# Industrial Mathematics Institute

1997:11

# Greedy algorithms and m-term approximation with regard to redundant dictionaries

V. Temlyakov

IMI
Preprint Series

Department of Mathematics
University of South Carolina

# GREEDY ALGORITHMS AND $M$-TERM APPROXIMATION WITH REGARDS TO REDUNDANT DICTIONARIES[1]

V.N. Temlyakov

Department of Mathematics, University of South Carolina, Columbia, SC 29208

[1]This research was supported by the National Science Foundation Grant DMS 9622925 and by ONR Grant N00014-96-1-1003

# GREEDY ALGORITHMS

V.N. Temlyakov

Department of Mathematics, University of South Carolina, Columbia, SC 29208

Abstract

We study the efficiency of greedy type algorithms with regard to redundant dictionaries in Hilbert space. In Section 2 we prove a general result which gives a sufficient condition on a dictionary to guarantee that Pure Greedy Algorithm is near best in the sense of power decay of error of approximation. We discuss also some important examples in Section 2.

It is known (see [DT1]) that the Pure Greedy Algorithm for to some dictionaries has a saturation property. In Section 3 we construct an example which shows that a natural generalization of the Pure Greedy Algorithm also has a saturation property.

In Section 4 we discuss some new phenomena which occur in approximation by a greedy type algorithm with regards to a highly redundant dictionary.

## 1. Introduction

Nonlinear approximation is an important tool in many numerical algorithms. We consider in this paper one particular method of nonlinear approximation, namely, $m$-term approximation. The $m$-term approximation is used in image and signal processing as well as in the design of neural networks. One of the basic questions in nonlinear approximation is how to construct an algorithm which realizes best or near best approximation. This question was discussed in many papers for different settings of nonlinear approximation problem (see for instance [B], [DDGS1], [DDGS2], [DJP], [DMA], [DT1], [DT2], [J], [T1], [T2]). In this paper we present some recent results in studying the settings discussed in [DT1] and [DT2]. The major question we try to answer is: how does redundancy effect the efficiency of best $m$-term approximation and the efficiency of greedy type algorithms with regards to a given dictionary.

We shall confine ourselves to studying in this paper only approximation in Hilbert space. Let $H$ be a real, separable Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x \rangle^{1/2}$. We briefly recall some definitions and notations from [DT1] and [DT2]. We call a system $\mathcal{D}$ of elements (functions) from $H$ a *dictionary* if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$) and its linear span is dense in $H$.

We let $\Sigma_m(\mathcal{D})$ denote the collection of all functions in $H$ which can be expressed as a linear combination of at most $m$ elements of $\mathcal{D}$. Thus each function $s \in \Sigma_m := \Sigma_m(\mathcal{D})$ can be written in the form

$$s = \sum_{g \in \Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad |\Lambda| \le m, \tag{1.1}$$

with the $c_g \in \mathbb{R}$.

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

$$\sigma_m(f) := \sigma_m(f, \mathcal{D}) := \inf_{s \in \Sigma_m} \|f - s\|. \tag{1.2}$$

The quantity $\sigma_m(f, \mathcal{D})$ gives the best possible error of approximation of $f$ by a linear combination of $m$ elements from a given dictionary $\mathcal{D}$. We define now an algorithm (Pure Greedy Algorithm) which realizes the best $m$-term approximation in the particular case when $\mathcal{D}$ is an orthonormal basis for $H$.

We describe this algorithm for a general dictionary $\mathcal{D}$ (in which case it does not generally produce a best approximation). If $f \in H$, we let $g = g(f) \in \mathcal{D}$ be an element from $\mathcal{D}$ which maximizes $|\langle f, g \rangle|$. We shall assume for simplicity that such a maximizer exists; if not, some modifications are necessary in the algorithms that follow. We define

$$G(f) := G(f, \mathcal{D}) := \langle f, g \rangle g \tag{1.3}$$

and

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

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

$$\begin{aligned} G_m(f) &:= G_m(f, \mathcal{D}) := G_{m-1}(f) + G(R_{m-1}(f)) \\ R_m(f) &:= R_m(f, \mathcal{D}) := f - G_m(f) = R(R_{m-1}(f)). \end{aligned} \tag{1.4}$$

The above algorithm is greedy in the sense that at each iteration it approximates the residual $R_m(f)$ as best possible by a single function from $\mathcal{D}$. One of advantages of the Pure Greedy Algorithm is that it is simple – the repetition of one basic step.

In Section 2 we present some partial progress in the following general problem.

**Problem 1.1.** *Let $0 < r \le 1/2$ be given. Characterize dictionaries $\mathcal{D}$ which posses the property: for any $f \in H$ such that*

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

*we have*

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

We impose the restriction $r \le 1/2$ in Problem 1.1 because of the following result from [DT1]. We constructed in [DT1] a dictionary $\mathcal{D} = \{\varphi_k\}_{k=1}^{\infty}$ such that for the function $f = \varphi_1 + \varphi_2$ we have

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

It is clear that $\sigma_m(f, \mathcal{D}) = 0$ for $m \ge 2$. This example of dictionary shows that in general we cannot get better than $m^{-1/2}$ rate of approximation by the Pure Greedy Algorithm even if we impose extremely tough restrictions on $\sigma_m(f, \mathcal{D})$. We call this phenomenon a saturation property.

In Section 2 we give a sufficient condition on $\mathcal{D}$ to have the property formulated in Problem 1.1. We consider dictionaries which we call $\lambda$-quasiorthogonal.

**Definition 1.1.** *We say $\mathcal{D}$ is a $\lambda$-quasiorthogonal dictionary if for any $n \in \mathbb{N}$ and any $g_i \in \mathcal{D}$, $\quad i = 1, \dots, n$, there exists a collection $\varphi_j \in \mathcal{D}$, $\quad j = 1, \dots, M$, $\quad M \le N := \lambda n$, with the properties:*

$$g_i \in X_M := \operatorname{span}(\varphi_1, \dots, \varphi_M); \tag{1.5}$$

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

$$\max_{1 \le j \le M} |\langle f, \varphi_j \rangle| \ge N^{-1/2} \|f\|. \tag{1.6}$$

**Remark 1.1.** *It is clear that an orthonormal dictionary is a 1-quasiorthogonal dictionary.*

We shall prove in Section 2 the following theorem and its slight generalization on asymptotically $\lambda$-quasiorthogonal dictionary. Examples of asymptotically $\lambda$-quasiorthogonal dictionaries are also given in Section 2.

**Theorem 1.1.** *Let a given dictionary $\mathcal{D}$ be $\lambda$-quasiorthogonal and let $0 < r < (2\lambda)^{-1}$ be a real number. Then for any $f$ such that*

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

*we have*

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

In Section 3 we consider a generalization of the Pure Greedy Algorithm. We study the $n$-Greedy Algorithm which differs from the Pure Greedy Algorithm in the basic step: instead of finding a single element $g(f) \in \mathcal{D}$ with the largest projection of $f$ on it we are looking for $n$ elements $g_1(f), \dots, g_n(f) \in \mathcal{D}$ with the largest projection $G^n(f, \mathcal{D})$ of $f$ onto their span. It is clear that

$$\|f - G^n(f, \mathcal{D})\| \le \|f - G_n(f, \mathcal{D})\|. \tag{1.7}$$

However, we construct in Section 3 an example of a dictionary $\mathcal{D}$ and a nonzero function $f \in \Sigma_{6n}(\mathcal{D})$ such that

$$\|f - G^n_m(f, \mathcal{D})\| \ge C(nm)^{-1/2}\|f\|.$$

This relation implies that like the Pure Greedy Algorithm the $n$-Greedy Algorithm has a saturation property for (for details see Section 3).

Section 4 deals with approximation of functions in $L_2$. We consider the periodic one-variable case. In the linear theory of approximation there is a powerful discretization method which allows us to reduce an approximation problem for smooth functions to the corresponding problem in a finite dimensional subspace, for instance, in the space $\mathcal{T}(n)$ of trigonometric polynomials of degree $n$. In Section 4 we make an attempt to use the idea of discretization in the case of nonlinear approximation with regards to a highly redundant dictionary. The difficulty arises in studying nonlinear algorithms, for instance, Pure Greedy Algorithm. The standard way of studying a linear approximation problem for classes of smooth functions is the following. We expand a function $f$ into a series

$$f = \sum_s f_s,$$

and get some restrictions on $\|f_s\|$ from the assumption about smoothness of $f$. Then we deal with each $f_s$ separately and using the linearity of the operator under investigation we sum the corresponding errors. It is clear that this method does not work for a nonlinear algorithm. For instance, if we take a dictionary $\mathcal{D} = \{\varphi_k\}_{k=1}^\infty$ from Theorem 4.1 in [DT1] we have for $f = \varphi_1 + \varphi_2$

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

despite the relations

$$\varphi_i = G_1(\varphi_i, \mathcal{D}), \quad i = 1, 2.$$

In Section 4 we study among other problems the efficiency of the Pure Greedy Algorithm in the Hölder smoothness class $H^r_2$. We consider a highly redundant dictionary $\mathcal{TV}$ that consists of all trigonometric polynomials $t$ with $\|t\|_2 = 1$ and such that all nonzero Fourier coefficients of $t$ are of the same absolute value. We prove that redundancy helps very much in this particular case. We obtain an exponential decay of the error: for any $f \in H^r_2$ we have

$$\|f - G_m(f, \mathcal{TV})\|_2 \le C(r)e^{-A(rm)^{1/2}}$$

with absolute positive constant $A$.

## 2. Some special redundant dictionaries

In this section we prove Theorem 1.1 and discuss $\lambda$-quasiorthogonal dictionaries. We begin with a numerical lemma.

**Lemma 2.1.** *Let three positive numbers $\alpha < \gamma \le 1$, $A > 1$ be given and let a sequence of positive numbers $1 \ge a_1 \ge a_2 \ge \dots$ satisfy the condition: if for some $\nu \in \mathbb{N}$ we have*

$$a_\nu \ge A\nu^{-\alpha}$$

*then*

$$a_{\nu+1} \le a_\nu(1-\gamma/\nu). \tag{2.1}$$

*Then there exists $B = B(A,\alpha,\gamma)$ such that for all $n = 1,2,\dots$ we have*

$$a_n \le Bn^{-\alpha}.$$

*Proof.* We have $a_1 \le 1 < A$ which implies that the set

$$V := \{\nu : a_\nu \ge A\nu^{-\alpha}\}$$

does not contain $\nu = 1$. We prove now that for any segment $[n, n+k] \subset V$ we have $k \le C(\alpha,\gamma)n$. Indeed, let $n \ge 2$ be such that $n-1 \notin V$, which means

$$a_{n-1} < A(n-1)^{-\alpha}, \tag{2.2}$$

and $[n, n+k] \subset V$, which in turn means

$$a_{n+j} \ge A(n+j)^{-\alpha}, \quad j = 0,1,\dots,k. \tag{2.3}$$

Then by the condition (2.1) of the lemma we get

$$a_{n+k} \le a_n \prod_{\nu=n}^{n+k-1}(1-\gamma/\nu) \le a_{n-1}\prod_{\nu=n}^{n+k-1}(1-\gamma/\nu). \tag{2.4}$$

Combining (2.2) – (2.4) we obtain

$$(n+k)^{-\alpha} \le (n-1)^{-\alpha}\prod_{\nu=n}^{n+k-1}(1-\gamma/\nu). \tag{2.5}$$

Taking logarithms and using the inequalities

$$\ln(1-x) \le -x, \quad x \in [0,1);$$

$$\sum_{\nu=n}^{m-1}\nu^{-1} \ge \int_n^m x^{-1}dx = \ln(m/n),$$

we get from (2.5)

$$-\alpha \ln \frac{n+k}{n-1} \le \sum_{\nu=n}^{n+k-1} \ln(1-\gamma/\nu) \le -\sum_{\nu=n}^{n+k-1} \gamma/\nu \le -\gamma \ln \frac{n+k}{n}.$$

Hence

$$(\gamma-\alpha)\ln(n+k) \le (\gamma-\alpha)\ln n + \alpha \ln \frac{n}{n-1},$$

which implies

$$n+k \le 2^{\frac{\alpha}{\gamma-\alpha}} n$$

and

$$k \le C(\alpha,\gamma)n.$$

Let us take any $\mu \in \mathbb{N}$. If $\mu \notin V$ we have the desired inequality with $B = A$. Assume $\mu \in V$, and let $[n, n+k]$ be the maximal segment in $V$ containing $\mu$. Then

$$a_\mu \le a_{n-1} \le A(n-1)^{-\alpha} = A\mu^{-\alpha}\left(\frac{n-1}{\mu}\right)^{-\alpha}. \tag{2.6}$$

Using the inequality $k \le C(\alpha,\gamma)n$ proved above we get

$$\frac{\mu}{n-1} \le \frac{n+k}{n-1} \le C_1(\alpha,\gamma). \tag{2.7}$$

Substituting (2.7) into (2.6) we complete the proof of Lemma 2.1 with $B = AC_1(\alpha,\gamma)^\alpha$.
■

*Proof of Theorem 1.1..* Let $\nu(r,\lambda)$ be such that for $\nu > \nu(r,\lambda)$ we have

$$(\lambda(\nu+1))^{-1} \ge (r/2 + 3/(4\lambda))/\nu.$$

Take two positive numbers $C \ge \nu(r,\lambda)^r$ and $\kappa$ which will be chosen later.

We consider the sequence $a_\nu := 1$ for $\nu < \nu(r,\lambda)$ and $a_\nu := \|f_\nu\|^2, \quad \nu \ge \nu(r,\lambda)$, where

$$f_\nu := f - G_\nu(f,\mathcal{D}).$$

The assumption $\sigma_1(f,\mathcal{D}) \le 1$ implies

$$a_{\nu(r,\lambda)} := \|f_{\nu(r,\lambda)}\|^2 \le \|f_1\|^2 \le 1.$$

Let us assume that for some $\nu$ we have $a_\nu \ge C^2\nu^{-2r}$. We want to prove that for those same $\nu$ we have

$$a_{\nu+1} \le a_\nu(1-\gamma/\nu)$$

with some $\gamma > 2r$. We shall specify the numbers $C$ and $\kappa$ in this proof. The assumptions $C \ge \nu(r,\lambda)^r$ and $a_\nu \ge C^2\nu^{-2r}$ imply $\nu \ge \nu(r,\lambda)$ and $\|f_\nu\| \ge C\nu^{-r}$, or

$$\nu^{-r} \le C^{-1}\|f_\nu\|. \tag{2.8}$$

We know that $f_\nu$ has the form

$$f_\nu = f - \sum_{i=1}^{\nu} c_i \phi_i, \quad \phi_i \in \mathcal{D}, \quad i = 1, \dots, \nu.$$

Therefore, by the assumption of Theorem 1.1 we have

$$\sigma_{[(1+\kappa)\nu]+1}(f_\nu) \leq \sigma_{[\kappa\nu]+1}(f) < (\kappa\nu)^{-r},$$

where $[x]$ denotes the integer part of the number $x$. This inequality implies that there are $l := [(1+\kappa)\nu] + 1$ elements $g_1, \dots, g_l \in \mathcal{D}$ such that

$$\|f_\nu - \sum_{i=1}^{l} c_i g_i\| \leq (\kappa\nu)^{-r}. \tag{2.9}$$

Now we use the assumption that $\mathcal{D}$ is a $\lambda$-quasiorthogonal dictionary. We find $M \leq N = \lambda l$ elements $\varphi_j \in \mathcal{D}, \quad j = 1, \dots, M$, satisfying the properties (1.5) and (1.6). Denote by $u$ an orthogonal projection of $f_\nu$ onto $X_M = \operatorname{span}(\varphi_1, \dots \varphi_M)$ and set $v := f_\nu - u$. The property (1.5) and the inequality (2.9) imply

$$\|v\| \leq (\kappa\nu)^{-r},$$

and, therefore, by (2.8) we have

$$\|u\|^2 = \|f_\nu\|^2 - \|v\|^2 \geq \|f_\nu\|^2(1 - (C\kappa^r)^{-2}).$$

Making use of property (1.6) we get

$$\sup_{g \in \mathcal{D}} |\langle f_\nu, g\rangle| \geq \max_{1 \leq j \leq M} |\langle f_\nu, \varphi_j\rangle| = \max_{1 \leq j \leq M} |\langle u, \varphi_j\rangle| \geq N^{-1/2}\|u\|.$$

Hence,

$$\|f_{\nu+1}\|^2 \leq \|f_\nu\|^2 - \|u\|^2/N \leq \|f_\nu\|^2(1 - (1 - (C\kappa^r)^{-2})(\lambda([(1+\kappa)\nu] + 1))^{-1}).$$

It is clear that taking a small enough $\kappa > 0$ and a sufficiently large $C$ we can make for $\nu \geq \nu(r, \lambda)$

$$(1 - (C\kappa^r)^{-2})(\lambda([(1+\kappa)\nu] + 1))^{-1} \geq \gamma > 2r.$$

With the $C$ as chosen we get a sequence $\{a_\nu\}_{\nu=1}^\infty$ satisfying the hypotheses of Lemma 2.1 with $A = C^2, \quad \alpha = 2r, \quad \gamma > \alpha$. Applying Lemma 2.1 we obtain

$$\|f_n\| = a_n^{1/2} \leq C(r, \lambda) n^{-r}, \quad n = 1, 2, \dots \quad ,$$

which completes the proof of Theorem 1.1. ■

The above proof of Theorem 1.1 gives a slightly more general result, with a $\lambda$-quasiorthogonal dictionary replaced by an asymptoticaly $\lambda$-quasiorthogonal dictionary. We formulate the corresponding definition and statements.

**Definition 2.1.** *We say $\mathcal{D}$ is an asymptotically $\lambda$-quasiorthogonal dictionary if for any $n \in \mathbb{N}$ and any $g_i \in \mathcal{D}, \quad i = 1, \dots, n$, there exists a collection $\varphi_j \in \mathcal{D}, \quad j = 1, \dots, M, \quad M \le N(n)$, with the properties:*

$$\limsup_{n\to\infty} N(n)/n = \lambda;$$

$$g_i \in X_M := \operatorname{span}(\varphi_1, \dots, \varphi_M); \tag{1.5a}$$

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

$$\max_{1\le j\le M} |\langle f, \varphi_j\rangle| \ge N(n)^{-1/2}\|f\|. \tag{1.6a}$$

**Theorem 2.1.** *Let a given dictionary $\mathcal{D}$ be asymptotically $\lambda$-quasiorthogonal and let $0 < r < (2\lambda)^{-1}$ be a real number. Then for any $f$ such that*

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

*we have*

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

In the proof of this theorem we use the following Lemma 2.2 instead of Lemma 2.1.

**Lemma 2.2.** *Let four positive numbers $\alpha < \gamma \le 1$, $A > 1$, $\quad U \in \mathbb{N}$ be given and let a sequence of positive numbers $1 \ge a_1 \ge a_2 \ge \dots$ satisfy the condition: if for some $\nu \in \mathbb{N}, \quad \nu \ge U$ we have*

$$a_\nu \ge A\nu^{-\alpha}$$

*then*

$$a_{\nu+1} \le a_\nu(1 - \gamma/\nu).$$

*Then there exists $B = B(A, \alpha, \gamma, U)$ such that for all $n = 1, 2, \dots$ we have*

$$a_n \le Bn^{-\alpha}.$$

We proceed now to a discussion of $\lambda$-quasiorthogonal dictionaries.

**Proposition 2.1.** *Let a system $\{\varphi_1, \dots, \varphi_M\}$ and its linear span $X_M$ satisfy (1.6). If $M = N$ and $\dim X_M = N$, then $\{\varphi_j\}_{j=1}^N$ is an orthonormal system.*

*Proof.* Our proof is by contradiction. The system $\{\varphi_j\}_{j=1}^N$ is normalized and we assume that it is not orthogonal. Consider a system $\{v_j\}_{j=1}^N$ biorthogonal to $\{\varphi_j\}_{j=1}^N$ :

$$\langle \varphi_i, v_j\rangle = \delta_{i,j}, \quad 1 \le i, j \le N.$$

Our assumption implies that $\{v_j\}_{j=1}^N$ is also not orthogonal. Consider

$$u_j := v_j/\|v_j\|, \quad j = 1, 2, \dots N,$$

and form a vector

$$y_t := N^{-1/2} \sum_{i=1}^{N} r_i(t) u_i,$$

where the $r_n(t)$ are the Rademacher functions. Then for all $j = 1, 2, \dots, N$ and $t \in [0, 1]$ we have

$$|\langle y_t, \varphi_j \rangle| = N^{-1/2} |\langle u_j, \varphi_j \rangle| \leq N^{-1/2}; \tag{2.10}$$

and

$$\|y_t\|^2 = N^{-1} \sum_{i=1}^{N} \langle u_i, u_i \rangle + N^{-1} \sum_{i \neq j} r_i(t) r_j(t) \langle u_i, u_j \rangle =$$

$$= 1 + 2N^{-1} \sum_{1 \leq i < j \leq N} r_i(t) r_j(t) \langle u_i, u_j \rangle.$$

From this we get

$$\int_0^1 \|y_t\|^4 dt = 1 + 4N^{-2} \sum_{1 \leq i < j \leq N} |\langle u_i, u_j \rangle|^2 > 1.$$

This inequality implies that for some $t^*$ we have $\|y_{t^*}\| > 1$ and by (2.10) for this $t^*$ we get for all $1 \leq j \leq N$

$$|\langle y_{t^*}, \varphi_j \rangle| < N^{-1/2} \|y_{t^*}\|,$$

which contradicts (1.6).

**Definition 2.2.** *For given $\mu, \gamma \geq 1$ a dictionary $\mathcal{D}$ is called $(\mu, \gamma)$-semistable if for any $g_i \in \mathcal{D}$, $\quad i = 1, \dots, n$, there exist elements $h_j \in \mathcal{D}$, $\quad j = 1, \dots, M \leq \mu n$, such that*

$$g_i \in \operatorname{span}\{h_1, \dots, h_M\}$$

*and for any $c_1, \dots, c_M$ we have*

$$\| \sum_{j=1}^{M} c_j h_j \| \geq \gamma^{-1/2} \left( \sum_{j=1}^{M} c_j^2 \right)^{1/2}. \tag{2.11}$$

**Proposition 2.2.** *A $(\mu, \gamma)$-semistable dictionary $\mathcal{D}$ is $\mu\gamma$-quasiorthogonal.*

*Proof.* It is clear from (2.11) that $\{h_1, \dots, h_M\}$ are linearly independent. Let $\psi_1, \dots, \psi_M$ be the biorthogonal system to $\{h_1, \dots, h_M\}$. We shall derive from (2.11) that for any $a_1, \dots, a_M$ we have

$$\| \sum_{j=1}^{M} a_j \psi_j \| \leq \gamma^{1/2} \left( \sum_{j=1}^{M} a_j^2 \right)^{1/2}. \tag{2.12}$$

Indeed, using the representation

$$g = \sum_{j=1}^{M} c_j(g) h_j$$

and (2.11) we get

$$\|\sum_{j=1}^{M} a_j\psi_j\| = \sup_{\|g\|\le 1}\langle\sum_{j=1}^{M} a_j\psi_j, g\rangle = \sup_{\|g\|\le 1}\sum_{j=1}^{M} a_j c_j(g) \le$$

$$\le \sup_{\|(c_1,\dots,c_M)\|\le\gamma^{1/2}}\sum_{j=1}^{M} a_j c_j = \gamma^{1/2}\Big(\sum_{j=1}^{M} a_j^2\Big)^{1/2}.$$

Take any $f\in \operatorname{span}\{h_1,\dots,h_M\} = \operatorname{span}\{\psi_1,\dots,\psi_M\}$. Let

$$f = \sum_{j=1}^{M} a_j(f)\psi_j.$$

Then

$$\langle f, h_j\rangle = a_j(f)$$

The inequality (2.12) implies

$$\max_{1\le j\le M}|a_j(f)| \ge (\gamma M)^{-1/2}\|f\| \ge (\gamma\mu n)^{-1/2}\|f\|.$$

The proof of Proposition 2.2 is complete. ■

We give now two concrete examples of asymptotically $\lambda$-quasiorthogonal dictionaries.

**Example 2.1.** *The dictionary* $\chi := \{f = |J|^{-1/2}\chi_J, \quad J\subset[0,1)\}$ *where* $\chi_J$ *is the characteristic function of an interval* $J$ *is an asymptotically* 2*-quasiorthogonal dictionary.*

*Proof.* The statement of this example follows from Remark 1.1 and from the known simple Lemma 2.3.

**Lemma 2.3.** *For any system of intervals* $J_i\subset[0,1), \quad i=1,\dots,n,$ *there exists a system of disjoint intervals* $J_i^d\subset[0,1), \quad i=1,\dots,2n+1, \quad [0,1)=\cup_{i=1}^{2n+1}J_i^d,$ *such that each* $J_i$ *can be represented as a union of some* $J_j^d$.

*Proof.* Our proof is by induction. Let $n=1$ and $J_1=[a,b)$. Take $J_1^d=[0,a)$, $J_2^d=[a,b)$, and $J_3^d=[b,1)$. Assume now that the statement is true for $n-1$. Consider $n$ intervals $J_1,\dots,J_{n-1},J_n$. Let $J_j^d=[a_j,a_{j+1}), \quad j=1,\dots,2n-1$ be the disjoint system of intervals corresponding to $J_1,\dots,J_{n-1}$ and let $J_n=[a,b)$. Then for at most two intervals $J_k^d$ and $J_l^d$ we have $a\in J_k^d$ and $b\in J_l^d$. If $k=l$ we split $J_k^d$ into three intervals $[a_k,a),[a,b)$, and $[b,a_{k+1})$. If $k\ne l$ we split each $J_k^d$ and $J_l^d$ into two intervals $[a_k,a),[a,a_{k+1})$ and $[a_l,b),[b,a_{l+1})$. In both cases the total number of intervals is $2n+1$. ■

Another corollary of Lemma 2.3 can be formulated as follows.

**Example 2.2.** *The dictionary $\mathcal{P}(r)$ that consists of functions of the form $f = p\chi_J$, $\|f\| = 1$, where $p$ is an algebraic polynomial of degree $r-1$ and $\chi_J$ is the characteristic function of an interval $J$, is asymptotically $2r$-quasiorthogonal.*

Theorems 1.1 and 2.1 work for small smoothness $r < (2\lambda)^{-1}$. It is known (see [DT1], Theorem 4.1) that there are dictionaries which have the saturation property for the Pure Greedy Algorithm. Namely, there is a dictionary $\mathcal{D}$ such that

$$\sup_{f\in\Sigma_2(\mathcal{D})} \|f - G_m(f,\mathcal{D})\|/\|f\| \ge Cm^{-1/2}.$$

We shall prove that the dictionary $\chi$ from Example 2.1 does not have the saturation property.

**Theorem 2.2.** *For any $f \in \Sigma_n(\chi)$ we have*

$$\|f - G_m(f,\chi)\| \le \left(1 - \frac{1}{2n+1}\right)^{m/2} \|f\|.$$

*Proof.* We prove a variant of Theorem 2.2 for functions of the form

$$f = \sum_{j=1}^{n} c_j g_{I_j}, \quad \cup_{j=1}^{n} I_j = [0,1), \quad g_J := |J|^{-1/2}\chi_J, \tag{2.13}$$

where the $I_1,\dots,I_n$ are disjoint.

**Lemma 2.4.** *For any $f$ of the form (2.13) we have*

$$\|f - G_m(f,\chi)\| \le (1-1/n)^{m/2}\|f\|.$$

*Proof.* We begin with the following lemma.

**Lemma 2.5.** *Let $I^1 = [a,b)$ and $I^2 = [b,d)$ be two adjacent intervals. Assume that a function $f$ is integrable on $I^1$ and equals a constant $c$ on $I^2$. Then we have the inequality ($g_I := |I|^{-1/2}\chi_I$)*

$$|\langle f, g_J\rangle| \le \max(|\langle f, g_{I^1}\rangle|, |\langle f, g_{I^1\cup I^2}\rangle|) \tag{2.14}$$

*for any $J = [a,y)$, $b \le y \le d$. Moreover, if the right hand side in (2.14) is nonzero we have a strict inequality in (2.14) for all $b < y < d$.*

*Proof.* Denote

$$A := \int_{I^1} f(x)dx.$$

Then we have

$$\langle f, g_J\rangle = |J|^{-1/2}(A + \int_b^y c\,dx) = (|I^1| + y - b)^{-1/2}(A + c(y-b)),$$

hence

$$\langle f, g_J \rangle = \frac{P + cy}{(Q+y)^{1/2}}, \quad b \le y \le d,$$

where $P = A - cb$ and $Q = |I^1| - b$. Let $z = (Q+y)^{1/2}$. Then

$$\frac{P + cy}{(Q+y)^{1/2}} = (P + c(z^2 - Q))/z = (P - cQ)/z + cz =: F(z).$$

In the cases $P - cQ = 0, \quad c \ne 0$ or $P - cQ \ne 0, \quad c = 0$ the statement is trivial. It remains to consider the case $P - cQ \ne 0, \quad c \ne 0$. Assume $P - cQ < 0, \quad c > 0$. Then

$$F'(z) = -\frac{P - cQ}{z^2} + c > 0$$

and the statement is true. Assume $P - cQ > 0, \quad c > 0$. Then

$$F''(z) = 2\frac{P - cQ}{z^3} > 0, \quad z > 0.$$

It follows that $F(z) > 0$ is a convex function and the statement is also true. ■

We use this lemma to prove one more lemma.

**Lemma 2.6.** *For each function $f$ of the form (2.13) the $\max_J |\langle f, g_J \rangle|$ is attained on an interval $J^*$ of the form $J^* = \cup_{j=k}^{l} I_j$.*

*Proof.* The function

$$F(x,y) := (y-x)^{-1/2} \int_x^y f(t)dt, \quad 0 \le x < y \le 1; \quad F(x,x) = 0, \quad 0 \le x \le 1;$$

is continuous on $Y := \{(x,y) : 0 \le x \le y \le 1\}$ for any $f$ of the form (2.13). This implies the existence of $J^*$ such that

$$|\langle f, g_{J^*} \rangle| = \max_J |\langle f, g_J \rangle|. \tag{2.15}$$

Clearly, $|\langle f, g_{J^*} \rangle| > 0$ if $f$ is nontrivial. We complete the proof by contradiction. Assume $J^* = [a,t)$ and, for instance, $t$ is an interior point of $I_s = [b,d)$. Apply Lemma 2.5 with $I^1 = [a,b), I^2 = [b,d), J = J^*$. We get strict inequality which contradicts (2.15). Hence, $t$ is an endpoint of one of the intervals $I_j$. The same argument proves that $a$ is also an endpoint of one of the intervals $I_j$. This comletes the proof of Lemma 2.6. ■

Lemma 2.6 implies that for $f$ of the form (2.13) all $R_j(f)$ (see (1.4)) are also of the form (2.13). Next, for $f$ of the form (2.13) we have

$$\max_J |\langle f, g_J \rangle| \ge \max_{I_j} |\langle f, g_{I_j} \rangle| \ge n^{-1/2} \|f\|.$$

Consequently,

$$\|R_m(f)\|^2 \le (1 - 1/n)\|R_{m-1}(f)\|^2 \le \cdots \le (1 - 1/n)^m \|f\|^2,$$

which completes the proof of Lemma 2.4. ■

The statement of Theorem 2.2 follows from Lemma 2.4 and Lemma 2.3.

## 3. An example for the $n$-Greedy Algorithm

We consider in this section a generalization of the Pure Greedy Algorithm. Take a fixed number $n \in \mathbb{N}$ and define the basic step of the $n$-Greedy Algorithm as follows. Find an $n$-term polynomial

$$p_n(f) := p_n(f, \mathcal{D}) = \sum_{n=1}^{n} c_i g_i, \quad g_i \in \mathcal{D}, \quad i = 1, \dots, n,$$

such that (we assume its existence)

$$\|f - p_n(f)\| = \sigma_n(f, \mathcal{D}).$$

Denote

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

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

$$\begin{aligned} G_m^n(f) :&= G_m^n(f, \mathcal{D}) := G_{m-1}^n(f) + G^n(R_{m-1}^n(f)) \\ R_m^n(f) :&= R_m^n(f, \mathcal{D}) := f - G_m^n(f) = R^n(R_{m-1}^n(f)). \end{aligned} \tag{3.1}$$

It is clear that a 1-Greedy Algorithm is a Pure Greedy Algorithm.

For a general dictionary $\mathcal{D}$, and for any $0 < \tau \leq 1$, we define the class of functions

$$\mathcal{A}_\tau^o(\mathcal{D}, M) := \{f \in H : f = \sum_{k \in \Lambda} c_k w_k, \quad w_k \in \mathcal{D},\ |\Lambda| < \infty \text{ and } \sum_{k \in \Lambda} |c_k|^\tau \leq M^\tau\},$$

and we define $A_\tau(\mathcal{D}, M)$ as the closure (in $H$) of $A_\tau^o(\mathcal{D}, M)$. Furthermore, we define $A_\tau(\mathcal{D})$ as the union of the classes $A_\tau(\mathcal{D}, M)$ over all $M > 0$. For $f \in A_\tau(\mathcal{D})$, we define the "quasinorm"

$$|f|_{\mathcal{A}_\tau(\mathcal{D})}$$

as the smallest $M$ such that $f \in A_\tau(\mathcal{D}, M)$.

We prove in this section that the $n$-Greedy Algorithm, like the Pure Greedy Algorithm has a saturation property.

**Theorem 3.1.** *For any orthonormal basis $\{\varphi_k\}_{k=1}^\infty$ there exists an element $g$ such that for the dictionary $\mathcal{D} = g \cup \{\varphi_k\}_{k=1}^\infty$ there is an element $f$ which has the property: for any $0 < \tau \leq 1$*

$$\|f - G_m^n(f)\| / |f|_{\mathcal{A}_\tau(\mathcal{D})} \geq C(\tau) n^{-1/\tau} (m+2)^{-1/2}.$$

*Proof.* Let $n \geq 2$ be given. Define

$$g := A n^{-1/2} \sum_{k=1}^{2n} \varphi_k + 1/3 \sum_{k=3n}^{\infty} (k(k+1))^{-1/2} \varphi_k,$$

with

$$A := \left(\frac{1}{2} - \frac{1}{54n}\right)^{1/2} \geq (1/3)^{1/2}.$$

Then

$$\|g\|^2 = 2A^2 + 1/(27n) = 1.$$

Take

$$f := An^{-1/2} \sum_{k=1}^{3n-1} \varphi_k + 2/3 \sum_{k=3n}^{6n-1} (k(k+1))^{-1/2} \varphi_k.$$

**1. First step.** We prove that for the dictionary $\mathcal{D} = g \cup \{\varphi_k\}_{k=1}^{\infty}$ we have

$$G^n(f, \mathcal{D}) = u := g + An^{-1/2} \sum_{k=2n+1}^{3n-1} \varphi_k.$$

First of all, it is easy to check that $f - u$ is orthogonal to $g$ and $\varphi_k, \quad k = 1, \dots, 3n-1$, and

$$\|f - u\|^2 = 1/9 \sum_{k=3n}^{\infty} \frac{1}{k(k+1)} = \frac{1}{27n}.$$

We shall prove that

$$\sigma_n(f, \mathcal{D})^2 \geq \frac{1}{27n}$$

and that the only approximant which provides equality in this estimate is $u$.

**1).** Assume that $g$ is not among the approximating elements. Then for $\Phi = \{\varphi_k\}_{k=1}^{\infty}$ we have

$$\sigma_n(f, \Phi)^2 = A^2(2n-1)/n + (4/9)(1/6n) > \frac{1}{27n}.$$

**2).** Assume $g$ is among approximating elements; then we should estimate

$$\delta := \inf_a \sigma_{n-1}(f - ag, \Phi)^2.$$

Denote

$$g_s := \sum_{k=s}^{\infty} (k(k+1))^{-1/2} \varphi_k.$$

We have

$$f - ag = (1-a)An^{-1/2} \sum_{k=1}^{2n} \varphi_k + An^{-1/2} \sum_{k=2n+1}^{3n-1} \varphi_k + (2-a)(g_{3n} - g_{6n})/3 - ag_{6n}/3.$$

If $|1 - a| \geq 1$ then

$$\sigma_{n-1}(f - ag, \Phi)^2 \geq (1-a)^2 A^2 > \frac{1}{27n}.$$

It remains to consider $0 < a < 2$. In this case the $n-1$ largest in absolute value coefficients of $f - ag$ are those of $\varphi_k, \quad k = 2n+1, \dots, 3n-1$. We have

$$\sigma_{n-1}(f - ag, \Phi)^2 = 2(1-a)^2 A^2 + ((2-a)^2 + a^2)/(54n). \tag{3.2}$$

It is clear that the right hand side of (3.2) is greater than or equal to $1/(27n)$ for all $a$, and equals $1/(27n)$ only for $a = 1$. This implies that the best $n$-term approximant to $f$ with regards to $\mathcal{D}$ is unique and coincides with $u$. This concludes the first step.

After the first step we get

$$f_1 := R^n(f) = (g_{3n} - 2g_{6n})/3.$$

**2. General step.** We prove now the following lemma.

**Lemma 3.1.** *Consider*

$$h_s := 1/3 \sum_{k=s}^{\infty} e_k (k(k+1))^{-1/2} \varphi_k, \quad e_k = \pm 1, \quad s \geq 3n.$$

*We have*

$$\sigma_n(h_s, \mathcal{D})^2 = 1/(9(s+n)),$$

*and the best $n$-term approximant with regards to $\mathcal{D}$ is unique and equal to*

$$v_n := 1/3 \sum_{k=s}^{s+n-1} e_k (k(k+1))^{-1/2} \varphi_k.$$

*Proof.* It is easy to verify that

$$\|h_s - v_n\|^2 = 1/(9(s+n)),$$

and that $v_n$ is the unique best $n$-term approximant with regard to $\Phi$. We prove now that for each $a$ we have

$$\sigma_{n-1}(h_s - ag, \Phi)^2 > 1/(9(s+n)).$$

We use the representation

$$h_s - ag = -aAn^{-1/2} \sum_{k=1}^{2n} \varphi_k - a/3 \sum_{k=3n}^{s-1} (k(k+1))^{-1/2} \varphi_k$$
$$+ 1/3 \sum_{k=s}^{\infty} (e_k - a)(k(k+1))^{-1/2} \varphi_k.$$

Let us assume that an $(n-1)$-term approximant to $h_s - ag$ with regards to $\Phi$ consists of $\mu$, $0 \le \mu \le n-1$, elements with indices $k \ge s$ and $n-1-\mu$, with indices $k < s$. Then for the error $e(a,\mu)$ of this approximation we get

$$e(a,\mu)^2 \ge a^2 A^2 (n+\mu+1)/n + a^2(1/(3n) - 1/s)/9 + (1-|a|)^2/(9(s+\mu)). \quad (3.3)$$

Taking into account that

$$\inf_a \sigma_{n-1}(h_s - ag, \Phi)^2 = \inf_{0\le\mu\le n-1} \inf_a e(a,\mu)^2$$

we conclude that we need to prove the corresponding lower estimate for the right hand side of (3.3) for all $\mu$ and $a$. We have

$$e(a,\mu)^2 \ge a^2/3 + (1-|a|)^2/(9(s+\mu)) \ge a^2/3 + (1-|a|)^2/(9(s+n-1)). \quad (3.4)$$

We use now the following simple relation: for $b, c > 0$ we have

$$\inf_a (a^2 b + (1-a)^2 c) = \frac{bc}{b+c} = c(1+c/b)^{-1}. \quad (3.5)$$

Specifying $b = 1/3$ and $c = 1/(9(s+n-1))$ we get for all $a$ and $\mu$

$$e(a,\mu)^2 \ge (9(s+n)-6)^{-1} > (9(s+n))^{-1}.$$

Lemma 3.1 is proved. ■

Applying Lemma 3.1 to the second step and to the following steps we obtain that

$$R_m^n(f) = 1/3 \sum_{k=3n+n(m-1)}^{\infty} e_k (k(k+1))^{-1/2} \varphi_k$$

and

$$\|R_m^n(f)\|^2 = 1/(9n(m+2)).$$

This relation and the estimate $\|f\| \le C$ imply (1.7) from Section 1.

In order to complete the proof of Theorem 3.1 it remains to note that

$$|f|_{\mathcal{A}_\tau(\mathcal{D})} \le C(\tau) n^{1/\tau - 1/2}.$$

■

## 4. Some examples of highly redundant dictionaries

In Sections 2 and 3 we studied dictionaries which differ only slighly from an orthonormal dictionary. It is clear that if $\mathcal{D}_1 \subset \mathcal{D}_2$ then for any $f$ we have

$$\sigma_m(f, \mathcal{D}_2) \le \sigma_m(f, \mathcal{D}_1).$$

However, the example of Section 3 shows that even a slight perturbation of an orthonormal dictionary can result in a dramatic change of efficiency of the corresponding greedy type algorithm.

In this section we consider some dictionaries that are far from orthogonal dictionaries. In order to help the reader we formulate several statements on approximation in $\mathbb{R}^n$ which are corollaries of the corresponding results in [DT2]. We shall use these results later in this section.

Let $B_2^n$ denote the unit Euclidean ball in $\mathbb{R}^n$. For a dictionary $\mathcal{D}$ and a set $F \subset \mathbb{R}^n$ we define

$$\sigma_m(F, \mathcal{D}) := \sup_{f\in F} \sigma_m(f, \mathcal{D}); \quad G_m(F, \mathcal{D}) := \sup_{f\in F} \|f - G_m(f, \mathcal{D})\|.$$

**Theorem 4.1.** *For any $\mathcal{D}$ in $\mathbb{R}^n$ with $|\mathcal{D}| = N$ we have*

$$\sigma_m(B_2^n, \mathcal{D}) \geq CN^{-\frac{m}{n-m}}, \quad m \leq n/2.$$

See [DT2], Corollary 2.2.

**Theorem 4.2.** *For any $N$ there exists a system $\mathcal{D}$, $|\mathcal{D}| = N$, such that*

$$\sigma_m(B_2^n, \mathcal{D}) \leq \min(1, (2(N^{1/n} - 1)^{-1})^m).$$

See [DT2], Theorem 3.1. Consider the system

$$\mathcal{V} := \{g = y/\|y\|_2, \quad y = (y_1, \dots, y_n) \neq (0, \dots, 0), \quad y_j = -1, 0, 1, \quad j = 1, \dots, n\}.$$

**Theorem 4.3.** *We have*

$$\sigma_m(B_2^n, \mathcal{V}) \leq n^{1/2} 3^{-m}.$$

See [DT2], Theorem 4.1.

**Theorem 4.4.** *We have the estimate*

$$G_m(B_2^n, \mathcal{V}) \leq \left(1 - \frac{1}{1 + \ln n}\right)^{m/2}.$$

See [DT2], Theorem 7.1.

**Theorem 4.5.** *For any $m \leq 3(1 + \ln n)/16$ we have*

$$G_m(B_2^n, \mathcal{V}) \geq 1/2.$$

See [DT2], Theorem 7.2.

In this section we are going to discuss some applications of the results about $m$-term approximation in $\mathbb{R}^n$ to approximation of functions. For simplicity of notation we consider approximation of functions of a single variable. Denote by $\mathcal{T}(n)$ the set of real trigonometric polynomials

$$t(x) = \sum_{k=0}^{n} (a_k(t) c_k(x) + b_k(t) s_k(x))$$

where $c_k(x) := \cos kx, \quad s_k(x) := \sin kx$ for $k = 1, 2, \dots$ and $c_0(x) := 1/2, \quad s_0(x) \equiv 0$. We set up a one-to-one correspondence between $\mathcal{T}(n)$ and $\mathbb{R}^{2n+1}$. Define $T_n : \mathbb{R}^{2n+1} \to \mathcal{T}(n)$ by

$$T_n(y_0, \dots, y_{2n}) = \sum_{k=0}^{n} (y_{2k} c_k(x) + y_{2k-1} s_k(x)),$$

where the term $y_{-1} s_0(x)$ disappears because $s_0(x) \equiv 0$. We keep this term for notational convenience. Considering the standard $L_2$-norm in $\mathcal{T}(n)$

$$\|t\|_2^2 := \|t\|_{L_2}^2 := \frac{1}{\pi} \int_0^{2\pi} |t(x)|^2 dx$$

we get by Parseval's Identity

$$\|T_n(y)\|_{L_2} = \|y\|_{l_2}.$$

The above standard construction allows us to reformulate the $l_2$ results in $\mathbb{R}^{2n+1}$ as the corresponding $L_2$ results in $\mathcal{T}(n)$. For example, Theorem 4.2 takes the form

**Theorem 4.6.** *For any $N$ there exists a system $\mathcal{D}, |\mathcal{D}| = N$, of trigonometric polynomials in $\mathcal{T}(n)$ such that for any $t \in \mathcal{T}(n)$ we have*

$$\sigma_m(t, \mathcal{D})_2 \le \min(1, (2(N^{1/(2n+1)} - 1)^{-1})^m)\|t\|_2.$$

Let us take an arbitrary increasing sequence $\{n_k\}_{k=1}^{\infty}$ of natural numbers and consider the sequence $N_k = 4^{2n_k+1}, \quad k = 1, 2, \dots$. Denote by $\mathcal{Q}(n_k)$ a system with $|\mathcal{Q}(n_k)| = N_k$ which is provided by Theorem 4.6. Then we have for any $t \in \mathcal{T}(n_k)$

$$\sigma_m(t, \mathcal{Q}(n_k))_2 \le (2/3)^m\|t\|_2, \quad m = 1, 2, \dots. \tag{4.1}$$

Consider the following system in $L_2$

$$\mathcal{Q} := \cup_{k=1}^{\infty}\mathcal{Q}(n_k).$$

It turns out that this system is good for approximation of functions in $L_2$ regardless of their smoothness.

**Proposition 4.1.** *For each function $f \in L_2$ and any $\epsilon > 0$ there exists $g \in \mathcal{Q}$ such that*

$$\|f - \langle f, g\rangle g\|_2 \le (2/3 + \epsilon)\|f\|_2.$$

*Proof.* Denote by $S_n$ the orthogonal projector onto $\mathcal{T}(n)$, i.e. $S_n(f)$ is the $n$-th Fourier sum of $f$. Find $k$ such that

$$\|f - S_{n_k}(f)\|_2 \le \epsilon\|f\|_2. \tag{4.2}$$

By (4.1) with $m = 1$ we find $g \in \mathcal{Q}(n_k)$ such that

$$\|S_{n_k}(f) - \langle S_{n_k}(f), g\rangle g\|_2 \le 2/3\|S_{n_k}(f)\|_2. \tag{4.3}$$

Approximate now $f$ by $\langle f, g\rangle g$. Denoting $U_n(f) := f - S_n(f)$ we get

$$\|f - \langle f, g\rangle g\|_2 = \|S_{n_k}(f) - \langle S_{n_k}(f), g\rangle g + U_{n_k}(f)\|_2 \le$$

$$\|S_{n_k}(f) - \langle S_{n_k}(f), g\rangle g\|_2 + \|U_{n_k}(f)\|_2 \le (2/3 + \epsilon)\|f\|_2,$$

which proves Proposition 4.1. ■

We say that a system $\mathcal{D}$ admits a Greedy type $q$-fast, $0 < q < 1$, algorithm if for each $f \in L_2$ we can find $g \in \mathcal{D}$ such that

$$\|f - \langle f, g\rangle g\|_2 \le q\|f\|_2.$$

Proposition 4.1 shows that for any $\epsilon > 0$ the system $\mathcal{Q}$ admits the Greedy type $(2/3 + \epsilon)$-fast algorithm. In particular, this implies

$$\sigma_m(f, \mathcal{Q})_2 \le (2/3)^m\|f\|_2.$$

Let us consider now one special simply defined system in $L_2$. Denote by $\mathcal{TV}$ the set of all trigonometric polynomials $t$, $\|t\|_2 = 1$, whose non-zero Fourier coefficients are equal in absolute value. The restriction of this system onto $\mathcal{T}(n)$ will be denoted $\mathcal{TV}(n)$. It is easy to see that the system $\mathcal{TV}(n)$ coincides with $T_n(\mathcal{V})$ with $\mathcal{V}$ defined for $\mathbb{R}^{2n+1}$. Recall that $\mathcal{V}$ was defined in the beginning of this section and its cardinality (in $\mathbb{R}^{2n+1}$) is $3^{2n+1} - 1$. Note that the above described system is not as big as the system $\mathcal{Q}$ is. We prove some results for $\mathcal{TV}$ which are qualitatively different from those for $\mathcal{Q}$.

**Proposition 4.2.** *For any $0 < q < 1$ the system $\mathcal{TV}$ does not admit a Greedy type $q$-fast algorithm.*

*Proof.* The statement of Proposition 4.2 can be derived from the following example which was constructed in the proof of Theorem 7.2 in [DT2]. Fix $n$ and consider

$$f = \sum_{k=1}^{n} z_k c_k(x),$$

where $z = (z_1, \dots, z_n)$ is defined as follows $z_1 := 1, z_k := k^{1/2} - (k-1)^{1/2}$, $k = 2, 3, \dots, n$. Then

$$\|z\|_2^2 = 1 + \sum_{k=2}^{n}(k^{1/2} - (k-1)^{1/2})^2 \geq 1 + \sum_{k=2}^{n}(\frac{1}{2k^{1/2}})^2 =$$

$$1 + \frac{1}{4}\sum_{k=2}^{n}\frac{1}{k} \geq 1 + \frac{1}{4}\int_2^{n+1}\frac{dx}{x} \geq \frac{1}{4}(1 + \ln n);$$

and for each $l \leq n$,

$$\sum_{k=1}^{l} z_k = l^{1/2},$$

which implies that for each $g \in \mathcal{TV}$ we have

$$|\langle f, g\rangle| \leq 1.$$

Therefore, for any $g \in \mathcal{TV}$ we have

$$\|f - \langle f, g\rangle g\|_2^2 = \|f\|_2^2 - \langle f, g\rangle^2 \geq (1 - 4/(1 + \ln n))\|f\|_2^2.$$

Taking $n$ such that $1 - 4/(1 + \ln n) > q$ completes the proof. ■

We study the efficiency of $\mathcal{TV}$ for classes of smooth functions. Define $H_2^r, \quad r > 0$, as the class of functions $f \in L_2$ which allow a representation

$$f(x) = \sum_{s=1}^{\infty} t_s(x), \qquad t_s \in \mathcal{T}(2^s), \quad \|t_s\|_2 \leq 2^{-rs}, \quad s = 1, 2, \dots.$$

**Theorem 4.7.** *There exist two absolute positive constants $A_1$ and $A_2$ such that*

$$C_1(r)e^{-A_1 m} \leq \sigma_m(H_2^r, \mathcal{TV})_2 \leq C_2(r)e^{-A_2 \rho m},$$

*where $\rho := \min(r, 1/2)$.*

*Proof.* Let us begin with the lower estimate. It is clear that for $f \in \mathcal{T}(n)$ we have

$$\sigma_m(f, \mathcal{TV})_2 = \sigma_m(f, \mathcal{TV}(n))_2.$$

Next, the set of trigonometric polynomials $t \in \mathcal{TV}(n)$ satisfying $\|t\|_2 \le (2n)^{-r}$ is embedded into $H_2^r$. Denote by $\mathcal{T}(n)_2$ the unit $L_2$-ball in $\mathcal{T}(n)$. For a given $m$ take $n = m$ and use Theorem 4.1 for $\mathbb{R}^{2n+1}$. This gives

$$\sigma_m(H_2^r, \mathcal{TV})_2 \ge \sigma_m(\mathcal{T}(m)_2, \mathcal{TV}(m))_2(2m)^{-r} = (2m)^{-r}\sigma_m(B_2^{2m+1}, \mathcal{V}) \ge$$

$$C(2m)^{-r}3^{-(2m+1)m/(m+1)} \ge C_1(r)e^{-A_1 m}.$$

We proceed to the upper estimate. For a fixed $n$ of the form $n = 2^l$ we represent $f$ in the form $f = S_n(f) + U_n(f)$ and get from the definition of the class $H_2^r$

$$\|U_n(f)\|_2 \le \sum_{s=l+1}^{\infty} 2^{-rs} \le C(r)2^{-rl},$$

and

$$\|S_n(f)\|_2 \le \|f\|_2 \le C(r).$$

We approximate $S_n(f)$ using Theorem 4.3. We get

$$\sigma_m(S_n(f), \mathcal{TV}(n))_2 \le C(r)(2n+1)^{1/2}3^{-m}.$$

Selecting $n$ such that

$$n^{-r} \asymp n^{1/2}3^{-m}$$

we obtain

$$\sigma_m(H_2^r, \mathcal{TV})_2 \le C_2(r)3^{-mr/(r+1/2)} \le C_2(r)e^{-A_2\rho m}.$$

Theorem 4.2 is proved. ∎

Let us discuss the efficiency of the Pure Greedy Algorithm with respect to the system $\mathcal{TV}$. We prove first that this algorithm is defined correctly, namely, we prove the existence theorem.

**Theorem 4.8.** *For any $f \in L_2$ there exists a function $g \in \mathcal{TV}$ such that*

$$\langle f, g \rangle = \sup_{g \in \mathcal{TV}} |\langle f, g \rangle|.$$

*Proof.* Let

$$f(x) = \sum_{k=0}^{\infty}(y_{2k}c_k(x) + y_{2k+1}s_k(x)).$$

The assumption $f \in L_2$ implies

$$\sum_{i=0}^{\infty} y_i^2 = \|f\|_2^2 < \infty.$$

Let $\{z_k\}_{k=1}^{\infty}$ denote the decreasing rearrangement of $\{|y_i|\}_{i=0}^{\infty}$. It is easy to see that the problem of finding $\sup_{g\in\mathcal{TV}} |\langle f, g\rangle|$ is equivalent to the following: find

$$\sup_n n^{-1/2} \sum_{k=1}^{n} z_k.$$

We prove the existence of a solution to this last problem for $z := \{z_k\}_{k=1}^{\infty} \in l_2$. It is sufficient to prove that

$$\lim_{n\to\infty} n^{-1/2} \sum_{k=1}^{n} z_k = 0. \tag{4.4}$$

Indeed, we have

$$n^{-1/2} \sum_{1\le k<n^{1/2}} z_k \le n^{-1/4} (\sum_{1\le k<n^{1/2}} z_k^2)^{1/2} \le n^{-1/4} \|z\|_{l_2}, \tag{4.5}$$

and

$$n^{-1/2} \sum_{n^{1/2}\le k\le n} z_k \le (\sum_{k\ge n^{1/2}} z_k^2)^{1/2}. \tag{4.6}$$

The relations (4.5) and (4.6) under assumption $z \in l_2$ imply (4.4). $\square$

**Theorem 4.9.** *There exist two absolute positive constants $A_3$ and $A_4$ such that*

$$C_3(r)e^{-A_3\rho m} \le G_m(H_2^r, \mathcal{TV})_2 \le C_4(r)e^{-A_4(rm)^{1/2}}$$

*where $\rho := \min(r, 1/2)$.*

*Proof.* We begin with the lower estimate. Let us use Theorem 4.5. We have

$$G_m(H_2^r, \mathcal{TV})_2 \ge (2n)^{-r} G_m(\mathcal{T}(n)_2, \mathcal{TV}(n))_2 = (2n)^{-r} G_m(B_2^{2n+1}, \mathcal{V}). \tag{4.7}$$

Define $n$ as the smallest integer satisfying the inequality $m < 3(1+\ln(2n+1))/16$. Then, for this $n$ using Theorem 4.5 we get

$$G_m(B_2^{2n+1}, \mathcal{V}) \ge 1/2$$

and by (4.7)

$$G_m(H_2^r, \mathcal{TV})_2 \ge (2n)^{-r}/2 \ge C_3(r)e^{-A_3rm}.$$

This gives the lower estimate for small $r$. The case $r > 1/2$ follows from Theorem 4.7.

We prove now the upper estimate. Using Theorem 4.4 we establish the following lemma.

**Lemma 4.1.** *Let $C_r$ denote a constant such that for $f$ we have*

$$\|U_n(f)\|_2 \le C_r n^{-r}.$$

*Denote $L(n) := 1 + \ln(2n + 1)$. Then for each such function $f$ with $\|f\|_2 \ge (2L(n))^{1/2} C_r n^{-r}$ we can find a $g \in \mathcal{TV}$ such that*

$$\|f - \langle f, g\rangle g\|_2^2 \le \left(1 - \frac{1}{2L(n)}\right) \|f\|_2^2.$$

*Proof.* Represent $f = S_n(f) + U_n(f)$ and find by Theorem 4.4 $g \in \mathcal{TV}(n)$ such that

$$\|S_n(f) - \langle S_n(f), g\rangle g\|_2^2 \le (1 - 1/L(n))\|S_n(f)\|_2^2. \tag{4.8}$$

Then

$$\|f - \langle f, g\rangle g\|_2^2 = \|S_n(f) - \langle S_n(f), g\rangle g\|_2^2 + \|U_n(f)\|_2^2 \le \tag{4.9}$$
$$(1 - 1/L(n))\|f\|_2^2 + (2L(n))^{-1}\|f\|_2^2 \le (1 - \frac{1}{2L(n)})\|f\|_2^2.$$

Lemma 4.1 is proved now. ■

We need now the following general property of the system $\mathcal{TV}$. For any set $Y \subset \mathbb{Z}$ we denote by $S_Y$ the orthogonal projector onto the subspace of trigonometric polynomials with frequencies in $Y$.

**Lemma 4.2.** *For any set $Y \subset \mathbb{Z}$ and any $f \in L_2$ we have*

$$\|S_Y(f - G(f, \mathcal{TV}))\|_2 \le \|S_Y(f)\|_2.$$

*Proof.* We prove this lemma by contradiction. Denote $h := G(f, \mathcal{TV})$ and assume that for some $Y$ we have

$$\|S_Y(f - h)\|_2 > \|S_Y(f)\|_2.$$

Let $X := \mathbb{Z} \setminus Y$. Then we have

$$\|f - S_X(h)\|^2 = \|S_X(f - h)\|^2 + \|S_Y(f)\|^2 < \tag{4.10}$$
$$\|S_X(f - h)\|^2 + \|S_Y(f - h)\|^2 = \|f - h\|^2.$$

Next, $S_X(h)$ has the form $ag$ with some $g \in \mathcal{TV}$. Therefore (4.10) contradicts the following minimizing property of $h$ :

$$\|f - h\| = \inf_{a \in \mathbb{R}, g \in \mathcal{TV}} \|f - ag\|.$$

This completes the proof of Lemma 4.2. ■

*Proof of Theorem 4.9 (continuation).* For a given $m$ find $n$ satisfying

$$\left(1 - \frac{1}{2L(n)}\right)^m \asymp L(n)n^{-2r}.$$

Denote

$$f_k := f - G_k(f, \mathcal{TV}), \quad k = 1, \dots, m.$$

Using the assumption $f \in H_2^r$ we get

$$\|U_n(f)\|_2 \le C_r n^{-r}.$$

If $\|f\|_2 \le (2L(n))^{1/2} C_r n^{-r}$ then we have

$$\|f - G_m(f, \mathcal{TV})\|_2 \le \|f\|_2 \le (2L(n))^{1/2} C_r n^{-r}.$$

If $\|f\|_2 \ge (2L(n))^{1/2} C_r n^{-r}$ then we apply Lemma 4.1 and get

$$\|f_1\| \le \left(1 - \frac{1}{2L(n)}\right)^{1/2} \|f\|_2.$$

Applying Lemma 4.2 we get

$$\|U_n(f_1)\|_2 \le \|U_n(f)\|_2 \le C_r n^{-r}.$$

Continuing this process we obtain

$$\|f - G_m(f, \mathcal{TV})\|_2 \le C(r) \min\{\left(1 - \frac{1}{2L(n)}\right)^{m/2}, (2L(n))^{1/2} n^{-r}\} \le C_4(r) e^{-A_4 (rm)^{1/2}}.$$

Theorem 4.9 is proved. ■

## References


[B] A. Barron, *Universal approximation bounds for superposition of a sigmoidal function*, IEEE Transactions on Information Theory **39** (1993), 930-945.

[D] V.V. Dubinin, *Greedy Algorithms and Applications*, Dissertation USC, 1997.

[DDGS1] C. Darken, M. Donahue, L. Gurvits, E. Sontag, *Rate of approximation results motivated by robust neural network learning*, 6th. ACM Conference on Computer Learning Theory, ACM, 1993, pp. 303-309.

[DDGS2] C. Darken, M. Donahue, L. Gurvits, E. Sontag, *Rates of Convex Approximation in non-hilbert Spaces*, Constructive Approximation **13** (1997), 187–220.

[DJP] R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, American Journal of Mathematics **114** (1992), 737–785.

[DMA] Geoffrey Davis, Stéphane Mallat, and Marco Avellaneda, *Adaptive Greedy Approximations*, Constructive Approximation **13** (1997), 57–98.

[DT1] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Advances in Computational Mathematics **5** (1996), 173–187.

[DT2] R.A. DeVore and V.N. Temlyakov, *Nonlinear approximation in finite dimensional spaces* (to appear).

[J] L. Jones, *A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training*, Annals of Statistics **20** (1992), 608–613.

[T1] V.N. Temlyakov, *Greedy Algorithm and $m$-term trigonometric approximation*, Constructive Approximation (to appear).

[T2] V.N. Temlyakov, *The best $m$-term approximation and greedy algorithms* (to appear).