Statistical Analysis of Audio Signals using Time-Frequency Analysis

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg
Hochschulkennziffer: 1180

vorgelegt von

Pavel Tafo
Master of Science
geboren in Douala, Kamerun

Marburg, 2023

Erstgutachter: Prof. Dr. Hajo Holzmann
Zweitgutachter: Prof. Dr. Christian Rieger
Tag der Einreichung: 26. Juni 2023
Tag der Disputation: 22. September 2023
Erscheinungsort und -jahr: Marburg, 2023

In loving memory of Massa Mo.

Acknowledgements

First and foremost, I would like to thank my supervisor, Hajo Holzmann, for the opportunity to work on this fantastic and horizon-expanding project and for his exceptional guidance. I am truly grateful for his support, encouragement, understanding and patience over the past few years. All of this has made this journey possible and enjoyable.

I would also like to thank Prof. Dr. Christian Rieger for making the second assessment of this thesis.

This work was supported financially and practically by the project LOEWE Nature 4.0, funded by the Hessian Ministry of Higher Education, Research and the Arts (HMWK), which I hereby acknowledge. This has introduced me to the field of nature monitoring and has presented me with engaging tasks, challenges, and a motivating use case. After spending several years listening to recordings of birds, I can now easily distinguish the blackbird from all other bird species with complete confidence. I extend my gratitude to all the working groups involved for their valuable contributions and the stimulating discussions. In particular, I thank Prof. Dr. Stephan Dahlke. His guidance and expertise have been invaluable in helping me grasp the concept of time-frequency analysis.

I also thank all current and former members of probability theory and statistics working group for exciting and inspiring exchanges, the always pleasant working atmosphere and making the last few years very enjoyable. In particular, Liane Bingel, Max Berger and Philipp Hermann.

On a more personal note I am deeply thankful to my parents for their immense love, support and leading my way so far. Special thanks to my mom and my sisters for encouraging me and always having my back. Last but not least, thank you Armine for always being there for me. You are a true source of strength and support.

Contents

Notations

$\mathbb{R}$	The set of real numbers.
$\mathbb{N}$	The set of natural numbers.
$\mathbb{C}$	The set of complex numbers.
$\mathbb{Z}, \mathbb{Z}_+$	The set of integers, positive integers.
B	Banach spaces.
B'	The dual space of a Banach space B.
$\mathcal{H}$	Hilbert spaces.
$\mathcal{S}(\mathbb{R}^d)$	The Schwartz space or space of rapidly decreasing functions on $\mathbb{R}^d$.
$\mathcal{S}'(\mathbb{R}^d)$	The space of tempered distributions or space of slowly growing functions on $\mathbb{R}^d$.
$L^p(\mathbb{R}^d)$	The Lebesgue space of measurable functions on $\mathbb{R}^d$ that are p-integrable, i.e., $\left\{f : \mathbb{R}^d \to \mathbb{C}\right\}$ for which $\|f\|_p < \infty$.
$L^p_m(\mathbb{R}^d)$	The Lebesgue space of measurable functions on $\mathbb{R}^d$ such that $fm \in L^p(\mathbb{R}^d)$.
$C^\infty(\mathbb{R}^d)$	The space of infinitely differentiable functions on $\mathbb{R}^d$.
$\mathbb{C}^L$	The space of periodic sequences of length $L \in \mathbb{N}$.
$C[0,T]$	The space of continuous function on $[0,T]$.
$\langle \cdot, \cdot \rangle$	The inner product.
D^α	The differential operator on $\mathbb{R}^d$ with $D^\alpha = \dfrac{\partial^{\alpha_1}}{\partial^{\alpha_1}_{x_1}} \cdots \dfrac{\partial^{\alpha_d}}{\partial^{\alpha_d}_{x_d}}$
X^β	The multiplication operator on $\mathbb{R}^d$ with $X^\beta = \prod_{i=1}^d X_i^{\beta_i}$

$\mathcal{F}f = \widehat{f}(\omega)$	Fourier transform of a function f.
$\mathcal{F}B$	Image of a Banachspace B under the $\mathcal{F}$ with $\|f\|_{\mathcal{F}B} = \|\mathcal{F}^{-1}f\|_B$.
$\|\cdot\|_B$	Norm on a Banach Space B.
$\|\cdot\|_p$	L^p-norm, $\|f\|_p = (\int_{\mathbb{R}^d} \|f(x)\|^p \, dx)^{\frac{1}{p}}$ for $p < \infty$
$\|\cdot\|_\infty$	Essential supremum, $\|f\|_\infty = \text{ess sup}_{x\in\mathbb{R}^d} \|f(x)\|$
$\|\cdot\|_2$	The Euclidean norm.
Λ	Set of index.
$\#\Lambda$	Cardinality of the set Λ.
$\ell^2(\Lambda)$	The space of square summable sequence space on Λ.
$\text{supp}(f)$	Support of a function $f : \overline{\{x \in \mathbb{R}^d : f(x) \neq 0\}}$.
$\text{sgn}(x)$	Sign of a real number a with $\text{sgn}(a) = a/\|a\|$.
$\mathbb{E}$	Expected value.
Cov	Covariance.
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2.
$a_n = \mathcal{O}(b_n)$	Big O-Notation. The sequence a_n is smaller than the sequence b_n up to a multiplicative constant.
$\overline{z}$	The complex conjugated of $z \in \mathbb{C}^d$.
$\Re(z)$	The real part of a complex number z.
$\Im(z)$	The imaginary part of a complex number z.
$\varphi_a(x)$	The Gaussian function, $\varphi_a(x) = \exp(-\pi\|x\|_2^2/a)$.

1. Introduction

Time-frequency analysis is a commonly used tool for investigating the time-varying spectral content of non-stationary signals. Unlike traditional frequency analysis methods such as Fourier analysis, which effectively represent frequency components but fail to capture the dynamic nature of non-stationary signals, time-frequency analysis provides a form of local Fourier analysis that simultaneously describes the temporal and spectral behavior of a signal. It relies on two important techniques, translation and modulation, to localize signal components in the time-frequency domain and derive a time-frequency representation. Gröchenig (2013) provides a comprehensive exploration of the mathematical aspects of time-frequency analysis.

Signal denoising is a prominent application of time-frequency analysis. An effective approach to signal denoising is to use the spectral information provided by time-frequency analysis. Noise and the underlying signal often have different spectral characteristics. Signals typically have a more concentrated and localized time-frequency content, whereas noise tends to spread randomly across the time-frequency domain or a particular part of it. For example, low frequency noise such as motors or human voices will primarily affect the low frequency portion of the observed signal. White noise affects all frequencies equally. By exploiting this prior knowledge, denoising algorithms such as thresholding can be designed to attenuate or eliminate the noise while preserving the essential spectral components of the signal. In the field of audio denoising, using this approach leads to a significant improvement in audio quality and is of great value in applications such as speech recognition, audio restoration and audio communication systems, see Yu et al. (2008) or Ashwin and Manoharan (2018).

A key aspect of this approach is the accurate localization of signal components in the time-frequency domain. This leads to a sparse representation of the signal, which helps to distinguish it from noise. The most popular time-frequency analysis method is the

short-time Fourier transform (STFT). It involves sliding a fixed window over the signal to locate the Fourier transform. It provides a fixed time-frequency resolution controlled by the length of the window. However, this approach often results in a trade-off between good time resolution and good frequency resolution. For any square-integrable signal a countable, redundant and energy preserving representation in the time-frequency domain is obtained as a non-orthogonal expansion of the signal using Gabor frames, a collection of time-frequency shifts sampled from the STFT. The STFT is well suited for signals with stable frequency content over long time periods. In such cases, an appropriate window size can be selected for analysis.

Modulation spaces provide a natural framework for time-frequency analysis. They extend the Fourier smoothness-and-decay principle to the time-frequency domain, allowing the smoothness of a signal to be measured from the decay characteristics of its STFT in time-frequency space, cf. Galperin and Samarah (2004). In certain cases, modulation spaces coincide with Sobolev spaces. In Dahlke et al. (2022), we provide nonparametric estimation of signals corrupted by stationary noise in the white noise model. The observed signal is an element in an appropriate modulation space. We constructed an adaptive and rate-optimal estimator on modulation spaces by thresholding the coefficients obtained from the Gabor expansion. The rate obtained using classical oracle inequalities from Donoho and Johnstone (1994), exhibit new features reflecting the implication of both the time and the frequency. The signals in Dahlke et al. (2022) were observed across the entire $\mathbb{R}^d$ because, first, signals such as audio signals naturally lack inherent restrictions to specific domains, cf. Goldenshluger et al. (2006). Second, unlike the theories of other function spaces such as Besov or Sobolev, the theory of Gabor expansions and modulation spaces is not fully developed for bounded domains.

A well-studied alternative to time-frequency analysis is the (orthogonal) wavelet transform (WT). As described in Daubechies (1990), the WT also provides time-frequency localization with a time-frequency resolution that varies with frequency. It effectively localizes high-frequency events in time and low-frequency events in frequency, making it a suitable choice for analyzing many real-world signals. These signals often have slowly oscillating content on longer scales, while high-frequency events tend to be abrupt or transient in nature. However, it's important to note that if the spectral content of the underlying signal is limited to a specific frequency range, using the WT alone may re-

sult in either reduced frequency resolution or compromised time resolution. Donoho and Johnstone (1994) show that wavelet denoising is close to optimal in the minimax sense on Besov spaces, specifically for Gaussian white noise scenarios. Mallat (2009) also discusses denoising wavelet frame expansions by thresholding. These results highlight the effectiveness of wavelet denoising techniques in dealing with noise in signals.

To address the resolution challenges posed by the Gabor and wavelet transforms, a hybrid approach called the Flexible Gabor-Wavelet Transform (GWT) has been introduced, as described in Nazaret and Holschneider (2003). The GWT uses a parameter $\alpha \in [0, 1]$ that dynamically adjusts the window length based on frequency, resulting in improved resolution compared to the STFT. It also addresses the loss of frequency resolution observed in the WT when analyzing higher frequencies. The smoothness space associated with this transform is called the α modulation space. It lies between modulation spaces and Besov spaces and provides a versatile framework for analyzing signals with varying degrees of smoothness in the time-frequency domain. The introduction of the α parameter allows for a flexible adjustment of the window length, striking a balance between time and frequency resolution.

In this thesis, we reformulate and extend the results presented in Dahlke et al. (2022). However, unlike the previous work, we derive all results from sequence space, thus providing a different perspective. In particular, the frame expansions used in this analysis are energy preserving, which means that the quadratic risk incurred during the denoising process of the time-frequency representation applies equally to the signal after synthesizing the coefficients. Furthermore, we extend the obtained results to obtain comparable results in the bounded domain. This allows for a broader applicability of the results beyond the unbounded setting. In addition, we generalize the estimation results to include the α modulation space in the one-dimensional setting. Through these extensions and generalizations, we aim to improve the understanding and applicability of time-frequency analysis techniques in various scenarios.

This thesis is organized as follows. Chapter 2 provides a comprehensive review of frames and frame expansions in both Hilbert spaces and Banach spaces. We examine the STFT and the use of Gabor frames for signal expansions. We review the concept of decay on product measure spaces using weight functions. We also introduce modulation spaces

and explore their characterization using Gabor expansions. Finally, we delve into the theory of the flexible Gabor-Wavelet transform and α-modulation spaces. In Chapter 3, we revisit the findings of Dahlke et al. (2022) concerning the nonparametric estimation of signals in modulation spaces using Gabor frames. Notably, our analysis is conducted entirely within the sequence space. We first constrain the signal representation by considering a finite subset of the representation and estimating the signal within this finite subsequence. We derive the minimax convergence rate and compare it to known results, such as those in the Sobolev case. Chapter 4 replicates the results of Chapter 3 but focuses on the utilization of α-modulation frames. We make comparisons with results obtained in Sobolev spaces as well as with those in Besov spaces. In Chapter 5, we present the outcomes of intensive computations that highlight the advantages of Gabor frames in audio denoising. These experiments demonstrate the effectiveness of the proposed approach, and we provide quantitative results to evaluate the denoising performance. We compare the estimated signals with the ground truth signals, using various performance metrics such as signal-to-noise ratio (SNR) and mean square error (MSE).

2. Frame Theory and Time-Frequency Representation

This chapter focuses on representing (audio) signals in the time-frequency domain using non-orthogonal expansions, with frames serving as a valuable tool for this purpose. By building upon time-frequency elements, frames provide a rich and redundant dictionary of vectors that can effectively represent (audio) signals in a sparse manner when designed appropriately. The use of frames allows the creation of a function space from which more general results can be derived. The chapter is structured as follows. First, we introduce frames in the context of signal representation in Hilbert spaces. Important concepts, facts and results about frames are reviewed, drawing on references such as Christensen (2016) and Mallat (2009). The notion of frames can be extended to Banach spaces, providing a non-orthogonal expansion for an even larger class of functions, as explored in works like Gröchenig (1991) and Megginson (1998). Next, the short-time Fourier transform is introduced, followed by Gabor's idea of frames for the $L^2(\mathbb{R}^d)$ space based on the short-time Fourier transform, as presented in Gabor (1946). Modulation spaces, which are smoothness classes that characterize their elements by the decay properties of their time-frequency representations, are also discussed. Finally, a more general smoothness class called α-modulation is introduced, which provides a generalization from modulation spaces, such as $L^2(\mathbb{R}^d)$, to Besov spaces. Throughout the chapter, relevant literature, notions, and theories are reviewed to provide a comprehensive understanding of the topic.

2.1. Frames in Hilbert Spaces

Introduced by Duffin and Schaeffer (1952) in the context of nonharmonic Fourier series, frames are families of vectors that provide a complete, stable, flexible, and potentially non-unique representation of elements in the underlying space. Frames soften the concept of a basis by providing the flexibility needed to represent a signal by unstructured families of vectors. Unlike a basis, elements of a frame can be linearly dependent, non-orthogonal, and overcomplete, i.e., have more elements than needed. This overcompleteness ensures the existence of such a representation and is more stable against noise in the coefficients.

Throughout this thesis, let $(\mathcal{H}, \|\cdot\|)$ be a (separable) Hilbert space with the inner product $\langle\cdot,\cdot\rangle$. For a countable index set Λ, $\ell^2(\Lambda)$ represents the set of square summable sequences on Λ. In the following, we consider a sequence $\{h_\lambda\}_{\lambda\in\Lambda}$ in $\mathcal{H}$.

Definition 1 (Frame and Riesz Basis, (Mallat, 2009, Definition 5.1))**.** A sequence $\{h_\lambda\}_{\lambda\in\Lambda}$ is a frame for $\mathcal{H}$ if there exist two constants $0 < A \leq B$ such that for all $f \in \mathcal{H}$

$$A\|f\|^2 \leq \sum_{\lambda\in\Lambda} |\langle f, h_\lambda\rangle|^2 \leq B\|f\|^2. \tag{2.1}$$

The numbers A, B are called frame bounds. They are redundancy factors and not unique. When $A = B$, the frame is said to be tight. If the $\{h_\lambda\}_{\lambda\in\Lambda}$ are linearly independent then the frame is called a *Riesz* basis. A frame which is not a Riesz basis is called *nonexact* or *redundant* or *overcomplete* frame, i.e. there exists nonzero sequences $\{c_\lambda\}_{\lambda\in\Lambda} \in \ell^2(\Lambda) \setminus \{0\}$ for which $\sum_{\lambda\in\Lambda} c_\lambda h_\lambda = 0$. A frame that ceases to be a frame when an arbitrary element is removed, it is called an *exact* frame. Hence exact frames are also Riesz basis.

We refer to the previous definition as the frame condition. It shows that frames define a stable signal representation since any bounded input signal produces a bounded output. Thus any perturbation of frame coefficients $\{\langle f, h_\lambda\rangle\}_{\lambda\in\Lambda}$ implies a modification of similar magnitude on the signal. For a better understanding of frames and reconstruction methods, we review some important associated operators on Hilbert spaces.

Definition 2 (Analysis and Synthesis Operator)**.**
For a sequence $\{h_\lambda\}_{\lambda\in\Lambda} \subseteq \mathcal{H}$ satisfying the frame condition (2.1) or at least the right hand side of the inequality (Bessel condition), the analysis operator T is given by

$$T : \mathcal{H} \to \ell^2(\Lambda), \qquad Tf = \{\langle f, h_\lambda\rangle\}_{\lambda\in\Lambda} \tag{2.2}$$

Its adjoint operator, the *synthesis operator* T^*, is given by

$$T^* : \ell^2(\Lambda) \to \mathcal{H}, \qquad T^*(\{c_\lambda\}_{\lambda\in\Lambda}) = \sum_{\lambda\in\Lambda} c_\lambda h_\lambda \tag{2.3}$$

Obviously $\|T\| \leq \sqrt{B}$, despite the frame vectors not having to be orthogonal in general. Hence both operators are bounded. Furthermore the existence of a well-defined, bounded and surjective synthesis operator, mapping from $\ell^2(\Lambda)$ onto $\mathcal{H}$, provides a characterization of frames for Hilbert spaces, see Christensen (2016, Section 5.5). This characterization of frames does not require any knowledge of the frame bounds. The frame condition (2.1) shows that the linear combination of coefficients obtained from the analysis operator is stable. The linear combination of frame vectors is also stable, see Mallat (2009, Theorem 5.1).

By composing T and T^* we obtain the *frame operator* S on $\mathcal{H}$

$$S : \mathcal{H} \to \mathcal{H}, \qquad Sf = T^*Tf = \sum_{\lambda\in\Lambda} \langle f, h_\lambda\rangle h_\lambda \tag{2.4}$$

For any $f \in \mathcal{H}$ the series defining the frame operator S converges unconditionally, i.e., regardless of the order of summation, see Christensen (2016, Lemma 2.1.1). The next lemma depicts important elementary properties of the frame operator in Hilbert spaces.

Lemma 1. *Let $\{h_\lambda\}_{\lambda\in\Lambda}$ be a frame with frame operator S and frame bounds A,B. Then the following holds:*

1. *The frame operator S maps $\mathcal{H}$ onto $\mathcal{H}$ and is bounded, invertible, self-adjoint, and positive*
2. *$\{S^{-1}h_\lambda\}_{\lambda\in\Lambda}$ is a frame with bounds B^{-1}, A^{-1}. The frame operator for $\{S^{-1}h_\lambda\}_{\lambda\in\Lambda}$ is S^{-1}, the inverse frame operator.*

Proof. Christensen (2016, Lemma 5.1.5) □

The frame $\{S^{-1}h_\lambda\}_{\lambda\in\Lambda}$ is referred to as the *canonical dual frame*. It plays an important role in signal in the reconstruction of any signal $f \in \mathcal{H}$. The most important frame result is the frame decomposition. It shows that if $\{h_\lambda\}_{\lambda\in\Lambda}$ is a frame for the Hilbert space $\mathcal{H}$, then every element in $\mathcal{H}$ has a representation as a linear combination of the frame elements h_λ. Thus frame can be viewed as a generalization of the concept of basis. From the frame operators, S and S^{-1}, we obtain the following representation formulas for $f \in \mathcal{H}$ as linear combination of the frame elements.

Theorem 2. *Let $\{h_\lambda\}_{\lambda\in\Lambda}$ be a frame for $\mathcal{H}$ with frame operator S. Every $f \in \mathcal{H}$ has non-orthogonal expansions*

$$f = \sum_{\lambda\in\Lambda} \langle f, S^{-1}h_\lambda\rangle h_\lambda \tag{2.5}$$

and

$$f = \sum_{\lambda\in\Lambda} \langle f, h_\lambda\rangle S^{-1}h_\lambda, \tag{2.6}$$

where both series converge unconditionnally in $\mathcal{H}$.

Proof. Gröchenig (2013, Theorem 5.1.3) □

The sequence $\left\{\langle f, S^{-1}h_\lambda\rangle\right\}_{\lambda\in\Lambda}$ is called frame coefficients. The canonical dual frame $\{S^{-1}h_\lambda\}_{\lambda\in\Lambda}$ plays the same role as the dual of a basis. Theorem 2 can be understood in the following way: while the dual analysis (eq. (2.5)) provides a non-orthogonal expansion of f with respect to the frame vectors $\{h_\lambda\}_{\lambda\in\Lambda}$, the dual synthesis (eq. (2.6)) reconstructs f from the frame coefficients using the dual frame as an expanding function. It also shows that all informations about the given $f \in \mathcal{H}$ is contained in the sequence $\{\langle f, S^{-1}h_\lambda\rangle\}_{k\in\Lambda}$. For an overcomplete frame $\{h_\lambda\}_{\lambda\in\Lambda} \subseteq \mathcal{H}$, there exist several choices of frames $\{g_\lambda\}_{\lambda\in\Lambda} \subseteq \mathcal{H}$ with $\{g_\lambda\}_{\lambda\in\Lambda} \neq \{S^{-1}h_\lambda\}_{\lambda\in\Lambda}$ for which

$$f = \sum_{\lambda\in\Lambda} \langle f, g_\lambda\rangle h_\lambda. \tag{2.7}$$

Any frame $\{g_\lambda\}_{\lambda\in\Lambda}$ satisfying (2.7) is called a *dual frame* of $\{h_\lambda\}_{\lambda\in\Lambda}$. It then becomes obvious that in contrast to orthogonal expansion, the coefficients in the frame expansion (2.5) and (2.6) are in general not unique. They are unique if and only if the corresponding frame is a Riesz basis of $\mathcal{H}$. This frame redundancy is useful in practical problems requiring robustness and error tolerance, like noise reduction.

2.1.1. Frames in Banach Spaces

Under certain conditions, the concept of frames can also be extended to provide non-orthogonal signal expansion in Banach spaces, see Gröchenig (1991) and Banach (1932), Megginson (1998). This will be useful later in this work.

Definition 3 (Fornasier (2004, Definition 1))**.** Let $(B, \|\,\|_B)$ be a separable Banach space and $(B_b, \|\,\|_{B_b})$ an associated sequence space. A sequence $\{g_\lambda\}_{\lambda\in\Lambda} \subseteq B'$ is Banach frame for B if the following properties hold.

- The analysis operator $Tf = \{\langle f, g_\lambda\rangle\}_{\lambda\in\Lambda}$, is bounded from B into B_b.
- Norm equivalence:
$$\|f\|_B \asymp \|\{\langle f, g_\lambda\rangle\}_{\lambda\in\Lambda}\|_{B_b}$$
- There exists a bounded synthesis operator T from B_b onto B such that
$$T^*(\{\langle f, g_\lambda\rangle\}_{\lambda\in\Lambda}) = f$$

A Banach frame allows a representation of any $f \in B$ in the associated sequence space. The sequence $\{g_\lambda\}_{\lambda\in\Lambda} \subseteq B$ is called *frame atomic decomposition* of the Banach space B if there exists a Banach frame $\{\tilde{g}_k\}_{\lambda\in\Lambda} \subseteq B$ such that for any $f \in B$ the series expansion

$$f = \sum_{\lambda\in\Lambda} \langle f, \tilde{g}_\lambda\rangle g_\lambda \tag{2.8}$$

converges unconditionally. We then refer to $\{\tilde{g}_k\}_{\lambda\in\Lambda}$ as the dual Banach frame.

We conclude this section with some important remarks on frames; proofs and further details can be found in Gröchenig (2013, Chapter 5), Christensen (2016), and Mallat (2009).

Remark 1.

1.) A sequence $\{h_\lambda\}_{\lambda\in\Lambda}$ of elements in $\mathcal{H}$ is a frame for $\mathcal{H}$ if and only if $\overline{\text{span}}\{h_\lambda\}_{\lambda\in\Lambda} = \mathcal{H}$. Therefore it is sufficient to check the frame condition on a dense set of $\mathcal{H}$, then it also holds on $\mathcal{H}$, with the same frame bounds, Christensen (2016, Lemma 5.1.7).

2.) The canonical dual frame $\{S^{-1}h_\lambda\}_{\lambda\in\Lambda}$ minimizes the l^2-norm of the frame coefficients among all dual frames, i.e. for any $\{c_\lambda\}_{\lambda\in\Lambda} \in \ell^2(\Lambda)$ satisfying $f = T^*\{c_\lambda\}_{\lambda\in\Lambda}$ it follows

$$\sum_{k\in\Lambda} |c_k|^2 \geq \sum_{k\in\Lambda} |\langle f, S^{-1}h_\lambda\rangle|^2 \tag{2.9}$$

See Gröchenig (2004, Proposition 5.1.4)

3.) The existence of multiple dual frames facilitates the reconstruction of the function f in situations where finding the inverse frame operator appears to be a difficult task. In the rest of the thesis, where no confusion can arise, we will refer to canonical dual frame just as dual frame.

4.) From a computational point of view, a tight frame offers a great advantage. The canonical dual of tight frame $\{h_\lambda\}_{\lambda\in\Lambda}$ with frame bound A is given as

$$S^{-1}h_\lambda = \{\frac{1}{A}\cdot h_\lambda\}_{\lambda\in\Lambda} \tag{2.10}$$

Not only do tight frames help facilitate the computation of the dual frame, they also help control the behavior of the dual frame, as the retain the linear properties of the frame.

5.) The optimal upper and lower frame bound are respectively the infimum over all upper frame bounds and the supremum over all lower frame bounds. By rewriting the frame condition as:

$$A\|f\|^2 \leq \langle Sf, f\rangle \leq B\|f\|^2, \qquad \forall f \in \mathcal{H}.$$

it results that A and B are the infimum and supremum values of the spectrum of the frame operator S, which correspond to the smallest and largest eigenvalues in finite dimension. If A, B are the optimal bounds for $\{h_\lambda\}_{\lambda\in\Lambda}$, then the bounds B^{-1}, A^{-1} are optimal for $\{S^{-1}h_\lambda\}$.

2.2. Time-Frequency Analysis

In this section, we will explore various aspects of the time-frequency representation of signals. The mathematical framework is built upon the Hilbert space $L^2(\mathbb{R}^d)$. Time-frequency analysis is best described as a localized version of the Fourier transform. The goal of time-frequency analysis is to provide a simultaneous description of the temporal and spectral behavior of a signal. This is essential when dealing with signals whose spectral characteristics change with time.

Before diving into the specific techniques used in time-frequency analysis, it is helpful to review some relevant results from Fourier analysis. For more results, details, and explanations of Time-frequency analysis, please refer to the book "Foundations of Time-Frequency Analysis" by Hans Gochenig (2013).

2.2.1. The Schwartz Space and Fourier Transform

The Schwartz space, denoted by $\mathcal{S}(\mathbb{R}^d)$, is an important building block in time-frequency analysis. It is defined as the space of rapidly decaying smooth functions on the Euclidean space $\mathbb{R}^d$. The functions in the Schwartz space have the property that any derivative of the function decays faster than any inverse power of $|x|$. Formally, the Schwartz class is defined as the set of all functions $f \in C^\infty(\mathbb{R}^d)$ such that the following condition holds:

$$\sup_{x\in\mathbb{R}^d} |D^\alpha X^\beta f(x)| < \infty \text{ for all } \alpha, \beta \in \mathbb{Z}_+^d. \tag{2.11}$$

Furthermore, it is dense in $L^p(\mathbb{R}^d)$ for $1 \leq p < \infty$, meaning that any function in $L^p(\mathbb{R}^d)$ can be approximated arbitrarily closely by a function in the Schwartz space. This makes it a convenient space for many mathematical operations, such as the Fourier transform, in the time-frequency analysis. In addition, the space of test functions, compact smooth functions, is also included in $\mathcal{S}(\mathbb{R}^d)$.

The dual of the Schwartz space, $\mathcal{S}'(\mathbb{R}^d)$, is called the space of tempered distributions, the space of continuous linear functionals on the Schwartz space. Tempered distributions can be characterized as the space of slowly growing functions. This means that any derivative of the function grows at most as fast as a polynomial. In particular, elements of the $L^p(\mathbb{R}^d)$ space for $p \geq 1$ are tempered distributions. This space is a convenient space for many mathematical operations, such as partial differential equations, Fourier analysis, and distributions theory.

Definition 4 (Fourier transform, Gröchenig (2013))**.** For $f \in \mathcal{S}(\mathbb{R}^d)$, the Fourier transform of f, denoted by $\widehat{f}(\omega)$, is given by:

$$\widehat{f}(\omega) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \omega x} \, dx$$

for a frequency $\omega \in \mathbb{R}^d$. The inverse Fourier transform is given by:

$$f(x) = \int_{\mathbb{R}^d} \widehat{f}(\omega) e^{2\pi i x \cdot k} \, dk \tag{2.12}$$

The Fourier transform is a bounded linear operator. To highlight this, we write $\mathcal{F}f$ instead of $\widehat{f}(\omega)$. It defines a continuous bijection on $\mathcal{S}(\mathbb{R}^d)$. By duality, the Fourier transform can be extended to $\mathcal{S}'(\mathbb{R}^d)$. The Fourier transform is a powerful tool for analyzing the frequency content of a signal. While $f(x)$ describes the temporal behavior, $\widehat{f}(\omega)$ describes the frequency behavior of the function f. For all frequencies $\omega \in \mathbb{R}^d$, we obtain the frequency domain representation. The inversion formula provides a synthesis process that reconstructs the original function f from its frequency domain representation.

Note that the Fourier transform is a unitary operator on the Hilbert space $L^2(\mathbb{R}^d)$. It preserves the inner product and the norm, i.e. $\langle f, g \rangle_{L^2(\mathbb{R}^d)} = \langle \widehat{f}(\omega), \widehat{g}(\omega) \rangle_{L^2(\mathbb{R}^d)}$ as well

as $\|f(x)\|_2 = \|\widehat{f}(\omega)\|_2$ for all $f, g \in L^2(\mathbb{R}^d)$, see Plancherel's theorem and Parseval's formula.

Another later useful remark is the uncertainty principle. This states that a signal that is well localized in time (i.e., has a short duration) cannot be well localized in frequency (i.e., has a narrow band of frequencies). Conversely, a signal that is well localized in frequency cannot be well localized in time. The mathematical representation of this principle is that a signal and its Fourier transform cannot both have finite support unless the signal is identically zero. In other words, a band-limited signal cannot be time-limited. For more details on the Fourier transform, see Gröchenig (2004, Chapter 1).

2.2.2. Short-Time Fourier Transform

The Fourier transform is indeed a powerful tool for analyzing signals by representing them in terms of their frequency content. However, it has a limitation in that it provides little information about the localization of frequencies in the time domain. That is, the Fourier transform does not provide information about the time localization of a particular frequency component of a signal. To overcome this limitation, time-frequency analysis methods such as the Short-Time Fourier Transform (STFT) have been developed to provide a more detailed representation of signals that includes both frequency and time information. The STFT represents a signal in the time-frequency domain, allowing analysis of the frequency content of the signal as it changes over time. To obtain information about the localization of frequencies in a signal, the signal is first localized to a neighborhood of a given time using a windowing function. This localized version of the signal is then transformed using the Fourier transform to obtain a frequency representation around that specific point in time. By using different window functions and varying the position of the window, a time-frequency representation of the signal can be obtained that provides information about the distribution of frequencies in the signal at different points in time. This time-frequency representation of a signal is useful for analyzing signals whose spectral content changes with time because it provides a simultaneous description of both the temporal and spectral behavior of the signal. The STFT provides a local view of the frequency content of a signal, which can be useful for identifying patterns or changes in the signal that may not be readily apparent in the time or frequency domain alone.

We begin by introducing two fundamental operators that play a central role in time-frequency analysis.

Definition 5. For $x \in \mathbb{R}^d$ the *translation* operator T_x is defined as

$$T_x f(t) = f(t - x)$$

For $\omega \in \mathbb{R}^d$ the *modulation* operator M_ω is defined as

$$M_\omega f(t) = e^{-2\pi i \omega t} f(t)$$

The two operators are related by the Fourier transform

$$\mathcal{F} M_\omega = T_\omega \mathcal{F} \qquad \text{and} \qquad \mathcal{F} T_x = M_{-x} \mathcal{F}.$$

The translation operator T_x or *time shift* shifts a signal in the time domain by a fixed amount x. This corresponds to a delay in the time domain. On the other hand, the modulation operator M_ω or *frequency shift* corresponds to a shift in the frequency domain by a fixed amount ω, since the Fourier transform of a modulated function is the shifted version of its Fourier transform. Similarly, the Fourier transform of a time-shifted signal is related to the original Fourier transform through the modulation operator. This relationship between time-shifting and frequency-shifting is a key aspect of the theory of time-frequency analysis.

Operators of the form $T_x M_\omega$ are called time-frequency shifts. The order of the operators is interchangeable due to the commutative relation $T_x M_\omega = e^{-2\pi i \omega x} M_\omega T_x$. Time-frequency shifts are unitary on $L^2(\mathbb{R}^d)$ and define isometries on $L^p(\mathbb{R}^d)$ for each $0 < p \leq \infty$, i.e.,

$$\|T_x M_\omega f\|_p = \|f\|_p. \tag{2.13}$$

Definition 6 (STFT)**.** Let $B \subseteq \mathcal{S}'(\mathbb{R}^d)$ be a Banach space. Fix a nonzero function $g \in B'$. The short-time Fourier transform or voice transform of a function $f \in B$ with

respect to g is defined as

$$V_g f(x,\omega) = \int_{\mathbb{R}^d} f(t)\overline{g(t-x)}e^{-2\pi i\omega t}dt, \qquad \text{for } x,\omega \in \mathbb{R}^d. \tag{2.14}$$

The STFT measures the content of the frequency band at ω in a neighborhood of time x. It parameterizes the localization with the window function g. For a fixed window function g, the STFT maps from $\mathbb{R}^{2d}$, the time-frequency plane, to a complex-valued weight. While $V_g f$ depends linearly on f, its properties depend on the window function g. The smoothness of g plays a crucial role, since discontinuities in the window would lead to singularities in the time-frequency plane. Using previously defined Fourier tools, we obtain the fundamental identity of the time-frequency analysis by rewriting definition 6 as

$$V_g f(x,\omega) = \langle f, M_\omega T_x g\rangle = \langle \widehat{f}(\omega), T_\omega M_{-x}\widehat{g}(\omega)\rangle \tag{2.15}$$

This form is most convenient for formal manipulations and reveals some of the deeper structures of the STFT. The STFT is well defined if the scalar product is well defined by some form of duality, such as for $g \in \mathcal{S}(\mathbb{R}^d)$ and $f \in \mathcal{S}'(\mathbb{R}^d)$. A nice window g allows a not so nice function f. It combines both f and $\widehat{f}(\omega)$ into a joint time-frequency representation. If the window function g is compact with its support centered at the origin, then $V_g f(x,\cdot)$ is the Fourier transform of a weighted segment of f centered in a neighborhood of x. Similarly, $V_g f(\cdot,\omega)$ is the local Fourier transform of $\widehat{f}(\omega)$, with $\widehat{g}(\omega)$ as the window function. The global time-frequency distribution of the function f is then given by $V_g f$.

For $f, g \in L^2(\mathbb{R})$ it follows that $V_g f \in L^2(\mathbb{R}^{2d})$. Furthermore, the STFT is uniformly continuous on $\mathbb{R}^{2d}$ and satisfies the following orthogonal relations

$$\langle V_{g_1} f_1, V_{g_2} f_2\rangle = \langle f_1, f_2\rangle\overline{\langle g_1, g_2\rangle} \tag{2.16}$$

for $f_1, f_2, g_1, g_2 \in L^2(\mathbb{R})$, see Gröchenig (2013, Theorem 3.2.1). This leads to the following equality for all $f, g \in L^2(\mathbb{R}^d)$:

$$\|V_g f\|_{L^2(\mathbb{R}^{2d})} = \|f\|_{L^2(\mathbb{R}^d)}\|g\|_{L^2(\mathbb{R}^d)}.$$

For $\|g\|_{L^2(\mathbb{R}^d)} = 1$, the STFT defines an isometry from $L^2(\mathbb{R})$ to $L^2(\mathbb{R}^{2d})$ and the function f is completely determined by $V_g f$. Furthermore, from the fundamental identity

of the time-frequency analysis (2.15) it follows that the only function orthogonal to all possible time-frequency shifts is the zero function. Therefore, for any fixed $g \in L^2(\mathbb{R}^d)$, the sequence of time-frequency shifts $\{M_\omega T_x\, g\}_{x,\omega\in\mathbb{R}^d}$ spans a dense subspace of $L^2(\mathbb{R}^d)$. This means that any function in $L^2(\mathbb{R}^d)$ can be approximated arbitrarily closely by a finite linear combination of time-frequency shifts. Any square integrable function can be reconstructed from its time-frequency representation. An inversion formula for the STFT is given below.

Corollary 3. *(Weak Inversion Formula for the STFT). Suppose that $g, \gamma \in L^2(\mathbb{R}^d)$ and $\langle g, \gamma\rangle \neq 0$. Then for all $f \in L^2(\mathbb{R}^d)$*

$$f = \frac{1}{\langle g, \gamma\rangle} \int\int_{\mathbb{R}^{2d}} V_g f(x,\omega) M_\omega T_x \gamma d_\omega d_x. \tag{2.17}$$

Proof. Gröchenig, 2013, corollary 3.2.3 □

The formula of weak inversion states that any square-integrable function can be expressed as a combination of time-frequency shifts using the Short-Time Fourier Transform as the weighting factor. The time-frequency shifts help to capture the variation of the function in both the time and frequency domains. This formula provides a way to represent any square-integrable function in terms of its time-frequency content, making it useful for various signal processing applications.

2.2.3. The Gaussian Function and Time-frequency Resolution

The Gaussian function plays an important role in time-frequency analysis. For $a > 0$ the non-normalized Gaussian function on $\mathbb{R}^d$ is denoted by

$$\varphi_a(x) = \exp(-\pi \|x\|_2^2 / a). \tag{2.18}$$

This is a Schwartz function and therefore decays rapidly. When used as a window function, the Gaussian function provides a great time localization because its essential

support is concentrated around a single point. In addition, its Fourier transform is also Gaussian,

$$\widehat{\varphi_a}(\omega) = a^{d/2}\varphi_{1/a}(x). \tag{2.19}$$

Therefore, it also provides great localization in frequency. The following results show a useful application of the Gaussian function in time-frequency analysis.

Lemma 4. *For all $a > 0$ we have*

1. *For $x, u, \omega, \eta \in \mathbb{R}^d$ we have*

$$\langle T_x M_\omega \varphi_a, T_u M_\eta \varphi_a \rangle = (a/2)^{d/2} e^{\pi i \langle u-x, \eta+\omega \rangle} \varphi_{2a}(u-x)\varphi_{\frac{2}{a}}(\eta-\omega) \tag{2.20}$$

2. *The set $\left\{T_x M_\omega \varphi_a : x, \omega \in \mathbb{R}^d\right\}$ spans a dense subspace of $L^2(\mathbb{R}^d)$*

3. *φ_a is separable, i.e.$\varphi_a(x_1, \ldots, x_d) = \varphi_a(x_1) \cdot \ldots \cdot \varphi_a(x_d)$ for $x \in \mathbb{R}^d$*

Proof. Gröchenig (2004, Lemma 1.5.2 & Lemma 1.5.3). iii) follows from (2.19). □

Time-frequency Resolution

An important concept in time-frequency analysis is the notion of time-frequency resolution. It describes the ability to distinguish between two closely spaced events in time-frequency space. This is not to be confused with the discrete resolution of a representation (image), which simply indicates the number of elements (pixels).

Suppose the window function γ and its Fourier $\widehat{\gamma}(\omega)$ have essential support T and $\Omega \subseteq \mathbb{R}^d$ respectively. T and Ω represent the time resolution and the frequency resolution, respectively. Therefore, the time-frequency resolution is given by $|T| \times |\Omega|$. The goal in time-frequency analysis is to make this resolution as small as possible. The uncertainty principle states that the time-frequency resolution cannot be arbitrarily small, Folland (1989) and Folland and Sitaram (1997). Good time resolution comes at cost of bad frequency resolution and vice versa.

Theorem 5 (Lieb Uncertainty Theorem)**.** *If $f, g \in L^2(\mathbb{R}^d)$ and $1 \leq p \leq 2$, then*

$$\int \int_{\mathbb{R}^{2d}} |V_g f(x,\omega)| \, dx d\omega \geq \left(\frac{2}{p}\right)^d (\|f\|_2 \|g\|_2)^p \tag{2.21}$$

Proof. Lieb (1990) □

In practical audio processing, a wide window is needed to achieve good time resolution, but this comes at the cost of poor frequency resolution, which is a manifestation of the Uncertainty Principle. A narrow window has the opposite trade-off. Gaussian window functions provide the best time-frequency resolution because they minimize the time-frequency resolution. They are therefore the most appropriate windowing function for the STFT.

2.2.4. Gabor Frames

The inversion formula of the short-time Fourier transform states that any square-integrable function $f \in L^2(\mathbb{R}^d)$ has a continuous expansion with respect to the uncountable sequence $\{M_\omega T_x \, g\}_{x,\omega \in \mathbb{R}^d}$, with $g \in L^2(\mathbb{R}^d)$. In other words, f can be expressed as an infinite series of these time-frequency shifts. However, the space $L^2(\mathbb{R}^d)$ is a separable Hilbert space, i.e. it has a countable dense subset. This means that a series expansion with respect to a countable subset of the time-frequency shifts is sufficient to represent any signal in $L^2(\mathbb{R}^d)$.

Consider a nonzero window function $g \in L^2(\mathbb{R}^d)$. For each pair $x, \omega \in \mathbb{R}^d$, the time-frequency shift $M_\omega T_x \, g$ captures the time-frequency information of a signal on a region of the time-frequency space with surface area E. By choosing a sufficiently dense lattice of time-frequency shifts, the entire time-frequency space $\mathbb{R}^{2d}$ can be covered by a countable number of time-frequency shifts. These shifts can overlap, i.e. the same region in the time-frequency space can be covered by several shifts. This lattice of time-frequency shifts forms a frame in the Hilbert space $L^2(\mathbb{R}^d)$, and by the properties of frames, any signal $f \in L^2(\mathbb{R}^d)$ can be represented as a series expansion with respect to this frame, allowing a comprehensive analysis of the time-frequency content of the signal.

We will now introduce Gabor frames. Introduced in the 1940s and named after Dennis Gabor, Gabor frames are formed by time-frequency shifts of a fixed window function g and are parametrized by lattice parameters $\alpha, \beta > 0$.

Definition 7. Given a nonzero window function $g \in L^2(\mathbb{R}^d)$ and lattice parameters $\alpha, \beta > 0$, a *Gabor system* is defined as the set of time-frequency shifts

$$\mathcal{G}(g, \alpha, \beta) = \left\{ T_{\alpha k} M_{\beta n} g : \, k, n \in \mathbb{Z}^d \right\}. \tag{2.22}$$

If $\mathcal{G}(g, \alpha, \beta)$ is a frame for $L^2(\mathbb{R}^d)$, then it is called *Gabor frame*.

In the literature, Gabor frames are also referred to as Weyl-Heisenberg frames, since they provide a suitable representation of the Heisenberg group. Gabor systems are parametrized by the window function g and the lattice density $\alpha\beta$. The author in Gröchenig (2013) provides conditions on the window function and the lattice density to ensure the existence of Gabor frames. The Wiener space, denoted $W(\mathbb{R}^d)$, is a Banach space consisting of almost periodic functions, i.e. locally bounded and globally in l_1. In particular, bounded functions with compact support are contained in the Wiener space. Hence, $W(\mathbb{R}^d)$ is a dense subspace of $L^p(\mathbb{R}^d)$ for $1 \leq p < \infty$. For $g \in W(\mathbb{R}^d)$, the synthesis operator based on the Gabor system $\mathcal{G}(g, \alpha, \beta)$ is well defined and bounded, see (Gröchenig, 2013, Proposition 6.2.2). Thus, $\mathcal{G}(g, \alpha, \beta)$ defines a frame for $T^*(l^2(\mathbb{Z}^{2d})) \subseteq L^2(\mathbb{R}^d)$. Furthermore, if the density, $\alpha\beta$, is small enough, $\alpha\beta \leq 1$, then the frame elements entirely span $L^2(\mathbb{R}^d)$, making $\mathcal{G}(g, \alpha, \beta)$ a frame for $L^2(\mathbb{R}^d)$. Moreover if $\alpha\beta = 1$, then $\mathcal{G}(g, \alpha, \beta)$ is a Riesz basis. For more details on these results, see Gröchenig (2013, Chapters 5,6 & 7).

The frame operator associated to a Gabor frame $\mathcal{G}(g, \alpha, \beta)$ is then given by

$$\begin{aligned} S_{g,g}^{\alpha,\beta} f &= \sum_{k,n \in \mathbb{Z}^d} \langle f, T_{\alpha k} M_{\beta n} g \rangle T_{\alpha k} M_{\beta n} g \\ &= \sum_{k,n \in \mathbb{Z}^d} \langle f, M_{\beta n} T_{\alpha k} g \rangle M_{\beta n} T_{\alpha k} g \end{aligned}$$

Note that the order of translation and modulation in the frame operator are interchangeable since the phase factor $e^{-2\pi i \omega}$ cancels out. For ease of reading, the indices are omitted in the Gabor frame operator whenever there is no possibility of confusion. The frame operator S and its inverse S^{-1} both commute with time-frequency shifts. Thus, if $\mathcal{G}(g, \alpha, \beta)$

is a frame for $L^2(\mathbb{R}^d)$, then there exists a dual window $\gamma = S^{-1}g$, such that the dual frame of $\mathcal{G}(g, \alpha, \beta)$ is given by $\mathcal{G}(\gamma, \alpha, \beta)$, see Gröchenig (2013, Proposition 5.2.1). It is the canonical dual frame. As a result, every $f \in L^2(\mathbb{R}^d)$ has the following non-orthogonal expansions

$$f = \sum_{k,n\in\mathbb{Z}^d} \langle f, T_{\alpha k} M_{\beta n}\, \gamma\rangle T_{\alpha k} M_{\beta n}\, g = \sum_{k,n\in\mathbb{Z}^d} V_\gamma f(\alpha k, \beta n) M_{\beta n} T_{\alpha k}\, g$$

and

$$f = \sum_{k,n\in\mathbb{Z}^d} \langle f, T_{\alpha k} M_{\beta n}\, g\rangle T_{\alpha k} M_{\beta n}\, \gamma = \sum_{k,n\in\mathbb{Z}^d} V_g f(\alpha k, \beta n) M_{\beta n} T_{\alpha k}\, \gamma,$$

with unconditionally convergence in $L^2(\mathbb{R}^d)$. This is a discrete version of the weak inversion formula for the STFT. It provides an explicit reconstruction of the signal f from samples of its short-time Fourier transform. We obtain a discrete time-frequency representation of signals. In addition, there are frame bounds $A, B > 0$, such that the following norm equivalences arise from the frame conditions:

$$A\|f\|_2^2 \leq \sum_{k,n\in\mathbb{Z}^d} |V_g f(\alpha k, \beta n)|^2 \leq B\|f\|_2^2 \tag{2.23}$$

$$B^{-1}\|f\|_2^2 \leq \sum_{k,n\in\mathbb{Z}^d} |V_\gamma f(\alpha k, \beta n)|^2 \leq A^{-1}\|f\|_2^2. \tag{2.24}$$

This equivalence provides a characterization for square integrable functions. A function f belongs to $L^2(\mathbb{R}^d)$ if and only if the l^2-norm of the STFT is finite. More results and details can be found in Gröchenig (2013, Chapter 5).

2.3. Modulation Spaces and their Characterization through Gabor Frames

In this section, we will consider function spaces characterized by time-frequency representations. It's well known that the smoothness and decay of a function are closely related to the decay and smoothness of its Fourier transform, and this relationship can

be extended to the short-time Fourier transform. The STFT is a critical component of time-frequency analysis because it provides a simultaneous representation of the magnitude of the function and its Fourier transform near a given point in time and frequency, respectively. This relationship means that the decay of the function is closely related to the decay of the STFT with respect to time and that the smoothness of the function is closely related to the decay of the STFT with respect to frequency. In this way, the STFT acts as a bridge between the time and frequency aspects of a function and provides a useful tool for relating these properties. Modulation spaces were invented by Feichtinger (1983) and are spaces of tempered distributions defined by the decay properties of their short-time Fourier transforms in the $L^p(\mathbb{R}^d)$ space. Just like Sobolev spaces, which can be characterized by their Fourier representations, and Besov spaces, which can be described by dyadic decompositions of the Fourier transform of their elements, modulation spaces are unique in that they are defined by the properties of the short-time Fourier transform. Modulation spaces are defined by imposing decay and integrability conditions on the STFT. To fully understand the concept of modulation spaces, it is important to first quantify decay in the time-frequency plane, $\mathbb{R}^{2d}$. This will provide a foundation for the introduction and characterization of modulation spaces.

2.3.1. Weight Functions

Weight functions are used to quantify decay in the time-frequency plane $\mathbb{R}^{2d}$. They are non-negative, locally bounded, and measurable, i.e. they can be integrated over any compact subset of $\mathbb{R}^{2d}$. They play a crucial role in characterizing the decay properties of functions in modulation spaces.

Definition 8. Let consider two weight functions v and m on $\mathbb{R}^{2d}$. We say that:

$$\begin{aligned} v \text{ is submultiplicative:} \quad & v(z_1 + z_2) \leq v(z_1)v(z_2) \\ m \text{ is } v\text{-moderate:} \quad & m(z_1 + z_2) \leq Cv(z_1)m(z_2) \\ m \text{ and } v \text{ are equivalent, } m \asymp v: \quad & C^{-1}m(z) \leq v(z) \leq Cm(z) \end{aligned}$$

for all $z, z_1, z_2 \in \mathbb{R}^{2d}$.

Note that Definition 8 implies $v(0) \geq 1$. The default class of weights on $\mathbb{R}^{2d}$ are polynomial type weights defined as

$$v_s(z) = (1 + \|z\|_2^2)^{s/2} \asymp (1 + \|x\|_2^2 + \|\omega\|_2^2)^{s/2} \tag{2.25}$$

for $s \geq 0$. The polynomial weight function v_s is submultiplicative, and both v_t and v_t^{-1} are v_s-moderate with $0 \leq t \leq s$. We say that a function $F(z)$ has a rapid decay on $\mathbb{R}^{2d}$ if the function decays faster than any polynomial weight function, i.e., for all $n \geq 0$ there exists a constant $C_n > 0$ such that

$$|F(z)|\, v_n(z) \leq C_n \tag{2.26}$$

Faster decay in the time-frequency plane can also be described with exponential weight functions defined as

$$v(z) = e^{\alpha |z|^\beta} \tag{2.27}$$

for $\alpha > 0$ and $0 \leq \beta < 1$. Throughout this thesis a particular class of weight functions plays an important role. For $0 \leq u, v \leq s$ we define the v_s-moderate weight function

$$m_{u,v}(x, \omega) = (1 + \|x\|_2^2)^{u/2} + (1 + \|\omega\|_2^2)^{v/2}$$

If $u = v$ the weight function is isotropic, i.e. it describes a similar decay in time and frequency. An anisotropic weight function, $u \neq v$, describes different decay properties in time and frequency.

Lemma 6. *Let $u, v \geq 0$.*

1. *If $u = v = s$, then $m_{s,s} \asymp v_s$.*
2. *$m_{u,v}$ is submultiplicative.*

We focus on polynomial weight functions in order to stay within the familiar setup of Schwartz functions and tempered distributions. In doing so, we can use $\mathcal{S}(\mathbb{R}^d)$ as a suitable class of test functions and windows, and thus the modulation spaces can be defined as subspaces of $\mathcal{S}'(\mathbb{R}^d)$. In the remainder of this work we will use v to denote a submultiplicative weight and m to denote a v-moderate weight. Furthermore, we will assume without loss of generality that v is continuous and symmetric in each coordinate, i.e.

$v(x,\omega) = v(-x,\omega) = v(x,-\omega) = v(-x,-\omega)$. It is useful to think of a weight function m as a pair (m,v), since one will prove that the results hold simultaneously for the class of all v-moderate weights. Several results depend only on the submultiplicative weight v, but not on individual v-moderate weights m. Next, we collect some results on the decay property, which connects the decay of a function f with the decay of its time-frequency representation.

Theorem 7. *Let $g \in \mathcal{S}(\mathbb{R}^d) \setminus \{0\}$, then*

1. *If $f \in \mathcal{S}'(\mathbb{R}^d)$, then there are constants $C > 0, N \geq 0$, such that*

$$|V_g f(x,\omega)| < C(1 + |x| + |\omega|)^N \quad \text{for all } x,\omega \in \mathbb{R}^d$$

2. *If $f \in \mathcal{S}'(\mathbb{R}^d)$, then*

$$f \in \mathcal{S}(\mathbb{R}^d) \iff V_g f \text{ has rapid decay.}$$

Proof. Gröchenig (2013, Section 11.2). □

Theorem 7 establishes the equivalence between rapid decay of the short-time Fourier transform and membership in the Schwartz space and gives a characterization of $\mathcal{S}(\mathbb{R}^d)$ via STFT. This leads to the following interpretations: The STFT of a slow growing function f has polynomial growth, i.e. also slow growing, whereas the STFT of a fast decaying function decays also fast.

2.3.2. Mixed-Norm Spaces

We will now examine mixed-norm spaces on $\mathbb{R}^{2d}$, which provide the necessary collection of (quasi-)Banach spaces and norms to function effectively. It's important to note that the notion of mixed-norm is not limited to the time-frequency plane, but can be applied to any product measure space.

Definition 9. Let m be a weight function on $\mathbb{R}^{2d}$ and let $0 < p, q < \infty$. The weighted mixed-norm space $L^{p,q}_m(\mathbb{R}^{2d})$ consists of all (Lebesgue) measurable functions on $\mathbb{R}^{2d}$, such that the (quasi-)norm

$$\|F\|_{L^{p,q}_m} = \left(\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |F(x,\omega)|^p m(x,\omega)^p dx \right)^{q/p} d\omega \right)^{1/q}$$

is finite. If $p = \infty$ or $q = \infty$, then the corresponding p-norm is replaced by the essential supremum.

This function space is obtained by taking a weighted L^p-norm with respect to the time and a L^q-norm with respect to the frequency. If $p = q$, then $L^{p,q}_m(\mathbb{R}^{2d}) = L^p_m(\mathbb{R}^{2d})$ is the usual weighted L^p-space. The mixed-norm spaces enjoy the same properties as L^p-spaces. They are quasi-Banach spaces (Banach spaces if $1 \leq p, q \leq \infty$) and invariant under translation T_z, $z \in \mathbb{R}^{2d}$. Furthermore, the Holder's inequality holds with $(L^{p,q}_m(\mathbb{R}^{2d}))' = L^{p',q'}_{1/m}(\mathbb{R}^{2d})$ and $\frac{1}{p} + \frac{1}{p'} = \frac{1}{q} + \frac{1}{q'} = 1$, see Gröchenig (2013, lemma 11.1.2). Also the convolution relation, that is $L^1_v * L^p_m \subseteq L^p_m$, extends to mixed-norm space. For more details on the theory of mixed-norm spaces, see Benedek and Panzone (1961).

Lemma 8. *Let m be v-moderate and let $0 < p, q \leq \infty$, then*

$$\|T_z F\|_{L^{p,q}_m} \leq C\, v(z) \|F\|_{L^{p,q}_m} \tag{2.28}$$

for $z \in \mathbb{R}^{2d}$.

Proof. See Gröchenig (2013, Lemma 11.1.2) for $1 \leq p, q \leq \infty$. The proof for $0 < p, q < 1$ is analogous. □

In the discrete setting, the discrete mixed-norm space $\ell^{p,q}_m(\mathbb{Z}^{2d})$ is defined by all sequences $a = a_{kn}{}_{k,n\in\mathbb{Z}^d}$ for which the (quasi-)norm

$$\|a\|_{\ell^{p,q}_m} = \left(\sum_{n\in\mathbb{Z}^d} \left(\sum_{k\in\mathbb{Z}^d} |a_{kn}|^p\, m(k,n)^p \right)^{q/p} \right)^{1/q}. \tag{2.29}$$

is finite. This space induces in particular the amalgam space $W(L^{p,q}_m)$, space of functions with local suprema in $\ell^{p,q}_m(\mathbb{Z}^{2d})$ with $\|f\|_{W(L^{p,q}_m)} = \|a\|_{\ell^{p,q}_m(\mathbb{Z}^{2d})}$ where the sequence a

represents the local suprema of f. Amalgam spaces establish an important connection between the weighted mixed-norm space $L^{p,q}_m(\mathbb{R}^{2d})$ and the discrete mixed-norm space $\ell^{p,q}_m(\mathbb{Z}^{2d})$.

2.3.3. Modulation Spaces

Now we come to the definition of modulation spaces. There is an extensive literature on modulation spaces $\mathcal{M}^{p,q}_m$ for the case $1 \leq p, q \leq \infty$, see Gröchenig (2013), or Feichtinger (1983). Galperin and Samarah (2004) extended this to the general case $0 \leq p, q \leq \infty$. For simplicity, we will focus on the case $p = q$, but it is worth noting that the results can also be applied to the case where $p \neq q$. In the following we consider a polynomial weight function v. Thus, we remain in the setting of Schwartz space. The following results are valid for all v-moderate weight function m.

Modulation spaces are defined by imposing decay and integrability conditions on the STFT as follows:

Definition 10. Fix a nonzero window $g \in \mathcal{S}(\mathbb{R}^d)$, a v-moderate weight function m on $\mathbb{R}^{2d}$, and $0 < p, q \leq \infty$. Then the modulation space $\mathcal{M}^{p,q}_m(\mathbb{R}^d)$ consists of all tempered distributions $f \in \mathcal{S}'(\mathbb{R}^d)$ such that $V_g f \in L^{p,q}_m(\mathbb{R}^{2d})$. The (quasi-)norm on $\mathcal{M}^p_m(\mathbb{R}^d)$ is defined as

$$\|f\|_{\mathcal{M}^{p,q}_m} = \|V_g f\|_{L^{p,q}_m} \tag{2.30}$$

If $p = q$, then we write $\mathcal{M}^p_m(\mathbb{R}^d)$ instead of $\mathcal{M}^{p,q}_m$, and if $m(z) \equiv 1$ on $\mathbb{R}^{2d}$, we then write $\mathcal{M}^p(\mathbb{R}^d)$.

Elements of the modulation space $\mathcal{M}^p_m(\mathbb{R}^d)$ have similar decay properties in the time-frequency domain, $\mathbb{R}^{2d}$. $\mathcal{M}^\infty_m(\mathbb{R}^d)$ consists of all (measurable) functions f with bounded weighted short-time Fourier transform, i.e. $|V_g f(z)| \cdot m(z) < C$. Modulation spaces are Banach spaces for $1 \leq p \leq \infty$. They inherit many properties of mixed-norm spaces. They are invariant under time-frequency shifts and $(\mathcal{M}^p_m)' = \mathcal{M}^{1-1/p}_{1/m}$ for $1 \leq p < \infty$. These properties, as well as further details on modulation spaces, can be found in Gröchenig (2013, Section 11.3). This makes modulation spaces a powerful tool for signal processing

and time-frequency analysis.

The Definition 10 of the modulation space is independent of the choice of the window function $g \in \mathcal{S}(\mathbb{R}^d)$, since different windows lead to equivalent norms. Galperin and Samarah (2004) extends the space of admissible windows to the following.

Definition 11. Let m be a v-moderate, $0 < p, q \leq \infty$, $r = \min(1, p)$ and $s = \min(1, p, q)$. For $r_1, s_1 > 0$, denote

$$w_{r_1,s_1}(x, \omega) = v(x, \omega) \cdot (1 + |x|)^{r_1} \cdot (1 + |\omega|)^{s_1} \tag{2.31}$$

Define the space of admissible windows $\mathfrak{M}_v^{p,q}$ for the modulation space $\mathcal{M}_m^{p,q}$ to be

$$\mathfrak{M}_v^{p,q} = \bigcup_{r_1 > d/r, s_1 > d/s, 1 \leq p_1 < \infty} \mathcal{M}_{w_{r_1,s_1}}^{p_1}. \tag{2.32}$$

For $g \in \mathfrak{M}_v^{p,q}$, $\|V_g f\|_{L_m^{p,q}}$ yields an equivalent (quasi-)norm on $\mathcal{M}_m^{p,q}$, i.e. different window functions $g \in \mathfrak{M}_v^{p,q}$ yield equivalent modulation space (quasi-)norm, see Galperin and Samarah (2004, Theorem 3.1). For $1 \leq p, q \leq \infty$ the class of admissible window functions can be extended from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{M}_v^1$. This is in contrast to the short-time Fourier transform (STFT), where the definition depends on the choice of the window function.

Another reason for the usefulness of modulation spaces is that many well-known function spaces can be identified with modulation spaces for certain weights, in particular the Sobolev space $W_p^s(\mathbb{R}^d)$,

$$W_p^s(\mathbb{R}^d) = \left\{ f \in \mathcal{S}'(\mathbb{R}^d) : D^\alpha f \in L^p(\mathbb{R}^d), \qquad |\alpha| \leq s \right\} \tag{2.33}$$

The following observations have been proven in Gröchenig (2013, Prop. 11.3.1)

1) $\mathcal{M}^2(\mathbb{R}^d) = L^2(\mathbb{R}^d)$.

2) If $m(x, \omega) = m(x)$, then $\mathcal{M}_m^2(\mathbb{R}^d) = L_m^2(\mathbb{R}^d)$.

3) If $m(x, \omega) = m(\omega)$, then $\mathcal{M}_m^2(\mathbb{R}^d) = \mathcal{F}L_m^p(\mathbb{R}^d)$.

4) For $m = m_{0,v}$, the space $\mathcal{M}^2_m(\mathbb{R}^d)$ corresponds to the Sobolev space $W^v_2(\mathbb{R}^d)$, the Bessel potential space.

The Schwartz space $\mathcal{S}(\mathbb{R}^d)$, which consists of all smooth and fast decaying functions on $\mathbb{R}^d$, can be characterized using modulation spaces. Since smoothness implies rapid decay in $\mathbb{R}^{2d}$, see Theorem 7, a Schwartz function is contained in all polynomial modulation spaces. Conversely, any tempered distribution that decays at most polynomially is contained in some polynomial modulation space. This is expressed by the following relationships:

$$\mathcal{S}(\mathbb{R}^d) = \bigcap_{s\geq 0} \mathcal{M}^\infty_{v_s} \quad \text{and} \quad \mathcal{S}'(\mathbb{R}^d) = \bigcup_{s\geq 0} \mathcal{M}^\infty_{1/v_s} \tag{2.34}$$

where v_s are polynomial weight functions and $\mathcal{M}^\infty_{v_s}$ are modulation spaces. Gröchenig (2013) has also proved that $\mathcal{S}(\mathbb{R}^d)$ is a dense subspace of $\mathcal{M}^p_m(\mathbb{R}^d)$ if there exists a polynomial that grows faster that the weight function m, that is $|m(z)| \leq C(1+|z|)^N$, and $0 < p < \infty$. This means that the Schwartz class $\mathcal{S}$ is the largest natural class of windows admissible for all modulation spaces $\mathcal{M}^{p,q}_m$ with $0 < p, q \leq \infty$.
These and other properties, such as the embedding property, $\mathcal{M}^{p_1}_{m_1}(\mathbb{R}^d) \subseteq \mathcal{M}^{p_2}_{m_2}(\mathbb{R}^d)$ for $p_1 \leq p_2$ and $m_2 \leq C\, m_1$, can be found in Gröchenig (2013) or Galperin and Samarah (2004). A generalization of modulation spaces for non-polynomial weights is given in Gröchenig (2013).

Remark. Polynomial weight functions play a major role in this work. To further emphasize this, we denote polynomial modulation spaces by $\mathcal{M}^{p,q}_{u,v} := \mathcal{M}^{p,q}_{m_{u,v}}$, for $u, v \geq 0$.

2.3.4. Gabor Expansion for Modulation Spaces

Modulation spaces arise naturally in the study of Gabor expansions. In this section we show that Gabor frames can also be frames for the modulation space $\mathcal{M}^p_m(\mathbb{R}^d)$. This would give a non-orthogonal expansion of functions $f \in \mathcal{M}^p_m(\mathbb{R}^d)$ in terms of time-frequency elements. This proves once again why modulation spaces are the correct

function spaces for quantitative time-frequency analysis.

We consider a Gabor system $\mathcal{G}(g, \alpha, \beta)$ and define a restriction of the weight function to the lattice $\tilde{m}(k, n) = m(\alpha k, \beta n)$. The next result gives a sufficient condition for frames on $\mathcal{M}^p_m(\mathbb{R}^d)$.

Theorem 9. *If $g \in \mathfrak{M}^{p,q}_v$, then the analysis operator T_g and the synthesis operator T_g^* are respectively bounded from $\mathcal{M}^{p,q}_m$ into $l^{p,q}_{\tilde{m}}(\mathbb{Z}^{2d})$ and from $l^{p,q}_{\tilde{m}}(\mathbb{Z}^{2d})$ into $\mathcal{M}^{p,q}_m$. Furthermore, if $p < \infty$, then the synthesis operator converges unconditionally in $\mathcal{M}^p_m(\mathbb{R}^d)$.*

Proof. Galperin and Samarah (2004, Theorem 3.5 & 3.6) □

Theorem 9 leads to the boundedness of the Gabor operator $S^{\alpha,\beta}_{g,\gamma}$ on $\mathcal{M}^p_m(\mathbb{R}^d)$, with $g, \gamma \in \mathfrak{M}^{p,q}_v$. Let $g \in \mathfrak{M}^{p,q}_v$ be an admissible window, and assume that there exists $\gamma \in \mathfrak{M}^{p,q}_v$ with $S^{\alpha,\beta}_{g,\gamma} = I$. As a result, each $f \in \mathcal{M}^p_m(\mathbb{R}^d)$ has the following non-orthogonal expansions

$$\begin{aligned} f &= \sum_{k,n \in \mathbb{Z}^d} \langle f, T_{\alpha k} M_{\beta n}\, \gamma \rangle T_{\alpha k} M_{\beta n}\, g \\ &= \sum_{k,n \in \mathbb{Z}^d} \langle f, T_{\alpha k} M_{\beta n}\, g \rangle T_{\alpha k} M_{\beta n}\, \gamma, \end{aligned}$$

with unconditional convergence in $\mathcal{M}^p_m(\mathbb{R}^d)$ if $p < \infty$. Furthermore, there exist constants $A, B > 0$ such that

$$A \left\| f \right\|_{\mathcal{M}^p_m(\mathbb{R}^d)} \leq \sum_{k,n \in \mathbb{Z}^d} \langle f, T_{\alpha k} M_{\beta n}\, g \rangle \tilde{m}(k, n) \leq B \left\| f \right\|_{\mathcal{M}^p_m(\mathbb{R}^d)} \tag{2.35}$$

See Galperin and Samarah (2004, Theorem 3.7) for a proof. The (quasi-)norm equivalence 2.35 implies that a function belongs to $\mathcal{M}^p_m(\mathbb{R}^d)$ if its Gabor coefficients belong to $l^{p,q}_{\tilde{m}}(\mathbb{Z}^{2d})$. Several modulation spaces can thus be characterized by the decay of the Gabor coefficients. For the norm equivalence, it is sufficient to check whether $g \in \mathfrak{M}^{p,q}_v$. In this thesis, we will mainly work with Gaussian functions and B-splines as window function. A Gaussian function belongs to the Schwartz space and is therefore admissible for modulation spaces. We have shown in Dahlke et al. (2022) that for splines of order

k the norm equivalence holds for $p > \frac{1}{k}$.

For the expansions in terms of Gabor elements, it remains to check whether both the window and the reconstruction window are admissible. Gröchenig (2013) has shown that for $1 \leq p, q \leq \infty$, it is sufficient to check invertibility of $S_{g,g}^{\alpha,\beta}$ on $\mathcal{M}_v^1$. For example, $S_{g,g}^{\alpha,\beta}$ is invertible for tight frames. In this thesis we will mainly work with Gaussian functions and splines as window functions. The dual of a Gaussian is a Gaussian, so the condition is satisfied. Dahlke et al. (2022) has shown that for splines of order k, the Gabor operator is invertible for $p > \frac{1}{k}$.

2.4. Flexible Gabor-Wavelet Transform

The wavelet transform (WT) is another widely used transformation in (audio) signal processing, introduced in the 1980s by Grossman and Morlet, see Grossmann et al. (1985). It is based on the representation of the affine group over $L^2(\mathbb{R}^d)$. Wavelet theory in signal analysis is well represented in the literature, with the current state of the art described in Mallat (2009). Unlike the STFT, which is parametrized by time and frequency, the WT replaces frequencies with the notion of scale. The frequency shifts in the definition of STFT are replaced by dilations. The dilation operator, also known as the scaling operator, is a unitary operator on $L^2(\mathbb{R}^d)$ that changes the scale of a function by stretching or shrinking it along one or more axes. The shape is preserved. The essential support is either increased or decreased. The continuous wavelet transform of a function $f \in L^2(\mathbb{R}^d)$ with respect to an analyzing function g, also called a wavelet, is defined as follows

$$W_g f(x, a) = \langle f, T_x D_a\, g \rangle_{L^2(\mathbb{R}^d)} = \langle \hat{f}, M_{-x} D_{a^{-1}}\, \hat{g} \rangle_{L^2(\mathbb{R}^d)} \tag{2.36}$$

with the dilation operator defined as $D_a f(t) = a^{-d/2} f(t/a)$ for $a > 0$. Although the WT is not an explicit time-frequency representation, it still has some frequency localization. Furthermore, if the wavelet g is admissible, i.e., satisfies additional orthogonality relations, then there exists a reconstruction wavelet h with $W_h^* W_g = \mathcal{I}$ such that any function $f \in L^2(\mathbb{R}^d)$ can be reconstructed from its wavelet transform.

The STFT and WT have distinct advantages and disadvantages and may not be equally relevant in different applications. By scaling the analyzing function up and down, the wavelet transform acts as a mathematical microscope. If the wavelet g is centered at the origin with variance 1, then the support of $T_x D_a g$ is a neighborhood of x of size a. As a result, $W_g f(x,a) = \langle f, T_x D_a g \rangle$ encodes local information of f at x. This property allows the wavelet transform to be used to analyze pointwise and microlocal regularity. In contrast, the STFT lacks this microlocal property because the size of the window function remains constant. The main disadvantage of the wavelet transform is its poor angular resolution, i.e., its ability to distinguish small frequency details. With the scale a going from 0 to ∞, $\langle \hat{f}, M_{-x} D_{a^{-1}} \hat{g} \rangle$ gives information about a conical set of frequencies. In contrast, the STFT isolates the central frequency.

Similar to the theory of the STFT, the continuous WT can be discretized. This leads to the definition of wavelet frames, see Gröchenig (2013, Chapter 10). Wavelet frames encode information about the smoothness and singularity properties of distributions and provide a characterization for Besov spaces. Gabor frames encode time-frequency information and characterize modulation spaces. The WT is particularly effective for studying signals that are mostly smooth with isolated singularities, while the Gabor transform is better suited for analyzing periodic structures.

In several relevant contributions, e.g. Nazaret and Holschneider (2003), Fornasier and Feichtinger (2006), an "intermediate" time-frequency transform between wavelet and short-time Fourier transform is considered. This allows a flexible characterization of function spaces between modulation spaces and Besov spaces.

Definition 12 (Gabor-Wavelet transform)**.** Let $\alpha \in [0,1]$ and $c > 0$. For any $f \in L^2(\mathbb{R}^d)$ and a nonzero analyzing function $g \in L^2(\mathbb{R}^d)$, the flexible Gabor-Wavelet transform (or α-transform or GWT) is defined by

$$V_g^\alpha f(x,\omega) = \langle f, T_x M_\omega D_{\eta_\alpha^{-1}(\omega)} g \rangle \tag{2.37}$$

with $\eta_\alpha(\omega) = c(1+|\omega|)^\alpha$ for $x, \omega \in \mathbb{R}^d$.

The GWT analyzes the function f around x at the scale $\eta_\alpha(\omega)^{-1}$. The parameter $\alpha \in [0,1]$ controls the relationship between the frequency and the dilation factor. For $\alpha = 0$, the dilation operator becomes obsolete and the transform $V_g^\alpha f$ coincides with the STFT $V_g f$, while for $\alpha \to 1$ we obtain a slight modification of the wavelet transform. Especially the intermediate case $\alpha = 1/2$ appears in the literature as the Fourier-Bros-Iagolnitzer transform Bros and Iagolnitzer (1975). Additionally, for an analyzing function $g \in \mathcal{S}(\mathbb{R}^d)$, $V_g^\alpha f$ defines a continuous mapping from $\mathcal{S}(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ and can be extended to $L^2(\mathbb{R}^d)$ by continuity, see Nazaret and Holschneider (2003, Theorem 2.1).

Several authors provide characterizations of function spaces based on the GWT. Nazaret and Holschneider (2003) use the flexible Gabor-Wavelet Transform to characterize L^2-Sobolev spaces, which are both modulation and Besov spaces, and to construct anisotropic Banach spaces of functions. For an admissible Banach space B in the time-frequency plane $\mathbb{R}^{2d}$, a Banach space $\mathcal{B}(\mathbb{R}^d)$ of distribution is defined as follows

$$\mathcal{B}(\mathbb{R}^d) := \left\{ f \in \mathcal{S}'(\mathbb{R}^d) : \qquad V_g^\alpha(f) \in B \right\} \tag{2.38}$$

A norm on $\mathcal{B}(\mathbb{R}^d)$ is then given by $\|f\|_{\mathcal{B}(\mathbb{R}^d)} = \|V_g^\alpha(f)\|_B$. See Nazaret and Holschneider (2003, Definition 4.7) for more details. Fornasier (2004) has shown that for the choice of the Banach space B as a certain weighted Lebesgue mixed norm $L_m^{p,q}(\mathbb{R}^{2d})$ space, the corresponding Banach space $\mathcal{B}(\mathbb{R}^d)$ is a so-called α-modulation space. A smoothness space based on the behavior of the spectral decomposition of its constituents. Dahlke et al. (2008) have shown that the GWT can be interpreted as the voice transform associated with a square-integrable representation modulo quotients of the affine Weyl-Heisenberg group, hence the Banach space $\mathcal{B}(\mathbb{R}^d)$ represents the so-called coorbit space, cf. Feichtinger and Gröchenig (1989).

2.4.1. α-Modulation Spaces

We now consider the one dimensional setting $d = 1$. The smoothness spaces associated with the flexible Gabor-Wavelet transform are the α-modulation spaces, which can be characterized by the GWT transform coefficients. They were introduced by Gröbner based on the Littlewood-Paley theory, which uses a decomposition of a function f into

a sum of frequencies localized functions, see Feichtinger and Gröbner (1985) for more details. This defines a Fourier analytical characterization of smoothness spaces between modulation spaces and Besov spaces. A proper definition of these spaces involves notions like spectral decomposition, Fourier multiplier, Littlewood-Paley theorem, admissible covering, BAPUs (bounded admissible partitions of unity), decomposition spaces, and Wiener amalgrams, notions for which we refer to the literature for further details, cf. Feichtinger (1987), Feichtinger and Gröbner (1985) or Fornasier (2004).

We now review the construction of α-modulation spaces. An α-modulation space measures the smoothness of its elements by examining the behavior of their spectral decomposition. They are constructed as follows: given a parameter $\alpha \in [0, 1]$, an admissible α-covering of the frequency space, $\mathcal{I}_\alpha$, is constructed, see Definition 15. The parameter α determines the size of the intervals. For $\alpha = 0$, they are set to the same length as in the Gabor case, whereas for $\alpha \to 1$ the length grows exponentially with the frequency in a manner defined by α. A set of $\mathcal{F}L_1$-integrable Schwartz functions $\{\psi_I^\alpha\}_{I\in\mathcal{I}_\alpha}$, called bounded partition of the unity (BAPU), is then defined on this coverage of the frequency space, see Definition 16. Their sum gives a characteristic function of the frequency space $\mathbb{R}$.

The previous setting gives rise to a spectral decomposition of any tempered distribution in entire analytic functions. Then any $f \in \mathcal{S}'(\mathbb{R})$ can be written as

$$f = \sum_{I\in\mathcal{I}_\alpha} \mathcal{F}^{-1}(\psi_I^\alpha \mathcal{F} f), \tag{2.39}$$

where each summand is a tempered distribution whose Fourier transform is compactly supported, and thus an entire analytic function. See the Paley-Wiener-Schwartz theorem for more details. The convergence is defined in the space of tempered distribution $\mathcal{S}'(\mathbb{R})$. This band-limited approximation of function gives rise to a notion of smoothness for $f \in \mathcal{S}'(\mathbb{R})$ based on the behavior of its spectral elements.

Definition 13. Given $1 \le p, q \le \infty$, $s \in \mathbb{R}$, and $0 \le \alpha \le 1$, let $\mathcal{I}_\alpha$ be an α-covering of $\mathbb{R}$ and let $\{\psi_I^\alpha\}_{I\in\mathcal{I}_\alpha}$ be a corresponding BAPU. The α-modulation space $\mathcal{M}^{p,q}_{0,s:\alpha}(\mathbb{R})$ is then defined for $q < \infty$ and $s \in \mathbb{R}$ as the set of tempered distributions $f \in \mathcal{S}'(\mathbb{R})$ satisfying

$\|f\|_{\mathcal{M}^{p,q}_{0,s:\alpha}} < \infty$ with

$$\|f\|_{\mathcal{M}^{p,q}_{0,s;\alpha}} := \left(\sum_{I \in \mathcal{I}_\alpha} \|\mathcal{F}^{-1}(\psi_I^\alpha \mathcal{F} f)\|_p^q (1 + |\omega_I|)^{sq} \right)^{1/q} \tag{2.40}$$

where $\omega_I \in I$ is an arbitrary frequency.

We choose this notation to emphasize the fact that the polynomial weight function depends only on the frequency. It also helps display the link to modulation spaces. For $p = q$, we write $\mathcal{M}^p_{0,s:\alpha}(\mathbb{R})$ instead of $\mathcal{M}^{p,q}_{0,s:\alpha}(\mathbb{R})$. The α-modulation norm is a discrete norm. The space $\mathcal{M}^p_{0,s:\alpha}(\mathbb{R})$ corresponds to the *decomposition space* subordinate to the covering $\mathcal{I}_\alpha$, with local component $\mathcal{F}L^p(\mathbb{R})$, and global component $l^q_m(\mathcal{I}_\alpha)$, where $m(I) = (1 + \omega_I)^s$. From the properties of the covering $\mathcal{I}_\alpha$ it follows that the definition of $\mathcal{M}^{p;q}_{s,\alpha}$ doesn't depend on any particular choice of $\{\omega_I\}_{I \in \mathcal{I}}$ either. In Feichtinger and Gröbner (1985), the authors showed that this space does not depend on a particular choice of the α-covering $\mathcal{I}_\alpha$ nor on the corresponding BAPU $\{\psi_I^\alpha\}_{I \in \mathcal{I}_\alpha}$.

For $b > 0$ and $\alpha \in [0, 1)$, an explicit example of α-covering was constructed in Fornasier and Feichtinger (2006). The following position function $p_\alpha : \mathbb{Z} \to \mathbb{R}$ and size function $s_\alpha : \mathbb{Z} \to \mathbb{R}^+$

$$\begin{aligned} p_\alpha(j) &= \operatorname{sgn}(j) \left((1 + (1 - \alpha) \cdot b \cdot |j|)^{\frac{1}{1-\alpha}} - 1 \right) \\ s_\alpha(j) &= b \cdot (1 + (1 - \alpha) \cdot b \cdot |j|)^{\frac{\alpha}{1-\alpha}} \end{aligned} \tag{2.41}$$

help identify the constituent intervals for an α-covering by

$$I_j^\alpha = \begin{cases} p_\alpha(j) + \operatorname{sgn}(j)[0, s_\alpha(j)] & \text{for } p_\alpha(j) \neq 0 \\ [-s_\alpha(j), s_\alpha(j)] & \text{else} \end{cases}$$

for $j \in \mathbb{Z}$. For $\alpha \to 1$ it holds

$$\mathcal{I}_1 = \{\operatorname{sgn}(j) \left((e^{b|j|} - 1) + [0, e^{b|j|+1}] \right)\}_{j \in \mathbb{Z} \setminus \{0\}} \cup \{[-e^b, e^b]\}$$

Without loss of generality, we can assume the previously defined position and size functions $p_\alpha(j)$ and $s_\alpha(j)$, as well as $\omega_{I_j} = p_\alpha(j)$. For $\alpha = 0$, the frequency intervals are

of equal length. Note that modulation spaces are independent of the window function. Thus, $\mathcal{M}^{p,q}_{0,s:0}$ coincides with the modulation spaces $\mathcal{M}^{p,q}_{0,s}$. Consider $b = \ln(2)$. For $\alpha \to 1$, we obtain a dyadic representation of the frequency domain with

$$I_j^1 = \begin{cases} [-2, 2] & \text{for } j = 0 \\ \operatorname{sgn}(j) \cdot [2^{j-1}, 2^{j+1}] & \text{else} \end{cases} \tag{2.42}$$

which is an α-covering. With this covering, definition 13 is equivalent to the Fourier analytical characterization of Besov spaces using a dyadic resolution of the unity in $\mathbb{R}$, see Triebel (1992, Theorem 1.3.4). Thus, the space $\mathcal{M}^{p,q}_{s,1}$ corresponds to the inhomogeneous Besov space $B^s_{p,q}(\mathbb{R})$.

The following result by Fornasier (2004) is an application of Feichtinger (1987, Theorem 4.3) showing that the discrete α-modulation norm and the continuous norm on the anisotropic Banach space $\left\{ f \in \mathcal{S}'(\mathbb{R}) : \quad V_g^\alpha(f) \in L^{p,q}_m(\mathbb{R}^2) \right\}$ are equivalent.

Theorem 10. *Assume $s \in \mathbb{R}$, $\alpha \in [0, 1]$, and $1 \leq p, q < \infty$. For a suitable band-limited $g \in \mathcal{S}(\mathbb{R}) \setminus \{0\}$*

$$\mathcal{M}^{p,q}_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R}) = \left\{ f \in \mathcal{S}'(\mathbb{R}) : \quad V_g^\alpha(f) \in L^{p,q}_{m_s}(\mathbb{R}^2) \right\} \tag{2.43}$$

with the following norm equivalence

$$\|f\|_{\mathcal{M}^{p,q}_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})} \asymp \|V_g^\alpha(f)\|_{L^{p,q}_{m_s}} \tag{2.44}$$

with $m_s(x, \omega) = (1 + \omega)^s$. For $p \cdot q = \infty$ the usual modification apply.

Proof. Fornasier and Feichtinger (2006, Theorem 3.5) □

In addition, Fornasier and Feichtinger (2006) provide a Banach frame and atomic decomposition for α-modulation spaces. This frame expansion for modulation spaces is based on the theory of localization of frames introduced by Gröchenig, see Gröchenig (2004). A set of frames is said to be (polynomially or exponentially) localized if its Gramian matrix has a (polynomial or exponentially) off-diagonal decay property, see Gröchenig (2004, Definition 7). A localized frame not only has a localized dual with the same off-diagonal

decay, but also defines a Banach frame for an associated Banach space $\mathcal{H}_m^p$, an abstract class of Banach spaces defined by

$$\mathcal{H}_m^p = \left\{ f \in \mathcal{H} \,:\, f = \sum c_n g_n \quad \text{for} \quad c \in \ell_m^p(\mathcal{N}) \right\} \tag{2.45}$$

with $\{g_n : n \in \mathcal{N}\}$ a Riesz basis of the Hilbert space $\mathcal{H}$, $\ell_m^p(\mathcal{N}) \subseteq \ell^2(\mathcal{N})$ and norm $\|f\|_{\mathcal{H}_m^p} = \|c\|_{l_m^p}$, cf. Gröchenig (2004, Definition 4). In other words, a polynomial (exponential) localized frame is also a Banach frame for an associated Banach space induced by a polynomial (sub-exponential) weight function, see Gröchenig (2004, Theorem 13).

Remark 2. The smoothness space, as defined in Theorem 10, consists of all functions whose voice transform has a certain decay. This highlights the dependence of the α-modulation construction on the theory of coorbit spaces and establishes a connection between coorbit spaces and α-modulation spaces, as described in Dahlke et al. (2008).

Let $\alpha \in [0, 1)$. For the pair of position and size functions (p_α, s_α) defined in (2.41) and an admissible analyzing function $h \in \mathcal{S}(\mathbb{R})$, we consider the following family

$$\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b) = \left\{ M_{p_\alpha(j)} D_{s_\alpha^{-1}(j)} T_{ak}\, h \right\}_{(j,k)\in\mathbb{Z}^2}, \qquad a, b > 0. \tag{2.46}$$

Fornasier and Feichtinger (2006) use the concept of intrinsic localization for frames on Hilbert spaces to show that this family is a polynomial self-localized frame for $L^2(\mathbb{R})$. The associated Banach space characterized by $\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b)$ is a α-modulation space, see Fornasier and Gröchenig (2005) and Fornasier and Feichtinger (2006, Theorem 2.1). The next results illustrate under which conditions this family defines a frame for $L^2(\mathbb{R}^d)$ and gives a frame expansion for elements of the α-modulation spaces.

Theorem 11. *Let $\alpha \in [0, 1)$ and $s \in \mathbb{R}$. For $h \in \mathcal{S}(\mathbb{R})$ with $\mathcal{F}h \neq 0$ and $b > 0$, there exists $0 < a_0 \leq 1$ small enough such that for all $0 < a \leq a_0$ the family $\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b)$ is a frame for $L^2(\mathbb{R})$.*

Proof. Fornasier and Feichtinger (2006, Theorem 1). □

For $f \in L^2(\mathbb{R})$ we obtain the following frame expansion

$$f = \sum_{\lambda \in \mathbb{Z}^2} \langle f, \tilde{h}_\lambda^\alpha \rangle h_\lambda^\alpha, \qquad h_\lambda^\alpha \in \mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b), \tag{2.47}$$

with dual frames $\tilde{h}_\lambda^\alpha$ and unconditional convergence in $L^2(\mathbb{R})$. Fornasier (2004) extended this results to characterize modulation spaces.

Theorem 12. *Let $\alpha \in [0, 1)$ and $s \in \mathbb{R}$. For $h \in \mathcal{S}$ with $\mathcal{F}h \neq 0$ and $b > 0$, there exists $0 < a_0 \leq 1$ small enough such that for all $0 < a \leq a_0$ the family*

$$\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b) = \left\{ M_{p_\alpha(j)} D_{s_\alpha^{-1}(j)} T_{ak}\, h \right\}_{(j,k) \in \mathbb{Z}^2} \tag{2.48}$$

is a Banach frame and atomic decomposition for the α-modulation space $\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})$ for all $p \in [1, \infty]$. The frame expansion (2.47) converges unconditionally in the modulation spaces $\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})$ for $p \in [1, \infty)$ with the following norm equivalence,

$$\|f\|_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})} \asymp \|\langle f, h_\lambda^\alpha \rangle\|_{\ell^p_{\bar{m}_{0,s:\alpha}}} \tag{2.49}$$

with

$$\bar{m}_{u,v:\alpha}(j,k) = (1 + |a \cdot k|)^u + (1 + |p_\alpha(j)|)^v.$$

Proof. Fornasier (2004, Theorem 3.4) □

The previous results hold not only for Fornasier and Feichtinger's example of position and size functions in eq. (2.41), but for any other pair (p_α, s_α) that satisfies the conditions given by the authors, see Fornasier and Feichtinger (2006, pp 10). This frame characterizes α-modulation spaces for $\alpha \in [0, 1)$.

When $\alpha = 0$, the dilation factor $s_\alpha^{-1}(\cdot)$ becomes a constant and the dilation operator loses its effect. This results in a relatively separated lattice $(b \cdot j, a \cdot k)$ with $(j, k) \in \mathbb{Z}^2$, forming a Gabor frame $h_{\lambda(\lambda \in \mathbb{Z}^2)}$. However, this frame characterization does not hold for the limiting case of $\alpha = 1$. Besov spaces can only be characterized by exponentially localized frames, and the frame obtained for the case $\alpha \to 1$ does not satisfy this requirement. Therefore, the frame characterization does not hold for the limiting case $\alpha = 1$.

2.5. Proofs

In this section we provide proofs for the results in this chapter.

2.5.1. Proof of Lemma 6

Proof. 1. From the submultiplicative property of v_s it follows:

$$\begin{aligned} m_s(z) &= (1+\|x\|_2^2)^{s/2} + (1+\|\omega\|_2^2)^{s/2} \\ &\leq 2\,(1+\|x\|_2^2+\|\omega\|_2^2)^{s/2} \\ &\leq 2^{\max(1,s/2)}\, v_s(z) \end{aligned}$$

$$\begin{aligned} v_s(z) &= (1+\|x\|_2^2+\|\omega\|_2^2)^{s/2} \\ &\leq (1+2\|x\|_2^2)^{s/2} + (1+2\|\omega\|_2^2)^{s/2} \\ &\leq 2^{s/2}\left((1+\|x\|_2^2)^{s/2} + (1+\|\omega\|_2^2)^{s/2}\right) \\ &\leq 2^{\max(1,s/2)}\, m_s(z) \end{aligned}$$

2. Using the equivalent definition $m_{u,v}(x,\omega) = (1+\|x\|_2)^u + (1+\|\omega\|_2)^v$, we obtain

$$\begin{aligned} m_{u,v}(z_1+z_2) &= (1+\|x_1+x_2\|_2)^u + (1+\|\omega_1+\omega_2\|_2)^v \\ &\leq (1+\|x_1\|_2+\|x_2\|_2)^u + (1+\|\omega_1\|_2+\|\omega_2\|_2)^v \\ &\leq (1+\|x_1\|_2)^u(1+\|x_2\|_2)^u + (1+\|\omega_1\|_2)^v(1+\|\omega_2\|_2)^v \\ &\leq \left((1+\|x_1\|_2)^u + (1+\|\omega_1\|_2)^v\right)\left((1+\|x_2\|_2)^u + (1+\|\omega_2\|_2)^v\right) \\ &= m_{u,v}(z_1)m_{u,v}(z_2) \end{aligned}$$

□

3. Estimation over Modulation Spaces

In this chapter we revisit and extend the results presented in Dahlke et al. (2022). We give nonparametric function estimation in the smoothness class $\mathcal{M}^p_m(\mathbb{R}^d)$ under global integrated squared error. Specifically, we consider a signal of interest observed in a Gaussian white noise model, which we transform into an equivalent infinite Gaussian sequence model using time-frequency analysis. The works of Brown and Low (1996) and Nussbaum (1996) provide valuable insights in this regard. Our approach involves optimal parameter estimation within a finite sequence submodel, which is at the heart of nonparametric statistics. We show that there is no significant loss by ignoring the remaining parameters. To reduce noise, we propose a spatially adaptive estimator based on soft/hard thresholding of Gabor frame coefficients. The redundancy of the frames plays a key role in minimizing the risk compared to orthogonal bases. Leveraging results from Donoho and Johnstone (1994), we find that nonlinear thresholding of Gabor coefficients achieves near-optimal convergence rates, in the minimax framework, over the modulation space $\mathcal{M}^p_s(\mathbb{R}^d)$, for $p \leq 2$. This outperforms linear methods. We derive sparsity-driven lower bounds using oracles, and obtain a lower bound in the compact setting through the geometric properties of the sequence space, as described in Donoho, Liu, and MacGibbon (1990). For the lower bound on $\mathbb{R}^d$, we use Tsybakov (2009) general theorem for deriving minimax lower bounds. We use Gaussian test sequences that minimize the overlap between two sequences. These results can be compared to results in Sobolev spaces for certain settings of the modulation spaces.

The chapter is structured as follows: First, we introduce our Gaussian white noise model, referred to as the analog model, and outline the minimax framework. Next, we demonstrate the transformation of the analog model into an equivalent infinite-dimensional Gaussian sequence model using frame expansion. In Section 3, we present the projection error obtained by considering a finite submodel, which we call the digital model. Section 4 provides lower bounds, while in the final section we establish the near-optimality of the

thresholding operators. Throughout this chapter, our primary focus is on Gabor frames, although the procedure remains applicable to other frames.

3.1. Analog Model

Consider the following Gaussian white noise model

$$\mathrm{d}\mathbf{Y}(x) = f(x)\ \mathrm{d}x + \varepsilon\,\mathrm{d}\mathbf{W}(x), \qquad x \in \mathbb{R}^d, \tag{3.1}$$

in which we observed a square-integrable analog signal $f \in L^2(\mathbb{R}^d)$ contaminated by a standard Wiener process $\mathbf{W}(x)$ with white noise intensity ε. Furthermore, we assume that the signal of interest f is an element of the smoothness space

$$\Theta^p_{u,v} = \left\{ f \in L^2(\mathbb{R}^d) | \quad \|f\|_{\mathcal{M}^p_{u,v}(\mathbb{R}^d)} \leq \bar{C} \right\}, \tag{3.2}$$

with $p \in (0, 2]$, $u, v \geq 0$ and the weight function

$$m_{u,v}(x, \omega) = (1 + |x|)^u + (1 + |\omega|)^v.$$

For $u = v$, we write Θ^p_u instead of $\Theta^p_{u,u}$.

The parameter space $\Theta^p_{u,v}$ is a bounded subset of $\mathcal{M}^p_{u,v}(\mathbb{R}^d)$. We denote the (quasi-)norm on $\Theta^p_{u,v}$ by

$$\|f\|_{\Theta^p_{u,v}} = \|f\|_{\mathcal{M}^p_{u,v}(\mathbb{R}^d)}.$$

The goal is to find an estimator $\hat{f} \in \Theta^p_{u,v}$ that minimizes the risk assessed by a given loss function. As discussed in Ma and Loizou (2011), various metrics are available in the audio domain to accurately measure audio or visual degradation, such as spectral loss, perceptual loss, or signal-to-noise ratio. For simplicity, we define the loss function as the mean squared distance, which is both straightforward and suitable for theoretical purposes. Given a parameter space Θ in (3.1), the risk of an estimator $\hat{f} := \hat{f}(y) \in \Theta$

based on the observations $y = \int \cdot d\mathbf{Y}(x)$, is given by the global mean integrated squared error (MISE)

$$r(\hat{f}, f) = \mathbb{E}_f[\|f - \hat{f}\|^2_{L^2(\mathbb{R}^d)}], \tag{3.3}$$

where $\mathbb{E}_f$ denotes the expected value if the underlying parameter is f. The performance of an estimator $\hat{f}$ over a parameter set Θ is measured by its maximum risk

$$r(\hat{f}, \Theta) = \sup_{f \in \Theta} r(\hat{f}, f). \tag{3.4}$$

An estimator is said to be optimal if its rate equals the lower bound or minimax risk, the least risk of all possible estimators. We define the minimax risk with

$$r_{\min}(\Theta) = \inf_{\hat{f}} r(\hat{f}, \Theta).$$

The minimax problem is separable. That is, the linear minimax risk $r^{\mathrm{L}}_{\min}$ is limited to linear procedures and the nonlinear minimax risk $r^{\mathrm{N}}_{\min}$ for the others.

3.2. Infinite-dimensional Gaussian Sequence Model

We recall the conventional nonparametric regression model

$$Y_i = f(X_i) + \xi_i, \qquad i = 1, ...n.$$

with $\mathbb{E}[\xi_i] = 0$ and $X_i \in \mathbb{R}^d$. Brown and Low (1996) and Nussbaum (1996) have shown that the analog model (3.1) arises as a large sample limit of the conventional nonparametric regression model where only finite countable observations are made. In this section, we establish a connection between the analog model (3.1) and the infinite-dimensional Gaussian sequence model based on frame expansion. We show that estimates obtained in the sequence model induce estimates in the analog model, with the same rate of convergence, and vice versa. This means that we can analyze the behavior of an estimator at the sequence level, and the results remain valid at the function level without loss of accuracy. This is a powerful result that allows us to simplify the analysis, by reducing to sequences that can be easily manipulated and analyzed. For more details, see Ibragimov

and Has'minskiĭ (1981).

In model (3.1), a continuous signal is observed. Therefore, we refer to it as the analog model. To obtain a sampled version of this model, we integrate a countable number of elements of a frame for the underlying function space with respect to the observed signal. This results in a Gaussian sequence model.

We define the index set $\Lambda = \{(k,n) : k,n \in \mathbb{Z}^d\}$. For an admissible real-valued non-zero window function $h \in L^2(\mathbb{R}^d)$, we consider $\alpha, \beta > 0$, so that the set of Gabor frame elements $\{h_\lambda\}_{\lambda\in\Lambda} = \{T_{\alpha k} M_{\beta n} h\}_{\lambda\in\Lambda} = \mathcal{G}(h,\alpha,\beta)$ defines frames for the Hilbert space $L^2(\mathbb{R}^d)$, with lower and upper frame bounds A and B respectively, and $0 < A \leq B$. The dual frame is denoted by $\{\tilde{h}_k\}_{\lambda\in\Lambda} := \mathcal{G}(\tilde{h},\alpha,\beta)$.

Each function $f \in L^2(\mathbb{R}^d)$ has a representation as a linear combination of the frame elements $\{h_\lambda\}_{\lambda\in\Lambda}$,

$$f = \sum_{\lambda\in\Lambda} \langle f, \tilde{h}_\lambda \rangle h_\lambda,$$

with unconditional convergence. Furthermore, from Theorem 9 there exist constants $\tilde{C}_1, \tilde{C}_2 > 0$ such that the following (quasi-)norm equivalence holds

$$\tilde{C}_1 \|f\|_{\Theta^p_{u,v}} \leq \|\langle f, \tilde{h}_\lambda \rangle_{\lambda\in\Lambda}\|_{\ell^p_{\bar{m}_{u,v}}} \leq \tilde{C}_2 \|f\|_{\Theta^p_{u,v}}, \tag{3.5}$$

with $\bar{m}_{u,v}(k,n) = m_{u,v}(\alpha k, \beta n)$. See eq. (2.35).

Due to the stability of the frame operator, the frame coefficients can be used to obtain a stable signal approximation of f using the synthesis operator. This means that for any estimator $\hat{F} \in \ell^2\{\Lambda\}$ of the frame coefficients $F = \{\langle f, \tilde{h}_\lambda \rangle\}_{\lambda\in\Lambda}$, there exists an equivalent estimator $\hat{f} = T^* \hat{F} \in L^2(\mathbb{R}^d)$ of the function f that achieves the same rate, and vice versa. In other words, we have $\hat{f} = T^* \hat{F}$ or equivalently $\hat{F} = T\hat{f}$. One can then optimize an estimate of the frame coefficients obtained from noisy observations in the following infinite dimensional Gaussian model

$$Y[\lambda] = F[\lambda] + W[\lambda], \qquad \lambda \in \Lambda \tag{3.6}$$

with

$$
\begin{aligned}
F[\lambda] &= \int f \cdot \tilde{h}_\lambda \, dx = \langle f, \tilde{h}_\lambda \rangle \\
W[\lambda] &= \int \varepsilon \cdot \tilde{h}_\lambda \, d\mathbf{W}(x) \sim \mathcal{N}\big(0, \varepsilon^2 \, \|\tilde{h}\|_{L^2}^2\big), \\
Y[\lambda] &= \int \tilde{h}_\lambda \, d\mathbf{Y} \sim \mathcal{N}\big(F[\lambda], \varepsilon^2 \, \|\tilde{h}\|_{L^2}^2\big) \\
\mathrm{Cov}(Y[\lambda_1], Y[\lambda_2]) &= \epsilon^2 \langle \tilde{h}_{\lambda_1}, \tilde{h}_{\lambda_2} \rangle
\end{aligned}
$$

using the isometry of the time-frequency shifts in (2.13), $\|\tilde{h}_\lambda\|_{L^2}^2 = \|\tilde{h}\|_{L^2}^2$. The corresponding parameter set

$$
\bar{\Theta}_{u,v}^p := \bar{\Theta}_{u,v}^p(\Lambda) = \Big\{ F \in \ell^2(\Lambda) \quad | \qquad \|T^*F\|_{\Theta_{u,v}^p} < \bar{C} \Big\}.
$$

For $u = v$, we write $\bar{\Theta}_u^p$ instead of $\bar{\Theta}_{u,u}^p$. The (quasi-)norm equivalence (3.5) can be reformulated as follows

$$
\tilde{C}_1 \|T^*F\|_{\Theta_{u,v}^p} \leq \|F\|_{\bar{\Theta}_{u,v}^p} \leq \tilde{C}_2 \|T^*F\|_{\Theta_{u,v}^p} \tag{3.7}
$$

with $\|F\|_{\bar{\Theta}_{u,v}^p} = \|F\|_{\ell^p_{\tilde{m}_{u,v}}}$.

Remark 3. Note that if the frame element h_λ is a complex-valued square-integrable function, as is often the case in time-frequency analysis, then $Y[\lambda]$ is a complex-valued normally distributed random variable with

$$
\begin{aligned}
\Re(Y[\lambda]) &\sim \mathcal{N}\big(\Re(F[\lambda]), \varepsilon^2 \, \|\Re(\tilde{h}_\lambda)\|_{L^2}^2\big) \\
\Im(Y[\lambda]) &\sim \mathcal{N}\big(\Im(F[\lambda]), \varepsilon^2 \, \|\Im(\tilde{h}_\lambda)\|_{L^2}^2\big).
\end{aligned}
$$

We adapt the minimax framework as follows. The risk of a sequence estimator $\hat{F}$ in the infinite-dimensional Gaussian sequence model (3.6) is given by

$$
r(\hat{F}, F) = \mathbb{E}_F[\|F - \hat{F}\|_{\ell^2(\Lambda)}^2], \tag{3.8}
$$

and its performance over a set of parameters $\bar{\Theta}$ is measured by

$$
r(\hat{F}, \bar{\Theta}) = \sup_{F \in \bar{\Theta}} r(\hat{F}, F). \tag{3.9}
$$

The minimax risk is defined as

$$r_{\min}(\bar{\Theta}) = \min_{\hat{F}} r(\hat{F}, \bar{\Theta}).$$

Because of the isometry of the of the L^2-risk and l^2-risk, estimating the signal f with respect to the L^2-risk is then equivalent to estimating its frame coefficients $F = Tf$ with respect to the l^2-risk.

Lemma 13. *Consider the analog model* (3.1) *with* $f \in \Theta$, $\{h_\lambda\}_{\lambda\in\Lambda}$ *a frame for* Θ *with lower and upper bounds* $0 < A \leq B$, *and the corresponding infinite-dimensional Gaussian model* (3.6) *with* $F \in \bar{\Theta}$.

If $\hat{F} := \hat{F}(Y)$ *defines an estimator for* F *in* (3.6) *based on the observations* $Y = \{Y[\lambda]\}_{\lambda\in\Lambda}$, *then* $\hat{f} = T^*\hat{F}$ *also defines an estimator for* f *in* (3.1) *with*

$$r(\hat{F}, \bar{\Theta}) \sim r(\hat{f}, \Theta). \tag{3.10}$$

Remark 4. Lemma 13 shows a fundamental advantage of frames over orthogonal bases. The redundancy of frames often leads to a lower risk compared to orthogonal bases. With the redundancy factor, $0 < A \leq B$, we obtain

$$\frac{1}{B} \cdot r(\hat{F}, \bar{\Theta}) \leq r(\hat{f}, \Theta) \leq \frac{1}{A} \cdot r(\hat{F}, \bar{\Theta}).$$

Higher redundancy then reduces the risk. For a tight frame, each estimator reduces the risk by averaging several estimators, each obtained in different orthogonal bases.

3.3. Projection Estimation and Digital Model

In this section, we reduce the previously presented infinite-dimensional model to a finite multivariate normal mean model, which is the core of parametric statistical inference.

We refer to this step as reducing the discrete resolution of the observed signal. We show that if the resolution is high enough, restricting to finite observations leads to a similar risk as in the entire sequence space. This means that we can accurately approximate a signal using a finite subspace of the sequence space without losing too much information. We also provide an upper bound on the linear approximation error on a time-frequency ball. The results presented in this section are inspired by the work of Mallat (2009, Chapter 11).

The observed sequence $\{F[\lambda]\}_{\lambda\in\Lambda}$ in the infinite-dimensional Gaussian sequence model (3.6) is a discrete representation of the signal f. In practice, it is often necessary to work with a finite representation of the signal, either because it is not possible to measure or store an infinite number of observations or because it is not necessary to use such high resolution. A finite representation is obtained by projecting the full representation into a subspace of the sequence space. This reduces the resolution of the signal.

For a nonempty finite index set $\Lambda_0 \subset \Lambda$, we consider the following finite sequence model, which we call the digital model

$$Y[\lambda] = F[\lambda] + W[\lambda], \qquad \lambda \in \Lambda_0 \tag{3.11}$$

with the parameter set

$$\bar{\Theta}^p_{u,v}(\Lambda_0) = \left\{ F \in \ell^2(\Lambda_0) \quad | \qquad \|T^*F\|_{\Theta^p_{u,v}} < \bar{C} \right\},$$

where $F[\lambda]$, $W[\lambda]$ and $Y[\lambda]$ are defined as before.

It is obvious that $\bar{\Theta}^p_{u,v}(\Lambda_0) \subseteq \bar{\Theta}^p_{u,v}$. Thus, any estimator for this model is also a valid estimator when considering the parameter set $\bar{\Theta}^p_{u,v}$. The set $\bar{\Theta}^p_{u,v}(\Lambda_0)$ contains the orthogonal projection of the elements of $\bar{\Theta}^p_{u,v}$ from $\ell^2(\Lambda)$ to $\ell^2(\Lambda_0)$. We denote the projection operator $P_{\Lambda_0}F$, with $P_{\Lambda_0}F = \{F[\lambda]\}_{\lambda\in\Lambda_0}$.

Next, we show that if the subspace Λ_0 is sufficiently large, the risk of an estimator in the finite model does not significantly deteriorate when applied to the full model. An optimal estimator for the reduced model would then remain optimal in the full model.

The risk of a finite dimensional projection estimator $\hat{F}_0 \in \bar{\Theta}^p_{u,v}(\Lambda_0)$ in the infinite-dimensional Gaussian model (3.6) is given by

$$\begin{aligned} r(\hat{F}_0, F) &= \mathbb{E}_F[\|\hat{F}_0 - F\|^2_{\ell^2(\Lambda)}] \\ &= \mathbb{E}_F[\|\hat{F}_0 - P_{\Lambda_0}F\|_2^2 + \|P_{\Lambda_0}F - F\|^2_{\ell^2(\Lambda)}] \\ &= \mathbb{E}_F[\|\hat{F}_0 - P_{\Lambda_0}F\|_2^2] + \|P_{\Lambda_0}F - F\|^2_{\ell^2(\Lambda)} \\ &= r(\hat{F}_0, P_{\Lambda_0}F) + \|P_{\Lambda_0}F - F\|^2_{\ell^2(\Lambda)} \end{aligned}$$

Taking the maximum over the parameter set $\bar{\Theta}^p_{u,v}$ gives the following inequality

$$r(\hat{F}_0, \bar{\Theta}^p_{u,v}(\Lambda_0)) \leq r(\hat{F}_0, \bar{\Theta}^p_{u,v}) \leq r(\hat{F}_0, \bar{\Theta}^p_{u,v}(\Lambda_0)) + \varepsilon_{\bar{\Theta}^p_{u,v}}(\Lambda, \Lambda_0),$$

where $\varepsilon_{\bar{\Theta}^p_{u,v}}(\Lambda, \Lambda_0) = \sup_{F \in \bar{\Theta}^p_{u,v}} \|F - P_{\Lambda_0}F\|^2_{\ell^2(\Lambda)}$ is the maximum linear approximation error when orthogonal projecting elements of $\bar{\Theta}^p_{u,v}$ from $\ell^2(\Lambda)$ onto $\ell^2(\Lambda_0)$.

If the linear approximation error is of the same order as the subspace estimation risk, i.e. $r(\hat{F}_0, \bar{\Theta}^p_{u,v}(\Lambda_0)) \sim \varepsilon_{\bar{\Theta}^p_{u,v}}(\Lambda, \Lambda_0)$, then we can conclude that

$$r(\hat{F}_0, \bar{\Theta}^p_{u,v}(\Lambda_0)) \sim r(\hat{F}_0, \bar{\Theta}^p_{u,v}). \tag{3.12}$$

Therefore, any optimal sequence estimator in $\bar{\Theta}^p_{u,v}(\Lambda_0)$ induces an optimal function estimator in $\bar{\Theta}^p_{u,v}$. In other words, appropriately reducing the resolution of the signal does not affect the optimality of an estimator.

3.3.1. Linear Approximation Error on Time-Frequency Subset

Now we examine the difference between a sequence in the time-frequency plane and its projection onto a particular subspace. For $K_1, K_2 > 0$ we consider the following index set

$$\Lambda_{0,(K_1,K_2)} = \{\lambda \in \Lambda : \|k\|_2 \leq K_1, \quad \|n\|_2 \leq K_2\}, \tag{3.13}$$

with $\Lambda_{0,K} = \Lambda_{0,(K,K)}$. The parameters K_1 and K_2 can be interpreted as signal discrete resolution in time and frequency. This corresponds to a uniform discretization of a subset of the time-frequency space,

$$\left\{ (x, \omega) \in \mathbb{R}^{2d} \quad | \quad \begin{array}{ll} x = \alpha k, & \|k\|_2 \leq K_1, \\ \omega = \beta n, & \|n\|_2 \leq K_2 \end{array} \right\}. \tag{3.14}$$

Lemma 14. *Let $u, v \geq 0$. For $F \in \bar{\Theta}^p_{u,v}$ and $p \in (0, 2]$ the following inequality holds*

$$\sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^2 \leq \min(\alpha K_1, \beta K_2)^{-2\max(u,v)} \tilde{C}_2^2 \|T^* F\|^2_{\bar{\Theta}^p_{u,v}}, \tag{3.15}$$

as well as

$$\sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^2 \leq \max(\alpha K_1, \beta K_2)^{-2\min(u,v)} \tilde{C}_2^2 \|T^* F\|^2_{\bar{\Theta}^p_{u,v}}. \tag{3.16}$$

It then follows

$$\varepsilon_{\bar{\Theta}^p_{u,v}}(\Lambda, \Lambda_{0,K}) = O(C\, K^{-2\max(u,v)}) \tag{3.17}$$

with $C = \tilde{C}_2^2 \cdot \bar{C}^2 \cdot \min(\alpha, \beta)^{-2\max(u,v)}$.

The projection error is a function of resolution. A larger subspace is able to capture more of the details of the original signal, resulting in a smaller projection error. Thus, the projection error decreases as the resolution increases. Next, we give an interpretation of these results for $K_1 = K_2 = K$.

We observe that the projection error does not depend on the sparsity of the signal itself but rather on the sparsity of its representation in the time-frequency space. If the coefficients of a signal show a rapid decay in the time-frequency plane (i.e., its energy is concentrated in a small region in the time-frequency plane), the signal has a sparse representation and can therefore be compressed more efficiently, resulting in a smaller linear approximation error. In the isotropic design, i.e., $u = v = s$, the error rate is reduced to $O(K^{-2s})$. The projection error decreases as the decay rate increases. This means that a signal with a fast decay in the time-frequency plane has a sparse representation, resulting in a smaller projection error. In particular, for $u = 0, v = s$, this is similar to the approximation rate by piecewise polynomial functions observed on Sobolev spaces $W_2^\alpha(\mathbb{R}^d)$ with $\alpha > 2s$, see Birman and Solomyak (1966, Theorem 3.1) and Pinkus (1985). In the anisotropic design, where the decay rate is different in the time and frequency dimensions, a faster decay in either domain results in a smaller projection error. The anisotropic design allows more flexibility in the time-frequency representation of the signal.

Remark 5. The discrete resolution K plays the same role as the bandwidth h for kernel estimators. By choosing an appropriate value for K, we can balance the trade-off between the accuracy of the estimation and the computational complexity of the estimation process.

3.4. Oracles and Thresholding Estimation in the Digital Model

In this section, we present an adaptive nonlinear projection estimator based on coordinate-wise thresholding of the frame coefficients observed in the digital model. We extend the results of Donoho and Johnstone (1994) to complex-valued random variables and provide an upper bound on the risk of thresholding with a universal threshold.

Stein (1956) shows the fundamental role of shrinkage estimators in parametric and non-parametric inference. Donoho and Johnstone (1994) use oracle information to develop a spatially adaptive method based on shrinkage of the empirical coefficients. Furthermore shrinkage estimators are asymptotically minimax for a wide range of parameter sets, see Pinsker (1980) and Donoho, Liu, and MacGibbon (1990) provides. We begin this section with a review of diagonal estimators.

3.4.1. Diagonal Estimators

Let $\Lambda_0 \subset \mathbb{Z}^{2d}$ be a nonempty finite index set. In the digital model (3.11) with $F \in \bar{\Theta}(\Lambda_0)$, we consider shrinkage estimators of the form

$$\hat{F}_a = \{a_\lambda \cdot Y[\lambda]\}_{\lambda \in \Lambda_0}\,, \tag{3.18}$$

with $|a_\lambda| \leq 1$ and $a = \{a_\lambda\}_{\lambda \in \Lambda_0}$.

These estimators are called diagonal estimators because they attenuate the noisy frame coefficients independently. We say that an estimator is linear if each coefficient a_λ is a constant independent of $Y[\lambda]$ otherwise it is nonlinear. For an observed sequence $F \in \bar{\Theta}(\Lambda_0)$, the risk of an estimator of the form (3.18) is given by

$$\begin{aligned} r(\hat{F}_a, F) &= \mathbb{E}_F[\|\hat{F}_a - F\|_2^2] \\ &= \sum_{\lambda\in\Lambda_0} \mathbb{E}_F[|a_\lambda \cdot Y[\lambda] - F[\lambda]|^2] \\ &= \sum_{\lambda\in\Lambda_0} \mathbb{E}_F[|a_\lambda \cdot (F[\lambda] + W[\lambda]) - F[\lambda]|^2] \\ &= \sum_{\lambda\in\Lambda_0} (1-a_\lambda)^2 |F[\lambda]|^2 + a_\lambda^2 \epsilon^2 \|\tilde{h}\|_{L^2}^2 \end{aligned} \tag{3.19}$$

The linear diagonal oracle estimator $\hat{F}_{\text{inf}} = \left\{a_\lambda^{\text{inf}} \cdot Y[\lambda]\right\}_{\lambda\in\Lambda_0}$ with

$$a_\lambda^{\text{inf}} = \frac{|F[\lambda]|^2}{|F[\lambda]|^2 + \epsilon^2\|\tilde{h}\|_{L^2}^2}.$$

minimizes the risk (3.19) coordinatewise. It achieves the lowest risk among all estimators of the form (3.18). Its risk is given by

$$r_{\text{inf}}(F) = r(\hat{F}_{\text{inf}}, F) = \sum_{\lambda\in\Lambda_0} \frac{|F[\lambda]|^2 \cdot \epsilon^2\|\tilde{h}\|_{L^2}^2}{|F[\lambda]|^2 + \epsilon^2\|\tilde{h}\|_{L^2}^2}.$$

We denote the risk of the linear diagonal oracle over the parameter space $\bar{\Theta}$ with

$$r_{\text{inf}}(\bar{\Theta}) = \sup_{F\in\bar{\Theta}} r_{\text{inf}}(F).$$

The analysis of this class of estimators is simplified by restricting $a_\lambda \in \{0, 1\}$. This restriction leads to projection estimators. The risk (3.19) is minimized by the nonlinear oracle projector $\hat{F}_{\text{pr}} = \{a_\lambda^{\text{pr}} \cdot Y[\lambda]\}_{\lambda\in\Lambda_0}$ with

$$a_\lambda^{\text{pr}} = \begin{cases} 1 & if \quad |F[\lambda]| \geq \epsilon \|\tilde{h}\|_{L^2} \\ 0 & else \end{cases}$$

Its risk is defined by

$$r_{\mathrm{pr}}(F) = r(\hat{F}_{\mathrm{pr}}, F) = \sum_{\lambda \in \Lambda} \min(|F[\lambda]|^2, \epsilon^2 \|\tilde{h}\|_{L^2}^2) \tag{3.20}$$

We denote the oracle projector risk with

$$r_{\mathrm{pr}}(\bar{\Theta}) = \sup_{F \in \bar{\Theta}} r_{\mathrm{pr}}(F).$$

Using the following inequality for $(x, y) \in \mathbb{R}^2$

$$\min(x, y) \geq \frac{xy}{x+y} \geq \frac{1}{2} \min(x, y),$$

it can be shown that the linear diagonal oracle and the oracle projector are of the same order

$$r_{\mathrm{pr}}(F) \geq r_{\mathrm{inf}}(F) \geq \frac{1}{2} r_{\mathrm{pr}}(F).$$

An oracle estimator may not be available in practice, but the next section shows that a thresholding estimator comes close to an oracle in terms of risk.

3.4.2. Thresholding Estimation

We now turn to thresholding estimators. These are nonlinear diagonal estimators that effectively remove small coefficients while preserving large ones. They have been widely used in signal processing, image processing, and machine learning. The key insight is that thresholding provides a way to trade off bias and variance, resulting in a well-calibrated estimator that performs well in practice.

Definition 14. Let $v \in \mathbb{R}$. The soft- and hard thresholding operators at level $\mu > 0$ are respectively defined as

$$t_\mu^{\mathrm{s}}(v) = \frac{v}{|v|}(|v| - \mu)_+ \qquad \text{and} \qquad t_\mu^{\mathrm{h}}(v) = v\,\mathbf{1}(|v| \geq \mu) \tag{3.21}$$

While soft thresholding shrinks the magnitude of all signal coefficients, hard thresholding eliminates small magnitude values while preserving larger values. The choice of the

appropriate threshold μ is a critical statistical decision that has a significant impact on the overall performance of the estimation procedure. Choosing a large threshold can cause the estimator to become significantly biased. On the other hand, choosing a small threshold can lead to an increase in variance.

Remark 6. This definition of the thresholding operator is also valid for complex numbers, $z \in \mathbb{C}$. However, we decide to apply this operator separately to the real and imaginary parts of any complex number, since this simplifies the calculations while maintaining the theoretical results. For $z = u + iv$, we then have

$$t_\mu^{\mathrm{j}}(z) = t_\mu^{\mathrm{j}}(u) + i \cdot t_\mu^{\mathrm{j}}(v) \tag{3.22}$$

for $j \in \{\mathrm{s}, \mathrm{h}\}$. In practical experiments we will use Definition 14 as recommended by several authors for reasons of better perceptual sound quality when denoising audio signals.

Let $\Lambda_0 \subseteq \Lambda$ be a finite subset and $\bar{\Theta} \subseteq \ell^2(\Lambda)$. In the digital model (3.11) with parameter set $\bar{\Theta}(\Lambda_0)$, we define the following estimator based on thresholding of the complex frame coefficients at level $\mu \in \mathbb{R}$

$$\begin{aligned} \hat{F}_\mu^{\mathrm{j}} &= \left\{ t_\mu^{\mathrm{j}}(Y[\lambda]) \right\}_{\lambda \in \Lambda_0} \\ &= \left\{ t_\mu^{\mathrm{j}}(\Re(Y[\lambda])) + i\, t_\mu^{\mathrm{j}}(\Im(Y[\lambda])) \right\}_{\lambda \in \Lambda_0}, \end{aligned}$$

with $\mathrm{j} \in \{\mathrm{h}, \mathrm{s}\}$.

The risk of this thresholding estimator is given by

$$\begin{aligned} r(\hat{F}_\mu^{\mathrm{j}}, F) &= \mathbb{E}[\sum_{\lambda \in \Lambda_0} \left| t_\mu^{\mathrm{j}}(Y[\lambda]) - F[\lambda] \right|^2] \\ &= \mathbb{E}[\sum_{\lambda \in \Lambda_0} \left| t_\mu^{\mathrm{j}}(\Re(Y[\lambda])) - \Re(F[\lambda])) \right|^2 + \left| t_\mu^{\mathrm{j}}(\Im(Y[\lambda])) - \Im(F[\lambda])) \right|^2] \\ &= r(t_\mu^{\mathrm{j}}(\Re(Y)), \Re(F)) + r(t_\mu^{\mathrm{j}}(\Im(Y)), \Im(F)) \end{aligned}$$

We then have the following oracle inequalities, which are extensions to complex-valued random variables of classical results from Donoho and Johnstone (1994).

Proposition 15. *Let $\Lambda_0 \subseteq \Lambda$ be a nonempty finite subset with $\#\Lambda_0 \geq 2$. In the digital model (3.11) with $F \in \bar{\Theta}(\Lambda_0)$, we have for soft thresholding with the universal threshold μ_{uni} that*

$$\frac{1}{2} \cdot r_{pr}(F) \leq r(\hat{F}^{\text{s}}_{\mu_{\text{uni}}}, F) \leq 2 \cdot \big(2 \log(\#\Lambda_0) + 1\big)\big(\varepsilon^2 \|h\|^2_{L^2} + r_{pr}(F)\big), \tag{3.23}$$

where $\mu_{\text{uni}} = \varepsilon \|\tilde{h}\|_{L^2} \sqrt{2 \log(\#\Lambda_0)}$.

Similarly, for hard thresholding we have the same estimate with a different leading constant.

Proposition 15 gives an upper bound for the thresholding estimator $\hat{F}^{\text{j}}_{\mu_{\text{uni}}}$. The risk remains within a $4 \log(\#\Lambda_0)$ range of the risk of the nonlinear oracle projector. The The left-hand side is obvious, since thresholding operators are diagonal estimators and therefore have a greater risk than the linear oracle estimator. The factor 2 on the right-hand side results from denoising the real and imaginary parts separately.

Remark 7. The logarithmic factor $4 \log(\#\Lambda_0)$ is optimal among diagonal estimators, Mallat (2009, Theorem 11.7). This factor can be eliminated by using a more sophisticated thresholding scheme such as SureShrink from Donoho and Johnstone (1995).

3.4.3. Upper Bound on the Thresholding Risk

Next, we derive upper bounds for the thresholding estimator $\hat{F}^{\text{j}}_{\mu}$ in the digital model (3.11). Specifically, we consider the parameter set $\bar{\Theta}^p_{u,v}(\Lambda_{0,K})$, where K is a resolution chosen to be sufficiently large. The next theorem gives an upper bound for the nonlinear oracle projector. We obtain an upper bound for the threshold estimator using Proposition 15.

Theorem 16. *Let $p \in (0, 2]$, $s > 0$, and $K > 0$. In the digital model (3.11) with $F \in \bar{\Theta}^p_s(\Lambda_{0,K})$ and K sufficiently large, the nonlinear oracle projector error has the upper bound*

$$r_{pr}(F) \leq \text{const.} \cdot \|F\|_{\bar{\Theta}^p_s}^{\frac{2\,d\,p}{2\,d+s\,p}} \varepsilon^{\frac{2\,d\,(2-p)+2sp}{2\,d+s\,p}}. \tag{3.24}$$

In general, for $F \in \bar{\Theta}^p_{u,v}(\Lambda_{0,K})$, $u, v > 0$, *and* $K > 0$, *the nonlinear oracle estimator achieves minimax risk in the digital model* (3.11) *with*

$$r_{pr}(F) \leq \text{const.} \cdot \|F\|_{\bar{\Theta}^p_{u,v}}^{\frac{p\,d(v+u)}{d(v+u)+pvu}} \varepsilon^{\frac{(2-p)\,d\,(v+u)+2pvu}{d(v+u)+pvu}}. \tag{3.25}$$

Remark 8. Note that the risk of the oracle projector is asymptotically bounded. For K sufficiently large, we obtain $r_{\text{pr}}(F) \sim \varepsilon_{\bar{\Theta}^p_{u,v}}(\Lambda, \Lambda_{0,K})$. Hence, this rate of convergence holds for $\bar{\Theta}^p_{u,v}$. Using Proposition 15, we obtain the rate for the thresholding estimator $\hat{F}^{\text{j}}_{\mu_{\text{uni}}}$.

In the isotropic design, the rate approaches the parametric rate ε^2 as the decay s increases. This means that a sparse time-frequency representation (better compression) reduces the error. The majority of the frame coefficients $F[\lambda]$ contain only noise. This observation holds in the sparse situation as p approaches 0. As d increases, the rate converges to ε^{2-p}. For $p = 2$, the rate converges to $\varepsilon^{\frac{4s}{2d+2s}}$.

Similar observations are made for the anisotropic design. The rate converges to ε^2 for $\min(u, v) \to \infty$. For $u = 0$ and $p = 2$, the Sobolev case (2.33), we obtain the rate $\varepsilon^{\frac{4v}{2d+2v}}$. This doesn't quite agree with the rate $\varepsilon^{\frac{4v}{d+2v}}$, which is observed over Sobolev ellipsoids for $f \in L^2([0,1])]$, see Pinsker (1980).

The term $2d$, which appears in the isotropic upper bound or in the Sobolev case, reflects the implication of the frequency and time domains, respectively of dimension d. The risk of the oracle projector in (3.20) is estimated over the time-frequency sphere $\Lambda_{0,K}$ whose size is of the order of $2d$, i.e. $\#\Lambda_{0,K} = O(2d)$. By restricting the support of the representation of the signal, the factor $2d$ can be halved. Using the definition $F[\lambda] = \langle f, T_{\alpha k} M_{\beta n} h\rangle =$, we define the following two assumptions:

Assumption 1.

- **Time:** There exists a $k_0 \in \mathbb{R}^d$ such that $F[\lambda] = 0$, for all $\lambda = (k, n)$ with $\|k\|_2 > \|k_0\|_2$. Then $\#\Lambda_{0,K} = O(d)$. This means that the function f observed in the analog model and the dual window function $\tilde{h}$ are time-limited.
- **Frequency:** There exists a $n_0 \in \mathbb{R}^d$ such that $F[\lambda] = 0$, for all $\lambda = (k, n)$ with $\|n\|_2 > \|n_0\|_2$. Then $\#\Lambda_{0,K} = O(d)$. This means that the function f observed in the analog model and the dual window function $\tilde{h}$ are band-limited.

Note that the frequency assumption is a consequence of the fundamental identity of the time-frequency analysis (2.15). The previous theorem can then be reformulated as follows.

Theorem 17. *Let $p \in (0, 2]$, $s > 0$, and $K > 0$. Under either the time or frequency assumption 1, the upper bound of the nonlinear oracle projector risk in the finite digital model* (3.11) *with $F \in \bar{\Theta}^p_s(\Lambda_{0,K})$ and K sufficiently large is given by*

$$r_{pr}(F) \leq \text{const.} \cdot \|F\|_{\bar{\Theta}^p_s}^{\frac{d\,p}{d+s\,p}} \varepsilon^{\frac{d\,(2-p)+2sp}{d+s\,p}}. \tag{3.26}$$

In general, for $F \in \bar{\Theta}^p_{u,v}(\Lambda_{0,K})$, under the time assumption we obtain

$$r_{pr}(F) \leq \text{const.} \cdot \|F\|_{\bar{\Theta}^p_{u,v}}^{\frac{d\,p}{d+v\,p}} \varepsilon^{\frac{d\,(2-p)+2vp}{d+p\,v}}. \tag{3.27}$$

and under the frequency assumption

$$r_{pr}(F) \leq \text{const.} \cdot \|F\|_{\bar{\Theta}^p_{u,v}}^{\frac{d\,p}{d+u\,p}} \varepsilon^{\frac{d\,(2-p)+2up}{d+p\,u}}. \tag{3.28}$$

For a function $f \in L^2[0, 1]$ and a compact dual window function $\tilde{h}$ the time assumption 1 is satisfied. Thus, for $u = 0$ and $p = 2$ we observe the Sobolev rate $\varepsilon^{\frac{4v}{d+2v}}$. Kerkyacharian et al. (2001) and Kerkyacharian et al. (2008) studied the minimax rate in a similar design focusing on functions with inhomogeneous smoothness properties and disparity of the inhomogeneous aspect in different directions.

3.5. Lower Bounds on the Minimax Risk

In this section, we derive lower bounds for the minimax risk in the infinite-dimensional Gaussian model (3.6). The lower bound is derived using the general theorem for deriving minimax lower bounds from Tsybakov (2009). Under each of the assumptions in (1), we use oracle information to obtain a lower bound. In the theory of nonparametric function estimation, linear estimators play a special role. Over solid and orthosymmetric parameter sets, the linear minimax risk is within a constant factor of the minimax risk. Moreover the linear minimax is achieved by diagonal linear estimators, see Pinsker (1980) and Donoho, Liu, and MacGibbon (1990).

Theorem 18. *For $p \in (0,2]$ and $s > 0$, we consider the infinite-dimensional model* (3.6) *with $F \in \bar{\Theta}^p_s$. Then we have that*

$$\liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{2\,d\,(2-p)+2sp}{2\,d+s\,p}} \inf_{\hat{F}_\varepsilon} r(\hat{F}_\varepsilon, \bar{\Theta}^p_s) \right) > 0,$$

where $\hat{F}_\varepsilon$ is an estimator in (3.6) *based on the observation Y.*

In general, for $F \in \bar{\Theta}^p_{u,v}$ and $u, v > 0$, we obtain

$$\liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}} \inf_{\hat{F}_\varepsilon} r(\hat{F}_\varepsilon, \bar{\Theta}^p_{u,v}) \right) > 0.$$

The proof of the lower bound uses standard tools from decision theory such as Fano's lemma and the Varshamov-Gilbert bound, see Tsybakov (2009). We provide a construction of test sequences with sufficient distance between them. The main effort here is to find an adequate family of distant test sequences to obtain the term $2\,d$ that appears in the upper bound. For two sequences $F, G \in \bar{\Theta}_{u,v}$ the distance between them is given by $\|F - G\|_2^2 = \|F\|_2^2 + \|G\|_2^2 - o(F,G)$, where $o(F,G)$ represents the overlap between the two sequences. We work with test sequences derived from a Gaussian window function $\varphi \in \Theta_{u,v}$. Gaussian functions concentrate the energy around a single point in the time-frequency space and minimize the overlap between sequences, see Theorem 5. We estimate the overlap in the time and frequency domains to bound the ℓ^2-distance.

Theorem 19. *Let $p \in (0,2]$ and $s > 0$. We consider the infinite-dimensional model (3.6) with $F \in \bar{\Theta}^p_s$. Under either the time or the frequency assumption 1 we have that*

$$\liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{d\,(2-p)+2sp}{d+s\,p}} \inf_{\hat{F}_\varepsilon} r(\hat{F}_\varepsilon, \bar{\Theta}^p_s) \right) > 0, \tag{3.29}$$

where $\hat{F}_\varepsilon$ is an estimator in (3.6) based on the observation Y.

In general, for $F \in \bar{\Theta}^p_{u,v}(\Lambda_{0,K})$ and $u, v > 0$, we obtain under the time assumption

$$\liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{(2-p)d+2\,vp}{d+vp}} \inf_{\hat{F}_\varepsilon} r(\hat{F}_\varepsilon, \bar{\Theta}^p_{u,v}) \right) > 0, \tag{3.30}$$

and under the frequency assumption

$$\liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{(2-p)d+2\,up}{d+up}} \inf_{\hat{F}_\varepsilon} r(\hat{F}_\varepsilon, \bar{\Theta}^p_{u,v}) \right) > 0. \tag{3.31}$$

Unlike the proof of Theorem 18 we cannot use Gaussian test sequences, since they do not satisfy the requirements of the theorem. The proof of Theorem 19 relies on geometric properties of the parameter set. We use results from Donoho, Liu, and MacGibbon (1990) to prove that the oracle linear estimator achieves the desired lower bound on a solid and orthosymmetric subset of the parameter set.

3.6. Minimax Rate-Optimality and Adaptivity on Modulation Spaces

Finally, we show near-optimality of the thresholding based estimator in the analog model (3.1). For any finite index set $\Lambda_{0,K}$, the threshold estimator $\hat{F}^{\mathrm{s}}_{\mu_{\mathrm{uni}}}$ has a risk close to the risk of the nonlinear oracle projector in the digital model, and is thus nearly optimal on the set $\bar{\Theta}^p_{u,v}(\Lambda_{0,K})$. For a sufficiently large K, the projection error on $\bar{\Theta}^p_{u,v}(\Lambda_{0,K})$ and the estimation error are of the same order. The rate then translates to $\bar{\Theta}^p_{u,v}$ and to the

analog model (3.1) by Lemma 13.

For $\mu > 0$ and $K > 0$ we define the following estimator for the analog model (3.1)

$$\hat{f}_\mu^j = \sum_{\lambda \in \Lambda_{0,K}} t_\mu^{\mathrm{j}}(Y[\lambda]) \cdot \tilde{h}_\lambda = T^* \left\{ t_\mu^{\mathrm{j}}(Y[\lambda]) \right\}_{\lambda \in \Lambda_{0,K}}, \tag{3.32}$$

with $j \in \{\mathrm{s}, \mathrm{h}\}$.

Corollary 20. *For $p \in (0, 2]$ and $s > 0$, consider the analog model (3.1) with $f \in \Theta_s^p$. By choosing the universal threshold $\mu_{\mathrm{uni}} = \varepsilon \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_{0,K})}$ and taking $K \gtrsim \varepsilon^{-\frac{d(2-p)+sp}{2ds+s^2p}}$, the oracle projector $\hat{f}_{pr} = T^* \hat{F}_{pr}$ is rate-optimal, and we obtain for the threshold estimator $\hat{f}_{\mu_{\mathrm{uni}}}^j$ the bound*

$$r(\hat{f}_{\mu_{\mathrm{uni}}}^j, f) \leq \text{const.} \cdot \max\left(\|f\|_{\Theta_s^p}^2, \|f\|_{\Theta_s^p}^{\frac{2dp}{2d+sp}} \right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{2d(2-p)+2sp}{2d+sp}}. \tag{3.33}$$

In general, for $f \in \Theta_{u,v}^p$ and $u, v > 0$ we obtain by thresholding with the universal threshold with $K \gtrsim \varepsilon^{-\frac{(2-p)d(v+u)+2pvu}{\max(u,v)(d(v+u)+pvu)}}$ the bound

$$r(\hat{f}_{\mu_{\mathrm{uni}}}^{\mathrm{j}}, f) \leq \text{const.} \cdot \max\left(\|f\|_{\Theta_{u,v}^p}^2, \|f\|_{\Theta_{u,v}^p}^{\frac{dp(u+v)}{d(v+u)+pvu}} \right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{(2-p)d(v+u)+2pvu}{d(v+u)+pvu}}. \tag{3.34}$$

The constant depends on the properties of the frame and the thresholding method.

Next, in the analog model, we assume that $\mathrm{supp}(f) \subseteq [0,1]^d$ and redefine the parameter space as follows

$$\Theta_{u,v}^p = \left\{ f \in L^2(\mathbb{R}^d) \quad | \quad \mathrm{supp}(f) \subseteq [0,1]^d, \quad \|f\|_{\mathcal{M}_{u,v}^p(\mathbb{R}^d)} \leq \bar{C} \right\}. \tag{3.35}$$

Furthermore, we consider a window function $h \in L^2(\mathbb{R})$ with a time-limited dual window $\tilde{h}$. Under this condition, the time assumption 1 is satisfied. The reason is that if D denotes the support of $\tilde{h}_\lambda$, then $\tilde{h}_\lambda$ has support $D + x$. This will be disjoint to $[0,1]^d$ for x sufficiently large. We can reformulate Corollary 20 as follows.

Corollary 21. *For $p \in (0,2]$ and $s > 0$, consider the digital model (3.1) with $f \in \Theta_s^p$. By choosing the universal threshold $\mu_{\text{uni}} = \varepsilon \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_{0,K})}$ as well as taking $K \gtrsim \varepsilon^{-\frac{d\,(1-\frac{p}{2})+sp}{d\,s+s^2\,p}}$, the oracle projector $\hat{f}_{pr} = T^* \hat{F}_{pr}$ is rate-optimal, and we obtain for the threshold estimator $\hat{f}^j_{\mu_{\text{uni}}}$ the bound*

$$r(\hat{f}^j_{\mu_{\text{uni}}}, f) \leq \text{const.} \cdot \max\left(\|f\|^2_{\Theta^p_s}, \|f\|^{\frac{d\,p}{d+s\,p}}_{\Theta^p_s} \right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{d\,(2-p)+2sp}{d+s\,p}}. \tag{3.36}$$

In general, for $f \in \Theta^p_{u,v}$ and $u, v > 0$, we obtain by thresholding with the universal threshold with $K \gtrsim \varepsilon^{-\frac{d\,(1-\frac{p}{2})+vp}{\max(u,v)(d+v\,p)}}$ the bound

$$r(\hat{f}^{\text{j}}_{\mu_{\text{uni}}}, f) \leq \text{const.} \cdot \max\left(\|f\|^2_{\Theta^p_{u,v}}, \|f\|^{\frac{d\,p}{d+v\,p}}_{\Theta^p_{u,v}} \right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{d\,(2-p)+2v\,p}{d+v\,p}}. \tag{3.37}$$

The constant depends on the properties of the frame and the thresholding method.

Remark 9. For a band-limited signal f and a dual window function $\tilde{h}$, results similar to those presented in corollary 21 can be derived. It follows that

$$r(\hat{f}^{\text{j}}_{\mu_{\text{uni}}}, f) \leq \text{const.} \cdot \max\left(\|f\|^2_{\Theta^p_{u,v}}, \|f\|^{\frac{d\,p}{d+u\,p}}_{\Theta^p_{u,v}} \right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{d\,(2-p)+2u\,p}{d+u\,p}}. \tag{3.38}$$

with the universal threshold with $K \gtrsim \varepsilon^{-\frac{d\,(1-\frac{p}{2})+up}{\max(u,v)(d+u\,p)}}$ and similar lines of proof.

The universal threshold estimator is adaptive up to a logarithmic factor for the L^2-loss and the parameter space $\Theta^p_{u,v}$. Combined with the Theorems 19 and 18, we derive, up to the logarithmic factor, optimal rate for the thresholding estimator in both the isotropic and the anisotropic cases.

3.7. Conclusion

We conclude this section with a few remarks. The thresholding estimators can be adapted for colored noise W. In a situation where the noise is not white, the noise has a different

intensity for each frame element. Thus, there is no longer a single threshold for the entire signal, but the threshold is chosen individually for each frame element or each region of the time-frequency plane. The threshold becomes a function of the lattice index λ. Donoho and Johnstone (1994) reformulate his result used in proposition 15. This still leads to near-optimality of our thresholds. The redundancy of the frames helps to reduce the risk of the estimator. This redundancy is the biggest advantage of frames over orthogonal bases. However, one must keep in mind that artificially increasing the redundancy also affects compression, since more elements must be stored to efficiently retrieve a signal. This affects the projection error.

3.8. Proofs

This section provides the proofs for the results presented in this chapter.

3.8.1. Proof of Lemma 13

Proof. The synthesis operator maps into $L^2(\mathbb{R}^d)$, thus $\hat{f} \in L^2(\mathbb{R}^d)$.
For each $F \in \bar{\Theta}$ with $f = T^*F$, we have

$$\begin{aligned} r(\hat{F}, F) &= \mathbb{E}[\|F - \hat{F}\|^2_{\ell^2\{\Lambda\}}] \\ &\leq B \cdot \mathbb{E}[\|T^*(F - \hat{F})\|^2_{L^2(\mathbb{R}^d)}] \\ &= B \cdot r(\hat{f}, f) \end{aligned}$$

as well as

$$r(\hat{F}, F) \geq A \cdot r(\hat{f}, f)$$

Taking the supremum over $\bar{\Theta}$ leads to the claim. □

3.8.2. Proof of Lemma 14

Proof. To simplify the proof, we will assume without loss of generality that $\alpha = \beta = 1$. Let $\lambda \in \Lambda^c_{0,(K_1,K_2)}$. From Lemma 6 we obtain

$$\begin{aligned}\bar{m}_{u,v}(\lambda) &= (1 + \|\alpha k\|_2^2)^{u/2} + (1 + \|\beta n\|_2^2)^{v/2} \\ &\geq (1 + K_1^2)^{u/2} + (1 + K_2^2)^{v/2} \\ &\geq K_1^u + K_2^v\end{aligned}$$

Since the p-norm is monotonously decreasing, we have

$$\begin{aligned}\sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^2 &\leq \Big(\sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^p \Big)^{2/p} \\ &\leq \Big((K_1^u + K_2^v)^{-p} \sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} \bar{m}_{u,v}(\lambda)^p \, |F[\lambda]|^p \Big)^{2/p} \\ &\leq (K_1^u + K_2^v)^{-2} \Big(\sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} \bar{m}_{u,v}(\lambda)^p \, |F[\lambda]|^p \Big)^{2/p} \\ &\leq \min(K_1, K_2)^{-2\max(u,v)} \|F\|^2_{\bar{\Theta}^p_{u,v}},\end{aligned}$$

using $K_1^u + K_2^v \geq \min(K_1, K_2)^{\max(u,v)}$ in the last step. The second inequality is obtained by using $K_1^u + K_2^v \geq \max(K_1, K_2)^{\min(u,v)}$ in the last step.

In addition, we have

$$\begin{aligned}\varepsilon_{\Theta^p_{u,v}}(\Lambda, \Lambda_{0,K}) &= \sup_{F \in \bar{\Theta}^p_{u,v}} \|F - P_{\Lambda_0} F\|^2_{\ell^2(\Lambda)} \\ &\leq \sup_{F \in \bar{\Theta}^p_{u,v}} \sum_{\lambda \in \Lambda^c_{0,K}} |F[\lambda]|^2 \\ &\leq \sup_{F \in \bar{\Theta}^p_{u,v}} K^{-2\max(u,v)} \|F\|^2_{\bar{\Theta}^p_{u,v}} \\ &\leq \sup_{F \in \bar{\Theta}^p_{u,v}} \tilde{C}_2^2 \cdot K^{-2\max(u,v)} \|T^* F\|^2_{\bar{\Theta}^p_{u,v}} \\ &= O(C\, K^{-2\max(u,v)})\end{aligned}$$

with $C = \tilde{C}_2^2 \cdot \bar{C}^2$ using (3.7). □

3.8.3. Proof of Proposition 15

Proof. We only give a proof of the right side. The left side is obvious, since thresholding operators are diagonal estimators and therefore have a higher risk than the linear oracle estimator. The proof is based on previous results by Donoho and Johnstone (1994), Candes (2006, Proof of Theorem 5.1) and Johnstone and Paul (2013). These results are applied separately to the real and imaginary parts of the complex frame coefficients. The risk can be extended as follows:

$$\begin{aligned} r(\hat{F}^{\mathrm{j}}_{\mu_{\mathrm{uni}}}, F) &= r(t^{\mathrm{j}}_{\mu_{\mathrm{uni}}}(\Re(Y)), \Re(F)) + r(t^{\mathrm{j}}_{\mu_{\mathrm{uni}}}(\Im(Y)), \Im(F)) \\ &= \sum_{\lambda \in \Lambda_0} \Big(\mathbb{E}[\big|t^{\mathrm{j}}_{\mu_{\mathrm{uni}}}(\Re(Y[\lambda])) - \Re(F[\lambda]))\big|^2] + \mathbb{E}[\big|t^{\mathrm{j}}_{\mu_{\mathrm{uni}}}(\Im(Y[\lambda])) - \Im(F[\lambda]))\big|^2] \Big) \end{aligned}$$

To bound the expected values in the last equation for soft thresholding, we use the following result from Donoho and Johnstone (1994) or Candes (2006, Proof of Theorem 5.1). Let $W \sim \mathcal{N}(a, \sigma_0^2)$. It follows that

$$\mathbb{E}\big[(t^{\mathrm{s}}_{\mu}(W) - a)^2\big] \leq \min(a^2, \sigma_0^2 + \mu^2) + 2\,\frac{\sigma_0^3}{\mu}\,\varphi\Big(\frac{\mu}{\sigma_0}\Big), \tag{3.39}$$

where φ is the density of the standard normal distribution.

From remark 3 we have that $\Re(Y[\lambda]) \sim \mathcal{N}\big(\Re(F[\lambda]), \varepsilon^2\,\|\Re(\tilde{h})\|^2_{L^2}\big)$. With this we obtain for the real part

$$\begin{aligned} &\mathbb{E}[\big|t^{\mathrm{s}}_{\mu}(\Re(Y[\lambda])) - \Re(F[\lambda]))\big|^2] \\ &\leq \min\big((\Re(F[\lambda]))^2, \varepsilon^2 \|\Re(\tilde{h}_\lambda)\|^2_{\mathcal{L}^2} + \mu^2\big) + \sqrt{2}\,\frac{\varepsilon^3\,\|\Re(\tilde{h}_\lambda)\|^3_{\mathcal{L}^2}}{\mu}\,\exp\Big(-\frac{\mu^2}{2\,\varepsilon^2\,\|\Re(\tilde{h}_\lambda)\|^2_{L^2}}\Big) \\ &\leq \min\big((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2} + \mu^2\big) + \sqrt{2}\,\frac{\varepsilon^3\,\|\tilde{h}\|^3_{L^2}}{\mu}\,\exp\Big(-\frac{\mu^2}{2\,\varepsilon^2\,\|\tilde{h}\|^2_{\mathcal{L}^2}}\Big) \end{aligned}$$

By inserting the universal threshold $\mu_{\mathrm{uni}} = \varepsilon\,\|\tilde{h}\|_{\mathcal{L}^2}\,\sqrt{2\,\log(\#\Lambda_0)}$ we obtain

$$\begin{aligned} &\mathbb{E}[\big|t^{\mathrm{s}}_{\mu_{\mathrm{uni}}}(\Re(Y[\lambda]) - \Re(F[\lambda]))\big|^2] \\ \leq &(1 + 2\,\log(\#\Lambda_0)) \cdot \min\big((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{L^2}\big) + \frac{\varepsilon^2\,\|\tilde{h}\|^2_{L^2}}{\#\Lambda_0 \cdot \sqrt{\log(\#\Lambda_0)}} \end{aligned}$$

And this leads us to the following result,

$$\begin{aligned}
&\sum_{\lambda\in\Lambda_0} \mathbb{E}\big[\big|t^{\mathrm{s}}_{\mu_{\mathrm{uni}}}(\Re(Y[\lambda])) - \Re(F[\lambda]))\big|^2\big] \\
\leq &\sum_{\lambda\in\Lambda_0} \left((1+2\,\log(\#\Lambda_0)) \cdot \min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) + \frac{\varepsilon^2\,\|\tilde{h}\|^2_{\mathcal{L}^2}}{\#\Lambda_0 \cdot \sqrt{\log(\#\Lambda_0)}} \right) \\
\leq &(1+2\,\log(\#\Lambda_0)) \cdot \sum_{\lambda\in\Lambda_0} \left(\min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) + \frac{\varepsilon^2\,\|\tilde{h}\|^2_{\mathcal{L}^2}}{\#\Lambda_0 \cdot \sqrt{\log(\#\Lambda_0)}} \right) \\
= &(1+2\,\log(\#\Lambda_0)) \cdot \left(\frac{\varepsilon^2\,\|\tilde{h}\|^2_{\mathcal{L}^2}}{\sqrt{\log(\#\Lambda_0)}} + \sum_{\lambda\in\Lambda_0} \min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) \right) \\
\leq &(1+2\,\log(\#\Lambda_0)) \cdot \left(\varepsilon^2\,\|\tilde{h}\|^2_{\mathcal{L}^2} + r_{\mathrm{pr}}(F) \right).
\end{aligned}$$

The same bound applies to the imaginary part

$$\sum_{\lambda\in\Lambda_0} \left(\mathbb{E}\big[\big|t^{\mathrm{s}}_{\mu_{\mathrm{uni}}}(\Im(Y[\lambda])) - \Im(F[\lambda]))\big|^2\big] \right) \leq (1+2\,\log(\#\Lambda_0)) \cdot \left(\varepsilon^2\,\|\tilde{h}\|^2_{\mathcal{L}^2} + r_{\mathrm{pr}}(F) \right)$$

Combining both bounds gives the desired result for soft thresholding.

The bound for hard thresholding is obtained in a similar way using the following result from (Johnstone, 2013, Proposition 8.1). For $W \sim \mathcal{N}(a, \sigma_0^2)$ and $\mu/\sigma_0 > 4$, with a suitable constant $C > 0$ we have that

$$\mathbb{E}\big[(t^{\mathrm{h}}_{\mu}(W) - a)^2\big] \leq C\,\min(a^2, \mu^2) + C\,\sigma_0\,\mu\,\varphi(\mu/\sigma_0 - 1).$$

Using this we obtain

$$\begin{aligned}
&\mathbb{E}\big[\big|t^{\mathrm{h}}_{\mu}(\Re(Y[\lambda])) - \Re(F[\lambda]))\big|^2\big] \\
\leq &C \min\left(\Re(F[\lambda])^2, \mu^2\right) + C\mu\,\varepsilon\,\|\Re(\tilde{h}_\lambda)\|_{L^2}\,\varphi\Big(\frac{\mu}{(\varepsilon\,\|\Re(\tilde{h}_\lambda)\|_{L^2})} - 1\Big)
\end{aligned}$$

By inserting the universal threshold $\mu_{\text{uni}} = \varepsilon \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_0)}$ we obtain

$$
\begin{aligned}
&\mathbb{E}[\left|t^{\text{h}}_{\mu_{\text{uni}}}(\Re(Y[\lambda])) - \Re(F[\lambda]))\right|^2] \\
&\leq C(1 + 2\,\log(\#\Lambda_0)) \cdot \min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) + C(\varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2} \sqrt{\log(\#\Lambda_0)})\, \exp\Big(- (\sqrt{\log(\#\Lambda_0)} - 1)^2\Big) \\
&\leq C(1 + 2\,\log(\#\Lambda_0)) \cdot \min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) + C(\varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2} \sqrt{\log(\#\Lambda_0)})\, \exp\Big(- \log(\#\Lambda_0)\Big) \\
&\leq C\left((1 + 2\,\log(\#\Lambda_0)) \cdot \min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) + \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\, \frac{\sqrt{\log(\#\Lambda_0)}}{\#\Lambda_0}\right) \\
&\leq C \cdot (1 + 2\,\log(\#\Lambda_0)) \cdot \left(\min\left((F[\lambda])^2, \varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}\right) + \frac{\varepsilon^2 \|\tilde{h}\|^2_{\mathcal{L}^2}}{\#\Lambda_0}\right)
\end{aligned}
$$

This is also true for the imaginary part. The bound for hard thresholding follows by summing over the index sets. □

3.8.4. Proof of Theorem 16

Proof. Let assume for convenience that $\|\tilde{h}\|^2_{L^2} = 1$. Since $p \in (0, 2]$, it follows

$$\min(\epsilon^2, |F[\lambda]|^2) \leq \epsilon^{2-p}\, |F[\lambda]|^p$$

Also note that for $K > 0$

$$\#\Lambda_{0,K} = CK^{2d},$$

with $C = C(\alpha, \beta, d)$.

Let $0 < K_s \leq K$. It follows

$$
\begin{aligned}
r_{\text{pr}}(F) &= \sum_{\lambda \in \Lambda_{0,K}} \min(\varepsilon^2, |F[\lambda]|^2) \\
&= \sum_{\lambda \in \Lambda_{0,K_s}} \min(\varepsilon^2, |F[\lambda]|^2) + \sum_{\lambda \in \Lambda_{0,K} \setminus \Lambda_{0,K_s}} \min(\varepsilon^2, |F[\lambda]|^2) \\
&\leq \varepsilon^2\, CK_s^{2d} + \varepsilon^{2-p} \sum_{\lambda \in \Lambda^c_{0,K_s}} |F[\lambda]|^p \\
&\leq \varepsilon^2\, CK_s^{2d} + \varepsilon^{2-p} K_s^{-sp} \sum_{\lambda \in \Lambda^c_{0,K_s}} \bar{m}_s(\lambda)^p\, |F[\lambda]|^p \\
&\leq \varepsilon^2\, CK_s^{2d} + \varepsilon^{2-p} K^{-sp}\, \|F\|^p_{\bar{\Theta}^p_s},
\end{aligned}
$$

using $\bar{m}_s(\lambda) \geq K^s$ for $\lambda \in \Lambda_{0,K}^c$ in the second last step, lemma 14. Balancing both terms using

$$K = \texttt{const.} \cdot \|F\|_{\bar{\Theta}_s^p}^{\frac{p}{2d+sp}} \varepsilon^{-\frac{p}{2d+sp}}$$

leads to the results in the isotropic case.

The proof for the anisotropic case is analogous. Let $K_u, K_v > 0$ with $K_u \leq K$ and $K_v \leq K$.

$$\begin{aligned} r_{\mathrm{pr}}(F) &= \sum_{\lambda \in \Lambda_{0,K}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &= \sum_{\lambda \in \Lambda_{0,(K_u,K_v)}} \min(\varepsilon^2, |F[\lambda]|^2) + \sum_{\lambda \in \Lambda_{0,(K,K)} \setminus \Lambda_{0,(K_u,K_v)}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &\leq \varepsilon^2 \, C K_u^d K_v^d + \epsilon^{2-p} \sum_{\lambda \in \Lambda_{0,(K_u,K_v)}^c} |F[\lambda]|^p \\ &\leq \varepsilon^2 \, C K_u^d K_v^d + \epsilon^{2-p} (K_u^{-up} + K_v^{-vp}) \, \|F\|_{\bar{\Theta}_{u,v}^p}^p \end{aligned}$$

using $\bar{m}_s(\lambda) \geq K_u^u + K_v^v$ for $\lambda \in \Lambda_{0,(K_u,K_v)}^c$.
We obtain the result by balancing both terms with

$$\begin{aligned} K_u &= \texttt{const.} \cdot \|F\|_{\bar{\Theta}_{u,v}^p}^{\frac{pv}{d(u+v)+puv}} \varepsilon^{-\frac{pv}{d(u+v)+puv}} \\ K_v &= \texttt{const.} \cdot \|F\|_{\bar{\Theta}_{u,v}^p}^{\frac{pu}{d(u+v)+puv}} \varepsilon^{-\frac{pu}{d(u+v)+puv}}. \end{aligned}$$

□

3.8.5. Proof of Theorem 17

Proof. The proof of this lemma is the same as that of Theorem 16. Some explanations are therefore omitted. Under the time assumption, there exists a $0 < K_0 \leq K$ such that $F[\lambda] = 0$ for $\lambda \in \Lambda_{0,(K_0,K_s)}^c$ and for all $K_s > 0$.

Let assume for convenience that $\|\tilde{h}\|_{L^2}^2 = 1$. Note that

$$\#\Lambda_{0,(K_0,K)} = C K_0^d K^d$$

with $C = C(\alpha, \beta, d)$. For $K_s \leq K$, we obtain

$$\begin{aligned} r_{\mathrm{pr}}(F) &= \sum_{\lambda \in \Lambda_{0,K}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &= \sum_{\lambda \in \Lambda_{0,(K_0,K_s)}} \min(\varepsilon^2, |F[\lambda]|^2) + \sum_{\lambda \in \Lambda^c_{0,(K_0,K_s)}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &\leq \varepsilon^2 \, C K_0^d K^d + \sum_{\lambda \in \Lambda^c_{0,K_s}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &\leq \varepsilon^2 \, C K_0^d K^d + \varepsilon^{2-p} K^{-sp} \, \|F\|^p_{\bar{\Theta}^p_s} \end{aligned}$$

Balancing both terms using

$$K = \texttt{const.} \cdot \|F\|^{\frac{p}{d+sp}}_{\bar{\Theta}^p_s} \varepsilon^{-\frac{p}{d+sp}}$$

leads to the results in the isotropic case.

The proof under the time assumption is analogous. Let $K_u, K_v > 0$ with $K_u \leq K$ and $K_v \leq K$.

$$\begin{aligned} r_{\mathrm{pr}}(F) &= \sum_{\lambda \in \Lambda_{0,K}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &= \sum_{\lambda \in \Lambda_{0,(K_0,K_v)}} \min(\varepsilon^2, |F[\lambda]|^2) + \sum_{\lambda \in \Lambda^c_{0,(K_0,K_v)}} \min(\varepsilon^2, |F[\lambda]|^2) \\ &\leq \varepsilon^2 \, C K_0^d K_v^d + \epsilon^{2-p} \sum_{\lambda \in \Lambda^c_{0,(K_v,K_v)}} |F[\lambda]|^p \\ &\leq \varepsilon^2 \, C K_0^d K_v^d + \epsilon^{2-p} (K_0^{-up} + K_v^{-vp}) \, \|F\|^p_{\bar{\Theta}^p_{u,v}} \end{aligned}$$

and balancing both terms with

$$K_v = \texttt{const.} \cdot \|F\|^{\frac{p}{d+vp}}_{\bar{\Theta}^p_{u,v}} \varepsilon^{-\frac{p}{d+vp}}$$

gives us the result. The proof under the frequency assumption is analogous and therefore omitted. □

3.8.6. Proof of Theorem 18

Proof. We give a proof for the general anisotropic case. For $u = v = s$ one can derive a proof for the isotropic case. The proof relies on the lower bound derived from Fano's lemma, see Tsybakov (2009, Theorem 2.5).

The task is to construct $M \in \mathbb{N}$ test sequences $F_j \in \bar{\Theta}^p_{u,v}$, such that

1. $F_j \in \bar{\Theta}^p_{u,v}, \qquad j = 1, \dots, M,$
2. $\|F_j - F_k\|^2_{\ell^2} \geq 2 \cdot \varepsilon^{\frac{(2-p)d(u+v)+2puv}{d(u+v)+puv}}, \qquad j \neq k.$
3. $\mathrm{KL}(Y_j, Y_k) \lesssim \log(M),$

where $\mathrm{KL}(Y_j, Y_k)$ represents the Kullback-Leibler divergence between two observations Y_j, Y_k in the digital model with $F = F_j$ and $F = F_k$ respectively. Applying Tsybakov (2009, Theorem 2.5) yields the desired lower bound.

i) Construction of the test sequences.
The test sequences are constructed as follows. We work with Gaussian window functions to maximize the ℓ^2-distance between sequences, see Theorem 5. We recall from Lemma 4 that the set $\left\{T_u M_\eta \varphi : u, \eta \in \mathbb{R}^d\right\}$ spans a dense subspace of $L^2(\mathbb{R}^d)$. Hence, for $f \in L^2(\mathbb{R}^d)$ we have $f = \sum c_j T_{u_j} M_{\eta_j} \varphi$ with $u, \eta \in \mathbb{R}^d$. For $F = Tf$, we obtain

$$\begin{aligned} F[\lambda] &= \langle f, T_{\alpha k} M_{\beta n} \varphi \rangle \\ &= \sum c_j \langle T_{u_j} M_{\eta_j} \varphi, T_{\alpha k} M_{\beta n} \varphi \rangle \\ &= \sum c_j \, 2^{-d/2} e^{\pi i \langle \alpha k - u_j, \beta n + \eta \rangle} \varphi_2(\alpha k - u_j) \varphi_2(\beta n - \eta_j) \end{aligned}$$

with $u, \eta \in \mathbb{R}^d$, $\lambda = (k, n)$ and c_k.
Let $m_1, m_2 \in \mathbb{N}$. For a (large) fixed $r > 0$ and (small) $c > 0$, the test sequences are chosen from the following set

$$\Sigma = \left\{ F_\iota[\lambda] = c \cdot \varepsilon \cdot \sum_{j \in \Omega_j} \iota_j \, \langle T_{rj'} M_{rj''} \varphi, T_{\alpha k} M_{\beta n} \varphi \rangle, \qquad \iota \in \Omega, \, j \in \Omega_j, \, \lambda = (k, n) \right\}$$

with

$$\Omega_j = \left\{ j = (j', j'')^\top : j' \in (1, \dots, m_1)^d,\, j'' \in (1, \dots, m_2)^d \right\}$$
$$\Omega_\iota = \left\{ \iota = \left(\iota_j\right)_{j \in \Omega_j} \right\} \in \{0,1\}^{(m_1^d m_2^d)}.$$

Each $F_\iota \in \Sigma$ can be rewritten as follows

$$\begin{aligned}
F_\iota[\lambda] &= c\,\varepsilon \sum_{j \in \Omega_j} \iota_j \, \langle T_{rj'} M_{rj''} \varphi, T_{\alpha k} M_{\beta n} \varphi \rangle \\
&= c\,\varepsilon\, 2^{-d/2} \sum_{j \in \Omega_j} \iota_j \, e^{\pi i \langle \alpha k - rj', \beta n + rj'' \rangle} \, \varphi_2(\alpha k - rj') \varphi_2(\beta n - rj'') \\
&= c\,\varepsilon\, 2^{-d/2} \sum_{j \in \Omega_j} \iota_j \, e^{\pi i \langle \alpha k - rj', \beta n + rj'' \rangle} \, \varphi_2(\lambda - rj) \\
&= c\,\varepsilon\, 2^{-d/2} \sum_{j \in \Omega_j} \iota_j \, e^{\pi i \langle \alpha k - rj', \beta n + rj'' \rangle} \, T_{rj} \varphi_2(\lambda).
\end{aligned}$$

Now, we show that these sequences belong to the parameter space $\bar{\Theta}^p_{u,v}$. For $F_\iota \in \Sigma$ it follows that

$$\begin{aligned}
\|F\|^p_{\bar{\Theta}^p_{u,v}} &= \|F\|^p_{\ell^p_{m_{u,v}}} \\
&= c^p\, \varepsilon^p\, 2^{-dp/2} \, \| \sum_{j \in \Omega_j} \iota_j \, |e^{\pi i \langle \alpha k - rj', \beta n + rj'' \rangle}| \, T_{rj} \varphi_2(\lambda) \|^p_{\ell^p_{m_{u,v}}} \\
&\le c^p\, \varepsilon^p\, 2^{-dp/2} \| \sum_{j \in \Omega_j} T_{rj} \varphi_2(\lambda) \|^p_{\ell^p_{m_{u,v}}} \\
&\overset{(1)}{\le} \text{const.} \cdot c^p\, \varepsilon^p\, 2^{-dp/2} \cdot \sum_{j \in \Omega_j} \| T_{rj} \varphi_2(\lambda) \|^p_{\ell^p_{m_{u,v}}} \\
&\overset{(2)}{\le} \text{const.} \cdot c^p\, \varepsilon^p\, 2^{-dp/2} \cdot \sum_{j \in \Omega_j} m^p_{u,v}(rj) \| \varphi_2(\lambda) \|^p_{\ell^p_{m_{u,v}}} \\
&\le \text{const.} \cdot c^p\, \varepsilon^p\, 2^{-dp/2} \cdot \sum_{j \in \Omega_j} (1 + |rj'|)^{pu} + (1 + |rj''|)^{pv} \\
&= \text{const.} \cdot c^p\, \varepsilon^p\, 2^{-dp/2}\, m_1^d\, m_2^d\, (m_1^{pu} + m_2^{pv}) \\
&= \text{const.} \cdot \varepsilon^p\, (m_1^d\, m_2^d)\, (m_1^{pu} + m_2^{pv})
\end{aligned}$$

For $m_1 \asymp \varepsilon^{-\frac{pv}{d(v+u)+pvu}}$ and $m_2 \asymp \varepsilon^{-\frac{pu}{d(v+u)+pvu}}$, we obtain $F_\iota \in \bar{\Theta}^p_{u,v}$. Following observations have been used.
(1): The essential support of the Gaussian is concentrated around a point, $\lambda = 0$. Therefore, for $r > 0$ sufficiently large, any two functions in the set of Gaussian functions $\{T_{rj}\varphi_2(\lambda)\}_{j\in\Omega_j}$ are almost disjunct. So the norm of the finite sum is equivalent to the finite sum of the norm.
(2): The inequality arises from lemma 8 and lemma 6 ii).

ii) Bounding the ℓ^2 distance

We prove that $\|F_\iota - F_{\iota'}\|^2_{\ell^2} \asymp \varepsilon^{\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}}$ for $\iota \neq \iota'$. This gives us a lower bound and an upper bound for the ℓ^2 distance. The upper bound will be used later to bound the Kullback-Leibler divergence.

The ℓ^2-distance is given by

$$\begin{aligned}\|F_\iota - F_{\iota'}\|^2_{\ell^2} &= \sum_{\lambda\in\Lambda} |F_\iota[\lambda] - F_{\iota'}[\lambda]|^2 \\ &= c^2\,\varepsilon^2\,2^{-d} \sum_{\lambda\in\Lambda} \left| \sum_{j\in\Omega_j} (\iota_j - \iota'_j) e^{\pi i \langle \alpha k - rj', \beta n + rj''\rangle}\, T_{rj}\varphi_2(\lambda) \right|^2\end{aligned}$$

Since the Gaussian function is square integrable, we obtain the following upper bound

$$\begin{aligned}\|F_\iota - F_{\iota'}\|^2_{\ell^2} &\leq c^2\,\varepsilon^2\,2^{-d} \sum_{\lambda\in\Lambda} \left(\sum_{j\in\Omega_j} T_{rj}\varphi_2(\lambda) \right)^2 \\ &\overset{(1)}{\leq} \text{const.} \cdot c^2\,\varepsilon^2\,2^{-d} \sum_{\lambda\in\Lambda} \sum_{j\in\Omega_j} T_{rj}\varphi_2(\lambda)^2 \\ &\leq \text{const.} \cdot \varepsilon^2\, m_1^d m_2^d\end{aligned}$$

Introducing m_1 and m_2 according to contruction of the sequences lead to the desired

upper bound. For the lower bound we obtain

$$\begin{aligned}
|F_\iota - F_{\iota'}|^2 &\geq c^2\,\varepsilon^2\,2^{-d}\left|\sum_{\lambda\in\Lambda}\left(\sum_{j\in\Omega_j}(\iota_j-\iota'_j)e^{\pi i\langle \alpha k - rj',\beta n + rj''\rangle}\,T_{rj}\varphi_2(\lambda)\right)^2\right| \\
&\geq \text{const.}\cdot c^2\,\varepsilon^2\,2^{-d}\left|\sum_{\lambda\in\Lambda}\sum_{j\in\Omega_j}(\iota_j-\iota'_j)^2 e^{2\pi i\langle \alpha k - rj',\beta n + rj''\rangle}\,T_{rj}\varphi_2(\lambda)^2\right| \\
&= \text{const.}\cdot c^2\,\varepsilon^2\,2^{-d}\left|\sum_{j\in\Omega_j}(\iota_j-\iota'_j)^2\sum_{\lambda\in\Lambda} e^{2\pi i\langle \alpha k - rj',\beta n + rj''\rangle}\,T_{rj}\varphi_2(\lambda)^2\right| \\
&\geq \text{const.}\cdot c^2\,\varepsilon^2\,2^{-d}\sum_{j\in\Omega_j}(\iota_j-\iota'_j)^2
\end{aligned}$$

The lower bound depends on the Hamming-distance between the indices, $\|\iota-\iota'\|_2^2$. Given $m_1, m_2 \geq 8$, using the Varshamov-Gilbert bound (Tsybakov, 2009, Lemma 2.9) we may choose $M = \exp(c_1\, m_1^d\, m_2^d)$, $c_1 = (\log 2)/8 > 0$ indices $\left\{\iota \in \{0,1\}^{(m_1^d\, m_2^d)}\right\}$ of Hamming - distance $\|\iota - \tilde{\iota}\|_2^2 \geq m_1^d\, m_2^d/8$ for any two vectors $\iota, \tilde{\iota} \in \{0,1\}^{(m_1^d\, m_2^d)}$. It follows that

$$|F_\iota - F_{\iota'}|^2 \geq \text{const.}\cdot\varepsilon^2\, m_1^d\, m_2^d/8$$

Introducing m_1 and m_2 according to contruction of the sequences lead to the desired lower bound.

iii)Kullback-Leibler divergence

For $P \sim \mathcal{N}(\mu,\sigma^2)$ and $P \sim \mathcal{N}(\tilde{\mu},\tilde{\sigma}^2)$ the Kullback-Leibler divergence is given as

$$\text{KL}(P,Q) = \log\frac{\tilde{\sigma}}{\sigma} + \frac{(\mu-\tilde{\mu})^2+\sigma^2}{2\tilde{\sigma}^2} - \frac{1}{2},$$

see Belov and Armstrong (2011).

From $Y_j \sim \mathcal{N}\big(F_j,\varepsilon^2\big)$, the upper bound of the l_2-distance and the choice of $M = \exp(c_1\, m_1^d\, m_2^d)$, we have that

$$\text{KL}(Y_j,Y_k) = \frac{\|F_j - F_k\|_{l^2}^2}{\varepsilon^2} \lesssim \log(M).$$

□

3.8.7. Proof of Theorem 19

Proof. The proof is based on results of Donoho, Liu, and MacGibbon (1990) showing that on hyperrectangles, the linear minimax risk is within a factor of 1.25 of the minimax risk. Moreover, the linear minimax risk is achieved by diagonal estimators, see Johnstone (2002, Lemma 7.4).

The proof is as follows. First, we construct an orthosymmetric subspace $\bar{\bar{\Theta}} \subseteq \bar{\Theta}_s^p$. A set Θ is said to be solid and orthosymmetric if $\theta = (\theta_1, ..., \theta_k, ...) \in \Theta$ implies that $\xi \in \Theta$ with $|\xi_i| \leq |\theta_i|$ for all i. Orthosymmetric sets can be written as a union of hyperrectangles. We then construct a sequence $F^* \in \bar{\bar{\Theta}}$, such that the linear diagonal oracle estimator on F^* achieves a risk equal to the desired lower bound. It then follows that

$$r_{\inf}(F^*) \leq 1.25 \cdot r_{\min}(\mathcal{R}(F^*)) \leq r_{\min}(\bar{\bar{\Theta}}) \leq r_{\min}(\bar{\Theta}_s^p), \tag{3.40}$$

and we obtain the desired lower bound.

First, the proof under the time assumption.

i) Construction of $\bar{\bar{\Theta}}_s^p()$.
For $Z > 0$ we consider

$$\bar{\bar{\Theta}}_s^p(Z) = \left\{ F \in \ell^2(\Lambda) \quad | \qquad \|F\|_{\bar{\Theta}_s^p} \leq Z \right\}, \tag{3.41}$$

Let $G = a \cdot F$ with $F \in \bar{\Theta}_s^p$, $a = \{a_\lambda\}_{\lambda \in \Lambda}$ and $|a_\lambda| \leq 1$. It follows $\|G\|_{\bar{\Theta}_s^p} = \|a \cdot F\|_{\bar{\Theta}_s^p} \leq \|F\|_{\bar{\Theta}_s^p}$. Therefore, parameter sets of this form are orthosymmetric.

From (3.7) we have

$$\bar{\bar{\Theta}}_s^p\left(\frac{\bar{C}}{\tilde{C}_2}\right) \subseteq \bar{\Theta}_s^p \subseteq \bar{\bar{\Theta}}_s^p\left(\frac{\bar{C}}{\tilde{C}_1}\right)$$

ii) Construction of $F^* \in \bar{\bar{\Theta}}_s^p\left(\frac{\bar{C}}{\tilde{C}_2}\right)$
There exists a $K_0 > 0$ such that $F[\lambda] = 0$ for $\lambda = (k, n)$ and $k \geq K_0$. For $\Delta > 0$, consider the following $2d$-dimensional cube in the time-frequency space

$$D = \Lambda_{0,(K_0, \Delta \cdot K_0)} \setminus \{0, 0\}$$

It follows

$$\begin{aligned} m(\lambda) &\asymp (1 + \|\lambda\|_2^2)^{s/2} \\ &\leq (1 + dK_0^2 + d(\Delta K_0)^2)^{s/2} \\ &\leq d^{\frac{s}{2}}(1 + K_0 + \Delta K_0)^s \\ &\leq d^{\frac{s}{2}}(\Delta + 1)^s (K_0 + 1)^s \end{aligned}$$

as well as

$$\begin{aligned} \#D &= \left(\frac{2\Delta K_0}{\beta}\right)^d \left(\frac{2K_0}{\alpha}\right)^d \\ &\leq \left(\frac{4}{\alpha\beta}\right)^d \Delta^d K_0^{2d} \end{aligned}$$

and

$$\#D \geq \left(\frac{2\Delta K_0}{\beta}\right)^d$$

We construct a sequence F^* satisfying

$$|F^*[\lambda]| = \begin{cases} \epsilon \cdot \|\tilde{h}\|_{L^2} & \text{for } \lambda \in D \\ 0 & \text{else} \end{cases}$$

We have that

$$\sum_{\lambda \in \Lambda} F^*[\lambda]^p m_s(\lambda)^p \leq \left(\frac{4}{\alpha\beta}\right)^d d^{\frac{s}{2}} \left(\epsilon \cdot \|\tilde{h}\|_{L^2}\right)^p (\Delta + 1)^{d+sp} (K_0 + 1)^{2d+sp}$$

For

$$(\Delta + 1)^{d+sp} = \left(\left(\frac{4}{\alpha\beta}\right)^d d^{\frac{s}{2}} \left(\epsilon \cdot \|\tilde{h}\|_{L^2}\right)^p (K_0 + 1)^{2d+sp}\right)^{-1} \left(\frac{\bar{C}}{\tilde{C}_2}\right)^p \tag{3.42}$$

with $\|\tilde{h}\|_{L^2}$ small enough, the sequence F^* belongs to the orthosymmetric set $\bar{\bar{\Theta}}_s^p(\bar{C}/\tilde{C}_2)$

iii) Linear diagonal oracle risk for F^*

The linear diagonal oracle estimator for F^* has the following risk

$$\begin{aligned} r_{\inf}(F^*) &= \sum_{\lambda\in\Lambda} \frac{|F^*[\lambda]|^2\,(\varepsilon\cdot\|\tilde{h}\|_{L^2}^2)^2}{|F^*[\lambda]|^2+(\varepsilon\cdot\|\tilde{h}\|_{L^2}^2)^2} \\ &\geq \left(\frac{2\Delta K_0}{\beta}\right)^d \varepsilon^2 \\ &= \text{const.}\cdot\varepsilon^{\frac{d\,(2-p)+2sp}{d+s\,p}} \end{aligned}$$

With eq. (3.40) we obtain the lower bound.

Anisotropic case

The space

$$\bar{\bar{\Theta}}^p_{u,v}(Z) = \left\{F\in\ell^2(\Lambda) \quad | \qquad \|F\|_{\bar{\Theta}^p_{u,v}} \leq Z\right\}. \tag{3.43}$$

is an orthosymmetric subspace of $\bar{\Theta}^p_{u,v}$.

From (3.7) we have

$$\bar{\bar{\Theta}}^p_{u,v}(\frac{\bar{C}}{\tilde{C}_2}) \subseteq \bar{\Theta}^p_{u,v} \subseteq \bar{\bar{\Theta}}^p_{u,v}(\frac{\bar{C}}{\tilde{C}_1}).$$

For $\Delta>0$, let us consider the following $2d$-dimensional cube in the time-frequency space

$$D = \Lambda_{0,(K_0,\Delta\cdot K_0)} \setminus \{0,0\}$$

For $\lambda\in D$ we obtain

$$\begin{aligned} m(\lambda) &= (1+\|\alpha k\|_2^2)^{u/2} + (1+\|\beta n\|_2^2)^{v/2} \\ &\leq (1+dK_0^2)^{u/2} + (1+d(\Delta K_0)^2)^{v/2} \\ &\leq d^{\frac{\max(u,v)}{2}}\left((1+K_0)^u + (1+\Delta K_0)^v\right) \\ &\leq d^{\frac{\max(u,v)}{2}}\left((1+K_0)^u + (1+\Delta)^v(1+K_0)^v\right) \\ &\leq 2\cdot d^{\frac{\max(u,v)}{2}}(1+\Delta)^v\left((1+K_0)^{\max(u,v)}\right) \end{aligned}$$

as well as

$$\begin{aligned}\#D &= \left(\frac{2\Delta K_0}{\beta}\right)^d \left(\frac{2K_0}{\alpha}\right)^d \\ &\leq \left(\frac{4}{\alpha\beta}\right)^d \Delta^d K_0^{2d}\end{aligned}$$

and

$$\#D \geq \left(\frac{2\Delta K_0}{\beta}\right)^d$$

We construct a function F^* satisfying

$$|F^*[\lambda]| = \begin{cases} \epsilon \cdot \|\tilde{h}\|_{L^2} & \text{for } \lambda \in D \\ 0 & \text{else} \end{cases}$$

We have that

$$\sum_{\lambda\in\Lambda} |F^*[\lambda]|^p m_{u,v}(\lambda)^p \leq 2 \cdot \left(\frac{4}{\alpha\beta}\right)^d d^{\frac{\max(u,v)}{2}} \left(\epsilon \cdot \|\tilde{h}\|_{L^2}\right)^p (\Delta+1)^{d+vp} (K_0+1)^{2d+\max(u,v)p}$$

For

$$(\Delta-1)^{d+vp} \leq \left(2 \cdot \left(\frac{4}{\alpha\beta}\right)^d d^{\frac{\max(u,v)}{2}} \left(\epsilon \cdot \|\tilde{h}\|_{L^2}\right)^p (K_0+1)^{2d+\max(u,v)p}\right)^{-1} \left(\frac{\bar{C}}{\tilde{C}_2}\right)^p$$

with $\|\tilde{h}\|_{L^2}$ small enough, the sequence F^* belongs to the orthosymmetric set $\bar{\bar{\Theta}}_s^p\left(\bar{C}/\tilde{C}_2\right)$

The linear oracle estimator for F^* has the following risk

$$\begin{aligned} r_{\inf}(F^*) &= \sum_{\lambda\in\Lambda} \frac{|F^*[\lambda]|^2 (\varepsilon \cdot \|\tilde{h}\|_{L^2}^2)^2}{|F^*[\lambda]|^2 + (\varepsilon \cdot \|\tilde{h}\|_{L^2}^2)^2} \\ &\geq \left(\frac{2\Delta K_0}{\beta}\right)^d \varepsilon^2 \\ &= \text{const.} \cdot \varepsilon^{\frac{d\,(2-p)+2v\,p}{d+s\,p}}. \end{aligned}$$

The proof for under the frequency assumption is analogous and therefore omitted.

□

3.8.8. Proof of Corollary 20

Proof. We give a proof only for the general anisotropic case, since it is identical to the result in the isotropic case for $u = v$. A proof of this theorem is obtained by showing that the thresholding estimator achieves this rate on a finite subspace of the parameter set. This finite subspace is chosen so that the orthogonal projection error onto it is of the same order as the estimation risk. The rate then holds for the entire parameter set.

For $K > 0$, consider the nonempty finite index set $\Lambda_{0,K} \subseteq \Lambda_0$.

From Theorem 18 it follows

$$r_{\mathrm{pr}}(\bar{\Theta}^p_{u,v}(\Lambda_{0,K})) \sim \varepsilon^{\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}}.$$

We derive from Lemma 14

$$\varepsilon_{\bar{\Theta}^p_{u,v}}(\Lambda, \Lambda_{0,K}) = \mathcal{O}(C^2\,\varepsilon^{\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}})$$

by taking $K \gtrsim \varepsilon^{-\frac{(2-p)\,d\,(v+u)+2pvu}{\max(u,v)(d(v+u)+pvu)}}$.

Furthermore, we obtain for this choice of K and using $\#\Lambda_{0,K} \leq \text{const.} \cdot K^{2d}$

$$\log(\#\Lambda_{0,K}) \leq \text{const.} \cdot \log(\varepsilon^{-1})$$

By Proposition 15 the thresholding estimator $\hat{F}^j_{\mu_{\mathrm{uni}}}$ then achieves following rate on the finite subspace

$$r(\hat{F}^j_{\mu_{\mathrm{uni}}}, \bar{\Theta}^p_{u,v}(\Lambda_{0,K})) = \mathcal{O}(\log(\varepsilon^{-1})\,\varepsilon^{\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}}).$$

This is of the same order as the projection error, so the rate can be transferred to $\bar{\Theta}^p_{u,v}$. We obtain

$$r(\hat{F}^j_{\mu_{\mathrm{uni}}}, \bar{\Theta}^p_{u,v}) = \mathcal{O}(\log(\varepsilon^{-1})\,\varepsilon^{\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}}).$$

In particular, we obtain for $F \in \bar{\Theta}^p_{u,v}$

$$r(\hat{F}^{\mathrm{j}}_{\mu_{\mathrm{uni}}}, F) \leq \text{const.} \cdot \max\left(\|F\|^2_{\bar{\Theta}^p_{u,v}}, \|F\|^{\frac{pd(u+v)}{d(v+u)+pvu}}_{\bar{\Theta}^p_{u,v}}\right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{(2-p)d(u+v)+2\,puv}{d\,(u+v)+puv}}.$$

The factor $\max\left(\|F\|^2_{\bar{\Theta}^p_{u,v}}, \|F\|^{\frac{pd(u+v)}{d(v+u)+pvu}}_{\bar{\Theta}^p_{u,v}}\right)$ follows from the proof of Theorem 16 and Lemma 14.

The proof is completed with Lemma 13. □

3.8.9. Proof of Corollary 21

Proof. The proof of this corollary is analogous to the proof of Corollary 20. Again, we only provide a proof for the general anisotropic case. A proof of this theorem is obtained by showing that the thresholding estimator achieves this rate on a finite subspace of the parameter set. This finite subspace is chosen such that the orthogonal projection error onto it is of the same order as the estimation risk. The rate then holds for the entire parameter set.

For $K > 0$, consider the nonempty finite index set $\Lambda_{0,K} \subseteq \Lambda_0$.

From Theorem 19 it follows under the time assumption

$$r_{\mathrm{pr}}(\Theta^p_{u,v}(\Lambda_{0,K})) \sim \varepsilon^{\frac{d\,(2-p)+2vp}{d+v\,p}}.$$

We derive from Lemma 14

$$\varepsilon_{\Theta^p_{u,v}}(\Lambda, \Lambda_{0,K}) = \mathcal{O}(C^2\,\varepsilon^{\frac{d\,(2-p)+2vp}{d+v\,p}})$$

by taking $K \gtrsim \varepsilon^{-\frac{d\,(1-\frac{p}{2})+vp}{\max(u,v)(d+v\,p)}}$.

Also, for this choice of K and using $\#\Lambda_{0,K} \leq \text{const.} \cdot K^d$ we obtain

$$\log(\#\Lambda_{0,K}) \leq \text{const.} \cdot \log(\varepsilon^{-1})$$

Using Proposition 15 we obtain for the thresholding estimator $\hat{F}^j_{\mu_{\mathrm{uni}}}$

$$r(\hat{F}^j_{\mu_{\mathrm{uni}}}, \bar{\Theta}^p_{u,v}(\Lambda_{0,K})) = \mathcal{O}(\log(\varepsilon^{-1})\,\varepsilon^{\frac{d\,(2-p)+2vp}{d+v\,p}}).$$

This is of the same order as the projection error. Therefore we obtain

$$r(\hat{F}^j_{\mu_{\mathrm{uni}}}, \bar{\Theta}^p_{u,v}) = \mathcal{O}(\log(\varepsilon^{-1})\,\varepsilon^{\frac{d\,(2-p)+2vp}{d+v\,p}}).$$

In particular, we obtain for $F \in \bar{\Theta}^p_{u,v}$

$$r(\hat{F}^{\mathrm{j}}_{\mu_{\mathrm{uni}}}, F) \leq \text{const.} \cdot \max\left(\|F\|^2_{\bar{\Theta}^p_{u,v}}, \|F\|^{\frac{pd}{d+pv}}_{\bar{\Theta}^p_{u,v}}\right) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{(2-p)d+2\,vp}{d+pv}}.$$

The factor $\max\left(\|F\|^2_{\bar{\Theta}^p_{u,v}}, \|F\|^{\frac{pd}{d+pv}}_{\bar{\Theta}^p_{u,v}}\right)$ arises from the proof of Theorem 17 and Lemma 14.

Lemma 13 completes the proof.

□

4. Estimation over α-Modulation Spaces

In this chapter, we extend the results of the previous chapter to a broader class of functions. Specifically, we give nonparametric function estimation within the smoothness class $\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})$ for $\alpha \in [0,1)$. This class includes well-known spaces such as modulation spaces and Besov spaces, which allows us to make comparisons between the results obtained here and those previously obtained in Chapter 3. To accomplish this, we construct an infinite-dimensional sequence model using the α modulation frame expansion. This model serves as the basis for the development of an adaptive estimator in a finite digital framework based on the principle of thresholding the coefficients. We then derive the convergence rate for this estimator. The overall framework remains unchanged, allowing us to reuse numerous tools and results from the previous chapter. Consequently, some explanations and details may be omitted, assuming familiarity with the concepts and techniques established earlier. The results presented here focus on the one-dimensional case, while possible extensions to the multidimensional case are reserved for future investigations. It should also be noted that we only present results for compactly supported functions on the unit interval, since similar results can be derived for the entire space $\mathbb{R}$. This chapter follows an analogous structure to the previous one, thus providing a consistent and coherent progression of the material.

4.1. Analog Model

We recall the Gaussian white noise model (3.1)

$$\mathrm{d}\mathbf{Y}(x) = f(x)\ \mathrm{d}x + \varepsilon\,\mathrm{d}\mathbf{W}(x), \qquad x \in \mathbb{R},$$

where a square-integrable signal $f \in L^2(\mathbb{R})$ is observed. We assume that the function f is part of the smoothness space

$$\Theta^p_{0,s:\alpha} = \left\{ f \in L^2(\mathbb{R}) \quad | \quad \mathrm{supp}(f) \subseteq [0,1], \quad \|f\|_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})} \leq \bar{C} \right\}. \tag{4.1}$$

with $p \in [1,2]$, $s \geq 0$, $\alpha \in [0,1)$, and the weight function

$$m_{u,v}(x,\omega) = (1+|x|)^u + (1+|\omega|)^v.$$

The parameter space $\Theta^p_{0,s:\alpha}$ is a bounded subset of $\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})$. We denote the norm on $\Theta^p_{0,s:\alpha}$ by

$$\|f\|_{\Theta^p_{0,s:\alpha}} = \|f\|_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})}.$$

For $\alpha = 0$, in the one-dimensional setting we obtain the same parameter space as in Chapter 3, $\Theta^p_{0,s:0} = \Theta^p_{0,s}$. Again, the goal here is to find an estimator $\hat{f} \in \Theta^p_{0,s:\alpha}$ that minimizes the risk previously defined in (3.4). This is done using the framework and tools defined in the previous chapter. First, we construct an infinite-dimensional Gaussian sequence model using α-modulation frames. A digital model is then derived by projecting the representation of the signal into a finite subspace of $\ell^2(\mathbb{Z}^2)$. We propose an upper bound on the linear approximation error. Finally, we show that the oracle projector achieves the minimax convergence rate. When the resolution is high enough, the projection estimator based on coordinate-wise universal thresholding of the α-modulation frame coefficients is close to optimal in the minimax sense.

4.2. Gaussian Sequence Model

In this section, we construct an infinite-dimensional Gaussian sequence model based on frame expansions. As shown in Lemma 13, an estimator in this sequence model induces

an estimator in the analog model 4.1 that achieves the same risk, and vice versa.

Let $\Lambda = \{(j,k) : k,n \in \mathbb{Z}^2\}$. For $\alpha \in [0,1)$, $a,b > 0$ and, an admissible real-valued nonzero and compactly supported window function $h \in \mathcal{S}(\mathbb{R})$ with $\mathcal{F}h \neq 0$, we consider the family of functions defined in (2.46)

$$\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b) = \left\{ M_{p_\alpha(j)} D_{s_\alpha^{-1}(j)} T_{ak}\, h \right\}_{(j,k)\in\mathbb{Z}^2}, \qquad a,b > 0. \tag{4.2}$$

It follows from Theorem 11 that for sufficiently small $a > 0$, this family defines a frame for the Hilbert space $L^2(\mathbb{R})$, with lower and upper frame bounds A and B respectively, and $0 < A \leq B$. Each function $f \in L^2(\mathbb{R})$ has a representation in terms of the frame elements

$$f = \sum_{\lambda\in\Lambda} \langle f, h_\lambda^\alpha \rangle \tilde{h}_\lambda^\alpha. \tag{4.3}$$

with $h_\lambda^\alpha \in \mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b)$ and dual frames $\tilde{h}_\lambda^\alpha \in L^2(\mathbb{R})$.

in addition, Theorem 12 shows that the family $\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b)$ also defines a Banach frame for $\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})$. So we obtain the following norm equivalence $\|f\|_{\Theta^p_{0,s:\alpha}} \asymp \|\langle f, h_\lambda^\alpha \rangle_{\lambda\in\Lambda}\|_{\ell^p_{\bar{m}_{0,s:\alpha}}}$, i.e. there exists constants $0 < \tilde{C}_1 \leq \tilde{C}_2$ so that

$$\tilde{C}_1 \|f\|_{\Theta^p_{0,s:\alpha}} \leq \|\langle f, h_\lambda^\alpha \rangle_{\lambda\in\Lambda}\|_{\ell^p_{\bar{m}_{0,s}}} \leq \tilde{C}_2 \|f\|_{\Theta^p_{0,s:\alpha}} \tag{4.4}$$

with

$$\bar{m}_{u,v:\alpha}(j,k) = (1 + |a\cdot k|)^u + (1 + |p_\alpha(j)|)^v.$$

In particular, Theorem 12 also show the existence of a sequence $\{\tilde{h}_\lambda^\alpha\}_{\lambda\in\Lambda} \in \mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b)'$ such that the series expansion (4.3) unconditionally converges in $\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})$. Due to the isometry of the L^2- and l^2-risk and the stability of the frame operator, for any estimator $\hat{F} \in \ell^2\{\Lambda\}$ of the frame coefficients $F = \{\langle f, h_\lambda^\alpha \rangle\}_{\lambda\in\Lambda}$, there exists an equivalent estimator $\hat{f} = T^*\hat{F} \in L^2(\mathbb{R}^d)$ of the function f that achieves the same rate, and vice versa.

Using the commutation relation $D_a T_x = T_{ax} D_a$, and the isometry property of time-

frequency shifts (2.13), we obtain

$$\begin{aligned}\|\tilde{h}_\lambda^\alpha\|_{L^2}^2 &= \|M_{p_\alpha(j)} T_{aks_\alpha^{-1}(j)} D_{s_\alpha^{-1}(j)}\, h\|_{L^2}^2 \\ &= \|D_{s_\alpha^{-1}(j)}\, h\|_{L^2}^2 \\ &= \|h\|_{L^2}^2.\end{aligned}$$

An equivalent infinite-dimensional Gaussian sequence model is then given by

$$Y[\lambda] = F[\lambda] + W[\lambda], \qquad \lambda \in \Lambda \tag{4.5}$$

with

$$\begin{aligned}F[\lambda] &= \int \tilde{h}_\lambda^\alpha \, df = \langle f, \tilde{h}_\lambda^\alpha \rangle \\ W[\lambda] &= \int \tilde{h}_\lambda^\alpha \, d\mathbf{W} \sim \mathcal{N}\big(0, \varepsilon^2 \|h\|_{L^2}^2\big), \\ Y[\lambda] &= \int \tilde{h}_\lambda^\alpha \, d\mathbf{Y} \sim \mathcal{N}\big(F[\lambda], \varepsilon^2 \|h\|_{L^2}^2\big) \\ \mathrm{Cov}(Y[\lambda_1], Y[\lambda_2]) &= \epsilon^2 \langle h_{\lambda_1}^\alpha, h_{\lambda_2}^\alpha \rangle\end{aligned}$$

and the corresponding parameter set

$$\bar{\Theta}_{0,s:\alpha}^p := \bar{\Theta}_{0,s:\alpha}^p(\Lambda) = \left\{ F \in \ell^2\{\Lambda\} \quad | \qquad \|T^* F\|_{\Theta_{0,s:\alpha}^p} < \bar{C} \right\}.$$

The equivalence (4.4) can be reformulated as follows

$$\tilde{C}_1 \|T^* F\|_{\Theta_{0,s:\alpha}^p} \le \|F\|_{\bar{\Theta}_{0,s:\alpha}^p} \le \tilde{C}_2 \|T^* F\|_{\Theta_{0,s:\alpha}^p} \tag{4.6}$$

with $\|F\|_{\bar{\Theta}_{0,s:\alpha}^p} = \|F\|_{\ell_{\bar{m}_{0,s:\alpha}}^p}$.

4.3. Projection Error on ℓ_2-Subset

In this section, we define the digital model by reducing the infinite-dimensional Gaussian sequence model to a finite-dimensional Gaussian sequence model. We also derive an

upper bound on the projection error on a finite index subset.

For a nonempty finite index set $\Lambda_0 \subseteq \Lambda$, we consider the following digital model

$$Y[\lambda] = F[\lambda] + W[\lambda], \qquad \lambda \in \Lambda_0 \tag{4.7}$$

with the parameter set

$$\bar{\Theta}^p_{0,s:\alpha}(\Lambda_0) = \left\{ F \in \ell^2\{\Lambda_0\} \quad | \qquad \|T^* F\|_{\Theta^p_{0,s:\alpha}} < \bar{C} \right\}.$$

Now, we examine the linear projection error in the time-frequency plane. For $K_1, K_2 > 0$, we consider the set

$$\Lambda_{0,(K_1,K_2)} = \{\lambda \in \Lambda : |k| \leq K_1, \quad |j| \leq K_2\}, \tag{4.8}$$

with $\Lambda_{0,K} = \Lambda_{0,(K,K)}$.

This corresponds to the following subset of the time-frequency space

$$\left\{ (x,\omega) \in \mathbb{R}^{2d} \quad | \quad \begin{array}{l} x = x(\omega) = s_\alpha^{-1}(j)\, a\, k, \quad \|k\|_2 \leq K_1, \quad \|j\|_2 \leq K_2, \\ \omega = p_\alpha(j), \quad \|j\|_2 \leq K_2 \end{array} \right\}. \tag{4.9}$$

Under the time assumption 1, there is a $K_0 > 0$ such that we can restrict the time domain to $|k| \leq K_0$ without losing information.

Remark 10. Note that the area of the time-frequency plane $\mathbb{R}^2$ covered by this discrete set varies depending on the dilation parameter $s_\alpha^{-1}(\cdot)$. This parameter stretches or compresses the discrete time support. Low-frequency regions have larger time steps, resulting in lower time resolution, while high-frequency region have smaller time steps, resulting in higher time resolution.

Lemma 22. *Let $s \geq 0$. For $F \in \bar{\Theta}^p_{0,s:\alpha}$ and $p \in [1,2]$ we have:*

$$\sum_{\lambda \in \Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^2 \leq \left(b\,(1-\alpha)\,K_2\right)^{-\frac{2s}{1-\alpha}} \tilde{C}_2^2 \|T^* F\|^2_{\bar{\Theta}^p_{0,s:\alpha}}. \tag{4.10}$$

It then follows

$$\varepsilon_{\bar{\Theta}^p_{0,s:\alpha}}(\Lambda, \Lambda_{0,K}) = O\big(C \cdot (b\,(1-\alpha)\cdot K)^{-\frac{2s}{(1-\alpha)}}\big) \tag{4.11}$$

with $C = \tilde{C}_2^2 \cdot \bar{C}^2$

For $\alpha = 0$, we recall the results of Lemma 14 for the anisotropic case $u = 0$, and $v = s$. The error decreases with increasing decay rate s or radius K. In addition to the observations made in Lemma 14, we observe that the error diverges for $\alpha \to 1$. This isn't very surprising, given the construction of α-modulation spaces. For $\alpha \to 1$ the factor $s_\alpha^{-1}(\cdot)$ decreases, so the subset of the time-frequency plane covered by the discrete representation simply shrinks, see (4.9). As noted in Remark 10 the frequency-dependent time resolution increases with $\alpha \to 1$. Obviously, a fixed resolution will eventually become insufficient and capture less information.

4.4. Thresholding Estimator

Here we propose an estimator in the digital model (4.7). This estimator attenuates the noise by thresholding the frame coefficients. The extension of Donoho and Johnstone (1994)'s results provided in Proposition 15 is not frame specific and remains valid in this context. We give an upper bound on the risk of the nonlinear oracle projector. This in fact induces an upper bound on the risk of the thresholding estimator.

In the digital model (4.7) with parameter set $\bar{\Theta}^p_{0,s:\alpha}(\Lambda_0)$, we define the following estimator based on thresholding the complex frame coefficients at level $\mu \in \mathbb{R}$

$$\begin{aligned}\hat{F}^{\mathrm{j}}_\mu &= \big\{t^{\mathrm{j}}_\mu(Y[\lambda])\big\}_{\lambda\in\Lambda_0} \\ &= \big\{t^{\mathrm{j}}_\mu(\Re(Y[\lambda])) + i\,t^{\mathrm{j}}_\mu(\Im(Y[\lambda]))\big\}_{\lambda\in\Lambda_0}\,.\end{aligned}$$

with $\mathrm{j} \in \{\mathrm{h}, \mathrm{s}\}$. Proposition 15 shows that the risk of this estimator is within a constant factor of the risk of the nonlinear oracle projector $\hat{F}_{\mathrm{pr}}$. Next we derive an upper bound

on the risk of the nonlinear oracle projector on $\bar{\Theta}^p_{0,s:\alpha}(\Lambda_{0,K})$ for sufficiently large $K > 0$.

Theorem 23. *Let $p \in [1,2]$, $s \geq 0$ and $K > 0$. In the digital model (4.7) with $F \in \bar{\Theta}^p_{0,s:\alpha}(\Lambda_{0,K})$, and K sufficiently large, the risk of the nonlinear oracle projector error has the following upper bound,*

$$r_{pr}(F) \leq \text{const.} \cdot \|F\|^{\frac{(1-\alpha)\,p}{(1-\alpha)+s\,p}}_{\bar{\Theta}^p_{0,s:\alpha}} (1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\, \varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}. \tag{4.12}$$

For $\alpha = 0$ we get the same results as in the previous context for $d = 1$, $u = 0$ and $v = s$, see Theorem 17. The bound diverges for $\alpha \to 1$. We do not observe the expected rate of $\varepsilon^{\frac{2s}{1+2s}}$ obtained by Donoho and Johnstone (1998) on Besov spaces. The reason is that although $\mathcal{M}^p_{0,s:\alpha}$ coincides with the inhomogeneous Besov space for $\alpha \to 1$, the frames expansion in (2.47) is not valid for the limiting case of $\alpha = 1$. Besov spaces can only be characterized by exponentially localized frames.

4.5. Lower Bound

In this section, we derive a lower bound on the minimax risk in the infinite-dimensional Gaussian model (4.5). The derivation is analogous to that of Theorem 19 and relies on the geometric properties of the parameter set.

Theorem 24. *For $p \in [1,2]$ consider the model (4.5) with $F \in \bar{\Theta}^p_{0,s:\alpha}$. Then we have that*

$$\liminf_{\varepsilon \downarrow 0} \left((1-\alpha)^{\frac{sp}{(1-\alpha)+sp}}\, \varepsilon^{-\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}} \inf_{\hat{F}_\varepsilon} r(\hat{F}_\varepsilon, \bar{\Theta}^p_{0,s:\alpha}) \right) > 0,$$

where $\hat{F}_\varepsilon$ is an arbitrary estimator in (4.5) based on the observation Y.

For $\alpha = 0$ the rate coincides with that obtained in Theorem 19. This bound diverges for $\alpha \to 1$. Given this lower bound, the universal threshold estimator is thus optimal in the

digital model eq. (4.7).

4.6. Optimal Thresholding Estimation

Finally, we show the optimality of the thresholding estimator in the analog model (4.1). This is done in a similar way to Section 3.6. For any finite index set $\Lambda_{0,K}$, the threshold estimator has a risk close to the risk of the nonlinear oracle projector, and thus is nearly optimal on the set $\Theta^p_{0,s:\alpha}(\Lambda_{0,K})$. For K large enough, the projection error on $\Theta^p_{0,s:\alpha}(\Lambda_{0,K})$ and the estimation error are of the same order. The rate is then translated to $\Theta^p_{0,s:\alpha}$.

Corollary 25. *For $p \in [1,2]$, and $\alpha \in [0,1)$, consider model (4.1) with $f \in \Theta^p_{0,s:\alpha}$. Choosing the universal threshold $\mu_{\mathrm{uni}} = \varepsilon \, \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_{0,K})}$ and taking $K \gtrsim \Big((1-\alpha)^{\frac{s}{1-\alpha}} \, \varepsilon\Big)^{-\frac{(1-\alpha)(1-\frac{p}{2})+s\,p}{(1-\alpha)+s\,p} \cdot \frac{1-\alpha}{s}}$, we obtain for the threshold estimator with $\mathrm{j} \in \{\mathrm{s}, \mathrm{h}\}$ the bound*

$$r^{\mathrm{j}}_{\mu_{\mathrm{uni}}}(f) \leq \text{ const.} \cdot C \cdot \log\left((1-\alpha)^{\frac{s}{1-\alpha}} \, \varepsilon\right) \, (1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}} \, \varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}$$

with $C = \max\left(\|f\|^2_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})}, \|f\|^{\frac{(1-\alpha)\,p}{(1-\alpha)+s\,p}}_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})}\right)$

The constant depends on the properties of the frame and the thresholding method. Together with the lower bound obtained in Theorem 24 we obtain the optimality of the thresholding estimators.

4.7. Proofs

We will now provide proofs for the results in this chapter.

4.7.1. Proof of Lemma 22

Proof. The proof is similar to the proof of lemma 14. We recall the weight function $\bar{m}_{u,v}(j,k) = (1+|a\cdot k|^2)^{u/2} + (1+|p_\alpha(j)|^2)^{v/2}$ with $j,k \in \mathbb{Z}$ and the position function $p_\alpha(j) = \operatorname{sgn}(j)\left((1+(1-\alpha)\cdot b\cdot|j|)^{\frac{1}{1-\alpha}} - 1\right)$.

For $\lambda \in \Lambda^c_{0,(K_1,K_2)}$ we have

$$\begin{aligned}\bar{m}_{0,s}(\lambda) &\geq |p_\alpha(K_2)|^s \\ &\geq \left((1+b\cdot(1-\alpha)K_2)^{\frac{1}{1-\alpha}} - 1\right)^s \\ &\geq (b\cdot(1-\alpha)K_2)^{\frac{s}{1-\alpha}}\end{aligned}$$

Since the p-norm is monotonously decreasing, we have

$$\begin{aligned}\sum_{\lambda\in\Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^2 &\leq \Big(\sum_{\lambda\in\Lambda^c_{0,(K_1,K_2)}} |F[\lambda]|^p\Big)^{2/p} \\ &\leq (b\cdot(1-\alpha)K_2)^{-\frac{2s}{1-\alpha}}\Big(\sum_{\lambda\in\Lambda^c_{0,(K_1,K_2)}} \bar{m}_{u,v}(\lambda)^p\,|F[\lambda]|^p\Big)^{2/p} \\ &= (b\cdot(1-\alpha)K_2)^{-\frac{2s}{1-\alpha}}\|F\|^2_{\bar{\Theta}^p_{u,v}}.\end{aligned}$$

We also have

$$\begin{aligned}\varepsilon_{\Theta^p_{0,s:\alpha}}(\Lambda,\Lambda_{0,K}) &= \sup_{F\in\bar{\Theta}^p_{0,s:\alpha}} \|F - P_{\Lambda_0}F\|^2 \\ &\leq \sup_{F\in\bar{\Theta}^p_{0,s:\alpha}} \left((b\cdot(1-\alpha)K)^{-\frac{2s}{1-\alpha}}\|F\|^2_{\bar{\Theta}^p_{0,s:\alpha}}\right) \\ &\leq \sup_{F\in\bar{\Theta}^p_{0,s:\alpha}} \left(\tilde{C}_2^2\cdot(b\cdot(1-\alpha)K)^{-\frac{2s}{1-\alpha}}\|F\|^2_{\bar{\Theta}^p_{0,s:\alpha}}\right) \\ &= O\big(C\left((1-\alpha)\,K\right)^{-\frac{2s}{(1-\alpha)}}\big)\end{aligned}$$

with $C = \tilde{C}_2^2\cdot\bar{C}^2\cdot b^{-\frac{2s}{1-\alpha}}$ using (4.6).

□

4.7.2. Proof of Theorem 23

Proof. The proof of this lemma is similar to that of Theorem 17. Some explanations will therefore be omitted. There exists a $0 < K_0 \leq K$ such that $F[\lambda] = 0$ for $\lambda \in \Lambda^c_{0,(K_0,K_s)}$ and for all $K_s > 0$.

For convenience, let assume that $\|\tilde{h}\|^2_{L^2} = 1$. Note that for $K > 0$

$$\#\Lambda_{0,K} \leq CK_0K$$

where the constant C depends on the parameters, a and b. Let $0 < K_s \leq K$. It follows Let $K_u, K_v > 0$ with $K_u \leq K$ and $K_v \leq K$.

$$\begin{aligned}
r_{\mathrm{pr}}(F) &= \sum_{\lambda\in\Lambda_{0,K}} \min(\varepsilon^2, |F[\lambda]|^2) \\
&= \sum_{\lambda\in\Lambda_{0,(K_0,K_s)}} \min(\varepsilon^2, |F[\lambda]|^2) + \sum_{\lambda\in\Lambda^c_{0,(K_0,K_s)}} \min(\varepsilon^2, |F[\lambda]|^2) \\
&\leq \varepsilon^2\, CK_0K_s + \epsilon^{2-p} \sum_{\lambda\in\Lambda^c_{0,(K_s,K_s)}} |F[\lambda]|^p \\
&\leq \varepsilon^2\, CK_0K_s + \epsilon^{2-p}\, (b\cdot(1-\alpha)K_s)^{-\frac{sp}{1-\alpha}}\, \|F\|^p_{\bar{\Theta}^p_{0,s:\alpha}}
\end{aligned}$$

Balancing both terms with

$$K = \text{const.} \cdot \|F\|^{\frac{(1-\alpha)p}{1-\alpha+sp}}_{\bar{\Theta}^p_{0,s:\alpha}} (1-\alpha)^{-\frac{sp}{1-\alpha+sp}}\, \varepsilon^{-\frac{(1-\alpha)p}{1-\alpha+sp}}$$

gives the results. □

4.7.3. Proof of Theorem 24

Proof. The proof is similar to that of 19. Some explanations are therefore omitted.

i) Construction of $\bar{\bar{\Theta}}^p_{0,s:\alpha}$

For $Z > 0$ we consider

$$\bar{\bar{\Theta}}^p_{0,s:\alpha}(Z) = \left\{ F \in L^2(\mathbb{R}^d) \quad | \qquad \|F\|_{\bar{\Theta}^p_{0,s:\alpha}} \leq Z \right\}. \tag{4.13}$$

This parameter set is an orthosymmetric set.

From (4.6) we have

$$\bar{\bar{\Theta}}^p_{0,s:\alpha}(\frac{\bar{C}}{\tilde{C}_2}) \subseteq \Theta^p_{0,s:\alpha} \subseteq \bar{\bar{\Theta}}^p_{0,s:\alpha}(\frac{\bar{C}}{\tilde{C}_1}).$$

i) Construction of $F^* \in \bar{\bar{\Theta}}^p_{0,s:\alpha}$

From the time assumption, there exists a $K_0 > 0$ such that $F[\lambda] = 0$ for $\lambda = (j,k)$ and $k \geq K_0$. For $\Delta > 0$, we consider the following subset of the time-frequency space

$$D = \Lambda_{0,(K_0,\Delta\cdot K_0)} \setminus \{0,0\}$$

For $\lambda \in D$ we have

$$\begin{aligned}
\bar{m}_{0,s}(\lambda) &\leq 2 \cdot (1 + |p_\alpha(j)|^2)^{s/2} \\
&= 2 \cdot (1 + ((1 + (1-\alpha) \cdot b \cdot |j|)^{\frac{1}{1-\alpha}} - 1)^2)^{s/2} \\
&\leq 2 \cdot (1 + (1 + b \cdot (1-\alpha) \cdot |j|)^{\frac{2}{1-\alpha}})^{s/2} \\
&\leq 4 \cdot (1 + b \cdot (1-\alpha) \cdot |j|)^{\frac{s}{1-\alpha}} \\
&\leq 4 \cdot (1 + b \cdot \Delta)^{\frac{s}{1-\alpha}} (1 + K_0)^{\frac{s}{1-\alpha}} (2-\alpha)^{\frac{s}{1-\alpha}}
\end{aligned}$$

as well as

$$2(\Delta + 1)K_0 \leq \#D \leq 4(\Delta + 1)K_0^2$$

We construct a sequence F^* satisfying

$$|F^*[\lambda]| = \begin{cases} \epsilon \cdot \|\tilde{h}\|_{L^2} & \text{for } \lambda \in D \\ 0 & \text{else} \end{cases}$$

For b small enough, we have that

$$\sum_{\lambda\in\Lambda} |F^*[\lambda]|^p \bar{m}_{0,s}(\lambda)^p \leq 4^{p+1} \left(\epsilon \cdot \|\tilde{h}\|_{L^2}\right)^p \cdot (1+\Delta)^{\frac{sp+1-\alpha}{1-\alpha}} (1+K_0)^{\frac{sp+2(1-\alpha)}{1-\alpha}} (2-\alpha)^{\frac{sp}{1-\alpha}}$$

For any $\Delta > 0$ satisfying

$$(1+\Delta)^{\frac{sp+1-\alpha}{1-\alpha}} = \left(4^{p+1}\left(\epsilon\cdot\|\tilde{h}\|_{L^2}\right)^p\cdot(1+K_0)^{\frac{sp+2(1-\alpha)}{1-\alpha}}(2-\alpha)^{\frac{sp}{1-\alpha}}\right)^{-1}\left(\frac{\bar{C}}{\tilde{C}_2}\right)^p \quad (4.14)$$

with $\epsilon\cdot\|\tilde{h}\|_{L^2}$ small enough, the sequence F^* belongs to the orthosymmetric set $\bar{\bar{\Theta}}^p_{0,s:\alpha}\big(\bar{C}/\tilde{C}_2\big)$.

The linear oracle estimator for F^* has the following risk

$$\begin{aligned} r_{\inf}(F^*) &= \sum_{\lambda\in\Lambda}\frac{|F^*[\lambda]|^2\,(\varepsilon\cdot\|\tilde{h}\|^2_{L^2})^2}{|F^*[\lambda]|^2+(\varepsilon\cdot\|\tilde{h}\|^2_{L^2})^2} \\ &\geq (\Delta+1)K_0\,\varepsilon^2 \\ &\geq \text{const.}\cdot(1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\,\varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}} \end{aligned}$$

where the last equality is obtained by choosing Δ according to eq. (4.14).

□

4.7.4. Proof of Corollary 25

Proof. A proof of this theorem is obtained by showing that the thresholding estimator achieves this rate on a finite subspace of the parameter set. This finite subspace is chosen so that the orthogonal projection error onto it is of the same order as the estimation risk. The rate then holds for the entire parameter set.

For $K > 0$, let consider the nonempty finite index set $\Lambda_{0,K} \subseteq \Lambda_0$.

From Theorem 24 it follows

$$r_{\text{pr}}(\Theta^p_{0,s:\alpha}\cap I) = \text{const.}\cdot(1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\,\varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}.$$

By taking $K \gtrsim \left((1-\alpha)^{\frac{s}{1-\alpha}}\,\varepsilon\right)^{-\frac{(1-\alpha)(1-\frac{p}{2})+s\,p}{(1-\alpha)+s\,p}\cdot\frac{1-\alpha}{s}}$, we derive from Lemma 22

$$\varepsilon_{\Theta^p_{0,s:\alpha}}(\Lambda,\Lambda_{0,K}) = \mathcal{O}\big(C\,(1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\,\varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}\big).$$

Furthermore for this choice of K and using $\#\Lambda_{0,K} \leq \text{const.} \cdot K$ we obtain

$$\log(\#\Lambda_{0,(K_0,K)}) \leq \text{const.} \cdot \log\big(((1-\alpha)^{\frac{s}{1-\alpha}}\,\varepsilon)^{-1}\big)$$

By Proposition 15 the thresholding estimator then achieves following rate on the finite subspace

$$r_{\text{th}}(\Theta^p_{0,s:\alpha}(\Lambda_{0,K})) = \mathcal{O}(\log\big(((1-\alpha)^{\frac{s}{1-\alpha}}\,\varepsilon)^{-1}\big)\ (1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\ \varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}).$$

This is of the same order as the projection error, so the rate can be transferred to $\Theta^p_{0,s:\alpha}$. We obtain

$$r_{\text{th}}(\Theta^p_{0,s:\alpha}) = \mathcal{O}(\log\big(((1-\alpha)^{\frac{s}{1-\alpha}}\,\varepsilon)^{-1}\big)\ (1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\ \varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}).$$

More precisely for $f \in \Theta^p_{0,s:\alpha}$ we have

$$r(\hat{f}^{\text{j}}_{\mu_{\text{uni}}}, f) \leq \text{const.} \cdot C\ \log\big(((1-\alpha)^{\frac{s}{1-\alpha}}\,\varepsilon)^{-1}\big)\ (1-\alpha)^{-\frac{sp}{(1-\alpha)+sp}}\ \varepsilon^{\frac{(1-\alpha)(2-p)+2s\,p}{(1-\alpha)+s\,p}}.$$

with $C = \max\Big(\|f\|^2_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})}, \|f\|^{\frac{(1-\alpha)\,p}{(1-\alpha)+s\,p}}_{\mathcal{M}^p_{0,s+\alpha(1/p-1/2):\alpha}(\mathbb{R})}\Big)$. □

5. Simulations and Real World Application

This chapter aims to provide a comprehensive understanding of the practical applications and advantages of Gabor and α-modulation frames for audio denoising. We explore the influence of key tuning parameters in Gabor analysis and compare Gabor coefficient thresholding with other state-of-the-art denoising methods such as wavelet thresholding or spatial median filtering. The computational aspects of Gabor analysis have been extensively studied by various authors, in particular in the work of Feichtinger (2019). However, one of the main challenges lies in the computation of dual frame elements. Our focus is primarily on fine-tuning the essential parameters to achieve optimal denoising results, rather than delving into the numerical aspects. To perform our computations, we rely on existing analysis tools such as the LTFAT Toolbox or the website `www.gaborator.com` for computing Gabor frames and their dual frames. Unfortunately, there are no tools available for α-modulation frames, and their computation is very resource-intensive and time-consuming. Therefore, in this chapter we provide illustrations of the potential of α-modulation frames, leaving extensive simulations for future investigations. The chapter is structured as follows. First, we discuss the theoretical aspects of discrete Gabor frames and show how any continuous Gabor system, as defined in Section 2.2.4, can be transformed into a discrete Gabor system. We analyze the effects of key tuning parameters in Gabor analysis, including the grid and the choice of window function. We highlight the importance of achieving an optimal time-frequency representation for sparse signal representation and improved denoising results. We also compare the denoising results obtained using Gabor and wavelet coefficients, as well as the performance of various thresholding algorithms, including Donoho and Johnstone's universal threshold and more advanced approaches. We provide denoising examples using both synthetic and real-world data, and address the problem of musical noise often encountered when applying diagonal estimation methods to audio signals. Finally, we

highlight the potential of α-modulation frames. All of the data presented in this chapter, as well as the complete Python code used to generate the results, are available on Github[1] for replication.

5.1. Discrete Gabor System

For numerical purposes, we consider a continuous and bounded signal $f \in C[0,T]$. This space is a good model for continuous phenomena with finite duration $T \in \mathbb{R}$. Only a sampled version of the signal is observed. Therefore, we further restrict the space to $\mathbb{C}^L$ for signals of length $L \in \mathbb{N}$. The previous theory of Gabor frames was only valid in $L^2(\mathbb{R}^d)$, with window function in $\mathcal{S}(\mathbb{R}^d)$ and therefore needs to be adapted for discrete observations in $\mathbb{C}^L$.

In the finite discrete case, we consider signals of length $L \in \mathbb{N}$. The Hilbert space $\mathcal{H}$ is taken to be the space of periodic sequences $\mathbb{C}^L$. The Gabor analysis for $\mathbb{C}^L$ can be defined in a very similar way to that of $L^2(\mathbb{R})$. Søndergaard (2007) has shown that by sampling and periodization of the window function of a Gabor frame for $L^2(\mathbb{R})$, one obtains a Gabor frame for $\mathbb{C}^L$ with the same frame bounds. A Gabor frame for $L^2([0,T])$ is obtained by periodizing the window function of a Gabor frame for $L^2(\mathbb{R})$. Again, sampling the window function of a Gabor frame for $L^2([0,T])$ yields a Gabor frame for $\mathbb{C}^L$. Further results by (Strohmer, 1998), (Orr, 1993) and (Janssen, 1997) cover the transition from the continuous to the discrete setting.

A Gabor system in $\mathbb{C}^L$ is a family of sequences defined by

$$\mathcal{G}_L(g,a,b) = \{T_{na} M_{mb} g\}_{n=0,\dots,N-1, m=0,\dots,M-1} \tag{5.1}$$

where $a, b \in \mathbb{N}$, $Mb = Na = L$, $g \in \mathbb{C}^L$ is a periodic sequence, and the discrete translation

[1] `https://github.com/ptafo/Statistical-Analysis-of-Audio-Signals-using-Time-Frequency-Analysis`

and modulation operators are defined as

$$T_j(g) = \{g[k+j]\}_{k\in\mathbb{Z}}, \quad M_j(g) = \left\{e^{2\pi ikj/L} g[k]\right\}_{k\in\mathbb{Z}} \tag{5.2}$$

respectively.

Søndergaard (2007) shows that if the sequence $g \in \mathbb{C}^L$ is obtained by periodizing and sampling an admissible window function $\tilde{g} \in \mathcal{S}(\mathbb{R})$ for a Gabor frame in $L^2(\mathbb{R})$ then the Gabor system $\mathcal{G}_L(g, a, b)$ defines a Gabor frame for $\mathbb{C}^L$. More precisely, we have that if $\mathcal{G}(\tilde{g}, \alpha, \beta)$ is a Gabor frame for $L^2(\mathbb{R})$ with canonical dual window γ, then $\mathcal{G}_L(g, a, b)$ is a Gabor frame for $\mathbb{C}^L$ with $\alpha\beta = \frac{a}{M} = \frac{b}{N}$, $Mb = Na = L$, and $a, b, M, N, L \in \mathbb{N}$. In addition, both frames have the same frame bounds, and the canonical dual window is obtained by periodization and sampling, see (Søndergaard, 2007, Theorem 22). This not only proves the existence of Gabor frames in the discrete setting, but the frame bounds remain the same as in the continuous setting.

5.1.1. Characterization of the Discrete Gabor Analysis

Let $g \in \mathbb{C}^L$ be a window sequence. There are three main parameters that characterize a discrete Gabor system $\mathcal{G}_L(g, a, b)$ in $\mathbb{C}^L$:

- a : time step. It defines the number of time bands of equal length or time pixel count, $N = \frac{L}{a}$.
- b : frequency step. It defines the number of frequency bands of equal length or frequency pixel count, $M = \frac{L}{b}$.
- W : support length of the window sequence. The width of the window is given as $w = W/L$.

where $a, b, M, N, L, W \in \mathbb{N}$.

The density of the discrete Gabor system $\mathcal{G}_L(g, a, b)$ is defined analogously to that of the continuous Gabor system $\mathcal{G}(\tilde{g}, \alpha, \beta)$ from which it was derived, with $\alpha\beta = \frac{a}{M}$. The system $\mathcal{G}_L(g, a, b)$ is a frame for $\mathbb{C}^L$ only if the density satisfies

$$\alpha\beta = \frac{a}{M} = \frac{L}{N \cdot M} < 1. \tag{5.3}$$

This results in a discrete time-frequency representation, with a discrete resolution of $M \times N$ pixels. The redundancy of the frame is given by $\frac{N \cdot M}{L}$. The time-frequency resolution is affected by the length of the window sequence.

The choice of window length plays an important role, see Theorem 5. A longer window provides better spatial resolution in the frequency domain, but results in poor spatial resolution in the time domain. On the other hand, a shorter window length is preferable for finer time resolution, but performs poorly in the frequency domain. The overlap rate of the window is given by $\mathrm{OR} = 1 - \frac{L}{N \cdot W}$. The overlap rate increases as the support of the window sequence increases or as more time bands are considered. It intensifies the redundancy among the coefficients. A low density in the frame system as well as a high redundancy of the window both result in a longer computation time.

For our numerical experiments we use the LTFAT toolbox founded by Søndergaard, see Søndergaard et al. (2012) and Průša et al. (2014). It provides an efficient and fast implementation of Søndergaard (2007, Theorem 22) to generate Gabor frames for the sequence space $\mathbb{C}^L$.

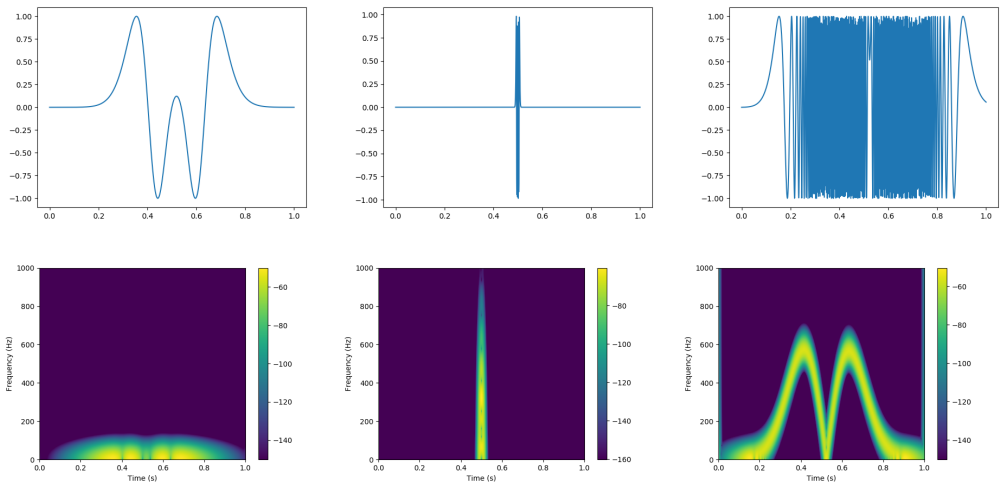
5.2. Parameter Tuning

In this section, we show how effective fine-tuning of the parameters of the discrete Gabor system allows it to adapt to the properties of the underlying signal. This increases the sparsity of the time-frequency representation of the signal and thus improves denoising. We also compare a simple approach using universal thresholding with a similar method in the wavelet case.

We consider the following family of signals

$$f_{A,B}(t) = \sin(2\pi \cdot B \cdot t \cdot e^{-A(t-0.5)^2}) \qquad t \in [0,1], \tag{5.4}$$

with $f_{A,B} \in \mathcal{M}^p_{v_s}(\mathbb{R})$ for $p \in (0,2]$ and $s \geq 0$. This class of functions describes an



(a) $f_{50,2}$
$\|\cdot\|_{L^2}^2 = 0.2202$

(b) $f_{50000,4}$
$\|\cdot\|_{L^2}^2 = 0.00825$

(c) $f_{40,200}$
$\|\cdot\|_{L^2}^2 = 0.4269$

Figure 5.1.: Signals (upper row) and spectrogram (lower row) of our test signals

apparent change of the frequency in the time and is similar to the Doppler functions used by Donoho and Johnstone (1994) and in many practical applications such as astrophysics or medical imaging, see e.g. Marple et al. (1998) and Matani et al. (1996). In particular, we focus on the following three signals, which exhibit three different patterns of variation in the time-frequency domain,

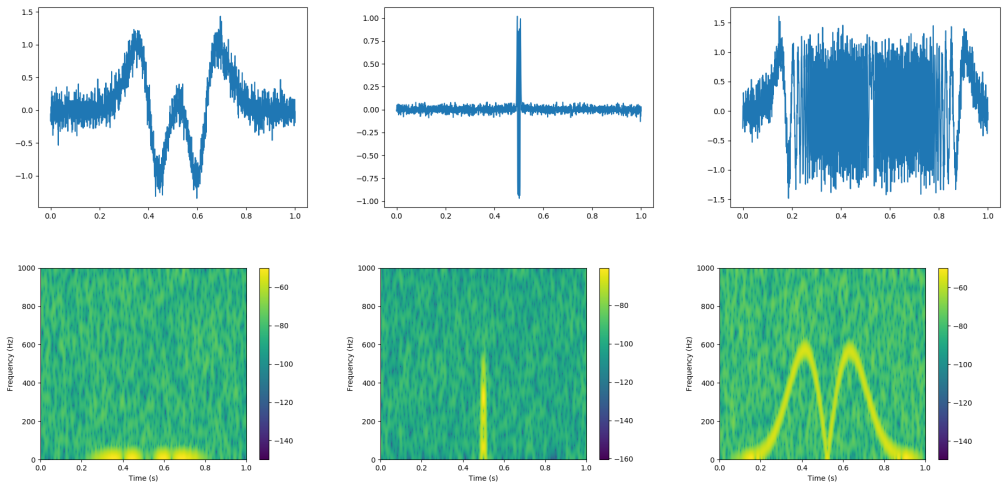
1. $f_{50,2}$, a signal with small frequency variation,

2. $f_{50000,4}$, a signal with small time variation,

3. $f_{40,200}$, a signal that varies equally in both the time and frequency domains,

see Figure 5.1. The lower part of Figure 5.1 displays the spectrograms of the considered signals, the absolute values of the time-frequency representation. It shows the energy of the signal.

The simulations were conducted as follows:



(a) $f_{50,2}$
$\sigma = 0.1484$

(b) $f_{50000,4}$
$\sigma = 0.02873$

(c) $f_{40,200}$
$\sigma = 0.2066$

Figure 5.2.: Noisy signals and spectrograms with SNR = 10 dB.

- For $\sigma > 0$, we generate for each signal discrete $L = 2000$ noisy observations from

$$Y_{A,B}[i] = f_{A,B}[i] + \frac{\sigma}{\sqrt{L}}\,\varepsilon_i, \qquad i = 1, ..., L,$$

with $f_{A,B}[i] = f(i/L)$ and $\varepsilon_i \sim \mathcal{N}(0,1)$. Each signal contains frequencies up to 1000 Hz. A noise level $\sigma > 0$ is chosen to obtain a signal-to-noise ratio of 10 decibels, SNR = 10 dB, i.e., the underlying signal has an energy level 10 times higher than the noise.

$$\text{SNR} = 10 \log_{10} \frac{\sum_i f^2[i]}{\sigma^2}\,\text{dB},$$

see Figure 5.2.

- We decompose each noisy signal $Y_{A,B}$ using a Gabor transform with different parameters as well as using wavelet transform. We use a discrete B-spline of order 4 as window sequence in the Gabor system. B-splines of order $k \geq 1$ are admissible window functions for continuous Gabor expansions and provide norm equivalence on $\mathcal{M}^p_m(\mathbb{R}^d)$ for $p > 1/k$, see (Dahlke et al., 2022, Theorem 1). Additionally, B-splines windows are compact and thus attractive from a numerical and computational point of view. For an adequate comparison, we compute the wavelet

transform using biorthogonal B-spline wavelet of order 4 ('bior4.4'). Simulations have shown that the best results in wavelet denoising are obtained using the maximum decomposition level. In all scenarios, we use $m = 2000$ replicates to compute the mean squared errors.

- We proceed to threshold the noisy frame coefficients using different thresholds (universal or optimal) and thresholding methods (soft or hard). While all coefficients of the Gabor representation are thresholded to reduce noise, only the detail coefficients of the wavelet transform are considered. The wavelet approximation coefficients are preserved.
- We construct an estimate of the signal by synthesizing the thresholded noisy frame coefficients.
- The signal-to-noise ratio in the resulting estimate $\hat{f}$ is used to evaluate the performance.
$$\text{SNR} = 10 \log_{10} \frac{\sum_i f^2[i]}{\sum_i (f[i] - \hat{f}[i])^2} = 10 \log_{10} \frac{\sum_i f^2[i]}{L \times \text{MSE}} \quad \text{dB},$$
- All results are based on 1000 repetitions.

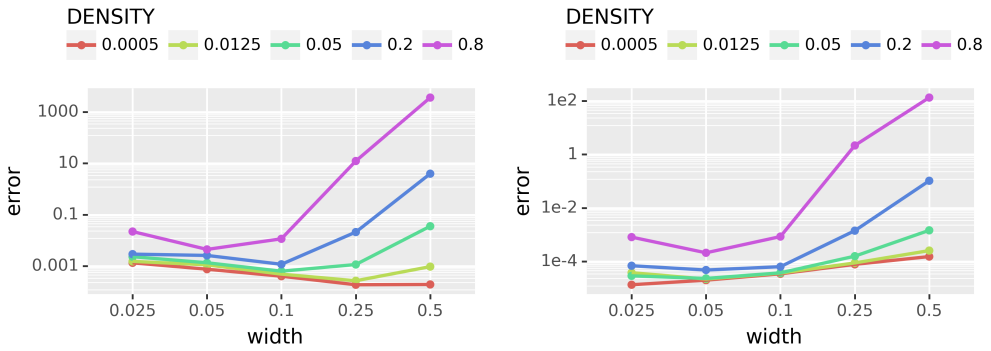
Regular lattice grid

Here we report results for the case $M = N$.

N	2000	400	200	100	50
M	2000	400	200	100	50
density	0.0005	0.0125	0.05	0.2	0.8
redundancy	2000	80	20	5	1.25

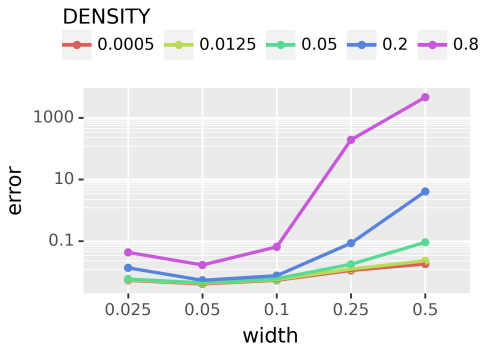
Table 5.1.: Number of time and frequency bands ,the grid density and redundancy in a quadratic time-frequency representation.

Figure 5.3 shows the effect of grid densities and different window lengths according to Table 5.1 and Table 5.2 on the mean squared error (MSE). We use hard thresholding with an MSE-optimal threshold which was computed over a fine grid of thresholds. The following observations can be made:



(a) $f_{50,2}$

(b) $f_{50000,4}$



(c) $f_{40,200}$

Figure 5.3.: Effect of the window width $w = W/L$ and the grid density with $M = N$ on the three signals for samples of size $n = 2000$, using hard thresholding with the optimal threshold.

W	50	100	200	500	1000
width	0.025	0.05	0.1	0.25	0.5

Table 5.2.: Window length and width.

density \ width	0.025	0.05	0.1	0.25	0.5
0.0005	98	99	99.5	99.8	99.9
0.0125	90	95	97.5	99	99.5
0.05	80	90	95	98	99
0.2	60	80	90	96	98
0.8	19.99	60	80	92	96

Table 5.3.: Window's overlap rate in percentage.

- A lower density by itself leads to a lower error, regardless of the window used or its width. This observation is consistent with the results of Lemma 4. Lower density implies more redundancy in the signal representation. However, this increased redundancy comes at the cost of longer computation times. For all three cases, a density of 0.05 or 0.2 leads to errors comparable to the lowest density.

- The error as a function of window width is shown in Figure 5.3. The curve reflects the time-frequency characteristics of each signal.

  For $f_{50,2}$, the signal is concentrated in a few frequency bands. Therefore, a longer window provides better performance because it provides better frequency localization.

  For $f_{50000.4}$, the signal is concentrated in a few time bands. Therefore, a shorter window will give better performance by providing better time localization.

  In contrast, the signal $f_{40,200}$ is almost evenly distributed over the time-frequency plane. Thus, the choice of $w = 0.05$, with equal support in time and frequency, gives the best results in this case.

- The overlap rate increases the redundancy between the coefficients. Therefore, thresholding is improved by a high overlap rate of the window sequence. This is even more obvious for signals contained in multiple time bands, such as $f_{50,2}$ and $f_{40,200}$. Table 5.3 lists the overlap rate for each combination of density and width.

Non-quadratic time-frequency representation

We also analyze the error of an irregular grid, i.e., an unequal number of frequency and time bands, $M \neq N$. Given the characteristics of some functions with signals in a single time or frequency fragment, it may be useful to consider a gain of the time-frequency representation in either the time or frequency plane. Figure 5.4 shows the MSE for four different scenarios according to Table 5.4.

Scenario	S1	S2	S3	S4
N	400	400	100	100
M	400	100	400	100
density	0.0125	0.05	0.05	0.3125

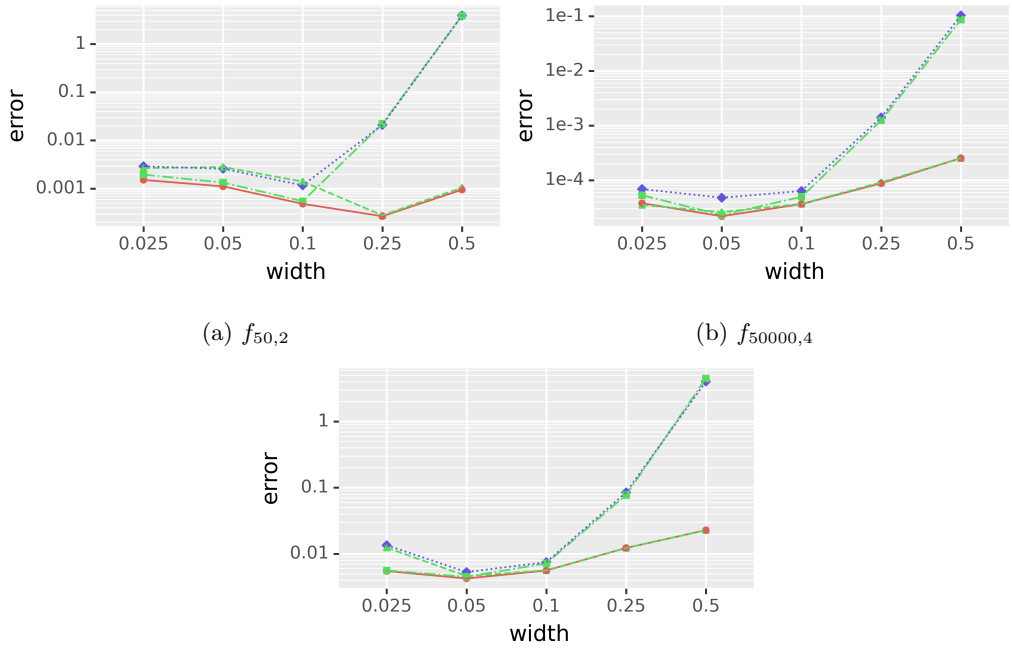
Table 5.4.: Gabor irregular grid density

In all cases, we observe that S2 outperforms S3 for a longer window, while S3 outperforms S2 for a shorter window. As the window size increases, the time domain must be more finely discretized to compensate for the loss of time resolution, while the frequency domain must be more finely discretized as the window size decreases to compensate for the loss of frequency resolution.

For $f_{50.2}$, S3 performs as well as S1, indicating that the loss of local time domain information in S3 does not affect the denoising method because the signal is only present in a small frequency domain. In addition, S2 and S4 have similar performances, indicating that increasing the local information in the time domain does not improve the denoising of $f_{50,2}$.

For $f_{50000,4}$ we observe the opposite results. S2 performs as well as S1 and better than S3, which in turn performs almost as well as S4. This is not surprising since the signal is in a small time fragment. Therefore, increasing the local information in the frequency domain does not lead to better results.

To strike a balance between computation time and MSE performance, we chose $M = N = 100$ with $\alpha \cdot \beta = 0.2$ in the following simulations. We also chose the following values for the window width.



(a) $f_{50,2}$

(b) $f_{50000,4}$

(c) $f_{40,200}$

Figure 5.4.: Effect of window width w and grid density. For the three signals for samples of size $n = 2000$, using hard thresholding with the optimal threshold. —S1, – –S2, –·–S3, · · · S4

signal	$f_{50,2}$	$f_{50000,4}$	$f_{40,200}$
w	0.1	0.025	0.05
overlap rate	90%	60%	80%

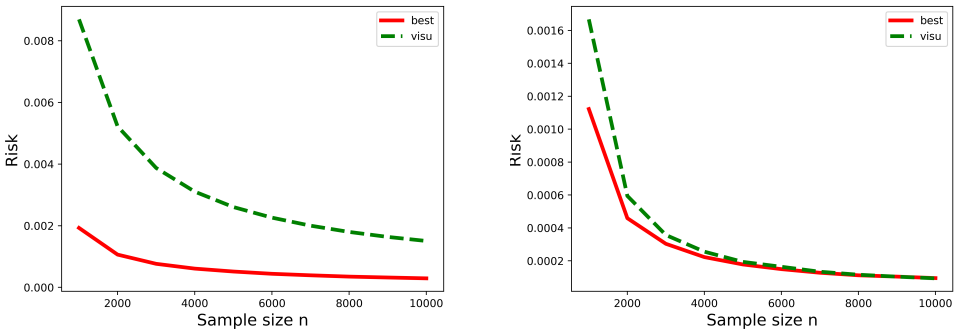
Choice of threshold and thresholding method

Next, in a repeated simulation for various sample sizes, we examine the MSE of soft and hard thresholds, both for the universal threshold of Donoho and Johnstone and for an MSE-optimal threshold computed over a fine grid of thresholds. The results are shown in Figure 5.5. Two important observations can be made from the results. First, hard thresholding outperforms soft thresholding. This may not be surprising, since soft thresholding aims to produce smoother (easier on the eye) images than hard thresholding. Soft thresholding reduces the amplitude of the signal and distorts the discontinuities in the signal, while hard thresholding is better at preserving discontinuities. However, we evaluate the performance of our method based on the reconstructed signal and not on the time-frequency representation. Second, the universal threshold seems to be a very reasonable choice for hard thresholding.

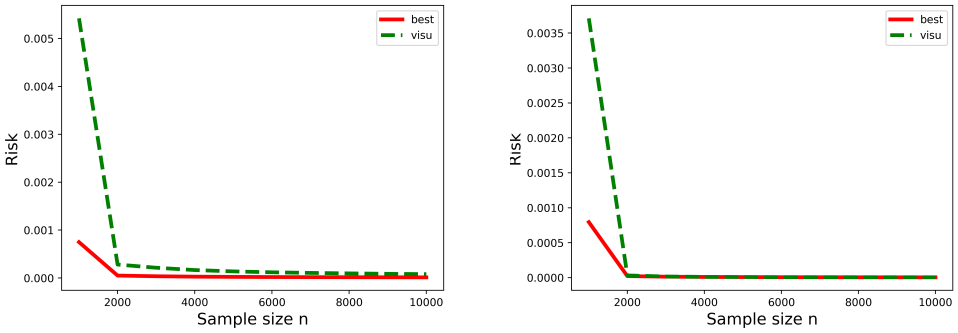
Comparison of Gabor and wavelet methods

Next, we compare the performance of our Gabor thresholding estimators with that of wavelet shrinkage methods that use biorthogonal B-splines of order 4 for signal reconstruction.

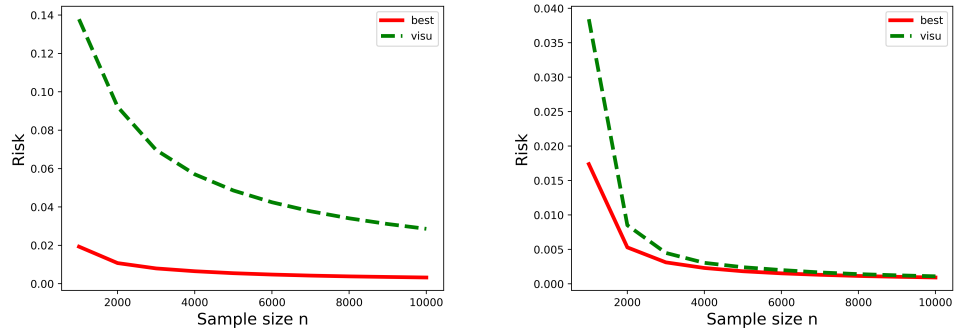
To visually compare the denoising performance, we plot the reconstructions obtained by hard thresholding with the universal threshold in Figure 5.6a. The reconstructed signal of $f_{40,200}$ appears to be of better quality with the Gabor frame-based method compared to the wavelet shrinkage method. On the other hand, for the signal $f_{50000.4}$, which contains a single spike, we expect the wavelet shrinkage method to perform better, although it is not yet apparent from the plots.



(a) $f_{50,2}$

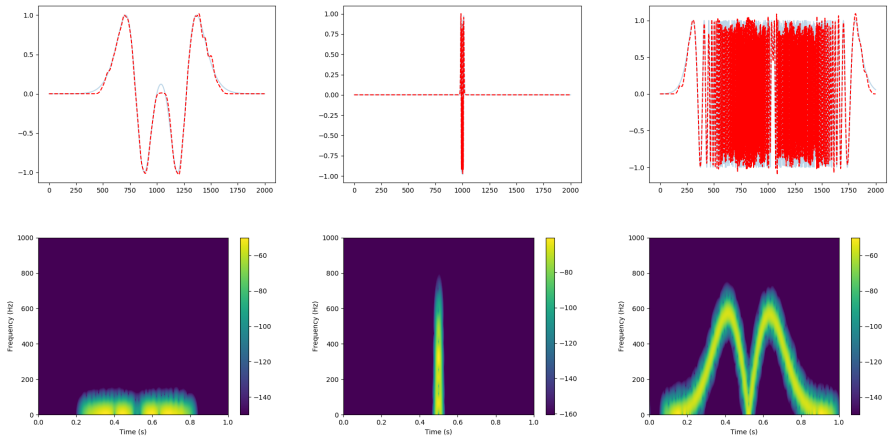


(b) $f_{50000,4}$



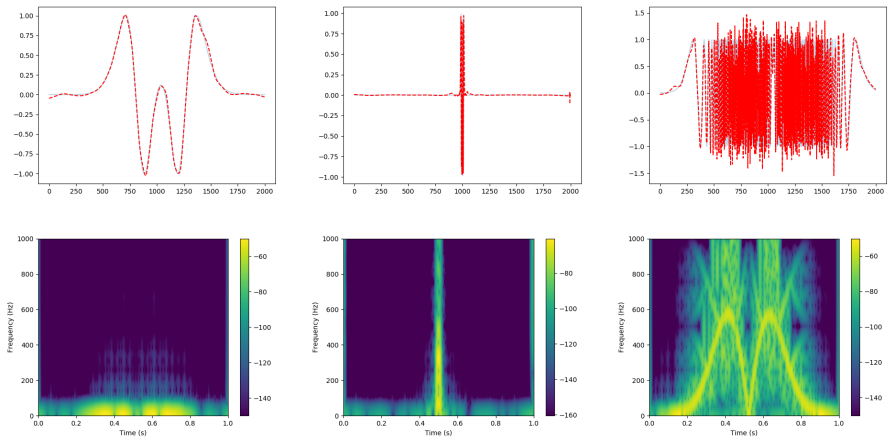
(c) $f_{40,200}$

Figure 5.5.: Left figures: MSE for soft thresholding, optimal threshold versus universal threshold; right figures: MSE for hard thresholding, optimal threshold versus universal threshold. The line style −− denotes the universal threshold. Different sample sizes are examined.



$f_{50,2}$ $f_{50000,4}$ $f_{40,200}$

(a) First row: Gabor frame based method, and the second row has the spectrograms of the Gabor-frame based estimates.



$f_{40,200}$ $f_{50,2}$ $f_{50000,4}$

(b) First row: Wavelet denoising with db4 and maximum level of decomposition and the second row has the spectrograms of the wavelet transform based estimates.

Figure 5.6.: Reconstructions for a given sample of size $n = 2000$ using hard thresholding with universal threshold.

Comparison MSE

We compare the MSE of both methods in a repeated simulation for different sample sizes. The results are shown in Figure 1 and Figure 2. For the $f_{40,200}$ signal, it is clear that our Gabor frame-based method outperforms wavelet shrinkage. This is also true for the $f_{50,2}$ signal, but to a lesser extent, except for small sample sizes. On the other hand, for the $f_{50000,4}$ signal, wavelet shrinkage seems to outperform our method, especially for small and moderate sample sizes.

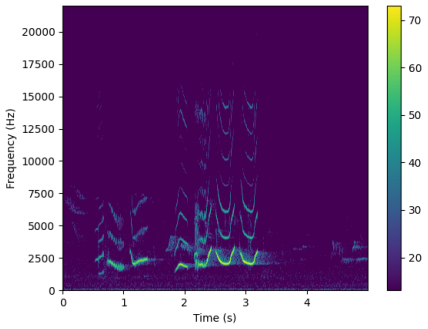
5.3. Denoising on Real Data

In this section, we demonstrate the performance of our method in reducing additive stationary noise on five real-world examples: a recording of a common blackbird, a piano melody, an ECG, a guitar recording, and a sample of human speech[2]. Each recording is 5 seconds long. The ECG has a sampling rate of 2 kHz while the other signals have a sampling rate of 44.1 kHz. Because these signals are non-stationary, their frequencies change over time, and some signals show frequency jumps and changes in vibration amplitude. Spectrograms are shown in Figure 5.7, using a Gaussian window sequence and an appropriate choice of relatively dense grid and window width, resulting in 75% overlap. These spectrograms illustrate the different spectral characteristics of these signals.

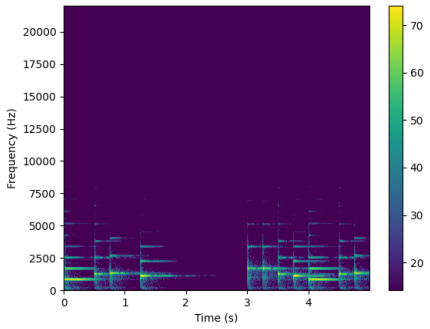
Each signal in the real data examples is contaminated with Gaussian white noise based on a specified signal-to-noise ratio. Current methods for dealing with such signals in audio processing are described in Yu et al. (2008). For comparison, we have also included a wavelet based thresholding approach and a spatial denoising method called the nonlinear median filter.

We applied the discrete wavelet transform using the Daubechies 4 wavelet family, db4, at the highest decomposition level, where only the detail wavelet coefficients were thresholded. Similar results were observed for other closely related wavelet families. For

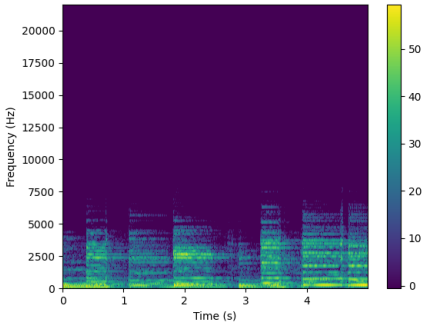
[2]obtained from *https://samplefocus.com/*



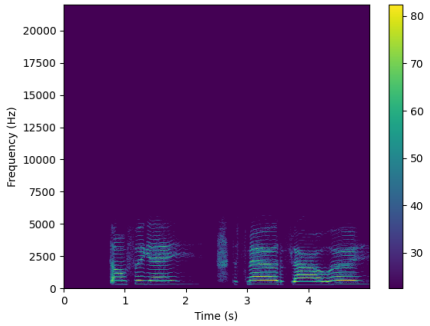
(a) Blackbird



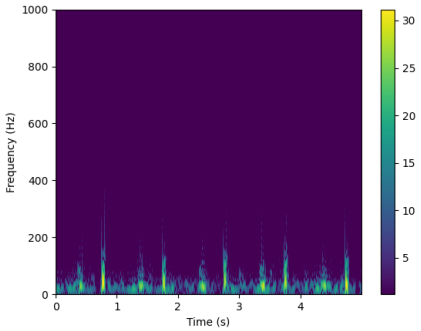
(b) Melody



(c) Guitar



(d) Human speech



(e) ECG

Figure 5.7.: Spectrograms of a recording of a common blackbird, a melody played on a piano, a guitar, a sample of human speech, and an electrocardiogram signal

the Gabor analysis, we computed both the STFT with a rectangular window sequence STFTrect and the STFT with a B-spline of order 4 window $\text{STFT}_{\text{Bspline4}}$. A rectangular window sequence is simply a sequence of ones. It extracts a chunk of the signal without weights. The remaining discrete Gabor parameters were chosen based on prior knowledge of such signals, but they are not optimal. For all signals, we set $a = b = 200$ and $W = 2000$. Better results could be obtained with better parameter tuning. The nonlinear median filter gives the best results with a 5×5 grid.

We ran simulations to evaluate three different thresholding methods: VisuShrink, the universal threshold proposed by Donoho and Johnstone (1994) (which we analyzed theoretically in Section 3.4), and SureShrink, which is determined by minimizing Stein's unbiased risk estimate Donoho and Johnstone (1995). In addition, we tested BayesShrink, an approach based on empirical Bayes introduced by Johnstone and Silverman (2005). While VisuShrink uses a simple global threshold, SureShrink and BayesShrink use coefficient-specific thresholds.

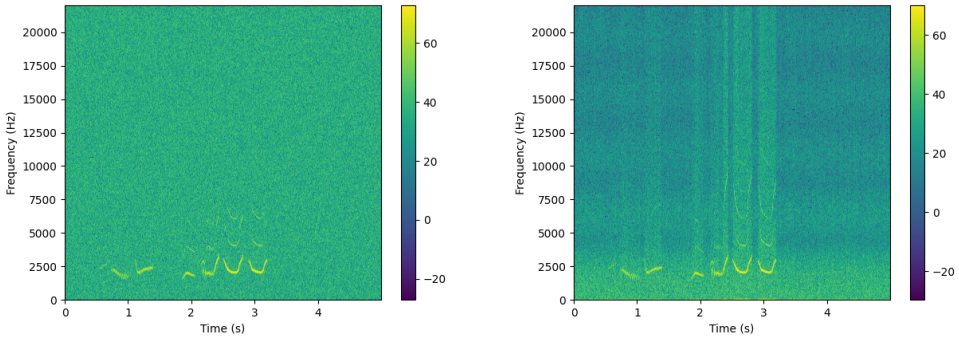
When dealing with acoustic signals, the use of diagonal estimation methods with thresholding can lead to the introduction of artificial musical noise, as pointed out by Yu et al. (2008). To mitigate this problem, block thresholding methods can be used, as suggested by Cai and Silverman (2001). Therefore, we have also included block thresholding versions of the above thresholding schemes.

Figure 5.8 shows the spectrograms of the noisy Blackbird signal for a specific instance of noise, along with some of the reconstruction techniques discussed earlier. It can be observed that all time-frequency based methods produce much clearer spectrograms compared to the wavelet approach and median filtering. However, Visushrink seems to produce an overly clear image.

For a formal comparison of the reconstruction quality of the denoised signal for each method, we considered

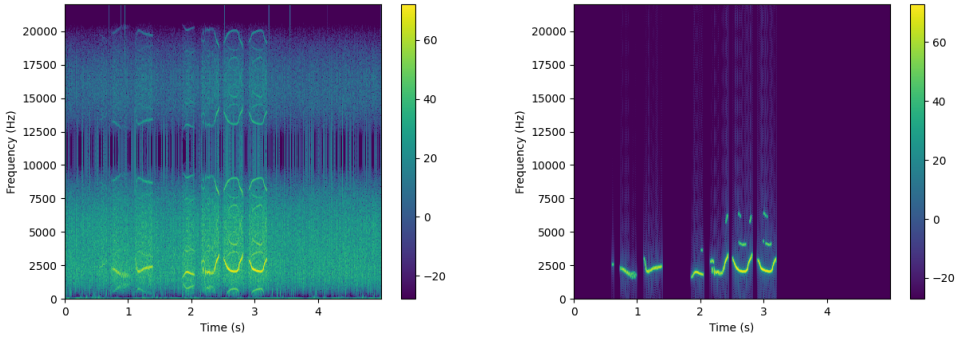
$$\text{SNR} = 10 \log_{10} \frac{\sum_n f^2[n]}{\sum_n (f[n] - \hat{f}[n])^2} \text{dB},$$

which we compared over 10000 iterations.



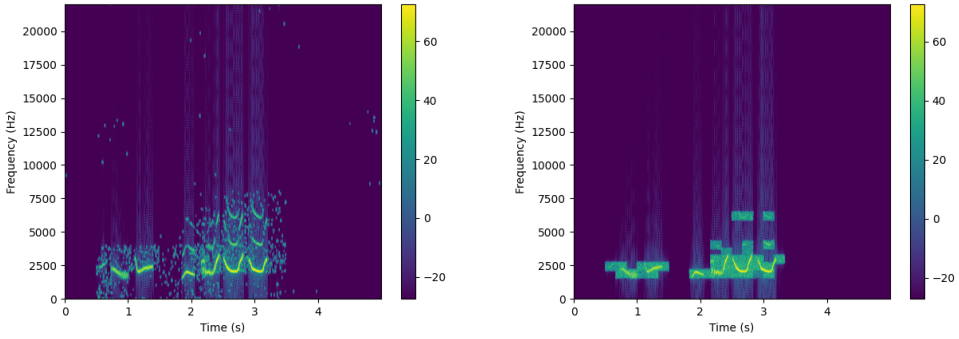
(a) blackbird, noisy recording

(b) Median filtering



(c) Wavelet thresholding

(d) VisuShrink



(e) SureShrink

(f) Block SureShrink

Figure 5.8.: Spectrograms of the noisy and denoised blackbird recordings

The following observations can be made:

- The results shown in Table 1 indicate the superior performance of the time-frequency based methods for VisuShrink. We still observe a better performance of the hard thresholding compared to the soft thresholding. Overall, the B-spline of order 4 window sequence provides better time-frequency localization and thus performs better than just taking a chunk of the signal. Gabor methods perform significantly better than wavelet. For the ECG signal, the two approaches are comparable. The ECG signal consists of singularities, so the wavelet transform performs well as expected. The rectangular window is sufficient to adequately represent such a signal.
- In Table 2, we compare global thresholding using Visushrink against frequency/scale level thresholding using Sureshrink and Bayesshrink. First, Visushrink outperforms the other two methods for the melody and the blackbird signal. The reason for this seems to be the improved frequency sparsity of these two signals. Second, the guitar and human speech signals have more content in each frequency band. Thus, a better performance of the frequency/scale level threshold can be observed.
- Although there is no performance improvement in block thresholding, Table 3 shows that the observations made in Table 1 remain valid.
- Table 4 shows the results using a 5×5 and 1×5 blocks to reduce musical noise. Bayesshrink shows the best results. Gabor outperforms wavelet. For the ECG, both methods are comparable.

5.4. Simulations with α-Modulation Frames

The numerical implementation of α-modulation frames presents a significantly greater challenge than that of Gabor frames. This challenge is primarily due to the irregular grid, which eliminates the benefits of the fast Fourier transform commonly used in discrete Gabor analysis implementations. In addition, there is no explicit method for constructing dual α-modulation frames. However, Fornasier and Feichtinger introduces the Local

Dual Algorithm (LDA), which constructs an explicit dual frame for a band-limited window function g. This is discussed in the work of Fornasier and Feichtinger, in particular in Remark 3. Furthermore, for window functions that are not band-limited, the family of duals can be approximated by the duals of band-limited window functions. To obtain comparable results, we retain the use of the previously used B-spline window function of order 4. To approximate the dual frames, we used the frame algorithm as described in Gröchenig (2013, Algorithm 5.1.1). This is an iterative process to approximate the dual frames. It is important to note that this approximation process is very time and resource consuming, especially for high frequency signals where the grid becomes progressively finer. Therefore, the approximated frames may not be close enough to the actual dual frames.

5.4.1. Characterization of Discrete α-Modulation Frames

Let $g \in \mathbb{C}^L$ be a window sequence, $b > 0$ and p_α and s_α as defined in (2.41).
Using the commutation relation $D_a T_x = T_{ax} D_a$, we obtain

$$\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b) = \left\{ M_{p_\alpha(j)} T_{ak s_\alpha^{-1}(j)} D_{s_\alpha^{-1}(j)}\, h \right\}_{(j,k)\in\Lambda}$$

Furthermore, using the previously defined discrete versions of the modulation and translation operators, we obtain the discrete α-modulation frames, defined as

$$\mathcal{G}_\alpha(h, p_\alpha, s_\alpha, a, b) = \left\{ M_{p_\alpha(j)} T_{ak s_\alpha^{-1}(j)} D_{s_\alpha^{-1}(j)}\, h \right\}_{j=0,\dots,M-1, k=0,\dots,s_\alpha(j)\cdot N}$$

where $a, b > 0$, $M = p_\alpha^{-1}(L)$, and $N = \frac{L}{a}$.

There are three main parameters that characterize a discrete α-modulation system:

- b : frequency factor. It defines the number of frequency bands $M = p_\alpha^{-1}(L)$, which is a decreasing function of α. The frequency bands have exponentially growing size $s_\alpha(j)$.
- a: time factor. It induces the equally spaced discrete time support of each frequency band. Each of them has as time step $s_\alpha^{-1}(j)/a$, which is a decreasing function of α.

- W : Support length of the window sequence.

with $a, b, M, N, L, W \in \mathbb{N}$.

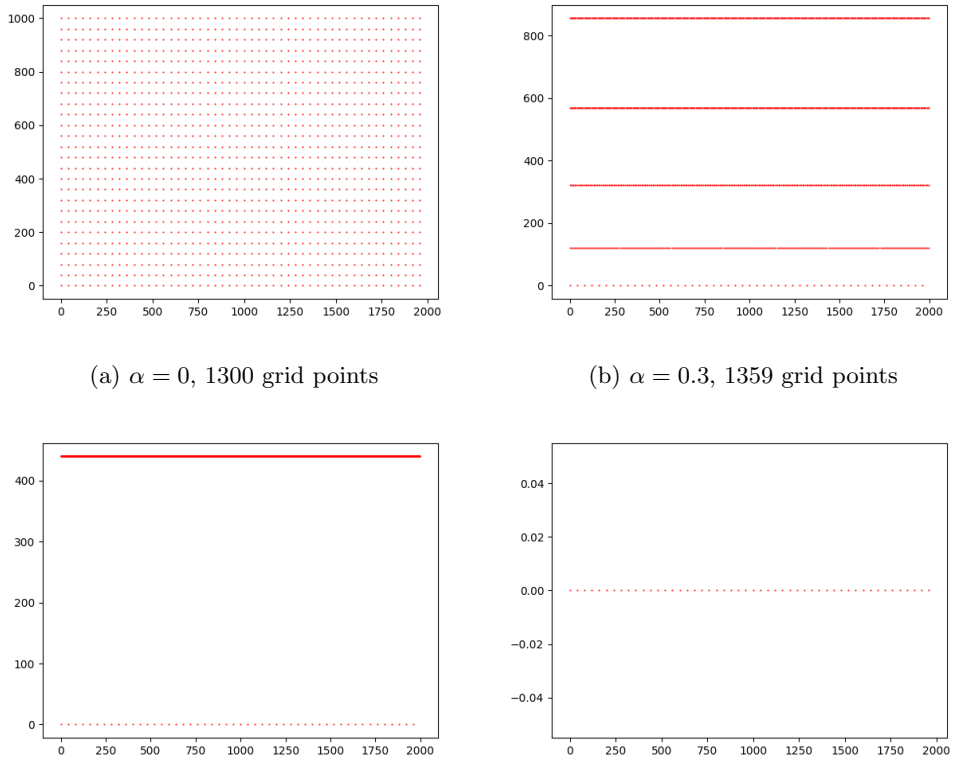
The length of the window sequence is particularly important, especially when $\alpha \to 1$. The dilation factor adjusts the length of the window sequence according to the frequency band, while maintaining a constant overlap rate of the sequence in each frequency band. This results in higher frequency resolution at low frequencies and higher time resolution at high frequencies, as shown in the following equation:

$$\mathrm{OR}_j = 1 - \frac{L}{s_\alpha(j) \cdot \frac{L}{a} \cdot W \cdot s_\alpha^{-1}(j)} = 1 - \frac{a}{W}$$

Figure 5.9 shows the lattice depending on the parameter α. When $\alpha = 0$, the lattice is evenly spaced but has only a few points. As we increase α, the number of points in the lattice grows exponentially until it peaks at about $\alpha = 0.5$, after which it begins to decrease even though fewer frequency bands are involved. Computing the dual frames using the frame algorithm becomes very difficult, especially for cases where $\alpha \approx 0.5$. Despite the difficulty, we are most interested in these cases. However, computing the dual frames using the frame algorithm for such cases requires an enormous amount of computing power. Increasing the time factor a would reduce the number of grid points, but this may violate the conditions of the Theorem 11. In our experiment, we were able to approximate and store dual frames for signals of length $L = 2000$ and few values of α.

We run simulations similar to those in Section 5.2 to fine tune the parameters. The results shown in Figures 5.10, 5.11, and 5.12 show different patterns for each signal. The signal $f_{50,2}$ has an almost constant frequency content, so it is better to analyze it using Gabor frames. So a lower value of $\alpha \approx 0$ is best. $f_{50000,4}$ presents a transient event at time $t = 0.5$ and is therefore better analyzed using wavelets. Therefore, a high value of $\alpha \approx 1$ is best. The signal $f_{40,200}$ contains both time-varying frequencies and a transient event. A value of $\alpha \approx 0.5$ is best.

The choice of the parameter $\alpha \in [0, 1]$ should depend on the singularities present in the signal. A low value of α is appropriate for signals with few or no singularities, while a



(a) $\alpha = 0$, 1300 grid points

(b) $\alpha = 0.3$, 1359 grid points

(c) $\alpha = 0.5$, 2050 grid points

(d) $\alpha = 0.7$, 50 grid points

Figure 5.9.: Lattice grid for α-modulation analysis with $a = 40$, $b = 40$ and $L = 2000$.

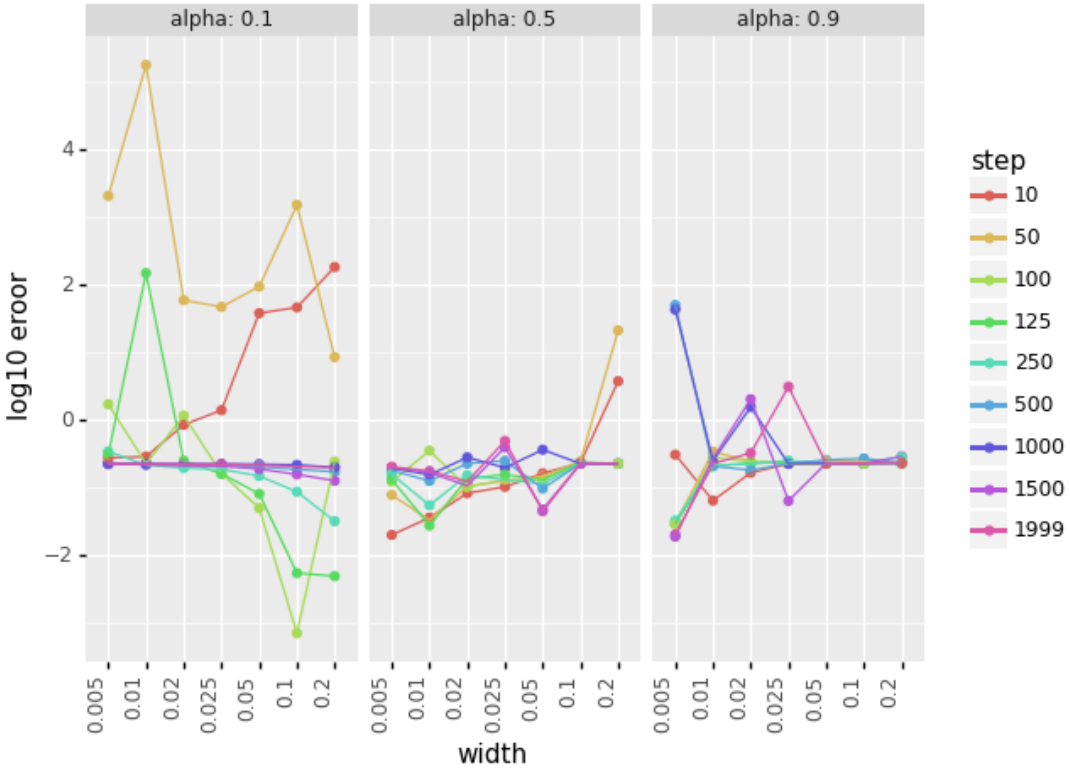


Figure 5.10.: Effect of the parameter $\alpha \in [0, 1)$, of the window width $w = W/L$ and the grid with $a = b =$ step on the signal $f_{50,2}$ for samples of size $n = 2000$, using hard thresholding with the universal threshold.

high value of α is preferable for signals with many singularities. The spectrograms shown in Figure 5.13 support this view. Signal $f_{50,2}$ has the most sparse representation with Gabor frames, $\alpha \approx 0$, while $f_{50000,4}$ is best represented with wavelet frames, $\alpha \approx 1$. On the other hand, $f_{40,200}$ is best represented with $\alpha \approx 0.4$.

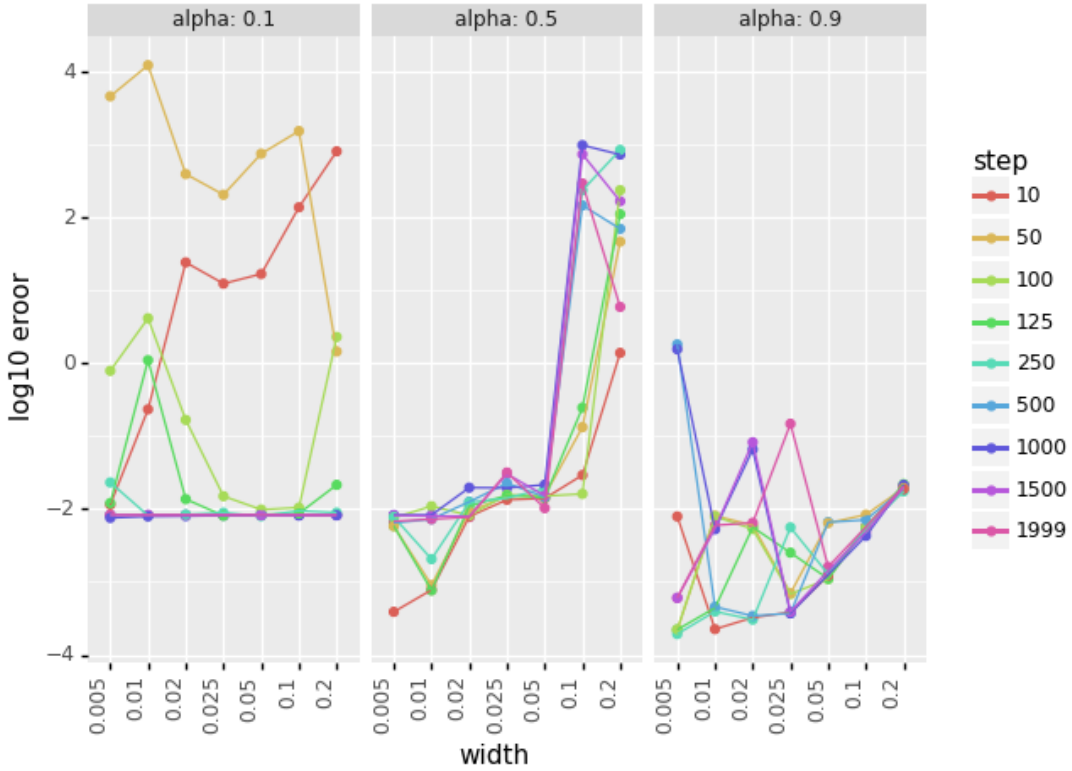


Figure 5.11.: Effect of the parameter $\alpha \in [0, 1)$, of the window width $w = W/L$ and the grid with $a = b =$ step on the signal $f_{50000,4}$ for samples of size $n = 2000$, using hard thresholding with the universal threshold.

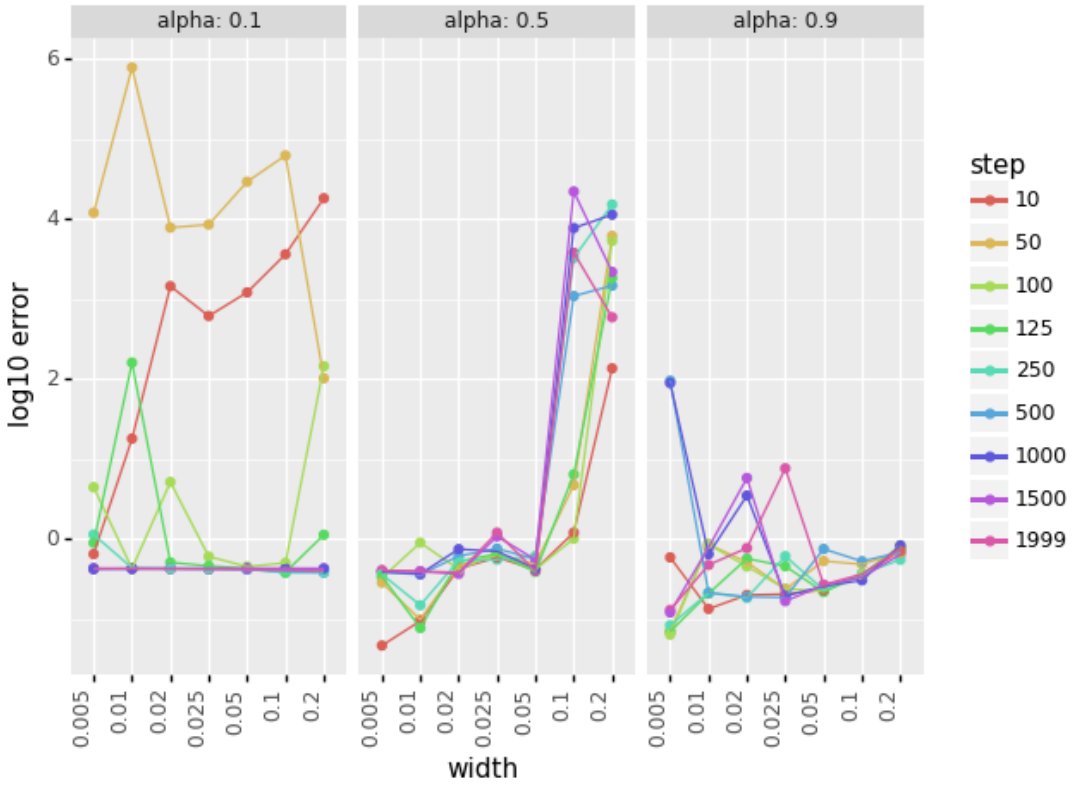
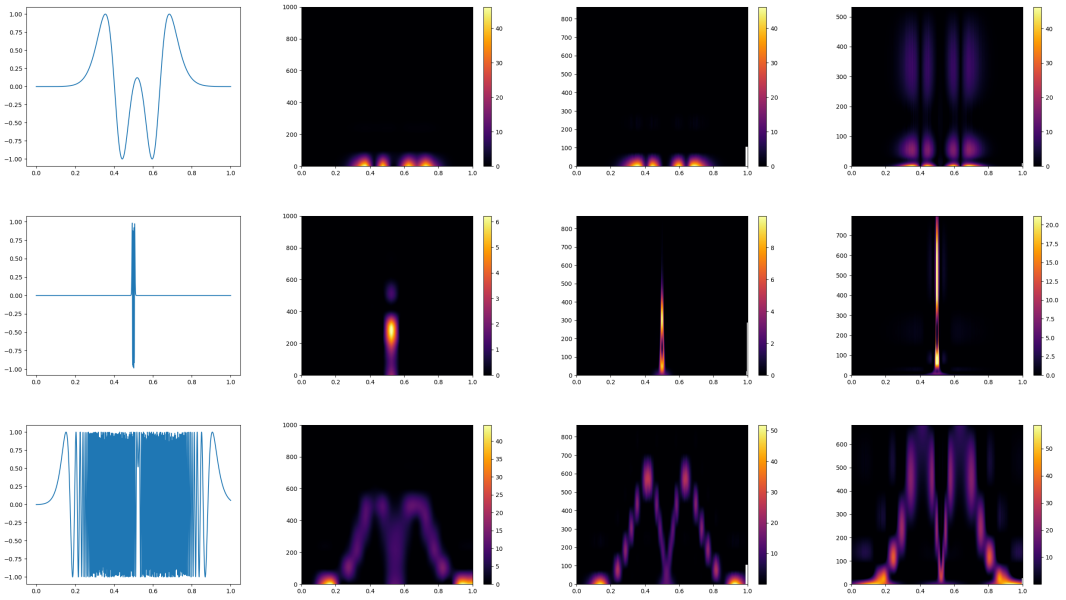


Figure 5.12.: Effect of the parameter $\alpha \in [0, 1)$, of the window width $w = W/L$ and the grid with $a = b =$ step on the signal $f_{40,200}$ for samples of size $n = 2000$, using hard thresholding with the universal threshold.



(a) α =0.0 (b) α =0.4 (c) α =0.8

Figure 5.13.: Spectrograms for different values of α. First row: $f_{50,2}$, second row: $f_{50000,4}$, third row: $f_{40,200}$.

Bibliography

Ashwin, J. and N. Manoharan (2018, 01). Audio denoising based on short time fourier transform. *Indonesian Journal of Electrical Engineering and Computer Science 9*, 89–92.

Banach, S. (1932). Théorie des opérations linéaires.

Belov, D. I. and R. D. Armstrong (2011). Distributions of the kullback–leibler divergence with applications. *British Journal of Mathematical and Statistical Psychology 64*(2), 291–309.

Benedek, A. and R. Panzone (1961). The space l_p with mixed norm. *Duke Mathematical Journal 28*(3), 301–324.

Birman, M. S. and M. Z. Solomyak (1966). Approximation of functions by piecewise polynomial functions. *Dokl. Akad. Nauk SSSR 171*, 1015–1018.

Bros, J. and D. Iagolnitzer (1975). Tuboides et structure analytique des distributions. II. Support essentiel et structure analytique des distributions. Sémin. Goulaouic-Lions-Schwartz 1974-1975, Équat. dériv. part. lin. non- lin., Exposé XVIII, 33 p. (1975).

Brown, L. D. and M. G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist 24*(6), 2384–2398.

Cai, T. T. and B. W. Silverman (2001). Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā: The Indian Journal of Statistics, Series B*, 127–148.

Candes, E. J. (2006). Modern statistical estimation via oracle inequalities. *Acta numerica 15*, 257–325.

Christensen, O. (2016). *An Introduction to Frames and Riesz Bases* (Second edition ed.). Applied and Numerical Harmonic Analysis. [Cham]: Birkhäuser.

Dahlke, S., M. Fornasier, H. Rauhut, G. Steidl, and G. Teschke (2008). Generalized coorbit theory, Banach frames, and the relation to α-modulation spaces. *Proceedings of the London Mathematical Society. Third Series 96*(2), 464–506.

Dahlke, S., S. Heuer, H. Holzmann, and P. Tafo (2022). Statistically optimal estimation of signals in modulation spaces using gabor frames. *IEEE Transactions on Information Theory 68*(6), 4182–4200.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory 36*(5), 961–1005.

Donoho, D. L. and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association 90*(432), 1200–1224.

Donoho, D. L. and I. M. Johnstone (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics 26*(3), 879 – 921.

Donoho, D. L. and J. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika 81*(3), 425–455.

Donoho, D. L., R. C. Liu, and B. MacGibbon (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist. 18*, 1416–1437.

Duffin, R. J. and A. C. Schaeffer (1952). A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society 72*, 341–366.

Feichtinger, H. (1983, 01). Modulation spaces on locally compact abelian group.

Feichtinger, H. G. (1987). Banach spaces of distributions defined by decomposition methods. II. *Mathematische Nachrichten 132*, 207–237.

Feichtinger, H. G. (2019). Gabor expansions of signals: computational aspects and open questions. In *Landscapes of Time-Frequency Analysis*, pp. 173–206. Springer.

Feichtinger, H. G. and K. Gröchenig (1989). Banach spaces related to integrable group representations and their atomic decompositions, i. *Journal of Functional Analysis 86*(2), 307–340.

Feichtinger, H. G. and P. Gröbner (1985). Banach spaces of distributions defined by decomposition methods. I. *Mathematische Nachrichten 123*, 97–120.

Folland, G. B. (1989). *Harmonic analysis in phase space*. Annals of Mathematics Studies. Princeton, NJ: Princeton University Press.

Folland, G. B. and A. Sitaram (1997). The uncertainty principle: a mathematical survey. *Journal of Fourier Analysis and Applications 3*(3), 207–238.

Fornasier, M. (2004). Banach frames for alpha-modulation spaces.

Fornasier, M. and H. G. Feichtinger (2006). Flexible Gabor-wavelet atomic decompositions for L^2 -Sobolev spaces. *Annali di Matematica Pura ed Applicata. Serie Quarta 185*(1), 105–131.

Fornasier, M. and K. Gröchenig (2005). Intrinsic localization of frames. *Constructive Approximation 22*(3), 395–415.

Gabor, D. (1946). Theory of communications. *J. lEE (London) 93*(3), 429–457.

Galperin, Y. V. and S. Samarah (2004). Time-frequency analysis on modulation spaces $m^{p,q}_m$, $0 < p, q \leq \infty$. *Applied and Computational Harmonic Analysis 16*(1), 1–18.

Goldenshluger, A., A. Tsybakov, A. Zeevi, et al. (2006). Optimal change-point estimation from indirect observations. *The Annals of Statistics 34*(1), 350–372.

Gröchenig, K. (2004). Localization of frames, banach frames, and the invertibility of the frame operator. *Journal of Fourier Analysis and Applications 10*.

Gröchenig, K. (2013). *Foundations of time-frequency analysis*. Springer Science & Business Media.

Gröchenig, K. H. (1991). Describing functions: frames versus atomic decompositions. *Monatshefte für Mathematik 112*(1), 1–41.

Grossmann, A., J. Morlet, and T. Paul (1985). Transforms associated to square integrable group representations. I: General results. *J. Math. Phys. 26*, 2473–2479.

Ibragimov, I. A. and R. Z. Has'minskiĭ (1981). *Statistical estimation, asymptotic theory*, Volume 16 of *Applications of Mathematics*. Springer-Verlag.

Janssen, A. J. (1997). From continuous to discrete weyl-heisenberg frames through sampling. *Journal of Fourier Analysis and Applications 3*(5), 583–596.

Johnstone, I. M. (2002). Function estimation and gaussian sequence model. *Unpublished manuscript.*.

Johnstone, I. M. (2013). Gaussian estimation: Sequence and wavelet models. *Unpublished manuscript*.

Johnstone, I. M. and D. Paul (2013, October). Adaptation in a class of linear inverse problems.

Johnstone, I. M. and B. W. Silverman (2005). Empirical bayes selection of wavelet thresholds. *The Annals of Statistics 33*(4), 1700–1752.

Kerkyacharian, G., O. Lepski, and D. Picard (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probability Theory and Related Fields 121*(2), 137–170.

Kerkyacharian, G., O. Lepski, and D. Picard (2008). Nonlinear estimation in anisotropic multi-index denoising. Sparse case. *Theory of Probability & Its Applications 52*(1), 58–77.

Lieb, E. H. (1990). Integral bounds for radar ambiguity functions and wigner distributions. *Journal of Mathematical Physics 31*(3), 594–599.

Ma, J. and P. C. Loizou (2011). Snr loss: A new objective measure for predicting speech intelligibility of noise-suppressed speech. *Speech Communication 53*(3), 340–354.

Mallat, S. (2009). *A wavelet tour of signal processing. The sparse way.* (3rd ed. ed.). Amsterdam: Elsevier/Academic Press.

Marple, S. L. j., T. Brotherton, and D. Jones (1998). The application of advanced time-frequency analysis techniques to Doppler ultrasound. In *Time-frequency and wavelets in biomedical signal processing*, pp. 73–100. New York, NY: IEEE.

Matani, A., O. Oshiro, and K. Chihara (1996). Doppler signal processing of blood flow using a wavelet transform. *Japanese Journal of Applied Physics 35*, 3131.

Megginson, R. E. (1998). *An introduction to Banach space theory*, Volume 183 of *Grad. Texts Math.* New York, NY: Springer.

Nazaret, B. and M. Holschneider (2003). An interpolation family between Gabor and wavelet transformations: Application to differential calculus and construction of anisotropic Banach spaces. In *Nonlinear hyperbolic equations, spectral theory, and*

wavelet transformations. A volume of Advances in Partial Differential Equations, pp. 363–394. Basel: Birkhäuser.

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and gaussian white noise. *Annals of Statistics 24*(6), 2399–2430.

Orr, R. S. (1993). Derivation of the finite discrete gabor transform by periodization and sampling. *Signal Processing 34*(1), 85–97.

Pinkus, A. (1985). *n-Widths in Approximation Theory*. Berlin: Springer-Verlag.

Pinsker, M. S. (1980). Optimal filtration of square integrable signals in gaussian noise. *Problems of Information Transmission 16*, 120–133.

Průša, Z., P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs (2014). The large time-frequency analysis toolbox 2.0. In M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad (Eds.), *Sound, Music, and Motion*, Lecture Notes in Computer Science, pp. 419–442. Springer.

Søndergaard, P. L. (2007). Gabor frames by sampling and periodization. *Advances in Computational Mathematics 27*(4), 355–373.

Søndergaard, P. L., B. Torrésani, and P. Balazs (2012). The linear time-frequency analysis toolbox. *International Journal of Wavelets, Multiresolution Analysis and Information Processing 10*(4).

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pp. 197–206. University of California Press, Berkeley-Los Angeles, Calif.

Strohmer, T. (1998). Numerical algorithms for discrete gabor expansions. In *Gabor Analysis and Algorithms: Theory and Applications*, pp. 267–294. Birkhäuser Boston.

Triebel, H. ([1992] ©1992). *Theory of function spaces II*, Volume 107 of *Monographs in Mathematics*. Birkhäuser/Springer, Cham.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Ser. Stat. New York, NY: Springer.

Yu, G., S. Mallat, and E. Bacry (2008). Audio denoising by time-frequency block thresholding. *IEEE Transactions on Signal processing 56*(5), 1830–1839.

Appendix

A. Definitions

Definition 15 (α-covering, Feichtinger and Gröbner (1985, Definition 2.1))**.** A countable set $\mathcal{I}$ of intervals $I \subset \mathbb{R}$ is called an admissible covering of $\mathbb{R}$ if

(a) $\mathbb{R} = \bigcup_{I \in \mathcal{I}}$, and

(b) $\#\{I \in \mathcal{I} : x \in I\} \leq 2$ for all $x \in \mathbb{R}$.

Furthermore, if there exists a constant $0 \leq \alpha \leq 1$ such that $|I| \equiv (1 + |\xi|)^\alpha$ for all $I \in \mathcal{I}_\alpha$ and all $\xi \in I$, then $\mathcal{I}_\alpha$ is called an α-covering.

Definition 16 (BAPU, Feichtinger and Gröbner (1985, Definition 2.2))**.** Given any Banach algebra $(B, \| \|_B)$ of bounded, complex-valued functions on a space X, and an admissible covering $\mathcal{I}$ of X, a family $\Psi = (\psi_I)_{I \in \mathcal{I}}$ is called a bounded admissible partition of unity in B, if the following is satisfied:

1. $\sup_{I \in \mathcal{I}} \|\phi_i\|_B < \infty$;
2. $\operatorname{supp} \psi_I \subset$ for all $I \in \mathcal{I}$;
3. $\sum_{I \in \mathcal{I}} \psi_I(\xi) = 1$ for all $\xi \in X$.

B. Tables and Figures

audio	SNR	soft thresholding			hard thresholding		
		STFT_{rect}	$\text{STFT}_{Bspline4}$	wavelet_{db4}	STFT_{rect}	$\text{STFT}_{Bspline4}$	wavelet_{db4}
blackbird	-10	1.769	3.663	0.072	3.864	**8.695**	0.272
	1	7.716	10.452	3.459	11.966	**15.079**	6.456
	10	13.334	15.938	9.023	17.336	**20.317**	13.485
	30	26.166	27.203	23.160	30.222	**30.981**	27.848
melody	-10	2.337	3.967	0.271	5.363	**7.662**	0.801
	1	8.008	10.063	4.219	12.515	**14.520**	7.141
	10	13.389	15.488	9.516	18.089	**20.183**	13.878
	30	27.770	28.721	24.366	33.582	**35.186**	30.633
ECG	-10	0.090	0.340	0.598	0.691	**1.798**	1.583
	1	3.659	4.080	5.013	**8.107**	7.474	7.878
	10	9.228	9.363	10.250	**16.116**	15.984	14.445
	30	23.421	23.318	23.395	**28.572**	27.365	26.932
guitar	-10	0.531	0.971	0.538	1.935	**2.421**	1.279
	1	4.585	5.110	3.536	7.926	**8.961**	5.343
	10	9.601	10.393	7.755	14.298	**15.605**	11.227
	30	24.905	25.043	22.081	31.472	**32.869**	28.314
human speech	-10	1.772	1.998	0.248	3.617	**4.926**	0.656
	1	6.498	8.102	4.150	10.421	**13.033**	7.500
	10	12.133	13.968	9.721	17.027	**19.340**	14.124
	30	27.174	27.623	24.600	31.989	**34.132**	30.523

Table 1.: Comparison between thresholding of Gabor coefficients and wavelet coefficients using the VisuShrink universal threshold. Rectangular and B-splines of order 4 window for Gabor and db4 wavelet with maximal level of decomposition for wavelet. Soft- and hard thresholding. All values are rounded up to the third decimal place and in decibel ($10 \log_{10} x$, dB).

audio	SNR	Visushrink		Sureshrink		Bayesshrink		median filter
		$\text{STFT}_{bspline4}$	wavelet_{db4}	$\text{STFT}_{bspline4}$	wavelet_{db4}	$\text{STFT}_{bspline4}$	wavelet_{db4}	
blackbird	-10	**3.663**	0.072	1.009	-0.031	0.000	-0.025	-2.772
	1	**10.452**	3.459	1.977	-0.001	2.283	-0.002	5.012
	10	**15.938**	9.023	3.797	-0.000	10.960	-0.000	7.705
	30	**27.203**	23.160	10.770	0.000	25.669	5.729	8.539
melody	-10	**3.967**	0.271	0.541	-0.042	0.000	-0.021	-2.276
	1	**10.063**	4.219	1.089	-0.002	1.805	-0.002	7.868
	10	**15.488**	9.516	1.652	-0.000	8.056	-0.000	13.88
	30	**28.721**	24.366	3.383	0.000	27.220	15.151	17.936
ECG	-10	0.340	0.598	0.000	**1.057**	0.000	-0.114	-2.276
	1	4.080	5.013	0.000	1.808	1.655	1.493	**8.218**
	10	9.363	10.250	0.234	2.625	9.351	6.824	**14.724**
	30	23.318	23.395	2.793	7.484	**25.993**	16.912	18.795
guitar	-10	**0.971**	0.538	0.021	-0.029	0.000	-0.047	-2.337
	1	5.110	3.536	0.349	-0.025	2.798	-0.022	**7.391**
	10	10.393	7.755	1.793	-0.010	8.252	1.147	**12.434**
	30	25.043	22.081	5.240	0.076	**27.645**	11.568	15.082
human speech	-10	**1.998**	0.248	1.161	-0.008	0.000	-0.018	-2.285
	1	**8.102**	4.150	3.173	-0.002	3.389	-0.002	7.994
	10	13.968	9.721	4.577	0.000	9.347	-0.000	**14.19**
	30	27.623	24.600	10.682	0.003	**28.916**	9.123	18.215

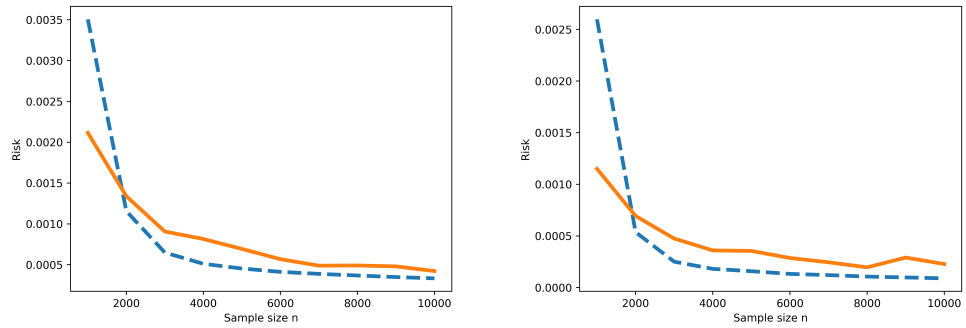
Table 2.: Comparison between soft thresholding of B-spline of order 4 Gabor coefficients and db4 wavelet coefficients using different methods: Visushrink, Sureshrink, Bayesshrink and Median filter. All values are rounded up to the third decimal place and in decibel ($10 \log_{10} x$, dB).

audio	SNR	soft thresholding			hard thresholding		
		STFT_{rect}	$\text{STFT}_{bspline4}$	wavelet_{db4}	STFT_{rect}	$\text{STFT}_{bspline4}$	wavelet_{db4}
blackbird	-10	0.000	0.002	-0.031	0.000	**0.062**	-0.025
	1	1.279	2.320	0.202	2.747	**4.781**	0.378
	10	4.129	6.735	1.056	6.220	**10.355**	1.876
	30	14.349	17.749	6.613	18.056	**22.492**	8.441
melody	-10	0.000	0.000	-0.035	0.000	0.000	-0.032
	1	0.994	2.614	0.164	3.180	**5.935**	0.275
	10	5.074	6.897	0.403	8.427	**10.085**	0.533
	30	15.820	18.491	2.762	18.836	**23.250**	4.182
ECG	-10	0.000	0.000	-0.122	**0.013**	0.000	-0.088
	1	1.146	1.364	0.564	2.354	**2.470**	1.399
	10	2.733	2.857	2.702	3.853	3.658	**4.413**
	30	11.385	11.579	7.627	16.760	**17.375**	8.669
guitar	-10	0.000	0.000	-0.103	0.000	**0.000**	-0.127
	1	0.045	0.549	0.209	0.279	**1.514**	0.672
	10	2.032	3.142	1.717	4.137	**5.587**	2.776
	30	12.160	13.888	8.009	16.891	**18.840**	10.228
human speech	-10	0.000	0.000	-0.023	0.000	**0.000**	-0.025
	1	0.764	0.926	0.468	2.013	**2.217**	0.803
	10	3.597	4.510	1.708	6.375	**8.371**	2.737
	30	14.426	16.686	8.313	19.161	**21.054**	10.842

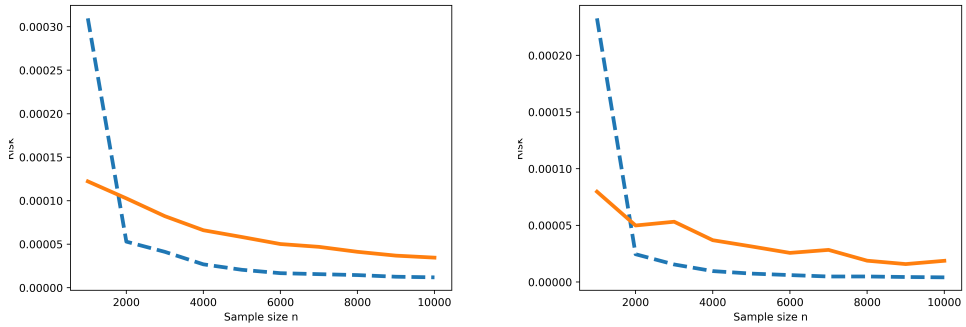
Table 3.: Comparison between block thresholding of Gabor coefficients and wavelet coefficients using the universal threshold VisuShrink. Rectangular and B-splines of order 4 window for the STFT and db4 wavelet coefficients with maximal level of decomposition. 5×5 Blocks for Gabor and 1×5 blocks for wavelet. Soft- and hard thresholding. All values are rounded up to the third decimal place and in decibel ($10 \log_{10} x$, dB).

audio	SNR	Visushrink		Sureshrink		Bayesshrink		Median filter
		$\text{STFT}_{bspline4}$	wavelet_{db4}	$\text{STFT}_{bspline4}$	wavelet_{db4}	$\text{STFT}_{bspline4}$	wavelet_{db4}	
blackbird	-10	0.002	-0.031	-5.332	-7.810	**4.656**	0.438	-2.772
	1	2.320	0.202	9.163	3.452	**13.110**	4.509	5.012
	10	6.735	1.056	18.194	9.267	**18.732**	10.083	7.705
	30	17.749	6.613	24.936	13.784	**32.113**	24.883	8.539
melody	-10	0.000	-0.035	-5.773	-7.980	**5.849**	0.412	-2.276
	1	2.614	0.164	8.550	2.554	**14.037**	3.226	7.868
	10	6.897	0.403	18.694	6.495	**19.932**	8.541	13.88
	30	18.491	2.762	27.083	11.348	**36.439**	25.051	17.936
ECG	-10	0.000	-0.122	1.010	-7.095	**1.829**	1.211	-2.276
	1	1.364	0.564	4.438	4.505	**8.923**	6.767	8.218
	10	2.857	2.702	6.329	9.475	**17.410**	10.978	14.724
	30	11.579	7.627	12.246	12.978	**28.310**	26.259	18.795
guitar	-10	0.000	-0.103	-2.904	-6.863	**2.580**	1.687	-2.337
	1	0.549	0.209	8.783	4.434	**9.239**	5.869	7.391
	10	3.142	1.717	12.635	9.750	**16.954**	11.091	12.434
	30	13.888	8.009	19.900	15.106	**35.335**	26.708	15.082
human speech	-10	0.000	-0.023	-7.126	-8.469	**3.632**	1.269	-2.285
	1	0.926	0.468	6.121	3.065	**11.158**	6.495	7.994
	10	4.510	1.708	16.515	11.045	**18.694**	12.303	14.19
	30	16.686	8.313	26.752	18.745	**36.383**	27.934	18.215

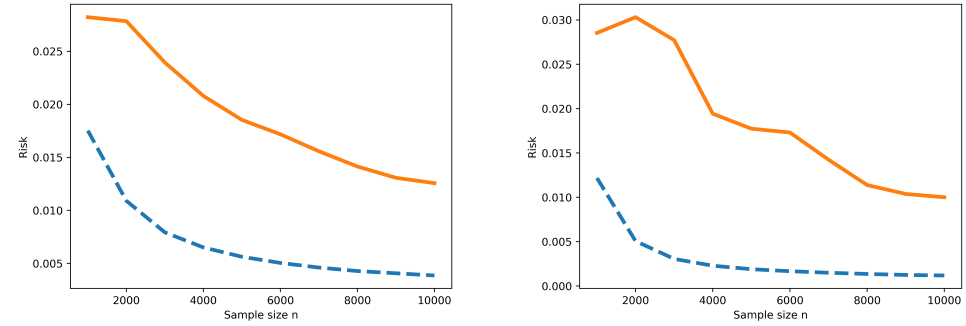
Table 4.: Comparison between block soft thresholding of B-splines of order 4 Gabor coefficients and db4 wavelet coefficients using different methods: Visushrink, Sureshrink, Bayesshrink and Median filter. 5×5 Blocks for Gabor and 1×5 blocks for wavelet. All values are rounded up to the third decimal place and in decibel ($10 \log_{10} x$, dB).



(a) $f_{50,2}$

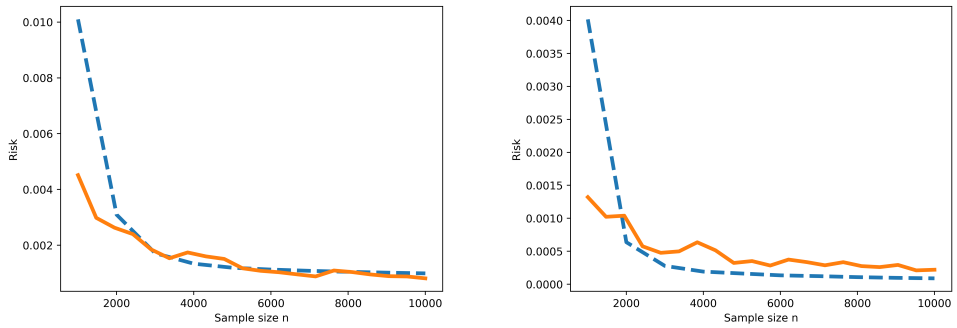


(b) $f_{50000,4}$

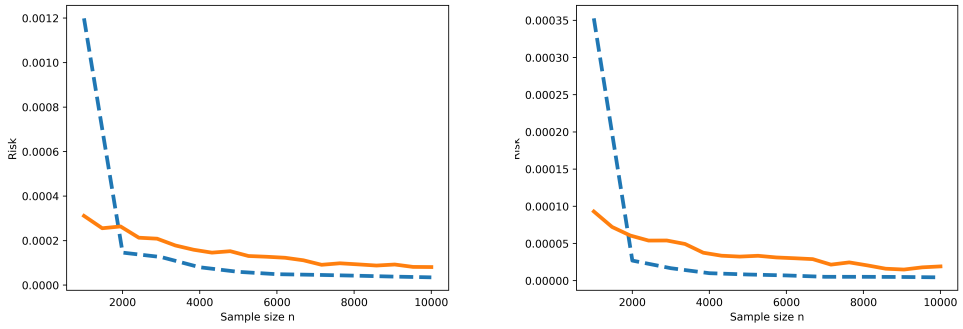


(c) $f_{40,200}$

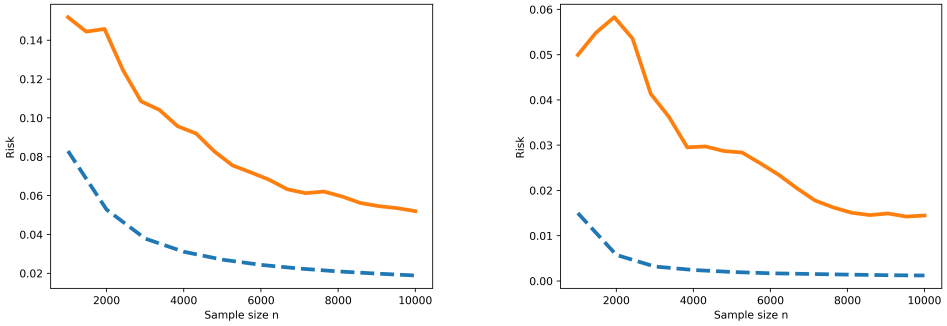
Figure 1.: Left figures: MSE of the competing methods using optimal soft thresholding; right figures: MSE of the competing methods using optimal hard thresholding. Line style −−: Gabor frame based method, —— wavelet shrinkage. Different sample sizes are investigated.



(a) $f_{50,2}$



(b) $f_{50000,4}$



(c) $f_{40,200}$

Figure 2.: Left figures: MSE of the competing methods using universal soft thresholding; right figures: MSE of the competing methods using universal hard thresholding. Line style −−: Gabor frame based method, —— wavelet shrinkage. Different sample sizes are investigated.

Zusammenfassung(deutsch)

Die Zeit-Frequenz-Analyse ist ein häufig verwendetes Instrument zur Untersuchung des zeitlich veränderlichen spektralen Inhalts nicht-stationärer Signale. Im Gegensatz zu herkömmlichen Frequenzanalysemethoden wie der Fourier-Analyse, die zwar Frequenzkomponenten effektiv darstellen, aber die dynamische Natur nichtstationärer Signale nicht erfassen können, bietet die Zeit-Frequenz-Analyse eine Form der lokalen Fourier-Analyse, die gleichzeitig das zeitliche und spektrale Verhalten eines Signals beschreibt. Sie stützt sich auf zwei wichtige Techniken, Translation und Modulation, um Signalkomponenten im Zeit-Frequenz-Bereich zu lokalisieren und eine Zeit-Frequenz-Darstellung abzuleiten. Gröchenig (2013) bietet eine umfassende Erkundung der mathematischen Aspekte der Zeit-Frequenz-Analyse.

Die Signalentrauschung ist eine wichtige Anwendung der Zeit-Frequenz-Analyse. Ein effektiver Ansatz zur Signalentrauschung ist die Nutzung der spektralen Informationen, die die Zeit-Frequenz-Analyse liefert. Rauschen und das zugrunde liegende Signal haben oft unterschiedliche spektrale Eigenschaften. Signale haben in der Regel einen konzentrierteren und lokalisierten Zeit-Frequenz-Inhalt, während Rauschen dazu neigt, sich zufällig über den Zeit-Frequenz-Bereich oder einen bestimmten Teil davon zu verteilen. Niederfrequentes Rauschen, wie z. B. Motoren oder menschliche Stimmen, wirkt sich beispielsweise hauptsächlich auf den niederfrequenten Teil des beobachteten Signals aus. Weißes Rauschen wirkt sich auf alle Frequenzen gleichermaßen aus. Mit diesem Vorwissen können Entrauschungsalgorithmen wie das "Thresholding" so konzipiert werden, dass das Rauschen reduziert oder eliminiert wird, während die wesentlichen Spektralkomponenten des Signals erhalten bleiben. Im Bereich der Audioverarbeitung führt dieser Ansatz zu einer erheblichen Verbesserung der Audioqualität, siehe Yu et al. (2008) oder Ashwin and Manoharan (2018).

Ein wichtiger Aspekt dieses Ansatzes ist die genaue Lokalisierung der Signalkomponen-

ten im Zeit-Frequenz-Bereich. Dies führt zu einer dünnbesetzten Darstellung des Signals, die hilft, es von Rauschen zu unterscheiden. Die beliebteste Methode der Zeit-Frequenz-Analyse ist die short-time Fourier transform (STFT). Dabei wird ein festes Fenster über das Signal geschoben, um die Fourier-Transformation zu ermitteln. Sie bietet eine feste Zeit-Frequenz-Auflösung, die durch die Länge des Fensters bestimmt wird. Dieser Ansatz führt jedoch häufig zu einem Kompromiss zwischen guter Zeitauflösung und guter Frequenzauflösung. Für jedes quadrat-integrierbare Signal erhält man eine abzählbare, redundante und energieerhaltende Darstellung des Signals unter Verwendung von Gabor-Frames, einer Sammlung von Zeit-Frequenz-Verschiebungen. Die STFT eignet sich gut für Signale mit stabilem Frequenzinhalt über lange Zeiträume. In solchen Fällen kann eine geeignete Fenstergröße für die Analyse gewählt werden.

Modulationsräume bieten einen natürlichen Rahmen für die Zeit-Frequenz-Analyse. Sie ermöglichen es, die Glätte eines Signals anhand der Abklingcharakteristik seiner STFT im Zeit-Frequenz-Raum zu messen, vgl. Galperin and Samarah (2004). In bestimmten Fällen stimmen die Modulationsräume mit Sobolev-Räumen überein. In Dahlke et al. (2022) liefern wir eine nichtparametrische Schätzung von Signalen, die durch stationäres Rauschen im Modell des weißen Rauschens verfälscht sind. Das beobachtete Signal ist ein Element in einem geeigneten Modulationsraum. Wir haben einen adaptiven und ratenoptimalen Schätzer für Modulationsräume konstruiert, indem wir die aus der Gabor-Expansion erhaltenen Koeffizienten kürzen. Die mit Hilfe klassischer Orakelungleichungen aus Donoho and Johnstone (1994) erhaltene Rate weist neue Merkmale auf, die die Einbeziehung sowohl der Zeit als auch der Frequenz widerspiegeln. Die Signale in Dahlke et al. (2022) wurden über den gesamten $\mathbb{R}^d$ beobachtet, weil erstens Signale wie Audiosignale natürlich keine inhärenten Beschränkungen auf bestimmte Domänen aufweisen, vgl. Goldenshluger et al. (2006). Zweitens ist die Theorie der Gabor-Expansionen und Modulationsräume im Gegensatz zu den Theorien anderer Funktionsräume wie Besov oder Sobolev für begrenzte Bereiche nicht vollständig entwickelt.

Eine gut untersuchte Alternative zur Zeit-/Frequenzanalyse ist die (orthogonale) Wavelet Transformation (WT). Wie in Daubechies (1990) beschrieben, bietet die WT auch eine Zeit-Frequenz-Lokalisierung mit einer Auflösung, die sich mit der Frequenz ändert. Sie lokalisiert effektiv hochfrequente Ereignisse in der Zeit und niederfrequente Ereignisse in der Frequenz, was sie zu einer geeigneten Wahl für die Analyse vieler realer Sig-

nale macht. Diese Signale haben oft langsam oszillierende Inhalte auf längeren Skalen, während hochfrequente Ereignisse eher abrupt oder vorübergehend auftreten. Es ist jedoch zu beachten, dass die WT zu einer reduzierten Frequenz- oder Zeitauflösung führen kann, wenn der spektrale Inhalt des zugrunde liegenden Signals auf einen bestimmten Frequenzbereich beschränkt ist. Donoho and Johnstone (1994) zeigen, dass die Wavelet-Entrauschung auf Besov-Räumen nahezu optimal ist, insbesondere für Szenarien mit weißem Gaußschen Rauschen. Mallat (2009) erörtert auch die Entrauschung von Wavelet Koeffizienten. Diese Ergebnisse unterstreichen die Wirksamkeit von Wavelet-Entrauschungstechniken im Umgang mit Rauschen in Signalen.

Um den Auflösungsproblemen der Gabor- und Wavelet-Transformationen zu begegnen, wurde ein hybrider Ansatz, die Flexible Gabor-Wavelet-Transformation (GWT), eingeführt, wie in Nazaret and Holschneider (2003) beschrieben. Die GWT verwendet einen Parameter $\alpha \in [0, 1]$, der die Fensterlänge dynamisch auf der Grundlage der Frequenz anpasst, was zu einer verbesserten Auflösung im Vergleich zur STFT führt. Außerdem wird der bei der WT beobachtete Verlust an Frequenzauflösung bei der Analyse höherer Frequenzen ausgeglichen. Der mit dieser Transformation verbundene Glättungsraum wird als α-Modulationsraum bezeichnet. Er liegt zwischen Modulationsräumen und Besov-Räumen und bietet einen vielseitigen Rahmen für die Analyse von Signalen mit unterschiedlichen Glättungsgraden im Zeit-Frequenz-Bereich. Die Einführung des α-Parameters ermöglicht eine flexible Anpassung der Fensterlänge, wodurch ein Gleichgewicht zwischen Zeit- und Frequenzauflösung erreicht wird.

In dieser Arbeit werden die in Dahlke et al. (2022) vorgestellten Ergebnisse neu formuliert und erweitert. Im Gegensatz zu der vorherigen Arbeit leiten wir jedoch alle Ergebnisse aus dem Sequenzraum ab und bieten somit eine andere Perspektive. Insbesondere sind die in dieser Analyse verwendeten Koeffizienten energieerhaltend, was bedeutet, dass das quadratische Risiko, das während des Entrauschungsprozesses der Zeit-Frequenz-Darstellung entsteht, auch für das Signal nach der Synthese der Koeffizienten gilt. Darüber hinaus erweitern wir die erzielten Ergebnisse, um vergleichbare Ergebnisse im beschränkten Bereich zu erhalten. Dies ermöglicht eine breitere Anwendbarkeit der Ergebnisse über den unbeschränkten Bereich hinaus. Darüber hinaus verallgemeinern wir die Schätzergebnisse, um den α-Modulationsraum in der eindimensionalen Umgebung einzubeziehen. Durch diese Erweiterungen und Verallgemeinerungen wollen

wir das Verständnis und die Anwendbarkeit von Zeit-Frequenz-Analysetechniken in verschiedenen Szenarien verbessern.

Diese Arbeit ist wie folgt gegliedert. Kapitel 2 gibt einen umfassenden Überblick über Frames in Hilbert- und Banach-Räumen. Wir untersuchen die STFT und die Verwendung von Gabor-Frames für Signalexpansionen. Wir besprechen das Konzept des Zerfalls in Produktmaßräumen unter Verwendung von Gewichtsfunktionen. Wir führen auch Modulationsräume ein und untersuchen ihre Charakterisierung mit Hilfe von Gabor-Frames. Schließlich befassen wir uns mit der Theorie der flexiblen Gabor-Wavelet-Transformation und α-Modulationsräumen. In Kapitel 3 greifen wir die Erkenntnisse von Dahlke et al. (2022) über die nichtparametrische Schätzung von Signalen in Modulationsräumen unter Verwendung von Gabor-Frames wieder auf. Insbesondere wird unsere Analyse vollständig im Sequenzraum durchgeführt. Wir schränken zunächst die Signalauflösung ein, indem wir eine endliche Teilmenge der Darstellung betrachten und das Signal innerhalb dieser endlichen Teilsequenz schätzen. Wir leiten die Minimax-Konvergenzrate ab und vergleichen sie mit bekannten Ergebnissen, wie z.B. denen im Sobolev-Fall. Kapitel 4 wiederholt die Ergebnisse aus Kapitel 3, konzentriert sich aber auf die Verwendung von α-Modulationsframes. Wir vergleichen sie mit Ergebnissen, die in Sobolev-Räumen und Besov-Räumen erzielt wurden. In Kapitel 5 präsentieren wir die Ergebnisse intensiver Berechnungen, die die Vorteile von Gabor-Frames bei der Audio-Entrauschung hervorheben. Diese Experimente zeigen die Effektivität des vorgeschlagenen Ansatzes, und wir liefern quantitative Ergebnisse. Wir vergleichen die geschätzten Signale mit den zugrundeliegenden Signalen. Dabei werden verschiedene Leistungskennzahlen wie das Signal-Rausch-Verhältnis (SNR) und den mittleren quadratischen Fehler (MSE) verwendet.