Lebesgue type inequalities for greedy approximation

V. Temlyakov

University of South Carolina

Banff 2011

1. Introduction

We say a set of functions $\mathcal{D}$ from a Hilbert space H is a dictionary if each $g \in \mathcal{D}$ has norm one $(\|g\| := \|g\|_H = 1)$ and the closure of span$\mathcal{D}$ coincides with H. 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. For a function $f \in H$ we define its best m-term approximation error

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

Pure Greedy Algorithm (PGA)

Pure Greedy Algorithm (PGA) We define $f_0 := f$. Then for each $m \geq 1$, we inductively define:
(1) $\varphi_m \in \mathcal{D}$ is any satisfying (we assume existence)

$$\langle f_{m-1}, \varphi_m \rangle = \sup_{g \in \mathcal{D}} \langle f_{m-1}, g \rangle;$$

(2)

$$f_m := f_{m-1} - \langle f_{m-1}, \varphi_m \rangle \varphi_m;$$

(3)

$$G_m(f, \mathcal{D}) := \sum_{j=1}^{m} \langle f_{j-1}, \varphi_j \rangle \varphi_j.$$

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.

Orthogonal Greedy Algorithm (OGA). We define $f_0 := f$. Then for each $m \geq 1$ we inductively define:

(1) $\varphi_m \in \mathcal{D}$ is any element satisfying (we assume existence)

$$|\langle f_{m-1}, \varphi_m \rangle| = \sup_{g \in \mathcal{D}} |\langle f_{m-1}, g \rangle|;$$

(2) $G_m(f, \mathcal{D}) := P_{H_m}(f), \quad$ where $\quad H_m := \operatorname{span}(\varphi_1, \ldots, \varphi_m);$

(3) $$f_m := f - G_m(f, \mathcal{D}).$$

Taking care of existence

Let a weakness parameter $t \in (0,1]$ be given.
Weak Orthogonal Greedy Algorithm (WOGA). We define $f_0^{o,t} := f$. Then for each $m \geq 1$ we inductively define:
(1) $\varphi_m^{o,t} \in \mathcal{D}$ is any element satisfying

$$|\langle f_{m-1}^{o,t}, \varphi_m^{o,t} \rangle| \geq t \sup_{g \in \mathcal{D}} |\langle f_{m-1}^{o,t}, g \rangle|.$$

WOGA

(2) Let $H_m^t := \mathrm{span}(\varphi_1^{o,t}, \ldots, \varphi_m^{o,t})$ and let $P_{H_m^t}(f)$ denote an operator of orthogonal projection onto H_m^t. Define

$$G_m^{o,t}(f, \mathcal{D}) := P_{H_m^t}(f).$$

(3) Define the residual after mth iteration of the algorithm

$$f_m^{o,t} := f - G_m^{o,t}(f, \mathcal{D}).$$

In the case $t = 1$, $k = 1, 2, \ldots$, WOGA coincides with the Orthogonal Greedy Algorithm (OGA).

Examples

It is clear that for an orthonormal basis $\mathcal{B}$ of a Hilbert space H we have for each f for both PGA and OGA

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

There is a nontrivial classical example of a redundant dictionary, having the same property: PGA and OGA realize the best m-term approximation for each individual function. We describe that dictionary now. Let Π be a set of functions from $L_2([0,1]^2)$ of the form $u(x_1)v(x_2)$ with the unit L_2-norm. Then for this dictionary and $H = L_2([0,1]^2)$ we have for each $f \in H$

$$\|f - G_m(f, \Pi)\| = \sigma_m(f, \Pi).$$

The Lebesgue inequality

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

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

where $S_n(f)$ is the nth 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 $\| \cdot \|_\infty$.

2. Incoherent dictionaries

We consider dictionaries that have become popular in signal processing. Denote

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

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.

First results

The first general Lebesgue type inequality for OGA for the M-coherent dictionary has been obtained in [Gilbert, Muthukrishnan, Strauss, 2003]. They proved that

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

The constants in this inequality were improved in [Tropp, 2004] (see also [Donoho, Elad, Temlyakov, 2004]):

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

Further results

The following results have been obtained in [Donoho, Elad, Temlyakov, 2006]:

Theorem 2.1. 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

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

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

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

Further results

Theorem 2.2. Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Then for any $S \le 1/(2M)$ we have the following inequalities for OGA

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

and the following inequalities for PGA

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

New results

These inequalities were improved in [Temlyakov and Zheltov, 2010].

Theorem 2.3. Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Then for any $S \leq 1/(2M)$ we have the following inequalities for OGA

$$\|f_S^o\|^2 \leq \sigma_{S-k}(f_k^o)^2 + 5MS\|f_k^o\|^2, \quad 0 \leq k \leq S. \tag{2.1}$$

and the following inequalities for PGA

$$\|f_S\|^2 \leq \sigma_S(f)^2 + 7MS\|f\|^2. \tag{2.2}$$

New results

It was pointed out in [Donoho, Elad, T., 2006] that the inequality $\|f_{[m \log m]}\| \leq 24\sigma_m(f)$ for OGA from the above Corollary is almost (up to a $\log m$ factor) perfect Lebesgue inequality. However, we are paying a big price for it in the sense of a strong assumption on m. It was mentioned in [Donoho, Elad, T., 2006] that it was not known if the assumption $m \leq 0.05M^{-2/3}$ can be substantially weakened. It was shown in [T. and Zheltov, 2010] that it can be substantially weakened.

New results

Theorem 2.4. Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. For any $\delta \in (0, 1/4]$ set $L(\delta) := [1/\delta] + 1$. Assume m is such that $20Mm^{1+\delta}2^{L(\delta)} \leq 1$. Then we have

$$\|f^o_{m(2^{L(\delta)+1}-1)}\| \leq \sqrt{3}\sigma_m(f).$$

Very recently Livshitz, 2010 improved the above Lebesgue-type inequality. He proved that

$$\|f^o_{2m}\| \leq 3\sigma_m(f)$$

for $m \leq (20M)^{-1}$. His proof is different from the proof of Theorem 2.4. It is much more technically involved.

New results

We now demonstrate the use of the inequality (2.2). The following result on PGA from [Temlyakov and Zheltov, 2010] is a corollary of (2.2).

Theorem 2.5. Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. For any $r > 0$ and $\delta \in (0,1]$ set $L(r,\delta) := [r/\delta] + 1$. Let f be such that

$$\sigma_m(f) \le m^{-r}\|f\|, \quad m \le 2^{-L(r,\delta)}(14M)^{-1/(1+\delta)}.$$

Then for all n such that $n \le (14M)^{-1/(1+\delta)}$ we have

$$\|f_n\| \le C(r,\delta)n^{-r}\|f\|.$$

Exact recovery by WOGA

We proceed to exact recovery of sparse signals by the WOGA(t). We present a result in a general setting where we do not assume that the dictionary $\mathcal{D}$ is finite.

Theorem 2.6 Let $\mathcal{D}$ be an M-coherent dictionary. The WOGA(t) recovers exactly any $f \in \Sigma_m(\mathcal{D})$ with $m < \frac{t}{1+t}(1 + M^{-1})$.

3. Banach spaces

We present here a generalization of the concept of M-coherent dictionary to the case of Banach spaces. Let $\mathcal{D}$ be a dictionary in a Banach space X. We define the coherence parameter of this dictionary in the following way

$$M(\mathcal{D}) := \sup_{g \neq h; g,h \in \mathcal{D}} \sup_{F_g} |F_g(h)|.$$

where F_g is the norming (peak) functional of g: $F_g(g) = \|g\|_X$, $\|F_g\|_{X'} = 1$. We note that, in general, a norming functional F_g is not unique. This is why we take $\sup_{F_g}$ over all norming functionals of g in the definition of $M(\mathcal{D})$.

Dual dictionary

We do not need $\sup_{F_g}$ in the definition of $M(\mathcal{D})$ if for each $g \in \mathcal{D}$ there is a unique norming functional $F_g \in X'$. Then we define $\mathcal{D}' := \{F_g, g \in \mathcal{D}\}$ and call $\mathcal{D}'$ a dual dictionary to a dictionary $\mathcal{D}$. It is known that the uniqueness of the norming functional F_g is equivalent to the property that g is a point of Gateaux smoothness:

$$\lim_{u \to 0} (\|g + uy\| + \|g - uy\| - 2\|g\|)/u = 0$$

for any $y \in X$. In particular, if X is uniformly smooth then F_f is unique for any $f \neq 0$.

Quasi-Orthogonal Greedy Algorithm

We considered in [T, 2006] the following greedy algorithm. Weak Quasi-Orthogonal Greedy Algorithm (WQOGA). Let $t \in (0,1]$. Denote $f_0 := f_0^{q,t} := f$ (here and below index q stands for quasi-orthogonal) and find $\varphi_1 := \varphi_1^{q,t} \in \mathcal{D}$ such that

$$|F_{\varphi_1}(f_0)| \geq t \sup_{g \in \mathcal{D}} |F_g(f_0)|.$$

Next, we find c_1 satisfying

$$F_{\varphi_1}(f - c_1\varphi_1) = 0.$$

Denote $f_1 := f_1^{q,t} := f - c_1\varphi_1$.

Quasi-Orthogonal Greedy Algorithm

We continue this construction in an inductive way. Assume that we have already constructed residuals $f_0, f_1, \ldots, f_{m-1}$ and dictionary elements $\varphi_1, \ldots, \varphi_{m-1}$. Now, we pick an element $\varphi_m := \varphi_m^{q,t} \in \mathcal{D}$ such that

$$|F_{\varphi_m}(f_{m-1})| \geq t \sup_{g \in \mathcal{D}} |F_g(f_{m-1})|.$$

Quasi-Orthogonal Greedy Algorithm

Next, we look for $c_1^m, \ldots, c_m^m$ satisfying

$$F_{\varphi_j}(f - \sum_{i=1}^{m} c_i^m \varphi_i) = 0, \quad j = 1, \ldots, m. \tag{3.1}$$

If there is no solution to (3.1) then we stop, otherwise we denote $f_m := f_m^{q,t} := f - \sum_{i=1}^m c_i^m \varphi_i$ with $c_1^m, \ldots, c_m^m$ satisfying (3.1).

Running WQOGA

Remark 3.1. We note that (3.1) has a unique solution if $\det \|F_{\varphi_j}(\varphi_i)\|_{i,j=1}^m \neq 0$. We apply WQOGA in the case of a dictionary with the coherence parameter $M := M(\mathcal{D})$. Then by a simple well known argument on the linear independence of the rows of the matrix $\|F_{\varphi_j}(\varphi_i)\|_{i,j=1}^m$ we conclude that (3.1) has a unique solution for any $m < 1 + 1/M$. Thus, in the case of an M-coherent dictionary $\mathcal{D}$, we can run WQOGA for at least $[1/M]$ iterations.

Exact recovery by WQOGA

The following result was obtained in [T, 2006].
Theorem 3.1 Let $t \in (0,1]$. Assume that $\mathcal{D}$ has coherence parameter M. Let $S < \frac{t}{1+t}(1+1/M)$. Then for any f of the form

$$f = \sum_{i=1}^{S} a_i \psi_i,$$

where ψ_i are distinct elements of $\mathcal{D}$, we have that $f_S^{q,t} = 0$.

Generalization of dual dictionary

We will discuss a more general setting. Instead of a pair $(\mathcal{D}, \mathcal{D}')$ of a dictionary $\mathcal{D}$ and its dual dictionary $\mathcal{D}'$ we now consider a pair $(\mathcal{D}, \mathcal{W})$ of a dictionary $\mathcal{D}$ and a set $\mathcal{W}$ of normalized elements w indexed by elements from $\mathcal{D}$. We define

$$\mathcal{W} := \{w_g \in X', \ \|w_g\|_{X'} = 1, \ g \in \mathcal{D}\}$$

and define the coherence parameter of the pair $(\mathcal{D}, \mathcal{W})$ in the following way

$$M(\mathcal{D}, \mathcal{W}) := \sup_{g \neq h; g,h \in \mathcal{D}} |w_g(h)|.$$

Incoherent pairs

We assume that the pair $(\mathcal{D}, \mathcal{W})$ satisfies the condition

$$w_g(g) \geq 1 - \delta, \qquad g \in \mathcal{D}, \tag{3.2}$$

with some $\delta \in [0, 1)$. If $\delta = 0$ then w_g is a norming functional of g.

Generalization of WQOGA

For a pair $(\mathcal{D}, \mathcal{W})$ we define an analog of WQOGA in the following way.

Weak Projective Greedy Algorithm (WPGA) Let $t \in (0,1]$. Denote $f_0 := f_0^{p,t} := f$ (here and below index p stands for projective) and find $\varphi_1 := \varphi_1^{p,t} \in \mathcal{D}$ such that

$$|w_{\varphi_1}(f_0)| \geq t \sup_{g \in \mathcal{D}} |w_g(f_0)|.$$

Generalization of WQOGA

Next, we find c_1 satisfying

$$w_{\varphi_1}(f - c_1\varphi_1) = 0.$$

Denote $f_1 := f_1^{p,t} := f - c_1\varphi_1$.

We continue this construction in an inductive way. Assume that we have already constructed residuals $f_0, f_1, \ldots, f_{m-1}$ and dictionary elements $\varphi_1, \ldots, \varphi_{m-1}$.

Generalization of WQOGA

Now, we pick an element $\varphi_m := \varphi_m^{p,t} \in \mathcal{D}$ such that

$$|w_{\varphi_m}(f_{m-1})| \geq t \sup_{g \in \mathcal{D}} |w_g(f_{m-1})|.$$

Next, we look for $c_1^m, \ldots, c_m^m$ satisfying

$$w_{\varphi_j}(f - \sum_{i=1}^{m} c_i^m \varphi_i) = 0, \quad j = 1, \ldots, m. \tag{3.3}$$

If there is no solution to (3.1) then we stop, otherwise we denote $f_m := f_m^{p,t} := f - \sum_{i=1}^{m} c_i^m \varphi_i$ with $c_1^m, \ldots, c_m^m$ satisfying (3.1).

Running WPGA

The following remark is an analog of Remark 3.1.

Remark 3.2. The system (3.3) has a unique solution if $\det ||w_{\varphi_j}(\varphi_i)||_{i,j=1}^{m} \neq 0$. We apply WPGA in the case of a pair $(\mathcal{D}, \mathcal{W})$ with the coherence parameter $M := M(\mathcal{D}, \mathcal{W})$. Then by a simple well known argument on the linear independence of the rows of the matrix $||w_{\varphi_j}(\varphi_i)||_{i,j=1}^{m}$ we conclude that (3.3) has a unique solution for any $m < 1 + (1-\delta)/M$. In this case we can run WPGA for at least $[(1-\delta)/M]$ iterations.

Property of WPGA

We begin with an auxiliary statement.

Lemma 3.1 Let $t \in (0,1]$. Assume that the pair $(\mathcal{D},\mathcal{W})$ has coherence parameter $M := M(\mathcal{D},\mathcal{W})$ and satisfies (3.2). Let $S < \frac{t}{1+t}(1+(1-\delta)/M)$. Then for any f of the form

$$f = \sum_{i=1}^{S} a_i \psi_i,$$

where ψ_i are distinct elements of $\mathcal{D}$, we have that $\varphi_1^{p,t} = \psi_j$ with some $j \in [1,S]$.

Exact recovery by WPGA

Theorem 3.2. Let $t \in (0,1]$. Assume that the pair $(\mathcal{D},\mathcal{W})$ has coherence parameter $M := M(\mathcal{D},\mathcal{W})$ and satisfies (3.2). Let $S < \frac{t}{1+t}(1+(1-\delta)/M)$. Then for any f of the form

$$f = \sum_{i=1}^{S} a_i \psi_i,$$

where ψ_i are distinct elements of $\mathcal{D}$, we have that $f_S^{p,t} = 0$.

Projection

As an analog of best m-term approximation consider the following projective best m-term approximation. For a set $\mathcal{D}_m = \{g_i\}_{i=1}^m$ of any m distinct elements of $\mathcal{D}$ we define projection $P_{\mathcal{D}_m}(f)$ as follows:

$$P_{\mathcal{D}_m}(f) = \sum_{i=1}^{m} c_i g_i$$

with c_i satisfying

$$F_{g_i}(f - P_{\mathcal{D}_m}(f)) = 0, \quad i = 1, 2, \ldots, m.$$

Projective best m-term approximation

In the case of incoherent dictionary such projection always exists and it is unique provided $m < 1 + 1/M(\mathcal{D})$. Define projective best m-term approximation as follows:

$$\sigma_m^q(f) := \inf_{\mathcal{D}_m} \|f - P_{\mathcal{D}_m}(f)\|.$$

It is clear that for any norm $\|\cdot\|_Y$ one has

$$\sigma_m(f)_Y \leq \sigma_m^q(f)_Y.$$

Lebesgue-type inequality

Introduce the following norm induced by the dictionary $\mathcal{D}$: for $f \in X$

$$\|f\|_{\mathcal{D}} := \sup_{g \in \mathcal{D}} |F_g(f)|.$$

The following Lebesgue-type inequality was obtained in [Savu, 2009]. For QOGA

$$\|f_m^q\|_{\mathcal{D}} \leq 4.5\sigma_m^q(f)_{\mathcal{D}}, \quad m < 1/(3M(\mathcal{D})). \tag{3.4}$$

Lebesgue-type inequality

It was proved in [Savu, 2009] that (3.4) implies the following Lebesgue-type inequality in the X-norm.

$$\|f_m^q\|_X \leq Cm\sigma_m^q(f)_X, \quad m < 1/(4M(\mathcal{D})).$$

In the case of Hilbert space it was proved in [Savu, 2009] that (3.4) implies for OGA

$$\|f_m\| \leq 9m^{1/2}\sigma_m(f).$$