Abstract

We present a new method for performing nonlinear approximation with redundant dictionaries. The method constructs an $m$−term approximation of the signal by thresholding with respect to a weighted version of its canonical expansion coefficients, thereby accounting for dependency between the coefficients. The main result is an associated strong Jackson embedding, which provides an upper bound on the corresponding reconstruction error. To complement the theoretical results, we compare the proposed method to the pure greedy method and the Windowed-Group Lasso by denoising music signals with elements from a Gabor dictionary.

Keywords: Weighted thresholding, nonlinear approximation, time-frequency analysis, Gabor frames, modulation spaces, social sparsity.

1. Introduction

Let $X$ be a Banach space equipped with a norm $\| \cdot \|_X$. We consider the problem of approximating a possibly complicated function $f \in X$ using linear combinations of simpler functions $\mathcal{D} := \{g_k\}_{k \in \mathbb{N}}$. We assume $\mathcal{D}$ forms a complete dictionary for $X$ such that $\|g_k\|_X = 1$, for all $k \in \mathbb{N}$, and
span\{g_k\}_{k \in \mathbb{N}} \text{ is dense in } X. \text{ A natural way of performing the approximation}

is to construct an \( m \)-term approximation \( f_m \) to \( f \) using a linear combination

of at most \( m \) elements from \( \mathcal{D} \) \cite{1, 2, 3}. \text{ This leads us to consider the set}

\[
\Sigma_m(\mathcal{D}) := \left\{ \sum_{k \in \Delta} c_k g_k \bigg| \Delta \subset \mathbb{N}, |\Delta| \leq m \right\}, \quad m \in \mathbb{N}.
\]

We note that \( \Sigma_m(\mathcal{D}) \) is nonlinear since a sum of two elements from \( \Sigma_m(\mathcal{D}) \)

will in general need \( 2^m \) terms in its representation by \( \{g_k\}_{k \in \mathbb{N}} \). \text{ We measure the approximation error associated to } \Sigma_m(\mathcal{D}) \text{ by}

\[
\sigma_m(f, \mathcal{D})_X := \inf_{h \in \Sigma_m(\mathcal{D})} \|f - h\|_X, \quad f \in X. \tag{1}
\]

One of the main challenges of nonlinear approximation theory is to characterize the elements \( f \in X \), which have a prescribed rate of approximation \( \alpha > 0 \) \cite{4, 5, 6}. \text{ This is usually done by defining an approximation space} \( \mathcal{A} \subseteq X \) with the property

\[
\sigma_m(f, \mathcal{D})_X = \mathcal{O} \left( m^{-\alpha} \right), \quad \forall f \in \mathcal{A}. \tag{2}
\]

It is often difficult to characterize the elements of \( \mathcal{A} \) directly and a standard

approach is therefore to construct a simpler space \( \mathcal{K} \subseteq X \) together with a continuous embedding \( \mathcal{K} \hookrightarrow \mathcal{A} \). The space \( \mathcal{K} \) is referred to as a smoothness or sparseness space and the continuous embedding as a Jackson embedding \cite{7, 8}.

In the special case \( \{g_k\}_{k \in \mathbb{N}} \) forms an orthonormal basis for a Hilbert space \( \mathcal{H} \), it follows from Parseval’s identity that the best \( m \)-term approximation to \( f \) is obtained by thresholding the (unique) expansion coefficients and keeping only the \( m \) largest coefficients. \text{ For a redundant dictionary} \( \mathcal{D} \), \text{ the problem of constructing the best} \( m \)-term approximation is in general computationally intractable \cite{9}. For this reason, various algorithms have been constructed to produce fast and good (but not necessarily the best) \( m \)-term approximations \cite{10, 11, 12}. \text{ By ”good approximations” we mean an algorithm } A_m : f \rightarrow f_m \text{ for which there exists an approximation space } \mathcal{T} \subset X \text{ with}

\[
\|f - f_m\|_X = \mathcal{O} \left( m^{-\alpha} \right), \quad \forall f \in \mathcal{T}.
\]

An associated embedding of the type \( \mathcal{K} \hookrightarrow \mathcal{T} \) is referred to as a strong Jackson embedding. \text{ Whereas a standard Jackson embedding only shows that}
the error of best \( m \)-term approximation decays as in (2), a strong Jackson embedding also provides an associated constructive algorithm with this rate of approximation [12].

Traditionally, the expansion coefficients are processed individually with an implicit assumption of independency between the coefficients [11]. However, for many signals such an assumption is not valid as the coefficients are often correlated and organised in structured sets [13, 14]. For instance, it is known that many music signals are generated by components, which are sparse in either time or frequency [15], resulting in sparse and structured time-frequency representations (cf. Fig. 1 on page 14). In this article we present a thresholding algorithm, which incorporates a weight function to account for dependency between expansion coefficients. The main result is given in Theorem 3.1 and provides a strong Jackson embedding for the proposed method.

The idea of exploiting the structure of the expansion coefficients is somewhat similar to the one found in social sparsity [16, 17, 18]. Social sparsity can be seen as a generalization of the classical Lasso [19] (also known as basis pursuit denoising, [20]) where a weighted neighbourhood of each coefficient is considered for deciding whether or not to keep the coefficient. However, whereas social sparsity searches through the dictionary for sparse reconstruction coefficients, the proposed method considers weighted thresholding of the canonical frame coefficients. We compare the proposed algorithm to the Windowed-Group-Lasso (WGL) [14, 21] from social sparsity and the greedy thresholding approach [11, 12] from nonlinear approximation theory by denoising music signals expanded in a Gabor dictionary [22, 23].

The structure of this article is as follows. In Section 2 we introduce the necessary tools from nonlinear approximation theory and in Section 3 we present the proposed algorithm and prove Theorem 3.1. In Section 4 we provide the link to modulation spaces and Gabor frames and in Section 5 we describe the numerical experiments. Finally, in Section 6 we give the conclusions.

2. Elements from nonlinear approximation theory

In this section we define the concepts from nonlinear approximation theory that we will use throughout this article. We refer the reader to [8, 24, 25, 12] for further details. For \( \{a_m\}_{m \in \mathbb{N}} \subset \mathbb{C} \), we let \( \{a_m^*\}_{m \in \mathbb{N}} \) denote a non-decreasing rearrangement of \( \{a_m\}_{m \in \mathbb{N}} \) such that \( |a_m^*| \geq |a_{m+1}^*| \) for all \( m \in \mathbb{N} \).
Definition 2.1. Given \( \tau \in (0, \infty) \), \( q \in (0, \infty] \), we define the Lorentz space \( \ell^\tau_q \) as the collection of \( \{a_m\}_{m \in \mathbb{N}} \subset \mathbb{C} \) satisfying that

\[
\|\{a_m\}_{m \in \mathbb{N}}\|_{\ell^\tau_q} = \left\{ \begin{array}{ll}
\left( \sum_{m=1}^{\infty} \left[ m^{1/\tau} |a_m^*|^q \frac{1}{m} \right] \right)^{1/q}, & 0 < q < \infty \\
\sup_{m \geq 1} m^{1/\tau} |a_m^*|, & q = \infty
\end{array} \right.
\]

is finite.

We note that \( \| \cdot \|_{\ell^\tau_q} = \| \cdot \|_{\ell^\tau} \) for any \( \tau \in (0, \infty) \). The Lorentz spaces are rearrangement invariant (quasi-)Banach spaces (Banach spaces if \( 1 \leq q \leq \tau < \infty \)) satisfying the continuous embeddings

\[
\ell^\tau_{q_1} \hookrightarrow \ell^\tau_{q_2} \quad \text{if} \quad \tau_1 < \tau_2 \quad \text{or} \quad \tau_1 = \tau_2 \quad \text{and} \quad q_1 \leq q_2,
\]

see [26, 24] for details. Let \( \mathcal{D} := \{g_k\}_{k \in \mathbb{N}} \) be a complete dictionary for a Banach space \( X \) and define \( \{\sigma_m(f, \mathcal{D})_X\}_{m \in \mathbb{N}} \) as in (1). We will use the following approximation spaces [8].

Definition 2.2. Given \( \alpha \in (0, \infty) \), \( q \in (0, \infty] \), we define

\[
\mathcal{A}_q^\alpha(\mathcal{D}, X) := \left\{ f \in X \mid \|f\|_{\mathcal{A}_q^\alpha(\mathcal{D}, X)} := \|\{\sigma_m(f, \mathcal{D})_X\}_{m \in \mathbb{N}}\|_{\ell^{1/\alpha}_q} + \|f\|_X < \infty \right\}.
\]

The quantity \( \| \cdot \|_{\mathcal{A}_q^\alpha(\mathcal{D}, X)} \) forms a (quasi-)norm on \( \mathcal{A}_q^\alpha(\mathcal{D}, X) \), and from (3) we immediately obtain the continuous embeddings

\[
\mathcal{A}_q^{\alpha_1}(\mathcal{D}, X) \hookrightarrow \mathcal{A}_q^{\alpha_2}(\mathcal{D}, X) \quad \text{if} \quad \alpha_1 > \alpha_2 \quad \text{or} \quad \alpha_1 = \alpha_2 \quad \text{and} \quad q_1 \leq q_2.
\]

We note the \( f \in \mathcal{A}_q^{\alpha}(\mathcal{D}, X) \) implies the decay in (2) as desired. Following the approach in [11], we define smoothness spaces as follows.

Definition 2.3. Given \( \tau \in (0, \infty) \), \( q \in (0, \infty] \), \( M > 0 \), we let

\[
\mathcal{K}_q^\tau(\mathcal{D}, X, M) := \text{clos}_X \left\{ \sum_{k \in \Delta} c_k g_k \in X \mid \Delta \subset \mathbb{N}, |\Delta| < \infty, \|\{c_k\}_{k \in \Delta}\|_{\ell^\tau_q} \leq M \right\}.
\]

We then define \( \mathcal{K}_q^\tau(\mathcal{D}, X) := \bigcup_{M > 0} \mathcal{K}_q^\tau(\mathcal{D}, X, M) \) with

\[
|f|_{\mathcal{K}_q^\tau(\mathcal{D}, X)} := \inf \left\{ M > 0 \mid f \in \mathcal{K}_q^\tau(\mathcal{D}, X, M) \right\}.
\]
It can be shown that $|·|_{K^\tau_q(D,X)}$ is a (semi-quasi-)norm on $K^\tau_q(D,X)$ and a (quasi-)norm if $\tau \in (0,1)$.

**Remark 2.1.** We note that for general $D$ and $X$, $f \in K^\tau_q(D,X)$ does not imply the existence of \{c_k\}_{k \in \mathbb{N}} \in \ell^r_q$ with $f = \sum_{k \in \mathbb{N}} c_k g_k$. All realizations of $K^\tau_q(D,X)$ considered in this article will, however, guarantee the existence of such reconstruction coefficients (cf. Proposition 2.1 below).

**Example 2.1.** Let $B = \{g_k\}_{k \in \mathbb{N}}$ be an orthonormal basis for a Hilbert space $H$. Given $\alpha \in (0,\infty), q \in (0,\infty]$ and $0 < \tau = (\alpha + 1/2)^{-1} < 2$, we have the characterization

$$A^\alpha_q(B,H) = K^\tau_q(B,H) = \{f \in H \mid |f|_{K^\tau_q(D,X)} = \|\{\langle f,g_k\rangle\}_{k \in \mathbb{N}}\|_{\ell^r_q} < \infty\},$$

where the first equality is with equivalent (quasi-)norms [27, 11].

Denoting the space of finite sequences on $\mathbb{N}$ by $\ell^0$, we define the reconstruction operator $R : \ell^0 \to X$ by

$$R : \{c_k\}_{k \in \Delta} \to \sum_{k \in \Delta} c_k g_k, \quad \{c_k\}_{k \in \Delta} \in \ell^0. \quad (4)$$

Recalling that $\|g_k\|_X = 1$, for all $k \in \mathbb{N}$, we can extend this operator to a bounded operator from $\ell^1$ to $X$ since

$$\|R\{c_k\}_{k \in \Delta}\|_X \leq \sum_{k \in \Delta} |c_k| \|g_k\|_X = \|\{c_k\}_{k \in \Delta}\|_{\ell^1}, \quad \forall \{c_k\}_{k \in \Delta} \in \ell^0. \quad (5)$$

Following the approach in [12] we introduce the following class of dictionaries.

**Definition 2.4.** Let $D = \{g_k\}_{k \in \mathbb{N}}$ be a dictionary in a Banach space $X$. Given $\tau \in (0,\infty), q \in (0,\infty]$, we say that $D$ is $\ell^r_q$-hilbertian if the reconstruction operator $R$ given in (4) is bounded from $\ell^r_q$ to $X$.

It follows from (5) and (3) that every $D$ is $\ell^r_q$-hilbertian if $\tau < 1$. According to [12] Proposition 3 we have the following characterization.
Proposition 2.1. Assume $\mathcal{D}$ is $\ell^p_1$-hilbertian with $p \in (1, \infty)$. Let $\tau \in (0, p)$ and $q \in (0, \infty]$. For all $f \in \mathcal{K}_q^\tau(\mathcal{D}, X)$, there exists some $c := c_{\tau,q}(f) \in \ell_q^\tau$ with $f = Rc$ and $\|c\|_{\ell_q^\tau} = |f|_{\mathcal{K}_q^\tau(\mathcal{D}, X)}$. If $1 < q \leq \tau < \infty$, then $c$ is unique. Consequently

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D}, X)} = \min_{c \in \ell_q^\tau, f = Rc} \|c\|_{\ell_q^\tau},$$

and

$$\mathcal{K}_q^\tau(\mathcal{D}, X) = \left\{ \sum_{k \in \mathbb{N}} c_k g_k \in X \left| \|\{c_k\}_{k \in \mathbb{N}}\|_{\ell_q^\tau} < \infty \right. \right\}$$

is a (quasi-)Banach space with $\mathcal{K}_q^\tau(\mathcal{D}, X) \hookrightarrow X$.

In the next section we describe the proposed algorithm using the framework presented in this section.

3. Weighted thresholding

Let $\mathcal{D} := \{g_k\}_{k \in \mathbb{N}}$ be a complete dictionary for a Banach space $X$ and let $R$ be the reconstruction operator defined in (4). Given $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$, we let $\pi : \mathbb{N} \rightarrow \mathbb{N}$ denote a bijective mapping such that $\{|c_{\pi(k)}|\}_{k \in \mathbb{N}}$ is non-increasing, i.e. $\{|c_{\pi(k)}|\}_{k \in \mathbb{N}} = \{c^*_k\}_{k \in \mathbb{N}}$. For $f \in X$, $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$, and $m \in \mathbb{N}$, a standard way of constructing an $m$-term approximant to $f$ from $\mathcal{D}$ is by thresholding

$$f_m := f_m(\pi, \{c_k\}_{k \in \mathbb{N}}, \mathcal{D}) := R\{c_{\pi(k)}\}_{k=1}^m = \sum_{k=1}^m c_{\pi(k)} g_{\pi(k)}. \quad (6)$$

For all practical purpose we choose $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$ as reconstruction coefficients for $f$ such that $f = R\{c_k\}_{k \in \mathbb{N}}$ (cf. Remark 2.1). With this choice, the approximants $\{f_m\}_{m \in \mathbb{N}}$ converge to $f$ as $m \rightarrow \infty$.

The thresholding approach in (6) chooses the $m$ elements from $\mathcal{D}$ corresponding to the $m$ largest of the coefficients $\{c_k\}_{k \in \mathbb{N}}$. As mentioned in the introduction, many real world signals have an inherent structure between the expansion coefficients, which should be accounted for in the approximation procedure. Therefore, we would like to construct an algorithm which preserves local coherence, such that a small coefficient $c_1$ might be preserved, in exchange for a larger (isolated) coefficient $c_2$, if $c_1$ belongs to a neighbourhood with many large coefficients. This leads us to consider banded Toeplitz matrices.
Let $\Lambda$ denote a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$, i.e., the non-zero entries $\Lambda_{(i,j)}$ satisfy $|i - j| \leq \Omega$. We assume the scalar on the diagonal of $\Lambda$ is positive and we define $\{c_k^\Lambda\}_{k \in \mathbb{N}} := \Lambda(\{c_k\}_{k \in \mathbb{N}})$ for $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$.

**Example 3.1.** With $\Omega = 2$ we obtain

$$
\{c_k^\Lambda\}_{k \in \mathbb{N}} = \begin{bmatrix}
\lambda_0 & \lambda_1 & \lambda_2 & 0 & 0 & 0 & \cdots \\
\lambda_{-1} & \lambda_0 & \lambda_1 & \lambda_2 & 0 & 0 & \cdots \\
\lambda_{-2} & \lambda_{-1} & \lambda_0 & \lambda_1 & \lambda_2 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix} \begin{bmatrix}
|c_1|
|c_2|
|c_3|
|c_4|
\vdots
\end{bmatrix},
$$

with $\lambda_0 > 0$ and $\lambda_l \geq 0$ for all $l \in \{-2, -1, 1, 2\}$. We note that

$$
c_k^\Lambda = \lambda_{-2} |c_{k-2}| + \cdots + \lambda_0 |c_k| + \cdots + \lambda_2 |c_{k+2}| = \sum_{j=k-\Omega}^{k+\Omega} \lambda_{j-k} |c_j|,
$$

for all $k \in \mathbb{N}$.

Given $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$, we let $\pi_\Lambda : \mathbb{N} \to \mathbb{N}$ denote a bijective mapping such that the sequence $\{c_{\pi_\Lambda(k)}^\Lambda\}_{k \in \mathbb{N}}$ is non-increasing. If $\Lambda$ is the identity map, we write $\pi$ instead of $\pi_\Lambda$ to be consistent with the notation of (6). We shall use the following technical result in the proof of Theorem 3.1.

**Lemma 3.1.** Let $\Lambda$ be a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$ and let $p \in (0, \infty)$, $q \in (0, \infty]$. There exists a constant $C > 0$, such that for any non-increasing sequence $\{c_k\}_{k \in \mathbb{N}} \in \ell^p_q$ we have the estimate

$$
\|\{c_{\pi_\Lambda(k)}\}_{k=m+1}^\infty\|_{\ell^p_q} \leq C \|\{c_k\}_{k=m+1-\Omega}^\infty\|_{\ell^p_q}, \quad \forall m \geq \Omega.
$$

**Proof.** Fix $m \in \mathbb{N}$ and let $a := \{c_{\pi_\Lambda(k)}\}_{k=m+1}^\infty$ and $b := \{c_k\}_{k=m+1}^\infty$. If $a$ contains the same coefficients as $b$, then there is nothing to prove. If this is not the case, then there exists $k' \leq m$, with $|c_{k'}| \in a$, and $k'' \geq m + 1$, with $|c_{k''}| \notin a$, satisfying

$$
\sum_{j=k'-\Omega}^{k'+\Omega} \lambda_{j-k'} |c_j| \leq \sum_{j=k''-\Omega}^{k''+\Omega} \lambda_{j-k''} |c_j| \Rightarrow |c_{k'}| \leq \frac{1}{\lambda_0} \sum_{j=k''-\Omega}^{k''+\Omega} \lambda_{j-k''} |c_j|.
$$
Denoting the maximum value of the $\lambda_i$'s by $\lambda_{\text{max}}$ and the maximum value of the $|c_j|$'s, for $j \in [k'' - \Omega, k'' + \Omega]$, by $|c_{\tilde{k}}|$, we thus get

$$|c_{k'}| \leq \frac{\lambda_{\text{max}}}{\lambda_0} (2\Omega + 1) |c_{\tilde{k}}|.$$ 

Since $\tilde{k} \in [k'' - \Omega, k'' + \Omega]$, and $k'' \geq m + 1$, then $\tilde{k} \in [m + 1 - \Omega, \infty)$. We conclude that $c_{\tilde{k}} \in \{c_k\}_{k=m+1-\Omega}^\infty$. The lemma follows directly from this observation.

Given $f \in X$, a set of reconstruction coefficients $\{c_k\}_{k \in \mathbb{N}} \subset \mathbb{C}$, and $m \in \mathbb{N}$, we generalize the notation of (6) and construct an $m$-term approximant to $f$ by

$$f_m^\Lambda := f_m^\Lambda(\pi_\Lambda, \{c_k\}_{k \in \mathbb{N}}, \mathcal{D}) := R\{c_{\pi_\Lambda(k)}\}_{k=1}^m = \sum_{k=1}^m c_{\pi_\Lambda(k)} g_{\pi_\Lambda(k)}. \quad (7)$$

If $\Lambda$ is the identity map, we just obtain the approximant $f_m$ given in (6). If not, we obtain an approximant which chooses the elements of $\{g_k\}_{k \in \mathbb{N}}$ corresponding to the indices of the $m$ largest of the weighted coefficients $\{c_k^\Lambda\}_{k \in \mathbb{N}}$. We generalize the notation of [12] and define weighted thresholding approximation spaces as follows.

**Definition 3.1.** Let $\Lambda$ be a banded non-negative Toeplitz matrix with band-width $\Omega \in \mathbb{N}$. Given $\alpha \in (0, \infty), q \in (0, \infty]$, we define

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

with

$$|f|_{\mathcal{T}_q^\alpha(\mathcal{D}, X, \Lambda)} := \inf_{\pi_\Lambda, \{c_k\}_{k \in \mathbb{N}}} \left\{ \begin{array}{ll}
\sum_{m=1}^\infty m^\alpha \left( \frac{m}{m-1} \right) \left( \sup_{m \geq 1} m^\alpha \left( \frac{\|f - f_m^\Lambda\|_X}{m} \right) \right)^q, & 0 < q < \infty \\
\inf_{\pi_\Lambda, \{c_k\}_{k \in \mathbb{N}}} \left( \sup_{m \geq 1} m^\alpha \left( \frac{\|f - f_m^\Lambda\|_X}{m} \right) \right)^q, & q = \infty
\end{array} \right..$$

When $\Lambda$ is the identity mapping, we write $\mathcal{T}_q^\alpha(\mathcal{D}, X)$ instead of $\mathcal{T}_q^\alpha(\mathcal{D}, X, \Lambda)$.

**Remark 3.1.** We note that the expression for $|f|_{\mathcal{T}_q^\alpha(\mathcal{D}, X, \Lambda)}$ cannot be written using the Lorentz norm, since the sequence $\{\|f - f_m^\Lambda\|_X\}_{m \in \mathbb{N}}$ might not be non-increasing.

8
In order to prove Theorem 3.1 we need to impose further assumptions on the dictionary \( \mathcal{D} \). We call \( \mathcal{D} \) an atomic decomposition (AD) \([28, 29]\) for \( X \), with respect to \( \ell^\tau_q \), if there exists a sequence \( \{\tilde{g}_k\}_{k \in \mathbb{N}} \), in the dual space \( X' \), such that

1. There exist \( 0 < C' \leq C'' < \infty \) with
   \[
   C' \|\{(f, \tilde{g}_k)\}_{k \in \mathbb{N}}\|_{\ell^r_q} \leq \|f\|_X \leq C'' \|\{(f, \tilde{g}_k)\}_{k \in \mathbb{N}}\|_{\ell^r_q}, \quad \forall f \in X.
   \]

2. The reconstruction operator \( R \) given in (4) is bounded from \( \ell^r_q \) onto \( X \) and we have the expansions
   \[
   R(\{(f, \tilde{g}_k)\}_{k \in \mathbb{N}}) = \sum_{k \in \mathbb{N}} (f, \tilde{g}_k) g_k = f, \quad \forall f \in X.
   \]

**Example 3.2.** Standard examples of ADs are Gabor frames for modulation spaces \([22]\) and wavelets for Besov spaces \([6]\). However, many other examples have been constructed in the general framework of decomposition spaces \([30, 31, 32]\). For instance curvelets, shearlets and nonstationary Gabor frames (or generalized shift-invariant systems) \([32, 33, 34]\).

We can now present the main result of this article.

**Theorem 3.1.** Let \( \Lambda \) be a banded non-negative Toeplitz matrix with bandwidth \( \Omega \in \mathbb{N} \) and let \( p \in (1, \infty) \), \( \tau \in (0, p) \), \( q \in (0, \infty) \). If \( \mathcal{D} = \{g_k\}_{k \in \mathbb{N}} \) is \( \ell^p_1 \)-hilbertian and forms an AD for \( X \), with respect to \( \ell^r_q \), then
   \[
   K^r_q(\mathcal{D}, X) \hookrightarrow T^\alpha_q(\mathcal{D}, X, \Lambda) \hookrightarrow A^\alpha_q(\mathcal{D}, X),
   \]
   with \( \alpha = 1/\tau - 1/p > 0 \).

**Proof.** The embedding \( T^\alpha_q(\mathcal{D}, X, \Lambda) \hookrightarrow A^\alpha_q(\mathcal{D}, X) \) follows directly from the definitions of these spaces. Let us now show that \( K^r_q(\mathcal{D}, X) \hookrightarrow T^\alpha_q(\mathcal{D}, X, \Lambda) \). Since the Lorentz spaces are rearrangement invariant, we may assume the canonical coefficients \( d := \{d_k\}_{k \in \mathbb{N}} := \{(f, \tilde{g}_k)\}_{k \in \mathbb{N}} \) form a non-increasing sequence. Given \( f \in K^r_q(\mathcal{D}, X) \), the \( \ell^p_1 \)-hilbertian property of \( \mathcal{D} \) implies
   \[
   \|f\|_X = \|Rd\|_X \leq C_1 \|d\|_{\ell^p_1}.
   \]
Defining \( f_m^A(\pi_A, d, \mathcal{D}) = R\{d_{\pi_A(k)}\}_{k=1}^m \) as in (7), we thus get
\[
\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D}, X, \Lambda)} = \inf_{\pi_A \in \mathcal{C}_q} \left( \sum_{m=1}^{\infty} \left[ m^\alpha \|f - f_m^A\|_X \right]^q \frac{1}{m} \right)^{1/q} + \|f\|_X \\
\leq \left( \sum_{m=1}^{\infty} \left[ m^\alpha \|f - f_m^A(\pi_A, d, \mathcal{D})\|_X \right]^q \frac{1}{m} \right)^{1/q} + C_1 \|d\|_{\ell_q^p}.
\] (8)

Now, since
\[
\|f - f_m^A(\pi_A, d, \mathcal{D})\|_X = \|R\{d_{\pi_A(k)}\}_{k=m+1}^\infty\|_X \\
\leq C_1 \|\{d_{\pi_A(k)}\}_{k=m+1}^\infty\|_{\ell_q^p} \leq C_1 \|d\|_{\ell_q^p},
\]
we get the following estimate for the first \( \Omega \) terms in (8)
\[
\sum_{m=1}^{\Omega} \left[ m^\alpha \|f - f_m^A(\pi_A, c, \mathcal{D})\|_X \right]^q \frac{1}{m} \leq C_2 \|d\|_{\ell_q^p}^q.
\] (9)

For \( m \geq \Omega + 1 \), Lemma 3.1 implies
\[
\|f - f_m^A(\pi_A, d, \mathcal{D})\|_X \leq C_1 \|\{d_{\pi_A(k)}\}_{k=m+1}^\infty\|_{\ell_q^p} \leq C_3 \|\{d_k\}_{k=m+1}^\infty\|_{\ell_q^p} \\
= C_3 \|d - \{d_k\}_{k=1}^{m-\Omega}\|_{\ell_q^p} = C_3 \sigma_{m-\Omega}(d, \mathcal{B})_{\ell_q^p},
\]
with \( \mathcal{B} \) denoting the canonical basis of \( \ell_q^p \). Hence,
\[
\sum_{m=\Omega+1}^{\infty} \left[ m^\alpha \|f - f_m^A(\pi_A, d, \mathcal{D})\|_X \right]^q \frac{1}{m} \leq C_3 \sum_{m=\Omega+1}^{\infty} \left[ m^\alpha \sigma_{m-\Omega}(d, \mathcal{B})_{\ell_q^p} \right]^q \frac{1}{m} \\
= C_3 \sum_{m=\Omega+1}^{\infty} \left[ \frac{(m + \Omega)^\alpha \sigma_m(d, \mathcal{B})_{\ell_q^p} \right]^q \frac{1}{m + \Omega} \\
\leq C_4 \|\{\sigma_m(d, \mathcal{B})_{\ell_q^p}\}_{m=\Omega}^q \|_{\ell_q^{\sigma_q}(\mathcal{B}, \ell_q^p)}. \] (10)

Combining (9) and (10), then (8) yields
\[
\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D}, X, \Lambda)} \leq C_5 \left( \|\{\sigma_m(d, \mathcal{B})_{\ell_q^p}\}_{m=\Omega}^q \|_{\ell_q^{\sigma_q}(\mathcal{B}, \ell_q^p)} + \|d\|_{\ell_q^p} \right) = C_5 \|d\|_{\mathcal{A}^q_\alpha(\mathcal{B}, \ell_q^p)}.
\]

Applying Theorem 3.1, we thus get
\[
\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D}, X, \Lambda)} \leq C_6 |d|_{\mathcal{K}_{q}^\alpha(\mathcal{B}, \ell_q^p)} = C_6 \|d\|_{\ell_q^p} \leq C_7 \|f\|_X.
\] (11)
Finally, since $\mathcal{D}$ is $\ell_1^p$-hilbertian, Proposition 2.1 states that we can find $c \in \ell_\tau^q$ with $f = R c$ and $\|c\|_{\ell_\tau^q} = |f|_{K^\tau_\ell(D,X)}$, such that (11) yields

$$\|f\|_{T_\alpha^q(D,X,\Lambda)} \leq C_7 \|R c\|_X \leq C_8 \|c\|_{\ell_\tau^q} = C_8 |f|_{K^\tau_\ell(D,X)}.$$  

This completes the proof. \square

Remark 3.2. The assumption in Theorem 3.1 of $\mathcal{D}$ being $\ell_1^p$-hilbertian directly implies the boundedness of $R : \ell_\tau^q \rightarrow X$ in the definition of an AD since $\tau \in (0,p)$. It should also be noted that if $\Lambda$ is the identity operator, then Theorem 3.1 holds without the assumption of an AD — this was proven in [12, Theorem 6]. However, in contrast to the proof presented here, there is no constructive way of obtaining the sparse expansion coefficients in the proof of [12, Theorem 6].

The Jackson embedding in Theorem 3.1 is strong in the sense that it provides an associated constructive algorithm which obtains the approximation rate. Given $f \in X$, the algorithm goes as follows:

1. Calculate the canonical coefficients $\{d_k\}_{k \in \mathbb{N}} = \{\langle f, \tilde{g}_k \rangle\}_{k \in \mathbb{N}}$.
2. Construct the weighted coefficients $\{d_k^\Lambda\}_{k \in \mathbb{N}}$ according to $\Lambda$ by

$$d_k^\Lambda = \sum_{j=k-\Omega}^{k+\Omega} \lambda_{j-k} |d_j|, \quad k \in \mathbb{N}.$$  

3. Choose $\pi_\Lambda : \mathbb{N} \rightarrow \mathbb{N}$ such that $\{d_k^\Lambda\}_{k \in \mathbb{N}}$ is non-increasing.
4. Construct an $m$-term approximation to $f$ by

$$f_m^\Lambda(\pi_\Lambda, \{d_k\}_{k \in \mathbb{N}}, \mathcal{D}) = \sum_{k=1}^{m} d_{\pi_\Lambda(k)}^\Lambda g_{\pi_\Lambda(k)}.$$  

With this construction, Theorem 3.1 states that

$$\|f - f_m^\Lambda(\pi_\Lambda, \{d_k\}_{k \in \mathbb{N}}, \mathcal{D})\|_X = \mathcal{O}(m^{-\alpha}), \quad \forall f \in K^\tau_\ell(D,X),$$  

with $\alpha = 1/\tau - 1/p > 0$. In the next section we consider the particular case where the dictionary is a Gabor frame and the smoothness space is a modulation space.
4. Modulation spaces and Gabor frames

In this section we choose $X$ as the modulation space $M^p := M^p(\mathbb{R}^d)$, with $1 \leq p < \infty$, consisting of all $f \in S'(\mathbb{R}^d)$ satisfying

$$
\|f\|_{M^p} := \left( \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_\gamma f(x, y)|^p \, dx \, dy \right)^{1/p} < \infty.
$$

Here, $V_\gamma f(x, y)$ denotes the short-time Fourier transform of $f$ with respect to a window function $\gamma \in S(\mathbb{R}^d) \setminus \{0\}$. It can be shown that $M^p$ is independent of the particular choice of window function and different choices yield equivalent norms [22]. Given $g \in S(\mathbb{R}^d) \setminus \{0\}$ and lattice parameters $a, b \in (0, \infty)$, we consider the Gabor system

$$
g_{m,n}(t) := g(t - na) e^{2\pi imb \cdot t}, \quad t \in \mathbb{R}^d.
$$

Assuming $\{g_{m,n}\}_{m,n}$ forms a frame for $L^2(\mathbb{R}^d)$ (see [23] for details), there exists a dual frame $\{\tilde{g}_{m,n}\}_{m,n}$ such that $\{g_{m,n}\}_{m,n} \in \mathbb{Z}^d$ is an AD for $M^p$ with respect to $\ell^p$ for all $1 \leq p < \infty$ [22]. We have the following version of [5, Proposition 3].

**Proposition 4.1.** Let $\Lambda$ be a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$. Let $1 \leq \tau < p < \infty$ and $g \in S(\mathbb{R}^d) \setminus \{0\}$. If $D = \{g_{m,n}\}_{m,n \in \mathbb{Z}^d}$ forms a Gabor frame for $L^2(\mathbb{R}^d)$ and $\|g_{m,n}\|_{M^p} = 1$ for all $m, n \in \mathbb{Z}^d$, then

$$
V_{\tau} = K_{\tau}^p(D, M^p) \hookrightarrow T_{\tau}^p(\Lambda, D, M^p) \hookrightarrow A_{\alpha}^p(D, M^p), \quad \alpha = 1/\tau - 1/p,
$$

where the first equality is with equivalent norms.

**Proof.** We first note that $D$ simultaneously forms an AD for both $M^p$ and $M^\tau$. Since $D$ constitutes an AD for $M^p$, and $p > 1$, we get

$$
\|Re\|_{M^p} \leq C_1 \|c\|_{\ell^p} \leq C_2 \|c\|_{\ell^p}, \quad \forall c \in \ell^p,
$$

which shows that $D$ is a $\ell^p$-hilbertian dictionary for $M^p$. Hence, the embeddings in Proposition 4.1 follows from Theorem 3.1. Let us now prove that $M^\tau = K_{\tau}^p(D, M^p)$ with equivalent norms. According to Proposition 2.1 then

$$
K_{\tau}^p(D, M^p) = \left\{ \sum_{m \in \mathbb{Z}^d} \sum_{n \in \mathbb{Z}^d} c_{m,n} g_{m,n} \in M^p \mid \|\{c_{m,n}\}_{m,n \in \mathbb{Z}^d}\|_{\ell^p} < \infty \right\}
$$

with

$$
|f|_{K_{\tau}^p(D, M^p)} = \min_{c \in \ell^p, f = Re} \|c\|_{\ell^p}, \quad f \in K_{\tau}^p(D, M^p).
$$

12
Since $\mathcal{D}$ constitutes an AD for $M^\tau$, then for $f \in M^\tau$ we have

$$f = \sum_{m \in \mathbb{Z}^d} \sum_{n \in \mathbb{Z}^d} \langle f, \tilde{g}_{m,n} \rangle g_{m,n}, \quad \text{with} \quad \left\| \{\langle f, \tilde{g}_{m,n} \rangle \}_{m,n \in \mathbb{Z}^d} \right\|_{\ell^\tau} \leq C \|f\|_{M^\tau}. $$

As $\tau < p$ then $M^\tau \subseteq M^p$ (cf. [22]), which shows that $f \in \mathcal{K}_\tau^p(\mathcal{D}, M^p)$ and $|f|_{\mathcal{K}_\tau^p(\mathcal{D}, M^p)} \leq C\|f\|_{M^\tau}$. The converse embedding follows from

$$\|f\|_{M^\tau} = \|\text{Re}\|_{M^\tau} \leq C'\|c\|_{\ell^\tau} = C'|f|_{\mathcal{K}_\tau^p(\mathcal{D}, M^p)}, \quad f \in \mathcal{K}_\tau^p(\mathcal{D}, M^p).$$

This completes the proof. \hfill \qed

In the next section we apply the proposed method for denoising music signals with elements from a Gabor dictionary. For an introduction to Gabor theory in the discrete settings, we refer the reader to [36, 37].

5. Numerical experiments

For the implementation we use MATLAB 2017B and apply the routines from the following two toolboxes: The ”Large time-frequency analysis toolbox” (LTFAT) version 2.2.0 [38] available from

http://ltfat.sourceforge.net/

and the StrucAudioToolbox [16] available from

http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html

All Gabor transforms are constructed using 1024 frequency channels, a hop size of 256, and a Hanning window of length 1024 (this is the default settings in the StrucAudioToolbox). These settings lead to transforms of redundancy of four, meaning there are four times as many time-frequency coefficients as signal samples. The music signals we consider are part of the EBU-SQAM database [39], which consists of 70 test sounds sampled at 44 kHz. The database contains a large variety of different music sounds including single instruments, vocal, and orchestra. We measure the reconstructing error of an algorithm by the the relative root mean square (RMS) error

$$\text{RMS}(f, f_{\text{rec}}) := \frac{\|f - f_{\text{rec}}\|_2}{\|f\|_2}. $$

13
We begin by analyzing the first 524288 samples of signal 8 in the EBU-SQAM database, which consists of an increasing melody of 10 tones played by a violin. A noisy version of the signal is constructed by adding white Gaussian noise and the resulting spectrograms can be found in Fig. 1.

For the task of denoising we first compare the greedy thresholding approach from nonlinear approximation theory (cf. (6)) with the Windowed-Group-Lasso (WGL) from social sparsity. For the WGL we use the default settings of the StrucAudioToolbox, which applies a horizontal asymmetric neighbourhood for the shrinkage operator, see [16] for further details. The WGL constructs a denoised version of the spectrogram using only 74739 non-zero coefficients (out of a total of 10506244 coefficients). Using the same number of non-zero coefficients for the greedy thresholding approach we obtain the spectrograms shown in Fig. 2.

We note from Fig. 2 that the greedy thresholding approach includes more coefficients at higher frequencies than the WGL. On the other hand, the WGL includes more coefficients at lower frequencies, resulting in a smoother resolution for the fundamental frequencies and the first harmonics. This illustrates the way the WGL is designed, namely that a large isolated coefficient may be discarded in exchange for a smaller coefficient with large neighbours. The RMS error is $\approx 0.084$ for the WGL and $\approx 0.034$ for the greedy thresholding algorithm. To visualize the performance of the proposed algorithm we choose a rather extreme (horizontal) weight with

$$c_{m,n}^\Lambda = |c_{m,n-2}| + |c_{m,n-1}| + |c_{m,n}| + |c_{m,n+1}| + |c_{m,n+2}|.$$  (12)
We then choose the 74739 coefficients with largest weighted magnitudes according to (12). In Fig. 3 we have compared the resulting spectrogram against the spectrogram obtained using the greedy thresholding approach.

We can see from Fig. 3 that the weight in (12) enforces the structures at higher frequencies even further than the greedy approach. This might be desirable for some applications as the timbre of the instrument is determined by the harmonics. The RMS error associated with the weighted thresholding approach is \( \approx 0.074 \), which is higher than for the greedy approach but lower
than for the WGL. Applying a more moderate weight (for instance Weight 2 defined in (13) below) we obtain an RMS error of $\approx 0.031$, which is lower than for the greedy approach.

We now extend the experiment described above to the entire EBU-SQAM database. For the weighted thresholding we consider the following three weights

Weight 1: $c_{m,n}^A = |c_{m,n}| + (|c_{m-1,n}| + |c_{m+1,n}|)/2$,

Weight 2: $c_{m,n}^A = |c_{m,n}| + (|c_{m,n-1}| + |c_{m,n+1}|)/2$,  \hspace{1cm} (13)

Weight 3: $c_{m,n}^A = |c_{m,n}| + (|c_{m,n-1}| + |c_{m-1,n}| + |c_{m,n+1}| + |c_{m+1,n}|)/4$.

For each of the 70 test sounds in the EBU-SQAM database, we apply the WGL and calculate the associated RMS error and number of non-zero coefficients. Using the same number of non-zero coefficients, we then apply the greedy thresholding approach and the three weighted thresholding algorithms defined in (13). The resulting averaged values can be found in Table 1.

| Algorithm: | WGL | Greedy | Weight 1 | Weight 2 | Weight 3 |
|-----------|-----|--------|----------|----------|----------|
| Average RMS error: | 0.1031 | 0.0462 | 0.0511 | 0.0453 | 0.0472 |

The average number of coefficients was 68983, which corresponds to 6.5% of the total number of coefficients. We note that the RMS error associated with the WGL is roughly twice as large as for the various thresholding algorithms. We also note that the smallest RMS error is obtained by the weighted thresholding approach, which applies a horizontal weight (Weight 2). This is likely due to the horizontal structure of the harmonics as seen in Fig. 1.

Let us mention that it might be possible to reduce the error for the WGL by tuning the parameters instead of using the default settings (cf. [21] for a detailed analysis of the parameter settings for the WGL). On the other hand, the same holds true for the thresholding algorithms. In the experiment described above we have used the same number of non-zero coefficients as was chosen by the WGL. This is not likely to be optimal for the thresholding algorithms since the optimal number of non-zero coefficients usually depends on the sparsity of the signal (few coefficients for sparse signals and vice versa). Finally, we have not addressed the audio quality of the denoised sounds. In general, it is very hard to decide which algorithm sounds "the best" as this
depends on the application and the subjective opinion of the listener. Let us however mention that there are indeed audiable differences between the algorithms. For the violin music in Fig. 1, the WGL does the best job of removing the noise, but at the price of a poor timbre of the resulting sound. As we include more coefficients at higher frequencies, the original timbre of the instrument improves together with an increase in noise.

6. Conclusion

We have presented a new thresholding algorithm and proven an associated strong Jackson embedding under rather general conditions. The algorithm extends the classical greedy approach by incorporating a weight function, which exploits the structure of the expansion coefficients. In particular, the algorithm applies to approximation in modulation spaces using Gabor frames. As an application we have considered the task of denoising music signals and compared the proposed method with the greedy thresholding approach and the WGL from social sparsity. The numerical experiments show that the proposed method can be used both for improving the time-frequency resolution and for reducing the RMS error compared to the other algorithms. The experiments also show that the performance of the algorithm depends crucially on the choice of weight function, which should be adapted to the particular signal class under consideration.

References

[1] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen, Math. Ann. 63 (4) (1907) 433–476. doi:10.1007/BF01449770, URL http://dx.doi.org/10.1007/BF01449770

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

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

[4] P. P. Petrushev, Direct and converse theorems for spline and rational approximation and Besov spaces, in: Function spaces and applications (Lund, 1986), Vol. 1302 of Lecture Notes in Math., Springer, Berlin,
[5] K. Gröchenig, S. Samarah, Nonlinear approximation with local Fourier bases, Constr. Approx. 16 (3) (2000) 317–331. doi:10.1007/s003659910014 URL http://dx.doi.org/10.1007/s003659910014

[6] R. A. DeVore, B. Jawerth, V. Popov, Compression of wavelet decompositions, Amer. J. Math. 114 (4) (1992) 737–785. doi:10.2307/2374796 URL http://dx.doi.org/10.2307/2374796

[7] P. L. Butzer, K. Scherer, Jackson and Bernstein-type inequalities for families of commutative operators in Banach spaces, J. Approximation Theory 5 (1972) 308–342, collection of articles dedicated to J. L. Walsh on his 75th birthday, III.

[8] R. A. DeVore, G. G. Lorentz, Constructive approximation, Vol. 303 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993. doi:10.1007/978-3-662-02888-9

[9] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, Constructive Approximation 13 (1) (1997) 57–98. doi:10.1007/BF02678430 URL http://dx.doi.org/10.1007/BF02678430

[10] C. Darken, M. Donahue, L. Gurvits, E. Sontag, Rate of approximation results motivated by robust neural network learning, in: Proceedings of the Sixth Annual Conference on Computational Learning Theory, COLT ’93, ACM, New York, NY, USA, 1993, pp. 303–309. doi:10.1145/168304.168357 URL http://doi.acm.org/10.1145/168304.168357

[11] R. A. DeVore, V. N. Temlyakov, Some remarks on greedy algorithms, Advances in Computational Mathematics 5 (1) (1996) 173–187. doi:10.1007/BF02124742 URL http://dx.doi.org/10.1007/BF02124742

[12] R. Gribonval, M. Nielsen, Nonlinear approximation with dictionaries I. Direct estimates, J. Fourier Anal. Appl. 10 (1) (2004) 51–71. doi:
[13] M. Kowalski, Sparse regression using mixed norms, Appl. Comput. Harmon. Anal. 27 (3) (2009) 303–324.
URL https://doi.org/10.1016/j.acha.2009.05.006

[14] M. Kowalski, B. Torrésani, Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients, Signal, Image and Video Processing 3 (3) (2009) 251–264. doi:10.1007/s11760-008-0076-1
URL https://doi.org/10.1007/s11760-008-0076-1

[15] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, M. E. Davies, Sparse representations in audio and music: From coding to source separation, Proceedings of the IEEE 98 (6) (2010) 995–1005. doi:10.1109/JPROC.2009.2030345

[16] K. Siedenburg, M. Dörfler, Structured sparsity for audio signals, in: Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11), Paris, France, 2011.

[17] M. Kowalski, B. Torrésani, Structured Sparsity: from Mixed Norms to Structured Shrinkage, in: R. Gribonval (Ed.), SPARS’09 - Signal Processing with Adaptive Sparse Structured Representations, Inria Rennes - Bretagne Atlantique, Saint Malo, France, 2009.
URL https://hal.inria.fr/inria-00369577

[18] M. Kowalski, K. Siedenburg, M. Dörfler, Social sparsity! neighborhood systems enrich structured shrinkage operators, IEEE Transactions on Signal Processing 61 (10) (2013) 2498–2511. doi:10.1109/TSP.2013.2250967

[19] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society (Series B) 58 (1996) 267–288.

[20] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159. doi:10.1137/S003614450037906X
URL http://dx.doi.org/10.1137/S003614450037906X
[21] K. Siedenburg, M. Dörfler, Audio denoising by generalized time-frequency thresholding, in: Proceedings of the AES 45th Conference on Applications of Time-Frequency Processing, Helsinki, Finland, 2012.

[22] K. Gröchenig, Foundations of time-frequency analysis, Applied and Numerical Harmonic Analysis, Birkhäuser Boston, Inc., Boston, MA, 2001. doi:10.1007/978-1-4612-0003-1
URL http://dx.doi.org/10.1007/978-1-4612-0003-1

[23] O. Christensen, An introduction to frames and Riesz bases, 2nd Edition, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, [Cham], 2016. doi:10.1007/978-3-319-25613-9
URL http://dx.doi.org/10.1007/978-3-319-25613-9

[24] R. A. DeVore, Nonlinear approximation, in: Acta numerica, 1998, Vol. 7 of Acta Numer., Cambridge Univ. Press, Cambridge, 1998, pp. 51–150. doi:10.1017/S0962492900002816
URL http://dx.doi.org/10.1017/S0962492900002816

[25] R. A. DeVore, Nonlinear approximation and its applications, in: Multiscale, nonlinear and adaptive approximation, Springer, Berlin, 2009, pp. 169–201.
URL https://doi.org/10.1007/978-3-642-03413-8_6

[26] M. a. J. Carro, J. A. Raposo, J. Soria, Recent developments in the theory of Lorentz spaces and weighted inequalities, Mem. Amer. Math. Soc. 187 (877) (2007) xii+128. doi:10.1090/memo/0877
URL http://dx.doi.org/10.1090/memo/0877

[27] S. B. Stechkin, On absolute convergence of orthogonal series, Dokl. Akad. Nauk SSSR (N.S.) 102 (1955) 37–40.

[28] H. G. Feichtinger, K. H. Gröchenig, Banach spaces related to integrable group representations and their atomic decompositions. I, J. Funct. Anal. 86 (2) (1989) 307–340. doi:10.1016/0022-1236(89)90055-4
URL http://dx.doi.org/10.1016/0022-1236(89)90055-4

[29] H. G. Feichtinger, K. H. Gröchenig, Banach spaces related to integrable group representations and their atomic decompositions. II, Monatsh. Math. 108 (2-3) (1989) 129–148. doi:10.1007/BF01308667
URL http://dx.doi.org/10.1007/BF01308667
[30] H. G. Feichtinger, P. Gröbner, Banach spaces of distributions defined by decomposition methods. I, Math. Nachr. 123 (1985) 97–120. doi:10.1002/mana.19851230110. URL http://dx.doi.org/10.1002/mana.19851230110

[31] H. G. Feichtinger, Banach spaces of distributions defined by decomposition methods. II, Math. Nachr. 132 (1987) 207–237. doi:10.1002/mana.19871320116. URL http://dx.doi.org/10.1002/mana.19871320116

[32] L. Borup, M. Nielsen, Frame decomposition of decomposition spaces, J. Fourier Anal. Appl. 13 (1) (2007) 39–70. doi:10.1007/s00041-006-6024-y. URL http://dx.doi.org/10.1007/s00041-006-6024-y

[33] F. Voigtlaender, A. Pein, Analysis vs. synthesis sparsity for α-shearlets, ArXiv preprint arXiv:1702.03559v1. URL https://arxiv.org/abs/1702.03559

[34] E. S. Ottosen, M. Nielsen, A characterization of sparse non-stationary Gabor expansions, J. Fourier Anal. Appl. doi:10.1007/s00041-017-9546-6. URL https://doi.org/10.1007/s00041-017-9546-6

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

[36] P. L. Søndergaard, Finite discrete Gabor analysis, Ph.D. thesis, Technical University of Denmark (2007).

[37] T. Strohmer, Numerical algorithms for discrete Gabor expansions, in: Gabor analysis and algorithms, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 1998, pp. 267–294. doi:10.1007/978-1-4612-2016-9_9. URL http://dx.doi.org/10.1007/978-1-4612-2016-9_9

[38] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, P. Balazs, The Large Time-Frequency Analysis Toolbox 2.0, in: M. Aramaki, O. Derrien, R. Kronland-Martinet, S. Ystad (Eds.), Sound, Music, and Motion, Lecture Notes in Computer Science, Springer International Publishing,
2014, pp. 419–442. doi:{10.1007/978-3-319-12976-1_25}
URL http://dx.doi.org/10.1007/978-3-319-12976-1_25

[39] E. B. Union, Sound quality assessment material: Recordings for subjective tests ; User’s handbook for the EBU - SQAM compact disc (Sep 2008).
URL https://tech.ebu.ch/docs/tech/tech3253.pdf