IMI

Preprint Series

Industrial Mathematics Institute

2002:03

Nonlinear approximation with dictionaries. I. Direct estimates

R. Gribonval and M. Nielsen

Department of Mathematics
University of South Carolina

Nonlinear approximation with dictionaries. I. Direct estimates

Rémi Gribonval

IRISA-INRIA, Campus de Beaulieu, 35042 Rennes cedex, France
E-mail: remi.gribonval@inria.fr

Morten Nielsen

Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA
E-mail: nielsen@math.sc.edu

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 obtained by considering m-term approximation with algorithmic constraints: thresholding and Chebychev approximation classes are studied respectively. Several embeddings of the Jackson type hold between the mentioned approximation classes and associated generalized smoothness spaces. The general direct estimates are based on the geometry of the Banach space, and we discuss when stronger results can be obtained by assuming a certain structure of the dictionary. We give several examples of classical dictionaries where our results give some new Jackson type estimates and recover some known estimates in L^p spaces and modulation spaces.

Key Words: Redundant dictionaries, nonlinear m-term approximation, approximation class, thresholding algorithm, greedy algorithm, sparse decomposition, generalized smoothness class, Jackson inequality, hilbertian property

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 structure of the paper is as follows. In Section 1 we introduce the approximation classes we want to consider and compare; they are the classes associated with best m-term approximation, thresholding approximation, and finally Chebyshev approximation. We make a comparison of the classes and introduce generalized smoothness spaces at the end of the section.

In Section 2, we consider Jackson type embeddings of the smoothness spaces 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 geometric 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.

Section 3 contains a study of so-called hilbertian dictionaries. We analyze the smoothness spaces associated with such dictionaries and often we can give a complete characterization of the smoothness spaces in terms of sequence spaces. 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.

Several 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.

In Section 5 we briefly study inverse estimates and show that a Bernstein inequality along with the existence of an adaptive analysis operator gives a complete characterization of (all) the approximation classes in terms of generalized smoothness spaces.

Finally there is a conclusion where, among other things, encoding of generalized smoothness spaces is discussed. An appendix concludes the paper, in which we study structural properties of generalized smoothness spaces and of approximation classes.

1. APPROXIMATION CLASSES

Below is a description of the approximation classes we consider and compare in this paper. The first class is the benchmark for the rest; the class associated with best m-term approximation.

1.1. Best m-term approximation

The (nonlinear) set of all linear combinations of at most m elements from $\mathcal{D}$ is

$$\Sigma_m(\mathcal{D}) := \left\{ \sum_{k \in I_m} c_k g_k, I_m \subset \mathbb{N}, |I_m| \leq m, c_k \in \mathbb{C} \right\}. \tag{1}$$

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

$$\sigma_m\left(f, \mathcal{D}\right)_X := \inf_{h \in \Sigma_m(\mathcal{D})} \left\| f - h \right\|_X. \tag{2}$$

Best m-term approximation will serve as a benchmark for the various other approximation algorithms that we will consider in this paper. The *best m-term approximation classes* are defined as :

$$\mathcal{A}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} < \infty \right\} \tag{3}$$

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. Section A.1 in the appendix. 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].

1.2. 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 :

$$f_m(\pi, \{c_k\}, \mathcal{D}) := \sum_{k=1}^{m} c_k g_{\pi_k} \tag{4}$$

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] :

$$\mathcal{T}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} < \infty \right\}, \quad (5)$$

with

$$|f|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} := \inf_{\pi, \{c_k^\star\}} \left(\sum_{m=1}^{\infty} \left(\left[m^\alpha \left\| f - f_m(\pi, \{c_k^\star\}, \mathcal{D}) \right\|_X \right]^q \frac{1}{m} \right) \right)^{1/q}, \quad (6)$$

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 1. 1. 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.

1.3. Chebyshev approximation

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

$$\mathcal{V}_m(\pi, \mathcal{D}) := \operatorname{span}(g_{\pi_1}, \ldots, g_{\pi_m}) \quad (7)$$

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

$$\mathcal{C}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X, \|f\|_{\mathcal{C}_q^\alpha(\mathcal{D},X)} := \|f\|_X + |f|_{\mathcal{C}_q^\alpha(\mathcal{D},X)} < \infty \right\} \quad (8)$$

where

$$|f|_{\mathcal{C}_q^\alpha(\mathcal{D},X)} := \inf_\pi \left\| \{ \| f - P_{\mathcal{V}_m(\pi,\mathcal{D})} f \|_X \}_{m \geq 1} \right\|_{\ell_q^{1/\alpha}}, \tag{9}$$

with the obvious modification for $q = \infty$. It turns out, and we prove it in Section A.3 in the appendix, that $\mathcal{C}_q^\alpha(\mathcal{D}, X)$ is indeed a linear subspace of X, and the quantity $\|\cdot\|_{\mathcal{C}_q^\alpha(\mathcal{D},X)}$ is a (quasi)norm.

1.4. Characterization of the approximation classes

So far we do not claim that the quantity $\|\cdot\|_{\mathcal{T}_q^\alpha(\mathcal{D},X)}$ 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

$$\mathcal{T}_q^\alpha(\mathcal{D}, X) \subset \mathcal{C}_q^\alpha(\mathcal{D}, X) \subset \mathcal{A}_q^\alpha(\mathcal{D}, X) \subset X \tag{10}$$

together with the inequalities

$$|\cdot|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} \lesssim |\cdot|_{\mathcal{C}_q^\alpha(\mathcal{D},X)} \lesssim |\cdot|_{\mathcal{T}_q^\alpha(\mathcal{D},X)} \tag{11}$$

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 1. 2.

1. In most of this paper, $\mathcal{A}_q^\alpha(\mathcal{D}, X)$ will be denoted for short by $\mathcal{A}_q^\alpha(\mathcal{D})$, and similar shorthands will be used for the other classes.
2. We will reserve the notation $\|\cdot\|_V$ to the "nondegenerate" case when $\|f\|_V = 0 \Rightarrow f = 0$.

In a very special case Stechkin, DeVore, and Temlyakov have shown that all the above approximation classes are indeed identical and have a nice characterization.

Theorem 1.1 ([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 $0 < q \leq \infty$,*

$$\mathcal{A}_q^\alpha(\mathcal{B}, \mathcal{H}) = \mathcal{T}_q^\alpha(\mathcal{B}, \mathcal{H}) = \mathcal{K}_q^\tau(\mathcal{B}, \mathcal{H}) \tag{12}$$

with equivalent (quasi)norms, where

$$\mathcal{K}_q^\tau(\mathcal{B}, \mathcal{H}) := \left\{ f \in \mathcal{H}, |f|_{\mathcal{K}_q^\tau(\mathcal{B},\mathcal{H})} := \|\{\langle f, g_k \rangle\}_{k \geq 1}\|_{\ell_q^\tau} < \infty \right\}. \tag{13}$$

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.

Theorem 1.1 essentially characterizes the approximation classes in terms of the decay (or **sparsity**) of the coefficients $\{c_k(f)\}$ of f in a well structured basis. For $\mathcal{B}$ an orthonormal wavelet basis the decay properties (for $q = \tau$) are directly related to the smoothness of f in certain Besov spaces $B^\alpha_\tau(L^\tau(\mathbb{R}))$ [DJP92]. For $\mathcal{B}$ a local Fourier basis [CM91] with a certain restriction on the partition of the time axis, the decay (for $q = \tau$) corresponds to smoothness measured in terms of *modulation spaces* M_τ [GS00]. Modulation spaces were originally introduced by Feichtinger in the context of Gabor analysis and are discussed in details in [Grö00].

1.5. Abstract smoothness spaces

For general *redundant* dictionaries, there is not a unique decomposition $f = \sum_k c_k(f) g_k$, and it is not obvious what should be a good measure of the sparsity of "the" coefficients of f. For frames in a Hilbert space, there is a simple representation $f = \sum_k c_k(f) g_k$ where $c_k(f) = \langle f, \widetilde{g_k}\rangle$ with $\widetilde{\mathcal{D}} = \{\widetilde{g_k}\}$ the dual frame. It is a well known fact [DS52] that the **analysis coefficients** (or frame coefficients) $\{\langle f, \widetilde{g_k}\rangle\}$ minimize the ℓ^2 norm over all possible expansions $f = \sum_k c_k g_k$. However, analysis coefficients are in general not adapted to the aim of characterizing approximation classes in terms of sparsity. This is illustrated with the following simple example.

Example 1.1. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ two orthonormal bases in a Hilbert space $\mathcal{H}$, and $\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2$. Then $\mathcal{D}$ is a tight frame with the analysis coefficients $c_k(f) = \langle f, g_k\rangle /2$. On the one hand, as $\mathcal{A}^\alpha_q(\mathcal{D})$ is a linear space [DL93], it is clear that it contains at least $\mathcal{A}^\alpha_q(\mathcal{B}_1) + \mathcal{A}^\alpha_q(\mathcal{B}_2)$. On the other hand, for $f \in \mathcal{H}$ to be sparse in terms of analysis coefficients, f has to be in $\mathcal{K}^\tau_q(\mathcal{B}_1,\mathcal{H}) \cap \mathcal{K}^\tau_q(\mathcal{B}_2,\mathcal{H}) = \mathcal{A}^\alpha_q(\mathcal{B}_1) \cap \mathcal{A}^\alpha_q(\mathcal{B}_2)$ which shows there is, in general, no chance of getting a characterization of $\mathcal{A}^\alpha_q(\mathcal{D})$ in terms of sparsity of the frame coefficients.

It is thus preferable to define sparsity in terms of the best (*i.e.* sparsest) **synthesis coefficients** for f. Following DeVore and Temlyakov [DT96], we will measure sparsity using *abstract smoothness classes* $\mathcal{K}^\tau_q(\mathcal{D}, X)$ which are defined as follows. For $\tau \in (0,\infty)$ and $q \in (0,\infty]$ we let $\mathcal{K}^\tau_q(\mathcal{D}, X, M)$ denote the set

$$\mathrm{clos}_X \left\{ f \in X, f = \sum_{k \in I} c_k g_k, I \subset \mathbb{N}, |I| < \infty, \|\{c_k\}_{k\geq 1}\|_{\ell^\tau_q} \leq M \right\}.$$

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. 3. It will be shown in the appendix (Proposition A.9) that $|\cdot|_{\mathcal{K}_q^\tau(\mathcal{D},X)}$ is a (semi)-(quasi)norm on $\mathcal{K}_q^\tau(\mathcal{D},X)$.

The goal of this work is to generalize Theorem 1.1 to some redundant dictionaries. Based on examples in [GN01] we know that we need to require some structure of $\mathcal{D}$. In this paper, we focus our attention on the identification of the structure required to get continuous embeddings of the Jackson type

$$\mathcal{K}_q^\tau(\mathcal{D},X) \hookrightarrow \mathcal{T}_q^\alpha(\mathcal{D},X) \tag{14}$$

with $\tau = (\alpha+1/p)^{-1}$ for some $1 \leq p < \infty$. In a forthcoming paper [GN02a], we will discuss the Bernstein type

$$\mathcal{A}_q^\alpha(\mathcal{D},X) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D},X) \tag{15}$$

with $\tau = (\alpha + 1/r)^{-1}$ for some $p < r$. Examples from [GN01] show that it may not be possible to match p and r. We may even have to replace $\mathcal{T}$ with one of the larger classes $\mathcal{C}$ or $\mathcal{A}$ in the Jackson-type embedding, or $\mathcal{A}$ with one of the smaller classes in the Bernstein-type one.

2. GENERAL JACKSON-TYPE EMBEDDINGS

In this section we are interested in getting Jackson embeddings

$$\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}), \qquad \alpha = 1/\tau - 1/p \tag{16}$$

for some p. First we will see that a **universal** Jackson embedding holds (Theorem 2.1) 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

$$\mathcal{K}_\tau^\tau(\mathcal{D}) \hookrightarrow \mathcal{C}_\infty^\alpha(\mathcal{D}), \qquad \alpha = 1/\tau - 1/p \tag{17}$$

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 2.2).

2.1. Universal Jackson embedding

For *any* dictionary $\mathcal{D}$ in *any* Banach space X, the following standard Jackson estimates holds.

THEOREM 2.1. *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 $f \in \mathcal{K}_q^\tau(\mathcal{D})$*

$$\|f\|_{\mathcal{A}_q^\alpha(\mathcal{D})} \leq C|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \qquad \text{with } \alpha = 1/\tau - 1. \tag{18}$$

Proof. Let $f \in \mathcal{K}_q^\tau(\mathcal{D})$ and fix M an integer. Let $\epsilon > 0$, and $\boldsymbol{c}$ 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

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

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 (abstract) smoothness $\tau < 1$ can be approximated with a rate of approximation at least $\alpha = 1/\tau - 1$. This is depicted graphically on Figure 1. 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. For $1/\tau \leq 1$ 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 2.2. It is clear that the Jackson inequality given by Theorem 2.1 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. Geometric Jackson embedding

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

$$\sigma_m(f, \mathcal{D})_X \leq Cm^{-\alpha}|f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}, \qquad m \geq 1, \qquad \alpha = 1/\tau - 1/p \tag{19}$$

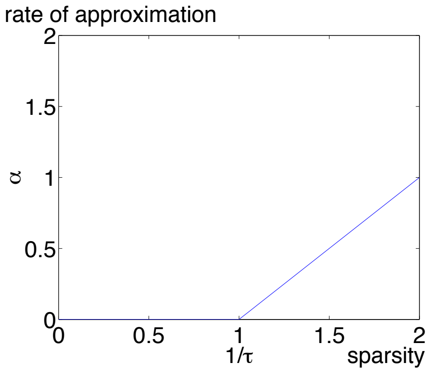

FIG. 1. The universal Jackson embedding line $\alpha = \max\{1/\tau - 1, 0\}$, with slope 1 that is valid for any Banach space X and any dictionary $\mathcal{D}$.

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 2.2. *In any infinite dimensional (separable) Hilbert space $\mathcal{H}$ there exists a dictionary $\mathcal{D}$ such that the Jackson inequality* (19) *fails for every $\tau > 1$.*

Temlyakov [Tem00] obtained a Jackson inequality (19) for $\mathcal{D}$ an arbitrary dictionary in a Banach space X, $1 < p \leq 2$ 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 2.3 (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

$$\|f - P_{\mathcal{V}_m(\pi(f),\mathcal{D})} f\|_X \;\leq\; C m^{-\alpha} |f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}, \qquad m \geq 1$$
$$\text{with } \alpha = 1/\tau - 1/p. \qquad (20)$$

Note that the Jackson inequality (20) 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**

$$\mathcal{K}_\tau^\tau(\mathcal{D}) \hookrightarrow \mathcal{C}_\infty^\alpha(\mathcal{D}), \qquad \alpha = 1/\tau - 1/p, \qquad 0 < \tau \leq 1 \qquad (21)$$

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

Remark 2. 1. *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 smoothness 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 2.1. 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 2. 2. 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 2 illustrates the improvement that can be obtained (compared to the universal Jackson embedding) from taking into account the geometry of the space X. 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 embedding line $\alpha = 1/\tau - 1/p$ for $0 < \tau \leq 1$, which improves the generic embedding line $\alpha = 1/\tau - 1$. 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 smoothness $0 < \tau \leq 1$. Note also that the embedding line is limited to the region where $1/\tau \geq 1$, which is a consequence of Theorem 2.2 as we will see in the next section.

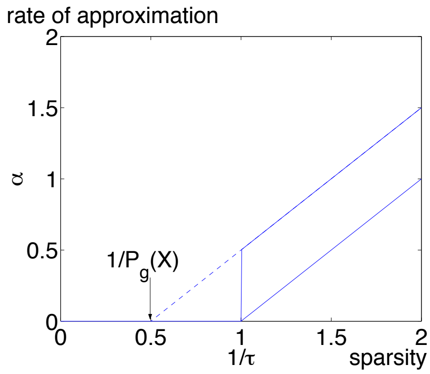

FIG. 2. The geometric Jackson embedding line $\alpha = 1/\tau - 1/p$ for $0 < \tau \leq 1$ (with $1 < p < P_g(X)$) compared to the generic embedding line $\alpha = 1/\tau - 1$.

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 2.2 which shows that $0 < \tau \leq 1$ is an essential limitation. For the proof of Theorem 2.2, we will need the following lemma.

LEMMA 2.1. *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 2.1. 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 $N \geq 1$

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

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

Proof of Theorem 2.2. 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 2.1, 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. $\square$

Remark 2. 3. Notice that the above arguments also show that the Jackson inequality (19) cannot be "repaired" for $\tau > 1$ by replacing $|\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$ with $\|\cdot\|_X + |\cdot|_{\mathcal{K}_\tau^\tau(\mathcal{D})}$.

Theorem 2.2 shows that a "complete" embedding line $\mathcal{K}_\tau^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_\infty^\alpha(\mathcal{D})$ cannot be expected in general unless we assume some structure on the dictionary $\mathcal{D}$. Structured dictionaries are discussed in the next section.

3. HILBERTIAN DICTIONARIES

In this section, we will see that no complete Jackson embedding can be obtained with $p > 1$ without assuming that $\mathcal{D}$ has a **hilbertian structure**.

DEFINITION 3.1. A dictionary $\mathcal{D}$ is called ℓ_q^τ-hilbertian if for any sequence $\boldsymbol{c} = \{c_k\}_{k\geq 1} \in \ell_q^\tau$, the series $\sum_{k\geq 1} c_k g_k$ is convergent in X and

$$\Big\| \sum_{k\geq 1} c_k g_k \Big\|_X \lesssim \|\boldsymbol{c}\|_{\ell_q^\tau}. \tag{22}$$

Remark 3. 1. Note that the convergence of $\sum_k c_k g_k$ in Definition 3.1 is necessarily unconditional provided that ℓ_q^τ is not one of the extremal nonseparable space such as ℓ^∞. Also notice that *any* dictionary is ℓ^τ-hilbertian for $0 < \tau \leq 1$.

First, we study hilbertian dictionary in more details and give a simple representation of the abstract smoothness class $\mathcal{K}_q^\tau(\mathcal{D})$ for such dictionaries.

Then, we will use this representation to prove a strong Jackson embedding of the type $\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{T}_q^\alpha(\mathcal{D})$ and get the following theorem

THEOREM 3.1. *Let $\mathcal{D}$ a dictionary in a Banach space X, and $p > 1$. The following properties are equivalent*

$$\forall\ \tau < p,\ \forall\ q,\ \forall\ \alpha < 1/\tau - 1/p \qquad \mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{T}_q^\alpha(\mathcal{D}), \tag{23}$$

$$\forall\ \tau < p,\ \forall\ q,\ \forall\ \alpha < 1/\tau - 1/p \qquad \mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{C}_q^\alpha(\mathcal{D}), \tag{24}$$

$$\forall\ \tau < p,\ \forall\ q,\ \forall\ \alpha < 1/\tau - 1/p \qquad \mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}), \tag{25}$$

$$\forall\ \tau < p \qquad \mathcal{D} \textit{ is } \ell_1^\tau - \textit{hilbertian}. \tag{26}$$

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. Properties of hilbertian dictionaries

For any dictionary $\mathcal{D} = \{g_k\}$ it makes sense to define the operator

$$T\ : \{c_k\} \mapsto \sum_k c_k g_k \tag{27}$$

on the space ℓ^0 of finite sequences $\boldsymbol{c} = \{c_k\}$. Some of the structure of $\mathcal{D}$ can be studied through the properties of T. In particular, the condition for $\mathcal{D}$ to be ℓ_q^τ-hilbertian can easily be verified to be equivalent to the requirement that T can be extended to a continuous linear operator from ℓ_q^τ to X. Notice that the ℓ_q^τ-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 3.2. For any dictionary $\mathcal{D}$ we define

$$P_s(\mathcal{D}, X) := \sup\{p : \mathcal{D} \text{ is } \ell_1^p\text{-hilbertian}\} \in [1, \infty].$$

Remark 3. 2. It is easy to deduce directly from the definition of cotype of a Banach space (see [LT79, Vol. II, Sec. 1.e.]), that if X has cotype $p_c \geq 2$, then $1 \leq P_s(\mathcal{D}, X) \leq p_c$.

EXAMPLE 3.1. In a Hilbert space $\mathcal{H}$, ℓ^2-hilbertian dictionaries are simply called *hilbertian* dictionaries. For any hilbertian dictionary, the operator $T : \ell^2 \to \mathcal{H}$ has a dense range provided that $\mathcal{D}$ is *complete.* If $\mathcal{D}$ is complete and hilbertian, the range of T is closed if, and only if, T is onto, that is to say if the dictionary is a frame [Chr02]. In such a case there exists a dual frame $\widetilde{\mathcal{D}} = \{\widetilde{g}_k\}$ such that for all $f \in \mathcal{H}$, $\|f\|_{\mathcal{H}} \asymp \|\{\langle f, g_k\rangle\}_{k\geq 1}\|_{\ell^2} \asymp \|\{\langle f, \widetilde{g}_k\rangle\}_{k\geq 1}\|_{\ell^2}$ and

$$f = \sum_k \langle f, g_k\rangle \widetilde{g}_k = \sum_k \langle f, \widetilde{g}_k\rangle g_k. \tag{28}$$

Sometimes these relations extend to other spaces as well: for "nice" shift invariant systems [AST01] there is a dual system $\widetilde{\mathcal{D}}$ such that for and any $f \in T\ell^p$, $1 < p < \infty$ the expansion (28) converges unconditionally in $L^p(\mathbb{R})$ and $\|f\|_{L^p(\mathbb{R})} \asymp \|\{\langle f, g_k\rangle\}_{k\geq 1}\|_{\ell^p} \asymp \|\{\langle f, \widetilde{g}_k\rangle\}_{k\geq 1}\|_{\ell^p}$; for "nice" Gabor systems, a similar result holds where $L^p(\mathbb{R})$ is replaced with the modulation space $M_p(\mathbb{R})$ [Grö00]. More generally, when $\mathcal{D}$ can be seen as a subset of the dual $X^\star$ and the relation $\|f\|_X \asymp \|\{\langle f, g_k\rangle\}_{k\geq 1}\|_{\ell^p}$ holds true, $\mathcal{D}$ is called a p-frame for X [Chr02].

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

PROPOSITION 3.1. *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 $|I_m| \leq m$ and every choice of signs

$$\|\sum_{k\in I_m} \pm g_k\|_X \leq C m^{1/p}. \tag{29}$$

Proof.

(i)⇒(ii) is obvious.

(ii)⇒(i) Let $\boldsymbol{c} \in \ell^p_1$ and π a permutation of $\mathbb{N}$ such that $c^\star_k = c_{\pi_k}$, and define a sequence $f_n = f_n(\boldsymbol{c}) := \sum_{k=1}^n c^\star_k g_{\pi_k} = \sum_{k=1}^n c_{\pi_k} g_{\pi_k}$. 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 (29), we can write for every $n \geq m$

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

By taking $m = 2^j$ and $n = 2^{j+1}$ with $j \geq 0$ we get

$$\|f_{2^{j+1}} - f_{2^j}\|_X \leq C 2^{j/p}|c^\star_{2^j}|$$

hence $\sum_{j=0}^{\infty} \|f_{2^{j+1}} - f_{2^j}\|_X \leq C \sum_{j=0}^{\infty} |c^\star_{2^j}| 2^{j/p} \leq \tilde{C}\|\boldsymbol{c}\|_{\ell^p_1}$. 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 $\|T\boldsymbol{c}\|_X \leq \|\boldsymbol{c}\|_{\ell^\infty} + \tilde{C}\|\boldsymbol{c}\|_{\ell^p_1} \leq (1+\tilde{C})\|\boldsymbol{c}\|_{\ell^p_1}$. 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^p_1 consisting of finite sequences, is linear and continuous. It follows by standard arguments that T extends to a bounded linear operator from ℓ^p_1 to X. □

3.2. Representation of the abstract smoothness class

The hilbertian structure of $\mathcal{D}$ makes it possible to get a nice representation of some of the smoothness classes $\mathcal{K}^\tau_q(\mathcal{D})$. This representation will be very useful in the proof of some Jackson-type estimates. The operator T below is the one defined at the beginning of Section 3.1.

PROPOSITION 3.2. *Assume $\mathcal{D}$ is ℓ^p_s-hilbertian, where ℓ^p_s is reflexive, and let $\ell^\tau_q \hookrightarrow \ell^p_s$. Then*

$$\mathcal{K}^\tau_q(\mathcal{D}) = T\ell^\tau_q \tag{30}$$

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

In order to prove Proposition 3.2, we need a lemma.

LEMMA 3.1. *Assume $\mathcal{D}$ is ℓ^p_s-hilbertian, where ℓ^p_s is reflexive, and let $\ell^\tau_q \hookrightarrow \ell^p_s$. For all $f \in \mathcal{K}^\tau_q(\mathcal{D})$, there exists some $\boldsymbol{c} \in \ell^\tau_q$ which realizes the smoothness norm*

$$|f|_{\mathcal{K}^\tau_q(\mathcal{D})} = \|\boldsymbol{c}\|_{\ell^\tau_q}, \text{ with } f = T\boldsymbol{c}. \tag{31}$$

Consequently

$$|f|_{\mathcal{K}^\tau_q(\mathcal{D})} = \min_{\boldsymbol{c}\in\ell^\tau_q, f=T\boldsymbol{c}} \|\boldsymbol{c}\|_{\ell^\tau_q}. \tag{32}$$

If in addition ℓ^τ_q is strictly convex, then $\boldsymbol{c} = \boldsymbol{c}_{\tau,q}(f)$ is unique.

Proof. By definition of $\mathcal{K}^\tau_q(\mathcal{D})$, for $f \in \mathcal{K}^\tau_q(\mathcal{D})$ there exists finite sequences $\boldsymbol{c}_n \in \ell^\tau_q$, $n = 1, 2, \ldots$ such that $\|\boldsymbol{c}_n\|_{\ell^\tau_q} \leq |f|_{\mathcal{K}^\tau_q(\mathcal{D})} + 1/n$ and $\|f - T\boldsymbol{c}_n\|_X \leq 1/n$. The sequence $\{\boldsymbol{c}_n\}_{n\geq 1}$ is bounded in ℓ^τ_q, hence it is also bounded in ℓ^p_s. As ℓ^p_s is a reflexive Banach space, it is weakly compact and there exists a subsequence $\boldsymbol{c}_{n_k}$ converging weakly in ℓ^p_s to some $\boldsymbol{c} \in \ell^p_s$. Applying Fatou's lemma twice gives $\|\boldsymbol{c}\|_{\ell^\tau_q} \leq |f|_{\mathcal{K}^\tau_q(\mathcal{D})}$.

From the weak convergence in ℓ^p_s and the continuity of $T : \ell^p_s \to X$ 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 $|f|_{\mathcal{K}^\tau_q(\mathcal{D})} \leq \|\boldsymbol{c}\|_{\ell^\tau_q}$.

By the strict convexity of ℓ^τ_q, if $\boldsymbol{c}_0 \neq \boldsymbol{c}_1$ both satisfy Eq. (31), we get $\|(\boldsymbol{c}_1 + \boldsymbol{c}_0)/2\|_{\ell^\tau_q} < |f|_{\mathcal{K}^\tau_q(\mathcal{D})}$. As $T((\boldsymbol{c}_0 + \boldsymbol{c}_1)/2) = f$ this contradicts (32). □

Proof of Proposition 3.2. The equality $\mathcal{K}^\tau_q(\mathcal{D}) = T\ell^\tau_q$ follows directly from Lemma 3.1. Then we observe that for any $f \in \mathcal{K}^\tau_q(\mathcal{D})$,

$$|f|_{\mathcal{K}^\tau_q(\mathcal{D})} = \|\boldsymbol{c}\|_{\ell^\tau_q} \geq \|\boldsymbol{c}\|_{\ell^p_s} \geq C^{-1}\|f\|_X$$

where the last inequality comes from the continuity of $T : \ell^p_s \to X$. The conclusion is reached using Proposition A.9 proved in the appendix. □

Let us state one simple and straightforward Corollary of Proposition 3.2.

Corollary 3.1. *Assume $\mathcal{D}_1$ and $\mathcal{D}_2$ are both ℓ^p_s-hilbertian dictionaries, where ℓ^p_s is reflexive, and let $\ell^\tau_q \hookrightarrow \ell^p_s$. Then for $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$*

$$\mathcal{K}^\tau_q(\mathcal{D}) = \mathcal{K}^\tau_q(\mathcal{D}_1) + \mathcal{K}^\tau_q(\mathcal{D}_2).$$

3.3. Strong Jackson embedding

For the standard Jackson embedding (16) to hold for any $0 < \tau < p$, $0 < \alpha = 1/\tau - 1/p < \infty$ with $p > 1$, the dictionary needs some hilbertian structure : this is essentially Theorem 3.1, but we can actually prove some slightly stronger results. Some of the statements in Theorem 3.1 are almost trivial : from (10) and (11) it is obvious that (23) ⇒ (24) ⇒ (25). Moreover, using the same technique as [GN01, Proposition 4.1], we easily obtain (25) ⇒ (26) as follows:

Proof. From the double embedding $\mathcal{K}^\tau_1(\mathcal{D}) \hookrightarrow \mathcal{A}^\alpha_\infty(\mathcal{D}) \hookrightarrow X$, $\tau < (\alpha + 1/p)^{-1}$, we have $\|\cdot\|_X \lesssim |\cdot|_{\mathcal{K}^\tau_1(\mathcal{D})}$. Thus we can check for $I_m \subset \mathbb{N}$ of cardinality $|I_m| = m$, $\|\sum_{k \in I_m} \pm g_k\|_X \leq C m^{1/\tau}$ which by Proposition 3.1 gives that $\mathcal{D}$ is ℓ^τ_1-hilbertian. But as α can be arbitrarily close to 0, τ can be arbitrarily close to p and this gives the result. □

Next we will show that not only the ℓ^p_1-hilbertian property of $\mathcal{D}$ is (almost) necessary for any Jackson embedding to hold for all $\alpha > 0$, it is also sufficient to get the *strong* Jackson embedding (23) for all $\alpha > 0$. We will make use of the representation of $\mathcal{K}_q^\tau(\mathcal{D})$ given by Proposition 3.2. Notice that Theorem 3.2 also shows that (26) $\Rightarrow$ (23) which finally completes the proof of Theorem 3.1.

THEOREM 3.2. *For any $\tau < p$, $0 < q \le \infty$, and $1 < p < \infty$ there is a constant $C = C(\tau, q, p)$ such that for any ℓ_1^p-hilbertian dictionary $\mathcal{D}$ and any $f \in \mathcal{K}_q^\tau(\mathcal{D})$*

$$\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D})} \le \|T\|_{\ell_1^p}^X C|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \qquad \textit{with } \tau = (\alpha + 1/p)^{-1} \tag{33}$$

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 $f \in \mathcal{K}_q^\tau(\mathcal{D})$. As $\tau < p$ and $p > 1$ we have, for some small $\epsilon > 0$, $\ell_q^\tau \hookrightarrow \ell^{p-\epsilon} \hookrightarrow \ell_1^p$, where $1 < p - \epsilon$. Hence by Lemma 3.1 we can take $\boldsymbol{c} \in \ell_q^\tau$ a representation of f. 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 $\boldsymbol{c} = \{c_k\}_{k\ge 1}$ 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 (4)). Obviously for $m \ge 1$

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

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

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

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

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

Remark 3. 3. When $\mathcal{D}$ is only ℓ^1-hilbertian, we loose the representation of $\mathcal{K}_q^\tau(\mathcal{D})$ (Lemma 3.1) 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 $\{\pm g, g \in \mathcal{D}\}$, see Example 3.2. It does not seem possible to extend Theorem 3.2 to $p = 1$.

Notice that the Jackson embedding provided by Theorem 3.2 is **strong** : not only does it show that the best m-term error decays in $\mathcal{O}(m^{-\alpha})$ (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

$$\|f - \sum_{k=1}^{m} c_k^\star g_{\pi_k(f)}\|_X \lesssim m^{-\alpha} |f|_{\mathcal{K}_\tau^\tau(\mathcal{D})}, \qquad m \geq 1, \qquad \tau = (\alpha + 1/p)^{-1}. \tag{34}$$

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 further in Section 5.

3.4. Adaptive analysis operator and atomic decompositions

One should notice that the representation $\boldsymbol{c}$ that yields the rate of thresholding approximation $m^{-\alpha}$ in the thresholding-Jackson inequality (34) may depend on f in a *nonlinear* way. A natural question arising from the representation of $\mathcal{K}_q^\tau(\mathcal{D})$ in Lemma 3.1 is whether there exists a (linear) **adaptive analysis operator** $U : f \mapsto U(f) = \{g_k^\star(f)\}_{k\geq 1},\ g_k^\star \in X^\star$ that gives (near) sparsest representation of every f. In other words, does there exist U and, for all $0 < \tau < p, q \in (0, \infty]$, some $C = C(\tau, q)$ such that for all $f \in \mathcal{K}_q^\tau(\mathcal{D})$,

$$\|U(f)\|_{\ell_q^\tau} \leq C |f|_{\mathcal{K}_q^\tau(\mathcal{D})}$$

and $TU(f) = f$. That is to say, is there a (linear) right inverse U to T that simultaneously maps $\mathcal{K}_q^\tau(\mathcal{D})$ to ℓ_q^τ in the range $0 < \tau < p$.

The existence of such an operator is trivial for $\mathcal{D}$ a basis, and it is also known to exist for twice oversampled framelet dictionaries in $L^p(\mathbb{R})$ [GN02b], p-frames in shift invariant subspaces of $L^p(\mathbb{R}^d)$ [AST01], and Gabor dictionaries in the weighted modulation spaces $M_p^w(\mathbb{R})$ [GS00]. However, whether such a U exists certainly depends in general on the structure of the dictionary $\mathcal{D}$. This question has an important practical impact on the construction of numerical approximation algorithms that "adapt" automatically to the sparsity of f (see a discussion on this topic in [GN02b]).

Remark 3. 4. The operator U discussed above is closely related to the concept of atomic decompositions for a Banach space Y, and more generally of Banach frames thereof. A Banach frame is basically defined as an isomorphic analysis operator induced by a sequence of "coefficient functionals" (elements of the dual space X^*) for which there exists a bounded left inverse or **reconstruction operator**, we refer to [Chr02] for more details. In the case considered above, for $1 \leq \tau < p$, $\mathcal{K}_q^\tau(\mathcal{D})$ is a Banach space by Proposition 3.2, and if $U : f \mapsto U(f) = \{g_k^*(f)\}$ maps $\mathcal{K}_q^\tau(\mathcal{D})$ continuously into ℓ_q^τ, then for all $f = TU(f) \in \mathcal{K}_q^\tau(\mathcal{D})$

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq \|U(f)\|_{\ell_q^\tau} \leq C|f|_{\mathcal{K}_q^\tau(\mathcal{D})}$$

and U is thus isomorphic on $\mathcal{K}_q^\tau(\mathcal{D})$. This means that $\left(\mathcal{K}_q^\tau(\mathcal{D}), \{g_k^*\}\right)$ is a Banach frame with coefficient space ℓ_q^τ. Indeed $\left(\{g_k\}, \{g_k^*\}\right)$ is an atomic decomposition of $\mathcal{K}_q^\tau(\mathcal{D})$.

In the appendix, we discuss the theoretical impact of the existence of such a U to the interpolation of (abstract) smoothness spaces. Next we summarize and discuss the Jackson-type results obtained so far.

3.5. Geometry *versus* structure of the dictionary

Let us now compare the embedding lines obtained so far. First we consider the case where the geometric line "beats" the structural line for $\tau \leq 1$. Later we will consider the opposite situation. We have the following example.

Example 3.2. Let X be a Banach space with $P_g(X) > 1$ [*e.g.* a Hilbert space] 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 2.1 and Proposition 3.2 we see that $P_s(\mathcal{D}, X) = 1$. Thus, in this case the geometric embedding line is strictly better than the structural 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.

The leftmost graph in Figure 3 illustrates a situation similar to that of Example 3.2, with the only difference that the graph depicts the case where $1 < P_s(\mathcal{D}, X) < P_g(X)$. The structural embedding line is valid for a larger range of values of $1/\tau$ than the geometric embedding, but the geometric result is stronger on its range of validity.

Remark 3. 5. The geometric Jackson embedding is given by $\mathcal{K}_\tau^\tau(\mathcal{D}) \hookrightarrow \mathcal{C}_\infty^\alpha(\mathcal{D}) \hookrightarrow \mathcal{A}_\infty^\alpha(\mathcal{D})$, $\alpha = 1/\tau - 1/p$, and thus restricted to secondary index $q = \infty$. However for $0 < \tau_1 \leq \tau_2 \leq 1$ we can easily interpolate to get

$$\left(\mathcal{K}_{\tau_2}^{\tau_2}(\mathcal{D}), \mathcal{K}_{\tau_1}^{\tau_1}(\mathcal{D})\right)_{\theta,q} \hookrightarrow \left(\mathcal{A}_\infty^{\alpha_2}(\mathcal{D}), \mathcal{A}_\infty^{\alpha_1}(\mathcal{D})\right)_{\theta,q} = \mathcal{A}_q^\alpha(\mathcal{D}),$$

with $\alpha = (1-\theta)\alpha_2 + \theta\alpha_1$ (we used the fact that $\mathcal{A}_q^\alpha(\mathcal{D})$ is an interpolation family, see [DL93, Chap. 7]). Notice that Lemma A.1 (in the appendix) shows that $P_s(\mathcal{D}, X) > 1$ ensures that $\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \left(\mathcal{K}_{\tau_2}^{\tau_2}(\mathcal{D}), \mathcal{K}_{\tau_1}^{\tau_1}(\mathcal{D})\right)_{\theta,q}$ with $1/\tau = (1-\theta)/\tau_2 + \theta/\tau_1$, which gives us the embedding line with a general secondary index.

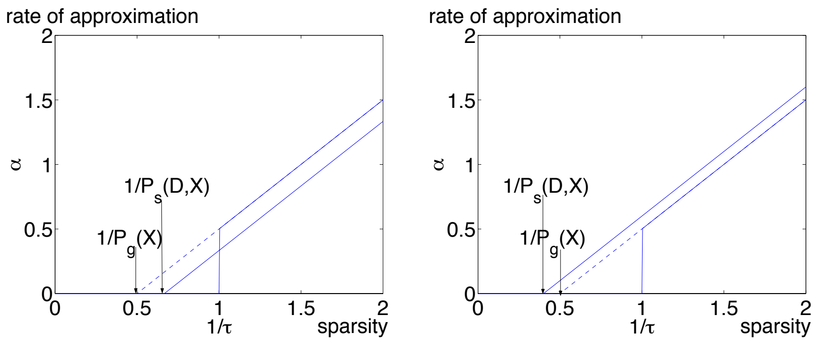

FIG. 3. Comparison of the structural Jackson embedding line and the geometric one. The leftmost graph depicts a situation where the structural embedding is valid for a larger range of values of $1/\tau$ than the geometric embedding, but the geometric result is stronger on its range of validity. The rightmost corresponds to the opposite situation, where the structural embedding line is better than the geometric line throughout its domain of validity.

In the same spirit, we interpolate the "extremal" geometric and structural embeddings and get with $r < 1 - 1/P_g(X)$, $1 < p < P_s(\mathcal{D}, X)$

$$\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \big(\mathcal{K}_1^p(\mathcal{D}), \mathcal{K}_1^1(\mathcal{D})\big)_{\alpha/r,q} \hookrightarrow \big(X, \mathcal{A}_\infty^r(\mathcal{D})\big)_{\alpha/r,q} \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}),$$

with $\alpha = (1/\tau - 1/p)\frac{r}{1-1/p} < r$. In the limit case where the embeddings hold with $r = 1 - 1/P_g(X)$ and $p = P_s(\mathcal{D}, X)$, we get the embedding segment

$$\alpha = \left(\frac{1}{\tau} - \frac{1}{P_s(\mathcal{D}, X)}\right) \frac{1 - 1/P_g(X)}{1 - 1/P_s(\mathcal{D}, X)}.$$

The opposite situation, where the structural embedding line is better than the geometric line throughout its domain, is also possible. This particular situation is illustrated on the rightmost graph in Figure 3. An explicit example is given by the following:

EXAMPLE 3.3. 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 $P_g(L^p(\mathbb{R}^d)) = \min\{2, p\}$ [LT79]. Moreover, it can be verified that $\mathcal{B}$ is ℓ^p-hilbertian ($1 < p \leq 2$) or ℓ_1^p-hilbertian ($2 < p < \infty$), hence $P_s(\mathcal{B}, X) = p$, $1 < p < \infty$.

- For $1 < p \leq 2$ 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 structural 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 structural embedding line is *strictly* better than the geometric line for $0 < \tau < p$, which corresponds exactly to the rightmost graph on Figure 3.

4. EXAMPLES OF HILBERTIAN DICTIONARIES

In this section we will consider the ℓ^p_1-hilbertian property of several classical types of dictionaries, first in L^p spaces, and then in other classical functional spaces (Besov spaces $B^\alpha_\tau(L^\tau(\mathbb{R}))$, modulation spaces $M^w_p(\mathbb{R})$). This will give further examples where the structural Jackson embedding result (Theorem 3.2) applies.

4.1. Dictionaries in L^p

First we consider a general lemma that will make it easier to check the ℓ^p-hilbertian property for many well-known dictionaries. We denote by Ω some σ-finite measure space, and we have the following lemma.

LEMMA 4.1. *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 $1 \leq p \leq r$ there is some constant $C = C(p)$ such that for every g_k, with $\theta_{r,p} = (1-1/p)/(1-1/r)$,*

$$\|g_k\|^{1-\theta_{r,p}}_{L^1(\Omega)} \leq C\|g_k\|_{L^p(\Omega)}. \tag{35}$$

Then $\mathcal{D}_p := \{g_k/\|g_k\|_{L^p(\Omega)}, k \in \mathbb{N}\}$ is an ℓ^p-hilbertian dictionary in $L^p(\Omega)$, $1 \leq p < r$.

Moreover, if $|\Omega| < \infty$ and $\mathcal{D}$ has dense span in $L^r(\Omega)$, then $\mathcal{D}_p$ has dense span in $L^p(\Omega)$, $1 \leq p \leq r$.

Proof. By assumption, T is continuous from ℓ^r to $L^r(\Omega)$. Then, as $\mathcal{D}_1$ is obviously 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 $C' = C'(p) < \infty$

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

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 $L^p(\Omega)$. ☐

It is easy to check the assumptions of Lemma 4.1 on classical dictionaries.

1 Wavelet bases and wavelet frames 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 $L^2(\mathbb{R}^d)$

$$\psi_{j,k}^{\ell}(x) := 2^{jd/2} \psi^{\ell}(2^j x - k), \qquad 1 \leq \ell \leq L,\ j \in \mathbb{Z}, k \in \mathbb{Z}^d. \tag{36}$$

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 $1 \leq p \leq \infty$,

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

so (35) 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 4.1 applies with $r = 2$, and we get that such systems are ℓ^p-hilbertian in $L^p(\mathbb{R}^d)$ for any $1 \leq p \leq 2$.

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 $\mathcal{D}_p$ a (bi)-orthogonal wavelet system normalized in $L^p(\mathbb{R}^d)$, Proposition 3.2 together with Theorem 3.2 shows that

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

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

Remark 4. 1. For wavelet like systems, one can sometimes extend the result obtained from Lemma 4.1 and show that such systems are ℓ^p_1-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].

2 **Uniformly bounded orthonormal systems in** $L^p(\Omega),\ |\Omega| < \infty$

Let $\mathcal{D} = \{g_k, k \in \mathbb{N}\}$ any uniformly bounded orthonormal system in $L^2(\Omega),\ |\Omega| < \infty$. A consequence of the Hölder inequality is that for $0 < p \leq \infty$, there exist $0 < A_p < \infty$ such that

$$1/A_p \leq \|g_k\|_{L^p(\Omega)} \leq A_p, \qquad k \in \mathbb{N}.$$

It follows that such systems satisfy the hypothesis (35) of Lemma 4.1 with $r = 2$, and we get that they are ℓ^p-hilbertian and complete in $L^p(\Omega)$ for any $1 \leq p \leq 2$. The prime examples of such systems are the **trigonometric system** and the **Walsh system** (see *e.g.* [GES91]) on $L^2[0, 1)$.

Using Proposition 3.2 and Theorem 3.2 together with the fact that such a system is simultaneously (quasi)normalized in all L^p, $1 \leq p \leq 2$, , one easily gets the following result.

PROPOSITION 4.3. *Let $\mathcal{D} = \{g_k, k \in \mathbb{N}\}$ a uniformly bounded orthonormal system in $L^2(\Omega),\ |\Omega| < \infty$. Assume that $\ell^\tau_q \hookrightarrow \ell^2$. We have, with equivalent norms, for all $1 \leq p \leq 2$ such that $\ell^\tau_q \hookrightarrow \ell^p$,*

$$\mathcal{K}^\tau_q\big(\mathcal{D}_p, L^p(\Omega)\big) = \mathcal{K}^\tau_q\big(\mathcal{D}, L^2(\Omega)\big) = \Big\{f \in L^p(\Omega) \cap L^2(\Omega), \{\langle f, g_k\rangle\}_{k\geq 1} \in \ell^\tau_q\Big\}, \tag{37}$$

and it follows that

$$\mathcal{K}^\tau_q(\mathcal{D}, L^2(\Omega)) \hookrightarrow \mathcal{T}^\alpha_q\big(\mathcal{D}, L^p(\Omega)\big), \qquad \alpha = 1/\tau - 1/p. \tag{38}$$

4.2. Dictionaries in L^p and modulation spaces

Not all interesting dictionaries live in L^p: in the following we concentrate on time-frequency dictionaries, for which the natural function spaces are the modulation spaces. We consider the ℓ^p-hilbertian property of such dictionaries, first in L^p spaces, then in modulation spaces. 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 a-moderate weight $w(x,y)$. For the trivial weight $w \equiv 1$, we denote $M_p(\mathbb{R})$ instead of $M_p^w(\mathbb{R})$.

1 **Local Fourier bases in $L^p(\mathbb{R})$**

Consider $\mathcal{D}$ an orthonormal **local Fourier basis** [CM91] in $L^2(\mathbb{R}) = M_2(\mathbb{R})$. For example, $\mathcal{D}$ may be the orthonormal Wilson bases of Daubechies-Jaffard-Journé [DJJ91] or, *e.g.*, a system of the form

$$g_{n,m}(x) := \sqrt{\frac{2}{\alpha_{n+1}-\alpha_n}}b_n(x)\sin\left(\pi m\frac{x-\alpha_n}{\alpha_{n+1}-\alpha_n}\right),\ n \in \mathbb{Z},\ m \in \mathbb{N}, \quad (39)$$

with appropriate "breakpoints" $\alpha_n < \alpha_{n+1}$ and "window" functions $b_n(x)$.

Provided that the window sizes $\alpha_{n+1}-\alpha_n$ are uniformly bounded from above and below, the same argument as for the uniformly bounded orthonormal systems on $L^2(\Omega)$, $|\Omega| < \infty$ applies: for $0 < p \leq \infty$ there exist $0 < A_p < \infty$ such that

$$1/A_p \leq \|g_{n,m}\|_{L^p(\mathbb{R})} \leq A_p, \qquad n \in \mathbb{Z},\ m \in \mathbb{N},$$

so that (35) is valid. Lemma 4.1 then ensures that $\mathcal{D}_p$ normalized in $L^p(\mathbb{R})$ is ℓ^p-hilbertian for $1 \leq p \leq 2$. Completeness follows from the fact that the expansion of f in a local Fourier basis can be rewritten as the expansion of a locally folded version of f in a standard trigonometric basis, see *e.g.* [AWW92]. Combining with Proposition 3.2 and Theorem 3.2, we get the following result.

Proposition 4.4. *Let $\mathcal{D} = \{g_{n,m}, n \in \mathbb{Z}, m \in \mathbb{N}\}$ be a local Fourier basis in $L^2(\mathbb{R})$, with window size uniformly bounded from above and below. Assume that $\ell^\tau_q \hookrightarrow \ell^2$. We have, with equivalent norms, for all $1 \leq p \leq 2$ such that $\ell^\tau_q \hookrightarrow \ell^p$,*

$$\mathcal{K}^\tau_q\left(\mathcal{D}_p, L^p(\mathbb{R})\right) = \mathcal{K}^\tau_q\left(\mathcal{D}, L^2(\mathbb{R})\right) = \left\{f \in L^p(\mathbb{R}) \cap L^2(\mathbb{R}), \{\langle f, g_{n,m}\rangle\}_{n,m} \in \ell^\tau_q\right\} \quad (40)$$

and it follows that

$$\mathcal{K}^\tau_q\left(\mathcal{D}, L^2(\mathbb{R})\right) \hookrightarrow \mathcal{T}^\alpha_q\left(\mathcal{D}, L^p(\mathbb{R})\right), \qquad \alpha = 1/\tau - 1/p. \quad (41)$$

2 Local Fourier bases in $M_p^w(\mathbb{R})$

The classical $L^p(\mathbb{R})$ spaces are not the natural spaces to study nonlinear approximation with local Fourier bases: one should instead consider them in modulation spaces. It is known that under some regularity conditions on the "window" functions b_n, for $1 \leq p \leq \infty$ [GS00, Theorem 2] and a-moderate weights w, we have the equivalence

$$\|f\|_{M_p^w(\mathbb{R})} \asymp \|\{\langle f, g_{n,m}\rangle w_{n,m}\}_{n,m}\|_{\ell^p(\mathbb{Z}\times\mathbb{N})} \tag{42}$$

with $w_{n,m} = w\big(\alpha_n, \frac{m}{2(\alpha_{n+1}-\alpha_n)}\big)$ and $\{g_{n,m}\}$ normalized in $L^2(\mathbb{R})$. The result also extends to $0 < p < 1$, [Sam98], but we will only consider the Banach space case here. It follows that for $1 \leq p \leq \infty$ there exist $0 < A_p < \infty$ such that

$$w_{n,m}/A_p \leq \|g_{n,m}\|_{M_p^w(\mathbb{R})} \leq A_p w_{n,m}, \qquad n \in \mathbb{Z}, m \in \mathbb{N},$$

and for $1 \leq p < \infty$, $\mathcal{D}$ is an unconditional basis for $M_p^w(\mathbb{R})$.

Indeeed, $\mathcal{D}^w := \big\{g_{n,m}^w := g_{n,m}/\|g_{n,m}\|_{M_2^w(\mathbb{R})}\big\}$ is simultaneously a (quasi)-normalized *greedy basis* [KT99] in all $M_p^w(\mathbb{R})$ spaces : it is both unconditional and *democratic*, *i.e.* for some $C < \infty$, any choice of signs and any two sets of indices $I, I' \subset \mathbb{Z}\times\mathbb{N}$ of same cardinality,

$$\|\sum_{n,m\in I} \pm g_{n,m}^w\|_{M_p^w(\mathbb{R})} \leq C\|\sum_{n,m\in I'} \pm g_{n,m}^w\|_{M_p^w(\mathbb{R})}.$$

From general results on nonlinear approximation with greedy bases [GN01, DKKT01], we recover [GS00, Theorem 3]: with equivalent norms, for all $0 < \tau < p$ with $1 \leq p \leq r < \infty$, we have

$$\mathcal{A}_q^\alpha\big(\mathcal{D}, M_p^w(\mathbb{R})\big) = \mathcal{T}_q^\alpha\big(\mathcal{D}, M_p^w(\mathbb{R})\big) = \mathcal{K}_q^\tau\big(\mathcal{D}_r^w, M_r^w(\mathbb{R})\big), \qquad \alpha = 1/\tau - 1/p, \tag{43}$$

as well as the characterization

$$\mathcal{K}_q^\tau\big(\mathcal{D}_r^w, M_r^w(\mathbb{R})\big), = \Big\{f \in M_p^w(\mathbb{R}) \cap M_r^w(\mathbb{R}), \{\langle f, g_{n,m}\rangle w_{n,m}\}_{n,m} \in \ell_q^\tau\Big\}$$

For the special case $\tau = q$, combining with (42), we get

$$\mathcal{A}_\tau^\alpha\big(\mathcal{D}, M_p^w(\mathbb{R})\big) = \mathcal{T}_\tau^\alpha\big(\mathcal{D}, M_p^w(\mathbb{R})\big) = M_\tau^w(\mathbb{R}), \qquad \alpha = 1/\tau - 1/p. \tag{44}$$

For the trivial weight $w \equiv 1$, denoting $M_p(\mathbb{R})$ instead of $M_p^w(\mathbb{R})$, this should be compared to the one-sided relation, valid for $0 < \tau < p$ with $1 \leq p \leq 2$,

$$M_\tau(\mathbb{R}) = \mathcal{K}_\tau^\tau(\mathcal{D}, L^2(\mathbb{R})) \hookrightarrow \mathcal{T}_\tau^\alpha(\mathcal{D}, L^p(\mathbb{R})), \qquad \alpha = 1/\tau - 1/p, \tag{45}$$

which is obtained by noticing that $M_2(\mathbb{R}) = L^2(\mathbb{R})$ and combining (42) with Proposition 4.4.

3 Gabor frames in $L^p(\mathbb{R})$, $1 \leq p \leq 2$

Consider a Gabor dictionary $\mathcal{D}$ consisting of the functions

$$g_{n,m}(x) := g(x - na)e^{2i\pi mbx}, \; n, m \in \mathbb{Z}. \tag{46}$$

Provided that g is an appropriate "window" function and (a, b) appropriate positive numbers (see, *e.g.*, [Dau92, Grö00, Chr02]), $\mathcal{D}$ is a frame in $L^2(\mathbb{R})$. It satisfies $\|g_{n,m}\|_{L^p(\mathbb{R})} = \|g\|_{L^p(\mathbb{R})}, 0 < p \leq \infty$. Lemma 4.1 then implies that the associated dictionaries $\mathcal{D}_p$ are ℓ^p-hilbertian in $L^p(\mathbb{R})$ for $1 \leq p \leq 2$. Again, we deduce easily the following result.

Proposition 4.5. *Let $\mathcal{D}$ a Gabor frame normalized in $L^2(\mathbb{R})$. Assume that $\ell^\tau_q \hookrightarrow \ell^2$. We have, with equivalent norms, for all $1 \leq p \leq 2$ such that $\ell^\tau_q \hookrightarrow \ell^p$,*

$$\mathcal{K}^\tau_q\left(\mathcal{D}_p, L^p(\mathbb{R})\right) = \mathcal{K}^\tau_q\left(\mathcal{D}, L^2(\mathbb{R})\right), \tag{47}$$

and it follows that

$$\mathcal{K}^\tau_q\left(\mathcal{D}, L^2(\mathbb{R})\right) \hookrightarrow \mathcal{T}^\alpha_q\left(\mathcal{D}, L^p(\mathbb{R})\right), \qquad \alpha = 1/\tau - 1/p. \tag{48}$$

4 Gabor frames in $M_p(\mathbb{R})$, $1 \leq p \leq 2$

As we already mentioned earlier, modulation spaces are better suited than L^p spaces when we consider nonlinear approximation properties of time-frequency systems. It is easily seen from the relation (42) that for a given a-moderate weight w, the family of modulation spaces $M^w_p(\mathbb{R})$, $0 < p \leq \infty$, is an interpolation family [BS88]. As a result, we can copy the proof of Lemma 4.1 to get an analogue result where $L^p(\Omega)$ is replaced with $M^w_p(\mathbb{R})$.

Lemma 4.2. *Let $1 < r < \infty$ and $\mathcal{D} = \{g_k, k \in \mathbb{N}\}$ an ℓ^r_1-hilbertian (normalized) dictionary in $X = M^w_r(\mathbb{R})$, and assume that every g_k is in $M^w_1(\mathbb{R})$. Suppose that for every $1 \leq p \leq r$ there is some constant $C = C(p)$ such that for every g_k, with $\theta_{r,p} = (1 - 1/p)/(1 - 1/r)$,*

$$\|g_k\|^{1-\theta_{r,p}}_{M^w_1(\mathbb{R})} \leq C\|g_k\|_{M^w_p(\mathbb{R})}. \tag{49}$$

Then $\mathcal{D}_p := \{g_k/\|g_k\|_{M_p^w(\mathbb{R})}, k \in \mathbb{N}\}$ *is an* ℓ^p*-hilbertian dictionary in* $M_p^w(\mathbb{R})$, $1 \le p < r$.

We apply this in the simplest case where $w \equiv 1$. Let $\mathcal{D} = \{g_{n,m},\ n,m \in \mathbb{Z}\}$ any Gabor frame generated by some window $g \in M_1(\mathbb{R}) \cap M_2(\mathbb{R})$. Using the relation $\|g_{n,m}\|_{M_p(\mathbb{R})} = \|g\|_{M_p(\mathbb{R})} < \infty, 1 \le p \le 2$, we can apply Lemma 4.2 to $\mathcal{D}$ because, as a frame, it is hilbertian. We get that $\mathcal{D}_p$ is ℓ^p-hilbertian in $M_p(\mathbb{R})$, $1 \le p \le 2$.

Proposition 4.6. *Let* $\mathcal{D}$ *a (normalized) Gabor frame. Assume that* $\ell^\tau_q \hookrightarrow \ell^2$*. For* $w \equiv 1$*, we have with equivalent norms, for all* $1 \le p \le 2$ *such that* $\ell^\tau_q \hookrightarrow \ell^p$,

$$\mathcal{K}^\tau_q\big(\mathcal{D}_p, M_p(\mathbb{R})\big) = \mathcal{K}^\tau_q\big(\mathcal{D}, M_2(\mathbb{R})\big), \tag{50}$$

and it follows that

$$\mathcal{K}^\tau_q\big(\mathcal{D}, M_2(\mathbb{R})\big) \hookrightarrow \mathcal{T}^\alpha_q\big(\mathcal{D}, M_p(\mathbb{R})\big), \qquad \alpha = 1/\tau - 1/p. \tag{51}$$

For a general a-moderate weight w, it is not clear under what conditions on g we can apply Lemma 4.2 to $\mathcal{D}^w$ a Gabor frame generated by g and normalized in $M_2^w(\mathbb{R})$. Moreover we have in general no expression of $\mathcal{K}^\tau_q\big(\mathcal{D}, M_2(\mathbb{R})\big)$ in terms of known smoothness spaces.

5 Gabor Banach frames in $M_p^w(\mathbb{R})$, $1 \le p < \infty$

Proposition 4.6 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) (see Eq. (46)). From the atomic decomposition theory for $M_p^w(\mathbb{R})$ (see [Grö00, Sec. 12.2.] for details) $\mathcal{D}$ constitutes a Banach frame for M_p^w, 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 a-moderate weights w and $1 \le p \le \infty$

$$\|f\|_{M_p^w(\mathbb{R})} \asymp \|\{\langle f, \widetilde{g}_{n,m}\rangle w_{n,m}\}_{n,m}\|_{\ell^p} \tag{52}$$

with $w_{n,m} = w\big(an, bm\big)$. Moreover, for $1 \le p < \infty$ the Gabor expansion

$$f = \sum_{n,m} \langle f, \widetilde{g}_{n,m}\rangle g_{n,m}$$

converges unconditionally in the norm of $M_p^w(\mathbb{R})$ for every $f \in M_p^w(\mathbb{R})$, and the synthesis operator $T\boldsymbol{c} = \sum_{n,m} c_{n,m} \frac{g_{n,m}}{\|g_{n,m}\|_{M_p^w}}$ is bounded from ℓ^p to M_p^w, i.e., $\mathcal{D}_p^w$ is ℓ^p-hilbertian [Grö00, Theorem 12.2.4].

By the Gabor expansion, we have $M_\tau^w(\mathbb{R}) \hookrightarrow \mathcal{K}_\tau^\tau(\mathcal{D}_p^w, M_p^w)$ for Gabor Banach frames and $1 \leq \tau < p < \infty$. The converse embedding, $\mathcal{K}_\tau^\tau(\mathcal{D}_p^w, M_p^w) \hookrightarrow M_\tau^w(\mathbb{R})$, follows by Lemma 3.1: we expand each $f \in \mathcal{K}_\tau^\tau(\mathcal{D}_p^w, M_p^w)$ as $f = \sum_{m,n} c_{m,n} g_{m,n}^w$ with $\|\boldsymbol{c}\|_{\ell^\tau} = \|f\|_{\mathcal{K}_\tau^\tau(\mathcal{D}_p^w, M_p^w)}$ and then apply the synthesis operator T to the sequence $\boldsymbol{c}$ to recover f.

Hence, for $1 \leq \tau < p < \infty$,

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

where the first equality is with equivalent norms. So in this case Theorem 3.2 gives [GS00, Proposition 3] as a corollary.

5. ON INVERSE ESTIMATES AND COMPLETE CHARACTERIZATIONS

We have shown with Theorem 3.1 and Theorem 3.2 that the ℓ_1^p-hilbertian property of $\mathcal{D}$ in X, with $p > 1$, is sufficient and almost necessary to get the following chain of embeddings

$$\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{T}_q^\alpha(\mathcal{D}) \hookrightarrow \mathcal{C}_q^\alpha(\mathcal{D}) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}), \tag{53}$$

with $p > 1$, $0 < \tau < p$, $\alpha = 1/\tau - 1/p$, and $0 < q \leq \infty$. In this short section we briefly discuss some conditions so that a Bernstein type embedding

$$\mathcal{A}_q^\alpha(\mathcal{D}) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D}), \qquad \tau = (\alpha + 1/r)^{-1} \tag{54}$$

holds with $r = p$ the same as in (53), in which case the chain of embeddings (53) collapses as in Theorem 1.1 and we get with equivalent norms :

$$\mathcal{A}_q^\alpha(\mathcal{D}) = \mathcal{T}_q^\alpha(\mathcal{D}) = \mathcal{K}_q^\tau(\mathcal{D}), \qquad \tau = (\alpha + 1/p)^{-1}.$$

In Section 5.1 we remind the reader that a Bernstein inequality is necessary to have a characterization of approximations spaces in terms of generalized smoothness spaces. In Section 5.2 we give some sufficient conditions on $\mathcal{D}$ and X (Corollary 5.1) so that (54) is valid.

5.1. Bernstein inequality

General results of approximation theory [DL93] relate the Bernstein embedding (54) to a Bernstein inequality for $\mathcal{K}_q^\tau(\mathcal{D})$ with exponent α :

$$|f_m|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq C m^\alpha \|f\|_X, \qquad m \geq 1, f_m \in \Sigma_m(\mathcal{D}). \tag{55}$$

Indeed, as shown by the following proposition, the Bernstein inequality must hold for (54) to be true.

PROPOSITION 5.1. *Let $\mathcal{D}$ an arbitrary dictionary in an arbitrary Banach space X. Suppose that for some $\alpha > 0$, some $0 < q, s \leq \infty$ and some $\tau > 0$ the embedding $\mathcal{A}_s^\alpha(\mathcal{D}) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D})$ holds. Then the Bernstein inequality* (55) *for $\mathcal{K}_q^\tau(\mathcal{D})$ with exponent α holds.*

Proof. By using $\mathcal{A}_s^\alpha(\mathcal{D}) \hookrightarrow \mathcal{K}_q^\tau(\mathcal{D})$ together with the Bernstein inequality for $\mathcal{A}_s^\alpha(\mathcal{D})$, see [DL93, Chap. 7; Theorem 9.3] we obtain for $f_m \in \Sigma_m(\mathcal{D})$:

$$|f_m|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq C\|f_m\|_{\mathcal{A}_s^\alpha(\mathcal{D})} \leq \tilde{C}m^\alpha \|f_m\|_X$$

which is the desired result. □

Remark 5. 1. In [GN01, Proposition 3.1] an example is given of a simple (complete and non-redundant) dictionary for which $\mathcal{A}_s^\alpha(\mathcal{D})$ cannot be embedded in any $\mathcal{K}_q^\tau(\mathcal{D})$.

Using a general theorem in [DL93, Chapter 7] one can derive from the Bernstein inequality (55) that for $0 < \gamma < \alpha$ and $0 < s \leq \infty$ there is a continuous embedding

$$\mathcal{A}_s^\gamma(\mathcal{D}) \hookrightarrow \left(X, \mathcal{K}_q^\tau(\mathcal{D})\right)_{\gamma/\alpha, s}. \tag{56}$$

Notice that this is not quite the embedding (54) we are looking for, since it is not clear in general whether the embedding

$$\left(X, \mathcal{K}_q^\tau(\mathcal{D})\right)_{\gamma/\alpha, s} \hookrightarrow \mathcal{K}_s^\eta(\mathcal{D}), \qquad \gamma = 1/\eta - 1/r \tag{57}$$

holds true. Next we consider a situation where a complete characterization is possible.

5.2. Bernstein inequality and adaptive analysis operator

In the special case where we have an adaptive analysis operator U (see Section 3.4) that maps continuously X into ℓ_∞^p, we have the following result (we leave the proof to the reader, it uses interpolation results that can be found in the appendix).

COROLLARY 5.1. *Let $\mathcal{D}$ an ℓ_1^p-hilbertian dictionary in X and $U : f \mapsto U(f) = \{g_k^\star(f)\}_{k \geq 1}$, with $g_k^\star \in X^\star$ a linear right inverse U to T that maps X into ℓ_∞^p. Assume that for some $0 < \tau < p$ and $0 < q \leq \infty$*

1.the operator U maps $\mathcal{K}_q^\tau(\mathcal{D})$ into ℓ_q^τ.
2.the Bernstein inequality for $\mathcal{K}_q^\tau(\mathcal{D})$ with exponent $\alpha = 1/\tau - 1/p$ holds;

Then we have, with equivalent norms, for $0 < \gamma < \alpha,\ 0 < s \leq \infty$,

$$\mathcal{A}_s^\gamma(\mathcal{D}) = \mathcal{K}_s^\eta(\mathcal{D}), \qquad \eta = (\gamma + 1/p)^{-1}.$$

If $p > 1$, we have in addition, with equivalent norms

$$\mathcal{A}_s^\gamma(\mathcal{D}) = \mathcal{T}_s^\gamma(\mathcal{D}) = \mathcal{K}_s^\eta(\mathcal{D}), \qquad \eta = (\gamma + 1/p)^{-1}.$$

Let us give a few examples.

EXAMPLE 5.1. For $\mathcal{D}$ a Schauder basis in X, the only linear right inverse to T is given by $g_k^\star, k \geq 1$, the coefficient functionals. Assuming that $P_s(\mathcal{D}, X) > 1$ (for example, it is sufficient to assume that X is super-reflexive [GG71]) we have by Proposition 3.1 that for any $0 < \tau < p$ and any $0 < q \leq \infty$, $|f|_{\mathcal{K}_q^\tau(\mathcal{D})} = \|\{g_k^\star(f)\}\|_{\ell_q^\tau}$ hence it is clear that U maps $\mathcal{K}_q^\tau(\mathcal{D})$ into ℓ_q^τ. Moreover for any such τ and q the Bernstein inequality comes for free as soon as we assume $U : X \to \ell_\infty^p$. Indeed for $f_m = \sum_{k \in I_m} g_k^\star(f) g_k$ with $|I_m| \leq m$ we have

$$|f_m|_{\mathcal{K}_q^\tau(\mathcal{D})} = \|\{g_k^\star(f)\}\|_{\ell_q^\tau} \leq m^{1/\tau - 1/p} \|\{g_k^\star(f)\}\|_{\ell_\infty^p} \leq C m^{1/\tau - 1/p} \|f_m\|_X.$$

COROLLARY 5.2. *Let $\mathcal{D} = \{g_k\}_{k \geq 1}$ a Schauder basis for X a Banach space and $\{g_k^\star\}_{k \geq 1}$ the coefficient functionals. Assume that the "sandwich" estimate*

$$C^{-1} \|\{g_k^\star(f)\}\|_{\ell_\infty^p} \leq \|f\|_X \leq C \|\{g_k^\star(f)\}\|_{\ell_1^p} \tag{58}$$

holds for $p > 1$ and some $C < \infty$. Then we have, with equivalent norms, for $0 < \tau < p$, $0 < q \leq \infty$,

$$\mathcal{A}_q^\alpha(\mathcal{D}) = \mathcal{T}_q^\alpha(\mathcal{D}) = \mathcal{K}_q^\tau(\mathcal{D}), \qquad \tau = (\alpha + 1/p)^{-1}. \tag{59}$$

Remark 5. 2. We refer the reader to [GN01, KP01, DKKT01] for more results and examples on nonlinear approximation with Schauder bases. Let us just mention here that there are examples of Schauder bases for which the sandwich estimate (58) fails, and for Schauder bases with some special structure, there is a converse to Corollary 5.1, that is to say (59) implies (58).

The next example demonstrates that an adaptive analysis operator may exist for redundant systems.

EXAMPLE 5.2. Consider $\mathcal{D}$ a (normalized) dictionary of twice oversampled framelets [DHRS01] in $X = L^p(\mathbb{R})$ based on a B-spline multiresolution analysis. The authors proved in [GN02b] that:

- $\mathcal{D}$ is ℓ_1^p-hilbertian in $L^p(\mathbb{R})$, $1 \leq p < \infty$.
- There is a Bernstein inequality for $\mathcal{D}$ which follows from Petrushev's results on approximation with free-knot splines [Pet88]. The exponent of the Bernstein inequality matches that of the Jackson inequality.
- There is an adaptive analysis operator U which follows from the fact that some "nice" biorthogonal wavelet has a finite expansion in $\mathcal{D}$. It maps $L^p(\mathbb{R})$ to ℓ_∞^p.

Thus we have, for $1 < p < \infty$, with $\tau = (\alpha + 1/p)^{-1}$,

$$\mathcal{A}_q^\alpha(\mathcal{D}, L^p(\mathbb{R})) = \mathcal{T}_q^\alpha(\mathcal{D}, L^p(\mathbb{R})) = \mathcal{K}_q^\tau(\mathcal{D}, L^p(\mathbb{R})).$$

We refer to [GN02b] for details and further related results.

Getting a Bernstein inequality is known to be a hard problem in general, and it is also generally hard to prove the existence of an adaptive analysis operator. Without such an operator, it seems that the interpolation properties of the family of smoothness spaces $\mathcal{K}_q^\tau(\mathcal{D})$ are difficult to deal with. In such a situation, even with a Bernstein inequality at hand, one is generally limited to getting an inverse result of the type (56).

In the case of Gabor Banach frames in modulation spaces, the adaptive analysis operator exists (cf Section 4 and [Grö00, Sec. 12.2.] for details), but no Bernstein inequality has been proved so far (see [GS00]). Also, for closed finitely generated shift invariant subspaces of $L^p(\mathbb{R}^d)$, there is an adaptive analysis operator [AST01], but no Bernstein inequality is known except in the special cases where the translates of the generators actually form a Schauder basis for the space and Corollary 5.2 applies. We will study Bernstein inequalities for some special structured dictionaries in a forthcoming paper [GN02a].

6. 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 generalized smoothness 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 generalized smoothness 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.

Let us use Corollary 3.1 to give an explicit example. From Corollary 3.1 we see that the generalized smoothness space associated with the union of two ℓ^p_s-hilbertian bases $\mathcal{B}_1$ and $\mathcal{B}_2$ for a Banach space X will consist of the vector sum of the two individual smoothness classes, $\mathcal{K}^\tau_q(\mathcal{B}_1 \cup \mathcal{B}_2) = \mathcal{K}^\tau_q(\mathcal{B}_1) + \mathcal{K}^\tau_q(\mathcal{B}_2)$. Thus, whenever those individual smoothness classes do not coincide, we gain by using the redundant dictionary $\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2$ in the sense that the domain for which there is a direct estimate is enlarged. This new larger domain can also (in theory, at least) be efficiently encoded provided that both $\mathcal{K}^\tau_q(\mathcal{B}_1)$ and $\mathcal{K}^\tau_q(\mathcal{B}_1)$ are "nice". Recall, that the number of bits needed to encode a compact class $K \subset X$ to within a distortion of $\varepsilon > 0$ is the Kolmogorov ε-entropy $\mathcal{E}^\varepsilon(K)$, see e.g. [CDDR01]. For $V \subset X$ normed by $\|\cdot\|_V$, denote $B_1[V] := \{f \in V : \|f\|_V \leq 1\}$, and let us assume that the unit balls $B_1[\mathcal{K}^\tau_q(\mathcal{B}_1)]$ and $B_1[\mathcal{K}^\tau_q(\mathcal{B}_1)]$ are both compactly embedded in X, then it is not hard to verify that $B_1[\mathcal{K}^\tau_q(\mathcal{B}_1 \cup \mathcal{B}_2)] \subset B_1[\mathcal{K}^\tau_q(\mathcal{B}_1)] + B_1[\mathcal{K}^\tau_q(\mathcal{B}_2)]$ is also compactly embedded in X and it follows easily that the Kolmogorov ε-entropy satisfies

$$\mathcal{E}^{2\varepsilon}(B_1[\mathcal{K}^\tau_q(\mathcal{B}_1 \cup \mathcal{B}_2)]) \leq \mathcal{E}^\varepsilon(B_1[\mathcal{K}^\tau_q(\mathcal{B}_1)]) + \mathcal{E}^\varepsilon(B_1[\mathcal{K}^\tau_q(\mathcal{B}_2)]).$$

Thus, asymptotically as $\varepsilon \to 0$, the order of $\mathcal{E}^\varepsilon(\mathcal{K}^\tau_q(\mathcal{B}_1 \cup \mathcal{B}_2))$ is no worse than the order of $\max\{\mathcal{E}^\varepsilon(\mathcal{K}^\tau_q(\mathcal{B}_1)), \mathcal{E}^\varepsilon(\mathcal{K}^\tau_q(\mathcal{B}_2)\}$. So, for the same number of bits (in the sense of order) we can in principle encode the larger smoothness space.

We have also considered the question of completely characterizing the different types of approximation classes in terms of generalized smoothness spaces. This question is closely related to obtaining an inverse estimate or Bernstein inequality. We prove that the existence of an adaptive analysis operator along with a Bernstein inequality leads to a complete characterization of the approximation classes associated with best m-term approximation in terms of generalized smoothness classes. So, for such dictionaries, the thresholding algorithm performs nearly optimally in the sense of the rate of approximation. Finally, we should mention that the problem of obtaining Bernstein estimates for structured redundant systems is studied in more detail in the 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.

APPENDIX

This appendix contains a short reminder on Lorentz spaces and results with detailed proofs on the structure of the generalized smoothness spaces and the approximation classes.

A.1. REMINDER: LORENTZ SPACES

First let us recall the Lorentz (quasi)norms, which are defined for $0 < \tau < \infty$ and $q \in (0,\infty]$ by,

$$\|\{a_m\}_{m=1}^{\infty}\|_{\ell_q^\tau} := \begin{cases} \left(\sum_{m=1}^{\infty} \frac{[m^{1/\tau} a_m^\star]^q}{m} \right)^{1/q}, & 0 < q < \infty \\ \sup_{m\in\mathbb{N}} m^{1/\tau} a_m^\star, & q = \infty, \end{cases} \tag{A.1}$$

where $\{a_k^\star\}$ denotes a decreasing rearrangement of $\{a_k\}$, *i.e.* $|a_k^\star| \geq |a_{k+1}^\star|$ for all $k \geq 1$. For $1 \leq q \leq \tau < \infty$, $\|\cdot\|_{\ell_q^\tau}$ is a norm for the Lorentz space

$$\ell_q^\tau = \{\{c_k\} : \|\{c_k\}\|_{\ell_q^\tau} < \infty\}.$$

Notice that $\|\cdot\|_{\ell_\tau^\tau} = \|\cdot\|_{\ell^\tau}$. It can be verified [LT79, Def. 2.b.8] that for $1 < \tau < q$, the quasi-norm $\|\cdot\|_{\ell_q^\tau}$ can be replaced by an equivalent norm on ℓ_q^τ. In such a case we always assume that we use the norm on ℓ_q^τ instead of the quantity defined by (A.1). For all values of τ, q, the Lorentz spaces ℓ_q^τ are (quasi)normed Banach spaces and satisfy the continuous embedding $\ell_{q_1}^{\tau_1} \hookrightarrow \ell_{q_2}^{\tau_2}$ provided that $\tau_1 < \tau_2$ or $\tau_2 = \tau_1$ with $q_1 \leq q_2$. It is useful to notice (see *e.g.* [DL93, Chap. 6]) that the ℓ_q^τ norm of a sequence can be estimated as

$$\|\{a_m\}_{m=1}^\infty\|_{\ell_q^\tau} \asymp \begin{cases} \left(\sum_{j=0}^{\infty} (2^{j/\tau} a_{2^j}^\star)^q\right)^{1/q}, & 0 < q < \infty \\ \sup_{j \geq 0} 2^{j/\tau} a_{2^j}^\star, & q = \infty. \end{cases} \tag{A.2}$$

A.2. STRUCTURE OF $\mathcal{A}(\mathcal{D})$

It is known [DL93, Chap. 7, Section 9] that $\mathcal{A}_q^\alpha(\mathcal{D})$ is a linear subspace of X and $\|\cdot\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)}$ is a (quasi)norm that makes it a complete metric space. The following proposition shows that if $q < \infty$, it is separable.

PROPOSITION A.7. *Let $f \in \mathcal{A}_q^\alpha(\mathcal{D})$, with $\mathcal{D}$ a dictionary in X a Banach space X, $\alpha > 0$, and $q \in (0, \infty]$. Let $f_m \in \Sigma_m(\mathcal{D})$ a sequence of (near) best m-term approximants to f,* i.e. *for some $C < \infty$ and all $m \geq 1$,*

$$\|f - f_m\|_X \leq C\sigma_m(f, \mathcal{D})_X. \tag{A.3}$$

If $q < \infty$ then $\|f - f_m\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} \to 0$. For all q,$\{f_m\}$ is bounded *in $\mathcal{A}_q^\alpha(\mathcal{D})$.*

Proof. Let us first prove some simple inequalities. First, we observe that

$$\sigma_l\,(f - f_m, \mathcal{D})_X \leq \|f - f_m\|_X \leq C\sigma_m\,(f, \mathcal{D})_X\,, \qquad l, m \geq 1. \tag{A.4}$$

Now for any $h_k \in \Sigma_k(\mathcal{D})$, we have $-f_m + h_k \in \Sigma_{m+k}(\mathcal{D})$ so

$$\sigma_{m+k}\,(f - f_m, \mathcal{D})_X \leq \|f - f_m - (-f_m + h_k)\|_X = \|f - h_k\|\,.$$

By taking the minimum over $h_k \in \Sigma_k(\mathcal{D})$ we get, for $k, m \geq 1$,

$$\sigma_{m+k}\,(f - f_m, \mathcal{D})_X \leq \sigma_k\,(f, \mathcal{D})_X$$

which we can write, for any $l = m + k \geq m + 1$, $m \geq 1$,

$$\sigma_l \left(f - f_m, \mathcal{D}\right)_X \leq \sigma_{l-m} \left(f, \mathcal{D}\right)_X . \tag{A.5}$$

Combining (A.4) and (A.5) we get for $q < \infty$

$$\begin{aligned}
|f - f_m|^q_{\mathcal{A}^\alpha_q(\mathcal{D},X)} &\asymp \sum_{l=1}^{2m} [l^\alpha \sigma_l \left(f - f_m, \mathcal{D}\right)_X]^q 1/l + \sum_{l=2m+1}^{\infty} [l^\alpha \sigma_l \left(f - f_m, \mathcal{D}\right)_X]^q 1/l \\
&\leq \sum_{l=1}^{2m} [l^\alpha \sigma_m \left(f, \mathcal{D}\right)_X]^q 1/l + \sum_{l=2m+1}^{\infty} [l^\alpha \sigma_{l-m} \left(f, \mathcal{D}\right)_X]^q 1/l \\
&\leq C m^{\alpha s} \sigma_m \left(f, \mathcal{D}\right)^q_X + \sum_{k=m+1}^{\infty} [(k+m)^\alpha \sigma_k \left(f, \mathcal{D}\right)_X]^q 1/(k+m) \\
&\leq C [m^\alpha \sigma_m \left(f, \mathcal{D}\right)_X]^q + \sum_{k=m+1}^{\infty} \left(\frac{k+m}{k}\right)^{\alpha s} [k^\alpha \sigma_k \left(f, \mathcal{D}\right)_X]^q 1/k \\
&\leq C [m^\alpha \sigma_m \left(f, \mathcal{D}\right)_X]^q + 2^{\alpha s} \sum_{k=m+1}^{\infty} [k^\alpha \sigma_k \left(f, \mathcal{D}\right)_X]^q 1/k.
\end{aligned}$$

The right hand side is easily seen to converge to zero as $m \to \infty$, because it is essentially the tail of a convergent series. For $q = \infty$ we derive $\|f_m\|_{\mathcal{A}^\alpha_q(\mathcal{D},X)} \leq C \|f\|_{\mathcal{A}^\alpha_q(\mathcal{D},X)}$ from the estimate

$$\begin{aligned}
|f - f_m|_{\mathcal{A}^\alpha_\infty(\mathcal{D},X)} &\asymp \sup_l l^\alpha \sigma_l(f - f_m, \mathcal{D}) \\
&= \max\left(\sup_{l \leq 2m} l^\alpha \sigma_l(f - f_m, \mathcal{D}), \sup_{l \geq 2m+1} l^\alpha \sigma_l(f - f_m, \mathcal{D})\right) \\
&\leq \max\left(\sup_{l \leq 2m} l^\alpha \sigma_m(f, \mathcal{D}), \sup_{k \geq m+1} \left(\frac{k+m}{k}\right)^\alpha \sup_{k \geq m+1} k^\alpha \sigma_k(f, \mathcal{D})\right) \\
&\leq 2^\alpha \sup_{k \geq m} k^\alpha \sigma_k(f, \mathcal{D}) \leq C \|f\|_{\mathcal{A}^\alpha_q(\mathcal{D},X)} \qquad \square.
\end{aligned}$$

A.3. STRUCTURE OF $\mathcal{C}(\mathcal{D})$

The general theory of approximation spaces [DL93] does not seem to apply to $\mathcal{C}^\alpha_q(\mathcal{D})$. In this section, we prove that $\mathcal{C}^\alpha_q(\mathcal{D})$ is nevertheless a linear subspace of X equipped with a (quasi)norm $\|\cdot\|_{\mathcal{C}^\alpha_q(\mathcal{D})}$.

Proposition A.8. *Let $\mathcal{D}$ a dictionary in a Banach space X, $\alpha > 0$ and $0 < q \le \infty$. The set $\mathcal{C}^\alpha_q(\mathcal{D})$ is a 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 $\left\|\{\|f - P_{\mathcal{V}_m(\pi,\mathcal{D})f\|}\}_{m\ge 1}\right\|_{\ell_q^{1/\alpha}} \le |f|_{\mathcal{C}^\alpha_q(\mathcal{D},X)} + \epsilon$ 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 $\psi_m \notin \{\phi_k, 1 \le k < 2n\}$; $\phi_{2n+1} = \pi_m$ where m is the smallest integer such that $\pi_m \notin \{\phi_k, 1 \le k < 2n+1\}$. As $\mathcal{V}_{2m}(\phi, \mathcal{D}) \supset \mathcal{V}_m(\pi, \mathcal{D}) + \mathcal{V}_m(\psi, \mathcal{D})$, we get for $j \ge 0$

$$\begin{aligned} \|f + g - P_{\mathcal{V}_{2^{j+1}}(\phi,\mathcal{D})}(f+g)\|_X &\le \|f - P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})} f + g - P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})} g\|_X \\ &\le \|f - P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})} f\|_X + \|g - P_{\mathcal{V}_{2^j}(\pi,\mathcal{D})} g\|_X. \end{aligned}$$

Using (A.2) and the fact that $\|\cdot\|_{\ell^\tau_q}$ is a (quasi)norm, we get easily that $|f+g|_{\mathcal{C}^\alpha_q(\mathcal{D})} \le C(\|f+g\|_X + |f|_{\mathcal{C}^\alpha_q(\mathcal{D})} + |f|_{\mathcal{C}^\alpha_q(\mathcal{D})})$. We conclude by letting ϵ go to zero. □.

A.4. STRUCTURE OF $\mathcal{K}(\mathcal{D})$

The (abstract) smoothness spaces also have a nice structure in general.

Proposition A.9. *Let $\mathcal{D}$ a dictionary in a Banach space X, $0 < \tau < \infty$ and $0 < q \le \infty$. Then $|\cdot|_{\mathcal{K}^\tau_q(\mathcal{D})}$ is a semi-(quasi)norm on the linear space $\mathcal{K}^\tau_q(\mathcal{D})$. Moreover, if $\|\cdot\|_X \lesssim |\cdot|_{\mathcal{K}^\tau_q(\mathcal{D})}$, then $|\cdot|_{\mathcal{K}^\tau_q(\mathcal{D})}$ is actually a (quasi)norm and $\mathcal{K}^\tau_q(\mathcal{D})$ is complete.*

Proof. It is easy to check $|\lambda f|_{\mathcal{K}^\tau_q(\mathcal{D})} = |\lambda|\, |f|_{\mathcal{K}^\tau_q(\mathcal{D})}$. Let us now check

$$|f_1 + f_2|_{\mathcal{K}^\tau_q(\mathcal{D})} \le C_{\tau,q}\left(|f_1|_{\mathcal{K}^\tau_q(\mathcal{D})} + |f_2|_{\mathcal{K}^\tau_q(\mathcal{D})}\right) \tag{A.6}$$

with $C = C(\tau, q)$ the constant of the (quasi)triangle inequality in ℓ^τ_q. Let $\epsilon > 0$. We can find finite approximants $\widehat{f}_j = T\boldsymbol{c}_j$ to f_j from $\mathcal{D}$, such that $\|\boldsymbol{c}_j\|_{\ell^\tau_q} \le |f_j|_{\mathcal{K}^\tau_q(\mathcal{D})} + \epsilon/2$ and $\|\widehat{f}_j - f_j\|_X \le \epsilon/2$. Their sum $\widehat{f} := \widehat{f}_1 + \widehat{f}_2 = T(\boldsymbol{c}_1 + \boldsymbol{c}_2)$ is a finite approximant to $f_1 + f_2$ from $\mathcal{D}$ such that $\|\widehat{f} - (f_1 + f_2)\|_X \le \epsilon$, and the (quasi)triangle inequality in ℓ^τ_q gives

$$\|\boldsymbol{c}_1 + \boldsymbol{c}_2\|_{\ell^\tau_q} \le C\big(|f_1|_{\mathcal{K}^\tau_q(\mathcal{D})} + |f_2|_{\mathcal{K}^\tau_q(\mathcal{D})} + \epsilon\big).$$

As we can let ϵ go to zero, this shows (A.6). With the additional assumption that $|\cdot|_{\mathcal{K}^\tau_q(\mathcal{D})} \gtrsim \|\cdot\|_X$, we get $|f|_{\mathcal{K}^\tau_q(\mathcal{D})} = 0 \Leftrightarrow f = 0$. Let us now turn to the completeness result. Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{K}_q^\tau(\mathcal{D})$. Because $|\cdot|_{\mathcal{K}_q^\tau(\mathcal{D})} \gtrsim \|\cdot\|_X$, $\{f_n\}$ is also a Cauchy sequence in X. Hence there exists $f \in X$, $\|f_n - f\|_X \to 0$. Let us prove that $|f_n - f|_{\mathcal{K}_q^\tau(\mathcal{D})} \to 0$. It will imply that f is the limit in $\mathcal{K}_q^\tau(\mathcal{D})$ of $\{f_n\}$.

Let $g_{n,l} = f_{n+l} - f_n$ and $\epsilon > 0$: as f_n is a Cauchy sequence in $\mathcal{K}_q^\tau(\mathcal{D})$, we can find N big enough so that $\forall n \geq N, l \geq 0$, $|g_{n,l}|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq \epsilon/2$. Now we can find finite approximants $T\boldsymbol{c}_{n,l}$ such that $\|g_{n,l} - T\boldsymbol{c}_{n,l}\|_X \leq 1/l$ and $\|\boldsymbol{c}_{n,l}\|_{\ell_q^\tau} \leq 2|g_{n,l}|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq \epsilon$. As a result, for all $n \geq N$ and $p \geq 0$,

$$\begin{aligned} \|(f - f_n) - T\boldsymbol{c}_{n,l}\|_X &\leq \|(f - f_n) - g_{n,l}\|_X + \|g_{n,l} - T\boldsymbol{c}_{n,l}\|_X \\ &\leq \|f - f_{n+l}\|_X + 1/l \end{aligned}$$

It is now clear that, for any fixed $n \geq N$, $\lim_{l\to\infty} \|(f - f_n) - T\boldsymbol{c}_{n,l}\|_X = 0$. Moreover, as $\|\boldsymbol{c}_{n,l}\|_{\ell_q^\tau} \leq \epsilon$, we have shown that for all $n \geq N$

$$|f - f_n|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq \epsilon. \qquad \square$$

A.5. INTERPOLATION OF SMOOTHNESS CLASSES

We use the results of Section 3 on hilbertian dictionaries to get one-sided embeddings of $\mathcal{K}_q^\tau(\mathcal{D})$ into the interpolation space between two associated smoothness spaces. We use the real-method of interpolation (for details, we refer to [DL93, Chap. 7]). The notations are that of Section 3.

LEMMA A.1. *Assume that $P_s(\mathcal{D}, X) > 1$ and that T is continuous from $\ell_{q_i}^{\tau_i}$ to X, $0 < \tau_1 < \tau_2 \leq P_s(\mathcal{D}, X)$. Then for any $0 < \theta < 1$, $0 < q \leq \infty$ we have with $1/\tau = (1-\theta)/\tau_2 + \theta/\tau_1$*

$$\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \left(\mathcal{K}_{q_2}^{\tau_2}(\mathcal{D}), \mathcal{K}_{q_1}^{\tau_1}(\mathcal{D})\right)_{\theta,q} \hookrightarrow \left(X, \mathcal{K}_{q_1}^{\tau_1}(\mathcal{D})\right)_{\theta,q}. \tag{A.7}$$

Proof. The rightmost embedding follows immediately from $\mathcal{K}_{q_2}^{\tau_2}(\mathcal{D}) \hookrightarrow X$. Let $f \in \mathcal{K}_q^\tau(\mathcal{D})$. As $\tau < \max(\tau_1, \tau_2) \leq P_s(\mathcal{D}, X)$ we can use Lemma 3.1 to find $\boldsymbol{c} \in \ell_q^\tau$ such that $f = T\boldsymbol{c}$ and $|f|_{\mathcal{K}_q^\tau(\mathcal{D})} = \|\boldsymbol{c}\|_{\ell_q^\tau}$. As T is continuous from $\ell_{q_i}^{\tau_i}$ to $\mathcal{K}_{q_i}^{\tau_i}(\mathcal{D})$, $i = 1, 2$, by interpolation T is also continuous from $\ell_q^\tau = \left(\ell_{q_2}^{\tau_2}, \ell_{q_1}^{\tau_1}\right)_{\theta,q}$ to $Y := \left(\mathcal{K}_{q_2}^{\tau_2}(\mathcal{D}), \mathcal{K}_{q_1}^{\tau_1}(\mathcal{D})\right)_{\theta,q}$. Hence, $|f|_Y = |T\boldsymbol{c}|_Y \leq C\|\boldsymbol{c}\|_{\ell_q^\tau} = C|f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. $\square$

Next we show that the existence of an adaptive analysis operator (see Section 3.4) is sufficient to get that $\mathcal{K}_q^\tau(\mathcal{D})$ forms an interpolation family.

LEMMA A.2. *Assume $U : f \mapsto U(f) = \{g_k^\star(f)\}_{k\geq 1}$, with $g_k^\star \in X^\star$, is a linear right inverse to T on X that continuously maps X into ℓ_∞^r and $\mathcal{K}_q^\tau$ into ℓ_q^τ, for some r, $\tau < P_s(\mathcal{D}, X)$ and $0 < q \leq \infty$. Then for any $0 < s \leq \infty$ and any $0 < \theta < 1$ such that $\eta < P_s(\mathcal{D}, X)$ with $1/\eta := (1-\theta)/r + \theta/\tau$, we have*

$$\Big(X, \mathcal{K}_q^\tau(\mathcal{D})\Big)_{\theta,s} \hookrightarrow \mathcal{K}_s^\eta(\mathcal{D}) \tag{A.8}$$

Proof. By interpolation U is continuous from $Y := \Big(X, \mathcal{K}_q^\tau(\mathcal{D})\Big)_{\theta,s}$ to ℓ_s^η. Assuming that $\eta < P_s(\mathcal{D}, X)$, we use the fact that U is a right inverse to T on X to get for all $f \in Y \subset X$, $f = TU(f) = \sum_k g_k^\star(f) g_k$ where the series is unconditionally convergent in X because $U(f) \in \ell_q^\eta$ and $\mathcal{D}$ is ℓ_q^η-hilbertian. It follows that $|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \leq \|U(f)\|_{\ell_q^\tau(\mathcal{D})} \leq C|f|_Y$. □
Notice by the way that we get that U is adaptive on the intermediate spaces.

REFERENCES

AST01. A. Aldroubi, Q. Sun, and W.-S. Tang. p-frames and shift invariant subspaces of l^p. *J. Fourier Anal. Appl.*, 7(1):1–21, 2001.

AWW92. P. Auscher, G. Weiss, and M. V. Wickerhauser. *Wavelets*, chapter Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets., pages 237–256. Academic Press, Boston, MA, 1992.

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.

DS52. R.J. Duffin and A.C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72:341–366, 1952.

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

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.

GG71. V. I. Gurariĭ and N. I. Gurariĭ. Bases in uniformly convex and uniformly smooth Banach spaces. *Izv. Akad. Nauk SSSR Ser. Mat.*, 35:210–215, 1971.

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. Technical Report 0202, IMI, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2002.

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. *The Annals of Statistics*, 13(2):435–475, 1985.

Jon87. L.K. Jones. On a conjecture of Huber concerning the convergence of PP-regression. *The Annals of Statistics*, 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.

Pet88. Pencho P. Petrushev. Direct and converse theorems for spline and rational approximation and Besov spaces. In *Function spaces and applications (Lund, 1986)*, pages 363–377. Springer, Berlin, 1988.

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 $L^p(\mathbb{R}^d)$. Technical report, 2000.

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. Technical Report 0010, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2000.

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

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.