CONSTRUCTIVE THEORY OF FUNCTIONS, Sozopol 2010:
In memory of Borislav Bojanov
(G. Nikolov and R. Uluchev, Eds.), pp. 182-191
Prof. Marin Drinov Academic Publishing House, Sofia, 2012

On Greedy Algorithms for Dictionary with Bounded Cumulative Coherence*

Eugene Livshitz

We discuss upper and lower estimates for the rate of convergence of Pure and Orthogonal Greedy Algorithms for dictionary with bounded cumulative coherence.

1. Introduction

Let H be a real, separable Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$. We say that a set $\mathcal{D}$, $\mathcal{D} \subset H$ is a dictionary if

$$g \in \mathcal{D} \quad \Rightarrow \|g\| = 1, \quad \text{and} \quad \overline{\operatorname{span}}\, \mathcal{D} = H.$$

Recently the following problem has been intensively studied in Approximation Theory and Numeral Analysis: for element $f \in H$ and $m \in \mathbb{N}$ to construct an m-term combination

$$f \to \sum_{k=1}^{m} c_k(f) g_k(f), \qquad c_k(f) \in \mathbb{R},\ g_k(f) \in \mathcal{D}$$

that provides a good approximation to f. Greedy Algorithms turn out to be effective for obtaining such m-term approximations (see tutorial [7] for details). Two most popular greedy algorithms are defined below.

Pure Greedy Algorithm (PGA). *Set* $f_0^{PGA} := f \in H$, $G_0^{PGA}(f, \mathcal{D}) := 0$. *For each* $m \geq 0$ *we inductively find* $g_{m+1}^{PGA} \in \mathcal{D}$ *such that*

$$|\langle f_m^{PGA}, g_{m+1}^{PGA} \rangle| = \sup_{g \in \mathcal{D}} |\langle f_m^{PGA}, g \rangle|$$

*This research is partially supported by Russian Foundation for Basic Research project 10-01-00442 and 11-01-00476.

and define

$$G_{m+1}^{PGA}(f,\mathcal{D}) := G_m^{PGA}(f,\mathcal{D}) + \langle f_m^{PGA}, g_{m+1}^{PGA}\rangle g_{m+1}^{PGA},$$

$$f_{m+1}^{PGA} := f - G_{m+1}^{PGA}(f,\mathcal{D}) = f_m^{PGA} - \langle f_m^{PGA}, g_{m+1}^{PGA}\rangle g_{m+1}^{PGA}.$$

Orthogonal Greedy Algorithm (OGA). *Set* $f_0^{OGA} := f \in H$, $G_0^{OGA}(f,\mathcal{D}) := 0$. *For each* $m \geq 0$ *we inductively find* $g_{m+1}^{OGA} \in \mathcal{D}$ *such that*

$$|\langle f_m^{OGA}, g_{m+1}^{OGA}\rangle| = \sup_{g\in\mathcal{D}} |\langle f_m^{OGA}, g\rangle|$$

and define

$$G_{m+1}^{OGA}(f,\mathcal{D}) := Proj_{g_1^{OGA},\dots,g_{m+1}^{OGA}}(f),$$

$$f_{m+1}^{OGA} := f - G_{m+1}^{OGA}(f,\mathcal{D}).$$

Thus for $f \in H$ and each $m \geq 1$ we construct m-term approximations $G_m^{PGA}(f,\mathcal{D})$ and $G_m^{OGA}(f,\mathcal{D})$.

In this article we study the rate of convergence of Greedy Algorithms for class $\mathcal{A}_0(\mathcal{D})$ that is a set of finite linear combination of elements from $\mathcal{D}$ and classes $\mathcal{A}^p(\mathcal{D})$, $1 \leq p < 2$, defined below. For $M \geq 0$ we define

$$\mathcal{A}^p(\mathcal{D},M) := \overline{\big\{\sum_{\lambda\in\Lambda} c_\lambda g^\lambda : \sum_{\lambda\in\Lambda} |c_\lambda|^p \leq M^p,\ c_\lambda \in \mathbb{R},\ g^\lambda \in \mathcal{D},\ \sharp\Lambda < \infty\big\}},$$

(where closure is taken in the norm of H). Set

$$\mathcal{A}^p(\mathcal{D}) := \bigcup_{M\geq 0} \mathcal{A}^p(\mathcal{D},M),$$

$$|f|_p := |f|_{\mathcal{A}^p(\mathcal{D})} := \inf\{M \geq 0 : f \in \mathcal{A}^p(\mathcal{D},M)\}, \qquad f \in \mathcal{A}^p(\mathcal{D}).$$

From results of DeVore, Temlyakov and Livshitz [1], [6], [5] it follows that Orthogonal Greedy Algorithm does provide the optimal rate of convergence $C|f|_1 m^{-1/2}$ in $\mathcal{A}^1(\mathcal{D})$, but Pure Greedy Algorithm does not. For narrower classes such as $\mathcal{A}_0(\mathcal{D})$ the rate of convergence of OGA could not be better than $Cm^{-1/2}$ and would not be optimal. In the same time if dictionary $\mathcal{D}$ satisfies some additional properties the rate of convergence of Greedy Algorithms (for some classes) could be essentially better. This area is called *Sparse Approximation* and has been intensively studied recently ([3], [4], [8], [2]). In this article results will be formulated using the notion of *cumulative coherence* of the dictionary introduced by Tropp [8]

$$\mu_1(\mathcal{D}) := \sup_{g\in\mathcal{D}} \sum_{\widetilde{g}\in\mathcal{D},\ \widetilde{g}\neq g} |\langle \widetilde{g}, g\rangle|. \tag{1}$$

The above-mentioned articles contain the following basic results of Sparse Approximation Theory.

Theorem A. *Let $\mathcal{D}$ be a dictionary with $\mu_1(\mathcal{D}) < 1/2$ and $f \in \mathcal{A}_0(\mathcal{D})$. Then*

$$G_m^{OGA}(f, \mathcal{D}) = f, \qquad m \geq m_0,$$

$$\|f - G_m^{PGA}(f, \mathcal{D})\| = \|f_m\| \leq C \exp(-c(f)m), \qquad m \geq 0.$$

For dictionaries with small $\mu_1(\mathcal{D})$ PGA provides optimal rate of convergence in $\mathcal{A}^p(\mathcal{D})$, $1 \leq p < 2$.

Theorem 1. *Let $\mathcal{D}$ be a dictionary with $\mu_1(\mathcal{D}) < 1/3$ and $f \in \mathcal{A}_1(\mathcal{D})$. Then*

$$\|f - G_m^{PGA}(f, \mathcal{D})\| = \|f_m\| \leq |f|_1 m^{-1/2}, \qquad m \geq 0.$$

Theorem 2. *Suppose $\mathcal{D}$ is a dictionary with $\mu_1(\mathcal{D}) < 1/3$, and $f \in \mathcal{A}^p(\mathcal{D})$, $1 \leq p < 2$. Then there exist $C_1 = C_1(p) > 0$ and $C_2 = C_2(\mu_1(\mathcal{D})) > 0$ such that for any $m \geq 1$*

$$\|f - G_m^{PGA}(f, \mathcal{D})\| = \|f_m\| \leq C_1 C_2 |f|_p m^{-1/p+1/2}.$$

In the same time for big (but finite) values of $\mu_1(\mathcal{D})$ Pure Greedy Algorithms can not always provide exponential rate of convergence, in fact, the rate of convergence could be worse than $Cm^{-1/2}$. We announce the following result.

Theorem 3. *There exists a dictionary $\mathcal{D}$ with $\mu_1(\mathcal{D}) < \infty$, $f_0 \in \mathcal{A}_0(\mathcal{D})$, $\beta > 0$ and $C > 0$ such that for any $m \geq 1$ we have*

$$\|f_0 - G_m^{PGA}(f_0, \mathcal{D})\| = \|f_m\| \geq Cm^{-1/2+\beta}.$$

Let us remind the following definition.

Definition. A dictionary $\mathcal{D} \in H$ is called minimal if for every $g \in \mathcal{D}$ we have

$$g \notin \overline{\operatorname{span}} \left(\mathcal{D} \setminus \{g\}\right).$$

We would like to stress that in all known nontrivial lower estimates for the rate of convergence of PGA (including Theorem 3) dictionary $\mathcal{D}$ is not minimal. We conjecture that minimality of the dictionary may significantly affect on the rate of convergence of PGA in $\mathcal{A}^1(\mathcal{D})$.

Open problem. *Is Pure Greedy Algorithm order-optimal for minimal dictionaries in $\mathcal{A}^1(\mathcal{D})$, that is, for any minimal dictionary $\mathcal{D}$ and $f \in \mathcal{A}^1(\mathcal{D})$, does the inequality*

$$\|f_0 - G_m^{PGA}(f, \mathcal{D})\| \leq Cm^{-1/2}$$

hold for all $m \geq 1$?

2. Properties of Dictionaries with Bounded Cumulative Coherence

It is easy to see that any dictionary with bounded cumulative coherence in separable Hilbert space is countable. We therefore may suppose that elements of a dictionary $\mathcal{D}$ are enumerated: $\mathcal{D} = \{g^\lambda\}_{\lambda\in\mathbb{N}}$.

Lemma 1. *Let $\mathcal{D}$ be a dictionary with $\mu_1(\mathcal{D}) < 1/2$, $N \in \mathbb{N}$, and $c_\nu \in \mathbb{R}$, $g^\nu \in \mathcal{D}$, $1 \le \nu \le N$. Then the following inequalities hold true:*

$$(1 - 2\mu_1(\mathcal{D}))\sum_{\nu=1}^{N} c_\nu^2 \le \Big\|\sum_{\nu=1}^{N} c_\nu g^\nu\Big\|^2 \le (1 + 2\mu_1(\mathcal{D}))\sum_{\nu=1}^{N} c_\nu^2 .$$

Proof. Without loss of generality we may assume that

$$|c_1| \ge |c_2| \ge \cdots \ge |c_N|.$$

We have

$$\Big\|\sum_{\nu=1}^{N} c_\nu g^\nu\Big\|^2 = \langle \sum_{\nu=1}^{N} c_\nu g^\nu, \sum_{\nu=1}^{N} c_\nu g^\nu\rangle = \sum_{\nu=1}^{N}\Big(c_\nu^2\langle g^\nu, g^\nu\rangle + 2c_\nu \sum_{\eta=\nu+1}^{N} c_\eta\langle g^\nu, g^\eta\rangle\Big).$$

Using (1) and monotony of $|c_\nu|$ we estimate

$$\Big|c_\nu^2\langle g^\nu, g^\nu\rangle + 2c_\nu \sum_{\eta=\nu+1}^{N} c_\eta\langle g^\nu, g^\eta\rangle - c_\nu^2\Big| \le 2c_\nu\Big|\sum_{\eta=\nu+1}^{N} c_\eta\langle g^\nu, g^\eta\rangle\Big|$$
$$\le 2c_\nu^2 \sum_{\eta=\nu+1}^{N} |\langle g^\nu, g^\eta\rangle| \le 2c_\nu^2\mu_1(D).$$

Hence

$$\Big|\Big\|\sum_{\nu=1}^{N} c_\nu g^\nu\Big\|^2 - \sum_{\nu=1}^{N} c_\nu^2\Big| \le 2\mu_1(\mathcal{D})\sum_{\nu=1}^{N} c_\nu^2,$$

which is exactly the claim of Lemma 1. □

Lemma 2. *Suppose $\Lambda \subset \mathbb{N}$ is a finite set of indices and $\epsilon > 0$. If for f the representation*

$$f = f_\epsilon + \sum_{\lambda\in\Lambda} c_\lambda g^\lambda, \qquad c_\lambda \in \mathbb{R},\ g^\lambda \in \mathcal{D} \tag{2}$$

holds, and, in addition,

$$\|f_\epsilon\| < \epsilon. \tag{3}$$

Then for $\lambda_0 \in \Lambda$ we have

$$\big|\langle f, g^{\lambda_0}\rangle - c_{\lambda_0}\big| < \mu_1(\mathcal{D})\max_{\lambda\in\Lambda}|c_\lambda| + \epsilon,$$

while for $\lambda_0 \notin \Lambda$ *there holds*

$$\big|\langle f, g^{\lambda_0}\rangle\big| < \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon.$$

Proof. Using representation (2) we write for $\lambda_0 \in \Lambda$

$$\begin{aligned}\langle f, g^{\lambda_0}\rangle - c_{\lambda_0} &= \langle f_\epsilon + \sum_{\lambda\in\Lambda} c_\lambda g^\lambda, g^{\lambda_0}\rangle - c_{\lambda_0}\langle g^{\lambda_0}, g^{\lambda_0}\rangle \\ &= \sum_{\lambda\in\Lambda,\ \lambda\neq\lambda_0} \langle c_\lambda g^\lambda, g^{\lambda_0}\rangle + \langle f_\epsilon, g^{\lambda_0}\rangle\end{aligned}$$

and for $\lambda_0 \notin \Lambda$

$$\langle f, g^{\lambda_0}\rangle = \langle f_\epsilon + \sum_{\lambda\in\Lambda} c_\lambda g^\lambda, g^{\lambda_0}\rangle = \sum_{\lambda\in\Lambda,\ \lambda\neq\lambda_0} \langle c_\lambda g^\lambda, g^{\lambda_0}\rangle + \langle f_\epsilon, g^{\lambda_0}\rangle.$$

To complete the proof we make use of (3) and Cauchy - Bunyakovsky - Schwarz inequality:

$$\begin{aligned}\Big|\sum_{\lambda\in\Lambda,\ \lambda\neq\lambda_0} \langle c_\lambda g^\lambda, g^{\lambda_0}\rangle + \langle f_\epsilon, g^{\lambda_0}\rangle\Big| &\leq \max_{\lambda\in\Lambda}|c_\lambda| \sum_{\lambda\in\Lambda,\ \lambda\neq\lambda_0} |\langle g^\lambda, g^{\lambda_0}\rangle| + \|f_\epsilon\|\|g^{\lambda_0}\| \\ &< \max_{\lambda\in\Lambda}|c_\lambda| \sum_{\widetilde{g}\in\mathcal{D},\ \widetilde{g}\neq g^{\lambda_0}} |\langle \widetilde{g}, g^{\lambda_0}\rangle| + \epsilon \\ &\leq \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda}|c_\lambda| + \epsilon.\end{aligned}$$

□

Lemma 3. *Let* $\mathcal{D}$ *be a dictionary with* $\mu_1(\mathcal{D}) < 1/3$*,* $f \in \mathcal{A}^p(\mathcal{D})$ *and* $m \geq 1$*. Assume that for* $n = m-1$*, finite* $\Lambda \subset \mathbb{N}$ *and* $\epsilon > 0$ *the following representation*

$$f_n = f - G_n^{PGA}(f,\mathcal{D}) = f_\epsilon + \sum_{\lambda\in\Lambda} c_{\lambda,n} g^\lambda, \qquad c_{\lambda,n}\in\mathbb{R},\ g^\lambda\in\mathcal{D},\ \|f_\epsilon\| < \epsilon \quad (4)$$

holds. If

$$\epsilon < \frac{1}{6}(1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_{\lambda,m-1}|, \quad (5)$$

then (4) *holds for* $n = m$ *with the same* Λ*,* f_ϵ*, and*

$$\sum_{\lambda\in\Lambda} |c_{\lambda,m}|^p \leq \sum_{\lambda\in\Lambda} |c_{\lambda,m-1}|^p - 2^{-p}(1 - 3\mu_1(\mathcal{D}))^p \max_{\lambda\in\Lambda} |c_{\lambda,m-1}|^p,$$

$$\max_{\lambda\in\Lambda} |c_{\lambda,m}| \leq \max_{\lambda\in\Lambda} |c_{\lambda,m-1}|. \quad (6)$$

Proof. From the definition of PGA it follows that for $m \geq 1$

$$f - G_m^{PGA}(f, \mathcal{D}) = f_m = f_{m-1} - G_1^{PGA}(f_{m-1}).$$

Therefore it suffices to prove the lemma for arbitrary $f \in \mathcal{A}_p(\mathcal{D})$ and $m = 1$.

For the sake of brevity we write c_λ instead of $c_{\lambda,0}$, $\lambda \in \Lambda$. From (5) we have

$$\begin{aligned}(1 - 2\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - 2\epsilon &> (1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - 3\epsilon \\ &\geq \frac{1}{2}(1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| > 0\,.\end{aligned} \tag{7}$$

By Lemma 2 we get

$$\max_{\lambda\in\Lambda} |\langle f, g^\lambda\rangle| > \max_{\lambda\in\Lambda} |c_\lambda| - \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| - \epsilon = (1 - \mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - \epsilon,$$

while for $\lambda \notin \Lambda$, using also (7) we obtain

$$|\langle f, g^\lambda\rangle| < \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon < (1 - \mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - \epsilon.$$

Therefore there exists $\lambda_0 \in \Lambda$ such that

$$|\langle f, g^{\lambda_0}\rangle| = \sup_{g\in\mathcal{D}} |\langle f, g\rangle| > (1 - \mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - \epsilon. \tag{8}$$

Using Lemma 2, we have

$$|\langle f, g^{\lambda_0}\rangle| < |c_{\lambda_0}| + \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon.$$

Combining last two inequalities, we obtain

$$(1 - \mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - \epsilon < |c_{\lambda_0}| + \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon\,,$$

hence

$$|c_{\lambda_0}| > (1 - 2\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - 2\epsilon.$$

Without loss of generality we may assume that $c_{\lambda_0} \geq 0$, that is

$$c_{\lambda_0} > (1 - 2\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - 2\epsilon. \tag{9}$$

Applying Lemma 2, (9) and (7), we obtain

$$\begin{aligned}\langle f, g^{\lambda_0}\rangle &> c_{\lambda_0} - \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| - \epsilon > (1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - 3\epsilon \\ &\geq \frac{1}{2}(1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda|,\end{aligned} \tag{10}$$

$$\langle f, g^{\lambda_0}\rangle < c_{\lambda_0} + \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon.$$

Hence by (9) and (7)

$$\begin{aligned} c_{\lambda_0} - \langle f, g^{\lambda_0}\rangle &\geq c_{\lambda_0} - \Big(c_{\lambda_0} + \mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon\Big) \\ &\geq -\Big(\mu_1(\mathcal{D}) \max_{\lambda\in\Lambda} |c_\lambda| + \epsilon\Big) \\ &\geq -c_{\lambda_0} + (1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| - 3\epsilon \\ &\geq -c_{\lambda_0} + \frac{1}{2}(1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| \,. \end{aligned} \tag{11}$$

Combining (10) and (11), we estimate

$$|c_{\lambda_0} - \langle f, g^{\lambda_0}\rangle| + \frac{1}{2}(1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda| \leq c_{\lambda_0},$$

and hence

$$|c_{\lambda_0} - \langle f, g^{\lambda_0}\rangle|^p \leq c_{\lambda_0}^p - \Big(\frac{1}{2}(1 - 3\mu_1(\mathcal{D})) \max_{\lambda\in\Lambda} |c_\lambda|\Big)^{1/p}. \tag{12}$$

If we set

$$c_{\lambda,1} = c_\lambda = c_{\lambda,0}, \qquad \lambda \in \Lambda \setminus \{\lambda_0\},$$
$$c_{\lambda_0,1} = c_{\lambda_0,0} - \langle f, g^{\lambda_0}\rangle,$$

then the claim of Lemma 3 will follow from (12). □

3. Proof of Theorem 1

Lemma 3 implies that for any $m \geq 0$

$$|f_m|_1 \leq |f|_1.$$

Using Lemma 3.5 from [1] and Lemma 3 we have for $m \geq 0$

$$|\langle f_m, g_{m+1}\rangle| = \sup_{g\in\mathcal{D}} |\langle f_m, g\rangle| \geq \frac{\|f_m\|^2}{|f_m|_1} \geq \frac{\|f_m\|^2}{|f|_1}.$$

By definition of PGA

$$\|f_{m+1}\|^2 = \|f_m\|^2 - \langle f_m, g_{m+1}\rangle^2 \leq \|f_m\|^2 - \left(\frac{\|f_m\|^2}{|f|_1}\right)^2 = \|f_m\|^2\left(1 - \frac{\|f_m\|^2}{|f|_1^2}\right).$$

Applying Lemma 3.4 from [1] with $a_m = \|f_{m-1}\|^2$ and $A = |f|_1^2$ and taking into account the inequality

$$a_1 = \|f_0\|^2 \leq |f|_1^2,$$

we obtain that for $\{a_m\}_{m=1}^{\infty}$ such that

$$a_{m+1} \le a_m\Big(1 - \frac{a_m}{|f|_1^2}\Big), \qquad a_1 \le A,$$

the following inequality

$$a_m \le Am^{-1}$$

holds. Thus for $m \ge 1$ we have

$$\|f_m\| = a_{m+1}^{1/2} \le |f|_1(m+1)^{-1/2} \le |f|_1 m^{-1/2}.$$

This completes the proof of the theorem. □

4. Proof of Theorem 2

Let $k \ge 1$ and $f \in \mathcal{A}_p$. For an arbitrary ϵ satisfying

$$0 < \epsilon < \frac{1}{6}(1 - 3\mu_1(\mathcal{D}))k^{-1/p}|f|_p\,, \tag{13}$$

there exists representation (2) such that inequality (3) holds and

$$\sum_{\lambda\in\Lambda} |c_\lambda|^p = |f|_p^p\,. \tag{14}$$

We claim that there exists n, $0 \le n \le k$ such that

$$\max_{\lambda\in\Lambda} |c_{\lambda,n}| \le c_1(p)c_2(\mu_1(\mathcal{D}))k^{-1/p}|f|_p\,, \tag{15}$$

$$\sum_{\lambda\in\Lambda} |c_{\lambda,n}|^p \le \sum_{\lambda\in\Lambda} |c_{\lambda,0}|^p = |f|_p^p\,. \tag{16}$$

For every $m = 1, \dots, k$, for $n = m-1$ the representation (4) holds (beginning with $c_{\lambda,0} := c_\lambda$) and either

$$\max_{\lambda\in\Lambda} |c_{\lambda,m-1}| \le k^{-1/p}|f|_p\,,$$

in which case we can set $n = m-1$, or

$$\max_{\lambda\in\Lambda} |c_{\lambda,m-1}| \le k^{-1/p}|f|_p\,.$$

Then taking into account (13) we have (5). Therefore, by Lemma 3 the representation (4) holds for $n = m$ and using (14) and (6) we have

$$\begin{aligned}
0 &\le \sum_{\lambda\in\Lambda} |c_{\lambda,m}|^p \le \sum_{\lambda\in\Lambda} |c_{\lambda,m-1}|^p - 2^{-p}(1 - 3\mu_1(\mathcal{D}))^p \max_{\lambda\in\Lambda} |c_{\lambda,m-1}|^p = \cdots \\
&= \sum_{\lambda\in\Lambda} |c_{\lambda,0}|^p - \sum_{n=1}^{m} 2^{-p}(1 - 3\mu_1(\mathcal{D}))^p \max_{\lambda\in\Lambda} |c_{\lambda,m-1}|^p \\
&\le |f|_p^p - m2^{-p}(1 - 3\mu_1(\mathcal{D}))^p \max_{\lambda\in\Lambda} |c_{\lambda,m}|^p\,,
\end{aligned}$$

whence

$$\max_{\lambda\in\Lambda}|c_{\lambda,m}| \le 2(1-3\mu_1(\mathcal{D}))^{-1}m^{-1/p}|f|_p \quad\text{and}\quad \sum_{\lambda\in\Lambda}|c_{\lambda,m}|^p \le |f|_p^p\,.$$

This provides (15) and (16) for $n = k$.

Using (15) and (16), we estimate

$$\begin{aligned}\sum_{\lambda\in\Lambda}|c_{\lambda,n}|^2 &= \sum_{\lambda\in\Lambda}|c_{\lambda,n}|^p\,|c_{\lambda,n}|^{2-p} \le \Big(\max_{\lambda\in\Lambda}|c_{\lambda,n}|\Big)^{2-p}\sum_{\lambda\in\Lambda}|c_{\lambda,n}|^p \\ &\le \big(c_3(p)c_4(\mu_1(\mathcal{D}))k^{-\frac{2-p}{p}}|f|_p^{2-p}\big)|f|_p^p = c_3(p)c_4(\mu_1(\mathcal{D})k^{-2/p+1}|f|_p^2\,.\end{aligned}$$

Applying Lemma 1 and (4), we obtain

$$\begin{aligned}\|f_k\| \le \|f_n\| \le \|\sum_{\lambda\in\Lambda}c_{\lambda,n}g^\lambda\| + \|f_\epsilon\| &\le \Big((1+2\mu_1(\mathcal{D}))\sum_{\lambda\in\Lambda}c_{\lambda,n}^2\Big)^{1/2} + \epsilon \\ &\le C_1(p)C_2(\mu_1(\mathcal{D}))k^{-1/p+1/2}|f|_p + \epsilon.\end{aligned}$$

Since $\epsilon > 0$ can be arbitrarily small, the last inequality completes the proof of Theorem 2. □

References

[1] R. A. DeVore and V. N. Temlyakov, Some remarks on Greedy Algorithms, *Adv. Comput. Math.* **5** (1996), 173–187.

[2] D. L. Donoho, M. Elad, and V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inform. Th.* **52** (2006), no. 1, 6–18.

[3] A. C. Gilbert, M. Muthukrishnan, and J. Strauss, Approximation of functions over redundant dictionaries using coherence, *in* "Proc. 14th Annual ACM-SIAM Symp. Discrete Algorithms", pp. 243-252, 2003.

[4] R. Gribonval and M. Nielsen, On the strong uniqueness of highly sparse expansions from redundant dictionaries, *in* "Proc. Int. Conf. Independent Component Anal. (ICA'04)", 2004.

[5] E. D. Livshitz, Optimality of the greedy algorithm for some function classes, *Math. Sbornik* **198** (2007), no. 5, 95–114.

[6] E. D. Livshitz and V. N. Temlyakov, Two lower estimates in greedy approximation, *Constr. Approx.* **19** (2003), 509–524.

[7] V. N. Temlyakov, Greedy approximation, *Acta Numer.* **17** (2008), 235–409.

[8] J. A. Tropp, Greed is good: algorithmic results for sparse approximation, *IEEE Trans. Inform. Th.* **50** (2004), no. 10, 2231–2242.

Eugene Livshitz
EverNote Corp.
Berezhkovskaya nab. 6, r. 501
121059, Moscow
RUSSIA
E-mail: livshitz@rambler.ru