Optimal sampling and reconstruction in high dimension

Albert Cohen

Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris

Münster, 22-06-2019

Motivation : high dimensional parametric PDE's

Partial differential equation $\mathcal{P}(u, y) = 0$ depending on a parameter vector $y \in Y \subset \mathbb{R}^d$ with $d >> 1$ or $d = \infty$.

The parameters may be deterministic (control, optimization, inverse problems) or random distributed according to a probability distribution ρ (uncertainty modeling and quantification, risk assessment, inverse problems).

Simple example : steady state diffusion equation

$$-\mathrm{div}(a \nabla u) = f,$$

on a physical domain D, with homogeneous Dirichlet boundary conditions $u_{|\partial D} = 0$, where $a = a(y)$ is parametrized by y.

Affine model : $a(y) = \bar{a} + \sum_{j \geq 1} y_j \psi_j$, with $y_j \in [-1, 1]$ uniformly distributed.

Lognormal model : $a(y) = \exp(\sum_{j \geq 1} y_j \psi_j)$, with i.i.d. $y_j \sim \mathcal{N}(0, 1)$.

Under suitable assumptions on $\bar{a}$ and $(\psi_j)_{j \geq 1}$ the problem is well posed in the Hilbert space $H_0^1(D)$ (Lax-Milgram) for a.e. $y \in Y$.

Motivation : high dimensional parametric PDE's

Partial differential equation $\mathcal{P}(u, y) = 0$ depending on a parameter vector $y \in Y \subset \mathbb{R}^d$ with $d >> 1$ or $d = \infty$.

The parameters may be deterministic (control, optimization, inverse problems) or random distributed according to a probability distribution ρ (uncertainty modeling and quantification, risk assessment, inverse problems).

Simple example : steady state diffusion equation

$$-\text{div}(a\nabla u) = f,$$

on a physical domain D, with homogeneous Dirichlet boundary conditions $u_{|\partial D} = 0$, where $a = a(y)$ is parametrized by y.

Affine model : $a(y) = \bar{a} + \sum_{j\geq 1} y_j \psi_j$, with $y_j \in [-1, 1]$ uniformly distributed.

Lognormal model : $a(y) = \exp(\sum_{j\geq 1} y_j \psi_j)$, with i.i.d. $y_j \sim \mathcal{N}(0, 1)$.

Under suitable assumptions on $\bar{a}$ and $(\psi_j)_{j\geq 1}$ the problem is well posed in the Hilbert space $H_0^1(D)$ (Lax-Milgram) for a.e. $y \in Y$.

Motivation : high dimensional parametric PDE's

Partial differential equation $\mathcal{P}(u,y)=0$ depending on a parameter vector $y \in Y \subset \mathbb{R}^d$ with $d >> 1$ or $d=\infty$.

The parameters may be deterministic (control, optimization, inverse problems) or random distributed according to a probability distribution ρ (uncertainty modeling and quantification, risk assessment, inverse problems).

Simple example : steady state diffusion equation

$$-\mathrm{div}(a\nabla u) = f,$$

on a physical domain D, with homogeneous Dirichlet boundary conditions $u_{|\partial D} = 0$, where $a = a(y)$ is parametrized by y.

Affine model : $a(y) = \overline{a} + \sum_{j\geq 1} y_j \psi_j$, with $y_j \in [-1,1]$ uniformly distributed.

Lognormal model : $a(y) = \exp(\sum_{j\geq 1} y_j\psi_j)$, with i.i.d. $y_j \sim \mathcal{N}(0,1)$.

Under suitable assumptions on $\overline{a}$ and $(\psi_j)_{j\geq 1}$ the problem is well posed in the Hilbert space $H_0^1(D)$ (Lax-Milgram) for a.e. $y \in Y$.

Motivation : high dimensional parametric PDE's

Partial differential equation $\mathcal{P}(u,y)=0$ depending on a parameter vector $y \in Y \subset \mathbb{R}^d$ with $d >> 1$ or $d = \infty$.

The parameters may be deterministic (control, optimization, inverse problems) or random distributed according to a probability distribution ρ (uncertainty modeling and quantification, risk assessment, inverse problems).

Simple example : steady state diffusion equation

$$-\mathrm{div}(a\nabla u) = f,$$

on a physical domain D, with homogeneous Dirichlet boundary conditions $u_{|\partial D} = 0$, where $a = a(y)$ is parametrized by y.

Affine model : $a(y) = \overline{a} + \sum_{j\geq 1} y_j \psi_j$, with $y_j \in [-1,1]$ uniformly distributed.

Lognormal model : $a(y) = \exp(\sum_{j\geq 1} y_j \psi_j)$, with i.i.d. $y_j \sim \mathcal{N}(0,1)$.

Under suitable assumptions on $\overline{a}$ and $(\psi_j)_{j\geq 1}$ the problem is well posed in the Hilbert space $H_0^1(D)$ (Lax-Milgram) for a.e. $y \in Y$.

Non-intrusive methods

Solution map for a general parametric PDE :

$$y \in Y \mapsto u(y) \in V.$$

For the diffusion equation $V = H_0^1(D)$.

The solution map is difficult to capture numerically (curse of dimensionality).

Objective : reconstruct the solution map, from "snapshots" : particular instances of solutions $u(y^i)$ for $i = 1, \dots, m$ computed by some numerical solver (non-intrusive).

In practice we query $y \mapsto u_h(y) \in V_h$ (finite element space).

Related objectives : numerical approximation of scalar quantities of interest

$$y \mapsto Q(y) = Q(u(y)) \in \mathbb{R}$$

or of averaged quantities $\overline{u} = \mathbb{E}(u(y))$ or $\overline{Q} = \mathbb{E}(Q(y))$.

Non-intrusive methods

Solution map for a general parametric PDE :

$$y \in Y \mapsto u(y) \in V.$$

For the diffusion equation $V = H_0^1(D)$.

The solution map is difficult to capture numerically (curse of dimensionality).

Objective : reconstruct the solution map, from "snapshots" : particular instances of solutions $u(y^i)$ for $i = 1, \dots, m$ computed by some numerical solver (non-intrusive).

In practice we query $y \mapsto u_h(y) \in V_h$ (finite element space).

Related objectives : numerical approximation of scalar quantities of interest

$$y \mapsto Q(y) = Q(u(y)) \in \mathbb{R}$$

or of averaged quantities $\overline{u} = \mathbb{E}(u(y))$ or $\overline{Q} = \mathbb{E}(Q(y))$.

Non-intrusive methods

Solution map for a general parametric PDE :

$$y \in Y \mapsto u(y) \in V.$$

For the diffusion equation $V = H_0^1(D)$.

The solution map is difficult to capture numerically (curse of dimensionality).

Objective : reconstruct the solution map, from “snapshots” : particular instances of solutions $u(y^i)$ for $i = 1, \dots, m$ computed by some numerical solver (non-intrusive).

In practice we query $y \mapsto u_h(y) \in V_h$ (finite element space).

Related objectives : numerical approximation of scalar quantities of interest

$$y \mapsto Q(y) = Q(u(y)) \in \mathbb{R}$$

or of averaged quantities $\overline{u} = \mathbb{E}(u(y))$ or $\overline{Q} = \mathbb{E}(Q(y))$.

Another motivation : reconstruction of acoustic fields (low dimension)

An acoustic pressure field $p(y, t)$ generated by a source is measured by n microphones at positions $y^1, \dots, y^m \in Y \subset \mathbb{R}^2$ or $\mathbb{R}^3$, for $t \in [0, T]$.

Fourier analysis in time $p(y^i, t) \mapsto \hat{p}(y^i, \omega)$ and focus at a frequency ω of interest.

One wants to reconstruct the function $u(y) := \hat{p}(y, \omega)$ on Y, from the observed data $u(y^i)$, $i = 1, \dots, n$.

Another motivation : reconstruction of acoustic fields (low dimension)

An acoustic pressure field $p(y,t)$ generated by a source is measured by n microphones at positions $y^1,\dots,y^m \in Y \subset \mathbb{R}^2$ or $\mathbb{R}^3$, for $t \in [0,T]$.

Fourier analysis in time $p(y^i,t) \mapsto \hat{p}(y^i,\omega)$ and focus at a frequency ω of interest.

One wants to reconstruct the function $u(y) := \hat{p}(y,\omega)$ on Y, from the observed data $u(y^i),\ i = 1,\dots,n$.

General features

Reconstruction of unknown function

$$u : y \in Y \mapsto u(y) \in \mathbb{R} \quad (\text{or } V \text{ or } V_h),$$

from scattered measurements $u^i = u(y^i)$ for $i = 1, \dots, m$ with $y^i \in Y \subset \mathbb{R}^d$.

For notational simplicity we consider scalar valued functions u.

Measurements are costly : one cannot afford to have $m >> 1$.

Measurements could be noisy : $u^i = u(y^i) + \eta_i$.

Analogies with statistical learning :

Non-parametric regression framework : from a random sample $(y^i, u^i)_{i=1,\dots,m}$ with unknown joint density, approximate $y \mapsto u(y)$.

Here active learning : the y^i are chosen by us (deterministically or randomly).

General questions : how should we sample ? how should we reconstruct ?

General features

Reconstruction of unknown function

$$u : y \in Y \mapsto u(y) \in \mathbb{R} \quad \text{(or } V \text{ or } V_h\text{)},$$

from scattered measurements $u^i = u(y^i)$ for $i = 1, \dots, m$ with $y^i \in Y \subset \mathbb{R}^d$.

For notational simplicity we consider scalar valued functions u.

Measurements are costly : one cannot afford to have $m >> 1$.

Measurements could be noisy : $u^i = u(y^i) + \eta_i$.

Analogies with statistical learning :

Non-parametric regression framework : from a random sample $(y^i, u^i)_{i=1,\dots,m}$ with unknown joint density, approximate $y \mapsto u(y)$.

Here active learning : the y^i are chosen by us (deterministically or randomly).

General questions : how should we sample ? how should we reconstruct ?

General features

Reconstruction of unknown function

$$u : y \in Y \mapsto u(y) \in \mathbb{R} \quad (\text{or } V \text{ or } V_h),$$

from scattered measurements $u^i = u(y^i)$ for $i = 1, \dots, m$ with $y^i \in Y \subset \mathbb{R}^d$.

For notational simplicity we consider scalar valued functions u.

Measurements are costly : one cannot afford to have $m >> 1$.

Measurements could be noisy : $u^i = u(y^i) + \eta_i$.

Analogies with statistical learning :

Non-parametric regression framework : from a random sample $(y^i, u^i)_{i=1,\dots,m}$ with unknown joint density, approximate $y \mapsto u(y)$.

Here active learning : the y^i are chosen by us (deterministically or randomly).

General questions : how should we sample ? how should we reconstruct ?

General features

Reconstruction of unknown function

$$u : y \in Y \mapsto u(y) \in \mathbb{R} \quad (\text{or } V \text{ or } V_h),$$

from scattered measurements $u^i = u(y^i)$ for $i = 1, \dots, m$ with $y^i \in Y \subset \mathbb{R}^d$.

For notational simplicity we consider scalar valued functions u.

Measurements are costly : one cannot afford to have $m >> 1$.

Measurements could be noisy : $u^i = u(y^i) + \eta_i$.

Analogies with statistical learning :

Non-parametric regression framework : from a random sample $(y^i, u^i)_{i=1,\dots,m}$ with unknown joint density, approximate $y \mapsto u(y)$.

Here active learning : the y^i are chosen by us (deterministically or randomly).

General questions : how should we sample ? how should we reconstruct ?

General features

Reconstruction of unknown function

$$u : y \in Y \mapsto u(y) \in \mathbb{R} \quad (\text{or } V \text{ or } V_h),$$

from scattered measurements $u^i = u(y^i)$ for $i = 1, \dots, m$ with $y^i \in Y \subset \mathbb{R}^d$.

For notational simplicity we consider scalar valued functions u.

Measurements are costly : one cannot afford to have $m >> 1$.

Measurements could be noisy : $u^i = u(y^i) + \eta_i$.

Analogies with statistical learning :

Non-parametric regression framework : from a random sample $(y^i, u^i)_{i=1,\dots,m}$ with unknown joint density, approximate $y \mapsto u(y)$.

Here active learning : the y^i are chosen by us (deterministically or randomly).

General questions : how should we sample ? how should we reconstruct ?

Approximability prior

The unknown function u is well approximated from some n-dimensional space V_n

$$e_n(u) := \min_{v \in V_n} \|u - v\| \leq \varepsilon(n),$$

where $\varepsilon(n)$ is a known bound and where

$$\|v\| := \|v\|_{L^2(Y,\rho)},$$

with ρ a probability measure on Y.

For certain parametric PDEs, one relevant choice is a sparse polynomial space

$$V_n = \mathbb{P}_{\Lambda_n} = \mathrm{span}\Big\{ y \to y^\nu = \prod_{j\geq 1} y_j^{\nu_j} \; : \; \nu = (\nu_j)_{j\geq 1} \in \Lambda_n \Big\},$$

where Λ_n is an index set such that $\#(\Lambda_n) = n$. Suitable choices of Λ_n obtained by best n-term truncation of $L^2(Y,\rho)$ orthonormal polynomial series provide with rates $\varepsilon(n) \sim n^{-s}$ that persist when $d = \infty$.

Sample result (Bachmayr-Cohen-DeVore-Migliorati 2015) for the affine and lognormal models : if $\sum_{j\geq 1} \kappa_j |\psi_j| < \infty$ with $(\kappa_j^{-1}) \in \ell^q$, then $\varepsilon(n) \sim n^{-s}$ with $s = \frac{1}{q}$.

Approximability prior

The unknown function u is well approximated from some n-dimensional space V_n

$$e_n(u) := \min_{v \in V_n} \|u - v\| \leq \varepsilon(n),$$

where $\varepsilon(n)$ is a known bound and where

$$\|v\| := \|v\|_{L^2(Y,\rho)},$$

with ρ a probability measure on Y.

For certain parametric PDEs, one relevant choice is a sparse polynomial space

$$V_n = \mathbb{P}_{\Lambda_n} = \mathrm{span}\Big\{y \to y^\nu = \prod_{j\geq 1} y_j^{\nu_j} \ : \ \nu = (\nu_j)_{j\geq 1} \in \Lambda_n\Big\},$$

where Λ_n is an index set such that $\#(\Lambda_n) = n$. Suitable choices of Λ_n obtained by best n-term truncation of $L^2(Y,\rho)$ orthonormal polynomial series provide with rates $\varepsilon(n) \sim n^{-s}$ that persist when $d = \infty$.

Sample result (Bachmayr-Cohen-DeVore-Migliorati 2015) for the affine and lognormal models : if $\sum_{j\geq 1} \kappa_j |\psi_j| < \infty$ with $(\kappa_j^{-1}) \in \ell^q$, then $\varepsilon(n) \sim n^{-s}$ with $s = \frac{1}{q}$.

Approximability prior

The unknown function u is well approximated from some n-dimensional space V_n

$$e_n(u) := \min_{v \in V_n} \|u - v\| \leq \varepsilon(n),$$

where $\varepsilon(n)$ is a known bound and where

$$\|v\| := \|v\|_{L^2(Y,\rho)},$$

with ρ a probability measure on Y.

For certain parametric PDEs, one relevant choice is a sparse polynomial space

$$V_n = \mathbb{P}_{\Lambda_n} = \mathrm{span}\Big\{y \to y^\nu = \prod_{j\geq 1} y_j^{\nu_j} \ : \ \nu = (\nu_j)_{j\geq 1} \in \Lambda_n\Big\},$$

where Λ_n is an index set such that $\#(\Lambda_n) = n$. Suitable choices of Λ_n obtained by best n-term truncation of $L^2(Y,\rho)$ orthonormal polynomial series provide with rates $\varepsilon(n) \sim n^{-s}$ that persist when $d = \infty$.

Sample result (Bachmayr-Cohen-DeVore-Migliorati 2015) for the affine and lognormal models : if $\sum_{j\geq 1} \kappa_j |\psi_j| < \infty$ with $(\kappa_j^{-1}) \in \ell^q$, then $\varepsilon(n) \sim n^{-s}$ with $s = \frac{1}{q}$.

Objectives

Use the samples $\{u(y^i) \ : \ i = 1, \dots, m\}$ to reconstruct an approximation $u_n \in V_n$ with certain optimality properties.

Instance optimality : $\|u - u_n\| \leq C e_n(u)$ for any u, for some fixed C.

Rate optimality : if $e_n(u) \leq C_0 n^{-s}$ for all n, then $\|u - u_n\| \leq C_1 n^{-s}$.

Budget optimality : this shoud be achieved with $m \sim n$ samples (up to log factors).

Progressivity : for a given or adaptively selected sequence of space

$$V_0 \subset V_1 \subset \cdots \subset V_n \cdots,$$

these objective should be met at each step with a cumulated sampling budget $\mathcal{O}(n)$ (previous samples should be recycled).

Objectives

Use the samples $\{u(y^i) \ : \ i = 1, \dots, m\}$ to reconstruct an approximation $u_n \in V_n$ with certain optimality properties.

Instance optimality : $\|u - u_n\| \leq Ce_n(u)$ for any u, for some fixed C.

Rate optimality : if $e_n(u) \leq C_0 n^{-s}$ for all n, then $\|u - u_n\| \leq C_1 n^{-s}$.

Budget optimality : this shoud be achieved with $m \sim n$ samples (up to log factors).

Progressivity : for a given or adaptively selected sequence of space

$$V_0 \subset V_1 \subset \cdots \subset V_n \cdots ,$$

these objective should be met at each step with a cumulated sampling budget $\mathcal{O}(n)$ (previous samples should be recycled).

Objectives

Use the samples $\{u(y^i) \ : \ i = 1, \dots, m\}$ to reconstruct an approximation $u_n \in V_n$ with certain optimality properties.

Instance optimality : $\|u - u_n\| \leq Ce_n(u)$ for any u, for some fixed C.

Rate optimality : if $e_n(u) \leq C_0 n^{-s}$ for all n, then $\|u - u_n\| \leq C_1 n^{-s}$.

Budget optimality : this shoud be achieved with $m \sim n$ samples (up to log factors).

Progressivity : for a given or adaptively selected sequence of space

$$V_0 \subset V_1 \subset \cdots \subset V_n \cdots ,$$

these objective should be met at each step with a cumulated sampling budget $\mathcal{O}(n)$ (previous samples should be recycled).

Objectives

Use the samples $\{u(y^i) \ : \ i = 1, \dots, m\}$ to reconstruct an approximation $u_n \in V_n$ with certain optimality properties.

Instance optimality : $\|u - u_n\| \leq C e_n(u)$ for any u, for some fixed C.

Rate optimality : if $e_n(u) \leq C_0 n^{-s}$ for all n, then $\|u - u_n\| \leq C_1 n^{-s}$.

Budget optimality : this shoud be achieved with $m \sim n$ samples (up to log factors).

Progressivity : for a given or adaptively selected sequence of space

$$V_0 \subset V_1 \subset \cdots \subset V_n \cdots,$$

these objective should be met at each step with a cumulated sampling budget $\mathcal{O}(n)$ (previous samples should be recycled).

Approximating the exact projection

The $L^2(Y,\rho)$-projection $P_n u$ of u has the accuracy $e_n(u)$.

It can be either described as

$$P_n u = \operatorname{argmin}\left\{ \int_Y |u(y) - v(y)|^2 d\rho(y) \ : \ v \in V_n \right\},$$

or

$$P_n u = \sum_{j=1}^{n} c_j L_j, \quad c_j := \int_Y u(y) L_j(y) d\rho(y),$$

where $(L_1, \dots, L_n)$ is an $L^2(Y,\rho)$-orthonormal basis of V_n.

Its exact computation is out of reach $\Longrightarrow$ replace the integrals by a discrete sum

$$\int_Y v(y) d\rho(y) \approx \frac{1}{m} \sum_{i=1}^{m} w(y^i) v(y^i).$$

where w is a weight function.

Approximating the exact projection

The $L^2(Y,\rho)$-projection $P_n u$ of u has the accuracy $e_n(u)$.

It can be either described as

$$P_n u = \text{argmin}\Big\{ \int_Y |u(y)-v(y)|^2 d\rho(y) \; : \; v \in V_n \Big\},$$

or

$$P_n u = \sum_{j=1}^{n} c_j L_j, \quad c_j := \int_Y u(y) L_j(y) d\rho(y),$$

where $(L_1,\dots,L_n)$ is an $L^2(Y,\rho)$-orthonormal basis of V_n.

Its exact computation is out of reach $\implies$ replace the integrals by a discrete sum

$$\int_Y v(y) d\rho(y) \approx \frac{1}{m} \sum_{i=1}^{m} w(y^i) v(y^i).$$

where w is a weight function.

Resulting approximation methods

Least-squares method :

$$u_n^{\mathrm{LS}} := \operatorname{argmin}\Big\{ \frac{1}{m} \sum_{i=1}^{m} w(y^i)|u(y^i) - v(y^i)|^2 \; : \; v \in V_n \Big\}.$$

Pseudo-spectral method :

$$u_n^{\mathrm{PS}} := \sum_{j=1}^{n} \tilde{c}_j L_j, \quad \tilde{c}_j := \frac{1}{m} \sum_{i=1}^{m} w(y^i) u(y^i) L_j(y^i).$$

Randomized sampling

Draw $(y^1, \dots, y^m)$ i.i.d. according to a sampling measure $d\sigma$.

Use weight w such that

$$w(y)d\sigma(y) = d\rho(y),$$

and therefore

$$\int_Y v(y)d\rho(y) = \int_Y w(y)v(y)d\sigma(y) = \mathbb{E}\Big(\frac{1}{m}\sum_{i=1}^m w(y^i)v(y^i)\Big).$$

The resulting approximations $u_n^{\rm LS}$ and $u_n^{\rm PS}$ should be compared to u in some probabilistic sense, for instance $\mathbb{E}(\|u - u_n\|^2)$.

Unweighted choice : $w = 1$ and $d\sigma = d\rho$ may lead to suboptimal results.

Optimality can be ensured by an appropriate choice of w and σ.

Randomized sampling

Draw $(y^1, \dots, y^m)$ i.i.d. according to a sampling measure $d\sigma$.

Use weight w such that

$$w(y)d\sigma(y) = d\rho(y),$$

and therefore

$$\int_Y v(y)d\rho(y) = \int_Y w(y)v(y)d\sigma(y) = \mathbb{E}\Big(\frac{1}{m}\sum_{i=1}^m w(y^i)v(y^i)\Big).$$

The resulting approximations $u_n^{\rm LS}$ and $u_n^{\rm PS}$ should be compared to u in some probabilistic sense, for instance $\mathbb{E}(\|u-u_n\|^2)$.

Unweighted choice : $w=1$ and $d\sigma = d\rho$ may lead to suboptimal results.

Optimality can be ensured by an appropriate choice of w and σ.

Randomized sampling

Draw $(y^1, \dots, y^m)$ i.i.d. according to a sampling measure $d\sigma$.

Use weight w such that

$$w(y)d\sigma(y) = d\rho(y),$$

and therefore

$$\int_Y v(y)d\rho(y) = \int_Y w(y)v(y)d\sigma(y) = \mathbb{E}\Big(\frac{1}{m}\sum_{i=1}^m w(y^i)v(y^i)\Big).$$

The resulting approximations u_n^{LS} and u_n^{PS} should be compared to u in some probabilistic sense, for instance $\mathbb{E}(\|u - u_n\|^2)$.

Unweighted choice : $w = 1$ and $d\sigma = d\rho$ may lead to suboptimal results.

Optimality can be ensured by an appropriate choice of w and σ.

Least-squares

The minimization problem is solved by using a given basis $L_1, \dots, L_n$ of V_n and searching

$$u_W = \sum_{j=1}^{n} c_j L_j.$$

The vector $\mathbf{c} = (c_1, \dots, c_n)^t$ is solution to the normal equations

$$\mathbf{G}\mathbf{c} = \mathbf{a},$$

with $\mathbf{G} = (G_{k,j})_{k,j=1,\dots,n}$ and $\mathbf{a} = (a_1, \dots, a_n)^t$, where

$$G_{k,j} := \frac{1}{m} \sum_{i=1}^{m} w(y^i) L_k(y^i) L_j(y^i) \quad \text{and} \quad a_k := \frac{1}{m} \sum_{i=1}^{m} w(y^i) u^i L_k(y^i).$$

The solution always exists and is unique if $\mathbf{G}$ is invertible.

Least-squares

The minimization problem is solved by using a given basis $L_1, \ldots, L_n$ of V_n and searching

$$u_W = \sum_{j=1}^{n} c_j L_j.$$

The vector $\mathbf{c} = (c_1, \ldots, c_n)^t$ is solution to the normal equations

$$\mathbf{G}\mathbf{c} = \mathbf{a},$$

with $\mathbf{G} = (G_{k,j})_{k,j=1,\ldots,n}$ and $\mathbf{a} = (a_1, \ldots, a_n)^t$, where

$$G_{k,j} := \frac{1}{m} \sum_{i=1}^{m} w(y^i) L_k(y^i) L_j(y^i) \quad \text{and} \quad a_k := \frac{1}{m} \sum_{i=1}^{m} w(y^i) u^i L_k(y^i).$$

The solution always exists and is unique if $\mathbf{G}$ is invertible.

Instance optimality

The approximation $u_n^{\rm LS}$ is the orthogonal projection of u onto V_n for the discrete norm

$$\|v\|_m^2 := \frac{1}{m}\sum_{i=1}^m w(y^i)|v(y^i)|^2.$$

Strategy : establish an equivalence with the continuous $L^2(Y,\rho)$ norm over V_n.

Let $(L_1,\dots,L_n)$ be an $L^2(Y,\rho)$-orthonormal basis of V_n so that the random matrix

$$\mathbf{G} = (G_{k,j}) := \Big(\frac{1}{m}\sum_{i=1}^m w(y^i)L_k(y^i)L_j(y^i)\Big),$$

satisfies $\mathbb{E}(\mathbf{G}) = \mathbf{I}$. Then

$$\|\mathbf{G}-\mathbf{I}\| \leq \frac{1}{2} \iff \frac{1}{2}\|v\|^2 \leq \|v\|_m^2 \leq \frac{3}{2}\|v\|^2, \quad v \in V_n,$$

where $\|\mathbf{X}\|$ is the spectral norm of a matrix $\mathbf{X}$.

When this holds one has

$$\|u-u_n^{\rm LS}\|^2 \leq e_n(u)^2+\|P_nu-u_n^{\rm LS}\|^2 \leq e_n(u)^2+2\|P_nu-u_n^{\rm LS}\|_m^2 \leq e_n(u)^2+2\|u-P_nu\|_m^2,$$

and $\mathbb{E}(\|u-P_nu\|_m^2) = e_n(u)^2 \implies$ instance optimality.

The approximation $u_n^{\rm LS}$ is the orthogonal projection of u onto V_n for the discrete norm

$$\|v\|_m^2 := \frac{1}{m}\sum_{i=1}^m w(y^i)|v(y^i)|^2.$$

Strategy : establish an equivalence with the continuous $L^2(Y,\rho)$ norm over V_n.

Let $(L_1,\dots,L_n)$ be an $L^2(Y,\rho)$-orthonormal basis of V_n so that the random matrix

$$\mathbf{G} = (G_{k,j}) := \Big(\frac{1}{m}\sum_{i=1}^m w(y^i)L_k(y^i)L_j(y^i)\Big),$$

satisfies $\mathbb{E}(\mathbf{G}) = \mathbf{I}$. Then

$$\|\mathbf{G}-\mathbf{I}\| \leq \frac{1}{2} \iff \frac{1}{2}\|v\|^2 \leq \|v\|_m^2 \leq \frac{3}{2}\|v\|^2, \quad v\in V_n,$$

where $\|\mathbf{X}\|$ is the spectral norm of a matrix $\mathbf{X}$.

When this holds one has

$$\|u-u_n^{\rm LS}\|^2 \leq e_n(u)^2+\|P_nu-u_n^{\rm LS}\|^2 \leq e_n(u)^2+2\|P_nu-u_n^{\rm LS}\|_m^2 \leq e_n(u)^2+2\|u-P_nu\|_m^2,$$

and $\mathbb{E}(\|u-P_nu\|_m^2) = e_n(u)^2 \implies$ instance optimality.

The approximation $u_n^{\rm LS}$ is the orthogonal projection of u onto V_n for the discrete norm

$$\|v\|_m^2 := \frac{1}{m}\sum_{i=1}^m w(y^i)|v(y^i)|^2.$$

Strategy : establish an equivalence with the continuous $L^2(Y,\rho)$ norm over V_n.

Let $(L_1,\dots,L_n)$ be an $L^2(Y,\rho)$-orthonormal basis of V_n so that the random matrix

$$\mathbf{G}=(G_{k,j}):=\Big(\frac{1}{m}\sum_{i=1}^m w(y^i)L_k(y^i)L_j(y^i)\Big),$$

satisfies $\mathbb{E}(\mathbf{G})=\mathbf{I}$. Then

$$\|\mathbf{G}-\mathbf{I}\|\leq\frac{1}{2}\iff\frac{1}{2}\|v\|^2\leq\|v\|_m^2\leq\frac{3}{2}\|v\|^2,\quad v\in V_n,$$

where $\|\mathbf{X}\|$ is the spectral norm of a matrix $\mathbf{X}$.

When this holds one has

$$\|u-u_n^{\rm LS}\|^2\leq e_n(u)^2+\|P_nu-u_n^{\rm LS}\|^2\leq e_n(u)^2+2\|P_nu-u_n^{\rm LS}\|_m^2\leq e_n(u)^2+2\|u-P_nu\|_m^2,$$

and $\mathbb{E}(\|u-P_nu\|_m^2)=e_n(u)^2\implies$ instance optimality.

The key ingredient to our analysis

Let $L_1, \dots, L_n$ be an orthonormal basis of V_n for the $L^2(Y, \rho)$ norm. We introduce

$$k_{n,w}(y) := w(y) \sum_{j=1}^{n} |L_j(y)|^2,$$

and

$$K_{n,w} := \|k_{n,w}\|_{L^\infty} = \sup_{y \in Y} w(y) \sum_{j=1}^{n} |L_j(y)|^2.$$

Both are independent on the choice orthonormal basis : only depends on (V_n, ρ, w).

Since $\int_Y k_{n,w} d\sigma = \sum_{j=1}^n \int_Y |L_j|^2 d\rho = n$, one has

$$K_{n,w} \geq n.$$

In the case $w = 1$, we obtain the inverse Christoffel function $k_n(y) := \sum_{j=1}^n |L_j(y)|^2$, which is the diagonal of the orthogonal projection kernel onto V_n, and such that

$$K_n := \|k_n\|_{L^\infty} = \max_{v \in V_n} \frac{\|v\|^2_{L^\infty}}{\|v\|^2}.$$

The key ingredient to our analysis

Let $L_1, \dots, L_n$ be an orthonormal basis of V_n for the $L^2(Y,\rho)$ norm. We introduce

$$k_{n,w}(y) := w(y) \sum_{j=1}^{n} |L_j(y)|^2,$$

and

$$K_{n,w} := \|k_{n,w}\|_{L^\infty} = \sup_{y \in Y} w(y) \sum_{j=1}^{n} |L_j(y)|^2.$$

Both are independent on the choice orthonormal basis : only depends on (V_n, ρ, w).

Since $\int_Y k_{n,w} d\sigma = \sum_{j=1}^n \int_Y |L_j|^2 d\rho = n$, one has

$$K_{n,w} \geq n.$$

In the case $w = 1$, we obtain the inverse Christoffel function $k_n(y) := \sum_{j=1}^n |L_j(y)|^2$, which is the diagonal of the orthogonal projection kernel onto V_n, and such that

$$K_n := \|k_n\|_{L^\infty} = \max_{v \in V_n} \frac{\|v\|_{L^\infty}^2}{\|v\|^2}.$$

The key ingredient to our analysis

Let $L_1, \dots, L_n$ be an orthonormal basis of V_n for the $L^2(Y, \rho)$ norm. We introduce

$$k_{n,w}(y) := w(y) \sum_{j=1}^{n} |L_j(y)|^2,$$

and

$$K_{n,w} := \|k_{n,w}\|_{L^\infty} = \sup_{y \in Y} w(y) \sum_{j=1}^{n} |L_j(y)|^2.$$

Both are independent on the choice orthonormal basis : only depends on (V_n, ρ, w).

Since $\int_Y k_{n,w} d\sigma = \sum_{j=1}^{n} \int_Y |L_j|^2 d\rho = n$, one has

$$K_{n,w} \geq n.$$

In the case $w = 1$, we obtain the inverse Christoffel function $k_n(y) := \sum_{j=1}^{n} |L_j(y)|^2$, which is the diagonal of the orthogonal projection kernel onto V_n, and such that

$$K_n := \|k_n\|_{L^\infty} = \max_{v \in V_n} \frac{\|v\|_{L^\infty}^2}{\|v\|^2}.$$

Deviation of $\mathbf{G}$ from $\mathbf{I}$: a concentration bound

Theorem (Cohen-Migliorati 2017, Doostan-Hampton 2015) :

Let $0 < \varepsilon < 1$ be arbitrary. Under the condition

$$m \geq cK_{n,w} \ln(2n/\varepsilon), \quad c := \frac{2}{3\ln(3/2) - 1},$$

one has the deviation bound

$$\Pr\left\{\|\mathbf{G} - \mathbf{I}\| \geq \frac{1}{2}\right\} \leq \varepsilon.$$

We set $u_n^{\rm LS} = 0$ when $\|G - \mathbf{I}\| \geq \frac{1}{2}$, and obtain the instance optimality bound

$$\mathbb{E}(\|u - u_n^{\rm LS}\|^2) \leq 3e_n(u)^2 + \varepsilon\|u\|^2.$$

Typical choice : take $\varepsilon = n^{-r}$ for $r > 0$ larger than the decay rate of $e_n(u)$ if known.

Gives stability condition $m \gtrsim K_{n,w} \ln(n)$, which imposes at least the regime $m \gtrsim n\ln(n)$, but can be much more demanding if $K_{n,w} >> n$.

Deviation of $\mathbf{G}$ from $\mathbf{I}$: a concentration bound

Theorem (Cohen-Migliorati 2017, Doostan-Hampton 2015) :

Let $0 < \varepsilon < 1$ be arbitrary. Under the condition

$$m \geq cK_{n,w} \ln(2n/\varepsilon), \quad c := \frac{2}{3\ln(3/2) - 1},$$

one has the deviation bound

$$\Pr\left\{\|\mathbf{G} - \mathbf{I}\| \geq \frac{1}{2}\right\} \leq \varepsilon.$$

We set $u_n^{\mathrm{LS}} = 0$ when $\|G - \mathbf{I}\| \geq \frac{1}{2}$, and obtain the instance optimality bound

$$\mathbb{E}(\|u - u_n^{\mathrm{LS}}\|^2) \leq 3e_n(u)^2 + \varepsilon\|u\|^2.$$

Typical choice : take $\varepsilon = n^{-r}$ for $r > 0$ larger than the decay rate of $e_n(u)$ if known.

Gives stability condition $m \gtrsim K_{n,w} \ln(n)$, which imposes at least the regime $m \gtrsim n\ln(n)$, but can be much more demanding if $K_{n,w} >> n$.

Deviation of $\mathbf{G}$ from $\mathbf{I}$: a concentration bound

Theorem (Cohen-Migliorati 2017, Doostan-Hampton 2015) :

Let $0 < \varepsilon < 1$ be arbitrary. Under the condition

$$m \geq cK_{n,w} \ln(2n/\varepsilon), \quad c := \frac{2}{3\ln(3/2)-1},$$

one has the deviation bound

$$\Pr\left\{\|\mathbf{G}-\mathbf{I}\| \geq \frac{1}{2}\right\} \leq \varepsilon.$$

We set $u_n^{\rm LS} = 0$ when $\|G - \mathbf{I}\| \geq \frac{1}{2}$, and obtain the instance optimality bound

$$\mathbb{E}(\|u - u_n^{\rm LS}\|^2) \leq 3e_n(u)^2 + \varepsilon\|u\|^2.$$

Typical choice : take $\varepsilon = n^{-r}$ for $r > 0$ larger than the decay rate of $e_n(u)$ if known.

Gives stability condition $m \gtrsim K_{n,w} \ln(n)$, which imposes at least the regime $m \gtrsim n\ln(n)$, but can be much more demanding if $K_{n,w} >> n$.

Where does the stability condition comes from

We may write

$$\mathbf{G} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{X}_i,$$

where $\mathbf{X}_i$ are i.i.d. copies of the $n \times n$ rank one random matrix

$$\mathbf{X} = w(y)(L_k(y)L_j(y))_{j,k=1,\dots,n},$$

with y distributed according to σ, which has expectation $\mathbb{E}(\mathbf{X}) = \mathbf{I}$.

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : if $\|\mathbf{X}\| \le K$ a.s., then

$$\Pr\left\{\left\|\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i - \mathbb{E}(\mathbf{X})\right\| \ge \delta\right\} \le 2n\exp\Big(-\frac{mc(\delta)}{K}\Big),$$

where $c(\delta) := (1+\delta)\ln(1+\delta) - \delta > 0$ (in particular $c(\frac{1}{2}) := c^{-1} = \frac{3\ln(3/2)-1}{2}$).

Here $K = \sup_{y \in Y} w(y)\sum_{j=1}^{n}|L_j(y)|^2 = K_{n,w}$.

Therefore $m \ge cK_{n,w}\ln(2n/\varepsilon) \implies \Pr\{\|G - I\| \ge \frac{1}{2}\} \le \varepsilon$.

Where does the stability condition comes from

We may write

$$\mathbf{G} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{X}_i,$$

where $\mathbf{X}_i$ are i.i.d. copies of the $n \times n$ rank one random matrix

$$\mathbf{X} = w(y)(L_k(y)L_j(y))_{j,k=1,\dots,n},$$

with y distributed according to σ, which has expectation $\mathbb{E}(\mathbf{X}) = \mathbf{I}$.

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : if $\|\mathbf{X}\| \leq K$ a.s., then

$$\Pr\left\{\left\|\frac{1}{m}\sum_{i=1}^{m} \mathbf{X}_i - \mathbb{E}(\mathbf{X})\right\| \geq \delta\right\} \leq 2n \exp\left(-\frac{mc(\delta)}{K}\right),$$

where $c(\delta) := (1+\delta)\ln(1+\delta) - \delta > 0$ (in particular $c(\frac{1}{2}) := c^{-1} = \frac{3\ln(3/2)-1}{2}$).

Here $K = \sup_{y\in Y} w(y)\sum_{j=1}^{n} |L_j(y)|^2 = K_{n,w}$.

Therefore $m \geq cK_{n,w}\ln(2n/\varepsilon) \implies \Pr\{\|G - I\| \geq \frac{1}{2}\} \leq \varepsilon$.

Where does the stability condition comes from

We may write

$$\mathbf{G} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i,$$

where $\mathbf{X}_i$ are i.i.d. copies of the $n \times n$ rank one random matrix

$$\mathbf{X} = w(y)(L_k(y)L_j(y))_{j,k=1,\dots,n},$$

with y distributed according to σ, which has expectation $\mathbb{E}(\mathbf{X}) = \mathbf{I}$.

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : if $\|\mathbf{X}\| \leq K$ a.s., then

$$\Pr\left\{\left\|\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i - \mathbb{E}(\mathbf{X})\right\| \geq \delta\right\} \leq 2n\exp\left(-\frac{mc(\delta)}{K}\right),$$

where $c(\delta) := (1+\delta)\ln(1+\delta) - \delta > 0$ (in particular $c(\frac{1}{2}) := c^{-1} = \frac{3\ln(3/2)-1}{2}$).

Here $K = \sup_{y\in Y} w(y)\sum_{j=1}^{n}|L_j(y)|^2 = K_{n,w}$.

Therefore $m \geq cK_{n,w}\ln(2n/\varepsilon) \implies \Pr\{\|G - I\| \geq \frac{1}{2}\} \leq \varepsilon$.

The unweighted case $w = 1$

The stability regime is described by the condition $m \gtrsim K_n \ln(n)$, with $K_n := \|k_n\|_{L^\infty}$.

We can estimate the inverse Christoffel function $k_n(y) = \sum_{j=1}^n |L_j(y)|^2$ in cases of practical interest.

A simple example : $Y = [-1, 1]$ and $V_n = \mathbb{P}_{n-1}$ the univariate polynomials.

(i) Distribution $\rho = \frac{dy}{\pi\sqrt{1-y^2}}$: the L_j are the Chebychev polynomials and $K_n = 2n + 1$. Up to log factors, the stability regime is $m \gtrsim n$.

(ii) Uniform distribution $\rho = \frac{dy}{2}$: the L_j are normalized Legendre polynomials and $K_n = \sum_{j=1}^n (2j-1) = n^2$. Up to log factors, the stability regime is $m \gtrsim n^2$.

These regimes are confirmed numerically.

The unweighted case $w = 1$

The stability regime is described by the condition $m \gtrsim K_n \ln(n)$, with $K_n := \|k_n\|_{L^\infty}$.

We can estimate the inverse Christoffel function $k_n(y) = \sum_{j=1}^n |L_j(y)|^2$ in cases of practical interest.

A simple example : $Y = [-1, 1]$ and $V_n = \mathbb{P}_{n-1}$ the univariate polynomials.

(i) Distribution $\rho = \frac{dy}{\pi\sqrt{1-y^2}}$: the L_j are the Chebychev polynomials and $K_n = 2n + 1$.
Up to log factors, the stability regime is $m \gtrsim n$.

(ii) Uniform distribution $\rho = \frac{dy}{2}$: the L_j are normalized Legendre polynomials and $K_n = \sum_{j=1}^n (2j - 1) = n^2$. Up to log factors, the stability regime is $m \gtrsim n^2$.

These regimes are confirmed numerically.

The unweighted case $w = 1$

The stability regime is described by the condition $m \gtrsim K_n \ln(n)$, with $K_n := \|k_n\|_{L^\infty}$.

We can estimate the inverse Christoffel function $k_n(y) = \sum_{j=1}^n |L_j(y)|^2$ in cases of practical interest.

A simple example : $Y = [-1, 1]$ and $V_n = \mathbb{P}_{n-1}$ the univariate polynomials.

(i) Distribution $\rho = \frac{dy}{\pi\sqrt{1-y^2}}$: the L_j are the Chebychev polynomials and $K_n = 2n + 1$. Up to log factors, the stability regime is $m \gtrsim n$.

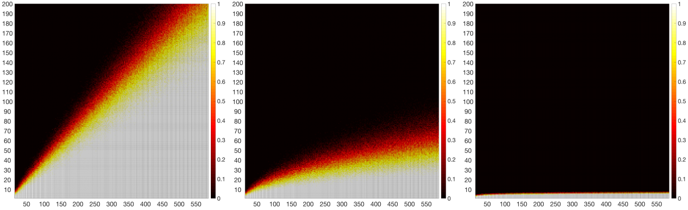
(ii) Uniform distribution $\rho = \frac{dy}{2}$: the L_j are normalized Legendre polynomials and $K_n = \sum_{j=1}^n (2j-1) = n^2$. Up to log factors, the stability regime is $m \gtrsim n^2$.

These regimes are confirmed numerically.

Illustration

Regime of stability : probability that $\kappa(\mathbf{G}) \leq 3$, white if 1, black if 0.

Left for $\rho = \frac{dy}{\pi\sqrt{1-y^2}}$, center : for $\rho = \frac{dy}{2}$ (with m on x axis, n on y axis).



Right : the gaussian case $Y = \mathbb{R}$ and $\rho = g(y)dy$, where $g(y) := \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$, for which the L_j are the Hermite polynomials.

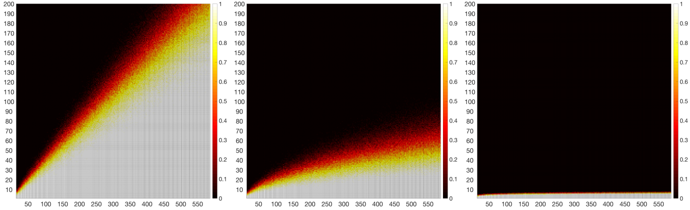
The unweighted theory cannot handle this case since $K_n = \infty$

A more ad-hoc analysis shows that stability holds if $m \gtrsim \exp(cn)$ and this regime is observed numerically.

Illustration

Regime of stability : probability that $\kappa(\mathbf{G}) \leq 3$, white if 1, black if 0.

Left for $\rho = \frac{dy}{\pi\sqrt{1-y^2}}$, center : for $\rho = \frac{dy}{2}$ (with m on x axis, n on y axis).



Right : the gaussian case $Y = \mathbb{R}$ and $\rho = g(y)dy$, where $g(y) := \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$, for which the L_j are the Hermite polynomials.

The unweighted theory cannot handle this case since $K_n = \infty$

A more ad-hoc analysis shows that stability holds if $m \gtrsim \exp(cn)$ and this regime is observed numerically.

Other examples

Local bases : Let V_n be the space of piecewise constant functions over a partition $\mathcal{P}_n$ of Y into n cells. An orthonormal basis is given by the functions $\rho(T)^{-1/2}\chi_T$.

If the partition is uniform with respect to ρ, i.e. $\rho(T) = \frac{1}{n}$ for all $T \in \mathcal{P}_n$, then $K_n = n$.

Trigonometric system : with ρ the uniform measure on a torus, since L_j is the complex exponential, one has $K_n = n$.

Spectral spaces on Riemannian manifolds : let $\mathcal{M}$ be a compact Riemannian manifold without boundary and let V_n be spanned by the n first eigenfunctions L_j of the Laplace-Beltrami operator. Then under mild assumptions (doubling properties and Poincaré inequalities), $K_n = \mathcal{O}(n)$ (estimation based on analysis of the Heat kernel in Dirichlet spaces by Kerkyacharian and Petrushev).

Such spaces are therefore well suited for stable least-squares methods. Example : spherical harmonics. Note that individually the eigenfunctions do not satisfy $\|L_j\|_{L^\infty} = \mathcal{O}(1)$.

Other examples

Local bases : Let V_n be the space of piecewise constant functions over a partition $\mathcal{P}_n$ of Y into n cells. An orthonormal basis is given by the functions $\rho(T)^{-1/2}\chi_T$.

If the partition is uniform with respect to ρ, i.e. $\rho(T) = \frac{1}{n}$ for all $T \in \mathcal{P}_n$, then $K_n = n$.

Trigonometric system : with ρ the uniform measure on a torus, since L_j is the complex exponential, one has $K_n = n$.

Spectral spaces on Riemannian manifolds : let $\mathcal{M}$ be a compact Riemannian manifold without boundary and let V_n be spanned by the n first eigenfunctions L_j of the Laplace-Beltrami operator. Then under mild assumptions (doubling properties and Poincaré inequalities), $K_n = \mathcal{O}(n)$ (estimation based on analysis of the Heat kernel in Dirichlet spaces by Kerkyacharian and Petrushev).

Such spaces are therefore well suited for stable least-squares methods. Example : spherical harmonics. Note that individually the eigenfunctions do not satisfy $\|L_j\|_{L^\infty} = \mathcal{O}(1)$.

High dimensions : parametric PDE's

Prototype example : elliptic PDE's on some domain $D \subset \mathbb{R}^2$ or $\mathbb{R}^3$ with affine parametrization of the diffusion function by $y = (y_1, \dots, y_d) \in Y = [-1, 1]^d$

$$-\mathrm{div}(a \nabla u) = f, \quad a = \bar{a} + \sum_{j=1}^{d} y_j \psi_j,$$

with ellipticity assumption $0 < r < a < R$ for all $y \in Y$, so $y \mapsto u(y) \in V = H_0^1(D)$.

With $\Lambda \subset \mathbb{N}^d$, approximation by multivariate polynomial space

$$V_\Lambda := \left\{ \sum_{\nu \in \Lambda} v_\nu y^\nu, \ \ v_\nu \in V \right\} = V \otimes \mathbb{P}_\Lambda,$$

where $y^\nu = y_1^{\nu_1} \cdots y_d^{\nu_d}$.

We consider downward closed index sets : $\nu \in \Lambda$ and $\mu \leq \nu \Rightarrow \mu \in \Lambda$.

Basis of $\mathbb{P}_\Lambda$: tensorized orthogonal polynomials $L_\nu(y) = \prod_{j=1}^d L_{\nu_j}(y_j)$ for $\nu \in \Lambda$.

High dimensions : parametric PDE's

Prototype example : elliptic PDE's on some domain $D \subset \mathbb{R}^2$ or $\mathbb{R}^3$ with affine parametrization of the diffusion function by $y = (y_1, \dots, y_d) \in Y = [-1,1]^d$

$$-\mathrm{div}(a\nabla u) = f, \quad a = \bar{a} + \sum_{j=1}^{d} y_j \psi_j,$$

with ellipticity assumption $0 < r < a < R$ for all $y \in Y$, so $y \mapsto u(y) \in V = H_0^1(D)$.

With $\Lambda \subset \mathbb{N}^d$, approximation by multivariate polynomial space

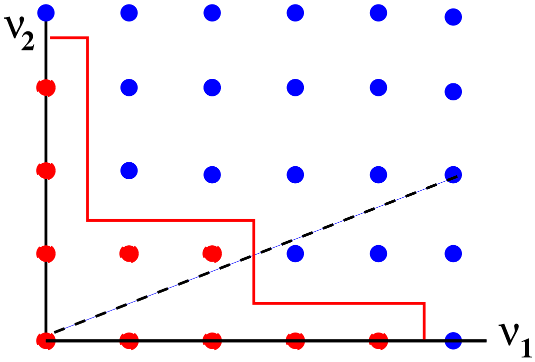
$$V_\Lambda := \left\{ \sum_{\nu\in\Lambda} v_\nu y^\nu, \ \ v_\nu \in V \right\} = V \otimes \mathbb{P}_\Lambda,$$

where $y^\nu = y_1^{\nu_1} \cdots y_d^{\nu_d}$.

We consider downward closed index sets : $\nu \in \Lambda$ and $\mu \le \nu \Rightarrow \mu \in \Lambda$.

Basis of $\mathbb{P}_\Lambda$: tensorized orthogonal polynomials $L_\nu(y) = \prod_{j=1}^d L_{\nu_j}(y_j)$ for $\nu \in \Lambda$.

Downward closed multivariate polynomials



Cohen-DeVore-Schwab (2011) + Bachmayr-Migliorati (2017) : approximation results.

Under suitable summability conditions on $(|\psi_j|)_{j\geq 1}$, there exists a sequence of downward closed sets $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n \dots$, with $n := \#(\Lambda_n)$ such that

$$\inf_{v\in V_n} \|u - v\|_{L^2(Y,V,\rho)} \leq Cn^{-s},$$

with $V_n := V_{\Lambda_n}$, where ρ is the uniform measure. The exponent $s > 0$ is robust with respect to the dimension d.

Chkifa-Cohen-Migliorati-Nobile-Tempone (2015) : estimate K_n for $\mathbb{P}_{\Lambda_n}$.

With $d\rho = \otimes^d(\frac{dx}{2})$ the uniform measure over Y, one has $K_n \leq n^2$ for all downward closed sets Λ_n such that $\#(\Lambda_n) = n$. Up to log factors, the stability regime is $m \gtrsim n^2$.

With the tensor-product Chebychev measure, improvement $K_n \leq n^\alpha$ with $\alpha := \frac{\ln 3}{\ln 2}$.

The theory and least-square method is not capable of handling lognormal diffusions :

$$a = \exp(b), \quad b = \sum_{j=1}^{d} y_j\psi_j, \quad y_j \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

which corresponds to the tensor product Gaussian measure over $Y = \mathbb{R}^d$.

Cohen-DeVore-Schwab (2011) + Bachmayr-Migliorati (2017) : approximation results.

Under suitable summability conditions on $(|\psi_j|)_{j\geq 1}$, there exists a sequence of downward closed sets $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n \ldots$, with $n := \#(\Lambda_n)$ such that

$$\inf_{v\in V_n} \|u - v\|_{L^2(Y,V,\rho)} \leq Cn^{-s},$$

with $V_n := V_{\Lambda_n}$, where ρ is the uniform measure. The exponent $s > 0$ is robust with respect to the dimension d.

Chkifa-Cohen-Migliorati-Nobile-Tempone (2015) : estimate K_n for $\mathbb{P}_{\Lambda_n}$.

With $d\rho = \otimes^d (\frac{dx}{2})$ the uniform measure over Y, one has $K_n \leq n^2$ for all downward closed sets Λ_n such that $\#(\Lambda_n) = n$. Up to log factors, the stability regime is $m \gtrsim n^2$.

With the tensor-product Chebychev measure, improvement $K_n \leq n^\alpha$ with $\alpha := \frac{\ln 3}{\ln 2}$.

The theory and least-square method is not capable of handling lognormal diffusions :

$$a = \exp(b), \quad b = \sum_{j=1}^{d} y_j \psi_j, \quad y_j \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

which corresponds to the tensor product Gaussian measure over $Y = \mathbb{R}^d$.

Cohen-DeVore-Schwab (2011) + Bachmayr-Migliorati (2017) : approximation results.

Under suitable summability conditions on $(|\psi_j|)_{j\geq 1}$, there exists a sequence of downward closed sets $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n \dots$, with $n := \#(\Lambda_n)$ such that

$$\inf_{v \in V_n} \|u - v\|_{L^2(Y,V,\rho)} \leq Cn^{-s},$$

with $V_n := V_{\Lambda_n}$, where ρ is the uniform measure. The exponent $s > 0$ is robust with respect to the dimension d.

Chkifa-Cohen-Migliorati-Nobile-Tempone (2015) : estimate K_n for $\mathbb{P}_{\Lambda_n}$.

With $d\rho = \otimes^d (\frac{dx}{2})$ the uniform measure over Y, one has $K_n \leq n^2$ for all downward closed sets Λ_n such that $\#(\Lambda_n) = n$. Up to log factors, the stability regime is $m \gtrsim n^2$.

With the tensor-product Chebychev measure, improvement $K_n \leq n^\alpha$ with $\alpha := \frac{\ln 3}{\ln 2}$.

The theory and least-square method is not capable of handling lognormal diffusions :

$$a = \exp(b), \quad b = \sum_{i=1}^{d} y_j \psi_j, \quad y_j \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

which corresponds to the tensor product Gaussian measure over $Y = \mathbb{R}^d$.

Optimal sampling

In the weighted least-square method, we sample according to $d\sigma$ such that $d\rho = w d\sigma$.

The stability condition is $m \gtrsim K_{n,w} \ln(n)$, where $K_{n,w} := \sup_{y \in Y} w(y) k_n(y)$.

Optimal choice : take

$$w(y) = \frac{n}{k_n(y)} \iff d\sigma := \frac{k_n}{n} d\rho = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\rho,$$

Then $d\sigma$ is a probability measure and we have $k_{n,w} = n$.

Therefore, up to log factors, the stability regime is $m \gtrsim n$ independently of ρ.

Optimal sampling

In the weighted least-square method, we sample according to $d\sigma$ such that $d\rho = wd\sigma$.

The stability condition is $m \gtrsim K_{n,w}\ln(n)$, where $K_{n,w} := \sup_{y\in Y} w(y)k_n(y)$.

Optimal choice : take

$$w(y) = \frac{n}{k_n(y)} \iff d\sigma := \frac{k_n}{n} d\rho = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\rho,$$

Then $d\sigma$ is a probability measure and we have $k_{n,w} = n$.

Therefore, up to log factors, the stability regime is $m \gtrsim n$ independently of ρ.

Optimal sampling

In the weighted least-square method, we sample according to $d\sigma$ such that $d\rho = w d\sigma$.

The stability condition is $m \gtrsim K_{n,w} \ln(n)$, where $K_{n,w} := \sup_{y\in Y} w(y) k_n(y)$.

Optimal choice : take

$$w(y) = \frac{n}{k_n(y)} \iff d\sigma := \frac{k_n}{n} d\rho = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\rho,$$

Then $d\sigma$ is a probability measure and we have $k_{n,w} = n$.

Therefore, up to log factors, the stability regime is $m \gtrsim n$ independently of ρ.

In the weighted least-square method, we sample according to $d\sigma$ such that $d\rho = w d\sigma$.

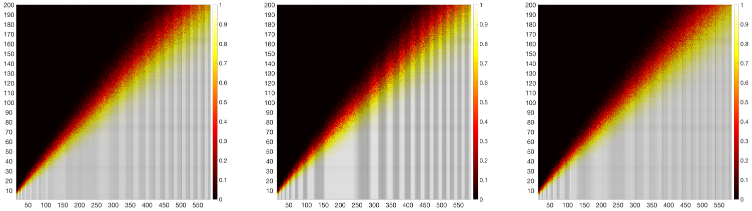
The stability condition is $m \gtrsim K_{n,w} \ln(n)$, where $K_{n,w} := \sup_{y \in Y} w(y) k_n(y)$.

Optimal choice : take

$$w(y) = \frac{n}{k_n(y)} \iff d\sigma := \frac{k_n}{n} d\rho = \frac{1}{n} \Big(\sum_{j=1}^{n} |L_j|^2 \Big) d\rho,$$

Then $d\sigma$ is a probability measure and we have $k_{n,w} = n$.

Therefore, up to log factors, the stability regime is $m \gtrsim n$ independently of ρ.



Stability regime for univariate polynomials with ρ Chebychev, uniform, and Gaussian (m on x axis, n on y axis).

Sampling the optimal density

The optimal sampling measure σ now depends on V_n :

$$d\sigma = d\sigma_n = \frac{k_n}{n} d\rho = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\rho.$$

In the case of parametric PDEs approximated with multivariate polynomials, $d\rho$ is a product measure (easy to sample), but $d\sigma_n$ is not.

Sampling strategies :

(i) Monte Carlo Markov Chain (MCMC) : generate by simple recursive rules a sample such that the the probability distribution asymptotically approaches $d\sigma_n$.

(ii) Conditional sampling : obtains first component by sampling the marginal $d\sigma_1(y_1)$, then the second component by sampling the conditional marginal probability $d\sigma_{y_1}(y_2)$ for this choice of the first component, etc...

(iii) Mixture sampling : draw uniform variable $j \in \{1, \dots, n\}$, then sample with probability $|L_j|^2 d\rho$.

Strategies (ii) and (iii) are more efficient on our cases of interests where the L_j have tensor product structure.

Sampling the optimal density

The optimal sampling measure σ now depends on V_n :

$$d\sigma = d\sigma_n = \frac{k_n}{n} d\rho = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\rho.$$

In the case of parametric PDEs approximated with multivariate polynomials, $d\rho$ is a product measure (easy to sample), but $d\sigma_n$ is not.

Sampling strategies :

(i) Monte Carlo Markov Chain (MCMC) : generate by simple recursive rules a sample such that the the probability distribution asymptotically approaches $d\sigma_n$.

(ii) Conditional sampling : obtains first component by sampling the marginal $d\sigma_1(y_1)$, then the second component by sampling the conditional marginal probability $d\sigma_{y_1}(y_2)$ for this choice of the first component, etc...

(iii) Mixture sampling : draw uniform variable $j \in \{1, \dots, n\}$, then sample with probability $|L_j|^2 d\rho$.

Strategies (ii) and (iii) are more efficient on our cases of interests where the L_j have tensor product structure.

Sampling on general domains

Optimal sampling may become unfeasible when $Y \subset \mathbb{R}^d$ is a domain with a general geometry : the $L_1, \dots, L_n$ have no simple expression and cannot be computed exactly.

General assumptions : χ_Y is easily computable $\Rightarrow$ sampling according to the uniform measure ρ is easy (sample uniformly on a bounding box, reject if $y \notin Y$).

An optimal two-step strategy (Cohen-Dolbeault, 2019) :

1. With $M \gtrsim K_n \ln(n)$ sample $z^1, \dots, z^M$ according to the uniform measure, and define

$$\tilde{\rho} := \frac{1}{M} \sum_{i=1}^{M} \delta_{z^i}.$$

Construct an orthonormal basis $\tilde{L}_1, \dots, \tilde{L}_n$ of V_n for the $L^2(X, \tilde{\rho})$ inner product and define $\tilde{k}_n = \sum_{j=1}^n |\tilde{L}_j|^2$.

2. With $m \gtrsim n \ln(n)$ sample $y^1, \dots, y^m$ according to

$$d\tilde{\sigma} = \frac{\tilde{k}_n}{n} d\tilde{\rho},$$

that is, select z^i with probability $p_i = \frac{\tilde{k}_n(z^i)}{Mn}$.

Sampling on general domains

Optimal sampling may become unfeasible when $Y \subset \mathbb{R}^d$ is a domain with a general geometry : the $L_1, \dots, L_n$ have no simple expression and cannot be computed exactly.

General assumptions : χ_Y is easily computable $\Rightarrow$ sampling according to the uniform measure ρ is easy (sample uniformly on a bounding box, reject if $y \notin Y$).

An optimal two-step strategy (Cohen-Dolbeault, 2019) :

1. With $M \gtrsim K_n \ln(n)$ sample $z^1, \dots, z^M$ according to the uniform measure, and define

$$\tilde{\rho} := \frac{1}{M} \sum_{i=1}^{M} \delta_{z^i}.$$

Construct an orthonormal basis $\tilde{L}_1, \dots, \tilde{L}_n$ of V_n for the $L^2(X, \tilde{\rho})$ inner product and define $\tilde{k}_n = \sum_{j=1}^{n} |\tilde{L}_j|^2$.

2. With $m \gtrsim n \ln(n)$ sample $y^1, \dots, y^m$ according to

$$d\tilde{\sigma} = \frac{\tilde{k}_n}{n} d\tilde{\rho},$$

that is, select z^i with probability $p_i = \frac{\tilde{k}_n(z^i)}{Mn}$.

Optimal sampling may become unfeasible when $Y \subset \mathbb{R}^d$ is a domain with a general geometry : the $L_1, \dots, L_n$ have no simple expression and cannot be computed exactly.

General assumptions : χ_Y is easily computable $\Rightarrow$ sampling according to the uniform measure ρ is easy (sample uniformly on a bounding box, reject if $y \notin Y$).

An optimal two-step strategy (Cohen-Dolbeault, 2019) :

1. With $M \gtrsim K_n \ln(n)$ sample $z^1, \dots, z^M$ according to the uniform measure, and define

$$\tilde{\rho} := \frac{1}{M} \sum_{i=1}^{M} \delta_{z^i}.$$

Construct an orthonormal basis $\tilde{L}_1, \dots, \tilde{L}_n$ of V_n for the $L^2(X, \tilde{\rho})$ inner product and define $\tilde{k}_n = \sum_{j=1}^n |\tilde{L}_j|^2$.

2. With $m \gtrsim n \ln(n)$ sample $y^1, \dots, y^m$ according to

$$d\tilde{\sigma} = \frac{\tilde{k}_n}{n} d\tilde{\rho},$$

that is, select z^i with probability $p_i = \frac{\tilde{k}_n(z^i)}{Mn}$.

Optimal sampling measure helps : Wozniakowski-Wasilkowski (2006), Krieg (2017)

We have

$$\|P_n u - u_n^{\rm PS}\|^2 = \sum_{j=1}^{n} |c_j - \tilde c_j|^2, \qquad \tilde c_j := \frac{1}{m}\sum_{i=1}^{m} w(y^i)L(y^i)u(y^i).$$

Variance analysis

$$\mathbb{E}(|c_j - \tilde c_j|^2) = \frac{1}{m}{\rm Var}(w(y)L_j(y)u(y)) \leq \frac{1}{m}\int_Y |w(y)|^2|L_j(y)|^2|u(y)|^2 d\sigma(y),$$

and therefore

$$\mathbb{E}(\|u_n - u_n^{\rm PS}\|^2) \leq \frac{1}{m}\int_Y w(y)\Big(\sum_{j=1}^{n}|L_j(y)|^2\Big)|u(y)|^2 d\rho(y).$$

Therefore, when using the optimal sampling measure, one finds that

$$\mathbb{E}(\|P_n u - u_n^{\rm PS}\|^2) \leq \frac{n}{m}\|u\|^2.$$

Multilevel strategy

For $l = 0, 1, \ldots, L$ set $n_l := 2^l$. Assume $u_{n_{l-1}} \in V_{n_{l-1}}$ has been constructed.

Draw $y^1, \ldots, y^{m_l}$ according to the measure σ_{n_l} with $m_l = \theta n_l$ for some $\theta > 1$.

Then define $u_{n_l} \in V_{n_l}$ by

$$u_{n_l} = u_{n_{l-1}} + \sum_{j=1}^{n_l} \tilde{c}_j L_j, \qquad \tilde{c}_j := \frac{1}{m_l} \sum_{i=1}^{m_l} w(y^i) L(y^i)(u(y^i) - u_{n_{l-1}}(y^i)).$$

One then has

$$\mathbb{E}(\|u - u_{n_L}\|^2) \leq \|u - P_{n_L} u\|^2 + \frac{n_l}{m_l} \mathbb{E}(\|u - u_{n_{L-1}}\|^2) = e_{n_L}(u)^2 + \theta^{-1} \mathbb{E}(\|u - u_{n_{L-1}}\|^2)$$

and we obtain by recursion $\mathbb{E}(\|u - u_{n_L}\|^2) \leq \sum_{l=0}^{L} \theta^{l-L} e_{n_l}(u)^2 + \theta^{-L-1} \mathbb{E}(\|u\|^2)$.

Assuming rate $e_n(u) \leq Cn^{-s}$ and taking $\theta > 2^{2s}$ we retrieve rate optimality.

The sampling budget is optimal : $m_0 + \cdots + m_L \leq 2\theta n_L$.

Recent work by D. Krieg : instance optimality achievable if $e_n(u)$ is known.

General defect : dimension n_l grows geometrically.

Multilevel strategy

For $l = 0, 1, \dots, L$ set $n_l := 2^l$. Assume $u_{n_{l-1}} \in V_{n_{l-1}}$ has been constructed.

Draw $y^1, \dots, y^{m_l}$ according to the measure σ_{n_l} with $m_l = \theta n_l$ for some $\theta > 1$.

Then define $u_{n_l} \in V_{n_l}$ by

$$u_{n_l} = u_{n_{l-1}} + \sum_{j=1}^{n_l} \tilde{c}_j L_j, \qquad \tilde{c}_j := \frac{1}{m_l} \sum_{i=1}^{m_l} w(y^i) L(y^i)(u(y^i) - u_{n_{l-1}}(y^i)).$$

One then has

$$\mathbb{E}(\|u - u_{n_L}\|^2) \leq \|u - P_{n_L} u\|^2 + \frac{n_l}{m_l} \mathbb{E}(\|u - u_{n_{L-1}}\|^2) = e_{n_L}(u)^2 + \theta^{-1} \mathbb{E}(\|u - u_{n_{L-1}}\|^2)$$

and we obtain by recursion $\mathbb{E}(\|u - u_{n_L}\|^2) \leq \sum_{l=0}^{L} \theta^{l-L} e_{n_l}(u)^2 + \theta^{-L-1} \mathbb{E}(\|u\|^2)$.

Assuming rate $e_n(u) \leq C n^{-s}$ and taking $\theta > 2^{2s}$ we retrieve rate optimality.

The sampling budget is optimal : $m_0 + \cdots + m_L \leq 2\theta n_L$.

Recent work by D. Krieg : instance optimality achievable if $e_n(u)$ is known.

General defect : dimension n_l grows geometrically.

For $l = 0, 1, \dots, L$ set $n_l := 2^l$. Assume $u_{n_{l-1}} \in V_{n_{l-1}}$ has been constructed.

Draw $y^1, \dots, y^{m_l}$ according to the measure σ_{n_l} with $m_l = \theta n_l$ for some $\theta > 1$.

Then define $u_{n_l} \in V_{n_l}$ by

$$u_{n_l} = u_{n_{l-1}} + \sum_{j=1}^{n_l} \tilde{c}_j L_j, \qquad \tilde{c}_j := \frac{1}{m_l} \sum_{i=1}^{m_l} w(y^i) L(y^i)(u(y^i) - u_{n_{l-1}}(y^i)).$$

One then has

$$\mathbb{E}(\|u - u_{n_L}\|^2) \leq \|u - P_{n_L} u\|^2 + \frac{n_l}{m_l} \mathbb{E}(\|u - u_{n_{L-1}}\|^2) = e_{n_L}(u)^2 + \theta^{-1} \mathbb{E}(\|u - u_{n_{L-1}}\|^2)$$

and we obtain by recursion $\mathbb{E}(\|u - u_{n_L}\|^2) \leq \sum_{l=0}^{L} \theta^{l-L} e_{n_l}(u)^2 + \theta^{-L-1} \mathbb{E}(\|u\|^2)$.

Assuming rate $e_n(u) \leq C n^{-s}$ and taking $\theta > 2^{2s}$ we retrieve rate optimality.

The sampling budget is optimal : $m_0 + \dots + m_L \leq 2\theta n_L$.

Recent work by D. Krieg : instance optimality achievable if $e_n(u)$ is known.

General defect : dimension n_l grows geometrically.

For $l = 0, 1, \dots, L$ set $n_l := 2^l$. Assume $u_{n_{l-1}} \in V_{n_{l-1}}$ has been constructed.

Draw $y^1, \dots, y^{m_l}$ according to the measure σ_{n_l} with $m_l = \theta n_l$ for some $\theta > 1$.

Then define $u_{n_l} \in V_{n_l}$ by

$$u_{n_l} = u_{n_{l-1}} + \sum_{j=1}^{n_l} \tilde{c}_j L_j, \qquad \tilde{c}_j := \frac{1}{m_l} \sum_{i=1}^{m_l} w(y^i) L(y^i)(u(y^i) - u_{n_{l-1}}(y^i)).$$

One then has

$$\mathbb{E}(\|u - u_{n_L}\|^2) \le \|u - P_{n_L} u\|^2 + \frac{n_l}{m_l} \mathbb{E}(\|u - u_{n_{L-1}}\|^2) = e_{n_L}(u)^2 + \theta^{-1} \mathbb{E}(\|u - u_{n_{L-1}}\|^2)$$

and we obtain by recursion $\mathbb{E}(\|u - u_{n_L}\|^2) \le \sum_{l=0}^{L} \theta^{l-L} e_{n_l}(u)^2 + \theta^{-L-1} \mathbb{E}(\|u\|^2)$.

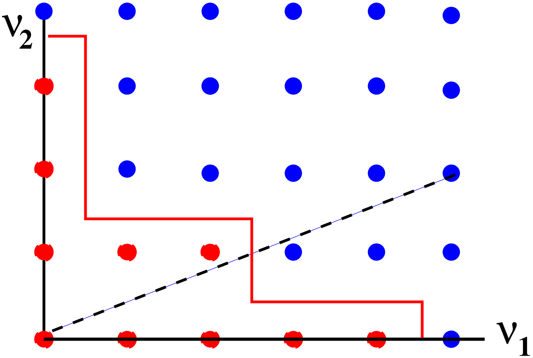
Assuming rate $e_n(u) \le C n^{-s}$ and taking $\theta > 2^{2s}$ we retrieve rate optimality.

The sampling budget is optimal : $m_0 + \cdots + m_L \le 2\theta n_L$.

Recent work by D. Krieg : instance optimality achievable if $e_n(u)$ is known.

General defect : dimension n_l grows geometrically.

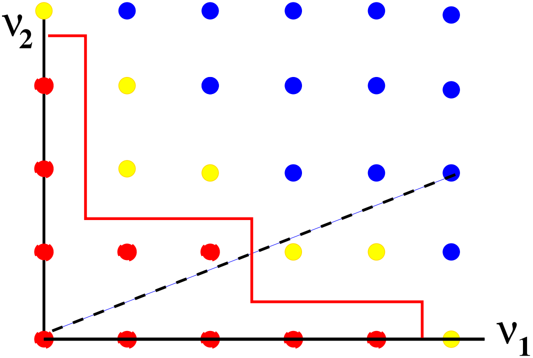
Update adaptively the polynomial space $\Lambda_{n-1} \longrightarrow \Lambda_n$, while increasing the amount of sample necessary for stability $m = m(n) \sim n \ln(n)$.



Problem : the optimal measure $\sigma = \sigma_n$ changes as we vary n. How should we recycle the previous samples ?

For certain simple cases $\sigma_n \sim \sigma^*$ as $n \longrightarrow \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

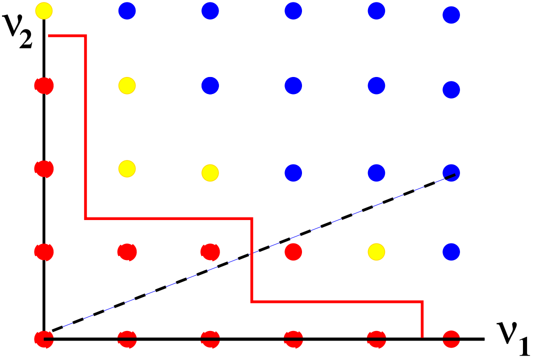
Update adaptively the polynomial space $\Lambda_{n-1} \rightarrow \Lambda_n$, while increasing the amount of sample necessary for stability $m = m(n) \sim n \ln(n)$.



Problem : the optimal measure $\sigma = \sigma_n$ changes as we vary n. How should we recycle the previous samples ?

For certain simple cases $\sigma_n \sim \sigma^*$ as $n \rightarrow \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

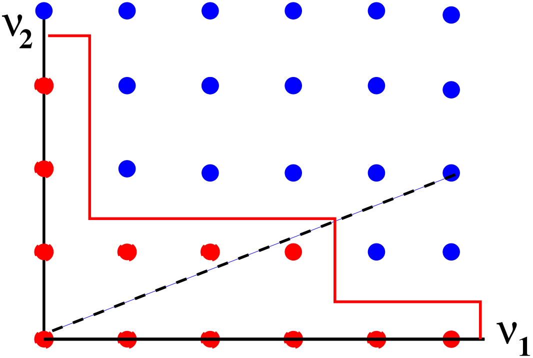
Update adaptively the polynomial space $\Lambda_{n-1} \longrightarrow \Lambda_n$, while increasing the amount of sample necessary for stability $m = m(n) \sim n \ln(n)$.



Problem : the optimal measure $\sigma = \sigma_n$ changes as we vary n. How should we recycle the previous samples ?

For certain simple cases $\sigma_n \sim \sigma^*$ as $n \longrightarrow \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

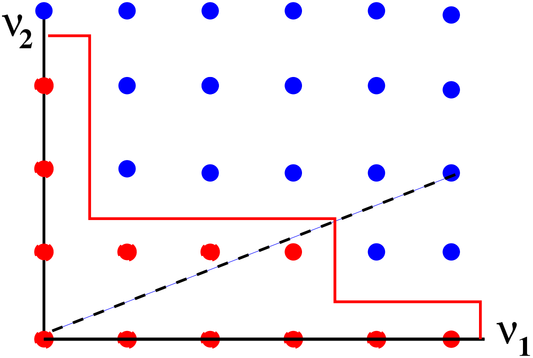
Update adaptively the polynomial space $\Lambda_{n-1} \longrightarrow \Lambda_n$, while increasing the amount of sample necessary for stability $m = m(n) \sim n \ln(n)$.



Problem : the optimal measure $\sigma = \sigma_n$ changes as we vary n. How should we recycle the previous samples ?

For certain simple cases $\sigma_n \sim \sigma^*$ as $n \longrightarrow \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

Update adaptively the polynomial space $\Lambda_{n-1} \longrightarrow \Lambda_n$, while increasing the amount of sample necessary for stability $m = m(n) \sim n \ln(n)$.

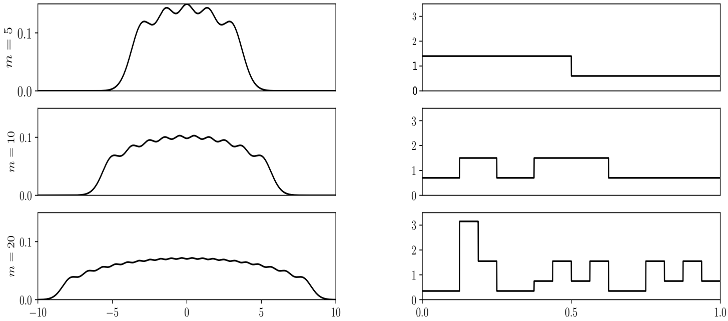


Problem : the optimal measure $\sigma = \sigma_n$ changes as we vary n. How should we recycle the previous samples ?

For certain simple cases $\sigma_n \sim \sigma^*$ as $n \longrightarrow \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

Example

Sampling densities σ_n for $n = 5, 10, 20$.



Left : Hermite polynomials of degrees $0, \ldots, m-1$ and ρ standard Gaussian.

Right : Haar wavelets selected by random tree refinement and ρ uniform.

Sequencial sampling

Observe that

$$d\sigma_n = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\rho = \Big(1 - \frac{1}{n}\Big) d\sigma_{n-1} + \frac{1}{n} d\nu_n \quad \text{where } d\nu_n = |L_n|^2 d\rho.$$

We use this mixture property to generate the sample in an incremental manner.

Assume that the sample $S_{n-1} = \{y^1, \dots, y^{m(n-1)}\}$ have been generated by independent draw according to the distribution $d\sigma_{n-1}$.

Then we generate a new sample $S_n = \{y^1, \dots, y^{m(n)}\}$ as follows :

For each $i = 1, \dots, m(n)$, pick Bernoulli variable $b_i \in \{0, 1\}$ with probability $\{\frac{1}{n}, 1 - \frac{1}{n}\}$.

If $b_i = 0$, generate y^i according to $d\nu_n$.

If $b_i = 1$, pick x_i incrementally inside S_{n-1}. If S_{n-1} has been exhausted generate y^i according to $d\sigma_{n-1}$.

Sequencial sampling

Observe that

$$d\sigma_n = \frac{1}{n}\left(\sum_{j=1}^{n} |L_j|^2\right) d\rho = \left(1-\frac{1}{n}\right) d\sigma_{n-1} + \frac{1}{n} d\nu_n \quad \text{where } d\nu_n = |L_n|^2 d\rho.$$

We use this mixture property to generate the sample in an incremental manner.

Assume that the sample $S_{n-1} = \{y^1, \dots, y^{m(n-1)}\}$ have been generated by independent draw according to the distribution $d\sigma_{n-1}$.

Then we generate a new sample $S_n = \{y^1, \dots, y^{m(n)}\}$ as follows :

For each $i = 1, \dots, m(n)$, pick Bernoulli variable $b_i \in \{0, 1\}$ with probability $\{\frac{1}{n}, 1 - \frac{1}{n}\}$.

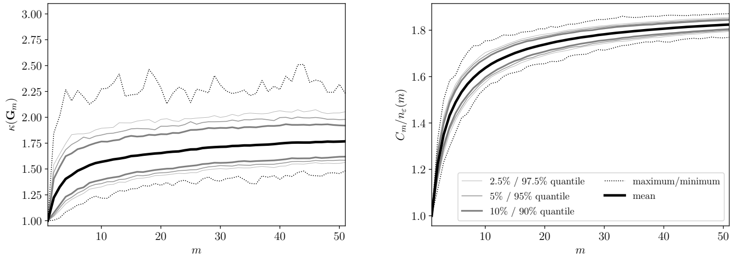
If $b_i = 0$, generate y^i according to $d\nu_n$.

If $b_i = 1$, pick x_i incrementally inside S_{n-1}. If S_{n-1} has been exhausted generate y^i according to $d\sigma_{n-1}$.

Optimality of the sequencial sampling algorithm

Arras-Bachmayr-Cohen (2018) : the total number of sample C_n used at stage n satisfies $\mathbb{E}(C_n) \sim n\ln(n)$ and $C_n \lesssim n\ln(n)$ with high probability for all values of n. With high probability, the matrix $\mathbf{G}$ satisfies $\kappa(\mathbf{G}) \leq 3$ for all values of n.

Example : hermite polynomials and Gaussian measure).



Left : Condition number $\kappa(\mathbf{G})$

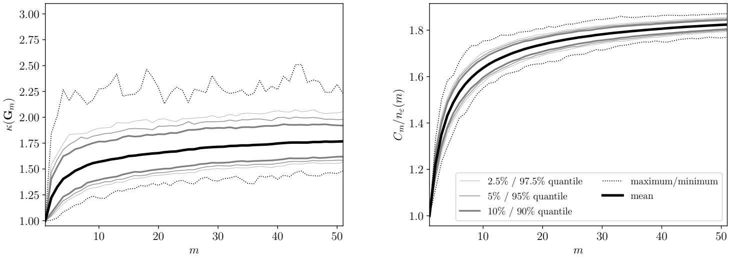
Right : Ratio between total sampling cost C_n and $m(n) \sim n\log n$.

Alternative strategy (Migliorati) : use a deterministic mixture.

Optimality of the sequencial sampling algorithm

Arras-Bachmayr-Cohen (2018) : the total number of sample C_n used at stage n satisfies $\mathbb{E}(C_n) \sim n \ln(n)$ and $C_n \lesssim n \ln(n)$ with high probability for all values of n. With high probability, the matrix $\mathbf{G}$ satisfies $\kappa(\mathbf{G}) \leq 3$ for all values of n.

Example : hermite polynomials and Gaussian measure).



Left : Condition number $\kappa(\mathbf{G})$

Right : Ratio between total sampling cost C_n and $m(n) \sim n \log n$.

Alternative strategy (Migliorati) : use a deterministic mixture.

Conclusions

Appropriate sampling yields optimal non-intrusive methods under the regime $m \sim n$.

Applicable to any measure ρ and spaces V_n, in any dimension.

Optimality can be preserved in a sequencial framework.

Convergence results are in expectation.

Perspectives

Similar convergence results with high probability ?

Convergence results in the uniform sense ?

Adaptive weighted least-squares strategies for the selection of index sets Λ_n.

Extend the optimal sampling measure theory to more general sensing systems.

Similar convergence results with deterministic sampling ?

Conclusions

Appropriate sampling yields optimal non-intrusive methods under the regime $m \sim n$.

Applicable to any measure ρ and spaces V_n, in any dimension.

Optimality can be preserved in a sequencial framework.

Convergence results are in expectation.

Perspectives

Similar convergence results with high probability ?

Convergence results in the uniform sense ?

Adaptive weighted least-squares strategies for the selection of index sets Λ_n.

Extend the optimal sampling measure theory to more general sensing systems.

Similar convergence results with deterministic sampling ?

References

A. Cohen, R. DeVore and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic PDEs*, Analysis and Application, 2011.

A. Cohen and R. DeVore, *High dimensional approximation of parametric PDEs*, Acta Numerica, 2015.

M. Bachmayr, A. Cohen and G. Migliorati, *Sparse polynomial approximation of parametric elliptic PDEs. Part I : affine coefficients*, M2AN 2017.

M. Bachmayr, A. Cohen, R. DeVore and G. Migliorati, *Sparse polynomial approximation of parametric elliptic PDEs. Part II : lognormal coefficients*, M2AN 2017.

G.W. Wasilkowski and H. Wozniakowski, *The power of standard information for multivariate approximation in the randomized setting*, Math of Comp. 2006.

D. Krieg, *Optimal Monte Carlo methods for L^2-approximation*, preprint 2018.

A. Cohen and G. Migliorati, *Optimal weighted least-squares methods*, SMAI J. of Comp. Math. 2017.

A. Doostan and M. Hadigol, *Least squares polynomial chaos expansion : A review of sampling strategies*, Computer Methods in Applied Mechanics and Engineering 2018.

B. Arras, M. Bachmayr and A. Cohen, *Sequential sampling for optimal weighted least squares approximations in hierarchical spaces*, to appear in SIMODS 2019

The unknown function u satisfies the Helmholtz equation

$$\Delta u + \lambda^2 u = 0,$$

over $Y \subset \mathbb{R}^2$ with unknown boundary condition, and where the spatial frequency λ is linked with with the considered temporal frequency ω.

Vekua theory : u belongs to the space V_λ generated by the plane waves

$$e_k(y) = e^{ik \cdot y}, \quad k \in \mathbb{R}^2 \text{ such that } |k| = \lambda,$$

which are particular solutions of $\Delta v + \lambda^2 v = 0$ over $\mathbb{R}^2$.

Angular discretization : we perform least-squares in the m dimensional space

$$V_n := \mathrm{Span}\{y \mapsto e_k(y) \ : \ k := \lambda(\cos(2j\pi/n), \sin(2j\pi/n)), \quad j = 0, \dots, n-1\}.$$

The unknown function u satisfies the Helmholtz equation

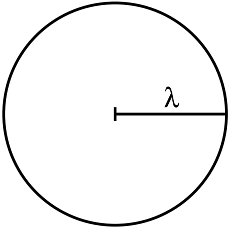
$$\Delta u + \lambda^2 u = 0,$$

over $Y \subset \mathbb{R}^2$ with unknown boundary condition, and where the spatial frequency λ is linked with with the considered temporal frequency ω.

Vekua theory : u belongs to the space V_λ generated by the plane waves

$$e_k(y) = e^{ik\cdot y}, \quad k \in \mathbb{R}^2 \text{ such that } |k| = \lambda,$$

which are particular solutions of $\Delta v + \lambda^2 v = 0$ over $\mathbb{R}^2$.



Angular discretization : we perform least-squares in the m dimensional space

$$V_n := \mathrm{Span}\{y \mapsto e_k(y) \ : \ k := \lambda(\cos(2j\pi/n), \sin(2j\pi/n)), \quad j = 0, \dots, n-1\}.$$

The unknown function u satisfies the Helmholtz equation

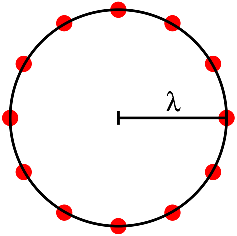
$$\Delta u + \lambda^2 u = 0,$$

over $Y \subset \mathbb{R}^2$ with unknown boundary condition, and where the spatial frequency λ is linked with with the considered temporal frequency ω.

Vekua theory : u belongs to the space V_λ generated by the plane waves

$$e_k(y) = e^{ik\cdot y}, \quad k \in \mathbb{R}^2 \text{ such that } |k| = \lambda,$$

which are particular solutions of $\Delta v + \lambda^2 v = 0$ over $\mathbb{R}^2$.



Angular discretization : we perform least-squares in the m dimensional space

$$V_n := \mathrm{Span}\{y \mapsto e_k(y) \;:\; k := \lambda(\cos(2j\pi/n), \sin(2j\pi/n)), \;\; j = 0, \dots, n-1\}.$$

Hipmair-Perugia-Moiola (2010) :if u belongs to the Sobolev space H^p,

$$\inf_{v\in V_n} \|u-v\|_{L^2} \le C_p n^{-p} \|v\|_{H^p}.$$

Fast decay of the approximation error with the number n of plane waves when u is a smooth solution of Helmholtz equation.

Chardon-Cohen-Daudet (2013) : for this space V_n and if Y is a disk, one has

$$K_n \sim n^2,$$

if $\rho = \frac{dy}{|Y|}$ is the uniform measure over Y, and

$$K_n \sim n,$$

if $\rho = (1-\alpha)\frac{dy}{|Y|} + \alpha\frac{ds}{|\partial Y|}$ combination of the uniform measures over Y and over its boundary ∂Y : distributing part of the microphones along the boundary improves the trade-off between the number of microphones and the quality of approximation.

Hipmair-Perugia-Moiola (2010) :if u belongs to the Sobolev space H^p,

$$\inf_{v \in V_n} \|u - v\|_{L^2} \le C_p n^{-p} \|v\|_{H^p}.$$

Fast decay of the approximation error with the number n of plane waves when u is a smooth solution of Helmholtz equation.

Chardon-Cohen-Daudet (2013) : for this space V_n and if Y is a disk, one has

$$K_n \sim n^2,$$

if $\rho = \frac{dy}{|Y|}$ is the uniform measure over Y, and

$$K_n \sim n,$$

if $\rho = (1-\alpha)\frac{dy}{|Y|} + \alpha \frac{ds}{|\partial Y|}$ combination of the uniform measures over Y and over its boundary ∂Y : distributing part of the microphones along the boundary improves the trade-off between the number of microphones and the quality of approximation.

Experimental result

α : proportion of microphones on the boundary
L : number of plane waves $(= n = \dim(V_n))$