arXiv:2304.13332v2 [math.NA] 23 Feb 2024

Entropy-based convergence rates of greedy algorithms

Yuwen Li[*]

*School of Mathematical Sciences, Zhejiang University,
Hangzhou, Zhejiang 310058, China
liyuwen@zju.edu.cn*

Jonathan W. Siegel

*Department of Mathematics, Texas A&M University,
College Station, TX 77845, USA
jwsiegel@tamu.edu*

We present convergence estimates of two types of greedy algorithms in terms of the entropy numbers of underlying compact sets. In the first part, we measure the error of a standard greedy reduced basis method for parametric PDEs by the entropy numbers of the solution manifold in Banach spaces. This contrasts with the classical analysis based on the Kolmogorov n-widths and enables us to obtain direct comparisons between the algorithm error and the entropy numbers, where the multiplicative constants are explicit and simple. The entropy-based convergence estimate is sharp and improves upon the classical width-based analysis of reduced basis methods for elliptic model problems. In the second part, we derive a novel and simple convergence analysis of the classical orthogonal greedy algorithm for nonlinear dictionary approximation using the entropy numbers of the symmetric convex hull of the dictionary. This also improves upon existing results by giving a direct comparison between the algorithm error and the entropy numbers.

1. Introduction

Greedy algorithms are ubiquitous in advanced scientific computing and computational mathematics. Successful greedy-type numerical algorithms include adaptive finite element[2,13,41,42,49] and wavelet[14,29] methods for boundary value problems, certified model reduction technique for parametric differential equations,[3,35,52] and nonlinear dictionary approximation[4,23,27,37,56,64] in signal processing, machine

---

[*]Corresponding author.

learning and statistics. The aforementioned algorithms use local extreme criteria and iterations to search a global quasi-optimizer. It is known that many greedy algorithms have guaranteed error controls, solid convergence analysis, and optimal computational complexity as advantages. In this paper, we are focused on two applications of greedy algorithms, one for the *Reduced Basis Method* (RBM) and one for approximating functions. We shall show that these two types of greedy algorithms originally designed for seemingly unrelated problems can be analyzed in a unified way.

The RBM is a class of popular numerical methods developed in recent decades for efficiently solving *parametric* PDEs.[15,35,46,52] In a Banach space X, let $\mathcal{P}(u_\mu, \mu) = 0$ be the model problem with the solution $u_\mu \in X$, where μ is a parameter in a compact set $\mathcal{U}$. A standard RBM is split into two stages. In the *offline* stage, the RBM builds a highly accurate finite-dimensional subspace $X_n = \text{span}\{u_{\mu_1}, \ldots, u_{\mu_n}\} \subset X$ to uniformly approximate the compact solution manifold

$$\mathcal{M} = u(\mathcal{U}) = \{u_\mu \in X : \mu \in \mathcal{U}\}.$$

The offline construction of X_n is potentially very expensive but is only implemented once during the offline setup stage of the RBM. Then the *online* module of the RBM is able to rapidly produce accurate Galerkin approximations to u_μ based on X_n for many instances of the parameter μ.

There are two main types of computational methods for constructing X_n. One is called the Proper Orthogonal Decomposition (POD),[5] which utilizes a (costly) singular value decomposition to extract the low-rank structure of a given high-quality dataset. The other, is an iterative greedy algorithm for building the RBM subspace X_n, which relies on a posteriori error estimates. Compared with POD, greedy algorithms are computationally more feasible and have certified error bounds and convergence rates. The work[47] provided a priori convergence rates of greedy RBMs for single-parameter problems. In general, the convergence of greedy algorithms for the RBM is given by comparing the numerical error of the algorithm $\sigma_n(\mathcal{M})$ against the Kolmogorov n-width $d_n(\mathcal{M})$, which measures the performance of the best possible n-dimensional subspace for approximating the solution manifold $\mathcal{M}$. A direct comparison between $\sigma_n(\mathcal{M})$ and $d_n(\mathcal{M})$ was developed in Ref. [10]. Later on, the works[7,19] showed for the first time that

$$d_n(\mathcal{M}) = \mathcal{O}(n^{-s}) \Longrightarrow \sigma_n(\mathcal{M}) = \mathcal{O}(n^{-s}), \tag{1.1}$$

where $s > 0$ and $\mathcal{O}$ is Landau's big O notation. In other words, greedy-type RBMs are rate-optimal for polynomial decaying $d_n(\mathcal{M})$.

As far as we know, convergence results for the greedy RBM in the literature are all based on comparison with the Kolmogorov n-width. In this paper, we derive convergence analysis of greedy-type RBMs in terms of $\varepsilon_n = \varepsilon_n(\text{co}(\mathcal{M}))$, the entropy number[38] of the symmetric convex hull $\text{co}(\mathcal{M})$. Just as the n-width $d_n = d_n(\mathcal{M})$, ε_n is also an asymptotically small quantity for compact sets (see Refs. [11] and [45]). Unlike the rate-optimality result (1.1), our analysis leads to direct and optimal

comparisons of the type

$$\sigma_n(\mathcal{M}) \le C n^t \varepsilon_n(\mathrm{co}(\mathcal{M})), \quad t \in \left[\frac{1}{2}, 1\right], \tag{1.2}$$

where C is an explicit constant and the exponent t depends upon the Banach space in which the error is measured. Specifically, $t = (1/2) + |1/2 - 1/p|$ for the spaces L_p. As $n \to \infty$ (1.2) is sharper than the direct comparisons based on $d_n(\mathcal{M})$ given in Refs. [10] and [19]. Moreover, in the context of solving parametric PDEs, it is easier to calculate the entropy numbers than the Kolmogorov n-width of relevant sets, which is another advantage of bounding $\sigma_n(\mathcal{M})$ in terms of ε_n, see Subsection 2.2 for more details.

On the other hand, there are a variety of greedy algorithms for constructing nonlinear approximations (see Ref. [20]) of a single target function in the fields of machine learning, statistics and signal processing. The historically first of these algorithms is known as projection pursuit regression in statistics,[28,36] the matching pursuit algorithm in signal procession,[17,48] and the Pure Greedy Algorithm (PGA) in approximation theory.[23] Other variants of these greedy algorithms have been proposed which enjoy significantly better convergence guarantees for general dictionaries, including the Relaxed Greedy Algorithm (RGA)[18,23,37,51] and the Orthogonal Greedy Algorithm (OGA).[4,23,50] We refer to the article[64] and the book[60] for a thorough introduction to greedy algorithms for function approximation.

In recent decades, greedy algorithms have also been used to solve PDEs, see[1,26,40] and the recent work[55] which uses RGAs and OGAs to solve PDEs with neural networks. These algorithms adaptively select basis functions from a redundant dictionary $\mathbb{D} \subset X$ and use a sparse linear combination

$$f_n = \sum_{i=1}^{n} c_i g_i, \quad g_i \in \mathbb{D}$$

of dictionary elements to approximate a target function f. Among these greedy algorithms, the RGA and OGA achieve the worst case optimal convergence rate $\mathcal{O}(n^{-1/2})$ for target functions in the convex hull of the dictionary $\mathbb{D}$,[23,51] while the PGA in general may perform worse.[43] In Ref. [56], it is shown that for dictionaries whose convex hull has small entropy numbers, the orthogonal greedy algorithm may converge faster, specifically for $f \in \mathrm{co}(\mathbb{D})$ and $s > 1/2$,

$$\varepsilon_n(\mathrm{co}(\mathbb{D})) = \mathcal{O}(n^{-s}) \Longrightarrow \|f - f_n\|_X = \mathcal{O}(n^{-s}). \tag{1.3}$$

As mentioned before, the OGA converges with rate $\mathcal{O}(n^{-1/2})$ and no better for a general dictionary $\mathbb{D}$ (see Ref. [23]). The result (1.3) is an improved convergence rate for OGAs under additional assumptions, which hold for many popular dictionaries, e.g., the `ReLU`$_k$ dictionary in shallow neural networks (see Ref. [57]).

The second main contribution of our work is a novel direct comparison for the OGA:

$$\|f - f_n\|_X \le C_f \varepsilon_n(\mathrm{co}(\mathbb{D})), \tag{1.4}$$

4   *Y. Li & J. W. Siegel*

where C_f is an explicit uniform constant depending on the target function f. We remark that (1.3) is indeed a *rate* comparison while our estimate (1.4) is *direct* and implies (1.3) as special cases. In addition, the validity of (1.3) relies on the polynomial decay of $\varepsilon_n(\mathrm{co}(\mathbb{D}))$, while the estimate (1.4) is non-asymptotic and is always valid for general dictionaries, e.g., when $\varepsilon_n(\mathrm{co}(\mathbb{D})) = \mathcal{O}(e^{-\alpha n^s})$ is exponentially diminishing.

Throughout the rest of this section, we briefly introduce the Kolmogorov n-width, the entropy numbers, and the details of greedy algorithms for the RBM and function approximation.

1.1. *Greedy Algorithms for RBMs*

Let K be a compact set in a Banach space X equipped with the norm $\|\bullet\| = \|\bullet\|_X$. The Kolmogorov n-width of K is defined as

$$d_n(K) = d_n(K)_X := \inf_{X_n} \sup_{f \in K} \mathrm{dist}(f, X_n),$$

where

$$\mathrm{dist}(f, X_n) := \inf_{g \in X_n} \|f - g\|$$

is the distance from f to X_n, and the infimum is taken over all n-dimensional subspaces of X.

On the other hand, Kolmogorov proposed and investigated the (dyadic) entropy numbers of K (see Ref. 38), which are given by

$$\varepsilon_n(K) = \varepsilon_n(K)_X := \inf\{\varepsilon > 0 : K \text{ is covered by } 2^n \text{ balls of radius } \varepsilon\}.$$

It is known that $\{d_n(K)\}_{n \geq 1}$ and $\{\varepsilon_n(K)\}_{n \geq 1}$ are decreasing sequences and $\lim_{n \to 0} d_n(K) = \lim_{n \to 0} \varepsilon_n(K) = 0$ for any compact set K (see Ref. 45). When K is compact, the symmetric or absolutely convex hull of K,

$$\mathrm{co}(K) := \overline{\left\{ \sum_i c_i g_i : \sum_i |c_i| \leq 1, \ g_i \in K \text{ for each } i \right\}},$$

is also compact. The key quantity used in our convergence analysis is the entropy number $\varepsilon_n(\mathrm{co}(K))$, while in traditional analyses the Kolmogorov n-widths $d_n(K)$ are typically used.

In the offline stage, the greedy algorithm for RBMs selects a sequence $\{f_n\}_{n \geq 1}$ from K in an adaptive way as follows.

Algorithm 1.1 (Greedy Algorithm for RBMs). Set $n = 1$ and $X_0 = \{0\}$.

Step 1: Compute

$$f_n = \arg\max_{f \in K} \mathrm{dist}(f, X_{n-1}).$$

Step 2: Set $X_n = \mathrm{span}\{f_1, \ldots, f_n\}$. Set $n = n + 1$. Go to Step 1.

In practice, K is often the solution manifold $\mathcal{M}$ of a parametric PDE corresponding to a range of varying parameters, and the computational cost of the exact greedy approach (Algorithm 1.1) might still be exceedingly high. In such cases, a weak greedy algorithm driven by a posterior error estimation is able to relax the `argmax` criterion in Algorithm 1.1 and to realize the next more feasible algorithm. This modification is analogous to the weak greedy algorithms for function approximation,[62] which we discuss in more detail in Subsection 1.2.

Algorithm 1.2 (Weak Greedy Algorithm for RBMs). Set $n = 1$, $\gamma \in (0,1]$ and $X_0 = \{0\}$.

> Step 1: Select $f_n \in K$ such that
>
> $$\text{dist}(f_n, X_{n-1}) \geq \gamma \max_{f \in K} \text{dist}(f, X_{n-1}).$$
>
> Step 2: Set $X_n = \text{span}\{f_1, \ldots, f_n\}$. Set $n = n + 1$. Go to Step 1.

We remark that Step 1 of Algorithm 1.2 is not directly implemented but practically achieved by an a posteriori error estimation.[25,35,53] The constant $\gamma \in (0,1]$ actually depends on the quality of a posteriori error estimators. The numerical error of Algorithm 1.2 is measured by

$$\sigma_n(K) = \sigma_n(K)_X := \sup_{f \in K} \text{dist}(f, X_n).$$

By definition the simple bound

$$d_n(K) \leq \sigma_n(K) \tag{1.5}$$

is always true. We note that Algorithm 1.2 with $\gamma = 1$ reduces to Algorithm 1.1 and the iterates $\{f_n\}_{n \geq 1}$ are not uniquely determined by the above greedy selection processes. When X is a Hilbert space, let P_U be the orthogonal projection onto a subspace $U \subset X$. In this case we have

$$\text{dist}(f, X_n) = \|f - P_{X_n}f\|,$$
$$\sigma_n(K) = \max_{f \in K} \|f - P_{X_n}f\|.$$

1.2. *Greedy Algorithms for Approximating Functions*

Let the dictionary $\mathbb{D}$ be a collection of elements in a real Hilbert space X equipped with inner product $\langle \cdot, \cdot \rangle$. For learning algorithms based upon wavelets, shallow neural networks, statistical regression or compressed sensing, it is often essential to construct nonlinear approximants to a target function $f \in X$. To achieve this goal, Refs. 28 and 48 developed the projection pursuit or matching pursuit algorithm:

$$g_n = \arg \max_{g \in \mathbb{D}} |\langle g, f - f_{n-1} \rangle|,$$
$$f_n = f_{n-1} + \langle f - f_{n-1}, g_n \rangle g_n,$$

where $f_0 = 0$ and $n = 1, 2, \ldots$ is the iteration counter. Note that if the dictionary elements are normalized in X, then this algorithm is adding the single term which minimizes the error in each step and validates the name "greedy algorithm." The theoretical analysis of this algorithm is notoriously complex,[23,36,58] and in any case it does not achieve an optimal rate of convergence.[43] As an alternative, a variety of different greedy algorithms which achieve better convergence behavior have been introduced, for example the following RGA[4,23,37] and its variants:[51]

$$(\alpha_n, \beta_n, g_n) = \arg\min_{\alpha,\beta\in\mathbb{R}, g\in\mathbb{D}} \|f - \alpha f_{n-1} - \beta g_n\|,$$

$$f_n = \alpha_n f_{n-1} + \beta_n g_n.$$

In this work, the algorithm under consideration is the OGA which has been developed and analyzed in, e.g., Refs. 4, 18, 23, 50, 60 and 56.

Algorithm 1.3 (Orthogonal Greedy Algorithm). Set $n = 1$ and $f_0 = 0$.

Step 1: Compute the optimizer

$$g_n = \arg\max_{g\in\mathbb{D}} |\langle g, f - f_{n-1}\rangle|.$$

Step 2: Set $X_n = \mathrm{span}\{g_1, \ldots, g_n\}$ and compute

$$f_n = P_{X_n} f = P_n f.$$

Set $n = n + 1$. Go to Step 1.

In Step 1 of Algorithm 1.3, it is possible that the argmax may not exist or may be impossible to compute. To overcome this, one could relax the selection criterion to

$$|\langle g_n, f - f_{n-1}\rangle| \geq \gamma \max_{g\in\mathbb{D}} |\langle g, f - f_{n-1}\rangle|$$

with $\gamma \in (0, 1]$. This modified algorithm is closer to what is typically implemented in practice and is called the weak OGA.[62] Weak versions of other greedy algorithms have also been introduced and analyzed in e.g., Ref. 60.

In our error estimates, C is a generic and uniform constant that may change from line to line but is independent of n and underlying compact sets or target functions. By $A \lesssim B$ we denote $A \leq CB$, and $A \approx B$ is equivalent to $A \lesssim B$ and $B \lesssim A$.

The rest of this paper is organized as follows. In Section 2, we present an entropy-based convergence analysis of greedy RBMs and corresponding applications in Hilbert spaces. In Section 3, we derive similar optimal convergence estimates for RBMs in Banach spaces. Section 4 is devoted to an optimal direct comparison estimate for the convergence of OGAs in Hilbert spaces. Concluding remarks are presented in Section 5.

2. Convergence of RBMs in Hilbert spaces

Throughout this section, we assume that X is a Hilbert space unless confusion arises. The next lemma is the main tool of our entropy-based convergence analysis for greedy algorithms.

Lemma 2.1. *Let* $V_n = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}$ *be the volume of an* ℓ_2-*unit ball in* $\mathbb{R}^n$. *Let* K *be a compact set in a Hilbert space* X. *Then for any* $v_1, \dots, v_n \in K$, *it holds that*

$$\left(\prod_{k=1}^{n} \left\| v_k - P_{k-1} v_k \right\|\right)^{\frac{1}{n}} \le (n! V_n)^{\frac{1}{n}} \varepsilon_n(\mathrm{co}(K))_X.$$

where $P_0 = 0$ *and* P_k *is the orthogonal projection onto* $\mathrm{span}\{v_1, \dots, v_k\}$.

Proof. Consider the symmetric simplex

$$S = \mathrm{co}\big(\{v_1, \dots, v_n\}\big) = \left\{\sum_{i=1}^{n} c_i v_i : \sum_{i=1}^{n} |c_i| \le 1\right\}.$$

We assume that $\{v_k\}_{k=1}^{n}$ are linearly independent otherwise the inequality automatically holds. Let X_n be the n-dimensional subspace spanned by $v_1, \dots, v_n$. We identify X_n with the Euclidean space $\mathbb{R}^n$ and S with a n-dimensional skewed simplex in $\mathbb{R}^n$. It is straightforward to see that

$$\{v_k\}_{k=1}^{n} \mapsto \{v_k - P_{k-1} v_k\}_{k=1}^{n}$$

is the Gram-Schmidt orthogonalization of $v_1, ..., v_n$. As a result, the volume of S is given by

$$|S| = \frac{2^n}{n!} \prod_{k=1}^{n} \left\| v_k - P_{k-1} v_k \right\|. \tag{2.1}$$

The definition of $\varepsilon_n = \varepsilon_n(\mathrm{co}(K))_X$ implies that S is covered by 2^n balls of radius ε_n. Therefore by (2.1) and a volume comparison between S and the union of balls, we have

$$\frac{2^n}{n!} \prod_{k=1}^{n} \left\| v_k - P_{k-1} v_k \right\| \le 2^n \varepsilon_n^n V_n.$$

The proof is complete. $\qquad \square$

2.1. *Convergence estimates*

Using Lemma 2.1, we immediately obtain a direct comparison result between the RBM error and the entropy number on convex compact sets.

Theorem 2.1. *Let* K *be a compact set in a Hilbert space* X. *For the weak greedy algorithm (Algorithm 1.2) we have*

$$\sigma_n(K)_X \le \gamma^{-1} (n! V_n)^{\frac{1}{n}} \varepsilon_n(\mathrm{co}(K))_X. \tag{2.2}$$

8   *Y. Li & J. W. Siegel*

Proof. For Algorithm 1.2 and $n \geq 1$, we have

$$\gamma \sigma_{n-1}(K) \leq \left\| f_n - P_{X_{n-1}} f_n \right\| \leq \sigma_{n-1}(K). \tag{2.3}$$

It then follows from (2.3) and Lemma 2.1 that

$$\left(\prod_{k=1}^n \sigma_{k-1}(K) \right)^{\frac{1}{n}} \leq \gamma^{-1} \left(\prod_{k=1}^n \left\| f_k - P_{X_{k-1}} f_k \right\| \right)^{\frac{1}{n}} \tag{2.4}$$

$$\leq \gamma^{-1} (n! V_n)^{\frac{1}{n}} \varepsilon_n(\mathrm{co}(K))_X.$$

Combining (2.4) with the fact

$$\sigma_0(K) \geq \sigma_1(K) \geq \cdots \geq \sigma_{n-1}(K) \geq \sigma_n(K)$$

completes the proof. $\qquad \square$

Using the well-known Stirling's formula

$$\lim_{n \to \infty} n! \left/ \sqrt{2\pi n} \left(\frac{n}{\mathrm{e}} \right)^n \right. = 1, \quad \lim_{n \to \infty} V_n \left/ \frac{1}{\sqrt{n\pi}} \left(\frac{2\pi \mathrm{e}}{n} \right)^{\frac{n}{2}} \right. = 1,$$

we obtain a corollary about the convergence rate of greedy-type RBMs.

Corollary 2.1. *Let K be a compact set in a Hilbert space X. For Algorithm 1.2 there exists an absolute constant C independent of K and n, such that*

$$\sigma_n(K)_X \leq C \sqrt{n} \varepsilon_n(\mathrm{co}(K))_X.$$

A natural question is whether the convergence rate of $\sigma_n(K)$ in Corollary 2.1 is sharp and whether the factor $\sqrt{n}$ is removable. In the following we give a negative answer to this question.

Proposition 2.1. *There exists a Hilbert space X and a symmetric convex compact set $K \subset X$ such that Algorithm 1.1 satisfies*

$$\sigma_n(K)_X \eqsim \sqrt{n} \varepsilon_n(K)_X.$$

Proof. On a Lipschitz domain $\Omega \subset \mathbb{R}^d$, we consider the $\mathtt{ReLU}_k$ activation function $\sigma(x) = \max(x^k, 0)$ with $k \geq 1$ and the dictionary

$$\mathbb{P}_k^d = \left\{ \sigma(\omega \cdot x + b) : \omega \in \mathbb{S}_1^d, \ b \in [\underline{b}, \bar{b}] \right\},$$

where $\mathbb{S}_1^d$ is the d-dimensional unit sphere centered at 0, and $\underline{b}$, $\bar{b}$ are constants depending on Ω. We set $K = \mathrm{co}(\mathbb{P}_k^d)$ and $X = L_2(\Omega)$. It has been shown in Ref. [57] that

$$d_n(\mathrm{co}(\mathbb{P}_k^d))_{L_2(\Omega)} = \mathcal{O}(n^{-\frac{2k+1}{2d}}), \tag{2.5a}$$

$$\varepsilon_n(\mathrm{co}(\mathbb{P}_k^d))_{L_2(\Omega)} = \mathcal{O}(n^{-\frac{1}{2} - \frac{2k+1}{2d}}). \tag{2.5b}$$

Combining (2.5) and Corollary 2.1 we arrive at

$$d_n(\mathrm{co}(\mathbb{P}_k^d))_{L_2(\Omega)} \leq \sigma_n(\mathrm{co}(\mathbb{P}_k^d))_{L_2(\Omega)} \lesssim \sqrt{n} \varepsilon_n(\mathrm{co}(\mathbb{P}_k^d))_{L_2(\Omega)},$$

and obtain the convergence rate

$$\sigma_n(\mathrm{co}(\mathbb{P}_k^d))_{L_2(\Omega)} = \mathcal{O}(n^{-\frac{2k+1}{2d}}).$$

As a result, the error estimate in Corollary 2.1 is not improvable. □

In the following we discuss the advantage of our direct comparisons over the classical ones. The first direct comparison result for greedy-type RBMs is derived in Ref. [10]:

$$\sigma_n(K) \lesssim n2^n d_n(K),$$

which is useful if $d_n(K) = o(n^{-1}2^{-n})$. Later on, this estimate has been improved in Refs. [7] and [19]. Currently the sharpest direct comparison in the RBM literature (see Ref. [19]) is

$$\sigma_{2n}(K) \lesssim \sqrt{d_n(K)}. \tag{2.6}$$

Next we show that our direct comparison result in Theorem 2.1 is indeed sharper than (2.6) as $n \to \infty$. In doing so, we need the well-known Carl's inequality (see Refs. [11] and [45]).

Lemma 2.2 (Carl's inequality). *For every $\alpha > 0$, there exists a constant $C(\alpha) > 0$ such that for any compact set K in a Banach space X,*

$$\varepsilon_n(K) \leq C(\alpha)n^{-\alpha} \max_{1 \leq i \leq n}(i^\alpha d_{i-1}(K)).$$

For a compact set K with $d_n(K) \lesssim n^{-s}$ and $s > 0$, Lemma 2.2 implies that $\varepsilon_n(K) \lesssim n^{-s}$, i.e., the entropy number decays as fast as the polynomial-decaying Kolmogorov n-width. Therefore combining the fact (see Ref. [45])

$$d_n(K) = d_n(\mathrm{co}(K)) \tag{2.7}$$

with Lemma 2.2 and Corollary 2.1, we have

$$d_n(K) \lesssim n^{-s} \Longrightarrow \sigma_n(K) \lesssim \sqrt{n}\varepsilon_n(\mathrm{co}(K)) \lesssim n^{\frac{1}{2}-s}. \tag{2.8}$$

In contrast, the estimate (2.6) under the same assumption yields

$$d_n(K) \lesssim n^{-s} \Longrightarrow \sigma_n(K) \lesssim n^{-\frac{s}{2}}.$$

Thus our direct comparisons in Theorem 2.1 are asymptotically sharper than the classical result (2.6) when $s > 1$.

2.2. *Examples*

In the rest of this section, we discuss applications of our results to the second order elliptic equation

$$-\nabla \cdot (a\nabla u(a)) = f \quad \text{in } \Omega \subset \mathbb{R}^d, \tag{2.9a}$$

$$u(a) = 0 \quad \text{on } \partial\Omega. \tag{2.9b}$$

Here $f \in H^{-1}(\Omega)$ is fixed, the diffusion coefficient $a \in L_\infty(\Omega)$ is the varying input parameter belonging to the function class $\mathcal{A}$ that will be specified later, and $u(a) \in H_0^1(\Omega)$ is the weak solution corresponding to a. First we present a simple lemma for calculating the entropy numbers of the solution manifold $\mathcal{M} = u(\mathcal{A})$ of (2.9).

Lemma 2.3. *Let $0 < \alpha \le 1$ and $\Phi : X \to Y$ be an α-Hölder continuous mapping between Banach spaces X and Y such that for all $x_1, x_2 \in X$,*

$$\|\Phi(x_1) - \Phi(x_2)\|_Y \le L\|x_1 - x_2\|_X^\alpha.$$

Then for any compact set $K \subset X$ we have

$$\varepsilon_n(\Phi(K))_Y \le L\varepsilon_n(K)_X^\alpha.$$

Proof. Let K be covered by 2^n balls $\{B_\varepsilon(x_i)\}_{i=1}^{2^n}$ of radius $\varepsilon = \varepsilon_n(K)_X$, where $B_\varepsilon(x_i)$ is centered at $x_i \in X$. Then the Hölder continuity of Φ implies that $\Phi(K)$ can be covered by balls $\{B_{L\varepsilon^\alpha}(\Phi(x_i))\}_{i=1}^{2^n}$, suggesting that $\varepsilon_n(\Phi(K))_Y \le L\varepsilon^\alpha$. $\qquad\square$

For parametric PDEs, it is generally much more difficult to calculate the n-width $d_n(u(\mathcal{A}))$ of the solution manifold $\mathcal{M} = u(\mathcal{A})$ than computing $\varepsilon_n(u(\mathcal{A}))$. To remedy this situation, Cohen and DeVore[15,16] developed a technical tool for estimating $d_n(u(\mathcal{A}))$ in terms of $d_n(\mathcal{A})$.

Theorem 2.2 (Theorem 1 from Ref. [16]). *For a pair of complex Banach spaces X and Y, suppose u is a holomorphic mapping from an open set $O \subset X$ into Y and u is uniformly bounded on O:*

$$\sup_{a \in O} \|u(a)\|_Y < \infty.$$

If $K \subset O$ is a compact set, then for any $\alpha > 1$ and $\beta < \alpha - 1$ we have

$$\sup_{n \ge 1} n^\alpha d_n(K)_X < \infty \implies \sup_{n \ge 1} n^\beta d_n(u(K))_Y < \infty \ .$$

2.2.1. *Example 1*

Let $C^s(\Omega) = C^{k,\alpha}(\Omega)$ with regularity index

$$s := k + \alpha$$

be the space of functions on Ω whose k-th derivatives are Hölder continuous with exponent $\alpha \in (0, 1]$. In the first example, we focus on the following function class of diffusion coefficients

$$\mathcal{A} = \left\{ a \in L_\infty(\Omega) : \inf_{x \in \Omega} a(x) > M_0 > 0, \ \|a\|_{C^{k,\alpha}(\Omega)} \le M_1 \right\},$$

where constants M_0, M_1 are fixed. In practice, $a \in \mathcal{A}$ has additional affinely parametrized structures allowing an efficient online implementation in RBMs, see,

e.g., Refs. [35] and [52]. Otherwise, researchers often utilize the empirical interpolation method[3] to construct affinely parametrized approximations to a and then apply RBMs to the approximate model.

Using Theorem 2.2, the holomorphy of the solution operator $u : a \mapsto u(a)$, and $d_n(\mathcal{A})_{L_\infty(\Omega)} = \mathcal{O}(n^{-\frac{s}{d}})$ (see Refs. [44] and [16]), it has been proved in Ref. [16] that

$$d_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim n^{-\frac{s}{d}+1+t} \tag{2.10}$$

for any $t > 0$. Then as a result of (2.10) and the classical convergence estimate of greed-type RBMs (see Ref. [7]), for $s > d$,

$$\sigma_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim n^{-\frac{s}{d}+1+t}. \tag{2.11}$$

However, (2.11) fails to indicate any convergence of RBMs for approximating the solution process $a \mapsto u(a)$ of (2.9) when $s \leq d$.

Our convergence results for RBMs shed some light on such cases. For (2.9) there exists a useful perturbation result (see Theorem 2.1 from Ref. [8])

$$\|u(a_1) - u(a_2)\|_{H^1(\Omega)} \leq C_\Omega \|\nabla u(a)\|_{L_p(\Omega)} \|a_1 - a_2\|_{L_q(\Omega)}, \tag{2.12}$$

where $p \geq 2$ and $q := 2p/(p-2)$. In particular, (2.12) with $p = 2$ implies the Lipschitz property of

$$u : L_\infty(\Omega) \to H^1(\Omega), \quad a \mapsto u(a).$$

Combining this fact with Lemma 2.3 and $\varepsilon_n(\mathcal{A})_{L_\infty(\Omega)} = \mathcal{O}(n^{-\frac{s}{d}})$ (see Ref. [38]), we have

$$\varepsilon_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim \varepsilon_n(\mathcal{A})_{L_\infty(\Omega)} \lesssim n^{-\frac{s}{d}}. \tag{2.13}$$

Next we need a lemma from Ref. [12] about the relations between the entropy numbers of K and $\mathrm{co}(K)$, see also Refs. [59] and [30] similar results.

Lemma 2.4 (Proposition 6.2 from Ref. [12]). *Let K be a compact set in a Banach space X of type $p \in (1, 2]$, and assume that*

$$\varepsilon_n(K)_X \lesssim n^{-\alpha}$$

for some $\alpha > 1 - 1/p$. Then there exists a constant $c(\alpha, p)$ such that

$$\varepsilon_n(\mathrm{co}(K))_X \leq c(\alpha, p) n^{-1+\frac{1}{p}} (\log(n+1))^{-\alpha+1-\frac{1}{p}}.$$

In particular, the Hilbert space $X = H^1(\Omega)$ is of type $p = 2$. As a result of Lemma 2.4 and (2.13), when $d/2 < s \leq d$ we have

$$\varepsilon_n(\mathrm{co}(u(\mathcal{A})))_{H^1(\Omega)} \lesssim n^{-\frac{1}{2}} (\log(n+1))^{-\frac{s}{d}+\frac{1}{2}}. \tag{2.14}$$

Therefore using Corollary 2.1 and (2.14) we obtain the convergence of the weak greedy algorithm for (2.9):

$$d_n(u(\mathcal{A}))_{H^1(\Omega)} \leq \sigma_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim (\log(n+1))^{-\frac{s}{d}+\frac{1}{2}},$$

where the function class $\mathcal{A}$ is of lower regularity $s \in (d/2, d]$.

2.2.2. *Example 2*

In the second example, we consider the same boundary value problem (2.9) on $\Omega = [0,1]^2$ with a different set of diffusion coefficients, which corresponds to DeVore's geometric model in Ref. 21. Let

$$\mathcal{H} = \big\{ \phi \in C^{k,\alpha}[0,1] : 0 \leq \phi(x) \leq 1 \text{ for } x \in [0,1] \big\}.$$

The graph of a function $\phi \in \mathcal{H}$ partitions Ω into subregions Ω_+^ϕ and Ω_-^ϕ, defined by

$$\Omega_+^\phi = \{(x,y) \in \Omega, \ y \geq \phi(x)\} \tag{2.15}$$

and $\Omega_-^\phi = \Omega \backslash \Omega_+^\phi$. The diffusion coefficient a is a member of

$$\mathcal{A} = \{a = a_\phi \in L_\infty(\Omega) : a|_{\Omega_+^\phi} = 2, \ a|_{\Omega_-^\phi} = 1 \text{ for } \phi \in \mathcal{H}\}. \tag{2.16}$$

Functions in $\mathcal{A}$ have discontinuities along the curve represented by $\phi \in \mathcal{H}$, posing a real challenge for reducing the model problem (2.9). In fact, there is no convergence result of the RBM for the geometric model in the classical literature.

Elementary arguments show that for $q \geq 1$ the mapping $\phi \mapsto a_\phi$ is $1/q$-Hölder continuous from $L_\infty[0,1]$ to $L_q(\Omega)$. Therefore using Lemma 2.3 we can estimate the metric entropy of $\mathcal{A}$ as

$$\varepsilon_n(\mathcal{A})_{L_q(\Omega)} \leq \varepsilon_n(\mathcal{H})^{\frac{1}{q}}_{L_\infty[0,1]} \eqsim n^{-\frac{s}{q}}. \tag{2.17}$$

Moreover, there exists an exponent $P > 2$ depending on Ω such that $\nabla u(a) \in L_P(\Omega)$ (see Refs. 32 and 8). Then the perturbation result (2.12) with $p = P$, $q = Q := 2P/(P-2) \in (2, \infty)$ implies that

$$u : L_Q(\Omega) \to H^1(\Omega), \quad a \mapsto u(a)$$

is a Lipschitz mapping. Then using (2.17) and Lemma 2.3 again, we arrive at

$$\varepsilon_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim \varepsilon_n(\mathcal{A})_{L_Q(\Omega)} \lesssim n^{-\frac{s}{Q}}.$$

Combining it with Corollary 2.1 and Lemma 2.4 shows that

$$\sigma_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim \log(n+1)^{-\frac{s}{Q}+\frac{1}{2}}, \quad s > Q/2,$$

indicating the convergence of the RBM Algorithm 1.2 with $K = u(\mathcal{A})$. In addition, we obtain the novel decay result of the n-width

$$d_n(u(\mathcal{A}))_{H^1(\Omega)} \lesssim \log(n+1)^{-\frac{s}{Q}+\frac{1}{2}}.$$

3. Convergence of RBMs in Banach spaces

In this section, we derive direct comparison and convergence estimates for the weak greedy algorithm (Algorithm 1.2), where X is only assumed to be a general Banach space. First we introduce the Banach-Mazur distance $d(X,Y)$ (see Ref. 67) between two isomorphic Banach spaces X and Y:

$$d(X,Y) = \inf \big\{ \|T\| \|T^{-1}\| : T \text{ is an isomorphism from } X \text{ to } Y \big\}.$$

Let ℓ_2^n denote the space $\mathbb{R}^n$ under the Euclidean ℓ_2-norm. For all $n \geq 1$ we consider the number

$$\delta_n(X) := \sup_{Y_n \subset X} d(Y_n, \ell_2^n) \leq \sqrt{n}, \tag{3.1}$$

where the supremum is taken over all n-dimensional Banach subspaces Y_n in X, and the upper bound in (3.1) has been proven in Section III.B.9 of Ref. [66]. The quantity $\delta_n(X)$ measures how far an n-dimensional subspace of X is away from the Hilbert space ℓ_2^n. With the help of (3.1), we are able to establish the key lemma for our analysis in Banach spaces.

Lemma 3.1. *Let K be a compact set in a Banach space X. For any $v_1, \ldots, v_n \in K$ with $X_k = \mathrm{span}\{v_1, \ldots, v_k\}$ and $X_0 = \{0\}$, we have*

$$\Big(\prod_{k=1}^n \mathrm{dist}(v_k, X_{k-1}) \Big)^{\frac{1}{n}} \leq \delta_n(X)(n!V_n)^{\frac{1}{n}} \varepsilon_n(\mathrm{co}(K))_X.$$

Proof. Let T be an isomorphism from X_n to ℓ_2^n and

$$\|v\|_T := \|Tv\|_{\ell_2^n} \text{ for any } v \in X_n.$$

Then we have the norm equivalence

$$\|T^{-1}\|^{-1} \|v\|_X \leq \|v\|_T \leq \|T\| \|v\|_X, \quad \forall v \in X_n. \tag{3.2}$$

In addition, by the definition of $\delta_n(X)$ we can choose T such that

$$\|T\| \|T^{-1}\| = \delta_n(X). \tag{3.3}$$

Let $X_{n,T}$ denote the Hilbert space, which is the same as X_n as a set, and is equipped with the inner product

$$(\bullet, \bullet)_T := (T\bullet, T\bullet)_{\ell_2^n}.$$

As a result of (3.2) it holds that

$$\mathrm{dist}(v_k, X_{k-1}) \leq \|T^{-1}\| \|v_k - Q_{k-1} v_k\|_T, \tag{3.4}$$

where Q_k is the orthogonal projection onto X_k with respect to $(\bullet, \bullet)_T$. Let $S = \mathrm{co}\{v_1, \ldots, v_n\}$. Then applying our previous estimate in Lemma 2.1 to the Hilbert space $X_{n,T}$ and using (3.4), we have

$$\left(\prod_{k=1}^n \mathrm{dist}(v_k, X_{k-1}) \right)^{\frac{1}{n}} \leq \|T^{-1}\| \left(\prod_{k=1}^n \|v_k - Q_{k-1} v_k\|_T \right)^{\frac{1}{n}}$$
$$\leq \|T^{-1}\| (n!V_n)^{\frac{1}{n}} \varepsilon_n(S)_{X_{n,T}}. \tag{3.5}$$

On the other hand, the lower bound in (3.2) leads to

$$\varepsilon_n(S)_{X_{n,T}} \leq \|T\| \varepsilon_n(S)_X \leq \|T\| \varepsilon_n(\mathrm{co}(K))_X. \tag{3.6}$$

Therefore combining (3.5) with (3.6) and (3.3) completes the proof. $\qquad \square$

The estimate (3.1) on the Banach-Mazur distance can be improved in special Banach spaces like $L_p(d\mu)$ with $1 < p < \infty$. For instance, it has been shown in Ref. [66] that

$$\delta_n(X) \leq n^{|\frac{1}{2} - \frac{1}{p}|} \quad \text{for } X = L_p(d\mu), \ p \in [1, \infty]. \tag{3.7}$$

Following the same proof of Theorem 2.1, we obtain the following error bounds of Algorithm 1.2 in Banach spaces by Lemma 3.1.

Theorem 3.1. *Let K be a compact set in a Banach space X. For the weak greedy algorithm (Algorithm 1.2) we have*

$$\sigma_n(K)_X \leq \gamma^{-1} \delta_n(X)(n! V_n)^{\frac{1}{n}} \varepsilon_n(\text{co}(K))_X.$$

In particular, the asymptotic error estimates hold:

$$\sigma_n(K)_X \lesssim \begin{cases} \gamma^{-1} n \varepsilon_n(\text{co}(K))_X, & \text{for a general Banach space } X, \\ \gamma^{-1} n^{\frac{1}{2} + |\frac{1}{2} - \frac{1}{p}|} \varepsilon_n(\text{co}(K))_X, & X = L_p(d\mu), \ p \in [1, \infty]. \end{cases}$$

As a byproduct of (1.5) and Theorem 3.1, we also have a direct comparison between the Kolmogorov n-width and entropy number in general Banach spaces, which generalizes Proposition 2 from Ref. [57] for Hilbert spaces.

Corollary 3.1. *For a compact set K in a Banach space X we have*

$$d_n(K)_X \leq \delta_n(X)(n! V_n)^{\frac{1}{n}} \varepsilon_n(\text{co}(K))_X, \quad n \geq 1.$$

In the end of this section, we shall show that the factor $n^{\frac{1}{2} + |\frac{1}{2} - \frac{1}{p}|}$ in Theorem 3.1 cannot be removed when $p \geq 2$. To that end, we need the following lemma concerning the metric entropy of the ℓ_1^n unit ball in ℓ_p. We remark that we do not know whether Theorem 3.1 is tight when $1 \leq p < 2$.

Lemma 3.2 (Theorem 1 from Ref. [54]). *Let B_m^1 be an ℓ_1-unit ball in $\mathbb{R}^m$. For $m/2 < s < m$ and $1 \leq p \leq \infty$ it holds that*

$$\varepsilon_s(B_m^1)_{\ell_p} \approx \log(m/s) s^{-1 + \frac{1}{p}}.$$

Proposition 3.1. *There exists a symmetric and convex compact set K in the Banach space $X = \ell_p$ with $2 \leq p \leq \infty$ such that Algorithm 1.1 satisfies*

$$\sigma_n(K)_X \approx n^{1 - \frac{1}{p}} \varepsilon_n(K)_X. \tag{3.8}$$

Proof. Let $\{x_i\}_{i=1}^{\infty}$ be a sequence of real numbers decreasing to 0. We consider the compact set

$$K = \text{co}\left(\cup_{i=1}^{\infty} \{f_i\} \right) \subset X = \ell_p$$

where $f_i = x_i e_i$ and e_i is the i-th unit vector in ℓ_p. Clearly the greedy algorithm 1.1 selects $f_1, f_2, \ldots, f_n, \ldots$ as the iterates and

$$\sigma_n(K) = \|f_n\|_{\ell_p} = |x_n|. \tag{3.9}$$

Given an exponent $\alpha > 1$, let $x_j = 2^{-k\alpha}$ for $2^{k-1} \leq j \leq 2^k - 1$. By construction $x_n = \mathcal{O}(n^{-\alpha})$ and we shall show that

$$\varepsilon_N(K)_{\ell_p} = \mathcal{O}(N^{-1+\frac{1}{p}-\alpha}) \text{ for } N = 3 \cdot 2^n. \tag{3.10}$$

We consider the following subsets of K:

$$K_0 = \text{co}\big(\cup_{i=1}^{2^n} \{f_i\}\big),$$
$$K_k = \text{co}\big(\cup_{i=2^{n+k-1}}^{2^{n+k}-1} \{f_i\}\big), \quad 1 \leq k \leq n,$$
$$K_\infty = \text{co}\big(\cup_{i=2^{2n}}^{\infty} \{f_i\}\big).$$

Using the elementary fact $\varepsilon_{s+t}(A + B) \leq \varepsilon_s(A) + \varepsilon_t(B)$ (see Refs. 45 and 57) we have

$$\varepsilon_N(K)_{\ell_p} \leq \varepsilon_{2^n}(K_0)_{\ell_p} + \sum_{k=1}^{n} \varepsilon_{2^{n-k}}(K_k)_{\ell_p} + \varepsilon_{2^n}(K_\infty)_{\ell_p}. \tag{3.11}$$

Using the formula for the volume of an ℓ_p ball of radius ε in $\mathbb{R}^m$ (see Refs. 24 and 65)

$$V_m^p(\varepsilon) = 2^m \varepsilon^m \frac{\Gamma(1+\frac{1}{p})^m}{\Gamma(1+\frac{m}{p})},$$

we estimate $\varepsilon_m(K_0)_{\ell_p}$ by calculating the volume of K_0, with $m = 2^n$:

$$\varepsilon_m(K_0)_{\ell_p} \approx \left(\frac{2^m}{m!} |x_1| \cdots |x_m| \frac{\Gamma(1+\frac{m}{p})}{2^m \Gamma(1+\frac{1}{p})^m} \right)^{\frac{1}{m}} \approx m^{-1+\frac{1}{p}-\alpha}. \tag{3.12}$$

For $1 \leq k \leq n$, Lemma 3.2 implies that

$$\varepsilon_{2^{n-k}}(K_k)_{\ell_p} \lesssim 2^{-(n+k)\alpha} \varepsilon_{2^{n-k}}\big(B_{2^{n+k}}^1\big)_{\ell_p}$$
$$\lesssim 2^{-(n+k)\alpha} \sqrt{k} 2^{(-1+\frac{1}{p})(n-k)}$$
$$\lesssim \sqrt{k} 2^{-k(\alpha-1)} 2^{-n(\alpha+1-\frac{1}{p})}.$$

As a result, we obtain the following bound

$$\sum_{k=1}^{n} \varepsilon_{2^{n-k}}(K_k)_{\ell_p} \lesssim \sum_{k=1}^{\infty} \left(\sqrt{k} 2^{-k(\alpha-1)} \right) 2^{-n(\alpha+1-\frac{1}{p})} \lesssim 2^{-n(\alpha+1-\frac{1}{p})}. \tag{3.13}$$

Since the ℓ_p distance from 0 to the elements of K_∞ is less than $2^{-2n\alpha}$, one can simply estimate $\varepsilon_{2^n}(K_\infty)_{\ell_p}$ by

$$\varepsilon_{2^n}(K_\infty)_{\ell_p} \lesssim 2^{-2n\alpha}. \tag{3.14}$$

Therefore combining (3.12), (3.13) and (3.14), we confirm (3.10) and thus

$$\varepsilon_n(K)_{\ell_p} \lesssim n^{-\alpha-1+\frac{1}{p}} \quad \text{for } n \geq 1. \tag{3.15}$$

Finally (3.8) with $p \geq 2$ follows from Theorem 3.1 and (3.9), (3.15). $\qquad\square$

4. Convergence analysis of OGAs

Let X be a real Hilbert space and $\mathbb{D} \subset X$ be a fixed dictionary or collection of functions. To analyze the convergence behavior of the OGA, we consider the following norm and space:[4,23]

$$\|f\|_{\mathcal{L}_1(\mathbb{D})} := \inf\left\{ \sum_i |c_i| : f = \sum_{g_i \in \mathbb{D}} c_i g_i \right\},$$

$$\mathcal{L}_1(\mathbb{D}) := \left\{ f \in X : \|f\|_{\mathcal{L}_1(\mathbb{D})} < \infty \right\}.$$

Here $\|f\|_{\mathcal{L}_1(\mathbb{D})}$ is sometimes known as the variation of f with respect to $\mathbb{D}$ in the classical literature (see Ref. [39]). For the OGA 1.3 with $f \in \mathcal{L}_1(\mathbb{D})$ and $\|\mathbb{D}\| := \sup_{g \in \mathbb{D}} \|g\| \le 1$, the classical error estimate (see Refs. [23] and [4]) reads

$$\|f - f_n\| \le \|f\|_{\mathcal{L}_1(\mathbb{D})}(n+1)^{-\frac{1}{2}}. \tag{4.1}$$

Recently, the work[56] developed a rate-optimal convergence estimate of the OGA based on the entropy numbers of $\mathrm{co}(\mathbb{D})$. For example, convergence of the OGA based on a sufficiently regular dictionary like $\mathbb{P}_k^d$ would be much faster than $\mathcal{O}(n^{-\frac{1}{2}})$.

Theorem 4.1 (Theorem 1 from Ref. [56]). *Let $f \in \mathcal{L}_1(\mathbb{D})$ and $t > 0$. For Algorithm 1.3 it holds that*

$$\varepsilon_n(\mathrm{co}(\mathbb{D})) \lesssim n^{-\frac{1}{2}-t} \implies \|f - f_n\| \lesssim \|f\|_{\mathcal{L}_1(\mathbb{D})} n^{-\frac{1}{2}-t}.$$

4.1. *Direct comparison estimate*

Although the OGA and the greedy-type RBM are not closely related research areas, we show that the main lemma 2.1 used in the analysis of greedy RBMs can be used to derive a new type of clean and transparent *direct* comparison between $\|f - f_n\|$ and $\varepsilon_n(\mathrm{co}(\mathbb{D}))$. It turns out that our analysis greatly simplifies the argument in Ref. [56] and leads to an improvement of Theorem 4.1. To this end, we need the following upper bound for recursively related sequences. We note that an equivalent result has previously appeared in a slightly different form in the literature, see Lemma 3.1 in Ref. [62] or Lemma 3.4 in Ref. [23]. For the readers convenience we provide the complete proof here.

Lemma 4.1. *Let $\{a_n\}_{n\ge 0}$ and $\{b\}_{n\ge 1}$ be non-negative sequences satisfying*

$$a_n \le a_{n-1}(1 - b_n a_{n-1}) \text{ for } n \ge 1. \tag{4.2}$$

and $b_0 = 1/a_0$. Then we have

$$a_n \le \frac{1}{b_0 + b_1 + \cdots + b_n}.$$

Proof. We prove the lemma by induction. First we note $\{a_n\}_{n\ge 0}$ is a decreasing sequence. By definition we have $a_0 \le 1/b_0$ for $n = 0$. For the time being assume

$$a_{n-1} \le \frac{1}{b_0 + \cdots + b_{n-1}}. \tag{4.3}$$

If $a_{n-1} \leq \frac{1}{b_0 + \cdots + b_n}$, then the monotonicity of $\{a_n\}_{n \geq 1}$ implies

$$a_n \leq a_{n-1} \leq \frac{1}{b_0 + \cdots + b_n}.$$

If $a_{n-1} > \frac{1}{b_0 + \cdots + b_n}$, then by (4.2) and the induction assumption (4.3), we obtain

$$
\begin{aligned}
a_n &\leq \frac{1}{b_0 + \cdots + b_{n-1}} \left(1 - \frac{b_n}{b_0 + \cdots + b_n} \right) \\
&= \frac{1}{b_0 + \cdots + b_n}.
\end{aligned}
$$

Therefore the induction is complete and Lemma 4.1 is true. $\qquad\square$

Theorem 4.2. *For $f \in \mathcal{L}_1(\mathbb{D})$ and Algorithm 1.3 we have*

$$\|f - f_n\| \leq \frac{(n! V_n)^{\frac{1}{n}}}{\sqrt{n}} \|f\|_{\mathcal{L}_1(\mathbb{D})} \varepsilon_n(\mathrm{co}(\mathbb{D})). \tag{4.4}$$

Proof. Let $r_n = f - f_n$ in Algorithm 1.3. Following the analysis in Ref. 56 (see also the analysis in Section 3 of Ref. 31), we have

$$
\begin{aligned}
\|r_n\|^2 &\leq \left\| r_{n-1} - \frac{\langle r_{n-1}, g_n - P_{n-1} g_n \rangle}{\|g_n - P_{n-1} g_n\|^2} (g_n - P_{n-1} g_n) \right\|^2 \\
&= \|r_{n-1}\|^2 - \frac{|\langle r_{n-1}, g_n - P_{n-1} g_n \rangle|^2}{\|g_n - P_{n-1} g_n\|^2}.
\end{aligned}
\tag{4.5}
$$

We remark $\|g_n - P_{n-1} g_n\| \neq 0$ otherwise $g_n \in X_{n-1} = \mathrm{span}\{g_1, \ldots, g_{n-1}\}$ and $\arg\max_{g \in \mathbb{D}} |\langle r_{n-1}, g \rangle| = |\langle r_{n-1}, g_n \rangle| = 0$, which implies $r_{n-1} = 0$ and termination of the OGA. On the other hand, using $r_{n-1} \perp X_{n-1}$ we have

$$
\begin{aligned}
\|r_{n-1}\|^2 &= \langle f, r_{n-1} \rangle \leq \|f\|_{\mathcal{L}_1(\mathbb{D})} |\langle r_{n-1}, g_n \rangle| \\
&= \|f\|_{\mathcal{L}_1(\mathbb{D})} |\langle r_{n-1}, g_n - P_{n-1} g_n \rangle|.
\end{aligned}
\tag{4.6}
$$

Therefore combining (4.5) and (4.6) leads to

$$\|r_n\|^2 \leq \|r_{n-1}\|^2 - \frac{\|r_{n-1}\|^4}{\|f\|_{\mathcal{L}_1(\mathbb{D})}^2 \|g_n - P_{n-1} g_n\|^2}. \tag{4.7}$$

With the notation

$$
\begin{aligned}
a_n &= \|f - f_n\|^2 / \|f\|_{\mathcal{L}_1(\mathbb{D})}^2, \\
b_n &= \|g_n - P_{n-1} g_n\|^{-2},
\end{aligned}
$$

the inequality (4.7) reduces to the recurrence relation

$$a_n \leq a_{n-1}(1 - b_n a_{n-1}). \tag{4.8}$$

By the definition of $\|f\|_{\mathcal{L}_1(\mathbb{D})}$ we have $a_0 \leq 1$. It then follows from (4.8) and Lemma 4.1 that

$$a_n \leq \frac{1}{1 + b_1 + \cdots + b_n}. \tag{4.9}$$

18 *Y. Li & J. W. Siegel*

Using (4.9), a mean value inequality and the key Lemma 2.1, we obtain

$$
\begin{aligned}
a_n &\leq \frac{1}{1 + n(b_1 \cdots b_n)^{\frac{1}{n}}} \\
&= \frac{1}{1 + n\big(\prod_{k=1}^{n} \|g_k - P_{k-1}g_k\| \big)^{-\frac{2}{n}}} \\
&\leq \frac{1}{1 + n(nV_n)^{-\frac{2}{n}} \varepsilon_n(\mathrm{co}(\mathbb{D}))^{-2}} \\
&\leq \frac{(nV_n)^{\frac{2}{n}}}{n} \varepsilon_n(\mathrm{co}(\mathbb{D}))^2.
\end{aligned}
$$

The proof is complete. $\qquad\square$

We remark that the classical Theorem 4.1 is valid under an asymptotic assumption while Theorem 4.2 is a non-asymptotic and unconditional estimate. In addition, the constant in the upper bound of (4.4) as $n \to \infty$ behaves like

$$
\lim_{n \to \infty} \frac{(n! V_n)^{\frac{1}{n}}}{\sqrt{n}} = \sqrt{\frac{2\pi}{\mathrm{e}}}.
$$

Therefore Theorem 4.2 yields

$$
\|f - f_n\| \lesssim \|f\|_{\mathcal{L}_1(\mathbb{D})} \varepsilon_n(\mathrm{co}(\mathbb{D})),
$$

which is indeed stronger than Theorem 4.1.

4.2. *Best n-term approximation*

Consider the set of n-sparse elements

$$
\Sigma_n(\mathbb{D}) = \Big\{ \sum_{i=1}^{n} c_i \tilde{g}_i : c_i \in \mathbb{R}, \ \tilde{g}_i \in \mathbb{D}, \ i = 1, \ldots, n \Big\}.
$$

It is natural to compare the error of greedy algorithms with the best approximation error

$$
E_n(f, \mathbb{D}) = \inf_{g \in \Sigma_n(\mathbb{D})} \|f - g\|
$$

by n-sparse elements. When $\mathbb{D}$ is an orthonormal set in X, the equality $\|f - f_n\| = E_n(f, \mathbb{D})$ automatically holds (see Ref. [64]). In general, such convergence analysis using $E_n(f, \mathbb{D})$ often requires near orthogonality assumption on $\mathbb{D}$. For example, on a M-coherent dictionary $\mathbb{D}$, the following Lebesgue inequality is true (see Section 2.6 of Ref. [64]):

$$
\|f - f_{\lfloor n \log n \rfloor}\| \leq c_1 E_n(f, \mathbb{D}),
$$

for $n \leq c_2 M^{-\frac{2}{3}}$ with c_1, c_2 being explicit constants. We also refer to Ref. [61] for Lebesgue-type inequality of generalized OGAs in Banach spaces. On the other hand,

the entropy-based estimate (4.4) applies to any normalized dictionary in Hilbert spaces.

In addition, we note that Theorem 4.4 yields the following direct comparison between $E_n(f, \mathbb{D})$ and $\varepsilon_n(\mathrm{co}(\mathbb{D}))$:

$$E_n(f, \mathbb{D}) \leq \frac{(n! V_n)^{\frac{1}{n}}}{\sqrt{n}} \|f\|_{\mathcal{L}_1(\mathbb{D})} \varepsilon_n(\mathrm{co}(\mathbb{D})), \tag{4.10a}$$

$$E_n(\mathbb{D}) := \sup_{f \in \mathrm{co}(\mathbb{D})} E_n(f, \mathbb{D}) \leq \frac{(n! V_n)^{\frac{1}{n}}}{\sqrt{n}} \varepsilon_n(\mathrm{co}(\mathbb{D})). \tag{4.10b}$$

When the number $|\mathbb{D}|$ of elements in $\mathbb{D}$ is finite, an indirect comparison between $E_n(\mathbb{D})$ and $\varepsilon_n(\mathrm{co}(\mathbb{D}))$ in the converse direction of (4.10b) can be found in Section 7.4 of Ref. 61. In particular, combining Theorem 7.4.3 of Ref. 61 with Theorem 4.4, we obtain the following novel comparison between the convergence rate of $E_n(\mathbb{D})$ and the error of OGAs whose target functions belonging to $\mathrm{co}(\mathbb{D})$.

Corollary 4.1. *Assume there exists a constant $s > 0$ such that*

$$E_n(\mathbb{D}) \lesssim n^{-s}, \quad n \leq |\mathbb{D}|,$$

where the cardinality $|\mathbb{D}|$ is finite. Then for f_n generated by the OGA (Algorithm 1.3) with $n \leq |\mathbb{D}|$, we have

$$\sup_{f \in \mathrm{co}(\mathbb{D})} \|f - f_n\| \leq C(s)(\log(2|\mathbb{D}|/n))^s n^{-s},$$

where $C(s)$ is a constant depending only on s.

We note that the above result is independent of the coherence property of $\mathbb{D}$. Moreover, Corollary 4.1 is not a Lebesgue-type inequality but is in a similar fashion to the classical n-width based convergence estimate (1.1) for the RBM. It confirms that the OGA collectively achieves the best n-term approximation rate up to a logarithmic factor on finite dictionaries.

4.3. *General target functions*

In general, if f is not contained in $\mathcal{L}_1(\mathbb{D})$, one can use the next corollary to estimate the error of the OGA.

Corollary 4.2. *For all $f \in X$ and any $h \in \mathcal{L}_1(\mathbb{D})$, Algorithm 1.3 satisfies*

$$\|f - f_n\|^2 \leq \|f - h\|^2 + 4 \frac{(n! V_n)^{\frac{2}{n}}}{n} \|h\|_{\mathcal{L}_1(\mathbb{D})}^2 \varepsilon_n(\mathrm{co}(\mathbb{D}))^2.$$

Proof. We use the same notation as in the proof of Theorem 4.2. We note (4.5) still holds and

$$\begin{aligned}
\|r_{n-1}\|^2 &= \langle h, r_{n-1} \rangle + \langle f - h, r_{n-1} \rangle \\
&\leq \|h\|_{\mathcal{L}_1(\mathbb{D})} |\langle r_{n-1}, g_n - P_{n-1} g_n \rangle| + \|f - h\| \|r_{n-1}\|.
\end{aligned}$$

Then combining the above inequality and a mean value inequality yields

$$\frac{1}{2}\left(\|r_{n-1}\|^2 - \|f - h\|^2\right) \leq \|h\|_{\mathcal{L}_1(\mathbb{D})}|\langle r_{n-1}, g_n - P_{n-1}g_n\rangle|. \qquad (4.11)$$

Let $c_n = (\|r_n\|^2 - \|f - h\|^2)/(4\|h\|_{\mathcal{L}_1(\mathbb{D})}^2)$. Using (4.5) and (4.11) we obtain

$$c_n \leq c_{n-1}(1 - b_n c_{n-1}).$$

The rest of the proof is same as Theorem 4.2. $\qquad\square$

For $t > 0$ and $g \in X$ the K-functional

$$K(t, g) = \inf_{h \in \mathcal{L}_1(\mathbb{D})} \left(\|g - h\|_X + t\|h\|_{\mathcal{L}_1(\mathbb{D})}\right)$$

measures how well the element of X can be approximated by $\mathcal{L}_1(\mathbb{D})$ with an approximant of small $\mathcal{L}_1(\mathbb{D})$ norm (see Refs. 6, 22 and 9). For the index pair (θ, ∞) with $\theta \in (0, 1)$, the interpolation norm and space between X and $\mathcal{L}_1(\mathbb{D})$ are

$$\|g\|_\theta = \|g\|_{[X, \mathcal{L}_1(\mathbb{D})]_{\theta,\infty}} := \sup_{0 < t < \infty} t^{-\theta} K(t, g),$$

$$X_\theta = \left\{g \in X : \|g\|_\theta < \infty\right\},$$

respectively. As in Refs. 4 and 56, we also obtain an explicit and improved error estimate of the OGA in terms of the interpolation norm $\|f\|_\theta$.

Proposition 4.1. *For all $f \in X_\theta$ and $\theta \in (0, 1)$ Algorithm 1.3 satisfies*

$$\|f - f_n\| \leq 2^\theta \frac{(n! V_n)^{\frac{\theta}{n}}}{n^{\frac{\theta}{2}}} \|f\|_\theta \varepsilon_n(\mathrm{co}(\mathbb{D}))^\theta.$$

Proof. By Corollary 4.2, we obtain for all $h \in \mathcal{L}_1(\mathbb{D})$

$$\|f - f_n\| \leq \|f - h\| + 2\frac{(n! V_n)^{\frac{1}{n}}}{\sqrt{n}} \|h\|_{\mathcal{L}_1(\mathbb{D})} \varepsilon_n(\mathrm{co}(\mathbb{D})).$$

As a result, setting $t = 2\frac{(n! V_n)^{\frac{1}{n}}}{\sqrt{n}} \varepsilon_n(\mathrm{co}(\mathbb{D}))$, we have

$$\|f - f_n\| \leq K(t, f) \leq t^\theta \|f\|_\theta$$

and complete the proof. $\qquad\square$

5. Concluding Remarks

We have derived direct and optimal convergence estimates of the greedy-type RBM and the OGA based on Kolmogorov's entropy numbers. Constants in our upper bounds are explicit and simple. A future research direction is to extend our analysis to POD-greedy algorithms[33,34] for time-dependent parametric PDEs, which is a combination of greedy algorithms and POD for temporal compression of snapshots. In addition, the generalization of the OGA in Banach spaces is called the Chebyshev Greedy Algorithm (CGA).[18,63,64] It would also be interesting to investigate if our analysis in Section 4 and Lemma 3.1 can be applied to the CGA.

Acknowledgements

References

1. A. Ammar, B. Mokdad, F. Chinesta and R. Keunings, A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids, *Journal of non-Newtonian fluid Mechanics* **139** (2006) 153–176.
2. I. Babuška and W. C. Rheinboldt, Error estimates for adaptive finite element computations, *SIAM J. Numer. Anal.* **15** (1978) 736–754.
3. M. Barrault, Y. Maday, N. C. Nguyen and A. T. Patera, An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, *C. R. Math. Acad. Sci. Paris* **339** (2004) 667–672.
4. A. R. Barron, A. Cohen, W. Dahmen and R. A. DeVore, Approximation and learning by greedy algorithms, *Ann. Statist.* **36** (2008) 64–94.
5. P. Benner, S. Gugercin and K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, *SIAM Rev.* **57** (2015) 483–531.
6. J. Bergh and J. Löfström, *Interpolation spaces. An introduction*, Grundlehren der Mathematischen Wissenschaften, No. 223 (Springer-Verlag, Berlin-New York, 1976).
7. 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.
8. A. Bonito, R. A. DeVore and R. H. Nochetto, Adaptive finite element methods for elliptic problems with discontinuous coefficients, *SIAM J. Numer. Anal.* **51** (2013) 3106–3134.
9. S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, volume 35 of *Texts in Applied Mathematics, 15* (Springer, New York, 2008), 3 edition.
10. A. Buffa, Y. Maday, A. T. Patera, C. Prud'homme and G. Turinici, *A priori* convergence of the greedy algorithm for the parametrized reduced basis method, *ESAIM Math. Model. Numer. Anal.* **46** (2012) 595–603.
11. B. Carl, Entropy numbers, *s*-numbers, and eigenvalue problems, *J. Functional Analysis* **41** (1981) 290–306.
12. B. Carl, I. Kyrezi and A. Pajor, Metric entropy of convex hulls in Banach spaces, *J. London Math. Soc. (2)* **60** (1999) 871–896.
13. J. M. Cascon, C. Kreuzer, R. H. Nochetto and K. G. Siebert, Quasi-optimal convergence rate for an adaptive finite element method, *SIAM J. Numer. Anal.* **46** (2008) 2524–2550.
14. A. Cohen, W. Dahmen and R. DeVore, Adaptive wavelet methods for elliptic operator equations: convergence rates, *Math. Comp.* **70** (2001) 27–75.
15. A. Cohen and R. DeVore, Approximation of high-dimensional parametric PDEs, *Acta Numer.* **24** (2015) 1–159.
16. A. Cohen and R. DeVore, Kolmogorov widths under holomorphic mappings, *IMA J. Numer. Anal.* **36** (2016) 1–12.

17. G. Davis, S. Mallat and M. Avellaneda, Adaptive greedy approximations, *Constr. Approx.* **13** (1997) 57–98.
18. A. V. Dereventsov and V. N. Temlyakov, A unified way of analyzing some greedy algorithms, *J. Funct. Anal.* **277** (2019) 108286, 30.
19. R. DeVore, G. Petrova and P. Wojtaszczyk, Greedy algorithms for reduced bases in Banach spaces, *Constr. Approx.* **37** (2013) 455–466.
20. R. A. DeVore, Nonlinear approximation, in *Acta numerica, 1998* (Cambridge Univ. Press, Cambridge, 1998), volume 7 of *Acta Numer.*, pp. 51–150.
21. R. A. DeVore, The theoretical foundation of reduced basis methods, in *Model Reduction and Approximation (SIAM Computational Science & Engineering)* (2014).
22. R. A. DeVore and G. G. Lorentz, *Constructive approximation*, volume 303 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* (Springer-Verlag, Berlin, 1993).
23. R. A. DeVore and V. N. Temlyakov, Some remarks on greedy algorithms, *Adv. Comput. Math.* **5** (1996) 173–187.
24. P. G. L. Dirichlet, Sur une nouvelle méthode pour la détermination des intégrales multiples, *Journal de Mathématiques Pures et Appliquées* **4** (1839) 164–168.
25. P. Edel and Y. Maday, Dual natural-norm a posteriori error estimators for reduced basis approximations to parametrized linear equations, *Math. Models Methods Appl. Sci.* .
26. L. E. Figueroa and E. Süli, Greedy approximation of high-dimensional Ornstein–Uhlenbeck operators, *Foundations of Computational Mathematics* **12** (2012) 573–623.
27. J. H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Statist.* **29** (2001) 1189–1232.
28. J. H. Friedman and W. Stuetzle, Projection pursuit regression, *Journal of the American statistical Association* **76** (1981) 817–823.
29. T. Gantumur, H. Harbrecht and R. Stevenson, An optimal adaptive wavelet method without coarsening of the iterands, *Math. Comp.* **76** (2007) 615–629.
30. F. Gao, Metric entropy of convex hulls, *Israel J. Math.* **123** (2001) 359–364.
31. Y. Gao, T. Qian, L.-F. Cao and V. Temlyakov, Aspects of 2D-adaptive Fourier decompositions, *arXiv preprint* arXiv:1710.09277.
32. P. Grisvard, *Singularities in boundary value problems*, Research in Applied Mathematics, 22 (Springer-Verlag, Berlin, 1992).
33. B. Haasdonk, Convergence rates of the POD-greedy method, *ESAIM Math. Model. Numer. Anal.* **47** (2013) 859–873.
34. B. Haasdonk and M. Ohlberger, Reduced basis method for finite volume approximations of parametrized linear evolution equations, *M2AN Math. Model. Numer. Anal.* **42** (2008) 277–302.
35. J. S. Hesthaven, G. Rozza and B. Stamm, *Certified reduced basis methods for parametrized partial differential equations*, SpringerBriefs in Mathematics (Springer, Cham; BCAM Basque Center for Applied Mathematics, Bilbao, 2016), bCAM SpringerBriefs.
36. L. K. Jones, On a conjecture of Huber concerning the convergence of projection pursuit regression, *The Annals of statistics* (1987) 880–882.
37. L. K. Jones, A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, *Ann. Statist.* **20** (1992) 608–613.
38. A. N. Kolmogorov and V. M. Tihomirov, ε-entropy and ε-capacity of sets in functional space, *Amer. Math. Soc. Transl. (2)* **17** (1961) 277–364.
39. V. Kurková and M. Sanguineti, Bounds on rates of variable-basis and neural-network

approximation, *IEEE Trans. Inform. Theory* **47** (2001) 2659–2665.

40. C. Le Bris, T. Lelievre and Y. Maday, Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations, *Constructive Approximation* **30** (2009) 621–651.

41. Y. Li, Some convergence and optimality results of adaptive mixed methods in finite element exterior calculus, *SIAM J. Numer. Anal.* **57** (2019) 2019–2042.

42. Y. Li, Quasi-optimal adaptive mixed finite element methods for controlling natural norm errors, *Math. Comp.* **90** (2021) 565–593.

43. E. D. Livshits, Lower bounds for the rate of convergence of greedy algorithms, *Izvestiya: Mathematics* **73** (2009) 1197.

44. G. G. Lorentz, Metric entropy, widths, and superpositions of functions, *Amer. Math. Monthly* **69** (1962) 469–485.

45. G. G. Lorentz, M. v. Golitschek and Y. Makovoz, *Constructive approximation*, volume 304 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* (Springer-Verlag, Berlin, 1996), advanced problems.

46. Y. Maday, Reduced basis method for the rapid and reliable solution of partial differential equations, in *International Congress of Mathematicians. Vol. III* (Eur. Math. Soc., Zürich, 2006), pp. 1255–1270.

47. Y. Maday, A. T. Patera and G. Turinici, A priori convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations, *J. Sci. Comput.* **17** (2002) 437–446.

48. S. G. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing* **41** (1993) 3397–3415.

49. P. Morin, K. G. Siebert and A. Veeser, A basic convergence result for conforming adaptive finite elements, *Math. Models Methods Appl. Sci.* **18** (2008) 707–737.

50. Y. Pati, R. Rezaiifar and P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in *Proceedings of 27th Asilomar conference on signals, systems and computers* (1993), pp. 40–44. IEEE.

51. G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, *Applied and Computational Harmonic Analysis* **41** (2016) 852–866.

52. A. Quarteroni, A. Manzoni and F. Negri, *Reduced basis methods for partial differential equations*, volume 92 of *Unitext* (Springer, Cham, 2016), an introduction, La Matematica per il 3+2.

53. 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. Methods Eng.* **15** (2008) 229–275.

54. C. Schütt, Entropy numbers of diagonal operators between symmetric Banach spaces, *J. Approx. Theory* **40** (1984) 121–128.

55. J. W. Siegel, Q. Hong, X. Jin, W. Hao and J. Xu, Greedy training algorithms for neural networks and applications to pdes, *J. Comput. Phys.* **484** (2023) 112084.

56. J. W. Siegel and J. Xu, Optimal convergence rates for the orthogonal greedy algorithm, *IEEE Trans. Inform. Theory* **68** (2022) 3354–3361.

57. J. W. Siegel and J. Xu, Sharp bounds on the approximation rates, metric entropy, and n-widths of shallow neural networks, *Found. Comput. Math.* .

58. A. Sil'nichenko, Rate of convergence of greedy algorithms, *Mathematical Notes* **76** (2004) 582–586.

59. I. Steinwart, Entropy of $C(K)$-valued operators, *J. Approx. Theory* **103** (2000) 302–328.

60. V. Temlyakov, *Greedy approximation*, volume 20 of *Cambridge Monographs on Applied*

and Computational Mathematics (Cambridge University Press, Cambridge, 2011).
61. V. Temlyakov, *Multivariate approximation*, volume 32 of *Cambridge Monographs on Applied and Computational Mathematics* (Cambridge University Press, Cambridge, 2018).
62. V. N. Temlyakov, Weak greedy algorithms, *Adv. Comput. Math.* **12** (2000) 213–227.
63. V. N. Temlyakov, Greedy algorithms in Banach spaces, *Adv. Comput. Math.* **14** (2001) 277–292.
64. V. N. Temlyakov, Greedy approximation, *Acta Numer.* **17** (2008) 235–409.
65. X. Wang, Volumes of generalized unit balls, *Mathematics Magazine* **78** (2005) 390–395.
66. P. Wojtaszczyk, *Banach spaces for analysts*, volume 25 of *Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge, 1991).
67. P. Wojtaszczyk, On greedy algorithm approximating Kolmogorov widths in Banach spaces, *J. Math. Anal. Appl.* **424** (2015) 685–695.