ASPECTS OF NONLINEAR APPROXIMATION WITH DICTIONARIES

by

Morten Nielsen
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G
DK-9220 Aalborg Øst
E-mail: mnielsen@math.aau.dk

Preface

This doctoral thesis is concerned with certain aspects of m-term nonlinear approximation with a (possibly redundant) function dictionary. It is based on twelve selected papers [**A-L**] (see page 2) produced over a period of six years.

The thesis consists of four parts. The first part deals with nonlinear approximation using dictionaries with minimal assumptions on dictionary structure. The second part concerns approximation with structured wavelet-type dictionaries, where the wavelet structure leads to stronger and more refined results than for general dictionaries. The third part deals with approximation with time-frequency frame dictionaries in decomposition-type smoothness spaces. The final part concerns sparse representation of signals in a finite dimensional space.

I owe thanks to many friends, collaborators, and colleagues around the world. Most of all, I would like to thank my close collaborators Rémi Gribonval and Lasse Borup. I am indebted to my former colleagues in the Department of Mathematics at the University of South Carolina–Columbia; their approach to approximation theory significantly influenced this thesis. I would also like to thank my current colleagues in the Department of Mathematical Sciences at Aalborg University for providing a friendly and inspiring professional environment. In particular, Arne Jensen is acknowledged for his encouragement and constant support since my days as an undergraduate.

Aalborg, September 2007 Morten Nielsen

List of papers

This thesis is based on the following twelve papers:

[A] L. Borup, R. Gribonval, and M. Nielsen. Bi-framelet systems with few vanishing moments characterize Besov spaces. *Appl. Comput. Harmon. Anal.*, 17(1):3–28, 2004.

[B] L. Borup and M. Nielsen. Approximation with wave packets generated by a refinable function. *Proc. Amer. Math. Soc.*, 133(8):2409–2418 (electronic), 2005.

[C] L. Borup and M. Nielsen. Nonlinear approximation in α-modulation spaces. *Math. Nachr.*, 279(1-2):101–120, 2006.

[D] L. Borup and M. Nielsen. Frame decomposition of decomposition spaces. *J. Fourier Anal. Appl.*, 13(2):39–70, 2007.

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

[F] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

[G] R. Gribonval and M. Nielsen. Letter to the editor: on a problem of Gröchenig about nonlinear approximation with localized frames. *J. Fourier Anal. Appl.*, 10(4):433–437, 2004.

[H] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51–71, 2004.

[I] R. Gribonval and M. Nielsen. On approximation with spline generated framelets. *Constr. Approx.*, 20(2):207–232, 2004.

[J] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. II. Inverse estimates. *Constr. Approx.*, 24(2):157–173, 2006.

[K] R. Gribonval and M. Nielsen. Beyond sparsity: recovering structured representations by l^1-minimization and greedy algorithms. *Adv. Comput. Math. (in press)*, 2007.

[L] R. Gribonval and M. Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Appl. Comput. Harmon. Anal.*, 22(3):335–355, 2007.

Contents

CHAPTER 1

Introduction

The fundamental problem of approximation theory is to resolve a possibly complicated target function by simpler functions called approximants. By increasing the complexity of the approximants, we can hope to obtain a better resolution of the target function, and one of the main goals of constructive approximation theory is to obtain quantitative estimates for the trade-off between resolution and complexity.

In this thesis we study a specific approximation model. We choose a Banach space X whose elements are considered target functions, and we form approximants by taking linear combination of m elements from a fixed countable system $\mathcal{D} \subset X$, called a dictionary. We call such an approximant an m-term approximant with respect to $\mathcal{D}$, and we use m as a measure of its complexity. The Banach space norm is used to measure the distance (error) between the target function and the approximant in question.

We have great freedom in our choice of strategy to construct the m-term approximants using $\mathcal{D}$. The path we follow here is consider so-called best m-term approximation. We fix a target function and $m \geq 1$. Then we (formally) create the approximant by requiring that the m-terms we select give the smallest possible error among all possible m-term approximants. This approximation procedure is *nonlinear* in the sense that the approximants are *not* given by a linear operator on the space of target functions. The main theoretical importance of best m-term approximation is that it provides a benchmark that can be used to evaluate the quality of any other method (or algorithm) to construct approximants. Another strong argument in favor of this type of approximation is that the procedure makes sense for any dictionary $\mathcal{D}$, even dictionaries without any type of structure.

For best m-term approximation, we study the following problem. Given a dictionary $\mathcal{D}$, we would like to characterize the family of elements $\mathcal{A}^\alpha \subset X$ for which the error of the best m-term approximation from $\mathcal{D}$ decays at the rate $\mathcal{O}(m^{-\alpha})$. The approximation class $\mathcal{A}^\alpha$ consists of objects that can be compressed "well" by using relatively few elements from $\mathcal{D}$. From a practical point of view, we would like to obtain a large approximation class $\mathcal{A}^\alpha$ without increasing the size and complexity of $\mathcal{D}$ too much. One important question is what happens when we go from a non-redundant dictionary, such as a basis, to a redundant dictionary. Intuitively it is clear that we should be able to gain "something" by using an overcomplete dictionary $\mathcal{D}$ compared to, say, an orthonormal bases. Quantitative estimates of the trade-off between the structure of $\mathcal{D}$ and the "size" of $\mathcal{A}^\alpha$ are at the core of the work presented here.

The history of nonlinear approximation is very rich and here we mention only some of the highlights related to m-term approximation. Motivated by problems related to integral equations, E. Schmidt [**93**] was the first to study m-term approximation with a dictionary. Ismagilov [**69**] was the first to obtain error estimates in m-term approximation that showed an advantage over linear methods. He obtained estimates for best m-term trigonometric approximation in L_∞ for specific functions.

The systematic study of nonlinear approximation was initiated in the 1950's, where this area of research took off inspired by problems in spline approximation (approximation with piecewise polynomials). In particular, the work of Birman and Solomjak [**6**] on adaptive spline approximation was influential. Later Brudnyi [**11**] and Bergh and Peetre [**4**] introduced various abstract smoothness spaces in order to better understand adaptive spline approximation. A breakthrough came with Petrushev's [**90**] characterization of univariate free knot spline approximation in terms of Besov spaces. Petrushev was thus the first to explain the advantages of nonlinear approximation in terms of classical smoothness spaces. The m-term approximation approach was first used for multivariate splines by Oskolkov [**85**].

Another highlight came in the 1980's where wavelet bases and other multiscale methods were introduced. Wavelets are especially remarkable since they provide unconditional bases for many of the smoothness spaces studied in approximation theory such as Besov and Sobolev spaces. Moreover, wavelets provided a non-redundant version of the Littewood-Paley decomposition, simplifying the analysis of the operators studied in harmonic analysis, see [**25**,**82**]. The problem of characterizing best m-term approximation in L_p with wavelet bases was completely settled by DeVore, Jawerth and Popov [**33**]. Later, Temlyakov [**97**] showed that the approximation results with wavelets in L_p can be understood in terms of so-called greedy bases for L_p.

Wavelets have also turned out to be a very efficient tool in signal processing. In particular, they provide an efficient tool for compressing images. The reason is that the wavelet representation of a natural image is often sparse. Recently the search for more efficient methods to obtain sparse representations of natural images has shown that new (often redundant) decomposition systems can produce sparser representations of certain natural images than wavelets. One important new decomposition system is the curvelet frame introduced by Candès and Donoho [**12**]. Sparse representations of signals relative to a redundant dictionary is another topic considered in this thesis.

Let us describe the contents of this thesis in detail. The work presented is divided into four main categories.

1. Nonlinear approximation with general dictionaries. Based on the papers [**E**, **G**, **H**, **J**], and presented in Chapter 3. We study nonlinear approximation with general dictionaries in a Banach space. The notion of a sparseness class for a general dictionary is introduced and Jackson type estimates are obtained for very general dictionaries. It is also demonstrated that for some structured redundant dictionaries it may not be possible to obtain a complete characterization of the associated approximation classes. In the final part of the chapter, we consider a class of non-redundant dictionaries, the so-called greedy bases in a Banach space. For such dictionaries, a complete characterization of the approximation classes is obtained.

2. Approximation with wavelet-type dictionaries. Based on the papers [**A**, **B**, **I**], and presented in Chapter 4. The results of Chapter 3 show that it may be impossible to get characterizations of the approximation classes unless the underlying dictionary is very structured. In Chapter 4, we study wavelet frame dictionaries. Wavelet frames are (possibly) redundant dictionaries with the same structure as classical orthonormal wavelet bases. The wavelet structure enables us to obtain much more refined results than in Chapter 3. The main result is a complete characterization of the approximation classes associated with m-term wavelet frame approximation in L_p. The result is an extension of the result of DeVore, Jawerth and Popov [**33**] to the redundant wavelet frame case.

3. Approximation with time-frequency frames. Based on the papers [**C**, **D**], and presented in Chapter 5. A construction of tight frames for $L_2(\mathbb{R}^d)$ with flexible time-frequency localization is considered. The frames can be adapted to form atomic decompositions for a large family of smoothness spaces on $\mathbb{R}^d$, a class of so-called decomposition spaces. The decomposition space norm can be completely characterized by a sparseness condition on the frame coefficients. This leads to natural Jackson-type estimates for m-term approximation. In particular, we consider approximation with time-frequency frames in so-called α-modulation spaces. In the univariate case, greedy bases for α-modulation spaces are constructed and a complete characterization of the approximation classes associated with m-term approximation is given.

4. Sparse representation of signals. Based on the papers [**F**, **K**, **L**], and presented in Chapter 6. In this final chapter, we change the point of view slightly and consider a computational problem. Given a finite dimensional space, a redundant dictionary, and a target function; how do we efficiently compute a sparse representation of the target function relative to the dictionary? A sparse representation can be used to compute an efficient approximation to the target function. We give several sufficient conditions on the target function to ensure that a certain polynomial time algorithm will recover the sparsest representation.

Chapter 2 contains some basic notation and some fundamental results on approximation theory.

CHAPTER 2

Some basic notation and results

Let us introduce some notation that will be used throughout this thesis. Let X be a Banach space, and let $\mathcal{D} = \{g_k\}_{k\in F} \subset X$ be a countable collection of non-zero elements, called a dictionary. The nonlinear manifold of m-term approximants with respect to $\mathcal{D}$ is given by

$$\Sigma_m(\mathcal{D}) := \left\{ \sum_{k\in\Lambda} c_k g_k : g_k \in \mathcal{D}, \Lambda \subset F; |\Lambda| = m \right\}, \tag{2.1}$$

where the c_k's are arbitrary scalars in X. The error of best m-term approximation to $f \in X$ is defined as

$$\sigma_m(f,\mathcal{D})_X := \inf_{g\in\Sigma_m(\mathcal{D})} \|f-g\|_X. \tag{2.2}$$

Often we will use $\sigma_m(f,\mathcal{D})_X$ as a theoretical benchmark for approximation with $\mathcal{D}$. No matter which procedure (or algorithm) that is used to create an m-term approximant from $\mathcal{D}$, we cannot obtain a smaller approximation error than $\sigma_m(f,\mathcal{D})_X$. In general, we will be very happy with an algorithm if it produces m-term approximants with approximation error that is proportional to $\sigma_m(f,\mathcal{D})_X$. At present, no such algorithm is know for a general dictionary.

To get a more precise classification of the elements in X, we define $\|f\|_{\mathcal{A}_s^\alpha(\mathcal{D})} := \|f\|_X + |f|_{\mathcal{A}_s^\alpha(\mathcal{D})}$, where

$$|f|_{\mathcal{A}_s^\alpha(\mathcal{D})} := \begin{cases} \left(\sum_{m=1}^{\infty} [m^\alpha \sigma_m(f,\mathcal{D})_X]^q \frac{1}{m} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{m\geq 1} m^\alpha \sigma_m(f,\mathcal{D})_X, & q = \infty. \end{cases} \tag{2.3}$$

Then we define

$$\mathcal{A}_s^\alpha(\mathcal{D}) = \{ f \in X : \|f\|_{\mathcal{A}_s^\alpha(\mathcal{D})} < \infty \}.$$

The class $\mathcal{A}_q^\alpha(\mathcal{D},X)$ is thus basically the set of target 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 [**35**, Chapter 7]. Moreover, $\mathcal{A}_s^\alpha(\mathcal{D})$ is continuously embedded[1] in X.

With the notation in place, we can precisely state one of the main problems studied in this thesis:

- Given a dictionary $\mathcal{D}$ in X, characterize the approximation class $\mathcal{A}_s^\alpha(\mathcal{D})$.

To approach this characterization problem, we need two fundamental estimates of approximation theory, the so-called Jackson and Bernstein inequalities. Let Y be a semi-normed space continuously embedded in the Banach space X. Suppose that for $r > 0$, the following two fundamental inequalities of approximation theory hold

$$\text{(Jackson inequality)} \qquad \sigma_m(f,\mathcal{D})_X \le Cm^{-r}|f|_Y, \qquad m = 1,2,\dots, \tag{2.4}$$

$$\text{(Bernstein inequality)} \qquad |S|_Y \le Cm^r\|f\|_X, \qquad S \in \Sigma_m(\mathcal{D}). \tag{2.5}$$

The importance of the Jackson and Bernstein estimate is made clear by the following fundamental result by DeVore and Popov, which gives a complete characterization of $\mathcal{A}_q^\alpha(\mathcal{D})$. We use $(X,Y)_{\theta,q}$ to denote the interpolation space (using the real interpolation method) between X and Y with parameters (θ,q), see [**3**].

THEOREM 1 ([**36**]). *If the Jackson inequality and the Bernstein inequality are both valid for some $r > 0$, then for each $0 < \gamma < r$ and $0 < q \le \infty$ the following relation holds between approximation spaces and interpolation spaces*

$$\mathcal{A}_q^\gamma(\mathcal{D}) = (X,Y)_{\gamma/r,q},$$

with equivalent norms.

In many cases it will be possible to characterize $(X,Y)_{\gamma/r,q}$ in terms of classical spaces. For example, this will be the case when we study approximation with wavelet-type systems in Chapter 4, where the approximation spaces are identified as Besov spaces.

Theorem 1 is actually valid in an even more general setup, see [**35**] for additional information. The proof of Theorem 1 is also quite illuminating; it shows that it makes sense to consider the two inequalities (2.4) and (2.5) separately. The Jackson inequality alone can be used to derive the following continuous embedding, which we will refer to as a Jackson embedding,

$$(X,Y)_{\gamma/r,q} \hookrightarrow \mathcal{A}_q^\gamma(\mathcal{D}). \tag{2.6}$$

Notice that the Jackson inequality gives a "minimal size" estimate of $\mathcal{A}_q^\gamma(\mathcal{D})$; it must at least as large as $(X,Y)_{\gamma/r,q}$. The Bernstein inequality on its own can be used to derive the Bernstein embedding

$$\mathcal{A}_q^\gamma(\mathcal{D}) \hookrightarrow (X,Y)_{\gamma/r,q}. \tag{2.7}$$

[1]We use the symbol $\hookrightarrow$ to denote a continuous embedding

The Bernstein embedding gives a "maximal size" estimate of $\mathcal{A}_q^\gamma(\mathcal{D})$; it cannot be larger than $(X,Y)_{\gamma/r,q}$.

By separating the Jackson and Bernstein estimates we can still derive interesting results in cases where one of the estimates cannot be obtained. We mention here that the Bernstein inequality, in general, is much more difficult to derive than the Jackson inequality, and in Chapter 3 we often study the Jackson embeddings separately.

Another aspect that has to be considered is how to choose the embedded space Y? If we have no good candidate for Y, then the Jackson/Bernstein approach is obviously not a possible path forward. In Chapter 4 where we consider wavelet-type approximation, it turns out that we have a good candidate for Y, namely a suitable Besov space. However, in Chapter 3, we consider the abstract problem of m-terms approximation with a dictionary. In that case we have to introduce a suitable abstract space Y that is built using the dictionary.

We conclude this chapter by recalling the definition of the Lorentz (quasi)-norms for $0 < \tau < \infty$ and $q \in (0,\infty]$. The Lorentz norms are used in Chapter 3 to build suitable candidates for smoothness/sparseness classes for a general dictionary. For any scalar sequence $\{a_m\}_{m\in\mathbb{N}}$ we define

$$\|\{a_m\}_{m=1}^\infty\|_{\ell_q^\tau} := \begin{cases} \left(\sum_{m=1}^\infty \frac{[m^{1/\tau}a_m^\star]^q}{m}\right)^{1/q}, & 0<q<\infty \\ \sup_{m\in\mathbb{N}} m^{1/\tau}a_m^\star, & q=\infty, \end{cases} \tag{2.8}$$

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

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

It can be verified **[35]** that for $1 < \tau \leq q$, 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 (2.8). 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 $q_1 \leq q_2$. The standard ℓ_τ-norm is defined by $\|\cdot\|_{\ell_\tau} := \|\cdot\|_{\ell_\tau^\tau}$.

CHAPTER 3

Approximation with general dictionaries

In this chapter, we consider the problem of characterizing approximation spaces for m-term approximation with a general dictionary in a Banach space. Depending on the amount of structure of the dictionary, we derive several types of Jackson embeddings. We begin by studying dictionaries with essentially no structure, and later we require additional structure to improve the estimates. The corresponding Bernstein estimates turn out to be harder to obtain, and we demonstrate that there may not be any such estimate even for nice dictionaries. In the finite dimensional case, we derive a Bernstein estimate for so-called incoherent dictionaries. The final part of the chapter is concerned with non-redundant dictionaries that form so-called greedy bases.

The problem of characterizing approximation spaces for m-term approximation with dictionaries has been considered by a number of authors in various settings. We mention here that general dictionaries are considered in [**12, 28–30, 37–39**]. Spline dictionaries are studied in [**6, 11, 71, 87, 89, 90**]. Approximation with Gabor systems is studied in [**67**], while wavelet-type systems are treated in [**2, 21, 33, 34, 77, 78, 97**]. The history of approximation with dictionaries is discussed in detail in the articles by DeVore [**31**] and by Temlyakov [**98**].

3.1. Sparseness classes

In order to study Jackson-type estimates, we need to introduce a suitable candidate for the embedded space Y that is needed in the Jackson inequality (2.4). In the study of nonlinear wavelet approximation in L_p, it is customary to take Y to be a Besov space, i.e., a classical smoothness space, see [**33**]. However, in a general Banach space, the equivalent of a smoothness spaces has to be defined in terms of the dictionary since no other structure is assumed. The definition below follows DeVore and Temlyakov [**37**] and is inspired by the classical principle that smoothness is often equivalent to sparseness of expansion coefficients. Thus, we use a sparseness space in place of a classical smoothness space.

A dictionary $\mathcal{D} = \{g_k\}$ is in X is called quasi-normalized if $\inf_k \|g_k\|_X > 0$ and $\sup_k \|g_k\|_X < \infty$. For a quasi-normalize dictionary $\mathcal{D}$, we define the *sparsity classes* $\mathcal{K}_q^\tau(\mathcal{D}, X)$ as follows. For $\tau \in (0,\infty)$ and $q \in (0,\infty]$ we let $\mathcal{K}_q^\tau(\mathcal{D}, X, M)$

denote the set

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

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

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D},X)} := \inf\{M, f \in \mathcal{K}_q^\tau(\mathcal{D}, X, M)\}.$$

It can be proved that $|\cdot|_{\mathcal{K}_q^\tau(\mathcal{D},X)}$ is a (semi)-(quasi)norm on $\mathcal{K}_q^\tau(\mathcal{D}, X)$. The definition of $\mathcal{K}_q^\tau(\mathcal{D}, X)$ in terms of a certain relative closure may seem rather technical, but the good news is that for many dictionaries there is a much simplified definition. We need the class of ℓ_q^τ-hilbertian dictionaries.

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

$$\left\| \sum_{k\geq 1} c_k g_k \right\|_X \lesssim \|\mathbf{c}\|_{\ell_q^\tau}.$$

We notice that the convergence of $\sum_k c_k g_k$ in Definition 1 is necessarily unconditional, provided that ℓ_q^τ is not one of the extremal non-separable spaces such as ℓ_∞. Also notice that *any* dictionary is ℓ_τ-hilbertian for $0 < \tau \leq 1$.

The hilbertian structure of $\mathcal{D}$ makes it possible to get a nice representation of the sparsity spaces $\mathcal{K}_q^\tau(\mathcal{D})$. For an arbitrary dictionary $\mathcal{D} = \{g_k\}$, we define the "reconstruction" operator

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

on the space ℓ_0 of finite sequences $\mathbf{c} = \{c_k\}$. We have the following representation result, giving a more direct interpretation of $\mathcal{K}_q^\tau(\mathcal{D})$ as a sparseness class.

THEOREM 2 ([**H**]). *Assume that $\mathcal{D}$ is ℓ_1^p-hilbertian for some $p > 1$. Let $\tau < p$ and $1 \leq q \leq \infty$. For all $f \in \mathcal{K}_q^\tau(\mathcal{D})$, there exists some $\mathbf{c} \in \ell_q^\tau$ which realizes the sparsity norm, i.e., $f = T\mathbf{c}$ and $\|\mathbf{c}\|_{\ell_q^\tau} = |f|_{\mathcal{K}_q^\tau(\mathcal{D})}$. In case $1 < \tau, q < \infty$, $\mathbf{c} = \mathbf{c}_{\tau,q}(f)$ is unique. Consequently*

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D})} = \min_{\mathbf{c}\in\ell_q^\tau, f=T\mathbf{c}} \|\mathbf{c}\|_{\ell_q^\tau}, \tag{3.1}$$

and

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

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

Let us mention an important application of Theorem 2. Suppose $\mathcal{D}_1$ and $\mathcal{D}_2$ are both ℓ_1^p-hilbertian dictionaries for some $p > 1$. Let $0 < \tau < p$ and

$1 \le q \le \infty$. Then for $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$

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

Thus, whenever the individual sparsity spaces $\mathcal{K}_q^\tau(\mathcal{D}_1)$ and $\mathcal{K}_q^\tau(\mathcal{D}_2)$ do not coincide, we gain by using the redundant dictionary $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ in the sense that the joint sparseness class is strictly larger that the individual sparseness classes. Consider, as a specific example, the case of $\mathcal{D}_1$ a nice univariate wavelet basis and $\mathcal{D}_2$ a local Fourier basis (see **[23]**) in $X = L^2(\mathbb{R})$. The individual sparsity spaces are respectively a Besov space $\mathcal{K}_\tau^\tau(\mathcal{D}_1, X) = B_{\tau,\tau}^\alpha(\mathbb{R})$ and a modulation space $\mathcal{K}_\tau^\tau(\mathcal{D}_2, X) = M^\tau(\mathbb{R})$, see [**67**, Theorem 2]. In this particular case, we have $\mathcal{K}_\tau^\tau(\mathcal{D}_1 \cup \mathcal{D}_2, X) = B_{\tau,\tau}^\alpha(\mathbb{R}) + M^\tau(\mathbb{R})$.

3.2. Jackson-type estimates

In this Section we consider two Jackson-Type estimates. The first estimate is valid for arbitrary quasi-normalized dictionaries. We have the following Jackson embedding.

THEOREM 3 (**[H]**). *For any $\tau < 1$ and $q \in (0, \infty]$, there is a constant $C = C(\tau, q)$ such that for $\mathcal{D}$ an* arbitrary *quasi-normalized dictionary in an* arbitrary *Banach space X,*

$$\mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}), \qquad \textit{with } \alpha = 1/\tau - 1.$$

Theorem 3 is not really satisfactory since substantial sparseness is needed before any useful estimate is obtained. The main problem is that without any assumptions on the dictionary, we just cover too many cases.

It was noted in **[H]** that Theorem 3 cannot be extended to the range $\tau > 1$. In fact, suppose $\mathcal{D} = \{g_k\}$ is a normalized dictionary in an arbitrary Banach space X and assume $g \in X$ is an accumulation point of $\mathcal{D}$. There exists an index sequence $\{k_n\}_{n \ge 0}$ such that $\|g - g_{k_n}\|_X \le 2^{-n}$. Note that $\|g_k\|_X = 1, k \ge 1$, implies $\|g\|_X = 1$. For all $N \ge 1$

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

It follows that $|g|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \le N^{1/\tau - 1}$ for all N. Hence, for all $\tau > 1$, $|g|_{\mathcal{K}_\tau^\tau(\mathcal{D})} = 0$. This clearly implies that we cannot extend Theorem 3 to get a Jackson inequality with any rate of decay for $\tau > 1$.

The above example also gives an idea of how to assess whether a given dictionary is "good" or "bad". The "good" dictionaries are the ones without (too many) aligned elements, while too much alignment (in the extreme case, with an accumulation point) is not beneficial. We also notice that alignment of the dictionary elements works against the dictionary being p-hilbertian for $p > 1$.

There is a much closer connection between the p-hilbertian property and Jackson-type estimates. The following complete characterization holds true.

THEOREM 4 (**[H]**). *Let $\mathcal{D}$ a dictionary in a Banach space X, and $p > 1$. Then properties* (3.2) *and* (3.3) *are equivalent*

$$\forall\, \tau < p,\ \forall\, q,\ \forall\, \alpha < 1/\tau - 1/p \qquad \mathcal{K}_q^\tau(\mathcal{D}) \hookrightarrow \mathcal{A}_q^\alpha(\mathcal{D}), \tag{3.2}$$

$$\forall\, \tau < p \qquad \mathcal{D} \text{ is } \ell_1^\tau - \text{hilbertian}. \tag{3.3}$$

Moreover, for $1 < p < \infty$, $\tau < p$, $0 < q \leq \infty$, there is a constant $C = C(\tau, q, p)$ such that for any *ℓ_1^p-hilbertian dictionary $\mathcal{D}$ in X,*

$$\|f\|_{\mathcal{A}_q^\alpha(\mathcal{D})} \leq C|f|_{\mathcal{K}_q^\tau(\mathcal{D})} \qquad \text{with } \tau = (\alpha + 1/p)^{-1}, \tag{3.4}$$

for all $f \in \mathcal{K}_q^\tau(\mathcal{D})$.

Notice that Theorem 4 extends Theorem 3 to the range $\tau > 1$ for hilbertian dictionaries. In fact, a more general version of Theorem 4 is proved in **[H]** for variations on the approximation space $\mathcal{A}_q^\alpha(\mathcal{D})$, defined in terms of thresholding approximation and Chebyshev approximation. We refer to **[H]** for the details.

3.3. Bernstein-type estimates

We now turn our attention to the problem of obtaining a Bernstein estimate for a redundant dictionary. We will demonstrate that Bernstein estimates are much more fragile than Jackson estimates for redundant dictionaries, and consequently the dictionary need to carry a lot of structure to support a Bernstein estimate.

Bernstein estimates for non-redundant dictionaries have been considered by several authors. The case of an orthonormal basis in a Hilbert space was studied by Stechkin **[94]** and DeVore and Temlyakov **[37]**. Bernstein estimates for L_p-approximation with non-redundant wavelet dictionaries were obtained by DeVore, Jawerth, and Popov in **[33]**, and an extension to approximation in a general Triebel-Lizorkin space was later obtained by Kyriazis **[75]**.

In a few specific cases, a Bernstein estimate has been derived for redundant dictionaries. For univariate rational approximation in L_p, a Bernstein estimate was obtained by Pekarskiĭ **[87]**. Petrushev derived a Bernstein estimate for L_p-approximation with free knot splines **[90]** in the univariate case. Bernstein estimates in L_p for redundant dictionaries based on refinable functions were considered by DeVore, Jawerth, and Popov in **[33]** and by Jia **[70]**.

In Chapter 4 we treat the case of dictionaries with wavelet structure based on a refinable function. For such dictionaries it is also possible to derive Bernstein estimates.

Localized frames and the Bernstein inequality. Let us consider the Bernstein inequality for a class of well structured dictionaries, the so-called localized frames in a Hilbert space. For a general discussion on frames in a Hilbert space, we refer to Christensen **[17]**. Localized frames were introduced by Gröchenig

[**66**], and they are basically frames that resemble an orthonormal basis in a certain sense. In fact, Gröchenig showed that a localized frame can always be written as a finite union of Riesz bases [**65**]. This closeness to orthonormal bases led Gröchenig to speculate that it should be possible to derive a Bernstein estimate for localized frames. However, as Theorem 5 below shows, localized frames are very far from supporting a Bernstein estimate in general.

Let us first introduce some notation. Let $\|\cdot\|_{\mathbb{R}^d}$ be a norm on $\mathbb{R}^d$, and let $K, \mathcal{N} \subset \mathbb{R}^d$ be two separated countable index sets, in the sense that $\inf_{k,\ell\in K, k\neq\ell} \|k-\ell\|_{\mathbb{R}^d} > 0$ and likewise for $\mathcal{N}$. Let $\mathcal{B} = \{e_n\}_{n\in\mathcal{N}}$ be a Riesz basis for $\mathcal{H}$ and $\tilde{\mathcal{B}} = \{\tilde{e}_n\}_{n\in\mathcal{N}}$ its dual basis. A frame $\mathcal{D} = \{g_k\}_{k\in K}$ in $\mathcal{H}$ is polynomially localized w.r.t. $\mathcal{B}$ with decay $s > 0$ if there exists a constant $C < \infty$ such that

$$\max\left(|\langle g_k, e_n\rangle|, |\langle g_k, \tilde{e}_n\rangle|\right) \leq C\left(1 + \|k-n\|_{\mathbb{R}^d}\right)^{-s}.$$

It is exponentially localized w.r.t. $\mathcal{B}$ if for some $\lambda > 0$ and $C < \infty$

$$\max\left(|\langle g_k, e_n\rangle|, |\langle g_k, \tilde{e}_n\rangle|\right) \leq C\exp\left(-\lambda\|k-n\|_{\mathbb{R}^d}\right).$$

An important property of localized frames is the equality *with equivalent norms*

$$\mathcal{K}_q^\tau(\mathcal{B}) = \mathcal{K}_q^\tau(\tilde{\mathcal{B}}) = \mathcal{K}_q^\tau(\mathcal{D}) = \mathcal{K}_q^\tau(\mathcal{D}), \qquad q \in (0,\infty] \tag{3.5}$$

which is valid for $d/s < \tau < 2$ when $\mathcal{D}$ is polynomially localized with decay $s > d$, and for $0 < \tau < 2$ when it is exponentially localized, see [**L**, **66**]. It is thus very easy to estimate the sparseness norm of a given element just by calculating the Fourier coefficients of the element w.r.t. the orthonormal bases.

The main result of [**I**] shows that a Bernstein estimate cannot be obtained for the class of localized frames. We have

THEOREM 5 ([**I**]). *Let $\mathcal{H}$ be an infinite dimensional separable Hilbert space with an orthonormal basis $\mathcal{B}$. Then there exist a frame $\mathcal{F}$ for $\mathcal{H}$ such that $\mathcal{F}$ is exponentially localized w.r.t. $\mathcal{B}$, and a sequence $\{h_k\}_{k\in\mathbb{N}} \subset \Sigma_2(\mathcal{F})$ such that for all $0 < \tau < 2$ and $q \in (0,\infty]$,*

$$\sup_k \frac{|h_k|_{\mathcal{K}_q^\tau(\mathcal{F})}}{\|h_k\|} = \infty.$$

Notice that the Bernstein estimate fails already for two-term expansions, so such localized frames are indeed very far from supporting any type of Bernstein estimate.

An interesting consequence of Theorem 5 is that for $0 < \alpha = 1/\tau - 1/2$, $q \in (0,\infty]$, we have

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

where the approximation space $\mathcal{A}_q^\alpha(\mathcal{F})$ is *strictly larger* than $\mathcal{K}_q^\tau(\mathcal{F})$.

The proof of Theorem 5 is based on a fairly simple idea. The frame $\mathcal{F}$ is defined as a union of $\mathcal{B}$ and a very slight perturbation of $\mathcal{B}$. The nearly co-linear elements in the frame can be used to obtain the wanted estimate. The same basic idea of using nearly co-linear elements to create "bad" dictionaries was also used in [**E**] to construct a non-redundant system in a Hilbert space for which the Bernstein estimate fails.

Incoherent dictionaries in a finite dimensional space. The example in the previous section shows that one cannot expect a Bernstein estimate for a redundant dictionary unless the dictionary has a lot of structure. In this section we discuss a Bernstein estimate for a class of dictionaries in a finite dimensional space. Of course, in finite dimensions, all norms are equivalent so the important aspect of the Bernstein inequality will be that the Bernstein constant only depends on a certain structure constant of the the dictionary. The structure constant measures the coherence between pairs of elements of the dictionary, i.e., it will detect if we have aligned elements in the dictionary.

For a general normalized dictionary $\mathcal{D} = \{g_k\}_{k\in F}$ in a Hilbert space $\mathcal{H}$ (not necessarily finite dimensional) the *coherence* is defined as

$$M(\mathcal{D}) := \sup_{k\neq l} |\langle g_k, g_l\rangle| \,. \tag{3.6}$$

The coherence naturally generalizes the measure of *mutual coherence* $M(\mathcal{B}_1, \mathcal{B}_2)$, see [**42**], defined for unions of two orthonormal bases to general dictionaries. Notice that in general $0 \leq M(\mathcal{D}) \leq 1$ (since we assume the dictionary is normalized). The extreme case are given by an an orthonormal basis $\mathcal{B}$ where $M(\mathcal{D}) = 0$, while any dictionary $\mathcal{D}$ containing two aligned elements has coherence $M(\mathcal{D}) = 1$. For a redundant dictionary $\mathcal{D}$ containing an orthonormal basis, it can easily be shown that $M(\mathcal{D}) \geq 1/\sqrt{N}$. We say that a redundant dictionary $\mathcal{D}$ containing an orthonormal basis is *perfectly incoherent* if $M(\mathcal{D}) = 1/\sqrt{N}$. We have the following Bernstein estimate.

THEOREM 6 ([**J**]). *Let $\mathcal{D}$ be a normalized dictionary in a finite dimensional Hilbert space $\mathcal{H}$ of dimension N, and assume that $\mathcal{D}$ contains an orthonormal basis $\mathcal{B}$. For any $0 < \tau < 2$, the Bernstein inequality for $\mathcal{K}_\tau^\tau(\mathcal{D})$ holds with exponent $\alpha = 2(1/\tau - 1/2)$:*

$$|f_m|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \leq Cm^\alpha \, \|f_m\|_{\mathcal{H}}\,, \qquad m \geq 1,\ f_m \in \Sigma_m(\mathcal{D}),$$

where

$$C = \max\left(\sqrt{2}, \left(2M(\mathcal{D})\right)^{2/\tau-1}\right). \tag{3.7}$$

Moreover, the exponent $\alpha = 2(1/\tau - 1/2)$ is sharp for the class of perfectly incoherent dictionaries.

The sharpness of Theorem 6 means that *among the class of perfectly incoherent dictionaries*, there is a subfamily for which the estimate cannot be improved.

It does not rule out that some other family of *particular* perfectly incoherent dictionaries satisfy an estimate with an improved exponent.

The exponent $\alpha = 2(1/\tau - 1/2)$ obtained in Theorem 6 is surprising since the corresponding exponent for an orthonormal basis is "twice as good", $\alpha = 1/\tau - 1/2$.

Let us take a closer look at a specific dictionary for which the Bernstein estimate cannot be improved. Consider $\mathcal{H} = \mathbb{C}^N$ with $N = P^2$, $P \in \mathbb{N}$. Let $\mathcal{D}_1 := \{\delta_n\}_{n=0}^{N-1}$ be the Dirac (standard) basis for $\mathcal{H}$ and let $\mathcal{D}_2 := \{e_n\}_{n=0}^{N-1}$ be the orthonormal Fourier basis for $\mathcal{H}$. One easily checks that $M(\mathcal{D}_1 \cup \mathcal{D}_2) = 1/\sqrt{N}$. We recall the identity

$$\sum_{k=0}^{P-1} \delta_{k\cdot P} - \sum_{k=0}^{P-1} e_{k\cdot P} = 0, \tag{3.8}$$

which is a consequence of the fact that the "Dirac comb" is invariant under the discrete Fourier transform. We form the dictionary $\mathcal{D} = \mathcal{D}_1 \cup (\mathcal{D}_2 \backslash \{e_0\})$ with $M(\mathcal{D}) = 1/\sqrt{N} = 1/P$. From (3.8) we get

$$e_0 = \sum_{k=0}^{P-1} \delta_{k\cdot P} - \sum_{k=1}^{P-1} e_{k\cdot P},$$

so $e_0 \in \Sigma_{2P-1}(\mathcal{D})$. Now consider an *arbitrary* expansion $e_0 = \sum_{k=0}^{N-1} c_k\delta_k + \sum_{l=1}^{N-1} d_l e_l$ of e_0 in $\mathcal{D}$. By the Hölder inequality we have, with $1 < \tau \le 2$ and $1/\tau + 1/\tau' = 1$,

$$\begin{aligned} 1 &= |\langle e_0, e_0\rangle| \le \sum_k |c_k||\langle \delta_k, e_0\rangle| \le \Big(\sum_k |c_k|^\tau\Big)^{1/\tau} \cdot \Big(\sum_k |\langle \delta_k, e_0\rangle|^{\tau'}\Big)^{1/\tau'} \\ &\le \Big(\sum_k |c_k|^\tau\Big)^{1/\tau} \cdot N^{1/\tau'} \cdot M(\mathcal{D}) = \Big(\sum_k |c_k|^\tau\Big)^{1/\tau} \cdot N^{1-1/\tau} \cdot N^{-1/2} \end{aligned}$$

thus $|e_0|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \ge N^{1/\tau-1/2} = P^{2(1/\tau-1/2)}$. The argument can easily be modified to hold for $\tau = 1$. It follows that we have found $e_0 \in \Sigma_{2P-1}(\mathcal{D})$ which satisfies

$$|e_0|_{\mathcal{K}_\tau^\tau(\mathcal{D})} \ge 2^{-\alpha}(2P-1)^\alpha \|e_0\|_{\mathcal{H}},$$

with $\alpha = 2(1/\tau - 1/2)$.

The reader can consult **[J]** for additional Bernstein estimates for dictionaries in an infinite dimensional Hilbert space built from incoherent finite dimensional dictionaries.

3.4. Greedy bases

We conclude this chapter by studying a special class of non-redundant dictionaries in a Banach space X. The dictionaries form so-called greedy bases for X, a notion that turn out to be fruitful in order to generalize known result on m-term approximation with orthonormal bases in a Hilbert space. Greedy bases and several related concepts were introduced in the abstract setting by Temlyakov and Konyagin [**73**]. Greedy basis are closely related to properties of wavelet bases in L_p, see e.g. [**97**]. Let us introduce some necessary notation.

DEFINITION 2. *Let $\mathcal{B} = \{g_k\}_{k\in\mathbb{N}}$ be a quasi-normed Schauder basis for the Banach space X with associated coefficient functionals $\{c_k(\cdot)\}_{k\in\mathbb{N}}$. For any $f \in X$ and $m \geq 1$, a* greedy m-term approximant *to f from $\mathcal{B}$ is any vector $G_m(f,\mathcal{B},\pi) := \sum_{k=1}^m c_k^\star g_{\pi(k)}$, where $\{c_k^\star\} = \{c_{\pi(k)}\}$ is a decreasing rearrangement of $\{c_k(f)\}$. The error associated to greedy m-term approximation to f form $\mathcal{B}$ is denoted by*

$$\gamma_m(f,\mathcal{B},\pi)_X := \|f - G_m(f,\mathcal{B},\pi)\|_X. \tag{3.9}$$

In the following, we suppress the permutation π and simply write $G_m(f,\mathcal{B})$ and $\gamma_m(f,\mathcal{B})_X$. Any statement on these quantities will be assumed to hold for all π such that $\{c_{\pi(k)}\}$ is a decreasing rearrangement of $\{c_k(f)\}$.

Greedy bases and the related notion of a quasi-greedy basis are defined as follows.

DEFINITION 3. *Let $\mathcal{B} = \{g_k\}_{k\in\mathbb{N}}$ be a quasi-normed Schauder basis for the Banach space X. We call $\mathcal{B}$ a* quasi-greedy *basis if for each $f \in X$ we have $\gamma_m(f,\mathcal{B})_X \to 0$ as $m \to \infty$. We call $\mathcal{B}$ a* greedy *basis if there exists a constant $C < \infty$ such that for each $f \in X$, we have for all m,*

$$\gamma_m(f,\mathcal{B}) \leq C\sigma_m(f)_X \tag{3.10}$$

Greedy bases and quasi-greedy bases were first introduced by Konyagin and Temlyakov [**73**]. The equivalent definition of a quasi-greedy basis given above is due to Wojtaszczyk [**104**].

We notice that quasi-greedy bases are exactly the bases for which decreasing rearrangements converge in norm. Greedy bases are even better; the partial sums corresponding to decreasing rearrangement give near-best m-term approximants.

A very nice characterization of greedy bases is given in [**73**]. A Schauder basis for X is a greedy basis if and only if it is *unconditional* and *democratic*. By democratic we mean the following:

DEFINITION 4. *A Schauder basis $\mathcal{B} = \{g_k\}_{k\in\mathbb{N}}$ is democratic if there exists a constant $C < \infty$ such that for every two finite sets $\Lambda, \Lambda' \subset \mathbb{N}$ of same cardinality $|\Lambda| = |\Lambda'|$ we have*

$$\Big\| \sum_{k\in\Lambda} g_k \Big\| \leq C \Big\| \sum_{k\in\Lambda'} g_k \Big\|.$$

Clearly, democracy implies that the basis is quasi-normed by taking $|\Lambda| = 1$ in the definition.

The most important example of a greedy basis is the L_p-normalized Haar system in $L_p((0,1)^d)$, $1 < p < \infty$, see [**95, 97**]. Greedy bases are clearly quasi-greedy but the two notions are not equivalent. An example of a conditional quasi-greedy basis is given in [**73**]. It turns out that most classical bounded systems in $L_p(0,1)$ fail to be quasi-greedy. It was proved by Temlyakov [**96**] that the trigonometric system in $L_p(0,1)$, $1 \le p \le \infty$, $p \ne 2$, fails to be quasi-greedy. Independently, and using a different approach, Córdoba and Fernández [**24**] proved the same result in the range $1 \le p < 2$. This negative result for the trigonometric system was extended to weighted spaces $L_p((0,1);w)$ by the author in [**84**]. The first example of a bounded conditional quasi-greedy basis in $L_p(0,1)$, $1 < p < \infty$, is given in [**83**].

Characterization of the approximation spaces. It turns out that the sparseness class $\mathcal{K}_s^\tau(\mathcal{B})$ defined in Section 3.1 is not quite flexible enough to obtain a characterization of the approximation classes $\mathcal{A}_s^\tau(\mathcal{B})$ for a greedy basis $\mathcal{B}$. We therefore introduce the notion of a generalized sparseness spaces $\mathcal{K}_s^\tau(w,\mathcal{B})$ by mimicking the definition in Section 3.1. This time we use weighted Lorentz norms, with weights $w = \{w_m\}$ that form a slowly increasing sequence, i.e., $w_{2m} \le C w_m$ for all m. We define

$$\|\{a_m\}_{m=1}^\infty\|_{\ell_s^\tau(w)} := \begin{cases} \left(\sum_{m=1}^\infty \frac{[w_m m^{1/\tau}|a_m^\star|]^s}{m} \right)^{1/s}, & 0 < s < \infty \\ \sup_{m\in\mathbb{N}} w_m m^{1/\tau}|a_m^\star|, & s = \infty. \end{cases} \tag{3.11}$$

Notice that for weights $w_m = m^{1/p}$, the weighted Lorentz spaces reduce to standard ones $\ell_s^{1/\alpha}(\{m^{1/p}\}) = \ell_s^{\tau_p}$, where $1/\tau_p = \alpha + 1/p$.

For any basis $\mathcal{B} = \{g_k\}_{k\ge1}$, we define a sequence $w(\mathcal{B}) = \{w_n\}_{n\ge0}$ with $w_0 = 0$ and for any $n \ge 1$:

$$w_n = \max\Big(w_{n-1}, \big\|\sum_{k=1}^n g_k\big\|\Big).$$

It is proved in [**E**] that for a quasi-greedy and democratic basis $\mathcal{B}$, $w(\mathcal{B})$ is a slowly increasing sequence. We now introduce the a variation on the approximation spaces $\mathcal{A}_s^\alpha(\mathcal{B})$, the so-called *greedy approximation spaces*

$$\mathcal{G}_s^\alpha(\mathcal{B}) := \Big\{ f \in X, \|f\|_{\mathcal{G}_s^\alpha(\mathcal{B})} := \|f\|_X + \big\|\{\gamma_m(f,\mathcal{B})_X\}_{m\ge1}\big\|_{\ell_s^{1/\alpha}} < \infty \Big\}.$$

For $\mathcal{B}$ a greedy basis, $\mathcal{G}_s^\alpha(\mathcal{B}) = \mathcal{A}_s^\alpha(\mathcal{B})$ with equivalent norms

$$\|\cdot\|_{\mathcal{G}_s^\alpha(\mathcal{B})} \asymp \|\cdot\|_{\mathcal{A}_s^\alpha(\mathcal{B})},$$

since $\sigma_m(f,\mathcal{B}) \asymp \gamma_m(f,\mathcal{B})$. We also mention that for a general conditional quasi-greedy bases $\mathcal{B}$, it is not known whether $\mathcal{G}_s^\alpha(\mathcal{B}) = \mathcal{A}_s^\alpha(\mathcal{B})$.

The following result gives a complete characterization of the greedy approximation spaces for a quasi-greedy democratic dictionary.

THEOREM 7 (**[E]**). *Let $\mathcal{B}$ be a quasi-greedy basis for a Banach space X. The following conditions are equivalent :*

(1) *$\mathcal{B}$ is democratic.*
(2) *For any $\alpha > 0$ and $s \in (0, \infty]$,*

$$\mathcal{G}_s^\alpha(\mathcal{B}) = \left\{ f \in X, \left\| \{c_k(f)\} \right\|_{\ell_s^{1/\alpha}(w(\mathcal{B}))} < \infty \right\}, \tag{3.12}$$

with equivalent norms

$$\| \cdot \|_{\mathcal{G}_s^\alpha(\mathcal{B})} \asymp \|\{c_k(\cdot)\}\|_{\ell_s^{1/\alpha}(w(\mathcal{B}))}. \tag{3.13}$$

(3) *Relations (3.12) and (3.13) hold for some slowly growing sequence $w = \{w_m\}$ at* some *point α, s.*

Theorem 7 was proved for orthonormal bases in a Hilbert space by Stechkin **[94]** and DeVore and Temlyakov **[37]**. Variations on Theorem 7 have been proved (independently) by Kerkyacharian and Picard **[72]** and Dilworth et al. **[39]**. Garrigós and Hernández **[59]** have recently extended Theorem 7 to quasi-Banach spaces.

An interesting characterization of quasi-greedy democratic bases has been found by Dilworth et al. **[39]**. They prove that such bases are exactly the so-called *almost greedy bases*, see **[39]** for details. For conditional almost greedy bases, we have a characterization of $\mathcal{G}_s^\alpha(\mathcal{B})$ but not of $\mathcal{A}_s^\alpha(\mathcal{B})$. However, for a greedy basis $\mathcal{B}$, we deduce from Theorem 7 that

$$\mathcal{A}_s^\alpha(\mathcal{B}) = \mathcal{G}_s^\alpha(\mathcal{B}) = \mathcal{K}_s^{1/\alpha}(w(\mathcal{B}), \mathcal{B}). \tag{3.14}$$

CHAPTER 4

Approximation with wavelet-type systems

This chapter concerns m-term approximation in L_p with wavelet frame dictionaries. Wavelet frames are defined as follows. Given a finite collection of functions (wavelet generators) $\Psi = \{\psi^1, \psi^2, \ldots, \psi^L\} \subset L_2(\mathbb{R}^d)$ we use the notation $X(\Psi)$ to denote the corresponding "wavelet" system,

$$X(\Psi) := \{2^{jd/2}\psi^\ell(2^j \cdot -k) \mid j \in \mathbb{Z}, k \in \mathbb{Z}^d, \ell = 1, 2, \ldots, L\}.$$

A *wavelet bi-frame for* $L_2(\mathbb{R}^d)$ consists of two sequences of wavelets

$$\Psi = \{\psi^1, \psi^2, \ldots, \psi^L\} \subset L_2(\mathbb{R}^d) \quad \text{and} \quad \tilde{\Psi} = \{\tilde{\psi}^1, \tilde{\psi}^2, \ldots, \tilde{\psi}^L\} \subset L_2(\mathbb{R}^d)$$

for which the systems $X(\Psi)$ and $X(\tilde{\Psi})$ are Bessel systems, and satisfy the perfect reconstruction formula

$$f = \sum_{\ell=1}^{L} \sum_{j\in\mathbb{Z}, k\in\mathbb{Z}^d} \langle f, \tilde{\psi}^\ell_{j,k}\rangle \psi^\ell_{j,k}, \qquad \forall f \in L_2(\mathbb{R}^d), \tag{4.1}$$

where

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

This definition implies that both $X(\Psi)$ and $X(\tilde{\Psi})$ are frames for $L_2(\mathbb{R}^d)$ and in fact the roles of Ψ and $\tilde{\Psi}$ are interchangeable in (4.1). The special case with $\Psi = \tilde{\Psi}$ corresponds to a so-called tight wavelet frame. The first systematic study of wavelet frames (in the non-continuous case) can be found in **[26]**, but wavelet frames also connect back to Grossmann and Morlet's seminal work **[68]** on the continuous wavelet transform.

Wavelet frames have the same structure as an orthonormal wavelet basis, but in general they form redundant dictionaries in L_p. In the important special case where the wavelet frame is based on a multiresolution analysis (MRA), the expansion (4.1) is very attractive from a computational point of view since for such frames, the standard discrete wavelet algorithm can be used for numerical calculations. Wavelet frames based on an MRA are referred to as *framelets*.

Below we study approximation in L_p with $X(\Psi)$. We consider three cases depending on the properties of Ψ. In the first case, we assume that the wavelet generators Ψ are all nice in the sense that they have smoothness, a number of vanishing moments, and some decay at infinity. For compactly supported systems based on a multiresolution analysis, a complete characterization of m-term approximation in L_p is obtained for the dictionary $X(\Psi)$. The second case is where at least one of the functions in Ψ have few vanishing moments compared to its smoothness. For such systems we show that it is possible to oversample the dictionary $X(\Psi)$ and obtain a complete characterization of the corresponding approximation space. In the third case, we study approximation with completely general systems $X(\psi)$ based on an MRA. For such general systems we show that it is also possible to oversample the dictionary $X(\psi)$ and obtain a complete characterization of the corresponding approximation space.

Approximation with non-redundant wavelet systems have been considered in various setups by a number of authors, see **[19, 20, 22, 30, 32–34, 75, 97]**. For redundant system with wavelet structure, Jackson type estimates have been obtained in **[77, 78, 88]**.

Wavelet frames and extension principles. Let us briefly review the standard approach to constructing wavelet frames (framelets) based on a multiresolution analysis using so-called extension principles. The extension principles for constructing bi-frames were introduced independently in **[18]** and **[27]**.

Let $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_L)$ and $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_0, \tilde{\tau}_1, \ldots, \tilde{\tau}_L)$ be two sequences of $2\pi\mathbb{Z}^d$-periodic essentially bounded functions. Assume that τ_0 and $\tilde{\tau}_0$ both generate refinable functions[1]

$$\hat{\varphi}(2\xi) = \tau_0(\xi)\hat{\varphi}(\xi) \qquad \text{and} \qquad \hat{\tilde{\varphi}}(2\xi) = \tilde{\tau}_0(\xi)\hat{\tilde{\varphi}}(\xi),$$

satisfying

$$\lim_{\xi\to 0} \hat{\varphi}(\xi) = 1 \quad \text{and} \quad \lim_{\xi\to 0} \hat{\tilde{\varphi}}(\xi) = 1,$$

with

$$\operatorname{ess\,sup}_{\xi} \sum_{k\in\mathbb{Z}^d} |\hat{\varphi}(\xi-k)|^2 < \infty \quad \text{and} \quad \operatorname{ess\,sup}_{\xi} \sum_{k\in\mathbb{Z}^d} |\hat{\tilde{\varphi}}(\xi-k)|^2 < \infty,$$

where $\hat{\varphi}(\xi)$ is the Fourier transform of the function $\varphi(x)$. We associate the wavelets with $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{\tau}}$ as follows

$$\hat{\psi}^{\ell}(2\xi) = \tau_\ell(\xi)\hat{\varphi}(\xi), \qquad \hat{\tilde{\psi}}^{\ell}(2\xi) = \tilde{\tau}_\ell(\xi)\hat{\tilde{\varphi}}(\xi). \tag{4.2}$$

The spectrum $\sigma(\varphi)$ associated with φ is defined up to a null-set as

$$\sigma(\varphi) := \{\omega \in [-\pi, \pi]^d : \hat{\varphi}(\omega + 2\pi k) \neq 0, \text{ for some } k \in \mathbb{Z}^d\}.$$

[1]We define the Fourier transform by $\mathcal{F}(f)(\xi) = \hat{f}(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(x)e^{-ix\cdot\xi}dx, \quad f \in L_1 \cap L_2$.

The spectrum $\sigma(\tilde{\varphi})$ associated with $\tilde{\varphi}$ is defined likewise. Assuming that the systems $X(\Psi)$ and $X(\tilde{\Psi})$ are both Bessel systems, we define the *mixed fundamental function of the parent vectors* $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{\tau}}$ by

$$\Theta(\xi) := \sum_{j=0}^{\infty} \sum_{\ell=1}^{L} \tau_\ell(2^j\xi)\overline{\tilde{\tau}_\ell(2^j\xi)} \prod_{m=0}^{j-1} \tau_0(2^m\xi)\overline{\tilde{\tau}_0(2^m\xi)}.$$

The following theorem proven in **[27]** is the main tool for creating bi-framelet systems, the theorem is called the *Mixed Oblique Extension Principle*.

THEOREM 8 (**[27]**; Mixed OEP). *Let $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{\tau}}$ be the combined mask of the systems $X(\Psi)$ and $X(\tilde{\Psi})$, respectively. Assume that the systems $X(\Psi)$ and $X(\tilde{\Psi})$ are Bessel systems. Suppose there exists a 2π-periodic function Θ satisfying*

a) *Θ is essentially bounded, continuous at the origin, and $\Theta(0) = 1$.*
b) *If $\xi \in \sigma(\varphi) \cap \sigma(\tilde{\varphi})$ and $\nu \in \{0, \pi\}^d$ such that $\xi + \nu \in \sigma(\varphi) \cap \sigma(\tilde{\varphi})$, then*

$$\Theta(2\xi)\tau_0(\xi)\overline{\tilde{\tau}(\xi+\nu)} + \sum_{\ell=1}^{L} \tau_\ell(\xi)\overline{\tilde{\tau}_\ell(\xi+\nu)} = \begin{cases} \Theta(\xi), & \text{if } \nu = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Then $X(\Psi)$, $X(\tilde{\Psi})$ is a bi-framelet system.

In many interesting cases the spectra $\sigma(\varphi)$ and $\sigma(\tilde{\varphi})$ are both equal to $[-\pi, \pi]^d$. For example, if the integer translates of the scaling functions φ and $\tilde{\varphi}$ are Riesz sequences, this is the case.

When $X(\Psi) = X(\tilde{\Psi})$, Theorem 8 gives the so-called Oblique Extension principle, see **[27]**. If, in addition, $\Theta \equiv 1$, Theorem 8 reduces to the Unitary Extension Principle, see **[91, 92]**.

The reader can consult **[18]** and **[27]** for many explicit examples on how to construct framelet systems using the different extension principles.

4.1. Wavelet frames with smoothness and vanishing moments

In this section we obtain approximation results in L_p for dictionaries $X(\Psi)$ under the assumption that the generators Ψ have a certain decay at infinity and a number of vanishing moments.

General Jackson estimates. Let us introduce the following two function classes.

DEFINITION 5. *For $N \in \mathbb{N}$ and $\gamma > 0$ we let $D_\gamma^N(\mathbb{R}^d)$ be the set of all functions f defined on $\mathbb{R}^d$ with N derivatives and decay γ,* i.e., *for which there exists a constant $c < \infty$ such that*

$$|\partial^\alpha f(x)| \le c(1+|x|)^{-\gamma} \qquad \text{for } x \in \mathbb{R}^d,\ \alpha \in \mathbb{N}^d,\ |\alpha| \le N, \tag{4.4}$$

where $|\alpha|$ is the usual length of a multi-index. Likewise, we let $M_\gamma^N(\mathbb{R}^d)$ denote the set of all functions f with N vanishing moments and decay, i.e., *for which*

$$\int_{\mathbb{R}^d} x^\alpha f(x)dx = 0 \qquad \text{for } \alpha \in \mathbb{N}^d,\ |\alpha| < N,$$

and

$$|f(x)| \le C(1+|x|)^{-d-N-\gamma} \qquad \text{for } x \in \mathbb{R}^d. \tag{4.5}$$

For notational convenience, let us define the function

$$\Lambda(x) = \Lambda(x,p,\gamma/d) := \begin{cases} p(1-x) & \text{for } x \le 1-1/p, \\ (x+1/p)^{-1} & \text{for } 1-1/p < x \le \gamma/d - 1/p, \\ d/\gamma & \text{for } \gamma/d - 1/p < x. \end{cases} \tag{4.6}$$

We have the general Jackson estimate. We refer to Triebel **[99]** for the definition of the Besov space $B^s_{p,q}(\mathbb{R}^d)$.

THEOREM 9 (**[A]**). *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a bi-frame. Suppose $\tilde{\Psi} \subset M_\gamma^{N_1}(\mathbb{R}^d)$ for some $N_1 \in \mathbb{N}$ and $\gamma > d$ and suppose there exist $\beta, \varepsilon > 0$ such that for all $\psi \in \Psi \cup \tilde{\Psi}$, $\psi \in C^\beta(\mathbb{R}^d)$ and $|\psi(x)| \le C(1+|x|)^{-d-\varepsilon}$. Then, we have the Jackson inequality*

$$\sigma_m(f, X(\Psi))_p \le Cm^{-\alpha}\|f\|_{B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d)}$$

for $p \in (1,\infty)$, $\Lambda\left(\frac{N_1}{d}\right) < \tau < p$, and $\alpha = 1/\tau - 1/p$, with $\Lambda(x) = \Lambda(x,p,d/\gamma)$ given by (4.6).

We refer to **[A]** for the detailed proof, but let us discuss the basic idea of the proof since it follows a fairly general and successful approach to obtaining Jackson estimates for redundant wavelet-type systems. The idea is to consider the following "change of basis" operator for $p \ge 1$,

$$\left(\mathbf{T}(c_{j,k})\right)_{j',k'} := \sum_{j\in\mathbb{Z},k\in\mathbb{Z}^d} c_{j,k}\langle \eta^p_{j,k}, \psi^{p'}_{j',k'}\rangle, \qquad \frac{1}{p}+\frac{1}{p'} = 1,$$

where $\{\eta_{j,k}\}_{j,k}$ is a nice smooth orthonormal wavelet such as a Meyer wavelet (if $d > 1$, we just pick one of the $2^d - 1$ orthonormal wavelet generators), and $\psi^p_{j,k} := 2^{jd(1/p-1/2)}\psi_{j,k}$ is $\psi_{j,k}$ "normalized" in L_p. The smoothness and vanishing moments of Ψ ensure that the matrix $[\langle \eta^p_{j,k}, \psi^{p'}_{j',k'}\rangle]$ is *almost diagonal* and induces a bounded operator on $\ell_\tau(\mathbb{Z}\times\mathbb{Z}^d)$ for τ in a suitable range. To prove the result we now expand $f \in B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d)$ in the Meyer wavelet system. This expansion is sparse according to classical results. Next we "change basis" to the wavelet frame, and use the estimates on $[\langle \eta^p_{j,k}, \psi^{p'}_{j',k'}\rangle]$ to conclude that the wavelet frame expansion of f is also sparse. The general Jackson estimate of Theorem 4 can then be used to conclude.

Theorem 9 can also be used to obtain the following characterization of the sparseness class $\mathcal{K}_\tau^\tau(L_p, X(\Psi))$.

$$B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d) = \left\{ f \in L_p, \{\langle f, \psi^{\ell,p'}_{j,k}\rangle\}_{j,k,\ell} \in \ell_\tau \right\} = \mathcal{K}_\tau^\tau(L_p, X(\Psi)), \tag{4.7}$$

for $\alpha = 1/\tau - 1/p$ and admissible τ (see the statement of Theorem 9).

The Jackson estimate provided by Theorem 9 is obtained by using the canonical frame expansion. The smoothness and vanishing moments of Ψ ensure a sparse wavelet frame expansion for a nice function f. Hence, we do not really use the redundancy in any way to obtain the estimate.

The "change of basis" approach to obtaining Jackson estimates for redundant wavelet-type systems has also been used by Petrushev [**88**], Kyriazis and Petrushev [**77, 78**], and Kyriazis [**76**] to obtain approximation results in Triebel-Lizorkin spaces. Approximation with tight wavelet frames in Sobolev spaces is considered in [**7**]. The general idea of using sparse wavelet-type frame expansions of smooth functions goes back to Frazier and Jawerth and their comprehensive study of the so-called φ-transform [**56, 57**].

Bernstein estimates and the approximation classes. We now turn our attention to Bernstein estimates that will lead to a complete characterization of the approximation spaces for $X(\Psi)$. In the following we denote by $W^s(L_\infty(\mathbb{R}^d))$ the Sobolev space consisting of functions with all s distributional derivatives in $L_\infty(\mathbb{R}^d)$. Given a function $\varphi \in L_\infty(\mathbb{R}^d)$, let

$$\Gamma = \{k \in \mathbb{Z}^d : |\{x \in (0,1)^d : \varphi(x-k) \neq 0\}| > 0\}.$$

We say that $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}^d}$ is a *locally linearly independent set* if the set $\{\varphi(\cdot - k)\}_{k\in\Gamma}$ is linearly independent. Also recall that a function $\varphi \in L_2(\mathbb{R}^d)$ is refinable if there exists suitable coefficients $\{c_k\}$ such that

$$\varphi(x) = \sum_{k\in\mathbb{Z}^d} c_k \varphi(2x-k), \qquad x \in \mathbb{R}^d.$$

For bi-framelet systems we have the following Bernstein inequality, see [**A**].

LEMMA 1. *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a bi-framelet system and assume that $X(\Psi)$ is based on a compactly supported refinable function φ where:*

(1) *$\varphi \in W^s(L_\infty(\mathbb{R}^d))$ with $s \geq 0$;*
(2) *(In case $d > 1$) $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}^d}$ is a locally linearly independent set;*
(3) *The functions $\tau_\ell(\xi)$, $1 \leq \ell \leq L$, in (4.2) are trigonometric polynomials.*

Then the Bernstein inequality

$$|S|_{B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d)} \leq C m^\alpha \|S\|_{L_p(\mathbb{R}^d)}, \qquad \forall S \in \Sigma_m(X(\Psi)), \qquad \forall m \geq 1 \tag{4.8}$$

holds true for each $0 < \alpha < s/d$, $0 < p \leq \infty$, with $1/\tau := \alpha + 1/p$ and $C = C(\alpha, p)$.

Let us again outline the idea of the proof. The fundamental assumption needed for the estimate is that φ is refinable, i.e., that the wavelet frame is based on a multiresolution analysis. By a result of Jia **[70]**, for each $0 < \alpha < s/d$, the Bernstein inequality

$$|S|_{B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d)} \leq Cm^{\alpha}\|S\|_{L_p(\mathbb{R}^d)}, \qquad \forall S \in \Sigma_m(X(\varphi)),$$

$1/\tau := \alpha + 1/p$, $0 < p \leq \infty$, holds true for the system

$$X(\varphi) := \{\varphi(2^j x - k)\}_{j\in\mathbb{Z},k\in\mathbb{Z}^d},$$

provided that $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}^d}$ is a locally linearly independent set with φ compactly supported. Now, since $X(\Psi)$ is based on φ, we have finite masks $\{b^\ell_k\}_k$ such that $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}^d}$ is a locally linearly independent set with φ compactly supported.

$$\psi^\ell(x) = \sum_{k\in\mathbb{Z}^d} b^\ell_k \varphi(2x - k).$$

Thus, for $j \in \mathbb{Z}$ and $i \in \mathbb{Z}^d$, we have

$$\psi^\ell(2^j x - i) = \sum_{k\in\mathbb{Z}^d} b^\ell_k \varphi\big(2^{j+1}x - 2i - k\big)$$

That is to say $\psi^\ell_{j,i} \in \Sigma_K(X(\varphi))$ for some uniform constant K depending only on the length of the finite masks used above. Take any $S \in \Sigma_m(X(\Psi))$, then $S \in \Sigma_{Km}(X(\varphi))$ and the estimate follows.

We can now combine the Jackson and Bernstein estimates and use Theorem 1 to obtain the main result of this section. A version of the result valid for spline based wavelet frames is proved in **[I]**.

THEOREM 10 (**[A]**). *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a wavelet bi-frame system and assume that $X(\Psi)$ is based on a compactly supported refinable function φ where:*

(1) *$\varphi \in W^s(L_\infty(\mathbb{R}^d))$ with $s \geq 0$;*
(2) *(In case $d > 1$) $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}^d}$ is a locally linearly independent set;*
(3) *The functions $\tau_\ell(\xi)$ in (4.2) are trigonometric polynomials;*
(4) *$\tilde{\Psi} \subset C^\beta(\mathbb{R}^d) \cap M^{N_1}_\gamma(\mathbb{R}^d)$ for some $\beta > 0$, $N_1 \in \mathbb{N}$ and $\gamma > d$.*

Let $p \in (1, \infty)$ and $\tau := (\alpha + 1/p)^{-1}$ where we assume

$$0 < \alpha < \min\left\{\frac{s}{d}, \frac{1}{\Lambda\left(\frac{N_1}{d}\right)} - \frac{1}{p}\right\}, \tag{4.9}$$

with $\Lambda(x) = \Lambda(x, p, d/\gamma)$ given by (4.6). Then, for each $0 < \beta < \alpha$, $q \in (0, \infty]$, we have the characterization

$$\mathcal{A}^\beta_q(L_p, X(\Psi)) = \left(L_p, B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d)\right)_{\beta/\alpha,q}. \tag{4.10}$$

The approximation spaces are thus essentially Besov spaces, similar to situation for non-redundant bi-orthogonal wavelets, see [33]. Perhaps it is not surprising that a nice smooth wavelet frame behaves the same as an orthonormal wavelet basis.

An interesting derived question is what happens when the framelets have few vanishing moments compared to the smoothness of the underlying scaling function. This case will be treated below.

4.2. Wavelet frames with few vanishing moments

For wavelet frames with few vanishing moments compared to the smoothness of the underlying multiresolution analysis scaling function, the approach to consider a "change of basis" matrix from a nice orthonormal wavelet system to the wavelet frame is no longer very useful. The matrix will have a relatively slow off diagonal decay, leading to a restricted Jackson estimate. However, it was first discovered in [**I**] that oversampling the wavelet frame dictionary is a feasible way to overcome the lack of vanishing moments.

Given a wavelet bi-frame $X(\Psi)$, $X(\tilde{\Psi})$ and $R \geq 1$, we let $X_R(\Psi)$ denote the oversampled system,

$$X_R(\Psi) := \{2^{jd/2}\psi^\ell(2^j \cdot -k/R) | j \in \mathbb{Z}, k \in \mathbb{Z}^d, \ell = 1, 2, \ldots, L\}.$$

Just as the non-oversampled system, the oversampled one $X_R(\Psi)$ is a frame in $L_2(\mathbb{R}^d)$. It can also be verified that the oversampled frame $X_R(\Psi)$ is ℓ_1^p-hilbertian in $L_p(\mathbb{R}^d)$, $1 < p < \infty$, after proper normalization, see [**I**] for a proof in the case $R = 2$.

Let us first state a very general case where we can derive a Jackson estimate for $X_R(\Psi)$. The following lemma was proved in [**A**].

LEMMA 2. *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a wavelet bi-frame system, $X(\{\psi^i\}_{i=1}^{2^d-1})$ a bi-orthogonal wavelet basis and $r > 0$ such that*

$$B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d) = \mathcal{K}^\tau_\tau(L_p, X(\{\psi^i\}_{i=1}^{2^d-1})),\ 0 < \alpha = 1/\tau - 1/p < r.$$

Assume that for $1 \leq i \leq 2^d - 1$ there exists sequences $\{d_k^{\ell,i}\}_{\ell \in E, k \in \mathbb{Z}^d} \in \ell_{1/(r+1)}$, such that

$$\psi^i(x) = \sum_{\ell=1}^{L} \sum_{k \in \mathbb{Z}^d} d_k^{\ell,i} \psi^\ell(x - k/R).$$

Then, for $1 < p < \infty$, and $0 < \alpha = 1/\tau - 1/p < r$, we have the Jackson inequality

$$\sigma_m(f, X_R(\Psi))_p \leq C m^{-\alpha} \|f\|_{B^{d\alpha}_{\tau,\tau}(\mathbb{R}^d)}.$$

We deal with the problem of how to obtain a sparse expansion

$$\psi^i(x) = \sum_{\ell=1}^{L} \sum_{k \in \mathbb{Z}^d} d_k^{\ell,i} \psi^\ell(x - k/R)$$

below.

We can now state our main result on approximation with the oversampled system $X_R(\Psi)$, where we obtain a complete characterization of the approximation spaces $\mathcal{A}_q^\beta(L_p, X_{2^N}(\Psi))$ even if the wavelet frame fails to have a significant number of vanishing moments.

THEOREM 11 (**[A]**). *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a bi-frame system with $X(\Psi)$ based on a compactly supported refinable function φ where:*

(1) *$\varphi \in W^s(L_\infty(\mathbb{R}^d))$ with $s \geq 0$;*
(2) *(In case $d > 1$) $\{\varphi(\cdot - k)\}_{k \in \mathbb{Z}^d}$ is a locally linearly independent set;*
(3) *The functions $\tau_\ell(\xi)$, $1 \leq \ell \leq L$, in* (4.2) *are trigonometric polynomials.*
(4) *$X(\Psi)$, $X(\tilde{\Psi})$ satisfy the assumptions of Lemma 2 for $R = 2^N$ with parameters s and r.*

Then, for $0 < \alpha < \min\{s/d, r\}$, $0 < \beta < \alpha$, $q \in (0, \infty]$, we have the characterization

$$\mathcal{A}_q^\beta(L_p, X_{2^N}(\Psi)) = \Big(L_p, B_{\tau,\tau}^{d\alpha}(\mathbb{R}^d)\Big)_{\beta/\alpha, q}. \tag{4.11}$$

Expanding orthonormal wavelets in a wavelet frame. One obvious criticism of Lemma 2 (and Theorem 11) is that the hypothesis is hard to check in the general case. However, it was noticed in **[I]** that most univariate MRA-based wavelet frames satisfy the hypothesis for $R = 1$. Let us explain why that is. Suppose $X(\Psi)$, $X(\tilde{\Psi})$ is an MRA-based univariate wavelet bi-frame system with combined mask $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_L)$ and $\tilde{\boldsymbol{\tau}}$. Let φ be a univariate scaling function generated by the refinement filter $\tau_0(\xi)$, and let $P(\xi) := \sum_{k \in \mathbb{Z}} |\hat{\varphi}(\xi - k)|^2$. The standard orthonormal wavelet ψ associated with the scaling function φ is defined by

$$\hat{\psi}(2\xi) = e^{-i\xi} \overline{\tau_0(\xi + \pi)} \frac{\hat{\varphi}(\xi)}{\sqrt{P(\xi)}}.$$

We wish to express ψ as a linear combination

$$\psi(\cdot) := \sum_{\ell=1}^{L} \sum_{k \in \mathbb{Z}} d_k^\ell \psi^\ell(\cdot - k/2), \qquad \Psi = \{\psi^1, \psi^2, \ldots, \psi^L\}, \tag{4.12}$$

where $\{d_k^\ell\} \in \bigcap_{\tau > 0} \ell_\tau$. In the frequency domain the problem is to find "nice" 2π-periodic functions $Q_\ell(\xi)$ such that

$$\hat{\psi}(\xi) = \sum_{\ell=1}^{L} Q_\ell(\xi/2) \tau_\ell(\xi/2) \hat{\varphi}(\xi/2).$$

We will look for Q_ℓ of the form $Q_\ell(\xi) = Q(\xi)\overline{\tau_\ell(\xi)}$. It is not difficult to see that the problem will be solved if Q_ℓ has fast decaying Fourier coefficients and Q satisfies

$$Q(\xi)\sum_{\ell=1}^{L}|\tau_\ell(\xi)|^2 = \frac{e^{-i\xi}\overline{\tau_0(\xi+\pi)}}{\sqrt{P(\xi)}}.$$

Hence, we define for $\xi \neq 0$

$$Q_\ell(\xi) := \frac{\tau_\ell(\xi)\cdot\overline{\tau_0(\xi+\pi)}}{\sum_{\ell=1}^{L}|\tau_\ell(\xi)|^2}\cdot\frac{e^{-i\xi}}{\sqrt{P(\xi)}}. \tag{4.13}$$

In the case where $\sum_{\ell=1}^{L}|\tau_\ell(\xi)|^2$ has no zeros, one verifies that Q_ℓ can be extended at $\xi = 0$. The factor

$$\frac{\tau_\ell(\xi)\cdot\overline{\tau_0(\xi+\pi)}}{\sum_{\ell=1}^{L}|\tau_\ell(\xi)|^2}$$

in (4.13) is a quotient of two trigonometric polynomials with no pole on the unit circle, so its Fourier coefficients decay exponentially. Thus, whenever $P^{-1/2}$ is C^∞, the sparse expansion (4.12) holds true.

An interesting follow-up question is whether is is possible to obtain the expansion (4.12) without any oversampling. This question has been studied in detail in **[I]**, and the answer is positive in certain cases. It is possible to obtain (4.12) without oversampling for certain wavelet frames, while for other systems it can be proved that oversampling by a factor two is indeed needed, see **[I]**.

Fast algorithm for near sparsest framelet expansion: univariate case. In a redundant dictionary, finding an expansion $f = \sum_k c_k g_k$ that (approximately) minimizes the ℓ_1 norm $\|\{c_k\}\|_{\ell_1}$ is a computationally intensive problem. For a twice oversampled wavelet frame dictionary, an $\mathcal{O}(N)$ algorithm based on an expansion similar to (4.12) was introduced in **[I]**.

The Euclidean algorithm can be used to solve for the Bezout relation that yields the expansion coefficients $\{d_k^\ell\}$ in Equation (4.12). Then, a near sparsest expansion of a function f in the framelet system can be obtained as follows. We put into brackets the computational complexity for a finite dimensional signal of size N.

(1) Using Mallat's algorithm, perform a fast expansion $[\mathcal{O}(N)]$

$$f(x) = \sum_{j,m}\langle f, \psi_{j,m}\rangle\psi_{j,m}(x);$$

(2) Using Equation (4.12), rewrite the above expansion in terms of framelets $[\mathcal{O}(NLK)]$

$$\begin{aligned} f(x) &= \sum_{j,m} \sum_{\ell=1}^{L} \sum_{n=0}^{K-1} \langle f, \psi_{j,m} \rangle d_n^\ell 2^{j/2} \psi^\ell(2^j x - m - n/2) \\ &= \sum_{j,k,\ell} \underbrace{\sum_{2m+n=k} \langle f, \psi_{j,m} \rangle d_n^\ell}_{c_{j,k,\ell}} 2^{j/2} \psi^\ell(2^j x - k/2) \end{aligned}$$

This expansion algorithm adapts to unknown sparsity of f just as a wavelet expansion does. The oversampled system is ℓ_1^p-hilbertian in L_p so it follows that the *thresholding algorithm*, which provides m-term approximants $A_m(f)$ by keeping the m largest L_p-normalized coefficients from the latter expansion, yields the optimal rate of approximation. That is to say, for all f and α if

$$\sigma_m(f, X_2(\Psi))_{L_p(\mathbb{R})} = \mathcal{O}(m^{-\alpha}), \qquad m \geq 1$$

then

$$\|f - A_m(f)\|_{L_p(\mathbb{R})} = \mathcal{O}(m^{-\alpha}), \qquad m \geq 1.$$

4.3. Approximation with wave packets

A closer analysis of the proof of Theorem 11 reveals a surprising fact. All that is needed to obtain a Jackson estimate for $X_R(\Psi)$ is that some nice wavelet can be represented sparsely by $X_R(\Psi)$. In [**B**], this idea was taken further. The difficult part of Theorem 11 is to obtain an expansion like in (4.12), but it turns out that this process can be simplified if we replace the orthonormal wavelet by a wavelet-type φ-transform generator (see [**56, 57**]). Such a generator turns out to be much easier to approximate.

Suppose $\varphi \in L_2(\mathbb{R}^d)$ is a refinable function. An associated wave packet is a function ψ given by

$$\hat{\psi}(2\xi) = m(\xi)\hat{\varphi}(\xi), \tag{4.14}$$

where m is a trigonometric polynomial. We can now state the main result of [**B**].

THEOREM 12 ([**B**]). *Given $s \geq 0$, let $\varphi \in W^s(L_\infty(\mathbb{R}^d))$ be a compactly supported refinable function with associated finite mask and let $\Psi = \{\psi^i\}_{i \in F}$ be a finite sequence of wave packets with associated trigonometric polynomials $\{m^i\}_{i \in F}$, where at least one, say m^{i_0}, satisfies*

$$m^{i_0}(\xi) \neq 0 \quad \text{for } 0 < |\xi| < r$$

for some $r > 0$. If $d > 1$ we suppose $\{\varphi(\cdot - k)\}_{k \in \mathbb{Z}^d}$ is a locally linearly independent set (condition is void if $d = 1$). Then there exists $K_0 \in \mathbb{N}_0$ such that for $K \geq K_0$,

$$\mathcal{A}_q^{\alpha/d}\big(L_p(\mathbb{R}^d), X_K(\Psi)\big) = \mathcal{A}_q^{\alpha/d}\big(L_p(\mathbb{R}^d), X_K(\psi^{i_0})\big) = \big(L_p(\mathbb{R}^d), B_{\tau,\tau}^{\gamma}(\mathbb{R}^d)\big)_{\alpha/\gamma, q},$$

for $1 < p < \infty$, $0 < \alpha < \gamma < s$, *and* $1/\tau := \gamma/d + 1/p$.

We remark that Theorem 12 holds some surprises in the case $d = 1$. For $d = 1$, (12) is satisfied by any non-trivial trigonometric polynomial m, since its zeroes are isolated. Thus, for $d = 1$, the conclusion of Theorem 12 holds true for any $\psi \in \mathrm{Span}\{\varphi(\cdot - k) \colon k \in \mathbb{Z}\}$, so $X(\psi)$ need not even be close to a frame or have dense span in $L_2(\mathbb{R})$. For $d > 1$ the condition (12) is more restrictive. In particular, (12) is not satisfied for separable wavelet (or framelet) systems. However, in the framelet case we can use the filter relation (4.3) to obtain an equivalent result, see Proposition 4.2 in [**B**].

CHAPTER 5

Approximation with time-frequency frames

In this chapter we consider a general construction of smoothness spaces, a subclass of so-called decomposition spaces, defined on $\mathbb{R}^d$ for which it is possible to find adapted tight frames for $L_2(\mathbb{R}^d)$. Each frame forms an atomic decomposition of the smoothness space, and the space can be completely characterized by a sparseness condition on the frame coefficients. It is therefore possible to compress the elements of such smoothness spaces using the frame, and the sparse expansions naturally leads to Jackson estimates for m-term approximation with the frame.

The second part of the chapter contains a case study of so-called α-modulation spaces. The family of α-modulation spaces was introduced by Gröbner [**63**] and tight frames for such spaces are obtained from the general construction. For univariate α-modulation, we construct orthonormal (greedy) bases for the α-modulation spaces and completely characterize the corresponding approximation spaces.

Several authors have considered function spaces built using ideas related to decomposition spaces. A very general method to construct decomposition spaces was introduced by Feichtinger and Gröbner [**49**] and Feichtinger [**46**]. Gröbner [**63**] used the decomposition space methods in [**49**] to define the α-modulation spaces as a family of intermediate spaces between modulation and Besov spaces. Banach frames for α-modulation spaces have been considered by Fornasier [**55**] and by Borup and Nielsen [**10**]. Group theoretical constructions of function spaces, including smoothness spaces, have been studied by Feichtinger and Gröchenig [**47**, **50–52**, **64**]. Frazier and Jawerth constructed frames (their so-called φ-transform) for Besov and Triebel-Lizorkin spaces in [**56**, **57**].

5.1. Decomposition spaces and sparse time-frequency representations

Here we consider a general family of smoothness spaces based on structured decompositions of the frequency space $\mathbb{R}^d$. This is a fairly standard approach to define smoothness spaces. For example, Besov spaces introduced by Besov in [**5**] correspond to a dyadic decomposition of $\mathbb{R}^d$, while the family of modulation spaces introduced by Feichtinger [**48**] correspond to a uniform decomposition $\mathbb{R}^d$.

Structured coverings and decomposition spaces. First we restrict the family of possible covering of the frequency space to so-called admissible coverings of $\mathbb{R}^d$.

DEFINITION 6. *A set $\mathcal{Q} := \{Q_i\}_{i\in I}$ of measurable subsets $Q_i \subset \mathbb{R}^d$ is called an admissible covering if $\mathbb{R}^d = \cup_{i\in I} Q_i$ and there exists $n_0 < \infty$ such that $\#\{j \in I\colon Q_i \cap Q_j \neq \emptyset\} \leq n_0$ for all $i \in I$.*

In order to define smoothness space we need well-behaved resolution of the identity adapted to a given admissible covering.

DEFINITION 7. *Given an admissible covering $\{Q_i\}_{i\in I}$ of $\mathbb{R}^d$. A corresponding bounded admissible partition of unity (BAPU) is a family of functions $\Psi = \{\psi_i\}_{i\in I}$ satisfying*

- *$supp(\psi_i) \subset Q_i$, $i \in I$,*
- *$\sum_{i\in I} \psi_i(\xi) = 1, \forall \xi \in \mathbb{R}^d$,*
- $\sup_{i\in I} |Q_i|^{1/p-1} \|\mathcal{F}^{-1}\psi_i\|_{L_p} < \infty, \forall p \in (0,1]$.

Given $\psi_i \in \Psi$, we define the multiplier $\psi_i(D)f := \mathcal{F}^{-1}(\psi_i \mathcal{F} f)$, $f \in L_2(\mathbb{R}^d)$. Also recall that a (quasi-)Banach sequence space Y on I is called solid if $|a_i| \leq |b_i|$ for all i implies that $\|\{a_i\}\|_Y \leq \|\{b_i\}\|_Y$.

We can now give the definition of a decomposition space on the Fourier side. For particular choices of coverings, the decomposition spaces yield classical spaces such as Besov and modulation spaces, see **[D, 46]**.

DEFINITION 8. *Let $\mathcal{Q} = \{Q_i\}_{i\in I}$ be an admissible covering of $\mathbb{R}^d$ for which there exists a BAPU Ψ. Let Y be a solid (quasi-)Banach sequence space on I satisfying that $\ell_0(I)$, the finite sequences on I, is dense in Y. Then for $p \in (0,\infty]$, we define the decomposition space $D(\mathcal{Q}, L_p, Y)$ as the set of functions $f \in \mathcal{S}'(\mathbb{R}^d)$ satisfying*

$$\|f\|_{D(\mathcal{Q},L_p,Y)} := \Big\| \big\{ \|\psi_i(D)f\|_{L_p} \big\}_{i\in I} \Big\|_Y < \infty, \tag{5.1}$$

We notice that Definition 8 can be used only for admissible coverings for which there exists an associated BAPU (see Definition 7). This requirement clearly imposes some structure of the admissible coverings. A special class of structured admissible coverings for which an associated BAPU exists was introduced in **[D]**. Structured coverings are obtained by applying a countable family of invertible affine transformations on $\mathbb{R}^d$ to some fixed neighborhood of the origin.

DEFINITION 9. *Given a family $\mathcal{T} = \{A_k \cdot + c_k\}_{k\in\mathbb{N}}$ of invertible affine transformations on $\mathbb{R}^d$. Suppose there exist two bounded open sets $P \subset Q \subset \mathbb{R}^d$, with P compactly contained in Q, such that*

$$\{P_T\}_{T\in\mathcal{T}} \quad \textit{and} \quad \{Q_T\}_{T\in\mathcal{T}} \quad \textit{are admissible coverings.} \tag{5.2}$$

Also assume that there exists a constant K such that

$$(A_kQ + c_k) \cap (A_{k'}Q + c_{k'}) \neq \emptyset \Rightarrow \|A_{k'}^{-1}A_k\|_{\ell_\infty(\mathbb{R}^{d\times d})} \leq K. \tag{5.3}$$

Then we call $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ a structured admissible covering *and $\mathcal{T}$ a structured family of affine transformations.*

Let us briefly discuss how to construct a BAPU for a structured family of affine transformations $\mathcal{T}$, we refer to **[D]** for the technical details.. Pick a non-negative function $\Phi \in C^\infty(\mathbb{R}^d)$ with $\Phi(\xi) = 1$ for $\xi \in P$ and $\text{supp}(\Phi) \subset Q$. For $T \in \mathcal{T}$, we let $g_T(\xi) := \Phi(T^{-1}\xi)$. We then define a BAPU by

$$\psi_T(\xi) := \frac{g_T(\xi)}{\sum_{T'\in\mathcal{T}} g_{T'}(\xi)}.$$

We also define an associated "square-root" of the BAPU by

$$\varphi_T(\xi) = \frac{g_T(\xi)}{\sqrt{\sum_{T'\in\mathcal{T}} g_{T'}^2(\xi)}}, \qquad T \in \mathcal{T}. \tag{5.4}$$

Tight frames for decomposition spaces. The system $\{\varphi_T\}_T$ defined by (5.4) can be used for an easy construction of tight frames for $L_2(\mathbb{R}^d)$ compatible with the structured admissible covering. Consider a structured admissible covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$. Suppose K_a is a cube in $\mathbb{R}^d$ (aligned with the coordinate axes) with side-length $2a$ satisfying $Q \subseteq K_a$. For $T\cdot = A\cdot + c$, we let $|T| = |\det A|$, and we define

$$e_{n,T}(\xi) := (2a)^{-\frac{d}{2}}|T|^{-\frac{1}{2}}\chi_{K_a}(T^{-1}\xi)e^{i\frac{\pi}{a}n\cdot T^{-1}\xi}, \qquad n \in \mathbb{Z}^d,\ T \in \mathcal{T}.$$

and

$$\hat{\eta}_{n,T} := \varphi_T e_{n,T} \qquad n \in \mathbb{Z}^d,\ T \in \mathcal{T},$$

with φ_T given by (5.4). We can also obtain an explicit representation of $\eta_{n,T}$ in direct space. Suppose $T = A\cdot + c$, and $\hat{\mu}_T(\xi) := \varphi_T(T\xi)$. Then

$$\eta_{n,T}(x) = (2a)^{-\frac{d}{2}}|T|^{1/2}\mu_T(\tfrac{\pi}{a}n + A^\top x)e^{ix\cdot c}.$$

It can be verified that for any $N \in \mathbb{N}$,

$$|\mu_T(x)| \leq C_N(1+|x|)^{-N},$$

with C_N independent of $T \in \mathcal{T}$. We notice that $\eta_{n,T}$ is obtained by translating, "dilating", and modulating a unit-scale element μ_T. In some sense, $\eta_{n,T}$ is a mix between a Gabor and a wavelet system. It turns out (see **[D]** for details) that $\{\eta_{n,T}\}$ is a tight frame for $L_2(\mathbb{R}^d)$. That is, we have a canonical expansion

$$f = \sum_{T\in\mathcal{T}}\sum_{n\in\mathbb{Z}^d}\langle f, \eta_{n,T}\rangle\eta_{n,T}., \qquad \forall f \in L_2(\mathbb{R}^d). \tag{5.5}$$

We also claim that $\{\eta_{n,T}\}$ is adapted to characterize the associated family of decomposition spaces $D(\mathcal{Q}, L_p, Y)$ for suitable weights Y. We need the following class of moderate weights relative to an admissible covering $\mathcal{Q}$.

DEFINITION 10. *Let $\mathcal{Q} = \{Q_i\}_{i\in I}$ be an admissible covering of $\mathbb{R}^d$. A strictly positive function w on $\mathbb{R}^d$ is called $\mathcal{Q}$-moderate if there exists $C > 0$ such that $w(x) \leq Cw(y)$ for all $x, y \in Q_i$ and all $i \in I$. A strictly positive $\mathcal{Q}$-moderate weight on I (derived from w) is a sequence $v_i = w(x_i)$, $i \in I$, with $x_i \in Q_i$ and w a $\mathcal{Q}$-moderate function.*

For Y a solid (quasi-)Banach sequence space on I, we define the weighted space $Y_v := \{\{d_i\}_{i\in I} \colon \{d_i v_i\}_{i\in I} \in Y\}$.

To simplify the statement of Theorem 13 below, we let

$$\eta^p_{n,T} := |T|^{1/2-1/p}\eta_{n,T}$$

denote the function $\eta_{n,T}$ "normalized" in $L_p(\mathbb{R}^d)$, $p \in (0, \infty]$. We have the following characterization result.

THEOREM 13 (**[D]**). *Let $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ be a structured admissible covering. Let Y be a symmetric (quasi-)Banach sequence space on $\mathcal{T}$, and let v be a $\mathcal{Q}$-moderate weight. Then, for $0 < p \leq \infty$, we have the characterization*

$$\|f\|_{D(\mathcal{Q},L_p,Y_v)} \asymp \left\| \left\{ \Big(\sum_{n\in\mathbb{Z}^d} |\langle f, \eta^p_{n,T}\rangle|^p \Big)^{1/p} \right\}_{T\in\mathcal{T}} \right\|_{Y_v},$$

with the usual modification for $p = \infty$.

The problem of obtaining general structured admissible coverings is studied in detail in **[D]**. The approach followed there is to consider coverings given by balls in a suitable homogeneous space over $\mathbb{R}^d$. We refer to **[D]** for a number of examples.

Jackson estimates. Let us introduce a family of sparseness spaces $S^\beta_{p,q}$ associated with a certain type of admissible covering and a special class of weights. The spaces $S^\beta_{p,q}$ have a very simple characterizations in terms of the frame coefficients relative to $\{\eta_{n,T}\}$.

Let $\mathcal{T}$ be a structured countable family of invertible affine transformations with associated admissible covering $\mathcal{Q}$. Given $\beta \in \mathbb{R}$ and a $\mathcal{Q}$-moderate function w, define $v_{w,\beta} := \{(w(b_T))^\beta\}_{A_T\cdot+b_T\in\mathcal{T}}$. We let $S^\beta_{p,q}(\mathcal{T}, w)$ denote the decomposition space $D(\mathcal{Q}, L_p, (\ell_q)_{v_{w,\beta}})$ for $\beta \in \mathbb{R}$, $0 < p \leq \infty$, and $0 < q < \infty$.

To simplify further, let us suppose that there exits a constant $\delta > 0$ such that $w(b_T) \asymp w_\delta(b_T) := |T|^{\frac{1}{\delta}}$ for $T \in \mathcal{T}$. Notice that,

$$|\langle f, \eta^\tau_{n,T}\rangle| = |T|^{1/p-1/\tau}|\langle f, \eta^p_{n,T}\rangle|, \qquad \text{for } 0 < \tau, p \leq \infty.$$

Then, according to Theorem 13, we have the characterization

$$\|f\|_{S^\beta_{p,q}} \asymp \Big(\sum_T |T|^{\beta q/\delta}\Big(\sum_{n\in\mathbb{Z}^d} |\langle f, \eta^p_{n,T}\rangle|^p\Big)^{q/p}\Big)^{1/q}.$$

$$= \Big(\sum_{T\in\mathcal{T}}\Big(\sum_{n\in\mathbb{Z}^d} |\langle f, \eta^r_{n,T}\rangle|^p\Big)^{q/p}\Big)^{1/q}, \qquad \text{when } \frac{\beta}{\delta} = \frac{1}{p} - \frac{1}{r}. \tag{5.6}$$

We can use this simple characterization to obtain Jackson estimates for functions in $\mathcal{S}^\gamma_{\tau,\tau}(\mathcal{T}, w_\delta)$. For $f \in \mathcal{S}^\gamma_{\tau,\tau}(\mathcal{T}, w_\delta)$, let $\{\theta_m\}_{m\in\mathbb{N}}$ be a decreasing rearrangement of the frame coefficients

$$\{|\langle f, \eta^r_{n,T}\rangle|\}_{n,T}, \tag{5.7}$$

where r is given by $\gamma/\delta = 1/\tau - 1/r$. We let f^F_m be the m-term approximation of f obtained by extracting from the canonical frame expansion of f the terms corresponding to the m largest coefficients from (5.7). Assume $\beta \in \mathbb{R}$, and $p > 0$ satisfy $(\gamma - \beta)/\delta = 1/\tau - 1/p > 0$. Then using the techniques of Chapter 3, and the fact that $\beta/\delta = 1/p - 1/r$, the approximation error in $S^\beta_{p,p}$ obeys

$$\|f - f^F_m\|_{S^\beta_{p,p}} \le C'\|f\|_{\mathcal{S}^\gamma_{\tau,\tau}} \cdot m^{-(\gamma-\beta)/\delta}, \tag{5.8}$$

see **[D]** for details.

The estimate (5.8) immediately leads to a Jackson inequality for nonlinear m-term approximation with $\mathcal{D} = \{\eta_{n,T}\}$ since $f^F_m \in \Sigma_m(\mathcal{D})$. We have

$$\sigma_m(f, \mathcal{F}(\mathcal{T}))_{S^\beta_{p,p}} \le C\|f\|_{\mathcal{S}^\gamma_{\tau,\tau}} \cdot m^{-(\gamma-\beta)/\delta}, \qquad m \ge 1,$$

for $1/\tau - 1/p = (\gamma - \beta)/\delta$.

5.2. Case study: α-modulation spaces

In this section we study an interesting special case of the general construction of the smoothness spaces $S^\beta_{p,q}$. The family of α-modulation spaces $M^{\beta,\alpha}_{p,q}(\mathbb{R}^d)$ was introduced by Gröbner **[63]**. They form a family of "intermediate" spaces between the classical modulation spaces and the Besov spaces.

For $0 < \alpha < 1$, define $b_k = k|k|^{\alpha/(1-\alpha)}$, $k \in \mathbb{Z}^d \setminus \{0\}$, and let $\mathcal{T} = \{T_k\}_{k\in\mathbb{Z}^d\setminus\{0\}}$ be given by

$$T_k\xi = |k|^{\alpha/(1-\alpha)}\xi + b_k,$$

with $|\cdot|$ the Euclidean norm on $\mathbb{R}^d$. An associated structured covering is given by the Euclidean balls

$$\big\{B\big(b_k, \rho|k|^{\alpha/(1-\alpha)}\big)\big\}_{k\in\mathbb{Z}^d},$$

with ρ a suitable positive constant. We have the geometric property that for $Q \in \mathcal{Q}$,

$$|Q| \asymp (1+|x|)^{\alpha d}, \qquad x \in Q.$$

A dyadic covering corresponds to the limit case $\alpha = 1$, while a uniform covering corresponds to $\alpha = 0$. For $0 < \alpha < 1$ we obtain an intermediate type "polynomial" covering. This type of covering was first considered by Päivärinta and Somersalo in [**86**] to study pseudodifferential operators, and it was proven in [**10**] that (5.6) with a suitable weight w gives a characterization of the α-modulation space $M^{\beta,\alpha}_{p,q}(\mathbb{R}^d)$. In fact, for $w = 1 + |\cdot|$, we have

$$S^{\beta}_{p,q}(\mathcal{T}, 1+|\cdot|) = M^{\beta,\alpha}_{p,q}(\mathbb{R}^d), \qquad 0 \le \alpha \le 1, \beta \in \mathbb{R}, 0 < p, q < \infty.$$

It is proved in [**D**] that the system $\{\eta_{n,T}\}$ forms a Banach frame for $M^{\beta,\alpha}_{p,q}(\mathbb{R}^d)$ for $1 \le p, q < \infty$. Other constructions of Banach frames for α-modulation spaces can be found in [**10, 55**].

For $0 < \alpha < 1$, (5.1) provides the Jackson estimate

$$\sigma_m(f, \{\eta_{n,T}\})_{M^{\beta,\alpha}_{p,p}(\mathbb{R}^d)} \le C\|f\|_{M^{\gamma,\alpha}_{\tau,\tau}(\mathbb{R}^d)} \cdot m^{-(\gamma-\beta)/(d\alpha)}, \qquad m \ge 1,$$

for $1/\tau - 1/p = (\gamma - \beta)/(d\alpha)$.

At present, no Bernstein estimate is known for the dictionary $\{\eta_{n,T}\}$. However, in the univariate case we can actually created a non-redundant basis for $M^{\beta,\alpha}_{p,q}(\mathbb{R})$ based on so-called brushlets.

Brushlets and univariate α-modulation spaces. We now consider the univariate α-modulation spaces. In the univariate case, we can construct so-called brushlet bases adapted to the α-modulation spaces. The brushlet systems are based on local Fourier bases as introduced by Coifman and Meyer in [**23**], and by Malvar in [**80**] for applications in signal processing. These systems were further developed by Wickerhauser in [**103**]. An atom from a local Fourier basis has perfect localization in time and is well localized in frequency. Laeng noticed in [**79**] that it is possible to map a local Fourier basis by the Fourier transform to a new basis with compact support in the frequency domain. In [**81**], Coifman and Meyer studied similar systems, called brushlets, using the bases introduced by Wickerhauser.

Let us introduce the brushlet system. Let $\mathbb{I}$ be a disjoint covering $\mathbb{R}$ consisting of pairwise disjoint half-open intervals $I = [\alpha_I, \alpha'_I)$, $\alpha_I < \alpha'_I$. We suppose that each interval in $\mathbb{I}$ has a unique adjacent interval in $\mathbb{I}$ to the left and to the right, and that there exists a constant $A > 1$ such that

$$A^{-1} \le \frac{|I|}{|I'|} \le A, \qquad \text{for all adjacent } I, I' \in \mathbb{I}.$$

We assign to each interval $I \in \mathbb{I}$ a cutoff radius $\varepsilon_I > 0$ at the left endpoint and a cutoff radius $\varepsilon'_I > 0$ at the right endpoint, satisfying

$$\begin{cases} \text{(i)} & \varepsilon'_I = \varepsilon_{I'} \text{ whenever } \alpha'_I = \alpha_{I'} \\ \text{(ii)} & \varepsilon_I + \varepsilon'_I \leq |I| \\ \text{(iii)} & \varepsilon_I \geq c|I|, \end{cases}$$

with $c > 0$ independent of I.

For each $I \in \mathbb{I}$, we construct a smooth bell function localized in a neighborhood of this interval. Take a non-negative ramp function $\rho \in C^r(\mathbb{R})$, for some $r \geq 1$, satisfying

$$\rho(\xi) = \begin{cases} 0 & \text{for } \xi \leq -1, \\ 1 & \text{for } \xi \geq 1, \end{cases}$$

with the property that

$$\rho(\xi)^2 + \rho(-\xi)^2 = 1 \qquad \text{for all } \xi \in \mathbb{R}.$$

Define for each $I = [\alpha_I, \alpha'_I) \in \mathbb{I}$ the *bell function*

$$b_I(\xi) := \rho\left(\frac{\xi - \alpha_I}{\varepsilon_I}\right)\rho\left(\frac{\alpha'_I - \xi}{\varepsilon'_I}\right).$$

Notice that $\operatorname{supp}(b_I) \subset [\alpha_I - \varepsilon_I, \alpha'_I + \varepsilon'_I]$ and $b_I(\xi) = 1$ for $\xi \in [\alpha_I + \varepsilon_I, \alpha'_I - \varepsilon'_I]$.

The set of local cosine functions

$$\hat{w}_{n,I}(\xi) = \sqrt{\frac{2}{|I|}}\, b_I(\xi) \cos\left(\pi\big(n + \tfrac{1}{2}\big)\frac{\xi - \alpha_I}{|I|}\right), \quad n \in \mathbb{N}_0, \quad I \in \mathbb{I},$$

constitute an orthonormal basis for $L_2(\mathbb{R})$, see e.g. **[1]**. We call the collection $\{w_{n,I} \colon I \in \mathbb{I}, n \in \mathbb{N}_0\}$ a *brushlet system*. The brushlets also have an explicit representation in the time domain. Define the set of *central bell functions* $\{g_I\}_{I \in \mathbb{I}}$ by

$$\hat{g}_I(\xi) := \rho\left(\frac{|I|}{\varepsilon_I}\xi\right)\rho\left(\frac{|I|}{\varepsilon'_I}(1 - \xi)\right),$$

such that $b_I(\xi) = \hat{g}_I\big(|I|^{-1}(\xi - \alpha_I)\big)$, and let

$$e_{n,I} := \frac{\pi\big(n + \frac{1}{2}\big)}{|I|}, \qquad I \in \mathbb{I},\ n \in \mathbb{N}_0.$$

Then,

$$w_{n,I}(x) = \sqrt{\frac{|I|}{2}}\, e^{i\alpha_I x}\big\{g_I\big(|I|(x + e_{n,I})\big) + g_I\big(|I|(x - e_{n,I})\big)\big\}.$$

By a straight forward calculation it can be verified that there exists a constant $C < \infty$ independent of $I \in \mathbb{I}$, such that

$$|g_I(x)| \leq C(1 + |x|)^{-r}, \tag{5.9}$$

with $r \geq 1$ given by the smoothness of the ramp function. Thus a brushlet $w_{n,I}$ essentially consists of two humps at $\pm e_{n,I}$.

Brushlet bases are extremely flexible since we have a lot of freedom choosing the covering $\mathbb{I}$. Brushlet systems based on dyadic coverings are considered in **[9]**, where it is shown that such system form greedy bases for L_p, $1 < p < \infty$. Here we are interested in coverings compatible with α-modulation spaces. We have the following definition.

DEFINITION 11. *A family $\mathbb{I}$ of intervals $I \in \mathbb{R}$ is called an an α-covering of $\mathbb{R}$ if there exists a constant $0 \leq \alpha \leq 1$, such that $|I| \asymp (1+|\xi|)^\alpha$ for all $I \in \mathbb{I}$, and all $\xi \in I$.*

For brushlets based on α-coverings of $\mathbb{R}$, we have the following complete characterization of m-term brushlet approximation in α-modulation space. We mention that it is possible to prove Bernstein estimates in this case due to the fact that the brushlet system is non-redundant unlike the redundant frame defined in Section 5.1. The proof of Theorem 14 is based on an application of Theorem 7, see **[C]** for details.

THEOREM 14 (**[C]**). *Let $\{w_{n,I}\}_{I\in\mathbb{I},n\in\mathbb{N}_0}$ be a brushlet system associated with a disjoint α-covering $\mathbb{I}$ for some $0 < \alpha \leq 1$, and let $\mathcal{B} = \big\{w_{n,I}/\|w_{n,I}\|_{M_p^{s,\alpha}(L_p)}\big\}_{I\in\mathbb{I},n\in\mathbb{N}}$. Then $\mathcal{B}$ constitutes a greedy basis for the α-modulation spaces $M^{s,\alpha}_{p,p}(\mathbb{R})$, $1 < p < \infty$, $s \in \mathbb{R}$, and*

$$\mathcal{A}^\gamma_q\big(M^{s,\alpha}_{p,p}(\mathbb{R}),\mathcal{B}\big) = \mathcal{K}^\tau_q\big(M^{s,\alpha}_{p,p}(\mathbb{R}),\mathcal{B}\big), \qquad \tau^{-1} = \gamma + p^{-1},\ \gamma > 0,\ 0 < q \leq \infty,$$

with equivalent norms. Moreover, for $\tau > 0$,

$$\mathcal{K}^\tau_\tau\big(M^{s,\alpha}_{p,p}(\mathbb{R}),\mathcal{B}\big) = M^{\beta,\alpha}_{\tau,\tau}(\mathbb{R}), \qquad \textit{with} \quad \beta = \frac{\alpha}{\tau} - \frac{\alpha}{p} + s.$$

CHAPTER 6

Sparse representations

In this final chapter we consider the computational problem of finding an efficient representation of a signal w.r.t. a redundant dictionary in a finite dimensional space. Given a redundant signal (or image) dictionary in a finite dimensional space and a signal, we would ideally like to find the best (or near) approximation to the signal with a prescribed number of atoms. However, it was proved by Davis et al. [**28**] that finding the best approximation of a signal from an arbitrary dictionary with a prescribed number of atoms is an *NP*-hard problem, so no efficient algorithm exists at present.

We therefore restrict our attention to the more tractable problem of finding a sparse representation of the signal relative to a redundant dictionary. In the early 1990's, the so-called Matching Pursuit and Basis Pursuit strategies were introduced with the purpose of getting good representations of signals relative to redundant dictionaries. It was experimentally noticed that for a simple redundant dictionary given by the union of a Dirac and Fourier orthonormal basis, when the signal has a sufficiently sparse expansion (in the sense of counting non-zero coefficients) in the dictionary, the Basis Pursuit algorithm can exactly recover it. This observation lead to the seminal contribution [**44**] by Donoho and Huo, where a mathematical explanation of the experimental facts was given.

Since then, mathematical problems related to sparse representations have attracted a great deal of attention, and several authors have contributed to the area, we mention here [**15, 41, 42, 53, 58, 100–102**]. See also the editorial [**62**]. Recovery of signals contaminated with noise is considered in [**43, 61, 102**].

Sparse representations are not only interesting from a theoretical point of view. It is now well established in signal processing that sparse representations are useful for applications as diverse as compression [**32, 33**], feature extraction [**54, 74**], and blind source separation [**60, 105**].

6.1. Sparse representations through optimization

Let us introduce some notation. A dictionary in $\mathcal{H} = \mathbb{R}^N$ (resp. $\mathcal{H} = \mathbb{C}^N$) is a family of $K \geq N$ unit (column) vectors $\{g_k\}$ that spans $\mathcal{H}$. We will use the matrix notation $\mathcal{D} = [g_1, \ldots, g_K]$ for a dictionary.

By a *representation* of s in $\mathcal{D}$ we mean a (column) vector $\alpha = (\alpha_k) \in \mathbb{R}^K$ (resp. in $\mathbb{C}^K$) such that

$$s = \mathcal{D}\alpha.$$

We notice that when $K > N$, the vectors of $\mathcal{D}$ are no longer linearly independent and the representation of s is not unique. The hope is that among all possible representations of s there is a *very sparse* representation, i.e., a representation with few non-zero coefficients. The trade-off is that we have to *search* all possible representations of s to find the sparse representations, and then determine whether there is a unique sparsest representation.

The traditional approach (following [**44**, **45**]) is to measure the sparsity of a representation $s = \mathcal{D}\alpha$ by two quantities: the ℓ_0 and the ℓ_1 norm of α, resp. The ℓ_0-"norm" simply counts the number of non-zero entries of a vector. We also notice that the optimal representation w.r.t. the ℓ_1-norm is closely related to the sparseness norm $\|\cdot\|_{\mathcal{K}_1^1(\mathcal{D})}$.

Our goal is to find the sparsest representation and this leads in a natural way to the following two minimization problems to determine the sparsest representation of s:

$$\text{minimize} \quad \|\alpha\|_0 \quad \text{subject to } s = \mathcal{D}\alpha, \tag{6.1}$$

and

$$\text{minimize} \quad \|\alpha\|_1 \quad \text{subject to } s = \mathcal{D}\alpha. \tag{6.2}$$

It turns out that the optimization problem (6.2) is much easier to handle than the combinatorial optimization (6.1). In fact, a solution to (6.2) can be calculated in polynomial time using the Basis Pursuit algorithm introduced by Donoho et al. [**16**]. The Basis Pursuit algorithm is based on linear programming techniques.

It is thus important to know the relationship between the solution(s) of (6.1) and (6.2), and to determine sufficient conditions for the two problems to have the same unique solution. This problem has been studied in details in [**44**] and later been refined in [**45**] in the special case where the dictionary $\mathcal{D}$ is the union of *two* orthonormal bases.

Recall that $M(\mathcal{D})$, defined by

$$M(\mathcal{D}) := \max_{k \neq k'} |\langle g_k, g_{k'} \rangle| \,,$$

is the coherence of the dictionary. The following result gives a sufficient condition for the optimization problems (6.2) and (6.1) to have the same unique solution.

THEOREM 15 ([**F**]). *For any dictionary, if*

$$\|\alpha\|_0 < \frac{1}{2}(1 + 1/M(\mathcal{D})) \tag{6.3}$$

then α is the (unique) solution to both the ℓ^0 and the ℓ^1 minimization problems.

We can apply the result as follows. For a given signal, run the Basis Pursuit algorithm, and obtain a minimizer of (6.2). If that minimizer has support that satisfies the condition of Theorem 15, then we know that the minimizer is the unique solution to both (6.2) and (6.1). In the negative case, no conclusion about the obtained representation can be made.

Next we consider the situation where the dictionary $\mathcal{D}$ is a union of L orthonormal bases. Taking unions of bases is an easy way to build dictionaries, and with care it is possible to obtain highly redundant dictionaries with values of $M(\mathcal{D})$ equal to $1/\sqrt{N}$, see **[F]**. The special case of a union of two orthonormal bases is considered in **[44,45]**. We have the following more general result for the union of $L \geq 2$ orthonormal bases.

THEOREM 16 (**[F]**). *Let $\mathcal{D} = [\mathcal{B}_1, \ldots, \mathcal{B}_L]$ be a union of $L \geq 2$ orthonormal bases. Consider a representation $\alpha = \begin{bmatrix} \alpha^1 \\ \ldots \\ \alpha^L \end{bmatrix}$ with $\alpha^l \in \mathbb{R}^N$ (resp. $\mathbb{C}^N$). Without loss of generality, we can assume that the bases $\mathcal{B}_l$ have been numbered so that $\|\alpha^1\|_0 \leq \ldots \leq \|\alpha^L\|_0$. If*

$$\sum_{l \geq 2} \frac{M\|\alpha^l\|_0}{1 + M\|\alpha^l\|_0} < \frac{1}{2(1 + M\|\alpha^1\|_0)}. \tag{6.4}$$

then α is the (unique) solution to both the ℓ_0 and the ℓ_1 minimization problems.

The hypothesis of Theorem 16 is perhaps a bit cumbersome to check for a given representation α, and the following non-trivial corollary proved in **[F]** gives a more explicit condition that is very easy to check.

COROLLARY 1. *For a dictionary that is the union of L orthonormal bases, if*

$$\|\alpha\|_0 < \left(\sqrt{2} - 1 + \frac{1}{2(L-1)}\right) \frac{1}{M(\mathcal{D})} \tag{6.5}$$

then α is the (unique) solution to both the ℓ_0 and the ℓ_1 minimization problems.

The following table gives numerical values of the constants

$$c(L) := \sqrt{2} - 1 + \frac{1}{2(L-1)}.$$

L	2	3	4	5	6	7
c	0.914	0.664	0.580	0.539	0.514	0.497

TABLE 1. Numerical values of the constant $\sqrt{2} - 1 + \frac{1}{2(L-1)}$ in Corollary 1 for small values of L.

By comparing $c(L)$ to the values given by the general estimate (6.3), we see that Corollary 1 is *stronger* only for $L \leq 6$.

Other sparseness measures. We are not restricted to using only ℓ_0 and ℓ_1 to measure sparseness of a representation α. For example, we can use the ℓ_τ-"norms" given by

$$\|\alpha\|_\tau := \sum_k |\alpha_k|^\tau,$$

for $0 < \tau < 1$. We can also consider a more general measure f,

$$\|\alpha\|_f := \sum_k f(|\alpha_k|), \tag{6.6}$$

where $f : [0,\infty) \to [0,\infty)$ is a fixed function. Obviously, we need to impose some restrictions on f to be able to claim that it measures sparseness. The following class of admissible sparseness measures is considered in [**L**].

DEFINITION 12. *We let $\mathcal{M}$ the set of all non-decreasing functions $f : [0,\infty) \to [0,\infty)$, not identically zero, with $f(0) = 0$ and such that $t \mapsto f(t)/t$ is non-increasing on $(0,\infty)$.*

One can verify that $\mathcal{M}$ contains both the ℓ_0 and ℓ_1 norm, and the intermediate ℓ_τ, $0 < \tau < 1$, norms. For $f \in \mathcal{M}$, it makes sense to study the relationship between the solution(s) of the computationally tractable problem (6.2) and the optimization problem

$$\text{minimize} \quad \|\alpha\|_f \quad \text{subject to } s = \mathcal{D}\alpha. \tag{6.7}$$

We have the following result that relates the solution(s) of (6.2) and (6.7).

THEOREM 17 ([**L**]). *Let $\mathcal{D}$ be an arbitrary dictionary in a finite dimensional space. Assume m is an integer such that for any x and y with $y = \mathcal{D}x$ and $\|x\|_0 \le m$, x is the unique ℓ_1-sparsest representation of y. Then, for any x and y such that $y = \mathcal{D}x$ and $\|x\|_0 \le m$, x is indeed the unique f-sparsest representation of y for any sparseness measure $f \in \mathcal{M}$. In particular, it is the ℓ_τ-sparsest representation for $0 \le \tau \le 1$.*

Let us consider an example. For a general dictionary $\mathcal{D}$, Theorem 15 shows that any integer

$$m < \frac{1}{2}(1 + 1/M(\mathcal{D}))$$

is admissible. Suppose a given signal s has a representation $s = \mathcal{D}\alpha$ with $\|\alpha\|_0 < \frac{1}{2}(1 + 1/M(\mathcal{D}))$, then we deduce that α is the unique minimizer of (6.7) for any sparseness measure $f \in \mathcal{M}$.

For $\mathcal{D}$ the union of L orthonormal bases, Theorem 16 (or rather Corollary 1) can be used to obtain suitable values of m.

6.2. Beyond sparsity: structured representations

In the previous section, and in several related recent papers [**42, 44, 45, 100**], sufficient conditions have been identified where algorithms such as Basis Pursuit actually compute an optimal representation of a given signal, in the sense that they solve the best approximation problem under a constraint on the size of the support of the signal. Typically, one calculates the coherence of $\mathcal{D}$

$$M(\mathcal{D}) = \max_{i \neq j} |\langle g_i, g_j \rangle|.$$

Then for signals X with a representation $X = \mathcal{D}(S)$ satisfying $|\operatorname{supp}(S)| < \lfloor \frac{1}{2}(1 + 1/M) \rfloor$, Basis Pursuit will recover the representation S. One serious problem with this type of results using the coherence is that they represent *worst case* estimates. For example, the coherence is close to one as soon as we have one pair of atoms that are approximately co-linear while the rest of the dictionary may be much nicer. For such dictionaries, only extraordinary sparse signal representations (perhaps with one or two non-zero coefficients) can be recovered by Basis Pursuit.

A more refined type of result can be obtained by considering the cumulative coherence introduced by Tropp **[100]**

$$\mu_1(\mathcal{D}, m) := \sup_{|\Lambda| = m} \sup_{j \notin \Lambda} \sum_{i \in \Lambda} |\langle g_i, g_j \rangle|.$$

However, the cumulative coherence also gives a *worst case* estimate that does not take into account the finer structure of the dictionary, and the derived estimates are often too weak for many applications.

One way to overcome these shortcomings is by shifting to a probabilistic viewpoint and consider *random dictionaries*. The probabilistic approach has been considered in a number of recent papers, see e.g. **[13, 14, 40]**. Random dictionaries are typically created by picking a number of unit vectors randomly from some larger ensemble. The results on sparse representations using random dictionaries are typically much better than the corresponding deterministic results. One problem is that the results are difficult to interpret when we consider a specific dictionary.

Here we are interested in obtaining more optimistic results for structured dictionaries. The idea is to use more of the structure of the dictionary. Since the point of view has changed slightly, let us reintroduce the notion of a dictionary. Let F and G be two finite index sets, and let $\mathcal{H} = \mathbb{C}^G$ be the signal space. A dictionary for $\mathcal{H}$ is a linear map $\mathcal{D}\colon \mathbb{C}^F \to \mathbb{C}^G$ from the coefficient space $\mathbb{C}^F$ onto the signal space (one can also replace $\mathbb{C}^F$ and $\mathbb{C}^G$ with $\mathbb{R}^F$ and $\mathbb{R}^G$). The atoms associated with $\mathcal{D}$ are the columns of the matrix representation of $\mathcal{D}$ wrt. the canonical bases for $\mathbb{C}^F$ and $\mathbb{C}^G$, i.e., $\mathcal{D} = [g_i]_{i \in F}$. We assume that the dictionary is normalized with respect to the ℓ_2 norm, i.e, that $\|g_i\| = 1$, for $i \in F$. The support of a coefficient sequence $S = (s_i)_{i \in F} \in \mathbb{C}^F$ is defined as $\operatorname{supp}(S) = \{i \in F : s_i \neq 0\} \subseteq F$.

With this notations the sparse approximation problem can be expressed as

$$\min_S \|X - \mathcal{D}(S)\| \quad \text{subject to} \quad |\operatorname{supp}(S)| \leq m \tag{6.8}$$

where $X = (x_i)_{i \in G} \in \mathbb{C}^G$ and $|I|$ denotes the cardinal of the set I.

We generalize this problem by considering, for any family $\mathcal{S}$ of subsets of F, the following *structured approximation problem*, or approximation with structure constraint $\mathcal{S}$

$$\min_S \|X - \mathcal{D}(S)\| \quad \text{subject to } \mathrm{supp}(S) \in \mathcal{S}. \tag{6.9}$$

A particular instance of the structured approximation problem is the sparse approximation problem (6.8), which corresponds to the family $\mathcal{S}_m = \{I \subseteq F : |I| \leq m\}$. That is to say, we simply put as a constraint a bound on the allowed number of nonzero coefficients in S. However, in many cases it also makes sense to consider families $\mathcal{S}$ taking into account not only the sparsity of I but also properties that may be related to the "geometry" of F and G. To study this problem more closely, we introduce the concept of *identifiable structures*.

DEFINITION 13. *A family $\mathcal{S}$ of subsets of F is called an* identifiable structure *if $\mathcal{D}(S) = \mathcal{D}(S')$ with supp(S), supp$(S') \in \mathcal{S}$ implies that $S = S'$.*

The significance of Definition 13 is the following:

(1) if a signal X satisfies the *model* $X = \mathcal{D}(S)$ with S supported on an identifiable structure $\mathcal{S}$, then the representation S is the *unique* representation of X supported on $\mathcal{S}$, and it can be recovered as the unique solution of the optimization problem (6.9).
(2) if an *algorithm* (supposedly computationally efficient) provides *some* representation $X = \mathcal{D}(S_{\mathrm{alg}})$ where S_{alg} is supported on an identifiable structure $\mathcal{S}$, then one can be sure that this representation is optimal within the class of representations supported by $\mathcal{S}$, thus bypassing the (generally hard) combinatorial optimization in (6.9).

Examples of identifiable structure can be found in [**K**]. In the recent paper [**8**], structured infinite representations in a wavelet/Gabor dictionary are studied.

Let us consider one example of an identifiable structure. For $I \subset F$, we define

$$P_1(I) = \sup_{Z \in \mathrm{Ker}(\mathcal{D}), Z \neq 0} \frac{\sum_{i \in I} |z_i|}{\|Z\|_1},$$

with $\|Z\|_1 := \sum_{i \in F} |z_i|$ and $\mathrm{Ker}(\mathcal{D})$ the null space of $\mathcal{D}$. The following is proved in [**L**].

LEMMA 3. *Let $X = \mathcal{D}(S)$. Let I be such that supp$(S) \subseteq I$, and suppose*

$$P_1(I) := \sup_{Z \in Ker(\mathcal{D}), Z \neq 0} \frac{\sum_{i \in I} |z_i|}{\|Z\|_1} < \frac{1}{2}. \tag{6.10}$$

Then S is the unique solution of $\min \sum_{i \in F} |s'_i|$ *subject to $X = \mathcal{D}(S')$.*

This leads us to define the class $\mathcal{S}_{\mathrm{LP}} := \left\{ I \subseteq F : P_1(I) < \frac{1}{2} \right\}$. We have the following result which concludes this thesis.

THEOREM 18 (**[K]**). *The structure class $\mathcal{S}_{LP}$ is identifiable: if a signal X has two representations S and S' satisfying supp(S), supp$(S') \in \mathcal{S}_{LP}$, then $S = S'$. Moreover, the unique representation of $X = \mathcal{D}(S)$ with supp$(S) \in \mathcal{S}_{LP}$ is the solution of the ℓ_1 minimization problem*

$$\min \sum_{i \in F} |s'_i| \qquad \textit{subject to } X = \mathcal{D}(S'),$$

and it can therefore be recovered by the Basis Pursuit algorithm.

References

[1] P. Auscher, G. Weiss, and M. V. Wickerhauser. Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets. In *Wavelets*, volume 2 of *Wavelet Anal. Appl.*, pages 237–256. Academic Press, Boston, MA, 1992.

[2] P. Bechler, R. Devore, A. Kamont, G. Petrova, and P. Wojtaszczyk. Greedy wavelet projections are bounded on BV. *Trans. Amer. Math. Soc.*, 359(2):619–635 (electronic), 2007.

[3] C. Bennett and R. Sharpley. *Interpolation of operators*, volume 129 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, 1988.

[4] J. Bergh and J. Peetre. On the spaces V_p $(0 < p \leq \infty)$. *Boll. Un. Mat. Ital. (4)*, 10:632–648, 1974.

[5] O. V. Besov. Investigation of a class of function spaces in connection with imbedding and extension theorems. *Trudy. Mat. Inst. Steklov.*, 60:42–81, 1961.

[6] M. Š. Birman and M. Z. Solomjak. Piecewise polynomial approximations of functions of classes ${W_p}^\alpha$. *Mat. Sb. (N.S.)*, 73 (115):331–355, 1967.

[7] L. Borup, R. Gribonval, and M. Nielsen. Tight wavelet frames in Lebesgue and Sobolev spaces. *J. Funct. Spaces Appl.*, 2(3):227–252, 2004.

[8] L. Borup, R. Gribonval, and M. Nielsen. Beyound coherence: Recovering structured time-frequency representations. *Preprint*, 2007.

[9] L. Borup and M. Nielsen. Approximation with brushlet systems. *J. Approx. Theory*, 123(1):25–51, 2003.

[10] L. Borup and M. Nielsen. Banach frames for multivariate α-modulation spaces. *J. Math. Anal. Appl.*, 321(2):880–895, 2006.

[11] J. A. Brudnyĭ. Spline approximation, and functions of bounded variation. *Dokl. Akad. Nauk SSSR*, 215:511–513, 1974.

[12] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[13] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[14] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Submitted*, 2006.

[15] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159 (electronic), 2001. Reprinted from SIAM J. Sci. Comput. **20** (1998), no. 1, 33–61 (electronic) [MR1639094 (99h:94013)].

[17] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2003.

[18] C. K. Chui, W. He, and J. Stöckler. Compactly supported tight and sibling frames with maximum vanishing moments. *Appl. Comput. Harmon. Anal.*, 13(3):224–262, 2002.

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

[20] A. Cohen, R. DeVore, G. Kerkyacharian, and D. Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(2):167–191, 2001.

[21] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space $\mathrm{BV}(\mathbf{R}^2)$. *Amer. J. Math.*, 121(3):587–628, 1999.

[22] A. Cohen, R. A. DeVore, and R. Hochmuth. Restricted nonlinear approximation. *Constr. Approx.*, 16(1):85–113, 2000.

[23] R. R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(3):259–261, 1991.

[24] A. Córdoba and P. Fernández. Convergence and divergence of decreasing rearranged Fourier series. *SIAM J. Math. Anal.*, 29(5):1129–1139 (electronic), 1998.

[25] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[26] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5):1271–1283, 1986.

[27] I. Daubechies, B. Han, A. Ron, and Z. Shen. Framelets: MRA-based constructions of wavelet frames. *Appl. Comput. Harmon. Anal.*, 14(1):1–46, 2003.

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

[29] C. de Boor, R. A. DeVore, and A. Ron. Approximation orders of FSI spaces in $L_2(\mathbf{R}^d)$. *Constr. Approx.*, 14(4):631–652, 1998.

[30] R. De Vore, G. Petrova, and V. Temlyakov. Best basis selection for approximation in L_p. *Found. Comput. Math.*, 3(2):161–185, 2003.

[31] R. A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[32] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Inform. Theory*, 38(2, part 2):719–746, 1992.

[33] R. A. DeVore, B. Jawerth, and V. Popov. Compression of wavelet decompositions. *Amer. J. Math.*, 114(4):737–785, 1992.

[34] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov. Hyperbolic wavelet approximation. *Constr. Approx.*, 14(1):1–26, 1998.

[35] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[36] R. A. DeVore and V. A. Popov. Interpolation spaces and nonlinear approximation. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 191–205. Springer, Berlin, 1988.

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

[38] R. A. DeVore and V. N. Temlyakov. Nonlinear approximation in finite-dimensional spaces. *J. Complexity*, 13(4):489–508, 1997.

[39] S. J. Dilworth, N. J. Kalton, D. Kutzarova, and V. N. Temlyakov. The thresholding greedy algorithm, greedy bases, and duality. *Constr. Approx.*, 19(4):575–597, 2003.

[40] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[41] D. L. Donoho. For most large underdetermined systems of equations, the minimal l_1-norm near-solution approximates the sparsest near-solution. *Comm. Pure Appl. Math.*, 59(7):907–934, 2006.

[42] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[43] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1):6–18, 2006.

[44] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, 2001.

[45] M. Elad and A. M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, 2002.

[46] H. G. Feichtinger. Banach spaces of distributions defined by decomposition methods. II. *Math. Nachr.*, 132:207–237, 1987.

[47] H. G. Feichtinger. Amalgam spaces and generalized harmonic analysis. In *Proceedings of the Norbert Wiener Centenary Congress, 1994 (East Lansing, MI, 1994)*, volume 52 of *Proc. Sympos. Appl. Math.*, pages 141–150, Providence, RI, 1997. Amer. Math. Soc.

[48] H. G. Feichtinger. Modulation spaces of locally compact abelian groups. In R. Radha, M. Krishna, and S. Thangavelu, editors, *Proc. Internat. Conf. on Wavelets and Applications*, pages 1–56. Allied Publishers, New Delhi, 2003.

[49] H. G. Feichtinger and P. Gröbner. Banach spaces of distributions defined by decomposition methods. I. *Math. Nachr.*, 123:97–120, 1985.

[50] H. G. Feichtinger and K. Gröchenig. A unified approach to atomic decompositions via integrable group representations. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 52–73. Springer, Berlin, 1988.

[51] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. I. *J. Funct. Anal.*, 86(2):307–340, 1989.

[52] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. II. *Monatsh. Math.*, 108(2-3):129–148, 1989.

[53] A. Feuer and A. Nemirovski. On sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 49(6):1579–1581, 2003.

[54] D. Field and B. Olshausen. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[55] M. Fornasier. Banach frames for α-modulation spaces. *Appl. Comput. Harmon. Anal.*, 22(2):157–174, 2007.

[56] M. Frazier and B. Jawerth. Decomposition of Besov spaces. *Indiana Univ. Math. J.*, 34(4):777–799, 1985.

[57] M. Frazier and B. Jawerth. A discrete transform and decompositions of distribution spaces. *J. Funct. Anal.*, 93(1):34–170, 1990.

[58] J. J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Trans. Inform. Theory*, 51(10):3601–3608, 2005.

[59] G. Garrigós and E. Hernández. Sharp Jackson and Bernstein inequalities for N-term approximation in sequence spaces with applications. *Indiana Univ. Math. J.*, 53(6):1739–1762, 2004.

[60] R. Gribonval. Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture. In *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP02)*, Orlando, FL., May 2002.

[61] R. Gribonval, R. Figueras i Ventura, and P. Vandergheynst. A simple test to check the optimality of a sparse signal approximation. *Signal Processing*, 86(3):496–510, 2006.

[62] R. Gribonval and M. Nielsen. Sparse approximations in signal and image processing (editorial). *Signal Process.*, 86(3):415–416, 2006.

[63] P. Gröbner. *Banachräume glatter Funktionen und Zerlegungsmethoden*. PhD thesis, University of Vienna, 1992.

[64] K. Gröchenig. Describing functions: atomic decompositions versus frames. *Monatsh. Math.*, 112(1):1–42, 1991.

[65] K. Gröchenig. Localized frames are finite unions of Riesz sequences. *Adv. Comput. Math.*, 18(2-4):149–157, 2003. Frames.

[66] K. Gröchenig. Localization of frames, Banach frames, and the invertibility of the frame operator. *J. Fourier Anal. Appl.*, 10(2):105–132, 2004.

[67] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–331, 2000.

[68] A. Grossman and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.*, 15:723–736, 1984.

[69] R. S. Ismagilov. Diameters of sets in normed linear spaces, and the approximation of functions by trigonometric polynomials. *Uspehi Mat. Nauk*, 29(3(177)):161–178, 1974.

[70] R. Q. Jia. A Bernstein-type inequality associated with wavelet decomposition. *Constr. Approx.*, 9(2-3):299–318, 1993.

[71] B. Karaivanov and P. Petrushev. Nonlinear piecewise polynomial approximation beyond Besov spaces. *Appl. Comput. Harmon. Anal.*, 15(3):177–223, 2003.

[72] G. Kerkyacharian and D. Picard. Entropy, universal coding, approximation, and bases properties. *Constr. Approx.*, 20(1):1–37, 2004.

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

[74] K. Kreutz-Delgado, B. Rao, K. Engan, and T.-W. Lee. Convex/Schur-convex (CSC) log-priors and sparse coding. In *6th Joint Symposium on Neural Computation, Institute for Neural Computation*, pages 65–71.

[75] G. Kyriazis. Non-linear approximation and interpolation spaces. *J. Approx. Theory*, 113(1):110–126, 2001.

[76] G. Kyriazis. Decomposition systems for function spaces. *Studia Math.*, 157(2):133–169, 2003.

[77] G. Kyriazis and P. Petrushev. New bases for Triebel-Lizorkin and Besov spaces. *Trans. Amer. Math. Soc.*, 354(2):749–776 (electronic), 2002.

[78] G. Kyriazis and P. Petrushev. On the construction of frames for Triebel-Lizorkin and Besov spaces. *Proc. Amer. Math. Soc.*, 134(6):1759–1770 (electronic), 2006.

[79] E. Laeng. Une base orthonormale de $L^2(\mathbf{R})$ dont les éléments sont bien localisés dans l'espace de phase et leurs supports adaptés à toute partition symétrique de l'espace des fréquences. *C. R. Acad. Sci. Paris Sér. I Math.*, 311(11):677–680, 1990.

[80] H. Malvar. Lapped transforms for efficient transform/subband coding. *IEEE Trans. Acoustics Speech Signal and Processing*, 38:969–978, 1990.

[81] F. G. Meyer and R. R. Coifman. Brushlets: a tool for directional image analysis and image compression. *Appl. Comput. Harmon. Anal.*, 4(2):147–187, 1997.

[82] Y. Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.

[83] M. Nielsen. An example of an almost greedy uniformly bounded orthonormal basis for $L_p(0,1)$. *J. Approx. Theory (to appear)*, 2007.

[84] M. Nielsen. Trigonometric quasi-greedy bases for $L_p(\mathbb{T};w)$. *Rocky Mountain J. Math. (to appear)*, 2007.

[85] K. I. Oskolkov. Polygonal approximation of functions of two variables. *Mat. Sb. (N.S.)*, 107(149)(4):601–612, 639, 1978.

[86] L. Päivärinta and E. Somersalo. A generalization of the Calderón-Vaillancourt theorem to L^p and h^p. *Math. Nachr.*, 138:145–156, 1988.

[87] A. A. Pekarskiĭ. Chebyshev rational approximation in a disk, on a circle and on a segment. *Mat. Sb. (N.S.)*, 133(175)(1):86–102, 144, 1987.

[88] P. Petrushev. Bases consisting of rational functions of uniformly bounded degrees or more general functions. *J. Funct. Anal.*, 174(1):18–75, 2000.

[89] P. Petrushev. Nonlinear n-term approximation from hierarchical spline bases. In *Constructive theory of functions*, pages 33–85. DARBA, Sofia, 2003.

[90] P. P. Petrushev. Direct and converse theorems for spline and rational approximation and Besov spaces. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 363–377. Springer, Berlin, 1988.

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

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

[93] E. Schmidt. Zur Theorie der linearen und nicht linearen Integralgleichungen Zweite Abhandlung. *Math. Ann.*, 64(2):161–174, 1907.

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

[95] V. N. Temlyakov. The best m-term approximation and greedy algorithms. *Adv. Comput. Math.*, 8(3):249–265, 1998.

[96] V. N. Temlyakov. Greedy algorithm and m-term trigonometric approximation. *Constr. Approx.*, 14(4):569–587, 1998.

[97] V. N. Temlyakov. Non-linear m-term approximation with regard to the multivariate Haar system. *East J. Approx.*, 4(1):87–106, 1998.

[98] V. N. Temlyakov. Nonlinear methods of approximation. *Found. Comput. Math.*, 3(1):33–107, 2003.

[99] H. Triebel. *Theory of function spaces*, volume 78 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1983.

[100] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[101] J. A. Tropp. Recovery of short, complex linear combinations via l_1 minimization. *IEEE Trans. Inform. Theory*, 51(4):1568–1570, 2005.

[102] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, 52(3):1030–1051, 2006.

[103] M. V. Wickerhauser. Smooth localized orthonormal bases. *C. R. Acad. Sci. Paris Sér. I Math.*, 316(5):423–427, 1993.

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

[105] M. Zibulevsky and B. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Comput.*, 13(4):863–882, 2001.

Dansk resumé

Denne afhandling giver bidrag til teorien om ikke-lineær approksimation med overkomplette funktionsbiblioteker. Den grundlæggende idé er, at approksimere en funktion (et signal) med m-leds partialsummer dannet udfra et funktionsbibliotek. Ved at gøre funktionsbiblioteket større end en basis kan man opnå den fordel at klassen af funktioner, der kan approksimeres med en given effektivitet, er større end en tilsvarende klasse opnået ved approksimation med et lineært uafhængigt funktionsbibliotek.

Afhandlings resultater falder indenfor 4 hovedområder. I det første undersøges approksimation med generelle funktionsbiblioteker i et Banachrum. Med minimale antagelser om struktur af funktionsbiblioteket udledes en række estimater for ikke-lineær approksimation. Der gives samtidig en række eksempler på klassiske estimater for ikke-redundante funktionsbiblioteker, som påvises ikke at være sande for tilsvarende overkomplette biblioteker.

I anden del er emnet approksimation med wavelet frames. Wavelet frames er overkomplette systemer, der har samme basale struktur som orthonormale wavelets. For sådanne frames gives en komplet karakterisation af approksimationsegenskaberne i L_p. Ligeledes betragtes konstruktive algoritmer til at opnå tyndt besatte wavelet frame representationer af signaler.

Tredie del omhandler approksimation med tids-frekvens biblioteker. For en generel klasse af glathedsrum konstrueres tilhørende tids-frekvens frames, der benyttes til at give en komplet karakterisation af glathedsrummene. Denne karakterisation leder naturligt til Jackson estimater for ikke-linær approksimation med tids-frekvens framen. Specielt behandles approksimation i såkaldte α-modulationsrum.

I den sidste del behandles representation af signaler i et endeligdimensionalt vektorrum relativt til et overkomplet funktionsbibliotek. Specielt studeres spørgsmålet om at finde tyndt besatte representationer af signaler. Dette kan anskues som et optimeringsproblem, og der opstilles tilstrækkelige betingelser som sikrer at en tyndt besat representation af et givet signal kan beregnes effektivt.