

Statistically optimal estimation of signals in modulation spaces using Gabor frames

Stephan Dahlke, Sven Heuer, Hajo Holzmann, and Pavel Tafo

Time-frequency analysis deals with signals for which the underlying spectral characteristics change over time. The essential tool is the short-time Fourier transform, which localizes the Fourier transform in time by means of a window function. In a white noise model, we derive rate-optimal and adaptive estimators of signals in modulation spaces, which measure smoothness in terms of decay properties of the short-time Fourier transform. The estimators are based on series expansions by means of Gabor frames and on thresholding the coefficients. The minimax rates have interesting new features, and the derivation of the lower bounds requires the use of test functions which approximately localize both in time and in frequency. Simulations and applications to audio recordings illustrate the practical relevance of our methods. We also discuss the best N -term approximation and the approximation of variational problems in modulation spaces by Gabor frame expansions.

Index Terms—denoising, Gabor frame, minimax estimation, short-time Fourier transform, time-frequency analysis, thresholding.

I. INTRODUCTION

Time-frequency analysis allows to deal with signals f for which the underlying frequencies change over time, as is common in many acoustic signals such as music [11] or bird songs [4], as well as in psychoacoustics [25] and wireless communications [28]. Gröchenig [17] provides a comprehensive

All authors are with the Faculty of Mathematics and Computer Science, Philipps-Universität Marburg, Marburg, Germany. H. Holzmann is the corresponding author
 {dahlke, heuersv, holzmann, tafo}@mathematik.uni-marburg.de

account of the mathematics of time-frequency analysis. The essential tool of time-frequency analysis is the short-time Fourier transform (STFT), which is defined by

$$\begin{aligned} V_{h_0}g(x, \omega) \\ = \int_{\mathbb{R}^d} g(t) \overline{h_0}(t-x) \exp(-2\pi i \langle \omega, t \rangle) dt \end{aligned} \quad (1)$$

where $g, h_0 \in \mathcal{L}^2(\mathbb{R}^d)$, $h_0 \neq 0$ is the so-called window function. Here, $\langle \omega, t \rangle$ denotes the Euclidean inner product of the vectors x and $\omega \in \mathbb{R}^d$, and \bar{z} is the complex-conjugate of $z \in \mathbb{C}$. Note that (1) is well-defined even for only locally integrable (e.g. bounded) signals g if the window h_0 has a compact support.

The STFT localizes the ordinary Fourier transform in time x by means of the window function h_0 . Modulation spaces measure the smoothness of signals by decay properties of their STFT in both time x and frequency ω . Similarly to Besov spaces and wavelet expansions, signals can be characterized as elements in modulation spaces by their Gabor frame expansions.

In this paper we investigate estimation of signals f observed in the white noise model

$$dY(x) = f(x) dx + \varepsilon dW(x), \quad x \in \mathbb{R}^d, \quad (2)$$

by Gabor frame expansions, where f is element in a suitable modulation space. We use estimators for f in (2) based on soft and hard thresholding of the Gabor coefficients. The analysis uses the classical oracle inequalities from Donoho and Johnstone [10], extended to complex-valued coefficients. See also Section 11 in Mallat [24] for a discussion of

denoising in frame expansions by thresholding. We show that our estimators achieve optimal rates in the minimax sense up to logarithmic factors for modulation spaces with commonly used weight functions. These rates appear to be new and not to correspond directly to the known rates over Sobolev or Besov spaces. In contrast to much literature on wavelet thresholding, the derivation of the lower bounds does not rely on a sequence space characterization of modulation spaces. Rather, we work directly with their definition via the STFT, and use Gaussian test functions which approximately localize simultaneously in time as well as in frequency. Our contributions to denoising are thus complementary to those of Wolfe, Godsill, and Ng [30] and Yu et al. [31], who study discrete Gabor expansions and focus on computational and practical aspects of denoising.

White noise models such as (2) are widely used in statistical analysis as stylized versions of more realistic nonparametric regression models. Formal approximation results of nonparametric regression by white noise on the rectangle $[0, 1]^d$ are e.g. established in Reiß [26]. In our setting, signals cannot be naturally restricted to certain domains. A further, more technical reason to use the white noise model on the whole of \mathbb{R}^d is that the theory of Gabor expansions and modulation spaces seems not to be developed fully for bounded domains as e.g. the theory of wavelet expansions and Besov spaces.

Some parts of the paper hold for general frames and are not restricted to the Gabor case. Indeed, given a frame, one can define an abstract smoothness space simply by collecting all functions for which the frame expansion coefficients are contained in a weighted ℓ_p space, say. Some of our results would then hold with respect to these kinds of spaces. However, it is one of the most important consequences of the famous coorbit theory that these abstract approximation spaces coincide with smoothness spaces, that is, the modulation spaces

constructed by means of the decay of a voice transform. It is one of the goals of this paper to indicate how this deep relation to the theory of function spaces can be exploited in practice.

The structure of the paper is as follows. We start in Section II by summarizing some facts on modulation spaces and Gabor expansions that we shall subsequently require. B-splines are attractive window functions from a numerical and computational point of view, and we briefly investigate the theoretical properties of B-spline windows. In Section III we introduce the thresholding estimators, and derive the minimax rates of convergence. Section IV deals with the best N -term approximation of functions in modulation spaces by Gabor frame expansions with emphasis on the sparse case. We also show how to approximate solutions to variational problems formulated in the corresponding sequence space using Gabor coefficients. Section V contains the results of extensive numerical experiments. As an illustration of our theoretical denoising results, in Section V-A we compare denoising with Gabor-based thresholding and wavelet thresholding both with universal threshold on various synthetic signals. Next, in Section V-B we investigate numerically the compression performance for various spline - and Gaussian window functions, both for a synthetic and a real-data signal. Finally, in Section V-C we give an illustration of the denoising performance of time-frequency based methods for various more sophisticated thresholding algorithms as well as two competing methods on several real-data examples. Proofs are deferred to Section VII.

II. MODULATION SPACES AND GABOR FRAMES

While Sobolev spaces measure smoothness of a signal f by decay properties of its Fourier transform, modulation spaces analogously rely on the short time Fourier transform in (1). In this section we gather the most relevant notions from the expositions of Gröchenig [17] and Galperin and

Samarah [16]. The formal definition of modulation spaces requires the notion of a weighted \mathcal{L}^p -space. The integer d will denote the dimension of the signal. Since the STFT involves both time and frequency, we introduce weighted \mathcal{L}^p -spaces on \mathbb{R}^{2d} . A weight function $v : \mathbb{R}^{2d} \rightarrow [0, \infty)$ is *submultiplicative* if $v(z_1 + z_2) \leq C v(z_1) v(z_2)$, $z_1, z_2 \in \mathbb{R}^{2d}$ for some constant $C > 0$, and $m : \mathbb{R}^{2d} \rightarrow [0, \infty)$ is *v-moderate* if $m(z_1 + z_2) \leq C v(z_1) m(z_2)$. A standard choice is $v_s(z) = m_s(z) = (1 + \|z\|_2^2)^{s/2}$ for a parameter $s \geq 0$, where $\|\cdot\|_2$ is the Euclidean norm. Given $p \in (0, \infty)$, the *weighted \mathcal{L}^p -space*, defined by

$$\mathcal{L}_m^p = \{g : \mathbb{R}^{2d} \rightarrow \mathbb{C} \text{ measurable} \mid \|g\|_{\mathcal{L}_m^p}^p = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |g(x, \omega)|^p m(x, \omega)^p dx d\omega < \infty\},$$

is a complete (semi-) normed space for $p \geq 1$, and a complete quasi-(semi-) normed space for $p \in (0, 1)$.

The *modulation space* $\mathcal{M}_m^p(\mathbb{R}^d)$ with weight function m and parameter $p > 0$ is defined as the set of tempered distributions $f \in \mathcal{S}'(\mathbb{R}^d)$, the dual of the Schwartz space $\mathcal{S}(\mathbb{R}^d)$, for which $V_{h_0} f \in \mathcal{L}_m^p(\mathbb{R}^{2d})$, and $h_0 \in \mathcal{S}(\mathbb{R}^d)$, $h_0 \neq 0$ is a fixed window function in the Schwartz space. See Gröchenig [17, Section 11.2] for the definition of the STFT of a tempered distribution. The modulation (quasi-) norm is given by

$$\begin{aligned} \|f\|_{\mathcal{M}_m^p(\mathbb{R}^d)} &= \|V_{h_0} f\|_{\mathcal{L}_m^p(\mathbb{R}^{2d})} \\ &= \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_{h_0} f(x, \omega)|^p m(x, \omega)^p dx d\omega \right)^{1/p}. \end{aligned} \quad (3)$$

If the weight function is $m = 1$ we write $\mathcal{M}_1^p(\mathbb{R}^d) = \mathcal{M}^p(\mathbb{R}^d)$ and $\mathcal{L}_1^p(\mathbb{R}^{2d}) = \mathcal{L}^p(\mathbb{R}^{2d})$. It turns out that the definition of the set $\mathcal{M}_m^p(\mathbb{R}^d)$ is independent of the window function, and the resulting modulation norms are equivalent, for a wide range of window functions, in particular all functions in the Schwartz space. See Gröchenig [17, Theorem 11.3.7 and Proposition 12.1.2] for $p \geq 1$ and Galperin and Samarah [16, Theorem 3.1]

for $p \in (0, 1)$. A standard choice is the Gaussian window function $\varphi_a(x) = \exp(-\pi \|x\|_2^2/a)$, $a > 0$. While by definition, modulation spaces consist of tempered distributions, for the weight functions m_s defined above and $m_{u,v}$ in (15) that we focus on, the modulation space is a subset of a Bessel potential space [17, Proposition 11.3.1] so that its elements actually correspond to functions. Modulation spaces can be characterized and its elements represented in terms of *Gabor frames*. A countable family $(e_\lambda)_{\lambda \in \Lambda}$ in a separable Hilbert space $(\mathcal{H}, \|\cdot\|)$ is a *frame* if for all $f \in \mathcal{H}$,

$$A \|f\|^2 \leq \sum_{\lambda \in \Lambda} |\langle f, e_\lambda \rangle|^2 \leq B \|f\|^2,$$

where $0 < A < B$ are the lower and upper frame bounds. The *synthesis operator* D associated with the frame is defined by $D\mathbf{c} = \sum_{\lambda \in \Lambda} c_\lambda e_\lambda$, where $\mathbf{c} = (c_\lambda)_{\lambda \in \Lambda}$ is a complex-valued sequence indexed by Λ , and the frame operator S by $Sf = \sum_{\lambda \in \Lambda} \langle f, e_\lambda \rangle e_\lambda$. The frame operator is a positive, invertible operator on \mathcal{H} . The family $(S^{-1} e_\lambda)_{\lambda \in \Lambda}$ is a frame called the *canonical dual frame*, and we have the representation $f = \sum_{\lambda \in \Lambda} \langle f, S^{-1} e_\lambda \rangle e_\lambda$. Given $\alpha, \beta > 0$ and a window function $h \in \mathcal{L}^2(\mathbb{R}^d)$, a *Gabor system* is the family of functions, indexed by the lattice $\Lambda = \alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$, defined by

$$h_\lambda(x) = \exp(2\pi i \langle \beta n, x - \alpha k \rangle) h(x - \alpha k), \quad (4)$$

with $\lambda = (\alpha k, \beta n) \in \Lambda$. For sufficiently small choices of $\alpha, \beta > 0$ and a suitable choice of h [17, Theorem 6.5.1], $(h_\lambda)_{\lambda \in \Lambda}$ is actually a frame in $\mathcal{L}^2(\mathbb{R}^d)$, called a *Gabor frame*, and the dual frame is also of the form (4) with the dual window $\tilde{h} = S^{-1}h$.

To characterize modulation spaces by Gabor frame expansions, we require the mixed norm sequence spaces: A sequence $\mathbf{c} = (c_\lambda)_{\lambda \in \Lambda}$ is in ℓ_m^p if [17, Def. 11.1.3]

$$\|\mathbf{c}\|_{\ell_m^p}^p = \sum_{\lambda \in \Lambda} |c_\lambda|^p m(\lambda)^p < \infty,$$

where m is the moderate weight function which is used in the definition of the weighted \mathcal{L}^p -space.

Under the assumption that the STFT of the window function h is element of a suitable amalgam space, see Gröchenig [17, Theorem 12.2.4], Galperin and Samarah [16, Theorem 3.7] we have that if $\mathbf{c} = (c_\lambda)_{\lambda \in \Lambda} \in \ell_m^p$ then $f = \sum_{\lambda \in \Lambda} c_\lambda h_\lambda \in \mathcal{M}_m^p(\mathbb{R}^d)$ and

$$\|f\|_{\mathcal{M}_m^p(\mathbb{R}^d)} \leq \text{const.} \|\mathbf{c}\|_{\ell_m^p}. \quad (5)$$

All Schwartz functions are admissible window functions. Moreover, if we let $\tilde{h} = S^{-1}h$ denote the dual window, for $f \in \mathcal{M}_m^p(\mathbb{R}^d)$ we have the expansion

$$f = \sum_{\lambda \in \Lambda} \langle f, \tilde{h}_\lambda \rangle h_\lambda \quad (6)$$

and for the canonical coefficients, also called moments, $(\langle f, \tilde{h}_\lambda \rangle)_{\lambda \in \Lambda}$ we have the bounds

$$\begin{aligned} \tilde{C}_1 \|f\|_{\mathcal{M}_m^p(\mathbb{R}^d)} &\leq \|(\langle f, \tilde{h}_\lambda \rangle)_{\lambda \in \Lambda}\|_{\ell_m^p} \\ &\leq \tilde{C}_2 \|f\|_{\mathcal{M}_m^p(\mathbb{R}^d)} \end{aligned} \quad (7)$$

with $f \in \mathcal{M}_m^p(\mathbb{R}^d)$ and for suitable constants $0 < \tilde{C}_1 < \tilde{C}_2$. It is one of the most important consequences of the coorbit theory that for p in $[1, \infty)$, the estimate in (7) is uniform, i.e. the constants C_1, C_2 do not depend on p . One would conjecture that the same is true for $p < 1$, or that the dependence shows at least a moderate behaviour, however, we were not able to find a rigorous statement of this form in the current literature. A dependence on p would of course not devalue our results. This would only imply that in some asymptotic results, such as in (16) in Theorem 6, it would take a longer time to see the asymptotics due to larger constants. However, our numerical experiments based on cardinal B-splines described in Section V strongly indicate that this seems not to be the case, so that an at most moderate dependence of the constant for small p appears to be plausible. From a numerical and computational perspective, window functions with compact support are more suitable choices. In particular, cardinal B-splines can be useful as window functions in large-data classification problems,

where computation time is an issue [19]. They are defined by convolutions, i.e., starting with $N_1 = \mathbf{1}_{[0,1]}$, the convolution product $N_k := N_{k-1} * N_1$ is called the *B-spline of order k* , $k \geq 2$. However, since B-splines are not contained in the Schwartz space, the question arises which modulation spaces can be characterized by means of B-spline window functions, i.e., for which range of p (7) still holds. The answer is given by the following theorem which, to the best of our knowledge, has not been stated explicitly in the literature before.

Theorem 1. *Let $d = 1$ and let m be a v -moderate weight with $v(x, \omega) = v(x)$ depending only on time. Also, let the window function h be a B-spline of order k . Then, the norm equivalence (7) holds for all $p > \frac{1}{k}$.*

Remark. It is a consequence of the general coorbit theory [14] that if one samples densely enough, the norm equivalence (7) holds. Indeed, one has to find suitable so-called U-dense sets such that certain integrability conditions are satisfied. In the setting of Theorem 1, these U-dense sets can be constructed by a dense enough sampling strategy.

The proof is provided at the end of Section VII. We shall investigate the numerical performance of B-splines as window functions further in the Section V.

III. THRESHOLDING ESTIMATORS, ORACLE INEQUALITIES AND RATES OF CONVERGENCE

Let us introduce threshold estimators based on Gabor frame expansions, that is systems of functions of the type (4) for suitable $\alpha, \beta > 0$ and window function h . The white noise model (2) can be interpreted as a Gaussian process, and by integrating a real-valued function $h \in \mathcal{L}^2(\mathbb{R}^d)$ we observe the Gaussian random variable

$$Y(h) = \int h \, dY \sim \mathcal{N}(\langle f, h \rangle_{\mathcal{L}^2}, \varepsilon^2 \|h\|_{\mathcal{L}^2}^2), \quad (8)$$

where $\mathcal{N}(a, \sigma^2)$ is the normal distribution with mean a and variance σ^2 , and $Z \sim \mathcal{N}(a, \sigma^2)$ means

that the random variable Z has the $\mathcal{N}(a, \sigma^2)$ -distribution. Furthermore, for $h_1, \dots, h_m \in \mathcal{L}^2(\mathbb{R}^d)$, the random variables $Y(h_1), \dots, Y(h_m)$ from (8) are jointly normally distributed with covariance $\text{Cov}(Y(h_j), Y(h_k)) = \varepsilon^2 \langle h_j, h_k \rangle$.

When applying (8) to a complex-valued square-integrable function $g = h_1 + i h_2$, where $h_1, h_2 \in \mathcal{L}^2(\mathbb{R}^d)$ are real-valued, $Y(g) = Y(h_1) + iY(h_2)$ is a complex-valued normally-distributed random variable, meaning that $(Y(h_1), Y(h_2))^\top$ is bivariate (real-valued) Gaussian. To construct estimators for f in (2), assume that the window h together with its dual window \tilde{h} are such that (7) holds, and estimate the coefficient

$$\vartheta_\lambda := \langle f, \tilde{h}_\lambda \rangle \quad \text{by} \quad Y(\tilde{h}_\lambda). \quad (9)$$

Note that even though the window \tilde{h} is real-valued in case of a regular grid, see Strohmer [28] for more information on the canonical dual window, the element of the frame \tilde{h}_λ in (4) will be complex-valued.

Soft and hard thresholding at level $\mu > 0$ are defined by

$$\begin{aligned} t_s(v; \mu) &= \text{sign}(v) (|v| - \mu) \mathbf{1}(|v| \geq \mu), \\ t_h(v; \mu) &= v \mathbf{1}(|v| \geq \mu), \end{aligned}$$

where $v \in \mathbb{R}$ and $\text{sign}(v)$ is the sign of v . For complex $z = u + iv \in \mathbb{C}$, $u, v \in \mathbb{R}$ we define $t_j(z; \mu) = t_j(u; \mu) + i t_j(v; \mu)$, $j \in \{h, s\}$. Then we have the following oracle inequalities, which are extensions to complex-valued random variables of classic results from Donoho and Johnstone [10].

Proposition 2. *Let $\Lambda_0 \subset \Lambda$ with $\#\Lambda_0 < \infty$. Then, in the Gaussian white noise model (2) we have for soft thresholding with universal threshold $\mu_{\text{uni}} = \varepsilon \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_0)}$ that*

$$\begin{aligned} & \mathbb{E} \left[\sum_{\lambda \in \Lambda_0} |t_s(Y(\tilde{h}_\lambda); \mu_{\text{uni}}) - \vartheta_\lambda|^2 \right] \\ & \leq (4 \log(\#\Lambda_0) + 2) \cdot (\varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2 \\ & \quad + \sum_{\lambda \in \Lambda_0} \min(\varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2, |\vartheta_\lambda|^2)). \quad (10) \end{aligned}$$

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

The proposition is obtained when thresholding real-and imaginary parts separately. We mention that Mallat [24], Section 11.3.3., recommends to threshold the complex modulus instead, mainly for reasons of better perceptual sound quality when denoising an audio signal. For the theoretical developments, however, the above result is quite adequate.

For a finite subset $\Lambda_0 \subset \Lambda$ we consider the thresholding estimators

$$\hat{f}_j(\cdot; \mu) = \sum_{\lambda \in \Lambda_0} t_j(Y(\tilde{h}_\lambda); \mu) h_\lambda, \quad j \in \{h, s\}. \quad (11)$$

From boundedness of the synthesis operator D_h [17, Prop. 5.1.1 (b)] we obtain that for $j \in \{h, s\}$,

$$\begin{aligned} & \|\hat{f}_j(\cdot; \mu) - f\|_{\mathcal{L}^2}^2 \quad (12) \\ & = \left\| \sum_{\lambda \in \Lambda_0} (t_j(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda) h_\lambda - \sum_{\lambda \in \Lambda_0^c} \vartheta_\lambda h_\lambda \right\|_{\mathcal{L}^2}^2 \\ & \leq B \left(\sum_{\lambda \in \Lambda_0} |t_j(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda|^2 + \sum_{\lambda \in \Lambda_0^c} |\vartheta_\lambda|^2 \right), \end{aligned}$$

where B is the upper frame constant of the Gabor frame. In the following, using the oracle inequalities from Proposition 2 we shall bound (12) for particular choices of the weight function. We also mention that interchanging the roles of the window function h and its canonical dual \tilde{h} would lead to the same theoretical results.

Isotropic weight function

First let us consider an isotropic weight function of the following form. For a parameter $s > 0$ and $\lambda = (x^\top, \omega^\top)^\top$ choose the weight function as

$$m_s(\lambda) = v_s(\lambda) = (1 + \|x\|_2^2 + \|\omega\|_2^2)^{s/2}, \quad (13)$$

and $v_s = m_s$, as mentioned in the introduction. Given $K > 0$ let $\Lambda_{0,K} = \{\lambda \in \Lambda \mid \|\lambda\|_2 \leq K\}$, and for large K consider the estimator in (11) which can be written as

$$\hat{f}_j(\cdot; \mu) = \sum_{\|\lambda\|_2 \leq K} t_j(Y(\tilde{h}_\lambda); \mu) h_\lambda, \quad j \in \{s, h\}.$$

Theorem 3. Consider model (2) with $f \in \mathcal{M}_{m_s}^p(\mathbb{R}^d)$ for $p \in (0, 2]$. Choosing the universal threshold $\mu_{\text{uni}} = \varepsilon \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_{0,K})}$ as well as taking $K \gtrsim \varepsilon^{-\frac{d(2-p)+sp}{2d+s+2p}}$ we have the bound

$$\begin{aligned} & \mathbb{E}[\|\hat{f}_j(\cdot; \mu) - f\|_{\mathcal{L}^2}^2] \\ & \leq \text{const.} \cdot \max\left(\|f\|_{\mathcal{M}_{m_s}^p}^2, \|f\|_{\mathcal{M}_{m_s}^{\frac{2d+p}{2d+s+p}}}\right) \\ & \quad \cdot \log(1/\varepsilon) \cdot \varepsilon^{-\frac{2d(2-p)+2sp}{2d+s}}, \quad j \in \{s, h\}, \end{aligned} \quad (14)$$

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

Remark. Let us discuss the rate obtained in (14). For $p = 2$ we obtain $\varepsilon^{\frac{4s}{2d+2s}}$ up to the logarithmic factor, which is reminiscent with the rate over Sobolev ellipsoids, but somewhat surprisingly involves the dimension as $2d$ instead of the ordinary d . The reason is that both the dimensions d of time x as well as of frequency ω influence the rate. As indicated above, the rates obtained in the white noise model (2) on the whole of \mathbb{R}^d cannot be directly compared to rates in the case of compact support. For small p , that is in the sparse situation the rate in (14) approaches the parametric rate ε^2 . The logarithmic factor $\log(1/\varepsilon)$ is probably not necessary, and could potentially be eliminated by adopting a more sophisticated thresholding scheme such as SureShrink from Donoho and Johnstone [9].

Let us complement the result by a lower bound.

Theorem 4. For a constant $C > 0$ and $p \in (0, 2]$ consider the ball in $\mathcal{M}_{m_s}^p(\mathbb{R}^d)$,

$$\mathcal{M}_{s,C}^p = \{f \in \mathcal{M}_{m_s}^p(\mathbb{R}^d) \mid \|f\|_{\mathcal{M}_{m_s}^p}^2 \leq C\}.$$

Then in model (2) we have that

$$\begin{aligned} & \liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{2d(2-p)+2sp}{2d+s+p}} \cdot \right. \\ & \quad \left. \inf_{\hat{f}_\varepsilon} \sup_{f \in \mathcal{M}_{s,C}^p} \mathbb{E}_f[\|\hat{f}_\varepsilon - f\|_{\mathcal{L}^2}^2] \right) > 0, \end{aligned}$$

where \hat{f}_ε is any estimator in (2) based on the observation Y , and \mathbb{E}_f denotes the expected value if the underlying parameter in (2) is f .

Remark. We provide a construction which is based directly on the definition of modulation spaces in (3), and does not rely on characterizations in terms of Gabor coefficients. While the proof uses standard tools from decision theory such as Fano's lemma and the Varshamov-Gilbert bound, see Tsybakov [29], the issue is to construct the hypothesis functions in order to obtain the term $2d$ which arises in the upper bound. Since there are no integrable functions which localize sharply, meaning with compact support, in both time and frequency domain [17, Theorem 2.3.3], we work with Gaussian test functions and estimate the overlaps in time and frequency domain.

Anisotropic weight function

Let us consider the more general situation of an anisotropic weight function of the following form. For $0 \leq u, v \leq s$ and $\lambda = (x^\top, \omega^\top)^\top$ choose

$$m_{u,v}(\lambda) = (1 + \|x\|_2^2)^{u/2} + (1 + \|\omega\|_2^2)^{v/2}, \quad (15)$$

it is then also v_s -moderate for v_s given in (13).

For a constant $C > 0$ let

$$\mathcal{M}_{u,v,C}^p = \{f \in \mathcal{M}_{m_{u,v}}^p(\mathbb{R}^d) \mid \|f\|_{\mathcal{M}_{m_{u,v}}^p}^2 \leq C\}$$

denote the ball in $\mathcal{M}_{m_{u,v}}^p(\mathbb{R}^d)$.

Theorem 5. Consider model (2) with $f \in \mathcal{M}_{m_{u,v}}^p(\mathbb{R}^d)$ for $p \in (0, 2]$. Choosing the universal threshold $\mu_{\text{uni}} = \varepsilon \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \log(\#\Lambda_{0,K})}$ as well as taking $K \gtrsim \varepsilon^{-\frac{(2-p)d(v+u)+2pvu}{\min(u,v)(d(v+u)+pvu)}}$ we have the upper bound

$$\begin{aligned} & \mathbb{E}[\|\hat{f}(\cdot; \mu) - f\|_{\mathcal{L}^2}^2] \\ & \leq \text{const.} \cdot \max\left(\|f\|_{\mathcal{M}_{m_{u,v}}^p}^2, \|f\|_{\mathcal{M}_{m_{u,v}}^{\frac{pd(v+u)}{d(v+u)+pvu}}}\right) \\ & \quad \cdot \log(1/\varepsilon) \cdot \varepsilon^{-\frac{(2-p)d(v+u)+2pvu}{d(v+u)+pvu}}, \end{aligned}$$

and furthermore the corresponding lower bound

$$\begin{aligned} & \liminf_{\varepsilon \downarrow 0} \left(\varepsilon^{-\frac{(2-p)d(v+u)+2pvu}{d(v+u)+pvu}} \cdot \right. \\ & \quad \left. \inf_{\hat{f}_\varepsilon} \sup_{f \in \mathcal{M}_{u,v,C}^p} \mathbb{E}_f[\|\hat{f}_\varepsilon - f\|_{\mathcal{L}^2}^2] \right) > 0. \end{aligned}$$

Remark. If $u = v = s$ in Theorem 5 we recover the result in Theorem 3. On the other hand, if $u = 0$

the rate reduces to ε^{2-p} and hence is independent of v , that is, a higher decay in the frequency (or in time) alone cannot be used in the estimator in the model (2). For the case $p = 2$, we then do not obtain a rate. This is somewhat surprising since for $u = 0$ and $p = 2$, $\mathcal{M}_{m_0, v}^2(\mathbb{R}^d)$ corresponds to a Sobolev space [17, 11.3.1]. Here the reason seems to be the choice of the model (2) on \mathbb{R}^d . Indeed, if we assume that the signal f as well as the window h_0 have compact support, then the support of STFT $V_{h_0}f(x, \omega)$ is also uniformly bounded in x for all ω , and hence f belongs to $\mathcal{M}_{m_u, v}^p$ for arbitrarily large u . For $p = 2$, fixed v and $u \rightarrow \infty$ we obtain the exponent $4v/(d + 2v)$, corresponding to the Sobolev case on bounded domains. The situation is somewhat reminiscent of anisotropic multi-index denoising [22, 23], though the rates that we obtain are different.

IV. COMPRESSION AND APPROXIMATION OF VARIATIONAL PROBLEMS

So far, we have discussed denoising algorithms based on Gabor frames. Another very important task in signal processing is of course compression. To this end, the signal is decomposed with respect to the underlying dictionary, the small coefficients are thrown away by thresholding, and then the signal is reconstructed. These kinds of algorithms are clearly very much related with best N -term approximation schemes. In this section, we briefly analyze the approximation rate of thresholding algorithms based on Gabor frames.

Classic results from wavelet theory [8] imply that the convergence rate of best N -term wavelet approximation schemes depends on the smoothness of the signal under consideration in a specific scale of Besov spaces. For functions in modulation spaces, similar results hold true for best N -term approximations by elements from a Wilson basis, see Gröchenig [17, Theorem 12.4.2], at least for non-weighted norms - that is $m = 1$ - and in the non-sparse case $p \geq 1$. Further results which also

treat the case $0 < p$ but involve a further parameter $q > p$ in the definition of modulation spaces are presented in Gröchenig and Samarah [18], Samarah and Al-Sa'di [27]. Here we give a simple upper bound for $p \in (0, 2)$ in terms of Gabor frames. A numerical illustration can be found in Section V-B.

Theorem 6. *Let $p \in (0, 2)$ and $\mu > 0$. Given $f \in \mathcal{M}^p(\mathbb{R}^d)$, setting $I_\mu = \{\lambda \in \Lambda \mid |\langle f, \tilde{h}_\lambda \rangle| \geq \mu\}$, $N_\mu = \#I_\mu$ and*

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

we have that $N_\mu < \infty$ and that

$$\|f - f_\mu\|_{\mathcal{L}^2}^2 \leq B \tilde{C}_2^2 \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}^2 N_\mu^{1-2/p}, \quad (16)$$

where \tilde{C}_2 is from (7).

Remark. To prove this compression rate, we only need the norm equivalence in (7). By Theorem 1, this holds whenever we use a B-spline of order $k > \frac{1}{p}$ as the window function. Therefore we can balance the computational complexity and the compression rate by using B-splines of suitable order, depending on the smoothness of the signal f in the modulation space. In Section V we show some simulations highlighting this effect.

Remark. Theorem 6 holds for general frames that characterize modulation spaces. It cannot be expected that the compression rate can be improved by increasing the overlap. This would more or less produce more significant coefficients with lower amplitude. Nevertheless, the quality of the compression is not completely independent of the characteristics of the frame. Indeed, the frame bounds and hence the conditioning of the frame influence the constants in (16). Therefore, in practice one would prefer a well-conditioned frame with good localization and moderate overlap.

Next, given $f \in \mathcal{L}^2(\mathbb{R}^d)$ and a parameter $\mu > 0$, we aim to find $g \in \mathcal{M}_m^p(\mathbb{R}^d)$ that solves the *variational problem*

$$\min_{g \in \mathcal{M}_m^p(\mathbb{R}^d)} (\|f - g\|_{\mathcal{L}^2}^2 + \mu \|g\|_{\mathcal{M}_m^p(\mathbb{R}^d)}^p). \quad (17)$$

Variational problems of the form (17) occur in many practical applications. A prominent example is given by the regularization of inverse problems by means of Tikhonov schemes, see, e.g., Engl et al. [12]. Usually, the penalty terms are given by smoothness norms such as Sobolev norms. Quite recently, to improve sparsity, also Besov norms in combination with wavelets have been used [6]. To the best of our knowledge, Gabor frames and modulation spaces have only rarely been used in this context.

Therefore, here we show how to balance the mean squared fit $\|f - g\|_{\mathcal{L}_2}^2$ of g to the data f and the smoothness of g as measured by the multiple of the modulation norm, $\mu \|g\|_{\mathcal{M}_m^p(\mathbb{R}^d)}^p$. For smoothness penalties from Sobolev or Besov norms, solutions of such variational problems have been extensively studied in terms of orthogonal wavelet expansions, see e.g. DeVore and Lucier [7], Chambolle et al. [3].

Theorem 7. *Let $p > 0$, and suppose that the window $h \in \mathcal{M}_v^{\min(1,p)}(\mathbb{R}^d)$. There is a constant $C > 0$ depending on the upper frame bound of $\{h_\lambda \mid \lambda \in \Lambda\}$ and on the constants in (5) such that if $\mathbf{c} = (c_\lambda)_{\lambda \in \Lambda} \in \ell_m^p$ and hence $g = \sum_{\lambda \in \Lambda} c_\lambda h_\lambda \in \mathcal{M}_m^p(\mathbb{R}^d)$ by (5), we have that*

$$\begin{aligned} & \|f - g\|_{\mathcal{L}_2}^2 + \mu \|g\|_{\mathcal{M}_m^p(\mathbb{R}^d)}^p \\ & \leq C \sum_{\lambda \in \Lambda} ((\langle f, \tilde{h}_\lambda \rangle - c_\lambda)^2 + \mu m(\lambda)^p |c_\lambda|^p). \end{aligned} \quad (18)$$

Remark 1. The right-hand side of (18) can be minimized coefficientwise by minimizing

$$E(c) = (\langle f, \tilde{h}_\lambda \rangle - c)^2 + \mu m(\lambda)^p |c|^p$$

in c for each $\lambda \in \Lambda$. Since for the minimizer $\tilde{\mathbf{c}} = (\tilde{c}_\lambda)_{\lambda \in \Lambda}$, the right-hand side of (18) is finite (it is finite for the choice $c_\lambda = 0$), we must have $\tilde{\mathbf{c}} \in \ell_m^p$. The choice

$$\hat{c}_\lambda = \langle f, \tilde{h}_\lambda \rangle \mathbf{1}_{|\langle f, \tilde{h}_\lambda \rangle|^2 \geq \mu m(\lambda)^p |\langle f, \tilde{h}_\lambda \rangle|^p}$$

satisfies $E(\hat{c}_\lambda) \leq 4E(\tilde{c}_\lambda)$, see Chambolle et al. [3, Section 3, p. 5] hence we also have that $\hat{\mathbf{c}} = (\hat{c}_\lambda)_{\lambda \in \Lambda} \in \ell_m^p$.

Remark 2. In contrast to orthogonal wavelet expansions, where the version in terms of the coefficients is of the same order, here we only have the upper bound as stated in (18). This can be explained as follows: When representing $g = \sum_{\lambda \in \Lambda} c_\lambda h_\lambda \in \mathcal{M}_m^p(\mathbb{R}^d)$ for general frame coefficients $(c_\lambda)_{\lambda \in \Lambda}$, we have the upper bound in (5) but the upper bound in (7) is only valid for the canonical frame coefficients $(\langle f, \tilde{h}_\lambda \rangle)_{\lambda \in \Lambda}$.

Modulation spaces at least for $p \geq 1$ can also be characterized in terms of the coefficients of Wilson basis, see Gröchenig [17, Theorem 12.3.1]. Thus, in this situation solutions to the variational problem can actually be characterized and not merely bounded up to constants by solutions involving Wilson basis coefficients. However, in the present paper our particular emphasis is on the sparse case $p < 1$.

V. NUMERICAL EXPERIMENTS

This section contains the results of extensive numerical experiments. As an illustration of our results in Section III, in Section V-A we compare denoising with Gabor-based thresholding and wavelet thresholding both with universal threshold on various synthetic signals. Next, in Section V-B we investigate numerically our results on compression from Section IV for various spline - and Gaussian window functions, both for a synthetic and a real-data signal. Finally, in Section V-C we give an illustration of the denoising performance of time-frequency based methods for several more sophisticated thresholding algorithms as well as two competing methods on three real-data examples.

A. Denoising

In this section we illustrate numerically our denoising based on Gabor thresholding, as discussed

in Section III, for various synthetic signals, and contrast it with wavelet thresholding methods. The aim of this section is thus not to optimally tune thresholding for Gabor expansions, but rather to compare a simple approach with universal thresholding with a similar method in the wavelet case, and to investigate which approach performs better for specific signals. In the subsequent Section V-C, for various real-data signals we compare several thresholding algorithms for Gabor expansions and also briefly discuss the issue of musical noise. For background on numerical Gabor analysis see e.g. Feichtinger [13].

To investigate denoising as discussed from a theoretical point of view in Section III, we shall consider the following family of functions of spatially varying frequency

$$f_{A,B}(t) = \sin(2\pi \cdot B \cdot t \cdot e^{-A(t-0.5)^2}), \quad t \in [0, 1], \quad (19)$$

which resemble the Doppler functions used by Donoho and Johnstone [10]. In particular we consider the following three signals: firstly, the signal $f_{50,2}$ with small frequency variation, see Figure 1(a); secondly, the signal $f_{50000,4}$ with small time variation, see Figure 1(b); and lastly the signal $f_{40,200}$ which varies equally in both time and frequency domains, see Figure 1(c).

For the wavelet shrinkage methods we use a B-spline of level 4 as the window function for the Gabor system and for comparison use wavelet thresholding with the biorthogonal B-spline wavelet of level 4 ('bior4.4').

Each signal includes frequencies up to 1000 Hz. We generate discrete observations from

$$Y_i = f(i/n) + \sigma \epsilon_i, \quad i = 1, \dots, n,$$

with $n = 2000$ and use a discretized version of the Gabor frames in (4). In repeated simulations we shall use $m = 2000$ iterations, e.g. to compute mean squared errors.

N	2000	400	200	100	50
M	2000	400	200	100	50
$\alpha \cdot \beta$	0.0005	0.0125	0.05	0.2	0.8

TABLE I: Number of frequency and time bands and the grid density

Effects of window length and grid density

In this setting the window function is discretized over $W \in \mathbb{N}$ observations within its support. We start by investigating the effects of the window width $w = \frac{W}{n} \in (0, 1]$ as well as the effect of the grid density $\alpha \cdot \beta$ of the frame in (4). In the discrete setting we refer to $\alpha = \lceil \frac{n}{N} \rceil$ as the step size in terms of samples in the time domain and to $\beta = \frac{1}{M}$ as the step size in the frequency domain, with $M, N \in \{1, \dots, n\}$. M and N respectively represent the number of frequency and time bands considered. This yields a time-frequency representation of the signal of the size $M \times N$. A Gabor frame can only exist if the density satisfies $\frac{n}{N \cdot M} < 1$, see [17][Theorem 7.5.3]. The interaction between grid density and window length affects the computational time as the overlap rate of the window function, $1 - \frac{1}{N \cdot w}$, increases. A dense grid as well as a long window function intensify the redundancy among the coefficients. Here we report results for the case $M = N$, additional results for distinct time - and frequency resolutions can be found in the supplement. Figure 3 shows the effect of different window lengths and grid densities according to Table I on the mean squared error (MSE) for our three signals for samples of sizes $n = 2000$, where we use hard thresholding with an MSE-optimal threshold which was computed over a fine grid of threshold values.

Let us interpret the results. In all three cases, and in particular for $f_{40,200}$, for suitable values of the window width the MSE for the density of 0.05 and even of 0.2 is almost as low, or at least of similar magnitude, than that for the grid with finest resolution 0.0005. Second, for the signals

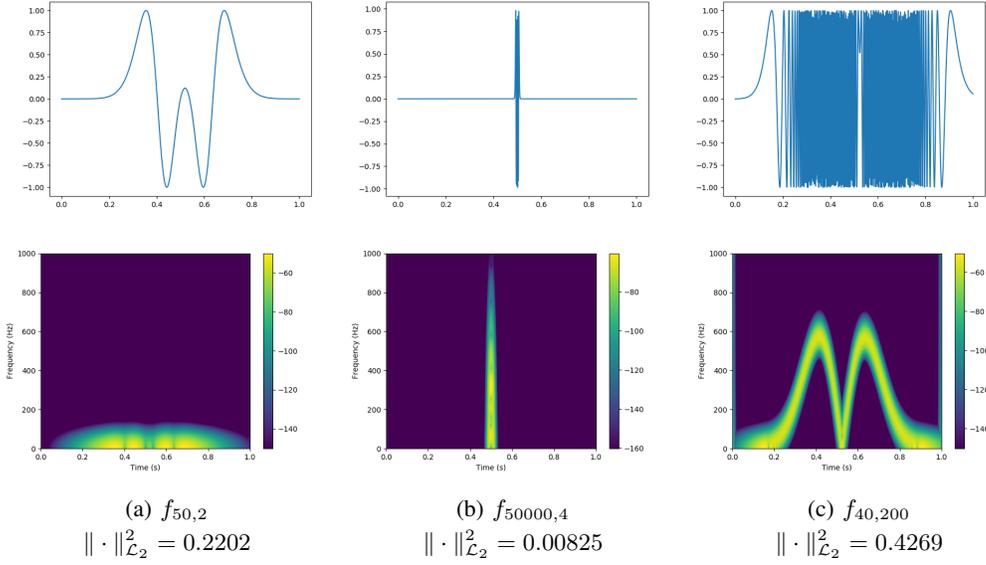


Fig. 1: Signals (upper row) and spectrogram (lower row) of our test signals

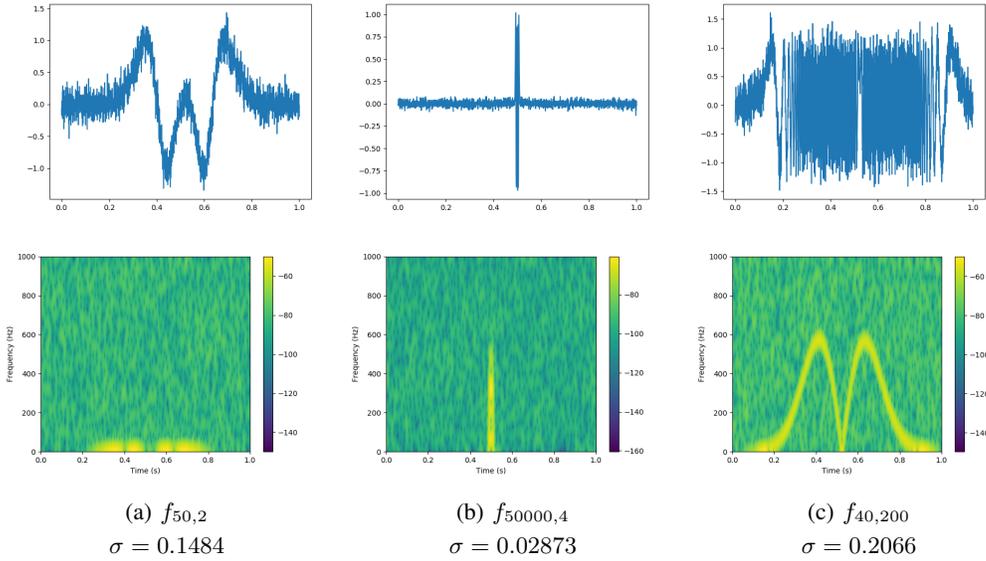


Fig. 2: Noisy signals with SNR = 10

$f_{50000,4}$ and $f_{40,200}$, and also for $f_{50,2}$ except for a very fine resolution, the MSE increases as the window width is increased, at least above 0.05 or 0.1. The significant improvement in the MSE justified the very high overlap rate, around 90%, needed to achieve lowest error. Then, to achieve reasonable computation times and still have good MSE properties, in all three cases in the following

simulations we choose $M = N = 100$, resulting in $\alpha \cdot \beta = 0.2$, and for the window width we take the following values

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

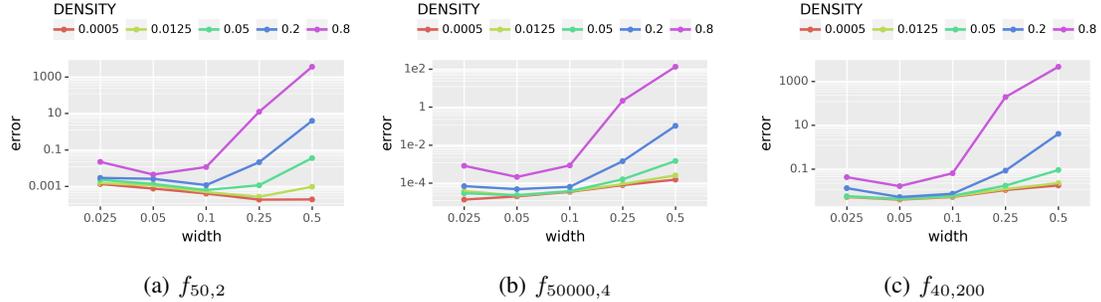


Fig. 3: Effect of the window width w and the grid density with $M = N$ for the three signals for samples of sizes $n = 2000$, using hard thresholding with the optimal threshold.

Choice of threshold and thresholding method

Next, in a repeated simulation for various sample sizes we investigate the MSE of soft - and hard thresholding, both for the universal threshold as well as for an MSE-optimal threshold, which was computed over a fine grid of threshold values. The results are contained in Figure 4, where the figures in the first row contain the MSE for soft thresholding while those in the second line have the MSE for hard thresholding. We observe that first, hard thresholding performs better than soft thresholding, and second, the universal threshold appears to be a very reasonable choice for hard thresholding.

Comparison of Gabor-frame and wavelet-based methods

Next we investigate the reconstruction performance of our thresholding estimators in (11), compared to wavelet shrinkage methods based on biorthogonal B-splines wavelet of order 4. Due to the previous results we chose to focus on hard thresholding with the universal threshold. First, for a visual impression, for a particular sample of size $n = 2000$ we plot the reconstructions of the signals in Figure 5, where the figures in the first row are from wavelet shrinkage, the figures in the second row from the Gabor-frame based method, and the last row has the spectrograms of the Gabor-frame based estimates. In particular for the signal $f_{40,200}$ the

Gabor-frame based method seems to give a better result, whereas for the signal $f_{50000,4}$ which consists of a single spike, we expect wavelet shrinkage to perform better. This is however not yet visible from the plots. Next we compare the MSE of both methods in a repeated simulation for various sample sizes. The results are given in Figure 6. For the signal $f_{40,200}$, our Gabor-frame based method clearly outperforms wavelet shrinkage. This is also true, though less substantially, for the signal $f_{50,2}$ except for small sample sizes. Finally, for the signal $f_{50000,4}$ wavelet shrinkage seems to be superior, in particular for small and moderate samples.

B. Compression rates

In Theorem 6, we have studied the achievable compression rate using the best N -term approximation with Gabor coefficients. Note that taking logarithms, (16) can be written as

$$\begin{aligned} & \log(\|f - f_\mu\|_{\mathcal{L}^2}^2) \\ & \leq \log(B \tilde{C}_2^2 \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}^2) + \log(N_\mu) \cdot (1 - 2/p). \end{aligned} \quad (20)$$

We now demonstrate this rate by compressing a synthetic and a real signal with different window functions and comparing the errors. The complete code is available on GitHub¹.

First, we analyse the Gaussian function

$$f(x) = \frac{1}{0.3 \cdot \pi} \exp\left(-\frac{1}{2} \left(\frac{x - 0.5}{0.15}\right)^2\right),$$

¹<https://github.com/Heuerv/Compression-Simulations>

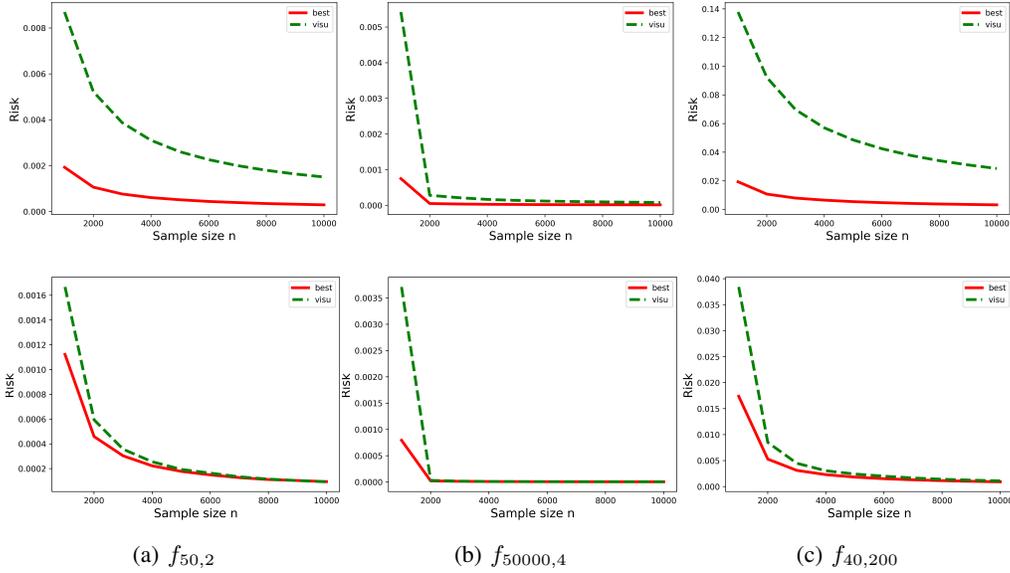


Fig. 4: Figures in first row: MSE for soft thresholding, optimal threshold versus universal threshold; Figures in second row: MSE for hard thresholding, optimal threshold versus universal threshold. Line style — denotes universal threshold. Various sample sizes are investigated.

for $x \in [0, 1]$ which is scaled in a way that f is close to zero at the boundary. We sample f at 2000 equidistant time stamps. A synthetic signal like this should be in all modulation spaces $\mathcal{M}^p([0, 1])$, $p > 0$, at least approximately taking into account the small discontinuities at the boundary. We use a Gabor transformation on this signal and reconstruct it using only N coefficients, with N being between 5% and 25% of the coefficients. This way, we eliminate distortions that result from the logarithmic scaling of N in (20) or the good time-frequency localisation of the signal and only focus on the “main slope” in the results. To achieve compatibility between the results, we always use 500 different values for N , spaced equidistantly in the chosen interval of 5% to 25%.

The exact results clearly depend on the step sizes in time and frequency and the window lengths of the spline windows. We found step sizes of 10 samples in time and 10 Hz in frequency as well as a universal window length of 180 samples for the splines to work fine for this signal. This

results in a very high redundancy in the frame. For a comparison of Gaussian window functions of different widths see the end of this section. Figure 7 (left) shows the results. Here, we can clearly see the advantage of a smooth window function if the signal is smooth as well. For the spline windows, we get better compression rates for higher spline orders. To quantify this, we take (20) and use linear regression on the log-scaled errors with respect to the log-scaled values for N and calculate p from the slope. In Table II we present the results and compare them to values of p we would expect with the relation $p > \frac{1}{k}$ from Theorem 1 in mind.

The deviation from the estimated p to $1/k$ with k equal to the order of the spline is quite small. Additional experiments showed that by modifying the window size, the difference could even be further reduced. It should also be noted, that the simulation with the Gaussian window took 1.66 seconds, while we needed 10.96 seconds for all ten spline windows. This shows that even with such a small signal (only 2000 samples) we get a time

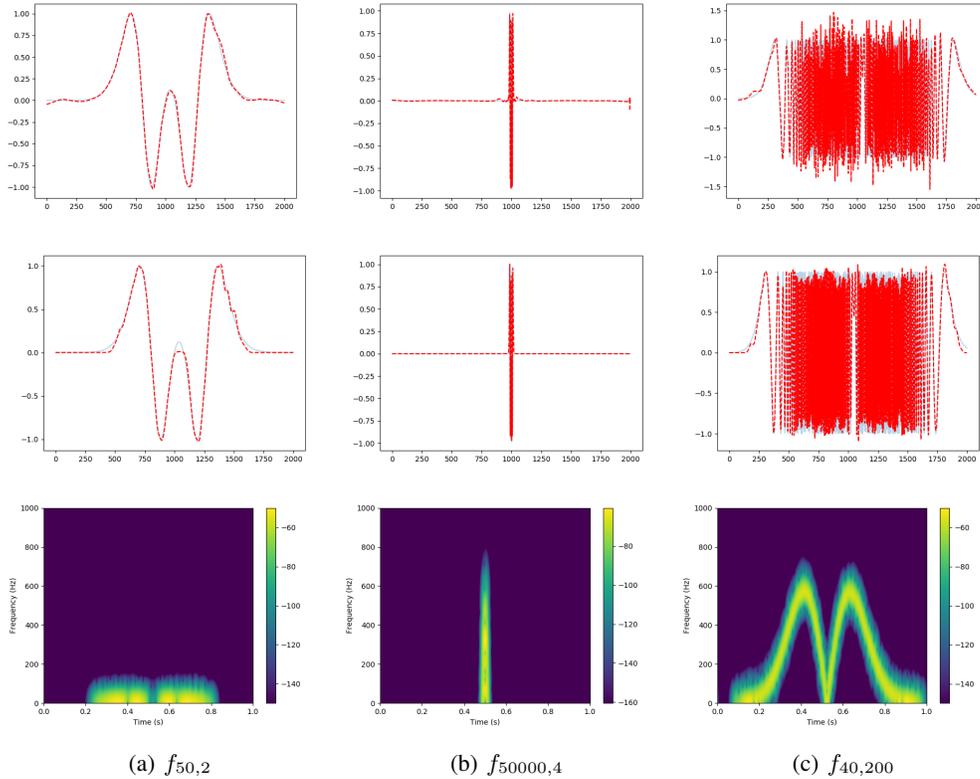


Fig. 5: Reconstructions for a particular sample of size $n = 2000$ using hard thresholding with universal threshold. Figures in the first row are from wavelet shrinkage, the figures in the second row from the Gabor-frame based method, and the last row has the spectrograms of the Gabor-frame based estimates.

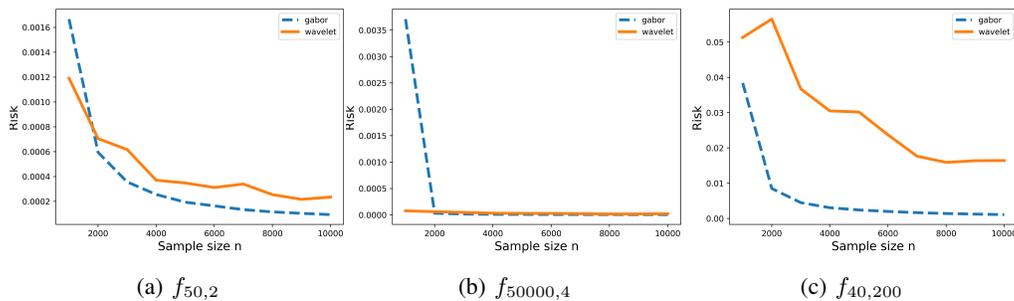


Fig. 6: MSE of the competing methods using universal hard thresholding. ---: Gabor-frame based method, — wavelet-shrinkage.

advantage by using the compactly supported spline windows, but at the expense of worse compression rates. All simulations in this section were carried out on an Intel i7 CPU with a clock speed of 2.93GHz.

We now turn to numerical illustrations using the

recording of a common blackbird² which will be further analyzed in Section V-C. We use the same procedure as for the synthetic signal and only change the parameters to step sizes of 130 samples

²obtained from <https://www.xeno-canto.org/>

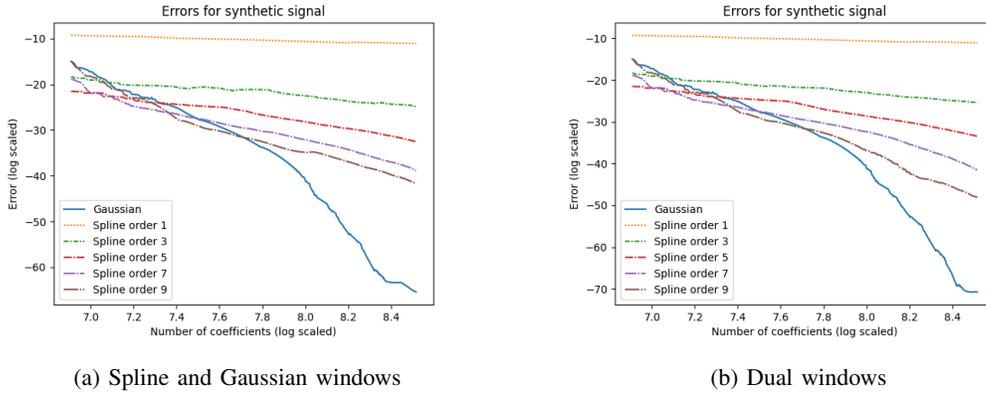


Fig. 7: Errors for best N -term approximation of synthetic signal with different window functions.

Window function	estimated p	$1/k$, k spline order	estimated p for dual window
Spline order 1	0.9318	1	0.9319
Spline order 2	0.4538	0.5	0.4481
Spline order 3	0.4170	0.3333	0.3918
Spline order 4	0.3068	0.25	0.2842
Spline order 5	0.2540	0.2	0.2336
Spline order 6	0.2066	0.1667	0.1809
Spline order 7	0.1715	0.1429	0.1509
Spline order 8	0.1366	0.125	0.1293
Spline order 9	0.1340	0.1111	0.0986
Spline order 10	0.1147	0.1	0.0925
Gaussian	0.0569	≈ 0	0.0531

TABLE II: Values of p from compressing a synthetic signal for spline and Gaussian window functions (second column) and for dual windows (forth column)

in time and 130 Hz in frequency as well as a window length of 800 samples for all splines. The compression results are visualised in Figure 8 (left).

Here, we can see the linear dependency in (20) much clearer. We also see that even the B-spline of order 2 already achieves the same compression rate as the Gaussian window function. Calculating p as before confirms this, see Table III.

We conclude that the signal itself is only in the modulation space \mathcal{M}^p for $p > 0.82$. Also, for this longer signal the time advantage is very clear: The simulations with the Gaussian window took 844 seconds or approximately 14 minutes, while the simulations with all four spline orders together took only 291 seconds or less than five minutes.

This shows how with real signals, using splines of rather low order yields the same compression rate as using smoother window functions while taking much less computation time. We should also mention that, given we always decompress the file with 500 different values for the number of coefficients N , we get real time decompression even with the Gaussian window.

Additional simulations with Gaussian windows

For the comparison of different spline windows to a Gaussian window we always used a Gaussian with a time-frequency ratio of 1. To justify this, we now look at the synthetic signal and do the same simulations using Gaussian windows with time-frequency ratio between 0.5 and 2. Figure 9 shows the results and again, we compute the parameter p

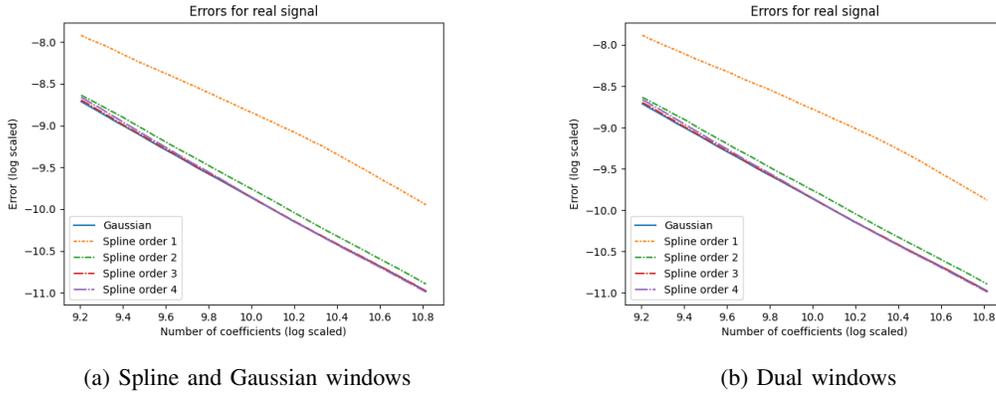


Fig. 8: Errors for best N -term approximation of blackbird recording with different window functions.

Window function	estimated p from compression rate	estimated p for dual window
Spline order 1	0.8844	0.8935
Spline order 2	0.8322	0.8319
Spline order 3	0.8317	0.8309
Spline order 4	0.8207	0.8205
Gaussian	0.8342	0.8342

TABLE III: Values of p from compressing the blackbird signal

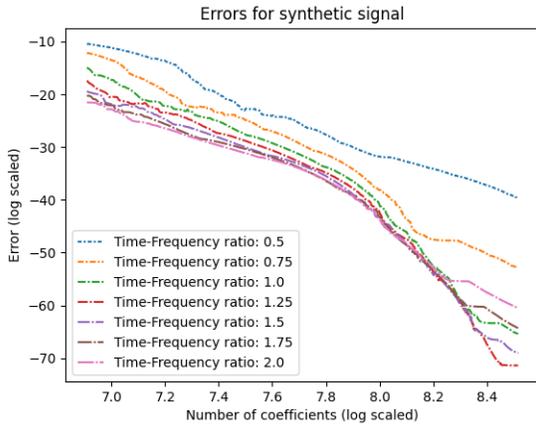


Fig. 9: Errors for best N -term approximation of blackbird recording with different window functions.

TFR of Gaussian	estimated p from compression rate
0.5	0.1031
0.75	0.0714
1	0.0569
1.25	0.0542
1.5	0.0578
1.75	0.0643
2	0.0717

TABLE IV: Values of p from compressing the synthetic signal with different Gaussians

of the modulation space from the slopes (see Table IV). As we can see, all Gaussians with a time-frequency ratio that is reasonably close to around 1.25 show very similar behaviour, rendering closer inspections of window length unnecessary for the

purpose of illustrating the result from Theorem 6.

Another thing to look at is the role of the window and the dual window in Theorem 6. We can test this by simply switching the windows for our test on the synthetic and the real signal. Figures 7 (right) and 8 (right) show the results, and the last columns in Tables 1 and III the corresponding estimates on p . As expected, the differences are quite small. In fact, with this setting we achieve slightly better compression rates on the synthetic signal with the dual window functions.

C. Denoising on real data

Finally we illustrate the reduction of additive stationary noise on three real-data examples, that is, recordings of a common blackbird as used in Section V-B, of a melody played on the piano and a sample of human speech³. All recordings are 5 seconds long with a sampling rate of 44.1 kHz. The various recordings can be found on GitHub⁴. Figure 10 gives the spectrograms with a Gaussian window function and suitable choices of a relatively dense grid and a window width resulting in 75% overlap.

Each signal is corrupted with Gaussian white noise according to a given signal-to-noise ratio (SNR). State of the art methods for such signals rely on spectral domain denoising [31]. A time-frequency representation of the noisy signal is computed and the resulting coefficients are processed by thresholding methods to attenuate the noise. For comparison we also included a wavelet- thresholding based approach, as well as the non linear median filter which is a spatial denoising method.

We performed simulations for three thresholding methods: VisuShrink, the universal threshold proposed by Donoho and Johnstone [10] which we analyzed theoretically in Section III, SureShrink which is computed by minimizing Stein’s unbiased risk estimate [9] and BayesShrink, a method based on empirical Bayes by Johnstone and Silverman [21]. Whereas VisuShrink provides a simple global threshold, SureShrink and BayesShrink deliver coefficient - dependent thresholds.

For acoustic signals, diagonal estimation methods with thresholding can introduce artificial musical noise [31], which can be ameliorated by block-thresholding methods [1]. Therefore, we also implemented block-thresholding versions of the above threshold schemes.

³obtained from <https://samplefocus.com/>

⁴[Link](https://github.com/ptafo/Statistically-optimal-estimation-of-signals-in-modulation-spaces-using-Gabor-frames) to recordings on GitHub under <https://github.com/ptafo/Statistically-optimal-estimation-of-signals-in-modulation-spaces-using-Gabor-frames>

Figure 11 displays the spectrograms of the noisy signal of the blackbird for a particular realization of the noise, and of several of the above reconstruction methods. While all the time-frequency based methods result in much clearer spectrograms as compared to the wavelet method and median filtering, Visushrink tends to produce a somewhat overly clear image appearance.

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. The results displayed in Table V indicate the superior performance of the time-frequency based methods. In terms of time-frequency based thresholding methods, Sureshrink for Gabor frames results in the best denoising performance. In particular for the blackbird signal but also for the melody played on a piano, the simple VisuShrink methods for Gabor frames which we focused on is also competitive. Block-thresholding for Gabor frames does not seem to improve noise removal for the specific signals investigated. Finally, while significantly inferior for the blackbird signal and the melody, wavelet thresholding with universal thresholding performs surprisingly well for the sample of human speech. The median filter shows the worst performance throughout.

Results comparing SureShrink and BayesShrink, both for the diagonal methods as well as for block thresholding are provided in the supplementary material.

VI. CONCLUSIONS AND OUTLOOK

We showed that for signals contained in modulation spaces, shrinkage methods for Gabor series expansions lead to optimal denoising in a white-noise model in the statistical minimax sense. While this could basically be expected, our results show interesting new phenomena in the minimax rates in par-

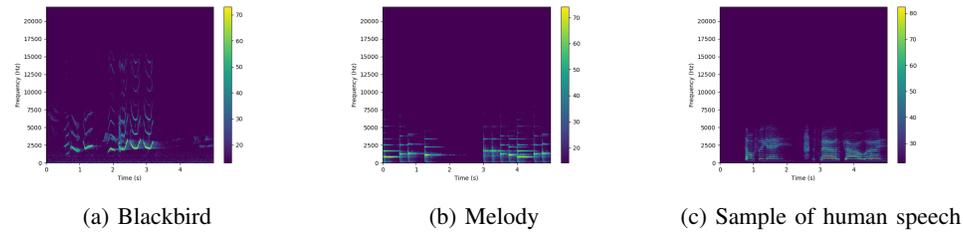


Fig. 10: Spectrograms of a recording of a common blackbird, of a melody played on the piano and of a sample of human speech

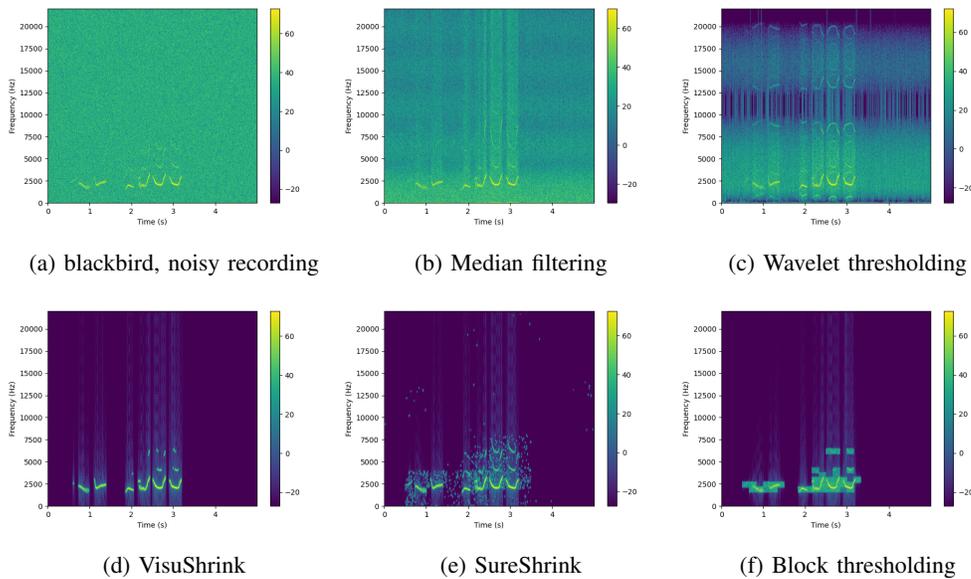


Fig. 11: Spectrograms of the noisy and denoised blackbird recordings

audio	SNR	VisuShrink	SureShrink	block th	wavelet	median filter
blackbird	-10	8.823	9.128	4.901	2.9	-2.772
	1	15.337	16.3	12.565	10.018	5.012
	10	20.785	21.973	19.274	16.559	7.705
melody	-10	6.758	8.35	5.545	3.938	-2.276
	1	14.16	15.708	13.505	11.346	7.868
	10	20.103	22.21	20.374	18.113	13.88
human speech	-10	4.075	6.576	5.531	4.096	-2.285
	1	11.894	14.281	12.739	11.195	7.994
	10	18.746	20.818	19.663	18.046	14.19

TABLE V: Comparison of the thresholding of the Gabor coefficients with the universal threshold VisuShrink, SureShrink, Time-Frequency block thresholding with SureShrink to remove musical noise, SureShrink thresholding of the wavelet coefficients and the spatial non-linear median filter. All values in decibel ($10 \log_{10} x$, dB).

ticular with an unexpected effect of the dimension. We also give rates for signal compression which cover the sparse case. Here, we highlight how much smoothness of the window function is required to

achieve optimal signal compression. In our numerical experiments we demonstrate the practical use of our methods for a range of synthetic and real acoustic signals. In the experiments, we also illustrate the advantage of compactly supported B-spline window functions in terms of computation time, which becomes relevant for large scale classification problems of acoustic signals such as bird songs [19].

Extensions of our methods to other statistical estimation problems such as density estimation or regression on compact domains should be relatively straightforward. On the methodological side of time-frequency theory, an analysis which covers α -modulation spaces [5] with a varying time and frequency resolution should be of theoretical and also of applied interest. Finally, in classification algorithms based on spectrograms, a thorough theoretical as well as numerical investigation of the roles of denoising or compression still seems to be lacking.

ACKNOWLEDGEMENTS

This work was supported by the LOEWE priority project Nature 4.0 — Sensing Biodiversity funded by the Hessian Ministry for Research and Arts, Hesse, Germany.

VII. PROOFS

A. Proofs of Proposition 2 and Theorem 3

Proof of Proposition 2. For (10), we show that for $\mu > 0$,

$$\begin{aligned} & \mathbb{E} \left[\sum_{\lambda \in \Lambda_0} |t_s(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda|^2 \right] \\ & \leq \sum_{\lambda \in \Lambda_0} \min(|\vartheta_\lambda|^2, \varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2 + 2\mu^2) \\ & \quad + (\#\Lambda_0) \frac{\varepsilon^3 \|\tilde{h}\|_{\mathcal{L}^2}^3}{\mu} \exp\left(-\frac{\mu^2}{2\varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2}\right). \end{aligned} \quad (21)$$

Inserting the universal threshold gives the result.

Since $\vartheta_\lambda := \langle f, \tilde{h}_\lambda \rangle = \langle f, \Re(\tilde{h}_\lambda) \rangle + i \langle f, \Im(\tilde{h}_\lambda) \rangle$ we have for $j \in \{s, h\}$ that

$$\begin{aligned} & \sum_{\lambda \in \Lambda_0} |t_j(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda|^2 \\ & = \sum_{\lambda \in \Lambda_0} \left((t_j(\Re(Y(\tilde{h}_\lambda)); \mu) - \langle f, \Re(\tilde{h}_\lambda) \rangle)^2 \right. \\ & \quad \left. + (t_j(\Im(Y(\tilde{h}_\lambda)); \mu) - \langle f, \Im(\tilde{h}_\lambda) \rangle)^2 \right). \end{aligned} \quad (22)$$

To bound the expected value of (22), we use the following fact, see Donoho and Johnstone [10] or Candes [2, Proof of Theorem 5.1] that for $W \sim \mathcal{N}(a, \sigma_0^2)$ we have that

$$\begin{aligned} & \mathbb{E}[(t_s(W; \mu) - a)^2] \\ & \leq \min(a^2, \sigma_0^2 + \mu^2) + 2 \frac{\sigma_0^3}{\mu} \varphi\left(\frac{\mu}{\sigma_0}\right), \end{aligned} \quad (23)$$

where φ is the density of the standard normal distribution. Applying this to (22) we obtain

$$\begin{aligned} & \mathbb{E} \left[\sum_{\lambda \in \Lambda_0} \left((t_s(\Re(Y(\tilde{h}_\lambda)); \mu) - \langle f, \Re(\tilde{h}_\lambda) \rangle)^2 \right. \right. \\ & \quad \left. \left. + (t_s(\Im(Y(\tilde{h}_\lambda)); \mu) - \langle f, \Im(\tilde{h}_\lambda) \rangle)^2 \right) \right] \\ & = \sum_{\lambda \in \Lambda_0} \mathbb{E} \left[(t_s(\Re(Y(\tilde{h}_\lambda)); \mu) - \langle f, \Re(\tilde{h}_\lambda) \rangle)^2 \right] \\ & \quad + \sum_{\lambda \in \Lambda_0} \mathbb{E} \left[(t_s(\Im(Y(\tilde{h}_\lambda)); \mu) - \langle f, \Im(\tilde{h}_\lambda) \rangle)^2 \right] \\ & \leq \sum_{\lambda \in \Lambda_0} \left(\min((\Re(\vartheta_\lambda))^2, \varepsilon^2 \|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2 + \mu^2) \right. \\ & \quad \left. + \frac{\varepsilon^3 \|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^3}{\mu} \exp\left(-\frac{\mu^2}{2\varepsilon^2 \|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2}\right) \right) \\ & \quad + \sum_{\lambda \in \Lambda_0} \left(\min((\Im(\vartheta_\lambda))^2, \varepsilon^2 \|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2 + \mu^2) \right. \\ & \quad \left. + \frac{\varepsilon^3 \|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^3}{\mu} \exp\left(-\frac{\mu^2}{2\varepsilon^2 \|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2}\right) \right) \\ & \leq \sum_{\lambda \in \Lambda_0} \min(|\vartheta_\lambda|^2, \varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2 + 2\mu^2) \\ & \quad + (\#\Lambda_0) \frac{\varepsilon^3 \|\tilde{h}\|_{\mathcal{L}^2}^3}{\mu} \exp\left(-\frac{\mu^2}{2\varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2}\right), \end{aligned}$$

which is (21).

For hard thresholding, for $W \sim \mathcal{N}(a, \sigma_0^2)$ and $\mu/\sigma_0 > 4$, with a suitable constant $C > 0$ we have that [20, Proposition 8.1]

$$\begin{aligned} & \mathbb{E}[(t_h(W; \mu) - a)^2] \\ & \leq C \min(a^2, \mu^2) + C \sigma_0 \mu \varphi(\mu/\sigma_0 - 1). \end{aligned} \quad (24)$$

Proceeding from (22), since

$$\Re(Y(\tilde{h}_\lambda)) \sim \mathcal{N}(\Re(\vartheta_\lambda), \varepsilon^2 \|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2)$$

we have that

$$\begin{aligned}
& \mathbb{E} \left[\sum_{\lambda \in \Lambda_0} |t_h(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda|^2 \right] \\
& \leq C \sum_{\lambda \in \Lambda_0} \left(\min \left((\Re(\vartheta_\lambda))^2, \mu^2 \right) \right. \\
& \quad \left. + \min \left((\Im(\vartheta_\lambda))^2, \mu^2 \right) \right) \\
& + C (\#\Lambda_0) \varepsilon \mu \cdot \left(\|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2} \varphi \left(\frac{\mu}{\varepsilon \|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}} - 1 \right) \right. \\
& \quad \left. + \|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2} \varphi \left(\frac{\mu}{\varepsilon \|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}} - 1 \right) \right) \\
& \leq C \sum_{\lambda \in \Lambda_0} \min \left(|\vartheta_\lambda|^2, 2\mu^2 \right) \\
& \quad + 2C\varepsilon (\#\Lambda_0) \|\tilde{h}\|_{\mathcal{L}^2} \mu \varphi \left(\frac{\mu}{\varepsilon \|\tilde{h}\|_{\mathcal{L}^2}} - 1 \right).
\end{aligned}$$

and the bound for hard thresholding follows by inserting the universal threshold. \square

Proof of Theorem 3. We may bound the second term in the oracle inequality (10) for functions in the modulation space associated to the weight function in (13) as follows.

Lemma 8. *For $f \in \mathcal{M}_{m_s}^p(\mathbb{R}^d)$, $p \in (0, 2]$ we have the bound*

$$\begin{aligned}
& \sum_{\|\lambda\|_2 \leq K} \min(\varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2, |\vartheta_\lambda|^2) \\
& \leq \text{const.} \cdot \|f\|_{\mathcal{M}_{m_s}^p}^{\frac{2d}{2d+sp}} \varepsilon^{\frac{2d(2-p)+2sp}{2d+sp}}.
\end{aligned}$$

Proof of Lemma 8. Note that we have $\|\tilde{h}\|_{\mathcal{L}^2}^2 \leq 1$, for notational convenience assume that $\|\tilde{h}\|_{\mathcal{L}^2}^2 = 1$. Since $p \in (0, 2]$ have that

$$\min(\varepsilon^2, |\vartheta_\lambda|^2) \leq \varepsilon^{2-p} |\vartheta_\lambda|^p.$$

Note that

$$\#\Lambda_{0,K} \leq C K^{2d},$$

where the constant C depends on α, β and d . Then for $0 < K_s \leq K$ we obtain

$$\begin{aligned}
& \sum_{\|\lambda\|_2 \leq K} \min(\varepsilon^2, |\vartheta_\lambda|^2) \\
& = \sum_{\|\lambda\|_2 \leq K_s} \min(\varepsilon^2, |\vartheta_\lambda|^2) + \sum_{K_s < \|\lambda\|_2 \leq K} \min(\varepsilon^2, |\vartheta_\lambda|^2) \\
& \leq \varepsilon^2 C K_s^{2d} + \varepsilon^{2-p} \sum_{K_s < \|\lambda\|_2} |\vartheta_\lambda|^p \\
& \leq \varepsilon^2 C K_s^{2d} + \varepsilon^{2-p} (1 + K_s^2)^{-sp/2} \sum_{K_s < \|\lambda\|_2} m_s(\lambda)^p |\vartheta_\lambda|^p \\
& \leq \varepsilon^2 C K_s^{2d} + \varepsilon^{2-p} K_s^{-sp} \tilde{C}_2^p \|f\|_{\mathcal{M}_{m_s}^p}^p
\end{aligned} \tag{25}$$

by using $m_s(\lambda)^p > (1 + K_s^2)^{sp/2}$ for $K_s < \|\lambda\|_2$ in the second last step, and (7) in the last step. Balancing both terms using

$$K_s = \text{const.} \cdot \|f\|_{\mathcal{M}_{m_s}^p}^{\frac{p}{2d+sp}} \varepsilon^{-\frac{p}{2d+sp}}$$

we get the result. \square

Now let us estimate the truncation term in (12).

Lemma 9. *For $f \in \mathcal{M}_{m_s}^p(\mathbb{R}^d)$, $p \in (0, 2]$ we have that*

$$\sum_{K < \|\lambda\|_2} |\vartheta_\lambda|^2 \leq K^{-2s} \tilde{C}^2 \|f\|_{\mathcal{M}_{m_s}^p}^2.$$

Proof of Lemma 9. Since the p -norm is monotonously decreasing, we have

$$\begin{aligned}
& \sum_{K < \|\lambda\|_2} |\vartheta_\lambda|^2 \\
& \leq \left(\sum_{K < \|\lambda\|_2} |\vartheta_\lambda|^p \right)^{2/p} \\
& \leq \left((1 + K^2)^{-sp/2} \sum_{K < \|\lambda\|_2} m_s(\lambda)^p |\vartheta_\lambda|^p \right)^{2/p}.
\end{aligned}$$

Applying (7), we obtain the claim. \square

To conclude the proof of the theorem we apply the expected value to (12). The first term is bounded by using the oracle inequality (10) resp. its version for hard thresholding, together with Lemma 8. The truncation error in (12) is estimated by using Lemma 9 together with the assumption $K \gtrsim \varepsilon^{\frac{d(2-p)+sp}{2d+sp}}$ in the theorem. \square

B. Proof of Theorem 4

Proof of Theorem 4. The proof relies on the lower bound derived from Fano's lemma, see Tsybakov [29, Theorem 2.5]. For $m \in \mathbb{N}$ of order $m \asymp \varepsilon^{-\frac{p}{ps+2d}}$ the task is to construct $M \in \mathbb{N}$ test functions $f_j \in \mathcal{M}_{s,C}^p$, where $M \asymp \exp(cm^{2d})$ for some $c > 0$, such that

$$\|f_j - f_k\|_{\mathcal{L}^2}^2 \asymp \varepsilon^2 m^{2d}, \quad j \neq k. \tag{26}$$

Then, if $Y^{(j)}$ and $Y^{(k)}$ are the observations in model (2) with $f = f_j$ and $f = f_k$, respectively, for the Kullback-Leibler divergence $\text{KL}(Y^{(j)}, Y^{(k)})$ we have that

$$\text{KL}(Y^{(j)}, Y^{(k)}) = \frac{\|f_j - f_k\|_{\mathcal{L}^2}^2}{\varepsilon^2} \lesssim \log(M),$$

so that by applying Tsybakov [29, Theorem 2.5] yields the lower bound of order $\varepsilon^2 m^{2d} = \varepsilon^{\frac{2d(2-p)+2sp}{2d+sp}}$, which is the statement of the theorem. Note that we need the upper bound in (26) for bounding the Kullback-Leibler divergence, and the lower bound for obtaining the rate.

To construct the test functions, we consider the Gaussian function

$$\varphi(x) = \exp(-\pi \|x\|_2^2).$$

Given $m \in \mathbb{N}$, using the Varshamov-Gilbert bound [29, Lemma 2.9] we may choose $M = \exp(c_1 m^{2d})$, $c_1 = (\log 2)/8 > 0$ indices

$$\iota = (\iota_{(k_1, \dots, k_{2d})})_{k_1, \dots, k_{2d}=1, \dots, m} \in \{0, 1\}^{(m^{2d})}$$

of Hamming - distance $\|\iota - \tilde{\iota}\|_2^2 \geq m^{2d}/8$. Note that given two vectors $\iota, \tilde{\iota} \in \{0, 1\}^{(m^{2d})}$, the Euclidean distance $\|\iota - \tilde{\iota}\|_2^2$ determines at how many positions ι and $\tilde{\iota}$ differ, that is, reduces to the Hamming distance. Recall the translation and modulation operators

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

$$M_\omega f(t) = \exp(2\pi i \langle \omega, t \rangle) f(t), \quad x, \omega, t \in \mathbb{R}^d.$$

Then, for (large) fixed $r > 0$ and (small) $c > 0$ we let

$$\begin{aligned} f_\iota &= c \cdot \varepsilon \sum_{k_1, \dots, k_{2d}=1}^m \left(\iota_{(k_1, \dots, k_{2d})}^\top \right. \\ &\quad \cdot T_{r(k_1, \dots, k_d)}^\top M_{r(k_{d+1}, \dots, k_{2d})}^\top \varphi \Big) \quad (27) \\ &= c \cdot \varepsilon \sum_{k', k''}^m \iota_{k', k''} T_{r k'} M_{r k''} \varphi, \end{aligned}$$

where we denote $k' = (k_1, \dots, k_d)^\top$ and $k'' = (k_{d+1}, \dots, k_{2d})^\top$, and $\sum_{k', k''}^m$ indicates that the coordinates in k' and k'' vary from 0 up to m .

Lemma 10. *We have that*

$$\begin{aligned} &(V_\varphi f_\iota)(x, \omega) \\ &= c \cdot \varepsilon \sum_{k', k''} \left(\iota_{k', k''} \exp(-2\pi i \langle \omega, r k' \rangle) 2^{-d/2} \right. \\ &\quad \cdot \exp(\pi i \langle x - r k', \omega - r k'' \rangle) \quad (28) \\ &\quad \cdot \exp(-\pi \|x - r k'\|_2^2 / 2) \exp(-\pi \|\omega - r k''\|_2^2 / 2) \Big). \end{aligned}$$

Proof of Lemma 10. The short time Fourier transform of the Gaussian function φ can be computed as

$$\begin{aligned} &V_\varphi(\varphi)(x, \omega) \\ &= 2^{-d/2} \exp(-\pi \|x\|_2^2 / 2) \exp(\pi i \langle x, \omega \rangle) \\ &\quad \cdot \exp(-\pi \|\omega\|_2^2 / 2). \end{aligned}$$

Moreover, from the relation $T_x M_\omega = \exp(-2\pi i \langle x, \omega \rangle) M_\omega T_x$, from

$$\begin{aligned} \langle T_x f, g \rangle_{\mathcal{L}^2} &= \langle f, T_{-x} g \rangle_{\mathcal{L}^2}, \\ \langle M_\omega f, g \rangle_{\mathcal{L}^2} &= \langle f, M_{-\omega} g \rangle_{\mathcal{L}^2} \end{aligned}$$

and from the representation $V_g f(x, \omega) = \langle f, M_\omega T_x g \rangle_{\mathcal{L}^2}$ we obtain

$$\begin{aligned} (V_g(T_s f))(x, \omega) &= \exp(-2\pi i \langle \omega, s \rangle) \\ &\quad \cdot (V_g f)(x - s, \omega), \\ (V_g(M_\eta f))(x, \omega) &= (V_g f)(x, \omega - \eta). \end{aligned}$$

Using these formulas and the linearity of the STFT gives (28). \square

Bounding the \mathcal{L}_2 -distance

Since $\|V_g f\|_{\mathcal{L}^2} = \|f\|_{\mathcal{L}^2} \|g\|_{\mathcal{L}^2}$ and $\|\varphi\|_{\mathcal{L}^2}^2 = 2^{-d/2}$, we obtain

$$\|f_\iota - f_{\iota'}\|_{\mathcal{L}^2}^2 = 2^{-d/2} \|V_\varphi f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2, \quad (29)$$

where $\iota, \iota' \in \{0, 1\}^{(m^{2d})}$.

Lemma 11. *For a sufficiently large (fixed) $r > 0$ and for ι and ι' of Hamming distance $\|\iota - \tilde{\iota}\|_2^2 \geq m^{2d}/8$ we have for constants $\tilde{c}_i > 0$ that*

$$\begin{aligned} \|V_\varphi f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2 &\geq \tilde{c}_1 \varepsilon^2 \|\iota - \iota'\|_2^2 \geq \tilde{c}_1 \varepsilon^2 m^{2d}/8, \quad (30) \\ \|V_\varphi f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2 &\leq \tilde{c}_2 \varepsilon^2 \|\iota - \iota'\|_2^2 \leq \tilde{c}_2 \varepsilon^2 m^{2d}. \quad (31) \end{aligned}$$

Proof of Lemma 11. First consider the lower bound (30). We have that

$$\begin{aligned} &|(V_\varphi f_\iota) - (V_\varphi f_{\iota'})|^2(x, \omega) \\ &\geq c^2 \varepsilon^2 2^{-d} \left(\sum_{k', k''}^m (\iota_{k', k''} - \iota'_{k', k''})^2 \right. \quad (32) \\ &\quad \cdot \exp(-\pi \|x - r k'\|_2^2) \exp(-\pi \|\omega - r k''\|_2^2) \\ &\quad - \sum_{\substack{k', k'' \neq (k'_1, k''_1) \\ (k', k'') \neq (k'_1, k''_1)}}^m |\iota_{k', k''} - \iota'_{k', k''}| |\iota_{k'_1, k''_1} - \iota'_{k'_1, k''_1}| \\ &\quad \cdot \exp(-\frac{\pi}{2} (\|x - r k'\|_2^2 + \|x - r k'_1\|_2^2)) \\ &\quad \cdot \exp(-\frac{\pi}{2} (\|\omega - r k''\|_2^2 + \|\omega - r k''_1\|_2^2)) \Big). \end{aligned}$$

Lemma 12. *Given $a, r > 0$ with $a^{1/2} \cdot r$ sufficiently large we have that for all $x, \omega \in \mathbb{R}^d$,*

$$\begin{aligned} &\sum_{(k', k'') \neq (0,0)} \left(\exp(-a (\|x\|_2^2 + \|x - r k'\|_2^2)) \right. \\ &\quad \cdot \exp(-a (\|\omega\|_2^2 + \|\omega - r k''\|_2^2)) \Big) \quad (33) \\ &\leq 4 \exp(-a \frac{r^2}{8}) \exp(-a \frac{\|x\|_2^2 + \|\omega\|_2^2}{2}). \end{aligned}$$

Here, we let the indices k' and k'' in the sum range through \mathbb{Z}^d .

Proof of Lemma 12. By rescaling we may assume that $a = 1$. The left side then reduces to

$$\begin{aligned} &e^{-(\|x\|_2^2 + \|\omega\|_2^2)} \sum_{(k', k'') \neq (0,0)} \left(\exp(-\|x - r k'\|_2^2) \right. \\ &\quad \cdot \exp(-\|\omega - r k''\|_2^2) \Big). \quad (34) \end{aligned}$$

Of course,

$$\begin{aligned} & \sum_{(k', k'') \neq (0, 0)} \exp(-\|x - r k'\|_2^2) \exp(-\|\omega - r k''\|_2^2) \\ & \leq \sum_{(k', k'')} \exp(-\|x - r k'\|_2^2) \exp(-\|\omega - r k''\|_2^2), \end{aligned}$$

and we bound the sum on the right side uniformly in x and ω . Indeed, since the function on the right side has period r in each coordinate of x and ω , it suffices to bound it on $[-r/2, r/2]^{2d}$, and we can bound

$$\begin{aligned} & \sum_{(k', k'')} \exp(-\|x - r k'\|_2^2) \exp(-\|\omega - r k''\|_2^2) \\ & \leq \sum_{(k', k'')} \exp(-r^2 \|k'\|_2^2/4) \exp(-r^2 \|k''\|_2^2/4) \\ & \leq 2, \end{aligned} \quad (35)$$

for sufficiently large r .

Now, concerning (34) if $\|x\|_2 \geq r/2$ then $\|x\|_2^2 \geq \|x\|_2^2/2 + r^2/8$ and hence $e^{-(\|x\|_2^2 + \|\omega\|_2^2)} \leq e^{-r^2/8} e^{-(\|x\|_2^2 + \|\omega\|_2^2)/2}$, which together with (35) implies (33). The case $|\omega| \geq r/2$ is similar.

If both $\|x\|_2 \leq r/2$ and $\|\omega\|_2 \leq r/2$, then if $k' \neq 0$,

$$\begin{aligned} \|x - r k'\|_2^2 & \geq (r \|k'\|_2 - \|x\|_2)^2 \\ & \geq r^2 (\|k'\|_2 - 1/2)^2 \\ & \geq r^2 (\|k'\|_2^2/2 - 1/4), \end{aligned}$$

and hence in view of (35), and similarly for $k'' \neq 0$, and hence we can bound

$$\begin{aligned} & \sum_{(k', k'') \neq (0, 0)} \exp(-\|x - r k'\|_2^2) \exp(-\|\omega - r k''\|_2^2) \\ & \leq 4 \exp(-r^2/4). \end{aligned}$$

□

We resume the proof of Lemma 11. Using (33) we may provide an upper bound on the second term in the bracket in (32) (which is subtracted) by

$$\begin{aligned} & \sum_{(k', k'') \neq (k'_1, k''_1)}^m \left(|l_{k', k''} - l'_{k', k''}| |l_{k'_1, k''_1} - l'_{k'_1, k''_1}| \right. \\ & \quad \cdot \exp(-\frac{\pi}{2} (\|x - r k'\|_2^2 + \|x - r k'_1\|_2^2)) \\ & \quad \cdot \exp(-\frac{\pi}{2} (\|\omega - r k''\|_2^2 + \|\omega - r k''_1\|_2^2)) \left. \right) \\ & \leq 4 \exp(-\pi r^2/16) \sum_{(k', k'')} \left((l_{k', k''} - l'_{k', k''})^2 \right. \\ & \quad \cdot \exp(-\frac{\pi}{4} \|x - r k'\|_2^2) \exp(-\frac{\pi}{2} \|\omega - r k''\|_2^2) \left. \right). \end{aligned}$$

Thus, we can provide a lower bound on the difference in brackets in (32) by

$$\begin{aligned} & \sum_{(k', k'')} (l_{k', k''} - l'_{k', k''})^2 \cdot \left(\exp(-\pi \|x - r k'\|_2^2) \right. \\ & \quad \cdot \exp(-\pi \|\omega - r k''\|_2^2) - 4 \exp(-\pi r^2/16) \\ & \quad \cdot \exp(-\frac{\pi}{4} \|x - r k'\|_2^2) \exp(-\frac{\pi}{2} \|\omega - r k''\|_2^2) \left. \right). \end{aligned}$$

Now, the integrals of $\exp(-\pi \|x - r k'\|_2^2)$, $\exp(-\pi \|\omega - r k''\|_2^2)$ and $\exp(-\frac{\pi}{4} \|x - r k'\|_2^2)$, $\exp(-\frac{\pi}{2} \|\omega - r k''\|_2^2)$ are positive and do not depend on r . Thus, by choosing r large enough we obtain the desired first inequality in (30), while the second follows from the assumption on ι and ι' .

For the upper bound, we obtain similarly that

$$\begin{aligned} & \sum_{(k', k'')} (l_{k', k''} - l'_{k', k''})^2 \cdot \left(\exp(-\pi \|x - r k'\|_2^2) \right. \\ & \quad \cdot \exp(-\pi \|\omega - r k''\|_2^2) + 4 \exp(-\pi r^2/16) \\ & \quad \cdot \exp(-\frac{\pi}{4} \|x - r k'\|_2^2) \exp(-\frac{\pi}{2} \|\omega - r k''\|_2^2) \left. \right), \end{aligned}$$

which yields the first inequality in (31), while the second follows from $\|\iota - \iota'\|_2^2 \leq m^{2d}$. □

Bounding the modulation norm

Lemma 13. *For the functions defined in (27), we have that*

$$\|f_\iota\|_{\mathcal{M}_{m_s}^p}^p \leq \text{const.} \cdot m^{ps+2d} \epsilon^p,$$

where the constant can be made small by decreasing the constant c in (27).

Proof. We have that

$$\begin{aligned} & \|f_\iota\|_{\mathcal{M}_{m_s}^p}^p \\ & = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_\varphi f_\iota(x, \omega)|^p m_s(x, \omega)^p dx d\omega \\ & \leq c^p \epsilon^p \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\sum_{k', k''}^m \exp(-\pi \|x - r k'\|_2^2/2) \right. \\ & \quad \cdot \exp(-\pi \|\omega - r k''\|_2^2/2) \left. \right)^p \\ & \quad \cdot (1 + \|x\|_2^2 + \|\omega\|_2^2)^{ps/2} dx d\omega. \end{aligned} \quad (36)$$

For $a, r > 0$ with $a^{1/2} \cdot r$ sufficiently large we have for all $t \in \mathbb{R}$ that

$$\begin{aligned} & \sum_{j=1}^m \exp(-a(t-rj)^2) \\ & \leq 4 \left(\sum_{j=2}^{m-1} \exp(-a(t-rj)^2) \mathbf{1}_{|t-rj| \leq r/2} \right. \\ & \quad + \exp(-a(t-r)^2) \mathbf{1}_{t \leq 3r/2} \\ & \quad \left. + \exp(-a(t-rm)^2) \mathbf{1}_{t \geq (m-1/2)r} \right). \end{aligned} \quad (37)$$

This follows from

$$\exp(-at^2) \geq 2 \sum_{j=1}^{\infty} \exp(-a(t-rj)^2), \quad t \leq r/2,$$

for $a^{1/2} \cdot r$ sufficiently large, which is immediate from the geometric series. The functions on the right of (37) have disjoint support, so that

$$\begin{aligned} & \left(\sum_{j=1}^m \exp(-a(t-rj)^2) \right)^p \\ & \leq 4^p \left(\sum_{j=2}^{m-1} \exp(-ap(t-rj)^2) \mathbf{1}_{|t-rj| \leq r/2} \right. \\ & \quad \left. + \exp(-ap(t-r)^2) \mathbf{1}_{t \leq 3r/2} \right. \\ & \quad \left. + \exp(-ap(t-rm)^2) \mathbf{1}_{t \geq (m-1/2)r} \right) \\ & \leq 4^p \sum_{j=1}^m \exp(-ap(t-rj)^2), \end{aligned}$$

and inserting this bound in each of the sums in (36) we bound the integral by

$$\begin{aligned} & \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\sum_{k', k''}^m \exp(-\pi \|x - rk'\|_2^2 / 2) \right. \\ & \quad \left. \cdot \exp(-\pi \|\omega - rk''\|_2^2 / 2) \right)^p \\ & \quad \cdot (1 + \|x\|_2^2 + \|\omega\|_2^2)^{ps/2} \, dx \, d\omega \\ & \leq \text{const.} \cdot \sum_{k', k''}^m \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \exp(-p\pi \|x - rk'\|_2^2 / 2) \\ & \quad \cdot \exp(-p\pi \|\omega - rk''\|_2^2 / 2) \\ & \quad \cdot (1 + |x_1^{ps}| + \dots + |x_d^{ps}| + |\omega_1^{ps}| + \dots + |\omega_d^{ps}|) \\ & \quad dx \, d\omega \\ & \leq \text{const.} \cdot \sum_{k', k''}^m \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \exp(-p\pi \|x - rk'\|_2^2 / 2) \\ & \quad \cdot \exp(-p\pi \|\omega - rk''\|_2^2 / 2) (1 + \|x\|_2^2 + \|\omega\|_2^2)^{ps/2} \\ & \quad dx \, d\omega \tag{38} \end{aligned}$$

Since for $b \geq 0$,

$$\int_{\mathbb{R}} \exp(-pt^2) |t+b|^{ps} \, dx \leq \text{const.} \cdot |b|^{ps},$$

we may further bound (38) by

$$\begin{aligned} & \text{const.} \cdot \sum_{k_1, \dots, k_{2d}=1}^m (1 + k_1^{ps} + \dots + k_{2d}^{ps}) \\ & \leq \text{const.} \cdot m^{2d-1} \cdot m^{ps+1} \\ & \leq \text{const.} \cdot m^{ps+2d}, \end{aligned}$$

which together with (36) proves the lemma. \square

Conclusion of the proof of Theorem 4. From the above paragraph we obtain as condition for $f_v \in \mathcal{M}_{s,C}^p$ in $d = 1$ that $m^{ps+2d} \varepsilon^p \lesssim 1$, which leads to $m \asymp \varepsilon^{-\frac{p}{ps+2d}}$.

Now (30) implies (26), and the conclusion follows as in the first paragraph of the proof. \square

As pointed out by a reviewer, the construction (27) can potentially be related to the so-called Piano-Reconstruction Theorem, Theorem 6.4 in Feichtinger and Gröchenig [15], which would indicate that computations could be transferred to the level of sequence spaces.

C. Proof of Theorem 5

Proof of Theorem 5. Let us first consider the upper risk bound. We state an extension of Lemma 8.

Lemma 14. For $f \in \mathcal{M}_{m_u, v}^p(\mathbb{R}^d)$, $p \in (0, 2]$ we have the bound

$$\begin{aligned} & \sum_{\|\lambda\|_2 \leq K} \min(\varepsilon^2 \|\tilde{h}\|_{\mathcal{L}^2}^2, |\vartheta_\lambda|^2) \\ & \leq \text{const.} \cdot \|f\|_{\mathcal{M}_{m_u, v}^p}^{\frac{pd(v+u)}{d(v+u)+pvu}} \varepsilon^{\frac{(2-p)d(v+u)+2pvu}{d(v+u)+pvu}}. \end{aligned} \tag{39}$$

Proof of Lemma 14. Again assume $\|\tilde{h}\|_{\mathcal{L}^2}^2 = 1$ and use that $\min(\varepsilon^2, |\vartheta_\lambda|^2) \leq \varepsilon^{2-p} |\vartheta_\lambda|^p$. For $K_u > 0$ and $K_v > 0$ with a computation similar to (25) we have that

$$\begin{aligned} & \sum_{\|\lambda\|_2 \leq K} \min(\varepsilon^2, |\vartheta_\lambda|^2) \\ & \leq \sum_{\|x\|_2 \leq K_u, \|\omega\|_2 \leq K_v} \min(\varepsilon^2, |\vartheta_\lambda|^2) \\ & \quad + \sum_{\|x\|_2 > K_u} \min(\varepsilon^2, |\vartheta_\lambda|^2) \\ & \quad + \sum_{\|\omega\|_2 > K_v} \min(\varepsilon^2, |\vartheta_\lambda|^2) \\ & \leq \varepsilon^2 C K_v^d K_u^d \\ & \quad + \tilde{C}_2^p \|f\|_{\mathcal{M}_{m_u, v}^p}^p \varepsilon^{2-p} (K_u^{-up} + K_v^{-vp}) \end{aligned}$$

Balancing the terms with

$$\begin{aligned} K_v &= \text{const.} \cdot \|f\|_{\mathcal{M}_{m_u, v}^p}^{\frac{pu}{d(v+u)+pvu}} \varepsilon^{-\frac{pu}{d(v+u)+pvu}}, \\ K_u &= \text{const.} \cdot \|f\|_{\mathcal{M}_{m_u, v}^p}^{\frac{pv}{d(v+u)+pvu}} \varepsilon^{-\frac{pv}{d(v+u)+pvu}}, \end{aligned}$$

we get the result. \square

Next we extend Lemma 9. The proof is analogous and therefore omitted.

Lemma 15. For $f \in \mathcal{M}_{m_u, v}^p(\mathbb{R}^d)$, $p \in (0, 2]$ we have that $\sum_{K < \|\lambda\|_2} |\vartheta_\lambda|^2 \leq K^{-2 \min(u, v)} \tilde{C}^2 \|f\|_{\mathcal{M}_{m_u, v}^p}^2$.

These combine to give the upper risk bound in Theorem 5.

Now let us turn to the lower risk bound. The proof is along the lines of that of Theorem 4. For $m_1, m_2 \in \mathbb{N}$

of order $m_1 \asymp \varepsilon^{-\frac{pv}{d(v+u)+pvu}}$ and $m_2 \asymp \varepsilon^{-\frac{pu}{d(v+u)+pvu}}$, we construct $M \in \mathbb{N}$ test functions $f_j \in \mathcal{M}_{u,v,C}^p$, where $M \asymp \exp(c m_1^d m_2^d)$ for some $c > 0$, such that

$$\|f_j - f_k\|_{\mathcal{L}^2}^2 \asymp \varepsilon^2 m_1^d m_2^d, \quad j \neq k.$$

The conclusion then follows as in the proof of Theorem 4.

Again we consider the Gaussian function $\varphi(x) = \exp(-\pi \|x\|_2^2)$, and given $m_1, m_2 \in \mathbb{N}$, using the Varshamov-Gilbert bound [29, Lemma 2.9] we may choose $M = \exp(c_1 m_1^d m_2^d)$, $c_1 = (\log 2)/8 > 0$ indices

$$\begin{aligned} \iota &= (\iota_{(k_1, \dots, k_{2d})})_{k_1, \dots, k_d=1, \dots, m_1, k_{d+1}, \dots, k_{2d}=1, \dots, m_2} \\ &\in \{0, 1\}^{(m_1^d m_2^d)} \end{aligned}$$

of Hamming - distance $\|\iota - \tilde{\iota}\|_2^2 \geq m_1^d m_2^d / 8$. For suitable (large) fixed $r > 0$ and some (small) fixed $c > 0$ we let

$$\begin{aligned} f_\iota &= c \cdot \varepsilon \sum_{k_1, \dots, k_d=1}^{m_1} \sum_{k_{d+1}, \dots, k_{2d}=1}^{m_2} \left(\iota_{(k_1, \dots, k_{2d})}^\top \right. \\ &\quad \left. \cdot T_r(k_1, \dots, k_d)^\top M_r(k_{d+1}, \dots, k_{2d})^\top \varphi \right) \quad (40) \\ &= c \cdot \varepsilon \sum_{k', k''} \iota_{k', k''} T_r k' M_r k'' \varphi, \end{aligned}$$

where we denote $k' = (k_1, \dots, k_d)^\top$ and $k'' = (k_{d+1}, \dots, k_{2d})^\top$.

Bounding the \mathcal{L}_2 -distance

Analogously to Lemma 11, we have that

Lemma 16. *For a sufficiently large (fixed) $r > 0$ and for ι and ι' of Hamming distance $\|\iota - \iota'\|_2^2 \geq m_1^d m_2^d / 8$ we have that*

$$\begin{aligned} \|V_\varphi f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2 &\geq \tilde{c}_1 \varepsilon^2 \|\iota - \iota'\|_2^2 \geq \tilde{c}_1 \varepsilon^2 m_1^d m_2^d / 8, \\ \|V_\varphi f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2 &\leq \tilde{c}_2 \varepsilon^2 \|\iota - \iota'\|_2^2 \leq \tilde{c}_2 \varepsilon^2 m_1^d m_2^d. \end{aligned} \quad (41)$$

Bounding the modulation norm

Lemma 17. *For the functions defined in (40), we have that*

$$\|f_\iota\|_{\mathcal{M}_{m_u, v}^p}^p \leq \text{const.} \cdot \varepsilon^p (m_1^d m_2^d) (m_1^{pu} + m_2^{pv}),$$

where the constant can be made small by decreasing the constant c in (40).

For the choices m_1 and m_2 above, the upper bound in Lemma 17 remains bounded, and inserting these choices in (41) gives the rate. \square

D. Proofs of Section IV

Proof of Theorem 6. Since $(\langle f, \tilde{h}_\lambda \rangle)_{\lambda \in \Lambda} \in \ell_m^p$ we have that N_μ is finite. By definition of I_μ we have that

$$\begin{aligned} N_\mu \mu^p &\leq \sum_{\lambda \in I_\mu} |\langle f, \tilde{h}_\lambda \rangle|^p \\ &\leq \sum_{\lambda \in \Lambda} |\langle f, \tilde{h}_\lambda \rangle|^p \\ &\leq \tilde{C}_2^p \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}^p, \end{aligned}$$

where we used (7) in the last step. Hence

$$\mu \leq N_\mu^{-1/p} \tilde{C}_2 \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}. \quad (42)$$

Further, for $\lambda \in I_\mu^c$ we have that

$$|\langle f, \tilde{h}_\lambda \rangle|^2 \leq \mu^{2-p} |\langle f, \tilde{h}_\lambda \rangle|^p \quad (43)$$

since $p \in (0, 2)$. Hence by boundedness of the synthesis operator

$$\begin{aligned} \|f - f_\mu\|_{\mathcal{L}^2}^2 &\leq B \sum_{\lambda \in I_\mu^c} |\langle f, \tilde{h}_\lambda \rangle|^2 \\ &\leq B \mu^{2-p} \sum_{\lambda \in I_\mu^c} |\langle f, \tilde{h}_\lambda \rangle|^p \end{aligned}$$

Using (7) we obtain

$$\|f - f_\mu\|_{\mathcal{L}^2}^2 \leq B \tilde{C}_2^p \mu^{2-p} \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}^p$$

and inserting the bound (42) yields the result. \square

Proof of Theorem 7. From (5) we have the bound

$$\|g\|_{\mathcal{M}_m^p(\mathbb{R}^d)}^p \leq \tilde{C} \cdot \sum_{\lambda \in \Lambda} |c_\lambda|^p m(\lambda)^p.$$

Further we may write

$$\begin{aligned} \|f - g\|_{\mathcal{L}^2}^2 &= \left\| \sum_{\lambda \in \Lambda} (\langle f, \tilde{h}_\lambda \rangle - c_\lambda) \tilde{h}_\lambda \right\|_{\mathcal{L}^2}^2 \\ &= \|D_h(\langle f, \tilde{h}_\lambda \rangle - c_\lambda)_{\lambda \in \Lambda}\|_{\mathcal{L}^2}^2, \end{aligned}$$

where D_h is the Gabor synthesis operator associated with $\{\tilde{h}_\lambda \mid \lambda \in \Lambda\}$. From Gröchenig [17, Prop. 5.1.1 (b)] we obtain the estimate

$$\|D_h(\langle f, \tilde{h}_\lambda \rangle - c_\lambda)_{\lambda \in \Lambda}\|_{\mathcal{L}^2}^2 \leq B \|(\langle f, \tilde{h}_\lambda \rangle - c_\lambda)_{\lambda \in \Lambda}\|_{\mathcal{L}^2}^2,$$

where B is the upper frame bound. Setting $C = \max(\tilde{C}, B)$ we obtain the result. \square

E. Proof of Theorem 1

Proof. Following Galperin and Samarah [16, Theorem 3.1], it is sufficient to show

$$h \in \bigcup_{\substack{r,s>1/p \\ 1 \leq p^* < \infty}} M_{w_{r,s}}^{p^*}$$

with

$$w_{r,s}(x, \omega) = v(x, \omega)(1 + |x|)^r(1 + |\omega|)^s.$$

We therefore need $V_h h \in \mathcal{L}_{w_{r,s}}^{p^*}$. Since [see, e.g., 17, Lemma 3.1.1]

$$V_g f(x, \omega) = e^{-2\pi i x \omega} V_g \hat{f}(\omega, -x),$$

we have

$$\begin{aligned} & |V_h h(x, \omega)| \\ &= \left| V_{\hat{h}} \hat{h}(\omega, -x) \right| \\ &= \left| \int_{\mathbb{R}} \hat{h}(t) \overline{\hat{h}(t - \omega)} e^{2\pi i x t} dt \right| \\ &= \left| \int_{\mathbb{R}} e^{-it/2 + i(t-\omega)/2} \left(\frac{\sin(t/2) \cdot \sin((t-\omega)/2)}{t/2 \cdot (t-\omega)/2} \right)^k e^{2\pi i x t} dt \right|. \end{aligned}$$

By defining

$$f_\omega(t) = \left(\frac{\sin(t/2) \cdot \sin((t-\omega)/2)}{t/2 \cdot (t-\omega)/2} \right)^k,$$

we can conclude that

$$\begin{aligned} & \|V_h h\|_{\mathcal{L}_{w_{r,s}}^{p^*}}^{p^*} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} |V_h h(x, \omega)|^{p^*} w_{r,s}(x, \omega)^{p^*} dx d\omega \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \hat{f}_\omega(-x) \right|^{p^*} w_{r,s}(x, \omega)^{p^*} dx d\omega \\ &= \int_{\mathbb{R}} (1 + |\omega|)^{s p^*} \\ & \quad \cdot \left(\int_{\mathbb{R}} \left| \hat{f}_\omega(x) \right|^{p^*} v(-x, \omega)(1 + |x|)^{r p^*} dx \right) d\omega. \end{aligned}$$

The sinc function $\sin(t)/t$ is bandlimited and its translation in time corresponds to a modulation of its Fourier transform, which keeps its support. Therefore, f_ω is bandlimited and the support of \hat{f}_ω is independent of ω . Also, we assumed v to be independent of ω as well. This means that we can bound the weights in the inner integral by some constants and get

$$\|V_h h\|_{\mathcal{L}_{w_{r,s}}^{p^*}}^{p^*} \lesssim \int_{\mathbb{R}} (1 + |\omega|)^{s p^*} \left\| \hat{f}_\omega \right\|_{\mathcal{L}^{p^*}}^{p^*} d\omega.$$

Assuming $2 \leq p^* < \infty$ and $\frac{1}{p^*} + \frac{1}{q} = 1$, we can use the Hausdorff-Young inequality to obtain

$$\|V_h h\|_{\mathcal{L}_{w_{r,s}}^{p^*}}^{p^*} \lesssim \int_{\mathbb{R}} (1 + |\omega|)^{s p^*} \|f_\omega\|_{\mathcal{L}^q}^{p^*} d\omega. \quad (44)$$

To bound the norm $\|f_\omega\|_{\mathcal{L}^q}$ we define the set

$$N_\omega = \left\{ t \in \mathbb{R} \mid \left| t - \frac{\omega}{2} \right| \leq \frac{|\omega|}{4} \right\}.$$

For all $t \in N_\omega$, since $\frac{|\omega|}{2} - |t| \leq \left| t - \frac{\omega}{2} \right|$, we have the inequality $|t| \geq \frac{|\omega|}{4}$. Therefore

$$\begin{aligned} \|f_\omega\|_{\mathcal{L}^q}^q &= 2 \int_{\mathbb{R}} \left| \frac{\sin t}{t} \right|^{kq} \left| \frac{\sin(t - \omega/2)}{t - \omega/2} \right|^{kq} dt \\ &\leq 2 \int_{N_\omega} \left| \frac{1}{t} \right|^{kq} \left| \frac{\sin(t - \omega/2)}{t - \omega/2} \right|^{kq} dt \\ & \quad + 2 \int_{N_\omega^c} \left| \frac{\sin t}{t} \right|^{kq} \left| \frac{1}{t - \omega/2} \right|^{kq} dt \\ &\leq 2 \cdot 4^{kq} |\omega|^{-kq} \int_{N_\omega} \left| \frac{\sin(t - \omega/2)}{t - \omega/2} \right|^{kq} dt \\ & \quad + 2 \cdot 4^{kq} |\omega|^{-kq} \int_{N_\omega^c} \left| \frac{\sin t}{t} \right|^{kq} dt \\ &\lesssim |\omega|^{-kq}, \end{aligned}$$

where we used the integrability of $\left| \frac{\sin(x)}{x} \right|^\alpha$ for $\alpha > 1$ in the last step.

Inserting this into (44), we get

$$\begin{aligned} \|V_h h\|_{\mathcal{L}_{w_{r,s}}^{p^*}}^{p^*} &\lesssim \int_{\mathbb{R}} (1 + |\omega|)^{s p^*} \|f_\omega\|_{\mathcal{L}^q}^{p^*} d\omega \\ &\lesssim \int_{|\omega|>1} \left(\frac{(1 + |\omega|)^s}{|\omega|^k} \right)^{p^*} d\omega \\ & \quad + \int_{|\omega|\leq 1} (1 + |\omega|)^{s p^*} \|f_\omega\|_{\mathcal{L}^q}^{p^*} d\omega \\ &\lesssim 2^{s p^*} \int_{\mathbb{R}} (1 + |\omega|)^{(s-k)p^*} d\omega \\ & \quad + \int_{|\omega|\leq 1} (1 + |\omega|)^{s p^*} \|f_\omega\|_{\mathcal{L}^q}^{p^*} d\omega. \end{aligned}$$

In the second summand we integrate a continuous function over a compact set, this integral is always finite. Therefore, we have $\|V_h h\|_{\mathcal{L}_{w_{r,s}}^{p^*}} < \infty$, if $(s - k)p^* < -1$. Additionally we assumed $s > 1/p$, so $1/p^* < k - s < k - 1/p$.

Since $2 \leq p^* < \infty$ can be arbitrarily large, we finally get the norm equivalence (7) for all B-spline window functions of order $k > 1/p$. \square

REFERENCES

- [1] 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.
- [2] Candes, E. J. (2006). Modern statistical estimation via oracle inequalities. *Acta Numerica* 15, 257–325.
- [3] Chambolle, A., R. A. DeVore, N.-Y. Lee, and B. J. Lucier (1998). Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing* 7(3), 319–335.
- [4] Connor, E., S. Li, and S. Li (2012). Automating identification of avian vocalizations using time-frequency information extracted from the Gabor transform. *The Journal of the Acoustical Society of America* 132, 507–17.
- [5] 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* 96(2), 464–506.
- [6] Daubechies, I., M. Defrise, and C. Mol (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraints. *Communications on Pure and Applied Mathematics* 57.
- [7] DeVore, R. and B. Lucier (1992). Fast wavelet techniques for near-optimal image processing. In *MILCOM*, Volume 92, pp. 1129–1135. IEEE.
- [8] DeVore, R. A., B. Jawerth, and V. Popov (1992). Compression of wavelet decompositions. *American Journal of Mathematics* 114(4), 737–785.
- [9] 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.
- [10] Donoho, D. L. and J. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81(3), 425–455.
- [11] Dörfler, M. (2001). Time-frequency analysis for music signals: A mathematical approach. *Journal of New Music Research* 30, 3–12.
- [12] Engl, H., M. Hanke, and A. Neubauer (2000). *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands.
- [13] Feichtinger, H. G. (2019). Gabor expansions of signals: computational aspects and open questions. In *Landscapes of Time-Frequency Analysis*, pp. 173–206. Springer.
- [14] Feichtinger, H. G. and K. Gröchenig (1989). Banach spaces related to integrable group representations and their atomic decompositions, i. *Journal of Functional Analysis* 86, 307–340.
- [15] Feichtinger, H. G. and K. Gröchenig (1992). Gabor wavelets and the heisenberg group: Gabor expansions and short time Fourier transform from the group theoretical point of view. *Wavelets: A tutorial in theory and applications* 2, 359–398.
- [16] Galperin, Y. V. and S. Samarah (2004). Time-frequency analysis on modulation spaces $M_m^{p,q}$, $0 < p, q \leq \infty$. *Applied and Computational Harmonic Analysis* 16(1), 1–18.
- [17] Gröchenig, K. (2001). *Foundations of Time-Frequency Analysis*. Springer Science & Business Media.
- [18] Gröchenig, K. and S. Samarah (2000). Non-linear approximation with local Fourier bases. *Constructive Approximation* 16(3), 317–331.
- [19] Heuer, S., P. Tafo, H. Holzmann, and S. Dahlke (2019). New aspects in birdsong recognition utilizing the Gabor transform. In *Proceedings of the 23rd International Congress on Acoustics. Aachen*.
- [20] Johnstone, I. M. (2013). Gaussian estimation: Sequence and wavelet models. *Unpublished manuscript*.
- [21] Johnstone, I. M. and B. W. Silverman (2005).

- Empirical bayes selection of wavelet thresholds. *The Annals of Statistics* 33(4), 1700–1752.
- [22] Kerkyacharian, G., O. Lepski, and D. Picard (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probability Theory and Related Fields* 121(2), 137–170.
- [23] 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.
- [24] Mallat, S. (2009). *A Wavelet Tour of Signal Processing. The Sparse Way*. Elsevier.
- [25] Necciari, T., P. Balazs, R. Kronland-Martinet, S. Ystad, B. Laback, S. Savel, and S. Meunier (2012). Auditory time-frequency masking: Psychoacoustical data and application to audio representations.
- [26] Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *The Annals of Statistics* 36(4), 1957–1982.
- [27] Samarah, S. and S. Al-Sa’di (2006). Characterization of modulation spaces by nonlinear approximation. *arXiv preprint math/0602250*.
- [28] Strohmer, T. (2001). Approximation of dual Gabor frames, window decay, and wireless communications. *Applied and Computational Harmonic Analysis* 11(2), 243–262.
- [29] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*, Volume 606. Springer series in statistics.
- [30] Wolfe, P. J., S. J. Godsill, and W.-J. Ng (2004). Bayesian variable selection and regularization for time–frequency surface estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66(3), 575–589.
- [31] Yu, G., S. Mallat, and E. Bacry (2008). Audio denoising by time-frequency block thresholding. *IEEE Transactions on Signal processing* 56(5), 1830–1839.
- Stephan Dahlke** Stephan Dahlke received his Ph.D. in Mathematics from the University of Bremen in 1989. From 1989 to 2002, he has been with the Universities of Berlin, Aachen, South Carolina and Bremen, respectively. Since 2002 he is a Full Professor of Numerical Analysis at the Philipps-Universität Marburg, Germany. He is the author of more than 100 scientific papers in international journals and proceedings. His research interests include Harmonic Analysis, Numerical Analysis, Regularity Theory of PDEs, and Data Sciences.
- Sven Heuer** Sven Heuer received his B.Sc. degree in mathematical economics in 2015 and his M.Sc. degree in mathematics in 2018 at Philipps-Universität Marburg, Germany. He is currently employed as a research assistant and Ph.D. candidate within the LOEWE priority project Nature 4.0 — Sensing Biodiversity, working on classification of audio signals with neural networks and time-frequency methods.
- Hajo Holzmam** Hajo Holzmam received his Ph.D. in Mathematics from the University of Göttingen in 2004. Since 2009 he is a Full Professor of Mathematical Statistics at the Philipps-Universität Marburg, Germany. He is the author of more than 50 scientific papers in international journals and proceedings. His research interests include nonparametric statistics and inverse problems, high-dimensional statistics and machine learning as well as latent variable models.
- Pavel Tafo** Pavel Tafo received his B.Sc. and M.Sc. degrees in mathematical economics in 2014 and 2018, respectively, at Philipps-Universität Marburg, Germany. He is currently employed as a research assistant and Ph.D. candidate within the LOEWE priority project Nature 4.0 — Sensing Biodiversity, working on time-frequency methods for audio analysis as well as on statistical methods for presence-only and presence-absence data.