Simultaneous approximation by greedy algorithms

D. Leviathan and V.N. Temlyakov
Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE
   2003

2. REPORT TYPE

3. DATES COVERED
   00-00-2003 to 00-00-2003

4. TITLE AND SUBTITLE
   Simultaneous approximation by greedy algorithms

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
   University of South Carolina, Department of Mathematics, Columbia, SC, 29208

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
   Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:
   a. REPORT
      unclassified
   b. ABSTRACT
      unclassified
   c. THIS PAGE
      unclassified

17. LIMITATION OF ABSTRACT
   Same as Report (SAR)

18. NUMBER OF PAGES
   18

19a. NAME OF RESPONSIBLE PERSON

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18
SIMULTANEOUS APPROXIMATION BY GREEDY ALGORITHMS

D. Leviatan\textsuperscript{1} and V.N. Temlyakov\textsuperscript{2}

\textbf{Abstract.} We study nonlinear \textit{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

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

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 $f^1, \ldots, f^N$ with a requirement that the dictionary elements $\varphi_j$ of these expansions are the same for all $f^i$, $i = 1, \ldots, N$. We study convergence and rate of convergence of such expansions which we call \textit{simultaneous} expansions.

1. \textbf{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 \textit{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

\begin{equation}
(1.1) \quad f = \sum_{k=1}^{\infty} c_k(f)\psi_k
\end{equation}

and we can identify $f$ with the set of its coefficients $\{c_k(f)\}_{k=1}^{\infty}$. The problem of \textit{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],
\textsuperscript{1}Part of this work was done while the first author visited the University of South Carolina in January 2003.
\textsuperscript{2}This research was supported by the National Science Foundation Grant DMS 0200187 and by ONR Grant N00014-96-1-1003.
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 \( D \) of functions (elements) from \( H \) is a dictionary if
each \( g \in D \) has norm one (\( \| g \| = 1 \)) and \( \text{span} D = H \). In [7], the second author and DeVore
studied the following greedy algorithm. If \( f \in H \), one lets \( g = g(f) \in D \) be the element
from \( D \) which maximizes \( |\langle f, g \rangle| \) (of course for this one makes an additional assumption that
such a maximizer always exists), and define

\[
G(f) := G(f, D) := \langle f, g \rangle g,
\]

and

\[
R(f) := R(f, D) := f - G(f).
\]

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

\[
G_m(f) := G_m(f, D) := G_{m-1}(f) + G(R_{m-1}(f))
\]

\[
R_m(f) := R_m(f, D) := f - G_m(f) = R(R_{m-1}(f)).
\]

For a given dictionary \( D \) we can introduce a norm associated with \( D \) as

\[
\| f \|_D := \sup_{g \in D} |\langle f, g \rangle|.
\]

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 \( \varphi_m^\tau \in D \) be any element satisfying

\[
|\langle f_{m-1}^\tau, \varphi_m^\tau \rangle| \geq t_m \| f_{m-1}^\tau \|_D;
\]
2. 
\[ f_m^\tau := f_{m-1}^\tau - \langle f_{m-1}^\tau, \varphi_m^\tau \rangle \varphi_m^\tau; \]

3. 
\[ G_m^\tau(f, D) := \sum_{j=1}^m \langle f_{j-1}^\tau, \varphi_j^\tau \rangle \varphi_j^\tau. \]

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 \( \varphi_m^\tau \) 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 \( f_{m-1}^\tau \) onto \( \varphi_m^\tau \). Therefore, the WGA provides for each \( f \in H \) an expansion into a series (a greedy expansion)

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

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 \( f_{j-1}^\tau \). It is easy to see that

\[ \|f_m^\tau\|^2 = \|f_{m-1}^\tau\|^2 - |c_m(f)|^2. \]

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 \( \tau \) for each \( f \in H \), and all Hilbert spaces \( H \) and dictionaries \( D \) (see \[40\] for the history of this problem). For a general dictionary \( D \), we define the class of functions

\[ \mathcal{A}_1^0(D, M) := \{ f \in H : f = \sum_{k \in \Lambda} c_k w_k, \ w_k \in D, \ #\Lambda < \infty \text{ and } \sum_{k \in \Lambda} |c_k| \leq M \} \]

and we define \( \mathcal{A}_1(D, M) \) as the closure (in \( H \)) of \( \mathcal{A}_1^0(D, M) \). Furthermore, we define \( \mathcal{A}_1(D) \) as the union of the classes \( \mathcal{A}_1(D, M) \) over all \( M > 0 \). For \( f \in \mathcal{A}_1(D) \), we define the norm \( |f|_{\mathcal{A}_1(D, \infty)} \), as the smallest \( M \) such that \( f \in \mathcal{A}_1(D, M) \).

For \( M = 1 \) we denote \( \mathcal{A}_1(D) := \mathcal{A}_1(D, 1) \). The rate of convergence of the PGA and the WGA for elements from \( \mathcal{A}_1(D) \) has been studied in \[7, 24, 39, 28, 41\]. The following result has been obtained in \[39\].
Theorem 1.1. Let $D$ be an arbitrary dictionary in $H$. Assume $\tau := \{t_k\}_{k=1}^{\infty}$ is a nonincreasing sequence. Then for $f \in A_1(D)$ we have

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

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

\[
n^{-1} \sum_{k=m+1}^{(l+1)n} t_k^2 \geq t^2,
\]

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

\[
\|f^n_{m} \| \leq \left(3n / \delta^2 \right)^{\frac{n}{m-\alpha}} (1 + lt^2) - \frac{\alpha}{\delta^2}
\]

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 $f^1, \ldots, f^n$, simultaneously with the same sequence $\{\varphi_j\}$ for all $f^i$, $i = 1, \ldots, N$. The first result in this direction has recently been obtained in [30]. The Vector Greedy Algorithms that are designed for the purpose of constructing $m$th 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 $f^i \in H$, $i = 1, \ldots, N$ be given. We write $f^{i,\nu,\tau} := f^i$. Then for each $m \geq 1$, we inductively define:

1. Let $\varphi_{m}^{\nu,\tau} \in D$ be any element satisfying

\[
\max_{i} |\langle f^{i,\nu,\tau}_{m-1}, \varphi_{m}^{\nu,\tau} \rangle| \geq t_m \max_{i} \|f^{i,\nu,\tau}_{m-1}\|_D,
\]

2. $f^{i,\nu,\tau}_{m} := f^{i,\nu,\tau}_{m-1} - \langle f^{i,\nu,\tau}_{m-1}, \varphi_{m}^{\nu,\tau} \rangle \varphi_{m}^{\nu,\tau}$, $i = 1, \ldots, N$,

3. $G^{\nu,\tau}_m(f^i, D) := \sum_{j=1}^{m} \langle f^{j,\nu,\tau}_{j-1}, \varphi_{j}^{\nu,\tau} \rangle \varphi_{j}^{\nu,\tau}$, $i = 1, \ldots, N$. 


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

\[
f^i = \sum_{j=1}^{\infty} b^i_j g_j, \quad g_j \in \mathcal{D},
\]

with the property

\[
\|f^i\|^2 = \sum_{j=1}^{\infty} |b^i_j|^2, \quad i = 1, \ldots, N.
\]

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 \( \tau := \{t_k\}_{k=1}^\infty, t_k = t, k = 1, \ldots, 0 < t < 1 \). Then for any vector of elements \( f^1, \ldots, f^N, f^i \in A_1(\mathcal{D}), i = 1, \ldots, N \), we have

\[
\sum_{i=1}^{N} \|f^i_{m,\nu,\tau}\|^2 \leq (N + mt^2)^{-t/(2N^2 + t)} N^{2N^2+3}. N^N.
\]

We will improve this estimate in Section 3, proving

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

\[
\sum_{i=1}^{N} \|f^i_{m,\nu,\tau}\|^2 \leq N^2(1 + mt^2 /N)^{N^2+1}. N^{N^2+1}.
\]

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 \( \varphi^{s,1,\tau}_m \in \mathcal{D} \) satisfying

\[
(1.9) \quad \left( \sum_{i=1}^{N} \|f^i_{m-1}, \varphi^{s,1,\tau}_m\|^2 \right)^{1/2} \geq t_m \max_i \|f^i_{m-1}\|_{\mathcal{D}}, \quad f^i_{m-1} := f^i_{m-1,\tau}.
\]

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

1. (SWGA2) We look for any \( \varphi^{s,2,\tau}_m \in \mathcal{D} \) satisfying

\[
(1.10) \quad \left( \sum_{i=1}^{N} \|f^i_{m-1}, \varphi^{s,2,\tau}_m\|^2 \right)^{1/2} \geq t_m \sup_{g \in \mathcal{D}} \left( \sum_{i=1}^{N} \|f^i_{m-1}, g\|^2 \right)^{1/2}, \quad f^i_{m-1} := f^i_{m-1,\tau}.
\]

Clearly, any \( \varphi_m \) 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
\[
a_0 \leq A, \quad a_m \leq a_{m-1}(1-t^2_m a_{m-1}/A), \quad m = 1, 2, \ldots,
\]
with \( 0 \leq t_k \leq 1, \ k = 1, 2, \ldots \). Then for each \( m \) we have
\[
a_m \leq A(1 + \sum_{k=1}^{m} t_k^2)^{-1}.
\]

We need the following modification of this lemma.

**Lemma 2.2.** Let \( A \geq 2 \) and \( 0 \leq \beta_n \leq 1, \ n = 1, 2, \ldots \). Suppose \( 1 \geq x_0 \geq x_1 \geq \cdots \geq 0 \), satisfy the recurrent inequalities
\[
x_n \leq x_{n-1} - \frac{\beta_n}{A} x_n^2,
\]
(2.1)

Then we have
\[
x_m \leq \frac{3}{2} A(1 + \sum_{n=1}^{m} \beta_n)^{-1}, \quad m = 1, 2, \ldots.
\]
(2.2)

**Proof.** We will use the following simple inequality
\[
(1 + x)^{-1} \leq 1 - \frac{2}{3} x, \quad 0 \leq x \leq 1/2.
\]
(2.3)

We rewrite (2.1) in the form
\[
x_n(1 + \frac{\beta_n}{A} x_n) \leq x_{n-1}.
\]
(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)
\[
x_{n-1}^{-1} \leq x_{n-1}^{-1}(1 + \frac{\beta_n}{A} x_n)^{-1} \leq x_{n-1}^{-1} - \frac{2 \beta_n}{3 A},
\]
or
\[
x_{n-1}^{-1} \geq x_{n-1}^{-1} + \frac{2 \beta_n}{3 A}.
\]
This implies
\[
x_m^{-1} \geq x_0^{-1} + \frac{2}{3 A} \sum_{n=1}^{m} \beta_n \geq 1 + \frac{2}{3 A} \sum_{n=1}^{m} \beta_n \geq \frac{2}{3 A}(1 + \sum_{n=1}^{m} \beta_n).
\]
Finally

\[ x_m \leq \frac{3}{2} A (1 + \sum_{n=1}^{m} \beta_n)^{-1}. \]

We are ready to prove Theorem 1.2

**Proof of Theorem 1.2.** Denote

\[ a_m := \|f_m^T\|^2, \quad y_m := |\langle f^T_{m-1}, \varphi^T_m \rangle|, \quad m = 1, 2, \ldots, \quad y_0 := 0. \]

Recalling (1.5)

\[ \|f_m^T\|^2 = \|f^T_{m-1}\|^2 - \langle f^T_{m-1}, \varphi^T_m \rangle, \]

which can be rewritten as

\[ a_m = a_m - y_m^2, \]

we conclude that \( y_m \leq 1, \ m \geq 0. \) Let the sequence \( \{b_n\} \) be defined by

\[ b_0 := n/\delta, \quad b_m := b_{m-1} + y_m, \quad m = 1, 2, \ldots. \]

Then, evidently, \( f_m^T \in A_1(\mathcal{D}, b_m). \) By Lemma 3.5 of [7], we get

\[ \sup_{g \in \mathcal{D}} |\langle f^T_{m-1}, g \rangle| \geq \|f_m^T\|^2 / b_{m-1}, \]

which in turn implies (by the definition of \( \varphi^T_m \))

\[ y_m \geq t_m a_{m-1} / b_{m-1}. \]

Denote

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

Then (2.5) and (2.6) imply

\[ x_{l+1} = x_l - z_l^2, \]

\[ w_{l+1} \leq w_l + z_l, \]

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

\[ z_l \geq \frac{x_{l+1}}{w_{l+1}}. \]
Now, combining (E1) and (E3) it follows that
\[
x_{l+1} \leq x_l - t^2 \left( \frac{x_{l+1}}{w_{l+1}} \right)^2,
\]
or
\[
x_{l+1} \left( 1 + t^2 \frac{x_{l+1}}{w_{l+1}^2} \right) \leq x_l.
\]
Again by the monotonicity of \( \{w_l\} \) we obtain
\[
\frac{x_{l+1}}{w_{l+1}^2} \left( 1 + t^2 \frac{x_{l+1}}{w_{l+1}^2} \right) \leq \frac{x_l}{w_l^2}.
\]
Hence, by Lemma 2.2 with \( A = 2, \beta_n = t^2, n = 1, 2, \ldots, \) we have
\[
\frac{x_l}{w_l^2} \leq 3(1 + lt^2)^{-1}.
\]
Also, (E1) and (E3) imply
\[
x_{l+1} \leq x_l - z_l t \frac{x_{l+1}}{w_{l+1}},
\]
or
\[
x_{l+1} \left( 1 + t \frac{z_l}{w_{l+1}} \right) \leq x_l.
\]
At the same time (E2) implies
\[
w_{l+1} \leq w_l \left( 1 + z_l/w_l \right).
\]
Thus, combining (2.9) and (2.10) we conclude that
\[
x_{l+1} \left( 1 + t \frac{z_l}{1 + z_l/w_l} \right) \leq x_l.
\]
Since \( z_l \leq n^{1/2} \) and \( w_l \geq w_0 := n^{1/2}/\delta \), it follows that \( z_l/w_l \leq \delta \) for all \( l \). For \( \alpha := t(1 - \delta) \) we apply (2.10) and the inequality
\[
(1 + x)^\alpha \leq 1 + \alpha x \leq 1 + t \frac{x}{1 + x}, \quad 0 \leq x \leq \delta,
\]
to obtain
\[
x_{l+1} w_{l+1}^\alpha \leq x_{l+1} w_l^\alpha (1 + z_l/w_l)^\alpha
\]
\[
\leq x_{l+1} \left( 1 + t \frac{z_l}{1 + z_l/w_l} \right) w_l^\alpha
\]
\[
\leq x_l w_l^\alpha \leq x_0 w_0^\alpha
\]
\[
\leq \left( n^{1/2}/\delta \right)^\alpha,
\]
where in the third inequality we applied (2.11). Hence, by (2.8) we obtain
\[ x_l^{2+\alpha} \leq 3^\alpha (1 + lt^2)^{-\alpha} x_l^{2\alpha} \]
\[ \leq (3n/\delta^2)^\alpha (1 + lt^2)^{-\alpha}, \]
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. □

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
\[ t^i_k = \begin{cases} 1, & k = \ln + i, \quad l = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \]
Then if \( f \in A_1(D) \), we have the upper estimate for the error of the WGA
\[ \|f_n\|^2 \leq (3n/\delta^2)^{\frac{\alpha}{2+\alpha}} (1 + ln^{-1})^{-\frac{\alpha}{2+\alpha}} = (3n^2/\delta^2)^{\frac{\alpha}{2+\alpha}} (l + 1)^{-\frac{\alpha}{2+\alpha}}, \quad 0 < \delta < 1, \]
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 \( \xi_n \) in
\[ \|f_n\|^2 \leq C(n)(l + 1)^{-\xi_n} \]
is indeed of order \( \xi_n \leq Cn^{-1/2} \).

To this end we use the construction of \( D_t \) 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
\[ \|f_n\|^2 = \|f_{n-1}\|^2 - t^2, \]
so that we can solve for $\alpha_{n+1} \geq 0$ as long as $N \leq [t^{-2}]$. Writing $f_n = a_ne_i + b_ne_j$, it follows that

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

so that, in particular, $a_n - b_n$ in decreasing. Also by virtue of the inequality

$$1 \leq x + (1 - x^2)^{1/2} \leq 2^{1/2}, \quad 0 \leq x \leq 1,$$

we see that

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

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),

$$(2.17) \quad \frac{1}{\sqrt{2}t} - 1 < N_t \leq \frac{1}{t}.$$

At this stage we modify the $N$th 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),

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

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

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

Evidently, $E(e_i, e_j, t)$ is a WGA with respect to the dictionary $\mathcal{D}(i, j) := \{e_i, g_1, g_2, \ldots, g_N\}$, with the "weakness" parameter $t$. It is worthwhile to note that the values $\{\alpha_k\}, \{a_k\}$ and $\{b_k\}, k = 1, \ldots, N_t$, 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 $N_t \leq 2^\mu$, and if $t = 2^{-\mu - 1}$, then $N_t > 2^{\mu + 1/2} - 1 \geq 2^\mu$.

We define a WGA with respect to the dictionary $D_t := \bigcup_{(i,j) \in S} D(i,j)$ 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 $g_1^0, \ldots, g_{N_t}^0$, and

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

with the property

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

We now obtain $g_1^1, \ldots, g_{2N_t}^1$, 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 $\mu$ iterations we have made $M_\mu$ steps, where

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

and obtained

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

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

$$f_n := f^\mu - \langle f^\mu, e_{2^\nu} \rangle e_{2^\nu}$$

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

Indeed,

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

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

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

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


Suppose that at the \( \nu \)th 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

\[
E(e_{i_1}, e_{2^{\nu}+1}, t_{\mu}), \ldots, E(e_{i_{2^{\mu} - 1}}, e_{2^{\nu}+2^{\mu} - 1}, t_{\mu}).
\]

Thus, we have performed \( 2^{\mu}(2^\mu - 1) = n - 1 \) steps of the WGA. Since \( i_{2^{\mu} - 1} < L_{\nu} \), we remove \( c_{\nu} e_{i_{L_{\nu}}} \) by a PGA as in the \( n \)th step. We now apply \( E(e_{i_2}, e_{2^{\nu}+2^{\mu} + 1}, t_{\mu}), \ldots, E(e_{i_{2^{\mu} + 1} - 2}, e_{2^{\nu}+2^{\mu+1} - 2}, t_{\mu}) \), and if \( i_{2^{\mu+1} - 2} < L_{\nu - 1} \), 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 \( \Lambda_{\nu} \), or that we are left with less than \( n - 1 \) elements. In the former case we have arrived at

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

With \( \lambda := \max \Lambda \), we then remove \( c_{\nu+1} e_{\lambda} \) in the \( n \)th 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

\[
f^{\nu+1} - 0(f^{\nu+1}, e_{i})e_{i}, \quad i < \lambda,
\]

as needed in order to have a total of \( n - 1 \) steps and in the \( n \)th 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,

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

and

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

Also

\[
M_{\nu+1} \geq M_{\nu} + (L_{\nu} - (\lfloor L_{\nu} / (2^\mu - 1) \rfloor))2^\mu + \lfloor L_{\nu}/(2^\mu - 1) \rfloor \]

\[
\geq M_{\nu} + L_{\nu} 2^\mu - [L_{\nu}/(2^\mu - 1)](2^\mu - 1) \geq M_{\nu} + L_{\nu}(2^\mu - 2).
\]

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)

\begin{equation}
M_\nu \geq \left(2(1-2^{-\mu+1})\right)^{\nu-\mu} 2^{-\mu}(2^\mu - 2) \geq C(\mu) 2^{\nu}, \quad \nu \geq \mu,
\end{equation}

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

\[ \|f_{M_\nu}\|^2 \geq h^{\nu-\mu}(1-2^{-\mu+1})^{\nu-\mu}\|f_{M_\mu}\|^2 \geq (1-2^{-\mu+1})^{2\nu-\mu+1} \]

\[ \geq C(\mu)2^{-C_1\nu^{2-\mu}} \geq C(\mu)M_\nu^{-C_22^{-\mu}}, \]

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 \( n^{-1/2} \leq \sqrt{22^{-\mu}} \), we conclude that the exponent of the power rate of decrease of \( \|f_{M_\mu}\|^2 \) is of order of \( n^{-1/2} \).

3. SIMULTANEOUS APPROXIMATION BY GREEDY ALGORITHMS

Given 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 \( F := (f^1, \ldots, f^N) \), \( f^k \in H \). It is a Hilbert space with the inner product

\[ \langle F_1, F_2 \rangle := \sum_{k=1}^{N} \langle f^k_1, f^k_2 \rangle. \]

Let \( \mathcal{D}_N \) be the collection

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

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 \((0, \ldots, 0, g, 0, \ldots, 0)\), where \( g \in \mathcal{D} \) is arbitrary and is in arbitrary position.) Also, all elements in \( \mathcal{D}_N \) are normalized.

We begin with \( F_0 := (f^1_0, \ldots, f^N_0) \) and a sequence \( 0 \leq t_k \leq 1 \) 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{equation}
G := G(F, g) := (\beta_1 g, \beta_2 g, \ldots, \beta_N g), \quad g \in \mathcal{D},
\end{equation}

\[ \beta_i := \langle f^i, g \rangle \left( \sum_{i=1}^{N} \|f^i, g\|^2 \right)^{-1/2}, \quad i = 1, \ldots, N. \]

For \( G \) of the form (3.1) the operation

\[ F_1 := F - \langle F, G \rangle G \]

means the same operation performed coordinate-wise

\[ f^i_1 := f^i - \langle f^i, g \rangle g, \quad i = 1, \ldots, N. \]

We note that

\begin{equation}
\|F\|_{\mathcal{D}_N} = \sup_{\alpha := (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}^N} \left| \sum_{i=1}^{N} \langle f^i, g_i \rangle \alpha_i \right| = \left( \sum_{i=1}^{N} \|f^i\|_{\mathcal{D}}^2 \right)^{1/2}.
\end{equation}
Lemma 3.1. For any $F \in H_N$ we have

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

Proof. On the one hand,

$$\sup_{g \in \mathcal{D}} |\langle F, G(F, g) \rangle| = \sup_{g \in \mathcal{D}} \left( \sum_{i=1}^{N} |\langle f_i, g \rangle|^2 \right)^{1/2}$$

$$\geq \max_i \sup_{g \in \mathcal{D}} |\langle f_i, g \rangle| = \max_i \|f_i\|_\mathcal{D},$$

and on the other, by (3.2),

$$\|F\|_{\mathcal{D}_N} = \left( \sum_{i=1}^{N} \|f_i\|_\mathcal{D}^2 \right)^{1/2} \leq N^{1/2} \max_i \|f_i\|_\mathcal{D}.$$ 

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

Given a weakness sequence $\tau = \{t_k\}_{k=1}^{\infty}$. The upper estimate for the VWGA, namely, for $\sum_{i=1}^{N} \|f_m^{i,s,\tau}\|^2$, can be obtained by Lemma 3.1 from the corresponding upper estimate for the WGA with the weakness sequence $\tau' := \{t_kN^{-1/2}\}_{k=1}^{\infty}$. 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 $f^1, \ldots, f^N, f^i \in A_1(\mathcal{D}), i = 1, \ldots, N$, we have

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

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

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

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 $f^i \in A_1(\mathcal{D}), i = 1, \ldots, N$ implies $(f^1, \ldots, f^N) \in A_1(\mathcal{D}_N, N)$. \qed

A similar proof yields
Theorem 3.2. Assume that for the weak sequence \( \tau = \{t_k\}_{k=1}^{\infty} \) there are a natural number \( n \) and a real number \( 0 < t \leq 1 \) such that
\[
n^{-1} \sum_{k=n+1}^{(l+1)n} t_k^2 \geq t^2, \quad l = 0, 1, 2, \ldots.
\]

Then for any \( 0 < \delta < 1 \) and all \( f_i \in A_1(D), i = 1, \ldots, N, \)
\[
\sum_{i=1}^{N} \| f_{i_{\ln}}^{s, \tau} \|^2 \leq N^2 (3n/\delta^2)^{\frac{r}{1+r}} (1 + lt^2)^{-\frac{r}{1+r}}
\]

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 \( x = \{x_k\}_{k=1}^{\infty}, x_k \geq 0, k = 1, 2, \ldots, \) for which there exists a sequence \( 0 = q_0 < q_1 < \ldots \) such that,
\[
\sum_{s=1}^{\infty} 2^s \Delta q_s < \infty,
\]
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 \( K \geq 1 \) and \( c \) we have \( 0 \leq y_k \leq cx_k, k \geq K, \) 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 \( \tau \), for each \( f \) and all Hilbert spaces \( H \) and dictionaries \( D \).

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 \( \tau \). 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 \( \tau' := \{t_kN^{-1/2}\} \notin \mathcal{V} \). 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 \( \tau \), for each vector of elements \( f^1, \ldots, f^N, N \) arbitrary, and all Hilbert spaces \( H \) and dictionaries \( D \).

Theorems 3.1 and 3.2 give estimates for the \( \ell_2^N \)-norm of the residual vector \( (\|f_m^1\|, \ldots, \|f_m^N\|) \). We wish to introduce greedy type algorithms that yield estimates for the \( \ell_\infty^N \)-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 $m$th iteration, $m = lN + i$, in the first step of the AWGA

1. (AWGA) We look for any $\varphi^n_m \in D$ satisfying

$$\left| \langle f^{i,a,\tau}_{m-1}, \varphi^n_m \rangle \right| \geq t \| f^{i,a,\tau}_{m-1} \|_D.$$  

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

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

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

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

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

**References**


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