Full length article

On the optimality of the Orthogonal Greedy Algorithm for $\mu$-coherent dictionaries

E.D. Livshitz

Moscow State University, Vorobyevy Gory, 1., 119992 Moscow, Russian Federation

Received 26 May 2010; received in revised form 15 November 2011; accepted 20 January 2012

Communicated by Martin Buhmann

Abstract

In this article, we continue to study the performance of Greedy Algorithms. We show that the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) provides an almost optimal approximation on the first $\frac{\mu^{-1}}{20}$ steps for $\mu$-coherent dictionaries.

Keywords: Orthogonal Greedy Algorithms; Orthogonal Matching Pursuit; Best $m$-term approximation; Lebesgue-type inequalities

1. Introduction

In this article, we continue to study the convergence of greedy algorithms with regard to dictionaries with small coherence (see [5,6,11,3,4,10,8]). The study of approximation with regard to incoherent dictionaries was mainly motivated by applications to compressed sensing. In [5,11,3], it was shown that the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) is effective for signal recovering. In this article, we discuss this problem from the point of view of Approximation Theory.

This research was supported in part by the Russian Foundation for Basic Research (grants 10-01-00442 and 11-01-00476).

E-mail addresses: evgliv@gmail.com, livshitz@rambler.ru.

0021-9045/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jat.2012.01.007
Let us recall the standard definitions of Greedy Algorithms theory. We say that a set $\mathcal{D}$ from a Hilbert space $H$ is a dictionary if
\[ \phi \in \mathcal{D} \Rightarrow \| \phi \| = 1, \quad \text{and} \quad \text{span}\mathcal{D} = H. \]

We study dictionaries with small values of coherence
\[ \mu := \sup_{\phi, \psi \in \mathcal{D}, \phi \neq \psi} | \langle \phi, \psi \rangle |. \]  

(1)

Dictionaries with coherence $\mu$ are called $\mu$-coherent. There are many constructions producing highly redundant dictionaries with small coherence. In $H = \mathbb{R}^M$ it is possible to obtain dictionaries consisting of $N$ elements with coherence $\lesssim \frac{\log N}{M^{1/2}}$ for explicit constructions, and $\lesssim \left( \frac{\log N}{M} \right)^{1/2}$ for standard probabilistic constructions [7,1,2,9].

ORTHOGONAL GREEDY ALGORITHM (OGA) Set $f_0 := f \in H$ and $G_0^{OGA}(f, \mathcal{D}) := 0$. For each $m \geq 0$, we inductively find a $g_{m+1} \in \mathcal{D}$ such that
\[ | \langle f_m, g_{m+1} \rangle | = \sup_{g \in \mathcal{D}} | \langle f_m, g \rangle | \]

and define
\[ G_{m+1}^{OGA}(f, \mathcal{D}) := \text{Proj}_{\text{span}(g_1, \ldots, g_{m+1})}(f), \]
\[ f_{m+1} := f - G_{m+1}^{OGA}(f, \mathcal{D}). \]

For a function $f \in H$, we define its best $m$-term approximation
\[ \sigma_m(f) := \sigma_m(f, \mathcal{D}) := \inf_{c_i \in \mathbb{R}, \phi_i \in \mathcal{D}, 1 \leq i \leq m} \left\| f - \sum_{i=1}^{m} c_i \phi_i \right\|. \]

Suppose that dictionary $\mathcal{D}$ is $\mu$-coherent and $m < \frac{1}{2}(\mu^{-1} + 1)$. It is well known (see [5,6]) that if $f \in H$ is $m$-sparse, that is $\sigma_m(f, \mathcal{D}) = 0$, then
\[ f = G_m^{OGA}(f, \mathcal{D}). \]

(2)

Moreover, Temlyakov and Zheltov showed that if $m \geq \frac{1}{2}(\mu^{-1} + 1)$, then equality (2) does not hold for all $\mu$-coherent dictionaries $\mathcal{D}$ and all $m$-sparse $f$:
\[ \exists \mathcal{D}, m \geq \frac{1}{2}(\mu^{-1}(\mathcal{D}) + 1), \quad f \in H : \| f - G_m^{OGA}(f, \mathcal{D}) \| > 0, \]
\[ \sigma_m(f, \mathcal{D}) = 0. \]

(3)

The stability of equality (2) has been intensively studied.

Following Temlyakov, we recall results connecting the error of Greedy approximation and the best $m$-term approximation Lebesgue type inequalities. These inequalities do hold not for all $m \in \mathbb{N}$, but only for $m \leq C(\mu)$; they provide an estimate for the quality of approximation of $A(m)$ iterations of OGA by the best $m$-term approximation:
\[ \| f - G_{A(m)}^{OGA}(f, \mathcal{D}) \| \leq B(m)\sigma_m(f, \mathcal{D}), \quad m \leq C(\mu), \]

with some $A(m) \in \mathbb{N}, B(m), C(\mu) \in \mathbb{R}$.
**Remark 1.** It is natural to assume that $A(m) \geq m$, $B(m) \geq 1$ and $C(\mu) < \frac{1}{2}(\mu^{-1} + 1)$ (see (3)).

The first Lebesgue type inequality for Greedy Algorithms was obtained by Gilbert et al. in [5]. They established (4) for an optimal $A(m) := m$, an order-optimal $C(\mu) := \frac{1}{20\mu^{2/3}}$, and fast growing $B(m) := 8m^{1/2}$.

Donoho et al. [4] obtained inequality (4) with optimal (up to a constant factor) $B(m) = 24$, but not optimal $A(m) := \lfloor m \log m \rfloor$ and $C(\mu) = \frac{1}{20\mu^{2/3}}$.

Recently, Temlyakov and Zheltov [10] proved inequality (4) with $A(m) := m \left[ 2 \sqrt{\log m} \right]$, $B(m) := 24$ and $C(\mu)$, which guarantees inequality $m2\sqrt{2\log m} \leq \frac{1}{26\mu}$. In other words, they proved

**Theorem 1.** For every $\mu$-coherent dictionary $\mathcal{D}$ and any function $f \in H$,

$$\left\| f - G_{m}^{OGA}(f, \mathcal{D}) \right\| \leq 3\sigma_{m}(f, \mathcal{D}), \quad \text{if } m2\sqrt{2\log m} \leq \frac{1}{26\mu}.$$  

It is easy to see that Theorem 1 is optimal (up to a sub-polynomial factor Lebesgue type inequality).

The aim of this article is to prove (4) with

$$A(m) := 2m, \quad B(m) := 2.7, \quad C(\mu) = \frac{1}{20\mu},$$

and thereby to obtain an accurate (up to a constant factor) Lebesgue type inequality (see the theorem below).

**Theorem 2.** For every $\mu$-coherent dictionary $\mathcal{D}$ and any function $f \in H$,

$$\left\| f - G_{2m}^{OGA}(f, \mathcal{D}) \right\| = \left\| f_{2m} \right\| \leq 2.7\sigma_{m}(f, \mathcal{D})$$

for all

$$1 \leq m \leq \frac{1}{20\mu}.$$  

The constants in (5) can be slightly improved, but we do not know the answer to the following problem.

**Open problem.** Is it possible for any $\epsilon > 0$ to prove inequality (4) with

$$A(m) \leq a(\epsilon)m, \quad B(m) \leq b(\epsilon), \quad C(\mu) \geq \frac{1}{c(\epsilon)} \frac{1}{2}(\mu^{-1} + 1),$$

$$a(\epsilon) \geq 1, \quad b(\epsilon) \geq 1, \quad c(\epsilon) \geq 1$$

such that at least one [two, three] of the following inequalities

$$a(\epsilon) < 1 + \epsilon, \quad b(\epsilon) < 1 + \epsilon, \quad c(\epsilon) < 1 + \epsilon$$

hold.
2. Notation

By the definition of the best $m$-term approximation, there exist $a_{j,0} \in \mathbb{R}$, $\psi_j \in D$, $1 \leq j \leq m$, and $\xi_0 \in H$ such that

$$f = f_0 = \sum_{j=1}^{m} a_{j,0}\psi_j + \xi_0, \quad \langle \xi_0, \psi_j \rangle = 0, \ 1 \leq j \leq m,$$

$$\|\xi_0\| \leq 1.01\sigma_m(f, D) = 1.01\sigma_m(f).$$  \hspace{1cm} (6)

Set

$$P_m(\cdot) := \text{Proj}_{\text{span}(\psi_1, \ldots, \psi_m)}(\cdot), \quad P_m^\perp(\cdot) := \text{Proj}_{\text{span}(\psi_1, \ldots, \psi_m)^\perp}(\cdot) = \cdot - P_m(\cdot),$$

$$\xi_n := P_m^\perp(f_n), \quad n \geq 0.$$  

Let the numbers $a_{j,n}, n \geq 0, 1 \leq j \leq m$, satisfy equalities

$$f_n = P_m(f_n) + P_m^\perp(f_n) = \sum_{j=1}^{m} a_{j,n}\psi_j + \xi_n.$$  \hspace{1cm} (7)

Define

$$T_1 := \left\{ i \in \{1, \ldots, 2m\} : g_i \in \{\psi_j\}_{j=1}^{m} \right\}.$$

$$T_2 := \{1, \ldots, 2m\} \setminus T_1.$$  

Then, for $n \geq 1$, we let

$$T_2^n := T_2 \cap \{1, \ldots, n\},$$

$$d_n := \langle f_{n-1}, g_n \rangle,$$  \hspace{1cm} (8)

$$D := \sum_{n \in T_2} d_n^2.$$  \hspace{1cm} (9)

3. Proof of Theorem 2

The idea of our method is to use representation (7) and an accurate estimate of the norm of $\xi_n$. We consider the cases $n \in T_1$ and $n \in T_2$ separately and prove (in Section 5) the following lemmas

**Lemma 1.** Let $n \in T_1$; then

$$\|\xi_n\|^2 \leq \|\xi_{n-1}\|^2 + 0.25D\mu.$$  

**Lemma 2.** Let $n \in T_2$; then

$$\|\xi_n\|^2 \leq \|\xi_{n-1}\|^2 - 0.62d_n^2.$$  

Combining these results, we easily obtain the following statement.
Lemma 3. The following estimates hold

\[ D^{1/2} \leq 1.31 \sigma_m(f), \]  
\[ \|\xi_2m\| \leq \|\xi_0\|. \]  

(10)  
(11)

In Section 5 we will also obtain the upper estimate for \( \|P_m(f_2m)\| \).

Lemma 4. The following estimate holds

\[ \left\| \sum_{j=1}^{m} a_{j,2m} \psi_j \right\|^2 \leq 1.63D. \]  

(12)

Now, using the announced Lemmas 3 and 4, we can obtain the proof of Theorem 2.

\[ \|f_2m\| = \left( \sum_{j=1}^{m} a_{j,2m} \psi_j + \xi_2m \right)^2 \leq \left\| \sum_{j=1}^{m} a_{j,2m} \psi_j \right\|^2 + \|\xi_0\| \]  
(13)

\[ = (1.63D)^{1/2} + 1.01 \sigma_m(f) \]  
(10)

\[ \leq (1.63)^{1/2}1.31 \sigma_m(f) + 1.01 \sigma_m(f) \]  
\[ \leq 2.7 \sigma_m(f). \]  

This completes the proof. □

4. Preliminary lemmas

By conditions of Theorem 2, we have

\[ \mu \leq m\mu \leq 1/20. \]  

(13)

From the definition of OGA, it follows that

\[ f_n = f - \text{Proj}_{\text{span}(g_1,\ldots,g_n)}(f) = \text{Proj}_{\text{span}(g_1,\ldots,g_n)\perp}(f), \]  

(14)

\[ \langle f_n, g_i \rangle = 0, \quad 1 \leq i \leq n. \]  

We first prove a well-known simple lemma that provides estimates for the inner product of

\[ h \in H \]  

with elements of dictionary \( \mathcal{D} \) via the coefficients of the expansion of \( h \) with regard to \( \mathcal{D} \).

Lemma 5. For any \( n, 1 \leq n \leq 2m, \) and

\[ h = \sum_{i=1}^{n} c_i \phi_i, \quad c_i \in \mathbb{R}, \phi_i \in \mathcal{D}, \]  

the following relations hold

\[ \max_{1 \leq i \leq n} |\langle h, \phi_i \rangle| \leq \max_{1 \leq i \leq n} |c_i|(1 + 2m\mu), \]  

(15)

\[ \max_{1 \leq i \leq n} |\langle h, \phi_i \rangle| \geq \max_{1 \leq i \leq n} |c_i|(1 - 2m\mu), \]  

(16)

\[ \max_{1 \leq i \leq n} |c_i| \leq K_1 \max_{1 \leq i \leq n} |\langle h, \phi_i \rangle|, \]  

(17)
where
\[ K_1 := \frac{1}{1 - 2m\mu} = \frac{10}{9}. \]  

**Proof.** For any \(1 \leq i \leq n\), using (1), we have
\[
\langle h, \phi_i \rangle = \langle c_i \phi_i, \phi_i \rangle + \left\langle \sum_{1 \leq j \leq n, j \neq i} c_j \phi_j, \phi_i \right\rangle 
\leq c_i + (n - 1) \left( \max_{1 \leq i \leq n} |c_i| \right) \mu \leq c_i + \left( \max_{1 \leq i \leq n} |c_i| \right) 2m\mu.
\]
Similarly,
\[
\langle h, \phi_i \rangle \geq c_i - \left( \max_{1 \leq i \leq n} |c_i| \right) 2m\mu.
\]
The last two inequalities imply (15) and (16). Inequality (17) follows from (16). \(\square\)

As a consequence, we can state the following

**Lemma 6.** Let \(n \leq 2m\), \(h \in H\), \(\phi_i \in D\), \(1 \leq i \leq n\). Suppose that
\[
\text{Proj}_{\text{span}(\phi_1, \ldots, \phi_n)}(h) = \sum_{i=1}^{n} c_i \phi_i.
\]
Then
\[
\max_{1 \leq i \leq n} |c_i| \leq K_1 \max_{1 \leq i \leq n} |\langle h, \phi_i \rangle|.
\]

**Proof.** Set
\[
h' := \text{Proj}_{\text{span}(\phi_1, \ldots, \phi_n)}(h) = \sum_{i=1}^{n} c_i \phi_i.
\]
It is clear that
\[
\langle h, \phi_i \rangle = \langle h', \phi_i \rangle, \quad 1 \leq i \leq n.
\]
Thus, the lemma follows from inequality (17) for \(h'\). \(\square\)

Let the numbers \(x_{i,n}, 1 \leq i \leq n\), satisfy the equality
\[
f_n = f_{n-1} - \sum_{i=1}^{n} x_{i,n} g_i. \]  
If \(D\) was an orthonormal basis \((\mu(D) = 0)\), then, by the definition of OGA, we would have
\[
f_n = f - \text{Proj}_{\text{span}(g_1, \ldots, g_n)}(f) = f - \text{Proj}_{\text{span}(g_1, \ldots, g_{n-1})}(f) - \text{Proj}_{g_n}(f)
= f_{n-1} - \text{Proj}_{g_n}(f) - \text{Proj}_{\text{span}(g_1, \ldots, g_{n-1})}(f)
= f_{n-1} - \text{Proj}_{g_n}(f_{n-1}) = f_{n-1} - \langle f_{n-1}, g_n \rangle g_n.
\]
Hence
\[ x_{i,n} = 0, \quad 1 \leq i \leq n - 1, \quad x_{n,n} = \langle f_{n-1}, g_n \rangle \]

The following lemma shows how the value of \( x_{i,n} \) depends on the coherence of the dictionary.

**Lemma 7.** For any \( n \leq 2m \), the following estimates hold:
\[
| x_{i,n} | \leq K_1 \mu |d_n|, \quad 1 \leq i \leq n - 1, \quad (20)
\]
\[
| x_{n,n} - d_n | \leq K_1 \mu |d_n|. \quad (21)
\]

**Proof.** Consider the element
\[ h := f_{n-1} - f_n - d_n g_n. \]

Then, taking (19) into account, we can write
\[
h = \sum_{i=1}^{n-1} x_{i,n} g_i + (x_{n,n} - d_n) g_n. \quad (22)
\]

By (1) and (14), for \( 1 \leq i \leq n - 1 \) we have
\[
| \langle h, g_i \rangle | = | \langle f_{n-1} - f_n - d_n g_n, g_i \rangle | \leq | \langle f_{n-1}, g_i \rangle | + | \langle f_n, g_i \rangle | + | d_n || \langle g_n, g_i \rangle | \leq \mu |d_n|,
\]
\[
| \langle h, g_n \rangle | \leq | \langle f_{n-1} - d_n g_n, g_n \rangle | + | \langle f_n, g_n \rangle | = | d_n - d_n | = 0.
\]

Hence
\[
\max_{1 \leq i \leq n} | \langle h, g_i \rangle | \leq \mu |d_n|
\]

and by Lemma 5 and (22) we have
\[
x_{i,n} \leq K_1 \mu |d_n|, \quad 1 \leq i \leq n - 1, \quad (20)
\]
\[
| x_{n,n} - d_n | \leq K_1 \mu |d_n|. \quad \Box
\]

Clearly, \( \lim_{n \to \infty} |d_n| = 0 \). But we cannot guarantee that the sequence \( \{|d_n|\} \) decreases. The following lemma provides an estimate for “non-monotonicity” of \( \{|d_n|\} \).

**Lemma 8.** For any \( 1 \leq l \leq n \leq 2m + 1 \),
\[
|d_n| \leq K_2 |d_l|,
\]

where
\[
K_2 := \exp(2m \mu K_1) \leq \exp(1/9) < 1.118. \quad (23)
\]

**Proof.** Using Lemma 7, for \( 1 \leq l \leq n \leq 2m \) we have
\[
|d_{n+1}| = | \langle f_{n+1}, g_{n+1} \rangle | \leq | \langle f_{n-1} - \sum_{i=1}^{n} x_{i,n} g_i, g_{n+1} \rangle |
\]
\[
\leq | \langle f_{n-1}, g_{n+1} \rangle | + \sum_{i=1}^{n} | x_{i,n} \langle g_i, g_{n+1} \rangle |
\]

\[
\leq K_1 \mu |d_n| + K_1 \mu |d_{n-1}| + \cdots + K_1 \mu |d_1| + K_1 \mu |d_0|
\]
\[
= K_1 \mu \sum_{i=1}^{n} |d_i| \leq K_1 \mu \sum_{i=1}^{n} |d_1| = K_1 \mu n |d_1| \leq K_1 \mu (n+1) |d_1| = K_2 |d_{n+1}|.
\]
\[(1) \quad \leq |d_n| + \mu \left( |x_{n,n}| + \sum_{i=1}^{n-1} |x_{i,n}| \right)\]

\[(20), (21) \quad \leq |d_n| + \mu (|d_n| + nK_1\mu |d_n|)\]

\[\leq |d_n| (1 + \mu (1 + 2m\mu K_1))\]

\[(13), (18) \quad \leq |d_n| \left( 1 + \left( 1 + \frac{2}{20} \frac{10}{9} \mu \right) \right) \leq |d_n| (1 + K_1\mu).\]

Hence for any \(n_i, 1 \leq l \leq n \leq 2m + 1\), we can write

\[|d_n| \leq |d_l|(1 + K_1\mu)^{n-l} \leq |d_l| \left( 1 + \frac{2\mu K_1}{2m} \right)^{2m} \leq |d_l| \exp (2\mu K_1) = K_2|d_l|. \quad \square\]

Now we obtain our main tool for the estimate of \(P_m(f_n) = \sum_{j=1}^{m} a_{j,n}\psi_j\).

**Lemma 9.** For any \(n \geq 1\), the following inequality

\[\max_{1 \leq j \leq m} |a_{j,n-1}| \leq K_1|d_n|. \quad (24)\]

holds.

**Proof.** For any \(l, 1 \leq l \leq m\), we have

\[\left| \left( \sum_{j=1}^{m} a_{j,n-1}\psi_j, \psi_l \right) \right| = \left| \left( \sum_{j=1}^{m} a_{j,n-1}\psi_j + \xi_{n-1}, \psi_l \right) \right| \leq |\langle f_{n-1}, \psi_l \rangle| \leq |d_n|. \quad (7)\]

Then, by Lemma 5,

\[\max_{1 \leq j \leq m} |a_{j,n-1}| \leq K_1 \left( \max_{1 \leq l \leq m} \left| \sum_{j=1}^{m} a_{j,n-1}\psi_j, \psi_l \right| \right) \leq K_1|d_n|. \quad \square\]

We end this section with the proof of a technical lemma that will be used in the proof of Lemma 2.

**Lemma 10.** Let \(1 \leq i \leq 2m\), \(i, n \in T_2\). Then

\[|\langle P_m^\perp (g_n), g_i \rangle| \leq 1.1\mu. \]

**Proof.** Let

\[P_m(g_n) = \sum_{j=1}^{m} c_j\psi_j. \]

Since \(n \in T_2\) and

\[g_n \neq \psi_j, \quad |\langle g_n, \psi_j \rangle| \leq \mu, \quad 1 \leq j \leq m, \]

it follows from Lemma 6 that

\[\max_{1 \leq j \leq m} |c_j| \leq K_1\mu. \]
Therefore, we have
\[
\langle (P_m^\perp g_n), g_i \rangle = \langle (g_n - P_m(g_n)), g_i \rangle \leq \langle (g_n), g_i \rangle + \langle (P_m(g_n)), g_i \rangle
\]
\[
\leq \mu + \left( \sum_{j=1}^{m} c_j \psi_j, g_i \right) \leq \mu + m \left( \max_{1 \leq j \leq m} |c_j| \right) \max_{1 \leq j \leq m} \langle \psi_j, g_i \rangle
\]
\[
\leq \mu + (m \mu) K_1 \mu \leq 1.1 \mu. \quad \Box
\]

5. Proof of the main lemmas

Let us estimate \( \|\xi_n\| \) for \( n \in T_1 \).

5.1. Proof of Lemma 1

Let
\[
t_n := \sharp T_2^n.
\] (25)
If \( T_2^n = \emptyset \), then \( \xi_n = \xi_{n-1} = \xi_0 \) and no prove is needed, so we can assume that \( t_n \geq 1 \). By Lemma 8,
\[
|d_n| \leq K_2 \min_{i \in T_2^n} |d_i|.
\] (26)
On the other hand, by definition (9), we have
\[
\left( \min_{i \in T_2^n} |d_i| \right)^2 \leq \sum_{i \in T_2^n} d_i^2 \leq \sum_{i \in T_2} d_i^2 = D.
\]
Combining this with (26), we obtain
\[
|d_n| \leq K_2 \left( \frac{D}{t_n} \right)^{1/2}.
\]
\[
d_n^2 t_n \leq K_2^2 D.
\] (27)
Define
\[
h := \sum_{i \in T_2^n} x_{i,n} g_i = \sum_{i \in T_2^{n-1}} x_{i,n} g_i.
\] (28)
According to the definition of \( \xi_n \), we have
\[
\xi_n = P_m^\perp (f_n) = P_m^\perp \left( f_{n-1} - \sum_{i=1}^{n} x_{i,n} g_i \right)
\]
\[
= P_m^\perp \left( f_{n-1} - \sum_{i \in T_2^n} x_{i,n} g_i \right) = \xi_{n-1} - P_m^\perp (h),
\]
\[
\|\xi_n\|^2 = \|\xi_{n-1} - P_m^\perp (h)\| \leq \|\xi_{n-1}\|^2 + 2|\langle \xi_{n-1}, P_m^\perp (h) \rangle| + \|P_m^\perp (h)\|^2
\]
\[
\leq \|\xi_{n-1}\|^2 + 2|\langle \xi_{n-1}, h \rangle| + \|h\|^2.
\] (29)
Thus to prove the lemma, we must estimate \(|\langle \xi_{n-1}, h \rangle|\) and \(\|h\|^2\). Using (14) and (28), we obtain
\[
\langle f_{n-1}, h \rangle = 0,
\]
\[
\|\langle \xi_{n-1}, h \rangle\| \overset{(7)}{=} \left\| f_{n-1} - \sum_{j=1}^{m} a_{j,n-1} \psi_j, h \right\| = \left\| \sum_{j=1}^{m} a_{j,n-1} \psi_j, h \right\| \overset{(28)}{\leq} \sum_{j=1}^{m} \left| a_{j,n-1} \right| \left\| \sum_{i \in T_2^{n-1}} x_{i,n} g_i \right\| \leq \sum_{j=1}^{m} \left| a_{j,n-1} \right| \sum_{i \in T_2^{n-1}} |x_{i,n} \langle \psi_j, g_i \rangle|. \tag{30}
\]
Applying (24) we obtain the estimate
\[
\sum_{j=1}^{m} |a_{j,n-1}| \leq m K_1 |d_n|.
\]
It follows from (1) and Lemma 7 that, for \(1 \leq j \leq m\),
\[
\sum_{i \in T_2^{n-1}} |x_{i,n} \langle \psi_j, g_i \rangle| \leq \mu \mu T_2^{n-1} \max_{i \in T_2^{n-1}} |x_{i,n}| \overset{(20)}{\leq} \mu \mu T_2^n K_1 \mu |d_n| \overset{(25)}{=} K_1 \mu^2 |d_n| t_n.
\]
Thus, we can continue (30) as
\[
\|\langle \xi_{n-1}, h \rangle\| \leq \sum_{j=1}^{m} |a_{j,n-1}| \sum_{i \in T_2} |x_{i,n} \langle \psi_j, g_i \rangle| \leq m \mu^2 K_1^2 d_n^2 t_n \leq m \mu^2 K_1^2 K_2^2 D = (K_1^2 K_2^2 / 20) D \mu \leq 0.078 D \mu.
\]
According to Lemma 7, we can write
\[
\|h\|^2 \overset{(28)}{=} \left\| \sum_{i \in T_2^{n-1}} x_{i,n} g_i \right\|^2 \overset{(1)}{=} \left( \max_{i \in T_2^{n-1}} x_{i,n}^2 \right) (\mu T_2^{n-1} + (\mu T_2^{n-1})^2 \mu) \overset{(20)}{\leq} K_1^2 \mu^2 d_n^2 \left( \mu T_2^{n-1} + (\mu T_2^{n-1})^2 \mu \right) \overset{(25)}{\leq} K_1^2 \mu^2 d_n^2 (t_n + t_n^2 \mu) \leq K_1^2 \mu^2 (d_n^2 t_n)(1 + 2m \mu) \overset{(27), (13)}{\leq} K_1^2 \mu \frac{1}{20} K_2^2 D \times 1.1 = 0.085 D \mu. \tag{31}
\]
Now using the estimates for \(|\langle \xi_{n-1}, h \rangle|\) and \(\|h\|^2\), we can continue inequality (29):
\[
\|\xi_{n}\|^2 \leq \|\xi_{n-1}\|^2 + 2 |\langle \xi_{n-1}, h \rangle| + \|h\|^2 \leq \|\xi_{n-1}\|^2 + 2 \times (0.078 D \mu) + 0.085 D \mu \leq \|\xi_{n-1}\|^2 + 0.25 D \mu.
\]
This estimate completes the proof of the lemma. \(\square\)

Now we proceed to the estimate of \(\|\xi_n\|\) for \(n \in T_2\).
5.2. Proof of Lemma 2

Just as in the proof of Lemma 1, we use the element

\[ h := \sum_{i \in T_2^{n-1}} x_{i,n} g_i. \]

Set

\[ \xi'_n := P_m^\perp(f_{n-1} - x_{n,n}g_n). \]

Then we can write

\[
\xi_n = P_m^\perp(f_n) = P_m^\perp(f_{n-1} - \sum_{i=1}^{n} x_{i,n} g_i) \\
= P_m^\perp(f_{n-1} - x_{n,n}g_n) - P_m^\perp\left(\sum_{i=1}^{n-1} x_{i,n} g_i\right) \\
= \xi'_n - P_m^\perp\left(\sum_{i \in T_2^{n-1}} x_{i,n} g_i\right) = \xi'_n - P_m^\perp(h),
\]

\[
\|\xi_n\|^2 = \|\xi'_n\|^2 - 2 \langle\xi'_n, P_m^\perp(h)\rangle + \|P_m^\perp(h)\|^2 \leq \|\xi'_n\|^2 + 2\|\langle\xi'_n, h\rangle\| + \|h\|^2. \tag{32}
\]

Therefore, to prove the lemma it suffices to obtain upper bounds for \(\|\xi'_n\|^2, |\langle\xi'_n, h\rangle|, \|h\|^2\). To estimate \(\|h\|^2\), we can use inequality (31) from Lemma 1.

\[
\|h\|^2 \leq K_1^2 \mu^2 d_n^2 \left(\varepsilon T_2^{n-1} + (\varepsilon T_2^{n-1})^2 \mu\right) \leq K_1^2 \mu^2 d_n^2 (2m + (2m)^2 \mu) \\
\leq 2K_1^2 d_n^2 (m\mu)^2 (1 + 2m\mu) \tag{18),(13} \leq 0.007 d_n^2. \tag{33}
\]

Then we proceed to the estimate of \(\|\xi'_n\|^2\).

Using (7), Lemma 9 and the inclusion \(n \in T_2\), we can write

\[
|\langle\xi_{n-1}, g_n\rangle - d_n| = \left|\langle f_{n-1} - \sum_{j=1}^m a_{j,n-1} \psi_j, g_n\rangle - d_n\right| \\
= \left|\langle f_{n-1}, g_n\rangle - \sum_{j=1}^m a_{j,n-1} \langle \psi_j, g_n\rangle - d_n\right| = \left|\sum_{j=1}^m a_{j,n-1} \langle \psi_j, g_n\rangle\right| \\
\leq \left(\max_{1 \leq j \leq m} |a_{j,n-1}|\right) m \max_{1 \leq j \leq m} |\langle \psi_j, g_n\rangle| \tag{24}, \forall n \in T_2 \leq K_1 |d_n| m\mu. \tag{34}
\]

Then, using Lemma 7, we obtain the estimate

\[
2x_{n,n} \langle\xi_{n-1}, g_n\rangle = 2(d_n + (x_{n,n} - d_n)) (d_n + (\langle\xi_{n-1}, g_n\rangle - d_n)) \\
\geq 2(|d_n| - K_1 m|d_n|)(|d_n| - K_1 |d_n| m\mu) \\
\geq 2d_n^2 (1 - K_1 m\mu) \geq 2d_n^2 (1 - 2K_1 m\mu) \tag{18),(13} \\
= \frac{16}{9} d_n^2.
\]
and, finally, obtain
\[
\|\xi_n\|^2 \leq \|P_m(f_{n-1} - x_{n,n}g_n)\|^2 = \|\xi_{n-1} - x_{n,n}P_m^+(g_n)\|^2
\]
\[
= \|\xi_{n-1}\|^2 - 2x_{n,n}\langle \xi_{n-1}, P_m^+(g_n) \rangle + x_{n,n}^2\|P_m^+(g_n)\|^2
\]
\[
\leq \|\xi_{n-1}\|^2 - 2x_{n,n}\langle \xi_{n-1}, g_n \rangle + x_{n,n}^2 \leq \|\xi_{n-1}\|^2 - \frac{16}{9}d_n^2 + x_{n,n}^2
\]
\[
(21) \leq \|\xi_{n-1}\|^2 - \frac{16}{9}d_n^2 + (|d_n| + K_1\mu|d_n|)^2
\]
\[
\leq \|\xi_{n-1}\|^2 + d_n^2 \left( (1 + K_1\mu\mu)^2 - \frac{16}{9} \right) \leq \|\xi_{n-1}\|^2 - 0.66d_n^2. \quad (35)
\]

It remains to estimate \(|\langle \xi_n', h \rangle|\). Equalities (14) imply that
\[
\langle f_{n-1}, h \rangle = 0.
\]

We have
\[
|\langle \xi_n', h \rangle| = |\langle P_m(f_{n-1}) - x_{n,n}P_m^+(g_n), h \rangle| = |\langle \xi_{n-1} - x_{n,n}P_m^+(g_n), h \rangle|
\]
\[
\overset{(7)}{=} \left| \left( f_{n-1} - \sum_{j=1}^m a_{j,n-1}\psi_j - x_{n,n}P_m^+(g_n), h \right) \right|
\]
\[
\leq \sum_{j=1}^m |\langle a_{j,n-1}\psi_j, h \rangle| + |x_{n,n}\langle P_m^+(g_n), h \rangle|
\]
\[
\leq \sum_{j=1}^m |a_{j,n-1}| \sum_{i \in T_{n-1}} |\langle \psi_j, x_{i,n}g_i \rangle| + \sum_{i \in T_{n-1}} |x_{n,n}x_{i,n}\langle P_m^+(g_n), g_i \rangle|
\]
\[
=: Z_1 + Z_2. \quad (36)
\]

Let us estimate the summands \(Z_1\) and \(Z_2\) separately.
\[
Z_1 \leq \sum_{j=1}^m |a_{j,n-1}| \sum_{i \in T_{n-1}} |x_{i,n}| |\langle \psi_j, g_i \rangle|
\]
\[
\leq \max_{1 \leq j \leq m} |a_{j,n-1}| \max_{i \in T_{n-1}} |x_{i,n}| \sum_{j=1}^m \sum_{i \in T_{n-1}} \mu
\]
\[
(24), (20) \leq K_1|d_n|K_1\mu|d_n|mn\mu \leq 2d_n^2K_1^2(m\mu)^2 \leq 0.007d_n^2.
\]

Using Lemmas 10 and 7, we find that
\[
Z_2 \leq |x_{n,n}| \sum_{i \in T_{n-1}} |x_{i,n}\langle P_m^+(g_n), g_i \rangle| \leq |x_{n,n}| \sum_{i \in T_{n-1}} |x_{i,n}|1.1\mu
\]
\[
\leq (|d_n| + K_1\mu|d_n|)(nK_1\mu|d_n|)1.1\mu
\]
\[
\leq d_n^2(1 + K_1\mu\mu)2.2m^2\mu^2 \leq 0.006d_n^2.
\]

Substituting estimate for \(Z_1\) and \(Z_2\) into (36) we obtain
\[
|\langle \xi_n', h \rangle| \leq Z_1 + Z_2 \leq 0.013d_n^2. \quad (37)
\]
Using inequalities (33), (35) and (37), we can continue estimate (32) and complete the proof:

\[ \| \xi_n \|^2 \leq \| \xi_n' \|^2 + 2|\langle \xi_n', P_m \perp h \rangle| + \| h \|^2 \leq \| \xi_{n-1} \|^2 - 0.66d_n^2 + 2(0.013d_n^2) + 0.007d_n^2 \leq \| \xi_{n-1} \|^2 - 0.62d_n^2. \]

Combining Lemmas 1 and 2, we easily obtain the proof of Lemma 3.

5.3. Proof of Lemma 3

Using Lemmas 1 and 2, we obtain

\[ (1.01\sigma_m(f))^2 \geq \| \xi_0 \|^2 \geq \| \xi_0 \|^2 - \| \xi_{2m} \|^2 = \sum_{n=1}^{2m} (\| \xi_{n-1} \|^2 - \| \xi_n \|^2) \]

\[ \geq \#T_1 (-0.25D\mu) + \sum_{n \in T_2} 0.62d_n^2 \]

\[ \geq m(-0.25D\mu) + 0.62D \geq 0.6D > 0. \]

Hence

\[ D^{1/2} \leq 1.01(0.6)^{-1/2}\sigma_m(f) \leq 1.31\sigma_m(f). \]

It remains to estimate \( \| P_m(f_{2m}) \| \).

5.4. Proof of Lemma 4

Using Lemmas 9 and 8, we can write for any \( l, 1 \leq l \leq 2m \),

\[ \max_{1 \leq j \leq 2m} |a_{j,2m}| \leq K_1|d_{2m+1}| \leq K_1K_2|d_l|. \]

(38)

Since \( \#T_2 \geq m \), using definition (9), we obtain

\[ \sum_{j=1}^{m} a_{j,2m}^2 \leq m \max_{1 \leq j \leq m} a_{j,2m}^2 \leq \sum_{l \in T_2} (K_1K_2|d_l|)^2 = (K_1K_2)^2D. \]

Applying a well-known inequality (see, for example, Lemma 2.1 from [4]) and substituting the values of \( K_1 \) and \( K_2 \) (see (18) and (23)), we find the estimates

\[ \left( \sum_{j=1}^{m} a_{j,2m}^2 \right)^2 \leq \left( \sum_{j=1}^{m} a_{j,2m}^2 \right) (1 + m\mu) \leq (K_1K_2)^2D \times 1.05 \leq 1.63D. \]

Acknowledgments

The author is grateful to Professor V.N. Temlyakov and Professor S.V. Konyagin for helpful discussions. Also, the author would like to thank anonymous referees for very useful suggestions which significantly improved this paper.
References


