus: $\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^N \Delta_n^{m(i_n)}$.
4. Sparse grid approximation: on an index set $\mathcal{I} \subset \mathbb{N}^N$

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} \Delta^{m(\mathbf{i})}[u]$$

Assumption: The set $\mathcal{I}$ is downward closed:
$\mathbf{i} \in \mathcal{I} \quad \Rightarrow \quad \mathbf{i} - \mathbf{e}_n \in \mathcal{I}$, $n = 1, \ldots, N$.

Collocation on a (generalized) Sparse Grid

Let $\mathbf{i} = [i_1, \dots, i_N] \in \mathbb{N}_+^N$ and $m(i) : \mathbb{N}_+ \to \mathbb{N}_+$ an increasing function

1. 1D polynomial interpolant operators: $\mathscr{U}_n^{m(i_n)}$ on $m(i_n)$ abscissas. We use either
   - Clenshaw-Curtis (extrema on Chebyshev polynomials)
   - Gauss points w.r.t. the weight ρ_n, assuming that the probability density factorizes as $\rho(\mathbf{y}) = \prod_{n=1}^N \rho_n(y_n)$
2. Detail operator: $\Delta_n^{m(i_n)} = \mathscr{U}_n^{m(i_n)} - \mathscr{U}_n^{m(i_n-1)}$, $\mathscr{U}_n^{m(0)} = 0$.
3. Hierarchical surplus: $\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^N \Delta_n^{m(i_n)}$.
4. Sparse grid approximation: on an index set $\mathcal{I} \subset \mathbb{N}^N$

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} \Delta^{m(\mathbf{i})}[u]$$

Assumption: The set $\mathcal{I}$ is downward closed:

$\mathbf{i} \in \mathcal{I} \quad \Rightarrow \quad \mathbf{i} - \mathbf{e}_n \in \mathcal{I}$, $n = 1, \dots, N$.

Collocation on a (generalized) Sparse Grid

Let $\mathbf{i} = [i_1, \dots, i_N] \in \mathbb{N}_+^N$ and $m(i) : \mathbb{N}_+ \to \mathbb{N}_+$ an increasing function

1. 1D polynomial interpolant operators: $\mathscr{U}_n^{m(i_n)}$ on $m(i_n)$ abscissas. We use either
   - Clenshaw-Curtis (extrema on Chebyshev polynomials)
   - Gauss points w.r.t. the weight ρ_n, assuming that the probability density factorizes as $\rho(\mathbf{y}) = \prod_{n=1}^N \rho_n(y_n)$
2. Detail operator: $\Delta_n^{m(i_n)} = \mathscr{U}_n^{m(i_n)} - \mathscr{U}_n^{m(i_n-1)}$, $\mathscr{U}_n^{m(0)} = 0$.
3. Hierarchical surplus: $\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^N \Delta_n^{m(i_n)}$.
4. Sparse grid approximation: on an index set $\mathcal{I} \subset \mathbb{N}^N$

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} \Delta^{m(\mathbf{i})}[u]$$

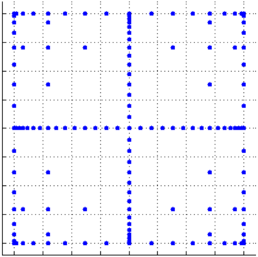
Assumption: The set $\mathcal{I}$ is downward closed:
$\mathbf{i} \in \mathcal{I} \;\Rightarrow\; \mathbf{i} - \mathbf{e}_n \in \mathcal{I},\ n = 1, \dots, N.$

Equivalent formulation

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} c(\mathbf{i}) \mathscr{U}_1^{m(i_1)} \otimes \cdots \otimes \mathscr{U}_N^{m(i_N)} u.$$

with $\quad c(\mathbf{i}) = \sum_{\substack{\mathbf{j} \in \{0,1\}^N \\ (\mathbf{i}+\mathbf{j}) \in \mathcal{I}}} (-1)^{j_1 + \ldots + j_N}, \quad$ and $\quad c(\mathbf{i}) = 0 \quad$ if $\quad \mathbf{i} + \mathbf{1} \in \mathcal{I}$

- linear combination of tensor grids (each with relatively few points!)



Theorem ([Back-Nobile-Tamellini-Tempone, 2010])

Let $\Lambda(\mathcal{I}, m) = \{\mathbf{p} \in \mathbb{N}^N : \ \mathbf{p} \leq m(\mathbf{i}) - 1, \ \mathbf{i} \in \mathcal{I}\}$. Then

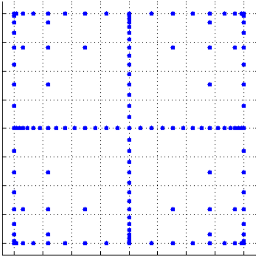
- $\mathcal{S}_{\mathcal{I}} : C^0(\Gamma) \to \mathbb{P}_{\Lambda(\mathcal{I},m)}(\Gamma)$
- $\mathcal{S}_{\mathcal{I}} v = v, \ \ \forall v \in \mathbb{P}_{\Lambda(\mathcal{I},m)}(\Gamma)$

Equivalent formulation

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i}\in\mathcal{I}} c(\mathbf{i}) \mathscr{U}_1^{m(i_1)} \otimes \cdots \otimes \mathscr{U}_N^{m(i_N)} u.$$

with $\quad c(\mathbf{i}) = \sum_{\substack{\mathbf{j}\in\{0,1\}^N \\ (\mathbf{i}+\mathbf{j})\in\mathcal{I}}} (-1)^{j_1+\ldots+j_N}, \quad$ and $\; c(\mathbf{i}) = 0 \;$ if $\; \mathbf{i}+\mathbf{1} \in \mathcal{I}$

- linear combination of tensor grids (each with relatively few points!)



Theorem ([Back-Nobile-Tamellini-Tempone, 2010])

Let $\Lambda(\mathcal{I}, m) = \{\mathbf{p} \in \mathbb{N}^N : \; \mathbf{p} \leq m(\mathbf{i}) - 1, \; \mathbf{i} \in \mathcal{I}\}$. Then

- $\mathcal{S}_{\mathcal{I}} : C^0(\Gamma) \to \mathbb{P}_{\Lambda(\mathcal{I},m)}(\Gamma)$
- $\mathcal{S}_{\mathcal{I}} v = v, \;\; \forall v \in \mathbb{P}_{\Lambda(\mathcal{I},m)}(\Gamma)$

Quasi-optimal sparse grid construction

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i}\in\mathcal{I}} \Delta^{m(\mathbf{i})}[u] \implies \|u - \mathcal{S}_{\mathcal{I}} u\| = \|\sum_{\mathbf{i}\notin\mathcal{I}} \Delta^{m(\mathbf{i})}[u]\| \leq \sum_{\mathbf{i}\notin\mathcal{I}} \|\Delta^{m(\mathbf{i})}[u]\|$$

One can use a knapsack problem-approach [Griebel-Knapek '09, Gerstner-Griebel '03, Bungartz-Griebel '04] to select the best $\mathcal{I}$: for each multiindex $\mathbf{i}$:

- Estimated error contribution (how much error decreases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta E(\mathbf{i}) \geq \|\Delta^{m(\mathbf{i})}[u]\|_V$$

- Estimated work contribution (how much the work, i.e. number of evaluations, increases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta W(\mathbf{i}) \quad \text{such that} \quad \sum_{\mathbf{i}\in\mathcal{I}} \Delta W(\mathbf{i}) \geq W(\mathcal{I})$$

where $W(\mathcal{I})$ is the total number of points in the sparse grid

Quasi-optimal sparse grid construction

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} \Delta^{m(\mathbf{i})}[u] \implies \|u - \mathcal{S}_{\mathcal{I}} u\| = \| \sum_{\mathbf{i} \notin \mathcal{I}} \Delta^{m(\mathbf{i})}[u] \| \leq \sum_{\mathbf{i} \notin \mathcal{I}} \|\Delta^{m(\mathbf{i})}[u]\|$$

One can use a knapsack problem-approach [Griebel-Knapek '09, Gerstner-Griebel '03, Bungartz-Griebel '04] to select the best $\mathcal{I}$: for each multiindex $\mathbf{i}$:

- Estimated error contribution (how much error decreases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta E(\mathbf{i}) \geq \|\Delta^{m(\mathbf{i})}[u]\|_V$$

- Estimated work contribution (how much the work, i.e. number of evaluations, increases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta W(\mathbf{i}) \quad \text{such that} \quad \sum_{\mathbf{i} \in \mathcal{I}} \Delta W(\mathbf{i}) \geq W(\mathcal{I})$$

where $W(\mathcal{I})$ is the total number of points in the sparse grid

Quasi-optimal sparse grid construction

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i} \in \mathcal{I}} \Delta^{m(\mathbf{i})}[u] \implies \|u - \mathcal{S}_{\mathcal{I}} u\| = \|\sum_{\mathbf{i} \notin \mathcal{I}} \Delta^{m(\mathbf{i})}[u]\| \leq \sum_{\mathbf{i} \notin \mathcal{I}} \|\Delta^{m(\mathbf{i})}[u]\|$$

One can use a knapsack problem-approach [Griebel-Knapek '09, Gerstner-Griebel '03, Bungartz-Griebel '04] to select the best $\mathcal{I}$: for each multiindex $\mathbf{i}$:

- Estimated error contribution (how much error decreases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta E(\mathbf{i}) \geq \|\Delta^{m(\mathbf{i})}[u]\|_V$$

- Estimated work contribution (how much the work, i.e. number of evaluations, increases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta W(\mathbf{i}) \quad \text{such that} \quad \sum_{\mathbf{i} \in \mathcal{I}} \Delta W(\mathbf{i}) \geq W(\mathcal{I})$$

where $W(\mathcal{I})$ is the total number of points in the sparse grid

Quasi-optimal sparse grid construction

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i}\in\mathcal{I}} \Delta^{m(\mathbf{i})}[u] \implies \|u - \mathcal{S}_{\mathcal{I}} u\| = \|\sum_{\mathbf{i}\notin\mathcal{I}} \Delta^{m(\mathbf{i})}[u]\| \leq \sum_{\mathbf{i}\notin\mathcal{I}} \|\Delta^{m(\mathbf{i})}[u]\|$$

One can use a knapsack problem-approach [Griebel-Knapek '09, Gerstner-Griebel '03, Bungartz-Griebel '04] to select the best $\mathcal{I}$: for each multiindex $\mathbf{i}$:

- Estimated error contribution (how much error decreases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta E(\mathbf{i}) \geq \|\Delta^{m(\mathbf{i})}[u]\|_V$$

- Estimated work contribution (how much the work, i.e. number of evaluations, increases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta W(\mathbf{i}) \quad \text{such that} \quad \sum_{\mathbf{i}\in\mathcal{I}} \Delta W(\mathbf{i}) \geq W(\mathcal{I})$$

where $W(\mathcal{I})$ is the total number of points in the sparse grid

Quasi-optimal sparse grid construction

$$\mathcal{S}_{\mathcal{I}} u = \sum_{\mathbf{i}\in\mathcal{I}} \Delta^{m(\mathbf{i})}[u] \implies \|u - \mathcal{S}_{\mathcal{I}} u\| = \|\sum_{\mathbf{i}\notin\mathcal{I}} \Delta^{m(\mathbf{i})}[u]\| \leq \sum_{\mathbf{i}\notin\mathcal{I}} \|\Delta^{m(\mathbf{i})}[u]\|$$

One can use a knapsack problem-approach [Griebel-Knapek '09, Gerstner-Griebel '03, Bungartz-Griebel '04] to select the best $\mathcal{I}$: for each multiindex $\mathbf{i}$:

- Estimated error contribution (how much error decreases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta E(\mathbf{i}) \geq \|\Delta^{m(\mathbf{i})}[u]\|_V$$

- Estimated work contribution (how much the work, i.e. number of evaluations, increases if $\mathbf{i}$ is added to $\mathcal{I}$)

$$\Delta W(\mathbf{i}) \quad \text{such that} \quad \sum_{\mathbf{i}\in\mathcal{I}} \Delta W(\mathbf{i}) \geq W(\mathcal{I})$$

where $W(\mathcal{I})$ is the total number of points in the sparse grid

Quasi-optimal sparse grid construction

Then estimate the profit of each $\mathbf{i}$ as

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}$$

and build the sparse grid using the set $\mathcal{I}_M$ of the M indices with the largest estimated profit.

$$\mathcal{I}_M := \{\mathbf{i} \in \mathbb{N}^N \quad P(\mathbf{i}) \geq P_M^{ord}\}$$

where $\{P_j^{ord}\}_j$ is the ordered sequence of profits.

If the set $\mathcal{I}_M$ is not downward closed (lower), take the smallest lower set $\tilde{\mathcal{I}}_M \supset \mathcal{I}_M$. This is equivalent to consider the modified profits $\tilde{P}(\mathbf{i}) = \max_{\mathbf{j} > \mathbf{i}} P(\mathbf{j})$. (see [Chkifa-Cohen-DeVore-Schwab M2AN '13])

Quasi-optimal sparse grid construction

Then estimate the profit of each $\mathbf{i}$ as

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}$$

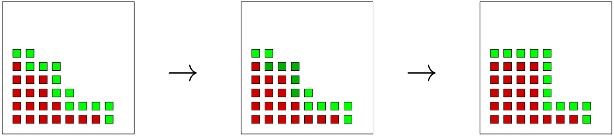
and build the sparse grid using the set $\mathcal{I}_M$ of the M indices with the largest estimated profit.

$$\mathcal{I}_M := \{\mathbf{i} \in \mathbb{N}^N \quad P(\mathbf{i}) \geq P_M^{ord}\}$$

where $\{P_j^{ord}\}_j$ is the ordered sequence of profits.

If the set $\mathcal{I}_M$ is not downward closed (lower), take the smallest lower set $\tilde{\mathcal{I}}_M \supset \mathcal{I}_M$. This is equivalent to consider the modified profits $\tilde{P}(\mathbf{i}) = \max_{\mathbf{j} > \mathbf{i}} P(\mathbf{j})$. (see [Chkifa-Cohen-DeVore-Schwab M2AN '13])

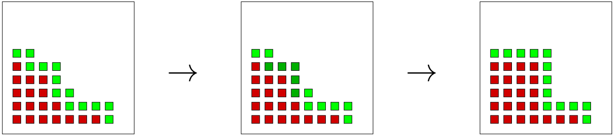
- A posteriori approach [Gerstner-Griebel '03, Klimke, PhD '06]. Given a set Λ, explore all the neighbor multi-indices (margin) and pick up those corresponding to the largest profits.



- A priori approach [Back-Nobile-Tamellini-Tempone '11]. Whenever possible, use *a-priori/a-posteriori* information to build the optimal set. Avoids the "exploration" cost that can be expensive in high dimension.

  The estimate of $\Delta E(\mathbf{i})$ can be related to the decay of the coefficients of the gPC expansion $u = \sum_{\mathbf{p}} u_{\mathbf{p}} \psi_{\mathbf{p}}$ of the solution onto an orthonormal polynomial basis (Legendre, Chebyshev, Hermite, ...) and to the Lebesgue constant of the interpolation scheme.

- A posteriori approach [Gerstner-Griebel '03, Klimke, PhD '06]. Given a set Λ, explore all the neighbor multi-indices (margin) and pick up those corresponding to the largest profits.



- A priori approach [Back-Nobile-Tamellini-Tempone '11]. Whenever possible, use *a-priori/a-posteriori* information to build the optimal set. Avoids the "exploration" cost that can be expensive in high dimension.

  The estimate of $\Delta E(\mathbf{i})$ can be related to the decay of the coefficients of the gPC expansion $u = \sum_{\mathbf{p}} u_{\mathbf{p}} \psi_{\mathbf{p}}$ of the solution onto an orthonormal polynomial basis (Legendre, Chebyshev, Hermite, ...) and to the Lebesgue constant of the interpolation scheme.

General convergence result

Theorem [Tamellini PhD thesis '12], [Nobile-Tamellini-Tempone, in preparation]

Let $\mathcal{S}_{\tilde{\mathcal{I}}_M} u$ be the quasi-optimal sparse grid approximation and $W_{\tilde{\mathcal{I}}_M}$ the total number of points in the sparse grid.

If $C(\tau) := \left(\sum_{\mathbf{i}\in\mathbb{N}^N} \tilde{P}(\mathbf{i})^\tau \Delta W(\mathbf{i})\right)^{\frac{1}{\tau}} < \infty$ for some $\tau < 1$

Then $\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}$

Proof

The proof uses

- Stechkin lemma: given a non-negative descreasing sequence $\{a_k\}_k$, then

$$\sum_{k=N+1}^{\infty} a_k \leq N^{1-\frac{1}{\tau}} \left(\sum_{k=1}^{\infty} a_k^{\tau} \right)^{\frac{1}{\tau}}, \quad 0 < \tau < 1.$$

- ordered repeated sequence of profits $\{\hat{P}_k\}_k = \{\underbrace{\tilde{P}_1, \ldots, \tilde{P}_1}_{\Delta W_1 \text{ times}}, \underbrace{\tilde{P}_2, \ldots, \tilde{P}_2}_{\Delta W_2 \text{ times}}, \ldots\}$.
- Let $W_M = \sum_{j=1}^{M} \Delta W_j$ and observe that $W_M \geq W_{\tilde{\mathcal{I}}_M}$. Then

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq \sum_{k=M+1}^{\infty} \Delta E_k \leq \sum_{k=W_M+1}^{\infty} \hat{P}_k \quad \text{[apply Stechkin]}$$

$$\leq \|\{\hat{P}_k\}_k\|_{l^\tau} W_M^{1-\frac{1}{\tau}} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}$$

Proof

The proof uses

- Stechkin lemma: given a non-negative descreasing sequence $\{a_k\}_k$, then

$$\sum_{k=N+1}^{\infty} a_k \leq N^{1-\frac{1}{\tau}} \left(\sum_{k=1}^{\infty} a_k^{\tau} \right)^{\frac{1}{\tau}}, \quad 0 < \tau < 1.$$

- ordered repeated sequence of profits $\{\hat{P}_k\}_k = \{\underbrace{\tilde{P}_1, \dots, \tilde{P}_1}_{\Delta W_1 \text{ times}}, \underbrace{\tilde{P}_2, \dots, \tilde{P}_2}_{\Delta W_2 \text{ times}}, \dots\}$.

- Let $W_M = \sum_{j=1}^{M} \Delta W_j$ and observe that $W_M \geq W_{\tilde{\mathcal{I}}_M}$. Then

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq \sum_{k=M+1}^{\infty} \Delta E_k \leq \sum_{k=W_M+1}^{\infty} \hat{P}_k \quad \text{[apply Stechkin]}$$

$$\leq \|\{\hat{P}_k\}_k\|_{l^\tau} W_M^{1-\frac{1}{\tau}} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}$$

Proof

The proof uses

- Stechkin lemma: given a non-negative descreasing sequence $\{a_k\}_k$, then

$$\sum_{k=N+1}^{\infty} a_k \leq N^{1-\frac{1}{\tau}} \left(\sum_{k=1}^{\infty} a_k^{\tau}\right)^{\frac{1}{\tau}}, \quad 0 < \tau < 1.$$

- ordered repeated sequence of profits $\{\hat{P}_k\}_k = \{\underbrace{\tilde{P}_1, \ldots, \tilde{P}_1}_{\Delta W_1 \text{ times}}, \underbrace{\tilde{P}_2, \ldots, \tilde{P}_2}_{\Delta W_2 \text{ times}}, \ldots\}$.

- Let $W_M = \sum_{j=1}^{M} \Delta W_j$ and observe that $W_M \geq W_{\tilde{\mathcal{I}}_M}$. Then

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma, V)} \leq \sum_{k=M+1}^{\infty} \Delta E_k \leq \sum_{k=W_M+1}^{\infty} \hat{P}_k \quad \text{[apply Stechkin]}$$

$$\leq \|\{\hat{P}_k\}_k\|_{l^\tau} W_M^{1-\frac{1}{\tau}} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}$$

Proof

The proof uses

- Stechkin lemma: given a non-negative descreasing sequence $\{a_k\}_k$, then

$$\sum_{k=N+1}^{\infty} a_k \leq N^{1-\frac{1}{\tau}} \left(\sum_{k=1}^{\infty} a_k^{\tau} \right)^{\frac{1}{\tau}}, \quad 0 < \tau < 1.$$

- ordered repeated sequence of profits $\{\hat{P}_k\}_k = \{\underbrace{\tilde{P}_1, \ldots, \tilde{P}_1}_{\Delta W_1 \text{ times}}, \underbrace{\tilde{P}_2, \ldots, \tilde{P}_2}_{\Delta W_2 \text{ times}}, \ldots\}$.

- Let $W_M = \sum_{j=1}^{M} \Delta W_j$ and observe that $W_M \geq W_{\tilde{\mathcal{I}}_M}$. Then

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq \sum_{k=M+1}^{\infty} \Delta E_k \leq \sum_{k=W_M+1}^{\infty} \hat{P}_k \quad \text{[apply Stechkin]}$$

$$\leq \|\{\hat{P}_k\}_k\|_{l^\tau} W_M^{1-\frac{1}{\tau}} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}$$

How to estimate ΔW and ΔE

1. $\Delta W(\mathbf{i})$: number of new points in $\bigotimes_{n=1}^N \Delta_n^{m(i_n)}$
   Count all points in $\mathscr{U}_1^{m(i_1)} \otimes \cdots \otimes \mathscr{U}_N^{m(i_N)}$ (non-nested case) or just the extra points added (nested case)
2. $\Delta E(\mathbf{i})$: use expansion on a suitable basis $u = \sum_{\mathbf{p}} u_{\mathbf{p}} \psi_{\mathbf{p}}$ (Legendre, Chebyshev, ...) and relate $\Delta E(\mathbf{i})$ with the decay of the Fourier coefficients $u_{\mathbf{p}}$.
   E.g. for Chebyshev expansion ($\|\psi_{\mathbf{p}}\|_\infty = 1$)

$$\Delta E(\mathbf{i}) \leq 2\, \mathbb{L}_{m(\mathbf{i})} \sum_{\mathbf{p} \geq m(\mathbf{i}-1)} \|u_{\mathbf{p}}\|_V,$$

where $\mathbb{L}_{m(\mathbf{i})} = \prod_{n=1}^N \mathbb{L}_{m(i_n)}$ and $\mathbb{L}_{m(i_n)} := \|\mathscr{U}_n^{m(i_n)}\|_{\mathcal{L}(C^0, L^2_\rho)}$ is the Lebesgue constant form C^0 to L^2_ρ.

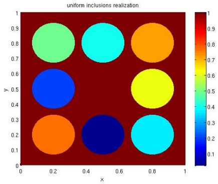
How to estimate ΔW and ΔE

1. $\Delta W(\mathbf{i})$: number of new points in $\bigotimes_{n=1}^{N} \Delta_n^{m(i_n)}$
   Count all points in $\mathscr{U}_1^{m(i_1)} \otimes \cdots \otimes \mathscr{U}_N^{m(i_N)}$ (non-nested case) or just the extra points added (nested case)
2. $\Delta E(\mathbf{i})$: use expansion on a suitable basis $u = \sum_{\mathbf{p}} u_{\mathbf{p}} \psi_{\mathbf{p}}$ (Legendre, Chebyshev, ...) and relate $\Delta E(\mathbf{i})$ with the decay of the Fourier coefficients $u_{\mathbf{p}}$.
   E.g. for Chebyshev expansion ($\|\psi_{\mathbf{p}}\|_\infty = 1$)

$$\Delta E(\mathbf{i}) \leq 2\, \mathbb{L}_{m(\mathbf{i})} \sum_{\mathbf{p} \geq m(\mathbf{i}-1)} \|u_{\mathbf{p}}\|_V,$$

where $\mathbb{L}_{m(\mathbf{i})} = \prod_{n=1}^{N} \mathbb{L}_{m(i_n)}$ and $\mathbb{L}_{m(i_n)} := \|\mathscr{U}_n^{m(i_n)}\|_{\mathcal{L}(C^0, L^2_\rho)}$ is the Lebesgue constant form C^0 to L^2_ρ.

Elliptic equation with random inclusions

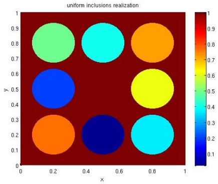


- $Y_i \sim \mathcal{U}[a_i, b_i]$, independent
- $\Gamma = \prod_{i=1}^{N} [a_i, b_i]$
- $u(\mathbf{y}) : \Gamma \to H_0^1(D)$

- The solution $u(\mathbf{y}) : \Gamma \to H_0^1(D)$ is analytic in a polydisk in the complex plane $\mathbb{C}^N$
- The solution can be expanded in Chebyshev series $u(\mathbf{y}) = \sum_{\mathbf{p}} u_{\mathbf{p}} \psi_{\mathbf{p}}(\mathbf{y})$ with $\|\psi_{\mathbf{p}}\|_\infty \leq 1$. Estimates on Chebyshev coefficients are available [Babuska-Nobile-Tempone '07] [Nobile-Tamellini-Tempone '13]

$$\|u_{\mathbf{p}}\|_{H_0^1(D)} \leq C e^{-\sum_{n=1}^{N} g_i p_i}$$

Elliptic equation with random inclusions



- $Y_i \sim \mathcal{U}[a_i, b_i]$, independent
- $\Gamma = \prod_{i=1}^{N}[a_i, b_i]$
- $u(\mathbf{y}) : \Gamma \to H_0^1(D)$

- The solution $u(\mathbf{y}) : \Gamma \to H_0^1(D)$ is analytic in a polydisk in the complex plane $\mathbb{C}^N$
- The solution can be expanded in Chebyshev series $u(\mathbf{y}) = \sum_{\mathbf{p}} u_{\mathbf{p}}\psi_{\mathbf{p}}(\mathbf{y})$ with $\|\psi_{\mathbf{p}}\|_\infty \leq 1$. Estimates on Chebyshev coefficients are available [Babuska-Nobile-Tempone '07] [Nobile-Tamellini-Tempone '13]

$$\|u_{\mathbf{p}}\|_{H_0^1(D)} \leq Ce^{-\sum_{n=1}^{N} g_i p_i}$$

Quasi-optimal sparse grid construction using (nested) Clenshaw-Curtis points

- $m(i) = 2^{i-1} + 1$ (doubling the points). $\Delta W(\mathbf{i}) \approx \prod_{n=1}^{N} 2^{i_n - 2}$
- Decay of Chebyshev coefficients: $\|u_{\mathbf{p}}\|_V \leq C e^{-g \sum_{n=1}^{N} p_n}$
- Lebesgue constant: $\mathbb{L}(i) \leq \frac{2}{\pi} \log(i+1) + 1$
- Error estimate: $\Delta E(\mathbf{i}) = C \prod_{n=1}^{N} e^{-\hat{g} m(i_n - 1)}$
- Profit estimate: $P(\mathbf{i}) = \tilde{C} e^{-\hat{g} \sum_{n=1}^{N} m(i_n - 1)}$

Theorem [Nobile-Tamellini-Tempone '13]

$$\left(\sum_{\mathbf{i} \in \mathbb{N}^N} P(\mathbf{i})^{\tau} \Delta W(\mathbf{i})\right)^{\frac{1}{\tau}} < \infty \quad \text{for all } 0 < \tau < 1$$

$$\Longrightarrow \quad \|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma, V)} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}, \quad \forall\, 0 < \tau < 1$$

By optimizing with respect to τ one can get the estimate

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma, V)} \leq C_1(N) \exp\{-N C_2 W_{\tilde{\mathcal{I}}_M}^{\frac{1}{N}}\}$$

Quasi-optimal sparse grid construction using (nested) Clenshaw-Curtis points

- $m(i) = 2^{i-1} + 1$ (doubling the points). $\Delta W(\mathbf{i}) \approx \prod_{n=1}^{N} 2^{i_n - 2}$
- Decay of Chebyshev coefficients: $\|u_{\mathbf{p}}\|_V \leq C e^{-g \sum_{n=1}^{N} p_n}$
- Lebesgue constant: $\mathbb{L}(i) \leq \frac{2}{\pi} \log(i+1) + 1$
- Error estimate: $\Delta E(\mathbf{i}) = C \prod_{n=1}^{N} e^{-\hat{g} m(i_n - 1)}$
- Profit estimate: $P(\mathbf{i}) = \tilde{C} e^{-\hat{g} \sum_{n=1}^{N} m(i_n - 1)}$

Theorem [Nobile-Tamellini-Tempone '13]

$$\left(\sum_{\mathbf{i}\in\mathbb{N}^N} P(\mathbf{i})^\tau \Delta W(\mathbf{i})\right)^{\frac{1}{\tau}} < \infty \quad \text{for all } 0 < \tau < 1$$

$$\Longrightarrow \quad \|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma, V)} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}, \quad \forall\, 0 < \tau < 1$$

By optimizing with respect to τ one can get the estimate

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma, V)} \leq C_1(N) \exp\{-N C_2 W_{\tilde{\mathcal{I}}_M}^{\frac{1}{N}}\}$$

Quasi-optimal sparse grid construction using (nested) Clenshaw-Curtis points

- $m(i) = 2^{i-1} + 1$ (doubling the points). $\Delta W(\mathbf{i}) \approx \prod_{n=1}^{N} 2^{i_n - 2}$
- Decay of Chebyshev coefficients: $\|u_{\mathbf{p}}\|_V \leq Ce^{-g\sum_{n=1}^{N} p_n}$
- Lebesgue constant: $\mathbb{L}(i) \leq \frac{2}{\pi}\log(i+1) + 1$
- Error estimate: $\Delta E(\mathbf{i}) = C\prod_{n=1}^{N} e^{-\hat{g}m(i_n - 1)}$
- Profit estimate: $P(\mathbf{i}) = \tilde{C}e^{-\hat{g}\sum_{n=1}^{N} m(i_n-1)}$

Theorem [Nobile-Tamellini-Tempone '13]

$$\left(\sum_{\mathbf{i}\in\mathbb{N}^N} P(\mathbf{i})^\tau \Delta W(\mathbf{i})\right)^{\frac{1}{\tau}} < \infty \quad \text{for all } 0 < \tau < 1$$

$$\implies \quad \|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}, \quad \forall\, 0 < \tau < 1$$

By optimizing with respect to τ one can get the estimate

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq C_1(N)\exp\{-NC_2 W_{\tilde{\mathcal{I}}_M}^{\frac{1}{N}}\}$$

Quasi-optimal sparse grid construction using (non-nested) Legendre points

- $m(i) = i$ (adding 1 point at the time). $\Delta W(\mathbf{i}) = \prod_{n=1}^{N} i_n$
- Decay of Chebyshev coefficients: $\|u_{\mathbf{p}}\|_V \leq Ce^{-g\sum_{n=1}^{N} p_n}$
- Lebesgue constant: $\mathbb{L}(i) = 1$.
- Error estimate: $\Delta E(\mathbf{i}) = C\prod_{n=1}^{N} e^{-\hat{g}m(i_n-1)}$
- Profit estimate: $P(\mathbf{i}) = \tilde{C}e^{-\hat{g}\sum_{n=1}^{N} m(i_n-1)}$

Theorem [Nobile-Tamellini-Tempone '13]

$$\left(\sum_{\mathbf{i}\in\mathbb{N}^N} P(\mathbf{i})^\tau \Delta W(\mathbf{i})\right)^{\frac{1}{\tau}} < \infty \quad \text{for all } 0 < \tau < 1$$

$$\implies \quad \|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}, \quad \forall\, 0 < \tau < 1$$

By optimizing with respect to τ one can get the estimate

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq \tilde{C}_1(N)\exp\{-N\tilde{C}_2 W_{\tilde{\mathcal{I}}_M}^{\frac{1}{2N}}\}$$

Quasi-optimal sparse grid construction using (non-nested) Legendre points

- $m(i) = i$ (adding 1 point at the time). $\Delta W(\mathbf{i}) = \prod_{n=1}^{N} i_n$
- Decay of Chebyshev coefficients: $\|u_{\mathbf{p}}\|_V \leq Ce^{-g\sum_{n=1}^{N} p_n}$
- Lebesgue constant: $\mathbb{L}(i) = 1$.
- Error estimate: $\Delta E(\mathbf{i}) = C\prod_{n=1}^{N} e^{-\hat{g}m(i_n-1)}$
- Profit estimate: $P(\mathbf{i}) = \tilde{C}e^{-\hat{g}\sum_{n=1}^{N} m(i_n-1)}$

Theorem [Nobile-Tamellini-Tempone '13]

$$\left(\sum_{\mathbf{i}\in\mathbb{N}^N} P(\mathbf{i})^\tau \Delta W(\mathbf{i})\right)^{\frac{1}{\tau}} < \infty \quad \text{for all } 0<\tau<1$$

$$\implies \quad \|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq C(\tau) W_{\tilde{\mathcal{I}}_M}^{1-\frac{1}{\tau}}, \quad \forall\, 0<\tau<1$$

By optimizing with respect to τ one can get the estimate

$$\|u - \mathcal{S}_{\tilde{\mathcal{I}}_M} u\|_{L^2_\rho(\Gamma,V)} \leq \tilde{C}_1(N)\exp\{-N\tilde{C}_2 W_{\tilde{\mathcal{I}}_M}^{\frac{1}{2N}}\}$$

Quasi optimal sparse grid in practice

We have the theoretical bound $\|u_{\mathbf{p}}\|_V \approx Ce^{-\sum_{n=1}^{N} g_n p_n}$, hence, in particular, for $\mathbf{p} = \mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$, $\|u_{\mathbf{e}_j}\|_V \approx Ce^{-g_j p_j}$ which implies that a 1D polynomial interpolation in the variable y_j only converges exponentially with rate g_j.

- The rates g_j can be estimated numerically by "1D analyses". (increase the polynomial degree in one variable at the time and fit the convergence rate).
- Once the rates g_j are available, we build the quasi optimal index set based on the profit estimate

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})} \propto \prod_{n=1}^{N} \frac{\mathbb{L}_{m(i_n)} e^{-g_{i_n} m(i_n - 1)}}{\Delta W(i_n)}$$

Quasi optimal sparse grid in practice

We have the theoretical bound $\|u_{\mathbf{p}}\|_V \approx Ce^{-\sum_{n=1}^{N} g_n p_n}$, hence, in particular, for $\mathbf{p} = \mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$, $\|u_{\mathbf{e}_j}\|_V \approx Ce^{-g_j p_j}$ which implies that a 1D polynomial interpolation in the variable y_j only converges exponentially with rate g_j.

- The rates g_j can be estimated numerically by "1D analyses". (increase the polynomial degree in one variable at the time and fit the convergence rate).
- Once the rates g_j are available, we build the quasi optimal index set based on the profit estimate

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})} \propto \prod_{n=1}^{N} \frac{\mathbb{L}_{m(i_n)} e^{-g_{i_n} m(i_n - 1)}}{\Delta W(i_n)}$$

Quasi optimal sparse grid in practice

We have the theoretical bound $\|u_{\mathbf{p}}\|_V \approx Ce^{-\sum_{n=1}^N g_n p_n}$,
hence, in particular, for $\mathbf{p} = \mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$,
$\|u_{\mathbf{e}_j}\|_V \approx Ce^{-g_j p_j}$ which implies that a 1D polynomial interpolation in the variable y_j only converges exponentially with rate g_j.

- The rates g_j can be estimated numerically by "1D analyses". (increase the polynomial degree in one variable at the time and fit the convergence rate).
- Once the rates g_j are available, we build the quasi optimal index set based on the profit estimate

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})} \propto \prod_{n=1}^N \frac{\mathbb{L}_{m(i_n)} e^{-g_{i_n} m(i_n - 1)}}{\Delta W(i_n)}$$

Quasi optimal sparse grid in practice

When working with random fields instread of inclusion problmes a good estimate for the Legendre/Chebyshev coefficients is (see [Bech-N.-Tamellini-Tempone M3AS '12, Cohen-DeVore-Schwab FoCM '10])

$$\|u_{\mathbf{p}}\|_V \approx C\frac{|\mathbf{p}|!}{\mathbf{p}!}\exp\{-\sum_{n=1}^{N} g_n p_n\}.$$

The rates g_n can be estimated again by 1D inexpensive analyses.

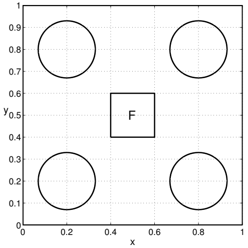
Quasi optimal sparse grid in practice

When working with random fields instread of inclusion problmes a good estimate for the Legendre/Chebyshev coefficients is (see [Bech-N.-Tamellini-Tempone M3AS '12, Cohen-DeVore-Schwab FoCM '10])

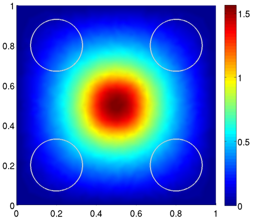
$$\|u_{\mathbf{p}}\|_V \approx C \frac{|\mathbf{p}|!}{\mathbf{p}!} \exp\{-\sum_{n=1}^{N} g_n p_n\}.$$

The rates g_n can be estimated again by 1D inexpensive analyses.

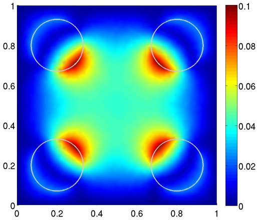
Isotropic test case – 4 random inclusions



- Conductivity coefficient: matrix k=1
  circular inclusions: $k|_{\Omega_i} \sim \mathcal{U}(0.01, 1.99)$
  $\rightarrow$ 4 iid uniform random variables
- forcing term $f = 100\mathbb{1}_F$
- zero boundary conditions
- quantity of interest $\psi(u) = \int_F u$



mean



std

Isotropic test case – 4 random inclusions

convergence plot for $\|\psi(u) - \mathcal{S}_{\mathcal{I}}\psi(u)\|_{L^2_\rho(\Gamma)}$ versus # pts sparse grid

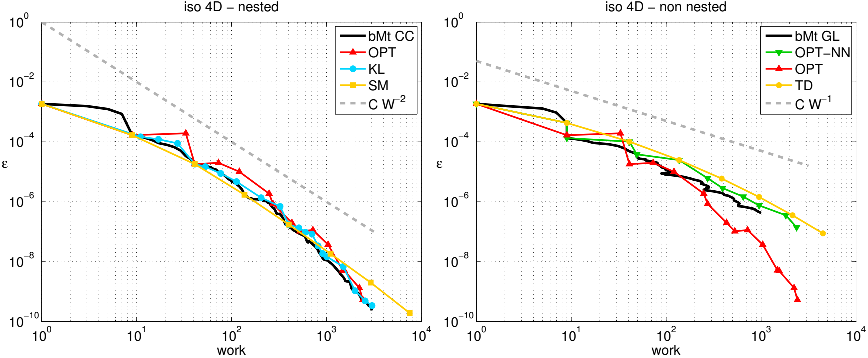


Figure: Results for the isotropic problem. **Left**: (nested) CC points **Right**: (non-nested) Gauss-Legendre points

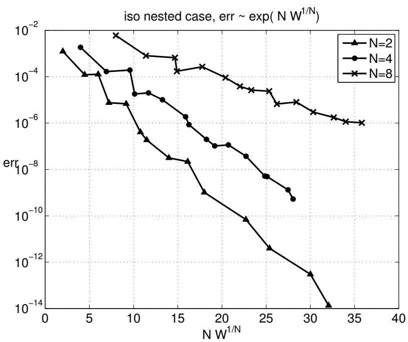
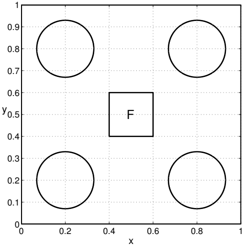
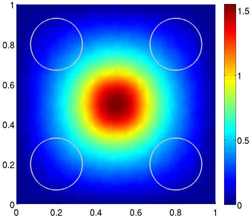


Figure: Results for the isotropic problem. Optimal sparse grids and their predicted convergence rates. **Top**: Nested CC points. **Bottom**: Non-nested Gauss Legendre points.

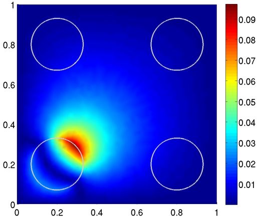
Anisotropic test case – 4 random inclusions



- Conductivity coefficient: matrix k=1 circular inclusions: $k|_{\Omega_i} \sim \gamma_i \mathcal{U}(-0.99, 0.99) \to$ 2 iid uniform random variables
- $\gamma_{1,2,3,4} = 1,\ 0.06,\ 0.0035,\ 0.0002$



mean



std

Anisotropic test case – 4 random inclusions

convergence plot for $\|\psi(u) - \mathcal{S}_{\mathcal{I}}\psi(u)\|_{L^2_\rho(\Gamma)}$ versus # pts sparse grid

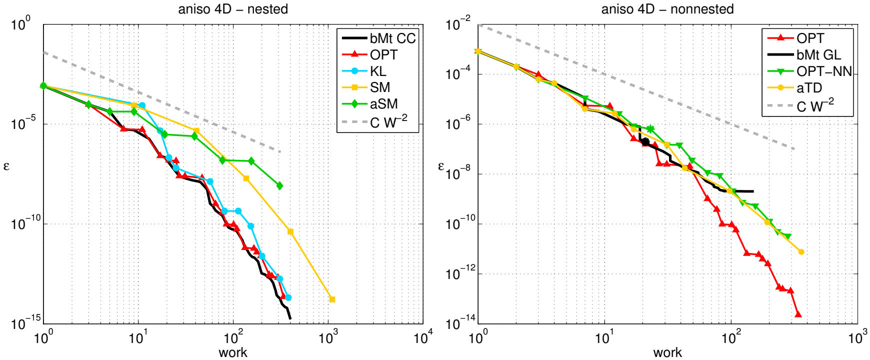


Figure: Results for the anisotropic problem. **Left**: (nested) CC points
Right: (non-nested) Gauss-Legendre points

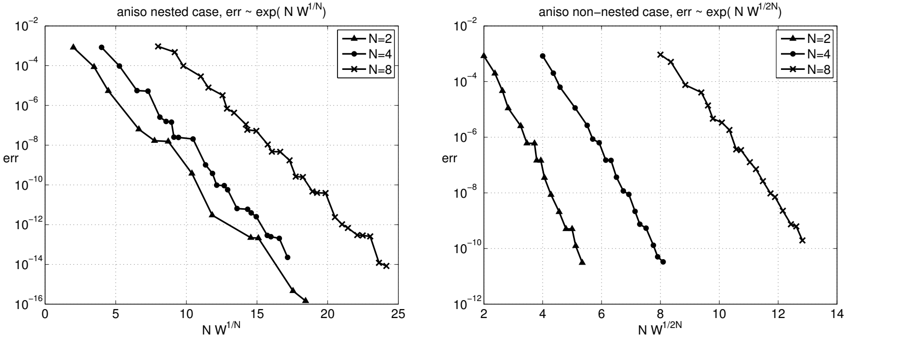
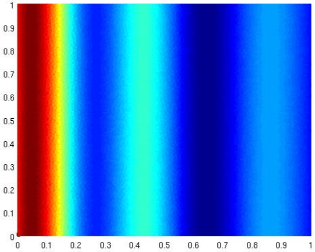


Figure: Results for the anisotropic problem. Optimal sparse grids and their predicted convergence rates. **Top**: Nested CC points. **Bottom**: Non-nested Gauss Legendre points.

Numerical test - 1D stationary lognormal field



$L = 1,\ D = [0, L]^2.$

$$\begin{cases} -\nabla \cdot a(\mathbf{y}, \mathbf{x}) \boldsymbol{\nabla} u(\mathbf{y}, \mathbf{x}) = 0 \\ u = 1 \text{ on } x = 0,\ h = 0 \text{ on } x = 1 \\ \text{no flux otherwise} \end{cases}$$

$$a(\mathbf{x}, \mathbf{y}) = e^{\gamma(\mathbf{x}, \mathbf{y})},\ \mu_\gamma(\mathbf{x}) = 0,$$

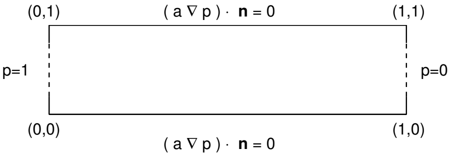
$$Cov_\gamma(\mathbf{x}, \mathbf{x}') = \sigma^2 e^{-\frac{|\mathbf{x}_1 - \mathbf{x}_1'|^2}{LC^2}}$$

We approximate γ as

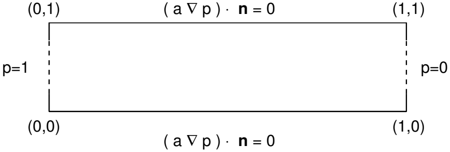
$$\gamma(\mathbf{y}, \mathbf{x}) \approx \mu(\mathbf{x}) + \sigma a_0 y_0 + \sigma \sum_{k=1}^{K} a_k \left[y_{2k-1} \cos\left(\frac{\pi}{L} k \mathbf{x}_1\right) + y_{2k} \sin\left(\frac{\pi}{L} k \mathbf{x}_1\right) \right]$$

with $y_i \sim \mathcal{N}(0, 1)$, i.i.d.

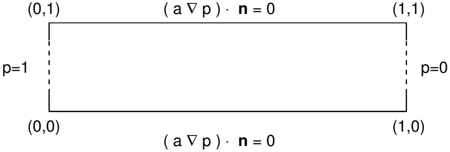
Given the Fourier series $\sigma^2 e^{-\frac{|z|^2}{LC^2}} = \sum_{k=0}^{\infty} c_k \cos\left(\frac{\pi}{L} kz\right)$, $a_k = \sqrt{c_k}$.



- Quantity of interest: effective permeability $\mathbb{E}[\Phi(u)]$, with $\Phi = \left[\int_0^L k(\cdot,x)\frac{\partial u(\cdot,x)}{\partial x}dx\right]$
- Convergence: $|\mathbb{E}[\Phi(\mathcal{S}_{\tilde{\mathcal{I}}_M} u)] - \mathbb{E}[\Phi(u)]|$
- We compare Monte Carlo estimate with quasi-optimal sparse grids based on Gauss-Hermite-Patterson points (nested Gauss-Hermite)
- Estimate of Hermite coefficients decay:
  - for the simpler problem $\nabla \cdot a(\mathbf{y})\nabla u(\mathbf{y},\mathbf{x}) = f$, $a(\mathbf{y}) = e^{b_0 + \sum_{n=1}^N y_n b_n}$, we have $\|u_{\mathbf{i}}\|_V = C\frac{b_n^{i_n}}{\sqrt{i_n!}}$.
  - Heuristic: use the same ansatz $\|u_{\mathbf{i}}\|_V \approx C\prod_{n=1}^N \frac{e^{-g_n i_n}}{\sqrt{i_n!}}$ but estimate the rates g_n numerically.

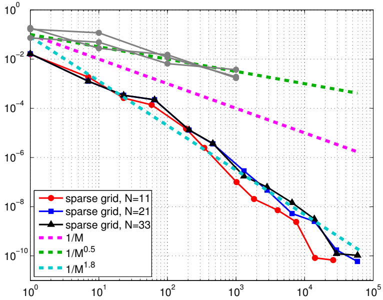


- Quantity of interest: effective permeability $\mathbb{E}[\Phi(u)]$, with $\Phi = \left[\int_0^L k(\cdot, x) \frac{\partial u(\cdot, x)}{\partial x} dx\right]$
- Convergence: $|\mathbb{E}[\Phi(\mathcal{S}_{\tilde{\mathcal{I}}_M} u)] - \mathbb{E}[\Phi(u)]|$
- We compare Monte Carlo estimate with quasi-optimal sparse grids based on Gauss-Hermite-Patterson points (nested Gauss-Hermite)
- Estimate of Hermite coefficients decay:
  - for the simpler problem $\nabla \cdot a(\mathbf{y}) \boldsymbol{\nabla} u(\mathbf{y}, \mathbf{x}) = f$, $a(\mathbf{y}) = e^{b_0 + \sum_{n=1}^N y_n b_n}$, we have $\|u_{\mathbf{i}}\|_V = C \frac{b_n^{i_n}}{\sqrt{i_n!}}$.
  - Heuristic: use the same ansatz $\|u_{\mathbf{i}}\|_V \approx C \prod_{n=1}^N \frac{e^{-g_n i_n}}{\sqrt{i_n!}}$ but estimate the rates g_n numerically.

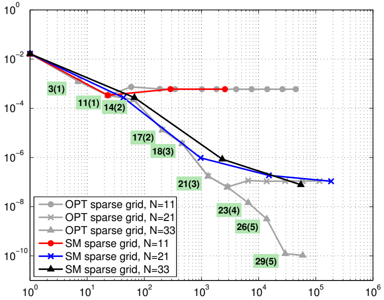


- Quantity of interest: effective permeability $\mathbb{E}[\Phi(u)]$, with $\Phi = \left[\int_0^L k(\cdot,x)\frac{\partial u(\cdot,x)}{\partial x}dx\right]$
- Convergence: $|\mathbb{E}[\Phi(\mathcal{S}_{\tilde{\mathcal{I}}_M} u)] - \mathbb{E}[\Phi(u)]|$
- We compare Monte Carlo estimate with quasi-optimal sparse grids based on Gauss-Hermite-Patterson points (nested Gauss-Hermite)
- Estimate of Hermite coefficients decay:
  - for the simpler problem $\nabla \cdot a(\mathbf{y})\boldsymbol{\nabla} u(\mathbf{y},\mathbf{x}) = f$, $a(\mathbf{y}) = e^{b_0 + \sum_{n=1}^N y_n b_n}$, we have $\|u_{\mathbf{i}}\|_V = C\frac{b_n^{i_n}}{\sqrt{i_n!}}$.
  - Heuristic: use the same ansatz $\|u_{\mathbf{i}}\|_V \approx C\prod_{n=1}^N \frac{e^{-g_n i_n}}{\sqrt{i_n!}}$ but estimate the rates g_n numerically.

Correlation length: $LC = 0.2$, Std: $\sigma = 0.3$ (c.o.v. $\sim 30\%$)



(a) Convergence of quasi-optimal sparse grid approximations.



(b) Convergence with respect to the reference solution with $N = 33$ r.vs.

- The quasi optimal construction automatically adds new variables when needed.
- No need to truncate a-priori the random field

"A quasi-optimal sparse grids procedure for groundwater flows" by J. Beck, F. Nobile, L. Tamellini and R. Tempone. To appear, LNCSE, Springer, 2013.

Partial conclusions

- Sparse grid polynomial approximation works well for elliptic problems with random coefficients for
  - inclusion type problems with few inclusions
  - smooth random fields with long correlation length
- On the other hand, these techniques suffer in the cases of
  - rough fields even with long correlation lenght; e.g. exponential covariance: $Cov_a(x,y) = \sigma^2 e^{|x-y|/l_c}$

$$a(\omega, x) = \bar{a}(x) + \sum_{n=1}^{\infty} \sqrt{\lambda_n} y_n(\omega) b_n(x), \qquad \lambda_n \sim n^{-2}$$

  - fields with short correlation length:

$$\lambda_n = O(1) \quad \text{for} \quad n \leq diam(D)/l_c$$

Idea. In the case of rough longly correlated fields, combine sparse grid polynomial approximation with Multi Level Monte Carlo.

Partial conclusions

- Sparse grid polynomial approximation works well for elliptic problems with random coefficients for
  - inclusion type problems with few inclusions
  - smooth random fields with long correlation length
- On the other hand, these techniques suffer in the cases of
  - rough fields even with long correlation lenght; e.g. exponential covariance: $Cov_a(x,y) = \sigma^2 e^{|x-y|/l_c}$

$$a(\omega, x) = \bar{a}(x) + \sum_{n=1}^{\infty} \sqrt{\lambda_n} y_n(\omega) b_n(x), \qquad \lambda_n \sim n^{-2}$$

  - fields with short correlation length:

$$\lambda_n = O(1) \quad \text{for} \quad n \leq diam(D)/l_c$$

Idea. In the case of rough longly correlated fields, combine sparse grid polynomial approximation with Multi Level Monte Carlo

Partial conclusions

- Sparse grid polynomial approximation works well for elliptic problems with random coefficients for
  - inclusion type problems with few inclusions
  - smooth random fields with long correlation length
- On the other hand, these techniques suffer in the cases of
  - rough fields even with long correlation lenght; e.g. exponential covariance: $Cov_a(x, y) = \sigma^2 e^{|x-y|/l_c}$

$$a(\omega, x) = \bar{a}(x) + \sum_{n=1}^{\infty} \sqrt{\lambda_n} y_n(\omega) b_n(x), \qquad \lambda_n \sim n^{-2}$$

  - fields with short correlation length:

$$\lambda_n = O(1) \quad \text{for} \quad n \leq diam(D)/l_c$$

Idea. In the case of rough longly correlated fields, combine sparse grid polynomial approximation with Multi Level Monte Carlo.

Partial conclusions

- Sparse grid polynomial approximation works well for elliptic problems with random coefficients for
  - inclusion type problems with few inclusions
  - smooth random fields with long correlation length
- On the other hand, these techniques suffer in the cases of
  - rough fields even with long correlation lenght; e.g. exponential covariance: $Cov_a(x,y) = \sigma^2 e^{|x-y|/l_c}$

$$a(\omega, x) = \bar{a}(x) + \sum_{n=1}^{\infty} \sqrt{\lambda_n} y_n(\omega) b_n(x), \qquad \lambda_n \sim n^{-2}$$

  - fields with short correlation length:

$$\lambda_n = O(1) \quad \text{for} \quad n \leq diam(D)/l_c$$

Idea. In the case of rough longly correlated fields, combine sparse grid polynomial approximation with Multi Level Monte Carlo.

Outline

1 Model problem: elliptic PDEs with random coefficients

2 Polynomial approximation by sparse grid collocation
- Quasi optimal sparse grid construction
- Convergence result for elliptic PDEs with random inclusions
- Numerical results

3 Combined MLMC / Sparse grid method
- MLMC with Sparse Grid Control Variate
- Variance analysis and algorithm tuning
- Numerical results

4 Conclusions

Darcy Problem with log-normal permeability

find $\mathbf{u}(x,\omega) : \overline{D} \times \Omega \to \mathbb{R}^d$ and $p(x,\omega) : \overline{D} \times \Omega \to \mathbb{R}$ such that almost everywhere in $\omega \in \Omega$ it holds :

$$\begin{cases} \mathbf{u}(x,\omega) = -a(x,\omega)\nabla p(x,\omega) & in \quad D, \\ \mathrm{div}(\mathbf{u}(x,\omega)) = f(x) & in \quad D, \\ + \text{boundary conditions} & on \quad \partial D. \end{cases}$$

- D is the bounded physical domain and $(\Omega, \mathcal{F}, P)$ is the probability space
- $\mathbf{u}$ and p are the Darcy velocity and pressure, respectively
- $a(x,\omega)$ is the permeability field modeled as a lognormally distributed random field $a(x,\omega) = e^{\gamma(x,\omega)}$
- $\gamma(x,\omega)$ is a Gaussian stationary random field

Matérn Covariance Function

Matérn Family

$$\mathrm{cov}_\gamma(x_1 - x_2) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu} \frac{\|x_1 - x_2\|}{L_c} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{\|x_1 - x_2\|}{L_c} \right)$$

- L_c is a correlation length
- Γ is the gamma function
- K_ν is the modified Bessel function of the second kind

- $\nu = 0.5$: $\mathrm{cov}_\gamma(x_1 - x_2) = \sigma^2 e^{-\frac{\|x_1 - x_2\|}{L_c}}$ (exponential covariance)
- $\nu \to \infty$: $\mathrm{cov}_\gamma(x_1 - x_2) = \sigma^2 e^{-\frac{\|x_1 - x_2\|^2}{L_c^2}}$ (Gaussian covariance)

Matérn Covariance Function: Regularity issues

Let $\nu = s + \alpha$ with $s = \lceil \nu - 1 \rceil \in \mathbb{N}$ and $\alpha = \nu - \lceil \nu - 1 \rceil \in (0, 1]$. Then the realizations of the random field are almost surely Hölder continuous, $\gamma(x, \omega) \in \mathcal{C}^{s,\beta}(\bar{D})$ with $\beta < \alpha$.

- For $\nu = 0.5$ the covariance function is only Lipschitz continuous and the field is almost surely Hölder continuous $\gamma(x, \omega) \in \mathcal{C}^{0,\alpha}(\bar{D})$ with $\alpha < 0.5$.
- For $\nu \to \infty$ the covariance function as well as the field are continuous with all their derivates, namely $\mathrm{cov}_\gamma(x), \gamma(x, \omega_i) \in C^\infty(\bar{D})$ $\forall \omega \in \Omega$

(see [Graham, Kuo, Nichols, Scheichl, Schwab, Sloan '13])

Matérn Covariance Function: Regularity issues

Let $\nu = s + \alpha$ with $s = \lceil \nu - 1 \rceil \in \mathbb{N}$ and $\alpha = \nu - \lceil \nu - 1 \rceil \in (0, 1]$. Then the realizations of the random field are almost surely Hölder continuous, $\gamma(x, \omega) \in \mathcal{C}^{s,\beta}(\bar{D})$ with $\beta < \alpha$.

- For $\nu = 0.5$ the covariance function is only Lipschitz continuous and the field is almost surely Hölder continuous $\gamma(x, \omega) \in \mathcal{C}^{0,\alpha}(\bar{D})$ with $\alpha < 0.5$.
- For $\nu \to \infty$ the covariance function as well as the field are continuous with all their derivates, namely $\mathrm{cov}_\gamma(x), \gamma(x, \omega_i) \in C^\infty(\bar{D})\ \forall \omega \in \Omega$

(see [Graham, Kuo, Nichols, Scheichl, Schwab, Sloan '13])

Matérn Covariance Function: Regularity issues

Let $\nu = s + \alpha$ with $s = \lceil \nu - 1 \rceil \in \mathbb{N}$ and $\alpha = \nu - \lceil \nu - 1 \rceil \in (0, 1]$. Then the realizations of the random field are almost surely Hölder continuous, $\gamma(x, \omega) \in \mathcal{C}^{s,\beta}(\bar{D})$ with $\beta < \alpha$.

- For $\nu = 0.5$ the covariance function is only Lipschitz continuous and the field is almost surely Hölder continuous $\gamma(x, \omega) \in \mathcal{C}^{0,\alpha}(\bar{D})$ with $\alpha < 0.5$.
- For $\nu \to \infty$ the covariance function as well as the field are continuous with all their derivates, namely $\text{cov}_\gamma(x), \gamma(x, \omega_i) \in C^\infty(\bar{D})\ \forall \omega \in \Omega$

(see [Graham, Kuo, Nichols, Scheichl, Schwab, Sloan '13])

MLMC

$Q = Q(p)$: QoI related to the solution of the PDE
$\mathcal{T}_{h_0}, \dots, \mathcal{T}_{h_L}$: sequence of increasingly fine triangulations
$Y_{h_l} = Q_{h_l} - Q_{h_{l-1}}$: difference of the QoI between two consecutive grids.

Telescopic sum + Linearity of expectation:

$$\mathbb{E}[Q_{h_L}] = \sum_{l=0}^{L} \mathbb{E}[Y_{h_l}], \quad Q_{h_{-1}} = 0$$

MLMC Estimator:

$$\hat{Q}^{MLMC}_{L,\{M_l\}} = \sum_{l=0}^{L} \frac{1}{M_l} \sum_{i=1}^{M_l} \left(Q_{h_l}(\omega_i) - Q_{h_{l-1}}(\omega_i) \right)$$

see [Teckentrup, Scheichl, Giles, Ullmann '12], [Charrier, Scheichl, Teckentrup '11]

MLMC – Mean Square Error (MSE)

$$e(\hat{Q}^{MLMC}_{L,\{M_l\}})^2 = \underbrace{\sum_{l=0}^{L} \frac{\mathbb{V}\mathrm{ar}(Y_{h_l})}{M_l}}_{(i)} + \underbrace{\mathbb{E}[Q_{h_L} - Q]^2}_{(ii)}$$

M_l: number of samples on each level. A good choice of M_l, for $l = 0, ..., L$, represents a crucial issue for the effectiveness of the method (see [Cliffe, Giles, Scheichl, Teckentrup, '11], [Barth, Schwab, Zollinger, '11])

(i): represents the variance of the estimator, i.e. the statistical error: it is expected to be significantly smaller than the variance of the standard MC estimator

(ii): represents the bias of the error, due to the finite element (FE) discretization

Control Variate

Idea: use the solution of an auxiliary problem with regularized coefficient as control variate

Problem: we do not know exactly the expected value of the control variate. However this can be computed efficiently by a Stochastic Collocation (SC) technique

Original Problem

$$\begin{cases} -\mathrm{div}(a\nabla p) = f & \text{in } D, \\ + \text{ boundary conditions} & \text{on } \partial D. \end{cases}$$

Auxiliary Problem

$$\begin{cases} -\mathrm{div}(a^\epsilon\nabla p^\epsilon) = f & \text{in } D, \\ + \text{ boundary conditions} & \text{on } \partial D. \end{cases}$$

- $a = e^{\gamma}$: random field obtained starting from the Matérn covariance function with parameter ν
- $a^\epsilon = e^{\gamma^\epsilon}$: regularized version of a

Control Variate

Regularized Gaussian random field obtained via convolution with a Gaussian kernel

$$\gamma^{\epsilon} = \gamma * \phi_{\epsilon}(x), \text{ where } \phi_{\epsilon} = \frac{1}{(2\pi\epsilon^2)^{\frac{d}{2}}} e^{-\frac{\|x\|^2}{2\epsilon^2}}$$

Quantity of interest defined via control variate

$$Q^{CV} = Q - (Q^{\epsilon} - \mathbb{E}[Q^{\epsilon}])$$

where $Q = Q(p)$ and $Q^{\epsilon} = Q(p^{\epsilon})$

Control Variate

Regularized Gaussian random field obtained via convolution with a Gaussian kernel

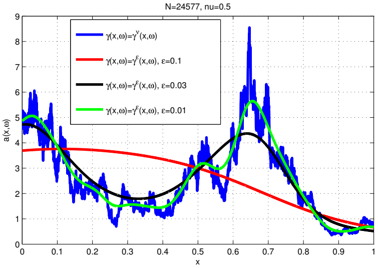
$$\gamma^{\epsilon} = \gamma * \phi_{\epsilon}(x), \text{ where } \phi_{\epsilon} = \frac{1}{(2\pi\epsilon^2)^{\frac{d}{2}}} e^{-\frac{\|x\|^2}{2\epsilon^2}}$$

Quantity of interest defined via control variate

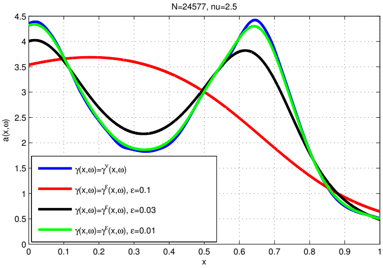
$$Q^{CV} = Q - (Q^{\epsilon} - \mathbb{E}[Q^{\epsilon}])$$

where $Q = Q(p)$ and $Q^{\epsilon} = Q(p^{\epsilon})$

Control Variate



Lipschitz continuous covariance function ($\nu = 0.5$)



Twice differentiable covariance function ($\nu = 2.5$)

MLMC with Control Variate

$\mathcal{T}_{h_0}, \dots, \mathcal{T}_{h_L}$: sequence of increasingly fine triangulations;

$Y_{h_l}^{CV} = Q_{h_l}^{CV} - Q_{h_{l-1}}^{CV}$: difference of the QoI between two consecutive grids.

Telescopic sum + Linearity of expectation:

$$\mathbb{E}[Q_{h_L}^{CV}] = \sum_{l=0}^{L} \mathbb{E}[Y_{h_l}^{CV}], \quad Q_{h_{-1}}^{CV} = 0$$

MLCV Estimator

$$\hat{Q}_{h_L,\{M_l\}}^{MLCV} = \sum_{l=0}^{L} \frac{1}{M_l} \sum_{i=1}^{M_l} \left(Q_{h_l}^{i} - Q_{h_{l-1}}^{i} - (Q_{h_l}^{\epsilon,i} - Q_{h_{l-1}}^{\epsilon,i}) \right) + \mathbb{E}[Q_{h_L}^{\epsilon,SC}]$$

MLMC with Control Variate

Mean Square Error (MSE)

$$e(\hat{Q}^{MLCV}_{h_L,\{M_l\}})^2 \leq \underbrace{\sum_{l=0}^{L} \frac{\mathbb{V}\mathrm{ar}(Y^{CV}_{h_l})}{M_l}}_{(i)} + \underbrace{2\mathbb{E}[Q^{\epsilon}_{h_L} - Q^{\epsilon,SC}_{h_L}]^2}_{(ii)} + \underbrace{2\mathbb{E}[Q_{h_L} - Q]^2}_{(iii)}$$

$\mathbb{E}[Q^{\epsilon,SC}_{h_L}]$: mean of the QoI $Q^{\epsilon}_{h_L}$ computed with a SC scheme on sparse grids.

(i): variance of the estimator

(ii): bias due to the SC approximation of the mean of the control variate $\mathbb{E}[Q^{\epsilon}]$

(iii): bias due to the finite element approximation

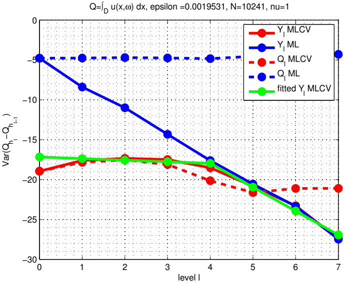
Remark: if ϵ tends to 0 the statistical error (i) vanishes; on the other hand keeping small the SC error (ii) becomes too costly.

MLMC with Control Variate

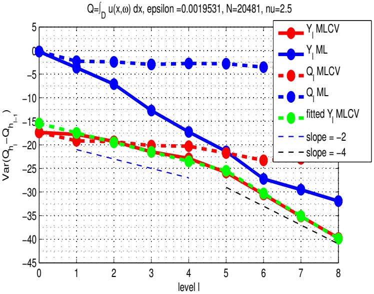
Error estimate (preliminary result)

$$\mathbb{E}\left[\left(Q - Q_{h_l} - (Q^{\epsilon} - Q^{\epsilon}_{h_l})\right)^2\right]^{\frac{1}{2}} \leq c h_l^{\min\{\nu,p\}} \min_{s=1,\dots,p} h_l^{\min\{\nu,s\}} \epsilon^{\min\{(\nu-s)_+,2\}}$$

Variance of the difference of the QoI between consecutive grids. The dashed lines represent the slopes h_l^2 and h_l^4.



$\nu = 1$, $\epsilon = 1/8^3$



$\nu = 2.5$, $\epsilon = 1/8^3$.

MLCV Algorithm

1. Given a prescribed *tol* select h_L in such a way to have $(iii) \leqslant tol^2$;
2. set $h_0 = O(|D|)$ and evaluate $(iv) = \mathbb{V}ar(Y^{CV}_{h_l})$ and $(v) = \mathbb{V}ar(Q^{CV}_{h_l})$; we can select among two basic strategies:
   Strategy 1: if $(iv) < (v)\ \forall\ l$ apply the MLCV scheme starting from level 0;
   Strategy 2: if $(iv) \approx (v)$ for $l = 0, ..., l_0$ set $h_0 = h_{l_0}$; and use the control variate only on level l_0 and a standard MLMC on subsequent levels, namely

$$\hat{Q}^{MLCV}_{h_L,\{M_l\}} = \frac{1}{M_{l_0}} \sum_{i=1}^{M_{l_0}} \left(Q^i_{h_{l_0}} - Q^{\epsilon,i}_{h_{l_0}} \right) + \sum_{l=l_0+1}^{L} \frac{1}{M_l} \sum_{i=1}^{M_l} \left(Q^i_{h_l} - Q^i_{h_{l-1}} \right) + \mathbb{E}[Q^{\epsilon,SC}_{h_{l_0}}]$$

3. according to the strategy selected compute the number of samples M_l for $l = 0, ..., L$ and the number of knots of the sparse grid M_{SG} by solving an optimization problem in such a way to have $(i) + (ii) \leqslant tol^2$

Optimization Problem

- ϵ is considered fixed
- C_l is the computational cost needed to solve one deterministic system on the grid of mesh size h_l; cost model: $C_l = 2C_{l-1} = ... = 2^l C_0 = \gamma 2^l h_0^{-1}$. For instance see [Cliffe, Giles, Scheichl, Teckentrup, '11]

Computational cost

- strategy 1: $C(M_l, M_{SG}) = 2M_0 C_0 + 2\sum_{l=1}^{L} M_l(C_l + C_{l-1}) + M_{SG} C_L$
- strategy 2: $C(M_l, M_{SG}) = 2M_{l_0} C_{l_0} + \sum_{l=l_0+1}^{L} M_l(C_l + C_{l-1}) + M_{SG} C_{l_0}$

Associated Error fitted model

- strategy 1: $e(M_l, M_{SG}) = \sum_{l=0}^{L} \frac{\min\left\{c_1 h_l^{2\min\{\nu,p\}}, c_2 h_l^{4\min\{\nu,p\}}\right\}}{M_l} + c_3 M_{SG}^{\alpha}$
- strategy 2: $e(M_l, M_{SG}) = \sum_{l=l_0}^{L} \frac{c_1 h_l^{4\min\{\nu,p\}}}{M_l} + c_2 M_{SG}^{\alpha}$

Perform a Lagrange optimization by considering

$$\mathcal{L}(M_l, M_{SG}, \lambda) = C(M_l, M_{SG}) - \lambda(e(M_l, M_{SG}) - tol^2)$$

Optimization Problem

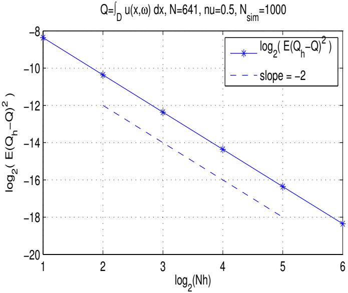
Values obtained (strategy one):

- $M_0 = \sqrt{-\lambda}\sqrt{\frac{c_1 h_0^{2\min\{\nu,p\}}}{2C_0 2^l}} = \sqrt{-\lambda}\sqrt{\frac{v_0}{2C_0}}$
- $M_l = \sqrt{-\lambda}\sqrt{\frac{\min\left\{c_1 h_l^{2\min\{\nu,p\}}, c_2 h_l^{4\min\{\nu,p\}}\right\}}{3C_0 2^l}} = \sqrt{-\lambda}\sqrt{\frac{v_l}{3C_0 2^l}}$ for $l = 1, ..., L$
- $M_{SG} = (-\lambda)^{\frac{1}{1-\alpha}} \left(-\alpha 2^{-L}\frac{c_3}{C_0}\right)^{\frac{1}{1-\alpha}}$

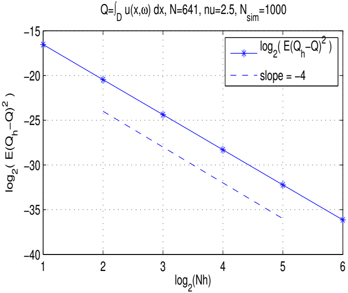
where λ has to be computed from:

$$\frac{1}{\sqrt{-\lambda}} = \frac{tol^2}{\sqrt{C_0}(\sqrt{2v_0} + \sum_{l=1}^{L}\sqrt{3v_l}2^l) + c_3(\frac{1}{\sqrt{-\lambda}})^{\frac{-1-\alpha}{1-\alpha}}(-\alpha 2^{-L}\frac{c_3}{C_0})^{\frac{\alpha}{1-\alpha}}}$$

Approximation error $\mathbb{E}[Q_{h_L} - Q]^2$

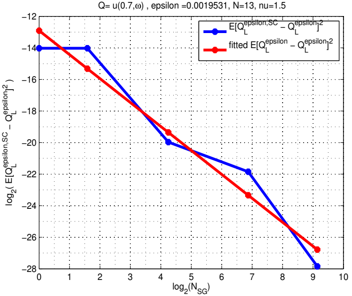


$\nu = 0.5$

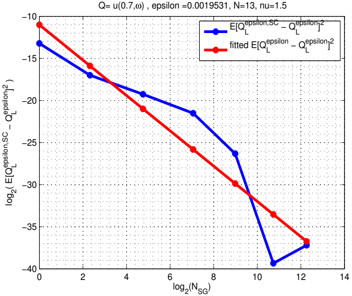


$\nu = 2.5$

SC error $\mathbb{E}[Q_{h_L}^{\epsilon,SC} - Q^{\epsilon}]^2$



$\nu = 0.5$; fitted slope $\alpha = -1.5$



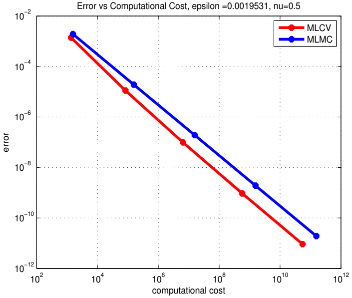
$\nu = 2.5$; fitted slope $\alpha = -2.2$

In both cases more than the 99% of the variability has been taken into account. The fitted rate is better than MC in both cases.

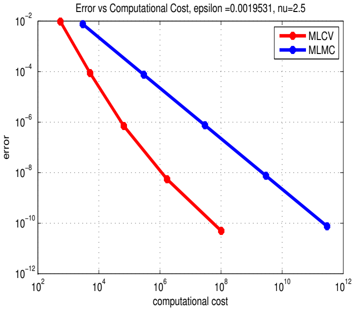
Error vs Cost : MLCV vs MLMC

Error and computational cost associated to several values of $tol = 10^{-1}, ..., 10^{-5}$. $\epsilon = 1/8^3$ in both cases.

error = statistical error + SC error



$\nu = 0.5$



$\nu = 2.5$

Outline

1. Model problem: elliptic PDEs with random coefficients
2. Polynomial approximation by sparse grid collocation
    - Quasi optimal sparse grid construction
    - Convergence result for elliptic PDEs with random inclusions
    - Numerical results
3. Combined MLMC / Sparse grid method
    - MLMC with Sparse Grid Control Variate
    - Variance analysis and algorithm tuning
    - Numerical results
4. Conclusions

Conclusions

- We have analyzed the convergence of quasi-optimal sparse grid approximations based on the selection of the M most profitable hierarchical surpluses. Convergence rates are related to summability properties of the profits, weighted by the corresponding works.
- Sharp a-priori / a-posteriori analysis of the decay of the polynomial chaos expansion of the solution allows to construct optimized sparse grids that provide effective approximations also in infinite dimensions for smooth fields.
- The "profit based" a-posteriori adaptive algorithm is also performing very efficiently, close to the best approximation.
- for rough, longly correlated fields, a good idea is to use a sparse grid polynomial approximation on a smoothed problem as a control variate in a MLMC algorith. Preliminary results show a considerable improvement of the overall complexity.

Conclusions

- We have analyzed the convergence of quasi-optimal sparse grid approximations based on the selection of the M most profitable hierarchical surpluses. Convergence rates are related to summability properties of the profits, weighted by the corresponding works.
- Sharp a-priori / a-posteriori analysis of the decay of the polynomial chaos expansion of the solution allows to construct optimized sparse grids that provide effective approximations also in infinite dimensions for smooth fields.
- The "profit based" a-posteriori adaptive algorithm is also performing very efficiently, close to the best approximation.
- for rough, longly correlated fields, a good idea is to use a sparse grid polynomial approximation on a smoothed problem as a control variate in a MLMC algorith. Preliminary results show a considerable improvement of the overall complexity.

Conclusions

- We have analyzed the convergence of quasi-optimal sparse grid approximations based on the selection of the M most profitable hierarchical surpluses. Convergence rates are related to summability properties of the profits, weighted by the corresponding works.
- Sharp a-priori / a-posteriori analysis of the decay of the polynomial chaos expansion of the solution allows to construct optimized sparse grids that provide effective approximations also in infinite dimensions for smooth fields.
- The "profit based" a-posteriori adaptive algorithm is also performing very efficiently, close to the best approximation.
- for rough, longly correlated fields, a good idea is to use a sparse grid polynomial approximation on a smoothed problem as a control variate in a MLMC algorith. Preliminary results show a considerable improvement of the overall complexity.

Conclusions

- We have analyzed the convergence of quasi-optimal sparse grid approximations based on the selection of the M most profitable hierarchical surpluses. Convergence rates are related to summability properties of the profits, weighted by the corresponding works.
- Sharp a-priori / a-posteriori analysis of the decay of the polynomial chaos expansion of the solution allows to construct optimized sparse grids that provide effective approximations also in infinite dimensions for smooth fields.
- The "profit based" a-posteriori adaptive algorithm is also performing very efficiently, close to the best approximation.
- for rough, longly correlated fields, a good idea is to use a sparse grid polynomial approximation on a smoothed problem as a control variate in a MLMC algorith. Preliminary results show a considerable improvement of the overall complexity.

References

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

J. Beck, F. Nobile, L. Tamellini, and R. Tempone.
A quasi-optimal sparse grids procedure for groundwater flows, LNCSE, 2013, in press.

L. Tamellini,
Polynomial approximation of PDEs with stochastic coefficients, PhD Thesis, Politecnico di Milano. March 2012.

J. Beck and F. Nobile and L. Tamellini and R. Tempone
On the optimal polynomial approximation of stochastic PDEs by Galerkin and Collocation methods, Math. Mod. Methods Appli. Sci. (M3AS), 22(9), 2012.

J. Bäck, F. Nobile, L. Tamellini and R. Tempone
Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison, LNCSE Springer, 76:43-62, 2011

I. Babuška, F. Nobile and R. Tempone.
A stochastic collocation method for elliptic PDEs with random input data, SIAM Review, 52(2):317–355, 2010

Thank you for your attention!