Multilevel discrete least squares polynomial apporixmation

Fabio Nobile

CSQI-MATH, EPFL, Switzerland

Joint work with: R. Tempone, S. Wolfers (KAUST), A-L. Haji Ali (Oxford)

Acknowledgements: L. Tamellini (CNR Pavia), A. Cohen, G. Migliorati (UPMC)

QUIET 2017
"Quantification of Uncertainty: Improving Efficiency and Technology",
SISSA, Trieste, Italy, July 18-21, 2017

Outline

PDEs with random parameters

Consider a differential problem

$$\mathcal{L}(\mathbf{y}; u) = \mathcal{G} \qquad (*)$$

depending on a set of random parameters $\mathbf{y} = (y_1, \dots, y_N) \in \Gamma \subset \mathbb{R}^N$ with joint probability measure μ on Γ.

We assume that $(*)$ has a unique solution $u(\mathbf{y})$ in some suitable function space V and we focus on a Quantity of Interest $Q : V \to \mathbb{R}$.

Goal: approximate the whole response function

$$\mathbf{y} \mapsto f(\mathbf{y}) := Q(u(\mathbf{y})) : \Gamma \to \mathbb{R}$$

by multivariate polynomials.

Possibly derive approximated statistics as $\mathbb{E}[f]$, $Var[f]$, etc.

PDEs with random parameters

Consider a differential problem

$$\mathcal{L}(\mathbf{y}; u) = \mathcal{G} \qquad (*)$$

depending on a set of random parameters $\mathbf{y} = (y_1, \dots, y_N) \in \Gamma \subset \mathbb{R}^N$ with joint probability measure μ on Γ.

We assume that $(*)$ has a unique solution $u(\mathbf{y})$ in some suitable function space V and we focus on a Quantity of Interest $Q : V \to \mathbb{R}$.

Goal: approximate the whole response function

$$\mathbf{y} \mapsto f(\mathbf{y}) := Q(u(\mathbf{y})) : \Gamma \to \mathbb{R}$$

by multivariate polynomials.

Possibly derive approximated statistics as $\mathbb{E}[f]$, $Var[f]$, etc.

PDEs with random parameters

Consider a differential problem

$$\mathcal{L}(\mathbf{y}; u) = \mathcal{G} \qquad (*)$$

depending on a set of random parameters $\mathbf{y} = (y_1, \ldots, y_N) \in \Gamma \subset \mathbb{R}^N$ with joint probability measure μ on Γ.

We assume that $(*)$ has a unique solution $u(\mathbf{y})$ in some suitable function space V and we focus on a Quantity of Interest $Q : V \to \mathbb{R}$.

Goal: approximate the whole response function

$$\mathbf{y} \mapsto f(\mathbf{y}) := Q(u(\mathbf{y})) : \Gamma \to \mathbb{R}$$

by multivariate polynomials.

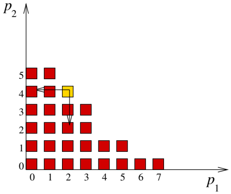
Possibly derive approximated statistics as $\mathbb{E}[f]$, $Var[f]$, etc.

Polynomial approximation on downward closed sets

Assume $f \in L^2_\mu(\Gamma)$. We seek an approximation of f in a finite dimensional polynomial subspace

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

with $\Lambda \subset \mathbb{N}^N$ a downward closed index set.



Definition. An index set Λ is downward closed if

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

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

1. Weighted discrete least squares approximation
2. Multilevel least squares approximation
3. Application to random elliptic PDEs
4. Conclusions

Weighted discrete least squares approximation

1. sample independently M points $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}) \in \Gamma^M$ from a distribution $\nu \ll \mu$, with density $\rho = \frac{d\nu}{d\mu}$
2. define the weight function $w(\mathbf{y}) = \frac{1}{\rho(\mathbf{y})}$
3. weighted discrete least squares approximation on V_Λ

$$\hat{\Pi}_M f = \operatorname*{argmin}_{v \in V_\Lambda} \|f - v\|_M, \quad \text{with} \quad \|g\|_M^2 = \frac{1}{M} \sum_{j=1}^{M} w(\mathbf{y}^{(j)}) g(\mathbf{y}^{(j)})^2$$

Remark: $\mathbb{E}[\|g\|_M^2] = \int_\Gamma w(\mathbf{y}) g(\mathbf{y})^2 \nu(d\mathbf{y}) = \int_\Gamma g(\mathbf{y})^2 \mu(d\mathbf{y}) = \|g\|_{L^2_\mu}^2$

Algebraic system: let $\{\phi_j\}_{j=1}^{|\Lambda|}$ be a basis of V_Λ orthonormal w.r.t. μ and $\hat{\Pi}_M f(\mathbf{y}) = \sum_{j=1}^{|\Lambda|} c_j \phi_j(\mathbf{y})$. Then $\mathbf{c} = (c_1, \ldots, c_{|\Lambda|})^T$ satisfies

$$G\mathbf{c} = \hat{\mathbf{f}}, \qquad G_{i,j} = (\phi_i, \phi_j)_M, \quad \hat{f}_i = (f, \phi_i)_M$$

Weighted discrete least squares approximation

1. sample independently M points $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}) \in \Gamma^M$ from a distribution $\nu \ll \mu$, with density $\rho = \frac{d\nu}{d\mu}$
2. define the weight function $w(\mathbf{y}) = \frac{1}{\rho(\mathbf{y})}$
3. weighted discrete least squares approximation on V_Λ

$$\hat{\Pi}_M f = \operatorname*{argmin}_{v \in V_\Lambda} \|f - v\|_M, \quad \text{with} \quad \|g\|_M^2 = \frac{1}{M} \sum_{j=1}^{M} w(\mathbf{y}^{(j)}) g(\mathbf{y}^{(j)})^2$$

Remark: $\mathbb{E}[\|g\|_M^2] = \int_\Gamma w(\mathbf{y}) g(\mathbf{y})^2 \nu(d\mathbf{y}) = \int_\Gamma g(\mathbf{y})^2 \mu(d\mathbf{y}) = \|g\|_{L^2_\mu}^2$

Algebraic system: let $\{\phi_j\}_{j=1}^{|\Lambda|}$ be a basis of V_Λ orthonormal w.r.t. μ and $\hat{\Pi}_M f(\mathbf{y}) = \sum_{j=1}^{|\Lambda|} c_j \phi_j(\mathbf{y})$. Then $\mathbf{c} = (c_1, \ldots, c_{|\Lambda|})^T$ satisfies

$$G\mathbf{c} = \hat{\mathbf{f}}, \qquad G_{i,j} = (\phi_i, \phi_j)_M, \quad \hat{f}_i = (f, \phi_i)_M$$

Weighted discrete least squares approximation

1. sample independently M points $(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}) \in \Gamma^M$ from a distribution $\nu \ll \mu$, with density $\rho = \frac{d\nu}{d\mu}$
2. define the weight function $w(\mathbf{y}) = \frac{1}{\rho(\mathbf{y})}$
3. weighted discrete least squares approximation on V_Λ

$$\hat{\Pi}_M f = \operatorname*{argmin}_{v \in V_\Lambda} \|f - v\|_M, \quad \text{with } \|g\|_M^2 = \frac{1}{M} \sum_{j=1}^{M} w(\mathbf{y}^{(j)}) g(\mathbf{y}^{(j)})^2$$

Remark: $\mathbb{E}[\|g\|_M^2] = \int_\Gamma w(\mathbf{y}) g(\mathbf{y})^2 \nu(d\mathbf{y}) = \int_\Gamma g(\mathbf{y})^2 \mu(d\mathbf{y}) = \|g\|_{L^2_\mu}^2$

Algebraic system: let $\{\phi_j\}_{j=1}^{|\Lambda|}$ be a basis of V_Λ orthonormal w.r.t. μ and $\hat{\Pi}_M f(\mathbf{y}) = \sum_{j=1}^{|\Lambda|} c_j \phi_j(\mathbf{y})$. Then $\mathbf{c} = (c_1, \dots, c_{|\Lambda|})^T$ satisfies

$$G\mathbf{c} = \hat{\mathbf{f}}, \qquad G_{i,j} = (\phi_i, \phi_j)_M, \quad \hat{f}_i = (f, \phi_i)_M$$

Optimality of discrete least squares approximation

Theorem [Cohen-Migliorati 2017] [Cohen-Davenport-Leviatan 2013]

For arbitrary $r > 0$ define

$$\kappa_r := \frac{1/2(1-\log 2)}{1+r}, \qquad K_{\Lambda,w} := \sup_{\mathbf{y}\in\Gamma}\left(w(\mathbf{y}) \sum_{j=1}^{|\Lambda|} \phi_i(\mathbf{y})^2 \right)$$

If $\frac{M}{\log M} \geq \frac{K_{\Lambda,w}}{\kappa_r}$, then

- $P(\|G-I\| \leq \frac{1}{2}) > 1 - 2M^{-r}$
- $\|f - \hat{\Pi}_M f\|_{L^2_\mu} \leq (1+\sqrt{2}) \inf_{v\in V_\Lambda} \|f-v\|_{L^\infty_{\sqrt{w}}}$ with prob. $> 1 - 2M^{-r}$
- $\mathbb{E}[f - \hat{\Pi}^c_M f\|^2_{L^2_\mu}] \leq C_M \inf_{v\in V_\Lambda} \|f-v\|^2_{L^2_\mu} + 2\|f\|^2_{L^2_\mu} M^{-r}$

  where $\hat{\Pi}^c_M f = \hat{\Pi}_M f \cdot \mathbf{1}_{\{\|G-I\|\leq\frac{1}{2}\}}$ and $C_M = (1 + \frac{4\kappa_r}{\log M}) \xrightarrow{M\to\infty} 1$

Sufficient number of points for stability

- Uniform measure: $\mu = \mathcal{U}(\prod_{i=1}^{N} \Gamma_i)$
  [Chkifa-Cohen-Migliorati-N.-Tempone 2015] When sampling from the same distribution ($\nu = \mu$ and $w = 1$) then

$$|\Lambda| \leq K_{\Lambda,1} \leq |\Lambda|^2$$

  Hence (unweighted) discrete least square is stable and optimally convergence under the condition

$$\frac{M}{\log M} \geq \frac{|\Lambda|^2}{\kappa_r} \qquad \text{(quadratic proportionality)}$$

Sufficient number of points for stability - optimal measure

[Cohen-Migliorati 2017] For arbitrary μ, when sampling from the optimal measure

$$\frac{d\nu^*}{d\mu}(\mathbf{y}) = \rho^*(\mathbf{y}) = \frac{1}{|\Lambda|}\sum_{j=1}^{|\Lambda|}\phi_j(\mathbf{y})^2, \quad \Longrightarrow \quad K_{\Lambda,w^*} = 1$$

weighted discrete least squares stable and optimal with

$$\frac{M}{\log M} \geq \frac{|\Lambda|}{\kappa_r} \qquad \text{(linear proportionality)}$$

Sampling algorithms

- Sampling algorithms from the optimal distribution are available (marginalization [Cohen-Migliorati 2017], acceptance rejection [HajiAli-N.-Tempone-Wolfers, 2017])
  However, the optimal distribution depends on Λ. Not good for adaptive algorithms
- Alternatively, for uniform measure μ (or more generally a product measure $\mu = \otimes_{j=1}^{N} \mu_j$, with μ_j doubling measure, i.e. $\mu_j(2I) = L\mu_j(I)$) one can sample from the arcsin (Chebyshev) distribution.

$$K_{\Lambda,w} \leq C^N |\Lambda|, \qquad \frac{M}{\log M} \geq \frac{C^N}{\kappa_r} |\Lambda|$$

  Still linear scaling but with a constant exponentially dependent on N. Advantage: the sampling measure does not depend on Λ. Good for adaptivity.

In both cases, the cost of computing $\hat{\Pi}_M f$ is linear in $|\Lambda|$ up to logarithmic terms.

Sampling algorithms

- Sampling algorithms from the optimal distribution are available (marginalization [Cohen-Migliorati 2017], acceptance rejection [HajiAli-N.-Tempone-Wolfers, 2017])
  However, the optimal distribution depends on Λ. Not good for adaptive algorithms
- Alternatively, for uniform measure μ (or more generally a product measure $\mu = \otimes_{j=1}^{N} \mu_j$, with μ_j doubling measure, i.e. $\mu_j(2I) = L\mu_j(I)$) one can sample from the arcsin (Chebyshev) distribution.

$$K_{\Lambda,w} \leq C^N |\Lambda|, \qquad \frac{M}{\log M} \geq \frac{C^N}{\kappa_r} |\Lambda|$$

  Still linear scaling but with a constant exponentially dependent on N.
  Advantage: the sampling measure does not depend on Λ. Good for adaptivity.

In both cases, the cost of computing $\hat{\Pi}_M f$ is linear in $|\Lambda|$ up to logarithmic terms.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Sampling algorithms

- Sampling algorithms from the optimal distribution are available (marginalization [Cohen-Migliorati 2017], acceptance rejection [HajiAli-N.-Tempone-Wolfers, 2017])
  However, the optimal distribution depends on Λ. Not good for adaptive algorithms
- Alternatively, for uniform measure μ (or more generally a product measure $\mu = \otimes_{j=1}^{N} \mu_j$, with μ_j doubling measure, i.e. $\mu_j(2I) = L\mu_j(I)$) one can sample from the arcsin (Chebyshev) distribution.

$$K_{\Lambda,w} \leq C^N |\Lambda|, \qquad \frac{M}{\log M} \geq \frac{C^N}{\kappa_r} |\Lambda|$$

  Still linear scaling but with a constant exponentially dependent on N.
  Advantage: the sampling measure does not depend on Λ. Good for adaptivity.

In both cases, the cost of computing $\hat{\Pi}_M f$ is linear in $|\Lambda|$ up to logarithmic terms.

Outline

Multilevel least squares approximation

In practice $f(\mathbf{y})$ can not be evaluated exactly as it implies the solution of a PDE.

- We introduce a sequence of approximations f_{n_ℓ}, $n_\ell \in \mathbb{N}$ with increasing cost, s.t.

$$\lim_{\ell \to \infty} \|f - f_{n_\ell}\|_{L^2_\mu} = 0$$

  (or possibly a stronger norm)

- Similarly, we introduce a sequence of nested downward closed sets

$$\Lambda_{m_0} \subset \Lambda_{m_1} \subset \ldots \subset \Lambda_{m_k} \subset \ldots$$

  such that

$$\lim_{k \to \infty} \inf_{v \in V_{\Lambda_k}} \|f - v\|_{L^2_\mu} = 0$$

  Correspondingly, for each Λ_{m_k} we introduce a weighted discrete least squares projector $\hat{\Pi}_{M_k}$ using $\frac{M_k}{\log M_k} = O(|\Lambda_{m_k}|)$ random points.

Multilevel least squares approximation

In practice $f(\mathbf{y})$ can not be evaluated exactly as it implies the solution of a PDE.

- We introduce a sequence of approximations f_{n_ℓ}, $n_\ell \in \mathbb{N}$ with increasing cost, s.t.

$$\lim_{\ell \to \infty} \|f - f_{n_\ell}\|_{L^2_\mu} = 0$$

(or possibly a stronger norm)

- Similarly, we introduce a sequence of nested downward closed sets

$$\Lambda_{m_0} \subset \Lambda_{m_1} \subset \ldots \subset \Lambda_{m_k} \subset \ldots$$

such that

$$\lim_{k \to \infty} \inf_{v \in V_{\Lambda_k}} \|f - v\|_{L^2_\mu} = 0$$

Correspondingly, for each Λ_{m_k} we introduce a weighted discrete least squares projector $\hat{\Pi}_{M_k}$ using $\frac{M_k}{\log M_k} = O(|\Lambda_{m_k}|)$ random points.

Multilevel least squares approximation

In practice $f(\mathbf{y})$ can not be evaluated exactly as it implies the solution of a PDE.

- We introduce a sequence of approximations f_{n_ℓ}, $n_\ell \in \mathbb{N}$ with increasing cost, s.t.
$$\lim_{\ell\to\infty} \|f - f_{n_\ell}\|_{L^2_\mu} = 0$$
(or possibly a stronger norm)
- Similarly, we introduce a sequence of nested downward closed sets
$$\Lambda_{m_0} \subset \Lambda_{m_1} \subset \ldots \subset \Lambda_{m_k} \subset \ldots$$
such that
$$\lim_{k\to\infty} \inf_{v\in V_{\Lambda_k}} \|f - v\|_{L^2_\mu} = 0$$
Correspondingly, for each Λ_{m_k} we introduce a weighted discrete least squares projector $\hat{\Pi}_{M_k}$ using $\frac{M_k}{\log M_k} = O(|\Lambda_{m_k}|)$ random points.

Multilevel least squares approximation

Multilevel formula: given maximum level $L \in \mathbb{N}$

$$S_L f = \sum_{k+\ell \le L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}})$$
$$= \sum_{\ell=0}^{L} \hat{\Pi}_{M_{L-\ell}}(f_{n_\ell} - f_{n_{\ell-1}})$$

- In the multilevel formula one might consider more general index sets $(k, \ell) \in \mathcal{I} \subset \mathbb{R}^2$. However, one can always recast to $k + \ell \le L$ by properly choosing $\{n_\ell\}$ and $\{m_k\}$.
- **Question**: How to properly choose $\{n_\ell\}$, $\{m_k\}$ and $\{M_k\}$?
- **Issue**: Since the least squares projection is random, we have to ensure that it is stable and optimally convergent on all levels. (Need union bound on failure probabilities)

Multilevel least squares approximation

Multilevel formula: given maximum level $L \in \mathbb{N}$

$$S_L f = \sum_{k+\ell \leq L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}})$$
$$= \sum_{\ell=0}^{L} \hat{\Pi}_{M_{L-\ell}}(f_{n_\ell} - f_{n_{\ell-1}})$$

- In the multilevel formula one might consider more general index sets $(k, \ell) \in \mathcal{I} \subset \mathbb{R}^2$. However, one can always recast to $k + \ell \leq L$ by properly choosing $\{n_\ell\}$ and $\{m_k\}$.
- **Question**: How to properly choose $\{n_\ell\}$, $\{m_k\}$ and $\{M_k\}$?
- **Issue**: Since the least squares projection is random, we have to ensure that it is stable and optimally convergent on all levels. (Need union bound on failure probabilities)

Multilevel least squares approximation

Multilevel formula: given maximum level $L \in \mathbb{N}$

$$S_L f = \sum_{k+\ell \leq L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}})$$
$$= \sum_{\ell=0}^{L} \hat{\Pi}_{M_{L-\ell}}(f_{n_\ell} - f_{n_{\ell-1}})$$

- In the multilevel formula one might consider more general index sets $(k, \ell) \in \mathcal{I} \subset \mathbb{R}^2$. However, one can always recast to $k + \ell \leq L$ by properly choosing $\{n_\ell\}$ and $\{m_k\}$.
- **Question**: How to properly choose $\{n_\ell\}$, $\{m_k\}$ and $\{M_k\}$?
- **Issue**: Since the least squares projection is random, we have to ensure that it is stable and optimally convergent on all levels. (Need union bound on failure probabilities)

Multilevel least squares approximation

Multilevel formula: given maximum level $L \in \mathbb{N}$

$$S_L f = \sum_{k+\ell \leq L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}})$$
$$= \sum_{\ell=0}^{L} \hat{\Pi}_{M_{L-\ell}}(f_{n_\ell} - f_{n_{\ell-1}})$$

- In the multilevel formula one might consider more general index sets $(k,\ell) \in \mathcal{I} \subset \mathbb{R}^2$. However, one can always recast to $k+\ell \leq L$ by properly choosing $\{n_\ell\}$ and $\{m_k\}$.
- **Question**: How to properly choose $\{n_\ell\}$, $\{m_k\}$ and $\{M_k\}$?
- **Issue**: Since the least squares projection is random, we have to ensure that it is stable and optimally convergent on all levels. (Need union bound on failure probabilities)

Tuning of the ML least squares algorithm

- For the Multilevel algorithm to be effective, we have to rely on certain "mixed regularity"
- Let $(F, \|\cdot\|_F) \hookrightarrow (L^2_\mu, \|\cdot\|_{L^2_\mu})$ be a normed vector space of "smooth" functions (e.g. Hölder / Sobolev / analytic regularity)

- **Assumption 1 (regularity)**: $f, f_{n_\ell} \in F$ for all $\ell \in \mathbb{N}$
- **Assumption 2 (PDE discretization)**: the sequence $\{f_{n_\ell}\}$ is s.t.

$$\|f - f_{n_\ell}\|_{L^2_\mu} \lesssim n_\ell^{-\beta_w}, \qquad \|f - f_{n_\ell}\|_F \lesssim n_\ell^{-\beta_s}$$

and for a single $\mathbf{y} \in \Gamma$, the cost of computing $f_{n_\ell}(\mathbf{y})$ is

$$\text{Work}(f_{n_\ell}) \lesssim n_l^{\gamma}.$$

Tuning of the ML least squares algorithm

- For the Multilevel algorithm to be effective, we have to rely on certain "mixed regularity"
- Let $(F, \|\cdot\|_F) \hookrightarrow (L^2_\mu, \|\cdot\|_{L^2_\mu})$ be a normed vector space of "smooth" functions (e.g. Hölder / Sobolev / analytic regularity)

- **Assumption 1 (regularity)**: $f, f_{n_\ell} \in F$ for all $\ell \in \mathbb{N}$
- **Assumption 2 (PDE discretization)**: the sequence $\{f_{n_\ell}\}$ is s.t.

$$\|f - f_{n_\ell}\|_{L^2_\mu} \lesssim n_\ell^{-\beta_w}, \qquad \|f - f_{n_\ell}\|_F \lesssim n_\ell^{-\beta_s}$$

and for a single $\mathbf{y} \in \Gamma$, the cost of computing $f_{n_\ell}(\mathbf{y})$ is

$$\mathrm{Work}(f_{n_\ell}) \lesssim n_l^\gamma.$$

Tuning of the ML least squares algorithm

- For the Multilevel algorithm to be effective, we have to rely on certain "mixed regularity"
- Let $(F, \|\cdot\|_F) \hookrightarrow (L^2_\mu, \|\cdot\|_{L^2_\mu})$ be a normed vector space of "smooth" functions (e.g. Hölder / Sobolev / analytic regularity)

- **Assumption 1 (regularity)**: $f, f_{n_\ell} \in F$ for all $\ell \in \mathbb{N}$
- **Assumption 2 (PDE discretization)**: the sequence $\{f_{n_\ell}\}$ is s.t.

$$\|f - f_{n_\ell}\|_{L^2_\mu} \lesssim n_\ell^{-\beta_w}, \qquad \|f - f_{n_\ell}\|_F \lesssim n_\ell^{-\beta_s}$$

and for a single $\mathbf{y} \in \Gamma$, the cost of computing $f_{n_\ell}(\mathbf{y})$ is

$$\mathrm{Work}(f_{n_\ell}) \lesssim n_l^{\gamma}.$$

Tuning the ML least squares algorithm

- **Assumption 3 (polynomial approximability)**: the sequence $\{\Lambda_{m_k}\}$ is s.t.

$$\dim(V_{\Lambda_{m_k}}) = |\Lambda_{m_k}| \lesssim m_k^{\sigma}$$

$$\inf_{v \in V_{\Lambda_{m_k}}} \|f - v\|_{L^\infty_{\sqrt{w}}} \lesssim m^{-\alpha_p} \|f\|_F, \quad \forall f \in F$$

(Alternatively $\inf_{v \in V_{\Lambda_{m_k}}} \|f - v\|_{L^2_\mu} \lesssim m^{-\alpha_e} \|f\|_F, \quad \forall f \in F$)

Tuning the ML least squares algorithm

We now choose

$$n_\ell = C \exp\{\frac{\ell}{\gamma + \beta_s}\}, \quad \ell = 0, \dots, L \quad \text{(space discr.)}$$

$$m_k = C \exp\{\frac{k}{\sigma + \alpha_p}\}, \quad k = 0, \dots, L \quad \text{(Polynomial approx.)}$$

$$\frac{m_k}{\kappa_L} \leq \frac{M_k}{\log M_k} \leq \frac{2m_k^\sigma}{\kappa_L}, \quad k = 0, \dots, L \quad \text{(sample size with } r = L\text{)}$$

By taking $r = L$ we guarantee that

$$P(\exists k : \|G_k - I\| > \frac{1}{2}) \leq \sum_{k=0}^{L} P(\|G_k - I\| > \frac{1}{2}) \lesssim L^{-L}$$

Remark: This formulation is analogous to the anisotropic sparse approx.

$$S_L f = \sum_{(\sigma+\alpha_p)k+(\gamma+\beta_s)\ell \leq L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}}), \quad \text{with } n_\ell = Ce^\ell, \; m_k = Ce^k.$$

Tuning the ML least squares algorithm

We now choose

$$n_\ell = C \exp\{\frac{\ell}{\gamma + \beta_s}\}, \qquad \ell = 0, \dots, L \quad \text{(space discr.)}$$

$$m_k = C \exp\{\frac{k}{\sigma + \alpha_p}\}, \quad k = 0, \dots, L \quad \text{(Polynomial approx.)}$$

$$\frac{m_k}{\kappa_L} \leq \frac{M_k}{\log M_k} \leq \frac{2m_k^\sigma}{\kappa_L}, \quad k = 0, \dots, L \quad \text{(sample size with } r = L\text{)}$$

By taking $r = L$ we guarantee that

$$P(\exists k : \|G_k - I\| > \frac{1}{2}) \ \leq \ \sum_{k=0}^{L} P(\|G_k - I\| > \frac{1}{2}) \ \lesssim L^{-L}$$

Remark: This formulation is analogous to the anisotropic sparse approx.

$$S_L f = \sum_{(\sigma+\alpha_p)k+(\gamma+\beta_s)\ell\leq L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}}), \quad \text{with } n_\ell = Ce^\ell,\ m_k = Ce^k.$$

Tuning the ML least squares algorithm

We now choose

$$n_\ell = C\exp\{\frac{\ell}{\gamma+\beta_s}\}, \quad \ell = 0,\dots,L \quad \text{(space discr.)}$$

$$m_k = C\exp\{\frac{k}{\sigma+\alpha_p}\}, \quad k = 0,\dots,L \quad \text{(Polynomial approx.)}$$

$$\frac{m_k}{\kappa_L} \le \frac{M_k}{\log M_k} \le \frac{2m_k^\sigma}{\kappa_L}, \quad k = 0,\dots,L \quad \text{(sample size with } r = L\text{)}$$

By taking $r = L$ we guarantee that

$$P(\exists k : \|G_k - I\| > \frac{1}{2}) \le \sum_{k=0}^{L} P(\|G_k - I\| > \frac{1}{2}) \lesssim L^{-L}$$

Remark: This formulation is analogous to the anisotropic sparse approx.

$$S_L f = \sum_{(\sigma+\alpha_p)k+(\gamma+\beta_s)\ell\le L} (\hat{\Pi}_{M_k} - \hat{\Pi}_{M_{k-1}})(f_{n_\ell} - f_{n_{\ell-1}}), \quad \text{with } n_\ell = Ce^\ell,\ m_k = Ce^k.$$

Complexity result

Theorem [HajiAli-N.-Tempone-Wolfers 2017]

Given $\epsilon > 0$, we can chose $L \in \mathbb{N}$ such that

$$\|f - S_L f\|_{L^2_\mu} \leq \epsilon, \qquad \text{with prob.} \geq 1 - C\epsilon^{\log|\log \epsilon|}$$

$$\mathrm{Work}(S_L f) \lesssim \epsilon^{-\lambda} |\log \epsilon|^t \log|\log \epsilon|$$

with

$$\lambda = \begin{cases} \sigma/\alpha_p & \text{if } \gamma/\beta_s \leq \sigma/\alpha_p \\ \gamma/\beta_s & \text{if } \gamma/\beta_s > \sigma/\alpha_p \end{cases}$$

$$t = \begin{cases} 2 & \text{if } \gamma/\beta_s < \sigma/\alpha_p \\ 3 + \sigma/\alpha_p & \text{if } \gamma/\beta_s = \sigma/\alpha_p \\ 1 & \text{if } \gamma/\beta_s > \sigma/\alpha_p \end{cases}$$

Analogous result holds in expectation with α_p replaced by α_e.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Complexity result

Theorem [HajiAli-N.-Tempone-Wolfers 2017]

Given $\epsilon > 0$, we can chose $L \in \mathbb{N}$ such that

$$\|f - S_L f\|_{L^2_\mu} \leq \epsilon, \qquad \text{with prob.} \geq 1 - C\epsilon^{\log|\log\epsilon|}$$

$$\mathrm{Work}(S_L f) \lesssim \epsilon^{-\lambda} |\log\epsilon|^t \log|\log\epsilon|$$

with

$$\lambda = \begin{cases} \sigma/\alpha_p & \text{if } \gamma/\beta_s \leq \sigma/\alpha_p \\ \gamma/\beta_s & \text{if } \gamma/\beta_s > \sigma/\alpha_p \end{cases}$$

$$t = \begin{cases} 2 & \text{if } \gamma/\beta_s < \sigma/\alpha_p \\ 3 + \sigma/\alpha_p & \text{if } \gamma/\beta_s = \sigma/\alpha_p \\ 1 & \text{if } \gamma/\beta_s > \sigma/\alpha_p \end{cases}$$

Analogous result holds in expectation with α_p replaced by α_e.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Sketch of the proof

- **Bound on M_k**: use that $\sqrt{M_k} \leq \frac{M_k}{\log M_k} \leq \frac{2m_k^\sigma}{\kappa_L}$ and $\kappa_L \approx 1/(L+1)$

$$M_k \leq \frac{2}{\kappa_L} m_k^\sigma \log M_k \lesssim (L+1) e^{\frac{k\sigma}{\sigma+\alpha_p}}$$
$$\lesssim (L+1)\log(L+1) e^{\frac{k\sigma}{\sigma+\alpha_p}} (k+1)$$

- **Bound on total work:**

$$\mathrm{Work}(S_L f) \lesssim \sum_{\ell=0}^{L} M_{L-\ell} n_\ell^\gamma$$
$$\lesssim (L+1)\log(L+1) e^{\frac{L\sigma}{\sigma-\alpha_p}} \sum_{\ell=0}^{L} \exp\left\{-l\left(\frac{\sigma}{\sigma-\alpha_p} - \frac{\gamma}{\gamma+\beta_s}\right)\right\}(L-\ell+1)$$

hence, distinguish three cases $\gamma/\beta_s <, =, > \sigma/\alpha_p$

Sketch of the proof

- Bound on M_k: use that $\sqrt{M_k} \leq \frac{M_k}{\log M_k} \leq \frac{2m_k^\sigma}{\kappa_L}$ and $\kappa_L \approx 1/(L+1)$

$$M_k \leq \frac{2}{\kappa_L} m_k^\sigma \log M_k \lesssim (L+1) e^{\frac{k\sigma}{\sigma+\alpha_p}}$$
$$\lesssim (L+1)\log(L+1) e^{\frac{k\sigma}{\sigma+\alpha_p}} (k+1)$$

- Bound on total work:

$$\mathrm{Work}(S_L f) \lesssim \sum_{\ell=0}^{L} M_{L-\ell} n_\ell^\gamma$$

$$\lesssim (L+1)\log(L+1) e^{\frac{L\sigma}{\sigma-\alpha_p}} \sum_{\ell=0}^{L} \exp\left\{-l\left(\frac{\sigma}{\sigma-\alpha_p} - \frac{\gamma}{\gamma+\beta_s}\right)\right\} (L-\ell+1)$$

hence, distinguish three cases $\gamma/\beta_s <,=,> \sigma/\alpha_p$

Sketch of the proof

- Bound on the error in probability:

$$\begin{aligned}\|f - S_L f\|_{L^2_\mu} &= \|f - f_L + \sum_{\ell=0}^{L}(Id - \hat{\Pi}_{M_{L-\ell}})(f_\ell - f_{\ell-1})\|_{L^2_\mu} \\ &\leq \|f - f_L\|_{L^2_\mu} + \sum_{\ell=0}^{L} \|Id - \hat{\Pi}_{M_{L-\ell}}\|_{F \to L^2_\mu} \|f_\ell - f_{\ell-1}\|_F \\ &\lesssim e^{-\frac{L\beta_w}{\gamma+\beta_s}} + e^{-\frac{L\alpha}{\sigma+\alpha}} \sum_{\ell=0}^{L} \exp\left\{\ell\left(\frac{\alpha}{\sigma+\alpha_p} - \frac{\beta_s}{\gamma+\beta_s}\right)\right\}\end{aligned}$$

Again split the three cases $\gamma/\beta_s <, =, > \sigma/\alpha_p$ and notice that the first term $e^{-\frac{L\beta_w}{\gamma+\beta_s}}$ is always negligible as $\beta_w > \beta_s$.

Improved complexity in the case $\gamma/\beta_s > \sigma/\alpha$

In the case $\gamma/\beta_s > \sigma/\alpha$ and $\beta_w > \beta_s$ the complexity can be improved by taking

$$m_k = C \exp\left\{ \frac{k}{\sigma + \alpha_p} + \frac{L(\beta_w - \beta_s)}{\alpha(\gamma + \beta_s)} \right\}$$

In this case the complexity result becomes

$$\|f - S_L f\|_{L^2_\mu} \leq \epsilon, \qquad \text{with prob.} \geq 1 - C\epsilon^{\log|\log \epsilon|}$$

$$\mathrm{Work}(S_L f) \lesssim \epsilon^{-\lambda} |\log \epsilon|^t \log|\log \epsilon|$$

with $t = 1$ and

$$\lambda = \frac{\gamma}{\beta_w} + \left(1 - \frac{\beta_s}{\beta_w}\right)\frac{\sigma}{\alpha_p}$$

which always improves the single level rate $\lambda_{SL} = \frac{\gamma}{\beta_w} + \frac{\sigma}{\alpha_p}$.

Improved complexity in the case $\gamma/\beta_s > \sigma/\alpha$

In the case $\gamma/\beta_s > \sigma/\alpha$ and $\beta_w > \beta_s$ the complexity can be improved by taking

$$m_k = C \exp\left\{ \frac{k}{\sigma + \alpha_p} + \frac{L(\beta_w - \beta_s)}{\alpha(\gamma + \beta_s)} \right\}$$

In this case the complexity result becomes

$$\|f - S_L f\|_{L^2_\mu} \leq \epsilon, \qquad \text{with prob.} \geq 1 - C\epsilon^{\log|\log \epsilon|}$$

$$\mathrm{Work}(S_L f) \lesssim \epsilon^{-\lambda} |\log \epsilon|^t \log|\log \epsilon|$$

with $t = 1$ and

$$\lambda = \frac{\gamma}{\beta_w} + \left(1 - \frac{\beta_s}{\beta_w}\right) \frac{\sigma}{\alpha_p}$$

which always improves the single level rate $\lambda_{SL} = \frac{\gamma}{\beta_w} + \frac{\sigma}{\alpha_p}$.

Outline

Application to random elliptic PDEs

Consider

$$\begin{cases} -\operatorname{div}(a(\mathbf{y})\nabla u(\mathbf{y})) = g, & \text{in } D \subset \mathbb{R}^d \\ u(\mathbf{y}) = 0, & \text{on } \partial D \end{cases}$$

with $\mathbf{y} \in \Gamma = [-1, 1]^N$ and Q linear bounded functional in $L^2(D)$ (e.g. $Q(u) = \int_D u$).

Goal: approximate $f(\mathbf{y}) = Q(u(\mathbf{y}))$.

Assumptions:

- $0 > a_{min} \leq a(\mathbf{x}, \mathbf{y}) \leq a_{max}, \quad \forall (\mathbf{x}, \mathbf{y}) \in D \times \Gamma$.
- g and D sufficiently smooth.

Application to random elliptic PDEs

Consider

$$\begin{cases} -\operatorname{div}(a(\mathbf{y})\nabla u(\mathbf{y})) = g, & \text{in } D \subset \mathbb{R}^d \\ u(\mathbf{y}) = 0, & \text{on } \partial D \end{cases}$$

with $\mathbf{y} \in \Gamma = [-1, 1]^N$ and Q linear bounded functional in $L^2(D)$ (e.g. $Q(u) = \int_D u$).

Goal: approximate $f(\mathbf{y}) = Q(u(\mathbf{y}))$.

Assumptions:

- $0 > a_{min} \leq a(\mathbf{x}, \mathbf{y}) \leq a_{max}, \quad \forall (\mathbf{x}, \mathbf{y}) \in D \times \Gamma$.
- g and D sufficiently smooth.

Application to random elliptic PDEs

Consider

$$\begin{cases} -\operatorname{div}(a(\mathbf{y})\nabla u(\mathbf{y})) = g, & \text{in } D \subset \mathbb{R}^d \\ u(\mathbf{y}) = 0, & \text{on } \partial D \end{cases}$$

with $\mathbf{y} \in \Gamma = [-1,1]^N$ and Q linear bounded functional in $L^2(D)$ (e.g. $Q(u) = \int_D u$).

Goal: approximate $f(\mathbf{y}) = Q(u(\mathbf{y}))$.

Assumptions:

- $0 > a_{min} \leq a(\mathbf{x}, \mathbf{y}) \leq a_{max}, \quad \forall (\mathbf{x}, \mathbf{y}) \in D \times \Gamma$.
- g and D sufficiently smooth.

Application to random elliptic PDEs

Proposition

Let u_n be a finite element approximation of order $r \geq 1$ with maximal element diameter $h = n^{-1}$ and $f_n(\mathbf{y}) = Q(u_n(\mathbf{y}))$.

- If $a \in C^r(D \times \Gamma)$, then

$$\|f - f_n\|_{L^2_\mu(\Gamma)} \lesssim h^{r+1}, \qquad \|f - f_n\|_{C^{r-1}(\Gamma)} \lesssim h^2$$

- If $a \in C^{r,s}(D \times \Gamma) = \{v : D \times \Gamma \to \mathbb{R} : \|\partial_x^{\mathbf{r}} \partial_y^{\mathbf{s}} v\|_{C^0(D\times\Gamma)} < \infty,$
  $\forall |\mathbf{r}|_1 \leq r,\ |\mathbf{s}|_1 \leq s\}$, then

$$\|f - f_n\|_{C^p(\Gamma)} \lesssim h^{r+1}, \qquad \forall p = 0, \dots, s.$$

ML least squares complexity – mixed regularity

Consider the coefficient

$$a(\mathbf{x}, \mathbf{y}) = 1 + \|\mathbf{x}\|_2^r + \|\mathbf{y}\|_2^s \in C^{r-1,1}(D) \otimes C^{s-1,1}(\Gamma)$$

- smoother space: $F = C^{s-1,1}(\Gamma)$
- Spacial approximation: continuous finite elements of degree r
  - error: $\|f - f_n\|_{L^2_\mu} = O(n^{-(r+1)}) = \|f - f_n\|_{C^{s-1,1}} \implies \beta_w = \beta_s = r + 1$
  - cost: $\mathrm{Work}(f_n) = n^d$ with optimal solver $\implies \gamma = d$
- Polynomial approximation: $V_{\Lambda_m} = \mathbb{P}_m$= polynomial space of total degree m
  - error: $\|f - \Pi_{\mathbb{P}_m} f\|_{L^\infty} = O(m^{-s}), \implies \alpha_p = s$
  - cost: $\dim(V_{\Lambda_m}) = \binom{m+N}{N} \lesssim m^N, \implies \sigma = s$

ML least squares complexity – mixed regularity

- Complexity of Single Level

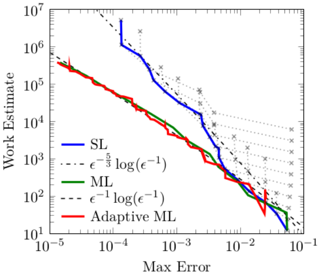
$$\text{Work}_{SL} = \mathcal{O}\left(\epsilon^{-\frac{d}{r+1}-\frac{N}{s}} \log \epsilon^{-1}\right)$$

- Complexity of Multi Level

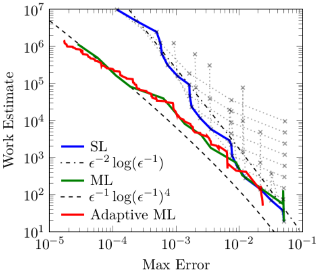
$$\text{Work}_{ML} = \mathcal{O}\left(\epsilon^{-\max\left\{\frac{d}{r+1},\frac{N}{s}\right\}} (\log \epsilon^{-1})^{t}\right)$$

with

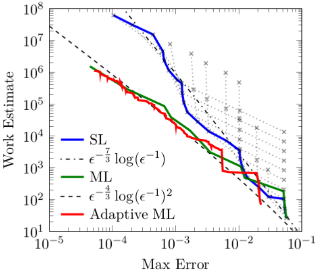
$$t = \begin{cases} 1, & \text{if } \frac{d}{r+1} > \frac{N}{s}, \\ 3 + \frac{d}{r+1}, & \text{if } \frac{d}{r+1} = \frac{N}{s}, \\ 2, & \text{if } \frac{d}{r+1} < \frac{N}{s} \end{cases}$$


Work Estimate
Max Error
SL
$\epsilon^{-\frac{5}{3}}\log(\epsilon^{-1})$
ML
$\epsilon^{-1}\log(\epsilon^{-1})$
Adaptive ML

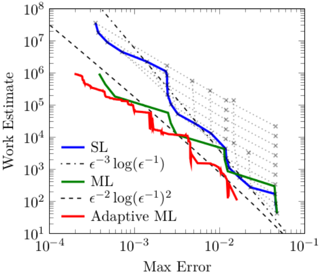
$N = 2$


Work Estimate
Max Error
SL
$\epsilon^{-2}\log(\epsilon^{-1})$
ML
$\epsilon^{-1}\log(\epsilon^{-1})^4$
Adaptive ML

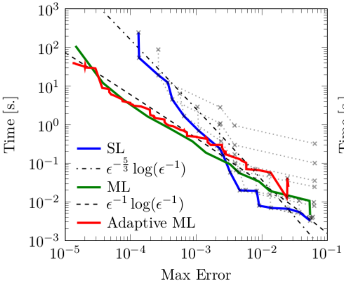
$N = 3$


Work Estimate
Max Error
SL
$\epsilon^{-\frac{7}{3}}\log(\epsilon^{-1})$
ML
$\epsilon^{-\frac{4}{3}}\log(\epsilon^{-1})^2$
Adaptive ML

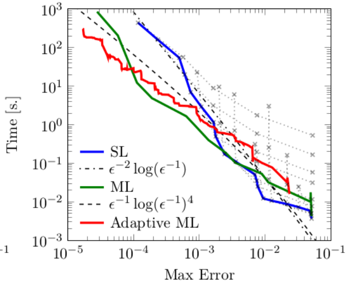
$N = 4$


Work Estimate
Max Error
SL
$\epsilon^{-3}\log(\epsilon^{-1})$
ML
$\epsilon^{-2}\log(\epsilon^{-1})^2$
Adaptive ML

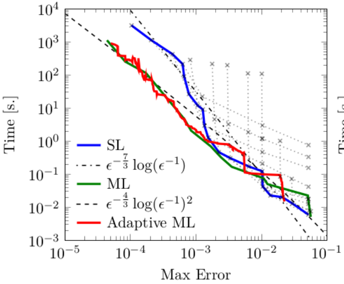
$N = 6$


Time [s.]
Max Error
SL
$\epsilon^{-\frac{5}{3}}\log(\epsilon^{-1})$
ML
$\epsilon^{-1}\log(\epsilon^{-1})$
Adaptive ML

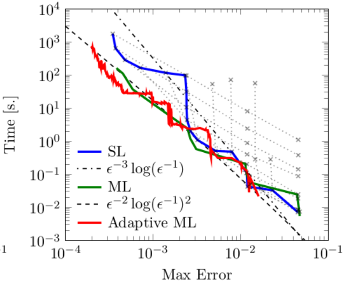
$N = 2$


Time [s.]
Max Error
SL
$\epsilon^{-2}\log(\epsilon^{-1})$
ML
$\epsilon^{-1}\log(\epsilon^{-1})^4$
Adaptive ML

$N = 3$


Time [s.]
Max Error
SL
$\epsilon^{-\frac{7}{3}}\log(\epsilon^{-1})$
ML
$\epsilon^{-\frac{4}{3}}\log(\epsilon^{-1})^2$
Adaptive ML

$N = 4$


Time [s.]
Max Error
SL
$\epsilon^{-3}\log(\epsilon^{-1})$
ML
$\epsilon^{-2}\log(\epsilon^{-1})^2$
Adaptive ML

$N = 6$

Outline

Conclusions

- We have derived a MultiLevel discrete least squares method for polynomial approximation of an output quantity of interest of a random PDE.
- The method uses the classical "Combination technique" and sparsifies sequences of polynomial approximations, obtained by weighted discrete least squares and sequences of spatial discretizations of the underlying PDE.
- In particular, we have proposed a way to select the number of sample points on each level, to guarantee the overall stability and accuracy of the ML formula with high probability.
- Currently working on adaptive algorithms for infinite dimensional problems.

Conclusions

- We have derived a MultiLevel discrete least squares method for polynomial approximation of an output quantity of interest of a random PDE.
- The method uses the classical "Combination technique" and sparsifies sequences of polynomial approximations, obtained by weighted discrete least squares and sequences of spatial discretizations of the underlying PDE.
- In particular, we have proposed a way to select the number of sample points on each level, to guarantee the overall stability and accuracy of the ML formula with high probability.
- Currently working on adaptive algorithms for infinite dimensional problems.

Conclusions

- We have derived a MultiLevel discrete least squares method for polynomial approximation of an output quantity of interest of a random PDE.
- The method uses the classical "Combination technique" and sparsifies sequences of polynomial approximations, obtained by weighted discrete least squares and sequences of spatial discretizations of the underlying PDE.
- In particular, we have proposed a way to select the number of sample points on each level, to guarantee the overall stability and accuracy of the ML formula with high probability.
- Currently working on adaptive algorithms for infinite dimensional problems.

Conclusions

- We have derived a MultiLevel discrete least squares method for polynomial approximation of an output quantity of interest of a random PDE.
- The method uses the classical “Combination technique” and sparsifies sequences of polynomial approximations, obtained by weighted discrete least squares and sequences of spatial discretizations of the underlying PDE.
- In particular, we have proposed a way to select the number of sample points on each level, to guarantee the overall stability and accuracy of the ML formula with high probability.
- Currently working on adaptive algorithms for infinite dimensional problems.

Thank you for your attention!

References

A-L. Haji Ali, F. Nobile, R. Tempone, S. Wolfers,
Multilevel weighted least squares polynomial approximation, arXiv:1707.00026.

A.-L. Haji-Ali, F. Nobile, L. Tamellini and R. Tempone.
Multi-Index Stochastic Collocation convergence rates for random PDEs with parametric regularity, FoCM 16(2016) 1555-1605.

A.-L. Haji-Ali, F. Nobile, L. Tamellini and R. Tempone.
Multi-Index Stochastic Collocation for random PDEs, CMAME 306(2016) 95–122.

A-L. Haji-Ali, F. Nobile, R. Tempone,
Multi index Monte Carlo: when sparsity meets sampling, Numer. Math. 132(2016) 767–806, published online.

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, ESAIM: M2AN, vol. 49, num. 3, p. 815-837, 2015.