# Industrial Mathematics Institute

2006:01

On the Lebesgue type inequalities for greedy approximation

D.L. Donoho, M. Elad and V.N. Temlyakov

IMI

Preprint Series

Department of Mathematics
University of South Carolina

# ON THE LEBESGUE TYPE INEQUALITIES FOR GREEDY APPROXIMATION

D.L. Donoho, M. Elad, and V. N. Temlyakov

Abstract. We study the efficiency of greedy algorithms with regard to redundant dictionaries in Hilbert spaces. We obtain upper estimates for the errors of the Pure Greedy Algorithm and the Orthogonal Greedy Algorithm in terms of the best $m$-term approximations. We call such estimates the Lebesgue type inequalities. We prove the Lebesgue type inequalities for dictionaries with special structure. We assume that the dictionary has a property of mutual incoherence (the coherence parameter of the dictionary is small). We develop a new technique that, in particular, allowed us to get rid of an extra factor $m^{1/2}$ in the Lebesgue type inequality for the Orthogonal Greedy Algorithm.

## 1. Introduction

A. Lebesgue proved the following inequality: for any $2\pi$-periodic continuous function $f$ one has

$$\|f - S_n(f)\|_\infty \leq (4 + \frac{4}{\pi^2} \ln n) E_n(f)_\infty, \tag{1.1}$$

where $S_n(f)$ is the $n$th partial sum of the Fourier series of $f$ and $E_n(f)_\infty$ is the error of the best approximation of $f$ by the trigonometric polynomials of order $n$ in the uniform norm $\| \cdot \|_\infty$. The inequality (1.1) relates the error of a particular method $(S_n)$ of approximation by the trigonometric polynomials of order $n$ to the best possible error $E_n(f)_\infty$ of approximation by the trigonometric polynomials of order $n$. By the Lebesgue type inequality we mean an inequality that provides an upper estimate for the error of a particular method of approximation of $f$ by elements of a special form, say, form $\mathcal{A}$, by the best possible approximation of $f$ by elements of the form $\mathcal{A}$. In the case of approximation with regard to bases (or minimal systems) the Lebesgue type inequalities are known both in linear and in nonlinear settings (see surveys [KT], [T3]). It would be very interesting to prove the Lebesgue type inequalities for redundant systems (dictionaries). However, there are substantial difficulties on this way. We begin our discussion with the Pure Greedy Algorithm, (PGA). We say a set of functions $\mathcal{D}$ from a Hilbert space $H$ is a *dictionary* if each $g \in H$ has norm one ($\|g\| := \|g\|_H = 1$) and the closure of span$\mathcal{D}$ coincides with $H$. We describe the PGA for a general dictionary $\mathcal{D}$. If $f \in H$, we let $g(f) \in \mathcal{D}$ be an element from $\mathcal{D}$ which maximizes $|\langle f, g \rangle|$. We will assume for simplicity that such a maximizer exists; if not suitable modifications are necessary (see Weak Greedy Algorithm in [T2]) in the algorithm that follows. We define

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

and

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

**Pure Greedy Algorithm (PGA).** *We define* $f_0 := R_0(f,\mathcal{D}) := f$ *and* $G_0(f,\mathcal{D}) := 0$. *Then, for each* $m \geq 1$, *we inductively define*

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

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

It is natural to compare performance of the PGA with the best $m$-term approximation with regard to a dictionary $\mathcal{D}$. We let $\Sigma_m(\mathcal{D})$ denote the collection of all functions (elements) in $H$ which can be expressed as a linear combination of at most $m$ elements of $\mathcal{D}$. Thus each function $s \in \Sigma_m(\mathcal{D})$ can be written in the form

$$s = \sum_{g\in\Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad \#\Lambda \leq m,$$

where the $c_g$ are real or complex numbers. In some cases, it may be possible to write an element from $\Sigma_m(\mathcal{D})$ in this form in more than one way. The space $\Sigma_m(\mathcal{D})$ is not linear: the sum of two functions from $\Sigma_m(\mathcal{D})$ is generally not in $\Sigma_m(\mathcal{D})$.

For a function $f \in H$ we define its best $m$-term approximation error

$$\sigma_m(f) := \sigma_m(f,\mathcal{D}) := \inf_{s\in\Sigma_m(\mathcal{D})} \|f - s\|.$$

It seems like there is no hope to prove a nontrivial Lebesgue type inequality for the PGA in the case of an arbitrary dictionary $\mathcal{D}$. This pessimism is based on the following result from [DT].

Let $\mathcal{B} := \{h_k\}_{k=1}^{\infty}$ be an orthonormal basis in a Hilbert space $H$. Consider the following element

$$g := Ah_1 + Ah_2 + aA\sum_{k\geq 3}(k(k+1))^{-1/2}h_k$$

with

$$A := (33/89)^{1/2} \quad \text{and} \quad a := (23/11)^{1/2}.$$

Then, $\|g\| = 1$. We define the dictionary $\mathcal{D} = \mathcal{B} \cup \{g\}$. It has been proved in [DT] that for the function

$$f = h_1 + h_2$$

we have

$$\|f - G_m(f,\mathcal{D})\| \geq m^{-1/2}, \quad m \geq 4.$$

It is clear that $\sigma_2(f,\mathcal{D}) = 0$.

Therefore, we look for conditions on a dictionary $\mathcal{D}$ that allow us to prove the Lebesgue type inequalities. The condition $\mathcal{D} = \mathcal{B}$ is an orthonormal basis for $H$ guarantees that

$$\|R_m(f,\mathcal{B})\| = \sigma_m(f,\mathcal{B}).$$

This is an ideal situation. The results that we will discuss here concern the case when we replace an orthonormal basis $\mathcal{B}$ by a dictionary that is, in a certain sense, not far from an orthonormal basis.

Let us begin with results that are close to known results from [T1]. We give a definition of a $\lambda$-quasiorthogonal dictionary with depth $D$. In the case $D = \infty$ this definition coincides with the definition of a $\lambda$-quasiorthogonal dictionary from [T1].

**Definition 1.** *We say $\mathcal{D}$ is a $\lambda$-quasiorthogonal dictionary with depth $D$ if for any $n \in [1, D]$ and any $g_i \in \mathcal{D}$, $\quad i = 1,\dots,n$, there exists a collection $\varphi_j \in \mathcal{D}$, $\quad j = 1,\dots,J$, $\quad J \le N := \lambda n$, with the properties:*

$$g_i \in X_J := \operatorname{span}(\varphi_1,\dots,\varphi_J), \quad i = 1,\dots,n,$$

*and for any $f \in X_J$ we have*

$$\max_{1\le j\le J} |\langle f,\varphi_j\rangle| \ge N^{-1/2}\|f\|.$$

**Remark 1.** *It is clear that an orthonormal dictionary is a 1-quasiorthogonal dictionary.*

The following theorem in the case $D = \infty$ has been established in [T1]. The corresponding proof from [T1] also works in the case $D < \infty$ and gives the following result.

**Theorem 1.** *Let a given dictionary $\mathcal{D}$ be $\lambda$-quasiorthogonal with depth $D$ and let $0 < r < (2\lambda)^{-1}$ be a real number. Then for any $f$ such that*

$$\sigma_m(f,\mathcal{D}) \le m^{-r}, \quad m = 1, 2, \dots, D,$$

*we have*

$$\|f_m\| = \|f - G_m(f,\mathcal{D})\| \le C(r,\lambda)m^{-r}, \quad m \in [1, D/2].$$

In this paper we consider dictionaries that have become popular in signal processing. Denote

$$M(\mathcal{D}) := \sup_{g\ne h; g,h\in\mathcal{D}} |\langle g,h\rangle|$$

the coherence parameter of a dictionary $\mathcal{D}$. For an orthonormal basis $\mathcal{B}$ we have $M(\mathcal{B}) = 0$. It is clear that the smaller the $M(\mathcal{D})$ the more the $\mathcal{D}$ resembles an orthonormal basis. However, we should note that in the case $M(\mathcal{D}) > 0$ the $\mathcal{D}$ can be a redundant dictionary. We show in Section 2 (see Proposition 2.1) that a dictionary with coherence $M := M(\mathcal{D})$ is a $(1 + 4\delta)$-quasiorthogonal dictionary with depth $\delta/M$, for any $\delta \in (0, 1/7]$. Therefore, Theorem 1 applies to $M$-coherent dictionaries. We will prove here a general Lebesgue type inequality for the PGA with regard to a $M$-coherent dictionary.

**Theorem 2.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Then for any $S \le 1/(2M)$ we have the following inequality*

$$\|f_S\|^2 \le 2\|f\|(\sigma_S(f,\mathcal{D}) + 5MS\|f\|). \tag{1.2}$$

As a direct corollary of this theorem we obtain the following inequality for functions $f$ that allow a $S$-sparse representation in $\mathcal{D}$ ($\sigma_S(f) = 0$):

$$\|f_S\| \le (10MS)^{1/2}\|f\|.$$

The inequality (1.2) is the first Lebesgue type inequality for the PGA in the case of incoherent dictionary $\mathcal{D}$.

We now proceed to a discussion of the Orthogonal Greedy Algorithm (OGA). If $H_0$ is a finite dimensional subspace of $H$, we let $P_{H_0}$ be the orthogonal projector from $H$ onto $H_0$. That is $P_{H_0}(f)$ is the best approximation to $f$ from $H_0$. As above we let $g(f) \in \mathcal{D}$ be an element from $\mathcal{D}$ which maximizes $|\langle f, g\rangle|$. We shall assume for simplicity that such a maximizer exists; if not suitable modifications are necessary (see Weak Orthogonal Greedy Algorithm in [T2]) in the algorithm that follows.

**Orthogonal Greedy Algorithm (OGA).** *We define $f_0^o := R_0^o(f) := R_0^o(f,\mathcal{D}) := f$ and $G_0^o(f) := G_0^o(f,\mathcal{D}) := 0$. Then for each $m \ge 1$, we inductively define*

$$\begin{aligned} H_m :=& H_m(f) := \operatorname{span}\{g(R_0^o(f)), \dots, g(R_{m-1}^o(f))\} \\ G_m^o(f) :=& G_m^o(f,\mathcal{D}) := P_{H_m}(f) \\ f_m^o := R_m^o(f) :=& R_m^o(f,\mathcal{D}) := f - G_m^o(f). \end{aligned}$$

It is clear from the definition of the OGA that at each step we have

$$\|f_m^o\|^2 \le \|f_{m-1}^o\|^2 - |\langle f_{m-1}^o, g(f_{m-1}^o)\rangle|^2.$$

We note the use of this inequality instead of the equality

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

that holds for the PGA allows us to prove an analogue of Theorem 1 for the OGA. The proof repeats the corresponding proof from [T1]. We formulate this as a remark.

**Remark 2.** *Theorem 1 holds for the OGA instead of the PGA (for $\|f_m^o\|$ instead of $\|f_m\|$).*

The first general Lebesgue type inequality for the OGA for the $M$-coherent dictionary has been obtained in [GMS]. They proved that

$$\|f_m^o\| \le 8m^{1/2}\sigma_m(f) \quad \text{for} \quad m < 1/(32M).$$

The constants in this inequality were improved in [Tr] (see also [DET]):

$$\|f_m^o\| \le (1+6m)^{1/2}\sigma_m(f) \quad \text{for} \quad m < 1/(3M). \tag{1.3}$$

We prove here an analogue of (1.2) for the OGA.

**Theorem 3.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Then for any $S \le 1/(2M)$ we have the following inequalities*

$$\|f_S^o\|^2 \le 2\|f_k^o\|(\sigma_{S-k}(f_k^o) + 3MS\|f_k^o\|), \quad 0 \le k \le S. \tag{1.4}$$

The inequality (1.4) can be used for improving (1.3) for small $m$. We prove here the following inequality.

**Theorem 4.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05M^{-2/3}$. Then for $l \ge 1$ satisfying $2^l \le \log m$ we have*

$$\|f^o_{m(2^l-1)}\| \le 6m^{2^{-l}}\sigma_m(f).$$

**Corollary 1.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05M^{-2/3}$. Then we have*

$$\|f^o_{[m\log m]}\| \le 24\sigma_m(f).$$

## 2. Proofs

We will use the following simple known lemma (see, for instance, [DET]).

**Lemma 2.1.** *Assume a dictionary $\mathcal{D}$ has mutual coherence $M$. Then we have for any distinct $g_j \in \mathcal{D}$, $j = 1, \dots, N$ and for any $a_j$, $j = 1, \dots, N$ the inequalities*

$$(\sum_{j=1}^N |a_j|^2)(1 - M(N-1)) \le \|\sum_{j=1}^N a_j g_j\|_2^2 \le (\sum_{j=1}^N |a_j|^2)(1 + M(N-1)).$$

*Proof.* We have

$$\|\sum_{j=1}^N a_j g_j\|_2^2 = \sum_{j=1}^N |a_j|^2 + \sum_{i\ne j} a_i \bar{a}_j \langle g_i, g_j\rangle.$$

Next,

$$|\sum_{i\ne j} a_i \bar{a}_j \langle g_i, g_j\rangle| \le M \sum_{i\ne j} |a_i a_j| = M(\sum_{i,j} |a_i a_j| - \sum_{i=1}^N |a_i|^2)$$

$$= M((\sum_{i=1}^N |a_i|)^2 - \sum_{i=1}^N |a_i|^2) \le (\sum_{i=1}^N |a_i|^2)M(N-1). \quad \square$$

We now proceed to one more technical lemma (see [DET]).

**Lemma 2.2.** *Suppose that $g_1,\dots,g_N$ are such that $\|g_i\|=1$, $i=1,\dots,N$; $|\langle g_i,g_j\rangle|\le M$, $1\le i\ne j\le N$. Let $H_N:=\operatorname{span}(g_1,\dots,g_N)$. Then for any $f$ we have*

$$\Big(\sum_{i=1}^N|\langle f,g_i\rangle|^2\Big)^{1/2}\ge\Big(\sum_{i=1}^N|c_i|^2\Big)^{1/2}(1-M(N-1)),$$

*where $\{c_i\}$ are from*

$$P_{H_N}(f)=\sum_{i=1}^N c_ig_i.$$

*Proof.* We have $\langle f-P_{H_N}(f),g_i\rangle=0$, $i=1,\dots,N$ and therefore

$$|\langle f,g_i\rangle|=|\langle P_{H_N}(f),g_i\rangle|=|\sum_{j=1}^N c_j\langle g_j,g_i\rangle|\ge|c_i|(1+M)-M\sum_{j=1}^N|c_j|.$$

Next, denoting $\sigma:=(\sum_{j=1}^N|c_j|^2)^{1/2}$ and using the inequality $\sum_{j=1}^N|c_j|\le N^{1/2}\sigma$ we get

$$\Big(\sum_{i=1}^N|\langle f,g_i\rangle|^2\Big)^{1/2}\ge\sigma(1-M(N-1)).$$

□

The following proposition is a direct corollary of Lemmas 2.1 and 2.2.

**Proposition 2.1.** *Let $\delta\in(0,1/7]$. Then any dictionary with mutual coherence $M$ is a $(1+4\delta)$-quasiorthogonal dictionary with depth $\delta/M$.*

*Proof.* Let $n\le\delta/M$. Consider any distinct $g_i\in\mathcal{D}$, $i=1,\dots,n$. Following the Definition 1 we specify $J=n$, $\varphi_j=g_j$, $j=1,\dots,n$. For any $f=\sum_{j=1}^n a_jg_j$ we have by Lemma 2.2

$$\max_{1\le j\le n}|\langle f,g_j\rangle|\ge n^{-1/2}(\sum_{j=1}^n|\langle f,g_j\rangle|^2)^{1/2}\ge n^{-1/2}(\sum_{j=1}^n|a_j|^2)^{1/2}(1-Mn).$$

Using the assumption $n\le\delta/M$ we get from here by Lemma 2.1

$$\max_{1\le j\le n}|\langle f,g_j\rangle|\ge n^{-1/2}\frac{1-\delta}{(1+\delta)^{1/2}}\|f\|\ge(n(1+4\delta))^{-1/2}\|f\|.$$

This completes the proof of Proposition 2.1. □

The proofs of Theorems 2 and 3 are similar. We combine these theorems in one Theorem 2.1 and carry out the detailed proof only for the OGA.

**Theorem 2.1.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Then for any $S \le 1/(2M)$ we have the following inequalities*

$$\|f^o_S\|^2 \le 2\|f^o_k\|(\sigma_{S-k}(f^o_k) + 3MS\|f^o_k\|), \quad 0 \le k \le S, \tag{2.1}$$

$$\|f_S\|^2 \le 2\|f\|(\sigma_S(f) + 5MS\|f\|).$$

*Proof.* Denote

$$d(f) := \sup_{g \in \mathcal{D}} |\langle f, g\rangle|. \tag{2.2}$$

For simplicity we assume that the maximizer in (2.2) exists. Then

$$\|f_m\|^2 = \|f_{m-1}\|^2 - d(f_{m-1})^2 \quad \text{and} \quad \|f^o_m\|^2 \le \|f^o_{m-1}\|^2 - d(f^o_{m-1})^2.$$

We carry out the proof for the OGA and later point out the necessary changes for the PGA. Let $k \in [0, S)$ be fixed. Assume $f^o_k \ne 0$. Denote by $g_1, \dots, g_{S-k} \subset \mathcal{D}$ the elements (distinct) that have the biggest inner products with $f^o_k$:

$$|\langle f^o_k, g_1\rangle| \ge |\langle f^o_k, g_2\rangle| \ge \dots \ge |\langle f^o_k, g_{S-k}\rangle| \ge \sup_{g \in \mathcal{D}, g \ne g_i, i=1,\dots,S-k} |\langle f^o_k, g\rangle|.$$

We define a natural number $s$ in the following way. If $\langle f^o_k, g_{S-k}\rangle \ne 0$ then we set $s := S-k$, otherwise $s$ is chosen such that $\langle f^o_k, g_s\rangle \ne 0$ and $\langle f^o_k, g_{s+1}\rangle = 0$. Let $m \in [k, k+s)$ and

$$f^o_m = f - P_{H_m}(f) = f^o_k - P_{H_m}(f^o_k), \quad H_m = \operatorname{span}(\varphi_1, \dots, \varphi_m), \quad \varphi_j \in \mathcal{D}.$$

We note that $\langle f^o_k, \varphi_l\rangle = 0$, $l \in [1, k]$. Therefore, each $g_i$, $i \in [1, s]$, is different from all $\varphi_l$, $l = 1, \dots, k$. There exists an index $i \in [1, m+1-k]$ such that $g_i \ne \varphi_j$, $j = 1, \dots, m$. For this $i$ we estimate

$$\langle f^o_m, g_i\rangle = \langle f^o_k, g_i\rangle - \langle P_{H_m}(f^o_k), g_i\rangle. \tag{2.3}$$

Let

$$P_{H_m}(f^o_k) = \sum_{j=1}^m c_j \varphi_j.$$

Clearly, $\|P_{H_m}(f^o_k)\| \le \|f^o_k\|$. Then by Lemma 2.1

$$(\sum_{j=1}^m |c_j|^2)^{1/2} \le \|f^o_k\|(1 - M(m-1))^{-1/2}.$$

We continue

$$|\langle P_{H_m}(f^o_k), g_i\rangle| \le M \sum_{j=1}^m |c_j| \le M m^{1/2} (\sum_{j=1}^m |c_j|^2)^{1/2} \le M S^{1/2} \|f^o_k\|(1 - MS)^{-1/2}. \tag{2.4}$$

Thus, we get from (2.3) and (2.4) that

$$d(f_m^o) \geq |\langle f_m^o, g_i\rangle| \geq |\langle f_k^o, g_i\rangle| - MS^{1/2}\|f_k^o\|(1-MS)^{-1/2}, \quad i \in [1, m+1-k].$$

Therefore,

$$(2.5) \qquad (\sum_{v=k}^{k+s-1} d(f_v^o)^2)^{1/2} \geq (\sum_{i=1}^{s}(|\langle f_k^o, g_i\rangle| - MS^{1/2}\|f_k^o\|(1-MS)^{-1/2})^2)^{1/2}$$

$$\geq (\sum_{i=1}^{s}|\langle f_k^o, g_i\rangle|^2)^{1/2} - MS\|f_k^o\|(1-MS)^{-1/2}.$$

Next, let

$$\sigma_{S-k}(f_k^o) = \|f_k^o - P_{H(n)}(f_k^o)\|, \quad P_{H(n)}(f_k^o) = \sum_{j=1}^{n} b_j\psi_j, \quad n \leq S-k,$$

where $\psi_j \in \mathcal{D}$, $j = 1, \dots, n$, are distinct. Then

$$\|P_{H(n)}(f_k^o)\| \geq \|f_k^o\| - \sigma_{S-k}(f_k^o)$$

and by Lemma 2.1

$$(2.6) \qquad \sum_{j=1}^{n}|b_j|^2 \geq (\|f_k^o\| - \sigma_{S-k}(f_k^o))^2(1+MS)^{-1}.$$

By Lemma 2.2

$$(2.7) \qquad \sum_{j=1}^{n}|\langle f_k^o, \psi_j\rangle|^2 \geq (\sum_{j=1}^{n}|b_j|^2)(1-MS)^2.$$

We get from (2.6) and (2.7)

$$\sum_{i=1}^{s}|\langle f_k^o, g_i\rangle|^2 \geq \sum_{j=1}^{n}|\langle f_k^o, \psi_j\rangle|^2 \geq (\|f_k^o\| - \sigma_{S-k}(f_k^o))^2(1+MS)^{-1}(1-MS)^2.$$

Finally, by (2.5) we get from here

$$(2.8) \qquad (\sum_{v=k}^{k+s-1} d(f_v^o)^2)^{1/2} \geq (\|f_k^o\| - \sigma_{S-k}(f_k^o))\frac{1-MS}{(1+MS)^{1/2}} - MS\|f_k^o\|(1-MS)^{-1/2}$$

and

$$\|f^o_S\|^2 \le \|f^o_k\|^2 - \sum_{v=k}^{k+s-1} d(f^o_v)^2 \le 2\|f^o_k\|(\|f^o_k\| - (\sum_{v=k}^{k+s-1} d(f^o_v)^2)^{1/2}).$$

We now use (2.8) to estimate

$$\|f^o_k\| - (\sum_{v=k}^{k+s-1} d(f^o_v)^2)^{1/2} \tag{2.9}$$

$$\le \sigma_{S-k}(f^o_k) + \|f^o_k\|\big(1 - \frac{1-MS}{(1+MS)^{1/2}} + \frac{MS}{(1-MS)^{1/2}}\big).$$

Let $MS \le 1/2$. Denote $x := MS$. Using the inequalities

$$\frac{1-x}{(1+x)^{1/2}} \ge 1 - \frac{3}{2}x, \quad x \le 1/2,$$

and

$$(1-x)^{-1/2} \le 2^{1/2} \le \frac{3}{2}, \quad x \le 1/2,$$

we continue (2.9)

$$\le \sigma_{S-k}(f^o_k) + 3MS\|f^o_k\|.$$

This completes the proof of Theorem 2.1 for the OGA. A few changes adapt the above proof for $k = 0$ to the PGA setting. As above we write

$$f_m = f - G_m(f); \quad G_m(f) = \sum_{j=1}^{m} b_j\psi_j, \quad \psi_j \in \mathcal{D},$$

and estimate $|\langle f_m, g_i\rangle|$ with $i \in [1, m+1]$ such that $g_i \ne \psi_j$, $j = 1, \dots, m$. Using instead of $\|P_{H_m}(f)\| \le \|f\|$ the inequality

$$\|G_m(f)\| \le \|f\| + \|f_m\| \le 2\|f\|$$

we obtain the following analogue of (2.4)

$$|\langle G_m(f), g_i\rangle| \le 2MS^{1/2}\|f\|(1-MS)^{-1/2}. \tag{2.10}$$

The rest of the proof is the same with (2.4) replaced by (2.10). □

We now show how one may combine the inequalities from Theorem 3 with the inequality (1.3). We formulate Theorem 4 here for convenience.

**Theorem 2.2.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05M^{-2/3}$. Then for $l \ge 1$ satisfying $2^l \le \log m$ we have*

$$\|f^o_{m(2^l-1)}\| \le 6m^{2^{-l}}\sigma_m(f). \tag{2.11}$$

*Proof.* We will prove (2.11) by induction on $l$. In the case $l = 1$ by (1.3) the inequality (2.11) holds because $0.05M^{-2/3} \le M^{-1}/3$. We now assume that (2.11) holds for $l - 1 \ge 1$. Using Theorem 3 with $S = m(2^l - 1)$, $k = m(2^{l-1} - 1)$ we get

$$\|f^o_{m(2^l-1)}\|^2 \le 2\|f^o_{m(2^{l-1}-1)}\|(\sigma_{m2^{l-1}}(f^o_{m(2^{l-1}-1)}) + 3Mm(2^l - 1)\|f^o_{m(2^{l-1}-1)}\|). \tag{2.12}$$

The application of Theorem 3 is justified by the inequality

$$MS \le M2^l m \le 0.05M^{1/3}(\log(1/M))\frac{2}{3} \le 1/2,$$

where we have used the estimate

$$x^a \ln(1/x) \le (ae)^{-1}, \quad x \in [0, 1], \tag{2.13}$$

for $a > 0$. By the induction assumption we obtain from (2.12)

$$\|f^o_{m(2^l-1)}\|^2 \le 12m^{2^{1-l}}\sigma_m(f)(\sigma_m(f) + 3Mm2^l 6m^{2^{1-l}}\sigma_m(f)). \tag{2.14}$$

We will prove that under our assumptions

$$18Mm^{1+2^{1-l}}2^l \le 2. \tag{2.15}$$

It is clear that (2.14) and (2.15) imply (2.11). So, it remains to establish (2.15). In the case $l = 2$ we have

$$18Mm^{3/2}4 \le 72(0.05)^{3/2} \le 2.$$

Let $l > 2$ be such that $2^l \le \log m$. Then we have

$$18Mm^{1+2^{1-l}}2^l \le 0.9M^{(1-2^{2-l})/3}\log m \le 0.6M^{1/6}\log(1/M).$$

We obtain from (2.13) that

$$x^{\frac{1}{6}}\ln(1/x) \le 6/e, \quad x \in [0, 1].$$

Therefore,

$$M^{1/6}\log(1/M) \le 6(\log e)/e \le 9/e.$$

It remains to note that

$$0.6(9/e) \le 2.$$

This completes the proof of (2.15) and the proof of Theorem 4. $\square$

**Corollary 2.1.** *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05 M^{-2/3}$. Then we have*

$$\|f^o_{[m \log m]}\| \le 24 \sigma_m(f).$$

*Proof.* Let $l$ be such that $2^l \le \log m < 2^{l+1}$. Then

$$m^{2^{-l}} = 2^{2^{-l} \log m} \le 4,$$

and

$$6 m^{2^{-l}} \le 24.$$

It remains to apply Theorem 2.2. $\square$

## References


[DET] D.L. Donoho, M. Elad, and V.N. Temlyakov, *Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise*, IMI-Preprints **06** (2004), 1–42.

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

[GMS] A.C. Gilbert, S. Muthukrishnan, and M.J. Strauss, *Approximation of functions over redundant dictionaries using coherence*, The 14th Annual ACM-SIAM Symposium on Discrete Algorithms (2003).

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

[T1] V.N. Temlyakov, *Greedy Algorithms and $m$-term Approximation With Regard to Redundant Dictionaries*, J. Approx. Theory **98** (1999), 117–145.

[T2] V.N. Temlyakov, *Weak greedy algorithms*, Advances in Comp. Math. **12** (2000), 213–227.

[T3] V.N. Temlyakov, *Nonlinear Methods of Approximation*, Found. Comput. Math. **3** (2003), 33–107.

[Tr] J.A. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory, **50** (2004), 2231–2242.