Optimal sampling in least-squares methods
Theory and practice

Albert Cohen

Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris

Moscow online school 05-05-2021

Agenda

1. Recovery in linear spaces : applicative settings and objectives.
2. Weighted least-squares methods : first estimation bounds.
3. Optimal sampling measure : theory and practical aspects.
4. Sparsification : towards optimal sampling budget ?
5. More general measurements : inverse problems and PDEs.

An ubiquitous numerical problem

Reconstruct an unknown multivariate function

$$u : x \mapsto u(x), \quad x = (x_1, \dots, x_d) \in D \subset \mathbb{R}^d,$$

from (noisy) observations $y^i \approx u(x^i)$ at sample points $x^i \in D$ for $i = 1, \dots, m$.

Distinction between two data acquisition settings :

Passive setting : we do not choose the x^i.

Active setting : we choose the x^i.

How should we sample ? How should we reconstruct ?

Passive aquisition setting

Input-output modeled by $(x, y) \in D \times \mathbb{R}$ is a random variable of unknown joint law.

We observe independant realizations (x^i, y^i) for $i = 1, \dots, m$. We search for a function that best explains y from x.

Applicative context : regression, machine learning, denoising...

The quadratic risk $\mathbb{E}(|y - v(x)|^2)$ is minimized among all functions v by $u(x) := \mathbb{E}(y|x)$ which is unknown.

For $\tilde{u} \neq u$, one has

$$\mathbb{E}(|y - \tilde{u}(x)|^2) = \mathbb{E}(|y - u(x)|^2) + \mathbb{E}(|\tilde{u}(x) - u(x)|^2) = \sigma^2 + \int_D |u(x) - \tilde{u}(x)|^2 d\mu,$$

where $d\mu$ is the unknown probability measure of x.

We thus measure performance of a reconstruction $\tilde{u}$ by $\|u - \tilde{u}\|_{L^2(D,\mu)}$.

Inherently noisy setting : $y^i = u(x^i) + \eta^i$, where η^i is a noise $\mathbb{E}(\eta|x) = 0$.

Active aquisition setting

We are allowed to query an unknown map $x \mapsto u(x)$, typically by running an experiment or a numerical simulation.

Each (offline) query $x^i \mapsto y^i = u(x^i)$ is costly (and could be noisy).

We want to compute an approximation map $x \mapsto \tilde{u}(x)$ that is much cheaper to evaluate (online) than u.

Applicative context : model reduction, data aquisition, inverse problems, design of computer experiments.

We measure performance by $\|u - \tilde{u}\|_{L^2(D,\mu)}$ where μ can be chosen by us, for example the Lebesgue measure.

Is there an optimal choice of the sample $(x^1, \dots, x^m)$? Easy to construct ?

We can invest some offline time designing the sample (prune from a larger sample).

When $d >> 1$ we want to avoid uniform grids (curse of dimensionality).

The function u may take its value in $\mathbb{R}$, or $\mathbb{R}^k$, or in an infinite dimensional space.

Optimal recovery

Let V be a general Banach space of functions defined on D, and let $\mathcal{K} \subset V$ a class that describes the prior information on u (for example smoothness).

We define the deterministic optimal recovery numbers

$$r_m^{\rm det}(\mathcal{K})_V := \inf_{\mathbf{x},\Phi_{\mathbf{x}}} \max_{u\in\mathcal{K}} \|u - \Phi_{\mathbf{x}}(u(x^1),\dots,u(x^m))\|_V,$$

where infimum is taken on all $\mathbf{x} = (x^1,\dots,x^m) \in D^m$ and maps $\Phi_{\mathbf{x}} : \mathbb{R}^m \to V$.

Randomized setting (random sampling) :

$$r_m^{\rm rand}(\mathcal{K})_V^2 := \inf_{\mathbf{x},\Phi_{\mathbf{x}}} \max_{u\in\mathcal{K}} \mathbb{E}_{\mathbf{x}}(\|u - \Phi_{\mathbf{x}}(u(x^1),\dots,u(x^m))\|_V^2),$$

where infimum is taken on all random variable $\mathbf{x} \in D^m$ and linear $\Phi_{\mathbf{x}} : \mathbb{R}^m \to V$.

Linear recovery : define $\rho_m^{\rm det}(\mathcal{K})_V$ and $\rho_m^{\rm rand}(\mathcal{K})_V$ similarly but with $\Phi_{\mathbf{x}}$ linear.

Obviously : $r_m^{\rm det}(\mathcal{K})_V \leq \rho_m^{\rm det}(\mathcal{K})_V$ and $r_m^{\rm rand}(\mathcal{K})_V \leq \rho_m^{\rm det}(\mathcal{K})_V$.

Also : $r_m^{\rm rand}(\mathcal{K})_V \leq r_m^{\rm det}(\mathcal{K})_V$ and $\rho_m^{\rm rand}(\mathcal{K})_V \leq \rho_m^{\rm det}(\mathcal{K})_V$.

Approximation

Error measure : $\|u - \tilde{u}\|_V$, where $V := L^2(D, \mu)$, or other Banach space of interest.

Most often, the reconstruction $\tilde{u}$ takes place within a family $V_n \subset V$ that can be parametrized by $n \leq m$ numbers.

So it is relevant to compare $\|u - \tilde{u}\|_V$ with

$$e_n(u)_V = \min_{v \in V_n} \|u - v\|_V.$$

We restrict our attention to linear families : V_n is a linear space with $n = \dim(V_n)$.

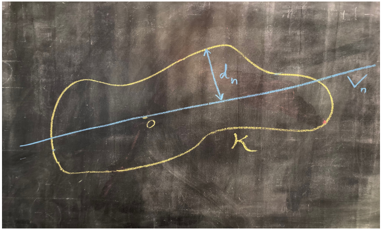
If V is a Hilbert space, $e_n(u) = \|u - P_{V_n} u\|_V$ with P_{V_n} the V-orthogonal projection.

Classical choices : algebraic polynomials, spline spaces, trigonometric polynomials, piecewise constant functions on a given partition of D.

Optimized choices : if our prior information is that $u \in \mathcal{K}$ where $\mathcal{K} \subset V$ is some compact class we are interested in spaces V_n that perform close to the Kolmogorov n-width, that is defined for a general Banach space V by

$$d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \max_{u \in \mathcal{K}} e_n(u)_V.$$

Kolmogorov n-widths



An optimal space achieving the infimum is not easy to construct.

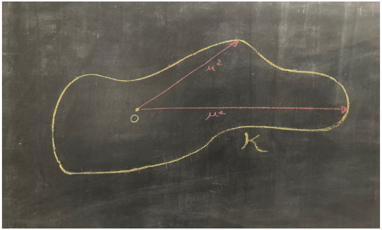
It can be emulated by reduced basis spaces $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

Greedy selection : given V_{k-1} pick next u^k such that

$$\|u - u^k\| = \max_{u \in \mathcal{K}} \|u - P_{V_{k-1}} u\|_V,$$

or in practice $\|u - u^k\| \geq \gamma \max_{u\in\mathcal{K}} \|u - P_{V_{k-1}} u\|_V$ for fixed $\gamma \in]0, 1[$.

Such algorithms have been proposed by Maday-Patera in the particular context of reduced order modeling, where the class $\mathcal{K}$ consists of solutions u to a PDE as we vary certain physical parameters (solution manifold).



An optimal space achieving the infimum is not easy to construct.

It can be emulated by reduced basis spaces $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

Greedy selection : given V_{k-1} pick next u^k such that

$$\|u - u^k\| = \max_{u \in \mathcal{K}} \|u - P_{V_{k-1}} u\|_V,$$

or in practice $\|u - u^k\| \geq \gamma \max_{u \in \mathcal{K}} \|u - P_{V_{k-1}} u\|_V$ for fixed $\gamma \in]0, 1[$.

Such algorithms have been proposed by Maday-Patera in the particular context of reduced order modeling, where the class $\mathcal{K}$ consists of solutions u to a PDE as we vary certain physical parameters (solution manifold).

Approximation performances

For the greedily generated spaces V_n, we would like to compare

$$\sigma_n(\mathcal{K})_V = \mathrm{dist}(\mathcal{K}, V_n)_V = \max_{u\in\mathcal{K}} \|u - P_{V_n} u\|_V,$$

with the n-widths $d_n(\mathcal{K})_V$ that correspond to the optimal spaces.

Direct comparison is deceiving.

Buffa-Maday-Patera-Turinici (2010) : $\sigma_n \leq n2^n d_n$.

For all $n \geq 0$ and $\varepsilon > 0$, there exists $\mathcal{K}$ such that $\sigma_n(\mathcal{K})_V \geq (1-\varepsilon)2^n d_n(\mathcal{K})_V$.

Comparison is much more favorable in terms of convergence rate.

Binev-Cohen-Dahmen-DeVore-Petrova-Wojtaszczyk (2013) : For any $s > 0$,

$$\sup_{n\geq 1} n^s d_n(\mathcal{K})_V < \infty \Rightarrow \sup_{n\geq 1} n^s \sigma_n(\mathcal{K})_V < \infty,$$

and

$$\sup_{n\geq 1} e^{cn^s} d_n(\mathcal{K})_V < \infty \Rightarrow \sup_{n\geq 1} e^{\tilde{c}n^s} \sigma_n(\mathcal{K})_V < \infty,$$

Nonlinear approximation

Approximation in linear spaces is known to be no so effective for several relevant model classes $\mathcal{K}$ in Banach spaces V : poor decay of $d_n(\mathcal{K})_V$.

Improved performance can be achieved by nonlinear approximation methods : the function u is approximated by simpler functions $v \in \Sigma_n$ that can be described by $\mathcal{O}(n)$ parameters, however Σ_n is not a linear space.

1. Rational fractions : $\Sigma_n = \left\{ \frac{p}{q} \, ; \, p, q \in \mathbb{P}_n \right\}$.

2. Neural networks : functions $v : \mathbb{R}^d \to \mathbb{R}^m$ of the form

$$v = A_k \circ \sigma \circ A_{k-1} \circ \sigma \circ A_{k-2} \circ \cdots \circ \sigma \circ A_1,$$

where $A_j : \mathbb{R}^{d_j} \to \mathbb{R}^{d_{j+1}}$ is affine and σ is a nonlinear (rectifier) function applied componentwise, for example $\sigma(x) = RELU(x) = \max\{x, 0\}$. Here Σ_n is the set of such functions when the total number of parameters does not exceed n.

3. Best n-term / sparse approximation in a basis $(e_k)_{k \geq 1}$: pick approximation from the set $\Sigma_n = \{\sum_{k \in E} c_k e_k \ : \ \#(E) \leq n\}$.

4. Piecewise polynomials, splines, finite elements on meshes generated after n steps of adaptive refinement (select and split an element in the current partition).

Example 3 and 4 : adaptively generated linear spaces $V_1 \subset V_2 \subset \cdots \subset V_n \ldots$

General objectives

Ideally we would like to combine

Instance optimality : achieve $\|u - \tilde{u}\|_V \leq Ce_n(u)_V$ for any u, for some fixed C.

Budget optimality : use $m \sim n$ samples (up to log factors).

Progressivity : when using $V_1 \subset V_2 \subset \dots V_n$ cumulated budget stays $m \sim n$.

In recent years, significant progresses have been made on randomized sampling and least-squares reconstruction strategies from various angles, allowing to reach the above (and other related) objectives.

Information based complexity : Wozniakowski, Wasilkowski, Kuo, Krieg, M. Ullrich, Kämmerer, Volkmer, Potts, T. Ullrich, Oettershagen, ...

Uncertainty quantification and model reduction : Doostan, Hampton, Narayan, Jakeman, Zhou, Nobile, Tempone, Chkifa, Webster, Harberstisch, Nouy, Perrin...

Approximation theory : Cohen, Davenport, Leviatan, Migliorati, Bachmayr, Arras, Adcock, Huybrechs, Temlyakov...

A simple example : interpolation by univariate polynomials

Consider $D = [-1, 1]$ and $V = \mathcal{C}(D)$ equipped with the max norm $\|\cdot\|_V = \|\cdot\|_{L^\infty}$.

Take $V_n = \mathbb{P}_{n-1}$ univariate polynomials of degree $n-1$.

With $(x^1, \dots, x^n) \in [-1, 1]$ pairwise distincts, reconstruct by the interpolation operator

$$\tilde{u} = I_n u \in \mathbb{P}_{n-1}, \quad s.t. \quad I_n u(x^i) = u(x^i), \quad i = 1, \dots, n.$$

Budget is optimal : $m = n$ points have been used.

Instance optimality : governed by Lebesgue constant $C_n = \max_{u \neq 0} \frac{\|I_n u\|_{L^\infty}}{\|u\|_{L^\infty}}$, since

$$\|u - I_n u\|_{L^\infty} \leq \|u - v\|_{L^\infty} + \|I_n v - I_n u\|_{L^\infty} \leq (1 + C_n)\|u - v\|_{L^\infty}, \quad v \in V_n,$$

thus bounded by $(1 + C_n) e_n(u)_{L^\infty}$.

Equispaced points are known to yield $C_n \sim 2^n$.

Chebychev points $\left\{ \cos\left(\frac{2k\pi}{2n+1}\right) \ : \ k = 1, \dots, n \right\}$ yield optimal value $C_n \sim \ln(n)$.

Limitations

Multivariate case : no general theory for optimal points on a general domain $D \subset \mathbb{R}^d$.

What about other types of spaces V_n ?

Fekete points : if V_n is a linear space with basis $(\phi_1, \dots, \phi_n)$, then the points

$$(x^1, \dots, x^n) = \text{argmax}\Big\{ \det(\phi_i(z_j))_{i,j=1,\dots,n} \; : \; (z_1, \dots, z_n) \in D^n \Big\},$$

yields $C_n \leq n$ but are not simply computable : non-convex optimization in $\mathbb{R}^{dn}$.

For univariate polynomials these points maximizes $\prod_{j\neq i} |x^i - x^j|$.

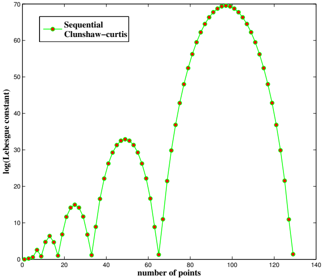
Progessivity : the Chebychev and Fekete points are not nested as $n \to n+1$!

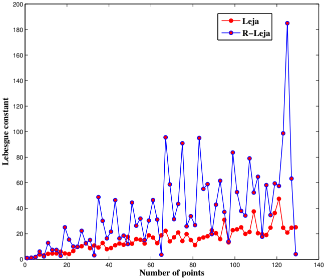
The Clenshaw-Curtis points $G_n = \Big\{ \cos\Big(\frac{k\pi}{n-1}\Big) \; : \; k = 0, \dots, n-1 \Big\}$ are partially nested :

$$G_3 \subset G_5 \subset G_9 \subset \cdots \subset G_{2^j+1} \subset G_{2^{j+1}+1} \subset \cdots$$

How to fill-in by intermediate points and preserve a well-behaved Lebesgue constant ?

Lebesgue constant for nested sets





Left : fill-in by increasing order.

Right (blue) : fill-in by Van der Corput enumeration $C_n \leq n^2$ (Chkifa, 2013).

Right (red) : greedy Fekete (Leja) $\max \prod_{j=1}^{k-1} |x - x^j| \to x^k$. Open problem : $C_n \sim n$?

The behaviour $C_n \sim \ln(n)$ does not seem achievable with nested sets.

From now on, $V = L^2(D,\mu)$. Notation : $\|v\| = \|v\|_{L^2(D,\mu)}$, and $e_n(u) = \|u - P_{V_n}u\|$.

The $L^2(D,\mu)$-projection

$$P_{V_n}u := \mathrm{argmin}\Big\{\int_D |u(x) - v(x)|^2 d\mu \ : \ v \in V_n\Big\},$$

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

$$\int_D v(x)d\mu \approx \frac{1}{m}\sum_{i=1}^m w(x^i)v(x^i).$$

where w is a weight function. This is the (weighted) least-squares method

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

In the noiseless case $y^i = u(x^i)$, the solution is the orthogonal projection of u onto V_n for the discrete (semi)-norm

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

that should in some sense be close to $\|v\|^2$.

Randomized sampling

Draw $(x^1, \dots, x^m)$ i.i.d. according to a sampling probability measure σ.

Use a weight w such that

$$w(x)d\sigma(x) = d\mu(x).$$

The random norm $\|v\|_m^2 = \frac{1}{m}\sum_{i=1}^m w(x^i)|v(x^i)|^2$ then satisfies, for any function v,

$$\mathbb{E}\left(\|v\|_m^2\right) = \mathbb{E}_\sigma(w(x)|v(x)|^2) = \int_D w(x)|v(x)|^2 d\sigma = \int_D |v(x)|^2 d\mu = \|v\|^2.$$

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

Optimality results will be achieved by appropriate choices of w and σ.

The weighted least-squares approximation u_n is now a random object. Its accuracy should be studied in some probabilistic sense, for instance $\mathbb{E}(\|u - u_n\|^2)$.

General strategy : study the probabilistic event E_δ of the equivalence

$$(1-\delta)\|v\|^2 \leq \|v\|_m^2 \leq (1+\delta)\|v\|^2, \quad v \in V_n,$$

for some $0 < \delta < 1$, for example $\delta = \frac{1}{2}$.

This is an instance ($p = 2$ and $w_i = m^{-1}w(x^i)$) of a Marcinkiewicz-Zygmund inequality :

$$(1-\delta)\int_D |v(x)|^p d\mu \leq \sum_{i=1}^m w_i |v(x^i)|^p \leq (1-\delta)\int_D |v(x)|^p d\mu, \quad v \in V_n.$$

Let $(L_1, \dots, L_n)$ be an $L^2(D,\mu)$-orthonormal basis of V_n and consider the random Gramian matrix

$$\mathbf{G} = (G_{k,j})_{k,j=1,\dots,n}, \quad G_{k,j} := \frac{1}{m}\sum_{i=1}^m w(x^i)L_k(x^i)L_j(x^i) = \langle L_k, L_j\rangle_m.$$

Then

$$E_\delta \iff (1-\delta)\mathbf{I} \leq \mathbf{G} \leq (1+\delta)\mathbf{I} \iff \|\mathbf{G}-\mathbf{I}\|_2 \leq \delta.$$

Note that $\mathbf{G} = \frac{1}{m}\sum_{j=1}^m \mathbf{X}^i$, where $\mathbf{X}^i$ are i.i.d. realizations of

$$\mathbf{X} = (w(x)L_k(x)L_j(x))_{k,j}, \qquad x \sim \sigma, \quad \text{so} \quad \mathbb{E}(\mathbf{G}) = \mathbf{I}$$

A first accuracy bound

Under the event $E_{1/2}$, one has $\frac{1}{2}\|v\|^2 \leq \|v\|_m^2 \leq \frac{3}{2}\|v\|^2$ for all $v \in V_n$, and so

$$\|u - u_n\|^2 = e_n(u)^2 + \|P_n u - u_n\|^2 \leq e_n(u)^2 + 2\|P_n u - u_n\|_m^2.$$

In addition $\|u - u_n\|_m^2 = \|P_n u - u_n\|_m^2 + \|P_n u - u\|_m^2$, and so

$$\|u - u_n\|^2 \leq e_n(u)^2 + 2\|u - P_n u\|_m^2.$$

Since $\mathbb{E}(\|u - P_n u\|_m^2) = e_n(u)^2$, we reach

$$\mathbb{E}(\|u - u_n\|^2 \chi_{E_{1/2}}) \leq 3e_n(u)^2.$$

We can test the validity of $E_{1/2}$ by checking if $\|\mathbf{G} - \mathbf{I}\|_2 \leq \frac{1}{2}$.

First choice : define $\tilde{u} = u_n$ if $E_{1/2}$ holds and $\tilde{u} = 0$ gives the estimate

$$\mathbb{E}(\|u - \tilde{u}\|^2) \leq 3e_n(u)^2 + \delta\|u\|^2, \quad \delta := \Pr(E_{1/2}^c).$$

Is δ small with $m \sim n$?

Key tools : Christoffel functions and matrix concentration.

Haberstisch-Nouy-Perrin (2019) : redraw $\{x^1, \dots, x^m\}$ until $E_{1/2}$ holds and take $\tilde{u} = u_n$

If $\delta = \Pr(E^c_{1/2})$ then the number of needed redraw k^* follows a Poisson law : one has $k^* > k$ with probability δ^k and $\mathbb{E}(k^*) = \frac{1}{1-\delta}$.

The resulting sample $x^1, \dots, x^m$ follows the law $\otimes^m \sigma$ conditionned to $E_{1/2}$ and therefore, by Bayes rule

$$\mathbb{E}(\|u - \tilde{u}\|^2) = \mathbb{E}(\|u - u_n\|^2 \,|\, E_{1/2}) = \Pr(E_{1/2})^{-1}\mathbb{E}(\|u - u_n\|^2 \chi_{E_{1/2}}),$$

which gives for all $u \in V$ (non uniform result : first fix u, then draw sample),

$$\mathbb{E}(\|u - \tilde{u}\|^2) \leq C e_n(u)^2, \quad C := \frac{3}{1-\delta}.$$

Assume V_n contains constants and that $M := \mu(D) = \int |1|^2 d\mu < \infty$. Then under $E_{1/2}$, we have $\frac{1}{m}\sum_{i=1}^m w(x^i) = \|1\|_m^2 \leq \frac{3M}{2}$, so both $\|\cdot\|$ and $\|\cdot\|_m$ dominated by $\|\cdot\|_{L^\infty}$.

Therefore, for the boosted sample $x^1, \dots, x^m$, we are ensured that for all $u \in \mathcal{C}(D)$,

$$\|u - u_n\| \leq \|u-v\| + \|v - u_n\|_m \leq \|u-v\| + \|u-v\|_m \leq C\|u-v\|_{L^\infty}, \quad C := \sqrt{M}(1 + \sqrt{3/2}),$$

and therefore (uniform result : first fix a deterministic sample, then pick any u)

$$\|u - \tilde{u}\| \leq C e_n(u)_{L^\infty}.$$

With $L_1, \dots, L_n$ an $L^2(D,\mu)$-orthonormal basis of V_n, define

$$k_n(x) := \sum_{j=1}^{n} |L_j(x)|^2,$$

the inverse of the Christoffel function, also defined as

$$k_n(x) = \max_{v \in V_n} \frac{|v(x)|^2}{\|v\|^2}.$$

We use the notation

$$K_n := \|k_n\|_{L^\infty} := \sup_{x \in D} \sum_{j=1}^{n} |L_j(x)|^2 = \max_{v \in V_n} \frac{\|v\|_{L^\infty}^2}{\|v\|^2}.$$

These quantities only depends on V_n and μ.

For the given weight w, we introduce

$$k_{n,w}(x) := w(x) k_n(x),$$

and $K_{n,w} := \|k_{n,w}\|_{L^\infty}$, which only depends on (V_n, μ, w).

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

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

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : let $\mathbf{G} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{X}^i$ where $\mathbf{X}^i$ are i.i.d. copies of an $n \times n$ symmetric matrix $\mathbf{X}$ such that $\mathbb{E}(\mathbf{X}) = \mathbf{I}$ and $\|\mathbf{X}\| \leq K$ a.s. Then

$$\Pr\left\{\|\mathbf{G}-\mathbf{I}\| \geq \delta\right\} \leq 2n\exp\Big(-\frac{mc_\delta}{K}\Big),$$

where $c_\delta := (1+\delta)\ln(1+\delta) - \delta > 0$.

In our case of interest,

$$\mathbf{X} = w(x)(L_k(x)L_j(x))_{j,k=1,\dots,n} = \mathbf{x}\mathbf{x}^T, \quad \mathbf{x} = (w(x)^{1/2}L_k(x))_{k=1,\dots,n},$$

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

$$K = \sup\|\mathbf{X}\| = \sup|\mathbf{x}|^2 = \sup_{x\in D} w(x)\sum_{j=1}^{n}|L_j(x)|^2 = K_{n,w}.$$

This gives the sampling budget condition

$$m \geq cK_{n,w}\ln(2n/\varepsilon) \implies \Pr(E_{1/2}^c) = \Pr\left\{\|G-I\| \geq \frac{1}{2}\right\} \leq \varepsilon,$$

with $c = c_{1/2}^{-1} \leq 10$. For the boosted sample, take $\varepsilon = \frac{1}{2}$, and so $m \geq 10K_{n,w}\ln(4n)$.

Optimal estimation and sampling budget

Using the boosted sample, we achieve near optimal non-uniform estimate

$$\mathbb{E}(\|u-\tilde{u}\|^2) \leq Ce_n(u)^2$$

as well as uniform estimate (assuming $\mu(D) < \infty$ and $\frac{1}{m}\sum_{i=1}^m w(x^i) < \infty$)

$$\|u-\tilde{u}\| \leq Ce_n(u)_{L^\infty}$$

under a sampling budget $m \sim K_{n,w} \geq n$ up to multiplicative logarithmic factor.

In the presence of noise of variance $\kappa(x)^2$, the estimation bound has an additional term

$$e_n(u)^2 + \frac{n}{m}\kappa^2, \qquad \kappa^2 = \int_D |\kappa(x)|^2 d\mu.$$

Unweighted least-squares : $w = 1$ and $\sigma = \mu$ requires $m \sim K_n = \max_{x\in D} \sum_{j=1}^n |L_j(x)|^2$

Sometimes $K_n >> n$. leading to an excessive sampling budget.

Illustration on univariate polynomials $V_n = \mathbb{P}_{n-1}$

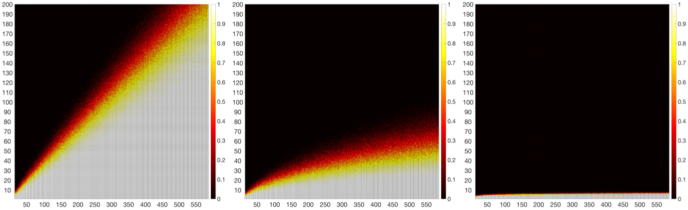
Regime of stability : probability that $\|\mathbf{G} - \mathbf{I}\| \leq \frac{1}{2}$, white if 1, black if 0.

Unweighted case requires at least $m \sim K_n$.

Left : $D = [-1, 1]$ with $d\mu = \frac{dx}{\pi\sqrt{1-x^2}}$ (Chebychev polynomials $K_n = 2n + 1 \sim n$).

Center : $D = [-1, 1]$ with $d\mu = \frac{dx}{2}$ (Legendre polynomials $K_n = n^2$)

Right : $D = \mathbb{R}$ with $d\mu = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ (Hermite polynomials $K_n = \infty$).



For the gaussian case, a more ad-hoc analysis shows that stability holds if $m \gtrsim \exp(cn)$

Narayan-Jakeman (2015), Doostan-Hampton (2015), Cohen-Migliorati (2017) : use sampling measure

$$d\sigma := \frac{k_n}{n} d\mu = \frac{1}{n}\Big(\sum_{j=1}^n |L_j|^2\Big) d\mu \implies w(x) = \frac{n}{k_n(x)}.$$

σ is a probability measure and we have $k_{n,w}(x) = w(x)k_n(x) = n$, thus $K_{n,w} = n$.

With this sampling strategy, optimal error bounds can be achieved with near optimal sampling budget $m \sim n$ up to logarithmic factors.

Observation by T. Ullrich (2020) : if μ has finite mass $\mu(D) = M < \infty$, one can also use $d\tilde{\sigma} := (\frac{1}{2M} + \frac{k_n}{2n}) d\mu$ ensuring both $K_{n,w} \leq 2n$ and $\frac{1}{m}\sum_{i=1}^m w(x_i) \leq 2M$.

Narayan-Jakeman (2015), Doostan-Hampton (2015), Cohen-Migliorati (2017) : use sampling measure

$$d\sigma := \frac{k_n}{n} d\mu = \frac{1}{n}\Big(\sum_{j=1}^n |L_j|^2\Big) d\mu \implies w(x) = \frac{n}{k_n(x)}.$$

σ is a probability measure and we have $k_{n,w}(x) = w(x)k_n(x) = n$, thus $K_{n,w} = n$.

With this sampling strategy, optimal error bounds can be achieved with near optimal sampling budget $m \sim n$ up to logarithmic factors.

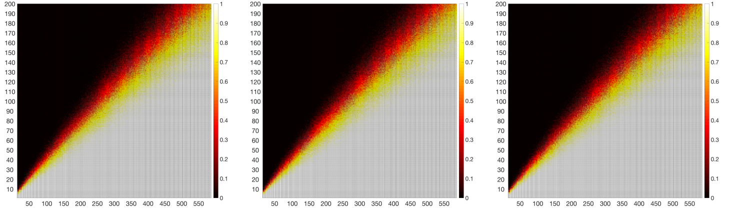
Observation by T. Ullrich (2020) : if μ has finite mass $\mu(D) = M < \infty$, one can also use $d\tilde{\sigma} := (\frac{1}{2M} + \frac{k_n}{2n}) d\mu$ ensuring both $K_{n,w} \le 2n$ and $\frac{1}{m}\sum_{i=1}^m w(x_i) \le 2M$.

Narayan-Jakeman (2015), Doostan-Hampton (2015), Cohen-Migliorati (2017) : use sampling measure

$$d\sigma := \frac{k_n}{n} d\mu = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\mu \implies w(x) = \frac{n}{k_n(x)}.$$

σ is a probability measure and we have $k_{n,w}(x) = w(x)k_n(x) = n$, thus $K_{n,w} = n$.

With this sampling strategy, optimal error bounds can be achieved with near optimal sampling budget $m \sim n$ up to logarithmic factors.



Stability regime for univariate polynomials with μ Chebychev, uniform, and Gaussian.

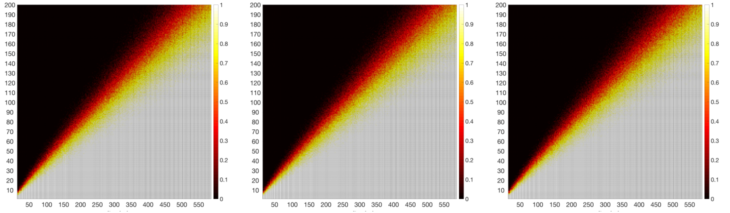
Observation by T. Ullrich (2020) : if μ has finite mass $\mu(D) = M < \infty$, one can also use $d\tilde{\sigma} := (\frac{1}{2M} + \frac{k_n}{2n})d\mu$ ensuring both $K_{n,w} \leq 2n$ and $\frac{1}{m}\sum_{i=1}^{m} w(x_i) \leq 2M$.

Narayan-Jakeman (2015), Doostan-Hampton (2015), Cohen-Migliorati (2017) : use sampling measure

$$d\sigma := \frac{k_n}{n} d\mu = \frac{1}{n}\Big(\sum_{j=1}^{n} |L_j|^2\Big) d\mu \implies w(x) = \frac{n}{k_n(x)}.$$

σ is a probability measure and we have $k_{n,w}(x) = w(x)k_n(x) = n$, thus $K_{n,w} = n$.

With this sampling strategy, optimal error bounds can be achieved with near optimal sampling budget $m \sim n$ up to logarithmic factors.



Stability regime for univariate polynomials with μ Chebychev, uniform, and Gaussian.

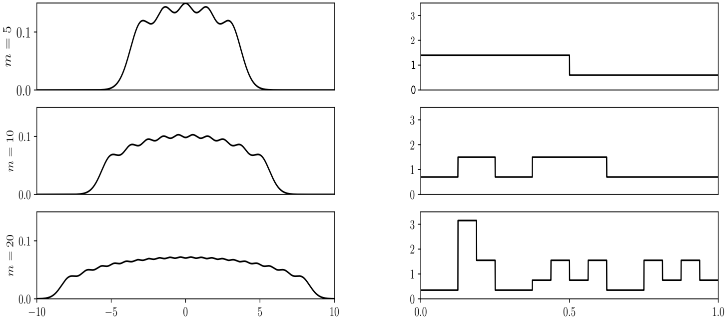
Observation by T. Ullrich (2020) : if μ has finite mass $\mu(D) = M < \infty$, one can also use $d\tilde{\sigma} := (\frac{1}{2M} + \frac{k_n}{2n})d\mu$ ensuring both $K_{n,w} \leq 2n$ and $\frac{1}{m}\sum_{i=1}^{m} w(x_i) \leq 2M$.

The optimal density is not fixed

When using a sequence $(V_n)_{n\geq 1}$ of approximation spaces

$$d\sigma = d\sigma_n := \frac{k_n}{n} d\mu.$$

Illustration : sampling densities σ_n for $n = 5, 10, 20$.



Left : Polynomials of degrees $0, \ldots, m-1$ and μ Gaussian.

Right : Piecewise constant functions on locally refined partitions and μ uniform.

Consider the space $V_n = \mathbb{P}_k$ of polynomials of total degree k on a multivariate domain $D \subset \mathbb{R}^d$, so that

$$n = \binom{k+d}{d}$$

and use the uniform probability measure $d\mu = |D|^{-1}dx$.

The local behaviour of k_n and thus of σ_n depends on closeness to the boundary of D and on the smoothness of this boundary.

Cohen-Dolbeault (2020) : For smooth domains $k_n(x) = \mathcal{O}(n^{\frac{d+1}{d}})$ on boundary, for Lipschitz domains $k_n(x) = \mathcal{O}(n^2)$ on exiting corners, for domains with cusps $k_n(x) = \mathcal{O}(n^r)$ at exiting cusps where r depends on the order of cuspitality.

Consider the space $V_n = \mathbb{P}_k$ of polynomials of total degree k on a multivariate domain $D \subset \mathbb{R}^d$, so that

$$n = \binom{k+d}{d}$$

and use the uniform probability measure $d\mu = |D|^{-1}dx$.

The local behaviour of k_n and thus of σ_n depends on closeness to the boundary of D and on the smoothness of this boundary.

Cohen-Dolbeault (2020) : For smooth domains $k_n(x) = \mathcal{O}(n^{\frac{d+1}{d}})$ on boundary, for Lipschitz domains $k_n(x) = \mathcal{O}(n^2)$ on exiting corners, for domains with cusps $k_n(x) = \mathcal{O}(n^r)$ at exiting cusps where r depends on the order of cuspitality.

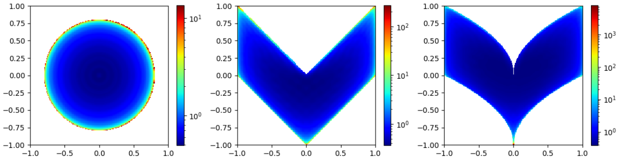
Consider the space $V_n = \mathbb{P}_k$ of polynomials of total degree k on a multivariate domain $D \subset \mathbb{R}^d$, so that

$$n = \binom{k+d}{d}$$

and use the uniform probability measure $d\mu = |D|^{-1}dx$.

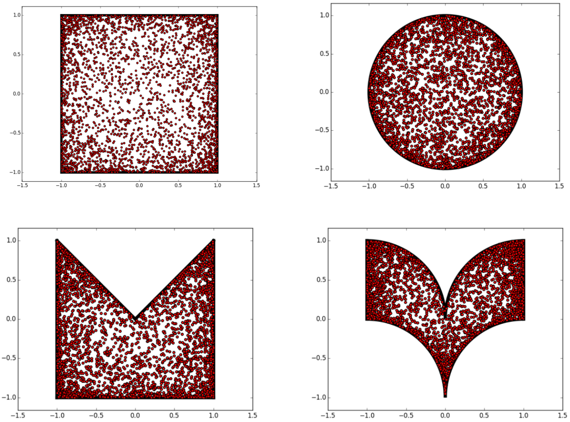
The local behaviour of k_n and thus of σ_n depends on closeness to the boundary of D and on the smoothness of this boundary.

Cohen-Dolbeault (2020) : For smooth domains $k_n(x) = \mathcal{O}(n^{\frac{d+1}{d}})$ on boundary, for Lipschitz domains $k_n(x) = \mathcal{O}(n^2)$ on exiting corners, for domains with cusps $k_n(x) = \mathcal{O}(n^r)$ at exiting cusps where r depends on the order of cuspitality.



Inverse Christoffel function $k_n(x)$ for $n = 231$ (total degree $k = 20$)

Examples of draw according to optimal sample distribution



Problem : generate efficiently i.i.d. samples according to the optimal sampling measure

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

This problem might be non-trivial in a multivariate setting $D \subset \mathbb{R}^d$.

In many relevant instances μ is a product measure (such as uniform, gaussian) and thus easy to sample, but $d\sigma_n$ is not. Sampling strategies :

(i) Rejection sampling : draw x^i according to μ and a uniform random variable z^i in $[0, M]$ where $M \geq \frac{\|k_n\|_{L^\infty}}{n}$. Reject x^i if $z^i > \frac{k_n(x^i)}{n}$.

(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...

Strategies (ii) is more efficient in cases where the L_j have tensor product structure.

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

Migliorati (2018) : one can also split the sample into n batches of size $\mathcal{O}(\ln(n))$ each of them sampled according to $d\nu_j = |L_j|^2 d\mu$, with same final estimation bounds.

Optimal sampling may become unfeasible when $D \subset \mathbb{R}^d$ 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_D is easily computable $\Rightarrow$ sampling according to the uniform measure μ is easy (sample uniformly on a bounding box, reject if $x \notin D$).

Migliorati, Adcock-Cadenas (2019), Cohen-Dolbeault (2020) : two-step strategies

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

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

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

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

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

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

Sequencial sampling

For a given hierarchy $V_1 \subset V_2 \subset \cdots \subset V_n$, note that

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

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

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

Then we generate a new sample $S_n = \{x^1, \dots, x^{m(n)}\}$ 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 new x^i according to $d\nu_n$.

If $b_i = 1$, recycle x^i incrementally from S_{n-1}.

Arras-Bachmayr-Cohen (2018) : the cumulated number of sample C_n used at stage n satisfies $C_n \sim n$ up to logarithmic factors with high probability for all values of n.

With high probability, the matrix $\mathbf{G}$ satisfies $\|\mathbf{G} - \mathbf{I}\| \leq \frac{1}{2}$ for all values of n.

Adaptive selection strategies ? See the lecture by Giovanni Migliorati.

Sparsification

Reducing further sampling budget to $\mathcal{O}(n)$: logarithmic factors removable ?

Batson-Spielman-Srivastava (2014) : let $\mathbf{x}_1, \dots, \mathbf{x}_m$ be $m \geq n$ be vectors of $\mathbb{R}^n$ such that

$$(1-\delta)\mathbf{I} \leq \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^T \leq (1+\delta)\mathbf{I}.$$

For any $c \geq 2$ there exists $S \subset \{1, \dots, m\}$ with $\#(S) \leq cn$ and weights s_i such that

$$\left(1 - \frac{1}{\sqrt{c}}\right)^2 (1-\delta)\mathbf{I} \leq \sum_{i \in S} s_i \mathbf{x}_i \mathbf{x}_i^T \leq (1+\delta)\left(1 + \frac{1}{\sqrt{c}}\right)^2 \mathbf{I}$$

Apply this to $\mathbf{x}_i = \left(\sqrt{\frac{w(x^i)}{m}} L_j(x^i)\right)_{j=1,\dots m}$ with $\{x^1, \dots, x^m\}$ a boosted sample.

Leads to a sample $(x^1, \dots, x^{2n})$ and weights $w_i = s_i \frac{w(x^i)}{m}$ such that

$$\alpha \|v\|^2 \leq \|v\|_{2n}^2 \leq \beta \|v\|^2, \qquad v \in V_n,$$

where $\|v\|_{2n}^2 = \sum_{i=1}^{2n} w_i |v(x^i)|^2$ and $\alpha = \frac{1}{2}\left(1 - \frac{1}{\sqrt{2}}\right)^2$, $\beta = \frac{3}{2}\left(1 + \frac{1}{\sqrt{2}}\right)^2$.

Based on these new samples and weights, we define a weighted least-squares estimate

$$\tilde{u} := \mathrm{argmin}\Big\{\frac{1}{2n}\sum_{i=1}^{2n} w_i|u(x^i)-v(x^i)|^2\Big\}.$$

for which we have for all $u \in \mathcal{C}(D)$

$$\|u-\tilde{u}\| \leq Ce_n(u)_{L^\infty},$$

assuming that μ is a finite measure.

The sparsification strategy of Batson-Spielman-Srivastava is performed by a deterministic greedy algorithm of total complexity $\mathcal{O}(mn^3)$: additional offline cost.

Temlyakov (2019) : comparison between deterministic linear optimal recovery numbers in L^2 and Kolmogorov n-width in L^∞ for any compact class $\mathcal{K}$ of $\mathcal{C}(D)$.

By optimizing the choice of V_n, one obtains

$$\rho_{2n}^{\mathrm{det}}(\mathcal{K})_{L^2} \leq Cd_{n-1}(\mathcal{K})_{L^\infty}.$$

Other results when $\mathcal{K}$ is the ball of a RKHS : Krieg-M.Ullrich, Nagel-Schäffer-T.Ullrich

We cannot prove $\mathbb{E}(\|u-\tilde{u}\|^2) \leq Ce_n(u)^2$ with the above strategy.

We miss the averaging property $\mathbb{E}(\|v\|_{2n}^2) = \|v\|^2$ for any $v \in V$.

Marcus-Spielman-Srivastava (2015) : if $\mathbf{x}_1, \dots, \mathbf{x}_m$ are m vectors from $\mathbb{R}^n$ of norm $|\mathbf{x}_i|^2 \leq \delta$ and such that

$$\alpha \mathbf{I} \leq \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T \leq \beta \mathbf{I}$$

then there exists a partition $S_1 \cup S_2 = \{1, \dots, m\}$ such that

$$\frac{1-5\sqrt{\delta/\alpha}}{2}\alpha \mathbf{I} \leq \sum_{i \in S_j} \mathbf{x}_i \mathbf{x}_i^T \leq \frac{1+5\sqrt{\delta/\alpha}}{2}\beta \mathbf{I}, \quad j = 1, 2.$$

Nitzan-Olevskii-Ulanovskii (2016) apply this process recursively in order to identify a $J \subset \{1, \dots, m\}$ such that $|J| \leq cn$ and

$$C^{-1}\alpha \mathbf{I} \leq \sum_{i \in J} \mathbf{x}_i \mathbf{x}_i^T \leq C\beta \mathbf{I}.$$

for some universal constant $C > 1$.

Cohen-Dolbeault (2021) : if the $\mathbf{x}_i$ have equal norms $|\mathbf{x}_i|^2 = \frac{n}{m}$, then iterative splitting delivers for some $L = \mathcal{O}(\ln(m/n))$ a partition $J_1 \cup J_2 \cup \cdots \cup J_{2^L} = \{1, \dots, m\}$ such that

$$c_0 \mathbf{I} \le \sum_{i \in J_k} \mathbf{x}_i \mathbf{x}_i^T \le C_0 \mathbf{I}, \quad k = 1, \dots, 2^L,$$

with (c_0, C_0) universal constants and $|J_k| \le C_0 n$ for all k.

Apply to $\mathbf{x}_i = \left(\sqrt{\frac{w(x^i)}{m}} L_j(x^i)\right)_{j=1,\dots m}$ with $Y = \{x^1, \dots, x^m\}$ the random boosted sample with $m \ge 10 n \ln(4n)$.

Let κ be the random variable taking value $k \in \{1, \dots, 2^L\}$ with probability $p_k = \frac{|J_k|}{m}$. Define weighted least-square estimate $\tilde{u}$ with random sample $X = \{x^i \in Y \,:\, i \in J_\kappa\}$.

$$\mathbb{E}_X\left(\frac{1}{\#(X)} \sum_{x^i \in X} w(x^i)|v(x^i)|^2\right) = \mathbb{E}_Y\left(\frac{1}{m} \sum_{i=1}^m w(x^i)|v(x^i)|^2\right) \le 2\|v\|^2, \quad v \in V.$$

This allows us to prove $\mathbb{E}(\|u - \tilde{u}\|^2) \le C e_n(u)^2$, with sample size $|X| \le C_0 n$.

Consequence : for any compact $\mathcal{K} \subset L^2$,

$$\rho^{\mathrm{rand}}_{C_0 n}(\mathcal{K})_{L^2} \le C d_n(\mathcal{K})_{L^2}.$$

Summary

We can improve sparsity of the sample up to near-optimality $m \sim n$.

This comes at the prize of computational feasability of the offline sample generation.

sampling complexity	sample cardinality m	offline complexity	$\mathbb{E}(\|u - \tilde{u}\|^2) \leq Ce_n(u)^2$	$\|u - \tilde{u}\|^2 \leq Ce_n(u)^2_\infty$
conditionned $\rho^{\otimes m} \mid E$	$10n \ln(4n)$	$\mathcal{O}(n^3 \ln(n))$	✓	✓
+ deterministic sparsification	$2n$	$\mathcal{O}(n^4 \ln(n))$	✗	✓
+ randomized sparsification	$C_0 n$	$\mathcal{O}(n^{cn}) \to \mathcal{O}(n^r)$?	✓	✓

Conflict between reducing sampling budget and limiting offline computational cost.

Haberstisch-Nouy-Perrin : cheap greedy sparsification but no theoretical guarantee.

Sparsification strategies do not seem to combine well with hierarchical sampling.

More general measurement models

Can we develop a similar sampling theory for other types of measurements

$$y^i = \ell_i(u), \qquad i = 1, \dots, m,$$

where ℓ_i are linear forms of some particular type ? Examples :

- Local averages $\ell_i(u) = \int_{\mathbb{R}^d} u(x)\varphi(x - x^i)$,

- Fourier samples $\ell_i(u) = \int_{\mathbb{R}^d} u(x)\exp(-i\omega^i \cdot x)$

- Radon samples $\ell_i(u) = \int_{L^i} u(s)ds$ where L^i are lines in $\mathbb{R}^2$,...

In all these examples, the linear forms are picked in a certain dictionnary where we want to make an optimal selection.

This may be viewed as apply point evaluation after a certain transformation.

$$y^i = \ell_i(u) = Ru(x^i), \qquad x^1, \dots, x^m \in D,$$

where D is now the transformed domain. For example $D = [0, \pi[\times \mathbb{R}$ for the Radon transform on $\mathbb{R}^2$.

We assume $u \mapsto Ru$ to be a "stable" representation of u for a Hilbert space V of interest, in the sense that for a certain measure μ

$$\|u\|_V^2 = \int_D |Ru(x)|^2 d\mu = \|Ru\|_{L^2(D,\mu)}^2.$$

This is the case in all above examples.

For picking the approximation $u_n \in V_n \subset V$, we now solve

$$\min_{v \in V_n} \sum_{i=1}^m w(x^i)|y^i - Rv(x^i)|^2.$$

The optimal sampling measure on the transformed domain is again defined by

$$d\sigma = \frac{k_n}{n} d\mu, \qquad k_n(x) = \sum_{j=1}^n |L_j(x)|^2,$$

however with $\{L_1, \dots, L_n\}$ now an orthonormal basis of $W_n := R(V_n)$.

With $\{x^1, \dots, x^m\}$ picked according to this sampling measure and $m \sim n$, we retrieve

$$\mathbb{E}(\|u - u_n\|_V^2) \leq C e_n(u)_V^2, \qquad e_n(u)_V = \min_{v \in V_n} \|u - v_n\|_V.$$

Choosing the error norm

Several possible choices of (V, μ) lead to different sampling strategies.

For the Fourier transform : $V = H^s(\mathbb{R}^d) \iff d\mu(\omega) = (1 + |\omega|^{2s})d\omega$.

For the Radon transform : taking $d\mu$ the Lebesgue measure,

$$\int_D |Ru(x)|^2 d\mu = \int_R \int_0^\pi |Ru(t,\theta)|^2 dt d\theta = \int_0^\pi \int_{\mathbb{R}} |\hat{u}(te_\theta)|^2 ds d\theta \sim \int_{\mathbb{R}^2} |\omega|^{-1} |\hat{u}(\omega)|^2 d\omega.$$

This leads to a very weak error norm $V = H^{-1/2}(\mathbb{R}^2)$.

If we want to control the error in $V = L^2(\mathbb{R}^2)$, we have

$$\|u\|_V^2 \sim \int_0^\pi |R(\theta,\cdot)|^2_{H^{1/2}(\mathbb{R})} d\theta.$$

Sobolev semi-norms may be viewed as weighted L^2 norms after applying the finite difference operator : for $0 < s < 1$

$$|v|^2_{H^s(\mathbb{R})} = \int_{\mathbb{R}\times\mathbb{R}} \frac{|v(t) - v(t')|^2}{|t - t'|^{1+2s}} dt dt' = \int_{\mathbb{R}^2} |V|^2 d\mu, \quad V(t,t') = v(t) - v(t').$$

Similar definitions for $s \geq 1$ using higher-order finite differences.

Solving PDEs by least-squares minimization ?

Consider a PDE set in some physical domain D (could include time variable), in general form

$$\mathcal{R}u(x) = 0,$$

where the residual $\mathcal{R}u$ accounts for the PDE, boundary condition, intial condition...
For example $\mathcal{R}u = (f + \Delta u, (u - g)_{|\partial D})$.

Discrete least-square collocation methods : approximation $u_n \in V_n$ defined by solving

$$\min_{v \in V_n} \frac{1}{m} \sum_{i=1}^{m} w(x^i) |\mathcal{R}v(x^i)|^2.$$

Recently applied in the framework of DNN (Physics Informed Neural Networks, Karniadakis-Mishra...) with unit weights and uniformly random or QMC points x^i.

In the case of a residual of the form $\mathcal{R}u = f - \mathcal{A}u$ for some linear operator $\mathcal{A}$, our results suggest using the optimal sampling measure $d\sigma = \frac{k_n}{n} dx$ and weight $w = \frac{n}{k_n}$ where $k_n(x) = \sum_{j=1}^{n} |L_j(x)|^2$ with $\{L_1, \dots, L_n\}$ an orthonormal basis of $W_n := \mathcal{A}(V_n)$

What bothers me here : the L^2 norm of the residual $\|\mathcal{R}v\|_{L^2}$ is rarely a good way to measure of the error $u - v$. Negative smoothness (dual) norms are often more natural, for example H^{-1} in the case of the Laplace equation. But these norms cannot be simply emulated by point evaluations.

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

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

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, *Optimal weighted least-squares methods*, SMAI J. of Comp. Math. 2017.

C. Haberstich, A. Nouy, and G. Perrin, *Boosted optimal weighted least-squares*, Math. Comp. 2021.

G. Migliorati, *Adaptive approximation by optimal weighted least-squares methods*, SINUM, 2019.

A. Marcus, D. Spielman and N. Srivastava, *Interlacing families II : Mixed characteristic polynomials and the Kadison-Singer problem*, Annals of Maths., 2015.

S. Nitzan, A. Olevskii and A. Ulanovskii, *Exponential frames on unbounded sets*, Proc. of the AMS, 2016.

V. N. Temlyakov, *On optimal recovery in* L^2, 2020.

N. Nagel, M. Schäfer and T. Ullrich, *A new upper bound for sampling numbers*, 2020.

A. Cohen and M. Dolbeault, *Optimal pointwise sampling for* L^2 *approximation*, 2021.