INDUSTRIAL
MATHEMATICS
INSTITUTE

2003:26

Simultaneous greedy approximation in Banach spaces

D. Leviatan and V.N. Temlyakov

IMI
Preprint Series

Department of Mathematics
University of South Carolina

SIMULTANEOUS GREEDY APPROXIMATION IN BANACH SPACES

D. Leviatan and V. N. Temlyakov[1]

Abstract. We study nonlinear m-term approximation with regard to a redundant dictionary $\mathcal{D}$ in a Banach space. It is known that in the case of Hilbert space H the Pure Greedy Algorithm (or, more generally, the Weak Greedy Algorithm) provides for each $f \in H$ and any dictionary $\mathcal{D}$ an expansion into a series

$$f = \sum_{j=1}^{\infty} c_j(f)\varphi_j(f), \quad \varphi_j(f) \in \mathcal{D}, \quad j = 1, 2, \dots$$

with the Parseval property: $\|f\|^2 = \sum_j |c_j(f)|^2$. The Orthogonal Greedy Algorithm (or, more generally, the Weak Orthogonal Greedy Algorithm) has been introduced in order to enhance the rate of convergence of greedy algorithms. Recently, we have studied analogues of the PGA and WGA for a given finite number of functions $f^1, \dots, f^N$ with a requirement that the dictionary elements φ_j of these expansions are the same for all f^i, $i = 1, \dots, N$. We have studied convergence and rate of convergence of such expansions which we call *simultaneous* expansions. The goal of this paper is twofold. First, we work in a Hilbert space and enhance the convergence of the simultaneous greedy algorithms by introducing an analogue of the orthogonalization process, and we give estimates on the rate of convergence. Then, we study simultaneous greedy approximation in a more general setting, namely, in uniformly smooth Banach spaces.

1. Introduction and historical survey

In this paper we continue the investigation of simultaneous greedy approximation. Greedy-type approximation is a vast area of research. We refer the reader to the following two surveys [D], [T3] that contain discussions of greedy approximation with regard to a dictionary. A new ingredient in the present paper, is a move from approximating a single element f to the simultaneous approximation of a set of elements $f^1, \dots, f^N$. This step has already been taken in the earlier papers [LuT], [LeT], and [T4], where the approximation in a Hilbert space has been considered. The goal of this paper is twofold. First, we work in a Hilbert space and enhance the convergence of the simultaneous greedy algorithms by introducing an analogue of the orthogonalization process, and we give estimates on the rate of convergence. Secondly, we study simultaneous greedy approximation in a more general setting, namely, in uniformly smooth Banach spaces.

[1]This research was supported by the National Science Foundation Grant DMS 0200187.

Two different approaches to the problem of simultaneous approximation have been developed in the papers [LuT], [LeT], [T4] in the case of Hilbert space. Here, we give generalizations of both approaches in the case of Banach spaces. We begin with a brief discussion of the two existing approaches.

First, recall some notations and definitions from the theory of approximation with regard to redundant systems. Let H be a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x \rangle^{1/2}$. We say a set $\mathcal{D}$ of functions (elements) from H is a dictionary if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$) and $\overline{\operatorname{span}} \mathcal{D} = H$. For a given dictionary $\mathcal{D}$ we can introduce a norm associated with $\mathcal{D}$ as

$$\|f\|_{\mathcal{D}} := \sup_{g \in \mathcal{D}} |\langle f, g \rangle|.$$

The Weak Greedy Algorithm (see [T1]) is defined as follows. Let the sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \le t_k \le 1$, be given.

Weak Greedy Algorithm (WGA). *Let $f_0^\tau := f$. Then for each $m \ge 1$, we inductively define:*

1. *Let $\varphi_m^\tau \in \mathcal{D}$ be any element satisfying*

$$|\langle f_{m-1}^\tau, \varphi_m^\tau \rangle| \ge t_m \|f_{m-1}^\tau\|_{\mathcal{D}};$$

2.

$$f_m^\tau := f_{m-1}^\tau - \langle f_{m-1}^\tau, \varphi_m^\tau \rangle \varphi_m^\tau;$$

3.

$$G_m^\tau(f, \mathcal{D}) := \sum_{j=1}^{m} \langle f_{j-1}^\tau, \varphi_j^\tau \rangle \varphi_j^\tau.$$

We note that in a particular case $t_k = t$, $k = 1, 2, \dots$, this algorithm was considered in [J], and the special case where $t_k = 1$, $k = 1, 2, \dots$, is the Pure Greedy Algorithm (PGA). Thus, the WGA is a generalization of the PGA in the direction of making it easier to construct an element φ_m^τ at the m-th greedy step. The term *weak* in the definition means that in step 1., we do not shoot for the optimal element of the dictionary which realizes the $\mathcal{D}$-norm, rather we are satisfied with a weaker property than being optimal. The obvious reason for this is that, in general, we do not know that such an optimal element exists. Another practical reason is that the weaker the assumption the easier it is to satisfy, and therefore easier to realize in practice. Note that the WGA includes, in addition to the first (greedy) step, a second step (see 2. and 3. in the above definition) where we update the approximant by adding to it, the orthogonal projection of the residual f_{m-1}^τ onto φ_m^τ. It will become apparent that for applications in simultaneous greedy approximation, it is important to have a theory of weak greedy approximation with arbitrary weakness sequence τ. However, we remark that in the case of the WGA we do not have a complete theory on the rate of convergence.

In order to formulate what is known on the rate of convergence, we define the class of functions

$$\mathcal{A}_1^o(\mathcal{D}, M) := \{f \in H : f = \sum_{k\in\Lambda} c_k w_k, \quad w_k \in \mathcal{D},\ \#\Lambda < \infty \text{ and } \quad \sum_{k\in\Lambda} |c_k| \le M\}$$

and $\mathcal{A}_1(\mathcal{D}, M)$ as the closure (in H) of $\mathcal{A}_1^o(\mathcal{D}, M)$. We will also use a brief notation $A_1(\mathcal{D}) := \mathcal{A}_1(\mathcal{D}, 1)$. The following result has been obtained in [T1] for nonincreasing weakness sequences.

Theorem 1.1 ([T1]). *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^\infty$ is a nonincreasing sequence. Then for $f \in A_1(\mathcal{D})$ we have*

$$\|f - G_m^\tau(f, \mathcal{D})\| \le (1 + \sum_{k=1}^m t_k^2)^{-t_m/2(2+t_m)}. \tag{1.1}$$

For a slightly modified WGA the Weak Orthogonal Greedy Algorithm (WOGA) we have a much better developed general theory. The WOGA was introduced by the second author (see [T1], and see [DT] for the Orthogonal Greedy Algorithm), in order to enhance the rate of convergence of the algorithm. It is defined as follows.

Weak Orthogonal Greedy Algorithm (WOGA). *Let $f_0^{o,\tau} := f$. Then for each $m \ge 1$, we inductively define:*

1. *Let $\varphi_m^{o,\tau} \in \mathcal{D}$ be any element satisfying*

$$|\langle f_{m-1}^{o,\tau}, \varphi_m^{o,\tau}\rangle| \ge t_m \|f_{m-1}^{o,\tau}\|_{\mathcal{D}};$$

2.
$$H_m^\tau(f) := \operatorname{span}\{\varphi_1^{o,\tau}, \dots, \varphi_m^{o,\tau}\};$$

3.
$$G_m^{o,\tau}(f, \mathcal{D}) := P_{H_m^\tau(f)}(f_{m-1}^{o,\tau}),$$

where $P_{H_m^\tau(f)}(g)$ denotes the orthogonal projection of $g \in H$ onto $H_m^\tau(f)$;
4.
$$f_m^{o,\tau} := f_{m-1}^{o,\tau} - G_m^{o,\tau}(f, \mathcal{D}).$$

It was proved in [T1] that

$$\sum_{k=1}^\infty t_k^2 = \infty,$$

is sufficient in order that

$$\lim_{m\to\infty} \|f - G_m^{o,\tau}(f, \mathcal{D})\| = 0.$$

It also has been shown in [T1] that

Theorem 1.2 ([T1]). *For every $f \in A_1(\mathcal{D})$ we have*

$$\|f - G_m^{o,\tau}(f,\mathcal{D})\| \le \left(1 + \sum_{k=1}^m t_k^2\right)^{-1/2}.$$

The above result suggests the following straightforward coordinatewise strategy for simultaneous approximation. This strategy has been used in [LuT].

Vector Weak Orthogonal Greedy Algorithm (VWOGA). *Let a vector of elements $f^i \in H$, $i = 1,\dots,N$, be given. We define $f_0^{i,v,\tau,o} := f^i$, $i = 1,\dots,N$. Then for each $m \ge 1$ we inductively define:*

1. *i_m is such that*
$$\|f_{m-1}^{i_m,v,\tau,o}\| \ge \|f_{m-1}^{i,v,\tau,o}\|, \qquad i = 1,\dots,N,$$
2. *$\varphi_m^{v,o,\tau} \in \mathcal{D}$ is any element satisfying*
$$|\langle f_{m-1}^{i_m,v,\tau,o}, \varphi_m^{v,\tau,o}\rangle| \ge t_m \|f_{m-1}^{i_m,v,\tau,o}\|_{\mathcal{D}},$$
3. $$G_m^{v,o,\tau}(f^i,\mathcal{D}) := P_{H_m^{v,\tau}}(f^i), \quad \textit{where} \quad H_m^{v,\tau} := \operatorname{span}\{\varphi_1^{v,\tau,o},\dots,\varphi_m^{v,\tau,o}\},$$
4. $$f_m^{i,v,\tau,o} := f^i - G_m^{v,\tau,o}(f^i,\mathcal{D}).$$

It is clear that for each coordinate element f^i a realization of the VWOGA is the WOGA with $\tau^i = \{t_k^i\}_{k=1}^\infty$ such that $t_k^i = t_k$ if $i_k = i$ and $t_k^i = 0$ otherwise. It was shown in [LuT] that in this case Theorem 1.2 implies the following estimate.

Theorem 1.3 ([LuT]). *Let $\mathcal{D}$ be an arbitrary dictionary in H and $\tau = \{t\}$, $0 < t \le 1$. If $f^i \in A_1(\mathcal{D})$, for all $1 \le i \le N$, then we have*

$$\|f^i - G_m^{v,o,\tau}(f^i,\mathcal{D})\| \le \min(1, \left(\frac{N}{mt^2}\right)^{1/2}), \qquad i = 1,\dots,N.$$

It is clear that the restriction that τ is a nonincreasing weakness sequence in Theorem 1.1 prevents the use of coordinatewise strategy in the case of the WGA. In order to overcome this difficulty the following two methods have been designed in [LuT] and [LeT].

Vector Weak Greedy Algorithm (VWGA). *Let a vector of elements $f^i \in H$, $i = 1,\dots,N$, be given. We write $f_0^{i,v,\tau} := f^i$, $i = 1,\dots,N$. Then for each $m \ge 1$, we inductively define:*

1. *Let $\varphi_m^{v,\tau} \in \mathcal{D}$ be any element satisfying*

$$\max_i |\langle f_{m-1}^{i,v,\tau}, \varphi_m^{v,\tau}\rangle| \ge t_m \max_i \|f_{m-1}^{i,v,\tau}\|_{\mathcal{D}},$$

2.
$$f_m^{i,v,\tau} := f_{m-1}^{i,v,\tau} - \langle f_{m-1}^{i,v,\tau}, \varphi_m^{v,\tau}\rangle \varphi_m^{v,\tau}, \quad i = 1, \dots, N,$$

3.
$$G_m^{v,\tau}(f^i, \mathcal{D}) := \sum_{j=1}^{m} \langle f_{j-1}^{i,v,\tau}, \varphi_j^{v,\tau}\rangle \varphi_j^{v,\tau}, \quad i = 1, \dots, N.$$

The following estimate of the rate of convergence of VWGA has been obtained in [LuT].

Theorem 1.4 ([LuT]). *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k = t$, $k = 1, \dots,$ $0 < t < 1$. Then for any vector of elements $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $1 \le i \le N$, we have*

$$\sum_{i=1}^{N} \|f_m^{i,v,\tau}\|^2 \le \left(1 + \frac{mt^2}{N}\right)^{-t/(2N+t)} N^{\frac{2N+2t}{2N+t}}.$$

Comparing Theorem 1.1 with $\tau = \{t\}$ with Theorem 1.4 we see that the exponent $\frac{t}{2N+t}$ of decay is seriously affected by the number N of simultaneously approximated elements. Also, simultaneous approximation brings an extra factor $N^{\frac{2N+2t}{2N+t}} \asymp N$. In [LeT] we improve the exponent of decay replacing $\frac{t}{2N+t}$ by $\frac{t}{2N^{1/2}+t}$ but we pay with a bigger constant N^2 instead of N. Here is the corresponding theorem from [LeT].

Theorem 1.5 ([LeT]). *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^{\infty}$ is a nonincreasing sequence. Then for any vector of elements $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $1 \le i \le N$, we have*

$$\sum_{i=1}^{N} \|f_m^{i,v,\tau}\|^2 \le N^2 \Big(1 + \frac{1}{N} \sum_{k=1}^{m} t_k^2\Big)^{\frac{-t_m}{2N^{1/2}+t_m}}. \tag{1.2}$$

Recently an estimate that improves the estimates in both Theorems 1.4 and 1.5, has been obtained in [T4]. This estimate combines the good features of the estimates of Theorems 1.4 and 1.5. It has the exponent from Theorem 1.5, and the constant N as in Theorem 1.4.

Theorem 1.6 ([T4]). *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k = t \in (0, 1]$, $k = 1, 2, \dots$. Then for any vector of elements $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $1 \le i \le N$, we have*

$$\sum_{i=1}^{N} \|f_m^{i,s,\tau}\|^2 \le N \left(1 + \frac{mt^2}{N}\right)^{\frac{-t}{2N^{1/2}+t}}.$$

Theorem 1.7 ([T4]). *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^{\infty}$ is a nonincreasing sequence. Then for any vector of elements $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $1 \le i \le N$, we have*

$$\sum_{i=1}^{N} \|f_m^{i,s,\tau}\|^2 \le CN(N + \sum_{k=1}^{m} t_k^2)^{\frac{-t_m}{2N^{1/2}+t_m}}$$

with an absolute constant $C = e^{2/e} < 3$.

We conclude this section with some comments on proofs of Theorems 1.4–1.7. The proof of Theorem 1.4 of [LuT], is an adaptation of the proof of Theorem 1.1 of [T1] to the vector case. This proof is independent of Theorem 1.1. The proof of Theorem 1.5 from [LeT] directly uses Theorem 1.1. In [LeT] we interpret a simultaneous approximation of $f^1, \dots, f^N$ in H with respect to $\mathcal{D}$, as an approximation of $F = (f^1, \dots, f^N)$ in the Hilbert space $H_N := \underbrace{H \times \cdots \times H}_{N \text{ times}}$, with respect to a special dictionary $\mathcal{D}_N \subset H_N$ built from $\mathcal{D}$. The proof of Theorems 1.6 and 1.7 from [T4] is more like that of Theorem 1.4. It is a modification of the proof of Theorem 1.1. Thus, we have two methods of analyzing the efficiency of simultaneous approximation. In the first ([LuT], [T4]), we stay within the space H with a dictionary $\mathcal{D}$, and analyze the rate of convergence for each coordinate f^i. In the second ([LeT]), we consider a new Hilbert space H_N with a new dictionary $\mathcal{D}_N$. In the latter case we approximate the vector $F = (f^1, \dots, f^N) \in H_N$. The above mentioned results show that the two methods of analysis provide the same rate of convergence with the former giving a better constant as a function on N. We have decided to present in this paper the generalization of both methods to the case of Banach spaces. The new results are formulated and proved in the coming sections.

2. Simultaneous approximation in Banach spaces

In this section we will present some results on simultaneous approximation in Banach spaces. Results on simultaneous approximation will be obtained from the corresponding results on approximation of a single element, that is, we follow the line of [LuT], [T4]. We note that there are two natural generalizations of the Pure Greedy Algorithm to the case of Banach space X: the X-Greedy Algorithm and the Dual Greedy Algorithm (see [T3, Section 1]). However, there are no general results on convergence and rate of convergence of the above two algorithms, therefore we will not discuss these two algorithms here. Instead, we will discuss two modifications of the Weak Greedy Algorithm the Weak Orthogonal Greedy Algorithm and the Weak Relaxed Greedy Algorithm that have been successfully generalized to the case of Banach spaces. It will be convenient for us to work in this section with symmetrized dictionaries.

Let X be a Banach space with norm $\|\cdot\|$. We say that a set of elements (functions) $\mathcal{D}$ from X is a dictionary if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$),

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

and $\overline{\operatorname{span}}\mathcal{D} = X$. Finally, we will use the same notation $A_1(\mathcal{D}) := \mathcal{A}_1(\mathcal{D}, 1)$, from the introduction, this time for the Banach space X.

We begin with the definitions of two types of greedy algorithms with regard to $\mathcal{D}$. For an element $f \in X$ we denote by F_f a norming (peak) functional for f:

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

The existence of such a functional is guaranteed by Hahn-Banach theorem. Let $\tau := \{t_k\}_{k=1}^{\infty}$ be a given sequence of nonnegative numbers $t_k \leq 1$, $k = 1, \dots$. We first define the Weak Chebyshev Greedy Algorithm (WCGA) which is a natural generalization of the Weak Orthogonal Greedy Algorithm, to Banach spaces (see [T2]).

Weak Chebyshev Greedy Algorithm (WCGA). *Denote $f_0^c := f_0^{c,\tau} := f$. Then for each $m \geq 1$, we inductively define*

1. *Let $\varphi_m^c := \varphi_m^{c,\tau} \in \mathcal{D}$ be any element satisfying*

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

2. *Set*

$$\Phi_m := \Phi_m^\tau := \operatorname{span}\{\varphi_j^c\}_{j=1}^m,$$

and define $G_m^c := G_m^{c,\tau}$ to be the best approximant to f from Φ_m.

3. *Denote*

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

We also define the generalization to Banach spaces (see [T2]) of the Weak Relaxed Greedy Algorithm that was studied in [T1] in the case of a Hilbert space. We refer the reader to [B], [DGDS], [J1] for related algorithms.

Weak Relaxed Greedy Algorithm (WRGA). *Let $f_0^r := f_0^{r,\tau} := f$ and $G_0^r := G_0^{r,\tau} := 0$. Then for each $m \geq 1$, we inductively define*

1. *Let $\varphi_m^r := \varphi_m^{r,\tau} \in \mathcal{D}$ be 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. *Denote*

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

Remark 2.1. It follows from the definitions of WCGA and WRGA that the sequences $\{\|f^c_m\|\}$ and $\{\|f^r_m\|\}$ are nonincreasing.

We repeat that the term *weak* in both definitions means that in step 1., we do not shoot for the optimal element of the dictionary which realizes the corresponding sup, rather we are satisfied with a weaker property than being optimal. Again, the obvious reason for this is that, in general, we do not know that such an optimal element exists, and for the practical reason that the weaker the assumption the easier it is to satisfy, and therefore easier to realize in practice. Applications of the WCGA and of the WRGA in simultaneous approximation provide further justification for studying the *weak* version instead of the *pure* version, namely, $\tau = \{1\}$, of greedy algorithms.

It is clear that in the case of WRGA, it is natural to assume that f belongs to the closure of convex hull of $\mathcal{D}$ (in our notation $A_1(\mathcal{D})$). It has been proved in [T1] that in the case of a Hilbert space the WRGA yields, for the class $A_1(\mathcal{D})$, an approximation error of the order

$$(1+\sum_{k=1}^{m} t_k^2)^{-1/2},$$

that is, just like the WOGA.

Following [T2] we consider here approximation in uniformly smooth Banach spaces. For a Banach space X we define the modulus of smoothness

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

The Banach space is called uniformly smooth if

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

It is easy to see that the modulus of smoothness $\rho(u)$ is an even convex function satisfying the inequalities

$$\max(0,u-1) \le \rho(u) \le u, \quad u \in (0,\infty).$$

It has been established in [DGDS] that the approximation error of an algorithm analogous to our WRGA with $t_k = 1$, $k = 1,2,\dots$, for the class $A_1(\mathcal{D})$ can be expressed in terms of the modulus of smoothness, namely, if $\rho(u) \le \gamma u^q$, $1 < q \le 2$, then the error is of $O(m^{1/q-1})$. The following rate of convergence of the WCGA and the WRGA has been established in [T2].

Theorem 2.1 ([T2]). *Let X be a uniformly smooth Banach space with a modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$, and let $\tau := \{t_k\}_{k=1}^\infty$, $0 \le t_k \le 1$, $k = 1,2,\dots$, be given. Then for any $f \in A_1(\mathcal{D})$ we have*

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

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

where the constant $C(q,\gamma)$ *may depend only on* q *and* γ*.*

We first follow [LuT], and study two *vector* versions of the WCGA and the WRGA.

Vector Weak Chebyshev Greedy Algorithm (VWCGA). *Given* $F := (f^1,\dots,f^N)$*, we let* $f_0^{i,c} := f_0^{i,v,c,\tau} := f^i$*,* $i=1,\dots,N$*. Then for each* $m \ge 1$*, we inductively define*

1. *Let* i_m *be such that*

$$\|f_{m-1}^{i_m,v,c,\tau}\| = \max_{1\le i\le N}\|f_{m-1}^{i,v,c,\tau}\|,$$

2. *Let* $\varphi_m^c := \varphi_m^{v,c,\tau} \in \mathcal{D}$ *be any element satisfying*

$$F_{f_{m-1}^{i_m,c}}(\varphi_m^c) \ge t_m \sup_{g\in\mathcal{D}} F_{f_{m-1}^{i_m,c}}(g).$$

3. *Define*

$$\Phi_m := \Phi_m^\tau := \mathrm{span}\{\varphi_j^c\}_{j=1}^m,$$

and define $G_m^{i,c} := G_m^{i,c,\tau}$ *to be the best approximant to* f^i *from* Φ_m*,* $i=1,\dots,N$*.*

4. *Denote*

$$f_m^{i,c} := f_m^{i,v,\tau,c} := f - G_m^{i,c}.$$

Vector Weak Relaxed Greedy Algorithm (VWRGA). *Given* $F := (f^1,\dots,f^N)$*, we let* $f_0^{i,r} := f_0^{i,v,r,\tau} := f^i$*, and* $G_0^{i,r} := G_0^{i,v,\tau,r} := 0$*,* $i=1,\dots,N$*. Then for each* $m\ge 1$*, we inductively define*

1. *Let* i_m *be such that*

$$\|f_{m-1}^{i_m,v,r,\tau}\| = \max_{1\le i\le N}\|f_{m-1}^{i,v,r,\tau}\|,$$

2. *Let* $\varphi_m^r := \varphi_m^{v,r,\tau} \in \mathcal{D}$ *be any element satisfying*

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

3. *Find* $0 \le \lambda_m^i \le 1$ *such that*

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

and define

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

4. *Denote*

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

We prove here the following rate of convergence of the VWCGA and the VWRGA.

Theorem 2.2. *Let X be a uniformly smooth Banach space with a modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Then for a sequence $\tau := \{t_k\}_{k=1}^{\infty}$, $0 \le t_k \le 1$, $k = 1, 2, \dots,$ we have for any $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$, that*

$$\|f_m^{i,v,\tau,b}\| \le C(q,\gamma) \min\left\{1, \left(\frac{1}{N}\sum_{k=1}^{m} t_k^p\right)^{-1/p}\right\}, \quad p := \frac{q}{q-1}, \tag{2.1}$$

with a constant $C(q,\gamma)$ which may depend only on q and γ, and where b stands for either c or r.

Note that in the special case where X is a Hilbert space, and the special weakness sequence τ such that $t_k = t \in (0,1]$, $k = 1, 2, \dots,$ Theorem 2.2 for $b = c$ is Theorem 1.3 which has been proved in [LuT] with $C(q,\gamma) = 1$.

Proof. The inequality

$$\|f_m^{i,v,\tau,b}\| \le 1$$

readily follows by the assumption $f^i \in A_1(\mathcal{D})$ and the trivial observation that the sequences $\{\|f_m^{i,v,\tau,b}\|\}$, $1 \le i \le N$, are decreasing. Therefore we only have to prove the estimate

$$\|f_m^{i,v,\tau,b}\| \le C(q,\gamma)(\frac{1}{N}\sum_{k=1}^{m} t_k^p)^{-1/p}, \quad i = 1, \dots, N.$$

To this end, let m be given. For each $l \in [1, N]$ denote $E_l := \{j \mid i_j = l\} \subseteq [1, m]$ (see VWCGA 1., VWRGA 1., respectively). In other words,

$$\|f_{j-1}^{l,v,\tau,b}\| = \max_{1 \le i \le N} \|f_{j-1}^{i,v,\tau,b}\|, \quad j \in E_l.$$

Evidently,

$$\sum_{k=1}^{m} t_k^p = \sum_{i=1}^{N} \sum_{k \in E_i} t_k^p$$

whence there is an $l_0 \in [1, N]$ such that

$$\sum_{k \in E_{l_0}} t_k^p \ge \frac{1}{N}\sum_{k=1}^{m} t_k^p.$$

Let $n_0 := \max\{k \mid k \in E_{l_0}\}$, and put $E'_{l_0} := E_{l_0} \setminus \{n_0\}$. Then we have

$$\max_i \|f_m^{i,v,\tau,b}\| \le \max_i \|f_{n_0-1}^{i,v,\tau,b}\| \le \|f_{n_0-1}^{l_0,v,\tau,b}\|.$$

If we restrict our attention to l_0, we see that the VWCGA, respectively, the VWRGA, are the application of the WCGA, respectively, the WRGA, with the weakness sequence $\tau^{l_0} := \{t_j^{l_0}\}$, given by

$$t_j^{l_0} = \begin{cases} t_j, & \text{if } j \in E_{l_0} \\ 0, & \text{otherwise,} \end{cases}$$

to f^{l_0}. Therefore we conclude from Theorem 2.1, that

$$\begin{aligned} \|f_{n_0-1}^{l_0}\| &\le C(q,\gamma) \left(1 + \sum_{k \in E'_{l_0}} t_k^p\right)^{-1/p} \\ &\le C(q,\gamma) \left(1 + \sum_{k \in E_{l_0}} t_k^p - 1\right)^{-1/p} \\ &\le C(q,\gamma) \left(\frac{1}{N} \sum_{k=1}^{m} t_k^p\right)^{-1/p}. \end{aligned}$$

This completes the proof of Theorem 2.2.

Remark 2.2. It follows from the proof of Theorem 2.2 that the constant $C(q,\gamma)$ in Theorem 2.2 is the same as the corresponding constant in Theorem 2.1. It is known (see [T1]) that in the case of Hilbert space the corresponding constant in Theorem 2.1 is equal to 1 for the WOGA and is equal to 2 for the WRGA. Therefore in Theorem 2.2 we may take $C(q,\gamma) = 1$ in the case of the VWOGA and $C(q,\gamma) = 2$ in the case of the VWRGA in a Hilbert space.

3. Simultaneous orthogonal greedy algorithms

In this section we study simultaneous greedy algorithms in Hilbert and Banach spaces, along the lines of [LeT].
We begin with a Hilbert space H and define

Orthogonal Vector Weak Greedy Algorithm (OVWGA). *Given $F := (f^1, \dots, f^N)$, $f^i \in H$, $1 \le i \le N$, we let $f_0^{i,v,o,\tau} := f^i$, $1 \le i \le N$. Then for each $m \ge 1$, we inductively define:*

1. *Let $\varphi_m^{v,o,\tau} \in \mathcal{D}$ be any element satisfying*

$$\max_i |\langle f_{m-1}^{i,v,o,\tau}, \varphi_m^{v,o,\tau}\rangle| \ge t_m \max_i \|f_{m-1}^{i,v,o,\tau}\|_{\mathcal{D}}; \tag{3.1}$$

2.

$$H_m^{v,\tau}(F) := \operatorname{span}\{\varphi_1^{v,o,\tau}, \dots, \varphi_m^{v,o,\tau}\}; \tag{3.2}$$

3.

$$G_m^{i,v,o,\tau}(F,\mathcal{D}) := P_{H_m^\tau(F)}(f^i), \quad i = 1, \dots, m,$$

where $P_{H_m^{v,\tau}(F)}(g)$ *denotes the orthogonal projection of* $g \in H$ *onto* $H_m^{v,\tau}(F)$*;*
4.
$$f_m^{i,v,o,\tau} := f^i - G_m^{i,v,o,\tau}(F, \mathcal{D}),$$

As we have done in [LeT], we may modify Step 1 in the definition of the OVWGA to in the following two ways. In the first step of the Weak Simultaneous Orthogonal Greedy Algorithm 1 (WSOGA1) we take
1. $\varphi_m^{s1,o,\tau} \in \mathcal{D}$ to be any element satisfying

$$\left(\sum_{i=1}^N |\langle f_{m-1}^{i,s1,o,\tau}, \varphi_m^{s1,o,\tau}\rangle|^2\right)^{1/2} \geq t_m \max_i \|f_{m-1}^{i,s1,o,\tau}\|_{\mathcal{D}}, \tag{3.3}$$

and we define $H_m^{s1,\tau}(F)$ in an analogous way to (3.2).
Similarly, in the first step of the Weak Simultaneous Orthogonal Greedy Algorithm 2 (WSOGA2) we take
1. $\varphi_m^{s2,o,\tau} \in \mathcal{D}$ to be any element satisfying

$$\left(\sum_{i=1}^N |\langle f_{m-1}^{i,s2,o,\tau}, \varphi_m^{s2,o,\tau}\rangle|^2\right)^{1/2} \geq t_m \sup_{g\in\mathcal{D}} \left(\sum_{i=1}^N |\langle f_{m-1}^{i,s2,o,\tau}, g\rangle|^2\right)^{1/2}, \tag{3.4}$$

and again, we define $H_m^{s2,\tau}(F, \mathcal{D})$ in an analogous way to (3.2). Clearly, any φ_m satisfying either (3.1) or (3.4) also satisfies (3.3).

We prove

Theorem 3.1. *Let* $\mathcal{D}$ *be an arbitrary dictionary in* H*, and let* $\tau := \{t_k\}_{k=1}^\infty$ *be a weakness sequence. Then for any vector of elements* $f^1, \dots, f^N$*,* $f^i \in A_1(\mathcal{D})$*,* $1 \leq i \leq N$*, and for* s *standing for either* v *or* $s1$ *or* $s2$*, we have*

$$\sum_{i=1}^N \|f_m^{i,s,o,\tau}\|^2 \leq N^2 \left(1 + \frac{1}{N}\sum_{k=1}^m t_k^2\right)^{-1}. \tag{3.5}$$

Note that (3.5) provides an estimate on the rate of convergence which is significantly better than (1.2), and without the assumption on the monotonicity of the weakness sequence τ.

In particular for the weakness sequence where $t_k = t$, $k = 1, 2 \dots$, we obtain as an immediate consequence, the same order of the rate of convergence as [LuT, Theorem 10] (see Theorem 1.3 of the present paper), namely,

Corollary 3.1. *Let* $\mathcal{D}$ *be an arbitrary dictionary in* H*, and let* $\tau := \{t_k\}_{k=1}^\infty$*, with* $t_k = t$*,* $k = 1, 2 \dots$ *. Then for any vector of elements* $f^1, \dots, f^N$*,* $f^i \in A_1(\mathcal{D})$*,* $i = 1, \dots, N$*, we have*

$$\left(\sum_{i=1}^N \|f_m^{i,v,o,\tau}\|^2\right)^{1/2} \leq N \left(1 + \frac{mt^2}{N}\right)^{-1/2}.$$

As has been alluded to in the introduction, given are a Hilbert space H and a dictionary $\mathcal{D}$. For $N \geq 2$, let $H_N := \underbrace{H \times \cdots \times H}_{N \text{ times}}$, i.e., the general element in H_N is $F := (f^1, \dots, f^N)$, $f^k \in H$. It is a Hilbert space with the inner product

$$\langle F_1, F_2 \rangle := \sum_{k=1}^{N} \langle f_1^k, f_2^k \rangle.$$

Let $\mathcal{D}_N$ be the collection

$$\{(\alpha_1 g_1, \dots, \alpha_N g_N) \mid g_k \in \mathcal{D}, \quad \sum_{k=1}^{N} \alpha_k^2 = 1\}.$$

Then it is easy to see that $\overline{\operatorname{span}} \mathcal{D}_N = H_N$. (Actually, H_N is spanned even by linear combinations of elements of the form $(0, \dots, 0, g, 0, \dots, 0)$, where $g \in \mathcal{D}$ is arbitrary and is in arbitrary position.) Also, all elements in $\mathcal{D}_N$ are normalized. Finally,

$$\|F\|_{\mathcal{D}_N} = \sup_{\substack{\alpha := (\alpha_1, \dots, \alpha_N) \\ \|\alpha\|_2 = 1 \\ g_1, \dots, g_N \in \mathcal{D}}} \Big|\sum_{i=1}^{N} \langle f^i, g_i \rangle \alpha_i\Big| = \Big(\sum_{i=1}^{N} \|f^i\|_{\mathcal{D}}^2\Big)^{1/2}. \tag{3.6}$$

Proof of Theorem 3.1. Given $F = (f^1, \dots, f^N)$, $f^i \in A_1(\mathcal{D})$, we see that $F \in \mathcal{A}_1(\mathcal{D}_N, N)$. We let s stand for either v or $s1$ or $s2$, and we set $f_0^{i,s,o,\tau} := f^i$, $i = 1, \dots, N$ and $F_0^{s,o,\tau} := F$. At stage $m \geq 1$ we select $\varphi_m^{s,o,\tau}$ satisfying (3.1), or (3.3), or (3.4), as the case may be, and we set

$$\Phi_m^{s,o,\tau} := (\beta_1 \varphi_m^{s,o,\tau}, \dots, \beta_N \varphi_m^{s,o,\tau}),$$

where

$$\beta_i := \langle f_{m-1}^{i,s,o,\tau}, \varphi_m^{s,o,\tau} \rangle \Big(\sum_{j=1}^{N} |\langle f_{m-1}^{j,s,o,\tau}, \varphi_m^{s,o,\tau} \rangle|^2\Big)^{-1/2}, \quad i = 1, \dots, N.$$

Then with $F_{m-1}^{s,o,\tau} := \big(f_{m-1}^{1,s,o,\tau}, \dots, f_{m-1}^{N,s,o,\tau}\big)$, it was proved in [LeT, (3.2) and the proof of Lemma 3.1] that

$$\langle F_{m-1}^{s,o,\tau}, \Phi_m^{s,o,\tau} \rangle \geq t_m N^{-1/2} \|F_{m-1}^{s,o,\tau}\|_{\mathcal{D}_N}. \tag{3.7}$$

Next we write

$$G_m^{s,o,\tau}(F, \mathcal{D}_N) := \big(G_m^{1,s,o,\tau}(F, \mathcal{D}), \dots, G_m^{N,s,o,\tau}(F, \mathcal{D})\big),$$

and we observe that

$$\begin{aligned} F_m^{s,o,\tau} :=& F_{m-1}^{s,o,\tau} - G_m^{s,o,\tau}(F, \mathcal{D}_N) \\ =& \big(f^1 - G_m^{1,s,o,\tau}(F, \mathcal{D}), \dots, f^N - G_m^{N,s,o,\tau}(F, \mathcal{D})\big), \end{aligned}$$

is perpendicular to $\underbrace{H_m^{s,\tau}(F,\mathcal{D}) \times \cdots \times H_m^{s,\tau}(F,\mathcal{D})}_{N \text{ times}}$. Hence, in particular, it is perpendicular to

$$H_m^{s,\tau}(F,\mathcal{D}_N) := \operatorname{span}\{\Phi_1^{s,o,\tau},\dots,\Phi_m^{s,o,\tau}\} \\ \subseteq H_m^{s,\tau}(F,\mathcal{D}) \times \cdots \times H_m^{s,\tau}(F,\mathcal{D}),$$

and we conclude that

$$\|F_m^{s,o,\tau}\| \le \|F_{m-1}^{s,o,\tau} - P_{H_m^{s,\tau}(F,\mathcal{D}_N)}(F_{m-1}^{s,o,\tau})\|.$$

Thus, by virtue of Theorem 1.2 we obtain by (3.7),

$$\|F - G_m^{s,o,\tau}(F,\mathcal{D}_N)\| = \|F_m^{s,o,\tau}\| \le N(1 + \sum_{k=1}^m t_k^2/N)^{-1/2},$$

and (3.5) is proven. $\square$

Remark 3.1. For the OVWGA case, one can give a simpler proof of Theorem 3.1, using a simpler $\Phi_m^{s,o,\tau}$, so that one does not have to rely on [LeT]. This proof generalizes to more general Banach spaces (see below). We give the separate proof for Hilbert spaces as it crystalizes the ideas.

Proof of Theorem 3.1 *for the case* $s = v$. As above, given $F = (f^1,\dots,f^N)$, $f^i \in A_1(\mathcal{D})$, we have that $F \in \mathcal{A}_1(\mathcal{D}_N, N)$. We set $f_0^{i,v,o,\tau} := f^i$, $i = 1,\dots,N$ and $F_0^{v,o,\tau} := F$. At stage $m \ge 1$ we select $\varphi_m^{v,o,\tau}$ satisfying (3.1), and for an appropriate $1 \le i_m \le N$ such that

$$|\langle f_{m-1}^{i_m,v,o,\tau}, \varphi_m^{v,o,\tau}\rangle| = \max_i |\langle f_{m-1}^{i,v,o,\tau}, \varphi_m^{v,o,\tau}\rangle|, \tag{3.8}$$

we set

$$\Phi_m^{v,o,\tau} := (\underbrace{0,\dots,0,\varphi_m^{v,o,\tau},0\dots,0}_{N}),$$

where the nonzero entry is at the i_mth place.

Then, with $F_{m-1}^{v,o,\tau} := \big(f_{m-1}^{1,v,o,\tau},\dots,f_{m-1}^{N,v,o,\tau}\big)$, it readily follows by (3.8) that

$$\begin{aligned} |\langle F_{m-1}^{v,o,\tau}, \Phi_m^{v,o,\tau}\rangle| &= |\langle f_{m-1}^{i_m,v,o,\tau}, \varphi_m^{v,o,\tau}\rangle| \\ &\ge t_m \max_i \|f_{m-1}^{i,v,o,\tau}\|_{\mathcal{D}} \\ &\ge t_m N^{-1/2} \|F_{m-1}^{v,o,\tau}\|_{\mathcal{D}_N}. \end{aligned} \tag{3.9}$$

As before, we write

$$G_m^{v,o,\tau}(F,\mathcal{D}_N) := \big(G_m^{1,v,o,\tau}(F,\mathcal{D}),\dots,G_m^{N,v,o,\tau}(F,\mathcal{D})\big),$$

and we observe that

$$F_m^{v,o,\tau} := F_{m-1}^{v,o,\tau} - G_m^{v,o,\tau}(F,\mathcal{D}_N),$$

is perpendicular to $\underbrace{H_m^{v,\tau}(F,\mathcal{D}) \times \cdots \times H_m^{v,\tau}(F,\mathcal{D})}_{N \text{ times}}$. Hence, in particular, it is perpendicular to

$$H_m^{v,\tau}(F,\mathcal{D}_N) := \operatorname{span}\{\Phi_1^{v,o,\tau},\ldots,\Phi_m^{v,o,\tau}\}.$$

Thus, again by virtue of Theorem 1.2 we obtain by (3.9),

$$\|F - G_m^{v,o,\tau}(F,\mathcal{D}_N)\| = \|F_m^{v,o,\tau}\| \le N(1+\sum_{k=1}^m t_k^2/N)^{-1/2},$$

and (3.5) is established. $\square$

We now approach the question of simultaneous approximation in a uniformly smooth Banach space X with the norm $\|\cdot\|_X$, and $\rho(u) \le \gamma u^q$, $1 < q \le 2$. We consider the N-tuple $V := (x^1,\ldots,x^N)$, $x^i \in X$, as an element of the space $\ell_2(X)$, namely, equipped with the norm

$$\|V\| := \left(\sum_{i=1}^N \|x^i\|_X^2\right)^{1/2}.$$

A functional $\mathcal{F}$, on $\ell_2(X)$, has the representation

$$\mathcal{F} := (F_1,\ldots,F_N),$$

where $F_i \in X^*$, and $\mathcal{F}(V) := \sum_{i=1}^N F_i(x^i)$. Evidently

$$\|\mathcal{F}\| = \left(\sum_{i=1}^N \|F_i\|_{X^*}^2\right)^{1/2},$$

and the norming functional $\mathcal{F}_V$, of $V \neq 0$, is given by

$$\mathcal{F}_V = (\alpha_1 F_{x^1},\ldots,\alpha_N F_{x^N}),$$

where F_{x^i} is the norming functional of x^i, and $\alpha_i = \|x^i\|_X/\|V\|$. Clearly, $\|\mathcal{F}_V\| = 1$. Also we put $\mathcal{F}_0 = 0$ since it may appear.

We define the vector analogue of the WCGA, denoted by CVWGA, as follows

Chebyshev Vector Weak Greedy Algorithm (CVWGA). *Given $F := (f^1,\ldots,f^N)$, we let $f_0^{v,i} := f_0^{v,i,\tau} := f^i$, $i = 1,\ldots,N$. Then for each $m \ge 1$, we inductively define*
1. *Let $\varphi_m^v := \varphi_m^{v,\tau} \in \mathcal{D}$ be any element such that*

$$\max_{1\le i\le N} |F_{f_{m-1}^i}(\|f_{m-1}^i\|\varphi_m^v)| \ge t_m \max_{1\le i\le N} \sup_{g\in\mathcal{D}} |F_{f_{m-1}^i}(\|f_{m-1}^i\| g)|, \tag{3.10}$$

where in the case $f^i_{m-1}=0$ for all $1\le i\le N$, the process stops.
2. *Let i_m be such that*

$$|F_{f^{i_m}_{m-1}}(\|f^{i_m}_{m-1}\|\varphi^v_m)|=\max_{1\le i\le N}|F_{f^i_{m-1}}(\|f^i_{m-1}\|\varphi^v_m)|,$$

and set

$$\Phi^v_m:=\Phi^{v,\tau}_m:=\underbrace{(0,\dots,0,\varphi^{v,\tau}_m,0\dots,0)}_{N},$$

where the nonzero entry is at the i_mth place.
3. *Let $G^v_m:=G^{v,\tau}_m$ be the best approximation to F from span$\{\Phi^v_1,\dots,\Phi^v_m\}$. Denote*

$$(f^{1,v,\tau}_m,\dots,f^{N,v,\tau}_m):=F-G^v_m.$$

Theorem 3.3. *Let X be a uniformly smooth Banach space with a modulus of smoothness $\rho(u)\le\gamma u^q$, $1<q\le 2$. Then for a sequence $\tau:=\{t_k\}_{k=1}^\infty$, $0\le t_k\le 1$, $k=1,2,\dots,$ we have for any $f^i\in A_1(\mathcal{D})$, $i=1,\dots,N$, that*

$$(3.11)\qquad \Big(\sum_{i=1}^N\|f^{i,v,\tau}_m\|^2\Big)^{1/2}\le C(q,\gamma)N\min\left\{1,\left(\frac{1}{N^{p/2}}\sum_{k=1}^m t_k^p\right)^{-1/p}\right\},\quad p:=\frac{q}{q-1},$$

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

Remark 3.2. Note that this greedy process is somewhat more simultaneous (vector) than the VWCGA, that we have defined in Section 2, but we pay a price in that we have a factor N instead of the (smaller) factor $N^{1/p}$ (compare with (2.1)).

Proof of Theorem 3.3. By a theorem of Figiel (see [P, Theorem 2.1]), we know that the space $Y:=\ell_2(X)$ is a uniformly smooth Banach space with $\rho_Y(u)\le C(q,\gamma)u^q$. We perform a WCGA, with weakness sequence $\tau_N:=\{t_k/N^{1/2}\}_{k=1}^\infty$, with respect to the dictionary $\mathcal{D}_N$ (see the beginning of this section with H replaced by X) in Y, beginning with the initial data $F\in\mathcal{A}_1(\mathcal{D}_N,N)$.
Denote the result at step m by

$$F^v_m:=F^{v,\tau}_m:=(f^{1,v,\tau}_m,\dots,f^{N,v,\tau}_m).$$

If $F^v_{m-1}\ne 0$, then

$$(3.12)\qquad \mathcal{F}_{F^v_{m-1}}(\Phi^v_m)=\frac{\|f^{i_m}_{m-1}\|_X}{\|F^v_{m-1}\|}F_{f^{i_m}_{m-1}}(\varphi^v_m).$$

Take $V \in \mathcal{D}_N$, then $V = (\alpha_1 g^1, \dots, \alpha_N g^N)$, where $g^i \in \mathcal{D}$ and $\sum_{i=1}^N \alpha_i^2 = 1$. Hence, by (3.10) and (3.12),

$$
\begin{aligned}
t_m |\mathcal{F}_{F^v_{m-1}}(V)| &\leq t_m \sum_{i=1}^N \frac{\|f^i_{m-1}\|_X}{\|F^v_{m-1}\|} |\alpha_i| |F_{f^i_{m-1}}(g^i)| \\
&\leq \frac{1}{\|F^v_{m-1}\|} \sum_{i=1}^N |\alpha_i| t_m |F_{f^i_{m-1}}(\|f^i_{m-1}\|_X g^i)| \\
&\leq \frac{1}{\|F^v_{m-1}\|} |F_{f^{i_m}_{m-1}}(\|f^{i_m}_{m-1}\|_X \varphi^v_m)| \sum_{i=1}^N |\alpha_i| \\
&\leq N^{1/2} \frac{\|f^{i_m}_{m-1}\|_X}{\|F^v_{m-1}\|} |F_{f^{i_m}_{m-1}}(\varphi^v_m)| \\
&= N^{1/2} |\mathcal{F}_{F^v_{m-1}}(\Phi^v_m)|,
\end{aligned}
$$

where we applied the inequality

$$\sum_{i=1}^N |\alpha_i| \leq N^{1/2} \sum_{i=1}^N \alpha_i^2 = N^{1/2}.$$

Therefore we conclude by Theorem 2.1 that

$$\|F^v_m\| \leq C(q,\gamma) N \min \left\{ 1, \left(\sum_{k=1}^m (t_k / N^{1/2})^p \right)^{-1/p} \right\},$$

which implies (3.11). □

References

[B] Andrew R. Barron, *Universal approximation bounds for superposition of n sigmoidal functions*, IEEE Transactions on Information Theory **39** (1993), 930– 945.

[DT] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Advances in comp. Math. **5** (1996), 173–187.

[DGDS] M. Donahue, L. Gurvits, C. Darken, E. Sontag, *Rate of convex approximation in non-Hilbert spaces*, Constr. Approx. **13** (1997), 187–220.

[J] L. Jones, *On a conjecture of Huber concerning the convergence of projection pursuit regression*, Annals of Stat. **15** (1987), 880–882.

[J1] L. Jones, *A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training*, Annals of Stat. **20** (1992), 608–613.

[LeT] D. Leviatan and V.N. Temlyakov, *Simultaneous approximation by greedy algorithms*, Advances in Comp. Math. (to appear).

[LuT] A. Lutoborski and V.N. Temlyakov, *Vector Greedy Algorithms*, J. Complexity **19** (2003), 458-473.

[P] Gilles Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math. **20** (1975), 326-249.

[T1] V.N. Temlyakov, *Weak greedy algorithms*, Advances in Comp. Math. **12** (2000), 213–227.
[T2] V. N. Temlyakov, *Greedy algorithms in Banach spaces*, Advances in Comp. Math. **14** (2001), 277–292.
[T3] V. N. Temlyakov, *Nonlinear Methods of Approximation*, Found. Comput. Math. **3** (2003), 33–107.
[T4] V. N. Temlyakov, *A remark on simultaneous greedy approximation*, East J. Approx. (2004) (to appear).