From linear to nonlinear n-width : optimality in reduced modelling

Albert Cohen

Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris

Collaborators : Ron DeVore, Guergana Petrova, Przemyslaw Wojtaszczyk

IHP, 08-11-2021

Announcing two great conferences, hopefully in physical format

1. Curves and Surfaces : Arcachon, France, June 20-24, 2022

Registration open : cs2022.sciencesconf.org

2. Foundation of Computational Mathematics : Whistler, Canada, June 12-21, 2023.

Registration opening should be officially announced in coming months.

Save the dates !

Agenda

1. Linear widths
2. Widths of parametrized PDEs and reduced bases
3. Nonlinear widths
4. Widths and sampling numbers

We are interested in approximating general functions $u \in V$, where V is a Banach space, by simpler functions v picked from a linear subspace $V_n \subset V$ of finite dimension n.

Classical Banach spaces : Lebesgue $L^p(\Omega)$, Sobolev $W^{m,p}(\Omega)$ for $\Omega \subset \mathbb{R}^d$.

Classical linear subspaces : algebraic or trigonometric polynomials of some prescribed degree, splines or finite elements on some given mesh, span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k \geq 1}$ of V.

Model class reflecting the properties the target function : $u \in \mathcal{K}$, where $\mathcal{K}$ is a compact set of V. In parametrized PDEs, the set $\mathcal{K}$ is the solution manifold that gathers all solutions $u(y) \in V$ as the parameter vector y varies.

Best choice of approximation spaces for this model class ?

The space V_n approximate $\mathcal{K}$ with uniform accuracy

$$\mathrm{dist}(\mathcal{K}, V_n)_V := \max_{u \in \mathcal{K}} \min_{v \in V_n} \|u - v\|_V$$

A.N. Kolmogorov (1936) defines the linear n-width of $\mathcal{K}$ in the metric V as

$$d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \mathrm{dist}(\mathcal{K}, V_n)_V$$

We are interested in approximating general functions $u \in V$, where V is a Banach space, by simpler functions v picked from a linear subspace $V_n \subset V$ of finite dimension n.

Classical Banach spaces : Lebesgue $L^p(\Omega)$, Sobolev $W^{m,p}(\Omega)$ for $\Omega \subset \mathbb{R}^d$.

Classical linear subspaces : algebraic or trigonometric polynomials of some prescribed degree, splines or finite elements on some given mesh, span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k \geq 1}$ of V.

Model class reflecting the properties the target function : $u \in \mathcal{K}$, where $\mathcal{K}$ is a compact set of V. In parametrized PDEs, the set $\mathcal{K}$ is the solution manifold that gathers all solutions $u(y) \in V$ as the parameter vector y varies.

Best choice of approximation spaces for this model class ?

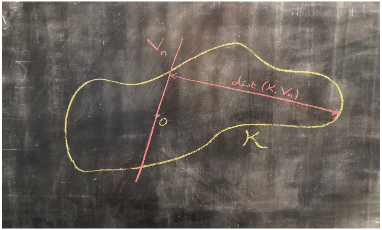
The space V_n approximate $\mathcal{K}$ with uniform accuracy

$$\mathrm{dist}(\mathcal{K}, V_n)_V := \max_{u \in \mathcal{K}} \min_{v \in V_n} \|u - v\|_V$$

A.N. Kolmogorov (1936) defines the linear n-width of $\mathcal{K}$ in the metric V as

$$d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \mathrm{dist}(\mathcal{K}, V_n)_V$$

Intuition



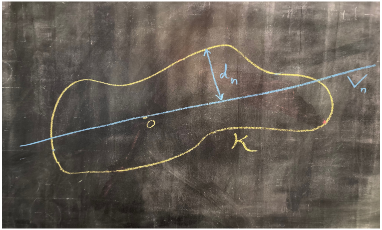
The optimal space achieving the infimum in

$$d_n(\mathcal{K})_V = \inf_{\dim(V_n)=n} \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u - v\|_V.$$

may not exist. One often assumes it exists in order to avoid limiting arguments.

The quantity $d_n(\mathcal{K})_V$ can be viewed as a benchmark/bottleneck for numerical methods applied to the elements from $\mathcal{K}$ that create approximations from linear spaces : interpolation, projection, least squares, Galerkin methods for solving PDEs...

Intuition



The optimal space achieving the infimum in

$$d_n(\mathcal{K})_V = \inf_{\dim(V_n)=n} \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_V.$$

may not exist. One often assumes it exists in order to avoid limiting arguments.

The quantity $d_n(\mathcal{K})_V$ can be viewed as a benchmark/bottleneck for numerical methods applied to the elements from $\mathcal{K}$ that create approximations from linear spaces : interpolation, projection, least squares, Galerkin methods for solving PDEs...

An analog concept in the stochastic framework : PCA

Assume that V is a Hilbert space and u is a random variable taking its value in V.

Optimal spaces in the mean-square sense.

$$\kappa_n^2 = \kappa_n(u)_V^2 := \min_{\dim(V_n)=n} \mathbb{E}\Big(\|u - P_{V_n} u\|_V^2\Big).$$

The space achieving the minimum is easily characterized by principal component analysis : consider the covariance operator

$$v \mapsto Rv = \mathbb{E}(\langle u, v\rangle u),$$

which is compact, when assuming that $\mathbb{E}(\|u\|_V^2) < \infty$. Diagonalized in the Karhunen-Loeve basis $(\varphi_k)_{k\geq 1}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \to 0$.

Then $V_n := \mathrm{span}\{\varphi_1, \dots, \varphi_n\}$ and $\kappa_n^2 = \sum_{k>n} \lambda_k$.

Note that $\kappa_n(u)_V^2 \leq d_n(\mathcal{K})_V^2$ when u is supported in $\mathcal{K}$.

An analog concept in the stochastic framework : PCA

Assume that V is a Hilbert space and u is a random variable taking its value in V.

Optimal spaces in the mean-square sense.

$$\kappa_n^2 = \kappa_n(u)_V^2 := \min_{\dim(V_n)=n} \mathbb{E}\Big(\|u - P_{V_n} u\|_V^2\Big).$$

The space achieving the minimum is easily characterized by principal component analysis : consider the covariance operator

$$v \mapsto Rv = \mathbb{E}(\langle u, v\rangle u),$$

which is compact, when assuming that $\mathbb{E}(\|u\|_V^2) < \infty$. Diagonalized in the Karhunen-Loeve basis $(\varphi_k)_{k\geq 1}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \to 0$.

Then $V_n := \mathrm{span}\{\varphi_1, \dots, \varphi_n\}$ and $\kappa_n^2 = \sum_{k>n} \lambda_k$.

Note that $\kappa_n(u)_V^2 \leq d_n(\mathcal{K})_V^2$ when u is supported in $\mathcal{K}$.

Variants to n-width : realization of the approximation

The best approximation $u_n = \mathrm{argmin}\{\|u - v\|_V \ : \ v \in V_n\}$ is the orthogonal projection if V is a Hilbert space.

For a general Banach space, the map $u \mapsto u_n$ is not linear, and may even not be continuous (non-uniqueness of best approximation).

This motivates alternate definitions of widths where we impose linearity or continuity of the approximation process.

Approximation numbers are defined as

$$a_n(\mathcal{K})_V := \inf_L \max_{u \in \mathcal{K}} \|u - Lu\|_V,$$

with infimum taken over all linear maps L such that $\mathrm{rank}(L) \leq n$.

In a general Banach space $d_n \leq a_n \leq \sqrt{n} d_n$ and right equality may hold.

On the other hand one can prove that

$$d_n(\mathcal{K})_V := \inf_F \max_{u \in \mathcal{K}} \|u - F(u)\|_V,$$

with infimum taken over all continuous maps F such that $\mathrm{rank}(F) \leq n$.

Variants to n-width : realization of the approximation

The best approximation $u_n = \mathrm{argmin}\{\|u - v\|_V \ : \ v \in V_n\}$ is the orthogonal projection if V is a Hilbert space.

For a general Banach space, the map $u \mapsto u_n$ is not linear, and may even not be continuous (non-uniqueness of best approximation).

This motivates alternate definitions of widths where we impose linearity or continuity of the approximation process.

Approximation numbers are defined as

$$a_n(\mathcal{K})_V := \inf_L \max_{u \in \mathcal{K}} \|u - Lu\|_V,$$

with infimum taken over all linear maps L such that $\mathrm{rank}(L) \le n$.

In a general Banach space $d_n \le a_n \le \sqrt{n} d_n$ and right equality may hold.

On the other hand one can prove that

$$d_n(\mathcal{K})_V := \inf_F \max_{u \in \mathcal{K}} \|u - F(u)\|_V,$$

with infimum taken over all continuous maps F such that $\mathrm{rank}(F) \le n$.

Variants to n-width : realization of the approximation

The best approximation $u_n = \text{argmin}\{\|u - v\|_V \ : \ v \in V_n\}$ is the orthogonal projection if V is a Hilbert space.

For a general Banach space, the map $u \mapsto u_n$ is not linear, and may even not be continuous (non-uniqueness of best approximation).

This motivates alternate definitions of widths where we impose linearity or continuity of the approximation process.

Approximation numbers are defined as

$$a_n(\mathcal{K})_V := \inf_L \max_{u \in \mathcal{K}} \|u - Lu\|_V,$$

with infimum taken over all linear maps L such that $\text{rank}(L) \leq n$.

In a general Banach space $d_n \leq a_n \leq \sqrt{n} d_n$ and right equality may hold.

On the other hand one can prove that

$$d_n(\mathcal{K})_V := \inf_F \max_{u \in \mathcal{K}} \|u - F(u)\|_V,$$

with infimum taken over all continuous maps F such that $\text{rank}(F) \leq n$.

Behaviour of n-widths of smoothness classes

Typical compact sets in $V = L^p(\Omega)$ are balls of smoothness spaces. The behaviour of n-width is well understood for such sets. Example : $V = L^\infty(I)$ where $I = [0,1] \subset \mathbb{R}$

and

$$\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I)) = \{u \;:\; \max\{\|u\|_{L^\infty}, \|u'\|_{L^\infty}\} \leq 1\},$$

Then one can prove

$$d_n(\mathcal{K})_V = \frac{1}{2n}, \quad n \geq 1, .$$

More generally when $V = W^{t,p}(\Omega)$ for some bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ and $\mathcal{K}$ is the unit ball of $W^{s,p}(\Omega)$ with $s > t$, one can prove

$$cn^{-(s-t)/d} \leq d_n(\mathcal{K})_V \leq Cn^{-(s-t)/d}, \quad n \geq 1.$$

Curse of dimensionality : exponential growth in d of the needed n to reach accuracy ε.

Proof of upper bound : use a standard approximation method (piecewise polynomials, finite elements, or splines, on uniform partitions of Ω)

Proof of lower bound ? Two systematic approaches.

Lemma : let $B_W = \{u \in W \ : \ \|u\|_V \leq 1\}$ be the unit ball of a subspace $W \subset V$ of dimension $n+1$, then $d_n(B_W)_V = 1$.

Proof : trivial if V is a Hilbert space. Follows from Borsuk-Ulam antipodality theorem in the Banach space case : for any continuous application F from an n-sphere $S_n = \partial B_W$ to an n dimensional space V_n, there exists $x \in S_n$ such that $F(x) = F(-x)$.

It follows that $d_n(\mathcal{K})_V \geq r$ if $\mathcal{K}$ contains the rescaled ball rB_W of an $n+1$-dimensional space W. In other words

$$d_n(\mathcal{K})_V \geq b_n(\mathcal{K})_V, \quad n \geq 1,$$

where the Bernstein n-width $b_n(\mathcal{K})_V$ is defined as the largest $r \geq 0$ such that there exists $W \subset V$ of dimension $n+1$ with $rB_W \subset \mathcal{K}$.

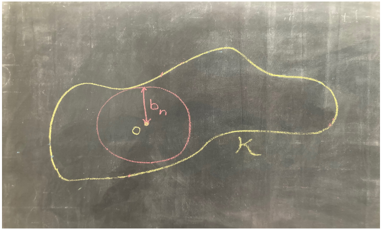
Lemma : let $B_W = \{u \in W \ : \ \|u\|_V \leq 1\}$ be the unit ball of a subspace $W \subset V$ of dimension $n+1$, then $d_n(B_W)_V = 1$.

Proof : trivial if V is a Hilbert space. Follows from Borsuk-Ulam antipodality theorem in the Banach space case : for any continuous application F from an n-sphere $S_n = \partial B_W$ to an n dimensional space V_n, there exists $x \in S_n$ such that $F(x) = F(-x)$.

It follows that $d_n(\mathcal{K})_V \geq r$ if $\mathcal{K}$ contains the rescaled ball rB_W of an $n+1$-dimensional space W. In other words

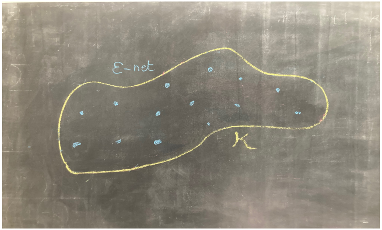
$$d_n(\mathcal{K})_V \geq b_n(\mathcal{K})_V, \quad n \geq 1,$$

where the Bernstein n-width $b_n(\mathcal{K})_V$ is defined as the largest $r \geq 0$ such that there exists $W \subset V$ of dimension $n+1$ with $rB_W \subset \mathcal{K}$.



Define $\varepsilon_n(\mathcal{K})_V$ as the smallest ε such that $\mathcal{K}$ can be covered by 2^n balls of radius ε :

$$\mathcal{K} \subset \bigcup_{i=1,\dots,2^n} B(u^i,\varepsilon), \qquad B(u^i,\varepsilon) := \{u \,:\, \|u-u^i\|_V \le \varepsilon\}.$$



Related to lossy coding : Elements of $\mathcal{K}$ can be encoded with n bits up to precision ε_n.

Carl's inequality : for all $s > 0$ one has

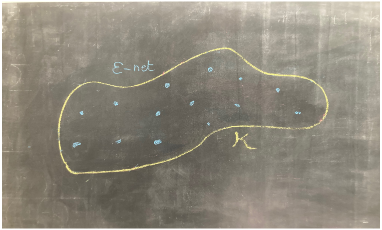
$$(n+1)^s \varepsilon_n(\mathcal{K})_V \le C_s \sup_{m=0,\dots,n} (m+1)^s d_m(\mathcal{K})_V, \quad n \ge 0$$

In particular

$$d_n(\mathcal{K})_V \lesssim n^{-s}, \quad n \ge 0 \implies \varepsilon_n(\mathcal{K})_V \lesssim n^{-s}, \quad n \ge 0.$$

Define $\varepsilon_n(\mathcal{K})_V$ as the smallest ε such that $\mathcal{K}$ can be covered by 2^n balls of radius ε :

$$\mathcal{K} \subset \bigcup_{i=1,\dots,2^n} B(u^i,\varepsilon), \qquad B(u^i,\varepsilon) := \{u \,:\, \|u-u^i\|_V \le \varepsilon\}.$$



Related to lossy coding : Elements of $\mathcal{K}$ can be encoded with n bits up to precision ε_n.

Carl's inequality : for all $s > 0$ one has

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \le C_s \sup_{m=0,\dots,n} (m+1)^s d_m(\mathcal{K})_V, \quad n \ge 0$$

In particular

$$d_n(\mathcal{K})_V \lesssim n^{-s}, \quad n \ge 0 \implies \varepsilon_n(\mathcal{K})_V \lesssim n^{-s}, \quad n \ge 0.$$

Reduced modeling for parametrized PDEs

Complex problems are often modelled by PDEs involving several physical parameters $y = (y^1, \dots, y^d) \in Y \subset \mathbb{R}^d$.

$$\mathcal{P}(u, y) = 0,$$

For each $y \in Y$, we assume well-posedness and therefore existence of a unique solution $u(y) \in V$.

In certain applications (optimization, inverse problems, uncertainty quantification), we may need to solve $y \mapsto u(y)$ for many instances of $y \in Y$: requires computational methods that are uniformly cheap and efficient, uniformly over $y \in Y$.

We are interested in well approximating the solution manifold

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

which we assume to be compact.

Reduced modeling usually involves two steps :

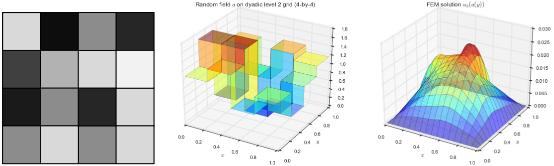
1. In a (costly) offline stage, we search for spaces V_n of dimension n that approximate as best as possible the set $\mathcal{K}$ (benchmark $d_n(\mathcal{K})_V$). These spaces are quite different from classical finite element spaces.

2. In a (cheap) online stage, for any required $y \in Y$ we may compute an accurate approximation $u_n(y) \in V_n$ of $u(y)$, for example by the Galerkin method.

An instructive example : consider the steady-state diffusion equation

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

on a $2d$ domain Ω (+ boundary conditions), with piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-c, c]^d$.



How large is the n-width of $\mathcal{K} = \{u(y) \; : \; y \in Y\} \subset V = H^1(\Omega)$?

Solutions $u(y)$ are bounded in H^s iff $s < 3/2$ and $d_n(\mathcal{U}(H^s))_{H^1} \sim n^{-(s-1)/2} \gtrsim n^{-1/4}$.

In fact $d_n(\mathcal{K})_{H^1}$ decreases faster than $\mathcal{O}(\exp(-cn^{1/d}))$: approximate by power series

$$\max_{y \in Y} \Big\| u(y) - \sum_{|\nu| \le k} u_\nu y^\nu \Big\|_{H^1} \le C \exp(-ck), \quad y^\nu = y_1^{\nu_1} \dots y_d^{\nu_d},$$

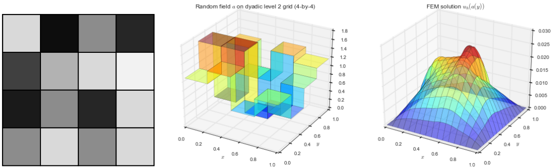
and use $V_n = \mathrm{span}\{u_\nu \; : \; |\nu| \le k\}$ of dimension $n = \binom{k+d}{k}$.

Estimating n-width of solution manifolds

An instructive example : consider the steady-state diffusion equation

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

on a 2d domain Ω (+ boundary conditions), with piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-c, c]^d$.



How large is the n-width of $\mathcal{K} = \{u(y) \; : \; y \in Y\} \subset V = H^1(\Omega)$?

Solutions $u(y)$ are bounded in H^s iff $s < 3/2$ and $d_n(\mathcal{U}(H^s))_{H^1} \sim n^{-(s-1)/2} \gtrsim n^{-1/4}$.

In fact $d_n(\mathcal{K})_{H^1}$ decreases faster than $\mathcal{O}(\exp(-cn^{1/d}))$: approximate by power series

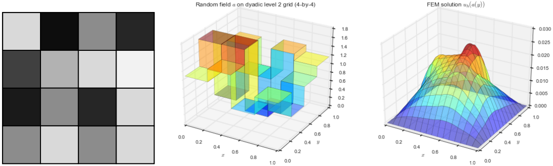
$$\max_{y\in Y} \Big\| u(y) - \sum_{|\nu|\leq k} u_\nu y^\nu \Big\|_{H^1} \leq C \exp(-ck), \quad y^\nu = y_1^{\nu_1} \dots y_d^{\nu_d},$$

and use $V_n = \mathrm{span}\{u_\nu \; : \; |\nu| \leq k\}$ of dimension $n = \binom{k+d}{k}$.

An instructive example : consider the steady-state diffusion equation

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

on a 2d domain Ω (+ boundary conditions), with piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-c, c]^d$.



How large is the n-width of $\mathcal{K} = \{u(y) \; : \; y \in Y\} \subset V = H^1(\Omega)$?

Solutions $u(y)$ are bounded in H^s iff $s < 3/2$ and $d_n(\mathcal{U}(H^s))_{H^1} \sim n^{-(s-1)/2} \gtrsim n^{-1/4}$.

In fact $d_n(\mathcal{K})_{H^1}$ decreases faster than $\mathcal{O}(\exp(-cn^{1/d}))$: approximate by power series

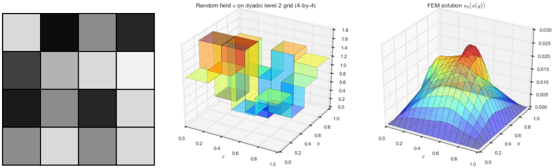
$$\max_{y\in Y} \Big\| u(y) - \sum_{|\nu|\leq k} u_\nu y^\nu \Big\|_{H^1} \leq C \exp(-ck), \quad y^\nu = y_1^{\nu_1} \dots y_d^{\nu_d},$$

and use $V_n = \mathrm{span}\{u_\nu \; : \; |\nu| \leq k\}$ of dimension $n = \binom{k+d}{k}$.

An instructive example : consider the steady-state diffusion equation

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

on a $2d$ domain Ω (+ boundary conditions), with piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-c, c]^d$.



How large is the n-width of $\mathcal{K} = \{u(y) \ : \ y \in Y\} \subset V = H^1(\Omega)$?

Solutions $u(y)$ are bounded in H^s iff $s < 3/2$ and $d_n(\mathcal{U}(H^s))_{H^1} \sim n^{-(s-1)/2} \gtrsim n^{-1/4}$.

In fact $d_n(\mathcal{K})_{H^1}$ decreases faster than $\mathcal{O}(\exp(-cn^{1/d}))$: approximate by power series

$$\max_{y \in Y} \Big\| u(y) - \sum_{|\nu| \leq k} u_\nu y^\nu \Big\|_{H^1} \leq C \exp(-ck), \quad y^\nu = y_1^{\nu_1} \dots y_d^{\nu_d},$$

and use $V_n = \mathrm{span}\{u_\nu \ : \ |\nu| \leq k\}$ of dimension $n = \binom{k+d}{k}$.

A general result for infinite dimensional parameter dependence

Theorem (Cohen-DeVore, 2016) : Let V_1 and V_2 be two complex valued Banach spaces and $\mathcal{K}_1 \subset V_1$ be a compact set. Let

$$F : V_1 \to V_2,$$

be a map that is holomorphic on an open neighbourhood of $\mathcal{K}_1$. Then, with $\mathcal{K}_2 := F(\mathcal{K}_1)$, one has for all $s > 1$

$$\sup_{n\geq 0} n^s d_n(\mathcal{K}_1)_{V_1} < \infty \implies \sup_{n\geq 0} n^t d_n(\mathcal{K}_2)_{V_2} < \infty, \quad t < s-1.$$

Note that if F was a continuous linear map, one would simply have

$$d_n(\mathcal{K}_2)_{V_2} \leq C d_n(\mathcal{K}_1)_{V_1}, \quad C = \|F\|_{V_1 \to V_2}.$$

The proof goes by expanding $a \in \mathcal{K}_1$ in a suitable basis $a = a(y) = \sum_{j\geq 1} y_j \psi_j$ with decay properties on the $\|\psi_j\|_{V_1}$ and then approximate $F(a(y))$ by polynomials in y. This induces a loss of 1 in the rate of decay. Open problem : same rate $t = s$?

This result applies to elliptic equations such as $-\text{div}(a\nabla u) = f$ for the map $F : a \to u$ with $V_1 = L^\infty$ and $V_2 = H^1$. Also applies to parabolic equations, nonlinear problems such as Navier-Stokes equations, and to these problems set on parametrized domains. It does not apply to hyperbolic equations.

A general result for infinite dimensional parameter dependence

Theorem (Cohen-DeVore, 2016) : Let V_1 and V_2 be two complex valued Banach spaces and $\mathcal{K}_1 \subset V_1$ be a compact set. Let

$$F : V_1 \to V_2,$$

be a map that is holomorphic on an open neighbourhood of $\mathcal{K}_1$. Then, with $\mathcal{K}_2 := F(\mathcal{K}_1)$, one has for all $s > 1$

$$\sup_{n\geq 0} n^s d_n(\mathcal{K}_1)_{V_1} < \infty \implies \sup_{n\geq 0} n^t d_n(\mathcal{K}_2)_{V_2} < \infty, \quad t < s-1.$$

Note that if F was a continuous linear map, one would simply have

$$d_n(\mathcal{K}_2)_{V_2} \leq C d_n(\mathcal{K}_1)_{V_1}, \quad C = \|F\|_{V_1\to V_2}.$$

The proof goes by expanding $a \in \mathcal{K}_1$ in a suitable basis $a = a(y) = \sum_{j\geq 1} y_j \psi_j$ with decay properties on the $\|\psi_j\|_{V_1}$ and then approximate $F(a(y))$ by polynomials in y. This induces a loss of 1 in the rate of decay. Open problem : same rate $t = s$?

This result applies to elliptic equations such as $-\text{div}(a\nabla u) = f$ for the map $F : a \to u$ with $V_1 = L^\infty$ and $V_2 = H^1$. Also applies to parabolic equations, nonlinear problems such as Navier-Stokes equations, and to these problems set on parametrized domains. It does not apply to hyperbolic equations.

A general result for infinite dimensional parameter dependence

Theorem (Cohen-DeVore, 2016) : Let V_1 and V_2 be two complex valued Banach spaces and $\mathcal{K}_1 \subset V_1$ be a compact set. Let

$$F : V_1 \to V_2,$$

be a map that is holomorphic on an open neighbourhood of $\mathcal{K}_1$. Then, with $\mathcal{K}_2 := F(\mathcal{K}_1)$, one has for all $s > 1$

$$\sup_{n \geq 0} n^s d_n(\mathcal{K}_1)_{V_1} < \infty \implies \sup_{n \geq 0} n^t d_n(\mathcal{K}_2)_{V_2} < \infty, \quad t < s - 1.$$

Note that if F was a continuous linear map, one would simply have

$$d_n(\mathcal{K}_2)_{V_2} \leq C d_n(\mathcal{K}_1)_{V_1}, \quad C = \|F\|_{V_1 \to V_2}.$$

The proof goes by expanding $a \in \mathcal{K}_1$ in a suitable basis $a = a(y) = \sum_{j \geq 1} y_j \psi_j$ with decay properties on the $\|\psi_j\|_{V_1}$ and then approximate $F(a(y))$ by polynomials in y. This induces a loss of 1 in the rate of decay. Open problem : same rate $t = s$?

This result applies to elliptic equations such as $-\mathrm{div}(a \nabla u) = f$ for the map $F : a \to u$ with $V_1 = L^\infty$ and $V_2 = H^1$. Also applies to parabolic equations, nonlinear problems such as Navier-Stokes equations, and to these problems set on parametrized domains. It does not apply to hyperbolic equations.

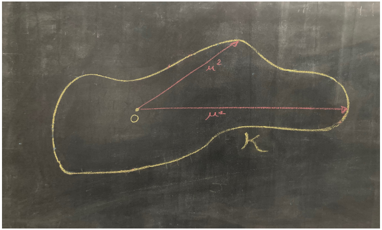
The reduced basis algorithm

Idea : use particular instances $u^i = u(y^i) \in \mathcal{K}$ for generating $V_n = \mathrm{span}\{u^1, \dots, u^n\}$.

Greedy selection in offline stage : having generated $u^1, \dots, u^{k-1}$, select next instance

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

where $P_{V_{k-1}}$ is the orthogonal projection. Here we assume V to be a Hilbert space.



In practice, weak selection $\|u - P_{V_{k-1}} u^k\|_V \geq \gamma \max_{u \in \mathcal{K}} \|u - P_{V_{k-1}} u\|_V$, for fixed $\gamma \in]0, 1[$, and maximization on a large finite training set $\widetilde{\mathcal{K}} \subset \mathcal{K}$.

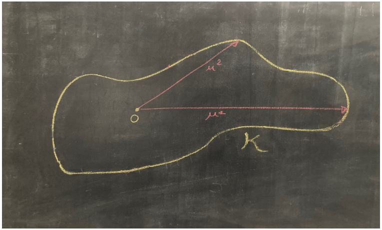
The reduced basis algorithm

Idea : use particular instances $u^i = u(y^i) \in \mathcal{K}$ for generating $V_n = \mathrm{span}\{u^1, \dots, u^n\}$.

Greedy selection in offline stage : having generated $u^1, \dots, u^{k-1}$, select next instance

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

where $P_{V_{k-1}}$ is the orthogonal projection. Here we assume V to be a Hilbert space.



In practice, weak selection $\|u - P_{V_{k-1}} u^k\|_V \geq \gamma \max_{u \in \mathcal{K}} \|u - P_{V_{k-1}} u\|_V$, for fixed $\gamma \in]0, 1[$, and maximization on a large finite training set $\widetilde{\mathcal{K}} \subset \mathcal{K}$.

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.

Theorem (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

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

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) : $\sigma_n \leq n 2^n d_n$.

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.

Theorem (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

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

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) : $\sigma_n \leq n 2^n d_n$.

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.

Theorem (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

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

Failure of linear reduced modeling

Linear reduced modeling for parametrized hyperbolic PDEs suffers from a slow decay of Kolmogorov n-width.

Simple example : consider the univariate linear transport equation

$$\partial_t u + a \partial_x u = 0,$$

with constant velocity $a \in \mathbb{R}$ and initial condition $u_0 = u(x,0) = \chi_{[0,1]}(x)$.

Parametrize the solution by the velocity $a \in [a_{\min}, a_{\max}]$ and consider the solution manifold at final time $T = 1$,

$$\mathcal{H} = \{\chi_{[a,a+1]} \; : \; a \in [a_{\min}, a_{\max}]\}.$$

It can be proved that for $1 \leq p < \infty$,

$$d_n(\mathcal{H})_{L^p} \sim n^{-1/p}.$$

In particular, we cannot hope for a good performance of reduced basis methods (not better than piecewise constant approximation on uniform meshes).

Failure of linear reduced modeling

Linear reduced modeling for parametrized hyperbolic PDEs suffers from a slow decay of Kolmogorov n-width.

Simple example : consider the univariate linear transport equation

$$\partial_t u + a\partial_x u = 0,$$

with constant velocity $a \in \mathbb{R}$ and initial condition $u_0 = u(x,0) = \chi_{[0,1]}(x)$.

Parametrize the solution by the velocity $a \in [a_{\min}, a_{\max}]$ and consider the solution manifold at final time $T = 1$,

$$\mathcal{H} = \{\chi_{[a,a+1]} \ : \ a \in [a_{\min}, a_{\max}]\}.$$

It can be proved that for $1 \leq p < \infty$,

$$d_n(\mathcal{H})_{L^p} \sim n^{-1/p}.$$

In particular, we cannot hope for a good performance of reduced basis methods (not better than piecewise constant approximation on uniform meshes).

Failure of linear reduced modeling

Linear reduced modeling for parametrized hyperbolic PDEs suffers from a slow decay of Kolmogorov n-width.

Simple example : consider the univariate linear transport equation

$$\partial_t u + a\partial_x u = 0,$$

with constant velocity $a \in \mathbb{R}$ and initial condition $u_0 = u(x,0) = \chi_{[0,1]}(x)$.

Parametrize the solution by the velocity $a \in [a_{\min}, a_{\max}]$ and consider the solution manifold at final time $T = 1$,

$$\mathcal{H} = \{\chi_{[a,a+1]} \, : \, a \in [a_{\min}, a_{\max}]\}.$$

It can be proved that for $1 \leq p < \infty$,

$$d_n(\mathcal{H})_{L^p} \sim n^{-1/p}.$$

In particular, we cannot hope for a good performance of reduced basis methods (not better than piecewise constant approximation on uniform meshes).

Nonlinear approximation

For such problems, one expects improved performance by nonlinear methods.

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

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

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

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

- 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.

Is there a natural notion of width describing optimal nonlinear approximation ?

Nonlinear approximation

For such problems, one expects improved performance by nonlinear methods.

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

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

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

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

- 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.

Is there a natural notion of width describing optimal nonlinear approximation ?

For such problems, one expects improved performance by nonlinear methods.

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

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

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

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

- 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.

Is there a natural notion of width describing optimal nonlinear approximation ?

Library widths

A library $\mathcal{L}_n$ is a finite collection of linear spaces $V_n \subset V$ of dimension at most n.

We approximate u by picking a space from $\mathcal{L}_n$, resulting in the error

$$e(u, \mathcal{L}_n)_V = \min_{V_n \in \mathcal{L}_n} \min_{v \in V_n} \|u - v\|_V.$$

Temlyakov (1998) defines the library width

$$d_{N,n}(\mathcal{K})_V := \inf_{\#(\mathcal{L}_n) \leq N} \max_{u \in \mathcal{K}} e(u, \mathcal{L}_n)_V.$$

Note that $d_{1,n} = d_n$.

The interesting regime is when $N >> n$. Typical choices that have been studied are $N = A^n$ or $N = n^{an}$ for some $A > 1$ or $a > 0$.

Remark : optimal library approximation amounts in splitting the set $\mathcal{K}$ into N different component $\mathcal{K}_j$, each of them being approximated by some optimal n-dimensional space V_n^j picked from the library.

This type of width is well adapted to describe optimality for best n-term approximation or adaptive refinements, but not for neural networks or rational fractions.

Library widths

A library $\mathcal{L}_n$ is a finite collection of linear spaces $V_n \subset V$ of dimension at most n.

We approximate u by picking a space from $\mathcal{L}_n$, resulting in the error

$$e(u, \mathcal{L}_n)_V = \min_{V_n \in \mathcal{L}_n} \min_{v \in V_n} \|u - v\|_V.$$

Temlyakov (1998) defines the library width

$$d_{N,n}(\mathcal{K})_V := \inf_{\#(\mathcal{L}_n) \leq N} \max_{u \in \mathcal{K}} e(u, \mathcal{L}_n)_V.$$

Note that $d_{1,n} = d_n$.

The interesting regime is when $N >> n$. Typical choices that have been studied are $N = A^n$ or $N = n^{an}$ for some $A > 1$ or $a > 0$.

Remark : optimal library approximation amounts in splitting the set $\mathcal{K}$ into N different component $\mathcal{K}_j$, each of them being approximated by some optimal n-dimensional space V_n^j picked from the library.

This type of width is well adapted to describe optimality for best n-term approximation or adaptive refinements, but not for neural networks or rational fractions.

Library widths

A library $\mathcal{L}_n$ is a finite collection of linear spaces $V_n \subset V$ of dimension at most n.

We approximate u by picking a space from $\mathcal{L}_n$, resulting in the error

$$e(u, \mathcal{L}_n)_V = \min_{V_n \in \mathcal{L}_n} \min_{v \in V_n} \|u - v\|_V.$$

Temlyakov (1998) defines the library width

$$d_{N,n}(\mathcal{K})_V := \inf_{\#(\mathcal{L}_n) \leq N} \max_{u \in \mathcal{K}} e(u, \mathcal{L}_n)_V.$$

Note that $d_{1,n} = d_n$.

The interesting regime is when $N >> n$. Typical choices that have been studied are $N = A^n$ or $N = n^{an}$ for some $A > 1$ or $a > 0$.

Remark : optimal library approximation amounts in splitting the set $\mathcal{K}$ into N different component $\mathcal{K}_j$, each of them being approximated by some optimal n-dimensional space V_n^j picked from the library.

This type of width is well adapted to describe optimality for best n-term approximation or adaptive refinements, but not for neural networks or rational fractions.

Naive idea : replace linear spaces V_n of dimension n by smooth manifolds $\mathcal{M}_n$ of dimension n in the definition of d_n.

This would lead to the quantity

$$\inf_{\dim(\mathcal{M}_n)=n} \max_{u\in\mathcal{K}} \min_{v\in V} \|u-v\|_V,$$

However its value is 0 even for $n=1$: space filling curves !

DeVore-Howard-Michelli (1989) : impose continuous selection by defining

$$\delta_n(\mathcal{K})_V := \inf_{D,E} \max_{u\in\mathcal{K}} \|u-D(E(u))\|_V,$$

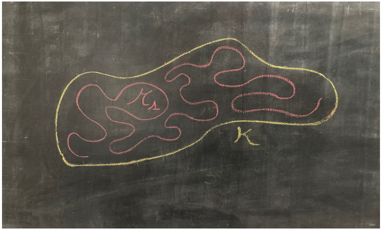
where infimum is taken on all continuous pairs $E : V \to \mathbb{R}^n$ and $D : \mathbb{R}^n \to V$.

Naive idea : replace linear spaces V_n of dimension n by smooth manifolds $\mathcal{M}_n$ of dimension n in the definition of d_n.

This would lead to the quantity

$$\inf_{\dim(\mathcal{M}_n)=n} \max_{u\in\mathcal{K}} \min_{v\in V} \|u-v\|_V,$$

However its value is 0 even for $n=1$: space filling curves !



DeVore-Howard-Michelli (1989) : impose continuous selection by defining

$$\delta_n(\mathcal{K})_V := \inf_{D,E} \max_{u\in\mathcal{K}} \|u - D(E(u))\|_V,$$

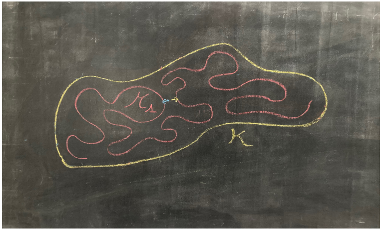
where infimum is taken on all continuous pairs $E : V \to \mathbb{R}^n$ and $D : \mathbb{R}^n \to V$.

Naive idea : replace linear spaces V_n of dimension n by smooth manifolds $\mathcal{M}_n$ of dimension n in the definition of d_n.

This would lead to the quantity

$$\inf_{\dim(\mathcal{M}_n)=n} \max_{u\in\mathcal{K}} \min_{v\in V} \|u-v\|_V,$$

However its value is 0 even for $n=1$: space filling curves !



DeVore-Howard-Michelli (1989) : impose continuous selection by defining

$$\delta_n(\mathcal{K})_V := \inf_{D,E} \max_{u\in\mathcal{K}} \|u-D(E(u))\|_V,$$

where infimum is taken on all continuous pairs $E : V \to \mathbb{R}^n$ and $D : \mathbb{R}^n \to V$.

Estimating nonlinear width

Both library and manifold widths match known rates of nonlinear approximation (DeVore-Popov, 1980-1990's) by wavelets or adaptive finite elements : if $V = L^p(\Omega)$ and $\mathcal{K} = \mathcal{U}(B^s_{q,q}(\Omega))$ for $\frac{1}{q} < \frac{1}{p} + \frac{s}{d}$, one has

$$d_{n,N}(\mathcal{K})_V \sim \delta_n(\mathcal{K})_V \sim n^{-s/d}.$$

Upper bounds obtained by these classical nonlinear approximation results.

Library widths satisfy Carl's inequality (for the regimes $N = A^n$ or $N = n^{an}$).

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_{m,N}(\mathcal{K})_V, \quad n \geq 0.$$

Manifold widths do not satisty Carl's inequality but are bounded by below by Bernstein widths (by the Borsuk-Ulam argument).

$$\delta_n(\mathcal{K})_V \geq b_n(\mathcal{K})_V.$$

For example, if $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$, one has $\delta_n(\mathcal{K})_V \sim n^{-1}$.

Yarotzki, Shen-Yang-Zhang (2020) : Neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} ! Parameter selection cannot be stable.

Estimating nonlinear width

Both library and manifold widths match known rates of nonlinear approximation (DeVore-Popov, 1980-1990's) by wavelets or adaptive finite elements : if $V = L^p(\Omega)$ and $\mathcal{K} = \mathcal{U}(B^s_{q,q}(\Omega))$ for $\frac{1}{q} < \frac{1}{p} + \frac{s}{d}$, one has

$$d_{n,N}(\mathcal{K})_V \sim \delta_n(\mathcal{K})_V \sim n^{-s/d}.$$

Upper bounds obtained by these classical nonlinear approximation results.

Library widths satisfy Carl's inequality (for the regimes $N = A^n$ or $N = n^{an}$).

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_{m,N}(\mathcal{K})_V, \quad n \geq 0.$$

Manifold widths do not satisty Carl's inequality but are bounded by below by Bernstein widths (by the Borsuk-Ulam argument).

$$\delta_n(\mathcal{K})_V \geq b_n(\mathcal{K})_V.$$

For example, if $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$, one has $\delta_n(\mathcal{K})_V \sim n^{-1}$.

Yarotzki, Shen-Yang-Zhang (2020) : Neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} ! Parameter selection cannot be stable.

Estimating nonlinear width

Both library and manifold widths match known rates of nonlinear approximation (DeVore-Popov, 1980-1990's) by wavelets or adaptive finite elements : if $V = L^p(\Omega)$ and $\mathcal{K} = \mathcal{U}(B^s_{q,q}(\Omega))$ for $\frac{1}{q} < \frac{1}{p} + \frac{s}{d}$, one has

$$d_{n,N}(\mathcal{K})_V \sim \delta_n(\mathcal{K})_V \sim n^{-s/d}.$$

Upper bounds obtained by these classical nonlinear approximation results.

Library widths satisfy Carl's inequality (for the regimes $N = A^n$ or $N = n^{an}$).

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_{m,N}(\mathcal{K})_V, \quad n \geq 0.$$

Manifold widths do not satisty Carl's inequality but are bounded by below by Bernstein widths (by the Borsuk-Ulam argument).

$$\delta_n(\mathcal{K})_V \geq b_n(\mathcal{K})_V.$$

For example, if $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$, one has $\delta_n(\mathcal{K})_V \sim n^{-1}$.

Yarotzki, Shen-Yang-Zhang (2020) : Neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} ! Parameter selection cannot be stable.

Estimating nonlinear width

Both library and manifold widths match known rates of nonlinear approximation (DeVore-Popov, 1980-1990's) by wavelets or adaptive finite elements : if $V = L^p(\Omega)$ and $\mathcal{K} = \mathcal{U}(B^s_{q,q}(\Omega))$ for $\frac{1}{q} < \frac{1}{p} + \frac{s}{d}$, one has

$$d_{n,N}(\mathcal{K})_V \sim \delta_n(\mathcal{K})_V \sim n^{-s/d}.$$

Upper bounds obtained by these classical nonlinear approximation results.

Library widths satisfy Carl's inequality (for the regimes $N = A^n$ or $N = n^{an}$).

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_{m,N}(\mathcal{K})_V, \quad n \geq 0.$$

Manifold widths do not satisty Carl's inequality but are bounded by below by Bernstein widths (by the Borsuk-Ulam argument).

$$\delta_n(\mathcal{K})_V \geq b_n(\mathcal{K})_V.$$

For example, if $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$, one has $\delta_n(\mathcal{K})_V \sim n^{-1}$.

Yarotzki, Shen-Yang-Zhang (2020) : Neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} ! Parameter selection cannot be stable.

Stable nonlinear widths

Cohen-DeVore-Petrova-Wojtaszczyk (2020) : for some fixed $L > 1$ define

$$\delta_{n,L}(\mathcal{K})_V := \inf_{D,E} \max_{u \in \mathcal{K}} \|u - D(E(u))\|_V,$$

where the infimum is taken on all pairs $E : V \to \mathbb{R}^n$ and $D : \mathbb{R}^n \to V$, that satisfy

$$\|D(x) - D(y)\|_V \leq L\|x - y\|_n \quad \text{and} \quad \|E(u) - E(v)\|_n \leq L\|u - v\|_V, \quad x, y \in \mathbb{R}^n, u, v \in V.$$

Here $\| \cdot \|_n$ is an arbitrary norm on $\mathbb{R}^n$.

This notion of stable width now satisfies Carl's inequality : for any $L > 1$,

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \leq C_s \sup_{m=0,\dots,n} (m+1)^s \delta_{m,L}(\mathcal{K})_V, \quad n \geq 0.$$

in addition to the lower bound by Gelfand width $\delta_{n,L}(\mathcal{K})_V \geq b_n(\mathcal{K})_V$

Open problem : with $V = L^p(\Omega)$ and $\mathcal{K} = \mathcal{U}(B^s_{q,q}(\Omega))$ for $\frac{1}{q} < \frac{1}{p} + \frac{s}{d}$, do we have $\delta_{n,L}(\mathcal{K})_V \sim n^{-s/d}$? Positive answer known only when $p = 2$.

Stable nonlinear widths

Cohen-DeVore-Petrova-Wojtaszczyk (2020) : for some fixed $L > 1$ define

$$\delta_{n,L}(\mathcal{K})_V := \inf_{D,E} \max_{u\in\mathcal{K}} \|u - D(E(u))\|_V,$$

where the infimum is taken on all pairs $E : V \to \mathbb{R}^n$ and $D : \mathbb{R}^n \to V$, that satisfy

$$\|D(x) - D(y)\|_V \le L\|x - y\|_n \quad \text{and} \quad \|E(u) - E(v)\|_n \le L\|u - v\|_V, \quad x, y \in \mathbb{R}^n, u, v \in V.$$

Here $\| \cdot \|_n$ is an arbitrary norm on $\mathbb{R}^n$.

This notion of stable width now satisfies Carl's inequality : for any $L > 1$,

$$(n+1)^s \varepsilon_n(\mathcal{K})_V \le C_s \sup_{m=0,\dots,n} (m+1)^s \delta_{m,L}(\mathcal{K})_V, \quad n \ge 0.$$

in addition to the lower bound by Gelfand width $\delta_{n,L}(\mathcal{K})_V \ge b_n(\mathcal{K})_V$

Open problem : with $V = L^p(\Omega)$ and $\mathcal{K} = \mathcal{U}(B^s_{q,q}(\Omega))$ for $\frac{1}{q} < \frac{1}{p} + \frac{s}{d}$, do we have $\delta_{n,L}(\mathcal{K})_V \sim n^{-s/d}$? Positive answer known only when $p = 2$.

When V is a Hilbert space, stable widths are strongly tied to entropy numbers.

Theorem : Let V be a Hilbert space, then for any $L > 1$, there exists a constant $c = c(L)$ such that, for any compact set $\mathcal{K}$,

$$\delta_{cn,L}(\mathcal{K})_V \le 3\varepsilon_n(\mathcal{K})_V.$$

With $L = 2$ one can take $c = 26$.

Together with Carl's inequality, this means that

$$\sup_{n\ge 0} n^s \delta_{n,L}(\mathcal{K})_V < \infty \iff \sup_{n\ge 0} n^s \varepsilon_n(\mathcal{K})_V < \infty,$$

for all $s > 0$.

We do not know if this result holds for Banach spaces. Proof for Hilbert spaces :

1. Consider $\mathcal{N}$ an ε_n-net of $\mathcal{K}$ with $\#(\mathcal{N}) = 2^n$.
2. Johnson-Lindenstrauss projection as encoder : $E = P_W$ where $\dim(W) \le cn$

$$L^{-1}\|u^i - u^j\|_V \le \|P_W(u^i - u^j)\|_V \le \|u^i - u^j\|_V, \quad u^i, u^j \in \mathcal{N}.$$

3. This gives an exact decoding map that is L-Lipschitz from $P_W\mathcal{N}$ to $\mathcal{N}$.
4. Extend this map from $W \sim \mathbb{R}^{cn}$ to V with same Lipschitz constant (Kirszbraun).

Stable widths and entropies

When V is a Hilbert space, stable widths are strongly tied to entropy numbers.

Theorem : Let V be a Hilbert space, then for any $L > 1$, there exists a constant $c = c(L)$ such that, for any compact set $\mathcal{K}$,

$$\delta_{cn,L}(\mathcal{K})_V \leq 3\varepsilon_n(\mathcal{K})_V.$$

With $L = 2$ one can take $c = 26$.

Together with Carl's inequality, this means that

$$\sup_{n\geq 0} n^s \delta_{n,L}(\mathcal{K})_V < \infty \iff \sup_{n\geq 0} n^s \varepsilon_n(\mathcal{K})_V < \infty,$$

for all $s > 0$.

We do not know if this result holds for Banach spaces. Proof for Hilbert spaces :

1. Consider $\mathcal{N}$ an ε_n-net of $\mathcal{K}$ with $\#(\mathcal{N}) = 2^n$.
2. Johnson-Lindenstrauss projection as encoder : $E = P_W$ where $\dim(W) \leq cn$

$$L^{-1}\|u^i - u^j\|_V \leq \|P_W(u^i - u^j)\|_V \leq \|u^i - u^j\|_V, \quad u^i, u^j \in \mathcal{N}.$$

3. This gives an exact decoding map that is L-Lipschitz from $P_W\mathcal{N}$ to $\mathcal{N}$.
4. Extend this map from $W \sim \mathbb{R}^{cn}$ to V with same Lipschitz constant (Kirszbraun).

Stable width of solution manifolds

For the linear transport equation manifold $\mathcal{H} = \Big\{\chi_{[a,a+1]} \,:\, a \in [a_{\min}, a_{\max}]\Big\}$ it is easily established that entropy numbers in L^p spaces have exponential decay

$$\varepsilon_n(\mathcal{H})_{L^p} \leq C \exp(-cn), \quad n \geq 0.$$

This implies in particular that $\delta_{n,L}(\mathcal{H})_{L^2} \leq \tilde{C} \exp(-\tilde{c}n)$ while $d_n(\mathcal{H})_{L^2} \sim n^{-1/2}$.

Similar results hold for manifolds resulting from more general hyperbolic equations.

A general result : if $F : V_1 \to V_2$ is a L-Lipschitz mapping between Banach spaces, then an ε-net of $\mathcal{K}_1 \subset V_1$ is mapped into an $L\varepsilon$-net of $\mathcal{K}_2 := F(\mathcal{K}_1)$ and therefore

$$\varepsilon_n(\mathcal{K}_2)_{V_2} \leq L\varepsilon_n(\mathcal{K}_1)_{V_1}, \quad n \geq 0.$$

This implies in particular that when V_2 is a Hilbert space

$$\sup_{n\geq 0} n^s \delta_{n,L}(\mathcal{K}_1)_{V_1} < \infty \implies \sup_{n\geq 0} n^s \delta_{n,L}(\mathcal{K}_2)_{V_2} < \infty.$$

Benchmark : develop concrete stable numerical methods that meet these rates.

Stable width of solution manifolds

For the linear transport equation manifold $\mathcal{H} = \left\{ \chi_{[a,a+1]} \,:\, a \in [a_{\min}, a_{\max}] \right\}$ it is easily established that entropy numbers in L^p spaces have exponential decay

$$\varepsilon_n(\mathcal{H})_{L^p} \leq C \exp(-cn), \quad n \geq 0.$$

This implies in particular that $\delta_{n,L}(\mathcal{H})_{L^2} \leq \tilde{C} \exp(-\tilde{c}n)$ while $d_n(\mathcal{H})_{L^2} \sim n^{-1/2}$.

Similar results hold for manifolds resulting from more general hyperbolic equations.

A general result : if $F : V_1 \to V_2$ is a L-Lipschitz mapping between Banach spaces, then an ε-net of $\mathcal{K}_1 \subset V_1$ is mapped into an $L\varepsilon$-net of $\mathcal{K}_2 := F(\mathcal{K}_1)$ and therefore

$$\varepsilon_n(\mathcal{K}_2)_{V_2} \leq L\varepsilon_n(\mathcal{K}_1)_{V_1}, \quad n \geq 0.$$

This implies in particular that when V_2 is a Hilbert space

$$\sup_{n\geq 0} n^s \delta_{n,L}(\mathcal{K}_1)_{V_1} < \infty \implies \sup_{n\geq 0} n^s \delta_{n,L}(\mathcal{K}_2)_{V_2} < \infty.$$

Benchmark : develop concrete stable numerical methods that meet these rates.

Recent results in this direction

Consider a dictionnary $\mathcal{D}$ in a Hilbert space V, such that $\|\varphi\|_V \leq 1$ for all $\varphi \in \mathcal{D}$, and the model class

$$\mathcal{K} := \Big\{ v = \sum c_j \varphi_j \ : \ \varphi_j \in \mathcal{D}, \ \sum |c_j| \leq 1 \Big\}.$$

The OMP algorithm recursively produces

$$u_n = \sum_{j=1}^{n} c_j \varphi_j = P_{V_n} u, \quad V_n = \mathrm{span}\{\varphi_1, \dots, \varphi_n\},$$

by selecting $\varphi_n \in \mathcal{D}$ maximizing $\frac{\langle u - P_{V_{n-1}} u, \varphi \rangle}{\|\varphi\|_V}$ over $\varphi \in \mathcal{D}$.

DeVore-Temlyakov (1998) : $u \in \mathcal{K} \implies \|u - u_n\|_V \leq n^{-1/2}$.

Siegel-Xu (2021) : consider smoothly parametrized dictionnaries

$$\mathcal{D} = \{\varphi = \mathcal{P}(\theta) \ : \ \theta \in \mathcal{M}\},$$

where $\mathcal{M}$ is d-dimensional compact manifold and $\mathcal{P}$ is $\mathcal{C}^r$. In this setting

$$u \in \mathcal{K} \implies \|u - u_n\|_V \lesssim n^{-s}, \quad s = \frac{1}{2} + \frac{r}{d} \quad \text{and} \quad \varepsilon_n(\mathcal{K})_V \lesssim n^{-s}.$$

For certain examples (ridge functions) one has exactly $\varepsilon_n(\mathcal{K})_V \sim n^{-s}$: optimal nonlinear approximation rate achived by greedy algorithm.

Consider a dictionnary $\mathcal{D}$ in a Hilbert space V, such that $\|\varphi\|_V \leq 1$ for all $\varphi \in \mathcal{D}$, and the model class

$$\mathcal{K} := \Big\{ v = \sum c_j \varphi_j \; : \; \varphi_j \in \mathcal{D}, \; \sum |c_j| \leq 1 \Big\}.$$

The OMP algorithm recursively produces

$$u_n = \sum_{j=1}^{n} c_j \varphi_j = P_{V_n} u, \quad V_n = \mathrm{span}\{\varphi_1, \dots, \varphi_n\},$$

by selecting $\varphi_n \in \mathcal{D}$ maximizing $\frac{\langle u - P_{V_{n-1}} u, \varphi \rangle}{\|\varphi\|_V}$ over $\varphi \in \mathcal{D}$.

DeVore-Temlyakov (1998) : $u \in \mathcal{K} \implies \|u - u_n\|_V \leq n^{-1/2}$.

Siegel-Xu (2021) : consider smoothly parametrized dictionnaries

$$\mathcal{D} = \{\varphi = \mathcal{P}(\theta) \; : \; \theta \in \mathcal{M}\},$$

where $\mathcal{M}$ is d-dimensional compact manifold and $\mathcal{P}$ is $\mathcal{C}^r$. In this setting

$$u \in \mathcal{K} \implies \|u - u_n\|_V \lesssim n^{-s}, \quad s = \frac{1}{2} + \frac{r}{d} \quad \text{and} \quad \varepsilon_n(\mathcal{K})_V \lesssim n^{-s}.$$

For certain examples (ridge functions) one has exactly $\varepsilon_n(\mathcal{K})_V \sim n^{-s}$: optimal nonlinear approximation rate achived by greedy algorithm.

Recent results in this direction

Consider a dictionnary $\mathcal{D}$ in a Hilbert space V, such that $\|\varphi\|_V \leq 1$ for all $\varphi \in \mathcal{D}$, and the model class

$$\mathcal{K} := \Big\{ v = \sum c_j \varphi_j \ : \ \varphi_j \in \mathcal{D}, \ \sum |c_j| \leq 1 \Big\}.$$

The OMP algorithm recursively produces

$$u_n = \sum_{j=1}^{n} c_j \varphi_j = P_{V_n} u, \quad V_n = \mathrm{span}\{\varphi_1, \dots, \varphi_n\},$$

by selecting $\varphi_n \in \mathcal{D}$ maximizing $\frac{\langle u - P_{V_{n-1}} u, \varphi \rangle}{\|\varphi\|_V}$ over $\varphi \in \mathcal{D}$.

DeVore-Temlyakov (1998) : $u \in \mathcal{K} \implies \|u - u_n\|_V \leq n^{-1/2}$.

Siegel-Xu (2021) : consider smoothly parametrized dictionnaries

$$\mathcal{D} = \{\varphi = \mathcal{P}(\theta) \ : \ \theta \in \mathcal{M}\},$$

where $\mathcal{M}$ is d-dimensional compact manifold and $\mathcal{P}$ is $\mathcal{C}^r$. In this setting

$$u \in \mathcal{K} \implies \|u - u_n\|_V \lesssim n^{-s}, \quad s = \frac{1}{2} + \frac{r}{d} \quad \text{and} \quad \varepsilon_n(\mathcal{K})_V \lesssim n^{-s}.$$

For certain examples (ridge functions) one has exactly $\varepsilon_n(\mathcal{K})_V \sim n^{-s}$: optimal nonlinear approximation rate achived by greedy algorithm.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x} \in D^m$ an $\Phi_\mathbf{x} : \mathbb{R}^m \to V$.

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

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

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x} \in D^m$ an $\Phi_\mathbf{x} : \mathbb{R}^m \to V$.

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x} \in D^m$ an $\Phi_\mathbf{x} : \mathbb{R}^m \to V$.

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

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

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x} \in D^m$ an $\Phi_{\mathbf{x}} : \mathbb{R}^m \to V$.

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x} \in D^m$ an $\Phi_\mathbf{x} : \mathbb{R}^m \to V$.

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x}\in D^m$ an $\Phi_\mathbf{x}:\mathbb{R}^m\to V$.

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Sampling numbers

Linear and nonlinear widths measure the approximability of a class $\mathcal{K}$ by linear or nonlinear families of complexity n (data compression).

Sampling numbers (IBC - optimal recovery) measure the approximability of a class $\mathcal{K}$ from n point evaluations (critical quantity when such evaluations are costly).

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

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 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 $\mathbf{x} \in D^m$ an $\Phi_\mathbf{x} : \mathbb{R}^m \to V$.

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

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

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

Gelfand numbers : replace point evaluation $u(x^i)$ by general linear measurement $\ell_i(u)$.

More realistic : the linear functionals ℓ_i are picked from a restricted dictionnary $\mathcal{D}$.

Adaptive sampling numbers $r_m^{\rm adap}(\mathcal{K})_V$: allow x^k to depend on $u(x^1),\dots,u(x^{k-1})$.

Are sampling numbers controlled by widths ?

Positive results for linear widths : let $V = L^2(D,\mu)$ for some measure μ.

Cohen-Dolbeault (2021) : $\rho^{\rm rand}_{C_0 n}(\mathcal{K})_V \leq C_1 d_n(\mathcal{K})_V$ for some fixed $C_0, C_1 > 1$.

Temlyakov (2020) : if $\mu(D) < \infty$, then $\rho^{\rm det}_{C_0 n}(\mathcal{K})_V \leq C_1 d_n(\mathcal{K})_{L^\infty}$ with $C_0 = 1 + \varepsilon$.

These results use weighted least squares as reconstruction and point sparsification techniques due to Spielman, Markus, Srivastava, Nitzan, Olevskii, Ulanovskii.

Negative results for nonlinear widths : there exists classes $\mathcal{K}$ such that sampling numbers $r^{\rm det}_n(\mathcal{K})_V$ and $r^{\rm rand}_n(\mathcal{K})_V$ decay much slower that quantities $\delta_{n,L}(\mathcal{K})_V$ or $\varepsilon_n(\mathcal{K})_V$ measuring nonlinear approximation capabilities.

Example : with $V = L^2([0,1])$, consider the class $\mathcal{K} := \{\chi_{[a,b]} \ : \ a, b \in [0,1]\}$.

Then $r^{\rm det}_n(\mathcal{K})_V \sim r^{\rm rand}_n(\mathcal{K})_V \sim d_n(\mathcal{K})_V \sim n^{-1/2}$, but $\delta_{n,L}(\mathcal{K})_V \sim \varepsilon_n(\mathcal{K})_V \sim \exp(-cn)$.

On the other hand, consider the class $\mathcal{K} := \{\chi_{[0,a]} \ : \ a \in [0,1]\}$.

Then $d_n(\mathcal{K})_V \sim n^{-1/2}$ but $r^{\rm det}_n(\mathcal{K})_V \sim \exp(-cn)$ (use adaptive dichotomy).

Question : identify relevant classes of functions for which sampling number behave much better than linear widths, closer to nonlinear widths and entropy numbers.

Are sampling numbers controlled by widths ?

Positive results for linear widths : let $V = L^2(D, \mu)$ for some measure μ.

Cohen-Dolbeault (2021) : $\rho^{\rm rand}_{C_0 n}(\mathcal{K})_V \leq C_1 d_n(\mathcal{K})_V$ for some fixed $C_0, C_1 > 1$.

Temlyakov (2020) : if $\mu(D) < \infty$, then $\rho^{\rm det}_{C_0 n}(\mathcal{K})_V \leq C_1 d_n(\mathcal{K})_{L^\infty}$ with $C_0 = 1 + \varepsilon$.

These results use weighted least squares as reconstruction and point sparsification techniques due to Spielman, Markus, Srivastava, Nitzan, Olevskii, Ulanovskii.

Negative results for nonlinear widths : there exists classes $\mathcal{K}$ such that sampling numbers $r^{\rm det}_n(\mathcal{K})_V$ and $r^{\rm rand}_n(\mathcal{K})_V$ decay much slower that quantities $\delta_{n,L}(\mathcal{K})_V$ or $\varepsilon_n(\mathcal{K})_V$ measuring nonlinear approximation capabilities.

Example : with $V = L^2([0,1])$, consider the class $\mathcal{K} := \{\chi_{[a,b]} \ : \ a, b \in [0,1]\}$.

Then $r^{\rm det}_n(\mathcal{K})_V \sim r^{\rm rand}_n(\mathcal{K})_V \sim d_n(\mathcal{K})_V \sim n^{-1/2}$, but $\delta_{n,L}(\mathcal{K})_V \sim \varepsilon_n(\mathcal{K})_V \sim \exp(-cn)$.

On the other hand, consider the class $\mathcal{K} := \{\chi_{[0,a]} \ : \ a \in [0,1]\}$.

Then $d_n(\mathcal{K})_V \sim n^{-1/2}$ but $r^{\rm det}_n(\mathcal{K})_V \sim \exp(-cn)$ (use adaptive dichotomy).

Question : identify relevant classes of functions for which sampling number behave much better than linear widths, closer to nonlinear widths and entropy numbers.

Are sampling numbers controlled by widths ?

Positive results for linear widths : let $V = L^2(D, \mu)$ for some measure μ.

Cohen-Dolbeault (2021) : $\rho^{\rm rand}_{C_0 n}(\mathcal{K})_V \leq C_1 d_n(\mathcal{K})_V$ for some fixed $C_0, C_1 > 1$.

Temlyakov (2020) : if $\mu(D) < \infty$, then $\rho^{\rm det}_{C_0 n}(\mathcal{K})_V \leq C_1 d_n(\mathcal{K})_{L^\infty}$ with $C_0 = 1 + \varepsilon$.

These results use weighted least squares as reconstruction and point sparsification techniques due to Spielman, Markus, Srivastava, Nitzan, Olevskii, Ulanovskii.

Negative results for nonlinear widths : there exists classes $\mathcal{K}$ such that sampling numbers $r^{\rm det}_n(\mathcal{K})_V$ and $r^{\rm rand}_n(\mathcal{K})_V$ decay much slower that quantities $\delta_{n,L}(\mathcal{K})_V$ or $\varepsilon_n(\mathcal{K})_V$ measuring nonlinear approximation capabilities.

Example : with $V = L^2([0,1])$, consider the class $\mathcal{K} := \{\chi_{[a,b]} \; : \; a, b \in [0,1]\}$.

Then $r^{\rm det}_n(\mathcal{K})_V \sim r^{\rm rand}_n(\mathcal{K})_V \sim d_n(\mathcal{K})_V \sim n^{-1/2}$, but $\delta_{n,L}(\mathcal{K})_V \sim \varepsilon_n(\mathcal{K})_V \sim \exp(-cn)$.

On the other hand, consider the class $\mathcal{K} := \{\chi_{[0,a]} \; : \; a \in [0,1]\}$.

Then $d_n(\mathcal{K})_V \sim n^{-1/2}$ but $r^{\rm det}_n(\mathcal{K})_V \sim \exp(-cn)$ (use adaptive dichotomy).

Question : identify relevant classes of functions for which sampling number behave much better than linear widths, closer to nonlinear widths and entropy numbers.

Are sampling numbers controlled by widths ?

Positive results for linear widths : let $V = L^2(D, \mu)$ for some measure μ.

Cohen-Dolbeault (2021) : $\rho^{\rm rand}_{C_0 n}(\mathcal{K})_V \le C_1 d_n(\mathcal{K})_V$ for some fixed $C_0, C_1 > 1$.

Temlyakov (2020) : if $\mu(D) < \infty$, then $\rho^{\rm det}_{C_0 n}(\mathcal{K})_V \le C_1 d_n(\mathcal{K})_{L^\infty}$ with $C_0 = 1 + \varepsilon$.

These results use weighted least squares as reconstruction and point sparsification techniques due to Spielman, Markus, Srivastava, Nitzan, Olevskii, Ulanovskii.

Negative results for nonlinear widths : there exists classes $\mathcal{K}$ such that sampling numbers $r^{\rm det}_n(\mathcal{K})_V$ and $r^{\rm rand}_n(\mathcal{K})_V$ decay much slower that quantities $\delta_{n,L}(\mathcal{K})_V$ or $\varepsilon_n(\mathcal{K})_V$ measuring nonlinear approximation capabilities.

Example : with $V = L^2([0,1])$, consider the class $\mathcal{K} := \{\chi_{[a,b]} \ : \ a, b \in [0,1]\}$.

Then $r^{\rm det}_n(\mathcal{K})_V \sim r^{\rm rand}_n(\mathcal{K})_V \sim d_n(\mathcal{K})_V \sim n^{-1/2}$, but $\delta_{n,L}(\mathcal{K})_V \sim \varepsilon_n(\mathcal{K})_V \sim \exp(-cn)$.

On the other hand, consider the class $\mathcal{K} := \{\chi_{[0,a]} \ : \ a \in [0,1]\}$.

Then $d_n(\mathcal{K})_V \sim n^{-1/2}$ but $r^{\rm det}_n(\mathcal{K})_V \sim \exp(-cn)$ (use adaptive dichotomy).

Question : identify relevant classes of functions for which sampling number behave much better than linear widths, closer to nonlinear widths and entropy numbers.

References

A. Pinkus, *n*-width in Approximation Theory, Springer 2012.

A. Cohen and R. DeVore, Approximation of high-dimensional PDEs, Acta Numerica, 2015.

R. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, Manuscripta Mathematica, 1989.

V. Temlyakov, Nonlinear Kolmogorov widths, Mathematical Notes, 1998.

A. Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk, Optimal stable nonlinear approximation, Foundation of Computational Mathematics, 2021.

Zuowei Shen, Haizhao Yang, and Shijun Zhang, Deep network approximation characterized by number of neurons, Communications in Computational Physics, 2020.

J. Siegel and J. Xu, Improved approximation properties of dictionaries and applications to neural networks, 2021.

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

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

THANKS !

References

A. Pinkus, *n*-width in Approximation Theory, Springer 2012.

A. Cohen and R. DeVore, Approximation of high-dimensional PDEs, Acta Numerica, 2015.

R. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, Manuscripta Mathematica, 1989.

V. Temlyakov, Nonlinear Kolmogorov widths, Mathematical Notes, 1998.

A. Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk, Optimal stable nonlinear approximation, Foundation of Computational Mathematics, 2021.

Zuowei Shen, Haizhao Yang, and Shijun Zhang, Deep network approximation characterized by number of neurons, Communications in Computational Physics, 2020.

J. Siegel and J. Xu, Improved approximation properties of dictionaries and applications to neural networks, 2021.

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

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

THANKS !