SIMULTANEOUS APPROXIMATION BY GREEDY ALGORITHMS

D. Leviatan[1] and V.N. Temlyakov[2]

Abstract. We study nonlinear m-term approximation with regard to a redundant dictionary $\mathcal{D}$ in a Hilbert space H. It is known that 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: $\|f\|^2 = \sum_j |c_j(f)|^2$. Following the paper of A. Lutoborski and the second author [21] we study analogs of the above expansions 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 study convergence and rate of convergence of such expansions which we call *simultanious* expansions.

1. Introduction

In this paper we study nonlinear approximation. The basic idea behind nonlinear approximation is that the elements used in the approximation do not come from a fixed linear space but are allowed to depend on the function being approximated. The classical problem in this regard is the problem of m-term approximation where one fixes a basis in the space, and seeks to approximate a target function f by a linear combination of m terms from that basis. When the basis is a wavelet basis or a basis of other waveforms, then this type of approximation is the starting point for compression algorithms. An important feature of approximation using a basis $\Psi := \{\psi_k\}_{k=1}^{\infty}$ of a Banach space X is that each function $f \in X$ has a unique representation

$$f = \sum_{k=1}^{\infty} c_k(f)\psi_k \tag{1.1}$$

and we can identify f with the set of its coefficients $\{c_k(f)\}_{k=1}^{\infty}$. The problem of m-term approximation with regard to a basis has been studied thoroughly and rather complete results have been established (see [2], [4]–[6], [9]–[11], [15], [19]–[23], [25]–[27], [31], [34]–[37],

[1]Part of this work was done while the first author visited the University of South Carolina in January 2003.
[2]This research was supported by the National Science Foundation Grant DMS 0200187 and by ONR Grant N00014-96-1-1003.

[42], [43]). In particular, it was established that the greedy type algorithm which forms a sum of m terms with the largest $\|c_k(f)\psi_k\|_X$ out of expansion (1.1), in many cases almost realizes the best m-term approximation for function classes ([5]), and even for individual functions ([35], [23]).

Recently, there has emerged another more complicated form of nonlinear approximation which we call highly nonlinear approximation. It takes many forms but has the basic ingredient that the basis is replaced by a larger system of functions that is usually redundant. We call such systems dictionaries. Redundancy on the one hand offers much promise for greater efficiency in terms of approximation rate, but on the other hand gives rise to highly nontrivial theoretical and practical problems. Approximation with regard to a redundant dictionary has been studied in [1], [3], [4], [7], [8], [12]–[14], [16]–[18], [24], [28]–[30], [32], [33], [38]–[42] and other papers. We refer the reader to surveys [4] and [42] for a discussion of approximation results for redundant dictionaries.

We 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 $\overline{\operatorname{span}}\mathcal{D} = H$. In [7], the second author and DeVore studied the following greedy algorithm. If $f \in H$, one lets $g = g(f) \in \mathcal{D}$ be the element from $\mathcal{D}$ which maximizes $|\langle f, g \rangle|$ (of course for this one makes an additional assumption that such a maximizer always exists), and define

$$G(f) := G(f, \mathcal{D}) := \langle f, g \rangle g, \tag{1.2}$$

and

$$R(f) := R(f, \mathcal{D}) := f - G(f). \tag{1.3}$$

Pure Greedy Algorithm (PGA). *Let $R_0(f) := R_0(f, \mathcal{D}) := f$ and $G_0(f) := 0$. Then, for each $m \geq 1$, we inductively define*

$$G_m(f) := G_m(f, \mathcal{D}) := G_{m-1}(f) + G(R_{m-1}(f))$$
$$R_m(f) := R_m(f, \mathcal{D}) := f - G_m(f) = R(R_{m-1}(f)).$$

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 [39]) is defined as follows. Let the sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 < t_k < 1$, be given.

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

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

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

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 $t_k = t$, $k = 1, 2, \ldots$, this algorithm was considered in [17]. 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. Note that the WGA includes, in addition to the first (greedy) step, a second step (see 2., 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. Therefore, the WGA provides for each $f \in H$ an expansion into a series (a greedy expansion)

$$f \sim \sum_{j=1}^{\infty} c_j(f) \varphi_j^\tau, \quad c_j(f) := \langle f_{j-1}^\tau, \varphi_j^\tau \rangle. \tag{1.4}$$

In general it is not an expansion into orthogonal series but it has some similar properties. The coefficients $c_j(f)$ of an expansion are obtained by the Fourier formulas with f replaced by the residuals f_{j-1}^τ. It is easy to see that

$$\|f_m^\tau\|^2 = \|f_{m-1}^\tau\|^2 - |c_m(f)|^2. \tag{1.5}$$

Therefore, for a convergent greedy expansion we get an analogue of the Parseval formula for orthogonal expansions:

$$\|f\|^2 = \sum_{j=1}^{\infty} |c_j(f)|^2.$$

The problem of convergence of the WGA is now settled in the following sense. In [40], a class $\mathcal{V}$ of sequences it has been introduced, such that the condition $\tau \notin \mathcal{V}$ is necessary and sufficient for the convergence of a Weak Greedy Algorithm with weakness sequence τ for each $f \in H$, and all Hilbert spaces H and dictionaries $\mathcal{D}$ (see [40] for the history of this problem). For a general dictionary $\mathcal{D}$, 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 we define $\mathcal{A}_1(\mathcal{D}, M)$ as the closure (in H) of $\mathcal{A}_1^o(\mathcal{D}, M)$. Furthermore, we define $\mathcal{A}_1(\mathcal{D})$ as the union of the classes $\mathcal{A}_1(\mathcal{D}, M)$ over all $M > 0$. For $f \in \mathcal{A}_1(\mathcal{D})$, we define the norm $|f|_{\mathcal{A}_1(\mathcal{D},\infty)}$, as the smallest M such that $f \in \mathcal{A}_1(\mathcal{D}, M)$.

For $M = 1$ we denote $A_1(\mathcal{D}) := \mathcal{A}_1(\mathcal{D}, 1)$. The rate of convergence of the PGA and the WGA for elements from $A_1(\mathcal{D})$ has been studied in [7], [24], [39], [28], [41]. The following result has been obtained in [39].

Theorem 1.1. *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*

$$\|f - G_m^\tau(f, \mathcal{D})\| \leq (1 + \sum_{k=1}^m t_k^2)^{-t_m/2(2+t_m)}. \tag{1.6}$$

While Theorem 1.1 is valid for nonincreasing weakness sequence, we obtain in Section 2 an upper estimate for the rate of convergence of the WGA for a class of weakness sequences which includes nonmonotone sequences.

Theorem 1.2. *Assume a weakness sequence $\tau = \{t_k\}_{k=1}^\infty$ has the property that there are a natural number n, and a real number $0 < t \leq 1$, such that the inequality*

$$n^{-1} \sum_{k=ln+1}^{(l+1)n} t_k^2 \geq t^2,$$

holds for all $l = 0, 1, 2, \dots$. Then if $f \in A_1(\mathcal{D})$, then for any $0 < \delta < 1$ we have

$$\|f_{ln}^\tau\|^2 \leq \left(3n/\delta^2\right)^{\frac{\alpha}{2+\alpha}} \left(1 + lt^2\right)^{-\frac{\alpha}{2+\alpha}}$$

with $\alpha := t(1 - \delta)$.

We also prove in Section 2 that Theorem 1.2 is sharp in a certain sense.

The main purpose of this paper is to construct greedy type (1.4) expansions for a given finite set of elements $f^1, \dots, f^N$, simultaneously with the same sequence $\{\varphi_j^\tau\}$ for all f^i, $i = 1, \dots, N$. The first result in this direction has recently been obtained in [30]. The Vector Greedy Algorithms that are designed for the purpose of constructing mth greedy approximants, simultaneously for a given finite number of elements, have been introduced and studied in [30]. Namely,

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

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

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

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

It was proved in [30] that under certain conditions on τ the VWGA converges. Therefore VWGA provides the convergent expansions

$$f^i = \sum_{j=1}^{\infty} b_j^i g_j, \qquad g_j \in \mathcal{D},$$

with the property

$$\|f^i\|^2 = \sum_{j=1}^{\infty} |b_j^i|^2, \qquad i = 1, \dots, N.$$

The following estimate of the rate of convergence of VWGA has been obtained in [30].

Theorem 1.3. *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 $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$, we have*

$$\sum_{i=1}^{N} \|f_m^{i,v,\tau}\|^2 \le \left(N + mt^2\right)^{-t/(2N+t)} N^{\frac{2N+3t}{2N+t}}.$$

We will improve this estimate in Section 3, proving

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

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

In addition to the VWGA we will consider in Section 3 two modifications of the VWGA. The modifications differ from the VWGA only in the first step. We modify this step in the following two ways. In the first step of the Simultaneous Weak Greedy Algorithm 1 (SWGA1)

1.(SWGA1) We look for any $\varphi_m^{s1,\tau} \in \mathcal{D}$ satisfying

$$\Big(\sum_{i=1}^{N} |\langle f_{m-1}^i, \varphi_m^{s1,\tau}\rangle|^2\Big)^{1/2} \ge t_m \max_i \|f_{m-1}^i\|_{\mathcal{D}}, \quad f_{m-1}^i := f_{m-1}^{i,s1,\tau}. \tag{1.9}$$

In the first step of the Simultaneous Weak Greedy Algorithm 2 (SWGA2)

1.(SWGA2) We look for any $\varphi_m^{s2,\tau} \in \mathcal{D}$ satisfying

$$\Big(\sum_{i=1}^{N} |\langle f_{m-1}^i, \varphi_m^{s2,\tau}\rangle|^2\Big)^{1/2} \ge t_m \sup_{g\in\mathcal{D}} \Big(\sum_{i=1}^{N} |\langle f_{m-1}^i, g\rangle|^2\Big)^{1/2}, \quad f_{m-1}^i := f_{m-1}^{i,s2,\tau}. \tag{1.10}$$

Clearly, any φ_m satisfying (1.8) or (1.10) also satisfies (1.9). Thus, any upper estimate for the SWGA1 yields an upper estimate for both the VWGA and the SWGA2. We prove in Section 3 an extension of Theorem 1.4 which holds for both variants of the Simultaneous Weak Greedy Algorithm (see Theorem 3.1).

2. Rate of convergence of WGA

The following lemma is due to [39].

Lemma 2.1. *Let $\{a_m\}_{m=0}^\infty$ be a sequence of nonnegative numbers satisfying the inequalities*

$$a_0 \le A, \qquad a_m \le a_{m-1}(1 - t_m^2 a_{m-1}/A), \quad m = 1, 2, \dots,$$

with $0 \le t_k \le 1$, $k = 1, 2, \dots$. Then for each m we have

$$a_m \le A(1 + \sum_{k=1}^m t_k^2)^{-1}.$$

We need the following modification of this lemma.

Lemma 2.2. *Let $A \ge 2$ and $0 \le \beta_n \le 1$, $n = 1, 2, \dots$. Suppose $1 \ge x_0 \ge x_1 \ge \dots \ge 0$, satisfy the recurrent inequalities*

$$x_n \le x_{n-1} - \frac{\beta_n}{A} x_n^2. \tag{2.1}$$

Then we have

$$x_m \le \frac{3}{2} A(1 + \sum_{n=1}^m \beta_n)^{-1}, \quad m = 1, 2, \dots. \tag{2.2}$$

Proof. We will use the following simple inequality

$$(1 + x)^{-1} \le 1 - \frac{2}{3}x, \quad 0 \le x \le 1/2. \tag{2.3}$$

We rewrite (2.1) in the form

$$x_n(1 + \frac{\beta_n}{A} x_n) \le x_{n-1}. \tag{2.4}$$

Clearly $x_{n-1} = 0$ implies $x_n = 0$. Thus it suffices to prove (2.2) for nonzero x_m. Using (2.3) we get from (2.4)

$$x_{n-1}^{-1} \le x_n^{-1}(1 + \frac{\beta_n}{A} x_n)^{-1} \le x_n^{-1} - \frac{2}{3}\frac{\beta_n}{A},$$

or

$$x_n^{-1} \ge x_{n-1}^{-1} + \frac{2}{3}\frac{\beta_n}{A}.$$

This implies

$$x_m^{-1} \ge x_0^{-1} + \frac{2}{3A}\sum_{n=1}^m \beta_n \ge 1 + \frac{2}{3A}\sum_{n=1}^m \beta_n \ge \frac{2}{3A}(1 + \sum_{n=1}^m \beta_n).$$

Finally

$$x_m \le \frac{3}{2} A(1 + \sum_{n=1}^{m} \beta_n)^{-1}. \quad \square$$

We are ready to prove Theorem 1.2

Proof of Theorem 1.2. Denote

$$a_m := \|f_m^\tau\|^2, \quad y_m := |\langle f_{m-1}^\tau, \varphi_m^\tau \rangle|, \quad m = 1, 2, \dots, \quad y_0 := 0.$$

Recalling (1.5)

$$\|f_m^\tau\|^2 = \|f_{m-1}^\tau\|^2 - \langle f_{m-1}^\tau, \varphi_m^\tau \rangle^2,$$

which can be rewritten as

$$a_m = a_{m-1} - y_m^2, \tag{2.5}$$

we conclude that $y_m \le 1$, $m \ge 0$. Let the sequence $\{b_n\}$ be defined by

$$b_0 := n/\delta, \quad b_m := b_{m-1} + y_m, \quad m = 1, 2, \dots. \tag{2.6}$$

Then, evidently, $f_m^\tau \in \mathcal{A}_1(\mathcal{D}, b_m)$. By Lemma 3.5 of [7], we get

$$\sup_{g \in \mathcal{D}} |\langle f_{m-1}^\tau, g \rangle| \ge \|f_{m-1}^\tau\|^2 / b_{m-1},$$

which in turn implies (by the definition of φ_m^τ)

$$y_m \ge t_m a_{m-1} / b_{m-1}. \tag{2.7}$$

Denote

$$x_l := a_{ln}, \quad z_l := (\sum_{k=ln+1}^{(l+1)n} y_k^2)^{1/2} \le n^{1/2}, \quad \text{and} \quad w_l := n^{-1/2} b_{ln}.$$

Then (2.5) and (2.6) imply

$$x_{l+1} = x_l - z_l^2, \tag{E1}$$

$$w_{l+1} \le w_l + z_l, \tag{E2}$$

and (2.7) together with (1.7) and the fact that $\{x_l\}$ is decreasing and $\{w_l\}$ is increasing, yields

$$z_l \ge t \frac{x_{l+1}}{w_{l+1}}. \tag{E3}$$

Now, combining (E1) and (E3) it follows that

$$x_{l+1} \le x_l - t^2 \left(\frac{x_{l+1}}{w_{l+1}} \right)^2,$$

or

$$x_{l+1} \left(1 + t^2 \frac{x_{l+1}}{w_{l+1}^2} \right) \le x_l.$$

Again by the monotonicity of $\{w_l\}$ we obtain

$$\frac{x_{l+1}}{w_{l+1}^2} \left(1 + t^2 \frac{x_{l+1}}{w_{l+1}^2} \right) \le \frac{x_l}{w_l^2}.$$

Hence, by Lemma 2.2 with $A = 2$, $\beta_n = t^2$, $n = 1, 2, \dots$, we have

$$\frac{x_l}{w_l^2} \le 3(1 + lt^2)^{-1}. \tag{2.8}$$

Also, (E1) and (E3) imply

$$x_{l+1} \le x_l - z_l t \frac{x_{l+1}}{w_{l+1}},$$

or

$$x_{l+1} \left(1 + t \frac{z_l}{w_{l+1}} \right) \le x_l. \tag{2.9}$$

At the same time (E2) implies

$$w_{l+1} \le w_l (1 + z_l / w_l). \tag{2.10}$$

Thus, combining (2.9) and (2.10) we conclude that

$$x_{l+1} \left(1 + t \frac{z_l / w_l}{1 + z_l / w_l} \right) \le x_l. \tag{2.11}$$

Since $z_l \le n^{1/2}$ and $w_l \ge w_0 := n^{1/2}/\delta$, it follows that $z_l / w_l \le \delta$ for all l. For $\alpha := t(1 - \delta)$ we apply (2.10) and the inequality

$$(1 + x)^\alpha \le 1 + \alpha x \le 1 + t \frac{x}{1 + x}, \quad 0 \le x \le \delta,$$

to obtain

$$\begin{aligned} x_{l+1} w_{l+1}^\alpha &\le x_{l+1} w_l^\alpha (1 + z_l / w_l)^\alpha \\ &\le x_{l+1} \big(1 + t \frac{z_l / w_l}{1 + z_l / w_l} \big) w_l^\alpha \\ &\le x_l w_l^\alpha \le x_0 w_0^\alpha \\ &\le (n^{1/2} / \delta)^\alpha, \end{aligned}$$

where in the third inequality we applied (2.11). Hence, by (2.8) we obtain

$$x_l^{2+\alpha} \leq 3^\alpha (1+lt^2)^{-\alpha} x^2 w_l^{2\alpha} \\ \leq (3n/\delta^2)^\alpha (1+lt^2)^{-\alpha},$$

and

$$x_l \leq (3n/\delta^2)^{\frac{\alpha}{2+\alpha}} (1+lt^2)^{-\frac{\alpha}{2+\alpha}}.$$

This completes the proof of Theorem 1.2. $\square$

An immediate consequence of Theorem 1.2 is

Corollary 2.1. *Let $n \geq 2$ and $1 \leq i \leq n$ be given, and set*

$$t_k^i = \begin{cases} 1, & k = ln + i, \quad l = 0, 1, 2, \ldots, \\ 0 & \textit{otherwise.} \end{cases} \tag{2.12}$$

Then if $f \in A_1(\mathcal{D})$, we have the upper estimate for the error of the WGA

$$\|f_{ln}\|^2 \leq (3n/\delta^2)^{\frac{\alpha}{2+\alpha}} (1+ln^{-1})^{-\frac{\alpha}{2+\alpha}} = (3n^2/\delta^2)^{\frac{\alpha}{2+\alpha}} (l+1)^{-\frac{\alpha}{2+\alpha}}, \quad 0<\delta<1, \tag{2.13}$$

with $\alpha = (1-\delta)n^{-1/2}$.

Thus, we see that the exponent $\frac{\alpha}{2+\alpha}$ in (2.13) decreases with n at the rate $n^{-1/2}$. We will show that for the particular case of a weakness sequence of the form (2.12) the dependence of the exponent ξ_n in

$$\|f_{ln}\|^2 \leq C(n)(l+1)^{-\xi_n}$$

is indeed of order $\xi_n \leq Cn^{-1/2}$.

To this end we use the construction of $\mathcal{D}_t$ from Section 2 of [29]. We begin with the Equalizer procedure. Namely, let H be a Hilbert space with an orthonormal basis $\{e_j\}_{j=1}^\infty$. For two elements e_i, e_j, $i \neq j$, and for a positive number $t \leq 1/3$ the following procedure is called "equalizer" and is denoted $E(e_i, e_j, t)$.

Equalizer $E(e_i, e_j, t)$**.** Set $f_0 := e_i$ and $g_1 := \alpha_1 e_i - (1-\alpha_1^2)^{1/2} e_j$ with $\alpha_1 := t$. Clearly, $\|g_1\| = 1$ and $\langle f_0, g_1 \rangle = t$. We define inductively the sequences $f_1, \ldots, f_N$; $g_2, \ldots, g_N$; and $\alpha_2 \geq 0, \ldots, \alpha_N \geq 0$, with N determined by the process. Let

$$f_n := f_{n-1} - \langle f_{n-1}, g_n \rangle g_n, \quad \text{and} \quad g_{n+1} := \alpha_{n+1} e_i - (1-\alpha_{n+1}^2)^{1/2} e_j,$$

where $\alpha_{n+1} \geq 0$ satisfies

$$\langle f_n, g_{n+1} \rangle = t, \quad n = 1, 2, \ldots.$$

Note that

$$\|f_n\|^2 = \|f_{n-1}\|^2 - t^2, \tag{2.14}$$

so that we can solve for $\alpha_{n+1} \geq 0$ as long as $N \leq [t^{-2}]$. Writing $f_n =: a_n e_i + b_n e_j$, it follows that

$$a_n = a_{n-1} - t\alpha_n, \quad b_n = b_{n-1} + t(1-\alpha_n^2)^{1/2}, \quad n \geq 2,$$
$$a_n - b_n = a_{n-1} - b_{n-1} - t(\alpha_n + (1-\alpha_n^2)^{1/2}), \quad n \geq 2, \tag{2.15}$$

so that, in particular, $a_n - b_n$ in decreasing. Also by virtue of the inequality $1 \leq x + (1-x^2)^{1/2} \leq 2^{1/2}$, $0 \leq x \leq 1$, we see that

$$a_{n-1} - b_{n-1} \leq a_n - b_n + \sqrt{2}t. \tag{2.16}$$

We proceed this way as long as

$$a_n - b_n \geq \sqrt{2}t,$$

arriving at $N = N_t$, such that

$$a_{N-1} - b_{N-1} \geq \sqrt{2}t \quad \text{and} \quad a_N - b_N < \sqrt{2}t.$$

Note that by (2.15) and (2.16),

$$\frac{1}{\sqrt{2}t} - 1 < N_t \leq \frac{1}{t}. \tag{2.17}$$

At this stage we modify the Nth step as follows. We take $g_N := 2^{-1/2}(e_i - e_j)$ and define

$$f_N = f_{N-1} - \langle f_{N-1}, g_N \rangle g_N.$$

It is clear that $a_N = b_N$, and by virtue of (2.16),

$$t \leq \langle f_{N-1}, g_N \rangle \leq 2t. \tag{2.18}$$

It follows from (2.14) and (2.17) that

$$\|f_{N-1}\|^2 \geq 1 - t + t^2,$$

and, in turn, by (2.18), we have

$$\|f_N\|^2 \geq \|f_{N-1}\|^2 - 4t^2 \geq \|f\|^2 - t - 3t^2.$$

Evidently, $E(e_i, e_j, t)$ is a WGA with respect to the dictionary $\mathcal{D}(i,j) := \{e_i, g_1, g_2, \dots, g_N\}$, with the "weakness" parameter t. It is worthwhile to note that the values $\{\alpha_k\}, \{a_k\}$ and $\{b_k\}$, $k = 1, \dots, N_t$, and the stopping stage N_t, depend only on t, and are independent of the choice of e_i and e_j. Also, N_t increases as t decreases, it is constant for a while and then jumps up by 1. Thus, we take $\mu \geq 3$, and $t = t_\mu$, $2^{-\mu-1} \leq t_\mu \leq 2^{-\mu}$, such that $N_t = 2^\mu$.

This can be done since by virtue of (2.17), if $t = 2^{-\mu}$, then $N_t \le 2^\mu$, and if $t = 2^{-\mu-1}$, then $N_t > 2^{\mu+1/2} - 1 \ge 2^\mu$.

We define a WGA with respect to the dictionary $\mathcal{D}_t := \cup_{(i,j)\in S}\mathcal{D}(i,j)$ where S is determined by the equalizer procedures $\{E(e_i, e_j, t)\}_{(i,j)\in S}^\infty$ defined above that will be used in the construction that follows. We begin with $f := e_1$ and apply $E(e_1, e_2, t)$, $t := t_\mu$. After $N_t = 2^\mu$ steps we obtain $g_1^0, \dots, g_{N_t}^0$, and

$$f^1 := c_1(e_1 + e_2), \quad h := 2c_1^2,$$

with the property

$$\|f^1\|^2 = h, \quad h \ge 1 - t - 3t^2.$$

We now obtain $g_1^1, \dots, g_{2N_t}^1$, by applying the equalizers $E(e_1, e_3, t)$ and $E(e_2, e_4, t)$. Thus after $2N_t$ additional steps of the WGA, we have

$$f^2 := c_2(e_1 + \cdots + e_4), \quad c_2 = c_1^2,$$

with the property

$$\|f^2\|^2 = 4c_2^2 = h^2.$$

After μ iterations we have made M_μ steps, where

$$M_\mu = N_t \sum_{k=0}^{\mu-1} 2^k = 2^\mu(2^\mu - 1) =: n - 1,$$

and obtained

$$f^\mu := c_\mu(e_1 + \cdots + e_{2^\mu}), \quad c_\mu = c_{\mu-1}c_1.$$

At the nth step ($n = 2^{2\mu} - 2^\mu + 1$), we remove $c_\mu e_{2^\mu}$ by the PGA step

$$\begin{aligned} f_n :&= f^\mu - \langle f^\mu, e_{2^\mu}\rangle e_{2^\mu} \\ &= c_\mu(e_1 + \cdots + e_{2^\mu-1}), \quad c_\mu^2 = h^\mu 2^{-\mu}. \end{aligned}$$

Indeed,

$$\sup_{g\in\mathcal{D}} \langle f^\mu, g\rangle = c_\mu = \langle f^\mu, e_{2^\mu}\rangle.$$

We proceed as follows to obtain $f^{\mu+1}$. We apply the equalizer procedure $E(e_1, e_{2^\mu+1}, t_\mu)$, $\dots, E(e_{2^\mu-1}, e_{2^\mu+2^\mu-1}, t_\mu)$, thus, we perform $2^\mu(2^\mu - 1) = n - 1$ additional steps of the WGA. We get

$$f^{\mu+1} := c_{\mu+1}(e_1 + \cdots + e_{2^\mu-1} + e_{2^\mu+1} + \cdots + e_{2^{\mu+1}-1}),$$

and we remove $c_{\mu+1}e_{2^\mu-1}$, to obtain f_{2n}.

Suppose that at the νth iteration, $(\nu \geq \mu+1)$, we have arrived at

$$f_{M_\nu} := c_\nu \sum_{i\in\Lambda_\nu} e_i, \quad c_\nu^2 = h^\nu 2^{-\nu}, \quad \Lambda_\nu = \{i_1 < i_2 < \cdots < i_{L_\nu}\} \subseteq [1, 2^\nu].$$

We begin performing the $(\nu+1)$st iteration by applying the equalizer procedure $E(e_{i_1}, e_{2^\nu+1}, t_\mu), \dots, E(e_{i_{2^\mu-1}}, e_{2^\nu+2^\mu-1}, t_\mu)$. Thus, we have performed $2^\mu(2^\mu-1) = n-1$ steps of the WGA. Since $i_{2^\mu-1} < i_{L_\nu}$, we remove $c_\nu e_{i_{L_\nu}}$ by a PGA as in the nth step. We now apply $E(e_{i_{2^\mu}}, e_{2^\nu+2^\mu}, t_\mu), \dots, E(e_{i_{2^{\mu+1}-2}}, e_{2^\nu+2^{\mu+1}-2}, t_\mu)$, and if $i_{2^{\mu+1}-2} < i_{L_\nu-1}$, we remove $c_\nu e_{i_{L_\nu-1}}$, and keep going until we can no longer continue. This means that either the $n-1$ st equalizer is applied to the last remaining element in Λ_ν, or that we are left with less than $n-1$ elements. In the former case we have arrived at

$$f^{\nu+1} := c_{\nu+1} \sum_{i\in\Lambda} e_i, \quad c_{\nu+1}^2 = h^{\nu+1} 2^{-\nu-1}, \quad \Lambda \subseteq [1, 2^{\nu+1}], \tag{2.19}$$

With $\lambda := \max\Lambda$, we then remove $c_{\nu+1}e_\lambda$ in the nth step, and denote $\Lambda_{\nu+1} := \Lambda \setminus \{\lambda\} \subseteq [1, 2^{\nu+1}]$. In the latter case we form equalizers for the remaining elements, and obtain (2.19). We now perform as many WGA steps of the form

$$f^{\nu+1} - 0\langle f^{\nu+1}, e_i\rangle e_i, \quad i < \lambda,$$

as needed in order to have a total of $n-1$ steps and in the nth step we remove $c_{\nu+1}e_\lambda$. As a result in both cases, after $M_{\nu+1}$ steps, we have

$$f_{M_{\nu+1}} := c_{\nu+1} \sum_{i\in\Lambda_{\nu+1}} e_i, \quad c_{\nu+1}^2 = h^{\nu+1} 2^{-\nu-1}, \quad \Lambda_{\nu+1} \subseteq [1, 2^{\nu+1}], \quad |\Lambda_{\nu+1}| =: L_{\nu+1}.$$

It is clear that we have removed at most $\lceil L_\nu/(2^\mu-1)\rceil$ elements e_i. Therefore,

$$L_{\nu+1} \geq 2(L_\nu - L_\nu/(2^\mu-1) - 1) = 2L_\nu\left(\frac{2^\mu-2}{2^\mu-1} - \frac{1}{L_\nu}\right) \geq 2L_\nu(1-2^{-\mu+1}), \tag{2.20}$$

and

$$\|f_{M_{\nu+1}}\|^2 = c_{\nu+1}^2 L_{\nu+1} \geq h^{\nu+1}2^{-\nu}L_\nu(1-2^{-\mu+1}) = h(1-2^{-\mu+1})\|f_{M_\nu}\|^2. \tag{2.21}$$

Also

$$\begin{aligned} M_{\nu+1} &\geq M_\nu + (L_\nu - (\lceil L_\nu/(2^\mu-1)\rceil))2^\mu + \lceil L_\nu/(2^\mu-1)\rceil \\ &\geq M_\nu + L_\nu 2^\mu - \lceil L_\nu/(2^\mu-1)\rceil(2^\mu-1) \geq M_\nu + L_\nu(2^\mu-2). \end{aligned}$$

Taking into account that

$$M_\mu = 2^{2\mu} - 2^\mu + 1, \quad \text{and} \quad L_\mu = 2^\mu - 1,$$

we get by (2.20)

$$M_\nu \geq \big(2(1-2^{-\mu+1})\big)^{\nu-\mu} 2^{-\mu}(2^\mu - 2) \geq C(\mu) 2^{c\nu}, \quad \nu \geq \mu, \tag{2.22}$$

with absolute constant $c > 0$, since $\mu \geq 3$. After M_ν steps we have by (2.21)

$$\begin{aligned}\|f_{M_\nu}\|^2 &\geq h^{\nu-\mu}(1-2^{-\mu+1})^{\nu-\mu}\|f_{M_\mu}\|^2 \geq (1-2^{-\mu+1})^{2\nu-\mu+1} \\ &\geq C(\mu) 2^{-C_1 \nu 2^{-\mu}} \geq C(\mu) M_\nu^{-C_2 2^{-\mu}},\end{aligned}$$

where we have applied the fact that $\|f_{M_\mu}\|^2 = h^\mu(1-2^{-\mu})$, and for the last inequality we used (2.22). Observing that $n^{-1/2} \leq \sqrt{2}2^{-\mu}$, we conclude that the exponent of the power rate of decrease of $\|f_{M_\nu}\|^2$ is of order of $n^{-1/2}$.

3. Simultaneous approximation by greedy algorithm

Given are a Hilbert space H and a dictionary $\mathcal{D}$. For $N \geq 2$, let $H_N := H \times \cdots \times H$, N 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.

We begin with $F_0 := (f_0^1, \dots, f_0^N)$ and a sequence $0 \leq t_m \leq 1$ and we want to construct weak greedy approximation from $\mathcal{D}$, simultaneously to all N functions. For a given F we are looking for an element $G \in \mathcal{D}_N$ of a special form

$$\begin{aligned}G &:= G(F,g) := (\beta_1 g, \beta_2 g, \dots, \beta_N g), \quad g \in \mathcal{D}, \\ \beta_i &:= \langle f^i, g \rangle \big(\sum_{i=1}^N |\langle f^i, g\rangle|^2\big)^{-1/2}, \quad i = 1, \dots, N.\end{aligned} \tag{3.1}$$

For G of the form (3.1) the operation

$$F_1 := F - \langle F, G \rangle G$$

means the same operation performed coordinatewise

$$f_1^i := f^i - \langle f^i, g \rangle g, \quad i = 1, \dots, N.$$

We note that

$$\|F\|_{\mathcal{D}_N} = \sup_{\substack{\alpha := (\alpha_1, \dots, \alpha_N) \\ \|\alpha\|_2 = 1 \\ g_1, \dots, g_N \in \mathcal{D}}} \big|\sum_{i=1}^N \langle f^i, g_i \rangle \alpha_i\big| = \big(\sum_{i=1}^N \|f^i\|_{\mathcal{D}}^2\big)^{1/2}. \tag{3.2}$$

Lemma 3.1. *For any $F \in H_N$ we have*

$$\sup_{g\in\mathcal{D}} |\langle F, G(F,g)\rangle| \ge \max_i \|f^i\|_{\mathcal{D}} \ge N^{-1/2}\|F\|_{\mathcal{D}_N}.$$

Proof. On the one hand,

$$\sup_{g\in\mathcal{D}} |\langle F, G(F,g)\rangle| = \sup_{g\in\mathcal{D}} \Big(\sum_{i=1}^N |\langle f^i, g\rangle|^2\Big)^{1/2} \ge \max_i \sup_{g\in\mathcal{D}} |\langle f^i, g\rangle| = \max_i \|f^i\|_{\mathcal{D}}, \tag{3.3}$$

and on the other, by (3.2),

$$\|F\|_{\mathcal{D}_N} = \Big(\sum_{i=1}^N \|f^i\|_{\mathcal{D}}^2\Big)^{1/2} \le N^{1/2} \max_i \|f^i\|_{\mathcal{D}}. \tag{3.4}$$

Combining (3.3) and (3.4), completes the proof of Lemma 3.1. □

Given a weakness sequence $\tau = \{t_k\}_{k=1}^\infty$. The upper estimate for the VWGA, namely, for $\sum_{i=1}^N \|f_m^{i,v,\tau}\|^2$, can be obtained by Lemma 3.1 from the corresponding upper estimate for the WGA with the weakness sequence $\tau' := \{t_k N^{-1/2}\}_{k=1}^\infty$. Actually we do better, we formulate two theorems which are valid for VWGA and for both SWGA1 and SWGA2. Thus let s stand for either v or $s1$ or $s2$.

Theorem 3.1. *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 $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$, we have*

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

Corollary 3.1. *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^\infty$, $t_k = t$, $k \ge 1$, $0 < t \le 1$. Then for any vector of elements $f^1, \dots, f^N$, $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$, we have*

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

Note that for $s = v$, Corollary 3.1 coincides with Theorem 1.4.

Proof. The proof follows from Theorem 1.1 and Lemma 3.1, when we observe that $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$ implies $(f^1, \dots, f^N) \in \mathcal{A}_1(\mathcal{D}_N, N)$. □

A similar proof yields

Theorem 3.2. *Assume that for the weakness sequence* $\tau = \{t_k\}_{k=1}^{\infty}$ *there are a natural number* n *and a real number* $0 < t \le 1$ *such that*

$$n^{-1}\sum_{k=ln+1}^{(l+1)n} t_k^2 \ge t^2, \quad l = 0, 1, 2, \dots.$$

Then for any $0 < \delta < 1$ *and all* $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$,

$$\sum_{i=1}^{N} \|f_{ln}^{i,s,\tau}\|^2 \le N^2\big(3n/\delta^2\big)^{\frac{r}{2+r}}\big(1 + lt^2\big)^{-\frac{r}{2+r}}$$

with $r := t(1-\delta)N^{-1/2}$.

We are in a position to discuss the convergence of the VWGA, SWGA1, and SWGA2. We denote by $\mathcal{V}$ the class of all sequences $x = \{x_k\}_{k=1}^{\infty}$, $x_k \ge 0$, $k = 1, 2, \dots$, for which there exists a sequence $0 = q_0 < q_1 < \dots$ such that,

$$\sum_{s=1}^{\infty} \frac{2^s}{\Delta q_s} < \infty,$$

where $\Delta q_s := q_s - q_{s-1}$, and

$$\sum_{s=1}^{\infty} 2^{-s} \sum_{k=1}^{q_s} x_k^2 < \infty.$$

Remark 3.1. It is clear from this definition that if $x \in \mathcal{V}$ and for some $K \ge 1$ and c we have $0 \le y_k \le cx_k$, $k \ge K$, then $y := \{y_k\}_{k=1}^{\infty} \in \mathcal{V}$. The following theorem has been proved in [40].

Theorem 3.3. *The condition* $\tau \notin \mathcal{V}$ *is necessary and sufficient for the convergence of the Weak Greedy Algorithm with a weakness sequence* τ, *for each* f *and all Hilbert spaces* H *and dictionaries* $\mathcal{D}$.

It is clear from Theorem 3.3 that the condition $\tau \notin \mathcal{V}$ is also necessary for convergence of the VWGA, SWGA1, and SWGA2 with the weakness sequence τ. It has been proved in [30] that this condition ($\tau \notin \mathcal{V}$) is also sufficient for the convergence of the VWGA. We note that $\tau = \{t_k\} \notin \mathcal{V}$ implies $\tau' := \{t_k N^{-1/2}\} \notin \mathcal{V}$. Thus Theorem 3.3 combined with Lemma 3.1 implies the following generalization of Theorem 3.3.

Theorem 3.4. *The condition* $\tau \notin \mathcal{V}$ *is necessary and sufficient for the convergence of each of the algorithms VWGA, SWGA1, SWGA2 with a weakness sequence* τ, *for each vector of elements* $f^1, \dots, f^N$, N *arbitrary, and all Hilbert spaces* H *and dictionaries* $\mathcal{D}$.

Theorems 3.1 and 3.2 give estimates for the ℓ_2^N-norm of the residual vector $(\|f_m^1\|, \dots, \|f_m^N\|)$. We wish to introduce greedy type algorithms that yield estimates for the ℓ_∞^N-norm of the residual vector. We define the Alternating Weak Greedy Algorithm for N elements (AWGA).

Again, it differs from the VWGA only at the first step (out of three) of each iteration. Let $t \in (0,1]$. At the mth iteration, $m = lN + i$, in the first step of the AWGA

1.(AWGA) We look for any $\varphi_m^{a,\tau} \in \mathcal{D}$ satisfying

$$|\langle f_{m-1}^{i,a,\tau}, \varphi_m^{a,\tau} \rangle| \geq t \|f_{m-1}^{i,a,\tau}\|_{\mathcal{D}}.$$

It is clear that for each i any realization of the AWGA for the ith component f^i can be viewed as a realization of the WGA with the weakness sequence $\tau^i := \{t_k^i\}_{k=1}^{\infty}$,

$$t_k^i = \begin{cases} 1, & k = lN + i, \quad l = 0, 1, 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3.5. *Given $f^i \in A_1(\mathcal{D})$, $i = 1, \dots, N$, the AWGA yields the estimates*

$$\|f_{lN}^i\|^2 \leq (3N^2/\delta^2)^{\frac{\alpha}{2+\alpha}} (1+l)^{-\frac{\alpha}{2+\alpha}}, \quad 0 < \delta < 1, \quad 1 \leq i \leq N,$$

with $\alpha = (1-\delta)N^{-1/2}$.

References

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

[2] A. Cohen, R.A. DeVore, and R. Hochmuth, *Restricted Nonlinear Approximation*, Constructive Approx. **16** (2000), 85–113.

[3] G. Davis, S. Mallat, and M. Avellaneda, *Adaptive greedy approximations*, Constr. Approx. **13** (1997), 57–98.

[4] R.A. DeVore, *Nonlinear Approximation*, Acta Numerica (1998), 51–150.

[5] R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, Amer. J. of Math. **114** (1992), 737–785.

[6] R.A. DeVore and V.N. Temlyakov, *Nonlinear approximation by trigonometric sums*, J. Fourier Anal. and Appl. **2** (1995), 29–48.

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

[8] R.A. DeVore and V.N. Temlyakov, *Nonlinear approximation in finite-dimensional spaces*, J. Complexity **13** (1997), 489–508.

[9] S.J. Dilworth, N.J. Kalton, D. Kutzarova, V.N. Temlyakov, *The Tresholding Greedy Algorithm, Greedy Bases, and Duality*, IMI-Preprints series **23** (2001), 1–23.

[10] D.L. Donoho, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmon. Anal. **1** (1993), 100–115.

[11] D.L. Donoho, *CART and Best-Ortho-Basis: A Connection*, Preprint (1995), 1–45.

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

[13] V.V. Dubinin, *Greedy Algorithms and Applications*, Ph.D. Thesis, University of South Carolina, 1997.

[14] J.H. Friedman and W. Stuetzle, *Projection pursuit regression*, J. Amer. Stat. Assoc. **76** (1981), 817–823.

[15] R. Gribonval and M. Nielsen, *Some remarks on non-linear approximation with Schauder bases*, East J. Approx. **7** (2001), 267–285.

[16] P.J. Huber, *Projection Pursuit*, Annals of Stat. **13** (1985), 435–475.

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

[18] 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.

[19] A. Kamont and V.N. Temlyakov, *Greedy approximation and the multivariate Haar system*, IMI-Preprint series **20** (2002), 1–24.

[20] B. S. Kashin and V. N. Temlyakov, *On best m-terms approximations and the entropy of sets in the space* L^1, Math. Notes **56** (1994), 57–86.

[21] B.S. Kashin and V.N. Temlyakov, *On estimating approximative characteristics of classes of functions with bounded mixed derivative*, Math. Notes **58** (1995), 922–925.

[22] G. Kerkyacharian and D. Picard, *Entropy, universal coding, approximation and bases properties*, University of Paris 6 and 7, Preprint **663** (2001), 1–32.

[23] S.V. Konyagin and V.N. Temlyakov, *A remark on greedy approximation in Banach spaces*, East J. on Approx. **5** (1999), 1–15.

[24] S.V. Konyagin and V.N. Temlyakov, *Rate of convergence of Pure Greedy Algorithm*, East J. on Approx. **5** (1999), 493–499.

[25] S.V. Konyagin and V.N. Temlyakov, *Convergence of Greedy Approximation I. General Systems*, IMI-Preprint series **08** (2002), 1–19.

[26] S.V. Konyagin and V.N. Temlyakov, *Convergence of Greedy Approximation II. The Trigonometric system*, IMI-Preprint series **09** (2002), 1–25.

[27] S.V. Konyagin and V.N. Temlyakov, *Greedy Approximation with regard to bases and general minimal systems*, Serdica Math. J. **28** (2002), 305–328.

[28] E.D. Livshitz, *On the rate of convergence of greedy algorithm*, Manuscript (2000).

[29] E.D. Livshitz and V.N. Temlyakov, *On convergence of Weak Greedy Algorithms,*, IMI-Preprint **13** (2000), 1–9.

[30] A. Lutoborski and V.N. Temlyakov, *Vector Greedy Algorithms*, IMI-Preprint **10** (2002), 1–16.

[31] P. Oswald, *Greedy algorithms and best m-term approximation with respect to biorthogonal systems*, Preprint (2000), 1–22.

[32] L. Rejtö and G.G. Walter, *Remarks on projection pursuit regression and density estimation*, Stochastic Analysis and Application **10** (1992), 213–222.

[33] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I*, Math. Annalen **63** (1906-1907), 433–476.

[34] V.N. Temlyakov, *Greedy algorithm and m-term trigonometric approximation*, Constr. Approx. **14** (1998), 569–587.

[35] V.N. Temlyakov, *The best m-term approximation and Greedy Algorithms*, Advances in Comp. Math. **8** (1998), 249–265.

[36] V.N. Temlyakov, *Nonlinear m-term approximation with regard to the multivariate Haar system*, East J. Approx. **4** (1998), 87–106.

[37] V.N. Temlyakov, *Greedy algorithms with regard to the multivariate systems with a special structure*, Constr. Approx. **16** (2000), 399–425.

[38] V.N. Temlyakov, *Greedy algorithms and m-term approximation with regard to redundant dictionaries*, J. Approx. Theory **98** (1999), 117–145.

[39] V.N. Temlyakov, *Weak greedy algorithms*, Advances in Comp. Math. **12** (2000), 213–227.

[40] V.N. Temlyakov, *A criterion for convergence of Weak Greedy Algorithms*, Advances in Comp. Math. **17** (2002), 269–280.

[41] V.N. Temlyakov, *Two lower estimates in greedy approximation*, IMI-Preprint series **07** (2001), 1–12.

[42] V.N. Temlyakov, *Nonlinear Methods of Approximation*, IMI-Preprint series **09** (2001), 1–57.

[43] P. Wojtaszczyk, *Greedy algorithms for general systems*, J. Approx. Theory **107** (2000), 293–314.

School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel
leviatan@math.tau.ac.il

Department of Mathematics, University of South Carolina, Columbia, SC 29208 USA
temlyak@math.sc.edu