Nonlinear approximation with dictionaries.
I. Direct estimates

Rémi Gribonval* and Morten Nielsen†

Abstract

We study various approximation classes associated with m-term approximation by elements from a (possibly redundant) dictionary in a Banach space. The standard approximation class associated with the best m-term approximation is compared to new classes defined by considering m-term approximation with algorithmic constraints: thresholding and Chebychev approximation classes are studied respectively. We consider embeddings of the Jackson type (direct estimates) of sparsity spaces into the mentioned approximation classes. General direct estimates are based on the geometry of the Banach space, and we prove that assuming a certain structure of the dictionary is sufficient and (almost) necessary to obtain stronger results. We give examples of classical dictionaries in L^p spaces and modulation spaces where our results recover some known Jackson type estimates, and discuss some new estimates they provide.

Keywords: nonlinear m-term approximation, dictionary, constrained approximation, thresholding algorithm, greedy algorithm, sparse decomposition, Jackson inequality, hilbertian property.
Mathematical Reviews Classification numbers: 41A17, 41A29, 41A46, 41A65

Introduction

Let X be a Banach space, and $\mathcal{D} = \{g_k, k \geq 1\}$ a countable family of unit vectors, $\|g_k\|_X = 1$, which will be called a **dictionary**. A dictionary with dense span is said to be **complete**. Our main purpose in this paper is to study **approximation classes** associated with m-term approximation, that is to say classes of elements $f \in X$ that can be approximated by m elements of $\mathcal{D}$ with some (theoretical) algorithm $f \mapsto A_m(f)$ at a certain rate, e.g., $\|f - A_m(f)\|_X = \mathcal{O}(m^{-\alpha})$.

The algorithm we will use as a "benchmark" is the one associated with best m-term approximation. The (nonlinear) set of all linear combinations of at most m elements from $\mathcal{D}$ is

$$\Sigma_m(\mathcal{D}) := \left\{ \sum_{k \in I_m} c_k g_k, I_m \subset \mathbb{N}, card(I_m) \leq m, c_k \in \mathbb{C} \right\}.$$

For any given $f \in X$, the error associated to the *best m-term* approximation to f from $\mathcal{D}$ is given by

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

The *best m-term approximation classes* are defined as :

$$\mathcal{A}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} < \infty \right\}$$

*IRISA-INRIA, Campus de Beaulieu, 35042 Rennes cedex, France, Phone: (+33) 299 842 506, Fax: (+33) 299 842 532, email: remi.gribonval@inria.fr
†Department of Mathematical Sciences, Aalborg University, Fredrik Bajers vej 7G, DK-9220 Aalborg East, Denmark, email: mnielsen@math.auc.dk

where $|\cdot|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} := \|\{\sigma_m(f,\mathcal{D})_X\}_{m\geq 1}\|_{\ell_q^{1/\alpha}}$ is defined using the Lorentz (quasi)norm, see e.g. [DL93]. The class $\mathcal{A}_q^\alpha(\mathcal{D},X)$ is thus basically the set of functions f that can be approximated at a given rate $\mathcal{O}(m^{-\alpha})$ $(0<\alpha<\infty)$ by a linear combination of m elements from the dictionary. The parameter $0<q\leq\infty$ is auxiliary and gives a finer classification of the approximation rate. It turns out that $\mathcal{A}_q^\alpha(\mathcal{D},X)$ is indeed a linear subspace of X, and the quantity $\|\cdot\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)}$ is a (quasi)norm, see *e.g.* [DL93, Chapter 7, Section 9].

Stechkin, DeVore, and Temlyakov have derived the following nice characterization.

Theorem 1 ([Ste55, DT96]) *If $\mathcal{B}$ is an orthonormal basis in a Hilbert space $\mathcal{H}$ then, for $0<\tau=(\alpha+1/2)^{-1}<2$ and $0<q\leq\infty$,*

$$\mathcal{A}_q^\alpha(\mathcal{B},\mathcal{H})=\mathcal{K}_q^\tau(\mathcal{B},\mathcal{H})$$

with equivalent (quasi)norms, where

$$\mathcal{K}_q^\tau(\mathcal{B},\mathcal{H}) := \left\{f\in\mathcal{H}, |f|_{\mathcal{K}_q^\tau(\mathcal{B},\mathcal{H})} := \|\{\langle f,g_k\rangle\}_{k\geq 1}\|_{\ell_q^\tau}<\infty\right\}.$$

In our previous paper [GN01], this result was extended to $\mathcal{B}$ a quasi-greedy basis in a Hilbert space (*e.g.* a Riesz basis), and similar results [KP01, DKKT01] were obtained whenever $\mathcal{B}$ is an almost-greedy basis in a general Banach space. We refer to [KT99, Woj00] for the notions of (quasi)-greedy bases and to [DKKT01] for the notion of almost-greedy bases.

The goal of the present paper is to generalize (part of) Theorem 1 to some redundant dictionaries. Based on examples in [GN01] we know that we need to require some structure of $\mathcal{D}$. We focus our attention on the identification of the structure required to get continuous embeddings of the Jackson type

$$\mathcal{K}_q^\tau(\mathcal{D},X)\hookrightarrow\mathcal{A}_q^\alpha(\mathcal{D},X)$$

with $\tau=(\alpha+1/p)^{-1}$ for some $1\leq p<\infty$. Let us comment on the definition of $\mathcal{K}_q^\tau(\mathcal{D},X)$ when $\mathcal{D}$ is not an orthonormal basis. In case of an orthonormal basis, $\mathcal{K}_q^\tau(\mathcal{D},X)$ measures the sparsity of the expansion of f, but for general *redundant* dictionaries, there is not a unique decomposition $f=\sum_k c_k(f)g_k$. For a redundant dictionary $\mathcal{D}$, we follow DeVore and Temlyakov [DT96], and define the *sparsity classes* $\mathcal{K}_q^\tau(\mathcal{D},X)$ as follows. For $\tau\in(0,\infty)$ and $q\in(0,\infty]$ we let $\mathcal{K}_q^\tau(\mathcal{D},X,M)$ denote the set

$$\mathrm{clos}_X\left\{f\in X, f=\sum_{k\in I}c_kg_k, I\subset\mathbb{N}, card(I)<\infty, \|\{c_k\}_{k\geq 1}\|_{\ell_q^\tau}\leq M\right\}.$$

Then we define $\mathcal{K}_q^\tau(\mathcal{D},X) := \cup_{M>0}\mathcal{K}_q^\tau(\mathcal{D},X,M)$ with $|f|_{\mathcal{K}_q^\tau(\mathcal{D},X)} := \inf\{M, f\in\mathcal{K}_q^\tau(\mathcal{D},X,M)\}$.

Remark 1 *It can be proved that $|\cdot|_{\mathcal{K}_q^\tau(\mathcal{D},X)}$ is a (semi)-(quasi)norm on $\mathcal{K}_q^\tau(\mathcal{D},X)$.*

The structure of the paper is as follows. In Section 1 we introduce the constrained approximation classes we want to consider; they are the classes associated with thresholding approximation and Chebyshev approximation. We make a comparison of the classes at the end of the section.

In Section 2, we consider general Jackson type embeddings of the sparsity classes into some of the approximation classes. Two types of embeddings are considered in this section; a universal embedding that holds for every type of dictionary in any space, and a result that applies to arbitrary dictionaries in Banach spaces with a modulus of smoothness of powertype. The two types of embeddings are compared at the end of the section.

The main results of the paper are contained in Section 3, where we study the so-called hilbertian dictionaries. We give a complete characterization of the sparsity spaces associated with such dictionaries in terms of sequence spaces, and a third type of Jackson embedding is considered, this one based on the hilbertian structure. At the end of Section 3 we discuss in detail how the different types of Jackson estimates are related depending on the structure of the Banach space and of the dictionary.

Examples of hilbertian dictionaries are given in Section 4 to illustrate how the Jackson estimate of Section 3 recovers some known results of nonlinear approximation in L^p spaces and in modulation spaces, and provide new estimates for some less classical dictionaries.

1 Constrained approximation classes

We defined the "benchmark" approximation class associated with best m-term approximation in the introduction. Below is a description of the other approximation classes that will be considered.

1.1 Thresholding approximation

Computing the best m-term approximant to a function f from an overcomplete dictionary is usually computationally intractable [DMA97, Jon97]. It may be much easier to build m-term approximants in an *incremental* way :

$$f_m(\pi, \{c_k\}, \mathcal{D}) := \sum_{k=1}^{m} c_k g_{\pi_k} \tag{1}$$

where $\pi : \mathbb{N} \to \mathbb{N}$ is injective. In [KT99], *greedy approximation* from a (Schauder) basis $\mathcal{D} = \mathcal{B}$ is compared to best m-term approximation. Greedy approximants can be written as $f_m(\pi, \{c_k^\star\}, \mathcal{D})$ where $c_k^\star = c_{\pi_k}(f)$ is a decreasing rearrangement of the (unique) coefficients $\{c_k(f)\}$ such that $f = \sum_{k=1}^{\infty} c_k(f) g_k$. They are obtained by *thresholding* the coefficients of f in the basis. In the recent paper [DKKT01], greedy approximants from a Schauder basis are compared to best m-term approximants, with the restriction that only coefficients obtained from the dual coefficient functionals are used (i.e., a weaker notion than best m-term approximation), see [DKKT01] for details. This leads to the concept of *almost-greedy* bases.

In a redundant dictionary, we can generalize the notion of greedy approximants by considering approximants of the form $f_m(\pi, \{c_k^\star\}, \mathcal{D})$ where $\{|c_k^\star|\}$ is decreasing. To avoid confusion with a different notion of "greedy algorithm" [FT74, Hub85, Jon87, DT96, Tem00], we will rather use the notion of *thresholding algorithm* and define *thresholding approximation classes* that generalize the "greedy approximation classes" $\mathcal{G}_q^\alpha(\mathcal{B})$ that we defined in [GN01] :

$$\mathcal{T}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} < \infty \right\},$$

with

$$|f|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} := \inf_{\pi, \{c_k^\star\}} \left(\sum_{m=1}^{\infty} \left(\left[m^\alpha \left\| f - f_m(\pi, \{c_k^\star\}, \mathcal{D}) \right\|_X \right]^q \frac{1}{m} \right) \right)^{1/q},$$

for $0 < q < \infty$, where $\{|c_k^\star|\}$ is required to be nonincreasing. In the case $q = \infty$ we simply put

$$|f|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} := \inf_{\pi, \{c_k^\star\}} \left(\sup_{m \geq 1} m^\alpha \left\| f - f_m(\pi, \{c_k^\star\}, \mathcal{D}) \right\|_X \right).$$

Remark 2 *Notice that the sum in the expression defining the quantity $|f|_{\mathcal{T}_q^\alpha(\mathcal{D},X)}$ is closely related to the Lorentz norm of $\{\|f - f_m(\pi, \{c_k^\star\}, \mathcal{D})\|_X\}_{m \geq 1}$ in $\ell_q^{1/\alpha}$, with the twist that the sequence $\{\|f - f_m(\pi, \{c_k^\star\}, \mathcal{D})\|_X\}_m$* might not *be decreasing.*

1.2 Chebyshev approximation

For each m, the Chebyshev projection $P_{\mathcal{V}_m(\pi,\mathcal{D})} f$ of f onto the (closed) finite dimensional subspace

$$\mathcal{V}_m(\pi, \mathcal{D}) := \text{span}(g_{\pi_1}, \ldots, g_{\pi_m})$$

is at least as good an m-term approximant to f as any incremental approximant $f_m(\pi, \{c_k\}, \mathcal{D}) \in \mathcal{V}_m(\pi, \mathcal{D})$. We define *Chebyshev (incremental) approximation classes* as

$$\mathcal{C}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{C}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{C}_q^\alpha(\mathcal{D},X)} < \infty \right\}$$

where
$$|f|_{\mathcal{C}^\alpha_q(\mathcal{D},X)} := \inf_\pi \big\|\{\big\|f - P_{\mathcal{V}_m(\pi,\mathcal{D})} f\big\|_X\}_{m\geq 1}\big\|_{\ell_q^{1/\alpha}},$$
with the obvious modification for $q=\infty$. It turns out, that $\mathcal{C}^\alpha_q(\mathcal{D},X)$ is indeed a linear subspace of X, and the quantity $\|\cdot\|_{\mathcal{C}^\alpha_q(\mathcal{D},X)}$ is a (quasi)norm.

Proposition 1 *Let $\mathcal{D}$ a dictionary in a Banach space X, $\alpha>0$ and $0<q\leq\infty$. The set $\mathcal{C}^\alpha_q(\mathcal{D})$ is a linear subspace of X, and $\|\cdot\|_{\mathcal{C}^\alpha_q(\mathcal{D})}$ is a (quasi)norm on $\mathcal{C}^\alpha_q(\mathcal{D})$.*

Proof. We let $f,g\in\mathcal{C}^\alpha_q(\mathcal{D},X)$ and fix some $\epsilon>0$. We consider two injections $\pi,\psi:\mathbb{N}\to\mathbb{N}$ such that $\big\|\{\|f-P_{\mathcal{V}_m(\pi,\mathcal{D})f\|}\}_{m\geq1}\big\|_{\ell_q^{1/\alpha}}\leq|f|_{\mathcal{C}^\alpha_q(\mathcal{D},X)}+\epsilon$ and a similar relation holds for g and ψ. We define an injection ϕ by induction: $\phi_1=\pi_1$; $\phi_{2n}=\psi_m$ where m is the smallest integer such that $\psi_m\notin\{\phi_k, 1\leq k<2n\}$; $\phi_{2n+1}=\pi_m$ where m is the smallest integer such that $\pi_m\notin\{\phi_k,1\leq k<2n+1\}$. As $\mathcal{V}_{2m}(\phi,\mathcal{D})\supset\mathcal{V}_m(\pi,\mathcal{D})+\mathcal{V}_m(\psi,\mathcal{D})$, we get for $j\geq0$
$$\begin{aligned}\|f+g-P_{\mathcal{V}_{2^{j+1}}(\phi,\mathcal{D})}(f+g)\|_X &\leq \|f-P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})}f+g-P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})}g\|_X\\ &\leq \|f-P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})}f\|_X+\|g-P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})}g\|_X.\end{aligned}$$
Now, the ℓ_q^τ (quasi)norm of a sequence $\{a_m\}$ can be estimated (see *e.g.* [DL93, Chap. 6]) as
$$\|\{a_m\}_{m=1}^\infty\|_{\ell_q^\tau}\asymp\begin{cases}\left(\sum_{j=0}^\infty(2^{j/\tau}a^\star_{2^j})^q\right)^{1/q}, & 0<q<\infty\\ \sup_{j\geq0}2^{j/\tau}a^\star_{2^j}, & q=\infty\end{cases},$$
hence we get $|f+g|_{\mathcal{C}^\alpha_q(\mathcal{D})}\leq C(\|f+g\|_X+|f|_{\mathcal{C}^\alpha_q(\mathcal{D})}+|f|_{\mathcal{C}^\alpha_q(\mathcal{D})}+2\epsilon)$. We let ϵ go to zero to conclude. □

1.3 Characterization of the approximation classes

We do not claim that the quantity $\|\cdot\|_{\mathcal{T}^\alpha_q(\mathcal{D},X)}$ is, in general, a (quasi)norm, nor do we claim that the corresponding classes are in general linear subspaces of X. However the following set inclusions hold
$$\mathcal{T}^\alpha_q(\mathcal{D},X)\subset\mathcal{C}^\alpha_q(\mathcal{D},X)\subset\mathcal{A}^\alpha_q(\mathcal{D},X)\subset X \tag{2}$$
together with the inequalities
$$|\cdot|_{\mathcal{A}^\alpha_q(\mathcal{D},X)}\lesssim|\cdot|_{\mathcal{C}^\alpha_q(\mathcal{D},X)}\lesssim|\cdot|_{\mathcal{T}^\alpha_q(\mathcal{D},X)} \tag{3}$$
where the notation $|\cdot|_W\lesssim|\cdot|_V$ denotes the existence of a constant $C<\infty$ such that $|f|_W\leq C|f|_V$ for all f. The value of the constant may vary from one occurrence in an equation to another. Throughout this paper we will use the notation $V\hookrightarrow W$, whenever $V\subset W$ and $|\cdot|_W\lesssim|\cdot|_V$. Let us insist on the fact that V (resp. W) is the *subset* of X where the functional $|\cdot|_V$ (resp. $|\cdot|_W$) is finite, which need not be a (semi)-(quasi)normed linear subspace of X.

Remark 3 *1. In most of this paper, $\mathcal{A}^\alpha_q(\mathcal{D},X)$ will be denoted for short by $\mathcal{A}^\alpha_q(\mathcal{D})$, and similar shorthands will be used for the other classes.*

2. We will reserve the notation $\|\cdot\|_V$ to the "nondegenerate" case when $\|f\|_V=0\Rightarrow f=0$.

2 General Jackson-type embeddings

In this section we are interested in getting Jackson embeddings $\mathcal{K}^\tau_q(\mathcal{D})\hookrightarrow\mathcal{A}^\alpha_q(\mathcal{D}), \alpha=1/\tau-1/p$ for some p. First we will see that a **universal** Jackson embedding holds (Theorem 2) with $p=1$ for any space X and any dictionary $\mathcal{D}$. In a second step we will discuss embeddings of the **Chebyshev-Jackson** type
$$\mathcal{K}^\tau_\tau(\mathcal{D})\hookrightarrow\mathcal{C}^\alpha_\infty(\mathcal{D}),\qquad\alpha=1/\tau-1/p$$

with some $p > 1$ that is given by the *geometry* of the unit ball of X. These embeddings hold for any dictionary in X and they imply standard Jackson embeddings. However, DeVore and Temlyakov [DT96] remarked that they seem to be restricted to $0 < \tau \leq 1$ and we will prove this fact (Theorem 3).

2.1 Universal Jackson embedding

In this section we obtain a universal Jackson inequality for any dictionary, but before we state the result let us introduce some notation that will be used throughout the paper. For *any* dictionary $\mathcal{D} = \{g_k\}$ in *any* Banach space X it makes sense to define the operator

$$T \ : \{c_k\} \mapsto \sum_k c_k g_k$$

on the space ℓ^0 of finite sequences $\boldsymbol{c} = \{c_k\}$.

Theorem 2 *For any $\tau < 1$ and $q \in (0, \infty]$, there is a constant $C = C(\tau, q)$ such that for $\mathcal{D}$ an* arbitrary *dictionary in an* arbitrary *Banach space X and any $f \in \mathcal{K}_q^\tau(\mathcal{D})$*

$$\|f\|_{\mathcal{A}_q^\alpha(\mathcal{D})} \leq C|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \qquad \textit{with } \alpha = 1/\tau - 1.$$

Proof. Let $f \in \mathcal{K}_q^\tau(\mathcal{D})$ and fix M an integer. Let $\epsilon > 0$, and $\boldsymbol{c} \in \ell^0$ a finite sequence such that $\|T\boldsymbol{c} - f\|_X \leq \epsilon\sigma_M(f, \mathcal{D})_X$ and $\|\boldsymbol{c}\|_{\ell_q^\tau} \leq (1+\epsilon)|f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. For any $1 \leq m \leq M$, let $\boldsymbol{c}_m$ the best m-term approximant to $\boldsymbol{c}$: we have the estimate

$$\begin{aligned} \sigma_m(f, \mathcal{D})_X &\leq \|T\boldsymbol{c}_m - f\|_X \leq \|T\boldsymbol{c}_m - T\boldsymbol{c}\|_X + \|T\boldsymbol{c} - f\|_X \\ &\leq \|\boldsymbol{c}_m - \boldsymbol{c}\|_{\ell^1} + \epsilon\sigma_M(f, \mathcal{D})_X \end{aligned}$$

which gives $(1-\epsilon)\sigma_m(f, \mathcal{D})_X \leq \sigma_m(\boldsymbol{c}, \mathcal{B})_{\ell^1}$ with $\mathcal{B}$ the canonical basis in ℓ^1. Taking partial sums we get

$$(1-\epsilon)\left(\sum_{m=1}^M \frac{[m^\alpha \sigma_m(f, \mathcal{D})_X]^q}{m}\right)^{1/q} \leq |\boldsymbol{c}|_{\mathcal{A}_q^\alpha(\mathcal{B}, \ell^1)} \leq C\,\|\boldsymbol{c}\|_{\ell_q^\tau} \leq C(1+\epsilon)|f|_{\mathcal{K}_q^\tau(\mathcal{D})}$$

with $\alpha = 1/\tau - 1$ and $C = C(\tau, q)$ given by the Hardy inequality. Letting ϵ go to zero, then M go to infinity, we eventually get $|\cdot|_{\mathcal{A}_q^\alpha(\mathcal{D})} \leq C|\cdot|_{\mathcal{K}_q^\tau(\mathcal{D})}$. We notice that, because $\tau < 1$, $\|.\|_X \leq |\cdot|_{\mathcal{K}_1^1(\mathcal{D})} \leq |\cdot|_{\mathcal{K}_q^\tau(\mathcal{D})}$, which gives the result. □

The **universal Jackson embedding** guarantees that any function with sparsity $\tau < 1$ can be approximated with a rate of approximation at least $\alpha = 1/\tau - 1$. For $\tau \geq 1$ we might find some space X, some dictionary $\mathcal{D}$ and some $f \in \mathcal{K}_\tau^\tau(\mathcal{D}, X)$ for which α is arbitrarily close to zero, this will be demonstrated in Theorem 3. It is clear that the Jackson embedding given by Theorem 2 may not be the best possible since the result is *too general*. In the following sections we will improve the embedding in cases where there is more structure either of the space X or of the dictionary $\mathcal{D}$.

2.2 Chebyshev-Jackson embedding

When the space X in which the approximation takes place has some geometric structure, a series of known results provides with improved Jackson-type estimates for arbitrary dictionaries. Maurey [Pis81] proved the Jackson inequality

$$\sigma_m(f, \mathcal{D})_X \leq Cm^{-\alpha}|f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}, \qquad m \geq 1, \qquad \alpha = 1/\tau - 1/p \tag{4}$$

for $\mathcal{D}$ an arbitrary dictionary in X a Hilbert space, $p = 2$ and $\tau = 1$ ($\alpha = 1/2$), and Jones [Jon92] proved that the *relaxed greedy algorithm* reaches this rate of approximation. DeVore and Temlyakov [DT96] extended this

Jackson inequality to $0 < \tau = (\alpha + 1/2)^{-1} \leq 1$ (*i.e.* $\alpha \geq 1/2$), for $\mathcal{D}$ an arbitrary dictionary in a Hilbert space. They also made the interesting remark that for $\alpha < 1/2$ "there seems to be no obvious analogue" to this result. This comment can be made rigorous; we have the following theorem that will be proved in Section 2.3.

Theorem 3 *In any infinite dimensional (separable) Hilbert space $\mathcal{H}$ there exists a dictionary $\mathcal{D}$ such that the Jackson inequality* (4) *fails for every $\tau > 1$ and $\alpha > 0$.*

Temlyakov [Tem00] obtained a Jackson inequality (4) for $\mathcal{D}$ an arbitrary dictionary in a Banach space X, with $1 < p \leq 2$ the powertype of the modulus of smoothness of X (see *e.g.* [LT79, Vol. II]), and $\tau = 1$ ($\alpha = 1 - 1/p$). Later on, the same autor extended this result to $0 < \tau = (\alpha + 1/p)^{-1} \leq 1$ (*i.e.* $\alpha \geq 1 - \frac{1}{p}$) using an idea from the proof of [DT96, Theorem 3.3], see [Tem01, Theorem 11.3]. Temlyakov's technique is constructive and uses a generalization of the *orthogonal greedy algorithm*, the so-called *weak Chebyshev greedy algorithm*.

Theorem 4 (Temlyakov) *Let X a Banach space with modulus of smoothness of powertype p, where $1 < p \leq 2$. For any $0 < \tau \leq 1$, there exists a constant $C = C(\tau, p)$ such that for any dictionary $\mathcal{D}$ in X, there is a constructive algorithm that selects, for any $f \in \mathcal{K}_\tau^\tau(\mathcal{D})$, a permutation $\pi(f)$ such that*

$$\begin{aligned} \|f - P_{\mathcal{V}_m(\pi(f),\mathcal{D})} f\|_X \quad &\leq \quad C m^{-\alpha} |f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}, \qquad m \geq 1 \\ &\quad \textit{with } \alpha = 1/\tau - 1/p. \end{aligned} \tag{5}$$

Note that the Jackson inequality (5) is not standard: the left hand-side is not $\sigma_m(f, \mathcal{D})_X$, but the error of approximation using what Temlyakov calls the *weak Chebyshev greedy algorithm*. Thus, the result is stronger than a standard Jackson inequality. To mark the difference we will call it a **Chebyshev-Jackson inequality**.

Using the easy fact that, for $0 < \tau \leq 1$, $\|\cdot\|_X \lesssim |\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$, Temlyakov's result can be restated in terms of a **Chebyshev-Jackson embedding**

$$\mathcal{K}_\tau^\tau(\mathcal{D}) \hookrightarrow \mathcal{C}_\infty^\alpha(\mathcal{D}), \qquad \alpha = 1/\tau - 1/p, \qquad 0 < \tau \leq 1 \tag{6}$$

with $1 < p \leq 2$ the powertype of the modulus of smoothness of X.

Remark 4 (Equivalent norms on X) *The Chebyshev-Jackson inequality proved by Temlyakov is of a geometric nature : it holds for any dictionary, but is intimately connected to the geometry of the unit ball of X. Notice that, if we replace the original norm $\|\cdot\|_X$ by an equivalent norm $|||\cdot|||_X$, the approximation spaces $\mathcal{A}_q^\alpha(\mathcal{D})$, $\mathcal{T}_q^\alpha(\mathcal{D})$, and $\mathcal{C}_q^\alpha(\mathcal{D})$ do not change, and their "norms" are simply changed to equivalent quantities. As for the sparsity spaces $\mathcal{K}_q^\tau(\mathcal{D})$ their definition only involves the topology of X, hence their "norm" remains identical under a change of equivalent norm on X. On the other hand, a change of norm on X can change drastically the powertype of its modulus of smoothness, as can be seen in finite dimension where all norms are equivalent but may have very different smoothness. Hence the powertype should be understood as the largest powertype over all equivalent norms on X.*

Keeping the above remark in mind, we have the following definition.

Definition 1 *We define $P_g(X)$ to be the largest real number such that some norm $|||\cdot|||_X$ equivalent to $\|\cdot\|_X$ has modulus of smoothness of powertype p for all $p < P_g(X)$.*

Remark 5 *Any Banach space has powertype 1, so $P_g(X) \geq 1$ always. It is also known [LT79, Vol. II, Theorem 1.e.16] that if X has type $p_t \leq 2$ then $1 \leq P_g(X) \leq p_t$.*

Figure 1 illustrates the improvement that can be obtained (compared to the universal Jackson embedding) from taking into account the geometry of the space X. As often (see [DeV98]), it is convenient to use $1/\tau$ rather than τ as a coordinate on the horizontal axis. For $1/\tau < 1$, the universally guaranteed α is given by a line of slope one $\alpha = 1/\tau - 1$. If $P_g(X) > 1$ then for any $1 < p < P_g(X)$ the space X has a modulus of smoothness of powertype p and we have the Chebyshev-Jackson embedding line $\alpha = 1/\tau - 1/p$ for $0 < \tau \leq 1$, which improves the universal embedding line. Notice that the value of α is improved by the amount $1 - 1/p$, *i.e.* taking into account the geometry of X made it possible to gain an extra factor $m^{-(1-1/p)}$ in the rate of approximation for any given sparsity $0 < \tau \leq 1$. Note also that the embedding line is limited to the region where $1/\tau \geq 1$, which is a consequence of Theorem 3 as we will see in the next section.

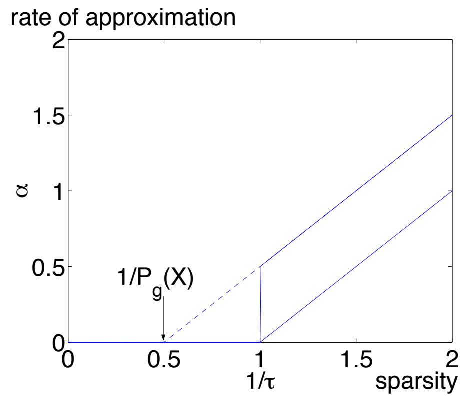

Figure 1: The Chebyshev-Jackson embedding line $\alpha = 1/\tau - 1/p$ for $0 < \tau \leq 1$ (with $1 < p < P_g(X)$) compared to the universal embedding line $\alpha = 1/\tau - 1$.

2.3 Limitations of the "geometric" embedding

The fact that the Chebyshev-Jackson estimate is restricted to $0 < \tau \leq 1$ could seem an artifact of the technique used to prove it, and one could wonder if a result giving a "complete" embedding line is possible. However, we have already mentioned Theorem 3 which shows that $0 < \tau \leq 1$ is an essential limitation. For the proof of Theorem 3, we will need the following lemma.

Lemma 1 *Let $\mathcal{D} = \{g_k\}$ a dictionary in an arbitrary Banach space X and assume $g \in X$ is an accumulation point of $\mathcal{D}$ (i.e. for every neightborhood B of g in X, there exists infinitely many values of k for which $g_k \in B$). Then for all $\tau > 1$, $|g|_{\mathcal{K}_\tau^\tau(\mathcal{D})} = 0$.*

This lemma shows in particular that if the dictionary has at least one accumulation point, then $|\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$ can at most be a *semi*-(quasi) norm.

Proof of Lemma 1. By standard arguments, there exists a sequence of $\{k_n\}_{n\geq 0}$ such that $\|g - g_{k_n}\|_X \leq 2^{-n}$. Note that $\|g_k\|_X = 1, k \geq 1$ implies $\|g\|_X = 1$. For all $N \geq 1$

$$\|g - \frac{1}{N}\sum_{n=1}^{N} g_{k_n}\|_X \leq \frac{1}{N}\sum_{n=1}^{N}\|g - g_{k_n}\|_X \leq \frac{1}{N}\sum_{n=1}^{N} 2^{-n} \leq \frac{1}{N}.$$

It follows that $|g|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq N^{1/\tau - 1}$ for all N, hence the result for $\tau > 1$. □.

Proof of Theorem 3. Let $\mathcal{B} = \{e_j\}_{j=0}^{\infty}$ an orthonormal basis of $\mathcal{H}$ and $V_j := \text{span}\{e_{2j}, e_{2j+1}\}$. Let $\mathcal{D} := \{g_{j,n}, j, n \geq 0\}$ where for each j $\{g_{j,n}\}_{n\geq 0}$ is a sequence of unit vectors from V_j. Clearly, $\{g_{j,n}\}_{n\geq 0}$ has at least one accumulation point $\overline{g}_j \in V_j, \|\overline{g}_j\| = 1$ for each j and, by Lemma 1, for any $\tau > 1$ and j, $|\overline{g}_j|_{\mathcal{K}_\tau^\tau(\mathcal{D})} = 0$. For any ℓ^2 sequence $\boldsymbol{c} = \{c_j\}_{j\geq 0}$, one can properly define $f := \sum_j c_j \overline{g}_j$ and check that $|f|_{\mathcal{K}_\tau^\tau(\mathcal{D})} = 0$. On the other hand, $\sigma_m(f, \mathcal{D})_{\mathcal{H}}$ can decrease arbitrarily slowly, as one can check that $\sigma_m(f, \mathcal{D})_{\mathcal{H}} = \sigma_m(f, \{\overline{g}_j, j \geq 0\})_{\mathcal{H}} = \sigma_m(\boldsymbol{c}, \tilde{\mathcal{B}})_{\mathcal{H}}$, where $\tilde{\mathcal{B}}$ is the canonical basis of ℓ^2. □

Remark 6 *Notice that the above arguments also show that the Jackson inequality (4) cannot be "repaired" for $\tau > 1$ by replacing $|\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$ with $\|\cdot\|_X + |\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$.*

3 Hilbertian dictionaries

Theorem 3 shows that a Jackson embedding line $\alpha = 1/\tau - 1/p$ cannot be expected to be "complete", *i.e.* valid even for $\tau > 1$, unless we assume some structure on the dictionary $\mathcal{D}$. In this section, we prove that getting a "complete" Jackson embedding line with $p > 1$ is almost equivalent to assuming that $\mathcal{D}$ has a **hilbertian structure**.

Definition 2 *A dictionary $\mathcal{D}$ is called ℓ^τ_q-hilbertian if for any sequence $\mathbf{c} = \{c_k\}_{k\geq 1} \in \ell^\tau_q$, the series $\sum_{k\geq 1} c_k g_k$ is convergent in X and*

$$\|\sum_{k\geq 1} c_k g_k\|_X \lesssim \|\mathbf{c}\|_{\ell^\tau_q}.$$

Remark 7 *Note that the convergence of $\sum_k c_k g_k$ in Definition 2 is necessarily unconditional, provided that ℓ^τ_q is not one of the extremal nonseparable spaces such as ℓ^∞. Also notice that* any *dictionary is ℓ^τ-hilbertian for $0 < \tau \leq 1$.*

First, we study hilbertian dictionaries in more details and give a simple representation of the sparsity class $\mathcal{K}^\tau_q(\mathcal{D})$ for such dictionaries. Then, we will use this representation to prove a strong Jackson embedding of the type $\mathcal{K}^\tau_q(\mathcal{D}) \hookrightarrow \mathcal{T}^\alpha_q(\mathcal{D})$, which we will call a **thresholding-Jackson embedding**, and get the following theorem, which will be proved in Section 3.3,

Theorem 5 *Let $\mathcal{D}$ a dictionary in a Banach space X, and $p > 1$. The following properties are equivalent*

$$\forall\ \tau < p,\ \forall\ q,\ \forall\ \alpha < 1/\tau - 1/p \qquad \mathcal{K}^\tau_q(\mathcal{D}) \hookrightarrow \mathcal{T}^\alpha_q(\mathcal{D}), \tag{7}$$

$$\forall\ \tau < p,\ \forall\ q,\ \forall\ \alpha < 1/\tau - 1/p \qquad \mathcal{K}^\tau_q(\mathcal{D}) \hookrightarrow \mathcal{C}^\alpha_q(\mathcal{D}), \tag{8}$$

$$\forall\ \tau < p,\ \forall\ q,\ \forall\ \alpha < 1/\tau - 1/p \qquad \mathcal{K}^\tau_q(\mathcal{D}) \hookrightarrow \mathcal{A}^\alpha_q(\mathcal{D}), \tag{9}$$

$$\forall\ \tau < p \qquad \mathcal{D} \textit{ is } \ell^\tau_1 - \textit{hilbertian}. \tag{10}$$

At the end of this section we will compare the embeddings provided by the geometry of X to the ones obtained from the structure of $\mathcal{D}$.

3.1 Characterization of ℓ^p_1-hilbertian dictionaries

Some of the structure of $\mathcal{D}$ can be studied through the properties of the operator T, introduced in Section 2.1. In particular, the condition for $\mathcal{D}$ to be ℓ^τ_q-hilbertian can easily be verified to be equivalent to the requirement that T can be extended to a continuous linear operator from ℓ^τ_q to X. Notice that the ℓ^τ_q-hilbertian property of $\mathcal{D}$ does not change under a change of equivalent norm on X. For the purpose of further discussion, we have the following definition.

Definition 3 *For any dictionary $\mathcal{D}$ we define $P_s(\mathcal{D}, X) := \sup\{p : \mathcal{D}$ is ℓ^p_1-hilbertian$\} \in [1, \infty]$.*

Remark 8 *It is easy to deduce directly from the definition of the cotype of a Banach space (see [LT79, Vol. II, Sec. 1.e.]), that if X has cotype $p_c \geq 2$, then $1 \leq P_s(\mathcal{D}, X) \leq p_c$.*

Let us give a simple characterization of dictionaries which are ℓ^p_1-hilbertian.

Proposition 2 *Let $\mathcal{D}$ a dictionary in a Banach space X, and $1 \leq p < \infty$. The following two properties are equivalent:*

(i) *$\mathcal{D}$ is ℓ^p_1-hilbertian.*

(ii) *There is a constant $C < \infty$, for every set of indices $I_m \subset \mathbb{N}$ of cardinality $card(I_m) \leq m$ and every choice of signs*

$$\|\sum_{k\in I_m} \pm g_k\|_X \leq C m^{1/p}. \tag{11}$$

Proof. It is obvious that (i) implies (ii), so let us prove the reciprocal. Let $\boldsymbol{c} \in \ell_1^p$ and π a permutation of $\mathbb{N}$ such that $c_k^\star = c_{\pi_k}$, and define a sequence $f_n = f_n(\boldsymbol{c}) := \sum_{k=1}^n c_k^\star g_{\pi_k} = \sum_{k=1}^n c_{\pi_k} g_{\pi_k}$. By an extremal point argument (write f_n in barycentric coordinates with respect to the system $\{\sum_1^n \pm g_{\pi_k}\}$ and use the triangle inequality) and the growth assumption (11), we can write for every $n \geq m$

$$\|f_n - f_m\|_X \leq C(n-m)^{1/p}|c_m^\star|.$$

By taking $m = 2^j$ and $n = 2^{j+1}$ with $j \geq 0$ we get $\|f_{2^{j+1}} - f_{2^j}\|_X \leq C2^{j/p}|c_{2^j}^\star|$, hence $\sum_{j=0}^\infty \|f_{2^{j+1}} - f_{2^j}\|_X \leq C\sum_{j=0}^\infty |c_{2^j}^\star| 2^{j/p} \leq \tilde{C}\|\boldsymbol{c}\|_{\ell_1^p}$. Hence we can define

$$T\boldsymbol{c} := \lim_{j\to\infty} f_{2^j} = f_1 + \sum_{j=0}^\infty (f_{2^{j+1}} - f_{2^j})$$

which satisfies $\|T\boldsymbol{c}\|_X \leq \|\boldsymbol{c}\|_{\ell^\infty} + \tilde{C}\|\boldsymbol{c}\|_{\ell_1^p} \leq (1+\tilde{C})\|\boldsymbol{c}\|_{\ell_1^p}$. It is easy to check that indeed $T\boldsymbol{c} = \lim f_n$ and the definition of $T\boldsymbol{c}$ does not depend on the choice of a particular decreasing rearrangement of $\boldsymbol{c}$. Now, for $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ two finite sequences and λ a scalar, it is clear that $T(\boldsymbol{c}_1 + \lambda\boldsymbol{c}_2) = T\boldsymbol{c}_1 + \lambda T\boldsymbol{c}_2$, hence T, restricted to the dense subspace ℓ^0 of ℓ_1^p consisting of finite sequences, is linear and continuous. It follows by standard arguments that T extends to a bounded linear operator from ℓ_1^p to X. □

3.2 Representation of the sparsity class

The hilbertian structure of $\mathcal{D}$ makes it possible to get a nice representation of the sparsity spaces $\mathcal{K}_q^\tau(\mathcal{D})$. The operator T below is the one defined in Section 2.1.

Proposition 3 *Assume $\mathcal{D}$ is ℓ_1^p-hilbertian, with $p > 1$. Let $\tau < p$ and $1 \leq q \leq \infty$. For all $f \in \mathcal{K}_q^\tau(\mathcal{D})$, there exists some $\boldsymbol{c} \in \ell_q^\tau$ which realizes the sparsity norm,* i.e. *$f = T\boldsymbol{c}$ and $\|\boldsymbol{c}\|_{\ell_q^\tau} = |f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. In case $1 < \tau, q < \infty$, $\boldsymbol{c} = \boldsymbol{c}_{\tau,q}(f)$ is unique. Consequently*

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D})} = \min_{\boldsymbol{c}\in\ell_q^\tau, f=T\boldsymbol{c}} \|\boldsymbol{c}\|_{\ell_q^\tau}, \qquad (12)$$

and

$$\mathcal{K}_q^\tau(\mathcal{D}) = T\ell_q^\tau = \left\{ f \in X, \exists \boldsymbol{c}, f = \sum_k c_k g_k, \|\boldsymbol{c}\|_{\ell_q^\tau} < \infty \right\}$$

is a (quasi)Banach space which is continuously embedded in X.

Proof. By definition of $\mathcal{K}_q^\tau(\mathcal{D})$, for $f \in \mathcal{K}_q^\tau(\mathcal{D})$ there exists finite sequences $\boldsymbol{c}_n$, $n = 1, 2, \ldots$ such that $\|\boldsymbol{c}_n\|_{\ell_q^\tau} \leq |f|_{\mathcal{K}_q^\tau(\mathcal{D})} + 1/n$ and $\|f - T\boldsymbol{c}_n\|_X \leq 1/n$. The sequence $\{\boldsymbol{c}_n\}_{n\geq 1}$ is bounded in ℓ_q^τ, hence it is also bounded in ℓ^r where $\max(1,\tau) < r < p$. As ℓ^r is a reflexive Banach space, it is weakly compact and there exists a subsequence $\boldsymbol{c}_{n_k}$ converging weakly in ℓ^r to some $\boldsymbol{c} \in \ell^r$. Applying Fatou's lemma twice gives the estimate from above $\|\boldsymbol{c}\|_{\ell_q^\tau} \leq |f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. ¿From the weak convergence in ℓ^r and the continuity of $T : \ell^r \to X$ we get that $T\boldsymbol{c}_{n_k}$ converges weakly to $T\boldsymbol{c}$ in X. As we already know its strong limit in X is f, we obtain $f = T\boldsymbol{c}$ which gives the estimate from below $|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq \|\boldsymbol{c}\|_{\ell_q^\tau}$, and (12) is proved. In case $1 < \tau, q < \infty$, the Lorentz space ℓ_q^τ is strictly convex, and if $\boldsymbol{c}_0 \neq \boldsymbol{c}_1$ both realize the sparsity norm, we get $\|(\boldsymbol{c}_1 + \boldsymbol{c}_0)/2\|_{\ell_q^\tau} < |f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. As $T((\boldsymbol{c}_0 + \boldsymbol{c}_1)/2) = f$ this contradicts (12).

The equality $\mathcal{K}_q^\tau(\mathcal{D}) = T\ell_q^\tau$ follows directly from (12). Then we observe that for any $f \in \mathcal{K}_q^\tau(\mathcal{D})$ and an associated sequence $\boldsymbol{c}$, $|f|_{\mathcal{K}_q^\tau(\mathcal{D})} = \|\boldsymbol{c}\|_{\ell_q^\tau} \geq \|\boldsymbol{c}\|_{\ell_1^p} \geq C^{-1}\|f\|_X$ where the last inequality comes from the continuity of $T : \ell_1^p \to X$. The conclusion is reached using Remark 1 from the introduction. □

3.3 Thresholding-Jackson embedding

We can now prove Theorem 5. Some of the statements in Theorem 5 are almost trivial : from (2) and (3) it is obvious that (7) $\Rightarrow$ (8) $\Rightarrow$ (9). Using the same technique as [GN01, Proposition 4.1], we easily obtain (9) $\Rightarrow$ (10) as follows:

Proof. From the double embedding $\mathcal{K}_1^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_\infty^\alpha(\mathcal{D}) \hookrightarrow X$, $\tau < (\alpha + 1/p)^{-1}$, we have $\|\cdot\|_X \lesssim |\cdot|_{\mathcal{K}_1^\tau(\mathcal{D})}$. Thus we can check for $I_m \subset \mathbb{N}$ of cardinality $card(I_m) = m$, $\|\sum_{k\in I_m} \pm g_k\|_X \leq C m^{1/\tau}$ which by Proposition 2 gives that $\mathcal{D}$ is ℓ_1^τ-hilbertian. But as α can be arbitrarily close to 0, τ can be arbitrarily close to p and this gives the result. □

So far we have proved that the ℓ_1^p-hilbertian property of $\mathcal{D}$ is (almost) necessary for any Jackson embedding to hold for all $\alpha > 0$. Next we complete the proof of Theorem 5 by showing that (10) $\Rightarrow$ (7). Notice that Theorem 6 is a bit stronger.

Theorem 6 *For any $1 < p < \infty$, $\tau < p$, $0 < q \leq \infty$, there is a constant $C = C(\tau, q, p)$ such that for* any *ℓ_1^p-hilbertian dictionary $\mathcal{D}$ in* any *Banach space X, for all $f \in \mathcal{K}_q^\tau(\mathcal{D})$*

$$\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D})} \leq \|T\|_{\ell_1^p}^X C |f|_{\mathcal{K}_q^\tau(\mathcal{D})} \qquad \text{with } \tau = (\alpha + 1/p)^{-1} \tag{13}$$

where $\|L\|_Y^X$ denotes the operator norm of a continuous linear operator $L : Y \to X$.

Proof. Let $0 < \tau = (\alpha + 1/p)^{-1} < p$, $q \in (0, \infty]$ and $f \in \mathcal{K}_q^\tau(\mathcal{D})$. By Proposition 3 we can take $\boldsymbol{c} \in \ell_q^\tau$ such that $f = T\boldsymbol{c}$ and $\|\boldsymbol{c}\|_{\ell_q^\tau} = |f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. Let $\{\boldsymbol{c}_m\}$ the best m-term approximants to $\boldsymbol{c}$ from the canonical basis $\mathcal{B}$ of the sequence space ℓ_q^τ : $\boldsymbol{c}_m$ is obtained by thresholding $\boldsymbol{c} = \{c_k\}_{k\geq 1}$ to keep its m largest coefficients. Let $f_m(\pi, \{c_k^\star\}, \mathcal{D}) := T\boldsymbol{c}_m$ (where $\{c_k^\star\}$ is a decreasing rearrangement of $\boldsymbol{c}$, see Equation (1)). For $m \geq 1$,

$$\|f - f_m(\pi, \{c_k^\star\}, \mathcal{D})\|_X = \|T\boldsymbol{c} - T\boldsymbol{c}_m\|_X \leq \|T\|_{\ell_1^p}^X \sigma_m(\boldsymbol{c}, \mathcal{B})_{\ell_1^p}$$

and $\|f\|_X \leq \|T\|_{\ell_1^p}^X \|\boldsymbol{c}\|_{\ell_1^p}$. From standard results (see *e.g.* [DT96]) we get

$$\|\boldsymbol{c}\|_{\ell_1^p} + \left\|\{\sigma_m(\boldsymbol{c}, \mathcal{B})_{\ell_1^p}\}_{m\geq 1}\right\|_{\ell_q^{1/\alpha}} \leq C\|\boldsymbol{c}\|_{\ell_q^\tau}$$

where $C = C(\tau, q, p)$. Eventually we obtain

$$\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D})} \leq \|T\|_{\ell_1^p}^X C \|\boldsymbol{c}\|_{\ell_q^\tau} = \|T\|_{\ell_1^p}^X C |f|_{\mathcal{K}_q^\tau(\mathcal{D})}. \qquad \square$$

Remark 9 *When $\mathcal{D}$ is only ℓ^1-hilbertian, we loose the representation of $\mathcal{K}_q^\tau(\mathcal{D})$ (Proposition 3) because the weak compactness argument breaks down. Indeed, we have essentially no other description of $\mathcal{K}_1^1(\mathcal{D})$ than the fact that it is the closure of the convex hull of $\{\pm g, g \in \mathcal{D}\}$* (c.f. *Example 1 in the next section). It does not seem possible to extend Theorem 6 to $p = 1$.*

Notice that the Jackson embedding provided by Theorem 6 is **strong** : not only does it show that the best m-term error decays like $\mathcal{O}(m^{-\alpha})$ (this would be the standard Jackson inequality), indeed there is a "thresholding algorithm" that takes as input an (adaptive) sparse representation $\boldsymbol{c}$ of $f \in \mathcal{K}_q^\tau(\mathcal{D})$ and provides the rate of approximation

$$\|f - \sum_{k=1}^m c_k^\star g_{\pi_k(f)}\|_X \lesssim m^{-\alpha} |f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}, \qquad m \geq 1, \qquad \tau = (\alpha + 1/p)^{-1}.$$

The above inequality is not a standard Jackson inequality, we will denote it a **thresholding-Jackson inequality**. The rate of Chebyshev or best m-term approximation could be even larger, we would need an inverse estimate (a Bernstein-type embedding) to eliminate this possibility. This will be discussed in a forthcoming paper [GN02a].

3.4 Geometry of X *versus* structure of the dictionary

Let us now compare the Jackson embedding lines obtained so far, i.e. the lines given by (6) and (13), respectively. The goal is always to use the *best* possible Jackson estimate, which ensures the fastest convergence of the approximation algorithm. The examples below will show that it depends on the particular situation which line, (6) or (13), provides the best Jackson estimate. This of course implies that we need both estimates depending on the situation.

First we consider the case where the Chebyshev-Jackson line (6) "beats" the thresholding-Jackson line (13) for $\tau \leq 1$. Later we will consider the opposite situation. We have the following example.

Example 1 *Let X be a Banach space with $P_g(X) > 1$ [*e.g.* a Hilbert space $\mathcal{H}$ with $P_g(\mathcal{H}) = 2$] and $\mathcal{D}$ a dictionary with at least one accumulation point g [*e.g.* such as constructed for the proof of Theorem 2.3]. Combining Lemma 1 and Proposition 3 we see that $P_s(\mathcal{D}, X) = 1$. Thus, in this case the Chebyshev-Jackson embedding line is strictly better than the thresholding-Jackson one for $1/\tau \geq 1$. Moreover, for $1/\tau < 1$, no Jackson type embedding makes sense as $|\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$ is only a* **semi**-*(quasi) norm.*

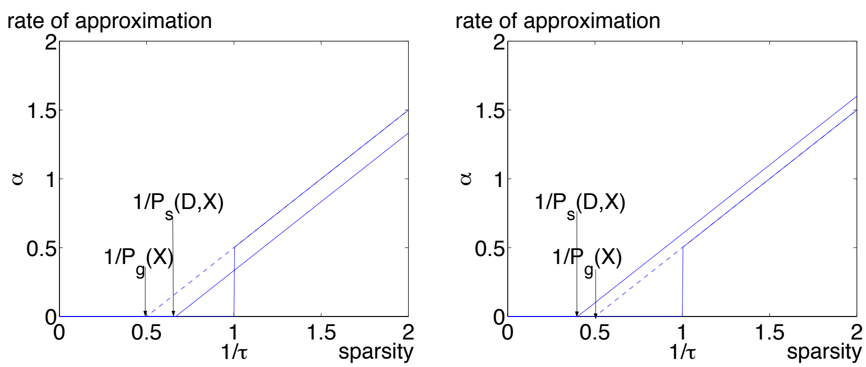

Figure 2: Comparison of the thresholding-Jackson embedding line and the Chebyshev-Jackson one. The leftmost graph depicts a situation where the thresholding-Jackson embedding is valid for a larger range of values of $1/\tau$ than the Chebyshev-Jackson one, but the latter is stronger on its range of validity. The rightmost corresponds to the opposite situation, where the thresholding-Jackson embedding line is better than the Chebyshev-Jackson one throughout its domain of validity.

The leftmost graph in Figure 2 illustrates a situation similar to that of Example 1, with the only difference that the graph depicts the case where $1 < P_s(\mathcal{D}, X) < P_g(X)$. The thresholding-Jackson embedding line is valid for a larger range of values of $1/\tau$ than the Chebyshev-Jackson one, but the latter is stronger on its range of validity. The opposite situation, where the thresholding-Jackson embedding line is better than the Chebyshev-Jackson one throughout its domain, is also possible. This particular situation is illustrated on the rightmost graph in Figure 2, and an explicit example will be given in Section 4 (see Example 2).

4 Examples of hilbertian dictionaries

In this section we will consider the ℓ_1^p-hilbertian property of several classical types of redundant dictionaries, first in L^p spaces, and then in other classical functional spaces (Besov spaces $B_\tau^\alpha(L^\tau(\mathbb{R}))$, modulation spaces

$M^p_w(\mathbb{R})$). We refer the reader to [Grö00, Chapters 11–12] for the basic definition and properties of weighted modulation spaces $M^p_w(\mathbb{R})$ with α-moderate weight $w(x,y)$. For the trivial weight $w \equiv 1$, we denote $M^p(\mathbb{R})$ instead of $M^p_w(\mathbb{R})$. For the definition of Besov spaces we refer to [Tri83].

At the end of the example section (section 4.5) we study the situation where the dictionary $\mathcal{D}$ is obtained by taking the union of two smaller (and possibly classical) ℓ^p_1-hilbertian dictionaries. This leads to new bigger nonclassical sparsity spaces for which there is a Jackson embedding.

4.1 Interpolation of hilbertian dictionaries

First we consider two general lemmas that will make it easier to check the ℓ^p-hilbertian property for many well-known dictionaries in classical functional spaces. For notational convenience, whenever we write $\mathcal{K}^\tau_q(\mathcal{D}, X)$, we always assume implicitly that $\mathcal{D}$ has been normalized in X. Lemma 2 deals with wavelet-type dictionaries in $L^p(\Omega)$ where Ω is some σ-finite measure space. Lemma 2 will be used with time-frequency dictionaries in modulation spaces $M^p_w(\mathbb{R})$.

Lemma 2 *Let $1 < r < \infty$ and $\mathcal{D} = \{g_k, k \in \mathbb{N}\}$ an ℓ^r_1-hilbertian (normalized) dictionary in $X = L^r(\Omega)$, and assume that every g_k is in $L^1(\Omega)$. Suppose that for every $1 \le p \le r$ there is some constant $C = C(p)$ such that for every g_k, with $\theta_{r,p} = (1-1/p)/(1-1/r)$,*

$$\|g_k\|^{1-\theta_{r,p}}_{L^1(\Omega)} \le C\|g_k\|_{L^p(\Omega)}. \tag{14}$$

Then $\mathcal{D}$ (properly normalized in $L^p(\Omega)$) is ℓ^p-hilbertian in $L^p(\Omega)$ for $1 \le p < r$. Moreover, if $|\Omega| < \infty$ and $\mathcal{D}$ has dense span in $L^r(\Omega)$, then $\mathcal{D}$ has dense span in $L^p(\Omega)$, $1 \le p \le r$.

Proof. By assumption, T is continuous from ℓ^r to $L^r(\Omega)$. Then, as $\mathcal{D}$ (normalized in $L^1(\Omega)$) is ℓ^1-hilbertian in $L^1(\Omega)$, T is also continuous from the weighted space $\ell^1(w)$ to $L^1(\Omega)$, where $w_k = \|g_k\|_{L^1(\Omega)}$. Hence T is continuous from the interpolation space $(\ell^1(w), \ell^r)_{\theta_{r,p},p}$ to the interpolation space $(L^1(\Omega), L^r(\Omega))_{\theta_{r,p},p} = L^p(\Omega)$ (for details on the real method of interpolation, we refer to [DL93, Chap. 7]). By Stein's theorem on interpolation of weighted ℓ^p spaces [BS88, p. 213], T is thus continuous from $\ell^p(w^{1-\theta_{r,p}})$ to $L^p(\Omega)$, hence for some constant $C' = C'(p) < \infty$

$$\begin{aligned}\Big\|\sum_k c_k \frac{g_k}{\|g_k\|_{L^p(\Omega)}}\Big\|_{L^p(\Omega)} &\le C'\Big\|\{c_k/\|g_k\|_{L^p(\Omega)}\}\Big\|_{\ell^p(w^{1-\theta_{r,p}})} \\ &\le C'\Big\|\{c_k \cdot (\|g_k\|^{1-\theta_{r,p}}_{L^1(\Omega)}/\|g_k\|_{L^p(\Omega)})\}\Big\|_{\ell^p} \le C'C\|\{c_k\}\|_{\ell^p}.\end{aligned}$$

The denseness claim follows from standard arguments using Hölder's inequality and the fact that, *e.g.*, the continuous functions are dense in both $L^r(\Omega)$ and $L^p(\Omega)$. □

Modulation spaces are better suited than L^p spaces when we consider nonlinear approximation properties of time-frequency dictionaries. The family of modulation spaces $M^p_w(\mathbb{R})$, for a given α-moderate weight w, is an interpolation family [BS88, Fei83, Grö00]. As a result, we can copy the proof of Lemma 2 to get an analogue result where $L^p(\Omega)$ is replaced with $M^p_w(\mathbb{R})$.

Lemma 3 *Let $1 < r < \infty$ and $\mathcal{D} = \{g_k, k \in \mathbb{N}\}$ an ℓ^r_1-hilbertian (normalized) dictionary in $X = M^r_w(\mathbb{R})$, and assume that every g_k is in $M^1_w(\mathbb{R})$. Suppose that for every $1 \le p \le r$ there is some constant $C = C(p)$ such that for every g_k, with $\theta_{r,p} = (1-1/p)/(1-1/r)$,*

$$\|g_k\|^{1-\theta_{r,p}}_{M^1_w(\mathbb{R})} \le C\|g_k\|_{M^p_w(\mathbb{R})}. \tag{15}$$

Then $\mathcal{D}$ (properly normalized in $M^p_w(\mathbb{R})$) is an ℓ^p-hilbertian dictionary in $M^p_w(\mathbb{R})$, $1 \le p < r$.

In Sections 4.2-4.4 we check the assumptions of these Lemmas on classical dictionaries and apply Theorem 6.

4.2 Wavelet type systems in $L^p(\mathbb{R}^d)$

Consider $\mathcal{D}$ a (bi)orthogonal wavelet basis [Coh92, Dau92, Mal98] or a (tight) wavelet frame [RS97b, RS97a, DHRS01] for $L^2(\mathbb{R}^d)$

$$\psi_{j,k}^{\ell}(x) := 2^{jd/2}\psi^{\ell}(2^j x - k), \qquad 1 \le \ell \le L,\ j \in \mathbb{Z}, k \in \mathbb{Z}^d.$$

As $\|\psi_{j,k}^{\ell}\|_{L^p(\mathbb{R}^d)} = 2^{jd(1/2-1/p)}\|\psi^{\ell}\|_{L^p(\mathbb{R}^d)}$, one can check, for $1 \le p \le \infty$,

$$\|\psi_{j,k}^{\ell}\|_{L^1(\mathbb{R}^d)}^{1-\theta_{2,p}} = 2^{jd(1/2-1/p)}\|\psi^{\ell}\|_{L^1(\mathbb{R}^d)}^{1-\theta_{2,p}} = \frac{\|\psi^{\ell}\|_{L^1(\mathbb{R}^d)}^{1-\theta_{2,p}}}{\|\psi^{\ell}\|_{L^p(\mathbb{R}^d)}}\|\psi_{j,k}^{\ell}\|_{L^p(\mathbb{R}^d)},$$

so (14) holds with $C(p) = \max_{\ell=1}^{L} \|\psi^{\ell}\|_{L^1(\mathbb{R})}^{1-\theta_{2,p}}/\|\psi^{\ell}\|_{L^p(\mathbb{R})}$. Thus Lemma 2 applies with $r = 2$, and we get that such systems are ℓ^p-hilbertian in $L^p(\mathbb{R}^d)$ for any $1 \le p \le 2$.

One needs to check in each case whether the $L^p(\mathbb{R}^d)$ normalized system is actually dense in $L^p(\mathbb{R}^d)$. This may be a highly nontrivial question in the frame case, see *e.g.* [Mey92, Chap. 4]. The (bi)orthogonal wavelet systems are dense in $L^p(\mathbb{R}^d)$, $1 < p < \infty$, assuming mild decay of the generators [Woj99, Pom00].

For $\mathcal{D}$ a (bi)orthogonal wavelet system normalized in $L^p(\mathbb{R}^d)$, Proposition 3 together with Theorem 6 shows that

$$\mathcal{K}_q^{\tau}(\mathcal{D}, L^p(\mathbb{R}^d)) \hookrightarrow \mathcal{T}_q^{\alpha}(\mathcal{D}), \qquad \tau = (\alpha + 1/p)^{-1}, \qquad 1 < p \le 2,$$

and if $\mathcal{D}$ and its dual system have sufficient smoothness and vanishing moments it is known (see, *e.g.*, [DJP92]) that $\mathcal{K}_{\tau}^{\tau}(\mathcal{D}, L^p(\mathbb{R}^d))$ can be identified with the Besov space $B_{\tau}^{d\alpha}(L^{\tau}(\mathbb{R}^d))$ for $\tau = (\alpha + 1/p)^{-1}$.

Remark 10 *For wavelet like systems, one can sometimes extend the result obtained from Lemma 2 and show that such systems are ℓ_1^p-hilbertian in $L^p(\mathbb{R}^d)$ for $1 < p < \infty$ using the special structure of the functions and the theory of Calderón-Zygmund operators, see [GN02b, Theorem 4.11].*

The following example shows that wavelet systems give an illustration of the rightmost graph on Figure 2.

Example 2 *Consider a normalized basis $\mathcal{B}$ of MRA wavelets (with isotropic dilation) in $X := L^p(\mathbb{R}^d)$, $1 < p < \infty$. It is known that the powertype of $L^p(\mathbb{R}^d)$ is $P_g(L^p(\mathbb{R}^d)) = \min\{2, p\}$ [LT79]. Moreover, it can be verified that $\mathcal{B}$ is ℓ^p-hilbertian ($1 < p \le 2$) or ℓ_1^p-hilbertian ($2 < p < \infty$), hence $P_s(\mathcal{B}, X) = p$, $1 < p < \infty$.*

- *For $1 < p \le 2$ we have $P_s(\mathcal{B}, X) = p = P_g(X)$.*
- *For $2 < p < \infty$, this gives $P_s(\mathcal{B}, X) = p > 2 = P_g(X)$.*

Hence for all $1 < p < \infty$, the thresholding-Jackson embedding line for dictionaries of this type is given by $\alpha = 1/\tau - 1/p$ for all $0 < \tau < p$. For $2 < p < \infty$ the thresholding-Jackson embedding line is strictly *better than the Chebyshev-Jackson embedding line, which corresponds exactly to the rightmost graph on Figure 2.*

4.3 Gabor frames in $L^p(\mathbb{R})$ and $M^p(\mathbb{R})$

Not all interesting dictionaries live in L^p: in the following we concentrate on time-frequency dictionaries, for which the natural function spaces are rather the modulation spaces. We consider the ℓ^p-hilbertian property of such dictionaries, both in $L^p(\mathbb{R})$ and in the modulation spaces $M^p(\mathbb{R})$ with the trivial weight $w \equiv 1$.

A Gabor dictionary $\mathcal{D}$ consists of the functions $g_{n,m}(x) := g(x - na)e^{2i\pi mbx}$, $n, m \in \mathbb{Z}$. Provided that g is an appropriate "window" function and $a, b > 0$ appropriate lattice parameters (see, *e.g.*, [Dau92, Grö00, Chr02]), $\mathcal{D}$ is a frame in $L^2(\mathbb{R}) = M^2(\mathbb{R})$. It satisfies the relations $\|g_{n,m}\|_{L^p(\mathbb{R})} = \|g\|_{L^p(\mathbb{R})}$ and $\|g_{n,m}\|_{M^p(\mathbb{R})} = \|g\|_{M^p(\mathbb{R})}$, for $0 < p \le \infty$. Hence, assuming that $g \in L^1(\mathbb{R}) \cap M^1(\mathbb{R})$, $\mathcal{D}$ is simultaneously (quasi)normalized in all $L^p(\mathbb{R})$ and $M^p(\mathbb{R})$. Moreover, as a frame, $\mathcal{D}$ is ℓ_1^2-hilbertian in $L^2(\mathbb{R}) = M^2(\mathbb{R})$, hence we can apply Lemma 2 and Lemma 3 to get the following results

Proposition 4 *Let $\mathcal{D}$ a Gabor frame normalized in $L^2(\mathbb{R})$, with window $g \in L^1(\mathbb{R}) \cap M^1(\mathbb{R})$. Let $0 < \tau < 2$ and $0 < q \leq \infty$. We have, with equivalent norms, for all $p \geq 1$ such that $\tau < p \leq 2$,*

$$\mathcal{K}_q^\tau\big(\mathcal{D}, L^p(\mathbb{R})\big) = \mathcal{K}_q^\tau\big(\mathcal{D}, M^p(\mathbb{R})\big) = \mathcal{K}_q^\tau\big(\mathcal{D}, M^2(\mathbb{R})\big) = \mathcal{K}_q^\tau\big(\mathcal{D}, L^2(\mathbb{R})\big).$$

It follows that

$$\mathcal{K}_q^\tau\big(\mathcal{D}, L^2(\mathbb{R})\big) \hookrightarrow \mathcal{T}_q^\alpha\big(\mathcal{D}, L^p(\mathbb{R})\big), \qquad \alpha = 1/\tau - 1/p.$$

and

$$\mathcal{K}_q^\tau\big(\mathcal{D}, L^2(\mathbb{R})\big) \hookrightarrow \mathcal{T}_q^\alpha\big(\mathcal{D}, M^p(\mathbb{R})\big), \qquad \alpha = 1/\tau - 1/p.$$

Notice that for the extremal cases of Proposition 4, corresponding to $p = 1$ and $p = 2$, we have $\mathcal{K}_1^1(\mathcal{D}, L^2(\mathbb{R})) \subset M^1(\mathbb{R})$ and $\mathcal{K}_2^2(\mathcal{D}, L^2(\mathbb{R})) = M^2(\mathbb{R}) = L^2(\mathbb{R})$, respectively.

For a general α-moderate weight w and an arbitrary window g, Lemma 3 cannot be applied directly since it is not clear whether we have the estimate (15) for the dictionary $\mathcal{D}$ normalized in $M_w^2(\mathbb{R})$. In the following section we deal with this situation, assuming that the window g has more structure.

4.4 Gabor Banach frames in $M_w^p(\mathbb{R})$, $1 \leq p < \infty$

Proposition 4 can be extended and becomes more interesting for Gabor frames with a bit more structure. Let $\mathcal{D} = \{g_{n,m},\ n, m \in \mathbb{Z}\}$ a Gabor frame generated by a "nice" window function g and with small enough lattice parameters a, b. From the atomic decomposition theory for $M_w^p(\mathbb{R})$ (see [Grö00, Sec. 12.2.] for details), $\mathcal{D}$ constitutes a Banach frame for M_w^p, that is to say there exist a dual window function $\widetilde{g}$ that generates a dual Gabor frame $\widetilde{\mathcal{D}} = \{\widetilde{g}_{n,m},\ n, m \in \mathbb{Z}\}$ such that for all α-moderate weights w and $1 \leq p \leq \infty$ $\|f\|_{M_w^p(\mathbb{R})} \asymp \|\{\langle f, \widetilde{g}_{n,m}\rangle w_{n,m}\}_{n,m}\|_{\ell^p}$ with $w_{n,m} = w\big(an, bm\big)$. Moreover, for $1 \leq p < \infty$ the Gabor expansion $f = \sum_{n,m} \langle f, \widetilde{g}_{n,m}\rangle g_{n,m}$ converges unconditionally in the norm of $M_w^p(\mathbb{R})$ for every $f \in M_w^p(\mathbb{R})$, and the synthesis operator $T\boldsymbol{c} = \sum_{n,m} c_{n,m} g_{n,m}$ is bounded from $\ell^p(w)$ to M_w^p [Grö00, Theorem 12.2.4], hence $\mathcal{D}$ (normalized in M_w^p) is ℓ^p-hilbertian .

By the Gabor expansion, we have $M_w^\tau(\mathbb{R}) \hookrightarrow \mathcal{K}_\tau^\tau(\mathcal{D}, M_w^p)$ for Gabor Banach frames and $1 \leq \tau < p < \infty$. The converse embedding, $\mathcal{K}_\tau^\tau(\mathcal{D}, M_w^p) \hookrightarrow M_w^\tau(\mathbb{R})$, follows by Proposition 3 : we expand each $f \in \mathcal{K}_\tau^\tau(\mathcal{D}, M_w^p)$ as $f = \sum_{m,n} c_{m,n} g_{m,n} / \|g_{n,m}\|_{M_w^p}$ with $\|\boldsymbol{c}\|_{\ell^\tau} = |f|_{\mathcal{K}_\tau^\tau(\mathcal{D}, M_w^p)}$, and then use the boundedness of the synthesis operator T to obtain $\|f\|_{M_w^\tau} \leq \|\boldsymbol{c}\|_{\ell^\tau}$. So Theorem 6 recovers [GS00, Proposition 3] as a corollary:

Proposition 5 *Let $\mathcal{D}$ a Gabor Banach frame, then for all α-moderate weights w and $1 \leq \tau < p < \infty$,*

$$M_w^\tau(\mathbb{R}) = \mathcal{K}_\tau^\tau(\mathcal{D}, M_w^p) \hookrightarrow \mathcal{T}_\tau^\alpha(\mathcal{D}, M_w^p), \qquad \alpha = 1/\tau - 1/p$$

where the first equality is with equivalent norms.

4.5 Union of dictionaries

To conclude this section on examples, let us mention a simple and straightforward corollary of Proposition 3.

Corollary 1 *Assume $\mathcal{D}_1$ and $\mathcal{D}_2$ are both ℓ_1^p-hilbertian dictionaries, with $p > 1$. Let $0 < \tau < p$ and $1 \leq q \leq \infty$. Then for $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$*

$$\mathcal{K}_q^\tau(\mathcal{D}, X) = \mathcal{K}_q^\tau(\mathcal{D}_1, X) + \mathcal{K}_q^\tau(\mathcal{D}_2, X).$$

Thus, whenever the individual sparsity spaces $\mathcal{K}_q^\tau(\mathcal{D}_1)$ and $\mathcal{K}_q^\tau(\mathcal{D}_2)$ do not coincide, we gain by using the redundant dictionary $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ in the sense that the domain(s) $\mathcal{T}_q^\alpha$, $\mathcal{C}_q^\alpha$ and $\mathcal{A}_q^\alpha$ for which there is a direct estimate are strictly enlarged.

Consider, for example, the case of $\mathcal{D}_1$ a wavelet basis and $\mathcal{D}_2$ a local Fourier basis [CM91, DJJ91] in $X = L^2(\mathbb{R})$. The individual sparsity spaces are respectively $\mathcal{K}_\tau^\tau(\mathcal{D}_1, X) = B_\tau^\alpha(L^\tau(\mathbb{R})) = \mathcal{A}_\tau^\alpha(\mathcal{D}_1, X)$ and $\mathcal{K}_\tau^\tau(\mathcal{D}_2, X) = M^\tau(\mathbb{R}) = \mathcal{A}_\tau^\alpha(\mathcal{D}_2, X)$ [GS00, Theorem 2]. In this particular case, we have $\mathcal{K}_\tau^\tau(\mathcal{D}_1 \cup \mathcal{D}_2, X)) = B_\tau^\alpha(L^\tau(\mathbb{R})) + M^\tau(\mathbb{R})$.

5 Conclusion

We have introduced and studied approximation classes associated with m-term thresholding and Chebychev approximation, respectively, with elements from a (possibly redundant) dictionary in a Banach space. The Chebyshev approximation class has been shown to be a linear (quasi)normed space, and the classes have been compared to the "benchmark" approximation class associated with the best m-term approximation. Different types of Jackson embedding results (direct estimates) of sparsity spaces into approximation classes have been studied in detail. We have considered three types of direct theorems, and how they are related. A completely general (and thus weak) estimate that applies to any situation has been derived, the second type of result is based on the geometry of the Banach space, while the third type of Jackson embedding relies on hilbertian properties of the dictionary. From the hilbertian property of a dictionary, we have also derived a simple representation of the sparsity spaces. Many examples are given with dictionaries in L^p and modulation spaces, and we have demonstrated how to apply the general theory to recover several well known results on nonlinear approximation with wavelet, local Fourier, and Gabor systems, respectively.

However, we should stress that the main attraction of the theory is not that it can recover already known results, but that it provides us with direct estimates for many new function classes that are often "bigger" than the classical smoothness spaces. One problem not addressed in this paper, is how to obtain a complete characterization of the different approximation classes in the spirit of Theorem 1. To get a similar characterization, as the one given by Theorem 1, we need a Bernstein embedding of the type $\mathcal{A}_q^\alpha(\mathcal{D}, X) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D}, X)$ for suitable values of α and τ. The problem of obtaining such an embedding is well known to be closely related to deriving an inverse (or Bernstein) inequality. We should also note that a complete characterization of the other types of approximation classes considered in this paper can be obtaining by proving weaker (and non-classical) embeddings of the type $\mathcal{T}_q^\alpha(\mathcal{D}, X) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D}, X)$ or $\mathcal{C}_q^\alpha(\mathcal{D}, X) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D}, X)$.

Examples in [GN01] show that a Bernstein inequality cannot hold unless the dictionary is very well structured. The problem of obtaining Bernstein estimates for structured redundant dictionaries is studied in detail in a forthcoming paper [GN02a] by the authors.

Acknowledgments

The authors would like to thank Ronald A. DeVore for introducing us to most of the problems discussed in this paper. We would also like to thank Hans G. Feichtinger and Karlheinz Gröchenig for discussions about modulation spaces while the first author visited the Gabor workshop, and Ole Christensen for helping us make some connections between our results and Banach frames.

References

[BS88] Colin Bennett and Robert Sharpley. *Interpolation of operators*. Academic Press Inc., Boston, MA, 1988.

[CDDR01] A. Cohen, W. Dahmen, I. Daubechies, and DeVore. R. Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11(2):192–226, 2001. doi:10.1006/acha.2001.0336.

[Chr02] Ole Christensen. *An introduction to frames and Riesz bases*. 2002. Manuscript.

[CM91] R. R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *Comptes-Rendus Acad. Sci. Paris(A)*, 312:259–261, 1991.

[Coh92] Albert Cohen. Biorthogonal wavelets. In Charles K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, volume 2, pages 123–152. Academic Press, Boston, MA, 1992.

[Dau92] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[DeV98] Ronald A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[DHRS01] Ingrid Daubechies, Bin Han, Amos Ron, and Zuowei Shen. Framelets: MRA-based constructions of wavelet frames. *Preprint*, 2001.

[DJJ91] I. Daubechies, S. Jaffard, and J.L. Journé. A simple Wilson orthonormal basis with exponential decay. *SIAM J. Math. Anal.*, 22:554–572, 1991.

[DJP92] Ronald A. DeVore, Björn Jawerth, and Vasil Popov. Compression of wavelet decompositions. *Amer. J. Math.*, 114(4):737–785, 1992.

[DKKT01] S.J. Dilworth, N.J. Kalton, D. Kutzarova, and V.N. Temlyakov. The thresholding greedy algorithm, greedy bases, and duality. Technical Report 0123, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2001.

[DL93] Ronald A. DeVore and George G. Lorentz. *Constructive approximation*. Springer-Verlag, Berlin, 1993.

[DMA97] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constr. Approx.*, 13(1):57–98, 1997.

[DT96] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Adv. Comput. Math.*, 5(2-3):173–187, 1996.

[Fei83] H. G. Feichtinger. Modulation spaces on locally compact abelian groups. Technical report, University of Vienna, 1983.

[FT74] J.H. Friedman and J.W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. on Computers*, C-23:881–889, 1974.

[GES91] B. Golubov, A. Efimov, and V. Skvortsov. *Walsh series and transforms*. Theory and applications, Translated from the 1987 Russian original by W. R. Wade. Kluwer Academic Publishers Group, Dordrecht, 1991.

[GN01] R. Gribonval and M. Nielsen. Some remarks on nonlinear approximation with Schauder bases. *East J. Approx.*, 7(3):267–285, 2001.

[GN02a] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. II. Inverse estimates. Technical report, IMI, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2002. in preparation.

[GN02b] R. Gribonval and M. Nielsen. On approximation with spline generated framelets. *Constr. Approx.* (to appear).

[Grö00] Karlheinz Gröchenig. *Foundations of Time-Frequency Analysis (Applied Numerical and Harmonic Analysis)*. Birkhauser, December 2000. ISBN: 0817640223.

[GS00] K. Gröchenig and S. Samarah. Nonlinear approximation with local fourier bases. *Constr. Approx.*, 16(3):317–332, 2000.

[Hub85] P. J. Huber. Projection pursuit. *Ann. Statist.*, 13(2):435–475, 1985.

[Jon87] L.K. Jones. On a conjecture of Huber concerning the convergence of PP-regression. *Ann. Statist.*, 15:880–882, 1987.

[Jon92] Lee K. Jones. A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training. *Ann. Statist.*, 20(1):608–613, 1992.

[Jon97] Lee K. Jones. The computational intractability of training sigmoidal neural networks. *IEEE Trans. Inform. Theory*, 43(1):167–173, 1997.

[KP01] G. Kerkyacharian and D. Picard. Entropy, universal coding, approximation and bases properties. Technical Report Preprint No. 663, Universities Paris 6 & 7, 2001.

[KT99] S.V. Konyagin and V.N. Temlyakov. A remark on greedy approximation in Banach spaces. *East J. Approx.*, 5(3):365–379, 1999.

[LT79] J. Lindenstrauss and L. Tzafriri. *Classical Banach Spaces II.* Springer, 1979.

[Mal98] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[Mey92] Yves Meyer. *Wavelets and operators.* Cambridge University Press, Cambridge, translated from the 1990 french original by d. h. salinger edition, 1992.

[Pis81] G. Pisier. Remarques sur un résultat non publié de B. Maurey. In *Seminar on Functional Analysis, 1980–1981*, pages Exp. No. V, 13. École Polytech., Palaiseau, 1981.

[Pom00] W. Pompe. Unconditional biorthogonal wavelet bases in $L^p(\mathbb{R}^d)$. *Colloq. Math.*, 92(1):19-34, 2002.

[RS97a] Amos Ron and Zuowei Shen. Affine systems in $L_2(\mathbb{R}^d)$. II. Dual systems. *J. Fourier Anal. Appl.*, 3(5):617–637, 1997. Dedicated to the memory of Richard J. Duffin.

[RS97b] Amos Ron and Zuowei Shen. Affine systems in $L_2(\mathbb{R}^d)$: the analysis of the analysis operator. *J. Funct. Anal.*, 148(2):408–447, 1997.

[Sam98] S. Samarah. *Approximation Theory and Modulation Spaces.* PhD thesis, University of Connecticut, 1998.

[Ste55] S. B. Stechkin. On absolute convergence of orthogonal series. *Dok. Akad. Nauk SSSR*, 102:37–40, 1955.

[Tem00] V.N. Temlyakov. Greedy algorithms in Banach spaces. *Adv. Comput. Math.*, 14(3):177-292, 2001.

[Tem01] V.N. Temlyakov. Nonlinear methods of approximation. Technical Report 0109, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2001.

[Tri83] H. Triebel. Theory of function spaces. Birkhäuser Verlag, Basel, 1983.

[Woj99] P. Wojtaszczyk. Wavelets as unconditional bases in $L_p(\mathbb{R})$. *J. Fourier Anal. Appl.*, 5(1):73–85, 1999.

[Woj00] P. Wojtaszczyk. Greedy algorithm for general biorthogonal systems. *J. Approx. Theory*, 107(2):293–314, 2000.