Approximate Weak Greedy Algorithms

R. Gribonval and M. Nielsen

IMI, Department of Mathematics, University of South Carolina
South Carolina 29208, USA
E-mail: remi@math.sc.edu, nielsen@math.sc.edu

We present a generalization of V. Temlyakov's weak greedy algorithm, and give a sufficient condition for norm convergence of the algorithm for an arbitrary dictionary in a Hilbert space. We provide two counter-examples to show that the condition cannot be relaxed for general dictionaries. For a class of dictionaries with more structure, we give a more relaxed necessary and sufficient condition for convergence of the algorithm.

We also provide a detailed discussion of how a "real-world" implementation of the weak greedy algorithm, where one has to take into account floating point arithmetic and other types of finite precision errors, can be modeled by the new algorithm.

Key Words: greedy algorithm, weak greedy algorithm, best m-term approximation, nonlinear approximation, numerical algorithm, computational complexity, redundant systems

1. INTRODUCTION

Given a set $\mathcal{D}$ of unit vectors with dense span in a separable Hilbert space $\mathcal{H}$, one can consider the problem of finding the best approximation of a given element $f_0 \in \mathcal{H}$ by a linear combination of m elements from $\mathcal{D}$. For $\mathcal{D}$ an orthonormal basis of $\mathcal{H}$ it is very easy to construct the best m-term approximation of f_0, but whenever $\mathcal{D}$ is redundant the construction is much more difficult. A greedy algorithm (known as Matching Pursuit in signal processing [MZ93], or Projection Pursuit in statistics [FS81]) provides an m-term approximation of f_0, which might be sub-optimal, by constructing a sequence $f_m \in \mathcal{H}$, $m \geq 1$, such that at each step

$$f_m = f_{m-1} - \langle f_{m-1}, g_m \rangle g_m, \quad g_m \in \mathcal{D}$$

with

$$|\langle f_{m-1}, g_m\rangle| = \sup_{g\in\mathcal{D}} |\langle f_{m-1}, g\rangle|. \tag{1}$$

The m-term approximant of f_0, denoted by G_m, is then defined as $G_m = f_0 - f_m$. Hence,

$$G_m = \sum_{k=1}^{m} \langle f_{k-1}, g_k\rangle g_k.$$

It was proved in [Hub85] that $f_m \to 0$ weakly, and norm convergence was proved in [Jo87]. However, the optimization step in (1) is very costly from a computational point of view, and more recently the convergence of the greedy algorithm was proved under the weaker condition

$$|\langle f_{m-1}, g_m\rangle| \geq t_m \sup_{g\in\mathcal{D}} |\langle f_{m-1}, g\rangle|, \tag{2}$$

provided that $\{t_m\}_{m\geq 1} \subset [0,1]$ complies with some additional condition. The greedy algorithm with the relaxed selection criterion (2) is called the weak greedy algorithm (WGA). In [Jo87] norm convergence of the WGA was already proved under the assumption that $\exists \tilde{t} > 0,\ \forall m : t_m \geq \tilde{t}$. Temlyakov improved this result considerably in [Tem00], proving norm convergence whenever $\sum_m t_m/m = \infty$.

In the present paper we propose a modification/generalization of the WGA which we call the Approximate Weak Greedy Algorithm (AWGA). The setup is as follows: let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot,\cdot\rangle$ and associated norm $\|f\| = \langle f, f\rangle^{1/2}$. We call $\mathcal{D} \subset \mathcal{H}$ a dictionary if each $g \in \mathcal{D}$ has norm one and $\mathrm{span}\{g : g \in \mathcal{D}\}$ is a dense subset of $\mathcal{H}$. Note that what we call a dictionary is generally called a *complete* dictionary.

Approximate Weak Greedy Algorithm (AWGA). Let $\{t_m\}_{m=1}^{\infty} \subset [0,1]$, $\{\varepsilon_m\}_{m=1}^{\infty} \subset [-1,1]$, and a dictionary $\mathcal{D}$ be given. For $f \in \mathcal{H}$ we define a sequence $\{f_m\}_{m=0}^{\infty}$ inductively by letting $f_0 = f$, and for $m \geq 1$ assume that $\{f_0, f_1, \ldots f_{m-1}\}$ have already been defined. Then:

1. Take any $g_m \in \mathcal{D}$ satisfying (2);

$$|\langle f_{m-1}, g_m\rangle| \geq t_m \sup_{g\in\mathcal{D}} |\langle f_{m-1}, g\rangle|,$$

2. Define

$$f_m = f_{m-1} - (1+\varepsilon_m)\langle f_{m-1}, g_m\rangle g_m. \tag{3}$$

3. Put

$$G_m = f - f_m = \sum_{j=1}^{m} (1 + \varepsilon_j)\langle f_{j-1}, g_j\rangle g_j. \tag{4}$$

Remark 1. 1. The above procedure for the choice $\varepsilon_m = 0$, $m \geq 1$, is the weak greedy algorithm introduced by V. Temlyakov in [Tem00].

Remark 1. 2. From (3) we obtain

$$\|f_m\|^2 = \|f_{m-1}\|^2 - (1 - \varepsilon_m^2)|\langle f_{m-1}, g_m\rangle|^2, \tag{5}$$

which shows that the error $\|f - G_m\|$ is decreasing since $|\varepsilon_m| \leq 1$. Conversely, whenever $f_m = f_{m-1} - c_m g_m$ and $\|f_m\| \leq \|f_{m-1}\|$, one can show that $c_m = (1 + \varepsilon_m)\langle f_{m-1}, g_m\rangle$ for some $\varepsilon_m \in [-1, 1]$. Hence, if $\{G_m\}$ is a sequence of approximants, with decreasing error $\|f - G_m\|$, that can be written as the partial sums $\sum_{j=1}^{m} c_j g_j$, then G_m can be obtained through some AWGA by choosing the associated t_m's small enough.

We are interested in norm convergence of the AWGA procedure for a given dictionary $\mathcal{D}$, i.e. whether $G_m \to f$ for every $f \in \mathcal{H}$ (or equivalently, $f_m \to 0$). If the procedure converges for every $f \in \mathcal{H}$ then we say that AWGA($\mathcal{D}$) is convergent.

In the following Section we give sufficient conditions on $\{\varepsilon_m\}$ and $\{t_m\}$ for AWGA($\mathcal{D}$) to converge with any dictionary $\mathcal{D}$, and we demonstrate by providing two counter-examples that the conditions cannot be relaxed in general. In Section 3, we show that the conditions can be improved for a class of dictionaries with some structure. One example of such a dictionary is an orthonormal basis.

One reason for introducing the parameters ε_m in the procedure is to provide an “algorithm” that takes into account the fact that, for most implementations of the weak greedy algorithm, we will only be able to compute the inner products appearing in the procedure to within a given relative error. Moreover, one is forced to use floating point arithmetic for all the computations. In Section 4 we will discuss the feasibility of using the AWGA to model a “real-world” implementation of the weak greedy algorithm.

2. CONVERGENCE OF AWGA IN GENERAL DICTIONARIES

In this section we will present conditions that ensure convergence of the AWGA in a general dictionary $\mathcal{D}$. We will also present two counterexamples to illustrate that the conditions cannot be relaxed without requiring some form of special structure of the dictionaries.

2.1. Sufficient Conditions for Convergence of AWGA

The main general result on convergence of the AWGA is the following.

THEOREM 2.1. *Let $\mathcal{D}$ be any dictionary. Suppose that* $\{\varepsilon_m\}_{m=1}^{\infty} \subset [-1, 1-\delta]$ *for some* $0 < \delta < 2$ *and*

$$\sum_{m=1}^{\infty} \frac{t_m(1-\varepsilon_m^2)}{m} = \infty. \tag{6}$$

Then AWGA($\mathcal{D}$) *is convergent.*

We will give a proof of Theorem 2.1 based on the technique introduced by V. Temlyakov in [Tem00], but before we get to the proof itself we have to state two lemmas. The first Lemma concerns weak convergence of the AWGA, and will also be used in Section 3.

LEMMA 2.1. *Suppose* $\{\varepsilon_m\}_{m=1}^{\infty} \subset [-1, 1]$ *and*

$$\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) = \infty. \tag{7}$$

Then there exists a subsequence $\{f_{m_k}\}_{k=0}^{\infty}$ *which converges weakly to zero.*

Proof. From Remark 1.2, the sequence $\{\|f_m\|^2\}_{m=1}^{\infty}$ is decreasing and thus convergent. Hence,

$$\begin{aligned} \|f_0\|^2 - \lim_{m\to\infty} \|f_m\|^2 &= \sum_{k=0}^{\infty} \left(\|f_k\|^2 - \|f_{k+1}\|^2\right) \\ &= \sum_{k=0}^{\infty} (1-\varepsilon_{k+1}^2)|\langle f_k, g_{k+1}\rangle|^2 \qquad (8) \\ &\geq \sum_{k=0}^{\infty} t_{k+1}^2(1-\varepsilon_{k+1}^2) \sup_{g\in\mathcal{D}} |\langle f_k, g\rangle|^2. \end{aligned}$$

By assumption, $\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) = \infty$ so we must have

$$\liminf_{m\to\infty} \sup_{g\in\mathcal{D}} |\langle f_m, g\rangle|^2 = 0,$$

and since the span of $\mathcal{D}$ is dense in $\mathcal{H}$ the result follows. ∎

We will also need the following property of ℓ^2-sequences due to V. Temlyakov and S. V. Konyagin, see [Tem00].

LEMMA 2.2. *Suppose* $\{\tau_n\}_{n=1}^{\infty} \subset [0,\infty)$ *satisfies*

$$\sum_{n=1}^{\infty} \frac{\tau_n}{n} = \infty.$$

Then for any $\{\alpha_n\}_{n=1}^{\infty} \in \ell^2$,

$$\liminf_{n\to\infty} \frac{|\alpha_n|}{\tau_n} \sum_{j=1}^{n} |\alpha_j| = 0.$$

We can now give a proof of the Theorem.

Proof of Theorem 2.1. First we notice by the Cauchy-Schwarz inequality that

$$\sum_{m=1}^{\infty} \frac{t_m(1-\varepsilon_m^2)}{m} \leq \left(\sum_{m=1}^{\infty} \frac{(1-\varepsilon_m^2)}{m^2}\right)^{1/2} \left(\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2)\right)^{1/2},$$

so $\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) = \infty$. Using Lemma 2.1 we see that it suffices to prove that $\{f_m\}$ is a norm convergent sequence or, equivalently, that it is strongly Cauchy. Suppose $m > n$. We have

$$\|f_n - f_m\|^2 = \|f_n\|^2 - \|f_m\|^2 - 2\langle f_n - f_m, f_m\rangle.$$

Denote $a_j = |\langle f_{j-1}, g_j\rangle|$ and let $\Delta_{n,m} = |\langle f_n - f_m, f_m\rangle|$. Clearly,

$$f_m - f_n = \sum_{j=n+1}^{m} (1+\varepsilon_j)\langle f_{j-1}, g_j\rangle g_j,$$

so we obtain the estimate

$$\begin{aligned}
\Delta_{n,m} &\leq \sum_{j=n+1}^{m} (1+\varepsilon_j)|\langle f_{j-1}, g_j\rangle||\langle f_m, g_j\rangle| \\
&\leq \frac{a_{m+1}}{t_{m+1}} \sum_{j=1}^{m+1} (1+\varepsilon_j) a_j \\
&= \frac{(1+\varepsilon_{m+1})}{(1+\varepsilon_{m+1})} \frac{a_{m+1}}{t_{m+1}} \sum_{j=1}^{m+1} (1+\varepsilon_j) a_j. \qquad (9)
\end{aligned}$$

Eq. (8) shows that

$$\sum_{j=1}^{\infty}(1+\varepsilon_j)^2 a_j^2 \le \frac{2-\delta}{\delta}\sum_{j=1}^{\infty}(1-\varepsilon_j^2)a_j^2 < \infty,$$

so we can use Lemma 2.2 with $\alpha_n = (1+\varepsilon_n)a_n$ and $\tau_n = t_n(1+\varepsilon_n)$ to conclude that

$$\liminf_{n\to\infty} \max_{m>n} \Delta_{n,m} = 0.$$

This together with the fact that $\{\|f_m\|\}_{m=1}^{\infty}$ is a convergent sequence shows that $\{f_m\}_{m=1}^{\infty}$ is strongly Cauchy. ■

2.2. Counter-examples for the AWGA with Weaker Conditions

With a result like Theorem 2.1 it is natural to wonder whether the condition on $\{t_m\}$ and $\{\varepsilon_m\}$ can be relaxed. Next we will show that this is not the case: the conditions cannot be relaxed for general dictionaries. We consider two different cases. First, we show

Theorem 2.2. *Suppose that*

$$\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) < \infty. \tag{10}$$

Then there exists a dictionary $\mathcal{D}$ for which AWGA($\mathcal{D}$) is divergent.

Proof. First, let us suppose $\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) < 1$. Let f_0 be any unit vector in $\mathcal{H}$. Define the sequences $g_1, g_2, \ldots$ and $f_1, f_2, \ldots$ recursively by

$$g_m : \|g_m\| = 1 \text{ and } |\langle f_{m-1}, g_m\rangle| = t_m\|f_{m-1}\|$$

$$f_m = f_{m-1} - (1+\varepsilon_m)\langle f_{m-1}, g_m\rangle g_m.$$

Clearly, $\{f_m\}$ is a sequence of residuals for the AWGA with any dictionary containing the vectors $\{g_m\}_{m\ge1}$. However,

$$\|f_m\|^2 = \|f_{m-1}\|^2 - t_m^2(1-\varepsilon_m^2)\|f_{m-1}\|^2,$$

so

$$\lim_{m\to\infty}\|f_m\|^2 = \prod_{k=1}^{\infty}\frac{\|f_k\|^2}{\|f_{k-1}\|^2} = \prod_{k=1}^{\infty}\left(1-t_k^2(1-\varepsilon_k^2)\right) > 0.$$

Now we consider the general case. We choose $M > 1$ such that

$$\sum_{m=M+1}^{\infty} t_m^2(1-\varepsilon_m^2) < 1.$$

Select a set of $M+1$ orthonormal vectors $\{u_1, \ldots, u_M, u_{M+1}\}$ in $\mathcal{H}$, and let $P_{\mathcal{V}}$ denote the orthogonal projection onto $\mathcal{V} = \mathrm{span}\{u_1, \ldots, u_M\}^{\perp}$.

We define the sequences $\{g_m\}_{m\geq 1}$ of unit vectors inductively as follows: do M steps of the AWGA for $f_0 = u_1 + \cdots + u_M + u_{M+1}$, where at step $m \leq M$ we go in the direction $g_m := u_m$ with coefficient $1+\varepsilon_m$ (this will be justified later). We use f_m to denote the residual at step m. Then, suppose f_{m-1}, $m-1 \geq M$, has been defined. We consider the following two possibilities:

1. If $\max_{j\leq M}\{|\langle f_{m-1}, u_j\rangle|\} > t_m\|P_{\mathcal{V}} f_{m-1}\|$ then we go in the direction $g_m := u_j$ for which $\max_{j\leq M}\{|\langle f_{m-1}, u_j\rangle|\}$ is obtained.
2. Otherwise, take g_m to be a unit vector in $\mathcal{V}$ for which $|\langle f_{m-1}, g_m\rangle| = t_m\|P_{\mathcal{V}} f_{m-1}\|$.

In both cases we define

$$f_m = f_{m-1} - (1+\varepsilon_m)\langle f_{m-1}, g_m\rangle g_m.$$

Finally, we let $\{h_j\}_{j\geq 1}$ be any dictionary for $\mathcal{V}$ containing all the vectors $\{g_k | g_k \in \mathcal{V}\}$. It is easy to see that $\{f_m\}$ is a sequence of residuals for the AWGA for the dictionary $\mathcal{D} = \{u_1, \ldots, u_M\} \cup \{h_1, h_2, \ldots\}$. Also, note that the special structure of $\mathcal{D}$ ensures that the first M steps of the AWGA are justified. However, by construction,

$$\lim_{m\to\infty} \|P_{\mathcal{V}} f_m\|^2 \geq \prod_{k=M+1}^{\infty} \left(1 - t_k^2(1-\varepsilon_k^2)\right) > 0,$$

where the estimate corresponds to the "worst case" scenario,

$$\{g_k |\, k \geq M+1\} \subset \mathcal{V}.$$

■

Next we consider the case where $\sum_m t_m/m < \infty$.

THEOREM 2.3. *Suppose that* $\{t_m\}_{m\geq 1}$ *is a* decreasing *sequence with*

$$\sum_{m=1}^{\infty} \frac{t_m}{m} < \infty. \tag{11}$$

Then for each sequence $\{\varepsilon_m\}_{m\geq 1} \subset [-1,1]$ *there exists a dictionary* $\mathcal{D}$ *for which AWGA*$(\mathcal{D})$ *diverges.*

The proof will be based on a modification of the so-called Equalizer procedure introduced by V. Temlyakov and E. Livshitz in [LT00]. The setup is as follows. Let $\{e_i\}_{i=1}^{\infty}$ be an orthonormal basis for $\mathcal{H}$, and let the sequence $\{\eta_m\}_{m=1}^{\infty} \subset [-1,1]$ satisfying $\sum_{m=1}^{\infty}(1+\eta_m) = \infty$ and $\tau \in (0,1]$ be given.

The idea of the Equalizer is to start at a basis vector e_i and then produce a sequence of vectors $\{f_m\} \subset (\mathbb{R}_+ e_i, \mathbb{R}_+ e_j)$ approaching the diagonal $\mathbb{R}_+(e_i+e_j)$ without loosing too much energy on the way. The last vector before the procedure crosses the diagonal will be f_{N-1} and f_N denotes the first vector to have crossed (or landed on) the diagonal. The technical details are as follows;

Equalizer $E(e_i, e_j, \tau, \{\eta_k\})$. Put $f_0 = e_i$. Define the sequences $g_1, \dots, g_N$; $\phi_1, \dots, \phi_N$ and $f_1, \dots, f_N$ inductively by:

$$g_m = \cos\phi_m e_i - \sin\phi_m e_j; \quad f_m = f_{m-1} - (1+\eta_n)\langle f_{m-1}, g_m\rangle g_m,$$

with $\phi_m \in [0, \pi/2]$ such that

$$\langle f_{m-1}, g_m\rangle = \tau\|f_m\|, \quad m = 1, 2, \dots$$

Notice that

$$\|f_m\|^2 = \|f_{m-1}\|^2 - \tau^2(1-\eta_m^2)\|f_{m-1}\|^2,$$

and

$$f_m = \|f_m\|(\cos\alpha_m e_i + \sin\alpha_m e_j), \tag{12}$$

for some α_m. Using the assumption $\sum_{k=1}^{\infty}(1+\eta_k) = \infty$ we will now show that, for sufficiently small values of τ, there exists $N > 1$ such that $\alpha_{N-1} < \pi/4$ but $\alpha_N \geq \pi/4$. We need the following Lemma to estimate the angles between the residuals produced by the Equalizer.

Lemma 2.3. *Let* $\Psi_m(\tau)$ *be the angle between* f_{m-1} *and* f_m *constructed by the Equalizer* $E(e_i, e_j, \tau, \{\eta_m\}_{m\geq 1})$. *Then*

$$\Psi_m(\tau) = \arccos\frac{1-\tau^2(1+\eta_m)}{\sqrt{1-\tau^2(1-\eta_m^2)}},$$

so

$$\Psi_m(\tau) = (1+\eta_m)\tau + \mathcal{O}(\tau^3), \qquad \textit{as } \tau \to 0.$$

Proof. From the definition of f_m in terms of f_{m-1} it follows that

$$\langle f_m, f_{m-1}\rangle = \|f_{m-1}\|^2 - \tau^2(1+\eta_m)\|f_{m-1}\|^2,$$

which combined with $\|f_m\|^2 = \|f_{m-1}\|^2 - \tau^2(1-\eta_m^2)\|f_{m-1}\|^2$ gives us

$$\cos\Psi_m(\tau) = \frac{1-\tau^2(1+\eta_m)}{\sqrt{1-\tau^2(1-\eta_m^2)}}.$$

The remainder of the Lemma follows from a Taylor expansion of Ψ_m about $\tau = 0$. ∎

We have the following easy corollary.

Corollary 2.1. *There exist $\tau_0 > 0$ and constants $c > 0; C < \infty$ such that for $\tau < \tau_0$, and $\alpha_m(\tau)$ defined by (12) using $E(e_i, e_j, \tau, \{\eta_m\}_{m\geq 1})$, we have*

$$c\sum_{k=1}^{m}(1+\eta_k)\tau \leq \alpha_m(\tau) \leq C\sum_{k=1}^{m}(1+\eta_k)\tau.$$

Remark 2. 1. It follows easily from Lemma 2.3 that the constants c and C in Corollary 2.1 satisfy $c, C \to 1$ as $\tau_0 \to 0$, i.e. we can get c and C as close to one as we wish by choosing τ_0 sufficiently small.

Let us assume that $\tau < \tau_0$ with τ_0 given by Corollary 2.1. The size of N can now be estimated in terms of τ and the η_m's as follows; since $\alpha_{N-1}(\tau) < \pi/4$ we have

$$\sum_{k=1}^{N-1}(1+\eta_k) \leq \frac{\pi}{4c\tau},$$

It follows that $\sum_{k=1}^{N}(1+\eta_k) \leq \pi/(4c\tau) + 2 \leq \tilde{C}/\tau$, using the assumption $\tau \in (0,1]$. From this we can estimate the norm of f_N; we have

$$\begin{aligned}\|f_N\|^2 &= \prod_{k=1}^{N}\frac{\|f_k\|^2}{\|f_{k-1}\|^2}\\ &= \prod_{k=1}^{N}\left(1-\tau^2(1-\eta_k^2)\right)\\ &\geq e^{-\alpha\tau}, \qquad (13)\end{aligned}$$

where $\alpha = 4\tilde{C}\log 2$, and we have used, see e.g. [Wic94, p. 3],

$$\sum_{k=1}^{N} \log(1-\tau^2(1-\eta_k^2)) \geq -2(\log 2)\sum_{k=1}^{N} \tau^2(1-\eta_k)(1+\eta_k) \geq -4\tilde{C}(\log 2)\tau.$$

For technical reasons, the proof of Theorem 2.3 will be much easier if we can make sure that the vector f_N is actually on the diagonal. We can consider f_N as a function of τ, and we will now show that for some $\tilde{\tau}$ with $\tau/2 \leq \tilde{\tau} \leq \tau$ the vector $f_N(\tilde{\tau})$ is on the diagonal.

Corollary 2.2. *Let N be such that $f_{N-1}(\tau)$ from $E(e_i, e_j, \tau, \{\eta_m\}_{m\geq 1})$ has not crossed the diagonal but $f_N(\tau)$ has. Then there exists a $\tau_0 \in (0,1)$ such that whenever $\tau \leq \tau_0$, there is a $\tilde{\tau}$ with $\tau/2 \leq \tilde{\tau} \leq \tau$ for which $f_N(\tilde{\tau})$ is on the diagonal.*

Proof. Use Corollary 2.1 and Remark 2.1, we see that whenever τ is small enough and $\alpha_N(\tau) > \pi/4$ we have $\alpha_N(\tau/2) < \pi/4$, so using the obvious continuity of $\alpha_N(\eta)$ as a function of η, we see that there is a $\tilde{\tau} \in (\tau/2, \tau)$ for which $\alpha_N(\tilde{\tau}) = \pi/4$. ■

Remark 2. 2. It is clear that $E(e_i, e_j, \tilde{\tau}, \{\eta_n\})$ defined as in the above Lemma is an AWGA in the Hilbert space span$\{e_i, e_j\}$ with regard to the dictionary $\{e_i, g_1, \ldots, g_N\}$ with weakness parameter $\tilde{\tau} : \tau/2 \leq \tilde{\tau} \leq \tau$. From now on, for $2\tau \leq \tau_0$, we will use the notation $\tilde{E}(e_i, e_j, \tau, \{\eta_n\})$ to denote the result of modifying $E(e_i, e_j, 2\tau, \{\eta_n\})$ according to Lemma 2.2. Hence, the output of $\tilde{E}(e_i, e_j, \tau, \{\eta_n\})$ consists of two sequences $f_0, f_1, \ldots, f_N$ and $g_1, g_2, \ldots, g_N$, where $\{f_m\}$ is a finite sequence of residuals for the AWGA in span$\{e_i, e_j\}$ with respect to the dictionary $\mathcal{D}_{i,j}^{\tau} = \{e_i, g_1, \ldots, g_N\}$ and weakness parameter $\tilde{\tau} \geq \tau$ (in particular, it is an AWGA with respect to the weakness parameter τ). Moreover, the vector f_N is right on the diagonal $\mathbb{R}_+(e_i + e_j)$.

Remark 2. 3. We also notice that $\{f_m\}$ remains a finite sequence of residuals for the AWGA in $\mathcal{H}$ with weakness parameter τ in any larger dictionary $\mathcal{D} : \mathcal{D}_{i,j}^{\tau} \subset \mathcal{D}$ for which any given elements $g \in \mathcal{D}_{i,j}^{\tau}$ and $u \in \mathcal{D} - \mathcal{D}_{i,j}^{\tau}$ share at most one nonzero coordinate in the orthonormal basis $\{e_i\}$. Also, note that for the resulting f_N we have

$$\|f_N\|^2 \geq e^{-\alpha\tilde{\tau}} \geq e^{-2\alpha\tau}. \tag{14}$$

With the above results we can now prove Theorem 2.3 using the same technique as Livshitz and Temlyakov [LT00].

Proof of Theorem 2.3. First, we notice that we only have to consider the case where

$$\sum_{m=1}^{\infty}(1+\varepsilon_m)=\infty,$$

since otherwise

$$\sum_{m=1}^{\infty}(1-\varepsilon_m^2)\leq 2\sum_{m=1}^{\infty}(1+\varepsilon_m)<\infty$$

and we are in the case covered by Theorem 2.2.

Let τ_0 be given by Corollary 2.2, and suppose $\tau_0/2\geq t_1\geq t_2\geq\cdots\geq 0$. Notice that since

$$\sum_{k=1}^{\infty}\frac{t_k}{k}=S<\infty,$$

we have

$$\sum_{\ell=0}^{\infty}t_{2^\ell}\leq 2S<\infty.$$

We define the AWGA and the dictionary $\mathcal{D}$ as follows. The idea is to equalize iteratively. Start with $f_0=e_1\in\operatorname{span}\{e_1,e_2\}$ and apply $\tilde{E}(e_1,e_2,t_1,\{\varepsilon_k\}_{k=1}^{\infty})$. After $m_1=N_{1,2}\geq 1$ steps we get $g_1^0,\ldots,g_{N_{1,2}}^0$ and

$$f_{m_1}=c_1(e_1+e_2),$$

with property, see (14),

$$\|f_{m_1}\|^2\geq\|f_0\|^2e^{-2\alpha t_1}.$$

Now apply $\tilde{E}(e_1,e_3,t_2,\{\varepsilon_k\}_{k=N_{1,2}+1}^{\infty})$ on the component c_1e_1 of f_{m_1} in $\operatorname{span}\{e_1,e_3\}$, using $N_{1,3}$ steps, and apply $\tilde{E}(e_2,e_4,t_2,\{\varepsilon_k\}_{k=N_{1,2}+N_{1,3}+1}^{\infty})$ on the component c_1e_2 in $\operatorname{span}\{e_2,e_4\}$ using $N_{2,4}$ steps. From this we obtain $g_1^1,\ldots,g_{N_{1,3}+N_{2,4}}^1$, with $m_2-m_1:=N_{1,3}+N_{2,4}\geq 2$, and

$$f_{m_2}=c_2(e_1+\cdots+e_4)$$

satisfying

$$\|f_{m_2}\|^2\geq\|f_{m_1}\|^2e^{-2\alpha t_2}.$$

After s iterations we get

$$f_{m_s} = c_s(e_1 + \cdots e_{2^s}),$$

and for $i = 1, 2, \ldots, 2^s$ we apply $\tilde{E}(e_i, e_{i+2^s}, t_{2^s}, \{\varepsilon_k\}_{k=N^i+1}^{\infty})$, where N^i is the largest index of an ε_k used by the previous application of the equalizer, on the component $c_s e_i$ of the residual along e_i in $\operatorname{span}\{e_i, e_{i+2^s}\}$. We use $N_{i,i+2^s}$ steps, and obtain unit vectors $g_1^s, \ldots, g_{N_{1,2^s+1}+\cdots+N_{2^s,2^{s+1}}}^s$, with $m_{s+1} - m_s := \sum_i N_{i,i+2^s} \geq 2^s$, and

$$f_{m_{s+1}} = c_{s+1}(e_1 + \cdots + e_{2^{s+1}})$$

satisfying

$$\|f_{m_{s+1}}\|^2 \geq \|f_0\|^2 e^{-2\alpha t_1} e^{-2\alpha t_2} \cdots e^{-2\alpha t_{2^s}} \geq \prod_{k=0}^{\infty} e^{-2\alpha t_{2^k}} \geq e^{-4\alpha S}.$$

Using Remark 2.3 we see that $\{f_{m_s}\}$ is actually a subsequence of residuals for the AWGA, with respect to the dictionary

$$\mathcal{D} = \bigcup_{k=1}^{\infty} e_k \cup \bigcup_{s\geq 0;\ell} g_\ell^s,$$

which fails to converge to zero.

For the general case, we notice that $t_m \to 0$ as $m \to \infty$ so we can find $L > 0$ such that $t_{2^L} \leq \tau_0/2$. Then we take $f = e_1 + \cdot + e_{2^L}$ and at the $2^L - 1$ first steps of the AWGA we go in the directions specified by $\{e_1, \ldots, e_{2^L-1}\}$. Then we use the procedure described above with $f = e_{2^L}$ to complete the proof. ∎

3. AWGA IN DICTIONARIES WITH SOME STRUCTURE

So far we have considered the AWGA($\mathcal{D}$) with no assumptions on the structure of the dictionary $\mathcal{D}$. One would expect that Theorem 2.1 can be improved provided that we have some control on the structure of the dictionary. This is indeed the case, and in this section we will give an example of a large class of dictionaries where we can improve the result. The prime example from this class is the dictionary with the most structure of all, the orthonormal basis. Let us state and prove the general result and then consider a number of examples. The reader should compare the result below to the negative result of Theorem 2.2.

THEOREM 3.1. *Suppose*

$$\mathcal{H} = \bigoplus_{j=0}^{\infty} \mathcal{W}_j,$$

with $\dim \mathcal{W}_j < \infty$ *for* $j = 0, 1, \ldots$, *and let*

$$\mathcal{D} = \bigcup_{j=0}^{\infty} \mathcal{D}_j$$

be a dictionary for $\mathcal{H}$ *for which* $\mathcal{D}_j \subset \mathcal{W}_j$ *is an arbitrary dictionary for* $\mathcal{W}_j$*. Then AWGA*($\mathcal{D}$) *is convergent provided that*

$$\sum_{m=1}^{\infty} t_m^2 (1 - \varepsilon_m^2) = \infty. \tag{15}$$

Proof. Let $P_{\mathcal{W}_j}$ denote the orthogonal projection onto $\mathcal{W}_j$. For a given function $f \in \mathcal{H}$ consider the sequence $\{\|P_{\mathcal{W}_j} f_m\|\}_{j=0}^{\infty} \in \ell^2(\mathbb{N})$ for $m = 1, 2, \ldots$. It follows from the orthogonality of the subspaces $\mathcal{W}_j$ and the definition of the AWGA that for each j, $\|P_{\mathcal{W}_j} f_m\|$ is decreasing as $m \to \infty$. Thus, by the Dominated Convergence Theorem, the sequence has an $\ell^2(\mathbb{N})$-limit, which we denote by $\{\gamma_j\}_j$, and

$$\lim_{m \to \infty} \|f_m\|^2 = \sum_{j=0}^{\infty} \gamma_j^2.$$

It follows from Lemma 2.1 that there exists a subsequence f_{m_k} that converges weakly to zero. Hence, for each j, the projections $P_{\mathcal{W}_j} f_{m_k}$ converges weakly to zero in $\mathcal{W}_j$ as $k \to \infty$. By assumption, $\dim \mathcal{W}_j < \infty$ so the weak convergence in $\mathcal{W}_j$ is also strong convergence and $\gamma_j = \lim_{k \to \infty} \|P_{\mathcal{W}_j} f_{m_k}\| = 0$. Hence, $\lim_{m \to \infty} \|f_m\| = 0$. ■

Remark 3. 1. By applying exactly the same technique as in the proof of Theorem 2.2, one can show that the condition (15) is sharp within this class of structured dictionaries.

Let us consider some examples of dictionaries that fit into the setup of the theorem. First up is the orthonormal basis.

EXAMPLE 3.1. Let $\mathcal{D} = \{e_j\}_{j=0}^{\infty}$ be an orthonormal basis for $\mathcal{H}$. Define $\mathcal{W}_j = \operatorname{span}\{e_j\}$. Clearly, Theorem 3.1 applies, so AWGA($\mathcal{D}$) converges provided $\sum_{m=1}^{\infty} t_m^2 (1 - \varepsilon_m^2) = \infty$.

The second example comes from the library of Walsh wavelet packet bases for $L^2[0,1)$. We remind the reader that the Walsh functions $\{W_n\}_{n=0}^{\infty}$ are the basic wavelet packets associated with the Haar Multiresolution Analysis, see [Wic94, HW96]. The Walsh functions form an orthonormal basis for $L^2[0,1)$ and the library of Walsh wavelet packet bases are obtained as follows; for every dyadic partition $\mathcal{P}$ of the "frequency axis" $\{0,1,\ldots\}$ with sets of the form

$$I_{n,j} = \{n2^j, n2^j+1, \ldots, (n+1)2^j - 1\}, \qquad \text{with } j,n \geq 0,$$

we have an orthonormal basis for $L^2[0,1)$ consisting of the family of functions

$$\bigcup_{I_{n,j}\in\mathcal{P}} \{2^{j/2} W_n(2^j x - k) | k = 0,1,\ldots,2^j - 1\}.$$

It can also be shown that for each set $I_{n,j}$ we have

$$\operatorname{span}\{2^j W_n(2^j x - k)\}_{k=0}^{2^j-1} = \operatorname{span}\{W_\ell\}_{\ell\in I_{n,j}}.$$

With these facts about the Walsh wavelet packets we can give the following fairly general setup where the Theorem works.

Example 3.2. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two orthonormal Walsh wavelet packet bases for $L^2[0,1)$. Define the dictionary $\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2$. Notice that $\mathcal{D}$ is a tight frame for $L^2[0,1)$ with frame bound 2. Using the remarks above, and the dyadic structure of the sets $I_{n,j}$ (the intersection of $I_{n,j}$ and $I_{\tilde{n},\tilde{j}}$ is either empty or one set is contained in the other), we see that it is always possible to find finite dimensional spaces $\mathcal{W}_j$, each spanned by elements from $\mathcal{B}_1$ and $\mathcal{B}_2$, such that

$$L^2[0,1) = \bigoplus_{j=0}^{\infty} \mathcal{W}_j.$$

We can thus apply the Theorem 3.1 to conclude that AWGA$(\mathcal{B}_1 \cup \mathcal{B}_2)$ converges provided that $\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) = \infty$. ■

Remark 3. 2. The reader can easily verify that the previous example can be generalized to dictionaries $\mathcal{D}$ being a union of a finite number of orthonormal Walsh wavelet packet bases. It is also possible to replace the Walsh wavelet packets with periodized versions of smoother wavelet packet bases. The periodization ensures the finite dimensionality of the spaces $\mathcal{W}_j$

defined as above [H-NW96]. An example of such smooth wavelet packets are the periodized wavelet packets associated with the Daubechies filters.

4. IMPLEMENTATION OF GREEDY ALGORITHMS

In this section we will analyze computational and approximation issues that occur in "real-life" implementations of greedy algorithms. We will give a description of the major modifications which were proposed to ensure sparsity of the approximations and improve computational efficiency. While such modifications do not fit in the pure greedy algorithm nor the WGA models, we will see that they are well modeled by AWGAs.

4.1. Computational Issues

It is known that greedy algorithms are unstable [DMA97], hence their numerical implementation is likely to be sensitive to the finite precision of the computations with floating point arithmetic. In particular, there is a need for a careful study of their convergence properties under finite precision assumptions. Moreover, because of the large size of the dictionary, the actual computation of *all* the inner-products $\langle f_m, g\rangle, g \in \mathcal{D}$, at each step of the algorithm, is "intractable" in most numerical implementations of greedy algorithms. As a result, numerical implementations do not compute all these inner products : at most steps, only a much smaller number is computed. It is also common to compute approximate values of the inner products in order to accelerate the computations. Hence implementations of greedy algorithms can be modeled as approximate weak greedy algorithms, as we will see right now with more details.

First, $\mathcal{D}$ might be an uncountable dictionary such as the Gabor multiscale dictionary [MZ93] or the multiscale dictionary of chirps [Bul99]. In such a situation one defines by discretization a suitable finite sub-dictionary $\mathcal{D}_d \subset \mathcal{D}$ such that

$$\forall f, \sup_{g\in\mathcal{D}_d} |\langle f, g\rangle| \geq \rho \sup_{g\in\mathcal{D}} |\langle f, g\rangle| \tag{16}$$

for some $\rho > 0$.

The numerical complexity of M iterations of the greedy algorithm in this dictionary $\mathcal{D}_d$ is essentially the cost of the computation of

$$\{\langle f_m, g\rangle, g \in \mathcal{D}_d, 0 \leq m \leq M-1\}.$$

As the completeness of $\mathcal{D}_d$ makes it no smaller than a basis, the number of inner products to be computed at each step is at least $\sharp\mathcal{D}_d \geq N = \dim \mathcal{H}$. Hence, the cost $\mathcal{C}(\mathcal{D}_d)$ of their computation cannot be less than N.

Actually, for the Gabor multiscale dictionary [Tor91, MZ93, QC94] one gets $\mathcal{C}(\mathcal{D}_d^g) = \mathcal{O}(N \log^2 N)$, while with local cosines [CM91], wavepackets [CMQW92] and the chirp dictionary [MH95, Bul99, Gri00], the corresponding costs are respectively $\mathcal{C}(\mathcal{D}_d^{lc}) = \mathcal{O}(N \log^2 N)$, $\mathcal{C}(\mathcal{D}_d^{wp}) = \mathcal{O}(N \log N)$, and $\mathcal{C}(\mathcal{D}_d^c) = \mathcal{O}(N^2 \log N)$. Such values of $\mathcal{C}(\mathcal{D}_d)$ show that the decomposition of high dimensional signals with greedy algorithms requires a large computational effort.

4.1.1. Adaptive sub-dictionaries

A more drastic modification of the algorithm can be introduced, which has strong connections with the notion of weak greedy algorithm. At each step, instead of choosing g_m by a costly optimization in the large dictionary $\mathcal{D}_d$, one only proceeds to a search in a much smaller *adaptive* sub-dictionary $\mathcal{D}_m \subset \mathcal{D}_d$ so that only a small number $\sharp\mathcal{D}_m \ll \sharp\mathcal{D}_d$ of inner products have to be computed.

A practical realization of this principle [Ber95, BM96, Gri99] was suggested in time-frequency dictionaries. The principle is to define $\mathcal{D}_m$ as a set of time-frequency atoms where $|\langle f_{m-1}, g\rangle|^2$ is locally maximum either in the time direction or in the frequency direction. The heuristics is that the location of such local maxima should not change two much within a few consecutive steps. Hence it allows to compute the locations only at reasonably spaced steps $\{m_p\}_{p\in\mathbb{N}}$: when $m_p \leq m \leq m_{p+1} - 1$, the search is done in $\mathcal{D}_m \subseteq \mathcal{D}_{m_p}$ and one computes only the inner products $\left\{\langle f_{m-1}, g\rangle, g \in \mathcal{D}_{m_p}\right\}$. Ideally, one would like to ensure that

$$\sup_{g\in\mathcal{D}_m} |\langle f_{m-1}, g\rangle| \geq \rho' \sup_{g\in\mathcal{D}_d} |\langle f_{m-1}, g\rangle| \tag{17}$$

for some $\rho' > 0$, but it is actually quite hard to check this condition for such adaptive sub-dictionaries as the sub-dictionaries of local maxima. Instead, the following condition

$$\sup_{g\in\mathcal{D}_m} |\langle f_{m-1}, g\rangle| \geq \rho_m \sup_{g\in\mathcal{D}_d} |\langle f_{m-1}, g\rangle| \tag{18}$$

is always true for some sequence $\{\rho_m\}_{m\in\mathbb{N}}$, $\rho_m \in [0, 1]$. Temlyakov results [Tem00] show that $\sum_m \rho_m/m = \infty$ is sufficient to ensure the convergence of such an implementation of a greedy algorithm.

With sub-dictionaries of local maxima one can easily check inequality (18), with $\rho_{m_p} = 1$ and $\rho_m = 0, m \notin \{m_p, p \in \mathbb{N}\}$. Temlyakov's condition thus becomes $\sum_p 1/m_p = \infty$, showing that m_p can be quite sparse and still ensuring convergence (*e.g.* $m_p \asymp p \log p$). In particular it gives a much weaker condition than the uniform boudedness of $m_{p+1} - m_p$ required by Bergeaud and Mallat [Ber95, BM96]. More recently, Livschitz and Temlyakov [LT00] showed that in such a 0/1 setting the m_p's can be even sparser.

4.1.2. Fast update of inner products

The reduction of numerical complexity in a strategy with sub-dictionaries also relies on the use of fast approximate computations of inner products by *updating procedures.* It was noted by Mallat and Zhang [MZ93] that from one step to the next one, for any $g \in \mathcal{D}$,

$$\langle f_m, g \rangle = \langle f_{m-1}, g \rangle - \langle f_{m-1}, g_m \rangle \langle g_m, g \rangle . \tag{19}$$

At the time of the computation of $\langle f_m, g \rangle$, the two numbers $\langle f_{m-1}, g \rangle$ and $\langle f_{m-1}, g_m \rangle$ are known, so this update essentially requires the computation of

$$\langle g_m, g \rangle = \int_{-\infty}^{+\infty} g_m(t)\overline{g(t)}dt. \tag{20}$$

In practice, one uses discrete atoms $g[n], n = 0, 1, \ldots, N-1, g \in \mathcal{D}_d$, so (20) is replaced by $\sum_{n=0}^{N-1} g_m[n]\overline{g}[n]$, which costs $\mathcal{O}(N)$.

In highly structured dictionaries such as the Gaussian time-frequency dictionaries in $L^2(\mathbb{R})$ [MZ93, Bul99], it is possible to derive analytic expressions for the inner products (20). With the discrete-time versions of these dictionaries, such analytic formulas are replaced by summation formulas [MZ93, Bul99, Gri99] which truncation give an *approximation* to $\langle g_m, g \rangle$ within a given relative error η. The computational cost $\mathcal{O}(1)$ of these approximate inner products is independent of the dimension N of the analyzed signal.

4.1.3. AWGA Model

Fast implementations of greedy algorithms are thus using approximate coefficients $\widetilde{\langle f_{m-1}, g \rangle}$ both for the choice of a "best" atom at each step

$$|\widetilde{\langle f_{m-1}, g_m \rangle}| \geq \rho_m \sup_{g \in \mathcal{D}_d} |\widetilde{\langle f_{m-1}, g \rangle}| \tag{21}$$

and for the update of the residual

$$f_m = f_{m-1} - \widetilde{\langle f_{m-1}, g_m \rangle} g_m. \tag{22}$$

An additional source of approximation is of course the finite precision of the computations with floating-point arithmetic [Wic94]. If one only models roundoff errors, the approximation of the inner products can be written as

$$\widetilde{\langle f_{m-1}, g\rangle} = \langle f_{m-1}, g\rangle + \|f_{m-1}\|\, \eta_m(g) \tag{23}$$

where the order of magnitude η_m of $\eta_m(g)$ is essentially determined by the machine precision and the dimension N. If the fast update procedure (19) is used with an approximate value of $\langle g_m, g\rangle$, η_m additionnaly depends on the precision of the latter approximation, as well as on the number of times this procedure was used, which may depend on m (*e.g.* $m - m_p$ times in the strategy with adaptive sub-dictionaries).

We notice that the roundoff error behaves like a relative error

$$\widetilde{\langle f_{m-1}, g\rangle} = (1 + \varepsilon_m(g))\, \langle f_{m-1}, g\rangle\,,\ \varepsilon_m(g) \in [-1, 1], \tag{24}$$

whenever $\widetilde{\langle f_{m-1}, g\rangle}$ is replaced by a thresholded value. The threshold $C_m\, \|f_{m-1}\|\, \eta_m$ is specified using some $C_m > 2$.

Let us analyze a given step and show that such an implementation of a greedy algorithm is indeed an AWGA.

To prove the "approximate" part of this statement is actually quite easy using (22) and (24) : it only requires showing $|\varepsilon_m(g_m)| \leq 1$. Let's note that if the atom g is such that $|\widetilde{\langle f_{m-1}, g\rangle}| = 0$, then $\varepsilon_m(g) = -1$ and $|\langle f_{m-1}, g\rangle| \leq (C_m + 1)\, \|f_{m-1}\|\, \eta_m$. In the opposite case, $|\langle f_{m-1}, g\rangle| \geq (C_m - 1)\, \|f_{m-1}\|\, \eta_m$, hence $|\varepsilon_m(g)| \leq (C_m - 1)^{-1} < 1$.

More interesting is the proof of the "weak" part. It is clear from the previous discussion that

$$\begin{aligned}\sup_{g \in \mathcal{D}_d, |\widetilde{\langle f_{m-1}, g\rangle}| \neq 0} |\langle f_{m-1}, g\rangle| \ &\geq\ (C_m - 1)\, \|f_{m-1}\|\, \eta_m \\ &\geq\ \frac{C_m - 1}{C_m + 1} \sup_{g \in \mathcal{D}_d, |\widetilde{\langle f_{m-1}, g\rangle}| = 0} |\langle f_{m-1}, g\rangle|\,.\end{aligned}$$

As a result

$$\sup_{g \in \mathcal{D}_d, |\widetilde{\langle f_{m-1}, g\rangle}| \neq 0} |\langle f_{m-1}, g\rangle| \geq \frac{C_m - 1}{C_m + 1} \sup_{g \in \mathcal{D}_d} |\langle f_{m-1}, g\rangle|$$

From (24) and $|\varepsilon_m(g)| \leq 1/(C_m - 1)$ when $|\widetilde{\langle f_{m-1}, g\rangle}| \neq 0$, one gets

$$\sup_{g \in \mathcal{D}_d, |\widetilde{\langle f_{m-1}, g\rangle}| \neq 0} |\widetilde{\langle f_{m-1}, g\rangle}| \geq \left(1 - \frac{1}{C_m - 1}\right) \frac{C_m - 1}{C_m + 1} \sup_{g \in \mathcal{D}_d} |\langle f_{m-1}, g\rangle|$$

which, using (21), becomes

$$|\widetilde{\langle f_{m-1}, g_m\rangle}| \geq \rho_m \frac{C_m - 2}{C_m + 1} \sup_{g\in\mathcal{D}_d} |\langle f_{m-1}, g\rangle| .$$

It follows that

$$\begin{aligned}|\langle f_{m-1}, g_m\rangle| &= \frac{1}{1+\varepsilon_m(g_m)} |\widetilde{\langle f_{m-1}, g_m\rangle}| \\ &\geq \frac{\rho_m}{1+\varepsilon_m(g_m)} \frac{C_m - 2}{C_m + 1} \sup_{g\in\mathcal{D}_d} |\langle f_{m-1}, g\rangle| ,\end{aligned}$$

so the choice of g_m is weak with weakness parameter

$$1 \geq t_m = \rho_m \frac{C_m - 1}{C_m} \frac{C_m - 2}{C_m + 1} \geq 0. \tag{25}$$

4.2. Modified correlation functions

Another family of modified greedy algorithms that fit in the AWGA model is the class of greedy algorithms which use a correlation function $C(f_{m-1}, g)$ in place of the inner product $\langle f_{m-1}, g\rangle$ to select an atom g_m at each step. The correlation function is used as the coefficient of the selected atom, and the next residual is $f_m = f_{m-1} - C(f_{m-1}, g_m)g_m$. An example, among some other [MC97], is the so-called *high-resolution pursuit* [JCMW98, GDR+96] where the correlation function satisfies the following property : for every f and g, there is some $\alpha : 0 \leq \alpha \leq 1$ such that $C(f, g) = \alpha \langle f, g\rangle$. Hence this modified greedy algorithm can be expressed as an AWGA.

5. CONCLUSION

We have defined and studied the class of Approximate Weak Greedy Algorithms, which generalize Greedy Algorithms by relaxing the method to construct greedy approximants. Any iterative construction of m-term approximants with decreasing error can be obtained through an algorithm of this class. We have established some necessary and some sufficient conditions for convergence of the procedure. In a special class of structured dictionaries, we were able to determine a sharp necessary and sufficient condition for convergence of AWGA (Theorem 3.1). However, with general dictionaries we have to make stronger assumptions to ensure convergence (Theorem 2.1), and there is still a small gap between the sufficient condition

(Theorem 2.1) and the necessary conditions given by the counter-examples (Theorems 2.2 and 2.3).

The main difference is that in Theorem 2.1 we have to assume that $\{\varepsilon_m\}$ is bounded away from 1, while in the second counter-example we make no assumption at all about this sequence. Nevertheless, the first counter-example (Theorem 2.2) shows that ε_m are not allowed to converge too fast to 1.

Our belief is that further study of greedy algorithms will be best done by changing the point of view: it is now an important question to characterize the family of dictionaries for which the condition

$$\sum_{m=1}^{\infty} t_m^2(1-\varepsilon_m^2) = \infty$$

is sufficient for convergence of the AWGA. We conjecture that this class contains the family of tight frames.

REFERENCES

Ber95. F. Bergeaud. *Représentations adaptatives d'images numériques,* Matching Pursuit. PhD thesis, Ecole Centrale Paris, 1995.

BM96. F. Bergeaud and S. Mallat. Matching Pursuit: Adaptative representations of images and sounds. *Computational and Applied Mathematics.* 15(2), Birkhauser, Boston, October 1996.

Bul99. A. Bultan. A four-parameter atomic decomposition of chirplets. *IEEE Trans. Signal Process.*, 47(3):731–745, March 1999.

CM91. R. R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *Comptes-Rendus Acad. Sci. Paris(A)*, 312:259–261, 1991.

CMQW92. R. R. Coifman, Y. Meyer, S. R. Quake, and M. V. Wickerhauser. Signal processing and compression with wavelet packets. In Y. Meyer and S. Roques, editors, *Progress in Wavelet Analysis and Applications*, 1992.

DMA97. G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constr. Approx.*, 13:57–98, 1997.

FS81. J.H. Friedman and W. Stuetzle. Projection pursuit regression. *J. Amer. Statist. Assoc.* 76:817-823.

GDR+96. R. Gribonval, Ph. Depalle, X. Rodet, E. Bacry, and S. Mallat. Sound signals decomposition using a high resolution matching pursuit. In *Proc. Int. Computer Music Conf. (ICMC'96)*, pages 293–296, August 1996.

Gri99. R. Gribonval. *Approximations non-linéaires pour l'analyse de signaux sonores.* PhD thesis, Université Paris IX Dauphine, September 1999.

Gri00. R. Gribonval. Fast ridge pursuit with a multiscale dictionary of Gaussian chirps. *submitted to IEEE Trans. Signal Process.*, 2000. preprint at http://www.math.sc.edu/∼imip/00papers/0002.ps.

HW96. E. Hernández and G. Weiss. *A First Course on Wavelets.* CRC Press, 1996.

H-NW96. N. Hess-Nielsen and M. V. Wickerhauser. Wavelets and time-frequency analysis. *Proc. IEEE*, 84(4):523–540, 1996.

Hub85. P. J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435-475, 1985.

JCMW98. S. Jaggi, W.C. Carl, S. Mallat, and A.S. Willsky. High resolution pursuit for feature extraction. *J. Applied and Computational Harmonic Analysis*, 5(4):428–449, October 1998.

Jo87. L. K. Jones. On a conjecture of Huber concerning the convergence of PP-regression. *The Annals of Statistics*, 15:880-882, 1987.

LT00. E.D. Livschitz and V.N. Temlyakov. On convergence of weak greedy algorithms. Technical Report 0013, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2000. http://www.math.sc.edu/∼imip/00papers/0013.ps.

MH95. S. Mann and S. Haykin. The chirplet transform : Physical considerations. *IEEE Trans. Signal Process.*, 43(11):2745–2761, November 1995.

MZ93. S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, December 1993.

MC97. M.R. McClure and L. Carin. Matching pursuits with a wave-based dictionary. *IEEE Trans. Signal Process.*, 45(12):2912–2927, December 1997.

QC94. S. Qian and D. Chen. Signal representation using adaptive normalized Gaussian functions. *Signal Process.*, 36(1):1–11, 1994.

Tem00. V.N. Temlyakov. Weak greedy algorithms. *Advances in Computational Mathematics*, 12(2,3):213–227, 2000.

Tor91. B. Torrésani. Wavelets associated with representations of the affine Weyl-Heisenberg group. *J. Math. Phys.*, 32:1273–1279, May 1991.

Wic94. M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A. K. Peters, 1994.