Greedy Approximation

Vladimir Temlyakov

Greedy approximation with respect to bases

Introduction

Let $\Psi := \{\psi\}_{k=1}^{\infty}$ be an orthonormal basis for a Hilbert space H. For any $f \in H$ there is a convergent (in H) orthogonal expansion

$$f = \sum_{k=1}^{\infty} \langle f, \psi_k \rangle \psi_k.$$

A classical way of approximation of f is to take a partial sum

$$S_n(f, \Psi) := \sum_{k=1}^{n} \langle f, \psi_k \rangle \psi_k.$$

For the error we have

$$\|f - S_n(f, \Psi)\|^2 = \sum_{k=n+1}^{\infty} |\langle f, \psi_k \rangle|^2.$$

Best m-term approximation

In nonlinear approximation we use the m-term approximation

$$\sum_{k \in \Lambda} \langle f, \psi_k \rangle \psi_k, \quad |\Lambda| = m.$$

It is clear that the optimal (from the point of view of the error) choice of Λ is the set of m biggest in absolute value coefficients $\langle f, \psi_k \rangle$. We can realize this choice by picking the biggest coefficients one by one. This results in the reordering (greedy reordering) of the orthogonal expansion:

$$f = \sum_{i=1}^{\infty} \langle f, \psi_{k_i} \rangle \psi_{k_i}, \quad |\langle f, \psi_{k_1} \rangle| \geq |\langle f, \psi_{k_2} \rangle| \geq \dots.$$

Major questions

1. How to work in a Banach space X instead of a Hilbert space H with a basis Ψ?

2. Let instead of an orthonormal basis Ψ we have a redundant system $\mathcal{D}$. How to approximate with regard to $\mathcal{D}$?

3. How to measure efficiency of a concrete algorithm?

Greedy approximation

Let a Banach space X with a normalized basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$, $\|\psi_k\| = 1$, $k = 1, 2, \dots$, be given. We consider the following greedy algorithm that we call the Thresholding Greedy Algorithm (TGA). For a given element $f \in X$ we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f)\psi_k.$$

Let an element $f \in X$ be given. We call a permutation ρ, $\rho(j) = k_j$, $j = 1, 2, \dots,$ of the positive integers decreasing and write $\rho \in D(f)$ if

$$|c_{k_1}(f)| \geq |c_{k_2}(f)| \geq \dots \quad .$$

In the case of strict inequalities here $D(f)$ consists of only one permutation. We define the m-th greedy approximant of f with regard to the basis Ψ corresponding to a permutation $\rho \in D(f)$ by the formula

$$G_m(f, \Psi) := G_m(f, \Psi, \rho) := \sum_{j=1}^{m} c_{k_j}(f)\psi_{k_j}.$$

Greedy versus best

In order to understand the efficiency of this algorithm we compare its accuracy with the best possible

$$\sigma_m(f, \Psi) := \sigma_m(f, \Psi)_X := \inf_{c_k, \Lambda; |\Lambda| = m} \|f - \sum_{k \in \Lambda} c_k \psi_k\|_X,$$

when an approximant is a linear combination of m terms from Ψ. The best we can achieve with the algorithm G_m is

$$\|f - G_m(f, \Psi, \rho)\| = \sigma_m(f, \Psi),$$

or a little weaker: for all elements $f \in X$

$$\|f - G_m(f, \Psi, \rho)\| \leq G\sigma_m(f, \Psi) \tag{1.1}$$

with a constant $G = C(X, \Psi)$ independent of f and m.

1. Trigonometric system

We proved in [T., 1998] the following results.

Theorem 2.1. For each $f \in L_p(\mathbb{T}^d)$ we have

$$\|f - G_m(f, \mathcal{T})\|_p \leq (1 + 3m^{h(p)})\sigma_m(f, \mathcal{T})_p, \quad 1 \leq p \leq \infty,$$

where $h(p) := |1/2 - 1/p|$.

Remark 2.1. There is a positive absolute constant C such that for each m and $1 \leq p \leq \infty$ there exists a function $f \neq 0$ with the property

$$\|G_m(f, \mathcal{T})\|_p \geq Cm^{h(p)}\|f\|_p.$$

Open problems

1. Characterize bases Ψ in L_p, $1 \le p \le \infty$, such that

$$\|f - G_m(f,\Psi)\|_p \le C(p) m^{h(p)} \sigma_m(f,\Psi)_p, \quad 1 \le p \le \infty,$$

where $h(p) := |1/2 - 1/p|$.

2. More general version of the above problem. Let $v(m)$ be an increasing function on m. Characterize bases Ψ in X such that

$$\|f - G_m(f,\Psi)\|_X \le C(X) v(m) \sigma_m(f,\Psi)_X.$$

Proof of Theorem 2.1. We treat separately the two cases $1 \le p \le 2$ and $2 \le p \le \infty$. Before splitting into these two cases we prove one auxiliary statement for $1 \le p \le \infty$.

Lemma 2.1. Let $\Lambda \subset \mathbb{Z}^d$ be a finite subset with cardinality $|\Lambda| = m$. Then for the operator S_Λ defined on $L_1(\mathbb{T}^d)$ by

$$S_\Lambda(f) := \sum_{k \in \Lambda} \hat{f}(k) e^{i(k,x)},$$

we have for all $1 \le p \le \infty$,

$$\|S_\Lambda(f)\|_p \le m^{h(p)} \|f\|_p. \tag{2.1}$$

Proof. For a given linear operator A denote by $\|A\|_{a \to b}$ the norm of this operator as an operator from $L_a(\mathbb{T}^d)$ to $L_b(\mathbb{T}^d)$. Then it is obvious that

$$\|S_\Lambda\|_{2 \to 2} = 1. \tag{2.2}$$

Consider

$$\mathcal{D}_\Lambda(x) := \sum_{k \in \Lambda} e^{i(k,x)}, \tag{2.3}$$

then

$$S_\Lambda(f) = f * \mathcal{D}_\Lambda := (2\pi)^{-d} \int_{\mathbb{T}^d} f(x-y) \mathcal{D}_\Lambda(y) dy$$

and for $p = 1$, or ∞ we have

$$\|S_\Lambda\|_{p \to p} \le \|\mathcal{D}_\Lambda\|_1 \le \|\mathcal{D}_\Lambda\|_2 = m^{1/2}. \tag{2.4}$$

The relations (2.2) and (2.4) and the Riesz-Thorin theorem imply (2.1).

We return to the proof of Theorem 2.1. We begin with the proof in the case $2 \le p \le \infty$. Take any function $f \in L_p(\mathbb{T}^d)$. Let t_m be a trigonometric polynomial which realizes the best m-term approximation to f in $L_p(\mathbb{T}^d)$. Denote by Λ the set of frequencies of t_m, i.e. $\Lambda := \{k : \hat{t}_m(k) \ne 0\}$. Then $|\Lambda| \le m$. Denote by Λ' the set of frequencies of $G_m(f)$. Then $|\Lambda'| = m$. Let us use the representation

$$f - G_m(f) = f - S_{\Lambda'}(f) = f - S_\Lambda(f) + S_\Lambda(f) - S_{\Lambda'}(f).$$

From this representation we derive

$$\|f - G_m(f)\|_p \le \|f - S_\Lambda(f)\|_p + \|S_\Lambda(f) - S_{\Lambda'}(f)\|_p. \tag{2.5}$$

We use Lemma 2.1 to estimate the first term in the right hand side of (2.5)

$$\|f - S_\Lambda(f)\|_p = \|f - t_m - S_\Lambda(f - t_m)\|_p \le (1 + m^{h(p)}) \sigma_m(f). \tag{2.6}$$

In estimating the second term in (2.5) we use the well-known inequality $\|f\|_2 \le \|f\|_p$ for $2 \le p \le \infty$ and the following lemma.

Lemma 2.2. Let $\Lambda \subset \mathbb{Z}^d$ be a finite subset with cardinality $|\Lambda| = n$. Then, for $2 \le p \le \infty$, we have

$$\|S_\Lambda(f)\|_p \le n^{h(p)}\|f\|_2. \tag{2.7}$$

Proof. For $p = \infty$ we have

$$\|S_\Lambda(f)\|_\infty \le \sum_{k\in\Lambda} |\hat{f}(k)| \le n^{1/2}(\sum_{k\in\Lambda} |\hat{f}(k)|^2)^{1/2} \le n^{1/2}\|f\|_2. \tag{2.8}$$

For $2 < p < \infty$ we use (2.8) and the following well-known inequality

$$\|g\|_p \le \|g\|_2^{2/p}\|g\|_\infty^{1-2/p}.$$

We continue estimating $\|S_\Lambda(f) - S_{\Lambda'}(f)\|_p$. Using Lemma 2.2 we get

$$\|S_\Lambda(f) - S_{\Lambda'}(f)\|_p = \|S_{\Lambda\setminus\Lambda'}(f) - S_{\Lambda'\setminus\Lambda}(f)\|_p \le \tag{2.9}$$

$$\|S_{\Lambda\setminus\Lambda'}(f)\|_p + \|S_{\Lambda'\setminus\Lambda}(f)\|_p \le m^{h(p)}(\|S_{\Lambda\setminus\Lambda'}(f)\|_2 + \|S_{\Lambda'\setminus\Lambda}(f)\|_2).$$

The definition of Λ' and the relations $|\Lambda'| = m$, $|\Lambda| \le m$ imply

$$\|S_{\Lambda\setminus\Lambda'}(f)\|_2 \le \|S_{\Lambda'\setminus\Lambda}(f)\|_2. \tag{2.10}$$

Finally, we have

$$\|S_{\Lambda'\setminus\Lambda}(f)\|_2 \le \|f - S_\Lambda(f)\|_2 \le \|f - t_m\|_2 \le \|f - t_m\|_p = \sigma_m(f). \tag{2.11}$$

Combining the relations (2.9)–(2.11) we get

$$\|S_\Lambda(f) - S_{\Lambda'}(f)\|_p \le 2m^{h(p)}\sigma_m(f). \tag{2.12}$$

The relations (2.5), (2.6) and (2.12) result in

$$\|f - G_m(f)\|_p \le (1 + 3m^{h(p)})\sigma_m(f).$$

This completes the proof of Theorem 2.1 in the case $2 \le p \le \infty$. We proceed to the case $1 \le p \le 2$. We keep the notation of the previous case. We start again with the inequality (2.5). Next, the inequality (2.6) holds for $1 \le p \le 2$ because it is based on Lemma 2.1 which covers the whole range $1 \le p \le \infty$ of the parameter p. Thus, it remains to estimate $\|S_\Lambda(f) - S_{\Lambda'}(f)\|_p$. Using the inequality $\|f\|_p \le \|f\|_2$ we get

$$\|S_\Lambda(f) - S_{\Lambda'}(f)\|_p = \|S_{\Lambda\setminus\Lambda'}(f) - S_{\Lambda'\setminus\Lambda}(f)\|_p \le$$

$$\|S_{\Lambda\setminus\Lambda'}(f)\|_p + \|S_{\Lambda'\setminus\Lambda}(f)\|_p \le \|S_{\Lambda\setminus\Lambda'}(f)\|_2 + \|S_{\Lambda'\setminus\Lambda}(f)\|_2. \tag{2.13}$$

In order to estimate $\|S_{\Lambda'\setminus\Lambda}(f)\|_2$ we use the part of the Hausdorff-Young theorem which states that

$$\|(\hat{f}(k))_{k\in\mathbb{Z}^d}\|_{l_{p'}} \le \|f\|_p, \quad 1 \le p \le 2, \quad p' := \frac{p}{p-1}.$$

We have

$$\begin{aligned}\|S_{\Lambda'\setminus\Lambda}(f)\|_2 &= \|(\hat{f}(k))_{k\in\Lambda'\setminus\Lambda}\|_{l_2}\\ &\le |\Lambda' \setminus \Lambda|^{1/p-1/2}\|(\hat{f}(k))_{k\in\Lambda'\setminus\Lambda}\|_{l_{p'}}\\ &\le m^{h(p)}\|(\hat{f}(k) - \hat{t}_m(k))_{k\in\mathbb{Z}^d}\|_{l_{p'}}\\ &\le m^{h(p)}\|f - t_m\|_p = m^{h(p)}\sigma_m(f).\end{aligned} \tag{2.14}$$

Gathering (2.5), (2.6), (2.10), (2.13) and (2.14) we get

$$\|f - G_m(f)\|_p \le (1 + 3m^{h(p)})\sigma_m(f),$$

which completes the proof of Theorem 2.1.

Haar system

Denote $\mathcal{H}_p := \{H_k^p\}_{k=1}^\infty$ the Haar basis on $[0,1)$ normalized in $L_p(0,1)$: $H_1^p = 1$ on $[0,1)$ and for $k = 2^n + l$, $l = 1, 2, \dots, 2^n$, $n = 0, 1, \dots$

$$H_k^p = \begin{cases} 2^{n/p}, & x \in [(2l-2)2^{-n-1}, (2l-1)2^{-n-1}) \\ -2^{n/p}, & x \in [(2l-1)2^{-n-1}, 2l2^{-n-1}) \\ 0, & \text{otherwise.} \end{cases}$$

Haar basis is a greedy basis

The following theorem establishes the existence of greedy bases for $L_p(0,1)$, $1 < p < \infty$.

Theorem 1.2 (T., 1998). Let $1 < p < \infty$ and a basis Ψ be L_p-equivalent to the Haar basis $\mathcal{H}_p$. Then for any $f \in L_p(0,1)$ and any $\rho \in D(f)$ we have

$$\|f - G_m(f, \Psi, \rho)\|_{L_p} \le C(p, \Psi)\sigma_m(f, \Psi)_{L_p}$$

with a constant $C(p, \Psi)$ independent of f, ρ, and m.

L_p-equivalence

In this theorem we use the following definition of the L_p-equivalence. We say that $\Psi = \{\psi_k\}_{k=1}^\infty$ is L_p-equivalent to $\mathcal{H}_p = \{H_k^p\}_{k=1}^\infty$ if for any finite set K and any coefficients c_k, $k \in K$, we have

$$C_1(p, \Psi)\| \sum_{k \in K} c_k H_k^p \|_{L_p} \le \| \sum_{k \in K} c_k \psi_k \|_{L_p} \le C_2(p, \Psi)\| \sum_{k \in K} c_k H_k^p \|_{L_p}$$

with two positive constants $C_1(p, \Psi), C_2(p, \Psi)$ which may depend on p and Ψ. For sufficient conditions on Ψ to be L_p-equivalent to $\mathcal{H}_p$ see [Frazier, Jawerth, 1990] and [DeVore, Konyagin, T., 1998].

Definition of greedy basis

Definition 1.1. We call a basis Ψ greedy basis if for every $f \in X$ there exists a permutation $\rho \in D(f)$ such that (1.1) holds.

The following proposition has been proved in [Konyagin, T., 1999].

Proposition 1.1. If Ψ is a greedy basis then (1.1) holds for any permutation $\rho \in D(f)$.

Theorem 1.2 establishes that each basis Ψ which is L_p-equivalent to the univariate Haar basis $\mathcal{H}_p$ is a greedy basis for $L_p(0,1)$, $1 < p < \infty$. We note that in the case of Hilbert space each orthonormal basis is a greedy basis with a constant $G = 1$ (see (1.1)).

Unconditional basis

We give now the definitions of unconditional and democratic bases.

Definition 1.2. A basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$ of a Banach space X is said to be unconditional if for every choice of signs $\theta = \{\theta_k\}_{k=1}^{\infty}$, $\theta_k = 1$ or -1, $k = 1, 2, \dots$, the linear operator M_θ defined by

$$M_\theta(\sum_{k=1}^{\infty} a_k \psi_k) = \sum_{k=1}^{\infty} a_k \theta_k \psi_k$$

is a bounded operator from X into X.

Democratic basis

Definition 1.3. We say that a basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$ is a democratic basis if for any two finite sets of indices P and Q with the same cardinality $|P| = |Q|$ we have

$$\|\sum_{k\in P} \psi_k\| \le D\|\sum_{k\in Q} \psi_k\|$$

with a constant $D := D(X, \Psi)$ independent of P and Q.

Characterization

We proved in [Konyagin, T., 1999] the following theorem.

Theorem 1.3. A basis is greedy if and only if it is unconditional and democratic.

This theorem gives a characterization of greedy bases. Further investigations ([T., 1998], [Cohen, DeVore, Hochmuth, 2000], [Kerkyacharian, Picard, 2004], [Gribonval, Nielsen, 2001], [Kamont, T., 2004]) showed that the concept of greedy bases is very useful in direct and inverse theorems of nonlinear approximation and also in applications in statistics.

2. 2. Almost greedy bases

Let us discuss a question of weakening the property of a basis of being a greedy basis. We begin with a concept of quasi-greedy basis introduced in [Konyagin, T., 1999].

Definition 2.1. We call a basis Ψ quasi-greedy basis if for every $f \in X$ and every permutation $\rho \in D(f)$ we have

$$\|G_m(f, \Psi, \rho)\|_X \le C\|f\|_X \tag{2.1}$$

with a constant C independent of f, m, and ρ.

P. Wojtaszczyk, 2000, proved the following theorem.

Theorem 2.1. A basis Ψ is quasi-greedy if and only if for any $f \in X$ and any $\rho \in D(f)$ we have

$$\|f - G_m(f, \Psi, \rho)\| \to 0 \quad \text{as} \quad m \to \infty. \tag{2.2}$$

Expansional best m-term approximation

We proceed to an intermediate concept of almost greedy basis. This concept was introduced and studied in [Dilworth, Kalton, Kutzarova, T., 2003]. Let

$$f = \sum_{k=1}^{\infty} c_k(f)\psi_k.$$

We define the following expansional best m-term approximation of f

$$\tilde{\sigma}_m(f) := \tilde{\sigma}_m(f,\Psi) := \inf_{\Lambda,|\Lambda|=m} \|f - \sum_{k\in\Lambda} c_k(f)\psi_k\|.$$

It is clear that $\sigma_m(f,\Psi) \le \tilde{\sigma}_m(f,\Psi)$.

Definition of almost greedy basis

It is also clear that for an unconditional basis Ψ we have

$$\tilde{\sigma}_m(f,\Psi) \le C\sigma_m(f,\Psi).$$

Definition 2.2. We call a basis Ψ almost greedy basis if for every $f \in X$ there exists a permutation $\rho \in D(f)$ such that

$$\|f - G_m(f,\Psi,\rho)\|_X \le C\tilde{\sigma}_m(f,\Psi)_X \tag{2.3}$$

holds with a constant independent of f, m.

The following proposition follows from [Dilworth, Kalton, Kutzarova, T., 2003].

Proposition 2.1. If Ψ is an almost greedy basis then (2.3) holds for any permutation $\rho \in D(f)$.

Characterization

The following characterization of almost greedy bases has been obtained in [Dilworth, Kalton, Kutzarova, T., 2003].

Theorem 2.1. Suppose Ψ is a basis of a Banach space. The following are equivalent:

A. Ψ is almost greedy.

B. Ψ is quasi-greedy and democratic.

C. For any (respectively, every) $\lambda > 1$ there is a constant $C = C_\lambda$ such that

$$\|f - G_{[\lambda m]}(f,\Psi)\| \le C_\lambda \sigma_m(f,\Psi).$$

Relations

We have discussed the following bases.

1. Unconditional;
2. Democratic;
3. Quasi-greedy;
4. Greedy;
5. Almost greedy.

We have formulated the following relations.

Unconditional + Democratic = Greedy

Quasi-greedy + Democratic = Almost greedy
We formulate some relations between the above bases.

Unconditional implies Quasi-greedy

Quasi-greedy does not imply Unconditional

Unconditional does not imply Democratic

Democratic does not imply Unconditional

Greedy implies Almost greedy

Almost greedy does not imply Greedy

These properties follow from [Konyagin, T., 1999].

Greedy approximation with respect to bases II

3. Weak Greedy Algorithms

The following weak type greedy algorithm was considered in [T., 1998]. Let $t \in (0,1]$ be a fixed parameter. For a given basis Ψ and a given $f \in X$ denote $\Lambda_m(t)$ any set of m indices such that

$$\min_{k \in \Lambda_m(t)} |c_k(f,\Psi)| \geq t \max_{k \notin \Lambda_m(t)} |c_k(f,\Psi)| \tag{3.1}$$

and define

$$G_m^t(f,\Psi) := \sum_{k \in \Lambda_m(t)} c_k(f,\Psi)\psi_k.$$

We call it the Weak Thresholding Greedy Algorithm (WTGA) with the weakness parameter t.

Stability for greedy basis

It was proved in [T., 1998] that in the case of $X = L_p$, $1 < p < \infty$, and Ψ is the Haar system $\mathcal{H}_p$ normalized in L_p we have for any $f \in L_p$

$$\|f - G_m^t(f,\mathcal{H}_p)\|_{L_p} \leq C(p,t)\sigma_m(f,\mathcal{H}_p)_{L_p}. \tag{3.2}$$

We noted in [Konyagin, T., 2002] that the proof of (3.2) from [T., 1998] works for any greedy basis instead of the Haar system $\mathcal{H}_p$. Thus for any greedy basis Ψ of a Banach space X and any $t \in (0,1]$ we have for each $f \in X$

$$\|f - G_m^t(f,\Psi)\|_X \leq C(\Psi,t)\sigma_m(f,\Psi)_X.$$

This means that for greedy bases we have more flexibility in constructing near best m-term approximants.

Stability for almost greedy basis

The following theorem is essentially due to Wojtaszczyk, 2000.

Theorem 3.1. Let Ψ be a quasi-greedy basis for a Banach space X. Then for any fixed $t \in (0,1]$ we have for each $f \in X$ that

$$G^t_m(f,\Psi) \to f \quad \text{as} \quad m \to \infty.$$

Theorem 3.2. (Konyagin, T., 2002) Let Ψ be an almost greedy basis. Then for $t \in (0,1]$ we have for any m

$$\|f - G^t_m(f,\Psi)\| \le C(t)\tilde{\sigma}_m(f,\Psi).$$

Generalization of weak greedy algorithm

In the paper [Kamont and T., 2004] the following generalization of the above Weak Thresholding Greedy Algorithm has been studied. We call it the Weak Thresholding Greedy Algorithm (WTGA) with a weakness sequence $\tau = \{t_k\}$. Let a weakness sequence $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k \in [0,1]$, $k = 1, \dots$ be given. We define the WTGA by induction. We take an element $f \in X$ and at the first step we let

$$\Lambda_1(\tau) := \{n_1\}; \qquad G^\tau_1(f,\Psi) := c_{n_1}\psi_{n_1}$$

with n_1 any satisfying

$$|c_{n_1}| \ge t_1 \max_n |c_n|$$

where we denote for brevity $c_n := c_n(f,\Psi)$.

WTGA

Assume we have already defined

$$G^\tau_{m-1}(f,\Psi) := G^{X,\tau}_{m-1}(f,\Psi) := \sum_{n \in \Lambda_{m-1}(\tau)} c_n\psi_n.$$

Then at the mth step we define $\Lambda_m(\tau) := \Lambda_{m-1}(\tau) \cup \{n_m\}$

$$G^\tau_m(f,\Psi) := G^{X,\tau}_m(f,\Psi) := \sum_{n \in \Lambda_m(\tau)} c_n\psi_n$$

with $n_m \notin \Lambda_{m-1}(\tau)$ any satisfying

$$|c_{n_m}| \ge t_m \max_{n \notin \Lambda_{m-1}(\tau)} |c_n|.$$

The first form of Lebesgue inequality

It is known that in the case of greedy basis and $\tau = \{t\}$, $t \in (0,1]$ the $G^\tau_m(\cdot,\Psi)$ realizes near best m-term approximation. There are two natural ways of adapting (1.1) to the case of nongreedy basis or to the case of general weakness sequence. In the first way (see [T., 1998], [Wojtaszczyk, 2000], [Oswald, 2000]) we write (1.1) in the form

$$\|f - G^\tau_m(f,\Psi)\| \le C(m,\tau,\Psi)\sigma_m(f,\Psi)$$

and look for the best (in the sense of order) constant $C(m,\tau,\Psi)$ in the above Lebesgue type inequality.

Fundamental functions

We now formulate the correspoding results from [Kamont and T., 2004]. For a basis Ψ we define the fundamental function $\varphi(m)$ and the function $\phi(m)$. We also need the following function

$$\varphi^s(n) := \sup_{|A|=n} \|\sum_{k\in A}\psi_k\|.$$

Define

$$\varphi(m) := \sup_{n\le m} \varphi^s(n);$$

$$\phi(m) := \inf_{|A|=m} \|\sum_{k\in A}\psi_k\|.$$

Characteristics of a basis

We now introduce some characteristics of a basis with respect to a weakness sequence τ. For a subset $V \subseteq [1,m]$ of integers we define

$$\phi(\tau,m,V) := \inf_{\{k_i\}} \|\sum_{i\in V} t_i\psi_{k_i}\|$$

where inf is taken over all sets $\{k_i\}$ of different indices. For two integers $1 \le n \le m$ we define

$$\phi(\tau,m,n) := \inf_{|V|=n, V\subseteq[1,m]} \phi(\tau,m,V),$$

and finally

$$\mu(\tau,m) := \sup_{n\le m} \frac{\varphi^s(n)}{\phi(\tau,m,n)}.$$

Lebesgue type inequality I

The following result has been proved in [Kamont and T., 2004].

Theorem 3.3. Let Ψ be a normalized unconditional basis for X. Then we have

$$\|f - G^\tau_m(f,\Psi)\| \le C(\Psi)\mu(\tau,m)\sigma_m(f,\Psi).$$

In Theorem 3.3 we compare efficiency of $G^\tau_m(\cdot,\Psi)$ with $\sigma_m(\cdot,\Psi)$. It is known in approximation theory that sometimes it is convenient to compare efficiency of an approximating operator which is characterized by m parameters with best possible approximation corresponding to smaller number of parameters $n \le m$. We use this idea in approximation by the WTGA.

The second form of Lebesgue inequality

Let us discuss a setting (see [Kamont and T., 2004]) when we write (1.1) in the form

$$\|f - G^\tau_{v_m}(f,\Psi)\| \le C(\Psi)\sigma_m(f,\Psi)$$

and look for the best (in the sense of order) sequence $\{v_m\}$ that is determined by the weakness sequence τ and the basis Ψ. We need some more notation. Define

$$\phi(\tau,N) := \phi(\tau,N,[1,N]) = \inf_{k_1,\dots,k_N} \|\sum_{j=1}^N t_j\psi_{k_j}\|.$$

Lebesgue type inequality II

Assume that $\phi(\tau,N)\to\infty$ as $N\to\infty$ and denote v_m the smallest N satisfying

$$\phi(\tau,N) \ge 2\varphi(m).$$

There is the following Lebesgue type inequality in this case ([Kamont and T., 2004]).

Theorem 3.4. For any normalized unconditional basis Ψ we have

$$\|f - G^\tau_{v_m}(f,\Psi)\| \le C(\Psi)\sigma_m(f,\Psi).$$

Example

Let Ψ be a normalized basis for $L_p([0,1))$. For the space $L_p([0,1)^d)$ we define

$$\Psi^d := \Psi\times\cdots\times\Psi$$

(d times);

$$\psi_{\mathbf{n}}(x) := \psi_{n_1}(x_1)\cdots\psi_{n_d}(x_d),$$

$x = (x_1,\dots,x_d)$, $\mathbf{n} = (n_1,\dots,n_d)$. In [Kerkyacharian, Picard, and T., 2006] we proved the following theorem.

Tensor product of greedy bases

Theorem 3.5. Let $1<p<\infty$ and let Ψ be a greedy basis for $L_p([0,1))$. Then for any Λ, $|\Lambda| = m$, we have for $2\le p<\infty$

$$C^1_{p,d} m^{1/p}\min_{\mathbf{n}\in\Lambda}|c_{\mathbf{n}}| \le \|\sum_{\mathbf{n}\in\Lambda} c_{\mathbf{n}}\psi_{\mathbf{n}}\|_p \le C^2_{p,d} m^{1/p}(\log m)^{h(p,d)}\max_{\mathbf{n}\in\Lambda}|c_{\mathbf{n}}|,$$

and for $1<p\le 2$

$$C^3_{p,d} m^{1/p}(\log m)^{-h(p,d)}\min_{\mathbf{n}\in\Lambda}|c_{\mathbf{n}}| \le \|\sum_{\mathbf{n}\in\Lambda} c_{\mathbf{n}}\psi_{\mathbf{n}}\|_p \le C^4_{p,d} m^{1/p}\max_{\mathbf{n}\in\Lambda}|c_{\mathbf{n}}|$$

where $h(p,d) := (d-1)|1/2-1/p|$.

Lebesgue type inequality

As a corollary of Theorems 3.3 and 3.5 we obtain the following inequalities for a greedy basis Ψ (for $L_p([0,1))$)

$$\|f - G_m^{L_p}(f, \Psi^d)\|_p \le C(\Psi, d, p)(\log m)^{h(p,d)} \sigma_m(f, \Psi^d)_p,$$

$$\|f - G_{v_m}^{L_p}(f, \Psi^d)\|_p \le C(\Psi, d) \sigma_m(f, \Psi^d)_p,$$

for $1 < p < \infty$, $v_m \ge C(\Psi, p, d) m (\log m)^{ph(p,d)}$.

4. Redundant systems

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

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

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 $m \ge 1$, 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 $m \ge 1$ 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) $G_m(f, \mathcal{D}) := P_{H_m}(f)$, where $H_m := \text{span}(\varphi_1, \dots, \varphi_m)$;

(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 $H_m^t := \operatorname{span}(\varphi_1^{o,t}, \dots, \varphi_m^{o,t})$ 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 $t = 1$, $k = 1, 2, \dots$, 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

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

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 $\|\cdot\|_\infty$.

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

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

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

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

Further results

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

Theorem 4.1. 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 4.1. Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05M^{-2/3}$. Then we have

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

Theorem 4.2. 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 4.3. 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 for OGA

$$\|f_S^o\|^2 \le \sigma_{S-k}(f_k^o)^2 + 5MS\|f_k^o\|^2, \quad 0 \le k \le S. \tag{4.1}$$

and the following inequalities for PGA

$$\|f_S\|^2 \le \sigma_S(f)^2 + 7MS\|f\|^2. \tag{4.2}$$

It was pointed out in [Donoho, Elad, T., 2006] that the inequality $\|f_{[m \log m]}\| \le 24\sigma_m(f)$ 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 $m \le 0.05M^{-2/3}$ can be substantially weakened. It was shown in [T. and Zheltov, 2010] that it can be substantially weakened.

Theorem 4.4. 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 $20Mm^{1+\delta}2^{L(\delta)} \le 1$. Then we have

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

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

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

for $m \le (20M)^{-1}$. His proof is different from the proof of Theorem 4.4. It is much more technically involved.

We now demonstrate the use of the inequality (4.2). The following result on PGA from [Temlyakov and Zheltov, 2010] is a corollary of (4.2).

Theorem 4.5. 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 4.6 Let $\mathcal{D}$ be an M-coherent dictionary. The WOGA(t) recovers exactly any $f \in \Sigma_m(\mathcal{D})$ with $m < \frac{t}{1+t}(1 + M^{-1})$.

Greedy approximation with respect to bases III

5. Lebesgue-type inequalities in L_p

We present here some Lebesgue-type inequalities for greedy approximation with respect to a quasi-greedy basis from T., Yang and Ye (2011). The main result of this paper is the following Lebesgue-type inequality for greedy approximation with respect to a quasi-greedy basis in the L_p spaces.

Theorem 5.1. Let $1 < p < \infty$, $p \neq 2$, and let Ψ be a quasi-greedy basis of the L_p space. Then for each $f \in L_p$ we have

$$\|f - G_m(f, \Psi)\|_{L_p} \leq C(p, \Psi) m^{|1/2-1/p|} \sigma_m(f, \Psi)_{L_p}.$$

Quasi-greedy constant

In our study of quasi-greedy bases we need the following known Lemma 5.1. It will be convenient to define the quasi-greedy constant K to be the least constant such that

$$\|G_m(f)\| \leq K\|f\| \quad \text{and} \quad \|f - G_m(f)\| \leq K\|f\|, \quad f \in X.$$

Lemma 5.1. Suppose Ψ is a quasi-greedy basis with a quasi-greedy constant K. Then, for any numbers a_j and any finite set of indices P, we have

$$(2K)^{-2} \min_{j\in P} |a_j| \| \sum_{j\in P} \psi_j \| \leq \| \sum_{j\in P} a_j \psi_j \| \leq 2K \max_{j\in P} |a_j| \| \sum_{j\in P} \psi_j \|.$$

Characteristic of non-democracy

We will use the notation

$$a_n(f) := |c_{k_n}(f)|$$

for the decreasing rearrangement of the coefficients of f. For a set of indices Λ we define the corresponding partial sum as follows

$$S_\Lambda(f) := \sum_{k\in\Lambda} c_k(f)\psi_k.$$

We will often use the following assumption: There exists an increasing function $v(m) := v(m, \Psi)$ such that for any two sets of indices A and B, $|A| = |B| = m$ we have

$$\| \sum_{k\in A} \psi_k \| \leq v(m) \| \sum_{k\in B} \psi_k \|. \tag{5.1}$$

Simple inequality

We begin with a theorem for a Banach space X. Later on we will specify this theorem for the L_p spaces.

Theorem 5.2. Let Ψ be a quasi-greedy basis of X satisfying assumption (5.1) with the property: For any set of indices Λ

$$\|S_\Lambda(f)\| \leq w(|\Lambda|)\|f\|,$$

where $w(m)$ is an increasing function on m. Then for each $f \in X$

$$\|f - G_m(f)\| \leq (1 + 2w(m) + (2K)^3 v(m) w(m)) \sigma_m(f).$$

Proof. Let, for a given $\epsilon > 0$, a polynomial

$$p_m(f) = \sum_{k\in P} b_k\psi_k, \quad |P| = m,$$

satisfy the inequality

$$\|f - p_m(f)\| \le \sigma_m(f) + \epsilon. \tag{5.2}$$

Denote by Q the set of indices picked by the greedy algorithm after m iterations

$$G_m(f) = \sum_{k\in Q} c_k(f)\psi_k.$$

We use the representation

$$f - G_m(f) = f - S_Q(f) = f - S_P(f) + S_P(f) - S_Q(f). \tag{5.3}$$

First, we bound

$$\begin{aligned}\|f - S_P(f)\| &= \|f - p_m(f) - S_P(f - p_m(f)\| \\ &\le (1 + w(m))\|f - p_m(f)\|. \end{aligned}\tag{5.4}$$

Second, we write

$$\begin{aligned}\|S_P(f) - S_Q(f)\| &= \|S_{P\setminus Q}(f) - S_{Q\setminus P}(f)\| \\ &\le \|S_{P\setminus Q}(f)\| + \|S_{Q\setminus P}(f)\|. \end{aligned}\tag{5.5}$$

We begin with estimating the second term in the right side of (5.5)

$$\|S_{Q\setminus P}(f)\| = \|S_{Q\setminus P}(f - p_m(f))\| \le w(m)\|f - p_m(f)\|. \tag{5.6}$$

For the first term we have by Lemma 5.1

$$\begin{aligned}\|S_{P\setminus Q}(f)\| &\le 2K \max_{k\in P\setminus Q} |c_k(f)| \|\sum_{k\in P\setminus Q} \psi_k\| \\ &\le 2K \min_{k\in Q\setminus P} |c_k(f)| v(m) \|\sum_{k\in Q\setminus P} \psi_k\| \\ &\le (2K)^3 v(m)\|S_{Q\setminus P}(f)\|. \end{aligned}\tag{5.7}$$

Combining (5.2) – (5.7) we obtain

$$\|f - G_m(f)\| \le (1 + 2w(m) + (2K)^3 v(m)w(m))\sigma_m(f).$$

This completes the proof.

We define the following expansional best m-term approximation of f with regard to Ψ

$$\tilde\sigma_m(f) := \tilde\sigma_m(f,\Psi) := \inf_{|\Lambda|=m} \|f - \sum_{k\in\Lambda} c_k(f)\psi_k\|.$$

It is clear that $\sigma_m(f) \le \tilde\sigma_m(f)$. It is known that for an unconditional basis Ψ we have

$$\tilde\sigma_m(f,\Psi) \le C(\Psi, X)\sigma_m(f,\Psi).$$

Inequality in terms of $\tilde{\sigma}_m(f)$

Theorem 5.3. Let Ψ be a quasi-greedy basis of X satisfying assumption (5.1). Then for each $f \in X$

$$\|f - G_m(f)\| \le C(\Psi, X)v(m)\tilde{\sigma}_m(f).$$

Proof. Let, for a given $\epsilon > 0$, a set of indices B be such that $|B| = m$ and

$$\|f - S_B(f)\| \le \tilde{\sigma}_m(f) + \epsilon. \tag{5.8}$$

Let as above

$$G_m(f) = \sum_{k \in Q} c_k(f)\psi_k.$$

Then

$$\|f - G_m(f)\| \le \|f - S_B(f)\| + \|S_{B \setminus Q}(f)\| + \|S_{Q \setminus B}(f)\|. \tag{5.9}$$

Our assumption that Ψ is quasi-greedy gives

$$\|S_{Q\setminus B}(f)\| = \|S_{Q\setminus B}(f - S_B(f))\|$$
$$= \|G_{|Q\setminus B|}(f - S_B(f))\| \le K\|f - S_B(f)\|. \tag{5.10}$$

Combining (5.8) – (5.10) and using (5.7) we obtain

$$\|f - G_m(f)\| \le (1 + K + 8K^4 v(m))\tilde{\sigma}_m(f).$$

L_p spaces

We use the brief notation $\|\cdot\|_p := \|\cdot\|_{L_p}$. We will use the following theorem fromT., Yang and Ye (2010). For $p = 2$ Theorem 5.4 was proved in P. Wojtaszczyk (2000).

Theorem 5.4. Let $\Psi = \{\psi_k\}_{k=1}^\infty$ be a quasi-greedy basis of the L_p space, $1 < p < \infty$. Then for each $f \in X$ we have for $2 \le p < \infty$

$$C_1(p)\sup_n n^{1/p}a_n(f) \le \|f\|_p \le C_2(p)\sum_{n=1}^\infty n^{-1/2}a_n(f)$$

and for $1 < p \le 2$

$$C_3(p)\sup_n n^{1/2}a_n(f) \le \|f\|_p \le C_4(p)\sum_{n=1}^\infty n^{1/p-1}a_n(f).$$

The following theorem is a corollary of the above Theorem 5.4.

Theorem 5.5. Let Ψ be a quasi-greedy basis of the L_p space, $1 < p < 2$, $2 < p < \infty$. Then for any set of indices Λ

$$\|S_\Lambda(f)\|_p \le C(p)|\Lambda|^{h(p)}\|f\|_p, \quad h(p) := |1/p - 1/2|.$$

Proof. Let $m := |\Lambda|$. Using Theorem 5.4 we get for $1 < p < 2$

$$\|S_\Lambda(f)\|_p \le C_4(p)\sum_{n=1}^m n^{1/p-1}a_n(S_\Lambda(f))$$

$$= C_4(p) \sum_{n=1}^{m} n^{1/p-3/2}(n^{1/2}a_n(S_\Lambda(f)))$$
$$\le C_5(p)m^{1/p-1/2} \sup_n n^{1/2}a_n(f)$$
$$\le C_5(p)C_3(p)^{-1}m^{1/p-1/2}\|f\|_p.$$

Again using Theorem 5.4 we obtain for $2 < p < \infty$

$$\|S_\Lambda(f)\|_p \le C_2(p) \sum_{n=1}^{m} n^{-1/2}a_n(S_\Lambda(f))$$
$$= C_2(p) \sum_{n=1}^{m} n^{-1/2-1/p}(n^{1/p}a_n(S_\Lambda(f)))$$
$$\le C_2(p) \sum_{n=1}^{m} n^{-1/2-1/p}(n^{1/p}a_n(f))$$
$$\le C_6(p)m^{1/2-1/p} \sup_n n^{1/p}a_n(f)$$
$$\le C_6(p)C_1(p)^{-1}m^{1/2-1/p}\|f\|_p.$$

It is pointed out in T., Yang and Ye (2010) that Theorem 5.4 implies the following inequality for a quasi-greedy basis Ψ of L_p

$$v(m,\Psi) \le C(p)m^{h(p)}, \quad 1 < p < \infty. \tag{5.11}$$

Using inequality (5.11) in Theorem 5.3 we obtain the following Theorem 5.6.

Theorem 5.6. Let Ψ be a quasi-greedy basis of L_p, $1 < p < \infty$. Then for each $f \in L_p$

$$\|f - G_m(f)\|_p \le C(\Psi,p)m^{h(p)}\tilde{\sigma}_m(f), \quad h(p) := |1/2 - 1/p|.$$

We now give a proof of Theorem 5.1.

Proof. The first part of the proof goes along the lines of proof of Theorem 5.2. We use the notation from that proof. By Theorem 5.5 we obtain

$$w(m) \le C(p)m^{h(p)}. \tag{5.12}$$

Thus (5.4) gives

$$\|f - S_P(f)\|_p \le (1 + C(p)m^{h(p)})\|f - p_m(f)\|_p. \tag{5.13}$$

Next, using Theorem 5.4 and our assumption that Ψ is a quasi-greedy basis of L_p we obtain for $1 < p < 2$

$$\|S_{Q\setminus P}(f)\|_p = \|S_{Q\setminus P}(f - p_m(f))\|_p$$
$$\le C_4(p) \sum_{n=1}^{m} n^{1/p-1}a_n(S_{Q\setminus P}(f - p_m(f)))$$
$$\le C_4(p) \sum_{n=1}^{m} n^{1/p-1}a_n(G_{|Q\setminus P|}(f - p_m(f)))$$

$$= C_4(p)\sum_{n=1}^{m} n^{1/p-3/2}(n^{1/2}a_n(G_{|Q\setminus P|}(f-p_m(f))))$$
$$\le C_7(p)m^{1/p-1/2}\sup_n n^{1/2}a_n(G_{|Q\setminus P|}(f-p_m(f)))$$
$$\le C_7(p)C_3(p)^{-1}m^{1/p-1/2}\|G_{|Q\setminus P|}(f-p_m(f))\|_p$$
$$\le C_8(p)Km^{1/p-1/2}\|f-p_m(f)\|_p. \tag{5.14}$$

In the same way we treat the case $2 < p < \infty$.

$$\|S_{Q\setminus P}(f)\|_p = \|S_{Q\setminus P}(f-p_m(f))\|_p$$
$$\le C_2(p)\sum_{n=1}^{m} n^{-1/2}a_n(S_{Q\setminus P}(f-p_m(f)))$$
$$\le C_2(p)\sum_{n=1}^{m} n^{-1/2}a_n(G_{|Q\setminus P|}(f-p_m(f)))$$
$$= C_2(p)\sum_{n=1}^{m} n^{-1/2-1/p}(n^{1/p}a_n(G_{|Q\setminus P|}(f-p_m(f))))$$
$$\le C_9(p)m^{1/2-1/p}\sup_n n^{1/p}a_n(G_{|Q\setminus P|}(f-p_m(f)))$$
$$\le C_9(p)C_1(p)^{-1}m^{1/2-1/p}\|G_{|Q\setminus P|}(f-p_m(f))\|_p$$
$$\le C_{10}(p)Km^{1/p-1/2}\|f-p_m(f)\|_p. \tag{5.15}$$

For the $S_{P\setminus Q}(f)$ we have for $1 < p < 2$

$$\|S_{P\setminus Q}(f)\|_p \le C_4(p)\sum_{n=1}^{m} n^{1/p-1}a_n(S_{P\setminus Q}(f))$$
$$\le C_4(p)\sum_{n=1}^{m} n^{1/p-1}a_n(S_{Q\setminus P}(f))$$
$$= C_4(p)\sum_{n=1}^{m} n^{1/p-1}a_n(S_{Q\setminus P}(f-p_m(f)))$$

which has been estimated in (5.14)

$$\le C_8(p)Km^{1/p-1/2}\|f-p_m(f)\|_p. \tag{5.16}$$

In the same way we obtain the bound in the case $2 < p < \infty$

$$\|S_{P\setminus Q}(f)\|_p \le C_2(p)\sum_{n=1}^{m} n^{-1/2}a_n(S_{P\setminus Q}(f))$$
$$\le C_2(p)\sum_{n=1}^{m} n^{-1/2}a_n(S_{Q\setminus P}(f))$$
$$= C_2(p)\sum_{n=1}^{m} n^{-1/2}a_n(S_{Q\setminus P}(f-p_m(f)))$$

which has been estimated in (5.15)

$$\le C_{10}(p)Km^{1/2-1/p}\|f-p_m(f)\|_p. \tag{5.17}$$

Combining (5.13) – (5.17) we complete the proof of Theorem 5.1.

Book

V.N. Temlyakov, Greedy Approximation, Cambridge University Press, is supposed to appear by the end of 2011.

Lebesgue type inequalities for greedy approximation

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

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

where the c_g are real or complex numbers. For a function $f \in H$ we define its best m-term approximation error

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

Pure Greedy Algorithm (PGA)

Pure Greedy Algorithm (PGA) We define $f_0 := f$. Then for each $m \ge 1$, 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 $m \ge 1$ 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) $G_m(f, \mathcal{D}) := P_{H_m}(f), \quad \text{where} \quad H_m := \text{span}(\varphi_1, \dots, \varphi_m);$

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

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 $H_m^t := \operatorname{span}(\varphi_1^{o,t}, \dots, \varphi_m^{o,t})$ 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 $t = 1,\ k = 1, 2, \dots$, 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 $f \in H$

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

The Lebesgue inequality

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

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

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 $m \leq 0.05M^{-2/3}$. Then for $l \geq 1$ satisfying $2^l \leq \log m$ we have

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

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

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

Theorem 2.2. 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

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

and the following inequalities for PGA

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

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

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

It was pointed out in [Donoho, Elad, T., 2006] that the inequality $\|f_{[m\log m]}\| \le 24\sigma_m(f)$ 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 $m \le 0.05M^{-2/3}$ can be substantially weakened. It was shown in [T. and Zheltov, 2010] that it can be substantially weakened.

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 $20Mm^{1+\delta}2^{L(\delta)} \le 1$. Then we have

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

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

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

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

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

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

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

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

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

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

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

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 $c_1^m, \dots, c_m^m$ 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

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

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 $f_0, f_1, \dots, f_{m-1}$ and dictionary elements $\varphi_1, \dots, \varphi_{m-1}$.

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 $c_1^m, \dots, c_m^m$ satisfying

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

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 $c_1^m, \dots, c_m^m$ 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 $f_S^{p,t} = 0$.

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 theorems were proved in Savu and Temlyakov, 2011.

Theorem 3.3. Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $m \leq 1/(3M)$ we have for the QOGA

$$\|f_m\|_{\mathcal{D}} \leq 13.5\sigma_m(f)_{\mathcal{D}}.$$

Theorem 3.4. Assume that $\mathcal{D}$ is an M-coherent dictionary in a Banach space X. There exists an absolute constant C such that for $m \leq 1/(3M)$ we have for the QOGA

$$\|f_m\|_X \leq C \inf_{g\in\Sigma_m(\mathcal{D})}(\|f-g\|_X + m\|f-g\|_{\mathcal{D}}).$$

Corollary 3.1. Using the inequality $\|g\|_{\mathcal{D}} \leq \|g\|_X$ we obtain from Theorem 3.4

$$\|f_m\|_X \leq C(1+m)\sigma_m(f)_X.$$

4. Free Relaxation

Weak Greedy Algorithm with Free Relaxation (WGAFR) Let $t \in (0,1]$ be a weakness parameter. We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

(1) $\varphi_m \in \mathcal{D}$ is any satisfying

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

WGAFR

(2) Find w_m and λ_m such that

$$\|f - ((1-w_m)G_{m-1} + \lambda_m \varphi_m)\| =$$

$$\inf_{\lambda, w} \|f - ((1-w)G_{m-1} + \lambda \varphi_m)\|$$

and define

$$G_m := (1-w_m)G_{m-1} + \lambda_m \varphi_m.$$

(3) Denote

$$f_m := f - G_m.$$

Robustness to noise

Theorem 4.1 (T., 2006). Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from H such that

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon / A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have for the WOGA and for the WGAFR

$$\|f_m\| \leq \max\Big(2\epsilon, C(A(\epsilon) + \epsilon)(1 + mt^2)^{-1/2}\Big).$$

Replacing greedy step by thresholding

Let a weakness parameter $t \in (0,1]$ be given.

Modified Weak Orthogonal Greedy Algorithm (MWOGA).

We define $f_0 := f_0^{M,t} := f$. Then for each $m \geq 1$ we inductively define:

(1) $\varphi_m := \varphi_m^{M,t} \in \mathcal{D}$ is any element satisfying

$$|\langle f_{m-1}, \varphi_m \rangle| \geq t \|f_{m-1}\|^2.$$

MWOGA

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

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

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

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

Rate of convergence of MWOGA

Theorem 4.2. Let $\mathcal{D}$ be an arbitrary dictionary in H. Then for each $f \in A_1(\mathcal{D})$ we have

$$\|f - G_m^{M,t}(f,\mathcal{D})\| \le (1+mt^2)^{-1/2}.$$

We note that in a particular case $t = 1$ the right hand side is equal to $(1+m)^{-1/2}$.

GART

Greedy Algorithm with Relaxation and Thresholding (GART)

We define $f_0 := f$ and $G_0 := 0$. Then for a given parameter $\delta \in (0, 1/2]$ we inductively define for $m \ge 1$

(1) $\varphi_m \in \mathcal{D}$ is any satisfying

$$|\langle f_{m-1}, \varphi_m\rangle| \ge \delta \|f_{m-1}\|. \tag{1}$$

If there is no $\varphi_m \in \mathcal{D}$ satisfying (1) then we stop.

(2) Find w_m and λ_m such that

$$\|f - ((1-w_m)G_{m-1} + \lambda_m\varphi_m)\|$$
$$= \inf_{\lambda,w} \|f - ((1-w)G_{m-1} + \lambda\varphi_m)\|$$

and define

$$G_m := (1-w_m)G_{m-1} + \lambda_m\varphi_m.$$

(3) Denote

$$f_m := f - G_m.$$

If $\|f_m\| \le \delta\|f\|$ then we stop, otherwise we proceed to the $(m+1)$th iteration.

GART. Error bound

Theorem 4.3 (T., 2006). Take a number $\epsilon \ge 0$ and two elements f, f^ϵ from H such that

$$\|f - f^\epsilon\| \le \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then the GART will stop after $m \le C\delta^{-2}\ln(1/\delta)$ iterations with

$$\|f_m\| \le \epsilon + \delta A(\epsilon).$$