Multilevel Polynomial Least Squares Approximation

Abdul-Lateef Haji-Ali[1], Fabio Nobile[2], Raúl Tempone[3], Sören Wolfers[3]

1 University of Oxford

2 École Polytechnique Fédérale de Lausanne

3 جامعة الملك عبدالله للعلوم والتقنية King Abdullah University of Science and Technology

MCQMC

July 4, 2018

MOTIATION: RANDOM PDE / UQ

- Consider PDE with uncertain parameter, e.g.:

$$\begin{cases} -\operatorname{div}(a_y \nabla u_y) = g & \text{in } U \\ u_y = 0 & \text{on } \partial U \end{cases} \qquad \forall y \in \Gamma \subset \mathbb{R}^d$$

- Approximate dependence of scalar quantity of interest on y

$$\begin{aligned} f : \Gamma &\to \mathbb{R} \\ y &\mapsto F(u_y) \end{aligned}$$

- Use approximation for:
  - Statistics (compute $\mathbb{E} f$, $\mathbb{V} f$,..., prob.density $p_F : \mathbb{R} \to \mathbb{R}_+$)
  - Performance guarantees (find $\max f$, $\min f$)
  - Inverse problems (with f being the likelihood or posteriori density)

Outline

- **Goal:** Expand f in $L^2_\mu(\Gamma)$-orthonormal basis:

$$f(y) = \sum_m c_m \phi_m(y)$$

OUTLINE

- **Goal:** Expand f in $L^2_\mu(\Gamma)$-orthonormal basis:

$$f(y) = \sum_m c_m \phi_m(y)$$

- MCQMC $\Rightarrow$ evaluate f in **random points** $(y_n)_{n=1}^N$
  - Could use Monte Carlo approximation of coefficients:

$$c_m \approx \frac{1}{N} \sum_{n=1}^{N} f(y_n)\phi_m(y_i)$$

Outline

- **Goal:** Expand f in $L^2_\mu(\Gamma)$-orthonormal basis:

$$f(y) = \sum_m c_m \phi_m(y)$$

- MCQMC $\Rightarrow$ evaluate f in **random points** $(y_n)_{n=1}^N$
  - Could use Monte Carlo approximation of coefficients
  - But will do something better: **Least squares**

Outline

- **Goal:** Expand f in $L^2_\mu(\Gamma)$-orthonormal basis:

$$f(y) = \sum_m c_m \phi_m(y)$$

- MCQMC $\Rightarrow$ evaluate f in **random points** $(y_n)_{n=1}^N$
  - Could use Monte Carlo approximation of coefficients
  - But will do something better: **Least squares**
- Hierarchical session $\Rightarrow$ evaluate f at different accuracies
  - Will use “Multilevel” as in Multilevel Monte Carlo

Outline

- **Goal:** Expand f in $L^2_\mu(\Gamma)$-orthonormal basis:

$$f(y) = \sum_m c_m \phi_m(y)$$

- MCQMC $\Rightarrow$ evaluate f in **random points** $(y_n)_{n=1}^N$
  - Could use Monte Carlo approximation of coefficients
  - But will do something better: **Least squares**
- Hierarchical session $\Rightarrow$ evaluate f at different accuracies
  - Will use "Multilevel" as in Multilevel Monte Carlo
- **Note:** y is random, but $f(y)$ for fixed y is not
  - Monte Carlo rates can be beaten if f is smooth enough

Naive Monte Carlo

Least squares

Multilevel least squares

Numerics

DEFINITION

- If $(\phi_m)_{m=1}^{\infty}$ is orthonormal basis of $L^2_\mu(\Gamma)$ and
$$f = \sum_m c_m \phi_m \in L^2_\mu(\Gamma),$$
then
$$c_m = \int_\Gamma \phi_m(y) f(y) \mu(\mathrm{d}y).$$

DEFINITION

- If $(\phi_m)_{m=1}^{\infty}$ is orthonormal basis of $L^2_\mu(\Gamma)$ and
$$f = \sum_m c_m \phi_m \in L^2_\mu(\Gamma),$$
then
$$c_m = \int_\Gamma \phi_m(y) f(y) \mu(\mathrm{d}y).$$
- For any $M < \infty$, let
$$\Pi^{\mathrm{MC}} f := \Pi^{\mathrm{MC}}_{M,(y_n)_{n=1}^N} f := \sum_{m=1}^{M} c_m^{\mathrm{MC}} \phi_m,$$
with
$$c_m^{\mathrm{MC}} := \frac{1}{N} \sum_{n=1}^{N} \phi_m(y_n) f(y_n)$$
where $(y_n)_{n=1}^N$ are sampled i.i.d. from μ

ACCURACY

- By orthonormality:

$$\|f - \Pi^{\mathrm{MC}} f\|_{L^2_\mu}^2 = \sum_{m=1}^{M} (c_m - c_m^{\mathrm{MC}})^2 + \sum_{m=M+1}^{\infty} c_m^2$$

- MC estimates converge at rate $N^{-1/2}$ as $N \to \infty$
- Joint convergence deteriorates as $M \to \infty$
- ► Need $\boldsymbol{\mathcal{O}(\varepsilon^{-2-\alpha})}$ evaluations for accuracy $\varepsilon > 0$
  (with $\alpha > 0$ depending on smoothness of f)

NAIVE MONTE CARLO

LEAST SQUARES

MULTILEVEL LEAST SQUARES

NUMERICS

MOTIVATION

- The empirical average

$$\frac{1}{N}\sum_{n=1}^{N} f(y_n)$$

solves discrete least squares minimization problem:

$$\arg\min_{c\in\mathbb{R}} \left\{ \|f - c\|_N^2 := \frac{1}{N}\sum_{n=1}^{N} |f(y_n) - c|^2 \right\}$$

MOTIVATION

- The empirical average

$$\frac{1}{N}\sum_{n=1}^{N} f(y_n)$$

solves discrete least squares minimization problem:

$$\underset{c\in\mathbb{R}}{\arg\min}\left\{\|f-c\|_N^2 := \frac{1}{N}\sum_{n=1}^{N}|f(y_n)-c|^2\right\}$$

- If $\phi_1 \equiv 1$, empirical average equals coefficient estimate c_1^{MC}

MOTIVATION

- The empirical average

$$\frac{1}{N}\sum_{n=1}^{N} f(y_n)$$

solves discrete least squares minimization problem:

$$\arg\min_{c\in\mathbb{R}} \left\{ \|f-c\|_N^2 := \frac{1}{N}\sum_{n=1}^{N} |f(y_n)-c|^2 \right\}$$

- If $\phi_1 \equiv 1$, empirical average equals coefficient estimate c_1^{MC}
- In general, the normalized MC coefficients

$$\frac{c_m^{\mathrm{MC}}}{\|\phi_m\|_N} = \frac{\frac{1}{N}\sum_{n=1}^{N} f(y_n)\phi_m(y_n)}{\|\phi_m\|_N}$$

solve the discrete least squares problems

$$\arg\min_{c\in\mathbb{R}} \|f-c\phi_m\|_N^2$$

DEFINITION

To find multiple coefficients, why not solve a single optimization problem for all?

DEFINITION

To find multiple coefficients, why not solve a single optimization problem for all?

- For finite-dimensional $V \subset L^2_\mu(\Gamma)$ let

$$\boxed{\Pi_V^{\mathrm{LS}} f := \Pi_{V,(y_n)_{n=1}^N}^{\mathrm{LS}} f := \operatorname*{arg\,min}_{v \in V} \|f - v\|_N^2}$$

- Coefficients of $\Pi_V^{\mathrm{LS}} f$ w.r.t. ONB $(\phi_m)_{m=1}^M$ of V:

$$\mathbf{c}^{\mathrm{LS}} = \mathbf{G}^{-1}\mathbf{c}^{\mathrm{MC}}$$

  with the **random Gramian matrix**

$$\mathbf{G} := (\langle \phi_i, \phi_j \rangle_N)_{i,j=1}^M$$

ACCURACY

Let Π_V^{best} be the $L^2_\mu(\Gamma)$-orthogonal projection onto $V \subset L^2_\mu$

- For $\Pi \in \{\Pi_V^{\text{LS}}, \Pi_V^{\text{MC}}\}$:

$$\underbrace{\|f - \Pi f\|^2_{L^2_\mu(\Gamma)}}_{\text{Total error}} = \underbrace{\|\Pi f - \Pi_V^{\text{best}} f\|^2_{L^2_\mu}}_{\text{Projection error}} + \underbrace{\|f - \Pi_V^{\text{best}} f\|^2_{L^2_\mu}}_{\text{Subspace error}}$$

  - ► No point in higher accuracy of projection than accuracy of best-approximation in subspace V

ACCURACY

Let Π_V^{best} be the $L^2_\mu(\Gamma)$-orthogonal projection onto $V \subset L^2_\mu$

- For $\Pi \in \{\Pi_V^{\text{LS}}, \Pi_V^{\text{MC}}\}$:

$$\underbrace{\|f - \Pi f\|^2_{L^2_\mu(\Gamma)}}_{\text{Total error}} = \underbrace{\|\Pi f - \Pi_V^{\text{best}} f\|^2_{L^2_\mu}}_{\text{Projection error}} + \underbrace{\|f - \Pi_V^{\text{best}} f\|^2_{L^2_\mu}}_{\text{Subspace error}}$$

  - No point in higher accuracy of projection than accuracy of best-approximation in subspace V
- For fixed V, normalization factor $\mathbf{G}^{-1}$ converges to identity at MC rate as $N \to \infty$
  - Projection error of Π_V^{LS} decays asymptotically at Monte Carlo rate, just as Π_V^{MC}

ACCURACY

Let Π_V^{best} be the $L^2_\mu(\Gamma)$-orthogonal projection onto $V \subset L^2_\mu$

- For $\Pi \in \{\Pi_V^{\text{LS}}, \Pi_V^{\text{MC}}\}$:

$$\underbrace{\|f - \Pi f\|^2_{L^2_\mu(\Gamma)}}_{\text{Total error}} = \underbrace{\|\Pi f - \Pi_V^{\text{best}} f\|^2_{L^2_\mu}}_{\text{Projection error}} + \underbrace{\|f - \Pi_V^{\text{best}} f\|^2_{L^2_\mu}}_{\text{Subspace error}}$$

  - No point in higher accuracy of projection than accuracy of best-approximation in subspace V
- For fixed V, normalization factor $\mathbf{G}^{-1}$ converges to identity at MC rate as $N \to \infty$
  - Projection error of Π_V^{LS} decays asymptotically at Monte Carlo rate, just as Π_V^{MC}
- **However: Proj. error of Π_V^{LS} "almost immediately" becomes as small as subsp. error; i.e. subsp. error acts as constant in front of MC rate**
  - Remember that for constant functions MC=LS, and indeed:

$$\text{Var}[f]^{1/2} N^{-1/2} = \|f - \mathbb{E}[f]\|_{L^2_\mu} N^{-1/2}$$

ACCURACY

(Remember: $M = \dim V$)

THEOREM (COHEN ET AL. '13, CHKIFA ET AL. '15)

For many μ there exist θ and C **independent of dimension of Γ** such that $\boldsymbol{N \geq CM^{\theta} \log M}$ suffices to achieve

Projection error $\leq 2 \times$ Subspace error

(Check references for precise statement)

- Uniqueness of least squares approximation requires $N \geq M$, i.e. $\theta \geq 1$
- For example, $\theta = 2$ for $\mu = \mathrm{Unif}(\Gamma = [-1,1]^d)$, but $\theta = \infty$ for normal distributions

PROOF SKETCH

Since Π^{LS} leaves V invariant, we have

$$\underbrace{\|\Pi^{\mathrm{LS}} f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}}_{\text{Projection error}} = \|\Pi^{\mathrm{LS}} f - \Pi^{\mathrm{LS}}\Pi^{\mathrm{best}} f\|_{L^2_\mu}$$

$$\leq \|\Pi^{\mathrm{LS}}\|_{L^2 \to L^2} \|f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}$$

$$\approx \|\mathbf{G}^{-1}\|_{\mathrm{op.}} \underbrace{\|f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}}_{\text{Subspace error}}.$$

PROOF SKETCH

Since Π^{LS} leaves V invariant, we have

$$\begin{aligned}\underbrace{\|\Pi^{\mathrm{LS}} f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}}_{\text{Projection error}} &= \|\Pi^{\mathrm{LS}} f - \Pi^{\mathrm{LS}}\Pi^{\mathrm{best}} f\|_{L^2_\mu} \\ &\le \|\Pi^{\mathrm{LS}}\|_{L^2\to L^2}\|f - \Pi^{\mathrm{best}} f\|_{L^2_\mu} \\ &\approx \|\mathbf{G}^{-1}\|_{\mathrm{op.}} \underbrace{\|f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}}_{\text{Subspace error}}.\end{aligned}$$

By **large deviation theory** for matrices, very quickly very likely

$$\|\mathbf{G}^{-1}\|_{\mathrm{op.}} \le \|\mathbf{I}\|_{\mathrm{op.}} + 1 = 2.$$

PROOF SKETCH

Since Π^{LS} leaves V invariant, we have

$$\underbrace{\|\Pi^{\mathrm{LS}} f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}}_{\text{Projection error}} = \|\Pi^{\mathrm{LS}} f - \Pi^{\mathrm{LS}}\Pi^{\mathrm{best}} f\|_{L^2_\mu}$$
$$\leq \|\Pi^{\mathrm{LS}}\|_{L^2 \to L^2} \|f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}$$
$$\approx \|\mathbf{G}^{-1}\|_{\mathrm{op.}} \underbrace{\|f - \Pi^{\mathrm{best}} f\|_{L^2_\mu}}_{\text{Subspace error}} .$$

By **large deviation theory** for matrices, very quickly very likely

$$\|\mathbf{G}^{-1}\|_{\mathrm{op.}} \leq \|\mathbf{I}\|_{\mathrm{op.}} + 1 = 2.$$

Precise statement involves supremum bounds of entries of $\mathbf{G}^{-1}$ and leads to the condition:

$$N/\log N \geq \sup_{y \in \Gamma} \sum_{m=1}^{M} \phi_m^2(y)$$

R.h.s. can be bounded on case-by-case basis by M^θ $\square$

IMPORTANCE SAMPLING

Draw samples from $\nu := \nu(V, \mu)$ with

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(y) := \frac{1}{M} \sum_{m=1}^{M} \phi_m^2(y)$$

and minimize **weighted** discrete ℓ^2-norm

$$\|f - v\|_N^2 := \frac{1}{N} \sum_{n=1}^{N} \frac{\mathrm{d}\mu}{\mathrm{d}\nu} |f(y_n) - v(y_n)|^2$$

(Weighting ensures $\|\cdot\|_N \approx \|\cdot\|_{L^2_\mu}$)

THEOREM (HAMPTON & DOOSTAN '15, COHEN & MIGLIORATI '16)

With importance sampling, $N \geq CM \log M$ suffices for same accuracy as before

(Check references for precise statement)

IMPORTANCE SAMPLING IN PRACTICE

- To sample from importance sampling distribution ν with

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(y) = \frac{1}{M}\sum_{m=1}^{M} \phi_m^2(y) :$$

(1) Randomly pick $m \in \{1, \dots, M\}$
(2) Sample y from $\phi_m^2 \mathrm{d}\mu$

  - Cost per sample independent of $\dim V = M$

---

see Haji-Ali, Nobile, Tempone, and Wolfers 2017

IMPORTANCE SAMPLING IN PRACTICE

- To sample from importance sampling distribution ν with

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(y) = \frac{1}{M}\sum_{m=1}^{M}\phi_m^2(y):$$

(1) Randomly pick $m \in \{1, \dots, M\}$
(2) Sample y from $\phi_m^2 \mathrm{d}\mu$

  - Cost per sample independent of $\dim V = M$

- If $\Gamma = \prod_{j=1}^{d}\Gamma_j$ and $\mu = \bigotimes_{j=1}^{d}\mu_j$, may use tensor basis

$$\phi_{\boldsymbol{m}} := \phi_{m_1} \otimes \cdots \otimes \phi_{m_d}$$

and **sample components $\phi_{m_j}^2 \mathrm{d}\mu_j$ in Step (2) independently** (using rejection sampling, exploiting bounds from theory of univariate orthogonal polynomials)

  - Cost per sample $\mathcal{O}(d)$

see Haji-Ali, Nobile, Tempone, and Wolfers 2017

BOUNDS ON ν FOR $\Gamma = [-1,1]^d$

For many measures on $[-1,1]^d$, the optimal importance sampling distribution **dominates the uniform distribution and is dominated by the arcsine distribution**, whose Lebesgue density is

$$p_{\sin}(y) := \prod_{j=1}^{d} \frac{1}{\pi\sqrt{(1+y^{(j)})(1-y^{(j)})}},$$

up to factors **independent of V**

see Haji-Ali, Nobile, Tempone, and Wolfers 2017

BOUNDS ON ν FOR $\Gamma = [-1,1]^d$

For many measures on $[-1,1]^d$, the optimal importance sampling distribution **dominates the uniform distribution and is dominated by the arcsine distribution**, whose Lebesgue density is

$$p_{\sin}(y) := \prod_{j=1}^{d} \frac{1}{\pi\sqrt{(1+y^{(j)})(1-y^{(j)})}},$$

up to factors **independent of V**

Upper bound suggests sampling from $p_{\sin}\mathrm{d}\lambda$ instead of ν

- #{Req. samples} increases by factor independent of V, but increasing exponentially w.r.t. d
- Easy sampling: $y^{(j)} := \sin(x^{(j)})$, $x^{(j)} \sim$ Unif$[-\pi,\pi]$
- Independence from V useful for adaptive algorithms

---

see Haji-Ali, Nobile, Tempone, and Wolfers 2017

INEXACT EVALUATIONS

REMEMBER

In our motivating example, evaluation of

$$f(y) = F(u_y)$$

requires solution of a PDE!

- Exact evaluations rarely possible
- Numerical solvers yield $f_h(y)$, $h > 0$ with $f_h(y) \to f(y)$ and increasing cost as $h \to 0$

SINGLE LEVEL APPROACH TO INEXACT EVALUATIONS

Idea: Apply least squares approximation to fixed f_h
Problem: Good approximation requires both a large subspace V ($\Rightarrow$ **many samples**) and small h ($\Rightarrow$ **expensive samples**)

EXAMPLE

If

- there is $\beta > 0$ such that $\|f - f_h\|_{L^2_\mu(\Gamma)} \lesssim h^\beta$
- there is $\alpha > 0$ and a sequence of subspaces $(V_k)_{k=1}^\infty$ such that

$$\|f_h - \Pi^{\text{best}}_{V_k} f_h\|_{L^2_\mu(\Gamma)} \lesssim (\dim V_k)^{-\alpha} \quad \forall h > 0, k \in \mathbb{N}$$

then accuracy $\varepsilon > 0$ requires $h \approx \varepsilon^{1/\beta}$ and $\dim V_k \approx \varepsilon^{-1/\alpha}$

- If evaluations of f_h require the work $\mathcal{O}(h^{-\gamma})$, the total work is

$$h^{-\gamma} \times (\dim V_k)^\theta \log \dim V_k \approx \varepsilon^{-\gamma/\beta - \theta/\alpha} |\log \varepsilon|$$

Naive Monte Carlo

Least squares

Multilevel least squares

Numerics

Definition

In analogy to Multilevel Monte Carlo (Heinrich ’01, Giles ’08,...), let $h_\ell \to 0$ and apply least squares approximation to each term in

$$f = f_{h_0} + \sum_{\ell=1}^{\infty} f_{h_\ell} - f_{h_{\ell-1}}$$

DEFINITION

In analogy to Multilevel Monte Carlo (Heinrich ’01, Giles ’08,...), let $h_\ell \to 0$ and apply least squares approximation to each term in

$$f = f_{h_0} + \sum_{\ell=1}^{\infty} f_{h_\ell} - f_{h_{\ell-1}}$$

- f_{h_0} is of same magnitude as f and is cheap to evaluate
  - Computing many samples makes sense
- $f_{h_\ell} - f_{h_{\ell-1}} \to 0$, and they are increasingly expensive
  - Computing many samples would be a waste

DEFINITION

In analogy to Multilevel Monte Carlo (Heinrich '01, Giles '08,...), let $h_\ell \to 0$ and apply least squares approximation to each term in

$$f = f_{h_0} + \sum_{\ell=1}^{\infty} f_{h_\ell} - f_{h_{\ell-1}}$$

- f_{h_0} is of same magnitude as f and is cheap to evaluate
  - Computing many samples makes sense
- $f_{h_\ell} - f_{h_{\ell-1}} \to 0$, and they are increasingly expensive
  - Computing many samples would be a waste

MULTILEVEL FORMULA

For increasing sequence of subspaces $(V_\ell)_{\ell=1}^{\infty}$, let

$$\Pi_L^{\mathrm{ML}} f := \Pi_{L}^{\mathrm{LS}} f_0 + \sum_{\ell=1}^{L} \Pi_{L-\ell}^{\mathrm{LS}} (f_\ell - f_{\ell-1})$$

$\Big(f_\ell := f_{h_\ell}, \Pi_\ell^{\mathrm{LS}} := \Pi_{V_\ell,(y_n)_{n=1}^{N_l}}^{\mathrm{LS}}$ with $N_\ell \approx (\dim V_l)^\theta \log \dim V_l\Big)$

CONVERGENCE ANALYSIS

THEOREM (HAJI-ALI, NOBILE, TEMPONE, WOLFERS '17)

If there is a normed space $F \subset L^2_\mu(\Gamma)$ and subspaces $(V_k)_{k=1}^\infty$ such that

- $\|f - f_h\|_F \lesssim h^\beta$
- $\|g - \Pi^{\text{best}}_{V_k} g\|_{L^2_\mu} \lesssim \|g\|_F (\dim V_k)^{-\alpha} \quad \forall g \in F, k \in \mathbb{N}$

there exists $L > 0$ and sequences $(V_{k_\ell})_{\ell=0}^L$ and $(h_\ell)_{\ell=0}^L$ such that the corresponding multilevel approximation requires the work

$$\varepsilon^{-\max\{\gamma/\beta,\theta/\alpha\}} |\log \varepsilon|^r$$

(with some $r = r(\alpha, \beta, \gamma, \theta) > 0$ whose value is known) to achieve

$$\mathbb{E}\|\Pi^{\text{ML}}_L f - f\|^2_{L^2_\mu(\Gamma)} \leq \varepsilon^2.$$

Proof sketch – Accuracy estimate

With $f_{-1} := 0$, we obtain

$$\Pi_L^{\mathrm{ML}} f - f_L = \sum_{\ell=0}^{L} (\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id})(f_\ell - f_{\ell-1}).$$

Hence,

$$\|\Pi_L^{\mathrm{ML}} f - f_L\|_{L^2_\mu} \le \sum_{\ell=0}^{L} \|\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id}\,\|_{F \to L^2} \|f_\ell - f_{\ell-1}\|_F.$$

PROOF SKETCH – ACCURACY ESTIMATE

With $f_{-1} := 0$, we obtain

$$\Pi_L^{\mathrm{ML}} f - f_L = \sum_{\ell=0}^{L} (\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id})(f_\ell - f_{\ell-1}).$$

Hence,

$$\|\Pi_L^{\mathrm{ML}} f - f_L\|_{L^2_\mu} \leq \sum_{\ell=0}^{L} \|\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id}\,\|_{F\to L^2} \|f_\ell - f_{\ell-1}\|_F.$$

- By assumption:

$$\|\Pi_\ell^{\mathrm{LS}} - \mathrm{Id}\,\|_{F\to L^2_\mu} \lesssim (\dim V_{k_l})^{-\alpha} \quad \text{and}$$

$$\|f_l - f_{l-1}\|_F \lesssim h_l^{-\beta}$$

- If we assume $\alpha = \beta = 1$ and use $\mathbf{\dim V_{k_\ell} := 2^\ell}$ **and** $\boldsymbol{h_\ell := 2^{-\ell}}$ we obtain $\|\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id}\,\|_{F\to L^2_\mu} \lesssim 2^{-(L-\ell)}$ and

$$\|f_\ell - f_{\ell-1}\|_F \lesssim 2^{-\ell}$$

PROOF SKETCH – ACCURACY ESTIMATE

With $f_{-1} := 0$, we obtain

$$\Pi_L^{\mathrm{ML}} f - f_L = \sum_{\ell=0}^{L} (\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id})(f_\ell - f_{\ell-1}).$$

Hence,

$$\|\Pi_L^{\mathrm{ML}} f - f_L\|_{L^2_\mu} \leq \sum_{\ell=0}^{L} \|\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id}\,\|_{F \to L^2} \|f_\ell - f_{\ell-1}\|_F.$$

- By assumption:

$$\|\Pi_\ell^{\mathrm{LS}} - \mathrm{Id}\,\|_{F \to L^2_\mu} \lesssim (\dim V_{k_l})^{-\alpha} \quad \text{and}$$

$$\|f_l - f_{l-1}\|_F \lesssim h_l^{-\beta}$$

- If we assume $\alpha = \beta = 1$ and use $\boldsymbol{\dim V_{k_\ell} := 2^\ell}$ **and** $\boldsymbol{h_\ell := 2^{-\ell}}$ we obtain $\|\Pi_{L-\ell}^{\mathrm{LS}} - \mathrm{Id}\,\|_{F \to L^2_\mu} \lesssim 2^{-(L-\ell)}$ and

$$\|f_\ell - f_{\ell-1}\|_F \lesssim 2^{-\ell}$$

Therefore,

$$\begin{aligned} \|\Pi_L^{\mathrm{ML}} f - f\|_{L^2_\mu} &\leq \|\Pi_L^{\mathrm{ML}} f - f_L\|_{L^2_\mu} + \|f_L - f\|_{L^2_\mu} \\ &\leq \sum_{\ell=0}^{L} 2^{-(L-\ell)} 2^{-\ell} + 2^{-L} \\ &= 2^{-L}(L+2) \end{aligned}$$

PROOF SKETCH – WORK ESTIMATE

The work at level ℓ is $N_{L-\ell}\,[\text{Work}(f_\ell) + \text{Work}(f_{\ell-1})]$, thus

$$\text{Work}(\Pi_L^{\text{ML}} f) = \sum_{\ell=0}^{L} N_{L-\ell}\,[\text{Work}(f_\ell) + \text{Work}(f_{\ell-1})]$$

PROOF SKETCH – WORK ESTIMATE

The work at level ℓ is $N_{L-\ell}\,[\text{Work}(f_\ell) + \text{Work}(f_{\ell-1})]$, thus

$$\text{Work}(\Pi_L^{\mathrm{ML}} f) = \textstyle\sum_{\ell=0}^{L} N_{L-\ell}\,[\text{Work}(f_\ell) + \text{Work}(f_{\ell-1})]$$

Assuming that also $\gamma = \theta = 1$, we get

$$\begin{aligned} N_{L-\ell} &= \dim V_{L-\ell} \log \dim V_{L-\ell} = 2^{L-\ell}(L-\ell) \\ \text{Work}(f_\ell) &\lesssim h_\ell^{-1} = 2^\ell, \end{aligned}$$

thus

$$\begin{aligned} \text{Work}(\Pi_L^{\mathrm{ML}} f) &\lesssim \textstyle\sum_{\ell=0}^{L} 2^{L-\ell} 2^{\ell} (L-\ell) \\ &\approx 2^L (L+1)^2. \end{aligned}$$

PROOF SKETCH – WORK ESTIMATE

The work at level ℓ is $N_{L-\ell}\,[\text{Work}(f_\ell) + \text{Work}(f_{\ell-1})]$, thus

$$\text{Work}(\Pi_L^{\text{ML}} f) = \sum_{\ell=0}^{L} N_{L-\ell}\,[\text{Work}(f_\ell) + \text{Work}(f_{\ell-1})]$$

Assuming that also $\gamma = \theta = 1$, we get

$$N_{L-\ell} = \dim V_{L-\ell} \log \dim V_{L-\ell} = 2^{L-\ell}(L-\ell)$$
$$\text{Work}(f_\ell) \lesssim h_\ell^{-1} = 2^\ell,$$

thus

$$\begin{aligned}\text{Work}(\Pi_L^{\text{ML}} f) &\lesssim \sum_{\ell=0}^{L} 2^{L-\ell} 2^\ell (L-\ell) \\ &\approx 2^L (L+1)^2.\end{aligned}$$

Denoting the accuracy estimate by $\varepsilon := 2^{-L}(L+2)$ we may conclude that

$$\begin{aligned}\text{Work}(\Pi_L^{\text{ML}}) &\lesssim \varepsilon^{-1} |\log \varepsilon|^3 \\ &= \varepsilon^{-\max\{\gamma/\beta, \theta/\alpha\}} |\log \varepsilon|^3\end{aligned}$$

□

Naive Monte Carlo

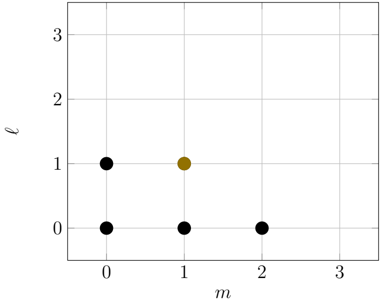
Least squares

Multilevel least squares

Numerics

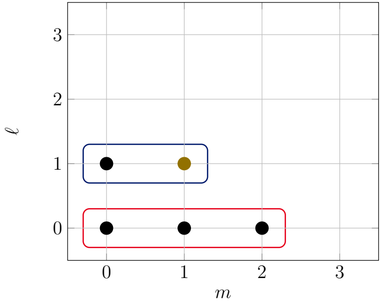
ADAPTIVE MULTILEVEL ALGORITHM – SETUP

- Must choose subspaces to approximate $f_\ell - f_{\ell-1}$ for all ℓ
- Fix basis and include or not each basis function for all ℓ
- **Subset selection in $\mathbb{N}^{d+1}$**. For example:
  - For $d = 1$, the multi-index $(1, 1)$ corresponds to coefficient of ϕ_1 in $f_1 - f_0$, etc.



ADAPTIVE MULTILEVEL ALGORITHM – SETUP

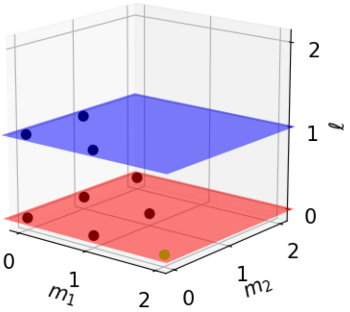
- Must choose subspaces to approximate $f_\ell - f_{\ell-1}$ for all ℓ
- Fix basis and include or not each basis function for all ℓ
- **Subset selection in $\mathbb{N}^{d+1}$**. For example:
  - For $d = 1$, the multi-index $(\mathbf{1}, \mathbf{1})$ corresponds to coefficient of $\phi_{\mathbf{1}}$ in $f_{\mathbf{1}} - f_0$, etc.



$$\Pi^{\mathrm{LS}}_{(\phi_0,\phi_1,\phi_2)} f_0 + \Pi^{\mathrm{LS}}_{(\phi_0,\phi_1)} (f_1 - f_0)$$

ADAPTIVE MULTILEVEL ALGORITHM – SETUP

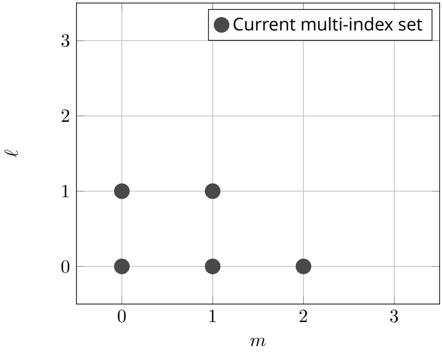
- Must choose subspaces to approximate $f_\ell - f_{\ell-1}$ for all ℓ
- Fix basis and include or not each basis function for all ℓ
- **Subset selection in $\mathbb{N}^{d+1}$**. For example:
  - For $d = 1$, the multi-index $(1, 1)$ corresponds to coefficient of ϕ_1 in $f_1 - f_0$, etc.
  - For $d = 2$, the multi-index $(2, 0, 0)$ corresponds to the coefficient of $\phi_2 \otimes \phi_0$ in f_0, etc.



$$\Pi^{\mathrm{LS}}_{(\phi_{(0,0)},\dots)} f_0 + \Pi^{\mathrm{LS}}_{(\phi_{(0,0)},\phi_{(0,1)},\phi_{(1,0)})}(f_1 - f_0)$$

ADAPTIVE MULTILEVEL ALGORITHM – EXECUTION

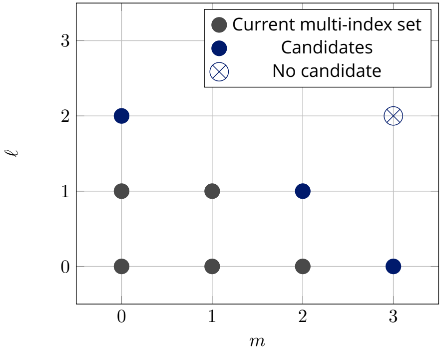
Mimic adaptive sparse grid algorithms (Gerstner'03, Hegland'03,...):



ADAPTIVE MULTILEVEL ALGORITHM – EXECUTION

Mimic adaptive sparse grid algorithms (Gerstner'03, Hegland'03,...):

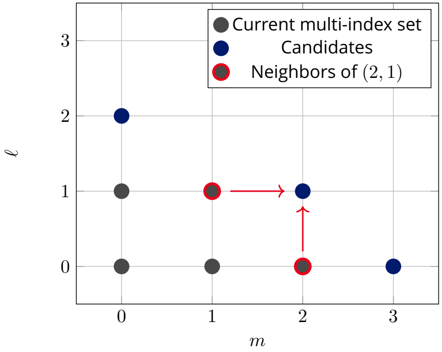
- Consider all neighbors of current set as candidates



ADAPTIVE MULTILEVEL ALGORITHM – EXECUTION

Mimic adaptive sparse grid algorithms (Gerstner'03, Hegland'03,...):

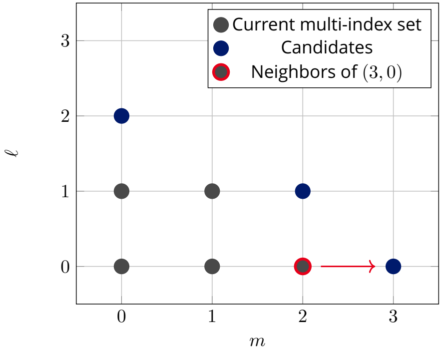
- Consider all neighbors of current set as candidates
- Estimate work and contribution of each candidate



ADAPTIVE MULTILEVEL ALGORITHM – EXECUTION

Mimic adaptive sparse grid algorithms (Gerstner'03, Hegland'03,...):

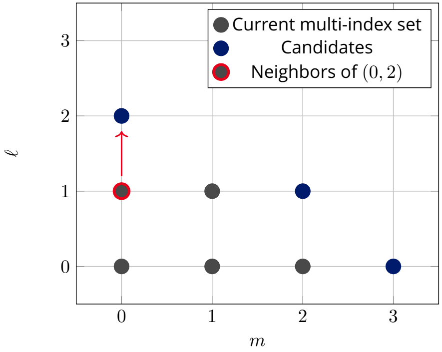
- Consider all neighbors of current set as candidates
- Estimate work and contribution of each candidate



ADAPTIVE MULTILEVEL ALGORITHM – EXECUTION

Mimic adaptive sparse grid algorithms (Gerstner'03, Hegland'03,...):

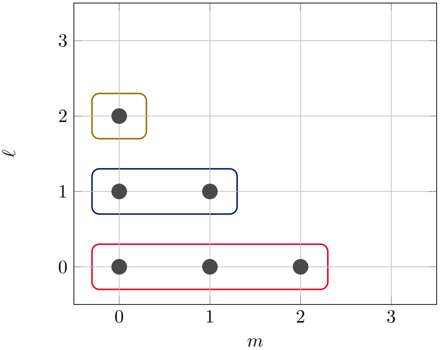
- Consider all neighbors of current set as candidates
- Estimate work and contribution of each candidate



ADAPTIVE MULTILEVEL ALGORITHM – EXECUTION

Mimic adaptive sparse grid algorithms (Gerstner'03, Hegland'03,...):

- Consider all neighbors of current set as candidates
- Estimate work and contribution of each candidate
- Add index that maximizes contribution to work ratio



$$\Pi_{(\phi_0,\phi_1,\phi_2)} f_0 + \Pi_{(\phi_0,\phi_1)}(f_1 - f_0) + \Pi_{(\phi_0)}(f_2 - f_1)$$

MIMCLIB

Multi-**I**ndex **M**onte **C**arlo **lib**rary

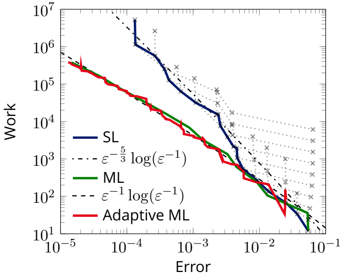
`https://github.com/haji-ali/mimclib`

- Open source
- Python interface, C core
- Provides lots of informative plots
- Supports all kinds of multilevel and multi-index algorithms
- Archives runs in database

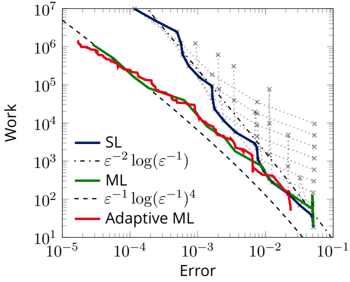
TOY PROBLEM

- Consider
$$\begin{cases} -\operatorname{div}(a_y \nabla u_y) = g & \text{in } U := [-1,1]^2 \\ u_y = 0 & \text{on } \partial U \end{cases}$$
with
$$a_y(x) = 1 + |y|_2^s + |x|_2^r \quad \forall y \in \Gamma := [-1,1]^d,\ x \in U$$
and $r, s \geq 1$ uneven
- Let $y \sim \mathrm{Unif}(\Gamma)$
- **Goal:** Approximate
$$y \mapsto f(y) := \int_U u_y(x)\, dx$$
- Since $a(y,x) := a_y(x) \in C^{s-1,1}(\Gamma) \otimes C^{r-1,1}(U)$ we have
$$\|f - f_h\|_{C^{s-1,1}=:F} \lesssim h^{r+1}$$

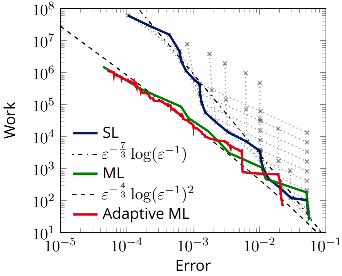
RESULTS (R=1,S=3)



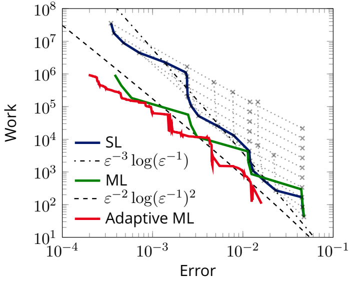
(a) $d = 2$



(b) $d = 3$



(c) $d = 4$



(d) $d = 6$

SUMMARY

- Least squares approximation:
  - Quasi-optimal approximations with d-independent oversampling
  - Importance sampling is easy and effective

Summary

- Least squares approximation:
  - Quasi-optimal approximations with d-independent oversampling
  - Importance sampling is easy and effective
- Multilevel least squares approximation:
  - Get rough idea of all coefficients from cheap f_0;
    adjust important coefficients with few expensive samples
  - Stronger assumption: mixed regularity

    $$\Downarrow$$

    Stronger result: only complexity of PDE solution or of least squares approximation remains