Course Notes Bilbao 2013
Numerical Methods in High Dimensions
Lecture 4: Stochastic and Parametric Equations

Ronald DeVore

April 17, 2013

Abstract

One of the important application domains where high dimensional problems arise is the numerical treatment of stochastic and parametric partial differential equations. In attempting to treat these problems numerically, we shall engage several new concepts for high dimensions including Reduced Basis (often call Reduced Modelling) and Greedy Algorithms.

1 Elliptic PDEs

We turn now to high dimensional problems that arise in solving stochastic and parametric PDES. We restrict our attention to elliptic problems where the theory is most fully developed.

1.1 Elliptic equations: general principles

We consider the elliptic equation

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

in a bounded Lipschitz domain $D \subset \mathbb{R}^d$, where $f \in L_2(D)$. Here, $a = a(x)$ is a scalar function which is assumed to be in $L^\infty(D)$ and satisfy the ellipticity assumption: there exist $0 < r < R$ such that

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

We could also 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.

There is a rich theory for existence and uniqueness for equations of (1.1) which we briefly recall. Central to this theory is the Sobolev space $H_0^1(D, a)$ (called the energy space) which is a Hilbert space equipped with the energy norm

$$\|v\|_{H_0^1(D,a)} := \|a|\nabla v|\|_{L^2(D)}. \tag{1.3}$$

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

$$\|v\|_{L_2(D)} \leq C_D \|v\|_{H_0^1(D,a)}, \tag{1.4}$$

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

If $a, \tilde{a}$ both satisfy the ellipticity assumption, then the norm for $H_0^1(a)$ and $H_0^1(\tilde{a})$ are equivalent. If we take $a = 1$, we obtain the classical space $H_0^1(D, 1)$ which in going further is simply denote by $H_0^1 = H_0^1(D)$. The dual of $H_0^1(D)$ consists of all linear functionals 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

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

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

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

where the gradient ∇ is taken with respect to the x variable. 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 satifies 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 ellipticity constants r, R, then the solutions u and $\tilde{u}$ with identical right side f will satisfy

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

1.2 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. This makes the implementation of (1.8) useless since it requires exact matching of the discontinuities. 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 [2] in which L_∞ perturbation is replaced by L_q perturbation with $q < \infty$.

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

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

provided $\nabla u \in L_p(D)$. In order for (1.9) to be relevant we need that ∇u is in L_p for some $p > 2$. It is possible to show that for every Lipschitz domain D, there is always a range $2 \leq p \leq P$, with $2 < P$, for which one has

CONDP: *For each $f \in W^{-1}(L_p(\Omega))$, the solution $u = u_f$ satisfies*

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

with the constant C_p independent of f.
Thus, the perturbation result (1.9) can be applied with $q = \frac{2p}{p-2}$ provided $f \in W^{-1}(L_p(D)$, which is a rather minimal assumption on f. In the special case $a = 1$ (the case of Laplace's equation), the validity of **CONDP** is a well studied problem in Harmonic Analysis. It is known that for each Lipschitz domain D, there is a $P > 2$ which depends on D such that CONDP holds for all $2 \le p \le P$ (see for example Jerison and Kenig [9]). In fact, one has in this setting $P > 4$ when $d = 2$ and $P > 3$ when $d = 3$. One can then treat the case of general a in (1.9) by a perturbation argument (see [2] for details).

In summary, there is always a P such that whenever $p \in [2, P]$, and $f \in W^{-1}(L_p(D))$, then

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

where $q := 2p/(p-1)$. For convex domains D in two or three dimensions, it is known that one can take $P = \infty$ (see [10] and [11]).

2 Parametric elliptic equations

In some application domains such as design, optimal control, and shape optimization, we are not interested in solving (1.1) for just one diffusion coefficient a but rather a family $\mathcal{A}$ of coefficients. Typically, the diffusion coefficients depend on a large number (possibly infinite) of parameters. For simplicity of the discussion, we fix the right side f and denote by u_a the solution to (1.1) for the diffusion coefficient a. We are interested in fast numerical methods for capturing u_a, $a \in \mathcal{A}$. This is typically a high dimensional problem because the number of parameters is large. We give three examples, which will guide our discussion.

2.1 Affine Dependence

In this setting, we are given a linearly independent family of functions $\psi_j(x)$, $j = 1, 2, \ldots$ defined on the physical domain D. We let U be the unit cube in ℓ_∞. Hence, $y \in U$ means that $y = (y_1, y_2, \ldots)$ with $|y_j| \le 1$. For any such $y \in U$, we define

$$a(x, y) = \overline{a}(x) + \sum_{j \ge 1} y_j \psi_j(x), \tag{2.1}$$

so that $\mathcal{A} = \{a(x, y) : y \in U\}$ is our family of diffusion coefficients. Of course, we shall also need additional assumptions to guarantee that the series in (7.3) converges. A typical assumption is that the sequence $(\|\psi_j\|_{L_\infty(D)})$ is in ℓ_p for some $p \le 1$. Note that we can always rearrange the indicies so that this sequence is monotonically decreasing.

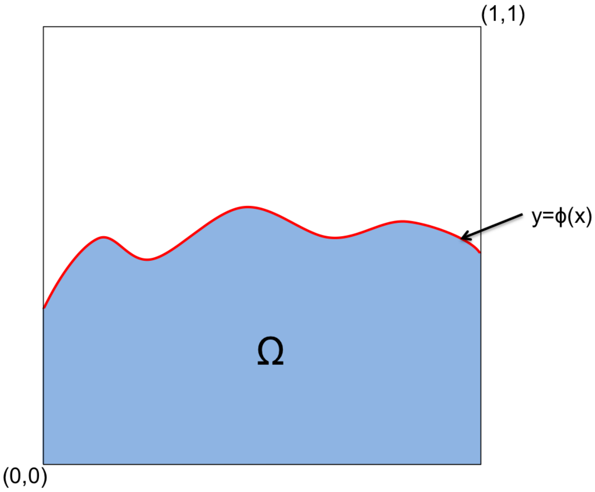

Figure 2.1: The region marked Ω corresponds to D_-.

2.2 A Geometrical Setting

Let $D = [0,1]^2$ for simplicity and let $\varphi(x)$, $x \in [0,1]$ be a $\mathrm{Lip}_M 1$ function 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 2.1). We consider diffusion coefficients

$$a_\phi(x) := \chi_{D_-} + 2\chi_{D_+}. \tag{2.2}$$

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

3 Comparisons between entropies and widths of $\mathcal{A}$ and $\mathcal{K}_{\mathcal{A}}$

In trying to understand how well numerical methods can perform in resolving $\mathcal{K}$, we follow our general principle of examining entropies and widths of $\mathcal{K}$. It turns out these are not so obvious because the solution operator is complicated. On the other hand, the widths and entropies of $\mathcal{A}$ are more transparent and so it is useful to have comparisons between the two. In this section, we shall discuss what is known in this regard.

3.1 Comparing entropies

One can utilize the perturbation results (1.9) to give a comparison of the entropies of the two classes $\mathcal{A}$ and $\mathcal{K} = \mathcal{K}_{\mathcal{A}}$. The results we derive below will show us that the entropy of these two sets are comparable and therefore the complexity of these sets are also comparable. Since the entropy of $\mathcal{A}$ is usually easier to compute, this will provide good information about $\mathcal{K}$.

We place ourselves in the following situation. We assume $f \in W^{-1}(L_p(D))$ for p in the range $[2, P]$ where **CONDP** is known to hold. Then we know that the perturbation result (1.11) holds for $q = 2p/(p-1)$. 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 $C\epsilon$ cover of $\mathcal{K}$ by the balls $B(u_{a_i}, C\epsilon)$ in the $H_0^1(D)$ topology. Therefore, we have

$$\epsilon_n(\mathcal{K}_{\mathcal{A}})_{H_0^1(D)} \le C\epsilon_n(\mathcal{A})_{L_q(D)}, \quad n \ge 1. \tag{3.1}$$

We next want to show that, in the case $q = 2$, by choosing f appropriately, we can reverse this last inequality and thereby see that the entropy of $\mathcal{K}$ is not noticeably better than that of $\mathcal{A}$. We illustrate how this argument would go in the one dimensional case in which case the PDE is simply

$$-[au']' = f, \quad u(0) = u(1) = 0. \tag{3.2}$$

If we denote by F any primitive of $-f$, then the solution u_a to (3.2) is any primitive G of F/a, provided both F and G are chosen so that u_a satisfies the boundary conditions. Since we are allowed to choose f, we are allowed to choose F subject to worrying about the boundary conditions imposed on u_a. We take F to be a smooth function which is $+1$ on $[1/6, 1/3]$, -1 on $[2/3, 5/6]$ and an odd function with respect to $x = 1/2$. We fix this F and consider only a which are even with respect to $x = 1/2$ in the following. Given any $a \in \mathcal{A}$, its restriction to $J := [1/6, 1/3]$ can always be extended to an even function $\bar{a}$ with respect to $x = 1/2$. For most parametric classes $\bar{a}$ is also in $\mathcal{A}$. We make this an assumption about $\mathcal{A}$:

Assumption 1: Whenever $a \in \mathcal{A}$, the function $\bar{a}$ defined above is always in $\mathcal{A}$.

By our construction, F/a is odd with respect to $x = 1/2$ and so $u_a = \int\limits_0^x [F(s)/a(s)]\, ds$ is the solution to (3.2). Note that returning to (3.2), 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 $a, \tilde{a}$

$$a - \tilde{a} = F/u_a' - F/u_{\tilde{a}}' = \frac{F}{u_a' u_{\tilde{a}}'}[u_{\tilde{a}}' - u_a']. \tag{3.3}$$

This gives the bound on J,

$$\|a - \tilde{a}\|_{L_2(J)} \le C\|u_a - u_{\tilde{a}}\|_{H_0^1(D)} \tag{3.4}$$

$$\epsilon_n(\mathcal{A})_{L_2(J)} \le C\epsilon_n(\mathcal{K})_{H_0^1(D)}, \quad n \ge 1. \tag{3.5}$$

Now, for most $\mathcal{A}$ one encounters in practice we can prove by simple rescaling that

Assumption 2: For the class $\mathcal{A}$, we have $\epsilon_n(\mathcal{A})_{L_2(J)} \sim \epsilon_n(\mathcal{A})_{L_2(D)}$

If both Assumptions 1 and 2 hold, we have

$$\epsilon_n(\mathcal{A})_{L_2(D)} \le C\epsilon_n(\mathcal{K})_{H_0^1(D)}, \quad n \ge 1, \tag{3.6}$$

and therefore we have reversed (3.1), in the case $q = 2$.

3.2 Comparing widths

One significant handicap to analyzing the Kolmogorov width of $\mathcal{K}$ is that we do not have any general comparison of the form $d_n(\mathcal{K})_{H_0^1(D)} \le Cd_n(\mathcal{A})_{L_q(D)}$ when the stability holds for L_q. It is an open question to establish properties of $\mathcal{A}$ under which such comparisons hold. Heuristically, this should be the case when the manifold $\mathcal{A}$ is smooth. We shall later give results that indicate that comparisons between widths of $\mathcal{A}$ and those of $\mathcal{K}$ can be obtained in the case of parametric models.

For nonlinear widths, the situation is much better and one can prove that for all $\mathcal{A}$ and q for which stability holds, we have

$$\delta_n(\mathcal{K})_{H_0^1(D)} \le C\delta_n(\mathcal{A})_{L_q(D)}, \tag{3.7}$$

provided the right side f is in $W^{-1}(L_p(\Omega))$ where p and q are related as before $q = \frac{2p}{p-2}$. We sketch how this is proved.

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

$$\sup_{a\in\mathcal{A}} \|a - M(b(a)\|_{L_q(D)} \le 2\epsilon. \tag{3.8}$$

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

$$\tilde{M}(\mathbf{z}) := u_{M(\mathbf{z})}. \tag{3.9}$$

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 $H_0^1(D)$.

We also need to construct a mapping from $\mathcal{K} \to \mathbb{R}^n$. We can write any $u \in \mathcal{K}$ as $u = u_a \in \mathcal{K}$. We would like to take $\tilde{b}(u) := b(a)$. This would work, but we face the problem that given $u_a \in \mathcal{K}$, there may be other $\tilde{a}$ such that $u_a = u_{\tilde{a}}$. That is, given $u \in \mathcal{K}$, the set $\mathcal{A}(u) := \{a \in \mathcal{A} :\ u_a = u\}$ contains more than one element. To get around this we assume we have a continuous selection that for each $u \in \mathcal{A}$, it assigns an $a(u) \in \mathcal{A}$.[1] Under this assumption, we can take $\tilde{b}(u) = b(a(u))$ which is obviously a continuous function from $\mathcal{K}$ into $\mathbb{R}^n$. Given $u \in \mathcal{K}$, we have

$$\|u - \tilde{M}(\tilde{b}(u))\|_{H_0^1(D)} = \|u_{a(u)} - u_{M(\tilde{b}(u))}\|_{H_0^1(D)} \le C\|a(u) - M(a(u))\|_{L_q(D)} \le 2C\epsilon \tag{3.10}$$

Since $\epsilon = \delta_n(\mathcal{A})_{L_q(D)}$, we have proven (3.7).

4 Numerical methods for parametric equations

Our main interest in this lecture is in numerical methods for solving a family of parametric equations. We wish to construct a numerical solver such that given a query $a \in \mathcal{A}$, it produces a function $\hat{u}_a$ which is a good approximation to u_a in the $H_0^1(D)$ norm. The standard method for solving such elliptic equations is the Galerkin method. Given a a linear space V_n of finite dimension n, it constructs $\hat{u}_a \in V_n$ as the solution to the discrete system of equations

$$\langle \hat{u}_a, v\rangle_a = \langle f, v\rangle, \quad \forall v \in V_n, \tag{4.1}$$

If we choose a basis $\varphi_1, \dots, \varphi_n$ 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

$$A\mathbf{c} = \mathbf{f} \tag{4.2}$$

[1] Originally, I thought I had a proof that such a continuous selection is always possible but I found a glitch in my argument so I make the existence of such a selection an assumption.

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 $f_i := \langle f, \varphi_i \rangle$, $i = 1, \dots, n$, 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:

$$\|u_a - \hat{u}_a\|_{H_0^1(D,a)}. \tag{4.3}$$

The central question we wish to engage is: what is a good choice for the finite dimensional space V_n?. Since we want V_n to be used for all $a \in \mathcal{A}$, it should be efficient at approximating all of the elements in $\mathcal{K} = \mathcal{K}_\mathcal{A}$. 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 $d_n(\mathcal{K})_{H_0^1(D)}$. Of course finding such a subspace may be difficult but it serves as a benchmark for the optimal performance we can expect.

Methods for finding a good subspace are known as *Reduced Basis Methods* (RBM) and are a well studied subject [3, 13, 14, 15, 16, 17]. 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 function $u_{a_i} \in \mathcal{K}$, $i = 1, \dots, n$. These functions are called *snapshots* of $\mathcal{K}$. The most popular method for finding these snapshot is a greedy algorithm which we now describe. While we are primarily interested in this algorithm in the case of a compact set $\mathcal{K}$ of a Hilbert space (in our case $\mathcal{K} = \mathcal{K}_\mathcal{A}$ 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 $\mathcal{F}$ be one of its compact subsets. For notational convenience only, we shall assume that the elements f of $\mathcal{F}$ satisfy $\|f\|_X \le 1$. We consider the following greedy algorithm for generating approximation spaces for $\mathcal{F}$. We first choose a function f_0 such that

$$\|f_0\| = \max_{f \in \mathcal{F}} \|f\|. \tag{4.4}$$

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

$$\operatorname{dist}(f_n, V_n)_X = \max_{f \in \mathcal{F}} \operatorname{dist}(f, V_n)_X, \tag{4.5}$$

and define

$$\sigma_n := \sigma_n(\mathcal{F})_X := \operatorname{dist}(f_n, V_n)_X := \sup_{f \in \mathcal{F}} \inf_{g \in V_n} \|f - g\|. \tag{4.6}$$

This greedy algorithm was introduced, for the case X is a Hilbert space in [13, 14]. This algorithm is very abstract. 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.

4.1 Weak greedy algorithm

We fix a constant $0 < \gamma \le 1$. At the first step of the algorithm, one chooses a function $f_0 \in \mathcal{F}$ such that

$$\|f_0\| \ge \gamma \sigma_0(\mathcal{F})_X := \max_{f \in \mathcal{F}} \|f\|. \tag{4.7}$$

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

$$\sigma_n(f)_X := \mathrm{dist}(f, V_n)_X. \tag{4.8}$$

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

$$\sigma_n(f_n)_X \geq \gamma \max_{f \in \mathcal{F}} \sigma_n(f)_X, \tag{4.9}$$

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, $(\sigma_n(\mathcal{F})_X)_{n\geq 0}$ is also monotone decreasing. Of course, neither the greedy algorithm or the weak greedy algorithm give a unique sequence $(f_n)_{n\geq 0}$, nor is the sequence $(\sigma_n(\mathcal{F})_X)_{n\geq 0}$ 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.

4.2 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 $\mathcal{F}$. For this purpose we would like to compare its performance with the best possible performance which is given by the Kolmogorov width $d_n(\mathcal{F})_X$ of $\mathcal{F}$. 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. A first result in this direction, in the case that X is a Hilbert space $\mathcal{H}$, was given in [3] where it was proved that

$$\sigma_n(\mathcal{F})_{\mathcal{H}} \leq C n 2^n d_n(\mathcal{F})_{\mathcal{H}}, \tag{4.10}$$

with C an absolute constant. While this is an interesting comparison, it is only useful if $d_n(\mathcal{F})_{\mathcal{H}}$ decays to zero faster than $n^{-1}2^{-n}$. Various improvements on (4.10) were given in [1], again in the Hilbert space setting. We mention two of these. It was shown that if $d_n(\mathcal{F})_{\mathcal{H}} \leq Cn^{-\alpha}$, $n = 1, 2, \dots$, then

$$\sigma_n(\mathcal{F})_{\mathcal{H}} \leq C'_\alpha n^{-\alpha}. \tag{4.11}$$

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 $d_n(\mathcal{F})_{\mathcal{H}} \leq Ce^{-cn^\alpha}$, $n = 1, 2, \dots$, then

$$\sigma_n(\mathcal{F})_{\mathcal{H}} \leq C'_\alpha e^{-c'_\alpha n^\beta}, \quad \beta = \frac{\alpha}{\alpha+1}, \quad n = 1, 2, \dots. \tag{4.12}$$

These results were improved in [8] 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

4.3 Results for a Banach space

The analysis of the greedy algorithm is quite simple and executed with elementary results from linear algebra. A core result to this analysis is the following lemma from [8].

Lemma 4.1 *Let $G = (g_{i,j})$ be a $K \times K$ lower triangular matrix with rows $\mathbf{g}_1, \dots, \mathbf{g}_K$, W be any m dimensional subspace of $\mathbb{R}^K$, and P be the orthogonal projection of $\mathbb{R}^K$ onto W. Then*

$$\prod_{i=1}^{K} g_{i,i}^2 \le \left\{ \frac{1}{m} \sum_{i=1}^{K} \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m \left\{ \frac{1}{K-m} \sum_{i=1}^{K} \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}, \tag{4.13}$$

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

Proof: We choose an orthonormal basis $\varphi_1, \dots, \varphi_m$ for the space W and complete it into an orthonormal basis $\varphi_1, \dots, \varphi_K$ 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

$$\prod_{j=1}^{m} \|\mathbf{c}_j\|_{\ell_2}^2 \le \left\{ \frac{1}{m} \sum_{j=1}^{m} \|\mathbf{c}_j\|_{\ell_2}^2 \right\}^m = \left\{ \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{K} \langle \mathbf{g}_i, \varphi_j \rangle^2 \right\}^m = \left\{ \frac{1}{m} \sum_{i=1}^{K} \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m. \tag{4.14}$$

Similarly,

$$\prod_{j=m+1}^{K} \|\mathbf{c}_j\|_{\ell_2}^2 \le \left\{ \frac{1}{K-m} \sum_{j=m+1}^{K} \|\mathbf{c}_j\|_{\ell_2}^2 \right\}^{K-m} = \left\{ \frac{1}{K-m} \sum_{i=1}^{K} \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}. \tag{4.15}$$

Now, Hadamard's inequality for the matrix C and relations (4.14) and (4.15) result in

$$(\det C)^2 \le \prod_{j=1}^{K} \|\mathbf{c}_j\|_{\ell_2}^2 \le \left\{ \frac{1}{m} \sum_{i=1}^{K} \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m \left\{ \frac{1}{K-m} \sum_{i=1}^{K} \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}. \tag{4.16}$$

The latter inequality and the fact that $\det G = \prod_{i=1}^{K} g_{i,i}$ and $|\det C| = |\det G|$ gives (4.13). □

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

Note that in general, the weak greedy algorithm does not terminate and we obtain an infinite sequence $f_0, f_1, f_2, \dots$. 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 $\sigma_N(\mathcal{F})_{\mathcal{H}} = 0$. 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,

$$f_n = P_{n+1} f_n = \sum_{j=0}^{n} a_{n,j} f_j^*, \quad a_{n,j} = \langle f_n, f_j^* \rangle, \ j \le n. \tag{4.17}$$

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 $\mathcal{F}$. The following two properties characterize any lower triangular matrix A generated by such a greedy algorithm. With the notation $\sigma_n := \sigma_n(\mathcal{F})_{\mathcal{H}}$, 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 (4.9). To see **P2**, we note that for $m \ge n$,

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

Remark 4.2 *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 $\mathcal{F}$ 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 4.3 *For the weak greedy algorithm with constant γ in a Hilbert space $\mathcal{H}$ and for any compact set $\mathcal{F}$, we have the following inequalities between $\sigma_n := \sigma_n(\mathcal{F})_{\mathcal{H}}$ and $d_n := d_n(\mathcal{F})_{\mathcal{H}}$, for any $N \ge 0$, $K \ge 1$, and $1 \le m < K$,*

$$\prod_{i=1}^{K} \sigma_{N+i}^2 \le \gamma^{-2K} \left\{ \frac{K}{m} \right\}^m \left\{ \frac{K}{K-m} \right\}^{K-m} \sigma_{N+1}^{2m} d_m^{2K-2m} \tag{4.18}$$

Proof: We consider the $K \times K$ matrix $G = (g_{i,j})$ which is formed by the rows and columns of A with indices from $\{N+1, \dots, N+K\}$. 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 $\operatorname{dist}(\mathcal{F}, \mathcal{H}_m) = d_m$. Then, $\operatorname{dist}(f_{N+i}, \mathcal{H}_m) \le d_m$, $i = 1, \dots K$. Let $\widetilde{W}$ be the linear space which is the restriction of $\mathcal{H}_m$ to the coordinates $N+1, \dots, N+K$. In general, $\dim(\widetilde{W}) \le m$.

Let W be an m dimensional space, $W \subset \operatorname{span}\{e_{N+1}, \dots, e_{N+K}\}$, 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,

$$\|P\mathbf{g}_i\|_{\ell_2} \le \|\mathbf{g}_i\|_{\ell_2} \le \sigma_{N+1}, \quad i = 1, \dots, K, \tag{4.19}$$

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

$$\|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2} \le \|\mathbf{g}_i - \widetilde{P}\mathbf{g}_i\|_{\ell_2} = \operatorname{dist}(\mathbf{g}_i, \widetilde{W}) \le \operatorname{dist}(f_{N+i}, \mathcal{H}_m) \le d_m, \quad i = 1, \dots, K. \tag{4.20}$$

It follows from Property **P1** that

$$\prod_{i=1}^{K} |a_{N+i,N+i}| \ge \gamma^K \prod_{i=1}^{K} \sigma_{N+i}. \tag{4.21}$$

We now apply Lemma 4.1 for this G and W, and use estimates (4.19), (4.20), and (4.21) to derive (4.18). The proof is completed. □

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

Corollary 4.4 *For the weak greedy algorithm with constant γ in a Hilbert space $\mathcal{H}$, we have the following:*

(i) *For any compact set $\mathcal{F}$ and $n \ge 1$, we have*

$$\sigma_n(\mathcal{F}) \le \sqrt{2}\gamma^{-1} \min_{1 \le m < n} d_m^{\frac{n-m}{n}}(\mathcal{F}). \tag{4.22}$$

In particular $\sigma_{2n}(\mathcal{F}) \le \sqrt{2}\gamma^{-1}\sqrt{d_n(\mathcal{F})}$, $n = 1, 2 \dots$.

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

(iii) *If $d_n(\mathcal{F}) \le C_0 e^{-c_0 n^\alpha}$, $n = 1, 2, \dots$, then $\sigma_n(\mathcal{F}) \le \sqrt{2C_0}\gamma^{-1}e^{-c_1 n^\alpha}$, $n = 1, 2 \dots$, where $c_1 = 2^{-1-2\alpha}c_0$,*

Proof: (i) We take $N = 0$, $K = n$ and any $1 \le m < n$ in Theorem 4.3, use the monotonicity of $(\sigma_n)_{n \ge 0}$ and the fact that $\sigma_0 \le 1$ to obtain

$$\sigma_n^{2n} \le \prod_{j=1}^{n} \sigma_j^2 \le \gamma^{-2n} \left\{\frac{n}{m}\right\}^m \left\{\frac{n}{n-m}\right\}^{n-m} d_m^{2n-2m}. \tag{4.23}$$

Since $x^{-x}(1-x)^{x-1} \le 2$ for $0 < x < 1$, we derive (4.22).

(ii) It follows from the monotonicity of $(\sigma_n)_{n \ge 0}$ and (4.18) for $N = K = n$ and any $1 \le m < n$ that

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

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

$$\sigma_{4s} \le \sqrt{2}\gamma^{-1}\sqrt{\sigma_{2s} d_s}. \tag{4.24}$$

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 (4.24), we have

$$\sigma_{4s} \leq \sqrt{2}\gamma^{-1}\sqrt{C_1(2s)^{-\alpha}}\sqrt{C_0 s^{-\alpha}} = \sqrt{2^{1-\alpha}C_0C_1}\gamma^{-1}s^{-\alpha}, \tag{4.25}$$

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

$$C_1 < 2^{3\alpha+1}\gamma^{-2}C_0 < 2^{5\alpha+1}\gamma^{-2}C_0, \tag{4.26}$$

which is the desired contradiction. If $M = 4s + q$, $q \in \{1, 2, 3\}$, then it follows from (4.25) 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

$$\sigma_{2n+1} \leq \sigma_{2n} \leq \sqrt{2}\gamma^{-1}\sqrt{d_n} \leq \sqrt{2C_0}\gamma^{-1}e^{-\frac{c_0}{2}n^\alpha} = \sqrt{2C_0}\gamma^{-1}e^{-c_0 2^{-1-\alpha}(2n)^\alpha}, \tag{4.27}$$

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 [8]) however there is some loss in the approximation rate. The precise results are as follows:

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

(ii) If for $\alpha > 0$, we have $d_n \leq C_0 n^{-\alpha}$, $n = 1, 2, \ldots$, then for any $0 < \beta < \min\{\alpha, 1/2\}$, we have $\sigma_n \leq C_1 n^{-\alpha+1/2+\beta}$, $n = 1, 2\ldots$, 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 $d_n \leq C_0 e^{-c_0 n^\alpha}$, $n = 1, 2, \ldots$, 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 [8]

4.4 Practical considerations in implementing 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 implementation of this algorithm is computationally not feasible, even in offline mode since it requires the 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 $\|f - P_n u_a\|_{H^{-1}(D)}$ which can be computed (since both f and $P_n u_a$ are available) without computing u_a. However, 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 the accuracy of the discretization we take for $\mathcal{A}$ will be inherited by $\mathcal{K}$ because of (1.8). If a discretization $\tilde{\mathcal{A}}$ of $\mathcal{A}$ has accuracy ϵ and the residual error

$$\max_{a\in\tilde{\mathcal{A}}} \|f - P_n u_a\|_{H^{-1}(D)} \geq 2\epsilon, \tag{4.28}$$

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 (4.28) is satisfied.

There are other issues, such as how fast one can compute the stiffness matrix for a given a, that will effect the performance of Reduced Basis Methods. The reader should check the literature for a discussion of this issue.

5 A priori guarantees

We can obtain an a priori guarantee of the numerical performance of Reduced Basis Methods based on the greedy algorithm provided we can bound the Kolmogorov width of $\mathcal{K}_{\mathcal{A}}$. We now discuss what is known in this regard for our two model classes of elliptic equations.

5.1 Affine model

We recall that for the affine model, we assume that

$$a(x,y) = \overline{a}(x) + \sum_{j\geq 1} y_j \psi_j(x), \tag{5.1}$$

where the y_j, $j = 1, \dots, d$, are parameters in $[-1, 1]$. We can always rearrange the indices so that the sequence $b_j := \|\psi_j\|_{L_\infty(D)}$, $j = 1, 2 \dots$, 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, mild 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

$$\sup_{y\in U} \|a(\cdot, y) - \sum_{j=1}^{n} y_j \psi_j\|_{L_\infty(D)} \leq \sum_{j=n+1}^{\infty} b_j \leq b_{n+1}^{1-p} \sum_{j=n+1}^{\infty} b_j^p \leq C n^{1-1/p}. \tag{5.2}$$

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

Note, however, we are not so much interested in approximating a which we know but rather the solution $u(x,y)$. We are therefore interested in seeing where these decay conditions on (b_j) translate into acompressible representation of $u(x,y)$. That this is indeed the case rests on analytic expansions of Banach spaced valued functions of an infinite number of variables, as we now describe.

Let $\mathcal{F}$ be the set of all sequences $\nu = (\nu_1, \nu_2, \dots)$ 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 [5], we showed the following theorem.

Theorem 5.1 *If $(b_j) \in \ell_p$ for some $p < 1$, then*

$$u(x,y) = \sum_{\nu\in\mathcal{F}} c_\nu(x) y^\nu, \tag{5.3}$$

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.

Remark: *This theorem shows that the compressibility of a in L_∞ translates into the same compressibility of u in $H_0^1(D)$.*

The line of reasoning for proving this theorem is the following. The mapping $y \to u(\cdot, y)$ 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

$$\int_D a(x,y)\nabla\partial_{y_j} u(x,y)\nabla v(x)dx + \int_D \psi_j(x)\nabla u(x,y)\nabla v(x)dx = 0. \tag{5.4}$$

One then shows that more generally,

$$\int_D a(x,y)\nabla\partial_y^\nu u(x,y)\nabla v(x)dx + \sum_{\{j:\,\nu_j\neq 0\}} \nu_j \int_D \psi_j(x)\nabla\partial_y^{\nu-e_j} u(x,y)\nabla v(x)dx = 0, \tag{5.5}$$

where e_j is the Kronecker sequence with value 1 at position j and 0 elsewhere. (5.5) is proved by induction on $|\nu|$ using the same idea as used in deriving (5.4). From (5.5) it is not difficult to prove

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

One now proves the representation (5.3) with $c_\nu(x) := \frac{D^\nu u(x,0)}{\nu!}$ (see [4] 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 $\mathcal{K}$. From the fact that $(\|c_\nu\|_{H_0^1(D)}) \in \ell_p$, one can use similar arguments to that in (5.2) to prove that one can take a set $\Lambda \subset \mathcal{F}$ with $\#(\Lambda) = n$ so that

$$\sup_{y\in U} \|u(\cdot, y) - \sum_{\nu\in\Lambda} c_\nu(x) y^\nu\|_{H_0^1(D)} \leq C n^{1-1/p}, \tag{5.6}$$

with an absolute constant C. This shows that the n-dimensional space $V := \text{span}\{c_\nu : \nu \in \Lambda\}$ approximates $\mathcal{K}$ with accuracy $Cn^{1-1/p}$ and therefore $d_n(\mathcal{K})_{H_0^1(D)} \leq Cn^{1-1/p}$. One important observation about this bound for the entropy is that we have broken the curse of dimensionality. Indeed, the parameters $y_1, y_2, \dots$ are infinite. In typical applications, the parameters are finite in number, say d, but then this result shows that the bound does not depend on d.

Given this bound for the entropy, we now know that the weak greedy algorithm for Reduced Basis, gives n snapshots with the same performance bound. We should point out that an alternative to this greedy algorithm, based on selecting index sets Λ, was given in [6] with the same performance bound.

5.2 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. Let us first consider $\mathcal{A}$ and try to understand its complexity. It makes no sense to consider the approximation of the functions $a \in \mathcal{A}$ 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 admissible range of q for the perturbation theorem is $2 \leq q \leq \infty$, see [2]. The best choice, for our purposes, is $q = 2$, since this is the weakest norm. Therefore, we concentrate on approximating the elements of $\mathcal{A}$ in the $L_2(D)$ norm in what follows.

We know from our general theory that if we measure the complexity of $\mathcal{A}$ and $\mathcal{K}$ in the sense of their entropy then (3.1) always holds. One can rather easily compute the entropy numbers of $\mathcal{A}$ in $L_q(D)$ for any $q \geq 2$. For $L_2(D)$, they satisfy

$$\epsilon_n(\mathcal{A})_{L_2(D)} \sim n^{-1/2}, \quad n \geq 1. \tag{5.7}$$

This means, we could compute the solution u_{a_i} for 2^n realizations of $a_i \in \mathcal{A}$ and then for any other $a \in \mathcal{A}$, $\|u_a - u_{a_i}\|_{H_0^1(D)} \leq Cn^{-1/2}$, for some value of i. This has huge offline cost but no online cost for resolving the parametric family $\mathcal{K}$

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

$$C_2 n^{-1/4} \leq d_n(\mathcal{A})_{L_2(D)} \leq C_1 n^{-1/4}, \quad n \geq 1. \tag{5.8}$$

To prove the upper estimate, we consider the dictionary $\mathcal{D}$ which consist of the functions χ_R, where $R = [(i-1)/n, i/n) \times [0, j/n]$, $1 \leq i, j \leq n$. Any function $a \in \mathcal{A}$ can be approximated by a sum $1+\sum_{R\in\Lambda}\chi_R$ with $\#(\Lambda)=n$ to accuracy $n^{-1/2}$ in $L_2(D)$. Since the space spanned by the $\mathcal{D}$ has dimension n^2, we obtain the upper estimate.

The lower estimate is a little more intricate. Let $V\subset L_2(D)$ be any fixed linear space of dimension $N\le n^2/2$ with n a two power and let $\varphi_1,\dots,\varphi_N$ be an orthonormal system for V. We assume $\operatorname{dist}(\mathcal{A},V)_{L_2(D)}\le\epsilon$ 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 $\Lambda\subset\{0,1,\dots,n-1\}$, the function $g_{j,\Lambda}:=j/n+\sum_{k\in\Lambda}\psi_k$ is in $\mathrm{Lip}_1 1$. The function

$$f_{j,\Lambda}:=a_{g_{j,\Lambda}}-a_{g_{j,\Lambda^c}} \tag{5.9}$$

can be approximated to accuracy 2ϵ by the space V. Each of these functions has support in the strip $j\le y\le j+1$ 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 $\Lambda=\{0,1,\dots,n-1\}$ and then the other $n-1$ choices according to the Haar patterns. In this way, we get n^2 orthogonal functions. We define the functions $h_1,\dots,h_{n^2}$ where each h_j is one of the functions $\sqrt{3n}f_{i,\Lambda}$ with the n^2 different choices of these function in (5.9). 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 $Nn^{-2}\le 1/2$. 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 $2\epsilon\sqrt{3n}\ge 1/\sqrt{2}$. In other words, $\epsilon\ge[2\sqrt{6}]^{-1}n^{-1/2}$. Since the only restriction on N is that $N\le n^2/2$, we obtain

$$d_{n^2}(\mathcal{A})_{L_2(D)}\ge Cn^{-1/2},\quad n\ge 1, \tag{5.10}$$

with C an absolute constant. The lower bound in (5.8) follows.

The above results describe how well we can approximate $\mathcal{A}$ and say nothing about approximating $\mathcal{K}$. Indeed, we have no direct estimate for $d_n(\mathcal{K})_{H_0^1(D)}$ in terms of $d_n(\mathcal{A})_{L_2(D)}$. So it is an open problem whether the results on the n-width of $\mathcal{A}$ can be transfered into results for the n-width of $\mathcal{K}$. This is actually a part of a more general question:

Open Problem: When can bounds on the n widths of $\mathcal{K}$ in $H_0^1(D)$, be derived from bounds on $d_n(\mathcal{A})_{L_2(D)}$?

6 Nonlinear methods in reduced bases

Our experience from the first lecture suggest that there could be a large benefit to using nonlinear methods of approximation in reduced basis. This idea is being employed by several researchers but as of yet there is no theory that demonstrate advantages of this approach. Here we wish to give some heuristic discussion of this topic.

For the **Affine Model Class**, there seems to be no advantage in using nonlinear methods since the manifold $\mathcal{K}_{\mathcal{A}}$ is provably smooth. On the other hand, the case of the **Geometric**

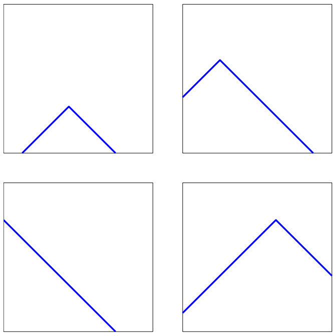

Figure 6.2: The basis functions $\phi_{i,j}$ with vertex $(i/n, j/n)$. The line segments have slope ±1.

Model seems ripe for the exploitation of nonlinear methods. We consider only this geometric example in what follows in this section. We have seen for example that the linear Kolmogorov widths satisfy

$$d_n(\mathcal{A})_{L_2(D)} \geq Cn^{-1/4}, \quad n \geq 1. \tag{6.1}$$

This is a good indication that the same lower bound holds for the widths of $\mathcal{K}$ in $H_0^1(D)$. On the other hand, we know that the entropy numbers of $\mathcal{K}$ behave like $n^{-1/2}$ which indicates nonlinear methods should provide this same rate of approximation.

Let us begin with our usual strategy of first trying to understand the nonlinear widths of $\mathcal{A}$ in $L_2(D)$. We have already discussed he dictionary $\mathcal{D}$ which consist of the n^2 functions χ_R, where $R = [(i-1)/n, i/n) \times [0, j/n]$, $1 \leq i, j \leq n$. We have pointed out that the function $\chi_D$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 $L_2(D)$.

We note, however, that this form of nonlinear approximation is not of the form considered in the definition of the nonlinear manifold width $\delta_n(\mathcal{K})_{L_2(D)}$. This can be remedied, as described in [7], 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 in the first lecture, that achieve the same rate $O(n^{-1/2})$. In other words, we have

$$\delta_n(\mathcal{A})_{L_2(D)} \leq Cn^{-1/2}. \tag{6.2}$$

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 6.2. We let $\mathcal{D}_0 := \{\phi_{i,j},\ 1 \leq i, j \leq n$. It is easy to

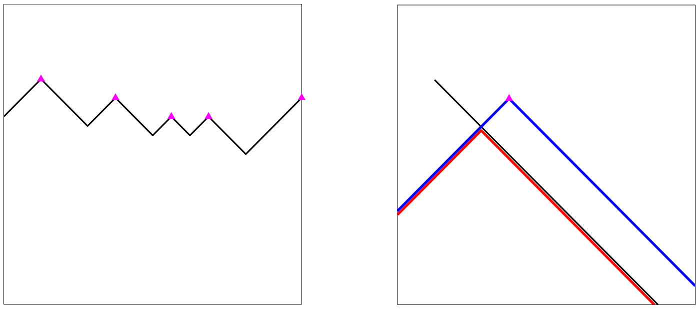

Figure 6.3: On the left is a typical piecewise linear function with slopes ±1 and on the right is a sample decomposition (the region below th

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 6.3.)

Of course, all of the discussion above is for $\mathcal{A}$ and does not carry over immediately to $\mathcal{K}$. **Open Problem:** Find n^2 snapshots of $\mathcal{K}$ such that any a can be approximated to accuracy $Cn^{-1/2}$ by using only n of these snapshots.

7 Stochastic equations

So far, we have only discussed parametric elliptic equations. There is a standard way to convert stochastic elliptic equation into parametric ones where the above results can be applied. We describe this briefly in the case of stochastic equations where the diffusion coefficient $a = a(x, \omega)$ is now a real valued random field on some probability space (Ω, Σ, P) but f remains a deterministic function. The solution $u = u(x, \omega)$ is now a random field associated to the same probability space. Stochasticity describes the uncertainty in the diffusion coefficient a. In order to ensure uniform ellipticity, one assumes

Assumption S: *There exist constants* $0 < a_{\min} \leq a_{\max}$ *such that*

$$a_{\min} \leq a(x, \omega) \leq a_{\max}, \quad (x, \omega) \in D \times \Omega. \tag{7.1}$$

There are two general numerical approaches to stochastic elliptic PDEs: Monte-Carlo (MC) methods and deterministic methods.

Monte-Carlo (MC) methods: These methods approximate quantities such as the mean ($\overline{u}(x) := \mathbb{E}(u(x)) = \int\limits_{\Omega} u(x, \omega) dP(\omega)$) or higher moments of u. Here and later $\mathbb{E}$ represents expectation. One takes n independent draws of a and computes the solution u_i, $i = 1, \dots, N$, corresponding to each of these draws and then uses the u_i to estimate the quantities of interest. For example, using standard stochastic estimates known as the law of large numbers, one proves that the average $\overline{u}_n := \frac{1}{n}\sum_{i=1}^n u_i$ gives an estimate in expectation

$$\mathbb{E}(\|\overline{u} - \overline{u}_n\|_V) \le (\mathbb{E}(\|u\|^2_{H^1_0(D)}))^{1/2} n^{-\frac{1}{2}} \tag{7.2}$$

i.e. Monte-Carlo approximations with N samples converge with rate $Cn^{-1/2}1/2$ in expectation provided that the solution u as a V-valued random function has finite second moments. Unfortunately, the rate $Cn^{-1/2}$ cannot be improved for MC. Similar bounds exist for other sochastic moments of $u(\cdot, \omega)$.

In practice, the u_i are computed approximately by space discretization, for example by the finite element method. But we will leave this issue aside in this lecture and instead focus on whether other methods could potentially outperform Monte-Carlo. Our benchmark is n which is the number of times we need to solve a corresponding elliptic equation.

Deterministic methods: These have been studied for several decades. In contrast to MC, these methods take advantage of the smooth dependence of u on a. We will consider the *spectral approach* which is based on the so-called Wiener generalized polynomial chaos expansion. The first step consists in representing a by a sequence of scalar random variables $(y_j)_{j\ge1}$, usually obtained through a decomposition of the oscillation $a - \overline{a}$ into an orthogonal basis $(\psi_j)_{j\ge1}$ of $L_2(D)$:

$$a(x,\omega) = \overline{a}(x) + \sum_{j\ge1} y_j(\omega)\psi_j(x). \tag{7.3}$$

Here, the reader can think of $\{\psi_j\}$ as his favorite basis, for example a wavelet basis or Fourier basis. In some approaches one tries to find the Karhunen-Loéve basis for the empirical process.

The solution is now viewed as a function $u(x,y)$ where $x \in D$ is the space variable and $y = (y_j)_{j\ge1}$ is a vector of "stochastic variables", and the objective is to compute a numerical approximation to $u(x,y)$. Any such approximation would give us access to all information about the solution u. Note that

$$y_j := \|\psi_j\|^{-2}_{L_2(D)} \int\limits_D (a - \overline{a})\psi_j, \quad j = 1, 2, \dots . \tag{7.4}$$

Of course, for each draw $\omega \in \Omega$, y is just a sequence of real numbers. So in the end we can consider parametric problems for real sequences y.

Up to a renormalization of the basis functions ψ_j, we may assume without loss of generality that for all $j \ge 1$ the random variables y_j are such that $\|y_j\|_{L_\infty(\Omega)} = 1$. Up to a change of the definition of a on a set of measure zero in Ω this is equivalent to

$$\sup_{\omega\in\Omega} |y_j(\omega)| = 1. \tag{7.5}$$

Now, we associate to the stochastic equation the parametric equation with diffusion coefficient

$$a(x,y) := \overline{a}(x) + \sum_{j=1}^{\infty} y_j\psi_j, \quad y_j \in [-1,2], \ j = 1, 2, \dots . \tag{7.6}$$

We can then apply the techniques we have developed for parametric equations and obtain methods to solve the stochastic equations uniformly in Ω. Notice that this leads us to the affine model we have heavily discussed for parametric equations. For this approach to be successful, we need that the deterministic expansion of $a(x,y)$ satisfy the assumption like $(\|\psi_j\|_{L_\infty}) \in \ell_p$. If this is the case with p sufficiently small then this breaks the barrier of $O(n^{-1/2})$, that occurs in Monte Carlo methods because of the law of large numbers. We do not go further into this here but refer the reader to the paper [4] and the references therein where these issues are discussed in detail.

References

[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), 1457–1472.

[2] A. Bonito, R. DeVore, and R. Nochetto, *Adaptive Finite Element Methods for Elliptic Problems with Discontinuous Coefficients*, preprint

[3] 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), 595-603.

[4] 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), 615–646.

[5] A. Cohen, R. DeVore and C. Schwab, *Analytic Regularity and Polynomial Approximation of Parametric Stochastic Elliptic PDEs*, Analysis and Applications, **9**(2011), 11–47.

[6] 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), 253–280.

[7] R. DeVore, G. Kyriazis, D. Leviatan, and V.M. Tikhomirov, *Wavelet compression and nonlinear n-widths*, Advances in Computational Math., **1**(1993) 197–214.

[8] R. DeVore, G. Petrova and P. Wojtaszczyk, *Greedy Algorithms for Reduced Bases in Banach Spaces*, Constructive Approximation, to appear.

[9] D. Jerrison and C.E. Kenig, *The inhomogeneous Dirichlet problem in Liptschitz domains* , J. Funct. Anal., **130**(1995), 161–219.

[10] P. Grisvard, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics, vol. 24, Pitman (Advanced Publishing Program), Boston, MA, 1985

[11] V. Maz'ya and J. Rossmann, *Elliptic equations in polyhedral domains*, Mathematical Surveys and Monographs, vol. 162, American Mathematical Society, Providence, RI, 2010.

[12] G.G. Lorentz, M. von Golitschek, and Y. Makovoz, *Constructive Approximation: Advanced Problems*, Springer Verlag, New York, 1996.

[13] 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), 437–446.

[14] Y. Maday, A. T. Patera, and G. Turinici, *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), 289–294.

[15] 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 application to transport and continuum mechanics*, Arch. Comput Method E, **15**(2008), 229–275.

[16] S. Sen, *Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems*, Numer. Heat Tr. B-Fund, **54**(2008), 369–389.

[17] K. Veroy, C. Prudhomme, D. V. Rovas, and A. T. Patera, *A Posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*, in: Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, 2003, Paper 2003-3847.