

Linear and nonlinear methods for model reduction

Diane Guignard



Joint work:

A. Bonito, R. DeVore, P. Jantsch, and G. Petrova (TAMU)
A. Cohen (Sorbonne Université)

Workshop - Mathematics of Reduced Order Models

February 17-21, 2020

ICERM

Outline

- 1 Introduction
- 2 Linear reduced methods
- 3 Nonlinear reduced methods
- 4 Conclusion

Outline

- 1 Introduction
- 2 Linear reduced methods
- 3 Nonlinear reduced methods
- 4 Conclusion

Introduction

Many real world applications lead to models with a large number of input parameters, such as weather forecast, optimal engineering design or option pricing.

Goal

Fast and efficient numerical approximation of a **high-dimensional function**

$$u : Y \subset \mathbb{R}^d \rightarrow V$$

with $d \gg 1$ (possibly infinite) and V a Banach space.

Typical example: u is the solution of some parametric/random PDE

Input Parameter

PDE

Solution

$$y \in Y \subset \mathbb{R}^d \quad \longrightarrow \quad \mathcal{P}(u, y) = 0 \quad \longrightarrow \quad u(y) \in V.$$

Introduction II

Approximation: for $y \in Y$

$$u(y) \in V \quad \approx \quad u_n(y) \in V_n.$$

Types of approximation:

- **Linear:** V_n is a linear space of dimension n , for instance
 - ▶ reduced basis space;
 - ▶ (Taylor) polynomial space.
- **Nonlinear:** V_n is a nonlinear space depending on n parameters, for instance
 - ▶ best n -term approximation from a dictionary;
 - ▶ some adaptive approximations.

Error: the space V is endowed with some norm $\|\cdot\|_V$ and the approximation error for $u(y) \in V$ is

$$\inf_{v_n \in V_n} \|u(y) - v_n\|_V.$$

Performance of a reduced model

Usually, we are not interested in a good approximation of $u(y)$ for **one fixed** y but for a certain model class \mathcal{K} .

In the **parametric PDE setting**: we consider the solution manifold

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

and the goal is to build V_n that *minimize* the worst error

$$\sup_{v \in \mathcal{M}} \inf_{v_n \in V_n} \|v - v_n\|_V.$$

Optimality: the best performance achievable by a **linear reduced model** is given by the Kolmogorov n -width

For any compact set $K \subset V$

$$d_n(K) := \inf_{\dim(V_n)=n} \sup_{v \in K} \inf_{v_n \in V_n} \|v - v_n\|_V.$$

Nonlinear methods can be performed better and widths for **nonlinear reduced model** can be defined in different ways.

Setting

- To get results that are immune to the so-called **curse of dimensionality**, we consider the case $d = \infty$ and we set $Y := [-1, 1]^{\mathbb{N}}$.
- We write \mathcal{F} the set of all finitely supported sequences $\nu = (\nu_1, \nu_2, \dots)$ with $\nu_j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$.
- We consider **Taylor polynomial approximations**: given a finite subset $\Lambda \subset \mathcal{F}$

$$u(y) \approx \sum_{\nu \in \Lambda} t_{\nu} y^{\nu},$$

for some $t_{\nu} \in V$ and $y^{\nu} := \prod_{j \geq 1} y_j^{\nu_j}$.

- We restrict ourselves to **lower sets** (or downward closed sets), namely sets for which

$$\nu \in \Lambda \quad \text{and} \quad \mu \leq \nu \quad \implies \quad \mu \in \Lambda.$$

Outline

- 1 Introduction
- 2 Linear reduced methods**
- 3 Nonlinear reduced methods
- 4 Conclusion

Goal

The design and analysis of (near) optimal linear reduced models is well developed in the framework of parametric PDEs, see for instance

- Polynomial basis: [Beck-Nobile-Tamellini-Tempone, 2012], [Chkifa-Cohen-DeVore-Schwab, 2013], [Tran-Webster-Zhang, 2017], [Bachmayr-Cohen-Migliorati, 2017].
- Reduced basis: [Maday-Patera-Turinici, 2002], [Roza-Huynh-Patera, 2008], [DeVore-Petrova-Wojtaszczyk, 2013].

The goal here is to:

- *move away* from the PDE setting (obtain approximation results without using the PDE theory);
- obtain sharp error estimates for all n (not only for n sufficiently large).

Class of anisotropic analytic functions

Norm: we want to approximate Banach space valued functions $u : Y \rightarrow V$ in the norm

$$\|u\|_{L_\infty(Y, V)} := \sup_{y \in Y} \|u(y)\|_V.$$

Sequence: let $\rho = (\rho_j)_{j \geq 1}$ be a non-decreasing sequence with $\rho_1 > 1$ and $\lim_{j \rightarrow \infty} \rho_j = \infty$.

Model class: for any $0 < p \leq \infty$, let $\mathcal{B}_{\rho, p}$ be the set of all $u \in L_\infty(Y, V)$ which can be represented uniquely by

$$u(y) = \sum_{\nu \in \mathcal{F}} t_\nu y^\nu$$

with uniform and unconditional convergence on Y , and such that

$$\|u\|_{\mathcal{B}_{\rho, p}} := \left(\sum_{\nu \in \mathcal{F}} [\rho^\nu \|t_\nu\|_V]^p \right)^{1/p} < \infty.$$

Measure of performance for $\mathcal{B}_{\rho,p}$

We approximate $u \in \mathcal{B}_{\rho,p}$ using Taylor polynomials: given a finite set $\Lambda \subset \mathcal{F}$

$$u(y) \approx \sum_{\nu \in \Lambda} t_\nu y^\nu, \quad t_\nu := t_\nu(u) := \frac{\partial_\nu u(0)}{\nu!}.$$

For a model class K of functions in $L_\infty(Y, V)$, the performance of a lower set Λ_n of cardinality n is controlled by

$$e_n(K) := \inf_{\#\Lambda \leq n} \sup_{u \in K} \sup_{y \in Y} \|u(y) - \sum_{\nu \in \Lambda} t_\nu y^\nu\|_V.$$

Given the model class $K = \mathcal{B}_{\rho,p}$ (or its unit ball), what can we say about

- the decay rate of $e_n(K)$ as n increases;
- the sharpness of the error bounds;
- the construction of a (near) optimal lower set Λ of cardinality n ?

Approximation of functions in $\mathcal{B}_{\rho,p}$

Let $(\delta_n)_{n \geq 1} := (\delta_n(\rho))_{n \geq 1}$ be a decreasing rearrangement of $(\rho^{-\nu})_{\nu \in \mathcal{F}}$ and let

$$\Lambda_n := \Lambda_{n,\rho} := \{\nu \in \mathcal{F} \text{ corresponding to the } n \text{ largest } \rho^{-\nu}\},$$

where ties are handled arbitrarily but so that Λ_n is a lower set of cardinality n .

Theorem

Let $1 \leq p \leq \infty$ and let p' be the conjugate of p . Then for all $u \in \mathcal{B}_{\rho,p}$ we have

$$\sup_{y \in Y} \|u(y) - \sum_{\nu \in \Lambda_n} t_\nu y^\nu\|_V \leq \|u\|_{\mathcal{B}_{\rho,p}} \begin{cases} \left(\sum_{k > n} \delta_k^{p'} \right)^{\frac{1}{p'}} & \text{if } 1 \leq p' < \infty \\ \delta_{n+1} & \text{if } p' = \infty. \end{cases}$$

Moreover, the set Λ_n is optimal in the sense that it minimizes the surrogate error

$$\sup_{u \in \mathcal{B}_{\rho,p}} \sum_{\nu \notin \Lambda} \|t_\nu\|_V$$

among all lower set Λ with $\#\Lambda = n$.

The sequence $\delta_n(\rho)$

From the previous results, the approximation error is controlled by the sequence

$$(\delta_n)_{n \geq 1} = (\delta_n(\rho))_{n \geq 1},$$

where δ_n is the n^{th} largest $\rho^{-\nu}$.

In order to compute δ_n or its decay as n increases, we study $\#\Lambda(\varepsilon, \rho)$, where

$$\Lambda(\varepsilon, \rho) := \{\nu \in \mathcal{F} : \rho^{-\nu} \geq \varepsilon\} = \{\nu \in \mathcal{F} : \rho^\nu \leq \varepsilon^{-1}\}.$$

Properties:

- $\#\Lambda(\varepsilon, \rho) < \infty$ whenever $\varepsilon > 0$;
- $\Lambda(\varepsilon, \rho)$ is a lower set since $\mu \leq \nu \Rightarrow \rho^{-\nu} \leq \rho^{-\mu}$;
- $\Lambda(\varepsilon, \rho) \subset \Lambda(\varepsilon', \rho)$ whenever $\varepsilon' \leq \varepsilon$;
- $\Lambda(\delta_n, \rho) \geq n$

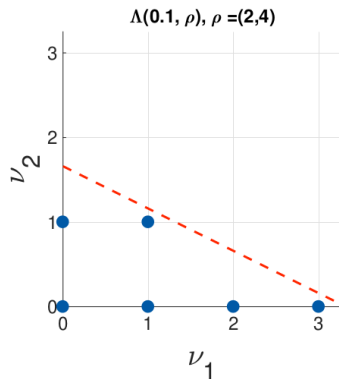
Remark: as a function of ε , $\#\Lambda(\varepsilon, \rho)$ is a piecewise constant function and $(\delta_n(\rho))_{n \geq 1}$ is the decreasing sequence of the breakpoints $\varepsilon_1, \varepsilon_2, \dots$ of $\#\Lambda(\varepsilon, \rho)$.

$\#\Lambda(\varepsilon, \rho)$ as lattice points in a simplex

There is a $D = D(\varepsilon)$ such that $\rho_j^{-1} < \varepsilon$ for $j > D$ and thus any $\nu \in \Lambda(\varepsilon, \rho)$ has support in $\{1, 2, \dots, D\}$.

$$\rho^{-\nu} \geq \varepsilon \iff \sum_{j=1}^D \nu_j \underbrace{\frac{\ln \rho_j}{\ln \varepsilon^{-1}}}_{\leq 1} \leq 1.$$

Hence, $\nu \in \Lambda(\varepsilon, \rho)$ if and only if ν is an integer lattice point in a simplex.



- Estimating the number of lattice points in such a simplex is a classical problem in number theory and combinatorics.
- Existing bounds are only asymptotic and not sharp for sets of small/moderate cardinality.

Specific sequences: polynomial growth

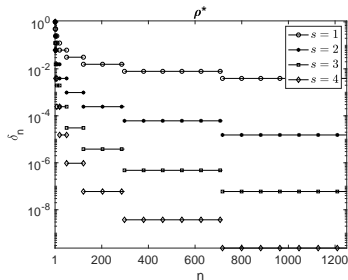
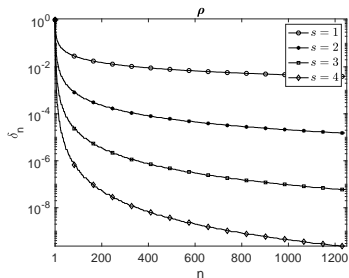
We can achieve better results for sequences $\rho := \rho(s)$, $s > 0$, of the form

$$\rho_j(s) := (j + 1)^s, \quad j \geq 1,$$

or the slightly modified sequence $\rho^* = \rho^*(s)$ defined by

$$\rho_j^* := 2^{ks}$$

for $j \in I_k$, where $I_1 := \{1, 2\}$ and $I_k := \{j : 2^{k-1} < j \leq 2^k\}$, $k \geq 2$.



Estimation of $\#\Lambda(\varepsilon, \rho)$: exact count

- Counting lattice point in the simplex described by the sequence $\rho(s)$ is directly related to counting the number of **multiplicative partitions** of integers [Canfield-Erdős-Pomerance, 1983],[Cohen-DeVore, 2015].

Exact counts are known only for some values of n and the computation is very intensive.

- For the modified sequence $\rho^*(s)$, the count is related to **additive partitions** of integers, which are easier to compute numerically.

Theorem

$$\#\Lambda(2^{-ms}, \rho^*(s)) = 1 + \sum_{k=1}^m \sum_{(N_1, \dots, N_k) \in \mathcal{Q}_k} \prod_{j=1}^k \binom{N_j - 1 + \#l_j}{N_j},$$

where

$$\mathcal{Q}_k := \{(N_1, \dots, N_k) \in \mathbb{N}_0^k : \sum_{j=1}^k jN_j = k\}$$

There is a one-to-one correspondance between the elements of \mathcal{Q}_k and additive partitions of k .

Estimation of $\#\Lambda(\varepsilon, \rho)$: upper bound

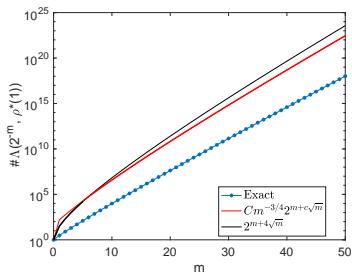
Theorem

For $m = 0$, $\#\Lambda(2^{-ms}, \rho^*(s)) = 1$, when $m = 1$, $\#\Lambda(2^{-ms}, \rho^*(s)) = 3$, and

$$\#\Lambda(2^{-ms}, \rho^*(s)) \leq \begin{cases} 2^{m+4\sqrt{m}}, & 2 \leq m \leq 5, \\ Cm^{-3/4}2^{m+c\sqrt{m}}, & m \geq 6. \end{cases}$$

where $C \approx 6.3$ and $c \approx 4$.

Comparison of the upper bound and the exact count (case $s = 1$)



Estimation of $\#\Lambda(\varepsilon, \rho)$: upper bound

Theorem

For $m = 0$, $\#\Lambda(2^{-ms}, \rho^*(s)) = 1$, when $m = 1$, $\#\Lambda(2^{-ms}, \rho^*(s)) = 3$, and

$$\#\Lambda(2^{-ms}, \rho^*(s)) \leq \begin{cases} 2^{m+4\sqrt{m}}, & 2 \leq m \leq 5, \\ Cm^{-3/4}2^{m+c\sqrt{m}}, & m \geq 6. \end{cases}$$

where $C \approx 6.3$ and $c \approx 4$.

Corollary

We have $\delta_n(\rho^*(s)) \leq 2^{-6s} n^{\frac{4s\sqrt{4+\log_2 n}}{\log_2 n}} n^{-s}$ and thus

$$\sup_{y \in Y} \|u(y) - \sum_{\nu \in \Lambda_n} t_\nu y^\nu\| \leq \|u\|_{B_{\rho^*,1}} 2^{-6s} n^{\frac{4s\sqrt{4+\log_2 n}}{\log_2 n}} n^{-s}$$

for any $u \in B_{\rho^*,1}$ (and similarly for $1 < p \leq \infty$).

Outline

- 1 Introduction
- 2 Linear reduced methods
- 3 Nonlinear reduced methods**
- 4 Conclusion

Motivational example

For $R > 1$, consider the rational function

$$f(y) = \frac{1}{R - y}, \quad y \in [-1, 1].$$

Using a truncated Taylor series about $y = 0$, the required number of terms n (degree $n - 1$) to achieve a prescribed accuracy ε is

ε	n
10^{-3}	19
10^{-4}	25
10^{-5}	31

$$R = 1.5$$

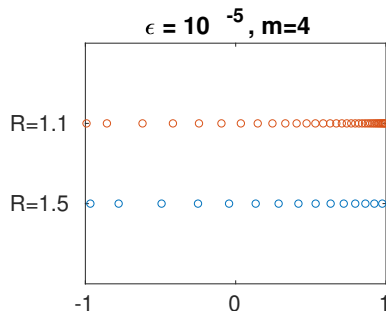
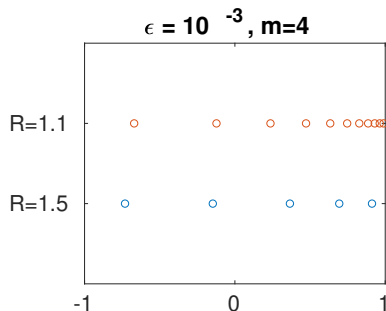
ε	n
10^{-3}	97
10^{-4}	121
10^{-5}	145

$$R = 1.1$$

Motivational example II

We can achieve the same accuracy with fewer terms if we use piecewise Taylor polynomials. Idea:

- Partition the interval $[-1, 1]$ into subintervals $[y_i, y_{i+1}]$, $i = 1, \dots, N$.
- Use a truncated Taylor series about $y = \frac{y_i + y_{i+1}}{2}$ with $m < n$ terms.



Motivation example III

Comparison of the required number of terms: n versus Nm .

ε	n
10^{-3}	19
10^{-4}	25
10^{-5}	31

$\varepsilon \setminus m$	4	5	6
10^{-3}	5	4	3
10^{-4}	9	6	4
10^{-5}	15	9	6

$$R = 1.5$$

ε	n
10^{-3}	97
10^{-4}	121
10^{-5}	145

$\varepsilon \setminus m$	4	5	6
10^{-3}	11	8	6
10^{-4}	20	12	9
10^{-5}	35	19	13

$$R = 1.1$$

Nonlinear reduced model

In some cases, the use of a linear space V_n is not possible (e.g. slow error decay, small target accuracy, data assimilation framework).

Often, linear methods are outperformed by numerical methods based on nonlinear approximations.

Library approximation

The idea is to replace the space V_n by a collection of spaces (aka library)

$$\mathcal{L}_{m,N} := \{V^1, \dots, V^N\}$$

with $\dim(V^j) \leq m < n$ for $j = 1, \dots, N$.

This idea is not new, see for instance

- [Eftang-Patera-Rønquist, 2010]
- [Maday-Stamm, 2013]
- [Zou-Kouri-Aquino,2019],

but to our knowledge, there is no unified study of nonlinear model reduction.

Parametric PDE

Let $D \subset \mathbb{R}^d$ be a bounded Lipschitz domain, $f \in L^2(D)$, and $Y = [-1, 1]^N$ the parameter space.

Elliptic diffusion model problem

Find $u : D \times Y \rightarrow \mathbb{R}$ such that

$$\begin{cases} -\operatorname{div}(a(x, y)\nabla u(x, y)) & = f(x) & x \in D, y \in Y \\ u(x, y) & = 0 & x \in \partial D, y \in Y, \end{cases}$$

where the diffusion coefficient a has the affine form

$$a(x, y) = \bar{a}(x) + \sum_{j \geq 1} y_j \psi_j(x)$$

and satisfies the uniform ellipticity assumption

$$0 < a_{\min} \leq a(x, y) \leq a_{\max} < \infty.$$

Then, for every $y \in Y$, there exists a unique solution $u(y) \in V := H_0^1(D)$.

Diffusion coefficient

The uniform ellipticity assumption is equivalent to

$$\left\| \frac{\sum_{j \geq 1} |\psi_j|}{\bar{a}} \right\|_{L^\infty(D)} < 1.$$

To prove results on polynomial approximation for $u(y)$, we need further assumptions on the diffusion coefficient.

Here we assume that there exists a non-decreasing sequence $(\rho_j)_{j \geq 1}$ with $\rho_1 \geq \kappa > 1$ and $(\rho_j^{-1})_{j \geq 1} \in \ell_q(\mathbb{N})$ such that

$$\delta := \left\| \frac{\sum_{j \geq 1} \rho_j |\psi_j|}{\bar{a}} \right\|_{L^\infty(D)} < 1.$$

Remark: with these assumptions, $u \in \mathcal{B}_{\rho,2}$ (see [Bachmayr-Cohen-Migliorati, 2017]).

Model class and nonlinear width

Model class: here, the model class is the solution manifold

$$\mathcal{M} = \{u(y) : y \in Y\} \subset V.$$

Nonlinear width: for library approximations, as suitable choice (due to [Temlyakov](#)) is the **library width**

$$d_{m,N}(\mathcal{M}) := \inf_{\mathcal{L}} \sup_{y \in Y} \inf_{V_j \in \mathcal{L}} \inf_{v_m \in V_j} \|u(y) - v_m\|_V$$

where the infimum is taken over all libraries with N spaces of dimension m .

The two extreme cases are

- $N = 1$: linear width, $d_{m,1}(\mathcal{M}) = d_m(\mathcal{M})$, might need $m \gg 1$
- $m = 1$: entropy, $d_{1,2^n} = \varepsilon_n(\mathcal{M})$, might need $N \gg 1$.

Can we get a better model using an intermediate value for m ?

Construction of a library: general idea

Piecewise Taylor polynomials

As in the motivational example, the idea is to partition the parameter domain

$$Y = \cup_{i=1}^N Q_i$$

and use a local Taylor polynomial with m terms on each Q_i

$$u(y) \approx \sum_{\nu \in \Lambda_m} \frac{\partial^\nu u(\bar{y}^i)}{\nu!} (y - \bar{y}^i)^\nu, \quad y \in Q_i.$$

Typical questions: for a fixed accuracy ε and a fixed number of terms m

- How large N needs to be?
- How to construct the partition?

Type of partition

For simplicity (mathematics and practical), we consider rectangular subdomains of the form

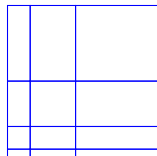
$$Q := Q_\lambda(\bar{y}) := \{y \in Y : |y_j - \bar{y}_j| \leq \lambda_j, \quad j \geq 1\} \subset Y$$

with **center** $\bar{y} \in Y$ and **half side-lengths** $(\lambda_j)_{j \geq 1}$.

Main idea

Control the error by making the side-length of the rectangles sufficiently small.

This requires a **local error estimate**.



Global error estimate

Following [Bachmayr-Cohen-Migliorati, 2017] we can derive the error estimate

Theorem

For each $m \geq 1$, there exists a set Λ_m with $\#\Lambda_m = m$ such that

$$E_m(Y) := \sup_{y \in Y} \|u(y) - \sum_{\nu \in \Lambda_m} t_\nu y^\nu\|_V \leq C \|(\rho_j^{-1})_{j \geq 1}\|_{\ell_q} m^{-r}, \quad r = \frac{1}{q} - \frac{1}{2},$$

for some constant $C = C(\delta, \rho, q)$.

Important remark: we can take

$$\Lambda_m := \{\nu \in \mathcal{F} \text{ corresponding to the } m \text{ largest } \rho^{-\nu}\}$$

to be a lower set and it can be computed *a priori* (only requires the sequence ρ).

Local error estimate

For the error on a subdomain $Q = Q_\lambda(\bar{y})$, we use a scaling - shifting argument to get

Corollary

If for $j \geq 1$

$$\tilde{\rho}_j := \frac{\rho_j - |\bar{y}_j|}{\lambda_j} \geq \kappa > 1 \quad \text{and} \quad \|(\tilde{\rho}_j^{-1})\|_{\ell_q} \leq \|(\rho_j^{-1})\|_{\ell_q}$$

then for each $m \geq 1$, there exists a polynomial P_m with m terms such that

$$E_m(Q) := \sup_{y \in Q} \|u(y) - P_m(y)\|_V \leq C \|(\tilde{\rho}_j^{-1})_{j \geq 1}\|_{\ell_q} m^{-r}, \quad r = \frac{1}{q} - \frac{1}{2}.$$

Sufficient condition to have $E_m(Q) \leq \varepsilon$:

$$C \|(\tilde{\rho}_j^{-1})\|_{\ell_q} m^{-r} \leq \varepsilon \quad \iff \quad \sum_{j \geq 1} \tilde{\rho}_j^{-q} \leq C^{-q} m^{rq} \varepsilon^q =: \eta.$$

Upper bound on the size of the library

Theorem

Given $\varepsilon > 0$ and $m \geq 1$, let J be the smallest integer satisfying

$$\sum_{j \geq J+1} \rho_j^{-q} \leq \frac{1}{2}\eta, \quad \text{and let } \sigma^q := \frac{1}{2J}\eta.$$

Then there exists a partition of Y into

$$N := N(\varepsilon, m) \leq \prod_{j=1}^J \left[\sigma^{-1} |\ln(1 - \rho_j^{-1})| + 1 \right]$$

hyperrectangles $(Q_i)_{i=1}^N$ such that $E_m(Q_i) \leq \varepsilon$ for $i = 1, \dots, N$.

Key points:

- Upper bound on the number of spaces.
- Explicit construction of the partition in the proof (tensor-based).

Sketch of the proof

- The *directions* $j > J$ are not partitioned (i.e., $\bar{y}_j = 0$ and $\lambda_j = 1$).
- For a given center \bar{y} , accuracy ε is reached using m terms if

$$\sum_{j=1}^J \tilde{\rho}_j^{-q} \leq \frac{1}{2}\eta, \quad \tilde{\rho}_j = \frac{\rho_j - |\bar{y}_j|}{\lambda_j}.$$

- A sufficient condition is to choose λ_j such that $\tilde{\rho}_j^{-q} = \frac{1}{2J}\eta$.
- Using this criteria and proceeding as in the motivational example, each direction $j = 1, \dots, J$ is partitioned into n_j subdomains.
- The collection of centers are points on a tensor product grid of the first J coordinates.
- The total number of cells is

$$N = \prod_{j=1}^J n_j.$$

Specific sequence: polynomial growth

Consider the sequence

$$\rho_j = Mj^s, \quad j \geq 1,$$

where $M > 1$ and for which $(\rho_j^{-1})_{j \geq 1} \in \ell_q(\mathbb{N})$ for any $q > 1/s$.

The bound on the required number of subdomains N reads

$$N(\varepsilon, m) \leq 2^{c(\varepsilon m^r)^{\frac{q}{1-qs}}} = 2^{\tilde{c}\left(\frac{n}{m}\right)^{\frac{qr}{qs-1}}}, \quad r = \frac{1}{q} - \frac{1}{2}$$

with n the number of terms needed to get accuracy ε with one cell.

Remarks:

- case $m = 1$: this bound is consistent with Carl's inequality [Pisier, 1989];
- improvement compared to previous bounds of the form $2^{c(n-m)}$.

Numerical example: characteristic function

Setup:

- Physical domain: $D = (0, 1)^2$.
- Forcing term: $f = 1$.
- Diffusion coefficient (piecewise constant): for a partition of D into square cells D_j

$$a(x, y) := 1 + \sum_{j=1}^{64} y_j c_j \chi_{D_j}(x), \quad c_j := (1 - a_{\min}) j^{-s}, \quad j = 1, \dots, 64,$$

for $s \in \{2, 4\}$ and $a_{\min} \in \{0.1, 0.05, 0.01\}$.

- Sequence:

$$\rho_j = \frac{1 - \frac{a_{\min}}{2}}{1 - a_{\min}} j^s$$

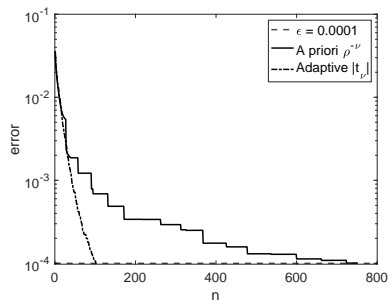
yielding $\delta = 1 - \frac{a_{\min}}{2}$.

Remark: the smaller a_{\min} the closer the coercivity constant to zero.

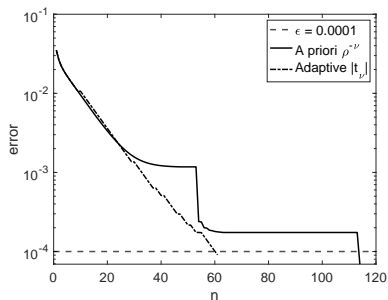
Numerical example: one cell

Selection of the lower set Λ_n :

- *a priori* (largest $\rho^{-\nu}$);
- *adaptive* (iteratively select the largest $\|t_\nu\|_V$ in the reduced margin).



$s = 2$ and $a_{\min} = 0.1$



$s = 4$ and $a_{\min} = 0.1$

Numerical example: multi cells

Number of terms m needed to achieve accuracy $\varepsilon = 10^{-4}$ for a given partition with N cells.

# of cells	$a_{min} = 0.1$		$a_{min} = 0.05$		$a_{min} = 0.01$	
	$s = 2$	$s = 4$	$s = 2$	$s = 4$	$s = 2$	$s = 4$
$N = 1$	102	61	185	128	666	603
$N = 4$	26	8	30	12	45	22
$N = 14$	22	5	24	5	29	7

Numerical example: data assimilation

For some unknown state $u(y^*)$, we are given the data

$$w_j = \ell_j(u(y^*)), \quad j = 1, \dots, L,$$

where the ℓ_j are linear functionals defined on V .

Measurement space: $W = \text{span}\{\omega_j : j = 1, \dots, L\}$ assumed to be of dimension L , where ω_j is the Riesz representant of ℓ_j .

Approximation: $\hat{u}_n \approx u(y^*)$ obtained by solving a least squares fit to the data from V_n [Binev-Cohen-Dahmen-DeVore-Petrova-Wojtaszczyk, 2011].

Performance:

$$\|u(y^*) - \hat{u}_n\|_V \leq \mu(W, V_n)\varepsilon_n,$$

where $\varepsilon_n := \text{dist}(\mathcal{M}, V_n)$ and the *inf-sup* constant $\mu(W, V_n) \geq 1$ can be view as the reciprocal of the angle between V_n and the space W .

Key observation: ε_n decreases as n increases while $\mu(W, V_n)$ increases as n increases and is ∞ if $n > L$.

Numerical example: data assimilation

Setup:

- $L = 20$ measurements that emulate point evaluation in D ;
- $s = 4$ and $a_{\min} = 0.1$.

One cell

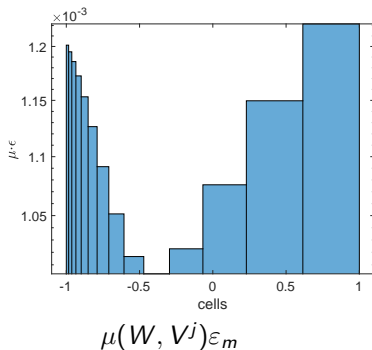
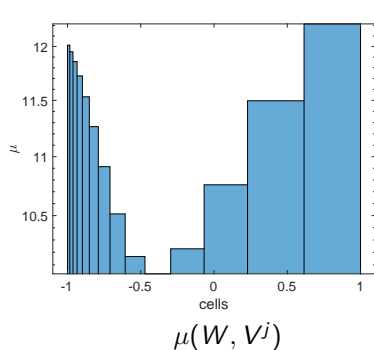
n	$\mu(W, V_n)$	$\mu(W, V_n)\varepsilon_n$	$\ u(y^*) - \hat{u}_n\ _V$
5	2.30600×10^2	3.81786×10^0	4.33473×10^{-2}
10	5.66581×10^9	5.82909×10^6	1.83385×10^4
15	7.07247×10^{11}	1.43338×10^8	6.68910×10^5

Numerical example: data assimilation

Setup:

- $L = 20$ measurements that emulate point evaluation in D ;
- $s = 4$ and $a_{\min} = 0.1$.

Multi-cell: $N = 14$ and $m = 5$ which ensures accuracy $\varepsilon_m \leq 10^{-4}$ on each cell.



Outline

- 1 Introduction
- 2 Linear reduced methods
- 3 Nonlinear reduced methods
- 4 Conclusion**

Concluding remarks

Linear reduced model:

- Approximation by Taylor polynomials described via lower sets.
- Sharp error bound for approximation of **general multivariate anisotropic functions** (model class $\mathcal{B}_{\rho,p}$) and for all values of n .
- A priori construction of an **optimal lower set** (in the surrogate norm).
- Exact count of lattice points in the simplex described by a sequence with algebraic growth ($\rho^*(s)$).

Nonlinear reduced model:

- **Library approximations** provide an alternative when standard linear reduced models fail to give satisfactory results.
- Step towards a more cohesive theory for nonlinear model reduction.
- Derivation of an upper bound on the size of the library, based on piecewise Taylor approximation with fixed number of terms, and design of an explicit partition of the parameter domain.