



INTERDISCIPLINARY
MATHEMATICS
INSTITUTE

2013:02

An inequality for the entropy
numbers and its application

V.N. Temlyakov

IMI

PREPRINT SERIES

COLLEGE OF ARTS AND SCIENCES
UNIVERSITY OF SOUTH CAROLINA

An inequality for the entropy numbers and its application

V.N. Temlyakov *

January 31, 2013

Abstract

We prove an inequality for the entropy numbers in terms of non-linear Kolmogorov's widths. This inequality is in a spirit of known inequalities of this type and it is adjusted to the form convenient in applications for m -term approximations with respect to a given system. Also, we obtain upper bounds for the m -term approximation by the Weak Relaxed Greedy Algorithm with respect to a system which is not a dictionary.

1 Introduction

This paper was motivated by the very recent paper [3]. The authors of [3] study the entropy and best m -term approximation of the ℓ_q -hulls of finite systems of elements in the L_p spaces. They conduct this study by probabilistic methods. In this context probabilistic methods were used in some earlier papers, for instance, in [2]. Here we demonstrate how known results from greedy approximation in Banach spaces combined with known technique of general inequalities for the entropy numbers allow us to obtain similar results. Moreover, we show that the use of a greedy algorithm allows us to provide a deterministic construction of good m -term approximants.

*University of South Carolina and Steklov Institute of Mathematics. Research was supported by NSF grant DMS-1160841

A number of different widths are being studied in approximation theory: Kolmogorov widths, linear widths, Fourier widths, Gel'fand widths, Alexandrov widths and others. All these widths were introduced in approximation theory as characteristics of function classes (more generally compact sets) which give the best possible accuracy of algorithms with certain restrictions. For instance, Kolmogorov's n -width for centrally symmetric compact set F in a Banach space X is defined as follows

$$d_n(F, X) := \inf_L \sup_{f \in F} \inf_{g \in L} \|f - g\|_X$$

where \inf_L is taken over all n -dimensional subspaces of X . In other words the Kolmogorov n -width gives the best possible error in approximating a compact set F by n -dimensional linear subspaces.

There has been an increasing interest last decades in nonlinear m -term approximation with regard to different systems. In [4] we generalized the concept of classical Kolmogorov's width in order to use it in estimating best m -term approximation. For this purpose we introduced a nonlinear Kolmogorov's (N, m) -width:

$$d_m(F, X, N) := \inf_{\mathcal{L}_N, \#\mathcal{L}_N \leq N} \sup_{f \in F} \inf_{L \in \mathcal{L}_N} \inf_{g \in L} \|f - g\|_X,$$

where \mathcal{L}_N is a set of at most N m -dimensional subspaces L . It is clear that

$$d_m(F, X, 1) = d_m(F, X).$$

The new feature of $d_m(F, X, N)$ is that we allow to choose a subspace $L \in \mathcal{L}_N$ depending on $f \in F$. It is clear that the bigger N the more flexibility we have to approximate f . It turns out that from the point of view of our applications the two cases

$$N \asymp K^m, \tag{1.1}$$

where $K > 1$ is a constant, and

$$N \asymp m^{am}, \tag{1.2}$$

where $a > 0$ is a fixed number, play an important role.

It is known (see [6]) that the (N, m) -widths can be used for estimating from below the best m -term approximations. Let X be a Banach space and let B_X denote the unit ball of X with the center at 0. Denote by $B_X(y, r)$ a

ball with center y and radius r : $\{x \in X : \|x - y\| \leq r\}$. For a compact set A and a positive number ϵ we define the covering number $N_\epsilon(A, X)$ as follows

$$N_\epsilon(A, X) := \min\{n : \exists y^1, \dots, y^n : A \subseteq \cup_{j=1}^n B_X(y^j, \epsilon)\}.$$

It is convenient to consider along with the entropy $H_\epsilon(A, X) := \log N_\epsilon(A, X)$ (here and later $\log := \log_2$) the entropy numbers $\epsilon_k(A, X)$:

$$\epsilon_k(A, X) := \inf\{\epsilon : \exists y^1, \dots, y^{2^k} \in X : A \subseteq \cup_{j=1}^{2^k} B_X(y^j, \epsilon)\}.$$

There are several general results (see [1]) which give lower estimates of the Kolmogorov widths $d_n(F, X)$ in terms of the entropy numbers $\epsilon_k(F, X)$. The Carl's (see [1]) inequality states: for any $r > 0$ we have

$$\max_{1 \leq k \leq n} k^r \epsilon_k(F, X) \leq C(r) \max_{1 \leq m \leq n} m^r d_{m-1}(F, X). \quad (1.3)$$

We proved in [4] (see also [7], Section 3.5) the inequality

$$\max_{1 \leq k \leq n} k^r \epsilon_k(F, X) \leq C(r, K) \max_{1 \leq m \leq n} m^r d_{m-1}(F, X, K^m), \quad (1.4)$$

where we denote

$$d_0(F, X, N) := \sup_{f \in F} \|f\|_X.$$

This inequality is a generalization of inequality (1.3). We also discussed in [4] and in Section 3.5 of [7] the possibility of replacing K^m by $(Kn/m)^m$ in (1.4). The corresponding remarks (Remark 2.1 in [4] and Remark 3.5 in [7]) should read as follows.

Remark 1.1. *Examining the proof of (1.4) one can check that the following inequality holds*

$$n^r \epsilon_n(F, X) \leq C(r, K) \max_{1 \leq m \leq n} m^r d_{m-1}(F, X, (Kn/m)^m).$$

In Section 2 we prove an upper bound for $\epsilon_k(F, X)$ for all $k \leq n$.

In Section 3 we demonstrate how the general inequality from Theorem 2.1 can be used in estimating the entropy numbers of different compacts. In particular, Corollary 3.3 gives a new proof of the corresponding upper bounds from Theorem 1 in [3].

In Section 4 we study the Weak Relaxed Greedy Algorithm with respect to a system which is not a dictionary. In particular, results of Section 4 provide an algorithm which gives the same upper bounds for the best m -term approximation as those obtained in [3].

2 A general inequality

Theorem 2.1. *Let a compact $F \subset X$ and a number $r > 0$ be such that for some $n \in \mathbb{N}$*

$$d_{m-1}(F, X, (Kn/m)^m) \leq m^{-r}, \quad m \leq n.$$

Then for $k \leq n$

$$\epsilon_k(F, X) \leq C(r, K) \left(\frac{\log(2n/k)}{k} \right)^r.$$

Proof. Let $X(N, m)$ denote the union of not more than N subspaces L with $\dim L \leq m$. Consider a collection $\mathcal{K}(l) := \{X((Kn2^{-s-1})^{2^{s+1}}, 2^{s+1})\}_{s=1}^l$, $2^{l+1} \leq n$ and denote

$$H^r(\mathcal{K}(l)) := \{f \in X : \exists L_1(f), \dots, L_l(f) : L_s(f) \in X((Kn2^{-s-1})^{2^{s+1}}, 2^{s+1}),$$

and $\exists t_s(f) \in L_s(f)$ such that

$$\|t_s(f)\|_X \leq 2^{-r(s-1)}, \quad s = 1, \dots, l; \quad \|f - \sum_{s=1}^l t_s(f)\|_X \leq 2^{-rl}.\}$$

Lemma 2.1. *We have for $r > 0$*

$$\epsilon_{2^l}(H^r(\mathcal{K}(l)), X) \leq C(r, K) 2^{-rl} (\log(Kn2^{-l}))^r, \quad 2^{l+1} \leq n.$$

Proof. We use a well known result (see, for instance, [7], p. 145) to estimate $\epsilon_n(B_X, X)$ of the unit ball B_X in the d -dimensional space X :

$$\epsilon_n(B_X, X) \leq 3(2^{-n/d}). \tag{2.1}$$

Take any sequence $\{n_s\}_{s=1}^{l(r)}$ of $l(r) \leq l - 2$ nonnegative integers. We will specify $l(r)$ later. Construct ϵ_{n_s} -nets consisting of 2^{n_s} points each for all unit balls of the spaces in $X((Kn2^{-s-1})^{2^{s+1}}, 2^{s+1})$. Then the total number of the elements y_j^s in these ϵ_{n_s} -nets does not exceed

$$M_s := (Kn2^{-s-1})^{2^{s+1}} 2^{n_s}.$$

We now consider the set A of elements of the form

$$y_{j_1}^1 + 2^{-r} y_{j_2}^2 + \dots + 2^{-r(l(r)-1)} y_{j_{l(r)}}^{l(r)}, \quad j_s \in [1, M_s], \quad s = 1, \dots, l(r).$$

The total number of these elements does not exceed

$$M = \prod_{s=1}^{l(r)} M_s, \quad \log M \leq \sum_{s=1}^{l(r)} 2^{s+1} \log(Kn2^{-s-1}) + \sum_{s=1}^{l(r)} n_s.$$

It is easy to see that

$$\sum_{s=1}^{l(r)} 2^{s+1} \log(Kn2^{-s-1}) \leq C_1 2^{l(r)} \log(Kn2^{-l(r)}).$$

We now set

$$n_s := [(r+1)(l-s)2^{s+1}], \quad s = 1, \dots, l(r),$$

where $[x]$ denotes the integer part of a number x . We choose $l(r) \leq l-2$ as a maximal natural number satisfying

$$\sum_{s=1}^{l(r)} n_s \leq 2^{l-1}$$

and

$$C_1 2^{l(r)} \log(Kn2^{-l(r)}) \leq 2^{l-1}.$$

It is clear that

$$2^{l(r)} \geq C_2 2^l (\log(Kn2^{-l}))^{-1}. \quad (2.2)$$

Then we have

$$M \leq 2^{2^l}.$$

For the error $\epsilon(f)$ of approximation of $f \in H^r(\mathcal{K}(l))$ by elements of A we have

$$\begin{aligned} \epsilon(f) &\leq 2^{-rl} + \sum_{s=1}^{l(r)} \|t_s(f) - 2^{-r(s-1)} y_{j_s}^s\|_X + \sum_{s=l(r)+1}^l \|t_s(f)\|_X \\ &\leq C(r) 2^{-rl(r)} + \sum_{s=1}^{l(r)} 2^{-r(s-1)} \epsilon_{n_s}(B_{L_s}(f), X) \\ &\leq C(r) 2^{-rl(r)} + 3 \sum_{s=1}^{l(r)} 2^{-r(s-1)} 2^{-n_s/2^{s+1}} \leq C(r) 2^{-rl(r)}. \end{aligned}$$

Taking into account (2.2) we complete the proof of Lemma 2.1. \square

We continue the proof of Theorem 2.1. Without loss of generality assume

$$\max_{1 \leq m \leq n} m^r d_{m-1}(F, X, (Kn/m)^m) < 1/2.$$

Then for $s = 1, 2, \dots, l$; $l \leq [\log(n-1)]$ we have

$$d_{2^s}(F, X, (Kn2^{-s})^{2^s}) < 2^{-rs-1}.$$

This means that for each $s = 1, 2, \dots, l$, there is a collection \mathcal{L}_s of $(Kn2^{-s})^{2^s}$ 2^s -dimensional spaces $L_j^s, j = 1, \dots, (Kn2^{-s})^{2^s}$, such that for each $f \in F$ there exists a subspace $L_{j_s}^s(f)$ and an approximant $a_s(f) \in L_{j_s}^s(f)$ such that

$$\|f - a_s(f)\| \leq 2^{-rs-1}.$$

Consider

$$t_s(f) := a_s(f) - a_{s-1}(f), \quad s = 2, \dots, l. \quad (2.3)$$

Then we have

$$t_s(f) \in L_{j_s}^s(f) \oplus L_{j_{s-1}}^{s-1}(f), \quad \dim(L_{j_s}^s(f) \oplus L_{j_{s-1}}^{s-1}(f)) \leq 2^s + 2^{s-1} < 2^{s+1}.$$

Note that for K large enough

$$(Kn2^{-s})^{2^s} (Kn2^{-s+1})^{2^{s-1}} \leq (Kn2^{-s-1})^{2^{s+1}}.$$

Let $X((Kn2^{-s-1})^{2^{s+1}}, 2^{s+1})$ denote the collection of all $L_{j_s}^s \oplus L_{j_{s-1}}^{s-1}$ over various $1 \leq j_s \leq (Kn2^{-s})^{2^s}$; $1 \leq j_{s-1} \leq (Kn2^{-s+1})^{2^{s-1}}$. For $t_s(f)$ defined by (2.3) we have

$$\|t_s(f)\| \leq 2^{-rs-1} + 2^{-r(s-1)-1} \leq 2^{-r(s-1)}.$$

Next, for $a_1(f) \in L^1(f)$ we have

$$\|f - a_1(f)\| \leq 1/2$$

and from $d_0(F, X) \leq 1/2$ we get

$$\|a_1(f)\| \leq 1.$$

Take $t_1(f) = a_1(f)$. Then we have $F \subset H^r(\mathcal{K}(l))$ and Lemma 2.1 gives the required bound

$$\epsilon_{2^l}(F) \leq C(r, K)2^{-rl}(\log(Kn2^{-l}))^r, \quad 1 \leq l \leq [\log(n-1)].$$

It is clear that these inequalities imply the conclusion of Theorem 2.1. \square

3 Applications

We begin with an application which motivated a study of $d_m(F, X, N)$ with $N = (Kn/m)^m$. Let $\mathcal{D} = \{g_j\}_{j=1}^n$ be a system of normalized elements of cardinality $|\mathcal{D}| = n$ in a Banach space X . Consider best m -term approximations of f with respect to \mathcal{D}

$$\sigma_m(f, \mathcal{D})_X := \inf_{\{c_j\}; |\Lambda|=m} \left\| f - \sum_{j \in \Lambda} c_j g_j \right\|.$$

For a function class F set

$$\sigma_m(F, \mathcal{D})_X := \sup_{f \in F} \sigma_m(f, \mathcal{D})_X.$$

Then it is clear that for any system \mathcal{D} , $|\mathcal{D}| = n$,

$$d_m(F, X, \binom{n}{m}) \leq \sigma_m(F, \mathcal{D})_X.$$

Next,

$$\binom{n}{m} \leq (en/m)^m.$$

Thus Theorem 2.1 implies the following theorem.

Theorem 3.1. *Let a compact $F \subset X$ be such that there exists a normalized system \mathcal{D} , $|\mathcal{D}| = n$, and a number $r > 0$ such that*

$$\sigma_m(F, \mathcal{D})_X \leq m^{-r}, \quad m \leq n.$$

Then for $k \leq n$

$$\epsilon_k(F, X) \leq C(r) \left(\frac{\log(2n/k)}{k} \right)^r. \quad (3.1)$$

Remark 3.1. *Suppose that a compact F from Theorem 3.1 belongs to an n -dimensional subspace $X_n := \text{span}(\mathcal{D})$. Then in addition to (3.1) we have for $k \geq n$*

$$\epsilon_k(F, X) \leq C(r) n^{-r} 2^{-k/n}. \quad (3.2)$$

Proof. Inequality (3.2) follows from Theorem 3.1 with $X = X_n$, $k = n$, inequality (2.1) and a simple well known inequality

$$\epsilon_{k_1+k_2}(A, X_n) \leq \epsilon_{k_1}(A, X_n) \epsilon_{k_2}(B_{X_n}, X_n), \quad (3.3)$$

where A is a compact and B_{X_n} is a unit ball of X_n . \square

As a corollary of Theorem 3.1 and Remark 3.1 we obtain the following classical bound.

Corollary 3.1. *For any $0 < q \leq \infty$ and $\max(1, q) \leq p \leq \infty$ we have*

$$\epsilon_k(B_q^n, \ell_p^n) \leq C(q, p) \begin{cases} (\frac{\log(2n/k)}{k})^{1/q-1/p}, & k \leq n \\ 2^{-k/n} n^{1/p-1/q}, & k \geq n. \end{cases}$$

Proof. Indeed, it is well known and easy to check that for a sequence of nonnegative numbers $x_1 \geq x_2 \geq \dots \geq x_n$ we have for $0 < q \leq p$

$$\left(\sum_{j=m+1}^n x_j^p \right)^{1/p} \leq m^{\frac{1}{p}-\frac{1}{q}} \left(\sum_{j=1}^n x_j^q \right)^{1/q}. \quad (3.4)$$

Therefore, for $0 < q \leq p$

$$\sigma_m(B_q^n, \{e_j\}_{j=1}^n)_{\ell_p^n} \leq m^{\frac{1}{p}-\frac{1}{q}}, \quad m \leq n,$$

where $\{e_j\}_{j=1}^n$ is a canonical basis for \mathbb{R}^n . Applying Theorem 3.1 and Remark 3.1 we obtain Corollary 3.1. \square

For a normalized system \mathcal{D} define $A_q(\mathcal{D})$, $q > 0$, as a closure in X of the set

$$\{x : x = \sum_j c_j g_j, g_j \in \mathcal{D}, \sum_j |c_j|^q \leq 1\}.$$

Corollary 3.2. *Let $1 < p < \infty$. For a normalized system \mathcal{D} of cardinality $|\mathcal{D}| = n$ we have*

$$\epsilon_k(A_1(\mathcal{D}), L_p) \leq C(p) \left(\frac{\log(2n/k)}{k} \right)^{\max(\frac{1}{2}, \frac{1}{p})-1}, \quad k \leq n. \quad (3.5)$$

Proof. It is known (see [2] and [7]) that

$$\sigma_m(A_1(\mathcal{D}), \mathcal{D})_{L_p} \leq C(p) m^{\max(\frac{1}{2}, \frac{1}{p})-1}. \quad (3.6)$$

It remains to apply Theorem 3.1. \square

Corollary 3.3. *Let \mathcal{D} be a normalized system of cardinality $|\mathcal{D}| = n$. Then for $0 < q \leq 1$ and $1 < p < \infty$ we have*

$$\epsilon_k(A_q(\mathcal{D}), L_p) \leq C(q, p) \begin{cases} (\frac{\log(2n/k)}{k})^{1/q-\max(\frac{1}{2}, \frac{1}{p})}, & k \leq n \\ 2^{-k/n} n^{\max(\frac{1}{2}, \frac{1}{p})-1/q}, & k \geq n. \end{cases}$$

Proof. We estimate $\sigma_m(A_q(\mathcal{D}), \mathcal{D})_{L_p}$. If $q = 1$ then the bound is given by (3.6). If $q < 1$ then we use (3.4) with $p = 1$ and by (3.6) we get

$$\sigma_{2m}(A_q(\mathcal{D}), \mathcal{D})_{L_p} \leq C(q, p)m^{\max(\frac{1}{2}, \frac{1}{p}) - \frac{1}{q}}.$$

Applying Theorem 3.1 and Remark 3.1 we obtain Corollary 3.3. □

We note that Corollary 3.3 gives the same upper bounds as in Theorem 1 of [3]. It is proved in [3] that these bounds are best possible up to a constant.

4 A greedy algorithm

In Section 3 we showed how best m -term approximations can be used for estimating the entropy numbers. Here we note that m -term approximations are very important by themselves in the context of sparse approximation. In this context an important problem is to provide an algorithm that builds a good m -term approximation. We discuss a greedy algorithm in this section. The theory of greedy approximation is well developed (see [7]). A typical problem of greedy approximation is a problem of m -term approximation with respect to a dictionary. We say that a set of elements (functions) \mathcal{D} from a Banach space X is a dictionary, respectively, symmetric dictionary, if each $g \in \mathcal{D}$ has norm bounded by one ($\|g\| \leq 1$),

$$g \in \mathcal{D} \quad \text{implies} \quad -g \in \mathcal{D},$$

and the closure of $\text{span } \mathcal{D}$ is X . We denote the closure (in X) of the convex hull of \mathcal{D} by $A_1(\mathcal{D})$. In this section we discuss greedy algorithms with regard to a system \mathcal{D} that is not a dictionary. Here, we will discuss a variant of the Weak Relaxed Greedy Algorithm (WRGA). Let X be a real Banach space and let $\mathcal{D} := \{g\}$ be a system of elements $g \in X$ such that $\|g\| \leq 1$ and $g \in \mathcal{D}$ implies $-g \in \mathcal{D}$. Usually, in the theory of greedy algorithms we consider approximation with regard to a dictionary \mathcal{D} . One of the properties of a dictionary \mathcal{D} is that the closure of $\text{span } \mathcal{D}$ is equal to X . In this section we do not assume that the system \mathcal{D} is a dictionary. In particular, we do not assume that the closure of $\text{span } \mathcal{D}$ is X . This setting is motivated by applications in Learning Theory (see Chapter 4 of [7]).

For a nonzero element $f \in X$ we let F_f denote a norming (peak) functional for f :

$$\|F_f\| = 1, \quad F_f(f) = \|f\|.$$

The existence of such a functional is guaranteed by Hahn-Banach theorem.

Let $\tau := \{t_k\}_{k=1}^\infty$ be a given weakness sequence of numbers $t_k \in [0, 1]$, $k = 1, \dots$.

Weak Relaxed Greedy Algorithm (WRGA). We define $f_0^r := f_0^{r,\tau} := f$ and $G_0^r := G_0^{r,\tau} := 0$. Then, for each $m \geq 1$ we have the following inductive definition.

(1) $\varphi_m^r := \varphi_m^{r,\tau} \in \mathcal{D}$ is any element satisfying

$$F_{f_{m-1}^r}(\varphi_m^r - G_{m-1}^r) \geq t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}^r}(g - G_{m-1}^r).$$

(2) Find $0 \leq \lambda_m \leq 1$ such that

$$\|f - ((1 - \lambda_m)G_{m-1}^r + \lambda_m\varphi_m^r)\| = \inf_{0 \leq \lambda \leq 1} \|f - ((1 - \lambda)G_{m-1}^r + \lambda\varphi_m^r)\|$$

and define

$$G_m^r := G_m^{r,\tau} := (1 - \lambda_m)G_{m-1}^r + \lambda_m\varphi_m^r.$$

(3) Let

$$f_m^r := f_m^{r,\tau} := f - G_m^r.$$

For a Banach space X we define the modulus of smoothness

$$\rho(u) := \sup_{\|x\|=\|y\|=1} \left(\frac{1}{2}(\|x + uy\| + \|x - uy\|) - 1 \right).$$

The uniformly smooth Banach space is the one with the property

$$\lim_{u \rightarrow 0} \rho(u)/u = 0.$$

The following theorem was proved in [5] (see also Theorem 6.17 on p. 348 in [7]) for \mathcal{D} being a dictionary.

Theorem 4.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Then, for a sequence $\tau := \{t_k\}_{k=1}^\infty$, $t_k \leq 1$, $k = 1, 2, \dots$, we have for any $f \in A_1(\mathcal{D})$ that*

$$\|f_m^{r,\tau}\| \leq C_1(q, \gamma) \left(1 + \sum_{k=1}^m t_k^p \right)^{-1/p}, \quad p := \frac{q}{q-1},$$

with a constant $C_1(q, \gamma)$ which may depend only on q and γ .

We prove here an analog of the above theorem when we do not assume that \mathcal{D} is a dictionary and only assume that $\mathcal{D} = \{g\}$ is a symmetric system with a property $\|g\| \leq 1$.

Theorem 4.2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Then, for a sequence $\tau := \{t_k\}_{k=1}^\infty$, $t_k \leq 1$, $k = 1, 2, \dots$, we have for any $f \in X$ that*

$$\|f_m^{r,\tau}\| \leq \inf_{\phi \in A_1(\mathcal{D})} \|f - \phi\| + C_2(q, \gamma) \left(1 + \sum_{k=1}^m t_k^p\right)^{-1/p}, \quad p := \frac{q}{q-1},$$

with a constant $C_2(q, \gamma)$ which may depend only on q and γ .

Remark 4.1. *In case of a Hilbert space H there are stronger results for similar greedy algorithms with $\tau = \{1\}$ (see [7], p. 99, Theorem 2.28):*

$$\|f_m\|_H^2 \leq \left(\inf_{\phi \in A_1(\mathcal{D})} \|f - \phi\|_H\right)^2 + Cm^{-1}.$$

Proof. Proof of Theorem 4.2 is similar to the proof of Theorem 4.1. Denote

$$b := \inf_{\phi \in A_1(\mathcal{D})} \|f - \phi\|.$$

We use the following lemma.

Lemma 4.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Then, for a given $f \in A_1(\mathcal{D})$ we have*

$$\begin{aligned} \|f_m^{r,\tau}\| &\leq \inf_{0 \leq \lambda \leq 1} (\|f_{m-1}^{r,\tau}\| - \lambda t_m (\|f_{m-1}^{r,\tau}\| - b)) \\ &\quad + 2\|f_{m-1}^{r,\tau}\| \rho\left(\frac{2\lambda}{\|f_{m-1}^{r,\tau}\|}\right), \quad m = 1, 2, \dots \end{aligned}$$

Proof. We have

$$f_m^r := f - ((1 - \lambda_m)G_{m-1}^r + \lambda_m \varphi_m^r) = f_{m-1}^r - \lambda_m(\varphi_m^r - G_{m-1}^r)$$

and

$$\|f_m^r\| = \inf_{0 \leq \lambda \leq 1} \|f_{m-1}^r - \lambda(\varphi_m^r - G_{m-1}^r)\|.$$

We have from the definition of the modulus of smoothness for any λ

$$\begin{aligned} & \|f_{m-1}^r - \lambda(\varphi_m^r - G_{m-1}^r)\| + \|f_{m-1}^r + \lambda(\varphi_m^r - G_{m-1}^r)\| \leq \\ & 2\|f_{m-1}^r\| \left(1 + \rho\left(\frac{\lambda\|\varphi_m^r - G_{m-1}^r\|}{\|f_{m-1}^r\|}\right)\right). \end{aligned} \quad (4.1)$$

Next we get for $\lambda \geq 0$

$$\begin{aligned} & \|f_{m-1}^r + \lambda(\varphi_m^r - G_{m-1}^r)\| \geq F_{f_{m-1}^r}(f_{m-1}^r + \lambda(\varphi_m^r - G_{m-1}^r)) = \\ & \|f_{m-1}^r\| + \lambda F_{f_{m-1}^r}(\varphi_m^r - G_{m-1}^r) \geq \|f_{m-1}^r\| + \lambda t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}^r}(g - G_{m-1}^r). \end{aligned}$$

Using Lemma 6.10, p. 343, from [7] we continue

$$= \|f_{m-1}^r\| + \lambda t_m \sup_{\phi \in A_1(\mathcal{D})} F_{f_{m-1}^r}(\phi - G_{m-1}^r) \geq \|f_{m-1}^r\| + \lambda t_m (\|f_{m-1}^r\| - b).$$

Using the trivial estimate $\|\varphi_m^r - G_{m-1}^r\| \leq 2$ we obtain from (4.1)

$$\begin{aligned} & \|f_{m-1}^r - \lambda(\varphi_m^r - G_{m-1}^r)\| \\ & \leq \|f_{m-1}^r\| - \lambda t_m (\|f_{m-1}^r\| - b) + 2\|f_{m-1}^r\| \rho\left(\frac{2\lambda}{\|f_{m-1}^r\|}\right), \end{aligned} \quad (4.2)$$

which proves Lemma 4.1. \square

Set

$$a_m := \|f_m^r\| - b.$$

Note that

$$0 \leq a_m \leq 2.$$

Using monotonicity of $\rho(u)/u$ we derive from Lemma 4.1

$$a_m \leq a_{m-1} \inf_{\lambda \in [0,1]} (1 - \lambda t_m + 2\rho(2\lambda/a_{m-1})). \quad (4.3)$$

For $\rho(u) \leq \gamma u^q$ it gives

$$a_m \leq a_{m-1} \inf_{\lambda \in [0,1]} (1 - \lambda t_m + 2\gamma(2\lambda/a_{m-1})^q). \quad (4.4)$$

Denote λ_1 the solution of the equation

$$\frac{1}{2}\lambda t_m = 2\gamma \left(\frac{2\lambda}{a_{m-1}}\right)^q, \quad \lambda_1 = \left(\frac{t_m a_{m-1}^q}{2^{q+2\gamma}}\right)^{\frac{1}{q-1}}.$$

If $\lambda_1 \leq 1$ then

$$\begin{aligned} \inf_{\lambda \in [0,1]} (1 - \lambda t_m + 2\gamma(2\lambda/a_{m-1})^q) &\leq 1 - \lambda_1 t_m + 2\gamma(2\lambda_1/a_{m-1})^q \\ &= 1 - \frac{1}{2}\lambda_1 t_m = 1 - C_3(q, \gamma)t_m^p a_{m-1}^p, \quad p := \frac{q}{q-1}. \end{aligned}$$

If $\lambda_1 > 1$ then for all $\lambda \leq \lambda_1$ we have

$$\frac{1}{2}\lambda t_m \geq 2\gamma \left(\frac{2\lambda}{a_{m-1}} \right)^q.$$

Specifying $\lambda = 1$ we get

$$\inf_{\lambda \in [0,1]} (1 - \lambda t_m + 2\gamma(2\lambda/a_{m-1})^q) \leq 1 - \frac{1}{2}t_m \leq 1 - C_4(q, \gamma)t_m^p a_{m-1}^p.$$

Setting $C_5 := C_5(q, \gamma) := \min(C_3(q, \gamma), C_4(q, \gamma))$ we obtain

$$a_m \leq a_{m-1}(1 - C_5 t_m^p a_{m-1}^p). \quad (4.5)$$

It is known (see [7], p. 345) that inequalities (4.5) imply

$$a_m \leq C_6(q, \gamma) \left(1 + \sum_{n=1}^m t_n^p \right)^{1/p}.$$

This completes the proof of Theorem 4.2. \square

It is known (see, for instance, [2], Lemma B.1) that in the case $X = L_p$ we have

$$\rho(u) \leq u^p/p \quad \text{if } 1 \leq p \leq 2 \quad \text{and} \quad \rho(u) \leq (p-1)u^2/2 \quad \text{if } 2 \leq p < \infty.$$

Therefore, in this case Theorem 4.2 gives: for any $f \in L_p$

$$\|f_m^{r,\tau}\|_{L_p} \leq \inf_{\phi \in A_1(\mathcal{D})} \|f - \phi\|_{L_p} + C(p) \left(1 + \sum_{k=1}^m t_k^s \right)^{-1/s}, \quad (4.6)$$

where $s := \max(\frac{p}{p-1}, 2)$. It was proved in [3] that for $0 < v \leq 1$,

$$\sigma_m(f, \mathcal{D})_{L_p} \leq \inf_{\phi \in A_v(\mathcal{D})} \|f - \phi\|_{L_p} + C(p)m^{\max(1/p, 1/2) - 1/v}. \quad (4.7)$$

The proof in [3] is probabilistic and does not provide a deterministic algorithm for constructing a good m -term approximation. We note that inequality (4.6) shows that in case $v = 1$ the greedy algorithm WRGA with $\tau = \{t\}$ provides the rate of approximation as in (4.7).

References

- [1] B. Carl, Entropy numbers, s -numbers, and eigenvalue problem, *J. Func. Analysis*, **41** (1981), 290–306.
- [2] M. Donahue, L. Gurvits, C. Darken and E. Sontag, Rate of convex approximation in non-Hilbert spaces, *Constructive Approx.*, **13** (1997), 187–220.
- [3] F. Gao, C-K. Ing, Y. Yang, Metric entropy and sparse linear approximation of ℓ_q -hulls for $0 < q \leq 1$, *J. Approx. Theory*, **166** (2013), 42–55.
- [4] V.N. Temlyakov, Nonlinear Kolmogorov’s widths, *Matem. Zametki*, **63** (1998), 891–902.
- [5] V.N. Temlyakov, Greedy algorithms in Banach spaces, *Adv. Comput. Math.*, **14** (2001), 277–292.
- [6] V.N. Temlyakov, Nonlinear methods of approximation, *Found. Comput. Math.*, **3** (2003), 33–107.
- [7] V.N. Temlyakov, *Greedy approximation*, Cambridge University Press, 2011.