Optimal linear and non-linear dimensionality reduction
Theory and algorithms

Albert Cohen

Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris

Oslo, 27-06-2024

Collaborators: P. Binev, W. Dahmen, R. DeVore, M. Dolbeault, O. Mula, G. Petrova, A. Somacal, P. Wojtaszczyk.

Curves and Surfaces 2026

Organized by SMAI-SIGMA

Taking place in St Malo, France, June 2026 (first or second week)

following Chamonix 1990, 1993, 1996; Saint-Malo 1999, 2002;
Avignon 2006, 2010;Paris 2014 ; Arcachon 2018, 2022

Save the date !

Dimensionality reduction - reduced modeling

Approximation of an unknown multivariate function $u \in V$ defined on some domain $\Omega \subset \mathbb{R}^d$ by simpler functions depending on finitely many parameters is used in various contexts.

Forward simulation : numerical computation of u solution to a given PDE.

Inverse problems : access to limited observation $\ell(u) = (\ell_1(u), \dots, \ell_m(u)) \in \mathbb{R}^m$.

Sampling : access to point values $u(x^1), \dots, u(x^m)$.

Here V is a Banach space equiped with norm $\|\cdot\| = \|\cdot\|_V$.

Prior information: $u \in \mathcal{K}$ compact class of V (smoothness class, family of solutions to parametrized PDEs...).

Computational strategies lead to approximation of u by $\tilde{u} \in V_n$ that can be described by $n \leq m$ parameters (or more generally $\mathcal{O}(n)$).

The set V_n can be a linear or nonlinear space

Approximation of an unknown multivariate function $u \in V$ defined on some domain $\Omega \subset \mathbb{R}^d$ by simpler functions depending on finitely many parameters is used in various contexts.

Forward simulation : numerical computation of u solution to a given PDE.

Inverse problems : access to limited observation $\ell(u) = (\ell_1(u), \dots, \ell_m(u)) \in \mathbb{R}^m$.

Sampling : access to point values $u(x^1), \dots, u(x^m)$.

Here V is a Banach space equiped with norm $\|\cdot\| = \|\cdot\|_V$.

Prior information: $u \in \mathcal{K}$ compact class of V (smoothness class, family of solutions to parametrized PDEs...).

Computational strategies lead to approximation of u by $\tilde{u} \in V_n$ that can be described by $n \leq m$ parameters (or more generally $\mathcal{O}(n)$).

The set V_n can be a linear or nonlinear space

Dimensionality reduction - reduced modeling

Approximation of an unknown multivariate function $u \in V$ defined on some domain $\Omega \subset \mathbb{R}^d$ by simpler functions depending on finitely many parameters is used in various contexts.

Forward simulation : numerical computation of u solution to a given PDE.

Inverse problems : access to limited observation $\ell(u) = (\ell_1(u), \dots, \ell_m(u)) \in \mathbb{R}^m$.

Sampling : access to point values $u(x^1), \dots, u(x^m)$.

Here V is a Banach space equiped with norm $\|\cdot\| = \|\cdot\|_V$.

Prior information: $u \in \mathcal{K}$ compact class of V (smoothness class, family of solutions to parametrized PDEs...).

Computational strategies lead to approximation of u by $\tilde{u} \in V_n$ that can be described by $n \leq m$ parameters (or more generally $\mathcal{O}(n)$).

The set V_n can be a linear or nonlinear space

Dimensionality reduction - reduced modeling

Approximation of an unknown multivariate function $u \in V$ defined on some domain $\Omega \subset \mathbb{R}^d$ by simpler functions depending on finitely many parameters is used in various contexts.

Forward simulation : numerical computation of u solution to a given PDE.

Inverse problems : access to limited observation $\ell(u) = (\ell_1(u), \dots, \ell_m(u)) \in \mathbb{R}^m$.

Sampling : access to point values $u(x^1), \dots, u(x^m)$.

Here V is a Banach space equiped with norm $\|\cdot\| = \|\cdot\|_V$.

Prior information: $u \in \mathcal{K}$ compact class of V (smoothness class, family of solutions to parametrized PDEs...).

Computational strategies lead to approximation of u by $\tilde{u} \in V_n$ that can be described by $n \leq m$ parameters (or more generally $\mathcal{O}(n)$).

The set V_n can be a linear or nonlinear space

Classical choice of approximation spaces.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k\geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k\geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k \geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k \geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Classical choice of approximation spaces.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k\geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Classical choice of approximation spaces.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k\geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k \geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Classical choice of approximation spaces.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k \geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k\geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Classical choice of approximation spaces.

Linear spaces:

- Algebraic polynomials: $V_n = \mathbb{P}_n$. When $d > 1$, we may set $V_n = \mathbb{P}_k$ with $n = \binom{k+d}{d}$.

- Span of the n first elements $\{e_1, \dots, e_n\}$ from a given basis $(e_k)_{k\geq 1}$ of V, for example trigonometric polynomials.

- Piecewise polynomials, splines or finite elements on a fixed partition of cardinality n.

Nonlinear spaces:

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

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

- 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 V_n is the set of such functions when the total number of parameters does not exceed n.

Standard questions in approximation theory

For a given family $(V_n)_{n\geq 0}$ of such classical spaces, a standard question is the study of the best approximation error

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

Typically, one tries to understand which properties of u (smoothness, sparsity...) ensure a certain rate of decay $e_n(u) \leq Cn^{-r}$.

This allows to understand for a given prior model class $\mathcal{K}$ the asymptotic behaviour of $\text{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} e_n(u)$.

A related standard question is the construction of near best approximations: simple computational map $u \mapsto u_n \in V_n$ such that $\|u - u_n\| \leq Ce_n(u)$ for some fixed $C \geq 1$.

Optimal dimensionality reduction

A less standard question is that of an optimal choice of V_n.

For a given model class $\mathcal{K}$, can we find a linear or nonlinear space V_n that best approximate $\mathcal{K}$, that is, make $\text{dist}(\mathcal{K}, V_n)_V$ as small as possible.

Such spaces may in certain cases significantly differ from the classical examples listed above, and may have no simple description.

Standard questions in approximation theory

For a given family $(V_n)_{n\geq 0}$ of such classical spaces, a standard question is the study of the best approximation error

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

Typically, one tries to understand which properties of u (smoothness, sparsity...) ensure a certain rate of decay $e_n(u) \leq Cn^{-r}$.

This allows to understand for a given prior model class $\mathcal{K}$ the asymptotic behaviour of $\mathrm{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} e_n(u)$.

A related standard question is the construction of near best approximations: simple computational map $u \mapsto u_n \in V_n$ such that $\|u - u_n\| \leq Ce_n(u)$ for some fixed $C \geq 1$.

Optimal dimensionality reduction

A less standard question is that of an optimal choice of V_n.

For a given model class $\mathcal{K}$, can we find a linear or nonlinear space V_n that best approximate $\mathcal{K}$, that is, make $\mathrm{dist}(\mathcal{K}, V_n)_V$ as small as possible.

Such spaces may in certain cases significantly differ from the classical examples listed above, and may have no simple description.

Standard questions in approximation theory

For a given family $(V_n)_{n\geq 0}$ of such classical spaces, a standard question is the study of the best approximation error

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

Typically, one tries to understand which properties of u (smoothness, sparsity...) ensure a certain rate of decay $e_n(u) \leq Cn^{-r}$.

This allows to understand for a given prior model class $\mathcal{K}$ the asymptotic behaviour of $\mathrm{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} e_n(u)$.

A related standard question is the construction of near best approximations: simple computational map $u \mapsto u_n \in V_n$ such that $\|u - u_n\| \leq Ce_n(u)$ for some fixed $C \geq 1$.

Optimal dimensionality reduction

A less standard question is that of an optimal choice of V_n.

For a given model class $\mathcal{K}$, can we find a linear or nonlinear space V_n that best approximate $\mathcal{K}$, that is, make $\mathrm{dist}(\mathcal{K}, V_n)_V$ as small as possible.

Such spaces may in certain cases significantly differ from the classical examples listed above, and may have no simple description.

Standard questions in approximation theory

For a given family $(V_n)_{n\geq 0}$ of such classical spaces, a standard question is the study of the best approximation error

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

Typically, one tries to understand which properties of u (smoothness, sparsity...) ensure a certain rate of decay $e_n(u) \leq Cn^{-r}$.

This allows to understand for a given prior model class $\mathcal{K}$ the asymptotic behaviour of $\mathrm{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} e_n(u)$.

A related standard question is the construction of near best approximations: simple computational map $u \mapsto u_n \in V_n$ such that $\|u - u_n\| \leq Ce_n(u)$ for some fixed $C \geq 1$.

Optimal dimensionality reduction

A less standard question is that of an optimal choice of V_n.

For a given model class $\mathcal{K}$, can we find a linear or nonlinear space V_n that best approximate $\mathcal{K}$, that is, make $\mathrm{dist}(\mathcal{K}, V_n)_V$ as small as possible.

Such spaces may in certain cases significantly differ from the classical examples listed above, and may have no simple description.

Standard questions in approximation theory

For a given family $(V_n)_{n\geq 0}$ of such classical spaces, a standard question is the study of the best approximation error

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

Typically, one tries to understand which properties of u (smoothness, sparsity...) ensure a certain rate of decay $e_n(u) \leq Cn^{-r}$.

This allows to understand for a given prior model class $\mathcal{K}$ the asymptotic behaviour of $\mathrm{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} e_n(u)$.

A related standard question is the construction of near best approximations: simple computational map $u \mapsto u_n \in V_n$ such that $\|u - u_n\| \leq Ce_n(u)$ for some fixed $C \geq 1$.

Optimal dimensionality reduction

A less standard question is that of an optimal choice of V_n.

For a given model class $\mathcal{K}$, can we find a linear or nonlinear space V_n that best approximate $\mathcal{K}$, that is, make $\mathrm{dist}(\mathcal{K}, V_n)_V$ as small as possible.

Such spaces may in certain cases significantly differ from the classical examples listed above, and may have no simple description.

The offline/online paradigm

We are interested in applications where we might need to search for many instances of $u \in \mathcal{K}$:

- Parametrized PDE's $\mathcal{P}(u, y) = 0$ giving rise to a solution map $y \mapsto u(y) \in V$ and manifold $\mathcal{K} := \{u(y) \ : \ y \in Y\}$.

- Applications requiring multiple queries of the solution map: optimization/control (y is deterministic), or uncertainty quantification (y is random) or inverse problems (identify y from observations of $u(y)$).

- Or we may want to recover many instances of u in a model class $\mathcal{K}$, from their observations $z = (\ell_1(u), \dots, \ell_m(u))$.

Offline stage: design a space V_n of moderate dimension n that is optimally taylored to the class $\mathcal{K}$. This can be computationally intensive but it is only done once.

Online stage: for each required parameter instance y (or.data z), compute an approximation $u_n(y)$ (or u_n) in V_n, by a hopefully fast computation : n numbers to compute.

The offline/online paradigm

We are interested in applications where we might need to search for many instances of $u \in \mathcal{K}$:

- Parametrized PDE's $\mathcal{P}(u, y) = 0$ giving rise to a solution map $y \mapsto u(y) \in V$ and manifold $\mathcal{K} := \{u(y) \ : \ y \in Y\}$.

- Applications requiring multiple queries of the solution map: optimization/control (y is deterministic), or uncertainty quantification (y is random) or inverse problems (identify y from observations of $u(y)$).

- Or we may want to recover many instances of u in a model class $\mathcal{K}$, from their observations $z = (\ell_1(u), \dots, \ell_m(u))$.

Offline stage: design a space V_n of moderate dimension n that is optimally taylored to the class $\mathcal{K}$. This can be computationally intensive but it is only done once.

Online stage: for each required parameter instance y (or.data z), compute an approximation $u_n(y)$ (or u_n) in V_n, by a hopefully fast computation : n numbers to compute.

The offline/online paradigm

We are interested in applications where we might need to search for many instances of $u \in \mathcal{K}$:

- Parametrized PDE's $\mathcal{P}(u, y) = 0$ giving rise to a solution map $y \mapsto u(y) \in V$ and manifold $\mathcal{K} := \{u(y) \; : \; y \in Y\}$.

- Applications requiring multiple queries of the solution map: optimization/control (y is deterministic), or uncertainty quantification (y is random) or inverse problems (identify y from observations of $u(y)$).

- Or we may want to recover many instances of u in a model class $\mathcal{K}$, from their observations $z = (\ell_1(u), \dots, \ell_m(u))$.

Offline stage: design a space V_n of moderate dimension n that is optimally taylored to the class $\mathcal{K}$. This can be computationally intensive but it is only done once.

Online stage: for each required parameter instance y (or.data z), compute an approximation $u_n(y)$ (or u_n) in V_n, by a hopefully fast computation : n numbers to compute.

The offline/online paradigm

We are interested in applications where we might need to search for many instances of $u \in \mathcal{K}$:

- Parametrized PDE's $\mathcal{P}(u, y) = 0$ giving rise to a solution map $y \mapsto u(y) \in V$ and manifold $\mathcal{K} := \{u(y) \ : \ y \in Y\}$.

- Applications requiring multiple queries of the solution map: optimization/control (y is deterministic), or uncertainty quantification (y is random) or inverse problems (identify y from observations of $u(y)$).

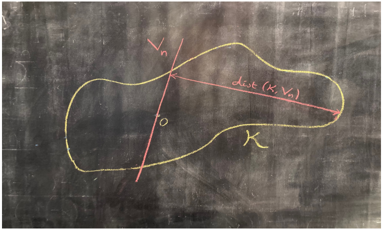
- Or we may want to recover many instances of u in a model class $\mathcal{K}$, from their observations $z = (\ell_1(u), \dots, \ell_m(u))$.

Offline stage: design a space V_n of moderate dimension n that is optimally taylored to the class $\mathcal{K}$. This can be computationally intensive but it is only done once.

Online stage: for each required parameter instance y (or.data z), compute an approximation $u_n(y)$ (or u_n) in V_n, by a hopefully fast computation : n numbers to compute.

Kolmogorov n-widths are defined as

$$d_n = d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \text{dist}(\mathcal{K}, V_n)_V, \qquad \text{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u - v\|_V,$$

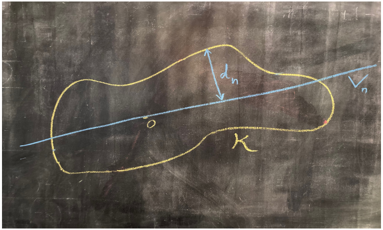


The quantity $d_n(\mathcal{K})_V$ can be viewed as a benchmark/bottleneck for numerical methods applied to the elements from $\mathcal{K}$ that create approximations from linear spaces: interpolation, projection, least squares, Galerkin methods for solving PDEs...

The optimal space V_n achieving the infimum may not exist (one often assumes it exists in order to avoid limiting arguments). Its exact construction is usually out of reach.

Kolmogorov n-widths are defined as

$$d_n = d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \operatorname{dist}(\mathcal{K}, V_n)_V, \qquad \operatorname{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_V,$$

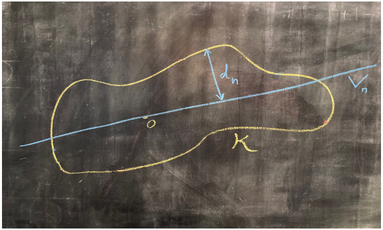


The quantity $d_n(\mathcal{K})_V$ can be viewed as a benchmark/bottleneck for numerical methods applied to the elements from $\mathcal{K}$ that create approximations from linear spaces: interpolation, projection, least squares, Galerkin methods for solving PDEs...

The optimal space V_n achieving the infimum may not exist (one often assumes it exists in order to avoid limiting arguments). Its exact construction is usually out of reach.

Kolmogorov n-widths are defined as

$$d_n = d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \operatorname{dist}(\mathcal{K}, V_n)_V, \qquad \operatorname{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_V,$$



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...

The optimal space V_n achieving the infimum may not exist (one often assumes it exists in order to avoid limiting arguments). Its exact construction is usually out of reach.

Optimality in linear and nonlinear dimensionality reduction

Kolmogorov n-widths are the benchmark for approximation by linear spaces.

$$d_n = d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \operatorname{dist}(\mathcal{K}, V_n)_V, \qquad \operatorname{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_V,$$

or equivalently

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

with infimum over all linear recovery maps $R : \mathbb{R}^n \to V$ and continuous encoding maps $E : V \to \mathbb{R}^n$.

Similar quantities can be defined with other prescriptions on E and R.

R \ E	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

Optimality in linear and nonlinear dimensionality reduction

Kolmogorov n-widths are the benchmark for approximation by linear spaces.

$$d_n = d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \operatorname{dist}(\mathcal{K}, V_n)_V, \qquad \operatorname{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_V,$$

or equivalently

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

with infimum over all linear recovery maps $R : \mathbb{R}^n \to V$ and continuous encoding maps $E : V \to \mathbb{R}^n$.

Similar quantities can be defined with other prescriptions on E and R.

R \ E	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

Optimality in linear and nonlinear dimensionality reduction

Kolmogorov n-widths are the benchmark for approximation by linear spaces.

$$d_n = d_n(\mathcal{K})_V := \inf_{\dim(V_n)=n} \operatorname{dist}(\mathcal{K}, V_n)_V, \qquad \operatorname{dist}(\mathcal{K}, V_n)_V := \max_{u\in\mathcal{K}} \min_{v\in V_n} \|u-v\|_V,$$

or equivalently

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

with infimum over all linear recovery maps $R : \mathbb{R}^n \to V$ and continuous encoding maps $E : V \to \mathbb{R}^n$.

Similar quantities can be defined with other prescriptions on E and R.

R \ E	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

Resulting quantities

E \ R	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

- $a_n = a_n(\mathcal{K})_V$ are the approximation numbers: $\inf \max_{u\in\mathcal{K}} \|u - Lu\|$, where the infimum is over all linear L of rank $\leq n$.

- Note that $a_n = d_n$ when V is a Hilbert space. Otherwise $a_n \geq d_n$.

- r_n and ρ_n are the sampling numbers: $\inf \max_{u\in\mathcal{K}} \|u - R(u(x^1),\dots,u(x^n))\|$, there the infimum is over all choices of sampling points $(x^1,\dots,x^n)$ and linear or continuous recovery maps.

- s_n are the sensing numbers: $\inf \max_{u\in\mathcal{K}} \|u - R(\ell_1(u),\dots,\ell_n(u))\|$, there the infimum is over all choices of linear functionals $\ell_1,\dots,\ell_n$ and continuous recovery maps.

- δ_n are the manifold widths. Benchmark for all general approximation methods.

Resulting quantities

R \ E	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

- $a_n = a_n(\mathcal{K})_V$ are the approximation numbers: $\inf \max_{u\in\mathcal{K}} \|u - Lu\|$, where the infimum is over all linear L of rank $\leq n$.

- Note that $a_n = d_n$ when V is a Hilbert space. Otherwise $a_n \geq d_n$.

- r_n and ρ_n are the sampling numbers: $\inf \max_{u\in\mathcal{K}} \|u - R(u(x^1), \dots, u(x^n))\|$, there the infimum is over all choices of sampling points $(x^1, \dots, x^n)$ and linear or continuous recovery maps.

- s_n are the sensing numbers: $\inf \max_{u\in\mathcal{K}} \|u - R(\ell_1(u), \dots, \ell_n(u))\|$, there the infimum is over all choices of linear functionals $\ell_1, \dots, \ell_n$ and continuous recovery maps.

- δ_n are the manifold widths. Benchmark for all general approximation methods.

Some variants

One may relax these definitions by imposing that the encoder is only defined on $\mathcal{K}$ and not on the whole of V.

This allows us, for example, to define sampling numbers for spaces such as $V = L^p(\Omega)$, assuming that all functions $u \in \mathcal{K}$ are continuous, that is, $\mathcal{K} \subset \mathcal{C}(\Omega)$.

On the other hand, one may want to strengthen these definitions by asking that the encoding and recovery map have some stability.

This leads to the notion of stable widths: for $L \geq 1$, we define

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

where the infimum is taken over all encoding and recovery maps such that

$$\|E(u) - E(v)\|_Z \leq L\|u - v\|_V \quad u, v \in V,$$

and

$$\|R(x) - R(y)\| \leq L\|x - y\|_Z, \quad x, y \in \mathbb{R}^n,$$

for some norm $\|\cdot\|_Z$ defined on $\mathbb{R}^n$.

Similarly, we may define the stable variants $(d_{n,L}, a_{n,L}, s_{n,L})$ to (a_n, d_n, s_n).

Some variants

One may relax these definitions by imposing that the encoder is only defined on $\mathcal{K}$ and not on the whole of V.

This allows us, for example, to define sampling numbers for spaces such as $V = L^p(\Omega)$, assuming that all functions $u \in \mathcal{K}$ are continuous, that is, $\mathcal{K} \subset \mathcal{C}(\Omega)$.

On the other hand, one may want to strengthen these definitions by asking that the encoding and recovery map have some stability.

This leads to the notion of stable widths: for $L \geq 1$, we define

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

where the infimum is taken over all encoding and recovery maps such that

$$\|E(u) - E(v)\|_Z \leq L\|u - v\|_V \quad u, v \in V,$$

and

$$\|R(x) - R(y)\| \leq L\|x - y\|_Z, \quad x, y \in \mathbb{R}^n,$$

for some norm $\|\cdot\|_Z$ defined on $\mathbb{R}^n$.

Similarly, we may define the stable variants $(d_{n,L}, a_{n,L}, s_{n,L})$ to (a_n, d_n, s_n).

One may relax these definitions by imposing that the encoder is only defined on $\mathcal{K}$ and not on the whole of V.

This allows us, for example, to define sampling numbers for spaces such as $V = L^p(\Omega)$, assuming that all functions $u \in \mathcal{K}$ are continuous, that is, $\mathcal{K} \subset \mathcal{C}(\Omega)$.

On the other hand, one may want to strengthen these definitions by asking that the encoding and recovery map have some stability.

This leads to the notion of stable widths: for $L \geq 1$, we define

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

where the infimum is taken over all encoding and recovery maps such that

$$\|E(u) - E(v)\|_Z \leq L\|u - v\|_V \quad u, v \in V,$$

and

$$\|R(x) - R(y)\| \leq L\|x - y\|_Z, \quad x, y \in \mathbb{R}^n,$$

for some norm $\|\cdot\|_Z$ defined on $\mathbb{R}^n$.

Similarly, we may define the stable variants $(d_{n,L}, a_{n,L}, s_{n,L})$ to (a_n, d_n, s_n).

One may relax these definitions by imposing that the encoder is only defined on $\mathcal{K}$ and not on the whole of V.

This allows us, for example, to define sampling numbers for spaces such as $V = L^p(\Omega)$, assuming that all functions $u \in \mathcal{K}$ are continuous, that is, $\mathcal{K} \subset \mathcal{C}(\Omega)$.

On the other hand, one may want to strengthen these definitions by asking that the encoding and recovery map have some stability.

This leads to the notion of stable widths: for $L \geq 1$, we define

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

where the infimum is taken over all encoding and recovery maps such that

$$\|E(u) - E(v)\|_Z \leq L\|u - v\|_V \quad u, v \in V,$$

and

$$\|R(x) - R(y)\| \leq L\|x - y\|_Z, \quad x, y \in \mathbb{R}^n,$$

for some norm $\|\cdot\|_Z$ defined on $\mathbb{R}^n$.

Similarly, we may define the stable variants $(d_{n,L}, a_{n,L}, s_{n,L})$ to (a_n, d_n, s_n).

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

Example : $V = L^\infty(I)$ where $I = [0,1] \subset \mathbb{R}$ and

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

Then

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

Estimating n-widths

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

Example : $V = L^\infty(I)$ where $I = [0,1] \subset \mathbb{R}$ and

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

Then

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

Estimating n-widths

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

Example : $V = L^\infty(I)$ where $I = [0,1] \subset \mathbb{R}$ and

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

Then

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

Example : $V = L^\infty(I)$ where $I = [0,1] \subset \mathbb{R}$ and

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

Then

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u \in \mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

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

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0, 1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

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

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \longrightarrow \infty$.

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

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

For a given class $\mathcal{K}$ of interest, we would like to estimate the above quantities from above and below.

Estimate from above: compute $\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V$ for particular admissible choices of E and R that are presumed to be near optimal.

Estimate from below is more tricky. We discuss further two possible ways.

For certains classes $\mathcal{K}$ and spaces V, all these quantities behave similar as $n \to \infty$.

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

$$\rho_n = r_n = a_n = d_n = s_n = \delta_n = \frac{1}{2n} \sim n^{-1}.$$

Upper bound: use piecewise linear reconstruction from n equispaced points.

General rate $n^{-\frac{s-t}{d}}$ for $V = W^{t,p}(Q)$ and $\mathcal{K} = \mathcal{U}(W^{s,p}(Q))$, with $Q = [0,1]^d$.

Thus for such prior classes, nonlinear approaches are not beneficial.

For other classes discussed further, one may have $\delta_n << d_n$.

Lower bounds: using continuity

Continuity of the encoder E essential in the definition of the nonlinear widths δ_n. Indeed, if we drop this assumption, then we would have

$$\delta_n(\mathcal{K}) = \inf_D \sup_{u\in\mathcal{K}} \inf_{x\in\mathbb{R}^n} \|u - D(x)\|,$$

which is the minimal distance of $\mathcal{K}$ to n-dimensional parametrized manifolds.

This quantity is 0 even for $n = 1$: space filling continuous curves.

Assuming continuity of E, we obtain lower bounds using Borsuk-Ulam theorem: if $W \subset V$ is any $n+1$ dimensional space, then any continuous map E from the n-sphere $S_n = \partial B_W$ to $\mathbb{R}^n$ admits a point $u^* \in S_n$ such that $E(u^*) = E(-u^*)$.

Thus, if $\mathcal{K}$ contains a rescaled ball rB_W of an $n+1$-dimensional space W, there exists $u^*, -u^* \in \mathcal{K}$ such that $R(E(u^*)) = R(E(-u^*))$ for any recovery map, and thus

$$\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V \geq r.$$

It follows that

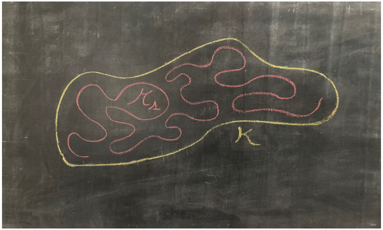
$$\delta_n \geq b_n,$$

Continuity of the encoder E essential in the definition of the nonlinear widths δ_n. Indeed, if we drop this assumption, then we would have

$$\delta_n(\mathcal{K}) = \inf_D \sup_{u\in\mathcal{K}} \inf_{x\in\mathbb{R}^n} \|u - D(x)\|,$$

which is the minimal distance of $\mathcal{K}$ to n-dimensional parametrized manifolds.

This quantity is 0 even for $n = 1$: space filling continuous curves.



Assuming continuity of E, we obtain lower bounds using Borsuk-Ulam theorem: if $W \subset V$ is any $n+1$ dimensional space, then any continuous map E from the n-sphere $S_n = \partial B_W$ to $\mathbb{R}^n$ admits a point $u^* \in S_n$ such that $E(u^*) = E(-u^*)$.

Thus, if $\mathcal{K}$ contains a rescaled ball rB_W of an $n+1$-dimensional space W, there exists $u^*, -u^* \in \mathcal{K}$ such that $R(E(u^*)) = R(E(-u^*))$ for any recovery map, and thus

$$\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V \geq r.$$

Continuity of the encoder E essential in the definition of the nonlinear widths δ_n. Indeed, if we drop this assumption, then we would have

$$\delta_n(\mathcal{K}) = \inf_D \sup_{u\in\mathcal{K}} \inf_{x\in\mathbb{R}^n} \|u - D(x)\|,$$

which is the minimal distance of $\mathcal{K}$ to n-dimensional parametrized manifolds.

This quantity is 0 even for $n = 1$: space filling continuous curves.

Assuming continuity of E, we obtain lower bounds using Borsuk-Ulam theorem: if $W \subset V$ is any $n+1$ dimensional space, then any continuous map E from the n-sphere $S_n = \partial B_W$ to $\mathbb{R}^n$ admits a point $u^* \in S_n$ such that $E(u^*) = E(-u^*)$.

Thus, if $\mathcal{K}$ contains a rescaled ball rB_W of an $n+1$-dimensional space W, there exists $u^*, -u^* \in \mathcal{K}$ such that $R(E(u^*)) = R(E(-u^*))$ for any recovery map, and thus

$$\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V \geq r.$$

It follows that

$$\delta_n \geq b_n,$$

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

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one finds that $b_n \geq \frac{1}{2n}$.

Lower bounds: using continuity

Continuity of the encoder E essential in the definition of the nonlinear widths δ_n. Indeed, if we drop this assumption, then we would have

$$\delta_n(\mathcal{K}) = \inf_D \sup_{u\in\mathcal{K}} \inf_{x\in\mathbb{R}^n} \|u - D(x)\|,$$

which is the minimal distance of $\mathcal{K}$ to n-dimensional parametrized manifolds.

This quantity is 0 even for $n = 1$: space filling continuous curves.

Assuming continuity of E, we obtain lower bounds using Borsuk-Ulam theorem: if $W \subset V$ is any $n+1$ dimensional space, then any continuous map E from the n-sphere $S_n = \partial B_W$ to $\mathbb{R}^n$ admits a point $u^* \in S_n$ such that $E(u^*) = E(-u^*)$.

Thus, if $\mathcal{K}$ contains a rescaled ball rB_W of an $n+1$-dimensional space W, there exists $u^*, -u^* \in \mathcal{K}$ such that $R(E(u^*)) = R(E(-u^*))$ for any recovery map, and thus

$$\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V \geq r.$$

It follows that

$$\delta_n \geq b_n,$$

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

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one finds that $b_n \geq \frac{1}{2n}$.

Continuity of the encoder E essential in the definition of the nonlinear widths δ_n. Indeed, if we drop this assumption, then we would have

$$\delta_n(\mathcal{K}) = \inf_D \sup_{u\in\mathcal{K}} \inf_{x\in\mathbb{R}^n} \|u - D(x)\|,$$

which is the minimal distance of $\mathcal{K}$ to n-dimensional parametrized manifolds.

This quantity is 0 even for $n = 1$: space filling continuous curves.

Assuming continuity of E, we obtain lower bounds using Borsuk-Ulam theorem: if $W \subset V$ is any $n+1$ dimensional space, then any continuous map E from the n-sphere $S_n = \partial B_W$ to $\mathbb{R}^n$ admits a point $u^* \in S_n$ such that $E(u^*) = E(-u^*)$.

Thus, if $\mathcal{K}$ contains a rescaled ball rB_W of an $n+1$-dimensional space W, there exists $u^*, -u^* \in \mathcal{K}$ such that $R(E(u^*)) = R(E(-u^*))$ for any recovery map, and thus

$$\max_{u\in\mathcal{K}} \|u - R(E(u))\|_V \geq r.$$

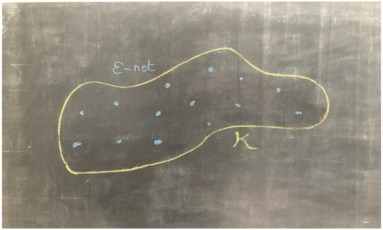
It follows that

$$\delta_n \geq b_n,$$

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

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one finds that $b_n \geq \frac{1}{2n}$.

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



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

$$(n+1)^s \varepsilon_n \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_m, \quad n \geq 0$$

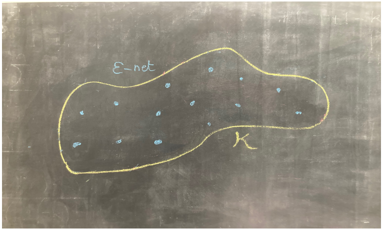
In particular

$$d_n \lesssim n^{-s}, \quad n \geq 0 \implies \varepsilon_n \lesssim n^{-s}, \quad n \geq 0.$$

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one can prove that $\varepsilon_n \geq cn^{-1}$.

Carl's inequality does not hold for δ_n but it holds for its stable version $\delta_{n,L}$.

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



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

$$(n+1)^s \varepsilon_n \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_m, \quad n \geq 0$$

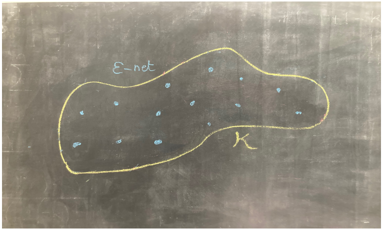
In particular

$$d_n \lesssim n^{-s}, \quad n \geq 0 \implies \varepsilon_n \lesssim n^{-s}, \quad n \geq 0.$$

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one can prove that $\varepsilon_n \geq cn^{-1}$.

Carl's inequality does not hold for δ_n but it holds for its stable version $\delta_{n,L}$.

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



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

$$(n+1)^s \varepsilon_n \leq C_s \sup_{m=0,\dots,n} (m+1)^s d_m, \quad n \geq 0$$

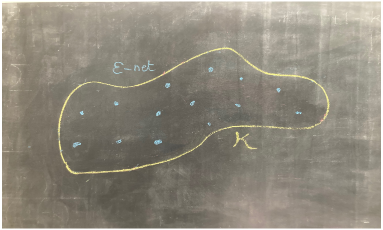
In particular

$$d_n \lesssim n^{-s}, \quad n \geq 0 \implies \varepsilon_n \lesssim n^{-s}, \quad n \geq 0.$$

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one can prove that $\varepsilon_n \geq cn^{-1}$.

Carl's inequality does not hold for δ_n but it holds for its stable version $\delta_{n,L}$.

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



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

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

In particular

$$d_n \lesssim n^{-s}, \quad n \ge 0 \implies \varepsilon_n \lesssim n^{-s}, \quad n \ge 0.$$

For $V = L^\infty(I)$ and $\mathcal{K} = \mathcal{U}(\mathrm{Lip}(I))$ one can prove that $\varepsilon_n \ge cn^{-1}$.

Carl's inequality does not hold for δ_n but it holds for its stable version $\delta_{n,L}$.

Approximation by neural networks

The family V_n of all neural networks $v : \mathbb{R}^d \to \mathbb{R}^m$ described by at most n parameters is an instance of nonlinear approximation.

For a given target function u, the search of an approximation $u_n \in V_n$ is usually done by solving an optimization problem using a large training set of points

$$u_n = \mathrm{argmin}_{v \in V_n} \sum_{i=1}^{M} |u(x^i) - v(x^i)|^2.$$

This is a non-convex optimization problem in the parameter space, often solved by stochastic gradient algorithms.

Approximation results by Yarotzki, Shen-Yang-Zhang (2020) for $d = 1$ and $m = 1$: neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} !

This means that the parameter selection that defines the encoding map E achieving such rates cannot be stable or even continuous.

Approximation by neural networks

The family V_n of all neural networks $v : \mathbb{R}^d \to \mathbb{R}^m$ described by at most n parameters is an instance of nonlinear approximation.

For a given target function u, the search of an approximation $u_n \in V_n$ is usually done by solving an optimization problem using a large training set of points

$$u_n = \operatorname{argmin}_{v \in V_n} \sum_{i=1}^{M} |u(x^i) - v(x^i)|^2.$$

This is a non-convex optimization problem in the parameter space, often solved by stochastic gradient algorithms.

Approximation results by Yarotzki, Shen-Yang-Zhang (2020) for $d = 1$ and $m = 1$: neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} !

This means that the parameter selection that defines the encoding map E achieving such rates cannot be stable or even continuous.

Approximation by neural networks

The family V_n of all neural networks $v : \mathbb{R}^d \to \mathbb{R}^m$ described by at most n parameters is an instance of nonlinear approximation.

For a given target function u, the search of an approximation $u_n \in V_n$ is usually done by solving an optimization problem using a large training set of points

$$u_n = \mathrm{argmin}_{v \in V_n} \sum_{i=1}^{M} |u(x^i) - v(x^i)|^2.$$

This is a non-convex optimization problem in the parameter space, often solved by stochastic gradient algorithms.

Approximation results by Yarotzki, Shen-Yang-Zhang (2020) for $d = 1$ and $m = 1$: neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} !

This means that the parameter selection that defines the encoding map E achieving such rates cannot be stable or even continuous.

Approximation by neural networks

The family V_n of all neural networks $v : \mathbb{R}^d \to \mathbb{R}^m$ described by at most n parameters is an instance of nonlinear approximation.

For a given target function u, the search of an approximation $u_n \in V_n$ is usually done by solving an optimization problem using a large training set of points

$$u_n = \mathrm{argmin}_{v \in V_n} \sum_{i=1}^{M} |u(x^i) - v(x^i)|^2.$$

This is a non-convex optimization problem in the parameter space, often solved by stochastic gradient algorithms.

Approximation results by Yarotzki, Shen-Yang-Zhang (2020) for $d = 1$ and $m = 1$: neural networks approximation of functions in $\mathrm{Lip}(I)$ converge in L^∞ with rate n^{-2} !

This means that the parameter selection that defines the encoding map E achieving such rates cannot be stable or even continuous.

The power of point value information for linear recovery

Recent results (M. Ullrich, T. Ullrich, Nagel, Krieg, Dolbeault...) reveal that when $V = L^2$, point value evaluation are enough for optimal rates of linear recovery.

Under mild assumptions on the $\mathcal{K}$, linear sampling number behave as well as n-widths.

$$a_n = d_n \lesssim n^{-s} \iff r_n \lesssim n^{-s}.$$

R \ E	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

$$\inf_{E,R} \max_{u \in \mathcal{K}} \|u - R(E(u))\|_V, .$$

The power of point value information for linear recovery

Recent results (M. Ullrich, T. Ullrich, Nagel, Krieg, Dolbeault...) reveal that when $V = L^2$, point value evaluation are enough for optimal rates of linear recovery.

Under mild assumptions on the $\mathcal{K}$, linear sampling number behave as well as n-widths.

$$a_n = d_n \lesssim n^{-s} \iff r_n \lesssim n^{-s}.$$

Ideas behind the proof:

(i) take a near optimal space V_n in the sense of n-widths

(ii) For E sample i.i.d. according to a well chosen measure that depends on V_n.

(iii) For R use a weighted least-square reconstruction.

Recent results (M. Ullrich, T. Ullrich, Nagel, Krieg, Dolbeault...) reveal that when $V = L^2$, point value evaluation are enough for optimal rates of linear recovery.

Under mild assumptions on the $\mathcal{K}$, linear sampling number behave as well as n-widths.

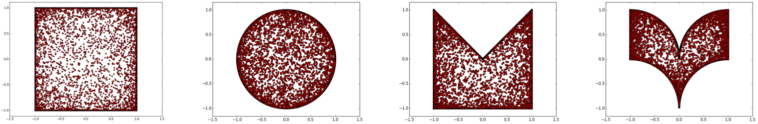
$$a_n = d_n \lesssim n^{-s} \iff r_n \lesssim n^{-s}.$$

Ideas behind the proof:

(i) take a near optimal space V_n in the sense of n-widths

(ii) For E sample i.i.d. according to a well chosen measure that depends on V_n.

(iii) For R use a weighted least-square reconstruction.



Optimal densities when V_n are bivariate polynomials of fixed total degree $k = 9$.

The power of point value information for linear recovery

Recent results (M. Ullrich, T. Ullrich, Nagel, Krieg, Dolbeault...) reveal that when $V = L^2$, point value evaluation are enough for optimal rates of linear recovery.

Under mild assumptions on the $\mathcal{K}$, linear sampling number behave as well as n-widths.

$$a_n = d_n \lesssim n^{-s} \iff r_n \lesssim n^{-s}.$$

Ideas behind the proof:

(i) take a near optimal space V_n in the sense of n-widths

(ii) For E sample i.i.d. according to a well chosen measure that depends on V_n.

(iii) For R use a weighted least-square reconstruction.

Point values are generally uneffective for optimal nonlinear recovery: $\delta_n << \rho_n$ for certain classes $\mathcal{K}$.

In DNN refered to as theory to practice gap (Adcock, Grohs, Voigtlaender...)

Achieving the accuracy of nonlinear spaces V_n may requires $m >> n$ point evaluations.

Cohen-DeVore-Petrova-Wojtaszczyk (2021) : when V is a Hilbert space, linear measurements are enough for optimal rates of stable nonlinear recovery.

Stable sensing numbers decay similar to stable manifold widths and entropy numbers.

$$\varepsilon_n \lesssim n^{-s} \iff \delta_{n,L} \lesssim n^{-s} \iff s_{n,L} \lesssim n^{-s}.$$

R \ E	point values	linear	nonlinear
linear	r_n	a_n	d_n
nonlinear	ρ_n	s_n	δ_n

$$\inf_{E,R} \max_{u \in \mathcal{K}} \|u - R(E(u))\|_V,.$$

Cohen-DeVore-Petrova-Wojtaszczyk (2021) : when V is a Hilbert space, linear measurements are enough for optimal rates of stable nonlinear recovery.

Stable sensing numbers decay similar to stable manifold widths and entropy numbers.

$$\varepsilon_n \lesssim n^{-s} \iff \delta_{n,L} \lesssim n^{-s} \iff s_{n,L} \lesssim n^{-s}.$$

Carl's inequality already tells us that $s_{n,L} \lesssim n^{-s} \implies \delta_{n,L} \lesssim n^{-s} \implies \varepsilon_n \lesssim n^{-s}$.

Conversely, in Hilbert spaces, we establish a direct comparison: with $L = 2$ and $c = 26$,

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

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

$$\frac{1}{2}\|u_i - u_j\|_V \leq \|P_W(u_i - u_j)\|_V \leq \|u_i - u_j\|_V, \quad u_i, u_j \in \mathcal{N}.$$

3. This gives an exact recovery map R that is 2-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).

This recovery procedure is not computationally feasible.

Cohen-DeVore-Petrova-Wojtaszczyk (2021) : when V is a Hilbert space, linear measurements are enough for optimal rates of stable nonlinear recovery.

Stable sensing numbers decay similar to stable manifold widths and entropy numbers.

$$\varepsilon_n \lesssim n^{-s} \iff \delta_{n,L} \lesssim n^{-s} \iff s_{n,L} \lesssim n^{-s}.$$

Carl's inequality already tells us that $s_{n,L} \lesssim n^{-s} \implies \delta_{n,L} \lesssim n^{-s} \implies \varepsilon_n \lesssim n^{-s}$.

Conversely, in Hilbert spaces, we establish a direct comparison: with $L = 2$ and $c = 26$,

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

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

$$\frac{1}{2}\|u_i - u_j\|_V \leq \|P_W(u_i - u_j)\|_V \leq \|u_i - u_j\|_V, \quad u_i, u_j \in \mathcal{N}.$$

3. This gives an exact recovery map R that is 2-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).

This recovery procedure is not computationally feasible.

The power of linear information for nonlinear recovery

Cohen-DeVore-Petrova-Wojtaszczyk (2021) : when V is a Hilbert space, linear measurements are enough for optimal rates of stable nonlinear recovery.

Stable sensing numbers decay similar to stable manifold widths and entropy numbers.

$$\varepsilon_n \lesssim n^{-s} \iff \delta_{n,L} \lesssim n^{-s} \iff s_{n,L} \lesssim n^{-s}.$$

Carl's inequality already tells us that $s_{n,L} \lesssim n^{-s} \implies \delta_{n,L} \lesssim n^{-s} \implies \varepsilon_n \lesssim n^{-s}$.

Conversely, in Hilbert spaces, we establish a direct comparison: with $L = 2$ and $c = 26$,

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

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

$$\frac{1}{2}\|u_i - u_j\|_V \leq \|P_W(u_i - u_j)\|_V \leq \|u_i - u_j\|_V, \quad u_i, u_j \in \mathcal{N}.$$

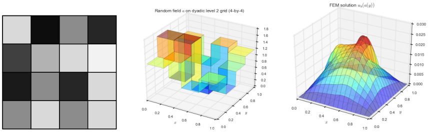
3. This gives an exact recovery map R that is 2-Lipschitz from $P_W\mathcal{N}$ to $\mathcal{N}$.
4. Extend this map from $W \sim \mathbb{R}^{cn}$ to V with same Lipschitz constant (Kirszbraun).

This recovery procedure is not computationally feasible.

An instructive example : consider the steady-state elliptic diffusion equation

$$-\mathrm{div}(a\nabla u) = f, \quad \text{on} \quad \Omega \subset \mathbb{R}^2, \quad u_{\partial\Omega} = 0,$$

with fixed f, and piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \ldots, y_d) \in Y = [-b, b]^d$, where $0 < b < a$.



How large is the Kolmogorov 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 $d_n(\mathcal{K})_{H^1} \lesssim exp(-cn^{1/d})$. This follows from the holomorphy of the map $y \mapsto u(y)$ that we can approximate by truncated power series

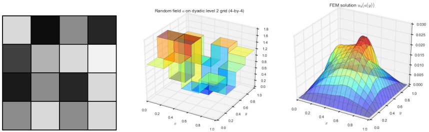
$$\max_{y\in Y} \Big\| u(y) - \sum_{|\nu|\leq k} u_\nu y^\nu \Big\| \leq C \exp(-ck), \quad y^\nu = y_1^{\nu_1} \ldots y_d^{\nu_d},$$

So $\mathcal{K}$ is approximated at this accuracy by $V_n = \mathrm{span}\{u_\nu \ : \ |\nu| \leq k\}$, $n = \binom{k+d}{k} \sim k^d$.

An instructive example : consider the steady-state elliptic diffusion equation

$$-\mathrm{div}(a\nabla u) = f, \quad \text{on} \quad \Omega \subset \mathbb{R}^2, \quad u_{\partial\Omega} = 0,$$

with fixed f, and piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-b, b]^d$, where $0 < b < a$.



How large is the Kolmogorov 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 $d_n(\mathcal{K})_{H^1} \lesssim exp(-cn^{1/d})$. This follows from the holomorphy of the map $y \mapsto u(y)$ that we can approximate by truncated power series

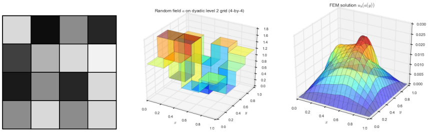
$$\max_{y\in Y} \Big\| u(y) - \sum_{|\nu|\le k} u_\nu y^\nu \Big\| \le C\exp(-ck), \quad y^\nu = y_1^{\nu_1} \dots y_d^{\nu_d},$$

So $\mathcal{K}$ is approximated at this accuracy by $V_n = \mathrm{span}\{u_\nu \,:\, |\nu| \le k\}$, $n = \binom{k+d}{k} \sim k^d$.

An instructive example : consider the steady-state elliptic diffusion equation

$$-\mathrm{div}(a\nabla u) = f, \quad \text{on} \quad \Omega \subset \mathbb{R}^2, \quad u_{\partial\Omega} = 0,$$

with fixed f, and piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-b, b]^d$, where $0 < b < a$.



How large is the Kolmogorov 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 $d_n(\mathcal{K})_{H^1} \lesssim exp(-cn^{1/d})$). This follows from the holomorphy of the map $y \mapsto u(y)$ that we can approximate by truncated power series

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

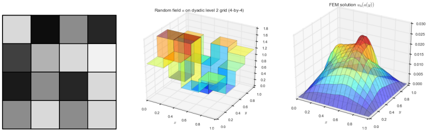
So $\mathcal{K}$ is approximated at this accuracy by $V_n = \mathrm{span}\{u_\nu \ : \ |\nu| \le k\}$, $n = \binom{k+d}{k} \sim k^d$.

Estimating n-width of solution manifolds

An instructive example : consider the steady-state elliptic diffusion equation

$$-\mathrm{div}(a\nabla u) = f, \quad \text{on} \quad \Omega \subset \mathbb{R}^2, \quad u_{\partial\Omega} = 0,$$

with fixed f, and piecewise constant diffusion function $a = a(y)$ having value $\overline{a} + y_j$ on subdomain Ω_j, where $y = (y_1, \dots, y_d) \in Y = [-b, b]^d$, where $0 < b < a$.



How large is the Kolmogorov 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 $d_n(\mathcal{K})_{H^1} \lesssim exp(-cn^{1/d})$. This follows from the holomorphy of the map $y \mapsto u(y)$ that we can approximate by truncated power series

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

So $\mathcal{K}$ is approximated at this accuracy by $V_n = \mathrm{span}\{u_\nu \ : \ |\nu| \le k\}$, $n = \binom{k+d}{k} \sim k^d$.

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 $T = 1$,

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

It is easily checked that

$$d_n = d_n(\mathcal{K})_{L^2} \sim n^{-1/2},$$

while

$$\delta_n = s_n = 0, \quad n \geq 1 \qquad \text{and} \qquad \varepsilon_n \sim 2^{-n/2}.$$

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 $T = 1$,

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

It is easily checked that

$$d_n = d_n(\mathcal{K})_{L^2} \sim n^{-1/2},$$

while

$$\delta_n = s_n = 0, \quad n \geq 1 \qquad \text{and} \qquad \varepsilon_n \sim 2^{-n/2}.$$

A multivariate class

More generally consider in $Q = [0,1]^d$ and with $s \geq 1$,

$$\mathcal{K} = \mathcal{K}_s := \{\chi_\Omega \;:\; \Omega \subset Q,\; \partial\Omega \text{ is } \mathcal{C}^s \text{ regular}\}.$$

Can be made a compact set of $L^2(Q)$ by imposing a uniform $\mathcal{C}^s$ bound on the local parametrizations of Ω.

One can prove that $d_n(\mathcal{K}_s)_{L^2} \sim n^{-\frac{1}{2d}}$ regardless of how large is s.

On the other hand,

$$s_n(\mathcal{K}_s)_{L^2} \sim \delta_n(\mathcal{K}_s)_{L^2} \sim \varepsilon_n(\mathcal{K}_s)_{L^2} \sim n^{-\frac{s}{2(d-1)}}.$$

Open problem: achievable by simple linear measurements and recovery strategies ?

Remark: sampling numbers seem to have intermediate rate $\rho_n \sim n^{-\frac{s}{2(d-1)+2s}}$.

A multivariate class

More generally consider in $Q = [0,1]^d$ and with $s \geq 1$,

$$\mathcal{K} = \mathcal{K}_s := \{\chi_\Omega \ : \ \Omega \subset Q, \ \partial\Omega \text{ is } \mathcal{C}^s \text{ regular}\}.$$

Can be made a compact set of $L^2(Q)$ by imposing a uniform $\mathcal{C}^s$ bound on the local parametrizations of Ω.

One can prove that $d_n(\mathcal{K}_s)_{L^2} \sim n^{-\frac{1}{2d}}$ regardless of how large is s.

On the other hand,

$$s_n(\mathcal{K}_s)_{L^2} \sim \delta_n(\mathcal{K}_s)_{L^2} \sim \varepsilon_n(\mathcal{K}_s)_{L^2} \sim n^{-\frac{s}{2(d-1)}}.$$

Open problem: achievable by simple linear measurements and recovery strategies ?

Remark: sampling numbers seem to have intermediate rate $\rho_n \sim n^{-\frac{s}{2(d-1)+2s}}$.

A multivariate class

More generally consider in $Q=[0,1]^d$ and with $s \geq 1$,

$$\mathcal{K}=\mathcal{K}_s := \{\chi_\Omega \;:\; \Omega \subset Q,\ \partial\Omega \text{ is } \mathcal{C}^s \text{ regular}\}.$$

Can be made a compact set of $L^2(Q)$ by imposing a uniform $\mathcal{C}^s$ bound on the local parametrizations of Ω.

One can prove that $d_n(\mathcal{K}_s)_{L^2} \sim n^{-\frac{1}{2d}}$ regardless of how large is s.

On the other hand,

$$s_n(\mathcal{K}_s)_{L^2} \sim \delta_n(\mathcal{K}_s)_{L^2} \sim \varepsilon_n(\mathcal{K}_s)_{L^2} \sim n^{-\frac{s}{2(d-1)}}.$$

Open problem: achievable by simple linear measurements and recovery strategies ?

Remark: sampling numbers seem to have intermediate rate $\rho_n \sim n^{-\frac{s}{2(d-1)+2s}}$.

For classes $\mathcal{K}$ such as solution manifolds of parametrized PDEs in Hilbert spaces: investing some offline computation of a near optimal approximation space V_n can be highly beneficial for fast online solvers, compared conventional approximation methods (finite elements, splines).

The reduced basis approach (Maday, Patera,...): $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

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

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

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

For classes $\mathcal{K}$ such as solution manifolds of parametrized PDEs in Hilbert spaces: investing some offline computation of a near optimal approximation space V_n can be highly beneficial for fast online solvers, compared conventional approximation methods (finite elements, splines).

The reduced basis approach (Maday, Patera,...): $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

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

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

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

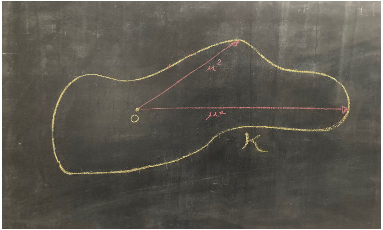
For classes $\mathcal{K}$ such as solution manifolds of parametrized PDEs in Hilbert spaces: investing some offline computation of a near optimal approximation space V_n can be highly beneficial for fast online solvers, compared conventional approximation methods (finite elements, splines).

The reduced basis approach (Maday, Patera,...): $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

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

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

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



For classes $\mathcal{K}$ such as solution manifolds of parametrized PDEs in Hilbert spaces: investing some offline computation of a near optimal approximation space V_n can be highly beneficial for fast online solvers, compared conventional approximation methods (finite elements, splines).

The reduced basis approach (Maday, Patera,...): $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

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

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

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

Binev-Cohen-Dahmen-DeVore-Petrova-Wojtaszczyk (2013): with $\sigma_n := \mathrm{dist}(\mathcal{K}, V_n)_V$,

$$d_n \lesssim n^{-s} \implies \sigma_n \lesssim n^{-s},$$

and

$$d_n \lesssim e^{-cn^s} \implies \sigma_n \lesssim e^{-\tilde{c}n^s}.$$

Open problem: similar practical realization of rate optimal E and nonlinear R ?

Can be thought as a learning problem over $\mathcal{K}$. DNN auto-encoders ?

For classes $\mathcal{K}$ such as solution manifolds of parametrized PDEs in Hilbert spaces: investing some offline computation of a near optimal approximation space V_n can be highly beneficial for fast online solvers, compared conventional approximation methods (finite elements, splines).

The reduced basis approach (Maday, Patera,...): $V_n = \mathrm{span}\{u^1, \dots, u^n\}$, with $u^i \in \mathcal{K}$.

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

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

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

Binev-Cohen-Dahmen-DeVore-Petrova-Wojtaszczyk (2013): with $\sigma_n := \mathrm{dist}(\mathcal{K}, V_n)_V$,

$$d_n \lesssim n^{-s} \implies \sigma_n \lesssim n^{-s},$$

and

$$d_n \lesssim e^{-cn^s} \implies \sigma_n \lesssim e^{-\tilde{c}n^s}.$$

Open problem: similar practical realization of rate optimal E and nonlinear R ?

Can be thought as a learning problem over $\mathcal{K}$. DNN auto-encoders ?

One of the objective of reduced modeling is the fast access to approximations of the solutions to general PDE's $\mathcal{P}(u) = 0$.

As a basic example, consider the elliptic problem: find $u \in V$ such that

$$a(u, v) = \ell(v), \quad v \in V,$$

in a Hilbert space V, under the standard Lax-Milgram assumptions. Equivalently

$$u = \mathrm{argmin}_{v \in V} J(v), \quad J(v) := \frac{1}{2} a(v, v) - \ell(v).$$

Approximation in a linear reduced model V_n by Galerkin: find $u_n \in V_n$ such that

$$a(u_n, v) = \ell(v), \quad v \in V_n \iff u_n = \mathrm{argmin}_{v \in V_n} J(v).$$

Cea's lemma ensures best approximation: $\|u - u_n\| \leq C \min_{v \in V_n} \|u - v\|$.

Computational time: dense $n \times n$ linear system with moderate n.

Assembling time: computation of matrix elements $a(\phi_j, \phi_j)$ can be the dominant part.

Linear reduced models in forward simulation

One of the objective of reduced modeling is the fast access to approximations of the solutions to general PDE's $\mathcal{P}(u) = 0$.

As a basic example, consider the elliptic problem: find $u \in V$ such that

$$a(u, v) = \ell(v), \quad v \in V,$$

in a Hilbert space V, under the standard Lax-Milgram assumptions. Equivalently

$$u = \mathrm{argmin}_{v \in V} J(v), \quad J(v) := \frac{1}{2} a(v, v) - \ell(v).$$

Approximation in a linear reduced model V_n by Galerkin: find $u_n \in V_n$ such that

$$a(u_n, v) = \ell(v), \quad v \in V_n \iff u_n = \mathrm{argmin}_{v \in V_n} J(v).$$

Cea's lemma ensures best approximation: $\|u - u_n\| \leq C \min_{v \in V_n} \|u - v\|$.

Computational time: dense $n \times n$ linear system with moderate n.

Assembling time: computation of matrix elements $a(\phi_j, \phi_j)$ can be the dominant part.

Linear reduced models in forward simulation

One of the objective of reduced modeling is the fast access to approximations of the solutions to general PDE's $\mathcal{P}(u) = 0$.

As a basic example, consider the elliptic problem: find $u \in V$ such that

$$a(u, v) = \ell(v), \quad v \in V,$$

in a Hilbert space V, under the standard Lax-Milgram assumptions. Equivalently

$$u = \mathrm{argmin}_{v \in V} J(v), \quad J(v) := \frac{1}{2} a(v, v) - \ell(v).$$

Approximation in a linear reduced model V_n by Galerkin: find $u_n \in V_n$ such that

$$a(u_n, v) = \ell(v), \quad v \in V_n \iff u_n = \mathrm{argmin}_{v \in V_n} J(v).$$

Cea's lemma ensures best approximation: $\|u - u_n\| \leq C \min_{v \in V_n} \|u - v\|$.

Computational time: dense $n \times n$ linear system with moderate n.

Assembling time: computation of matrix elements $a(\phi_j, \phi_j)$ can be the dominant part.

One of the objective of reduced modeling is the fast access to approximations of the solutions to general PDE's $\mathcal{P}(u) = 0$.

As a basic example, consider the elliptic problem: find $u \in V$ such that

$$a(u, v) = \ell(v), \quad v \in V,$$

in a Hilbert space V, under the standard Lax-Milgram assumptions. Equivalently

$$u = \mathrm{argmin}_{v \in V} J(v), \quad J(v) := \frac{1}{2} a(v, v) - \ell(v).$$

Approximation in a linear reduced model V_n by Galerkin: find $u_n \in V_n$ such that

$$a(u_n, v) = \ell(v), \quad v \in V_n \iff u_n = \mathrm{argmin}_{v \in V_n} J(v).$$

Cea's lemma ensures best approximation: $\|u - u_n\| \leq C \min_{v \in V_n} \|u - v\|$.

Computational time: dense $n \times n$ linear system with moderate n.

Assembling time: computation of matrix elements $a(\phi_j, \phi_j)$ can be the dominant part.

Linear reduced models in forward simulation

One of the objective of reduced modeling is the fast access to approximations of the solutions to general PDE's $\mathcal{P}(u) = 0$.

As a basic example, consider the elliptic problem: find $u \in V$ such that

$$a(u, v) = \ell(v), \quad v \in V,$$

in a Hilbert space V, under the standard Lax-Milgram assumptions. Equivalently

$$u = \mathrm{argmin}_{v \in V} J(v), \quad J(v) := \frac{1}{2} a(v, v) - \ell(v).$$

Approximation in a linear reduced model V_n by Galerkin: find $u_n \in V_n$ such that

$$a(u_n, v) = \ell(v), \quad v \in V_n \iff u_n = \mathrm{argmin}_{v \in V_n} J(v).$$

Cea's lemma ensures best approximation: $\|u - u_n\| \leq C \min_{v \in V_n} \|u - v\|$.

Computational time: dense $n \times n$ linear system with moderate n.

Assembling time: computation of matrix elements $a(\phi_j, \phi_j)$ can be the dominant part.

How can we adapt these approaches/results to nonlinear reduced models V_n ?

A systematic approach: for a fixed norm $\|\cdot\|_Z$, minimize the residual

$$u_n = \mathrm{argmin}_{v\in V_n} \|\mathcal{P}(v)\|_Z,$$

PINN's methods : use deep neural networks for V_n and the ℓ^2 norm on a sufficiently large training point sets $\sum_i |\mathcal{P}(v)(x^i)|^2$.

Alternative for elliptic problem: consider $u_n = \mathrm{argmin}_{v\in V_n} J(v)$.

This approximation satisfies Cea's estimate $\|u - u_n\| \leq C \min_{v\in V_n} \|u - v\|$.

All these approaches amount in solving non-convex optimization problems that can be computationally untractable, even for moderate values of n.

Polynomially mapped manifold (Haasdonk, Farhat, Willcox..): for some fixed $k \geq 1$ consider reduced models of the form

$$V_n := \Big\{ \sum_{|\nu|\leq k} x^\nu \varphi_\nu \ : \ (x_1, \dots, x_n) \in \mathbb{R}^n \Big\}.$$

The case $k = 1$ is linear reduced models. When $k > 1$ we introduce nonlinearity.

Solving $\min_{v\in V_n} J(v)$: an $n \times n$ polynomial system with coefficients $a(\varphi_\nu, \varphi_\mu)$.

How can we adapt these approaches/results to nonlinear reduced models V_n ?

A systematic approach: for a fixed norm $\|\cdot\|_Z$, minimize the residual

$$u_n = \mathrm{argmin}_{v\in V_n} \|\mathcal{P}(v)\|_Z,$$

PINN's methods : use deep neural networks for V_n and the ℓ^2 norm on a sufficienly large training point sets $\sum_i |\mathcal{P}(v)(x^i)|^2$.

Alternative for elliptic problem: consider $u_n = \mathrm{argmin}_{v\in V_n} J(v)$.

This approximation satisfies Cea's estimate $\|u-u_n\| \leq C \min_{v\in V_n} \|u-v\|$.

All these approaches amount in solving non-convex optimization problems that can be computationally untractable, even for moderate values of n.

Polynomially mapped manifold (Haasdonk, Farhat, Willcox..): for some fixed $k \geq 1$ consider reduced models of the form

$$V_n := \Big\{ \sum_{|\nu|\leq k} x^\nu \varphi_\nu \ : \ (x_1, \dots, x_n) \in \mathbb{R}^n \Big\}.$$

The case $k = 1$ is linear reduced models. When $k > 1$ we introduce nonlinearity.

Solving $\min_{v\in V_n} J(v)$: an $n \times n$ polynomial system with coefficients $a(\varphi_\nu, \varphi_\mu)$.

How can we adapt these approaches/results to nonlinear reduced models V_n ?

A systematic approach: for a fixed norm $\|\cdot\|_Z$, minimize the residual

$$u_n = \mathrm{argmin}_{v\in V_n}\|\mathcal{P}(v)\|_Z,$$

PINN's methods : use deep neural networks for V_n and the ℓ^2 norm on a sufficiently large training point sets $\sum_i |\mathcal{P}(v)(x^i)|^2$.

Alternative for elliptic problem: consider $u_n = \mathrm{argmin}_{v\in V_n} J(v)$.

This approximation satisfies Cea's estimate $\|u-u_n\| \leq C\min_{v\in V_n}\|u-v\|$.

All these approaches amount in solving non-convex optimization problems that can be computationally untractable, even for moderate values of n.

Polynomially mapped manifold (Haasdonk, Farhat, Willcox..): for some fixed $k\geq 1$ consider reduced models of the form

$$V_n := \Big\{ \sum_{|\nu|\leq k} x^\nu \varphi_\nu \ : \ (x_1,\dots,x_n)\in\mathbb{R}^n \Big\}.$$

The case $k=1$ is linear reduced models. When $k>1$ we introduce nonlinearity.

Solving $\min_{v\in V_n} J(v)$: an $n\times n$ polynomial system with coefficients $a(\varphi_\nu,\varphi_\mu)$.

How can we adapt these approaches/results to nonlinear reduced models V_n ?

A systematic approach: for a fixed norm $\|\cdot\|_Z$, minimize the residual

$$u_n = \mathrm{argmin}_{v\in V_n}\|\mathcal{P}(v)\|_Z,$$

PINN's methods : use deep neural networks for V_n and the ℓ^2 norm on a sufficiently large training point sets $\sum_i |\mathcal{P}(v)(x^i)|^2$.

Alternative for elliptic problem: consider $u_n = \mathrm{argmin}_{v\in V_n} J(v)$.

This approximation satisfies Cea's estimate $\|u-u_n\| \leq C\min_{v\in V_n}\|u-v\|$.

All these approaches amount in solving non-convex optimization problems that can be computationally untractable, even for moderate values of n.

Polynomially mapped manifold (Haasdonk, Farhat, Willcox..): for some fixed $k\geq 1$ consider reduced models of the form

$$V_n := \Big\{ \sum_{|\nu|\leq k} x^\nu \varphi_\nu \ : \ (x_1,\dots,x_n)\in\mathbb{R}^n \Big\}.$$

The case $k=1$ is linear reduced models. When $k>1$ we introduce nonlinearity.

Solving $\min_{v\in V_n} J(v)$: an $n\times n$ polynomial system with coefficients $a(\varphi_\nu,\varphi_\mu)$.

How can we adapt these approaches/results to nonlinear reduced models V_n ?

A systematic approach: for a fixed norm $\|\cdot\|_Z$, minimize the residual

$$u_n = \mathrm{argmin}_{v\in V_n}\|\mathcal{P}(v)\|_Z,$$

PINN's methods : use deep neural networks for V_n and the ℓ^2 norm on a sufficiently large training point sets $\sum_i |\mathcal{P}(v)(x^i)|^2$.

Alternative for elliptic problem: consider $u_n = \mathrm{argmin}_{v\in V_n} J(v)$.

This approximation satisfies Cea's estimate $\|u-u_n\| \leq C \min_{v\in V_n}\|u-v\|$.

All these approaches amount in solving non-convex optimization problems that can be computationally untractable, even for moderate values of n.

Polynomially mapped manifold (Haasdonk, Farhat, Willcox..): for some fixed $k \geq 1$ consider reduced models of the form

$$V_n := \Big\{\sum_{|\nu|\leq k} x^\nu \varphi_\nu \;:\; (x_1,\dots,x_n)\in\mathbb{R}^n\Big\}.$$

The case $k=1$ is linear reduced models. When $k>1$ we introduce nonlinearity.

Solving $\min_{v\in V_n} J(v)$: an $n\times n$ polynomial system with coefficients $a(\varphi_\nu,\varphi_\mu)$.

How can we adapt these approaches/results to nonlinear reduced models V_n ?

A systematic approach: for a fixed norm $\|\cdot\|_Z$, minimize the residual

$$u_n = \mathrm{argmin}_{v\in V_n}\|\mathcal{P}(v)\|_Z,$$

PINN's methods : use deep neural networks for V_n and the ℓ^2 norm on a sufficiently large training point sets $\sum_i |\mathcal{P}(v)(x^i)|^2$.

Alternative for elliptic problem: consider $u_n = \mathrm{argmin}_{v\in V_n} J(v)$.

This approximation satisfies Cea's estimate $\|u-u_n\| \leq C \min_{v\in V_n} \|u-v\|$.

All these approaches amount in solving non-convex optimization problems that can be computationally untractable, even for moderate values of n.

Polynomially mapped manifold (Haasdonk, Farhat, Willcox..): for some fixed $k\geq 1$ consider reduced models of the form

$$V_n := \Big\{ \sum_{|\nu|\leq k} x^\nu \varphi_\nu \;:\; (x_1,\dots,x_n)\in\mathbb{R}^n \Big\}.$$

The case $k=1$ is linear reduced models. When $k>1$ we introduce nonlinearity.

Solving $\min_{v\in V_n} J(v)$: an $n\times n$ polynomial system with coefficients $a(\varphi_\nu,\varphi_\mu)$.

From unknown $u \in V$ Hilbert space, we observe

$$\ell_i(u) = \langle u, \omega_i \rangle, \quad i = 1, \dots, m,$$

or equivalently $P_W u$ where $W = \mathrm{span}\{\omega_1, \dots, \omega_m\}$ is the measurement space.

Recovery in a linear reduced model space $V_n \subset V$

Can we recover up to the best approximation error $e_n(u) = \|u - P_{V_n} u\|$?

Best fit (least square) estimator : $\tilde{u} := \mathrm{argmin}\{\|P_W(u - v)\| \ : \ v \in V_n\}$

Maday-Patera-Penn-Yano (2015): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v \in V_n} \frac{\|v\|}{\|P_W v\|},$$

which is the inverse cosine of the angle between V_n and W. Then

$$\|u - \tilde{u}\| \le \mu e_n(u)_V.$$

The constant μ is computable (singular value analysis of a cross-grammian).

From unknown $u \in V$ Hilbert space, we observe

$$\ell_i(u) = \langle u, \omega_i \rangle, \quad i = 1, \dots, m,$$

or equivalently $P_W u$ where $W = \mathrm{span}\{\omega_1, \dots, \omega_m\}$ is the measurement space.

Recovery in a linear reduced model space $V_n \subset V$

Can we recover up to the best approximation error $e_n(u) = \|u - P_{V_n} u\|$?

Best fit (least square) estimator : $\tilde{u} := \mathrm{argmin}\{\|P_W(u - v)\| \; : \; v \in V_n\}$

Maday-Patera-Penn-Yano (2015): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v \in V_n} \frac{\|v\|}{\|P_W v\|},$$

which is the inverse cosine of the angle between V_n and W. Then

$$\|u - \tilde{u}\| \leq \mu e_n(u)_V.$$

The constant μ is computable (singular value analysis of a cross-grammian).

Linear reduced models and inverse problems

From unknown $u \in V$ Hilbert space, we observe

$$\ell_i(u) = \langle u, \omega_i \rangle, \quad i = 1, \dots, m,$$

or equivalently $P_W u$ where $W = \mathrm{span}\{\omega_1, \dots, \omega_m\}$ is the measurement space.

Recovery in a linear reduced model space $V_n \subset V$

Can we recover up to the best approximation error $e_n(u) = \|u - P_{V_n} u\|$?

Best fit (least square) estimator : $\tilde{u} := \mathrm{argmin}\{\|P_W(u - v)\| \ : \ v \in V_n\}$

Maday-Patera-Penn-Yano (2015): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v \in V_n} \frac{\|v\|}{\|P_W v\|},$$

which is the inverse cosine of the angle between V_n and W. Then

$$\|u - \tilde{u}\| \leq \mu e_n(u)_V.$$

The constant μ is computable (singular value analysis of a cross-grammian).

Linear reduced models and inverse problems

From unknown $u \in V$ Hilbert space, we observe

$$\ell_i(u) = \langle u, \omega_i \rangle, \quad i = 1, \dots, m,$$

or equivalently $P_W u$ where $W = \mathrm{span}\{\omega_1, \dots, \omega_m\}$ is the measurement space.

Recovery in a linear reduced model space $V_n \subset V$

Can we recover up to the best approximation error $e_n(u) = \|u - P_{V_n} u\|$?

Best fit (least square) estimator : $\tilde{u} := \mathrm{argmin}\{\|P_W(u - v)\| \ : \ v \in V_n\}$

Maday-Patera-Penn-Yano (2015): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v \in V_n} \frac{\|v\|}{\|P_W v\|},$$

which is the inverse cosine of the angle between V_n and W. Then

$$\|u - \tilde{u}\| \leq \mu e_n(u)_V.$$

The constant μ is computable (singular value analysis of a cross-grammian).

A nonlinear generalization

We now would like to recover $\tilde{u} \approx u$ in a nonlinear space V_n from the $\ell_i(u)$

Estimator $\tilde{u} := \text{argmin}\{\|P_W(u-v)\| \, : \, v \in V_n\}$ requires by non-convex optimization.

Cohen-Dolbeault-Mula-Somacal (2022): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v_1, v_2 \in V_n} \frac{\|v_1 - v_2\|}{\|P_W(v_1 - v_2)\|},$$

Then

$$\|u - \tilde{u}\| \leq (1 + 2\mu) e_n(u)_V.$$

Extensions : V Banach space, nonlinear measurement functionals.

The constant μ is sometimes difficult to estimate.

Compressed sensing example: V_n space of n-sparse vectors in $V = \mathbb{R}^N$ with $N >> n$.

Control of μ is equivalent to the so-called null-space property.

A nonlinear generalization

We now would like to recover $\tilde{u} \approx u$ in a nonlinear space V_n from the $\ell_i(u)$

Estimator $\tilde{u} := \operatorname{argmin}\{\|P_W(u - v)\| \;:\; v \in V_n\}$ requires by non-convex optimization.

Cohen-Dolbeault-Mula-Somacal (2022): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v_1, v_2 \in V_n} \frac{\|v_1 - v_2\|}{\|P_W(v_1 - v_2)\|},$$

Then

$$\|u - \tilde{u}\| \le (1 + 2\mu) e_n(u)_V.$$

Extensions : V Banach space, nonlinear measurement functionals.

The constant μ is sometimes difficult to estimate.

Compressed sensing example: V_n space of n-sparse vectors in $V = \mathbb{R}^N$ with $N >> n$.

Control of μ is equivalent to the so-called null-space property.

A nonlinear generalization

We now would like to recover $\tilde{u} \approx u$ in a nonlinear space V_n from the $\ell_i(u)$

Estimator $\tilde{u} := \mathrm{argmin}\{\|P_W(u-v)\| \ : \ v \in V_n\}$ requires by non-convex optimization.

Cohen-Dolbeault-Mula-Somacal (2022): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v_1, v_2 \in V_n} \frac{\|v_1 - v_2\|}{\|P_W(v_1 - v_2)\|},$$

Then

$$\|u - \tilde{u}\| \le (1 + 2\mu) e_n(u)_V.$$

Extensions : V Banach space, nonlinear measurement functionals.

The constant μ is sometimes difficult to estimate.

Compressed sensing example: V_n space of n-sparse vectors in $V = \mathbb{R}^N$ with $N >> n$.

Control of μ is equivalent to the so-called null-space property.

A nonlinear generalization

We now would like to recover $\tilde{u} \approx u$ in a nonlinear space V_n from the $\ell_i(u)$

Estimator $\tilde{u} := \operatorname{argmin}\{\|P_W(u-v)\| \ : \ v \in V_n\}$ requires by non-convex optimization.

Cohen-Dolbeault-Mula-Somacal (2022): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v_1, v_2 \in V_n} \frac{\|v_1 - v_2\|}{\|P_W(v_1 - v_2)\|},$$

Then

$$\|u - \tilde{u}\| \leq (1 + 2\mu) e_n(u)_V.$$

Extensions : V Banach space, nonlinear measurement functionals.

The constant μ is sometimes difficult to estimate.

Compressed sensing example: V_n space of n-sparse vectors in $V = \mathbb{R}^N$ with $N >> n$.

Control of μ is equivalent to the so-called null-space property.

A nonlinear generalization

We now would like to recover $\tilde{u} \approx u$ in a nonlinear space V_n from the $\ell_i(u)$

Estimator $\tilde{u} := \mathrm{argmin}\{\|P_W(u-v)\| \;:\; v \in V_n\}$ requires by non-convex optimization.

Cohen-Dolbeault-Mula-Somacal (2022): introduce the stability constant

$$\mu = \mu(V_n, W) := \max_{v_1, v_2 \in V_n} \frac{\|v_1 - v_2\|}{\|P_W(v_1 - v_2)\|},$$

Then

$$\|u - \tilde{u}\| \leq (1 + 2\mu) e_n(u)_V.$$

Extensions : V Banach space, nonlinear measurement functionals.

The constant μ is sometimes difficult to estimate.

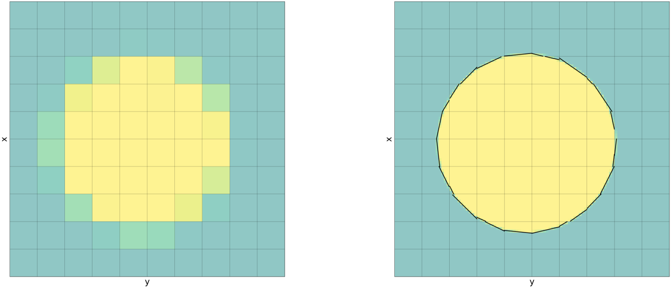
Compressed sensing example: V_n space of n-sparse vectors in $V = \mathbb{R}^N$ with $N >> n$.

Control of μ is equivalent to the so-called null-space property.

Application: shape recovery from local averages

Recover a characteristic function from the class $\mathcal{K}_s$ from cell-averages of sidelength h.

For simplicity, consider here dimension $d = 2$.

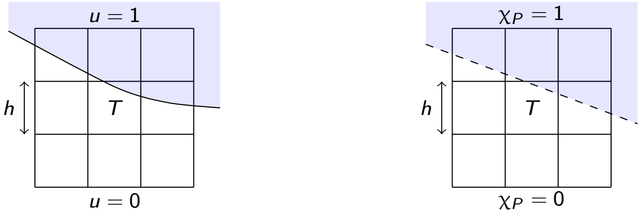


Linear reconstruction by piecewise constants (left) give L^2 rate $\sqrt{h} \sim n^{-1/4}$.

Nonlinear reconstruction by piecewise-linear interfaces.

Expect improved L^2 rate $h \sim n^{-1/2}$ for $s \geq 2$.

On each cell T, approximate u by $\tilde{u}|_T = \chi_P|_T$, where P is a half-plane computed from the average values of u on a 3×3 stencil S composed of T and 8 neighboring cells.



Subcell resolution (1d Harten 1992, 2d Arandiga-Cohen-Donat-Dyn-Matei 2003).

Volume of fluid, ELVIRA (Pilliod-Puckett 1997, Zaleski 1998).

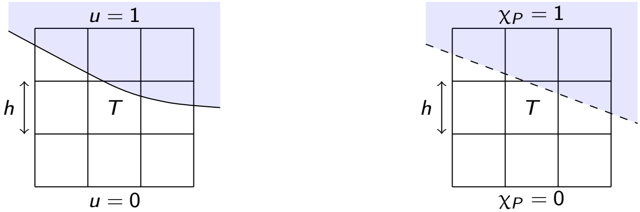
With $V_2 = \{\chi_P|_T,\ P \text{ half-plane}\}$, the local approximation error of $u \in \mathcal{K}_2$ in cells containing the interface $\partial\Omega$ is bounded by

$$\min_{v \in V_2} \|u - v\|_{L^1(T)} \leq Ch^3.$$

Half plane computed by least-squares fitting of the averages $\ell_j(u)$ for $j = 1, \dots, 9$.

A local approach

On each cell T, approximate u by $\tilde{u}|_T = \chi_P|_T$, where P is a half-plane computed from the average values of u on a 3×3 stencil S composed of T and 8 neighboring cells.



Subcell resolution (1d Harten 1992, 2d Arandiga-Cohen-Donat-Dyn-Matei 2003).

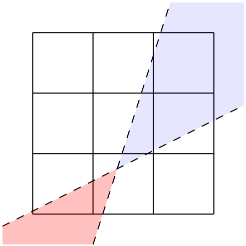
Volume of fluid, ELVIRA (Pilliod-Puckett 1997, Zaleski 1998).

With $V_2 = \{\chi_P|_T,\ P \text{ half-plane}\}$, the local approximation error of $u \in \mathcal{K}_2$ in cells containing the interface $\partial\Omega$ is bounded by

$$\min_{v \in V_2} \|u - v\|_{L^1(T)} \leq Ch^3.$$

Half plane computed by least-squares fitting of the averages $\ell_j(u)$ for $j = 1, \dots, 9$.

Stability



The optimal constant can be proved to be

$$\mu = \mu(V_2, W) := \max_{P,Q \text{ half-planes}} \frac{\|\chi_P - \chi_Q\|_{L^1(S)}}{\sum_{j=1}^{9} |\ell_j(\chi_P - \chi_Q)|} = \frac{3}{2}$$

This leads to the global second order reconstruction L^1 bound: for $u \in \mathcal{K}_2$

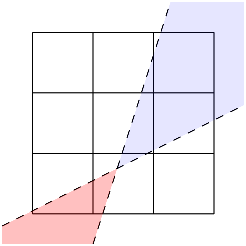
$$\|u - \tilde{u}\|_{L^1} \leq Ch^2,$$

and for the L^2 norm,

$$\|u - \tilde{u}\|_{L^2} \leq Ch = Cn^{-1/2}.$$

which is still not the optimal rate $n^{-\frac{s}{2(d-1)}} = n^{-1}$.

Stability



The optimal constant can be proved to be

$$\mu = \mu(V_2, W) := \max_{P,Q \text{ half-planes}} \frac{\|\chi_P - \chi_Q\|_{L^1(S)}}{\sum_{j=1}^{9} |\ell_j(\chi_P - \chi_Q)|} = \frac{3}{2}$$

This leads to the global second order reconstruction L^1 bound: for $u \in \mathcal{K}_2$

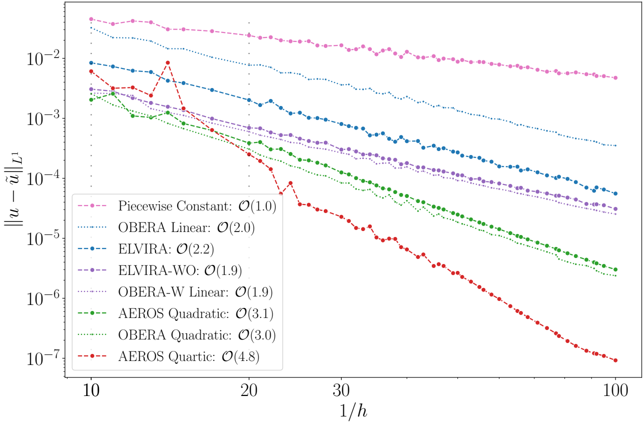
$$\|u - \tilde{u}\|_{L^1} \leq Ch^2,$$

and for the L^2 norm,

$$\|u - \tilde{u}\|_{L^2} \leq Ch = Cn^{-1/2}.$$

which is still not the optimal rate $n^{-\frac{s}{2(d-1)}} = n^{-1}$.

Cohen-Mula-Somacal (2024): higher order reconstructions by curved interfaces, treatment of corners...



Convergence for different reconstruction models (estimated rates in parenthesis).

Conclusions

A standard idea: reduce complexity of solving PDE's and inverse problems searching the approximation within a finite n-dimensional space.

A less standard idea: optimize the choice of the n-dimensional space. Theory is well settled. Provably optimal model reduction algorithms are available.

The nonlinear perspective: theoretical pillars are available. Provably optimal nonlinear model reduction algorithms are still lacking.

References

A. Pinkus, *n-width in Approximation Theory*, Springer 1985.

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

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

M. Dolbeault, D. Krieg and M. Ullrich, *A sharp upper bound for sampling numbers in* L^2, Applied and Computational Harmonic Analysis, 2023.

P. Grohs and F. Voigtlaender, *Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces*, Foundation of Computational Mathematics, 2023.

A. Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Optimal stable nonlinear approximation*, Foundation of Computational Mathematics, 2022.

A. Cohen, M. Dolbeault, O. Mula and A. Somacal, *Nonlinear approximation spaces for inverse problems*, Analysis and Application, 2023.

References

A. Pinkus, *n-width in Approximation Theory*, Springer 1985.

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

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

M. Dolbeault, D. Krieg and M. Ullrich, *A sharp upper bound for sampling numbers in* L^2, Applied and Computational Harmonic Analysis, 2023.

P. Grohs and F. Voigtlaender, *Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces*, Foundation of Computational Mathematics, 2023.

A. Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Optimal stable nonlinear approximation*, Foundation of Computational Mathematics, 2022.

A. Cohen, M. Dolbeault, O. Mula and A. Somacal, *Nonlinear approximation spaces for inverse problems*, Analysis and Application, 2023.

Thanks !