ETICS research school
3-7 oct. 2022

High-Dimensional Approximation

Part 1: Elements of approximation theory

Anthony Nouy

Centrale Nantes, Nantes Université, Laboratoire de Mathématiques Jean Leray

High dimensional problems

Many problems of computational science, statistics and probability require the approximation, integration or optimization of functions of many variables

$$u(x_1, \dots, x_d)$$

- High dimensional PDEs (Boltzmann, Schrödinger, Black-Scholes...)
- Multiscale problems
- Parameter-dependent or stochastic equations
- Statistical learning (density estimation, classification, regression)
- Probabilistic modelling
- ...

Approximation

The goal of approximation is to replace a target function u by a simpler function (easy to evaluate and to operate with).

An approximation is searched in a set of functions X_n, where n is related to some complexity measure, typically the number of parameters.

Approximation

We distinguish

- linear approximation when X_n is a finite-dimensional linear space (polynomials, trigonometric polynomials, fixed knot splines...)

$$X_n = \{\sum_{i=1}^{n} a_i \varphi_i : a_i \in \mathbb{R}\}$$

  where the φ_i form a basis of X_n.

- nonlinear approximation when X_n is a nonlinear set (rational functions, free knot splines, n-term approximation, neural networks, tensor networks...), e.g.

$$X_n = \{\sum_{i=1}^{n} a_i \varphi_i : a_i \in \mathbb{R}, \varphi_i \in \mathcal{D}\}$$

  for n-term approximation from a dictionary of functions $\mathcal{D}$, or

$$X_n = \{g(a) : a \in \mathbb{R}^n\}$$

  with some given nonlinear map g from $\mathbb{R}^n$ to X.

Error of best approximation

For a given function u from a normed vector space X and a given subset X_n, the error of best approximation

$$e_n(u)_X := E(u, X_n)_X = \inf_{v \in X_n} \|u - v\|_X$$

quantifies the best we can expect from X_n.

Fundamental problems in approximation

For a sequence $(X_n)_{n\geq 1}$ of sets of growing complexity, called an approximation tool, we would like to address the following questions.

- (universality) Does $e_n(u)_X$ converge to 0 for all functions u in X ?
- (expressivity) For a certain class of functions in X, determine how fast $e_n(u)_X$ converges to 0, or determine the complexity $n = n(\epsilon, u)$ such that $e_n(u) \leq \epsilon$. Typically,

  $$e_n(u)_X \leq M\gamma(n)^{-1}$$

  where γ is a strictly increasing function (growth function), and

  $$n(\epsilon, u) \geq \gamma^{-1}(\epsilon/M)$$

- (approximation classes) Characterize the class of functions for which a certain convergence type is achieved, e.g.

  $$\mathcal{A}^\gamma(X, (X_n)_{n\geq 1}) = \left\{ u : \sup_{n\geq 1} \gamma(n) e_n(u)_X < +\infty \right\}$$

  for some growth function γ.

Fundamental problems in approximation

- (proximinality) Determine if for all $u \in X$, there exists an element of best approximation $u_n \in X_n$ such that

$$\|u - u_n\|_X = e_n(u)_X.$$

- (algorithm) Construct an approximation $u_n \in X_n$ such that

$$\|u - u_n\|_X \leq C e_n(u)_X$$

with C independent of n or $C(n)e_n(u) \to 0$ as $n \to \infty$.

Algorithms depend on the available information, e.g. given by linear functionals such as point evaluations (interpolation, discrete least-squares), or equations satisfied by the function (variational/Galerkin methods).

Optimal approximation for a model class

If we know that the function u belongs to some class of functions K, we would like to find an approximation tool X_n presenting a good performance, or even the optimal performance for that class.

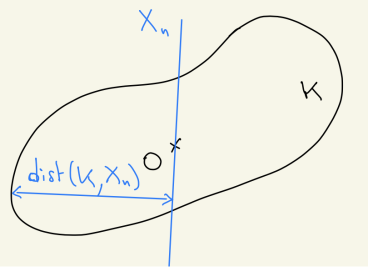
A fundamental problem is to quantify the best we can expect.

For that, we rely on different measures of complexity of K depending on the type of approximation (linear or nonlinear) and possibly on the properties of the approximation process (type of information, stability...)

Optimal linear approximation: Kolmogorov widths

For a compact subset K of a normed vector space X and a n-dimensional space X_n in X, we define the worst-case error

$$dist(K, X_n)_X = \sup_{u \in K} \inf_{v \in X_n} \|u - v\|_X$$

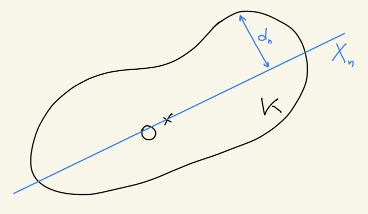


Optimal linear approximation: Kolmogorov widths

Then the Kolmogorov n-width of K is defined as

$$d_n(K)_X = \inf_{dim(X_n)=n} dist(K, X_n)_X$$

where the infimum is taken over all linear subspaces X_n of dimension n.



$d_n(K)_X$ measures how well the set K can be approximated (uniformly) by a n-dimensional space. It measures the ideal performance that we can expect from linear approximation methods.

Near to optimal spaces can be constructed by greedy algorithms (see in a next part).

Optimal linear approximation: weighted Kolmogorov widths

If K is equipped with a measure μ, a weighted Kolmogorov n-width is defined by

$$d_n^{(p,\mu)}(K)_X = \inf_{dim(X_n)=n} \left(\int_K E(u, X_n)_X^p d\mu(u) \right)^{1/p}.$$

If the measure is finite,

$$d_n^{(p,\mu)}(K)_X \leq \mu(K)^{1/p} d_n(K)_X.$$

For X a Hilbert space, $p = 2$ and μ the push-forward measure of a K-valued random variable $U \in L^2(\Omega; X)$, this is equivalent to

$$\inf_{dim(X_n)=n} \mathbb{E}(\|U - P_{X_n} U\|_X^2)^{1/2}$$

and an optimal space is given by Principal Component Analysis, that is a dominant eigenspace of the operator $v \mapsto \mathbb{E}((U, v)_X U)$ (see in a next part).

Optimal linear approximation: linear width

Another measure of complexity taking into account the approximation process is the linear width

$$a_n(K)_X = \inf_{A} \sup_{v \in K} \|v - Av\|_X$$

where the infimum is taken over all continuous linear maps $A : K \to X$ with rank at most n.

Equivalently,

$$a_n(K)_X = \inf_{g,a} \sup_{v \in K} \|v - g(a(v))\|_X$$

where both $a : K \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to X$ are linear maps.

For a Hilbert space X,

$$a_n(K)_X = d_n(K)_X$$

For a general Banach space X,

$$d_n(K)_X \leq a_n(K)_X \leq \sqrt{n} d_n(K)_X$$

Optimal performance for linear approximation from point evaluations

By restricting the information to point evaluations, the performance is characterized by sampling numbers.

For deterministic information, the worst-case optimal performance for the approximation of functions in K is measured through the (linear) sampling number

$$\rho_n(K)_X = \inf_{\boldsymbol{x}} \inf_{A} \sup_{f \in K} \|f - A(f(x_1), \ldots, f(x_n))\|_X$$

where the infimum is taken over all linear maps A and points $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, or equivalently

$$\rho_n(K)_X = \inf_{\boldsymbol{x}} \inf_{\varphi_1, \ldots, \varphi_n \in X} \sup_{f \in K} \|f - \sum_{i=1}^{n} f(x_i)\varphi_i\|_X$$

This quantifies the best we can expect from a linear algorithm using n samples for the approximation of functions in the class K.

Clearly,

$$\rho_n(K)_X \geq a_n(K)_X \geq d_n(K)_X$$

Optimal performance for linear approximation from point evaluations

For random information, the optimal performance can be measured in average mean squared error through the (linear) sampling number

$$\rho_n^{rand}(K)_X^2 = \inf_{\nu^n} \inf_{g} \sup_{f \in K} \mathbb{E}_{\boldsymbol{x} \sim \nu^n}(\|f - g(f(x_1), \dots, f(x_n))\|_X^2)$$

with an infimum taken over all measures ν^n on $\mathcal{X}^m$. Choosing for ν^n a dirac measure on an optimal deterministic set of points, we deduce that

$$d_n(K)_X \leq \rho_n(K)_X^{rand} \leq \rho_n(K)_X$$

The question is how far sampling numbers $\rho_n(K)_X$ or $\rho_n^{rand}(K)_X$ are from Kolmogorov widths $d_n(K)_X$, and how to generate optimal samples and algorithms in practice.

Optimal performance for linear approximation

A series of results have been recently obtained for L^2 approximation, comparing sampling numbers with Kolmogorov widths, e.g. [Cohen and Dolbeault 2021, Nagel, Schafer and Ullrich 2021, Temlyakov 2021, Dolbeault, Krieg and Ullrich 2022].

These results are based on constructive approaches for the approximation of functions in a given model class.

See in a next part.

Bounds of Kolmogorov widths $d_n(K)_X$

Upper bounds for $d_n(K)_X$ can be obtained by specific linear approximation methods. Proofs are sometimes constructive.

Lower bounds for $d_n(K)$ can be obtained using different techniques.

- Using diversity in K:

$$d_n(K)_X \geq d_n(S)_X$$

with S some subset of K whose Kolmogorov width can be bounded from below.

Example: if X is a Hilbert space and K contains a set of orthogonal vectors $S = \{u_1, \dots, u_m\}$ with norm $\|u_i\|_X = c_m$,

$$d_n(K)_X \geq d_n(S)_X = d_n(c_m B(\ell_1(\mathbb{R}^m)))_{\ell_2} = c_m \sqrt{1 - n/m}$$

where we used the fact that $d_n(S)_X$ is equal to the n-width of the balanced convex hull of S, which is isomorphic to $c_m B(\ell_1(\mathbb{R}^m))$, and a result of Stechkin (1954).

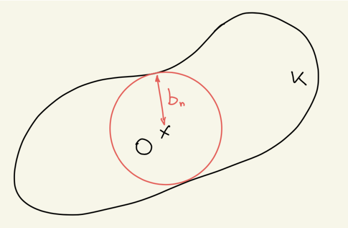
Bounds of Kolmogorov widths $d_n(K)_X$

- Using Bernstein width

$$b_n(K)_X = \sup_{\dim(X_{n+1})=n+1} \sup\{r : rB(X_{n+1}) \subset K\}$$

that is the largest $r > 0$ such that K contains the ball of radius r of some $(n+1)$-dimensional space

$$d_n(K)_X \geq b_n(K)_X$$

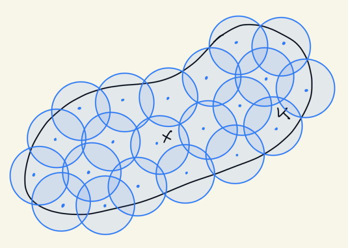


Bounds of Kolmogorov widths $d_n(K)_X$

- Using covering number $N_\epsilon(K)_X$ (minimal number of balls of radius ϵ for covering K) or entropy numbers

$$\epsilon_n(K)_X = \inf\{\epsilon : K \subset \bigcup_{i=1}^{2^n} B(u_i, \epsilon), u_i \in K\} = \inf\{\epsilon : \log_2(N_\epsilon(K)_X) \leq n\}$$

that is the smallest ϵ such that K can be covered by 2^n balls of radius ϵ. Any $u \in K$ can be encoded with n bits up to precision $\epsilon_n(K)$.



Carl's inequality: for all $s > 0$,

$$(n+1)^s \epsilon_n(K)_X \leq C_s \sup_{0 \leq m \leq n} (m+1)^s d_m(K)_X$$

Therefore, if $\epsilon_n(K)_X \gtrsim n^{-s}$, then $d_n(K)_X \lesssim n^{-r}$ can not hold with $r > s$.

Kolmogorov width of Sobolev balls

For $X = L^p(\mathcal{X})$, $\mathcal{X} = [0,1]^d$, $1 \leq p \leq \infty$, and K the unit ball of $W^{k,p}(\mathcal{X})$, it holds

$$d_n(K)_X \sim n^{-k/d}$$

and optimal performance is obtained e.g. by fixed knot splines (with degree adapted to the regularity).

We observe

- the curse of dimensionality : deterioration of the rate of approximation when d increases. Exponential growth with d of the complexity for reaching a given accuracy.
- the blessing of smoothness : improvement of the rate of approximation when k increases.

Kolmogorov width of mixed Sobolev balls

For $X = L^p(\mathcal{X})$, $\mathcal{X} = [0,1]^d$, $1 \leq p \leq \infty$, and K the unit ball of $MW^{k,p}(\mathcal{X})$ (Sobolev space with dominating mixed smoothness), that are functions u such that

$$\max_{|\alpha|_\infty \leq k} \|D^\alpha u\|_{L^p} \leq 1.$$

we have

$$d_n(K)_X \sim n^{-k} \log(n)^{k(d-1)}.$$

with optimal performance achieved by hyperbolic cross approximation (sparse expansion on tensor product of dilated splines) [Dung et al 2016].

Curse of dimensionality is milder but still present.

Optimal nonlinear approximation

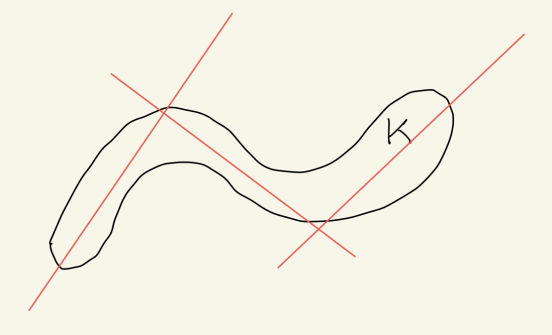
For evaluating the ideal performance of nonlinear methods for the approximation of functions from a class K, different notions of widths have been introduced.

Nonlinear Kolmogorov width

A measure of complexity closely related to n-term approximation and relevant for nonlinear model reduction is the nonlinear Kolmogorov width [Temlyakov 1998] or library width

$$d_n(K,N)_X = \inf_{\#\mathcal{L}_n = N} \sup_{u\in K} \inf_{V_n \in \mathcal{L}_n} E(u, V_n)_X$$

where the infimum is taken over all libraries $\mathcal{L}_n$ of N linear spaces of dimension n.



Choosing $N = N(n)$, this yields a width only depending on n. Interesting regimes are $N(n) = b^n$ or $N(n) = n^{\alpha n}$.

Nonlinear Kolmogorov width

It clearly holds

$$d_1(K, 2^n)_X \leq \epsilon_n(K)_X$$

Also, we have a Carl's type inequality: for all $r > 0$,

$$n^r \epsilon_n(K)_X \leq C(r, b) \max_{1 \leq k \leq n} k^r d_{k-1}(K, b^k)_X.$$

Therefore if for some $b > 0$, $d_{n-1}(K, b^n)_X \lesssim n^{-r}$, then $\epsilon_n(K)_X \lesssim n^{-r}$.

For unit balls K of Besov spaces $B^\alpha_q(L^\tau)$ compactly embedding in $L^p((0,1)^d)$, since $\epsilon_n(K) \gtrsim n^{-\alpha/d}$, we deduce that $d_n(K, b^n)_X \lesssim n^{-\beta}$ can not hold with $\beta > \alpha/d$.

Optimal nonlinear approximation: manifold approximation

Consider the approximation from a n-dimensional "manifold"

$$X_n = \{g(a) : a \in \mathbb{R}^n\}$$

parametrized by a nonlinear map $g : \mathbb{R}^n \to X$. We could consider the problem of finding the best manifold of dimension n for approximating functions from K:

$$\inf_g \sup_{u \in K} \inf_{a \in \mathbb{R}^n} \|u - g(a)\|_X := \eta_n$$

where the infimum is taken among all maps g from $\mathbb{R}^n$ to X.

For any compact set K, $\eta_n = 0$ for all $n \geq 1$. Indeed, K admits a countable dense subset $\{u_i\}_{i \in \mathbb{N}}$ (space-filling manifold). For $n = 1$, letting $g(a) = u_k$ for $a \in [k, k+1)$, we obtain $\eta_1 = 0$.

We can even provide a continuous parametrization, by considering a dense subset $\{u_i\}_{i \in \mathbb{Z}}$ and $g(a) = (a - k)u_{k+1} + (k + 1 - a)u_k$ for $a \in [k, k+1]$.

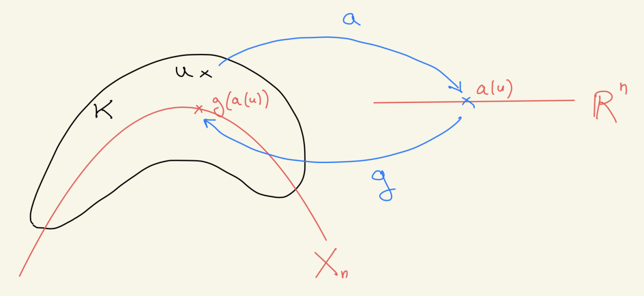
In general, the map which associates to $u \in K$ the coefficients $a(u)$ of its best approximation (if it exists) is not continuous, which makes the approximation process not reasonable.

Optimal nonlinear approximation: manifold width

The following definition of manifold width [DeVore, Howard, Michelli 1989] quantifies how well the set K can be approximated by n-dimensional nonlinear manifolds having continuous parametrization and a continuous parameter selection

$$\delta_n(K)_X = \inf_{g,a} \sup_{u \in K} \|u - g(a(u))\|_X$$

where the infimum is taken over all continuous functions a from K to $\mathbb{R}^n$ and all continuous functions g from $\mathbb{R}^n$ to K.



As for linear widths, the manifold width is lower bounded by the Bernstein width

$$\delta_n(K)_X \geq b_n(K)_X.$$

Manifold width of Sobolev balls

For $X = L^p(\mathcal{X})$, $\mathcal{X} = [0,1]^d$, and K the unit ball of Sobolev spaces $W^{s,q}$ or Besov spaces $B^s_q(L^\tau)$ which compactly embed in L^p

$$\delta_n(K)_X \sim n^{-s/d}$$

Rate $O(n^{-s/d})$ is achieved for a larger class of functions than for linear methods (functions with regularity measured in norms weaker than L^p) .

Optimal performance is achieved by free knot splines or best n-term approximation with a dictionary of tensor products of dilated splines.

Again, we observe the curse of dimensionality, which can not be avoided by such nonlinear methods.

Could extra regularity help ?

Consider $X = L^\infty(\mathcal{X})$ with $\mathcal{X} = [0,1]^d$ and

$$K = \{v \in C^\infty([0,1]^d) : \sup_\alpha \|D^\alpha u\|_{L^\infty} < \infty\},$$

It holds

$$K \subset B(W^{sd,\infty}) \quad \forall s > 0,$$

so that for all $s > 0$

$$d_n(K)_{L^\infty} \lesssim n^{-s}.$$

However,

$$\min\{n : d_n(K)_{L^\infty} \leq 1/2\} \geq c2^{d/2}.$$

The curse of dimensionality is still present.

Could extra regularity help ?

Consider the information based complexity measure of K

$$\delta_n^L(K)_{L^\infty} = \inf_{g,a} \sup_{u\in K} \|u - g(a(u))\|_{L^\infty} \leq a_n(K)_{L^\infty}$$

where the infimum is taken over all linear maps $a : K \to \mathbb{R}^n$ that extract n linear information $a_1(u), \dots a_n(u)$ from a function $u \in K$ (possibly selected adaptively) and over all nonlinear maps g.

It holds [Novak and Wozniakowski 2009]

$$\delta_n^L(K)_{L^\infty} = 1 \quad \text{for all } n = 0, 1, \dots, 2^{\lfloor d/2 \rfloor} - 1$$

or

$$\min\{n : \delta_n^L(K)_{L^\infty} < 1\} \geq 2^{\lfloor d/2 \rfloor}$$

Nonlinear methods can not help...

More assumptions of model classes K are needed...

Parameter dependent PDEs

Consider a parameter-dependent equation

$$\mathcal{P}(u(y); y) = 0, \quad u(y) \in X$$

with $y \in \mathcal{Y}$ some parameter.

The objective is to approximate the solution manifold (model reduction methods)

$$K = \{u(y) : y \in \mathcal{Y}\}$$

or to approximate explicitly the solution map $y \mapsto u(y)$.

As an example, consider the elliptic diffusion equation on a convex domain $D \subset \mathbb{R}^d$

$$-div(a(y)\nabla u(y)) = f$$

with $f \in H^{-1}$, $0 < \underline{a} \leq a(y) \leq \overline{a} < \infty$, and homogeneous Dirichlet boundary conditions.

The solutions

$$u(y) \in H_0^1 := X.$$

Parameter dependent PDEs

- Assuming $f \in L^2$ and $a(y)$ sufficiently smooth, we know that K is in some ball of $H^2(D)$, so that

$$d_n(K)_{H^1} \lesssim n^{-1/d}$$

with optimal performance achieved by splines (finite elements with uniform mesh).

- If $a(y) = a_0 + \sum_{i=1}^m a_i y_i$ with $(\|a_i\|_{L^\infty})_{i\geq 1} \in \ell_p$ for some $p > 1$, then

$$d_n(K)_{H^1} \leq C n^{-s}, \quad s = p^{-1} - 1$$

with constant C independent of d (no curse of dimensionality).

These rates are achieved by sparse polynomial expansions of $y \mapsto u(y)$, exploiting anisotropic analyticity of the solution map.

- More generally, letting $\mathcal{A} = \{a(y) : y \in \mathcal{Y}\}$, we have [Cohen and DeVore 2015]

$$\sup_{n\geq 1} n^s d_n(K)_{H^1} \lesssim \sup_{n\geq 1} n^r d_n(\mathcal{A})_{L^\infty}, \quad \forall s < r - 1.$$

- Optimal spaces X_n are data-dependent. Almost optimal spaces can be constructed using greedy algorithms (reduced basis methods) or sparse polynomial expansions.

- Similar results between nonlinear widths $\delta_n(K)_{H^1}$ and $\delta_n(\mathcal{A})_{L^q}$.

How to beat the curse of dimensionality ?

- No (reasonable) approximation tool is able to overcome the curse of dimensionality for standard regularity classes.

- The key is to make more assumptions on model classes of functions and to provide ad-hoc approximation tools .

- We would like flexible approximation tools that perform well for a wide range of applications (i.e. with sufficiently rich approximation classes)

References I

Approximation theory

A. Pinkus.
N-widths in Approximation Theory, volume 7.
Springer Science & Business Media, 2012.

R. A. DeVore and G. G. Lorentz.
Constructive approximation, volume 303.
Springer Science & Business Media, 1993.

R. A. DeVore.
Nonlinear approximation.
Acta Numerica, 7:51–150, 1998.

V. Temlyakov.
On optimal recovery in L2.
Journal of Complexity, 65:101545, 2021.

N. Nagel, M. Schäfer, and T. Ullrich.
A new upper bound for sampling numbers.
Foundations of Computational Mathematics, pages 1–24, 2021.

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

M. Dolbeault, D. Krieg, and M. Ullrich.
A sharp upper bound for sampling numbers in L_2, 2022.

High-dimensional approximation and model reduction

D. Dũng, V. N. Temlyakov, and T. Ullrich.
Hyperbolic Cross Approximation.
arXiv e-prints, page arXiv:1601.03978, Jan. 2016.

A. Cohen and R. DeVore.
Approximation of high-dimensional parametric pdes.
Acta Numerica, 24:1–159, 2015.

P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors.
Model Reduction and Approximation: Theory and Algorithms.
SIAM, Philadelphia, PA, 2017.

E. Novak and H. Woźniakowski.
Approximation of infinitely differentiable multivariate functions is intractable.
Journal of Complexity, 25(4):398–404, 2009.