Journal of Approximation Theory **111**, 128–138 (2001)
doi:10.1006/jath.2001.3566, available online at http://www.idealibrary.com on IDEAL®

A Counter-Example to the General Convergence of Partially Greedy Algorithms[1]

R. Gribonval

IMI, Department of Mathematics, University of South Carolina, Columbia, South Carolina 29208, U.S.A.
E-mail: remi@math.sc.edu

Communicated by Allan Pinkus

Received March 27, 2000; accepted in revised form January 5, 2001; published online May 17, 2001

In a separable Hilbert space $\mathscr{H}$, greedy algorithms iteratively define m-term approximants to a given vector from a complete redundant dictionary $\mathscr{D}$. With very large dictionaries, the pure greedy algorithm cannot be implemented and must be replaced with a weak greedy algorithm. In numerical applications, *partially greedy algorithms* have been introduced to reduce the numerical complexity. A conjecture about their convergence arises naturally from the observation of numerical experiments. We introduce, study and disprove this conjecture. © 2001 Academic Press

Key Words: nonlinear approximation; greedy algorithms; redundant dictionary; convergence.

1. INTRODUCTION

Given a complete dictionary $\mathscr{D}$ of unit vectors (or *atoms*) in a separable Hilbert space $\mathscr{H}$, one can consider the approximations of a vector R_1 by linear combinations of atoms taken from $\mathscr{D}$. When the dictionary is indeed an orthonormal basis, the best m-term approximation to R_1 can actually be constructed. But whenever the dictionary is redundant, there is no unique linear decomposition of R_1, and the best m-term approximation may be difficult to build. A greedy algorithm (known as Matching Pursuit in signal processing [8], or Projection Pursuit in statistics [6]) provides such an m-term approximation by constructing a sequence $R_m \in \mathscr{H}$, $m \geqslant 1$ such that at each step

$$R_m = \langle R_m, g_m \rangle g_m + R_{m+1}, \; g_m \in \mathscr{D} \tag{1}$$

[1] This research was supported by the National Science Foundation (NSF) Grant No. DMS-9872890.

0021-9045/01 $35.00

with

$$|\langle R_m, g_m \rangle| = \sup_{g \in \mathcal{D}} |\langle R_m, g \rangle|. \tag{2}$$

The resulting m-term expansion of R_1 is

$$R_1 = \langle R_1, g_1 \rangle\, g_1 + \langle R_2, g_2 \rangle\, g_2 + \cdots + \langle R_m, g_m \rangle\, g_m + R_{m+1}. \tag{3}$$

Weak (resp. strong) convergence of R_m to zero was proved by [6] (resp. [7]) for the case of $\mathcal{D} = \{\text{unit ridge functions in } L^2(P)\}$ with P a probability measure on a Euclidean space. When the dictionary $\mathcal{D}$ is very large, the choice (2) of the best atom from $\mathcal{D}$ may be so computationnaly costly that a sub-optimal choice has to be considered. In [7], the convergence of greedy algorithms (for $\mathcal{D} = \{\text{unit ridge functions}\}$) was also proved under the weaker sufficient condition

$$|\langle R_m, g_m \rangle| \geqslant t_m \sup_{g \in \mathcal{D}} |\langle R_m, g \rangle| \tag{4}$$

(see Eq. (2) in [7]) provided that t_m is greater or equal to some positive t. The results of [7] were extended to the case of general dictionaries in [9]. In [10] the stronger result of convergence whenever $\sum_m t_m/m = \infty$ is proved, and the remark was made that divergence may occur when $\sum_m t_m^2 < \infty$. An open question consists in filling the gap between these two conditions on t_m.

In [1, 2], a Matching Pursuit with sub-dictionaries of local maxima of $\mathcal{D}$ was defined for the approximation of images. The same technique was developped in [4, 5] in order to accelerate the analysis of sound signals. At the iteration m_p, a sub-dictionary $\mathcal{D}_p$ of local maxima is built that contains the best atom for this iteration. Between iterations m_p and $m_{p+1} - 1$, the best atom from $\mathcal{D}_p$ is selected. The convergence of such an algorithm was proved [1] under the restrictive assumption that $m_{p+1} - m_p$ is bounded. The proof used the convergence of weak gready algorithms [7, 9]. The results from [10] show that a weaker sufficient condition for the convergence of this algorithm is $\sum_p \frac{1}{m_p} = \infty$. For example if $m_p = p \log p$, there is convergence but $m_{p+1} - m_p$ is not bounded.

With a multiscale time-frequency dictionary of chirps [5], the author suggested a much weaker two-step choice of the "best" atom of $\mathcal{D}$. At first we only consider a complete sub-dictionary $\mathcal{D}^*$ of reasonable size, and select the best atom $g_m^* \in \mathcal{D}^*$ by the greedy procedure (2). Then we try to improve this choice by choosing a "locally optimal" atom $g_m \in \mathcal{D}$ with an interpolation method in a "neighborhood" of g_m^*. The final choice thus complies with

$$|\langle R_m, g_m\rangle| \geqslant |\langle R_m, g_m^*\rangle| = \sup_{g \in \mathscr{D}^*} |\langle R_m, g\rangle|. \tag{5}$$

Such a stepwise choice $g_m \in \mathscr{D}$ is generally much weaker than (4), if no additional assumption is made about $\mathscr{D}$ and $\mathscr{D}^*$. We call the corresponding class of algorithms "partially greedy algorithms".

A partially greedy algorithm is "stepwise better" than a pure greedy algorithm in $\mathscr{D}^*$. A natural question is whether such a choice of $g_m \in \mathscr{D}$ will improve the speed of convergence, compared to a pure greedy algorithm in $\mathscr{D}^*$. That is to say, if one is doing a pure greedy algorithm in $\mathscr{D}^*$ (which is known to converge), is it a good idea to get at each step a "better atom" in $\mathscr{D}$? Does it "take advantage" of the extra redundancy given by $\mathscr{D}$? The intuition tells us that it should converge, may be with an improvement of the speed of convergence. In finite dimension, the convergence of partially greedy algorithms follows from standard compactness arguments.

The goal of this article is to show that the intuition is false in infinite dimension. Not only the partially greedy algorithm can converge more slowly than a pure greedy algorithm in $\mathscr{D}^*$ (this is already known from a counter-example of DeVore and Temlyakov [3], where $\mathscr{D}^*$ is an orthonormal basis and $\mathscr{D} = \mathscr{D}^* \cup \{g_0\}$ with g_0 a "bad" vector), but it may not converge at all. We show this by building a counter-example.

In Section 2, we give a precise statement of the natural conjecture, and define the notion of (pure and partially) greedy sequence. In Section 3 we build a partially greedy sequence that is our counter-example. In Section 4 we make some comments about the implications of our result.

2. GREEDY ALGORITHMS AND SEQUENCES

The natural conjecture about partially greedy algorithms is the following

Conjecture 2.1. Let $\mathscr{D}^* \subset \mathscr{D}$ two complete dictionaries in a Hilbert space $\mathscr{H}$. Let $\{R_m\}_{m \geqslant 1}$ be such that, for all m, (1) holds with some $g_m \in \mathscr{D}$ chosen such that (5) is true. Then $\|R_m\|_2 \to 0$.

If this conjecture were true, then it should be true when $\mathscr{D}^* = \mathscr{B}$ is an orthonormal basis of $\mathscr{H}$ and $\mathscr{D} \supset \mathscr{B}$ is any dictionary containing this basis. It would imply the following result.

Conjecture 2.2. Let $\mathcal{B} = \{e_n, n \in \mathbb{N}\}$ an orthonormal basis of $\mathcal{H}$, and define for any $x \in \mathcal{H}$: $\|x\|_\infty := \sup_{n \in \mathbb{N}} |\langle x, e_n \rangle|$. Let $\{R_m\}_{m \geqslant 1}$ a sequence such that for all $m \geqslant m_0$,

$$\langle R_m, R_{m+1} \rangle = \|R_{m+1}\|_2^2 \tag{6}$$

and

$$\|R_m\|_2^2 - \|R_{m+1}\|_2^2 \geqslant \sup_{n \in \mathbb{N}} |\langle R_m, e_n \rangle|^2 = \|R_m\|_\infty^2 . \tag{7}$$

Then $\|R_m\|_2 \to 0$.

Proof. Let R_m comply with (6) and (7). We shall prove that (1) and (5) hold for $m \geqslant m_0$, for some dictionary $\mathcal{D}$ that we will specify later on. As the convergence is an asymptotic property, it is clear that we can replace, in Conjecture 2.1, "(1) and (5) hold for all m" by "(1) and (5) hold for all $m \geqslant m_0$". Thus, if Conjecture 2.1 is true, we will get $\|R_m\|_2 \to 0$.

Let us define

$$g_m := (R_m - R_{m+1}) / \|R_m - R_{m+1}\|_2 \tag{8}$$

and

$$\mathcal{D} := \mathcal{B} \cup \{g_m, m \in \mathbb{N}\}. \tag{9}$$

From the definition of g_m, $R_m = \delta_m g_m + R_{m+1}$ for some δ_m. From (6) we get $\langle R_{m+1}, g_m \rangle = 0$, thus $\delta_m = \langle R_m, g_m \rangle$, which shows (1).

Now (7) gives $|\langle R_m, g_m \rangle|^2 = \|R_m\|_2^2 - \|R_{m+1}\|_2^2 \geqslant \sup_{e_k \in \mathcal{B}} |\langle R_m, e_k \rangle|^2$ which shows (5). ∎

This enables us to only deal with properties of *sequences* in $\mathcal{H}$: we can forget about the algorithmic nature of the iterative decomposition. We will call *greedy sequence* any sequence R_m in $\mathcal{H}$ complying with (6). A greedy sequence that additionally complies with (7) will be called a *partially* greedy sequence, by opposition to a *pure* greedy sequence which is supposed to satisfy the stronger condition (which is equivalent to (2))

$$\|R_m\|_2^2 - \|R_{m+1}\|_2^2 \geqslant \sup_{g \in \mathcal{D}} |\langle R_m, g \rangle|^2.$$

We shall prove that Conjecture 2.1 is false by building a counter-example to Conjecture 2.2, that is to say a partially greedy sequence R_m such that $\|R_m\|_2$ is bounded from below by some $c > 0$.

3. A COUNTER-EXAMPLE

For convenience, our counter-example $\{R_m\}_{m \geqslant 1}$ will be defined through its normalization $S_m := R_m / \|R_m\|_\infty$ in $\|.\|_\infty$. Let us first show that, for any $\lambda > 0$ and any sequence $\{S_m\}_{m \geqslant 1}$ such that $\|S_m\|_\infty = 1$ there is a unique sequence

$$R_m = \|R_m\|_\infty \, S_m.$$

such that (6) holds and $\|R_1\|_\infty = \lambda$. It is done by proving that the sequence of norms $\{\|R_m\|_\infty\}_{m \geqslant 1}$ is defined by λ and the sequence $\{S_m\}_{m \geqslant 1}$, which comes from the fact that (6) is equivalent to

$$\|R_{m+1}\|_\infty / \|R_m\|_\infty = \langle S_{m+1}, S_m \rangle / \langle S_{m+1}, S_{m+1} \rangle. \tag{10}$$

Condition (7) on $\{R_m\}_{m \geqslant 1}$ then becomes a condition on $\{S_m\}_{m \geqslant 1}$

$$\langle S_m, S_m \rangle - \frac{\langle S_{m+1}, S_m \rangle^2}{\langle S_{m+1}, S_{m+1} \rangle} \geqslant 1. \tag{11}$$

Let us remark that for any greedy sequence, $\{\|R_m\|_2\}_{m \geqslant 1}$ is a decreasing positive sequence, hence it has a limit. For a partially greedy sequence, the fact that (7) holds for $m \geqslant m_0$ thus implies $\|R_m\|_\infty \to 0$. Suppose now that $\{R_m\}_{m \geqslant 1}$ is a counter-example to Conjecture 2.2, which is equivalent to $\lim \|R_m\|_2 > 0$: $\{R_m\}_{m \geqslant 1}$ corresponds to some "bad" sequence $\{S_m\}_{m \geqslant 1}$ complying with (11) such that

$$\|R_m\|_2 = \|R_m\|_\infty \|S_m\|_2 \geqslant c, \tag{12}$$

where $\|R_m\|_\infty$ is defined through (10) and $c > 0$ is some constant. Hence we must have $\|S_m\|_2 \to \infty$.

We are going to specify some badly behaved sequence $\{S_m\}_{m \geqslant 1}$ of vectors, such that $\|S_m\|_\infty = 1$, $\|S_m\|_2 \to \infty$, and (11)–(12) hold. Our setting is now $\mathscr{H} = \ell^2(\mathbb{N})$ and we define, for $0 < \varepsilon < 1$

$$S(\varepsilon) := \{(1-\varepsilon)^n\}_{n \geqslant 0}. \tag{13}$$

One can easily check that $\|S(\varepsilon)\|_\infty = 1$ and for all $0 < \varepsilon, \eta < 1$,

$$\langle S(\varepsilon), S(\eta) \rangle = \sum_{n=0}^{\infty} ((1-\varepsilon)(1-\eta))^n = (\varepsilon + \eta - \varepsilon\eta)^{-1} = \frac{\frac{1}{\varepsilon}\frac{1}{\eta}}{\frac{1}{\varepsilon} + \frac{1}{\eta} - 1}. \tag{14}$$

In particular

$$\|S(\varepsilon)\|_2^2 = (2\varepsilon - \varepsilon^2)^{-1}. \tag{15}$$

Using this family of "bad vectors" we can now build the counter-example we have announced.

PROPOSITION 3.1 *Let $\mathcal{H} = \ell^2(\mathbb{N})$ and $\mathcal{B}$ its canonical basis. Let $\alpha > 2$ and $\varepsilon_m = m^{-\alpha}$. Let $S_m = S(\varepsilon_m)$. Let $R_m = \|R_m\|_\infty S_m$ where $\{\|R_m\|_\infty\}_{m \geqslant 1}$ is inductively defined by* (10), *with an arbitrary initial value $\|R_1\|_\infty > 0$. Then $\{R_m\}_{m \geqslant 1}$ is a counter example, that is to say : there exists $m_0 \in \mathbb{N}$ such that for all $m \geqslant m_0$, relations* (6) *and* (7) *hold, and*

$$\exists c > 0, \forall m, \|R_m\|_2 \geqslant c. \tag{16}$$

Notations. We use the symbol $a_m \asymp b_m$ to denote the existence of constants c and C such that $c a_m \leqslant b_m \leqslant C a_m$ for m big enough. The notation $a_m \sim b_m$ means that $a_m / b_m \to 1$. Finally $a_m = \mathcal{O}(b_m)$ is written when a_m / b_m is bounded.

Proof. Using (14) we get $\langle S_m, S_m \rangle = m^{2\alpha}/(2m^\alpha - 1)$, $\langle S_{m+1}, S_{m+1} \rangle = (m+1)^{2\alpha}/(2(m+1)^\alpha - 1)$ and $\langle S_{m+1}, S_m \rangle = m^\alpha (m+1)^\alpha/(m^\alpha + (m+1)^\alpha - 1)$ from which we derive

$$\begin{aligned}
\langle S_m, S_m \rangle &- \frac{\langle S_{m+1}, S_m \rangle^2}{\langle S_{m+1}, S_{m+1} \rangle} \\
&= \frac{m^{2\alpha}}{2m^\alpha - 1} - \frac{2(m+1)^\alpha - 1}{(m+1)^{2\alpha}} \frac{m^{2\alpha}(m+1)^{2\alpha}}{(m^\alpha + (m+1)^\alpha - 1)^2} \\
&= \frac{m^{2\alpha}\left[(m^\alpha + (m+1)^\alpha - 1)^2 - (2(m+1)^\alpha - 1)(2m^\alpha - 1)\right]}{(2m^\alpha - 1)(m^\alpha + (m+1)^\alpha - 1)^2} \\
&= \frac{m^{2\alpha}((m+1)^\alpha - m^\alpha)^2}{(2m^\alpha - 1)(m^\alpha + (m+1)^\alpha - 1)^2} \\
&\sim \frac{m^\alpha}{8}((1 + 1/m)^\alpha - 1)^2 \asymp m^{\alpha - 2}
\end{aligned} \tag{17}$$

which proves that (11) is true for m greater than or equal to m_0 (for some m_0).

From (15) we know that $\|S_m\|_2^2 \sim 1/(2\varepsilon_m) \asymp m^\alpha$. It is thus sufficient to prove that $\|R_m\|_\infty \asymp m^{-\alpha/2}$ to obtain (12) and reach our conclusion. To get this result we use the sequence $v_m := \log(\|R_{m+1}\|_\infty / \|R_m\|_\infty)$, $m \geqslant 1$ and show that

$$v_m + \frac{\alpha}{2} \log \frac{m+1}{m} = \mathcal{O}(m^{-2}). \tag{18}$$

Indeed, from (10) we have $v_m = \log\langle S_{m+1}, S_m\rangle - \log\langle S_{m+1}, S_{m+1}\rangle$, thus using (14) we get

$$\begin{aligned} v_m + \log \frac{\varepsilon_m + \varepsilon_{m+1}}{2\varepsilon_{m+1}} &= \log \frac{1}{\varepsilon_m + \varepsilon_{m+1} - \varepsilon_m \varepsilon_{m+1}} + \log 2\varepsilon_{m+1}\left(1 - \frac{\varepsilon_{m+1}}{2}\right) \\ &\quad + \log \frac{\varepsilon_m + \varepsilon_{m+1}}{2\varepsilon_{m+1}} \\ &= \log\left(1 - \frac{\varepsilon_{m+1}}{2}\right) - \log\left(1 - \frac{\varepsilon_m \varepsilon_{m+1}}{\varepsilon_m + \varepsilon_{m+1}}\right) = \mathcal{O}(m^{-\alpha}). \end{aligned} \tag{19}$$

Moreover, one easily gets

$$\log \frac{\varepsilon_m + \varepsilon_{m+1}}{2\varepsilon_{m+1}} = \log\left[1 + \frac{1}{2}((1+1/m)^\alpha - 1)\right] = \frac{\alpha}{2m} + \mathcal{O}(m^{-2}). \tag{20}$$

As $1/m = \log(m+1) - \log m + \mathcal{O}(m^{-2})$, and $\alpha > 2$, (19) and (20) lead to (18). To finish with, (18) gives by a telescoping sum that

$$\log\left(\frac{\|R_m\|_\infty}{\|R_1\|_\infty}\right) + \frac{\alpha}{2} \log m = \sum_{k=1}^{m-1}\left(v_k + \frac{\alpha}{2} \log \frac{k+1}{k}\right) \tag{21}$$

has a limit $K \in \mathbb{R}$, which proves that

$$\|R_m\|_\infty \, m^{\alpha/2} \to C > 0. \tag{22}$$

4. COMMENTS AND CONSEQUENCES

This counter-example gives some additional information on the properties of partially greedy sequences. Let us state some (known) results about greedy sequences

Lemma 4.1 *Let* $\{R_m\}_{m \geqslant 1}$ *be any* partially *greedy sequence. If* $\sum \|R_m - R_{m+1}\|_2 < \infty$ *then* $\|R_m\|_2 \to 0$.

Proof. As $\mathscr{H}$ is complete, $\sum \|R_m - R_{m+1}\|_2 < \infty$ implies the (strong) convergence of $\{R_m\}_{m\geqslant 1}$ to some $R_\infty \in \mathscr{H}$. But we have seen that $\|R_m\|_\infty \to 0$ because of (7). Fatou's lemma thus gives $R_\infty = 0$.

The counter-example we have built must thus comply with $\sum \|R_m - R_{m+1}\|_2 = \infty$. On the other hand, it is easy to see that

LEMMA 4.2 *For every greedy sequence* $\{R_m\}_{m\geqslant 1}$,

$$\sum \|R_m - R_{m+1}\|_2^2 < \infty.$$

Proof. The stepwise orthogonal decomposition (1) implies a stepwise energy conservation $\|R_m\|_2^2 = \|R_m - R_{m+1}\|_2^2 + \|R_{m+1}\|_2^2$, which gives $\sum \|R_m - R_{m+1}\|_2^2 \leqslant \|R_1\|^2$ by a telescoping sequence argument. ∎

We currently know that, for any counter-example to Conjecture 2.2:

$$\sum \|R_m - R_{m+1}\|_2 = \infty \tag{23}$$

$$\sum \|R_m - R_{m+1}\|_2^2 < \infty. \tag{24}$$

What about the convergence of $\sum \|R_m - R_{m+1}\|_2^p$, $1 < p < 2$? The particular counter-example we built in Proposition 3.1 does show that such a convergence is not sufficient to ensure the strong convergence of R_m to zero.

LEMMA 4.3. *The counter-example* $\{R_m\}_{m\geqslant 1}$ *to Conjecture* 2.2 *built in Proposition* 3.1 *complies, for all* $p > 1$, *with*

$$\sum \|R_m - R_{m+1}\|_2^p < \infty.$$

Proof. We show that $\|R_m - R_{m+1}\|_2 \asymp m^{-1}$, which gives the result. Using the asymptotic rates (17) and (22) we get

$$\begin{aligned}\|R_m - R_{m+1}\|_2^2 &= \|R_m\|_2^2 - \|R_{m+1}\|_2^2\\ &= \|R_m\|_\infty^2 \left(\langle S_m, S_m\rangle - \langle S_{m+1}, S_m\rangle^2 / \langle S_{m+1}, S_{m+1}\rangle\right)\\ &\asymp m^{-\alpha} m^{\alpha-2} = m^{-2}. \quad ∎\end{aligned}$$

We know from [10] that, for any counter-example to Conjecture 2.2, $\sum t_m/m < \infty$, where we use (8) and (9) to define

$$t_m := \frac{|\langle R_m, g_m\rangle|}{\sup_{g\in\mathscr{D}} |\langle R_m, g\rangle|} = \frac{\|R_m - R_{m+1}\|_2}{\sup_{g\in\mathscr{D}} |\langle R_m, g\rangle|}. \tag{25}$$

We would like to know whether we can find any counter-example to Conjecture 2.2 such that $\sum t_m^2 = \infty$. It would show that $\sum t_m^2 = \infty$ is not a sufficient condition to ensure the convergence of weak greedy algorithms.

Let us start by studying the asymptotic behavior of t_m for any partially greedy sequence $\{R_m\}_{m \geqslant 1}$. It is actually easy to see that $t_m = \|R_m - R_{m+1}\|_2 / \sup_p |\langle R_m, g_p \rangle|$ where $g_p = (R_p - R_{p+1}) / \|R_p - R_{p+1}\|_2$. For each p, one can write

$$
\begin{aligned}
&|\langle R_m, g_p \rangle|^2 \\
&\quad = \frac{\left[\|R_m\|_\infty \|R_p\|_\infty \left(\langle S_m, S_p \rangle - \dfrac{\langle S_{p+1}, S_p \rangle}{\langle S_{p+1}, S_{p+1} \rangle} \langle S_m, S_{p+1} \rangle \right) \right]^2}{\|R_p - R_{p+1}\|_2^2} \\
&\quad = \|R_m\|_\infty{}^2 \, \frac{\left(\langle S_m, S_p \rangle - \dfrac{\langle S_{p+1}, S_p \rangle}{\langle S_{p+1}, S_{p+1} \rangle} \langle S_m, S_{p+1} \rangle \right)^2}{\langle S_p, S_p \rangle - \dfrac{\langle S_{p+1}, S_p \rangle}{\langle S_{p+1}, S_{p+1} \rangle} \langle S_p, S_{p+1} \rangle}
\end{aligned}
$$

so that

$$
t_m^2 = \frac{\|R_m - R_{m+1}\|_2^2}{\|R_m\|_\infty^2 \, K_m} \tag{26}
$$

with $K_m = \sup_p K_{m,p}$ and

$$
K_{m,p} = \frac{\left(\langle S_m, S_p \rangle - \dfrac{\langle S_{p+1}, S_p \rangle}{\langle S_{p+1}, S_{p+1} \rangle} \langle S_m, S_{p+1} \rangle \right)^2}{\langle S_p, S_p \rangle - \dfrac{\langle S_{p+1}, S_p \rangle}{\langle S_{p+1}, S_{p+1} \rangle} \langle S_p, S_{p+1} \rangle}. \tag{27}
$$

Let us now restrict the study to the very specific case of sequences $\{R_m\}_{m \geqslant 1}$ which associated sequence $\{S_m\}_{m \geqslant 1}$ can be written as $\{S(1/u_m)\}_{m \geqslant 1}$ (using definition (13)). We know that if $\{R_m\}_{m \geqslant 1}$ is a counter-example to Conjecture 2.2, then $\|S(1/u_m)\|_2 \to \infty$, which shows that $u_m \to \infty$ thanks to (15). Let us show that it implies $\sum t_m^2 < \infty$.

Lemma 4.4. *For any sequence $u_m \to \infty$, there exists $0 < \beta_1 < \beta_2 < \infty$, $\eta > 0$, and $m_0 \in \mathbb{N}$, such that for all $m \geqslant m_0$ there exists $p_m \in \mathbb{N}$ complying with*

$$
u_{p_m} \in [\beta_1 u_m, \beta_2 u_m] \tag{28}
$$

$$
u_{p_m+1} \notin [(1-\eta)/(1+\eta)\, u_m, (1+\eta)/(1-\eta)\, u_m] \tag{29}
$$

Proof. For every $\eta > 0$ and $x > 0$, denote $I_\eta(x)$ the interval $[(1-\eta)/(1+\eta)\, x, (1+\eta)/(1-\eta)\, x]$. Suppose the conclusion is false. The we know that for every $0 < \beta_1 < \beta_2 < \infty$, every $\eta > 0$ and every M, there exists $m \geqslant M$ such that for all p, $u_p \in [\beta_1 u_m, \beta_2 u_m] \Rightarrow u_{p+1} \in I_\eta(u_m)$. Let us take $\beta_1 = (1-\eta)/(1+\eta)$ and $\beta_2 = (1+\eta)/(1-\eta)$, for some arbitrary η. In this case, we know that, for some m, for all p, $u_p \in I_\eta(u_m) \Rightarrow u_{p+1} \in I_\eta(u_m)$. But we also know that $u_m \in I_\eta(u_m)$, so it becomes clear that by induction, $f(n) \in I_\eta(u_m)$ for all $n \geqslant m$, which is in contradiction with $u_m \to \infty$.

PROPOSITION 4.1. *For every partially greedy sequence which can be written as $R_m = \|R_m\|_\infty\, S(1/u_m)$ and is a counter-example to Conjecture 2.2, we have*

$$t_m \asymp \|R_m - R_{m+1}\|_2,$$

hence

$$\sum t_m^2 < \infty.$$

Proof. One can check that

$$K_{m,\,p} = u_m^2 \, \frac{2u_p - 1}{(u_p + u_m - 1)^2} \left(\frac{u_{p+1} - u_m}{u_{p+1} + u_m - 1} \right)^2. \tag{30}$$

It is then easy to show that

$$K_m = \sup_{p,\,m} K_{m,\,p} \leqslant \frac{u_m^2}{2u_m - 1} \sim u_m/2. \tag{31}$$

Moreover, using Lemma 4.4, one gets a sequence p_m such that

$$K_{m,\,p_m} \geqslant u_m^2 \, \frac{2\beta_1 u_m - 1}{((\beta_2 + 1)\, u_m - 1)^2}\, \eta^2 \sim \frac{2\beta_1 \eta^2}{\beta_2^2}\, u_m \tag{32}$$

which shows that $K_m \asymp u_m$. From (15) this becomes $K_m \asymp \|S(1/u_m)\|_2^2$ thus, using (26), we get

$$t_m^2 \asymp \|R_m - R_{m+1}\|_2^2 / (\|R_m\|_\infty^2 \, \|S(1/u_m)\|_2^2) = \|R_m - R_{m+1}\|_2^2 / \|R_m\|_2^2$$

which finally gives $t_m^2 \asymp \|R_m - R_{m+1}\|_2^2$ using (12). We get the square summability of t_m from Lemma 4.2. ∎

5. CONCLUSION

The family of potential counter-examples $R_m = \|R_m\|_\infty\, S(\varepsilon_m)$ that we have built does not discard the possibility that $\sum t_m^2 = \infty$ might be a sufficient condition to ensure the convergence of a weak greedy algorithm. These counter-examples show that too weak a choice of g_m in a greedy algorithm can prevent the algorithm from converging. However, some of our numerical experiments [5] do show convergence of a partially greedy algorithm in the multiscale time-frequency dictionary of Gaussian chirps $\mathscr{D}$, with an improvement of the speed of convergence compared to a pure greedy algorithm in the multiscale Gabor dictionary $\mathscr{D}^*$. One simple reason for the convergence is that numerical experiments use finite dimensional data. But this does not explain the improvement in the rate of convergence. The reason for the good behaviour of this algorithm still has to be investigated. It may be due to the particular structure of $\mathscr{D}^*$ and $\mathscr{D}$ and/or to the properties of the choice functional $R \mapsto g(R)$ which defines a particular set of partially greedy sequences $R_{m+1} = R_m - \langle R_m, g(R_m)\rangle\, g(R_m)$.

ACKNOWLEDGMENT

The author thanks V. Temlyakov for his interest in this research and his very valuable remarks.

REFERENCES

1. F. Bergeaud, “Représentations adaptatives d’images numériques, Matching Pursuit,” Ph.D. thesis, Ecole Centrale Paris, 1995.
2. F. Bergeaud and S. Mallat, Matching pursuit: Adaptive representations of images and sounds, *Comput. Appl. Math.* **15**, 2 (1996).
3. R. A. DeVore and V. N. Temlyakov, Some remarks on greedy algorithms, *Adv. Comput. Math.* **5** (1996), 173–187.
4. R. Gribonval, “Approximations non-linéaires pour l’analyse de signaux sonores,” Ph.D. thesis, Université Paris IX Dauphine, September 1999.
5. R. Gribonval, Fast matching pursuit with a multiscale dictionary of Gaussian chirps, *IEEE Trans. Signal Process.*, in press, 2001.
6. P. J. Huber, Projection pursuit, *Ann. Statist.* **13** (1985), 435–475.
7. L. K. Jones, On a conjecture of Huber concerning the convergence of PP-regression, *Ann. Statist.* **15** (1987), 880–882.
8. S. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE Trans. Signal Process.* **41** (1993), 3397–3415.
9. L. Rejtö and G. G. Walter, Remarks on projection pursuit regression and density estimation, *Stochastic Anal. Appl.* **10** (1992), 213–222.
10. V. N. Temlyakov, Weak greedy algorithms, *Adv. Comput. Math.* **12** (2000), 213–227.