Least-squares methods for high dimensional approximation

Albert Cohen

Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris

Columbia, 18-03-2019

Overview

1. Motivation for least square methods

Inverse problems : estimating state from pointwise data

Example : An acoustic pressure field $p(y,t)$ generated by a source is measured by m 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 $p(y^i,t) \mapsto \hat{p}(x^i,\omega)$ and focus at a frequency ω of interest.

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

Approximation of high dimensional parametric PDE's

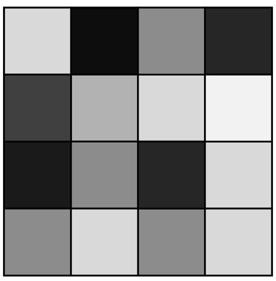
Partial differential equation $\mathcal{P}(u,y) = 0$ depending on a parameter vector $y \in Y \subset \mathbb{R}^d$ with $d >> 1$.

Simple example : steady state diffusion equation

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

on a physical domain D, with homogeneous Dirichlet boundary conditions $u_{|\partial D} = 0$.

Assume a diffusion function a that is piecewise constant on subdomains $D_1,\dots,D_d$, with values $y_1,\dots,y_d$, which define the parameter vector
$y = (y_1,\dots,y_d) \in Y = [y_{\min}, y_{\max}]^d$.



Non-intrusive methods

For each $y \in Y$, the PDE is well posed in some Hilbert space V : solution map

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

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

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). In the second case the solution $u(y)$ is a V-valued random variable.

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 $\overline{Q} = \mathbb{E}(Q(y))$.

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 $s = \frac{1}{q}$.

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

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

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 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 :

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

Pseudo-spectral method :

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

Questions : what prior spaces V_n should we use ? How should we sample to get instance/rate/budget optimality ?

2. Polynomial approximation of high-dimensional parametrized PDEs

References

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

A. Chkifa, A. Cohen and C. Schwab, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs*, Journal de Mathématiques Pures et Appliquées, 2015.

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.

Guiding example : elliptic PDEs

We consider the steady state diffusion equation

$$-\mathrm{div}(a\nabla u) = f \text{ on } D \subset \mathrm{R}^{\mathrm{m}} \text{ and } \mathrm{u}_{|\partial \mathrm{D}} = 0,$$

set on a domain $D \subset \mathbb{R}^m$, where $f = f(x) \in L^2(D)$ and $a \in L^\infty(D)$

Lax-Milgram lemma : assuming $a_{\min} := \min_{x\in D} a(x) > 0$, unique solution $u \in V = H_0^1(D)$ with

$$\|u\|_V := \|\nabla u\|_{L^2(D)} \le \frac{1}{a_{\min}} \|f\|_{V'}.$$

Proof of the estimate : multiply equation by u and integrate

$$a_{\min}\|u\|_V^2 \le \int_D a\nabla u \cdot \nabla u = -\int_D u\,\mathrm{div}(a\nabla u) = \int_D uf \le \|u\|_V \|f\|_{V'}.$$

We may extend this theory to the solution of the weak (or variational) formulation

$$\int_D a\nabla u \cdot \nabla v = \langle f, v\rangle, \quad v \in V = H_0^1(D),$$

if $f \in V' = H^{-1}(D)$

Parametrization

Assume diffusion coefficients in the form of an expansion

$$a = a(y) = \bar{a} + \sum_{j\ge 1} y_j \psi_j, \quad y = (y_j)_{j\ge 1} \in Y,$$

with $d >> 1$ or $d = \infty$ terms, where $\bar{a}$ and $(\psi_j)_{j\ge 1}$ are functions from L^∞,

Note that $a(y)$ is a function for each given y. We may also write

$$a = a(x, y) = \bar{a}(x) + \sum_{j\ge 1} y_j \psi_j(x), \quad x \in D, y \in Y,$$

where x and y are the spatial and parametric variable, respectively. Likewise, the corresponding solution $u(y)$ is a function $x \mapsto u(y, x)$ for each given y. We often ommit the reference to the spatial variable.

Up to a change of variable, we assume that all y_j range in $[-1, 1]$, therefore

$$y \in Y = [-1, 1]^d \text{ or } [-1, 1]^{\mathbb{N}}.$$

Uniform ellipticity assumption :

$$(UEA) \qquad 0 < r \le a(x, y) \le R, \quad x \in D,\ y \in Y$$

Then the solution map is bounded from Y to $V := H_0^1(D)$, that is, $u \in L^\infty(Y, V)$:

$$\|u(y)\|_V \le C_r := \frac{\|f\|_{V'}}{r}, \quad y \in Y,$$

Polynomial approximation

Approximations of the form

$$u_n(y) = \sum_{\nu\in\Lambda_n} c_\nu y^\nu,$$

with $\#(\Lambda_n) = n$ and $c_\nu \in V$.

Here $y^\nu = \prod_{j\geq 1} y_j^{\nu_j}$ for $\nu = (\nu_j)_{j\geq 1} \in \mathcal{F}$ finitely supported sequence.

Thus $u_n \in V_n = V_{\Lambda_n}$ where $V_\Lambda = V \otimes \mathbb{P}_\Lambda$ with $\mathbb{P}_\Lambda = \mathrm{span}\{y^\nu \,:\, \nu \in \Lambda\}$.

Strategy for theoretical approximation results :

Expand $y \mapsto u(y)$ as a polynomial series : Taylor, Legendre, Chebychev, Hermite...

Truncate the series by keeping n well chosen terms.

The choice of the truncation set Λ_n is critical.

Measure of performance

1. Uniform sense

$$\|u - u_n\|_{L^\infty(Y,V)} := \sup_{y\in Y} \|u(y) - u_n(y)\|_V,$$

2. Mean-square sense, for some probability measure ρ on Y,

$$\|u - u_n\|^2_{L^2(Y,V,\rho)} := \int_Y \|u(y) - u_n(y)\|_V^2 d\rho(y).$$

If y randomly distributed according to this measure, we have

$$\|u - u_n\|^2_{L^2(Y,V,\rho)} = \mathbb{E}(\|u(y) - u_n(y)\|_V^2).$$

Note that we always have

$$\mathbb{E}(\|u(y) - u_n(y)\|_V^2) \leq \|u - u_n\|^2_{L^\infty(Y,V)}.$$

A "worst case" estimate is more pessimistic than an "average" estimate.

Sparse approximation in ℓ^q spaces : fundamental observation (Stechkin)

Consider sequences $\mathbf{d} = (d_\nu)_{\nu\in\mathcal{F}}$ in $\ell^q(\mathcal{F})$ where $\mathcal{F}$ is a countable index set.

Best n-term approximation : we seek to approximate $\mathbf{d}$ by a sequence supported on a set of size n.

Best choice : $\mathbf{d}_n$ defined by leaving d_ν unchanged for the n largest $|d_\nu|$ and setting the others to 0.

Lemma : for $0 < p < q \leq \infty$, one has

$$\mathbf{d} \in \ell^p(\mathcal{F}) \implies \|\mathbf{d} - \mathbf{d}_n\|_{\ell^q} \leq C(n+1)^{-s}, \quad s = \frac{1}{p} - \frac{1}{q}, \quad C := \|\mathbf{d}\|_{\ell^p}.$$

Proof : introduce $(d_k^*)_{k\geq 1}$ the decreasing rearrangement of $(|d_\nu|)_{\nu\in\mathcal{F}}$, and combine

$$\|\mathbf{d} - \mathbf{d}_n\|_{\ell^q}^q = \sum_{k>n} |d_k^*|^q = \sum_{k>n} |d_k^*|^{q-p}|d_k^*|^p \leq C^p |d_{n+1}^*|^{q-p}$$

with

$$(n+1)|d_{n+1}^*|^p \leq \sum_{k=1}^{n+1} |d_k^*|^p \leq C^p.$$

Note that a large value of s corresponds to a value $p < 1$ (non-convex spaces).

From sequence approximation to Banach space valued function approximation

If a V-valued u has an expansion of the form $u(y) = \sum_{\nu\in\mathcal{F}} u_\nu \phi_\nu(y)$, in a given basis $(\phi_\nu)_{\nu\in\mathcal{F}}$, we use Steckhin's lemma to study the approximation of u by

$$u_n := \sum_{\nu\in\Lambda_n} u_\nu \phi_\nu,$$

where $\Lambda_n \subset \mathcal{F}$ corresponds to the n-largest $\|u_\nu\|_V$.

If $\sup_{y\in Y} |\phi_\nu(y)| = 1$, then by triangle inequality

$$\|u - u_n\|_{L^\infty(Y,V)} \leq \sum_{\nu\notin\Lambda_n} \|u_\nu \phi_\nu\|_{L^\infty(Y,V)} = \sum_{\nu\notin\Lambda_n} \|u_\nu\|_V,$$

If $(\phi_\nu)_{\nu\in\mathcal{F}}$ is an orthonormal basis of $L^2(Y, \rho)$, then by Parseval equality

$$\|u - u_n\|_{L^2(Y,V,\rho)}^2 = \sum_{\nu\notin\Lambda_n} \|u_\nu\|_V^2,$$

For concrete choices of bases a relevant question is thus : what smoothness properties of a function ensure that its coefficient sequence belongs to ℓ^p for small values of p ?

In the case of wavelet bases, such properties are characterized by Besov spaces.

In our present setting of high-dimensional functions $y \mapsto u(y)$ we rather use tensor-product polynomial bases instead of wavelet bases. Sparsity properties follows to the anisotropic features of these functions.

Return to the main guiding example

Steady state diffusion equation

$$-\mathrm{div}(a\nabla u) = f \quad \text{on} \quad D \subset \mathbb{R}^m \quad \text{and} \quad u_{|\partial D} = 0,$$

where $f = f(x) \in L^2(D)$ and the diffusion coefficients are given by

$$a = a(x,y) = \bar{a}(x) + \sum_{j\geq 1} y_j \psi_j(x),$$

where $\bar{a}$ and the $(\psi_j)_{j\geq 1}$ are given functions and $y \in Y := [-1,1]^{\mathbb{N}}$. Uniform ellipticity assumption :

$$(UEA) \qquad 0 < r \leq a(x,y) \leq R, \quad x \in D, \ y \in Y.$$

Equivalent expression of (UEA) : $\bar{a} \in L^\infty(D)$ and

$$\sum_{j\geq 1} |\psi_j(x)| \leq \bar{a}(x) - r, \quad x \in D,$$

or

$$\left\| \frac{\sum_{j\geq 1} |\psi_j|}{\bar{a}} \right\|_{L^\infty(D)} \leq \theta < 1.$$

Lax-Milgram : solution map is well-defined from Y to $V := H_0^1(D)$ with uniform bound

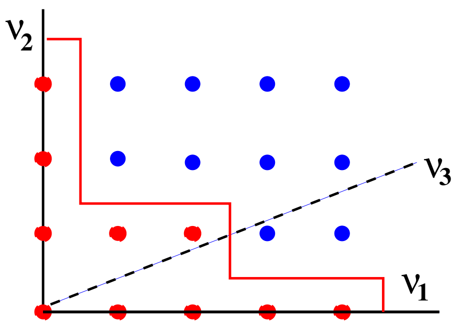
$$\|u(y)\|_V \leq C_r := \frac{\|f\|_{V'}}{r}, \quad y \in Y, \quad \text{where} \quad \|v\|_V := \|\nabla v\|_{L^2}.$$

Sparse polynomial approximations using Taylor series

We consider the expansion of $u(y) = \sum_{\nu\in\mathcal{F}} t_\nu y^\nu$, where

$$y^\nu := \prod_{j\geq 1} y_j^{\nu_j} \quad \text{and} \quad t_\nu := \frac{1}{\nu!}\partial^\nu u_{|y=0} \in V \quad \text{with} \quad \nu! := \prod_{j\geq 1} \nu_j! \quad \text{and} \quad 0! := 1.$$

where $\mathcal{F}$ is the set of all finitely supported sequences of integers (finitely many $\nu_j \neq 0$). The sequence $(t_\nu)_{\nu\in\mathcal{F}}$ is indexed by countably many integers.



Objective : identify a set $\Lambda \subset \mathcal{F}$ with $\#(\Lambda) = n$ such that u is well approximated by the partial expansion

$$u_\Lambda(y) := \sum_{\nu\in\Lambda} t_\nu y^\nu.$$

Best n-term approximation

Stechkin : if $(\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$ for some $p<1$, then for this Λ_n,

$$\sum_{\nu\notin\Lambda_n} \|t_\nu\|_V \leq Cn^{-s}, \quad s := \frac{1}{p}-1, \quad C := \|(\|t_\nu\|_V)\|_{\ell^p}.$$

Question : do we have $(\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$ for some $p<1$?

Cohen-DeVore-Schwab (2011) : under the uniform ellipticity assumption (UAE), then for any $p<1$,

$$(\|\psi_j\|_{L^\infty})_{j\geq 1} \in \ell^p(\mathbb{N}) \implies (\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F}).$$

We approximate $u(y)$ in $L^\infty(Y,V)$ with algebraic rate $\mathcal{O}(n^{-s})$ despite the curse of (infinite) dimensionality, due to the fact that y_j is less influencial as j gets large. Such approximation rates cannot be proved for the usual a-priori choices of Λ.

Same result for more general linear equations $Au=f$ with affine operator dependance : $A=\overline{A}+\sum_{j\geq 1} y_j A_j$ uniformly invertible over $y\in Y$, and $(\|A_j\|_{V\to W})_{j\geq 1} \in \ell^p(\mathbb{N})$, as well as other models (parabolic problems).

Key ingredient of proof : holomorphic extension of the solution map $z\mapsto u(z)$.

Idea of proof : extension to complex variable

Estimates on $\|t_\nu\|_V$ by complex analysis : extend $u(y)$ to $u(z)$ with $z=(z_j)\in\mathbb{C}^{\mathbb{N}}$.

Uniform ellipticity $\sum_{j\geq 1}|\psi_j| \leq \overline{a}-r$ implies that with $a(z)=\overline{a}+\sum_{j\geq 1} z_j\psi_j$,

$$0<r\leq \Re(a(x,z)) \leq |a(x,z)| \leq 2R, \quad x\in D,$$

for all $z\in\mathcal{U} := \{|z|\leq 1\}^{\mathbb{N}} = \otimes_{j\geq 1}\{|z_j|\leq 1\}$.

Lax-Milgram theory applies : $\|u(z)\|_V \leq C_0 = \frac{\|f\|_{V^*}}{r}$ for all $z\in\mathcal{U}$.

The function $u\mapsto u(z)$ is holomorphic in each variable z_j at any $z\in\mathcal{U}$: its first derivative $\partial_{z_j}u(z)$ is the unique solution to

$$\int_D a(z)\nabla\partial_{z_j}u(z)\cdot\nabla v = -\int_D \psi_j\nabla u(z)\cdot\nabla v, \quad v\in V.$$

Note that ∇ is with respect to spatial variable $x\in D$.

Extended domains of holomorphy : if $\omega=(\omega_j)_{j\geq 0}$ is any positive sequence such that for some $\delta>0$

$$\sum_{j\geq 1}\omega_j|\psi_j(x)| \leq \overline{a}(x)-\delta, \quad x\in D,$$

then u is holomorphic with uniform bound $\|u(z)\| \leq C_\delta = \frac{\|f\|_{V^*}}{\delta}$ in the polydisc

$$\mathcal{U}_\omega := \otimes_{j\geq 1}\{|z_j|\leq\omega_j\},$$

If $\delta<r$, we can take $\omega_j>1$.

Estimate on the Taylor coefficients

Use Cauchy formula. In 1 complex variable if $z \mapsto u(z)$ is holomorphic and bounded in a neighbourhood of disc $\{|z| \leq b\}$, then for all z in this disc

$$u(z) = \frac{1}{2i\pi} \int_{|z'|=b} \frac{u(z')}{z-z'} dz',$$

which leads by n differentiation at $z = 0$ to $|u^{(n)}(0)| \leq n! b^{-n} \max_{|z| \leq b} |u(z)|$.

Recursive application of this to all variables z_j such that $\nu_j \neq 0$, with $b = \omega_j$ gives

$$\|\partial^\nu u_{|z=0}\|_V \leq C_\delta \nu! \prod_{j \geq 1} \omega_j^{-\nu_j},$$

and thus

$$\|t_\nu\|_V \leq C_\delta \prod_{j \geq 1} \omega_j^{-\nu_j} = C_\delta \omega^{-\nu},$$

for any sequence $\omega = (\omega_j)_{j \geq 1}$ such that

$$\sum_{j \geq 1} \omega_j |\psi_j(x)| \leq \bar{a}(x) - \delta.$$

Optimization

Since ω is not fixed we have

$$\|t_\nu\|_V \leq C_\delta \inf\Big\{\omega^{-\nu} \ : \ \omega \ \text{s.t.} \ \sum_{j \geq 1} \omega_j |\psi_j(x)| \leq \bar{a}(x) - \delta, \ \ x \in D\Big\}.$$

We do not know the general solution to this problem, except in particular case, for example when the ψ_j have disjoint supports.

Instead design a particular choice $\omega = \omega(\nu)$ satisfying the constraint with $\delta = r/2$, for which we prove that

$$(\|\psi_j\|_{L^\infty})_{j \geq 1} \in \ell^p(\mathbb{N}) \implies (\omega(\nu)^{-\nu})_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F}),$$

therefore proving the main theorem.

A simple case

Assume that the ψ_j have disjoint supports. Then we maximize separately the ω_j so that

$$\sum_{j\geq 1} \omega_j |\psi_j(x)| \leq \bar{a}(x) - \frac{r}{2}, \quad x \in D,$$

which leads to

$$\omega_j := \min_{x\in D} \frac{\bar{a}(x) - \frac{r}{2}}{|\psi_j(x)|}.$$

We have, with $\delta = \frac{r}{2}$,

$$\|t_\nu\|_V \leq C_\delta \omega^{-\nu} = C_\delta b^\nu,$$

where $b = (b_j)$ and

$$b_j := \omega_j^{-1} = \max_{x\in D} \frac{|\psi_j(x)|}{\bar{a}(x) - \frac{r}{2}} \leq \frac{\|\psi_j\|_{L^\infty}}{R - \frac{r}{2}}.$$

Therefore $b \in \ell^p(\mathbb{N})$. From (UEA), we have $|\psi_j(x)| \leq \bar{a}(x) - r$ and thus $\|b\|_{\ell^\infty} < 1$. We finally observe that

$$b \in \ell^p(\mathbb{N}) \;\text{ and }\; \|b\|_{\ell^\infty} < 1 \iff (b^\nu)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F}).$$

Proof : factorize

$$\sum_{\nu\in\mathcal{F}} b^{p\nu} = \prod_{j\geq 1} \sum_{n\geq 0} b_j^{pn} = \prod_{j\geq 1} \frac{1}{1 - b_j^p}.$$

Improved summability results

One limitation of the previous result is that it depends on the ψ_j only through $\|\psi_j\|_{L^\infty}$, without taking their support into account. Improved results can be obtained, without relying on complex variable, by better exploiting the specific structure of PDE.

Recursive formula for the Taylor coefficients : with $e_j = (0, \dots, 0, 1, 0, \dots)$ the Kroeneker sequence of index j, the coefficient t_ν is solution to

$$\int_D \bar{a} \nabla t_\nu \nabla v = - \sum_{j:\ \nu_j \neq 0} \int_D \psi_j \nabla t_{\nu - e_j} \nabla v, \quad v \in V.$$

We introduce the quantities

$$d_\nu := \int_D \bar{a} |\nabla t_\nu|^2 \quad \text{and} \quad d_{\nu,j} := \int_D |\psi_j|\,|\nabla t_\nu|^2.$$

Recall that (UEA) implies that $\left\| \frac{\sum_{j\geq 1} |\psi_j|}{\bar{a}} \right\|_{L^\infty(D)} \leq \theta < 1$. In particular

$$\sum_{j\geq 1} d_{\nu,j} \leq \theta d_\nu.$$

We use here the equivalent norm $\|v\|_V^2 := \int_D \bar{a} |\nabla v|^2$.

Lemma : under (UEA), one has $\sum_{\nu\in\mathcal{F}} d_\nu = \sum_{\nu\in\mathcal{F}} \|t_\nu\|_V^2 < \infty$.

Proof

Taking $v = t_\nu$ in the recursion gives

$$d_\nu = \int_D \bar{a}|\nabla t_\nu|^2 = -\sum_{j:\,\nu_j\neq 0}\int_D \psi_j \nabla t_{\nu-e_j}\nabla t_\nu.$$

Apply Young's inequality on the right side gives

$$d_\nu \leq \sum_{j:\,\nu_j\neq 0}\Big(\frac{1}{2}\int_D |\psi_j|\,|\nabla t_\nu|^2 + \frac{1}{2}\int_D |\psi_j|\,|\nabla t_{\nu-e_j}|^2\Big) = \frac{1}{2}\sum_{j:\,\nu_j\neq 0} d_{\nu,j} + \frac{1}{2}\sum_{j:\,\nu_j\neq 0} d_{\nu-e_j,j}.$$

The first sum is bounded by θd_ν, therefore

$$\Big(1-\frac{\theta}{2}\Big)d_\nu \leq \frac{1}{2}\sum_{j:\,\nu_j\neq 0} d_{\nu-e_j,j}.$$

Now summing over all $|\nu| = k$ gives

$$\Big(1-\frac{\theta}{2}\Big)\sum_{|\nu|=k} d_\nu \leq \frac{1}{2}\sum_{|\nu|=k}\sum_{j:\,\nu_j\neq 0} d_{\nu-e_j,j} = \frac{1}{2}\sum_{|\nu|=k-1}\sum_{j\geq 1} d_{\nu,j} \leq \frac{\theta}{2}\sum_{|\nu|=k-1} d_\nu.$$

Therefore $\sum_{|\nu|=k} d_\nu \leq \kappa \sum_{|\nu|=k-1} d_\nu$ with $\kappa := \frac{\theta}{2-\theta} < 1$, and thus $\sum_{\nu\in\mathcal{F}} d_\nu < \infty$.

Rescaling

Now let $\omega = (\omega_j)_{j\geq 1}$ be any sequence with $\omega_j > 1$ such that $\sum_{j\geq 1}\omega_j|\psi_j| \leq \bar{a} - \delta$ for some $\delta > 0$, or equivalently such that $\Big\|\frac{\sum_{j\geq 1}\omega_j|\psi_j|}{\bar{a}}\Big\|_{L^\infty(D)} \leq \theta < 1$.

Consider the rescaled solution map $\tilde{u}(y) = u(\omega y)$ where $\omega y := (\omega_j y_j)_{j\geq 1}$ which is the solution of the same problem as u with ψ_j replaced by $\omega_j\psi_j$.

Since (UEA) holds for for these rescaled functions, the previous lemma shows that

$$\sum_{\nu\in\mathcal{F}}\|\tilde{t}_\nu\|_V^2 < \infty,$$

where

$$\tilde{t}_\nu := \frac{1}{\nu!}\partial^\nu \tilde{u}(0) = \frac{1}{\nu!}\omega^\nu\partial^\nu u(0) = \omega^\nu t_\nu.$$

This therefore gives the weighted ℓ^2 estimate

$$\sum_{\nu\in\mathcal{F}}(\omega^\nu\|t_\nu\|_V)^2 \leq C < \infty.$$

In particular, we retrieve the estimate $\|t_\nu\|_V \leq C\omega^{-\nu}$ that can be obtained by a complex variable approach (holomorphy of the solution map $y \mapsto u(y)$), however the above estimate is a bit stronger.

An alternate summability result

Applying Hölder's inequality gives

$$\sum_{\nu\in\mathcal{F}} \|t_\nu\|_V^p \le \Big(\sum_{\nu\in\mathcal{F}} (\omega^\nu \|t_\nu\|_V)^2\Big)^{p/2} \Big(\sum_{\nu\in\mathcal{F}} \omega^{-q\nu}\Big)^{1-p/2},$$

with $q = \frac{2p}{2-p} > p$, or equivalently $\frac{1}{q} = \frac{1}{p} - \frac{1}{2}$.
The sum in second factor is finite provided that $(\omega_j^{-1})_{j\ge1} \in \ell^q$. Therefore, the following result holds.

Bachmayr-Cohen-Migliorati (2017) : Let p and q be such that $\frac{1}{q} = \frac{1}{p} - \frac{1}{2}$. Assume that there exists a sequence $\omega = (\omega_j)_{j\ge1}$ of numbers larger than 1 such that

$$\sum_{j\ge1} \omega_j |\psi_j| \le \overline{a} - \delta,$$

for some $\delta > 0$ and

$$(\omega_j^{-1})_{j\ge1} \in \ell^q.$$

Then $(\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$.

The above conditions ensuring ℓ^p summability of $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ are significantly weaker than those in the first summability theorem especially for locally supported ψ_j.

Disjoint supports

Assume that the ψ_j have disjoint supports.

Then with $\delta = \frac{r}{2}$, we choose

$$\omega_j := \min_{x\in D} \frac{\overline{a}(x) - \frac{r}{2}}{|\psi_j(x)|} > 1.$$

so that $\sum_{j\ge1} \omega_j |\psi_j| \le \overline{a} - \delta$ holds.

We have

$$b_j := \omega_j^{-1} = \frac{|\psi_j(x)|}{\overline{a}(x) - \frac{r}{2}} \le \frac{\|\psi_j\|_{L^\infty}}{R - \frac{r}{2}}.$$

Thus in this case, our result gives for any $0 < q < \infty$,

$$(\|\psi_j\|_{L^\infty})_{j\ge1} \in \ell^q(\mathbb{N}) \implies (\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F}),$$

with $\frac{1}{q} = \frac{1}{p} - \frac{1}{2}$.

Similar improved results if the ψ_j have supports with limited overlap, such as wavelets.

No improvement in the case of globally supported functions, such as Fourier bases.

Other models

Model 1 : same PDE but no affine dependence, e.g. $a(x,y) = \overline{a}(x) + (\sum_{j\geq 0} y_j\psi_j(x))^2$. Assuming that $\overline{a}(x) \geq r > 0$ guarantees ellipticity uniformly over $y \in Y$.

Model 2 : similar problems + non-linearities, e.g.

$$g(u) - \mathrm{div}(a\nabla u) = f \quad \text{on} \quad D = D(y) \quad u_{|\partial D} = 0,$$

with same assumptions on a and f. Well-posedness in $V = H_0^1(D)$ for all $f \in V'$ is ensured for certain nonlinearities, e.g. $g(u) = u^3$ of u^5 in dimension $m = 3$ ($V \subset L^6$).

Model 3 : PDE's on domains with parametrized boundaries, e.g.

$$-\Delta v = f \quad \text{on} \quad D = D_y \quad u_{|\partial D} = 0.$$

where the boundary of D_y is parametrized by y, e.g.

$$D_y := \{(x_1, x_2) \in \mathbb{R}^2 \quad : \quad 0 < x_1 < 1 \quad \text{and} \quad 0 < x_2 < b(x_1, y)\},$$

where $b = b(x,y) = \overline{b}(x) + \sum_j y_j\psi_j(x)$ satisfies $0 < r < b(x,y) < R$. We transport this problem on the reference domain $[0,1]^2$ and study

$$u(y) := v(y) \circ \phi_y, \quad \phi_y : [0,1]^2 \to D_y, \quad \phi_y(x_1, x_2) := (x_1, x_2 b(x_1, y)).$$

which satisfies a diffusion equation with coefficient $a = a(x,y)$ non-affine in y.

Polynomial approximation for these models

In contrast to our guiding example (which we refer to as model 0), bounded holomorphic extension is generally not feasible in a complex domain containing the polydisc $\mathcal{Y} = \otimes_{j\geq 1}\{|z_j| \leq 1\}$. For this reason,Taylor series are not expected to converge.

Instead we consider the tensorized Legendre expansion

$$u(y) = \sum_{\nu\in\mathcal{F}} v_\nu L_\nu(y),$$

where $L_\nu(y) := \prod_{j\geq 1} L_{\nu_j}(y_j)$ and $(L_k)_{k\geq 0}$ are the Legendre polynomials normalized in $L^2\left([-1,1], \frac{dt}{2}\right)$.

Thus $(L_\nu)_{\nu\in\mathcal{F}}$ is an orthonormal basis for $L^2(Y, V, \rho)$ where $\rho := \otimes_{j\geq 1}\frac{dy_j}{2}$ is the uniform probability measure and we have

$$v_\nu = \int_Y u(y) L_\nu(y) d\rho(y).$$

We also consider the L^∞-normalized Legendre polynomials $P_k = (1+2k)^{-1/2} L_k$ and their tensorized version P_ν, so

$$u(y) = \sum_{\nu\in\mathcal{F}} w_\nu P_\nu(y),$$

where $w_\nu := \left(\prod_{j\geq 1}(1+\nu_j)^{1/2}\right) v_\nu$.

Main result

Chkifa-Cohen-Schwab (2014) : For models 0, 1, 2 and 3, and for any $p < 1$,

$$(\|\psi_j\|_X)_{j\geq 1} \in \ell^p(\mathbb{N}) \implies (\|v_\nu\|_V)_{\nu\in\mathcal{F}} \text{ and } (\|w_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F}).$$

with $X = L^\infty$ for models 0, 1, 2, and $X = W^{1,\infty}$ for model 3.

By the same application of Stechkin's argument as for Taylor expansions, best n-term truncations for the L^∞ normalized expansion converge rate $\mathcal{O}(n^{-s})$ in $L^\infty(Y, V)$ where $s = \frac{1}{p} - 1$.

Best n-term truncations for the L^2 normalized expansion converge with rate $\mathcal{O}(n^{-r})$ in $L^2(Y, V, \rho)$ where $r = \frac{1}{p} - \frac{1}{2}$.

In the particular case of our guiding example, model 0, we can obtain improved summability results for Legendre expansions, similar to Taylor expansions.

Key ingredient in the proof of the above theorem : estimates of Legendre coefficients for holomorphic functions in a "small" complex neighbourhood of Y.

Taylor vs Legendre expansions

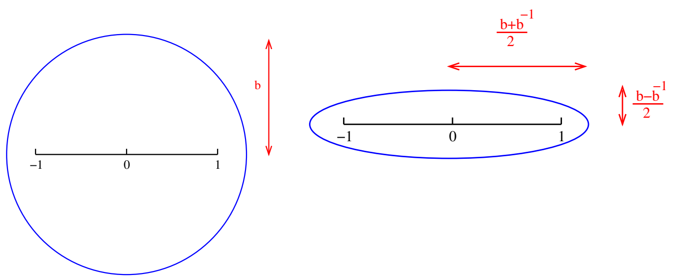
In one variable :

- If u is holomorphic in an open neighbourhood of the disc $\{|z| \leq b\}$ and bounded by M on this disc, then the n-th Taylor coefficient of u is bounded by

$$|t_n| := \left|\frac{u^{(n)}(0)}{n!}\right| \leq Mb^{-n}$$

- If u is holomorphic in an open neighbourhood of the domain $\mathcal{E}_b$ limited by the ellipse of semi axes of length $(b + b^{-1})/2$ and $(b - b^{-1})/2$, for some $b > 1$, and bounded by M on this domain, then the n-th Legendre coefficent w_n of u is bounded by

$$|w_n| \leq Mb^{-n}(1 + 2n)\phi(b), \qquad \phi(b) := \frac{\pi b}{b - 1}$$



Lognormal parametrization

We assume diffusion coefficients are given by

$$a = \exp(b),$$

with b a random function defined by an affine expansion of the form

$$b = b(y) = \sum_{j\geq 1} y_j \psi_j,$$

where (ψ_j) is a given family of functions from $L^\infty(D)$ and $y = (y_j)_{j\geq 1}$ a sequence of i.i.d. standard Gaussians $\mathcal{N}(0,1)$ variables.

Thus y ranges in $Y = \mathbb{R}^\mathbb{N}$ equipped with the probabilistic structure $(Y, \mathcal{B}(Y), \rho)$ where $\mathcal{B}(Y)$ is the cylindrical Borel Σ-algebra and ρ the tensorized Gaussian measure.

Commonly used stochastic model for diffusion in porous media.

The solution $u(y)$ is well defined in V for those $y \in Y$ such that $b(y) \in L^\infty(D)$, with

$$\|u(y)\|_V \leq \frac{1}{a_{\min}(y)} \|f\|_{V'} \leq \exp(\|b(y)\|_{L^\infty}) \|f\|_{V'}.$$

Main theoretical questions

1. Integrability : under which conditions is $y \mapsto u(y)$ Bochner measurable with values in V and satifies for $0 \leq k < \infty$.

$$\|u\|^k_{L^k(Y,V,\rho)} = \mathbb{E}(\|u(y)\|^k_V) < \infty,$$

In view of $\|u(y)\|_V \leq \exp(\|b(y)\|_{L^\infty}) \|f\|_{V'}$, this holds if $\mathbb{E}(\exp(k\|b(y)\|_{L^\infty}) < \infty$.

2. Approximability : if $u \in L^2(Y, V, \rho)$, consider the multivariate Hermite expansion

$$u = \sum_{\nu\in\mathcal{F}} u_\nu H_\nu, \quad H_\nu(y) := \prod_{j\geq 1} H_{\nu_j}(y_j) \quad \text{and} \quad u_\nu := \int_Y u(y) H_\nu(y) d\rho(y)$$

where $\mathcal{F}$ is the set of finitely supported integer sequences $\nu = (\nu_j)_{j\geq 1}$.

Best n-term approximation : $u_n = \sum_{\nu\in\Lambda_n} u_\nu H_\nu$, with Λ_n indices of n largest $\|u_\nu\|_V$.

Stechkin lemma : if $(\|u_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$ for some $0 < p < 2$ then

$$\|u - u_n\|_{L^2(Y,V,\rho)} \leq Cn^{-s}, \quad s := \frac{1}{p} - \frac{1}{2}, \quad C := \|(\|u_\nu\|_V)_{\nu\in\mathcal{F}}\|_{\ell^p}$$

Existing results

Integrability : sufficient conditions for $u \in L^k(Y, V, \rho)$ for all $0 \le k < \infty$ are known.

1. Smoothness : $C_b \in C^\alpha(D \times D)$ for some $\alpha > 0$ (Charrier).

2. Summability : $\sum_{j \ge 1} \|\psi_j\|_{L^\infty} < \infty$ (Schwab-Gittelson-Hoang)

3. $\sum_{j \ge 1} \|\psi_j\|_{L^\infty}^{2-\delta} \|\psi_j\|_{C^\alpha}^{\delta} < \infty$ for some $0 < \delta < 1$ (Dashti-Stuart)

Approximability :

Hoang-Schwab (2014) : for $0 < p \le 1$, if $(j\|\psi_j\|_{L^\infty}) \in \ell^p(\mathbb{N})$ then $(\|u_\nu\|_V) \in \ell^p(\mathcal{F})$.

Bachmayr-Cohen-DeVore-Migliorati (2017) : let $0 < p < 2$ and define $q := q(p) = \frac{2p}{2-p} > p$ (or equivalently $\frac{1}{q} = \frac{1}{p} - \frac{1}{2}$). Assume that there exists a positive sequence $\omega = (\omega_j)_{j \ge 1}$ such that

$$(\omega_j^{-1})_{j \ge 1} \in \ell^q(\mathbb{N}) \quad \text{and} \quad \sup_{x \in D} \sum_{j \ge 1} \omega_j |\psi_j(x)| < \infty.$$

Then $(\|u_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})$.

Main ingredient in the proof of the main result

1. Relate Hermite coefficients u_ν and partial derivatives $\partial^\mu u$. Base on 1-d Rodrigues formula : $H_n(t) = \frac{(-1)^n}{\sqrt{n!}} \frac{g^{(n)}(t)}{g(t)}$, where $g(t) := (2\pi)^{-1/2} \exp(-t^2/2)$. After some computation this leads to weighted ℓ^2 identity for any sequence $\omega := (\omega_j)_{j \ge 1}$.

$$\sum_{\|\mu\|_{\ell^\infty} \le r} \frac{\omega^{2\mu}}{\mu!} \int_Y \|\partial^\mu u(y)\|_V^2 \, d\rho(y) = \sum_{\nu \in \mathcal{F}} b_\nu \|u_\nu\|_V^2,$$

where $b_\nu := \sum_{\|\mu\|_{\ell^\infty} \le r} \binom{\nu}{\mu} \omega^{2\mu}$.

2. Prove finiteness of left hand side $\sum_{\|\mu\|_{\ell^\infty} \le r} \frac{\omega^{2\mu}}{\mu!} \int_Y \|\partial^\mu u(y)\|_V^2 \, d\rho(y)$ when

$$\sup_{x \in D} \sum_{j \ge 1} \omega_j |\psi_j(x)| =: K < C_r := r^{-1/2} \ln 2.$$

Use PDE : $\int_D a(y) \nabla \partial^\mu u(y) \cdot \nabla v = -\sum_{\nu \le \mu, \, \nu \ne \mu} \binom{\mu}{\nu} \int_D \psi^{\mu-\nu} a(y) \nabla \partial^\nu u(y) \cdot \nabla v$.

3. Derive ℓ^p estimate by mean of Hölder's inequality :

$$\Big(\sum_{\nu \in \mathcal{F}} \|u_\nu\|_V^p \Big)^{1/p} \le \Big(\sum_{\nu \in \mathcal{F}} b_\nu \|u_\nu\|_V^2 \Big)^{1/2} \Big(\sum_{\nu \in \mathcal{F}} b_\nu^{-q/2} \Big)^{1/q}.$$

We prove that the second factor is finite if $(\omega_j^{-1})_{j \ge 1} \in \ell^q(\mathbb{N})$ and r such that $\frac{2}{r+1} < p$.

In summary

The curse of dimensionality can be "defeated" by exploiting both smoothness and anisotropy in the different variables.

For certain models, this can be achieved by sparse polynomial approximations.

The way we parametrize the problem, or represent its solution, is crucial.

3. Numerical methods for polynomial approximation

From approximation results to numerical methods

The results so far are approximation results. They say that for several models of parametric PDEs, the solution map $y \mapsto u(y)$ can be accurately approximate (with rate n^{-s} for some $s > 0$) by multivariate polynomials having n terms.

These polynomials are computed by best n-term truncation of Taylor or Legendre or Hermite series, but this is not feasible in practical numercial methods.

Problem 1 : the best n-term index sets Λ_n are computationally out of reach. Their identification would require the knowledge of all coefficients in the expansion.

Objective : identify non-optimal yet good sets Λ_n.

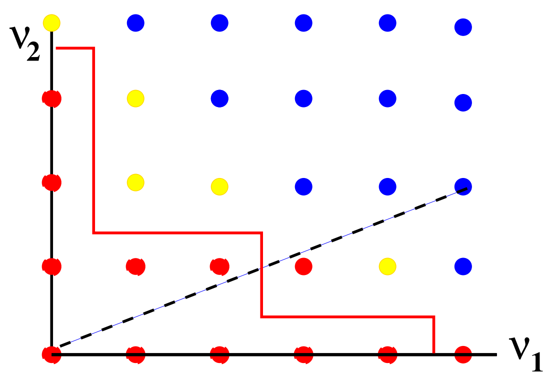
Problem 2 : the exact polynomial coefficients t_ν (or v_ν, w_ν, u_ν) of u for the indices $\nu \in \Lambda_n$ cannot be computed exactly.

Objective : numerical strategy for approximately computing polynomial coefficients.

Numerical methods : strategies to build the sets Λ_n

(i) Non-adaptive, based on the available a-priori estimates for the $\|t_\nu\|_V$ (or $\|v_\nu\|_V$, $\|w_\nu\|_V$, $\|u_\nu\|_V$). Take Λ_n to be the set corresponding to the n largest such estimates.

(ii) Adaptive, based on a-posteriori information gained in the computation $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n \cdots$.



Downward closed index sets

For adaptive algorithms it is critical that the index chosen sets are downward closed

$$\nu \in \Lambda \text{ and } \mu \leq \nu \implies \mu \in \Lambda,$$

where $\mu \leq \nu$ means that $\mu_j \leq \nu_j$ for all $j \geq 1$.

Such sets are also called downward closed (or lower) sets.

The sets corresponding to the n largest coefficients or estimates are generally not downward closed, however the same convergence rates established in the approximation theorems can be proved when imposing such a structure.

If Λ is downward closed, we consider the polynomial space

$$\mathbb{P}_\Lambda = \mathrm{span}\{y \to y^\nu \,:\, \nu \in \Lambda\} = \mathrm{span}\{L_\nu \,:\, \nu \in \Lambda\} = \mathrm{span}\{H_\nu \,:\, \nu \in \Lambda\}$$

and its V-valued version

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

After having selected Λ_n we search for a computable approximation of u in V_{Λ_n}.

Non-intrusive strategies to build the polynomial approximation

Based on snapshots $u(y^i)$ where $y^i \in Y$ for $i = 1, \dots, m$..

1. Pseudo spectral methods : computation of $\sum_{\nu \in \Lambda_n} v_\nu L_\nu$ by quadrature

$$v_\nu = \int_Y u(y) L_\nu(y) d\rho(y) \approx \sum_{i=1}^m w_i u(y^i) L_\nu(y^i).$$

2. Interpolation : with $m = n = \dim(\mathbb{P}_{\Lambda_n})$ search for $u_n = I_{\Lambda_n} u \in V_{\Lambda_n}$ such that

$$u_n(y^i) = u(y^i), \quad i = 1, \dots, n.$$

3. Least-squares : with $m \geq n$, search for $u_n \in V_{\Lambda_n}$ minimizing

$$\sum_{i=1}^m \|u(y^i) - u_n(y^i)\|_V^2.$$

4. Underdetermined least-squares : with $m < n$ search for $u_n \in V_{\Lambda_n}$ minimizing

$$\sum_{i=1}^m \|u(y^i) - u_n(y^i)\|_V^2 + \pi(u_n),$$

where π is a penalization functional. Compressed sensing : take for π the (weighted) ℓ^1 sum of V-norms of Legendre coefficients of u_n (promote sparse solutions).

Advantages of non-intrusive methods

Applicable to a broad range of models, in particular non-linear PDEs.

Adaptive algorithms seem to work well for the interpolation and least squares approach, however with no theoretical guarantees.

Additional prescriptions for non-intrusive methods :

(i) Progressive : enrichment $\Lambda_n \to \Lambda_{n+1}$ requires only one or a few new snapshots.

(ii) Stable : moderate growth with n of the norm of the reconstruction operator (Lebesgue constant in the case of interpolation).

Main issue : how to best choose the point y^i ?

In the following we focus on least-squares, for which interesting stability and accuracy results can be obtained in recent years using random sampling.

4. Least squares methods with random sampling

References

A. Cohen, M. Davenport and D. Leviatan, *On the stability and accuracy of least-squares approximation*, FoCM 2013.

A. Chkifa, A. Cohen, G. Migliorati, F. Nobile and R. Tempone, *Discrete least squares polynomial approximation with random evaluations - Application to parametric and stochastic elliptic PDEs*, M2AN 2015.

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.

A. Cohen and G. Migliorati, *Multivariate approximation in downward closed polynomial spaces*, in Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan, , Springer, 2018.

B. Arras, M. Bachmayr and A. Cohen, *Sequential sampling for optimal weighted least squares approximations in hierarchical spaces*, preprint 2018.

General context

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 $s = \frac{1}{q}$.

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

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

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 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.

Implementation of the least-squares method

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

$$u_n^{\rm LS} = \sum_{j=1}^n c_j L_j.$$

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.

When the y^i are random, then $\mathbf{G}$ is a random matrix and $u_n^{\rm LS}$ is a random function.

If $L_1, \dots, L_n$ is an orthonormal basis of V_n for the $L^2(Y, \rho)$ norm, then $\mathbb{E}(\mathbf{G}) = \mathbf{I}$.

Instance optimality of the least-square approximation

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

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

Equivalence with the continuous $L^2(Y,\rho)$ norm : the random Grammian

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

satisfies $\mathbb{E}(\mathbf{G}) = \mathbf{I}$. In addition,

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

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.

By convention, we set $u_n^{\rm LS}=0$ in the event where $\|\mathbf{G}-I\| \geq \frac{1}{2}$.

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

$$K_{n,w} \leq c \frac{m}{\log(2n/\varepsilon)}, \quad c := \frac{2\log(3/2) - 1}{2},$$

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 $K_{n,w} \lesssim \frac{m}{\log n}$, which imposes at least the regime $m \gtrsim n \log n$, 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 $\|\mathbf{X}\| \leq K$ 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 $c(\delta) := (1+\delta)\log(1+\delta) - \delta > 0$ (in particular $c(\frac{1}{2}) := c = \frac{3\log(3/2)-1}{2}$).

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

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

5. The Christoffel function and the sampling budget

The unweighted case $w = 1$

The stability regime is described by the condition $K_n = \|k_n\|_{L^\infty} \lesssim \frac{m}{\log n}$.

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 $m \gtrsim n$.

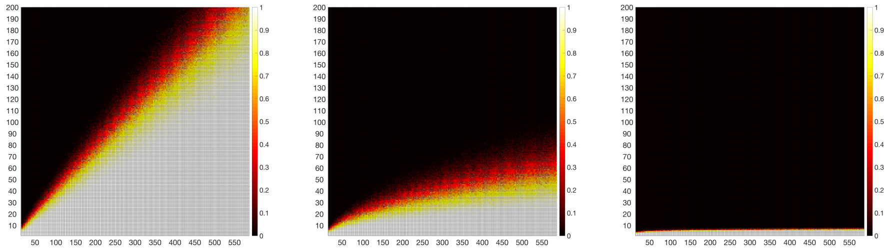
(ii) Uniform distribution $\rho = \frac{dy}{2}$: the L_j are normalized Legendre polynomials and $K_n = \sum_{j=1}^n (2j - 1) = n^2$. Up to log factors, the stability regime is $m \gtrsim n^2$.

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 $m/\log(m)$ 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 $\|L_j\|_{L^\infty} = \mathcal{O}(1)$.

Application to acoustic sampling

The unknown function u satisfies the Helmholtz equation

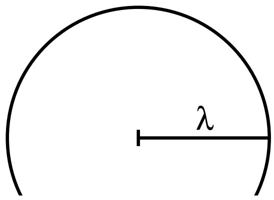
$$\Delta u + \lambda^2 u = 0,$$

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 $\mathbb{R}^2$.



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

$$K_n \sim n^2,$$

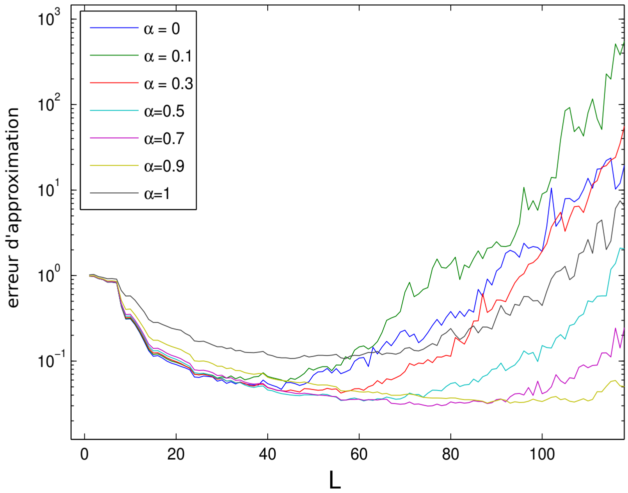
if $\rho = \frac{dy}{|Y|}$ is the uniform measure over Y, and

$$K_n \sim n,$$

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 $(= n = \dim(V_n))$



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

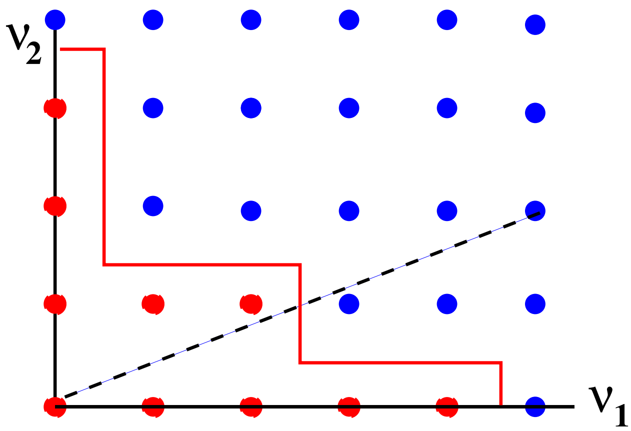
$$V_\Lambda := \left\{ \sum_{\nu\in\Lambda} v_\nu y^\nu, \quad v_\nu \in V \right\} = V \otimes \mathbb{P}_\Lambda,$$

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 $\nu \in \Lambda$.

Downward closed multivariate polynomials



Breaking the curse of dimensionality

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 any tensorized Jacobi measures. 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 distribution 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 $\alpha := \frac{\log 3}{\log 2}$.

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 $Y = \mathbb{R}^d$.

6. Weighted least-squares methods and optimal sampling

The optimal measure

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

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

The quantity $K_{n,w}$ is minimized by the choice

$$d\sigma(y) = \frac{\sum_{j=1}^{n} |L_j(y)|^2}{n} d\rho(y) \quad \text{and} \quad w(y) = \frac{n}{\sum_{j=1}^{n} |L_j(y)|^2}$$

which yields

$$K_{n,w} = n.$$

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

We thus obtain instance optimality with an optimal sampling budget.

Note that $\sigma = \sigma_n = \sigma(V_n, \rho)$ changes with n : issue for progressivity.

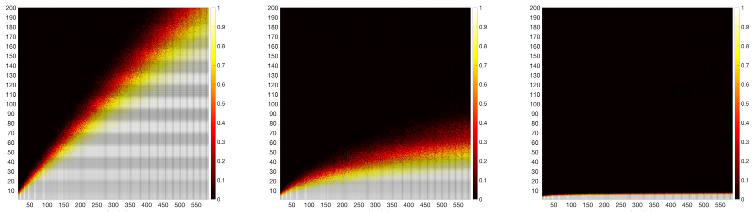
Sampling according to $d\sigma_n$ can be non-trivial, especially in high dimension.

Illustration

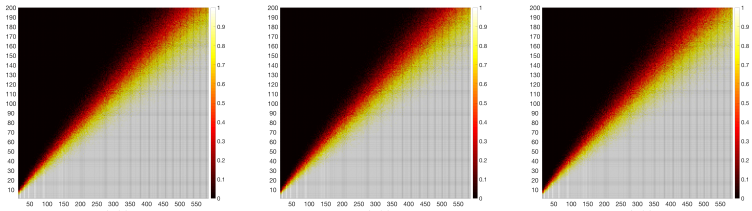
We take $V_n = \mathbb{P}_n$ univariate polynomials of degree n on $Y = [-1, 1]$

Plot : $\Pr(\kappa(\mathbf{G}) \leq 3)$ (white if 1, black if 0) with $m/\log(m)$ on x axis, n on y axis.

Left : $d\rho = \frac{dy}{\pi\sqrt{1-y^2}}$. Center : $d\rho = \frac{dy}{2}$. Right : $d\rho = (2\pi)^{-1/2}\exp(-y^2/2)$ on $Y = \mathbb{R}$.



Unweighted case : sampling budget $m \gtrsim n$, $m \gtrsim n^2$, $m \gtrsim \exp(n)$.



Optimal weighting : sampling budget $m \gtrsim n$.

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.

Pseudo-spectral methods

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 $l = 0, 1, \dots, L$ 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 $e_n(u) \leq Cn^{-s}$ and taking $\theta > 2^{2s}$ we retrieve rate optimality.

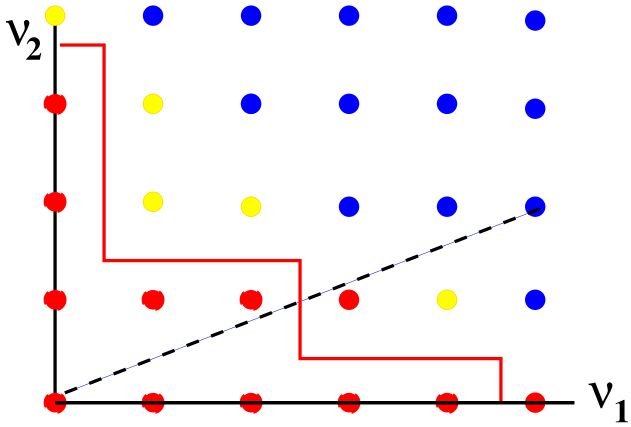
The sampling budget is optimal : $m_0 + \cdots + m_L \leq 2\theta n_L$.

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

General defect : dimension n_l grows geometrically.

Adaptivity

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

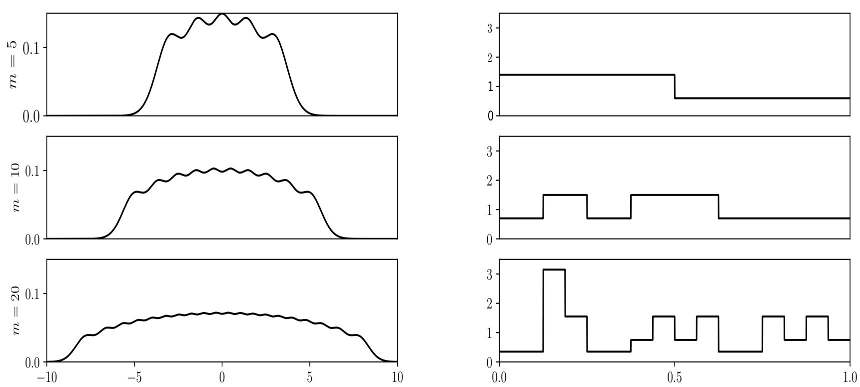


Problem : the optimal measure $\sigma = \sigma_n$ changes as we vary n. How should we recycle the previous samples ?

For certain simple cases $\sigma_n \sim \sigma^*$ as $n \to \infty$ (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 $0, \dots, m-1$ 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 $\{\frac{1}{n}, 1-\frac{1}{n}\}$.

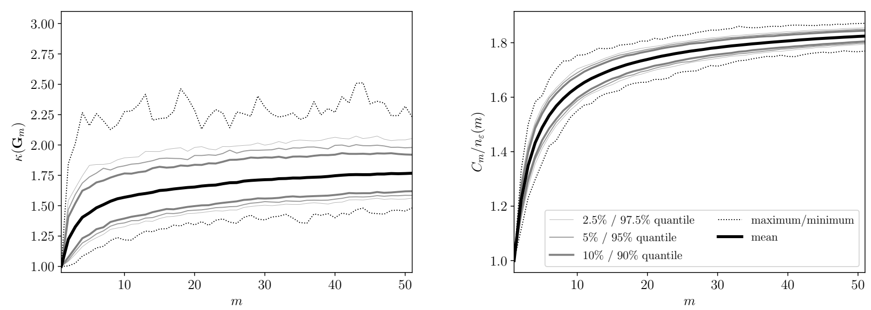
If $b_i = 0$, generate y^i according to $d\nu_n$.

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

Optimality of the sequencial sampling algorithm

Arras-Bachmayr-Cohen-Migliorati (2018) : the total number of sample C_n used at stage n satisfies $\mathbb{E}(C_n) \sim n\log(n)$ and $C_n \lesssim n\log(n)$ 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 (Giovanni 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.

Cases where the L_j and σ_n are not easily computable, e.g. for a general domain Y.

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

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