Greedy Algorithms for Reduced Bases in Banach Spaces*

Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk

April 5, 2012

Abstract

Given a Banach space X and one of its compact sets $\mathcal{F}$, we consider the problem of finding a good n dimensional space $X_n \subset X$ which can be used to approximate the elements of $\mathcal{F}$. The best possible error we can achieve for such an approximation is given by the Kolmogorov width $d_n(\mathcal{F})_X$. However, finding the space which gives this performance is typically numerically intractable. Recently, a new greedy strategy for obtaining good spaces was given in the context of the reduced basis method for solving a parametric family of PDEs. The performance of this greedy algorithm was initially analyzed in [2] in the case $X = \mathcal{H}$ is a Hilbert space. The results of [2] were significantly improved on in [1]. The purpose of the present paper is to give a new analysis of the performance of such greedy algorithms. Our analysis not only gives improved results for the Hilbert space case but can also be applied to the same greedy procedure in general Banach spaces.

Key words and phrases: greedy algorithms, convergence rates, reduced basis, general Banach space

AMS Subject Classification: 41A46, 41A25, 46B20, 15A15

1 Introduction

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 $\|f\|_X \leq 1$. We consider the following greedy algorithm for generating approximation spaces for $\mathcal{F}$. We first choose a function f_0 such that

$$\|f_0\| = \max_{f \in \mathcal{F}} \|f\|. \tag{1.1}$$

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

$$\mathrm{dist}(f_n, V_n)_X\| = \max_{f \in \mathcal{F}} \mathrm{dist}(f, V_n)_X, \tag{1.2}$$

and define

$$\sigma_n := \sigma_n(\mathcal{F})_X := \mathrm{dist}(f_n, V_n)_X := \sup_{f \in \mathcal{F}} \inf_{g \in V_n} \|f - g\|. \tag{1.3}$$

This greedy algorithm was introduced, for the case X is a Hilbert space, in the reduced basis method [5, 6] for solving a family of PDEs. Certain variants of this algorithm, known as *weak greedy algorithms*, described below, are now numerically implemented with great success in the reduced basis method. Our interest in this paper will be in the approximation properties of this algorithm and its weak variant.

We are interested in how well the space V_n approximates the elements of $\mathcal{F}$ and for this purpose we compare its performance with the best possible performance which is given by the Kolmogorov width $d_n(\mathcal{F})_X$ of $\mathcal{F}$ defined by

$$d_n := d_n(\mathcal{F})_X := \inf_{Y} \sup_{f \in \mathcal{F}} \ \mathrm{dist}(f, Y)_X, \tag{1.4}$$

where the infimum is taken over all n dimensional subspaces Y of X. We refer the reader to [4] for a general discussion of Kolmogorov widths. We also define

$$d_0 := d_0(\mathcal{F})_X := \max_{f \in \mathcal{F}} \|f\| = \sigma_0(\mathcal{F})_X,$$

which corresponds to approximating by zero dimensional spaces.

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 [2] where it was proved that

$$\sigma_n(\mathcal{F})_{\mathcal{H}} \leq C n 2^n d_n(\mathcal{F})_{\mathcal{H}}, \tag{1.5}$$

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

$$\sigma_n(\mathcal{F})_{\mathcal{H}} \leq C'_\alpha n^{-\alpha}. \tag{1.6}$$

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 $d_n(\mathcal{F})_{\mathcal{H}} \leq C e^{-c n^\alpha}$, $n = 1, 2, \ldots$, then

$$\sigma_n(\mathcal{F})_{\mathcal{H}} \leq C'_\alpha e^{-c'_\alpha n^\beta}, \quad \beta = \frac{\alpha}{\alpha + 1}, \quad n = 1, 2, \ldots. \tag{1.7}$$

In numerical implementations, the greedy algorithm is too demanding since at each iteration it requires finding an element from $\mathcal{F}$ which is at furthest distance from V_n. To circumvent this difficulty, one modifies the algorithm as follows. We fix a constant $0 < \gamma \leq 1$. At the first step of the algorithm, one chooses a function $f_0 \in \mathcal{F}$ such that

$$\|f_0\| \geq \gamma \sigma_0(\mathcal{F})_X.$$

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

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

is the best approximation error to f from V_n we now choose $f_n \in \mathcal{F}$ such that

$$\sigma_n(f_n)_X \geq \gamma \max_{f \in \mathcal{F}} \sigma_n(f)_X, \tag{1.8}$$

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.

In the present paper, we shall first prove a lemma that we use in our new analysis of the weak greedy algorithm. This new analysis gives a significant improvement of the previous results. We mention two of these:

The first, given in Corollary 3.3, is that

$$\sigma_{2n}(\mathcal{F})_{\mathcal{H}} \leq \sqrt{2} \gamma^{-1} \sqrt{d_n(\mathcal{F})_{\mathcal{H}}}, \quad n = 1, 2, \ldots. \tag{1.9}$$

This is the first direct comparison between $(\sigma_n(\mathcal{F})_{\mathcal{H}})_{n \geq 0}$ and $(d_n(\mathcal{F})_{\mathcal{H}})_{n \geq 0}$ for the special case X is a Hilbert space $\mathcal{H}$, which guarantees a specific rate of decay for $(\sigma_n(\mathcal{F})_{\mathcal{H}})_{n \geq 0}$ without any assumption of a decay rate for $(d_n(\mathcal{F})_{\mathcal{H}})_{n \geq 0}$. Notice that, in particular, this allows one to improve the sub-exponential results mentioned earlier (see Corollary 3.3).

The second part of our paper analyzes the performance of the greedy algorithm in a general Banach space. We prove estimates for the decay of $(\sigma_n(\mathcal{F})_X)_{n \geq 0}$ similar to those in the Hilbert space case, except that there is a loss of the order $O(\sqrt{n})$. We give examples which show that this loss in essence cannot be removed. However, our results for a general Banach space are still not definitive. For example, we have no result of the form (1.9) because of the $\sqrt{n}$ factor that appears in our Banach space results.

2 Main lemma

In this section, we shall prove a lemma for matrices that we employ in our analysis of weak greedy algorithms in both Hilbert and Banach spaces.

Lemma 2.1 *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*

$$\prod_{i=1}^{K} g_{i,i}^2 \leq \left\{ \frac{1}{m} \sum_{i=1}^{K} \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m \left\{ \frac{1}{K-m} \sum_{i=1}^{K} \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}, \tag{2.1}$$

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

$$\prod_{j=1}^m \|\mathbf{c}_j\|_{\ell_2}^2 \leq \left\{ \frac{1}{m} \sum_{j=1}^m \|\mathbf{c}_j\|_{\ell_2}^2 \right\}^m = \left\{ \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^K \langle \mathbf{g}_i, \varphi_j \rangle^2 \right\}^m = \left\{ \frac{1}{m} \sum_{i=1}^K \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m. \quad (2.2)$$

Similarly,

$$\prod_{j=m+1}^K \|\mathbf{c}_j\|_{\ell_2}^2 \leq \left\{ \frac{1}{K-m} \sum_{j=m+1}^K \|\mathbf{c}_j\|_{\ell_2}^2 \right\}^{K-m} = \left\{ \frac{1}{K-m} \sum_{i=1}^K \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}. \quad (2.3)$$

Now, Hadamard's inequality for the matrix C and relations (2.2) and (2.3) result in

$$(\det C)^2 \leq \prod_{j=1}^K \|\mathbf{c}_j\|_{\ell_2}^2 \leq \left\{ \frac{1}{m} \sum_{i=1}^K \|P\mathbf{g}_i\|_{\ell_2}^2 \right\}^m \left\{ \frac{1}{K-m} \sum_{i=1}^K \|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2}^2 \right\}^{K-m}. \quad (2.4)$$

The latter inequality and the fact that $\det G = \prod_{i=1}^K g_{i,i}$ and $|\det C| = |\det G|$ gives (2.1). $\quad\square$

3 A new analysis for the weak greedy algorithm in a Hilbert space

The purpose of this section is to obtain new results for the performance of the weak greedy algorithm in a Hilbert space that considerably improve on the analysis in [2] and [1]. This will be accomplished by making a finer comparison between $(\sigma_n(\mathcal{F})_{\mathcal{H}})_{n\geq 0}$ and $(d_n(\mathcal{F})_{\mathcal{H}})_{n\geq 0}$ than those given in [1]. We assume throughout this section that $X = \mathcal{H}$ is a Hilbert space and follow the notation from [1].

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 $\sigma_N(\mathcal{F})_{\mathcal{H}} = 0$. By $(f_n^*)_{n\geq 0}$ we denote the orthonormal system obtained from $(f_n)_{n\geq 0}$ 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,

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

4

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 adhere to this assumption throughout this section of the paper.

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 $\gamma \sigma_n \leq |a_{n,n}| \leq \sigma_n$.

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

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 (1.8). To see **P2**, we note that for $m \geq n$,

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

Remark 3.1 *If A is any matrix satisfying **P1** and **P2** with $(\sigma_n)_{n \geq 0}$ 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.*

The matrix representation A of the weak greedy algorithm was the basis of the analysis given in [1] and will also be critical in the proof of the next theorem.

Theorem 3.2 *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 $N \geq 0$, $K \geq 1$, and $1 \leq m < K$,*

$$\prod_{i=1}^{K} \sigma_{N+i}^2 \leq \gamma^{-2K} \left\{ \frac{K}{m} \right\}^m \left\{ \frac{K}{K-m} \right\}^{K-m} \sigma_{N+1}^{2m} d_m^{2K-2m} \tag{3.1}$$

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 $\{N+1, \ldots, N+K\}$. Each row $\mathbf{g}_i$ is the restriction of f_{N+i} to the coordinates $N+1, \ldots, N+K$. Let $\mathcal{H}_m$ be the m-dimensional Kolmogorov subspace of $\mathcal{H}$ for which $\mathrm{dist}(\mathcal{F}, \mathcal{H}_m) = d_m$. Then, $\mathrm{dist}(f_{N+i}, \mathcal{H}_m) \leq d_m$, $i = 1, \ldots K$. Let $\widetilde{W}$ be the linear space which is the restriction of $\mathcal{H}_m$ to the coordinates $N+1, \ldots, N+K$. In general, $\dim(\widetilde{W}) \leq m$. Let W be an m dimensional space, $W \subset \mathrm{span}\{e_{N+1}, \ldots, e_{N+K}\}$, 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,

$$\|P\mathbf{g}_i\|_{\ell_2} \leq \|\mathbf{g}_i\|_{\ell_2} \leq \sigma_{N+1}, \quad i = 1, \ldots, K, \tag{3.2}$$

5

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

$$\|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2} \le \|\mathbf{g}_i - \widetilde{P}\mathbf{g}_i\|_{\ell_2} = \mathrm{dist}(\mathbf{g}_i, \widetilde{W}) \le \mathrm{dist}(f_{N+i}, \mathcal{H}_m) \le d_m, \quad i = 1, \dots, K. \quad (3.3)$$

It follows from Property **P1** that

$$\prod_{i=1}^{K} |a_{N+i,N+i}| \ge \gamma^K \prod_{i=1}^{K} \sigma_{N+i}. \quad (3.4)$$

We now apply Lemma 2.1 for this G and W, and use estimates (3.2), (3.3), and (3.4) to derive (3.1). The proof is completed. $\qquad\square$

We next record some special cases of Theorem 3.2.

Corollary 3.3 *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*

$$\sigma_n(\mathcal{F}) \le \sqrt{2}\gamma^{-1} \min_{1 \le m < n} d_m^{\frac{n-m}{n}}(\mathcal{F}). \quad (3.5)$$

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 3.2, use the monotonicity of $(\sigma_n)_{n \ge 0}$ and the fact that $\sigma_0 \le 1$ to obtain

$$\sigma_n^{2n} \le \prod_{j=1}^{n} \sigma_j^2 \le \gamma^{-2n} \left\{\frac{n}{m}\right\}^m \left\{\frac{n}{n-m}\right\}^{n-m} d_m^{2n-2m}. \quad (3.6)$$

Since $x^{-x}(1-x)^{x-1} \le 2$ for $0 < x < 1$, we derive (3.5).
(ii) It follows from the monotonicity of $(\sigma_n)_{n \ge 0}$ and (3.1) 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

$$\sigma_{4s} \le \sqrt{2}\gamma^{-1}\sqrt{\sigma_{2s}d_s}. \quad (3.7)$$

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 (3.7), we have

$$\sigma_{4s} \le \sqrt{2}\gamma^{-1}\sqrt{C_1(2s)^{-\alpha}}\sqrt{C_0 s^{-\alpha}} = \sqrt{2^{1-\alpha}C_0 C_1}\gamma^{-1}s^{-\alpha}, \quad (3.8)$$

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

$$C_1(4s)^{-\alpha} < \sigma_{4s} \le \sqrt{2^{1-\alpha}C_0 C_1}\gamma^{-1}s^{-\alpha},$$

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 (3.8) and the monotonicity of $(\sigma_n)_{n\ge 0}$ that

$$C_1 2^{-3\alpha}s^{-\alpha} = C_1 2^{-\alpha}(4s)^{-\alpha} < C_1(4s + q)^{-\alpha} < \sigma_{4s+q} \le \sigma_{4s} \le \sqrt{2^{1-\alpha}C_0 C_1}\gamma^{-1}s^{-\alpha}.$$

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

$$\sigma_{2n+1} \le \sigma_{2n} \le \sqrt{2}\gamma^{-1}\sqrt{d_n} \le \sqrt{2C_0}\gamma^{-1}e^{-\frac{c_0}{2}n^{\alpha}} = \sqrt{2C_0}\gamma^{-1}e^{-c_0 2^{-1-\alpha}(2n)^{\alpha}}, \qquad (3.9)$$

from which (iii) easily follows. $\qquad\square$

Remark 3.4 *Note that one can obtain a better constant c_1 in* (iii) *if the minimum in* (3.5) *is computed. Namely, this gives*

$$\sigma_{2n} \le \sqrt{2}\gamma^{-1}C_0 \min_{1\le m<n} e^{-c_0 m^{\alpha}\frac{(n-m)}{n}} = \sqrt{2}\gamma^{-1}C_0 e^{-c_0 n^{\alpha}\{\max_{1\le m<n}\left(\frac{m}{n}\right)^{\alpha}\left(1-\frac{m}{n}\right)\}}.$$

Then, using the fact that $x^{\alpha}(1-x)$, $0 < x < 1$, has a maximum at $\frac{\alpha}{\alpha+1}$ results in a better constant.

4 Bounds for the greedy algorithm in Banach spaces

We will now derive bounds for the performance of the weak greedy algorithm in a general Banach space X. In this section, we will use the abbreviation $\sigma_n := \sigma_n(\mathcal{F})_X$ and $d_n := d_n(\mathcal{F})_X$. As in the Hilbert space case, we associate with the greedy procedure a lower triangular matrix $A = (a_{i,j})_{i,j=0}^{\infty}$ in the following way. For each $j = 0, 1, \ldots$, we let $\lambda_j \in X^*$ be the linear functional of norm one that satisfies

$$\text{(i) } \lambda_j(V_j) = 0, \quad \text{(ii) } \lambda_j(f_j) = \text{dist}(f_j, V_j)_X. \qquad (4.1)$$

The existence of such a functional is a simple consequence of the Hahn-Banach theorem (see [3, Chapt. IV, Cor.14.13]). We let A be the matrix with entries

$$a_{i,j} = \lambda_j(f_i).$$

From (ii) of (**??**) Its diagonal elements $a_{j,j}$ satisfy the inequality

$$\gamma \sigma_j \le a_{j,j} = \text{dist}(f_j, V_j)_X = \sigma_j, \tag{4.2}$$

because of the weak greedy selection property (1.8). Also, each entry $a_{i,j}$ satisfies

$$|a_{i,j}| = |\lambda_j(f_i)| = |\lambda_j(f_i - g)| \le \|\lambda_j\|_{X^*}\|f_i - g\| = \|f_i - g\|, \quad j < i,$$

for every $g \in V_j$, since $\lambda_j(V_j) = 0$. Therefore we have

$$|a_{i,j}| \le \text{dist}(f_i, V_j) \le \sigma_j, \quad j < i. \tag{4.3}$$

Theorem 4.1 *For the weak greedy algorithm with constant γ in a Banach space X and for any compact set $\mathcal{F}$ contained in the unit ball of X, we have the following inequalities between $\sigma_n := \sigma_n(\mathcal{F})_X$ and $d_n := d_n(\mathcal{F})_X$: for any $N \ge 0$, $K \ge 1$, and $1 \le m < K$,*

$$\prod_{i=1}^{K} \sigma_{N+i}^2 \le 2^K K^{K-m} \gamma^{-2K} \left\{ \sum_{i=1}^{K} \sigma_{N+i}^2 \right\}^m d_m^{2K-2m}. \tag{4.4}$$

Proof: As in the proof of Theorem 3.2, we consider the $K \times K$ matrix G which is formed by the rows and columns of A with indices from $\{N+1, \dots, N+K\}$. Let X_m be the Kolmogorov subspace of X for which $\text{dist}(\mathcal{F}, X_m) = d_m$. For each i, there is an element $h_i \in X_m$ such that

$$\|f_i - h_i\| = \text{dist}(f_i, X_m)_X \le d_m,$$

and therefore

$$|\lambda_j(f_i) - \lambda_j(h_i)| = |\lambda_j(f_i - h_i)| \le \|\lambda_j\|_{X^*}\|f_i - h_i\| \le d_m. \tag{4.5}$$

We now consider the vectors $(\lambda_{N+1}(h), \dots, \lambda_{N+K}(h))$, $h \in X_m$. They span a space $W \subset \mathbb{R}^K$ of dimension $\le m$. We assume that $\dim(W) = m$ (a slight notational adjustment has to be made if $\dim(W) < m$). It follows from (4.5) that each row $\mathbf{g}_i$ of G can be approximated by a vector from W in the ℓ_∞ norm to accuracy d_m, and therefore in the ℓ_2 norm to accuracy $\sqrt{K}d_m$. Let P be the orthogonal projection of $\mathbb{R}^K$ onto W. Hence, we have

$$\|\mathbf{g}_i - P\mathbf{g}_i\|_{\ell_2} \le \sqrt{K}d_m, \quad i = 1, \dots, K. \tag{4.6}$$

It also follows from (4.3) that

$$\|P\mathbf{g}_i\|_{\ell_2} \le \|\mathbf{g}_i\|_{\ell_2} \le \left\{ \sum_{j=1}^{i} \sigma_{N+j}^2 \right\}^{1/2},$$

and therefore

$$\sum_{i=1}^{K} \|P\mathbf{g}_i\|_{\ell_2}^2 \le \sum_{i=1}^{K} \sum_{j=1}^{i} \sigma_{N+j}^2 \le K \sum_{i=1}^{K} \sigma_{N+i}^2. \tag{4.7}$$

8

Next, we apply Lemma 2.1 for this G and W and use estimates (4.2), (4.6) and (4.7) to derive

$$
\begin{aligned}
\gamma^{2K} \prod_{i=1}^{K} \sigma_{N+i}^2 &\leq \left\{ \frac{K}{m} \sum_{i=1}^{K} \sigma_{N+i}^2 \right\}^m \left\{ \frac{K^2}{K-m} d_m^2 \right\}^{K-m} \\
&= K^{K-m} \left(\frac{K}{m} \right)^m \left(\frac{K}{K-m} \right)^{K-m} \left\{ \sum_{i=1}^{K} \sigma_{N+i}^2 \right\}^m d_m^{2(K-m)} \\
&\leq 2^K K^{K-m} \left\{ \sum_{i=1}^{K} \sigma_{N+i}^2 \right\}^m d_m^{2(K-m)},
\end{aligned}
$$

and the proof is completed. $\qquad\square$

In analogy with Corollary 3.3, we have the following special results for the weak greedy algorithm in a general Banach space.

Corollary 4.2 *Suppose that X is a Banach space. For the weak greedy algorithm with a constant γ, applied to a compact set $\mathcal{F}$ contained in the unit ball of X, the following holds for $\sigma_n := \sigma_n(\mathcal{F})_X$ and $d_n := d_n(\mathcal{F})_X$, $n = 1, 2, \ldots$,*
(i) For any such compact set $\mathcal{F}$ and $n \geq 1$, we have

$$
\sigma_n \leq \sqrt{2}\gamma^{-1} \min_{1 \leq m < n} n^{\frac{n-m}{2n}} \left\{ \sum_{i=1}^{n} \sigma_i^2 \right\}^{\frac{m}{2n}} d_m^{\frac{n-m}{n}}. \tag{4.8}
$$

In particular $\sigma_{2\ell} \leq 2\gamma^{-1}\sqrt{\ell d_\ell}$, $\ell = 1, 2 \ldots$.
(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.

Proof: The proofs are similar to those of Corollary 3.3 except that we use (4.4) in place of (3.1).
(i) We take $N = 0$, $K = n$, and any $1 \leq m < n$ in (4.4) and use the monotonicity of $(\sigma_n)_{n \geq 0}$ to obtain

$$
\sigma_n^{2n} \leq 2^n n^{n-m} \gamma^{-2n} \left\{ \sum_{i=1}^{n} \sigma_i^2 \right\}^m d_m^{2n-2m}. \tag{4.9}
$$

If we take a $2n$-th root of both sides, we arrive at (4.8). In particular, if $n = 2\ell$ and $m = \ell$, we have

$$
\sigma_{2\ell} \leq \sqrt{2}\gamma^{-1}(2\ell)^{1/4}\{\Sigma_{i=1}^{2\ell}\sigma_i^2\}^{1/4}\sqrt{d_\ell} \leq \sqrt{2}\gamma^{-1}(2\ell)^{1/4}(2\ell)^{1/4}\sqrt{d_\ell} = 2\gamma^{-1}\sqrt{\ell d_\ell},
$$

where we have used the fact that all $\sigma_i \leq 1$.

9

(ii) It follows from the monotonicity of $(\sigma_n)_{n\geq 0}$ and (4.4) for $N = K = n$ and any $1 \leq m < n$ that

$$\sigma_{2n} \leq \sqrt{2n}\gamma^{-1}\sigma_n^\delta d_m^{(1-\delta)}, \quad \delta := \frac{m}{n}. \tag{4.10}$$

Given our β, we define $m =: \lfloor \frac{2\beta}{2\beta+1}n \rfloor + 1$ ($m < n$ for $n \geq 2 > 2\beta + 1$). It follows that

$$\delta = \frac{m}{n} \in \left(\frac{2\beta}{2\beta+1}, \frac{2\beta}{2\beta+1} + \frac{1}{n} \right). \tag{4.11}$$

We next prove (ii) by contradiction. Suppose it is not true and M is the first value where $\sigma_M > C_1 M^{-\alpha+\beta+1/2}$. Clearly, because of the definition of C_1, and the fact that $\sigma_n \leq 1$, we must have $M > 7$. We first consider the case $M = 2n$, and therefore $n > 3$. From (4.10) we have

$$C_1(2n)^{-\alpha+\beta+1/2} < \sigma_{2n} \leq \sqrt{2n}\gamma^{-1}C_1^\delta n^{\delta(-\alpha+\beta+1/2)}C_0^{1-\delta}(\delta n)^{-\alpha(1-\delta)},$$

where we have used the fact that $\sigma_n \leq C_1 n^{-\alpha+\beta+1/2}$ and $d_m \leq C_0 m^{-\alpha}$. It follows that

$$C_1^{1-\delta} < C_0^{1-\delta}2^{\alpha-\beta}\gamma^{-1}\delta^{-\alpha(1-\delta)}n^{\frac{2\beta+1}{2}(\delta - \frac{2\beta}{2\beta+1})},$$

and therefore

$$C_1 < C_0 2^{\frac{\alpha-\beta}{1-\delta}}\gamma^{-\frac{1}{1-\delta}}\delta^{-\alpha}n^{\frac{2\beta+1}{2}\cdot\frac{\delta - \frac{2\beta}{2\beta+1}}{1-\delta}}.$$

Since for $n \geq 4 > 2(2\beta + 1)$, we have

$$\delta < \frac{4\beta+1}{2(2\beta+1)} < 1, \quad \text{and therefore} \quad \frac{1}{1-\delta} < 2(2\beta+1).$$

This gives

$$\frac{2\beta+1}{2} \cdot \frac{\delta - \frac{2\beta}{2\beta+1}}{1-\delta} < \frac{(2\beta+1)^2}{n}, \quad \text{and thus} \quad n^{\frac{2\beta+1}{2}(\delta - \frac{2\beta}{2\beta+1})} < n^{\frac{(2\beta+1)^2}{n}} < 2^{(2\beta+1)^2}.$$

Then, for $\beta < \min\{\alpha, 1/2\}$

$$\begin{aligned}
C_1 &< C_0 2^{2(\alpha-\beta)(2\beta+1)}\gamma^{-2(2\beta+1)}\left(\frac{2\beta}{2\beta+1}\right)^{-\alpha}2^{(2\beta+1)^2} \\
&< C_0 2^{2(2\alpha+1)}\gamma^{-4}\left(\frac{2\beta+1}{2\beta}\right)^\alpha < C_0 2^{2(4\alpha+1)}\gamma^{-4}\left(\frac{2\beta+1}{2\beta}\right)^\alpha, \tag{4.12}
\end{aligned}$$

which is the desired contradiction. Likewise, if $M = 2n + 1$ (since $M > 7$, we have $n > 3$), for $-\alpha + \beta + 1/2 < 0$ (which is the meaningful case),

$$\begin{aligned}
C_1 2^{-\alpha+\beta+1/2}(2n)^{-\alpha+\beta+1/2} &< C_1(2n+1)^{-\alpha+\beta+1/2} < \sigma_{2n+1} \leq \sigma_{2n} \\
&< \sqrt{2n}\gamma^{-1}C_1^\delta n^{\delta(-\alpha+\beta+1/2)}C_0^{1-\delta}(\delta n)^{-\alpha(1-\delta)},
\end{aligned}$$

and following the same argument as above we get

$$C_1 < C_0 2^{-\frac{1}{2(1-\delta)}}2^{2\frac{\alpha-\beta}{1-\delta}}\gamma^{-2(2\beta+1)}\left(\frac{2\beta}{2\beta+1}\right)^{-\alpha}2^{(2\beta+1)^2} < C_0 2^{2(4\alpha+1)}\gamma^{-4}\left(\frac{2\beta+1}{2\beta}\right)^\alpha,$$

where we have used that $2^{-\frac{1}{2(1-\delta)}} < 1$, and the proof is completed.

(iii) From (i), we have

$$\sigma_{2n+1} \leq \sigma_{2n} \leq 2\gamma^{-1}\sqrt{nd_n} \leq 2\gamma^{-1}\sqrt{C_0}\sqrt{n}e^{-\frac{c_0}{2}n^\alpha} < \sqrt{2C_0}\gamma^{-1}\sqrt{2n+1}e^{-c_02^{-1-2\alpha}(2n+1)^\alpha},$$

from which (iii) easily follows. $\qquad\square$

5   Lower bounds in a Banach space

It is natural to ask whether the factor $\sqrt{n}$ is necessary when proving results in a Banach space. Here, we shall provide examples which show that a loss of this type is indeed necessary. However, as it will be seen, there is still a small gap between what we have proved for direct estimates and what the examples below provide.

Let us begin by considering the space $X := \ell_\infty(\mathbb{N} \cup \{0\})$ equipped with its usual norm. We consider a monotone decreasing sequence $x_0 \geq x_1 \geq x_2 \geq \cdots$ of positive real numbers which converge to zero and define

$$f_j := x_j e_j, \quad j = 0, 1, \ldots,$$

where e_j, $j = 0, 1, \ldots$ are the usual coordinate vectors in $\mathbb{R}^{\mathbb{N}\cup\{0\}}$. Let $\mathcal{F} := \{f_0, f_1, \ldots\}$. From the monotonicity of the x_j's, the greedy algorithm for $\mathcal{F}$ in X can choose the elements from $\mathcal{F}$ in order $f_0, f_1, \ldots$. Hence,

$$\sigma_j = \sigma_j(\mathcal{F})_X = x_j, \quad j \geq 0.$$

We want to give an upper bound for the Kolmogorov width of $\mathcal{F}$. For this, we shall use the following result (see (7.2) of Chapter 14 in [4]) on s-widths of the unit ball b_1^m of ℓ_1^m in ℓ_∞^m:

$$d_s(b_1^m)_X \leq C \{\ln(m/s)\}^{1/2} s^{-1/2}, \quad 1 \leq s \leq m/2. \tag{5.1}$$

Let us now define the sequence $\{x_j\}_{j\geq 0}$ so that in position $2^{k-1} \leq j \leq 2^k - 1$ it has the constant value $2^{-k\alpha}$, for $k = 0, 1, \ldots$, where $\alpha > 1/2$. It follows that,

$$\sigma_n(\mathcal{F})_X = \mathcal{O}(n^{-\alpha}), \quad n = 1, 2, \ldots.$$

We shall now bound the N-width of $\mathcal{F}$ when $N = 2^{n+1}$ by constructing a good space X_N of dimension $\leq N$ for approximating $\mathcal{F}$. The space X_N will be the span of a set E of at most N vectors. First, we place in E all of the vectors, $e_1, \ldots, e_{2^n}$. Next, for each $k = 1, \ldots n$, we use (5.1) to choose a basis for the space of dimension 2^{n-k} whose vectors are supported on $[2^{n+k}, 2^{n+k+1} - 1]$ and this space approximates each of the f_j, $j = 2^{n+k}, \ldots, 2^{n+k+1} - 1$, in X to accuracy $C_0 2^{-(n+k)\alpha}\sqrt{k}2^{-(n-k)/2}$. We place these basis vectors in E. Notice that $|x_j| \leq 2^{-2n\alpha}$ for $j \geq 2^{2n}$. This means that for the space $X_N := \mathrm{span}(E)$ with dimension $\leq N$ we have

$$
\begin{aligned}
d_N(\mathcal{F})_X &\leq \mathrm{dist}(\mathcal{F}, X_N)_X \leq \max\left\{ 2^{-2n\alpha}, \max_{1\leq k\leq n} C_0 2^{-(n+k)\alpha} 2^{-(n-k)/2}\sqrt{k} \right\} \\
&= \max\left\{ 2^{-2n\alpha}, C_0 2^{-n(\alpha+1/2)} \cdot \max_{1\leq k\leq n} 2^{-k(\alpha-1/2)}\sqrt{k} \right\} \leq C_1 2^{-n(\alpha+1/2)}, \quad \alpha > 1/2.
\end{aligned}
$$

From the monotonicity of $(d_n(\mathcal{F})_X)_{n \geq 0}$, we obtain that

$$d_n(\mathcal{F})_X \leq C_2 n^{-\alpha - 1/2}, \quad n = 1, 2, \ldots.$$

This example shows that the factor $\sqrt{n}$ which appears in (ii) of Corollary 4.2 can in general not be removed.

References

[1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced bases* Methods, SIAM J. Math. Anal., **43** (2011), 1457–1472.

[2] 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.

[3] E. Hewitt, K. Stromberg, *Real and Abstract Analysis*, Springer Verlag, Berlin 1969

[4] G.G. Lorentz, M. von Golitschek, and Y. Makovoz, *Constructive Approximation: Advanced Problems*, Springer Verlag, vol. 304, New York, 1996.

[5] 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.

[6] 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.

Ronald DeVore, Department of Mathematics, Texas A&M University, College Station, TX, rdevore@math.tamu.edu

Guergana Petrova, Department of Mathematics, Texas A&M University, College Station, TX, gpetrova@math.tamu.edu

Przemyslaw Wojtaszczyk, Institute of Applied Mathematics, and Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland, wojtaszczyk@mimuw.edu.pl