Stable nonlinear manifold approximation using compositional networks

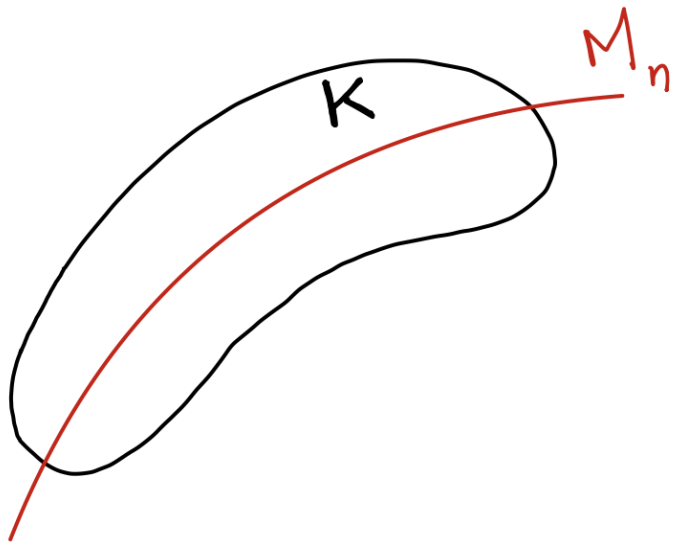
Anthony Nouy

Centrale Nantes, Nantes Université,
Laboratoire de Mathématiques Jean Leray

Joint work with Antoine Bensalah and Joel Soffo.

Introduction

We consider the problem of approximating a subset K of a normed space X by a low-dimensional set M_n.



A common setting (in statistics) if when K is the range of some vector or function-valued random variable.

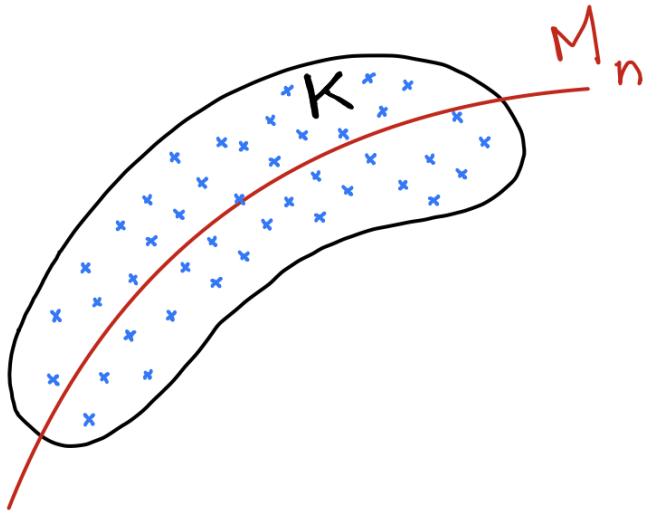
Another classical setting is the solution of forward or inverse problems for parameter-dependent equations, where

$$K = \{u(y) : y \in Y\} \quad \text{with} \quad R(u(y); y) = 0.$$

Introduction

The approximating sets M_n can be

- constructed offline by manifold approximation methods, using samples from K



- used online to compute approximations of elements in K with low computational complexity, or from limited information.

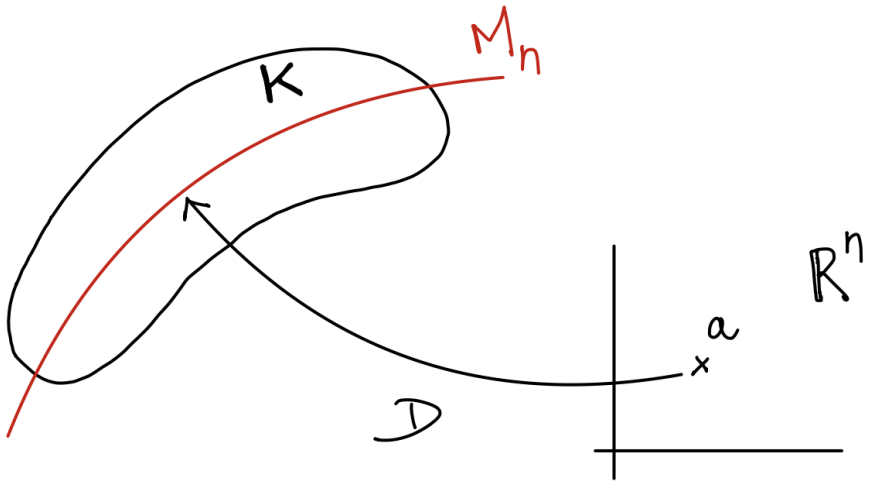
Encoder-Decoder

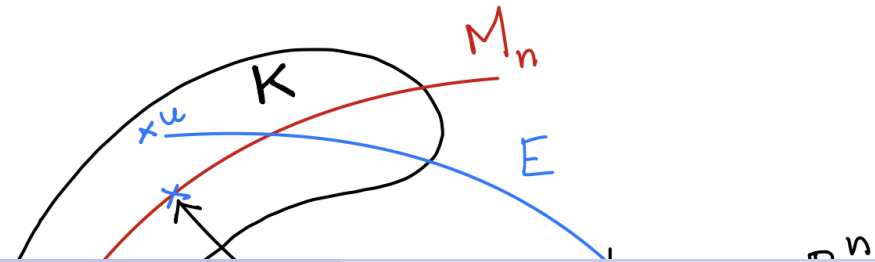
A large class of manifold approximation methods can be described by an encoder $E : K \to \mathbb{R}^n$ and a decoder $D : \mathbb{R}^n \to X$.

The decoder provides a parametrization of a n-dimensional "manifold"

$$M_n = \{D(a) : a \in \mathbb{R}^n\}.$$

The encoder is related to the approximation process (algorithm). It associates to $u \in K$ a parameter value $a = E(u) \in \mathbb{R}^n$.





Optimal performance

Manifold approximation methods can be classified in terms of the properties of their encoders and decoders.

The optimal performance of a given class $\mathcal{E}_n$ of encoders from X to $\mathbb{R}^n$ and a given class $\mathcal{D}_n$ of decoders from $\mathbb{R}^n \to X$ can be assessed in worst-case setting by

$$\inf_{D \in \mathcal{D}_n, E \in \mathcal{E}_n} \sup_{u \in K} \|u - D \circ E(u)\|_X.$$

If the set K is equipped with a measure ρ, the optimal performance can be measured in average sense by

$$\inf_{D \in \mathcal{D}_n, E \in \mathcal{E}_n} \left(\int_K \|u - D \circ E(u)\|_X^p d\rho(u) \right)^{1/p}.$$

These errors define measures of complexity (widths) of K.

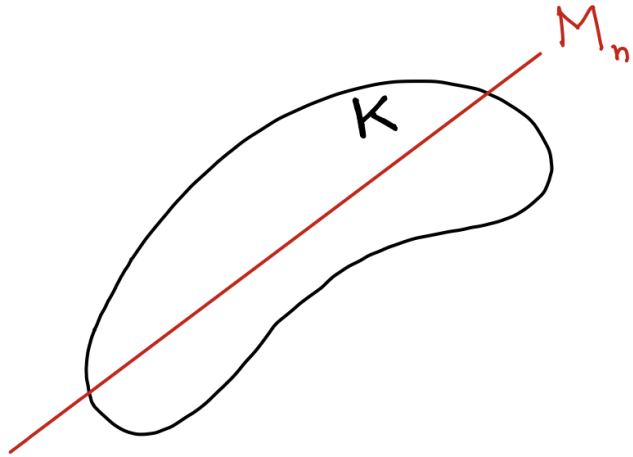
Linear approximation - Worst case setting

The range M_n of a **linear decoder** $D : \mathbb{R}^n \to X$ is a linear space with dimension at most n.

Restricting the **decoder and encoder to be linear** yields the **approximation numbers** (linear widths)

$$a_n(K)_X = \inf_{\text{rank}(A)=n} \sup_{u \in K} \|u - Au\|_X,$$

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



Restricting **only the decoder to be linear** yields the **Kolmogorov n-width**

$$d_n(K)_X = \inf_{D \in L(\mathbb{R}^n; X)} \sup_{u \in K} \inf_{a \in \mathbb{R}^n} \|u - D(a)\|_X = \inf_{\dim M_n = n} \sup_{u \in K} \inf_{v \in M_n} \|u - v\|_X.$$

For X a Hilbert space, $a_n(K)_X = d_n(K)_X$ and an optimal auto-encoder $D \circ E$ is given by the orthogonal projection P_{M_n} onto an optimal space M_n.

In practice optimal linear spaces in worst-case error are out of reach but near to optimal spaces M_n can be obtained by greedy algorithms, that generate an increasing sequence of spaces from samples in K [Buffa, Maday, Patera, Prud'homme, Turinici, Binev, Cohen, Dahmen, DeVore,...].

Linear approximation - Average setting

When X is a Hilbert space and the error is measured in average sense, it yields the **average Kolmogorov n-width**

$$d_n^{(p)}(K,\rho)_X^p = \inf_{D\in L(\mathbb{R}^n;X)} \int_K \inf_{a\in\mathbb{R}^n} \|u - D(a)\|_X^p d\rho(u) = \inf_{\dim M_n = n} \int_K \|u - P_{M_n} u\|_X^p d\rho(u).$$

For $p = 2$ (mean-squared error), an optimal space M_n is given by a dominant eigenspace of the (compact) operator

$$T(v) = \int_K u(u,v)_X d\rho(u), \quad v \in X,$$

and

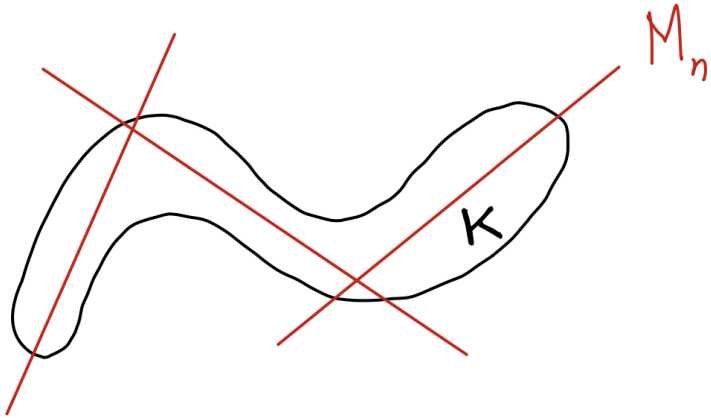
$$\int_K \|u - P_{M_n} u\|_X^2 d\rho(u) = d_n^{(2)}(K,\rho)_X^2 = \sum_{i>n} \lambda_i(T)$$

If ρ is a probability measure, assuming $\bar{u} = \int u\, d\rho(u) = 0$, T is the covariance operator of ρ and this corresponds to Principal Component Analysis (PCA).

Multi-space approximation

M_n can be chosen as a union of N linear spaces of dimension n

$$M_n = \bigcup_{k=1}^{N} V_k$$



A related measure of complexity is the **nonlinear Kolmogorov width** or **library width**

$$d_n(K, N)_X = \inf_{\#\mathcal{L}_n = N} \sup_{u \in K} \inf_{V \in \mathcal{L}_n} \|u - P_V u\|_X \qquad \text{[Temlyakov 1998]}$$

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

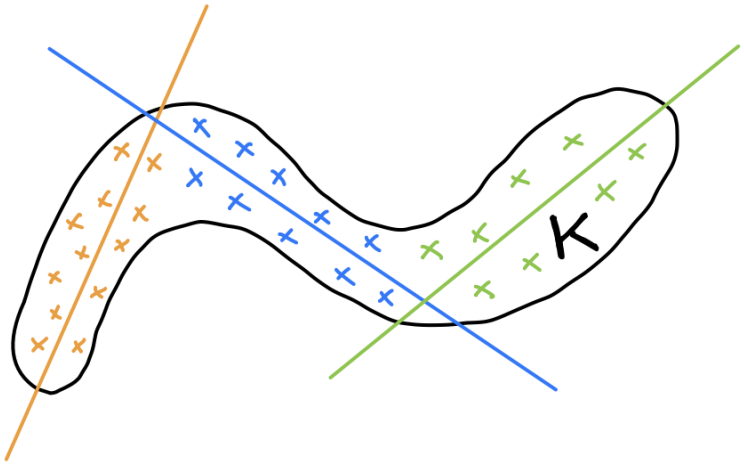
An encoder-decoder view is

$$E : K \to \{1, \dots, N\} \times \mathbb{R}^n, \quad D : \{1, \dots, N\} \times \mathbb{R}^n \to X \quad \text{with} \quad Range(D(k, \cdot)) = V_k$$

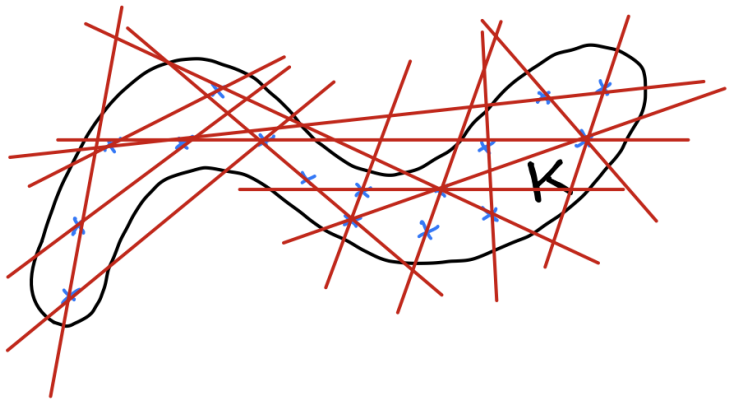
An optimal encoding requires a selection of a subspace V_k and a projection onto V_k.

Multi-space approximation

From samples of K, spaces can be obtained by a clustering approach (h-p method) [Eftang et al 2010][Bonito et al 2021][Guignard and Mula 2024]



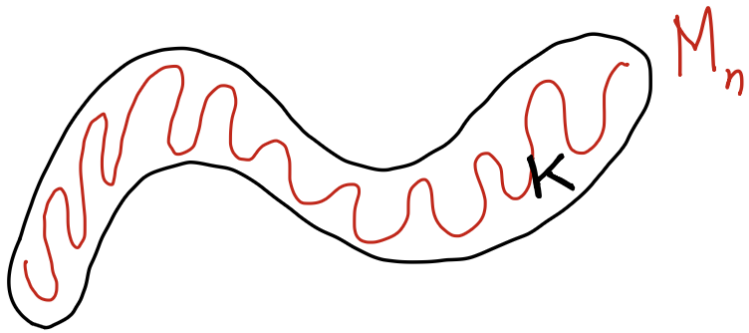
or generated from the dictionary of samples (related to n-term approximation) [Kaulmann and Haasdonk 2013] [Balabanov and Nouy 2021a][Nouy and Pasco 2024]. Computations made feasible by using randomized algebra.



Nonlinear continuous manifold approximation

Restricting both the **encoder and decoders to be continuous** (possibly nonlinear) yields the notion of **nonlinear manifold width** of [DeVore, Howard and Michelli 1989]

$$\delta_n(K)_X = \inf_{E \in C(X;\mathbb{R}^n)} \inf_{D \in C(\mathbb{R}^n;X)} \sup_{u \in K} \|u - D \circ E(u)\|_X.$$

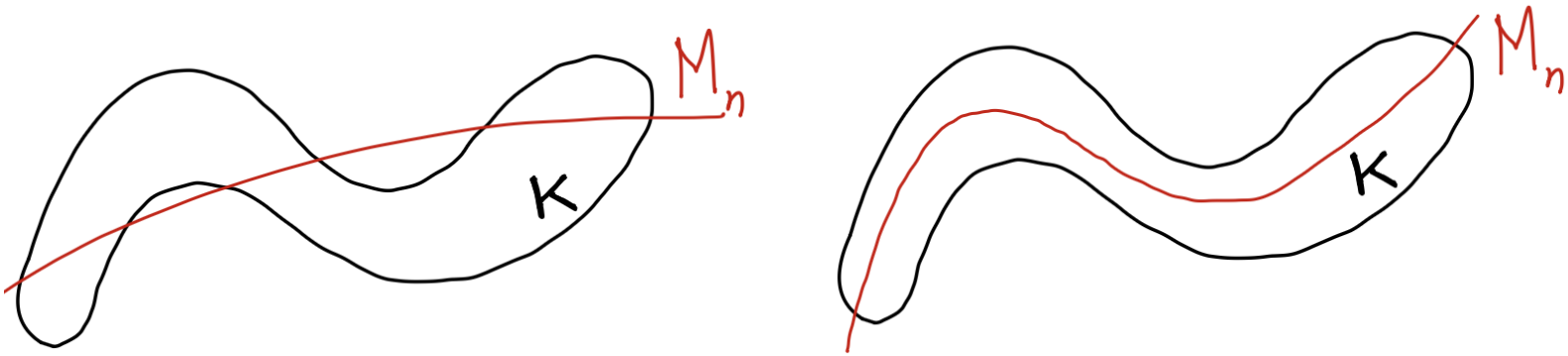


Nonlinear continuous manifold approximation

Further restricting **encoders and decoders to be Lipschitz continuous** yields the notion of **stable manifold width** [Cohen et al 2022]

$$\delta_n^L(K)_X = \inf_{Lip(E)\leq L} \inf_{Lip(D)\leq L} \sup_{u\in K} \|u - D \circ E(u)\|_X.$$

Lipschitz continuity ensures stability of the approximation process, a crucial property in practice.



(a) Low L (b) High L

However, implementation of optimal nonlinear (Lipschitz) encoders may be difficult or even infeasible, e.g. associated with NP-hard optimization problems.

Linear encoder and nonlinear decoder

Restricting the **encoder to be linear and continuous** yields the ***n*-th minimal error of linear information** (sometimes called sensing numbers)

$$e_n(K)_X = \inf_D \inf_{L_1,\dots,L_n} \sup_{u\in K} \|u - D(L_1(u),\dots,L_n(u))\|_X$$

where the infimum is taken over all linear forms $L_1,\dots,L_n$ and all nonlinear maps D.

This benchmark is relevant in many applications where the available information $E(u) = (L_1(u),\dots,L_n(u))$ is linear in u (point evaluations of functions, local averages of functions or more general linear functionals).

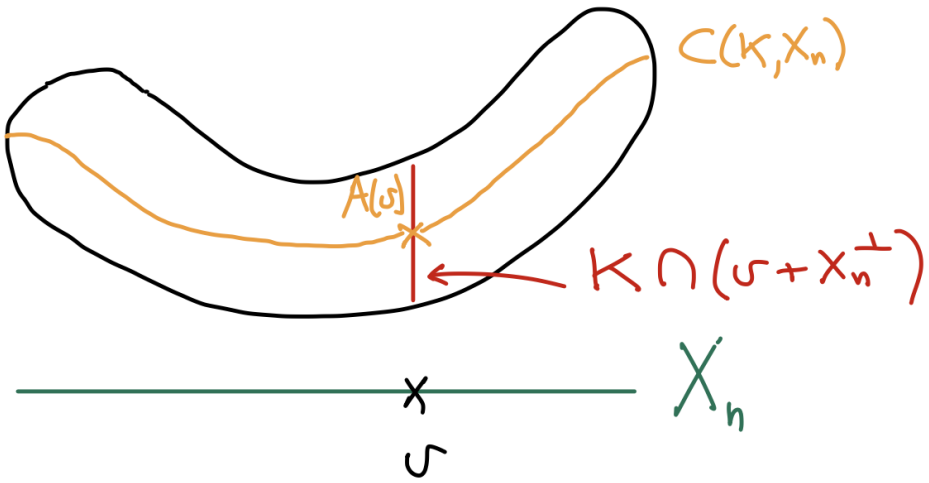
Linear encoder and nonlinear decoder

In a Hilbert setting and worst case setting, this corresponds to

$$e_n(K)_X = \inf_{\dim(X_n)=n} \inf_{A:X_n \to X} \sup_{u \in K} \|u - A(P_{X_n} u)\|_X$$

For a given X_n, an optimal algorithm $A : X_n \to X$ is such that $A(v)$ is the Chebychev center of the slice

$$K_v := K \cap (v + X_n^\perp) = \{u \in K : P_{X_n} u = v\}.$$



This yields an optimal n-dimensional manifold

$$M_n = C(K, X_n) := \{cen(K_v) : v \in P_{X_n} K\}.$$

Linear encoder and nonlinear decoder

In a mean squared setting, the minimal squared error is

$$e_n^{(2)}(K,\rho)_X^2 = \inf_{\dim(X_n)=n} \inf_{A:X_n\to X} \int_K \|u - A(P_{X_n}u)\|_X^2 \, d\rho(u)$$

An optimal algorithm A is obtained by averaging over slices:

$$A(v) = \frac{1}{\rho(K_v)} \int_{K_v} z \, d\rho_v(z)$$

with

$$K_v = K \cap (v + X_n^\perp)$$

and ρ_v is obtained by desintegration (conditional measure).

Linear encoder and nonlinear decoder

With $(\varphi_i)_{i\geq 1}$ an orthonormal basis of X and $X_n = span\{\varphi_1, \ldots, \varphi_n\}$, this is associated with the linear encoder

$$E(u) = ((u, \varphi_i))_{i=1}^n$$

and a decoder of the form

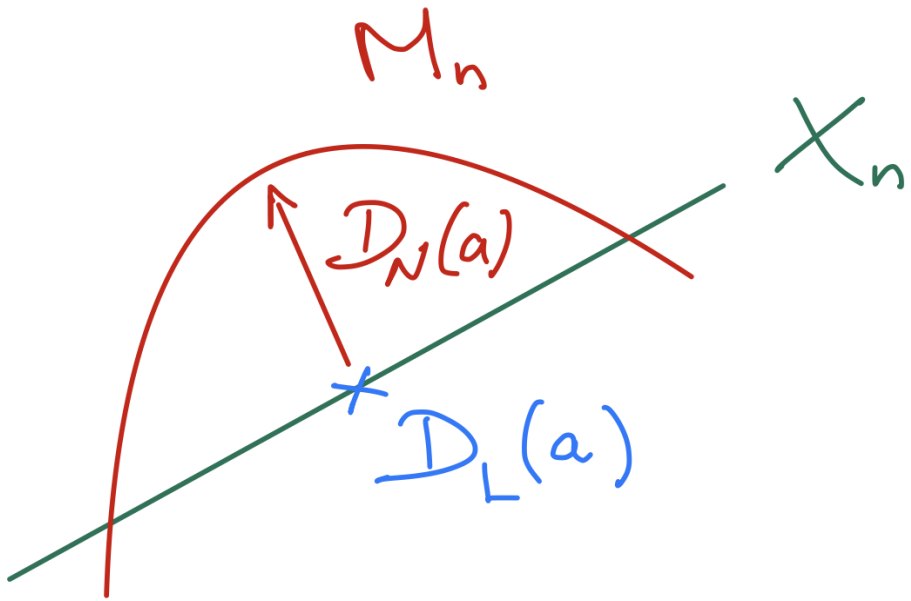
$$D(a) = A(\sum_{i=1}^n a_i \varphi_i) = \sum_{i=1}^n a_i \varphi_i + \sum_{i>n} g_i(a)\varphi_i = D_L(a) + D_N(a),$$

where the functions $g_i : \mathbb{R}^n \to \mathbb{R}$ are nonlinear maps.

D_L is a linear operator from $\mathbb{R}^n$ to $X_n := span\{\varphi_1, \ldots, \varphi_n\}$ such that $D_L(E(u)) = P_{X_n} u$.

D_N maps $\mathbb{R}^n$ to the complementary space of X_n in X.

For a feasible implementation, we truncate to the first N terms, so that the range of D is a nonlinear manifold $M_n \subset X_N = span\{\varphi_1, \ldots, \varphi_N\}$.



Linear encoder and nonlinear decoder

The structure of the decoder exploits the fact that for

$$u = \sum_{i=1}^{n} a_i(u)\varphi_i + \sum_{i>n} a_i(u)\varphi_i \in K,$$

the coefficients $a_i(u)$ for $i > n$ may be well approximated as functions $g_i(E(u))$ of the first few coefficients $E(u) = (a_i(u))_{i=1}^n$.

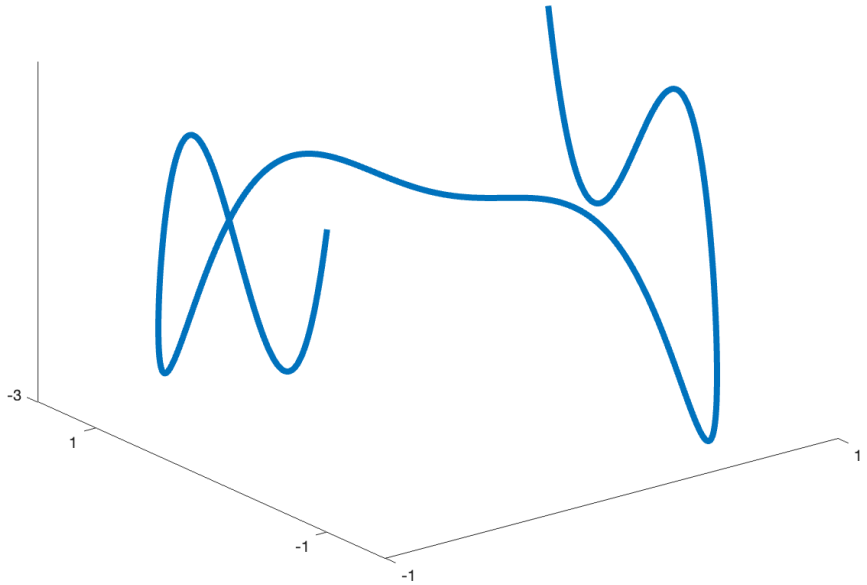
Natural choices for functions g_i are

- Quadratic polynomials [Barnett and Farhat 2022][Geelen et al 2023]
- Higher order polynomials [Geelen et al 2024]
- Neural networks or random forests [Barnett, Farhat and Maday 2023], [Cohen et al 2023]

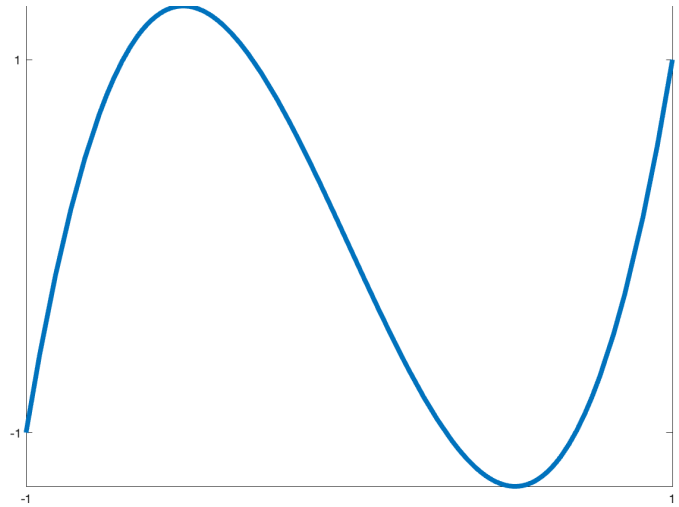
The relation between $a_i(u)$ and $E(u)$ may be highly nonlinear. Even highly expressive approximation tools may result in poor accuracy, due to the difficulty of learning with limited data.

Linear encoder and nonlinear decoder

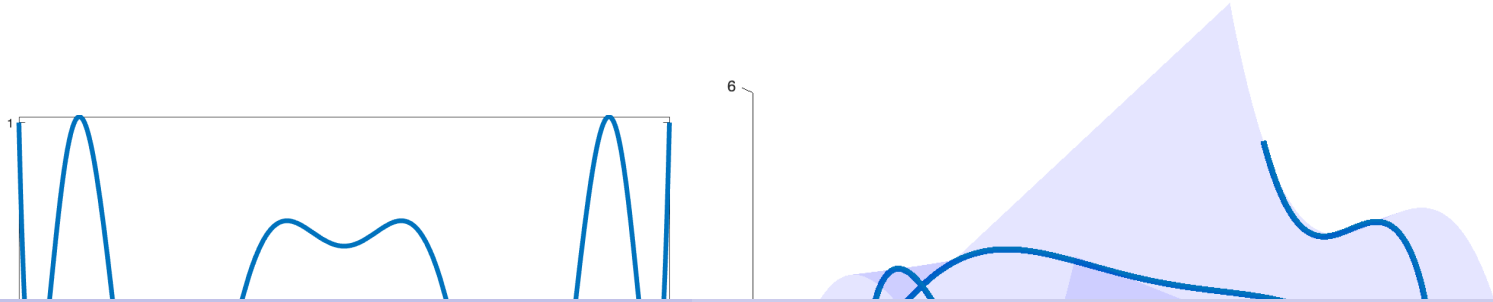
In many applications, a coefficient $a_i(u)$ for $i > n$ may have a highly nonlinear relation with the first n coefficients $a = E(u)$ but a much smoother relation when expressed in terms of a and additional coefficients $a_j(u)$ with $n < j < i$.



(c) $K = \{(a_1(t), a_2(t), a_3(t)) : t \in [0, 1]\}$



(d) a_2 as function of a_1



Decoder based on compositional polynomial network (CPN)

This suggests the following compositional structure of the decoder's functions

$$g_i(a) = f_i(a, (g_j(a))_{n<j\leq n_i}),$$

where the f_i are polynomial functions.

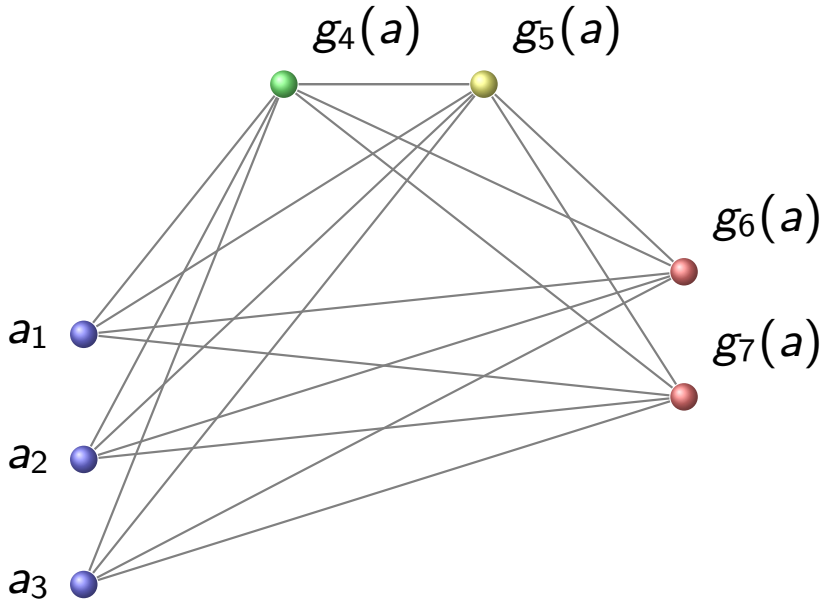


Figure: A compositional polynomial network (CPN) with $N = 7$ and $n = 3$, maximum number of compositions 3.

Decoder based on compositional polynomial network (CPN)

The variables $(a, (g_j(a))_{n<j\leq n_i})$ take values in a set of measure zero in $\mathbb{R}^{n_i}$, but it is still possible to learn polynomial functions f_i from a limited training set.

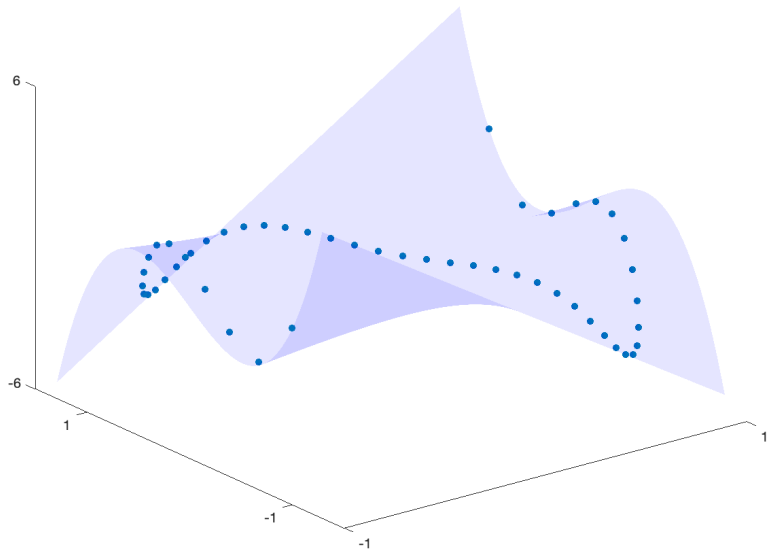


Figure: Learning a_3 as function of (a_1, a_2)

In practice, for high-dimensional approximation of $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ from samples, use of sparse polynomial approximation (CPN-S) or low-rank approximation (tensor networks) (CPN-LR) in $\mathbb{P}_p^{\otimes n_i}$.

Adaptive algorithm with prescribed precision and stability

In [Bensalah, Nouy and Soffo 2026], an adaptive algorithm is provided for the construction of an encoder $E(u) = ((u, \varphi_i)_X)_{i \in I}$ and decoder D that guarantees a prescribed precision ε in mean-squared sense

$$\int_K \|u - D(E(u))\|_X^2 d\rho(u) \leq \varepsilon^2$$

or in worst-case sense

$$\sup_{u \in K} \|u - D(E(u))\|_X \leq \varepsilon.$$

Also, the algorithm controls Lipschitz stability of the encoder-decoder pair

$$\|D(a) - D(\tilde{a})\|_X \leq L\|a - \tilde{a}\|_2, \quad \|E(u) - E(\tilde{u})\|_2 \leq \|u - \tilde{u}\|_X.$$

Numerical illustration: KdV

We consider the Korteweg-de Vries (KdV) equation

$$\frac{\partial u}{\partial t} + 4u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \quad on \quad [-\pi, \pi] \times [0, 1]$$

with periodic boundary conditions and some initial condition. We consider the manifold

$$K = \{u(\cdot, t) : t \in [0, 1]\}$$

We use 5000 samples $u_i = u(\cdot, t_i)$ as training samples.

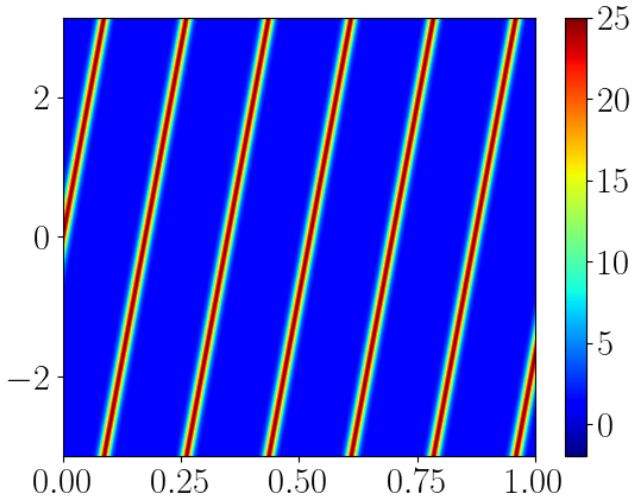


Figure: Function $u(x, t)$.

Method	p	n	N	$\mathbf{RE}_{\text{train}}$	$\mathbf{RE}_{\text{test}}$
Linear	/	5	5	0.435	0.450
Quadratic	2	5	20	0.094	0.099
Additive poly. + AM	5	5	43	0.081	0.084
Sparse	5	5	43	0.013	0.014
CPN-S ($\varepsilon = 10^{-4}$)	5	5	43	0.000072	0.000074

Table: Comparison of methods for the same manifold dimension $n = 5$. CPN-S with sparse polynomials of degree $p = 5$.

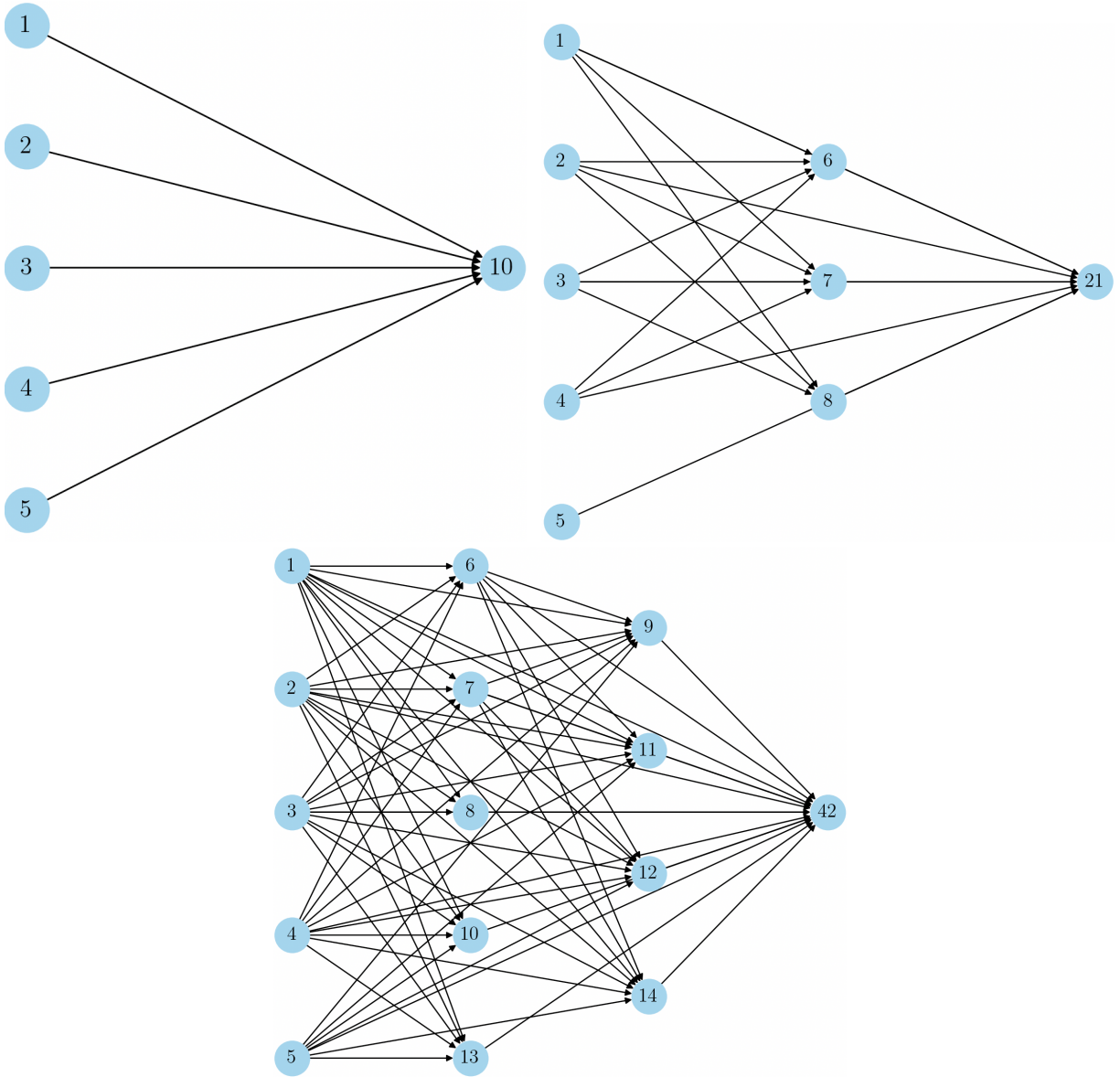


Figure: Compositional networks for g_{10}, g_{21} and g_{42} ($\epsilon = 10^{-4}$, $p = 5$)

Tolerance	n	N	$\mathbf{RE}_{\text{train}}$	$\mathbf{RE}_{\text{test}}$
$\varepsilon = 10^{-1}$	2	15	6.2×10^{-2}	6.4×10^{-2}
$\varepsilon = 10^{-2}$	2	25	6.67×10^{-3}	6.84×10^{-3}
$\varepsilon = 10^{-3}$	3	34	6.83×10^{-4}	7×10^{-4}
$\varepsilon = 10^{-4}$	5	43	7.170×10^{-5}	7.367×10^{-5}
$\varepsilon = 10^{-5}$	6	52	6.76×10^{-6}	6.916×10^{-6}
$\varepsilon = 10^{-6}$	11	61	7.689×10^{-7}	7.885×10^{-7}

Table: Results of CPN-S for $p = 5$ and different target precisions ε.

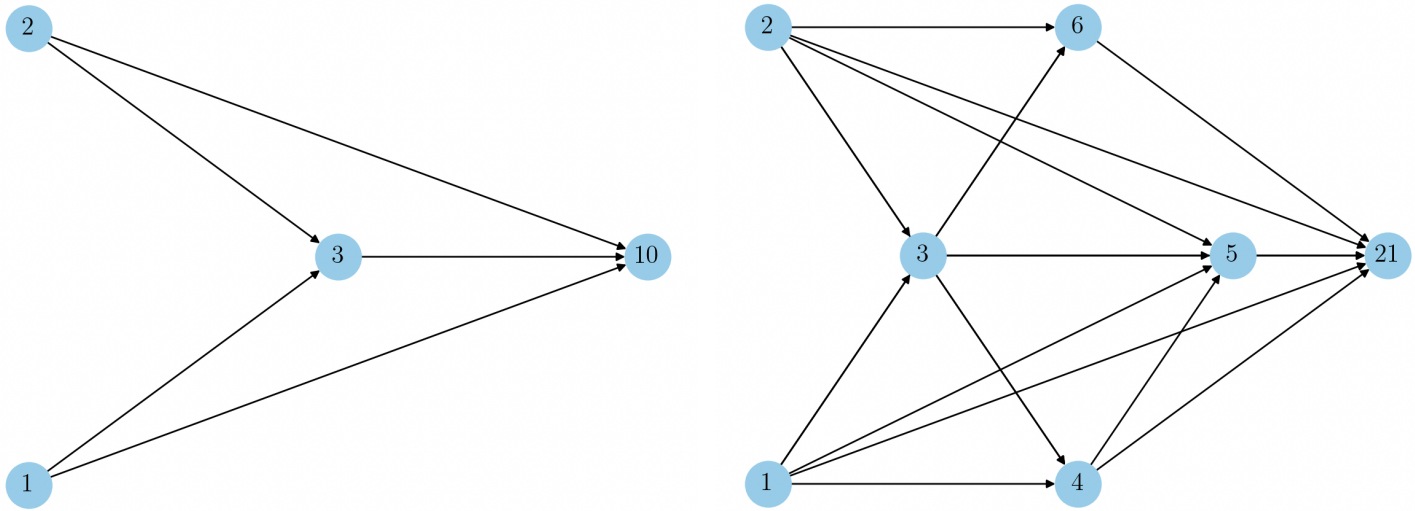


Figure: Compositional networks for g_{10} and g_{21} ($\epsilon = 10^{-2}$, $p = 5$)

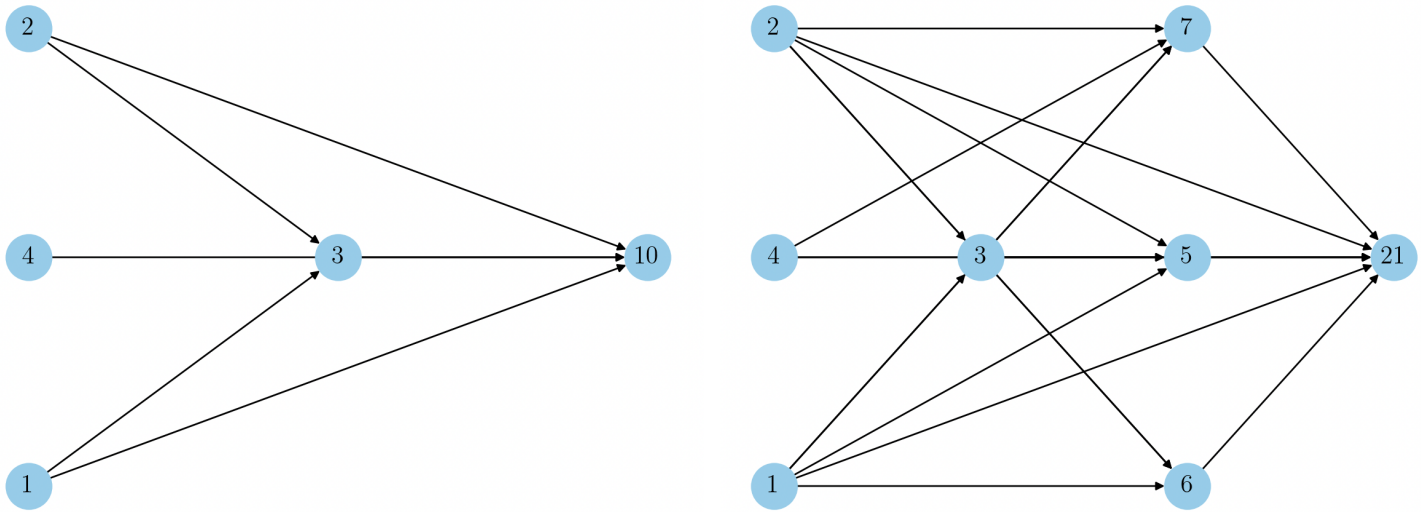


Figure: Compositional networks for g_{10} and g_{21} ($\epsilon = 10^{-3}$, $p = 5$)

Influence of Lipschitz stability

Method	L	n	N	RE_{train}	RE_{test}
CPN-S	2	14	43	7.21×10^{-5}	7.46×10^{-5}
	10	6	43	7.25×10^{-5}	7.49×10^{-5}
	100	5	43	7.30×10^{-5}	7.54×10^{-5}

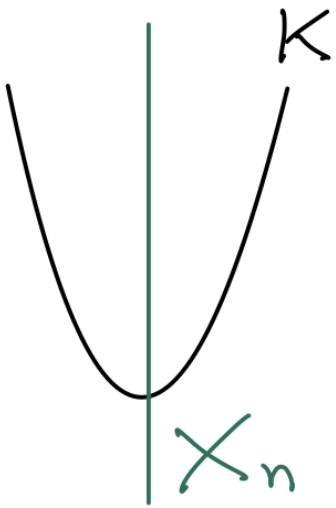
Table: (KdV) Results of CPN for different Lipschitz constants L, with $\epsilon = 10^{-4}$ and $p = 5$.

Optimal linear encoders and associated spaces

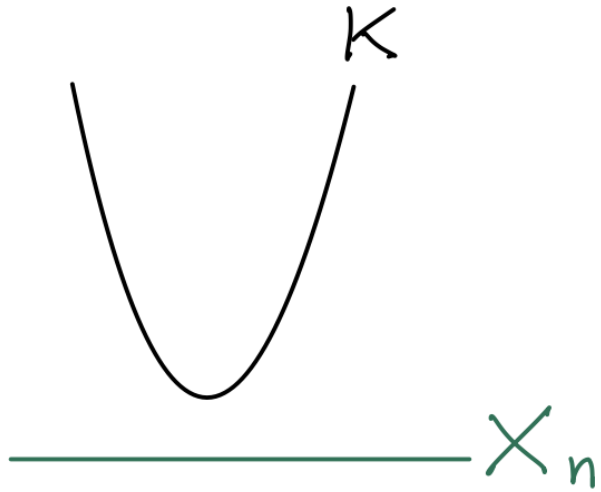
The encoder (or associated space X_n) can be optimized.

A natural choice is to consider the optimal or near-optimal spaces of linear methods, given by principal component analysis (optimal in mean-squared error) or greedy algorithms (close to optimal in worst case error) [Barnett and Farhat 2022, Geelen et al 2023, Barnett, Farhat and Maday 2023, Geelen et al 2024].

However, these may be far from optimal for nonlinear approximation.



$M_n = X_n$



$M_n = K$

Optimal linear encoders and associated spaces

For a decoder of the form

$$D(a) = \sum_{i=1}^{n} a_i \varphi_i + \sum_{i=n+1}^{N} \psi_i(a) \varphi_i$$

with given functions $\psi_i(a)$ (e.g. a polynomials), and an encoder of the form $E(u) = ((u, \varphi_i)_X)_{i=1}^n$, we can formulate an optimization problem over orthonormal systems

$$\min_{\varphi_1, \dots, \varphi_N} \|u - D(E(u))\|_p,$$

which is highly nonlinear [Geelen et al 2024].

In [Schwerdtner and Peherstorfer 2024], a greedy algorithm is proposed for a near optimal construction of X_n for quadratic manifold approximation (quadratic decoder), where the $\varphi_1, \dots, \varphi_n$ are selected greedily in a dictionary $\mathcal{D}$.

Optimal spaces guided by Lipschitz stability (worst case setting)

The benchmark for a linear encoder and a Lipschitz stable decoder is

$$e_n^L(K)_X = \inf_{Lip(E)\leq 1} \inf_{Lip(D)\leq L} \sup_{u\in K} \|u - D(E(u))\|_X$$

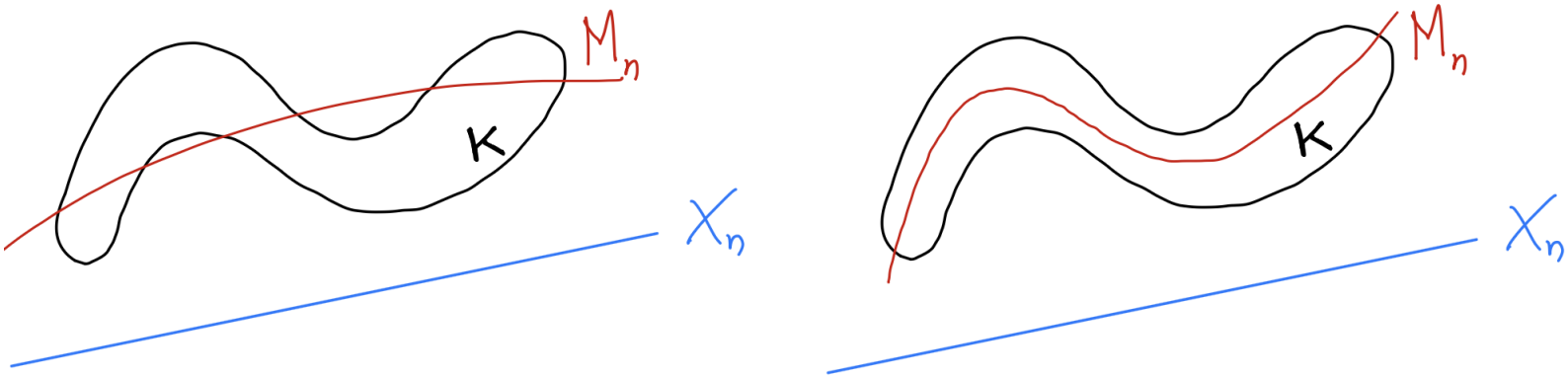
where the infimum is taken over linear 1-Lipschitz encoders and L-Lipschitz decoders.

Equivalently

$$e_n^L(K)_X = \inf_{\dim(X_n)=n} \inf_{Lip(A)\leq L} \sup_{u\in K} \|u - A(P_{X_n}u)\|_X$$

where

$$Lip(A) = \sup_{\substack{u,v\in K \\ P_{X_n}(u-v)\neq 0}} \frac{\|A(P_{X_n}u) - A(P_{X_n}u)\|_X}{\|P_{X_n}u - P_{X_n}v\|_X} \leq L$$



(a) Low L (b) High L

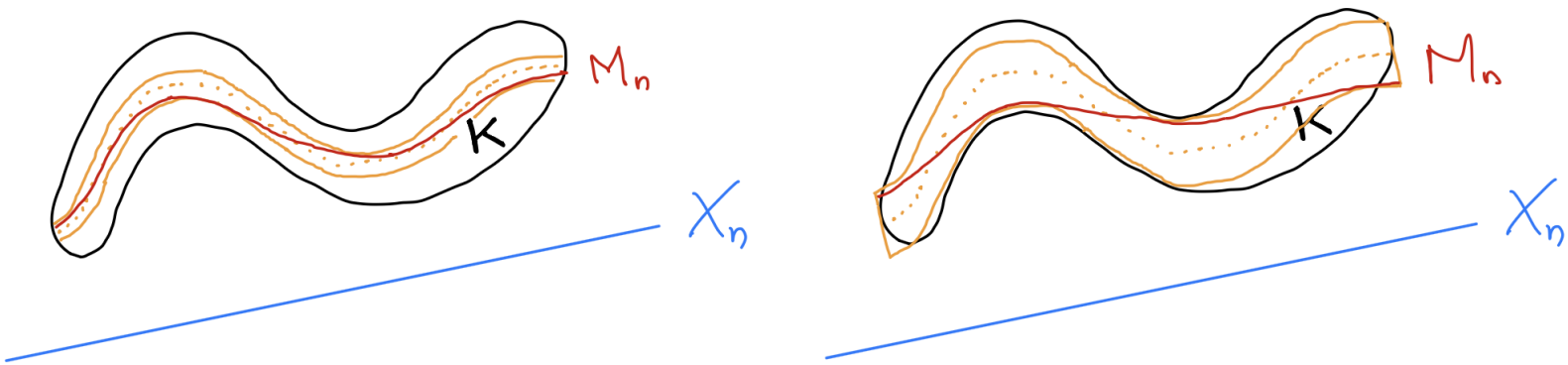
Optimal spaces guided by Lipschitz stability (worst case setting)

A dual version is

$$\tilde{e}_n^{\eta}(K)_X = \inf_{\dim(X_n)=n} \inf_{A} Lip(A)$$

where the infimum is taken over maps A such that

$$\sup_{u \in K} \|u - A(P_{X_n} u)\|_X \leq \eta$$

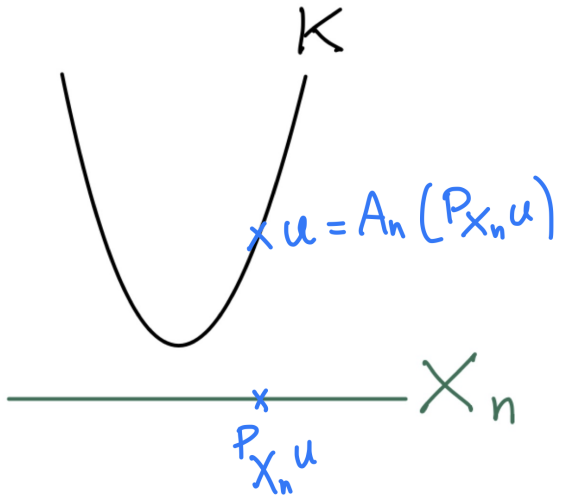


(c) Low η (d) High η

$e_n^L(K)_X$ or $\tilde{e}_n^{\eta}(K)_X$ can be seen as notions of "Lipschitz width" of K, using linear encoder. See related notion in [Petrova and Wojtaszczyk 2023] with arbitrary encoders.

Optimal spaces guided by Lipschitz stability (worst case setting)

Consider $\eta = 0$, assuming the existence of an exact recovery map A_n for some X_n. This requires K to be a n-dimensional manifold, and A_n is a global chart associated to X_n.



Then we obtain a measure of Lipschitz regularity of K related to X_n as

$$Lip(A_n) = \sup_{\substack{u,v \in K \\ P_{X_n}(u-v) \neq 0}} \frac{\|u - v\|_X}{\|P_{X_n}(u - v)\|_X} =: Lip(K, X_n)_X$$

By optimization over all space X_n, we obtain a notion of Lipschitz width of K

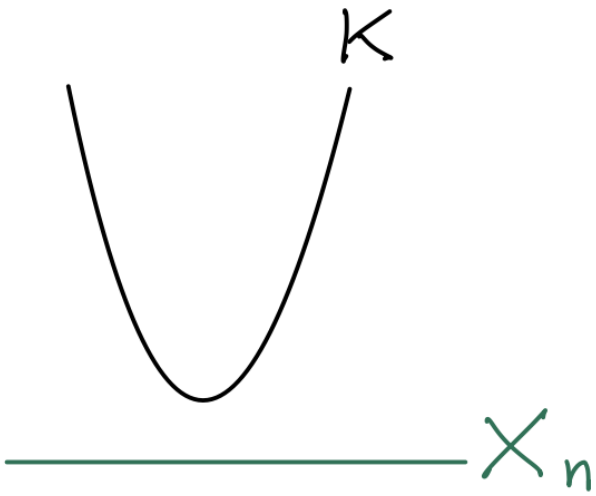
$$L_n(K)_X := \inf_{\dim(X_n)=n} Lip(K, X_n)_X$$

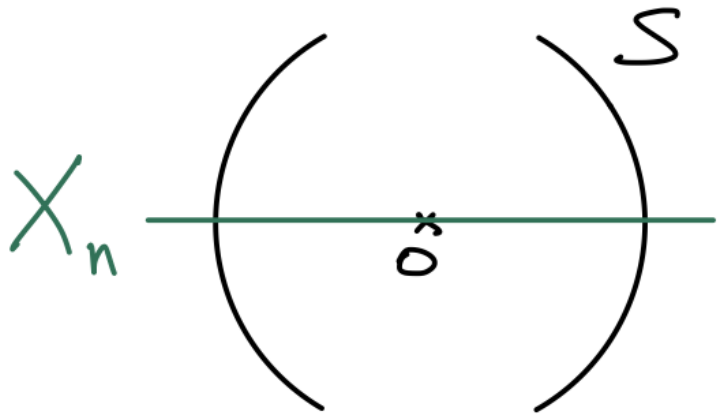
which defines an optimal space X_n in terms of stability of the corresponding chart A_n.

Optimal spaces guided by Lipschitz stability (worst case setting)

It holds

$$L_n(K)_X^2 = \frac{1}{1 - d_n(S)_X^2} \quad \text{with} \quad S = \{\frac{z}{\|z\|_X} : z \in (K - K) \setminus \{0\}\}.$$





$$\boxed{L_n(K)_X = \min_{X_n} Lip(K, X_n)_X} \iff \boxed{d_n(S)_X = \min_{X_n} d(S, X_n)_X}$$

Close to optimal spaces can be obtained by greedy algorithms.

Optimal spaces guided by Lipschitz stability (average setting)

In the average setting, possible relaxation by considering the average width $d^{(2)}(S,\pi)_X$, with π the push-forward measure on S of $\rho \otimes \rho$ through the map $(u,v) \mapsto (u-v)/\|u-v\|_X$.

Instead of

$$d_n(S) = \inf_{\dim(X_n)=n} \sup_{z \in S} \|z - P_{X_n} z\|_X$$

we consider

$$d^{(2)}(S,\pi)_X = \inf_{\dim(X_n)=n} \int_S \|z - P_{X_n} z\|_X^2 d\pi(z)$$

that is an eigenvalue problem.

In practice, we introduce estimators from finitely many samples in K.

Greedy selection from a dictionary (worst case setting)

An approximate solution of

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

can be obtained by restricting the space X_n to be generated from a dictionary $\mathcal{D} = \{\varphi_i\}_{i \in I}$.

We can construct a sequence of decreasing spaces $X_n = span\{\varphi_i\}_{i \in I_n}$ in a greedy fashion by setting

$$I_{n-1} = I_n \setminus \{j_n\}$$

with

$$j_n \in \min_j Lip(K, span\{\varphi_i\}_{i \in I_n \setminus j})_X$$

Illustration: wave propagation

$$\partial_{tt} s(x,t) - \Delta s(x,t) = 0, \quad (x,t) \in \Omega \times (0,T),$$
$$s(x,0) = \exp(-\alpha \|x - \xi\|^2), \ \partial_t s(x,0) = 0, \quad \partial_n s(x,t) = 0 \quad \text{on } \partial\Omega$$
$$\boxed{K = \{u = s(\cdot,t) : t \in (0,T)\}}$$

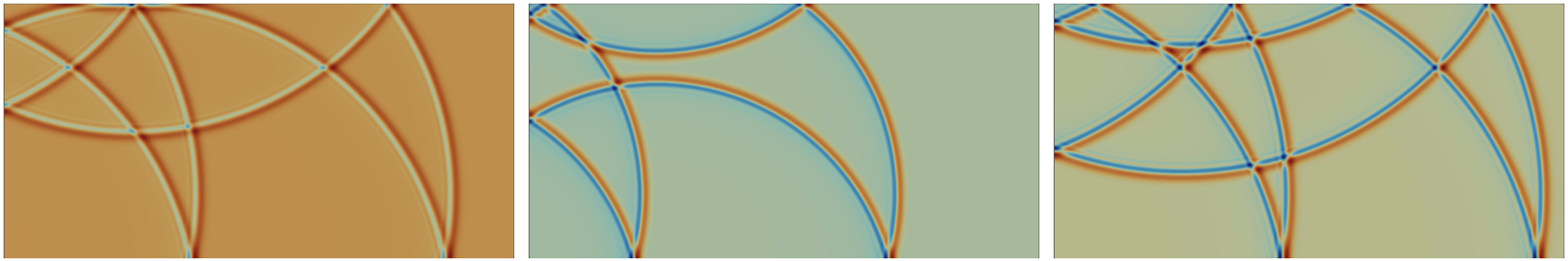


Figure: Some samples of the manifold K allowing the offline construction of M_n.

Method	p	n	N	RE_{train}	RE_{test}
Linear	/	9	/	8.05×10^{-1}	8.28×10^{-1}
Quadratic	2	9	54	4.08×10^{-1}	4.12×10^{-1}
Sparse poly.	6	9	84	2.65×10^{-2}	$3. \times 10^{-2}$
CPN-S ($\epsilon = 10^{-2}$)	6	9	84	8.95×10^{-3}	1.01×10^{-2}

Table: (2D wave equation) Comparison of methods for the same manifold dimension $n = 9$.

Manifold approximation using PCA basis with ordering guided by approximation properties (linear or quadratic approximation) or Lipschitz regularity. Use of sparse polynomial decoder.

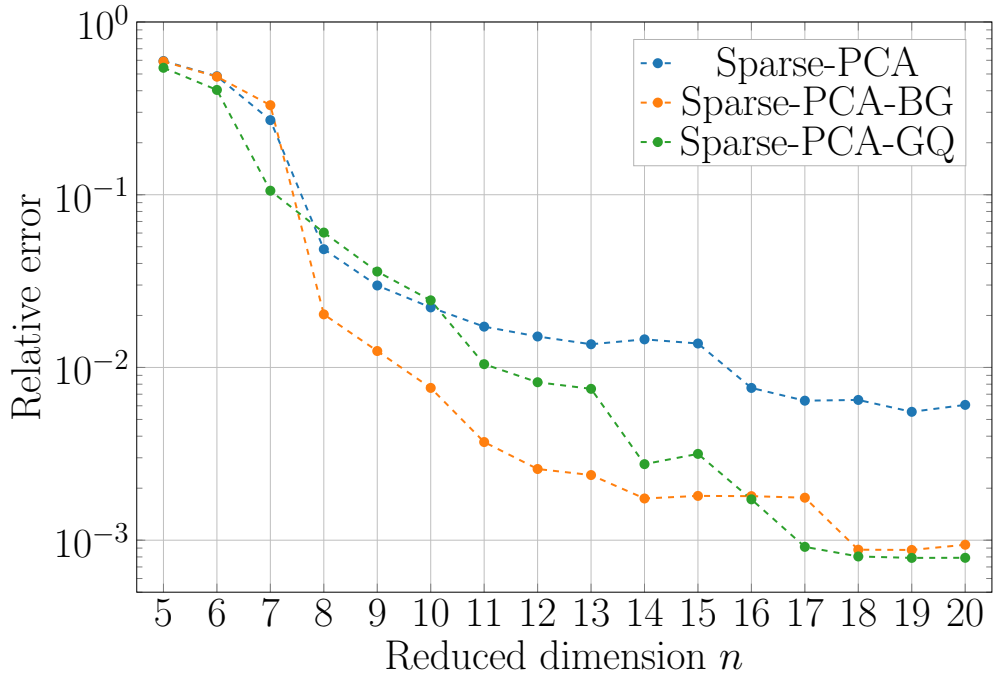


Figure: Error versus manifold dimension n. Optimal ordering for linear approximation (PCA), for quadratic approximation (PCA-GC) or for Lipschitz stability (PCA-BG).

Optimal spaces guided by Lipschitz stability

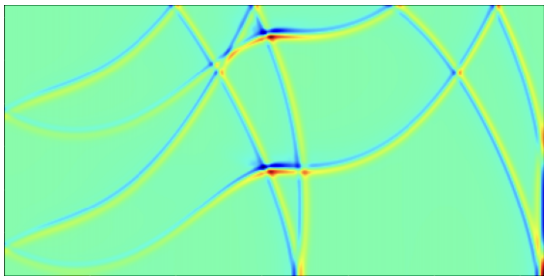
Previous notions may not be well defined for sets K that are not smooth n-dimensional manifolds described by global chart.

This requires to consider relaxations $\tilde{e}_n^{\eta}(K)_X$ that balance Lipschitz stability and approximation error, or n-dimensional approximations of the set K.

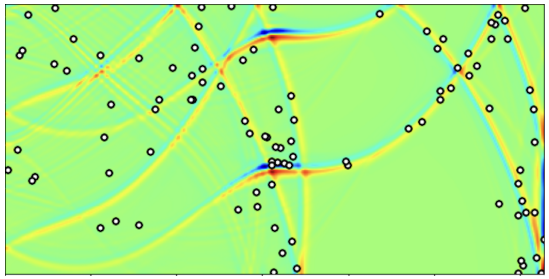
The design of robust strategies using (noisy) samples from K is a challenging problem.

Conclusions

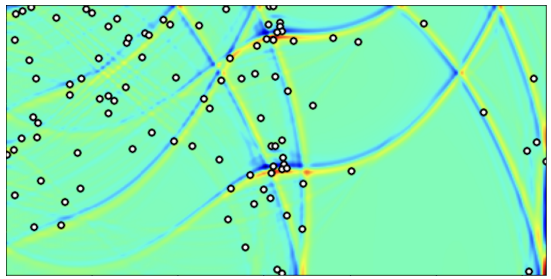
- Stable nonlinear manifold approximation with linear encoders and nonlinear decoder based on compositional polynomial networks, with control of error and stability [1].
- Optimal encoders guided by Lipschitz stability of decoders.
- Allow for efficient online approximation with low computational complexity, or from limited (adaptive) information, by combination of optimization algorithms on manifolds and optimal sampling [2,3].



(a) Ground truth



(b) Approximation at step 0



(c) Approximation at step 1

[1] A. Bensalah, A. Nouy, and J. Soffo. Nonlinear manifold approximation using compositional polynomial networks. *SIAM J. Sci. Comput.*, June 2026

[2] R. Gruhlke, A. Nouy and P. Trunschke. Optimal sampling for stochastic and natural gradient descent: arXiv:2402.03113.

[3] A. Nouy and A. Somacal. Natural gradient descent with momentum. arXiv, Apr. 2026.

References I

A. Bensalah, A. Nouy, and J. Soffo.
Nonlinear Manifold Approximation Using Compositional Polynomial Networks.
SIAM J. Sci. Comput., June 2026.

R. DeVore, G. Petrova, and P. Wojtaszczyk.
Greedy algorithms for reduced bases in Banach spaces.
Constructive Approximation, 37(3):455–466, 2013.

J. Barnett and C. Farhat.
Quadratic approximation manifold for mitigating the kolmogorov barrier in nonlinear projection-based model order reduction.
Journal of Computational Physics, 464:111348, Sept. 2022.

J. Barnett, C. Farhat, and Y. Maday.
Neural-network-augmented projection-based model order reduction for mitigating the Kolmogorov barrier to reducibility.
J. Comput. Phys., 492:112420, Nov. 2023.

R. Geelen, S. Wright, and K. Willcox.
Operator inference for non-intrusive model reduction with quadratic manifolds.
Computer Methods in Applied Mechanics and Engineering, 403:115717, Jan. 2023.

P. Schwerdtner and B. Peherstorfer.
Greedy construction of quadratic manifolds for nonlinear dimensionality reduction and nonlinear model reduction, 2024.

References II

R. Geelen, L. Balzano, S. Wright, and K. Willcox.
Learning physics-based reduced-order models from data using nonlinear manifolds.
Chaos, 34(3):033122, Mar. 2024.

A. Cohen, C. Farhat, Y. Maday, and A. Somacal.
Nonlinear compressive reduced basis approximation for PDE's.
Comptes Rendus. Mécanique, 351(S1):357–374, 2023.

A. Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk.
Optimal Stable Nonlinear Approximation.
Found. Comput. Math., 22(3):607–648, June 2022.

R. A. DeVore, R. Howard, and C. Micchelli.
Optimal nonlinear approximation.
Manuscripta mathematica, 63(4):469–478, 1989.

J. Creutzig.
Relations between Classical, Average, and Probabilistic Kolmogorov Widths.
J. Complexity, 18:287–303, Mar. 2002.

V. N. Temlyakov.
Nonlinear Kolmogorov widths.
Math. Notes, 63(6):785–795, June 1998.

J. L. Eftang, A. T. Patera, and E. M. Rønquist.
An "hp" certified reduced basis method for parametrized elliptic partial differential equations.
SIAM Journal on Scientific Computing, 32(6):3170–3200, 2010.

References III

S. Kaulmann and B. Haasdonk.

Online greedy reduced basis construction using dictionaries.

In *VI International Conference on Adaptive Modeling and Simulation (ADMOS 2013)*, pages 365–376, 2013.

O. Balabanov and A. Nouy.

Randomized linear algebra for model reduction—part ii: minimal residual methods and dictionary-based approximation.

Advances in Computational Mathematics, 47(2):1–54, 2021.

A. Nouy and A. Pasco.

Dictionary-based model reduction for state estimation.

Adv. Comput. Math., 50(3):1–31, June 2024.

A. Bonito, A. Cohen, R. DeVore, D. Guignard, P. Jantsch, and G. Petrova.

Nonlinear methods for model reduction.

ESAIM: M2AN, 55(2):507–531, Mar. 2021.

A. Cohen, W. Dahmen, O. Mula, and J. Nichols.

Nonlinear Reduced Models for State and Parameter Estimation.

SIAM/ASA Journal on Uncertainty Quantification, Feb. 2022.

D. Guignard and O. Mula.

Tree-Based Nonlinear Reduced Modeling.

In *Multiscale, Nonlinear and Adaptive Approximation II*, pages 267–298. Springer, Cham, Switzerland, Dec. 2024.

References IV

G. Petrova and P. Wojtaszczyk.
Lipschitz Widths.
Constr. Approx., 57(2):759–805, Apr. 2023.