List of Contributors

Ronald DeVore
Texas A&M University
College Station, Texas 77843
rdevore@math.tamu.edu

Chapter 1

The Theoretical Foundation of Reduced Basis Methods

*Ronald DeVore**

1.1 ▪ Introduction

The main theme of this volume is the efficient solution of families of stochastic or parametric partial differential equations. This article focuses on the theoretical underpinnings of such methods. It shows how concepts from approximation theory, such as entropy or widths, can help to quantify, a priori, how well such methods can perform. This theory is then used to analyze one of the primary numerical vehicles, Reduced Basis Methods (RBMs), for parametric equations. A particular emphasis is placed on understanding the performance of greedy algorithms for selecting basis in RBMs.

Reduced basis methods have met with much computational success that is amply described in other contributions of this volume. The present article sits at the other end of the spectrum since it is exclusively devoted to the theoretical aspects of this subject. The development of a theory for reduced bases is of great interest since it addresses one of the most challenging problems in modern numerical computation, namely the computational recovery of high dimensional functions. The theory we present here is far from complete and, indeed, one of the goals of the present exposition is to organize our thinking and illuminate some obvious questions whose solution may advance both the theoretical and the computational aspects of reduction methods.

This article will exclusively deal with linear elliptic problems. This restriction was imposed because, quite frankly, not enough is known theoretically in other settings to warrant much discussion. However, let us be clear that theoretical developments for other problems will be extremely interesting and could help advance other application domains. While this article will only treat parametric problems, the results put forward have relevance for stochastic problems via chaos expansions.

What is written here is a very personalized view and understanding of this subject. The form of this article has been strongly influenced by discussions with many individuals. I mention them here because any of them would justifiably be co-authors of this presentation.

*This research was supported by the Office of Naval Research Contracts ONR N00014-11-1-0712, ONR N00014-12-1-0561, ONR N00014-15-1-2181 and the NSF Grant DMS 1222715.

My first exposure to reduced bases occured many years ago when visiting Albert Cohen at Paris VI. I had the fortune to be there when Yvon Maday and his collaborators were proving their first results on greedy algorithms and magic points. It was clear this subject had a large intersection with approximation theory and yet seemed to be completely missed by the approximation community. Albert and I began reflecting on reduced modelling and naturally involved our collaborators Wolfgang Dahmen, Peter Binev, Guergana Petrova, and Przemek Wojtaszcsyk. I organized a small seminar on this subject at TAMU which includes (in addition to Guergana) Andrea Bonito, Bojan Popov, and Gerrit Welper. I thank all of these people for helping my understanding of the subject.

Subsequent to the writing of this Chapter, the survey article [7] was written and is already in print. Some topics considered in this chapter appear in expanded form in [7]. So if the reader finds the current exposition too terse on a certain topic, the chances are, it is dealt with in more detail in [7].

1.2 ▪ Elliptic PDEs

The focal point of this article is the study of numerical algorithms for solving a family of elliptic equations. Each of these elliptic equations is of the form

$$-\nabla\cdot(a\nabla u)=f \quad \text{in} \quad D, \qquad u|_{\partial D}=0, \tag{1.1}$$

where $D\subset\mathbb{R}^d$ is a bounded Lipschitz domain, and the right side f is in $H^{-1}(D)$ [2]. Here, $a=a(x)$ is a scalar function which is assumed to be in $L_\infty(D)$ and satisfies the ellipticity assumption: there exist $0<r<R$ such that

$$r\le a(x)\le R, \quad x\in D. \tag{1.2}$$

We could just as well consider the case where a is replaced by a positive definite matrix function $A(x)$ with a similar theory and results, only at the expense of more cumbersome notation. In this section, we begin by recalling what is known about the solution to (1.1) when a and f are fixed. The later sections of this paper will then turn to the question of efficiently solving a family of such problems.

There is a rich theory for existence and uniqueness for the equation (1.1) which we briefly recall. A much expanded discussion of this topic can be found in [7]. Central to this theory is the Sobolev space $H^1_0(D,a)$ (called the energy space) which is a Hilbert space equipped with the energy norm

$$\|v\|_{H^1_0(D,a)}:=\||a|\nabla v|\|_{L^2(D)}. \tag{1.3}$$

That this is a norm follows from a theorem of Poincaré which says that

$$\|v\|_{L_2(D)}\le C_D\|v\|_{H^1_0(D,a)}, \tag{1.4}$$

for every Lipschitz domain D and in particular for every polyhedral domain D.

If $a,\tilde{a}$ both satisfy the ellipticity assumption, then the norms for $H^1_0(a)$ and $H^1_0(\tilde{a})$ are equivalent. If we take $a=1$, we obtain the classical space $H^1_0(D,1)$, which in going further is simply denoted by $H^1_0=H^1_0(D)$. The dual of $H^1_0(D)$ consists of all linear functionals

[2] We use standard notation for Sobolev spaces throughout this chapter. The space $W^s(L_p(D))$ is the Sobolev space with smoothness index s in $L_p(D)$. For the special case $p=2$, this space is typically denoted by H^s in the numerical and PDE communities.

defined on this space and it is usually denoted by $H^{-1}(D)$ and its norm is defined by duality. Namely, if $\lambda \in H^{-1}(D)$, then

$$||\lambda||_{H^{-1}(D)} := \sup_{||v||_{H_0^1(D)} \leq 1} |\langle \lambda, v \rangle| \tag{1.5}$$

The solution u_a of (1.1) is defined in weak form as a function $u \in H_0^1(D)$ that satisfies

$$\int_D a(x) \nabla u_a(x) \cdot \nabla v(x) dx = \int_D f(x) v(x) dx. \text{ for all } v \in H_0^1(D). \tag{1.6}$$

This formulation shows that the Lax-Milgram theory applies. In particular, the ellipticity assumption is a sufficient condition for the existence and uniqueness of the solution u_a. Under this assumption, the solution satisfies the estimate

$$||u_a||_{H_0^1(D)} \leq C_0 \frac{||f||_{H^{-1}(D)}}{r}. \tag{1.7}$$

The same theory applies even if a is complex valued. Now the lower ellipticity condition replaces a by $Re(a)$ in (1.2) and the upper condition is that $|a|$ is uniformly bounded.

There is also a general principal of perturbation for elliptic equations which shows to some extent the smooth dependence of the solution on the diffusion coefficient a. If $a, \tilde{a}$ are two such coefficients with the same ellipticity constants r, R, then the solutions u and $\tilde{u}$ with identical right side f will satisfy

$$||u_a - u_{\tilde{a}}||_{H_0^1(D)} \leq C_0 \frac{||a - \tilde{a}||_{L_\infty(D)}}{r}. \tag{1.8}$$

The bound (1.8) shows that the mapping $a \to u_a$ is Lipschitz continuous. Actually, this mapping is in a certain sense analytic as will be explained in §1.3.1. This smooth dependence is at the heart of reduced modelling and so it will be of large concern to us as we proceed.

1.2.1 ▪ Other Perturbation Results

In some applications, the coefficients a, while bounded, are not continuous. In such applications, they may have discontinuities along curves or higher dimensional manifolds in $\mathbb{R}^d$. This makes (1.8), more or less, useless since it requires exact matching of the discontinuities of a and $\tilde{a}$. A related issue is that in numerical methods, the diffusion coefficient a is approximated by an $\tilde{a}$ and one will not have that $||a - \tilde{a}||_{L_\infty}$ is small since the discontinuity cannot be matched exactly. Thus, we need other perturbation results which are more amenable to such applications. Results of this type were given in [4] in which L_∞ perturbation is replaced by L_q perturbation for certain q with $q < \infty$, in the form of the following result.

For any $p \geq 2$, the functions u_a and $u_{\tilde{a}}$ satisfy

$$||u_a - u_{\tilde{a}}||_{H_0^1(D)} \leq r^{-1} ||\nabla u_a||_{L_p(D)} ||a - \tilde{a}||_{L_q(D)}, \quad q := \frac{2p}{p-2} \in [2, \infty] \tag{1.9}$$

provided $\nabla u_a \in L_p(D)$. Notice that the case $p = 2$ is (1.8). In order for (1.9) to be relevant for discontinuous $a, \tilde{a}$, we need that ∇u_a is in L_p for some $p > 2$. It is known that for every Lipschitz domain D, there is $P > 2$ such that for $2 \leq p \leq P$, one has

COND-p: *For each $f \in W^{-1}(L_p(D))$, the solution $u = u_1$ to (1.1) with $a = 1$ and right side f satisfies*

$$|u|_{W^1(L_p(D))} := ||\nabla u||_{L_p(D)} \le C_p ||f||_{W^{-1}(L_p(D))}, \tag{1.10}$$

with the constant C_p independent of f.

The assumption $f \in W^{-1}(L_p(\Omega))$ is a rather mild assumption on the right side f and leads to an L_q perturbation with $q := \frac{2p}{p-2}$.

In this special case $a = 1$ (the case of Laplace's equation), the validity of **COND-p** is a well studied problem in Harmonic Analysis (see for example Jerison and Kenig [18]). In fact, in this setting, one can take $P > 4$ when $d = 2$ and $P > 3$ when $d = 3$. The case of general a in (1.9) is proven by a perturbation argument (see [4] for details). It is also known that if D is convex then we can take $P = \infty$.

One can extend the above results from $a = 1$ to general a by using a perturbation result (see Proposition 1 in [4]).

Perturbation Property: If the diffusion coefficients $a, \tilde{a}$ satisfy the strong ellipticity condition for some r, R, then there is a P^* depending on this r, R and on the domain D such that whenever $p \in [2, P^*]$, and $f \in W^{-1}(L_p(D))$, then

$$||u_a - u_{\tilde{a}}||_{H_0^1(D)} \le C_1 ||f||_{W^{-1}(L_p(D))} ||a - \tilde{a}||_{L_q(D)}, \tag{1.11}$$

where $q := 2p/(p-1)$.

The strongest perturbation result occurs when we can take $P^* = \infty$. In this case, the L_2 norm appears on the right side of (1.11).

Let us emphasize that assumptions on f other than $f \in W^{-1}(L_p(D))$ may lead to $\nabla u \in L_p$ for a wider range of p. This would, in turn, give the perturbation for a wider range of q.

1.3 ▪ Parametric Elliptic Equations

We turn now to the principle topic of this article, namely, the solution of a family of elliptic equations. We are not interested in solving (1.1) for just one diffusion coefficient a but rather a family $\mathcal{A}$ of such coefficients. We always assume that the family $\mathcal{A}$ satisfies

Uniform Ellipticity Assumption: There exists r, R such that for each $a \in \mathcal{A}$, we have

$$r \le a(x) \le R, \quad x \in D. \tag{1.12}$$

This family is assumed to be a compact subset of either $L_\infty(D)$ or of an appropriate $L_q(D)$ space for which the following property holds:

L_q(D) stability of $\mathcal{A}$: We say this property holds for the given f, if there is a constant C_0 such that for all $a, \tilde{a} \in L_q(D)$ we have

$$||u_a - u_{\tilde{a}}||_{H_0^1(D))} \le C_0 ||a - \tilde{a}||_{L_q(D)}. \tag{1.13}$$

Of course this property always holds for $q = \infty$. We have already discussed in the previous section, the fact that this will also hold for a range of $Q \le q \le \infty$, with $2 \le Q < \infty$, under very mild restrictions on f. Notice that the most favorable range is when $Q = 2$.

Our goal is to build a black box solver such that when presented with any $a \in \mathcal{A}$, the solver will provide in a very fast manner an online computation of $u_a = u_{a,f}, a \in \mathcal{A}$. To begin the discussion, let us recall that there are already in existence Adaptive Finite Element solvers which when given $a \in \mathcal{A}$, will approximate in a rather efficient way the solution u_a. What governs the accuracy of these adaptive solvers is the smoothness of u_a which in turn is determined by properties of the physical domain D and the regularity of the right side f. Indeed, the performance of such adaptive methods is governed by the regularity of the solution u_a in a certain scale of Besov spaces corresponding to non linear approximation [2, 3] and this Besov regularity can be derived from the smoothness of f (see [11]). We do not wish to get into details here but only mention that the typical performance of this approach is to obtain convergence of order $O(n^{-\beta})$ for n computations, where β is typically small. For example, if $f \in L_2(D)$ and the domain is Lipschitz, then $\beta \leq 1$ if $d = 3$.

The motivation behind reduced basis methods is the smoothness of the u_a with varying a. We can view the set

$$\mathcal{U} := \mathcal{U}_{\mathcal{A}} := \mathcal{U}_{\mathcal{A},f} := \{u_a : a \in \mathcal{A}\}, \tag{1.14}$$

as a manifold in $H_0^1(D)$ and in light of our earlier discussion of perturbation for elliptic equations, this manifold inherits a certain smoothness from that of $\mathcal{A}$. If this smoothness is high enough, then it is reasonable to expect we can build solvers that perform better than the adaptive pde solvers since the latter never take advantage of the smoothness of this manifold. Our main interest is to quantify when this is indeed true.

We give two examples of sets $\mathcal{A}$ of diffusion coefficients, which will guide our discussion.

1.3.1 ▪ Affine Dependence

In this setting, we are given a family of functions $\psi_j(x)$, $j = 1, 2, \ldots$, defined on the physical domain D. We let U be the unit cube in $\ell_\infty := \ell_\infty(\mathbb{N})$. Hence, $y \in U$ means that $y = (y_1, y_2, \ldots)$ with $|y_j| \leq 1$. For any such $y \in U$, we define

$$a(x, y) = \overline{a}(x) + \sum_{j \geq 1} y_j \psi_j(x), \tag{1.15}$$

and take $\mathcal{A} = \{a(x, y) : y \in U\}$ as our family of diffusion coefficients. Of course, we shall also need additional assumptions to guarantee that the series in (1.15) converges. A typical assumption is that the sequence $(\|\psi_j\|_{L_\infty(D)})$ is in ℓ_p for some $p \leq 1$. We assume in going further that the indices have been rearranged so that this sequence $(\|\psi_j\|_{L_\infty(D)})$ is monotonically decreasing.

One may wonder why we consider an infinite number of parameters y_j in (1.15). The answer is twofold. First of all, a standard way of treating stochastic equations is to consider chaos expansions which can be converted to parametric equations but the number of parameters will be infinite. A second reason is that even when treating a parametric problem with a finite number of parameters m, we want to avoid convergence estimates that blowup with m. By treating the case of an infinite number of parameters one can sometimes obtain constants independent of m.

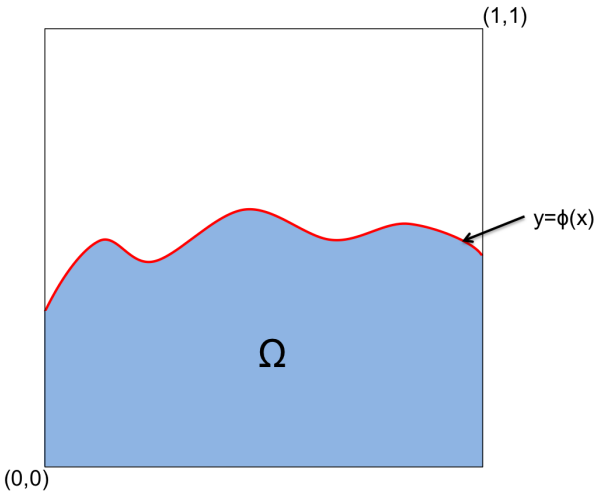

Figure 1.1. *The region marked* Ω *corresponds to* D_-.

1.3.2 ▪ A Geometrical Setting

Let $D = [0,1]^2$ for simplicity and let $\phi(x)$, $x \in [0,1]$ be a $\mathrm{Lip}_M 1$ function[3] taking values in $[0,1]$. Then the graph of φ separates D into two domains $D_\pm$ corresponding to the portion D_- of D below the graph and the portion D_+ above the graph (see Figure 1.3.2). We consider diffusion coefficients

$$a_\phi(x) := \chi_{D_-} + 2\chi_{D_+}. \tag{1.16}$$

These coefficients have a jump across the curve. The purpose of this toy example will be to see how to handle discontinuities in a.

These two examples sit at opposite extremes. The affine model is the most favorable for RBMs since, as we shall see, the manifold $\mathscr{M}$ of solutions is analytic. The geometric model, on the other hand, gives a manifold that is not very smooth because of the discontinuities in a. So, it provides a real challenge for RBMs. This model may however prove useful in certain application domains such as shape optimization.

1.4 ▪ Evaluating Numerical Methods

Numerical methods for solving partial differential equations are based on some form of approximation. Understanding the core results of approximation theory, not only suggests possible numerical procedures, but also determines the optimal performance a numerical method can have. This optimal performance is described by the concepts of entropy and widths which is the subject of this section.

1.4.1 ▪ Linear Methods

Methods of approximation are broadly divided into two classes: linear and non linear methods. In linear approximation, the approximation process takes place from a sequence

[3]The space $\mathrm{Lip}_M 1$ consists of all continuous functions satisfying $|f(x) - f(x')| \leq |x - x'|$ for all points $x \in [0,1]$.

of finite dimensional **linear** spaces X_n, $n = 1, 2, \ldots$, with increasing dimensions. By using the space $X_0 := \{0\}$ and, if necessary, repeating the spaces in this sequence, we can assume $\dim(X_n) \leq n$. Increasing n results in improved accuracy of the best approximations from X_n.

In our case of parametric equations, we want to choose the linear space X_n so that it approximates well all of the elements $u_a \in \mathcal{U}_{\mathcal{A}}$ in the norm of $H_0^1(D)$ (or perhaps $H_0^1(D,a)$). Once such a space X_n is found then we can build a numerical method for solving problems. For example, we could use the Galerkin solver corresponding to X_n. This would require the on-line assembly of the stiffness matrix and the solution of the corresponding matrix problem.

For each $a \in \mathcal{A}$, we have the error

$$E(u_a, X_n) := E(u_a, X_n)_{H_0^1(D)} := \inf_{g \in X_n} \|u_a - g\|_{H_0^1(D)}. \tag{1.17}$$

Notice that because of the **UEA**, the norm $\|\cdot\|_{H_0^1(D)}$ is equivalent to $\|\cdot\|_{H_0^1(D,a)}$ with constants depending only on r and R. Hence, $E(u_a, X_n)$ also can be used to measure the approximation error in $H_0^1(D,a)$. The effectiveness of the space X_n for our parametric problem is given by

$$E(\mathcal{U}_{\mathcal{A}}, X_n) := \sup_{a \in \mathcal{A}} E(u_a, X_n), \tag{1.18}$$

which is the *error on the class* $\mathcal{U}_{\mathcal{A}}$.

The best choice of a linear space X_n is the one which gives the smallest class error. This smallest error for the compact set $\mathcal{U}_{\mathcal{A}}$ is called the *Kolmogorov n-width* of $\mathcal{U}_{\mathcal{A}}$. We can define this width for any compact set K in any Banach space X by

$$d_n(K)_X := \inf_{\dim(Y)=n} \sup_{u \in K} \inf_{g \in Y} \|u - g\|_X, \quad n = 0, 1, \ldots. \tag{1.19}$$

So the n-width $d_n(\mathcal{U}_{\mathcal{A}})_{H_0^1(D)}$ gives the optimal performance we can achieve when using linear methods to solve the family of parametric equations (1.1) for all $a \in \mathcal{A}$. Determining d_n and finding an optimal or near optimal subspace is a difficult problem which we will return to.

1.4.2 ▪ Non Linear Methods

It is now well understood that non linear methods of approximation and numerical methods derived from them often produce superior performance when compared to linear methods. Classical non linear methods include approximation by rational functions, free knot splines, n-term approximation, and adaptive partitioning. The basic idea in non linear approximation is to replace the linear space X_n by a non linear space Σ_n depending on n parameters. Loosely speaking, one can view Σ_n as an n-dimensional manifold.

We discuss, in some detail, the case of n-term approximation in a Banach space X since it has promise in designing numerical algorithms for parametric equations. The starting point for this form of approximation is a collection $\mathcal{D} \subset X$ of functions which is called a *dictionary*. Given a dictionary, we define the set

$$\Sigma_n(\mathcal{D}) := \{\sum_{g \in \Lambda} c_g g : \ \Lambda \subset \mathcal{D}, \ \#(\Lambda) \leq n\}, \tag{1.20}$$

of all n-term linear combinations of elements from $\mathcal{D}$. The elements in Σ_n are said to be *sparse* of order n. Notice that the space Σ_n is not a linear space. If we add two elements

from Σ_n, we will generally need $2n$ terms to represent the sum. An important case is when the dictionary $\mathscr{D} = \{\varphi_j\}_{j=1}^{\infty}$ where the functions φ_j, $j = 1, 2, \ldots$, form a basis for X. In this case, any function in Σ_n is described by $2n$ parameters, namely the n indicies $j \in \Lambda$ and the n coefficients c_{φ_j}.

Suppose now that $X = \mathscr{H}$ is a Hilbert space and $\mathscr{D} = \{\varphi_j\}_{j=1}^{\infty}$ is an orthonormal basis for $\mathscr{H}$. It is very easy to describe the best approximation to a given function $v \in \mathscr{H}$ from Σ_n and the resulting error of approximation. We expand v in its unique series representation

$$v = \sum_{j=1}^{\infty} c_j(v)\varphi_j. \tag{1.21}$$

Given any sequence $(a_j)_{j\geq 1}$ of real numbers which tend to zero as $j \to \infty$, we denote by $(a_k^*)_{k\geq 1}$ the decreasing rearrangement of the $|a_j|$. Thus, a_k^* is the k-th largest of these numbers. For each k, we can find a λ_k such that $a_k^* = |a_{\lambda_k}|$ but the mapping $k \mapsto \lambda_k$ is not unique because of possible ties in the size of the entries. The following discussion is impervious to such differences. If we apply rearrangements to the coordinates $\{c_j(v)\}_{j\geq 1}$ and denote by $\Lambda_n := \Lambda_n(v) := \{j_1, \ldots, j_n\}$ the indices of a set of n-largest coefficients, then a best approximation to $u \in \mathscr{H}$ from Σ_n is given by the function

$$G_n v := \sum_{k=1}^{n} c_{j_k}(v)\varphi_{j_k} = \sum_{j\in\Lambda_n} c_j(v)\varphi_j, \tag{1.22}$$

and the resulting error of approximation is

$$\sigma_n(v)_{\mathscr{H}}^2 := \|v - G_n v\|_{\mathscr{H}}^2 = \sum_{j\notin\Lambda_n} |c_j(v)|^2 = \sum_{k>n} (c_k^*(v))^2. \tag{1.23}$$

In particular, $\sigma_0(v) = \|v\|_{\mathscr{H}}$. While best approximation from Σ_n is not unique, the approximation error $\sigma_n(v)_{\mathscr{H}}$ is uniquely defined. Also, note that $\sigma_n(v)_{\mathscr{H}} = \sigma_n(\tilde{v})_{\mathscr{H}}$ if v and $\tilde{v}$ have the same coefficients up to a permutation of the indices.

We can use (1.23) to characterize the functions $v \in \mathscr{H}$ which can be approximated with order $O(n^{-r})$, $r > 0$, in terms of the coefficients $c_j(v)$. Let us denote by $\mathscr{A}^r = \mathscr{A}_r((\Sigma_n)_{n=1}^{\infty}, \mathscr{H})$ this set of functions ($\mathscr{A}^r$ is called an *approximation class*) and equip it with the norm

$$\|v\|_{\mathscr{A}^r} := \sup_{n\geq 0} (n+1)^r \sigma_n(v)_{\mathscr{H}}. \tag{1.24}$$

Given an $r > 0$, we define p by the formula

$$\frac{1}{p} = r + \frac{1}{2}. \tag{1.25}$$

Notice that $p < 2$. The space $w\ell_p$ (*weak* ℓ_p) is defined as the set of all $\mathbf{a} = (a_j)_{j\geq 1}$ whose decreasing rearrangement $(a_k^*)_{k\geq 1}$satisfies

$$k^{1/p} a_k^* \leq M, \quad k \geq 1, \tag{1.26}$$

and the smallest $M = M(\mathbf{a})$ for which (1.26) is valid is the quasi-norm $\|\mathbf{a}\|_{w\ell_p}$ of $\mathbf{a}$ in this space. Notice that $w\ell_p$ contains ℓ_p and is slightly larger since it contains sequences whose rearrangement behaves like $k^{-1/p}$ which barely miss being in ℓ_p.

We claim that, with $\mathbf{c} := \mathbf{c}(v) := \{c_j(v)\}_{j\geq 1}$,

$$\mathcal{A}^r := \mathcal{A}^r(\mathcal{H}, (\Sigma_n)_{n\geq 1}) = \{v \,:\, \mathbf{c}(v) \in w\ell_p\}, \tag{1.27}$$

and $\|\mathbf{c}(u)\|_{w\ell_p}$ is equivalent to $\|u\|_{\mathcal{A}^r}$. Indeed, if $\mathbf{c}(v) \in w\ell_p$, then for any $n \geq 1$, we have

$$\sigma_n(v) = \sum_{k>n} (c_k^*(v))^2 \leq \|\mathbf{c}(v)\|_{w\ell_p}^2 \sum_{k>n} k^{-2r-1} \leq \frac{1}{2r} \|\mathbf{c}(v)\|_{w\ell_p}^2 n^{-2r}. \tag{1.28}$$

In addition,

$$\|v\|_{\mathcal{H}}^2 = \|\mathbf{c}(v)\|_{\ell^2}^2 \leq \|\mathbf{c}(v)\|_{w\ell^p}^2 \sum_{k\geq 1} k^{-2r-1} \leq (1 + \frac{1}{2r}) \|\mathbf{c}(v)\|_{w\ell_p}^2. \tag{1.29}$$

This shows that $\|v\|_{\mathcal{A}^r} \leq (1 + \frac{1}{2r})^{1/2} \|\mathbf{c}(v)\|_{w\ell_p}$.

To reverse this inequality, we note that for any $k \geq 1$, the monotonicity of $\mathbf{c}^*(v)$ gives

$$2^j (c_{2^{j+1}}^*(v))^2 \leq \sum_{k=2^j+1}^{2^{j+1}} (c_k(v)^*)^2 \leq \sigma_{2^j}(u)^2 \leq |v|_{\mathcal{A}^r}^2 2^{-2jr}. \tag{1.30}$$

For any n, we choose j so that $2^j \leq n < 2^{j+1}$. If $j > 0$, we obtain from the monotonicity of $\mathbf{c}^*(v)$ that

$$c_n^*(v) \leq c_{2^j}^*(v) \leq 2^{r+1/2} |v|_{\mathcal{A}^r} 2^{-(r+1/2)j} = 2^{1/p} |v|_{\mathcal{A}^r} 2^{-j/p} \leq 2^{2/p} |v|_{\mathcal{A}^r} n^{-1/p}. \tag{1.31}$$

On the other hand, we clearly have

$$c_1^*(v) \leq \|v\|_{\mathcal{H}} \leq \|v\|_{\mathcal{A}^r}. \tag{1.32}$$

This gives $\|\mathbf{c}(v)\|_{w\ell_p} \leq 2^{2/p} \|v\|_{\mathcal{A}^r}$ and completes the proof of the equivalence.

In numerical settings, one cannot implement n-term approximation in the form we have just presented since it would require the computation of all coefficients of v and a rearrangement of them. What is done in practice is one chooses a value N dependent on n and selects the best n-term approximation from the dictionary $\mathcal{D}_N := \{\varphi_j\}_{j=1}^N$. A typical choice of N is $N = n^A$ where A is a fixed integer.

There is another useful view of n-term approximation in this last setting. We can form from $\{\varphi_1, \ldots, \varphi_N\}$ all subsets $\{\varphi_i\}_{i\in\Lambda}$, $\#(\Lambda) = n$, which are linearly independent. Then each $X_\Lambda := \operatorname{span}\{x_i \,:\, i \in \Lambda\}$ is a linear space of dimension n. There are at most $\binom{N}{n} \leq [eN^{B-1}]^n$ such subspaces. Then, n-term approximation can be viewed as taking one of these linear spaces and using it to approximate v. The space chosen can depend on v.

1.4.3 ▪ Non Linear Widths

There have been several definitions of non linear widths that have been proposed to measure optimal performance of non linear methods. We mention the two that seem most relevant for the analysis of reduced basis methods. The first of these is the manifold width [12] which matches well numerical algorithms based on non linear approximation. Let X be a Banach space and K one of its compact subsets. To define this width, we consider two continuous functions. The first function b maps each element $x \in K$ into $\mathbb{R}^n$ and therefore picks out the parameters to be used in approximating x. The second function

M maps $\mathbb{R}^n$ into the set $\mathcal{M}$ (which we view as an n-dimensional manifold although we do not assume anything about the smoothness of the image $\mathcal{M}$). The manifold width of the compact set K is then defined by

$$\delta_n(K)_X := \inf_{M,b} \sup_{x\in K} \|x - M(b(x))\|_X. \tag{1.33}$$

For typical compact sets K of functions, the manifold widths behave like the entropy numbers defined below. For example, if K is the unit ball of any Besov or Sobolev space of smoothness s which compactly embeds into $L_p(\Omega)$ with $\Omega \subset \mathbb{R}^d$ a Lipschitz domain, then (see [13])

$$C_0 n^{-s/d} \le \delta_n(K)_{L_p(\Omega)} \le C_1 n^{-s/d}, \tag{1.34}$$

with C_0, C_1 independent of n. We see in (1.34) the curse of dimensionality in the appearance of d in the exponent of n. In order to obtain just moderate rates of convergence with $n \to \infty$ we need s to be comparable with d.

A second width, introduced by V. Temlyakov [26], fits the definition of n-term approximation. It considers any collection (called a *library*) $\mathcal{X} := \{X_j\}_{j=1}^N$, of n dimensional subspaces of X. The approximation error defined by

$$E(v,\mathcal{X})_X := \inf_{1\le j\le N} \operatorname{dist}(v, X_j)_X, \tag{1.35}$$

is another type of n-term approximation. This leads us to define *the library widths*

$$d^L_{n,N}(K)_X := \inf_{\mathcal{X}} \sup_{v\in K} E(v,\mathcal{X})_X, \tag{1.36}$$

with the infimum taken over all such collections $\mathcal{X}$. This is another type of non linear width. Typically, we would like to eliminate N from the above definition. Similar to the restrictions on dictionaries, one usually assumes that the the number N of bases is of the form $N = n^A$ for some fixed integer A. With this assumption $d_{n,N}$ now only depends on n.

Let us note that the definition of library widths includes approximation from a finite dictionary. Namely, if $\mathcal{D}$ is a dictionary with $\#(\mathcal{D}) = m$, then there are $\binom{m}{n}$ subspaces X_j of dimension $\le n$ that can be formed using n elements from the dictionary as a spanning set. Thus, with $N = \binom{m}{n}$, the library widths allows in the competition all n-term approximations from $\mathcal{D}$. However, the library width allows more general sequences of subspaces X_j in its definition since they do not have to be organized as coming from a fixed dictionary. When the subspaces X_j all come from a fixed dictionary of size m as discribed above then the corresponding width

$$d_{n,N}(K)_X := \inf_{\mathcal{D}} \sup_{v\in K} \sigma_n(K,\mathcal{D})_X, \tag{1.37}$$

is called the *dictionary width* of K.

1.4.4 ▪ Entropy Numbers

Another useful concept in our analysis of reduced basis methods will be the entropy numbers of a compact set $K \subset X$ where again X is a Banach space. If $\epsilon > 0$, we consider all possible coverings of $K \subset \bigcup_{i=1}^m B(x_i,\epsilon)$ using balls $B(x_i,\epsilon)$ of radius ϵ with centers $x_i \in X$.

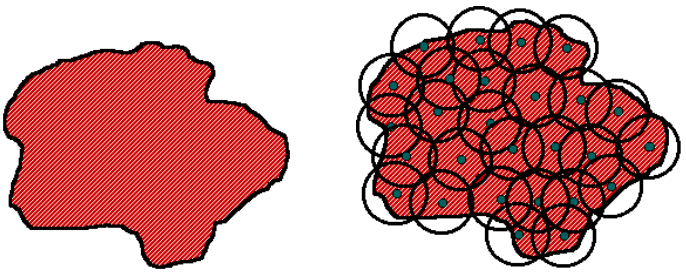

Figure 1.2. *A compact set K and its ϵ cover.*

The smallest number $m = N_\epsilon(K)_X$, for which such a covering exists, is called the covering number of K. The Kolmogorov entropy of K is then defined as

$$H_\epsilon(K)_X := \log_2(N_\epsilon(K))_X. \tag{1.38}$$

The Kolmogorov entropy measures the size or massivity of K. It has another important property of determining optimal encoding of the elements of K. Namely, if $x \in K$ then we can assign to x the binary bits of an index i for which $x \in B(x_i, \epsilon)$. Each x is then encoded to accuracy ϵ with $\leq \lceil H_\epsilon(K)_X \rceil$ bits and no other encoder can do better for K.

It is frequently more convenient to consider the entropy numbers

$$\epsilon_n(K)_X := \inf\{\epsilon : H_\epsilon(K)_X \leq n\}. \tag{1.39}$$

Typically, $\epsilon_n(K)_X$ decay like n^{-r} for standard compact sets. Not only does $\epsilon_n(K)_X$ tell us the minimal distortion we can achieve with n bit encoding, it also says that any numerical algorithm which computes an approximation to each of the elements of K to accuracy $\epsilon_n(K)_X$ will require at least n operations.

An important issue in optimal performance of the standard numerical algorithms for pdes are the entropy numbers of the classical smoothness spaces. If K is the unit ball $U(W^s(L_p(\Omega))$ of a Sobolev space, or a unit ball $U(B^s_q(L_p(\Omega))$ of a Besov space, then for any Lebesgue space $X = L_\mu(\Omega)$,

$$\epsilon_n(K)_X \geq C n^{-s/d}, \quad n = 0, 1, \ldots. \tag{1.40}$$

This result manifests the massivity of these compact sets as the dimension d increases and exhibits fully the curse of dimensionality.

1.4.5 ▪ Comparison of Widths

Concepts like n-widths are used to give bounds for the best possible performance of numerical algorithms. There are general comparisons between the different widths that are useful in making such evaluations. Let us mention those that will prove most useful for us. For any compact set K in a Banach space X, we always have (see [12]),

$$\delta_n(K)_X \leq d_n(K)_X, \quad n \geq 1. \tag{1.41}$$

It is also known that whenever $d_n(K)_X \leq C n^{-r}$, $n \geq 1$ with $r > 0$, then there is a constant C' such that

$$\epsilon_n(K)_X \leq C' n^{-r}, \quad n \geq 1. \tag{1.42}$$

This follows from Carl's inequality (see [20]).

In general, it is not possible to compare the non linear manifold widths to entropy. However, for the widths of (1.36), we have the same result as (1.42) (see [26]).

1.5 ▪ Comparing Widths and Entropies of $\mathcal{U}_{\mathcal{A}}$ with those of $\mathcal{A}$

Let us return our discussion to numerical methods for $\mathcal{U}_{\mathcal{A}}$. In trying to understand how well numerical methods can perform in resolving $\mathcal{U}_{\mathcal{A}}$, we would like to know the entropies and widths of $\mathcal{U}_{\mathcal{A}}$. Since we only know this set through the parameter set $\mathcal{A}$, the first question we would like to answer is whether we can bound the widths of $\mathcal{U}_{\mathcal{A}}$ by those of $\mathcal{A}$. The widths and entropies of $\mathcal{A}$ are usually more transparent and so such bounds give a good indication of how well reduced basis methods might perform. In this section, we shall discuss what is known about such comparisons.

1.5.1 ▪ Comparing Entropies

One can utilize the perturbation results (1.9) and (1.13) to give a comparison of the entropies of the two classes $\mathcal{A}$ and $\mathcal{U} = \mathcal{U}_{\mathcal{A}}$. We place ourselves in the following situation. We assume that $2 \le q \le \infty$ is a value for which the $L_q(D)$ stability is known to hold for $\mathcal{A}$. It follows that any ϵ cover of $\mathcal{A}$ in the $L_q(D)$ norm given by balls $B(a_i, \epsilon)$ will induce a $C_0\epsilon$ cover of $\mathcal{U}$ by the balls $B(u_{a_i}, C-0\epsilon)$ in the $H_0^1(D)$ topology. Therefore, we have

$$\epsilon_n(\mathcal{U}_{\mathcal{A}})_{H_0^1(D)} \le C_0\epsilon_n(\mathcal{A})_{L_q(D)}, \quad n \ge 1. \tag{1.43}$$

It now becomes an interesting question of whether the entropy numbers of $\mathcal{U}_{\mathcal{A}}$ could actually be much better than those of $\mathcal{A}$. We will now show why, in general, we cannot expect this to be the case. We consider the case $q = 2$ and $d = 1$. Our goal is to find an f and many classes $\mathcal{A}$ for which we can reverse (1.43) and thereby see that the entropy of $\mathcal{U}_{\mathcal{A}}$ is not noticeably better than that of $\mathcal{A}$, at least in general. We consider the one dimensional case $D = [0,1]$, where the PDE is simply

$$-[au']' = f, \quad u(0) = u(1) = 0. \tag{1.44}$$

We will specify a right side f as we proceed. Let $F(x) := -\int_0^x f(s)\,ds$. Then given a,

$$u_a = \int_0^x a^{-1}(s)[F(s) - c_a]\,ds, \quad c_a := \int_0^1 a^{-1}(s)F(s)\,ds. \tag{1.45}$$

Since we are allowed to choose f, we are allowed to choose F as follows. We take F to be a smooth function which is odd with respect to $x = 1/2$ and satisfies $F(0) = F(1/2) = 0$; F is increasing on $[0, 1/6]$; $F(x) = +1$ on $J := [1/6, 1/3]$ and F is decreasing on $[1/3, 1/2]$. We fix this F and assume the following about $\mathcal{A}$.

Assumptions on $\mathcal{A}$: Each $a \in \mathcal{A}$ is even with respect to $1/2$ and the class $\mathcal{A}_0$ of all $a \in \mathcal{A}$ restricted to J satisfies $\epsilon_n(\mathcal{A}_0)_{L_2(J)} \ge c_0\epsilon_n(\mathcal{A})_{L_2(D)}$ for an absolute constant c_0.

Returning to (1.45), we see that u_a' is $\ge 1/a$ on the interval $J := [1/6, 1/3]$. On J we can now write for any two such $a, \tilde{a}$

$$a - \tilde{a} = F/u_a' - F/u_{\tilde{a}}' = \frac{F}{u_a' u_{\tilde{a}}'}[u_{\tilde{a}}' - u_a']. \tag{1.46}$$

This gives the bound on J,

$$\|a-\tilde{a}\|_{L_2(J)} \le C\|u_a - u_{\tilde{a}}\|_{H_0^1(D)}, \tag{1.47}$$

and therefore,

$$\epsilon_n(\mathcal{A})_{L_2(J)} \le C\epsilon_n(K)_{H_0^1(D)}, \quad n \ge 1, \tag{1.48}$$

and therefore we have reversed (1.43), in the case $q = 2$.

1.5.2 ▪ Comparing Kolmogorov Widths

There is a general method (see [6]) for bounding the Kolmogorov n-width of $\mathcal{U}_{\mathcal{A}}$ in terms of the corresponding width of $\mathcal{A}$. Among other results, it gives the following theorem.

Theorem 1.1. *If $\mathcal{A}$ is any set contained in $L_\infty(D)$ whose Kolmgorov widths satisfies*

$$d_n(\mathcal{A})_{L_\infty(D)} \le C_0 n^{-\alpha}, \quad n \ge 1, \tag{1.49}$$

for some $\alpha > 1$ and constant C_0, then for each $\beta < \alpha - 1$,

$$d_n(\mathcal{U}_{\mathcal{A}})_{H_0^1(D)} \le C_1 n^{-\beta}, \quad n \ge 1, \tag{1.50}$$

for a constant C_1 depending only on C_0 and α.

We can give some ideas behind its proof which rests on the following result from [9]:

Theorem 1.2. *If $\mathcal{A}$ is defined by the affine model (1.15) where the ψ_j, $j = 1, 2, \ldots$, satisfy $(\|\psi_j\|_{L_\infty})_{j=1}^\infty \in \ell_p$ with $p < 1$, then $d_n(U_{\mathcal{A}}) \le M n^{-1+1/p}$.*

We can now sketch of how one proves Theorem 1.1. From our assumption on $\mathcal{A}$, we can find, for each $k \ge 0$, linear spaces Z_k of dimension 2^k which satisfy

$$\operatorname{dist}(\mathcal{A}, Z_k)_{L_\infty(D)} \le C_0 2^{-k\alpha}, \quad k = 0, 1, 2, \ldots. \tag{1.51}$$

Hence, for each $a \in \mathcal{A}$ there are functions $g_k \in Z_k$ such that

$$\|a - g_k\|_{L_\infty(D)} \le C_0 2^{-k\alpha}, \quad k = 0, 1, \ldots, \tag{1.52}$$

It follows that

$$a = g_0 + \sum_{k\ge 1}[g_k - g_{k-1}] = \sum_{k=0}^{\infty} h_k, \quad \text{and } \|h_k\|_{L_\infty(D)} \le 2C_0 2^{-k\alpha},\ k = 0, 1, \ldots, \tag{1.53}$$

where $h_k := g_k - g_{k-1}$ and $g_{-1} := 0$. Let us note that each h_k is in the linear space $Y_k := Z_{k-1} + Z_k$ which has dimension $m_k \le 2^k + 2^{k-1}$.

We now use the following consequence of Auerbach's basis theorem. There is a basis $\psi_{1,k} \ldots, \psi_{m_k,k}$ for Y_k, normalized so that $\|\psi_{j,k}\|_{C(D)} = 1$ for all j and such that its dual basis also has norm one. So each h_k can be written $h_k = \sum_{j=1}^{m_k} c_{j,k}\psi_{j,k}$ and the coefficients satisfy

$$\max_{1\le j\le m_k} |c_{j,k}| \le \|h_k\|_{L_\infty(D)} \le 2C_0 2^{-k\alpha}, \quad k = 1, 2, \ldots. \tag{1.54}$$

This gives us the decomposition

$$a = \sum_{j=1}^{\infty} y_j \psi_j, \quad , \; j = 1, 2, \ldots, \quad |y_j| \leq 1, \tag{1.55}$$

where each ψ_j is a renormalization of one of the $\psi_{i,k}$ and satisfies

$$\|\psi_j\|_{C(D)} \leq M_0 C_0 j^{-\alpha}, \quad j = 1, 2, \ldots. \tag{1.56}$$

Notice that a function ψ_j is possibly repeated in the sum (1.55).

We now choose K so that

$$M_0 \sum_{k>K} \|\psi_j\|_{C(D)} \leq r/2. \tag{1.57}$$

This means that $\sum_{j<K} \|\psi_j\|_{C(D)} \geq r/2$. It can be shown (details not given) that we can find a finite number, say N, of a_i each of the form $a_i = \sum_{1 \leq j \leq K} b_j \psi_j$, such that for any $a \in \mathscr{A}$, there is an a_i for which

$$a = a_i + \sum_{j=1}^{\infty} y_j \psi_j, \tag{1.58}$$

where $|y_j| \leq 1$ and moreover, each of the $a_i(x) \geq r/4$ for $x \in D$.

In other words, $\mathscr{U}_{\mathscr{A}}$ is contained in a union of a finite number N of $\mathscr{U}_{\mathscr{A}_i}$ where each set $\mathscr{A}_i$ is of the form (1.58). Because of (1.56), $(\|\psi_j\|_{L_\infty(D)} \in \ell_p$ for each $p > 1/\alpha$. This allows us to apply Theorem 1.2 and derive Theorem 1.1.

1.5.3 ▪ Comparing Non Linear Widths

We shall see in the sections that follow, that non linear numerical methods for reduced modeling are not as well developed as their linear counterparts. Nevertheless, it is very desirable to have bounds on the non linear widths of $\mathscr{U}_{\mathscr{A}}$ derived from the corresponding widths of $\mathscr{A}$. This would help us understand what non linear methods can possibly bring to the table and also perhaps help in the development of non linear methods for reduced modeling. In this section, we look at what we know about such comparisons.

Let us begin with the library width $d_{n,N}$ defined in (1.36). We assume that for each n, the library has $N = n^A$ bases with A a fixed integer. Then, building on Theorem 1.1, one can prove (see [6]) that whenever $\mathscr{A}$ satisfies

$$d_{n,N}(\mathscr{A})_{L_\infty(D)} \leq C n^{-\alpha}, \quad n \geq 1, \tag{1.59}$$

for some $\alpha > 1$, then for any $\beta < \alpha - 1$, we have

$$d_{n,N}(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \leq C n^{-\beta}, \quad n \geq 1. \tag{1.60}$$

For non linear manifold widths, general comparisons are not known. However, there is a setting, which is sometimes applicable, in which we can derive such comparisons. This setting rests on the following assumption.

Uniqueness Assumption: *We assume that the right side f and the class $\mathscr{A}$ of diffusion coefficients have the property that whenever $u_a = u_{\tilde{a}}$ with $a, \tilde{a}$ in $\mathscr{A}$, then $a = \tilde{a}$ almost everywhere.*

As we shall discuss in §1.6.2, this assumption is satisfied, for example, for the geometric model. Assume now that the **Uniqueness Assumption** is valid and that for the given f and a value of q, the stability bound (1.11) holds. Recall that this stability bound always holds for $q = \frac{2p}{p-2}$, and a certain range of $p \in [2, P]$ provided $f \in W^{-1}(L_p(D))$. Consider the mapping Φ from $\mathcal{U}_{\mathcal{A}}$ into $\mathcal{A}$ which takes $u = u_a$ into a. Since $\mathcal{A}$ is assumed to be compact in $L_q(D)$ and the mapping $a \to u_a$ is known to be continuous (see 1.11), we know from elementary principles that the mapping Φ is also continuous as a mapping from $H_0^1(D)$ to $L_q(D)$.

We now prove that under these assumptions, we have

$$\delta_n(\mathcal{U}_{\mathcal{A}})_{H_0^1(D)} \leq C\,\delta_n(\mathcal{A})_{L_q(D)}. \tag{1.61}$$

Given n and $\epsilon = \delta_n(\mathcal{A})_{L_q(D)}$, we can choose continuous mappings M, b as in the definition of non linear widths so that $b : \mathcal{A} \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to L_q(D)$ and

$$\sup_{a \in \mathcal{A}} ||a - M(b(a))||_{L_q(D)} \leq 2\epsilon. \tag{1.62}$$

Now, we want to construct appropriate mappings for $\mathcal{U}_{\mathcal{A}}$. We can take for $\mathbf{z} \in \mathbb{R}^n$,

$$\tilde{M}(\mathbf{z}) := u_{M(\mathbf{z})}. \tag{1.63}$$

Since M is continuous as a mapping into $L_q(D)$, the L_q stability (1.11) gives that $\tilde{M}$ is also continuous as a mapping into $H_0^1(D)$. Finally, we define $\tilde{b} : \mathcal{U}_{\mathcal{A}} \to \mathbb{R}_n$ by

$$\tilde{b}(u) = b(\Phi(u)), \quad u \in \mathcal{A}. \tag{1.64}$$

Since Φ is continuous from $\mathcal{U}_{\mathcal{A}}$ (in the $H_0^1(D)$ topology) to $\mathcal{A}$ in the L_q topology and b is continuous from $\mathcal{A}$ to $\mathbb{R}^n$, we have that $\tilde{b}$ is also continuous.

Given $u \in \mathcal{U}_{\mathcal{A}}$, we have

$$||u - \tilde{M}(\tilde{b}(u))||_{H_0^1(D)} = ||u_{\Phi(u)} - u_{M(\Phi(u))}||_{H_0^1(D)} \leq C||\Phi(u) - M(\Phi(u))||_{L_q(D)} \leq 2C\epsilon. \tag{1.65}$$

Since $\epsilon = \delta_n(\mathcal{A})_{L_q(D)}$, we have proven (1.61).

1.6 ▪ Widths of our Two Model Classes

The results of the preceding section were directed at giving general a priori guarantees about the performance of linear and non linear methods for reduced modeling. Since, the guarantees do not assume any particular structure of the set $\mathcal{A}$, it may be that they can be improved in settings where we assume a specific structure for $\mathcal{A}$. We now discuss what is known in this regard for our two model classes of elliptic equations.

1.6.1 ▪ Affine Model

We recall that for the affine model, we assume that

$$a(x, y) = \overline{a}(x) + \sum_{j \geq 1} y_j \psi_j(x), \tag{1.66}$$

where the y_j, $j \geq 1$, are parameters in $[-1,1]$. We can always rearrange the indices so that the sequence $b_j := \|\psi_j\|_{L_\infty(D)}$, $j = 1,2\ldots$, is decreasing. For canonical representation systems $\{\psi_j\}$, such as wavelets or Fourier, the rate of decrease of (b_j) to zero is related to the smoothness of $a(x,y)$as a function of x. Indeed, smoothness condtions on a translate into decay conditions on the (b_j).

Let us note that if $(b_j) \in \ell_p$, $p < 1$, then

$$\sup_{y\in U} \|a(\cdot,y) - \sum_{j=1}^{n} y_j \psi_j\|_{L_\infty(D)} \leq \sum_{j=n+1}^{\infty} b_j \leq b_{n+1}^{1-p} \sum_{j=n+1}^{\infty} b_j^p \leq C n^{1-1/p}. \tag{1.67}$$

Here we have used the fact that since (b_j) is decreasing and in ℓ_p, we must have $b_n^p \leq Cn^{-1}$, $n \geq 1$.

The result (1.67) shows that the Kolmogorov widths of $\mathcal{A}$ decays like $O(n^{1-1/p})$:

$$(b_j) \in \ell_p \quad \Longrightarrow \quad d_n(\mathcal{A}) \leq C n^{1-1/p}, \quad n \geq 1. \tag{1.68}$$

We could now use Theorem 1.1 to conclude that $d_n(\mathcal{U}_\mathcal{A})_{H_0^1(D)} \leq C n^{2-1/p-\varepsilon}$, for all $\varepsilon > 0$. However, we have already pointed out in Theorem 1.2 that this result is not optimal for the affine case. In fact, we used this stronger result in the derivation of Theorem 1.1. We give a little more details to illuminate how the stronger result Theorem 1.2 is proved.

Let $\mathcal{F}$ be the set of all sequences $\nu = (\nu_1, \nu_2, \ldots)$ such that ν has finite support and each entry in ν is a nonnegative integer. So $|\nu| = \sum_{j\geq 1} |\nu_j|$ is always finite when $\nu \in \mathcal{F}$. If $\alpha = (\alpha_j)_{j\geq 1}$ is a sequence of positive numbers, we define for all $\nu \in \mathcal{F}$

$$\alpha^\nu := \prod_{j\geq 1} \alpha_j^{\nu_j}.$$

In [9], we showed the following theorem.

Theorem 1.3. *If $(b_j) \in \ell_p$ for some $p < 1$, then*

$$u(x,y) = \sum_{\nu\in\mathcal{F}} c_\nu(x) y^\nu, \tag{1.69}$$

where the functions $c_\nu(x)$ are in $H_0^1(D)$ and $(\|c_\nu\|_{H_0^1(D)}) \in \ell_p$ for the same value of p.

The line of reasoning for proving this theorem is the following. The mapping $F : y \mapsto u(\cdot,y)$ takes U into $H_0^1(D)$. One shows this map is analytic and has a Taylor expansion as a function of y. The complications arise because y consists of an infinite number of variables and the mapping is Banach space valued. The proof of analyticity is not difficult. For a fixed $y \in U$, we know that for all $v \in H_0^1(D)$

$$\int_D a(x,y)\nabla u(x,y)\nabla v(x) dx = \int_D f(x)v(x)dx.$$

Differentiating this identity with respect to the variable y_j gives

$$\int_D a(x,y)\nabla \partial_{y_j} u(x,y)\nabla v(x) dx + \int_D \psi_j(x)\nabla u(x,y)\nabla v(x) dx = 0. \tag{1.70}$$

One then shows that more generally,

$$\int_D a(x,y)\nabla\partial_y^\nu u(x,y)\nabla v(x)dx + \sum_{\{j:\nu_j\neq 0\}} \nu_j \int_D \psi_j(x)\nabla\partial_y^{\nu-e_j} u(x,y)\nabla v(x)dx = 0, \quad (1.71)$$

where e_j is the Kronecker sequence with value 1 at position j and 0 elsewhere. The identity (1.71) is proved by induction on $|\nu|$ using the same idea as used in deriving (1.70). From (1.71) it is not difficult to prove

$$||\partial_y^\nu u(\cdot,y)||_V \leq C_0 \sum_{\{j:\nu_j\neq 0\}} \nu_j b_j (|\nu|-1)!b^{\nu-e_j} = C_0(\sum_{\{j:\nu_j\neq 0\}} \nu_j)(|\nu|-1)!b^\nu = C_0|\nu|!b^\nu,\ \nu\in\mathscr{F}.$$

One now proves the representation (1.69) with $c_\nu(x) := \frac{D^\nu u(x,0)}{\nu!}$ (see [8, 9] for details). The proof that $(||c_\nu||_{H_0^1(D)})\in\ell_p$ whenever $(||\psi_j||_{L_\infty(D)})\in\ell_p$ is far more difficult.

Now let us see how the above theorem gives an estimate for the Kolmogorov n-width of the class $\mathscr{A}$.

Corollary 1.4. *For the affine model $\mathscr{A}$, whenever $(||\psi_j||_{L_\infty})\in\ell_p$, $p<1$, the set $\mathscr{U}_\mathscr{A}$ has n-widths*

$$d_n(\mathscr{A})_{H_0^1(D)} \leq Cn^{1-1/p}, \quad n\geq 1, \quad (1.72)$$

with C depending only on p and the ellipticity constants r,R.

Indeed, from the fact that $(||c_\nu||_{H_0^1(D)})\in\ell_p$, one can use similar arguments to that in (1.67) to prove that there is a set $\Lambda\subset\mathscr{F}$ with $\#(\Lambda)=n$ so that

$$\sup_{y\in U}||u(\cdot,y)-\sum_{\nu\in\Lambda} c_\nu(x)y^\nu||_{H_0^1(D)} \leq Cn^{1-1/p}. \quad (1.73)$$

This shows that the n-dimensional space $V := \text{span}\{c_\nu : \nu\in\Lambda\}$ approximates $\mathscr{A}$ with accuracy $Cn^{1-1/p}$ and therefore $d_n(\mathscr{A})_{H_0^1(D)} \leq Cn^{1-1/p}$.

One important observation about this bound for the entropy is that we have broken the curse of dimensionality. Indeed, the parameters $y_1,y_2,\ldots$ are infinite. In typical applications, the parameters are finite in number, say d, but then this result shows that the exponent of n in this bound does not depend on d.

1.6.2 ▪ The Geometric Model

Although, as we shall see, the results about numerical performance for this example are not definitive, it is still instructive to discuss what is known and which questions are still unresolved in the case of the geometric model. Following our usual paradigm, let us first consider $\mathscr{A}$ and try to understand its complexity. It makes no sense to consider the approximation of the functions $a\in\mathscr{A}$ in the $L_\infty(D)$ norm since each of these functions is discontinuous and therefore any approximation would have to match these discontinuities exactly. On the other hand, we can approximate a in an $L_q(\Omega)$ norm and use the perturbation result (1.9). For the convex domain $D=[0,1]^2$, the best possible range of q for the perturbation theorem is $Q\leq q\leq\infty$, where the smallest value of Q depends on the constants r,R in the uniform ellipticity assumption (see [4]).

We know from our general theory that if we measure the complexity of $\mathscr{A}$ and $\mathscr{U}_\mathscr{A}$ in the sense of their entropy then (1.43) always holds. One can rather easily compute the

entropy numbers of $\mathscr{A}$ in $L_q(D)$ for any $q \geq 2$:

$$\epsilon_n(\mathscr{A})_{L_q(D)} \sim n^{-1/q}, \quad n \geq 1. \tag{1.74}$$

From our general comparison (1.43), this gives

$$\epsilon_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \leq C n^{-1/q}, \quad n \geq 1, \tag{1.75}$$

with the best bound holding for $q = Q$.

Let us next discuss what is known about linear widths for $\mathscr{A}$ and $\mathscr{U}_{\mathscr{A}}$. The following bounds for n-widths can be shown with appropriate constants $C_1, C_2 > 0$:

$$C_2 n^{-\frac{1}{2q}} \leq d_n(\mathscr{A})_{L_q(D)} \leq C_1 n^{-\frac{1}{2q}}, \quad n \geq 1, \tag{1.76}$$

with C_1, C_2 absolute constants. To prove the upper estimate, we consider the dictionary $\mathscr{D}_n$ which consist of the functions χ_R, where $R = [(i-1)/n, i/n) \times [0, j/n]$, $1 \leq i, j \leq n$. Given a function $a \in \mathscr{A}$, corresponding to the Lipschitz function φ, and a value $1 \leq i \leq n$, we let j_i be the largest integer such that

$$\frac{j_i}{n} \leq \varphi(x), \quad x \in [\frac{i-1}{n}, \frac{i}{n}]. \tag{1.77}$$

We consider the function $g_n := 1 + \sum_{i=1^n} \chi_{R_i}$ with $R_i := [(i-1)/n, i/n) \times [0, j_i/n]$, $1 \leq i \leq n$. The function g agrees with a except on a set of measure $\leq 1/n$. Hence,

$$\|a - g\|_{L_q(D)} \leq n^{-1/q}. \tag{1.78}$$

Since the space spanned by the $\mathscr{D}_n$ has dimension n^2, we obtain the upper estimate.

The lower estimate is a little more intricate. We first consider the case $q = 2$. Let $V \subset L_2(D)$ be any fixed linear space of dimension $N \leq n^2/2$ with n a two power and let $\varphi_1, \ldots, \varphi_N$ be an orthonormal system for V. We assume $\operatorname{dist}(\mathscr{A}, V)_{L_2(D)} \leq \epsilon$ and derive a bound from below for ϵ.

We will first construct some functions that can be approximated well by V. Let ψ_k be the piecewise linear function which is zero outside $I_k := [k/n, (k+1)/n]$ and is the hat function with height $1/(2n)$ on I_k. Then, for any $j > 0$ and any set $\Lambda \subset \{0, 1, \ldots, n-1\}$, the function $g_{j,\Lambda} := j/n + \sum_{k \in \Lambda} \psi_k$ is in $\operatorname{Lip}_1 1$. The function

$$f_{j,\Lambda} := a_{g_{j,\Lambda}} - a_{g_{j,\Lambda^c}}, \tag{1.79}$$

can be approximated to accuracy 2ϵ by the space V. Each of these functions has support in the strip $j \leq y \leq j+1$ and has norm $\|f_{j,\Lambda}\|^2_{L^2(D)} = 1/(3n)$. Obviously, these functions with different values of j are orthogonal. Moreover, for a fixed j, we can choose n different sets Λ such that these functions are also orthogonal. Indeed, we take $\Lambda = \{0, 1, \ldots, n-1\}$ and then the other $n-1$ choices corresponding to where the Walsh functions of order n are positive. In this way, we get n^2 orthogonal functions. We define the functions $h_1, \ldots, h_{n^2}$ where each h_j is one of the functions $\sqrt{3n} f_{i,\Lambda}$ with the n^2 different choices of these function in (1.79). Hence, these functions are an orthonormal system and each of these functions can be approximated to accuracy $2\epsilon\sqrt{3n}$.

We consider the $n^2 \times N$ matrix B whose i, j entry is $b_{i,j} := |\langle h_i, \varphi_j \rangle|^2$. Then, each of the N columns has sum at most 1. Hence, one of the rows i^* has sum at most $N n^{-2} \leq 1/2$.

This means that in approximating h_{i^*} by the elements of V in the $L_2(D)$ norm, we incur an error of at least $1/\sqrt{2}$. It follows that $2\epsilon\sqrt{3n} \geq 1/\sqrt{2}$. In other words, $\epsilon \geq [2\sqrt{6}]^{-1} n^{-1/2}$. Since the only restriction on N is that $N \leq n^2/2$, we obtain

$$d_{n^2}(\mathscr{A})_{L_2(D)} \geq C n^{-1/2}, \quad n \geq 1, \tag{1.80}$$

with C an absolute constant. The lower bound in (1.76) for $q = 2$ follows.

Let us mention, without proof, that the lower bound $d_n(\mathscr{A})_{L_q(D)} \geq n^{-\frac{1}{2q}}$, for $2 \leq q \leq \infty$, can be proved from the L_2 result by using an interpolation argument.

The above results describe how well we can approximate $\mathscr{A}$ and say nothing about approximating $\mathscr{U}_{\mathscr{A}}$. Indeed, our only direct estimate for $d_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)}$ in terms of $d_n(\mathscr{A})_{L_2(D)}$ is that given by Theorem 1.1. But multiplying $d_n(\mathscr{A})_{L_2(D)}$ by n does not give a decay to zero and therefore Theorem 1.1 gives a useless estimate. So, it remains an open problem to determine the n-width of $\mathscr{U}_{\mathscr{A}}$ for the geometric model.

1.7 ▪ Numerical Methods for Parametric Equations

Let us now turn our attention to numerical methods for solving a family of parametric equations with diffusion coefficients coming from the set $\mathscr{A}$. We wish to construct a numerical solver such that given a query $a \in \mathscr{A}$, it produces a function $\hat{u}_a$ which is a good approximation to u_a in the $H_0^1(D)$ norm.

1.7.1 ▪ Online and Offline Costs

This construction of a solver is decoupled into two tasks.

Offline costs: This is the cost of developing the numerical method which is tailor made for $\mathscr{A}$. For example, in linear methods, it is the computational cost needed to find a good n-dimensional subspace V_n contained in $H_0^1(D)$ which will be used to build an approximation to u_a when given a query $a \in \mathscr{A}$. For a non linear method based on n-term approximation from a dictionary of size N, it would be the cost in finding a good dictionary. Notice that the offline cost is a one time investment for the class $\mathscr{A}$ and does not include the task of actually finding an approximation to u_a given a query $a \in \mathscr{A}$.

We should mention that in some applications, we are not so much interested in finding an approximation $\hat{u}_a$ to u_a as we are in the evaluation of a linear functional ℓ on $H_0^1(D)$ to the solution u_a This is then the problem of developing a numerical method which approximates the real valued function $L(a) = \ell(u_a)$ given a query $a \in \mathscr{A}$. In this case, the offline costs would include the building of an approximation $\hat{L}$ to L. We will not say much more about this important second problem.

Online costs: This is the cost of implementing the solver which was built offline for finding an approximation $\hat{u}_a$ to u_a given a query a. For linear methods, this approximation is usually taken as the Galerkin projection onto V_n, although other projectors may also be reasonable in some scenarios. The Galerkin projector gives the best approximation to u_a from V_n in the $H_0^1(D,a)$ norm. Given the linear space V_n, the Galerkin projection constructs $\hat{u}_a \in V_n$ as the solution to the discrete system of equations

$$\langle \hat{u}_a, v \rangle_a = \langle f, v \rangle, \quad \forall v \in V_n, \tag{1.81}$$

where $\langle \cdot, \cdot \rangle_a$ is the $H_0^1(D,a)$ inner product. If we choose a basis $\varphi_1, \ldots, \varphi_n$ for V_n, then

$\hat{u}_a = \sum_{j=1}^n c_j \varphi_j$ where the coefficients $\mathbf{c} = (c_j)_{j=1}^n$ satisfy

$$A\mathbf{c} = \mathbf{f}, \tag{1.82}$$

where $A = (a_{ij})_{i,j=1}^n$, $a_{ij} := \langle \varphi_i, \varphi_j \rangle_a$, is the so-called stiffness matrix and $\mathbf{f} := (f_i)_{i=1}^n$, with $f_i := \langle f, \varphi_i \rangle$, $i = 1, \ldots, n$, is the discretization of the right side f. From the ellipticity assumption, the matrix A is positive definite and so the system is efficiently solved using standard numerical solvers for linear systems. The performance of this numerical method is usually measured by the error in the $H_0^1(D,a)$ norm:

$$\|u_a - \hat{u}_a\|_{H_0^1(D,a)} = \mathrm{dist}(u_a, V_n)_{H_0^1(D,a)} \approx \mathrm{dist}(u_a, V_n)_{H_0^1(D)}. \tag{1.83}$$

When the Galerkin projection is used, then give the query $a \in \mathscr{A}$, one must assemble the stiffness matrix and then solve the corresponding matrix problem. While the assembly of the matrix is a serious problem, we will largely ignore it here since we have nothing to add over what is already known.

Relevance of Entropy in Online - Offline Comparisons

Entropy suggests the following extreme setting for the online-offline comparisons. Let $\epsilon := C\epsilon_n(\mathscr{A})_{L_q(D)}$ and numerically find a cover $\{B(a_i, \epsilon)\}_{i=1}^N$, $N = 2^n$ for $\mathscr{A}$ and then compute offline the solutions u_{a_i}, $i = 1, \ldots, N$. Since $\mathscr{A}$ is known to us, we can usually find the cover of $\mathscr{A}$ by some form of piecewise polynomial approximation followed by quantization of the coefficients of the approximation. The offline costs would be proportional to Nm where m is the computational cost in employing an off the shelf solver to find an approximation $\hat{u}_{a_i}$ to u_{a_i} accurate to error ϵ.

For the online computation, given a query a, we find an approximation a_i to a using piecewise polynomial approximation and quantization. This is then followed by a look up table to find the approximation $\hat{u}_{a_i}$ as the approximation to u_a. The accuracy of this method is $C\epsilon_n(\mathscr{A})_{L_q(D))}$. The offline cost is exceedingly high but the online cost is very low since it does not involve a pde solve.

The point of this example is to show that it is not only the online cost of the solver that matters. One has to take into consideration the offline investment costs which are extreme in the above example. It is not clear exactly how this balancing should be done but it would be beneficial to quantify this in someway in order to advance the theory of reduced basis methods.

1.7.2 ▪ Finding a Good Linear Subspace

The central question in developing linear methods is how to find a good choice for the finite dimensional space V_n? Since we want V_n to be used for all $a \in \mathscr{A}$, it should be efficient at approximating all of the elements in $\mathscr{U}_{\mathscr{A}}$. Recall that all of the norms $\|\cdot\|_{H_0^1(D,a)}$ are equivalent to $\|\cdot\|_{H_0^1(D)}$. So essentially, the best choice for V_n is a subspace of $H_0^1(D)$ which achieves the Kolmogorov width $d_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)}$. Of course finding such a subspace may be difficult but it serves as a benchmark for the optimal performance we can expect.

One of the prominent methods for finding a good subspace are *Reduced Basis Methods* (RBM) and are a well studied subject [5, 21, 22, 24, 25, 27]. The general philosophy of such methods is that one is willing to spend high computational costs to determine offline a good subspace V_n. Typically, the space V_n is spanned by n functions $u_{a_i} \in \mathscr{U}_{\mathscr{A}}$,

$i = 1, \ldots, n$. These functions are called *snapshots* of $\mathcal{U}_{\mathcal{A}}$. The most popular method for finding these snapshots is the following intriguing *greedy algorithm* introduced first in [27]. While we are primarily interested in this algorithm in the case of a compact set K of a Hilbert space (in our case $K = \mathcal{U}_{\mathcal{A}}$ and the Hilbert space is $H_0^1(D)$), we will formulate the algorithm for any Banach space X.

Let X be a Banach space with norm $\|\cdot\| := \|\cdot\|_X$, and let K be one of its compact subsets. For notational convenience only, we shall assume that the elements f of K satisfy $\|f\|_X \leq 1$. We consider the following greedy algorithm for generating approximation spaces for K.

The Pure Greedy Algorithm: We first choose a function f_0 such that

$$\|f_0\| = \max_{f \in K} \|f\|. \tag{1.84}$$

Assuming $\{f_0, \ldots, f_{n-1}\}$ and $V_n := \operatorname{span}\{f_0, \ldots, f_{n-1}\}$ have been selected, we then take $f_n \in K$ such that

$$\operatorname{dist}(f_n, V_n)_X = \max_{f \in K} \operatorname{dist}(f, V_n)_X, \tag{1.85}$$

and define

$$\sigma_n := \sigma_n(K)_X := \operatorname{dist}(f_n, V_n)_X := \sup_{f \in K} \inf_{g \in V_n} \|f - g\|. \tag{1.86}$$

This greedy algorithm was introduced, for the case X is a Hilbert space in [21, 22]. In numerical settings, one cannot find the f_j exactly and also estimates for error needed in this algorithm are also not known precisely. This leads one to consider weaker forms of this algorithm which match better their application.

Weak greedy algorithm We fix a constant $0 < \gamma \leq 1$. At the first step of the algorithm, one chooses a function $f_0 \in K$ such that

$$\|f_0\| \geq \gamma \sigma_0(K)_X := \max_{f \in K} \|f\|.$$

At the general step, if $f_0, \ldots, f_{n-1}$ have been chosen, we set $V_n := \operatorname{span}\{f_0, \ldots, f_{n-1}\}$, and

$$\sigma_n(f)_X := \operatorname{dist}(f, V_n)_X.$$

We now choose $f_n \in \mathcal{F}$ such that

$$\sigma_n(f_n)_X \geq \gamma \max_{f \in K} \sigma_n(f)_X, \tag{1.87}$$

to be the next element in the greedy selection. Note that if $\gamma = 1$, then the weak greedy algorithm reduces to the greedy algorithm that we have introduced above.

Notice that similar to the greedy algorithm, $(\sigma_n(K)_X)_{n \geq 0}$ is also monotone decreasing. It is also important to note that neither the Pure Greedy Algorithm or the Weak Greedy Algorithm give a unique sequence $(f_n)_{n \geq 0}$, nor is the sequence $(\sigma_n(K)_X)_{n \geq 0}$ unique. In all that follows, the notation reflects any sequences which can arise in the implementation of the weak greedy selection for the fixed value of γ.

1.7.3 ▪ Performance of the Weak Greedy Algorithm

We are interested in how well the space V_n, generated by the weak greedy algorithm, approximates the elements of K. For this purpose we would like to compare its performance with the best possible performance which is given by the Kolmogorov width $d_n(K)_X$ of K. Of course, if $(\sigma_n)_{n\geq 0}$ decays at a rate comparable to $(d_n)_{n\geq 0}$, this would mean that the greedy selection provides essentially the best possible accuracy attainable by n-dimensional subspaces. Various comparisons have been given between σ_n and d_n. An early result in this direction, in the case that X is a Hilbert space $\mathcal{H}$, was given in [5] where it was proved that

$$\sigma_n(K)_{\mathcal{H}} \leq Cn2^n d_n(K)_{\mathcal{H}}, \tag{1.88}$$

with C an absolute constant. While this is an interesting comparison, it is only useful if $d_n(K)_{\mathcal{H}}$ decays to zero faster than $n^{-1}2^{-n}$.

Various improvements on (1.88) were given in [1], again in the Hilbert space setting. We mention two of these. It was shown that if $d_n(K)_{\mathcal{H}} \leq Cn^{-\alpha}$, $n = 1,2,\ldots$, then

$$\sigma_n(K)_{\mathcal{H}} \leq C'_\alpha n^{-\alpha}. \tag{1.89}$$

This shows that in the scale of polynomial decay the greedy algorithm performs with the same rates as n-widths. A related result was proved for sub-exponential decay. If for some $0 < \alpha \leq 1$, we have $d_n(K)_{\mathcal{H}} \leq Ce^{-cn^\alpha}$, $n = 1,2,\ldots$, then

$$\sigma_n(K)_{\mathcal{H}} \leq C'_\alpha e^{-c'_\alpha n^\beta}, \quad \beta = \frac{\alpha}{\alpha+1}, \quad n = 1,2,\ldots. \tag{1.90}$$

These results were improved in [14] and extended to the case of a general Banach space X as we are now discussing. We will outline what is known in this direction and sketch how these results are proved in the following section.

1.7.4 ▪ Results for a Banach Space

The analysis of the greedy algorithm is quite simple and executed with elementary results from linear algebra. We provide a little of the details since this may help develop the intuition of the reader. A core result for the analysis of greedy algorithms is the following lemma from [14].

Lemma 1.5. *Let $G = (g_{i,j})$ be a $K \times K$ lower triangular matrix with rows $\mathbf{g}_1,\ldots,\mathbf{g}_K$, W be any m dimensional subspace of $\mathbb{R}^K$, and P be the orthogonal projection of $\mathbb{R}^K$ onto W. Then*

$$\prod_{i=1}^{K} g_{i,i}^2 \leq \left\{ \frac{1}{m} \sum_{i=1}^{K} \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m \left\{ \frac{1}{K-m} \sum_{i=1}^{K} \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}, \tag{1.91}$$

where $\|\cdot\|_{\ell_2}$ is the Euclidean norm of a vector in $\mathbb{R}^K$.

Proof. We choose an orthonormal basis $\varphi_1,\ldots,\varphi_m$ for the space W and complete it into an orthonormal basis $\varphi_1,\ldots,\varphi_K$ for $\mathbb{R}^K$. If we denote by Φ the $K \times K$ orthogonal matrix whose j-th column is φ_j, then the matrix $C := G\Phi$ has entries $c_{i,j} = \langle \mathbf{g}_i, \varphi_j \rangle$. We denote by $\mathbf{c}_j$, the j-th column of C. It follows from the arithmetic geometric mean inequality

for the numbers $\{\|\mathbf{c}_j\|_{\ell_2}^2\}_{j=1}^m$ that

$$\prod_{j=1}^m \|\mathbf{c}_j\|_{\ell_2}^2 \le \left\{ \frac{1}{m} \sum_{j=1}^m \|\mathbf{c}_j\|_{\ell_2}^2 \right\}^m = \left\{ \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^K \langle \mathbf{g}_i, \varphi_j \rangle^2 \right\}^m = \left\{ \frac{1}{m} \sum_{i=1}^K \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m. \tag{1.92}$$

Similarly,

$$\prod_{j=m+1}^K \|\mathbf{c}_j\|_{\ell_2}^2 \le \left\{ \frac{1}{K-m} \sum_{j=m+1}^K \|\mathbf{c}_j\|_{\ell_2}^2 \right\}^{K-m} = \left\{ \frac{1}{K-m} \sum_{i=1}^K \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m} \tag{1.93}$$

Now, Hadamard's inequality for the matrix C and relations (1.92) and (1.93) result in

$$(\det C)^2 \le \prod_{j=1}^K \|\mathbf{c}_j\|_{\ell_2}^2 \le \left\{ \frac{1}{m} \sum_{i=1}^K \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m \left\{ \frac{1}{K-m} \sum_{i=1}^K \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}. \tag{1.94}$$

The latter inequality and the fact that $\det G = \prod_{i=1}^K g_{i,i}$ and $|\det C| = |\det G|$ gives (1.91).

□

Let us now see how this lemma is utilized to derive convergence results for the greedy algorithm. We will for the moment restrict ourselves to the case of a Hilbert space and the weak greedy algorithm with constant γ. Later, we shall say what changes are made when X is a general Banach space.

Note that in general, the weak greedy algorithm does not terminate and we obtain an infinite sequence $f_0, f_1, f_2, \ldots$. In order to have a consistent notation in what follows, we shall define $f_m := 0$, $m > N$, if the algorithm terminates at N, i.e. if $\sigma_N(K)_{\mathcal{H}} = 0$. By $(f_n^*)_{n\ge 0}$ we denote the orthonormal system obtained from $(f_n)_{n\ge 0}$ by Gram-Schmidt orthogonalization. It follows that the orthogonal projector P_n from $\mathcal{H}$ onto V_n is given by

$$P_n f = \sum_{i=0}^{n-1} \langle f, f_i^* \rangle f_i^*,$$

and, in particular,

$$f_n = P_{n+1} f_n = \sum_{j=0}^n a_{n,j} f_j^*, \quad a_{n,j} = \langle f_n, f_j^* \rangle,\ j \le n.$$

There is no loss of generality in assuming that the infinite dimensional Hilbert space $\mathcal{H}$ is $\ell_2(\mathbb{N} \cup \{0\})$ and that $f_j^* = e_j$, where e_j is the vector with a one in the coordinate indexed by j and is zero in all other coordinates, i.e. $(e_j)_i = \delta_{j,i}$.

We consider the lower triangular matrix

$$A := (a_{i,j})_{i,j=0}^\infty, \quad a_{i,j} := 0,\ j > i.$$

This matrix incorporates all the information about the weak greedy algorithm on K. The following two properties characterize any lower triangular matrix A generated by the weak greedy algorithm with constant γ. With the notation $\sigma_n := \sigma_n(K)_{\mathcal{H}}$, we have:

P1: The diagonal elements of A satisfy $\gamma\sigma_n \le |a_{n,n}| \le \sigma_n$.

P2: For every $m \ge n$, one has $\sum_{j=n}^{m} a_{m,j}^2 \le \sigma_n^2$.

Indeed, **P1** follows from

$$a_{n,n}^2 = \|f_n\|^2 - \|P_n f_n\|^2 = \|f_n - P_n f_n\|^2,$$

combined with the weak greedy selection property (1.87). To see **P2**, we note that for $m \ge n$,

$$\sum_{j=n}^{m} a_{m,j}^2 = \|f_m - P_n f_m\|^2 \le \max_{f\in K} \|f - P_n f\|^2 = \sigma_n^2.$$

Remark 1. *If A is any matrix satisfying* **P1** *and* **P2** *with $(\sigma_n)_{n\ge 0}$ a decreasing sequence that converges to 0, then the rows of A form a compact subset of $\ell_2(\mathbb{N}\cup\{0\})$. If K is the set consisting of these rows, then one of the possible realizations of the weak greedy algorithm with constant γ will choose the rows in that order and A will be the resulting matrix.*

Theorem 1.6. *For the weak greedy algorithm with constant γ in a Hilbert space $\mathcal{H}$ and for any compact set K, we have the following inequalities between $\sigma_n := \sigma_n(K)_{\mathcal{H}}$ and $d_n := d_n(K)_{\mathcal{H}}$, for any $N \ge 0$, $J \ge 1$, and $1 \le m < J$,*

$$\prod_{i=1}^{J} \sigma_{N+i}^2 \le \gamma^{-2J} \left\{\frac{J}{m}\right\}^m \left\{\frac{J}{J-m}\right\}^{J-m} \sigma_{N+1}^{2m} d_m^{2J-2m}. \tag{1.95}$$

Proof. We consider the $J \times J$ matrix $G = (g_{i,j})$ which is formed by the rows and columns of A with indices from $\{N+1,\dots,N+J\}$. Each row $\mathbf{g}_i$ is the restriction of f_{N+i} to the coordinates $N+1,\dots,N+K$. Let $\mathcal{H}_m$ be the m-dimensional Kolmogorov subspace of $\mathcal{H}$ for which $\operatorname{dist}(K,\mathcal{H}_m) = d_m$. Then, $\operatorname{dist}(f_{N+i},\mathcal{H}_m) \le d_m$, $i = 1,\dots J$. Let $\widetilde{W}$ be the linear space which is the restriction of $\mathcal{H}_m$ to the coordinates $N+1,\dots,N+J$. In general, $\dim(\widetilde{W}) \le m$. Let W be an m dimensional space, $W \subset \operatorname{span}\{e_{N+1},\dots,e_{N+J}\}$, such that $\widetilde{W} \subset W$ and P and $\widetilde{P}$ are the projections in $\mathbb{R}^K$ onto W and $\widetilde{W}$, respectively. Clearly,

$$\|P\mathbf{g}_i\|_{\ell_2} \le \|\mathbf{g}_i\|_{\ell_2} \le \sigma_{N+1}, \quad i = 1,\dots,J, \tag{1.96}$$

where we have used Property **P2** in the last inequality. Note that

$$\|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2} \le \|\mathbf{g}_i - \widetilde{P}\mathbf{g}_i\|_{\ell_2} = \operatorname{dist}(\mathbf{g}_i,\widetilde{W}) \le \operatorname{dist}(f_{N+i},\mathcal{H}_m) \le d_m, \quad i = 1,\dots,J. \tag{1.97}$$

It follows from Property **P1** that

$$\prod_{i=1}^{J} |a_{N+i,N+i}| \ge \gamma^J \prod_{i=1}^{J} \sigma_{N+i}. \tag{1.98}$$

We now apply Lemma 1.5 for this G and W, and use estimates (1.96), (1.97), and (1.98) to derive (1.95). □

Let us now indicate how one derives some of the performance results for the greedy algorithm from this theorem.

Corollary 1.7. *For the weak greedy algorithm with constant γ in a Hilbert space $\mathcal{H}$, we have the following:*

(i) *For any compact set K and $n \geq 1$, we have*

$$\sigma_n(K) \leq \sqrt{2}\gamma^{-1} \min_{1\leq m<n} d_m^{\frac{n-m}{n}}(K). \tag{1.99}$$

In particular $\sigma_{2n}(K) \leq \sqrt{2}\gamma^{-1}\sqrt{d_n(K)}$, $n = 1,2\ldots$.

(ii) *If $d_n(K) \leq C_0 n^{-\alpha}$, $n = 1,2,\ldots$, then $\sigma_n(K) \leq C_1 n^{-\alpha}$, $n = 1,2\ldots$, with $C_1 := 2^{5\alpha+1}\gamma^{-2}C_0$.*

(iii) *If $d_n(K) \leq C_0 e^{-c_0 n^\alpha}$, $n = 1,2,\ldots$, then $\sigma_n(K) \leq \sqrt{2C_0}\gamma^{-1}e^{-c_1 n^\alpha}$, $n = 1,2\ldots$, where $c_1 = 2^{-1-2\alpha}c_0$,*

Proof. (i) We take $N = 0, J = n$ and any $1 \leq m < n$ in Theorem 1.6, use the monotonicity of $(\sigma_n)_{n\geq 0}$ and the fact that $\sigma_0 \leq 1$ to obtain

$$\sigma_n^{2n} \leq \prod_{j=1}^{n} \sigma_j^2 \leq \gamma^{-2n} \left\{\frac{n}{m}\right\}^m \left\{\frac{n}{n-m}\right\}^{n-m} d_m^{2n-2m}. \tag{1.100}$$

Since $x^{-x}(1-x)^{x-1} \leq 2$ for $0 < x < 1$, we derive (1.99).

(ii) It follows from the monotonicity of $(\sigma_n)_{n\geq 0}$ and (1.95) for $N = J = n$ and any $1 \leq m < n$ that

$$\sigma_{2n}^{2n} \leq \prod_{j=n+1}^{2n} \sigma_j^2 \leq \gamma^{-2n} \left\{\frac{n}{m}\right\}^m \left\{\frac{n}{n-m}\right\}^{n-m} \sigma_n^{2m} d_m^{2n-2m}.$$

In the case $n = 2s$ and $m = s$ we have

$$\sigma_{4s} \leq \sqrt{2}\gamma^{-1}\sqrt{\sigma_{2s}d_s}. \tag{1.101}$$

Now we prove our claim by contradiction. Suppose it is not true and M is the first value where $\sigma_M(\mathcal{F}) > C_1 M^{-\alpha}$. Let us first assume $M = 4s$. From (1.101), we have

$$\sigma_{4s} \leq \sqrt{2}\gamma^{-1}\sqrt{C_1(2s)^{-\alpha}}\sqrt{C_0 s^{-\alpha}} = \sqrt{2^{1-\alpha}C_0C_1}\gamma^{-1}s^{-\alpha}, \tag{1.102}$$

where we have used the fact that $\sigma_{2s} \leq C_1(2s)^{-\alpha}$ and $d_s \leq C_0 s^{-\alpha}$. It follows that

$$C_1(4s)^{-\alpha} < \sigma_{4s} \leq \sqrt{2^{1-\alpha}C_0C_1}\gamma^{-1}s^{-\alpha},$$

and therefore

$$C_1 < 2^{3\alpha+1}\gamma^{-2}C_0 < 2^{5\alpha+1}\gamma^{-2}C_0,$$

which is the desired contradiction. If $M = 4s+q$, $q \in \{1,2,3\}$, then it follows from (1.102) and the monotonicity of $(\sigma_n)_{n\geq 0}$ that

$$C_1 2^{-3\alpha}s^{-\alpha} = C_1 2^{-\alpha}(4s)^{-\alpha} < C_1(4s+q)^{-\alpha} < \sigma_{4s+q} \leq \sigma_{4s} \leq \sqrt{2^{1-\alpha}C_0C_1}\gamma^{-1}s^{-\alpha}.$$

From this, we obtain

$$C_1 < 2^{5\alpha+1}\gamma^{-2}C_0,$$

which is the desired contradiction in this case. This completes the proof of (ii).

(iii) From (i), we have

$$\sigma_{2n+1} \le \sigma_{2n} \le \sqrt{2}\gamma^{-1}\sqrt{d_n} \le \sqrt{2C_0}\gamma^{-1}e^{-\frac{c_0}{2}n^\alpha} = \sqrt{2C_0}\gamma^{-1}e^{-c_0 2^{-1-\alpha}(2n)^\alpha}, \tag{1.103}$$

from which (iii) easily follows. □

Let us now comment on what happens when X is a general Banach space. The analysis is quite similar to that above (see [14]) however there is some loss in the approximation rate. The precise results are as follows:

(i) For any $n \ge 1$ we have $\sigma_{2n} \le 2\gamma^{-1}\sqrt{nd_n}$,

(ii) If for $\alpha > 0$, we have $d_n \le C_0 n^{-\alpha}$, $n = 1,2,\ldots$, then for any $0 < \beta < \min\{\alpha, 1/2\}$, we have $\sigma_n \le C_1 n^{-\alpha+1/2+\beta}$, $n = 1,2\ldots$, with

$$C_1 := \max\left\{C_0 4^{4\alpha+1}\gamma^{-4}\left(\frac{2\beta+1}{2\beta}\right)^\alpha, \max_{n=1,\ldots,7}\{n^{\alpha-\beta-1/2}\}\right\}.$$

(iii) If for $\alpha > 0$, we have $d_n \le C_0 e^{-c_0 n^\alpha}$, $n = 1,2,\ldots$, then $\sigma_n < \sqrt{2C_0}\gamma^{-1}\sqrt{n}e^{-c_1 n^\alpha}$, $n = 1,2\ldots$, where $c_1 = 2^{-1-2\alpha}c_0$. The factor $\sqrt{n}$ can be deleted by reducing the constant c_1.

In particular, we see that in the estimates (i) and (ii), we lose a factor $\sqrt{n}$ in approximation rate when compared with the Hilbert space case. It can be shown that in general, this loss cannot be avoided [14].

1.7.5 ▪ Practical Considerations in the Offline Implementation of Greedy Algorithms

Let us now return to the application of the above greedy algorithms to our parametric PDE problem. On first glance, it appears that the offline implementation of this algorithm is computationally not feasible, since it requires an accurate estimate of $\|u_a - P_{V_n}u_a\|_{H_0^1(D)}$ for all $a \in \mathcal{A}$. On the surface, this would require solving (1.1) for each a which is of course what we are trying to avoid. Fortunately, as is well known, this norm is equivalent to

$$S(a) := \|f - P_n u_a\|_{H^{-1}(D)}, \tag{1.104}$$

which can be computed (since both f and $P_n u_a$ are available) without computing u_a. (We do not discuss the role of the constants in this equivalence, even though they are an important issue.) We are still left with the problem of having to calculate this surrogate quantity for all a. What one does in practice is the following.

We know that whatever accuracy we have for the discretization of $\mathcal{A}$ in $L_\infty(D)$ (or $L_q(D)$) then this accuracy will be inherited by $\mathcal{U}_\mathcal{A}$ because of (1.8). Suppose a discretization $\tilde{\mathcal{A}}$ of $\mathcal{A}$ has accuracy ϵ and we find an $a^* \in \tilde{\mathcal{A}}$ such that

$$S(a^*) \ge C_0\epsilon, \tag{1.105}$$

with C_0 an appropriately large fixed constant (determined by the equivalency constants for the surrogate). Then, we are guaranteed that this discretization is accurate enough for the implementation of the weak greedy algorithm. Hence, we start with a coarse discretization of $\mathcal{A}$ and then decrease the resolution ϵ of the discretization until (1.105) is satisfied.

1.7.6 ▪ Summary on Performance of Reduced Bases

Let us summarize what we know about the performance of the weak greedy algorithm for our two sample model classes.

Affine Model Class (see §1.3.1): Assume that $(||\psi_j||_{L_\infty(D)}) \in \ell_p$, for some $p < 1$. Then, we know that

(i) $d_n(\mathscr{A})_{L_\infty(D)} \le Cn^{1-1/p}$, $n \ge 1$,
(ii) $d_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \le Cn^{1-1/p}$, $n \ge 1$,
(iii) The weak greedy algorithm generates spaces V_n such that $\sigma_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \le Cn^{1-1/p}$, $n \ge 1$.

Let us mention that an alternative to the construction of a good basis using the weak greedy algorithm is to utilize a selection based on monotone sets (also known as lower sets) as discussed in [10].

Geometric Model Class (see §1.3.2): Assume that the stability inequality (1.11) holds for a value of $q \in [2, \infty)$. Then, we know that

(i) $d_n(\mathscr{A})_{L_q(D)} \le Cn^{-\frac{1}{2q}}$, $n \ge 1$,
(ii) We do not know any estimate of the form

$$d_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \le Cn^{-\alpha}, \quad n \ge 1, \tag{1.106}$$

for a value of $\alpha > 0$.
(iii) If we could prove an estimate (1.106), then the weak greedy algorithm would generate spaces V_n such that $\sigma_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \le Cn^{-\alpha}$, $n \ge 1$.

1.8 ▪ Non Linear Methods in Reduced Bases

Generally speaking, there is a large benefit to using non linear approximation in the construction of numerical methods for pdes. Several interesting non linear approaches are emerging: hp-reduced basis for elliptic problems [15], [16]; adaptive parameter partitioning [17]; reduced basis selection from dictionaries [19]; and local greedy by parameter distance[23]. We are not aware of any definitive a priori analysis of such algorithms which would substantiate the use of non linear methods. In this section, we make some heuristic comments about the possible utilization of non linear methods in reduced modeling.

For the Affine Model, there seems to be no advantage in using non linear methods since the manifold $\mathscr{U}_{\mathscr{A}}$ is provably smooth. On the other hand, the case of the Geometric Model seems ripe for the exploitation of non linear methods. We consider only this geometric example in what follows in this section. We have seen that the linear Kolmogorov widths of $\mathscr{A}$ satisfy

$$d_n(\mathscr{A})_{L_2(D)} \ge Cn^{-1/4}, \quad n \ge 1. \tag{1.107}$$

This is a good indication that the same lower bound holds for the widths of $\mathscr{U}_{\mathscr{A}}$ in $H_0^1(D)$, although, as we have pointed out in the last section no such results have actually been proven.

1.8.1 ▪ Entropy Numbers for the Geometric Model

As we have already noted in (1.74), the entropy numbers $\epsilon_n(\mathscr{A})_{L_q(D)}$ behave like $n^{-1/q}$. It follows from out comparison (1.43) that the entropy numbers $\epsilon_n(\mathscr{U}_{\mathscr{A}})_{H_0^1(D)} \le Cn^{-1/q}$.

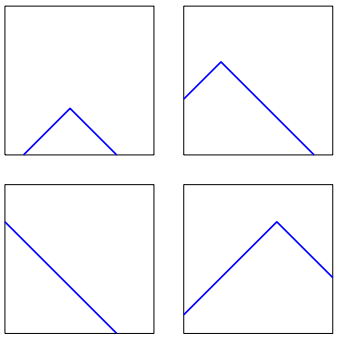

Figure 1.3. *The basis functions $\phi_{i,j}$ with vertex $(i/n, j/n)$. The line segments have slope ± 1.*

When $q = 2$, this shows a non linear performance of order $O(n^{-1/2})$ is expected for approximating $\mathcal{U}_{\mathcal{A}}$.

1.8.2 ▪ Non Linear n-widths for the Geometric Model

Let us begin with our usual strategy of first trying to understand the non linear widths of $\mathcal{A}$ in $L_2(D)$. We have already discussed the dictionary $\mathcal{D}$ which consists of the n^2 functions χ_R, where $R = [(i-1)/n, i/n) \times [0, j/n]$, $1 \le i, j \le n$. We have pointed out that each function $a \in \mathcal{A}$ can be approximated to accuracy $Cn^{-1/2}$ in $L_2(D)$ by using n elements of $\mathcal{D}$. Namely, any function $a \in \mathcal{A}$ can be approximated by a sum $1 + \sum_{R \in \Lambda} \chi_R$ with $\#(\Lambda) = n$ to accuracy $Cn^{-1/2}$ in $L_2(D)$. It follows that the dictionary widths defined in (1.36) satisfy

$$d_{n,n^2}(\mathcal{A})_{L_2(D)} \le Cn^{-1/2}, \quad n \ge 1. \tag{1.108}$$

One disadvantage in using the dictionary $\mathcal{D}$ when approximating the elements of $\mathcal{A}$ is that the dictionary elements themselves are not in $\mathcal{A}$. However, it is possible to introduce another dictionary $\mathcal{D}_0$ with n^2 functions that actually come from $\mathcal{A}$ and when using n-term approximation from $\mathcal{D}_0$ to approximate the elements of $\mathcal{A}$, it still achieves the bound $Cn^{-1/2}$ for error measured in $L_2(D)$. Namely, for each point $(i/n, j/n) \in D$, we associate the functions $\phi_{i,j}$ which is the characteristic of the region depicted in Figure 1.8.2 We let $\mathcal{D}_0 := \{\phi_{i,j},\ 1 \le i, j \le n$. It is easy to see that any $a \in \mathcal{A}$ can be approximated to accuracy $Cn^{-1/2}$ by $1 + \chi_S$, where S is the region under a piecewise linear function which always has slopes ± 1. Such a piecewise function can be written as a linear combination of n-terms from $\mathcal{D}_0$ (see Figure 1.8.2).

Given the above results for the dictionary n-width of $\mathcal{A}$, one expects correspondingly improved results for $d_{n,n^2}(\mathcal{U}_{\mathcal{A}})_{H_0^1(D)}$. Unfortunately, they do not follow from anything we know. The comparison (1.60) is to debilitating in this case since the factor n kills the decay rate $(n^{-1/2})$ in (1.108). We expect, however, that this is just a defect of our current state of knowledge and that the following problem will have a positive solution.

Open Problem: Find n^2 snapshots of $\mathcal{U}_{\mathcal{A}}$ such that any u_a can be approximated to accuracy $Cn^{-1/2}$ in $H_0^1(D)$ by using only n of these snapshots.

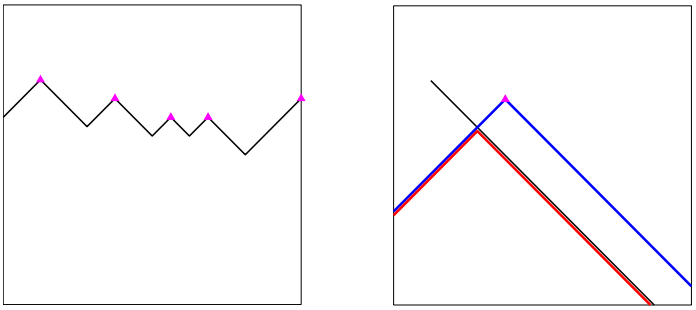

Figure 1.4. *On the left is a typical piecewise linear function with slopes* ± 1 *and on the right is a sample decomposition (the region below the red curve).*

Let us now turn our discussion to manifold widths. The above dictionary widths are not of the form of non linear approximation considered in the definition of the non linear manifold width $\delta_n(K)_{L_2(D)}$. This can be remedied, as described in [13], by using the famous Pontrjagin-Nöbling lemma on topological embeddings of complexes. We do not go into this in detail here but remark that it allows us to construct mappings b and M, of the form described above that achieve the same rate $O(n^{-1/2})$. In other words, we have

$$\delta_n(\mathscr{A})_{L_2(D)} \le Cn^{-1/2}. \tag{1.109}$$

With this result in hand, we can use our general theory to see that at least for certain right sides f, we have

$$\delta_n(\mathscr{U}_{\mathscr{A}}) \le \delta_n(\mathscr{A})_{L_2(D)} \le Cn^{-1/2}. \tag{1.110}$$

Indeed, this follows from (1.65) provided we show that $\mathscr{A}$ satisfies the uniqueness assumption.

The following simple argument (provided to me by Andrea Bonito) shows that the **Uniqueness Assumption** holds for our geometric class $\mathscr{A}$, whenever the right side f is nonzero almost everywhere. By a disjoint open finite covering (DOFC) of D, we mean a collection of open set D_j, $j = 1,...,K$ such that $D = \cup_{j=1}^K \overline{D_j}$, D_j open and $D_i \cap D_j = \emptyset$ for $i \neq j$. Define

$$\mathscr{A}_0 := \left\{ a = \sum_{j=1}^K c_j \chi_{D_j} : \ c_j \ge r > 0, \ j = 1,\dots,K \right\}.$$

Then, clearly $\mathscr{A} \subset \mathscr{A}_0$.

Now, for any $a \in \mathscr{A}_0$, define $u_a \in H_0^1(\Omega)$ as satisfying

$$\int_\Omega a \nabla u_a \cdot \nabla v = \int_\Omega f v, \qquad \forall v \in H_0^1(D).$$

Let $a \in \mathscr{A}_0$ and notice that for each set Ω_i of the (DOFC) and for each $x \in D_i$, there exists a ball $B(x,\delta)$ of radius $\delta > 0$ centered at x such that $B(x,\delta) \subset\subset D_i$. In this ball, a is constant and the interior regularity property for elliptic problem implies that $u_a \in$

$C^\infty(B(x,\delta))$. In particular a satisfies $-a\Delta u_a = f$ in $B(x,\delta)$, which in turn implies $\Delta u_a \neq 0$ a.e in D and

$$a = -f/\Delta u_a \qquad \text{a.e in } D.$$

Now, given a and $\hat{a}$ in $\mathscr{A}$ with $u_a = u_{\hat{a}} \in H_0^1(D)$ we realize using the above representation that $a = \hat{a}$ a.e.

While the above result is a nice theoretical consequence, it does not provide a reasonable numerical recipe for using non linear methods to solve the system of parametric pdes for the geometric model, even if used in conjunction with the n-term approximation from the dictionary of n^2 elements used to approximate $\mathscr{A}$. Indeed, given a query a, it would identify an $\hat{a}$ which is a good n-term approximation to a but ask to solve for $u_{\hat{a}}$. This would not allow the offline preparation of n^2 snapshots from which an n-term approximation would be constructed for u_a. This returns us to the Open Problem stated above (1.109).

Bibliography

[1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence Rates for Greedy Algorithms in Reduced Basis Methods*, SIAM J. Math. Anal., 43(2011), pp. 1457–1472. (Cited on p. 22)

[2] Peter Binev, W. Dahmen, and R. DeVore, *Adaptive Finite Element Methods with Convergence Rates*, Numerische Mathematik, 97(2004), pp. 219–268. (Cited on p. 5)

[3] Peter Binev, W. Dahmen, R. DeVore, and P. Petrushev, *Approximation Classes for Adaptive Methods*, Serdica Math. J., 28(2002), pp. 391–416. (Cited on p. 5)

[4] A. Bonito, R. DeVore, and R. Nochetto, *Adaptive Finite Element Methods for Elliptic Problems with Discontinuous Coefficients*, SIAM J. Numer. Anal. 51 (2013), pp. 3106–3134. (Cited on pp. 3, 4, 17)

[5] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme, and G. Turinici, *A Priori convergence of the greedy algorithm for the parameterized reduced basis*, M2AN Math. Model. Numer. Anal., 46(2012), pp. 595Ð-603. (Cited on pp. 20, 22)

[6] A. Cohen and R. DeVore, *Kolmogorov widths under holomorphic maps*, IMA Journal of Numerical Analysis, to appear. (Cited on pp. 13, 14)

[7] A. Cohen and R. DeVore, *Approximation of high dimensional parametric PDEs*, Acta Numerica, 24(2015), pp. 1–159. (Cited on p. 2)

[8] A. Cohen, R. DeVore and C. Schwab, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Foundations of Computational Mathematics, 10(2010), pp. 615–646. (Cited on p. 17)

[9] ———, *Analytic Regularity and Polynomial Approximation of Parametric Stochastic Elliptic PDEs*, Analysis and Applications, 9(2011), pp. 11–47. (Cited on pp. 13, 16, 17)

[10] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab, *Sparse Adaptive Taylor Approximation Algorithms for Parametric and Stochastic Elliptic PDEs*, M2AN Math. Model. Numer. Anal., 47(2013), pp. 253–280. (Cited on p. 27)

[11] S. Dahlke and R. DeVore, *Besov regularity for 2-D elliptic boundary value problems with variable coefficients*, Communication in PDEs, 22(1997) pp. 1–16. (Cited on p. 5)

[12] R. DeVore, R. Howard and C. Micchelli, *Optimal non-linear approximation*, Manuscripta Math., 63 (1989), pp. 469–478. 59. (Cited on pp. 9, 11)

[13] R. DeVore, G. Kyriazis, D. Leviatan, and V.M. Tikhomirov, *Wavelet compression and non linear n-widths*, Advances in Computational Math., 1(1993), pp. 197–214. (Cited on pp. 10, 29)

[14] R. DeVore, G. Petrova and P. Wojtaszczyk, *Greedy Algorithms for Reduced Bases in Banach Spaces*, Constructive Approximation, 37(2013), pp. 455-466. (Cited on pp. 22, 26)

[15] J.L. EFTANG, A.T. PATERA, AND E.M. RONQUIST, *An hp certified reduced basis method for parametrized elliptic partial differential equations*, SIAM Journal on Scientific Computing, 32(2010), pp. 3170–3200. (Cited on p. 27)

[16] J.L. EFTANG, D.J. KNEZEVIC, AND A.T. PATERA, *An hp certified reduced basis method for parametrized parabolic partial differential equations*, Mathematical and Computer Modelling of Dynamical Systems, 17(2011), pp. 395–422. (Cited on p. 27)

[17] B. HAASDONK, M. DIHLMANN, AND M. OHLBERGER, *A training set and multiple basis generation approach for parametrized model reduction based on adaptive grids in parameter space*, Mathematical and Computer Modelling of Dynamical Systems, 17(2011), pp. 423–442. (Cited on p. 27)

[18] D. JERRISON AND C.E. KENIG, *The inhomogeneous Dirichlet problem in Liptschitz domains*, J. Funct. Anal., 130(1995), pp. 161–219. (Cited on p. 4)

[19] S. KAULMANN, B. HAASDONK, B. MOITINHO DE ALMEIDA, J. P. B. DIEZ, *Online Greedy Reduced Basis Construction Using Dictionaries*, VI International Conference on Adaptive Modeling and Simulation (ADMOS 2013), pp. 365–376. (Cited on p. 27)

[20] G.G. LORENTZ, M. VON GOLITSCHEK, AND Y. MAKOVOZ, *Constructive Approximation: Advanced Problems*, Springer Verlag, New York, 1996. (Cited on p. 12)

[21] Y. MADAY, A.T. PATERA, AND G. TURINICI, *A priori convergence theory for reduced-basis approximations of single-parametric elliptic partial differential equations*, J. Sci. Comput., 17(2002), pp. 437–446. (Cited on pp. 20, 21)

[22] ———, *Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations*, C. R. Acad. Sci., Paris, Ser. I, Math., 335(2002), pp. 289–294. (Cited on pp. 20, 21)

[23] Y. MADAY, B. STAMM, *Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces*, SIAM Journal on Scientific Computing 35(2013), pp. 2417–2441. (Cited on p. 27)

[24] G. ROZZA, D.B.P. HUYNH, AND A.T. PATERA, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations Ñ application to transport and continuum mechanics*, Arch. Comput Method E, 15(2008), pp. 229–275. (Cited on p. 20)

[25] S. SEN, *Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems*, Numer. Heat Tr. B-Fund, 54(2008), pp. 369–389. (Cited on p. 20)

[26] V. TEMLYAKOV, *Non linear Kolmogorov widths*, Mathematical Notes, 63(1998), pp. 785–795. (Cited on pp. 10, 12)

[27] K. VEROY, C. PRUDÕHOMME, D. V. ROVAS, AND A. T. PATERA, *A Posteriori error bounds for reduced-basis approximation of parametrized noncoercive and non linear elliptic partial differential equations*, in: Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, 2003 pp. 2003–3847. (Cited on pp. 20, 21)

Index