Statistically optimal estimation of signals in modulation spaces using Gabor frames

Stephan Dahlke, Sven Heuer, Hajo Holzmann* and Pavel Tafo
Fachbereich Mathematik und Informatik
Philipps-Universität Marburg
{dahlke, heuersv, holzmann, tafo}@mathematik.uni-marburg.de

August 3, 2020

Abstract

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.

Keywords. Gabor frame, minimax estimation, short-time Fourier transform, time-frequency analysis, thresholding

1 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 (Doerfler, 2001) or bird songs (Connor et al., 2012), as well as in psychoacoustics (Necciari et al., 2012) and wireless communications (Strohmer,

*Corresponding author. Prof. Dr. Hajo Holzmann, Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Hans-Meerweinstr., 35043 Marburg, Germany

2001). Gröchenig (2013) provides a comprehensive account of the mathematics of time-frequency analysis. However, despite its major practical importance, the only contribution to a statistical treatment with the aim of denoising seems to be Wolfe et al. (2004).

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

$$V_{h_0} g(x,\omega) = \int_{\mathbb{R}^d} g(t)\, \overline{h_0}(t-x)\, \exp(-2\,\pi i\, \langle \omega, t\rangle)\, \mathrm{d}t, \qquad g \in \mathcal{L}^2(\mathbb{R}^d), \tag{1}$$

where $h_0 \in \mathcal{L}^2(\mathbb{R}^d)$ 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 $\overline{z}$ is the complex-conjugate of $z \in \mathbb{C}$. 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. Similar 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

$$\mathrm{d}Y(x) = f(x)\, \mathrm{d}x + \varepsilon\, \mathrm{d}W(x), \qquad x \in \mathbb{R}^d, \tag{2}$$

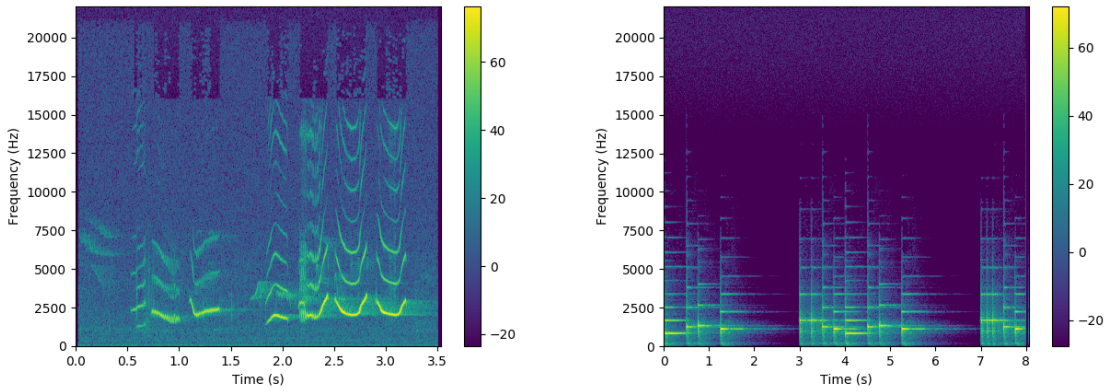
by Gabor frame expansions, where f is element in a suitable modulation space. White noise models are widely used in statistical analysis as stylized versions of more realistic nonparametric regression models. Formal approximation results of nonparametric regression by white noise on the rectangle $[0,1]^d$ are e.g. established in Reiß (2008). 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.

The estimators in (2) that we propose are based on soft and hard thresholding of the Gabor coefficients. The analysis uses the classical oracle inequalities from Donoho and Johnstone (1994), extended to complex-valued coefficients. 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. The lower bounds rely on Gaussian test functions, which approximately localize simultaneously in time as well as in frequency. Our contributions are thus complementary to Wolfe, Godsill, and Ng (2004), who study discrete Gabor expansions and focus on Bayesian regularization and computational aspects.

The structure of the paper is as follows. As motivation, in Section 2 we give an illustration of the performance of our method on a real-data example. Section 3 then summarizes the main 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 4 we introduce the thresholding estimators, and derive the minimax rates of convergence. Section 5 deals with the best N-term approximation of functions in modulations spaces by Gabor frame expansions with emphasis on the sparse case. We also show how to approximate solutions to variational problems in sequence space using Gabor coefficients. Finally, Section 6 contains the results of extensive numerical experiments. Proofs are deferred to Section 9.

2 A real data illustration

For motivation we illustrate our methods on two real-data examples. The first is a 3.46 seconds long recording of a common blackbird, sampled at 44.1 kHz, resulting in 152,605 samples, the second a simple melody which we played and recorded on a piano. Figure 1 gives the spectrograms, that is, time-frequency representations of the signals, with a Gaussian window function and suitable choices of grid size and window width. We then added Gaussian noise to the signals with various signal-to-



(a) Blackbird (b) Melody

Figure 1: Spectograms of two piano recordings

noise ratios (SNR). For examples of spectrograms of the noisy signals and our reconstructions see the supplement, Figure 10.

Iterating this procedure 10000 times, we then compared the mean squared errors between the original signal and the reconstruction based on noisy data for various signal-to-noise ratios, and for reconstructions both for our Gabor-frame expansion method as well as for wavelet shrinkage. We use hard thresholding in both cases. The results displayed in Table 1 indicate the superior performance of our Gabor-frame based method. The various recordings can be found on GitHub[1].

3 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 (2013) and Galperin and Samarah (2004). 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 $v : \mathbb{R}^{2d} \to [0, \infty)$ is *submultiplicative* if

[1] Link to recordings on GitHub under *https://github.com/ptafo/Statistically-optimal-estimation-of-signals-in-modulation-spaces-using-Gabor-frames*

	Reconstructing the blackbird			Reconstructing the melody	
SNR	Gabor frames	bior4.4 Wavelets	SNR	Gabor frames	bior4.4 Wavelets
-10	66.224	66.884	-10	58.843	65.354
10	51.997	54.792	10	47.469	52.525
30	37.241	39.413	30	32.580	35.504

Table 1: Mean squared errors of reconstructions by thresholding Gabor or wavelet expansions, both for the blackbird as well as for the melody. All values in decibel ($10 \log_{10} x$, dB).

$v(z_1 + z_2) \leq C\, v(z_1)\, v(z_2)$, $z_1, z_2 \in \mathbb{R}^{2d}$ for some constant $C > 0$, and $m : \mathbb{R}^{2d} \to [0, \infty)$ is *v-moderate* if $m(z_1 + z_2) \leq C\, v(z_1)\, m(z_2)$. A standard choice is $v_s(z) = m_s(z) = \big(1 + \|z\|_2^2\big)^{s/2}$ 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

$$\mathcal{L}^p_m = \{g : \mathbb{R}^{2d} \to \mathbb{C} \text{ measurable} \mid \|g\|^p_{\mathcal{L}^p_m} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |g(x,\omega)|^p\, m(x,\omega)^p \, \mathrm{d}x \, \mathrm{d}\omega < \infty\},$$

is a complete (semi-) normed space for $p \geq 1$, and a complete quasi-(semi-) normed space for $p \in (0, 1)$, see e.g. Benedek and Panzone (1961).

The *modulation space* $\mathcal{M}^p_m(\mathbb{R}^d)$ 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 $V_{h_0} f \in \mathcal{L}^p_m(\mathbb{R}^{2d})$, and $h_0 \in \mathcal{S}(\mathbb{R}^d)$ is a fixed window function in the Schwartz space. See Gröchenig (2013, Section 11.2) for the definition of the STFT of a tempered distribution. The modulation (quasi-) norm is given by

$$\|f\|_{\mathcal{M}^p_m(\mathbb{R}^d)} = \|V_{h_0} f\|_{\mathcal{L}^p_m(\mathbb{R}^{2d})} = \Big(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_{h_0} f(x,\omega)|^p\, m(x,\omega)^p \, \mathrm{d}x \, \mathrm{d}\omega \Big)^{1/p}.$$

If the weight function is $m = 1$ we write $\mathcal{M}^p_1(\mathbb{R}^d) = \mathcal{M}^p(\mathbb{R}^d)$ and $\mathcal{L}^p_1(\mathbb{R}^{2d}) = \mathcal{L}^p(\mathbb{R}^{2d})$.

It turns out that the definition of the set $\mathcal{M}^p_m(\mathbb{R}^d)$ 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 (2013, Theorem 11.3.7 and Proposition 12.1.2) for $p \geq 1$ and Galperin and Samarah (2004, Theorem 3.1) for $p \in (0, 1)$. A standard choice is the Gaussian window function $\varphi_a(x) = \exp\big(-\pi \|x\|_2^2/a\big)$, $a > 0$. While by definition, modulation spaces consist of tempered distributions, for the weight functions m_s defined above and $m_{u,v}$ in (13) that we focus on, the modulation space is a subset of a Bessel potential space (Gröchenig, 2013, 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 $f \in \mathcal{H}$,

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

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 $\mathbf{c} = (c_\lambda)_{\lambda \in \Lambda}$, 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 *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 on $\mathbb{R}$, a *Gabor frame* is the family of functions, indexed by the lattice $\Lambda = \alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$, defined by

$$h_\lambda(x) = \exp(2\pi i \langle \beta\, n, x - \alpha k \rangle)\, h(x - \alpha k), \quad \lambda = (\alpha k, \beta n) \in \Lambda. \tag{3}$$

For sufficiently small choices of $\alpha, \beta > 0$ and a suitable choice of h (Gröchenig, 2013, Theorem 6.5.1), $(h_\lambda)_{\lambda\in\Lambda}$ is indeed a frame in $\mathcal{L}^2(\mathbb{R}^d)$, and the dual frame is also of the form (3) 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^p_m if (Gröchenig, 2013, Def. 11.1.3)

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

where m is the 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 (2013, Theorem 12.2.4), Galperin and Samarah (2004, Theorem 3.6), we have that if $\mathbf{c} = (c_\lambda)_{\lambda\in\Lambda} \in \ell^p_m$ then $f = \sum_{\lambda\in\Lambda} c_\lambda h_\lambda \in \mathcal{M}^p_m(\mathbb{R}^d)$ and

$$\|f\|_{\mathcal{M}^p_m(\mathbb{R}^d)} \leq \text{const.}\, \|\mathbf{c}\|_{\ell^p_m}. \tag{4}$$

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

$$f = \sum_{\lambda\in\Lambda} \langle f, \tilde{h}_\lambda \rangle\, h_\lambda \tag{5}$$

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

$$\tilde{C}_1\, \|f\|_{\mathcal{M}^p_m(\mathbb{R}^d)} \leq \|(\langle f, \tilde{h}_\lambda \rangle)_{\lambda\in\Lambda}\|_{\ell^p_m} \leq \tilde{C}_2\, \|f\|_{\mathcal{M}^p_m(\mathbb{R}^d)}, \qquad f \in \mathcal{M}^p_m(\mathbb{R}^d), \tag{6}$$

for suitable constants $0 < \tilde{C}_1 < \tilde{C}_2$. From a numerical and computational perspective, windows functions with compact support might be more suitable choices. It particular, cardinal B-splines can be useful as window functions in large-data classification problems, where computation time is an issue (Heuer et al., 2019). 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 (6) 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* (6) *holds for all $p > \frac{1}{k}$.*

The proof is provided at the end of Section 9. We shall investigate the numerical performance of B-splines as window functions further in the Section 6.

4 Thresholding estimators, oracle inequalities and rates of convergence

Let us introduce threshold estimators based on Gabor frame expansions. 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

$$Y(h) = \int h \, \mathrm{d}Y \sim \mathcal{N}\big(\langle f, h\rangle_{\mathcal{L}^2}, \varepsilon^2 \, \|h\|^2_{\mathcal{L}^2}\big), \tag{7}$$

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 (7) are jointly normally distributed with covariance $\mathrm{Cov}(Y(h_j), Y(h_k)) = \varepsilon^2 \, \langle h_j, h_k\rangle$.

When applying (7) 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 (6) holds, and estimate the coefficient

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

Note that even if the window $\tilde{h}$ is real-valued, the element of the frame $\tilde{h}_\lambda$ in (3) will be complex-valued. Soft and hard thresholding at level $\mu > 0$ are defined by

$$t_{\mathrm{s}}(v; \mu) = \mathrm{sign}(v)\,(|v| - \mu)\,\mathbf{1}(|v| \geq \mu), \qquad \text{and} \qquad t_{\mathrm{h}}(v; \mu) = v\,\mathbf{1}(|v| \geq \mu),$$

where $v \in \mathbb{R}$ and $\mathrm{sign}(v)$ is the sign of v. For complex $z = u + iv \in \mathbb{C}$, $u, v \in \mathbb{R}$ we define $t_{\mathrm{j}}(z; \mu) = t_{\mathrm{j}}(u; \mu) + i\, t_{\mathrm{j}}(v; \mu)$, $\mathrm{j} \in \{\mathrm{h}, \mathrm{s}\}$. Then we have the following oracle inequalities, which are extensions to complex-valued random variables of classic results from Donoho and Johnstone (1994).

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 $\mu_{\mathrm{uni}} = \varepsilon \, \|\tilde{h}\|_{\mathcal{L}^2} \sqrt{2 \, \log(\#\Lambda_0)}$ that*

$$\mathbb{E}\Big[\sum_{\lambda \in \Lambda_0} \big| t_{\mathrm{s}}(Y(\tilde{h}_\lambda); \mu_{\mathrm{uni}}) - \vartheta_\lambda \big|^2 \Big] \leq \big(4 \, \log(\#\Lambda_0) + 2\big) \big(\varepsilon^2 \, \|\tilde{h}\|^2_{\mathcal{L}^2} + \sum_{\lambda \in \Lambda_0} \min(\varepsilon^2 \, \|\tilde{h}\|^2_{\mathcal{L}^2}, |\vartheta_\lambda|^2) \big). \tag{8}$$

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

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

$$\hat{f}_{\mathrm{j}}(\cdot; \mu) = \sum_{\lambda \in \Lambda_0} t_{\mathrm{j}}(Y(\tilde{h}_\lambda); \mu)\, h_\lambda, \quad \mathrm{j} \in \{\mathrm{h}, \mathrm{s}\}. \tag{9}$$

From boundedness of the synthesis operator D_h (Gröchenig, 2013, Prop. 5.1.1 (b)) we obtain that

$$\begin{aligned} \|\hat{f}_{\mathrm{j}}(\cdot; \mu) - f\|^2_{\mathcal{L}^2} &= \Big\| \sum_{\lambda \in \Lambda_0} \big(t_{\mathrm{j}}(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda \big)\, h_\lambda - \sum_{\lambda \in \Lambda_0^{\complement}} \vartheta_\lambda \, h_\lambda \Big\|^2_{\mathcal{L}^2} \\ &\leq B \Big(\sum_{\lambda \in \Lambda_0} \big| t_{\mathrm{j}}(Y(\tilde{h}_\lambda); \mu) - \vartheta_\lambda \big|^2 + \sum_{\lambda \in \Lambda_0^{\complement}} \big| \vartheta_\lambda \big|^2 \Big), \quad \mathrm{j} \in \{\mathrm{h}, \mathrm{s}\}, \end{aligned} \tag{10}$$

where B is the upper frame constant of the Gabor frame. In the following, using the oracles inequalities from Proposition 2 we shall bound (10) for particular choices of the weight function.

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

$$m_s(\lambda) = v_s(\lambda) = \big(1 + \|x\|_2^2 + \|\omega\|_2^2\big)^{s/2}, \tag{11}$$

and $v_s = m_s$, as mentioned in the introduction. Given $K > 0$ let $\Lambda_{0,K} = \{\lambda \in \Lambda \mid \|\lambda\|_2 \le K\}$, and for large K consider the estimator in (9) which can be written as

$$\hat{f}_{\mathrm{j}}(\cdot;\mu) = \sum_{\|\lambda\|_2 \le K} t_{\mathrm{j}}(Y(\tilde{h}_\lambda);\mu)\, h_\lambda, \qquad \mathrm{j} \in \{\mathrm{s},\mathrm{h}\}.$$

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

$$\mathbb{E}\big[\|\hat{f}_{\mathrm{j}}(\cdot;\mu) - f\|_{\mathcal{L}^2}^2\big] \le \text{const.} \cdot \max\Big(\|f\|^2_{\mathcal{M}^p_{m_s}}, \|f\|^{\frac{2\,d\,p}{2\,d+s\,p}}_{\mathcal{M}^p_{m_s}}\Big) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{2\,d\,(2-p)+2sp}{2\,d+s\,p}}, \quad \mathrm{j} \in \{\mathrm{s},\mathrm{h}\}, \tag{12}$$

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

Remark. Let us discuss the rate obtained in (12). For $p = 2$ we obtain $\varepsilon^{\frac{4s}{2\,d+2\,s}}$ up to the logarithmic factor, which is reminiscent with the rate over Sobolev ellipsoids, but somewhat surprisingly involves the dimension as $2\,d$ 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 (12) 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 (1995).

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* $\mathcal{M}^p_{m_s}(\mathbb{R}^d)$,

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

Then in model (2) *we have that*

$$\liminf_{\varepsilon \downarrow 0} \Big(\varepsilon^{-\frac{2\,d\,(2-p)+2sp}{2\,d+s\,p}} \inf_{\hat{f}_\varepsilon} \sup_{f \in \mathcal{M}^p_{s,C}} \mathbb{E}_f\big[\|\hat{f}_\varepsilon - f\|^2_{\mathcal{L}^2}\big]\Big) > 0,$$

where $\hat{f}_\varepsilon$ *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. While the proof uses standard tools from decision theory such as Fano's lemma and the Varshamov-Gilbert bound, see Tsybakov (2009), the issue is to construct the hypothesis functions in order to obtain the term $2\,d$ which arises in the upper bound. Since there are no integrable functions which localize sharply - with compact support - in both time and frequency domain (Gröchenig, 2013, 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

$$m_{u,v}(\lambda) = \big(1 + \|x\|_2^2\big)^{u/2} \cdot \big(1 + \|\omega\|_2^2\big)^{v/2}, \tag{13}$$

it is then also v_s -moderate for v_s given in (11). For a constant $C > 0$ let

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

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

Theorem 5. *Consider model* (2) *with* $f \in \mathcal{M}^p_{m_{u,v}}(\mathbb{R}^d)$ *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*

$$\mathbb{E}\big[\|\hat{f}(\cdot;\mu) - f\|^2_{\mathcal{L}^2}\big] \leq \text{ const.} \cdot \max\Big(\|f\|^2_{\mathcal{M}^p_{m_{u,v}}}, \|f\|^{\frac{p\,d\,(v+u)}{d(v+u)+pvu}}_{\mathcal{M}^p_{m_{u,v}}}\Big) \cdot \log(1/\varepsilon) \cdot \varepsilon^{\frac{(2-p)\,d\,(v+u)+2pvu}{d(v+u)+pvu}},$$

and furthermore the corresponding lower bound

$$\liminf_{\varepsilon \downarrow 0} \Big(\varepsilon^{-\frac{(2-p)\,d\,(v+u)+2pvu}{d(v+u)+pvu}} \inf_{\hat{f}_\varepsilon} \sup_{f \in \mathcal{M}^p_{u,v,C}} \mathbb{E}_f\big[\|\hat{f}_\varepsilon - f\|^2_{\mathcal{L}^2}\big] \Big) > 0.$$

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 $p = 2$, $\mathcal{M}^2_{m_{0,v}}(\mathbb{R}^d)$ corresponds to a Sobolev space (Gröchenig, 2013, 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 $\mathcal{M}^p_{m_{u,v}}$ for arbitrarily large u. For $p = 2$, fixed v and $u \to \infty$ 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 (Kerkyacharian et al., 2001, 2008), though the rates that we obtain are different.

5 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 (DeVore et al., 1992) 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 (2013, 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 (2000); Samarah and Al-Sa'di (2006). 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 6.1.

Theorem 6. *Given $f \in \mathcal{M}^p(\mathbb{R}^d)$, $p \in (0, 2)$ and $\mu > 0$, 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

$$\|f - f_\mu\|_{\mathcal{L}^2}^2 \leq \tilde{C}_2^2 \, \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}^2 \, N_\mu^{1-2/p}, \tag{14}$$

where $\tilde{C}_2$ is from (6).

Remark 1. To prove this compression rate, we only need the norm equivalence in (6). 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 6 we show some simulations highlighting this effect.

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*

$$\min_{g \in \mathcal{M}_m^p(\mathbb{R}^d)} \big(\|f - g\|_{\mathcal{L}_2}^2 + \mu \|g\|_{\mathcal{M}_m^p(\mathbb{R}^d)}^p \big). \tag{15}$$

Variational problems of the form (15) 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. (2000). 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 (Daubechies et al., 2004). 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 (1992); Chambolle et al. (1998).

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* (4) *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* (4)*, we have that*

$$\|f - g\|_{\mathcal{L}^2}^2 + \mu \|g\|_{\mathcal{M}_m^p(\mathbb{R}^d)}^p \leq C \sum_{\lambda \in \Lambda} \big((\langle f, \tilde{h}_\lambda \rangle - c_\lambda)^2 + \mu \, m(\lambda)^p \, |c_\lambda|^p \big). \tag{16}$$

Remark 2. The right-hand side of (16) 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 (16) is finite (it is finite for the choice $c_\lambda = 0$), we must have $\tilde{\mathbf{c}} \in \ell^p_m$. 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 $E(\hat{c}_\lambda) \leq 4\, E(\tilde{c}_\lambda)$, see Chambolle et al. (1998, Section 3, p. 5) hence we also have that $\hat{\mathbf{c}} = (\hat{c}_\lambda)_{\lambda\in\Lambda} \in \ell^p_m$.

Remark 3. 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 (16). The reason is that when representing $g = \sum_{\lambda\in\Lambda} c_\lambda\, h_\lambda \in \mathcal{M}^p_m(\mathbb{R}^d)$ for general frame coefficients $(c_\lambda)_{\lambda\in\Lambda}$, we have the upper bound in (4) but the upper bound in (6) 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 (2013, 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 $p < 1$.

6 Numerical illustrations

In this section we numerically illustrate our denoising methods from Section 4. We shall consider the following family of functions of spatially varying frequency

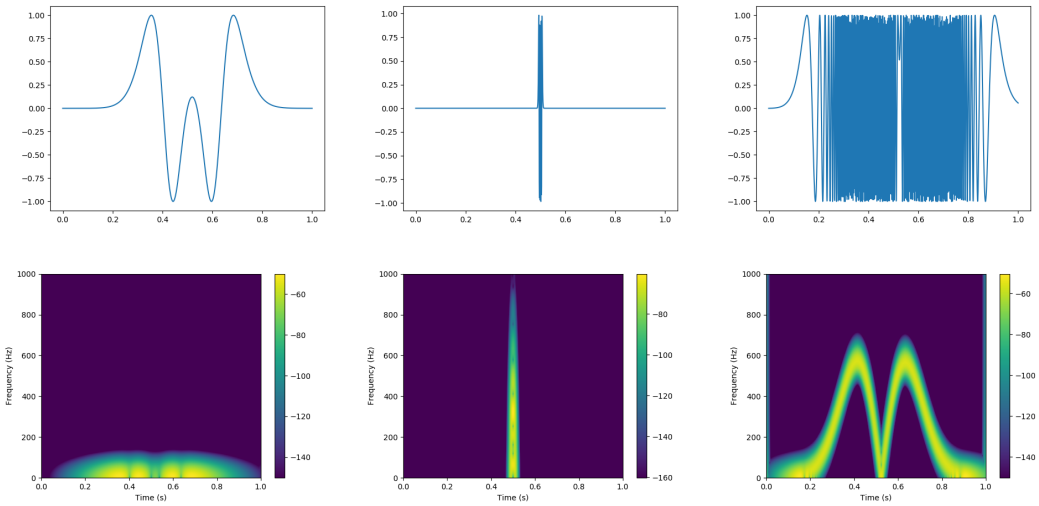
$$f_{A,B}(t) = \sin(2\pi \cdot B \cdot t \cdot e^{-A(t-0.5)^2}), \qquad t \in [0,1], \tag{17}$$

which resemble the Doppler functions used by Donoho and Johnstone (1994). In particular we consider the following three signals: firstly, the signal $f_{50,2}$ with small frequency variation, see Figure 2(a); secondly, the signal $f_{50000,4}$ with small time variation, see Figure 2(b); and lastly the signal $f_{40,200}$ which varies equally in both time and frequency domains, see Figure 2(c).
We investigate the numerical performance of the thresholding estimators in Section 4, and compare them to 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

$$Y_i = f(i/n) + \sigma\epsilon_i, \qquad i = 1, \ldots, n,$$

with $n = 2000$ and use a discretized version of the Gabor frames in (3). In repeated simulations we shall use $m = 2000$ iterations, e.g. to compute mean squared errors.

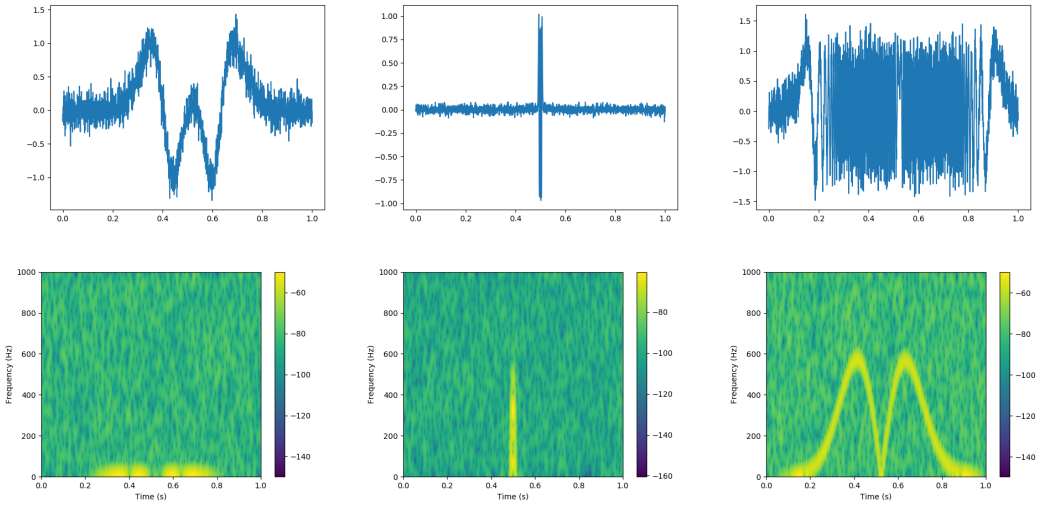


(a) $f_{50,2}$
$\|\cdot\|^2_{\mathcal{L}_2} = 0.2202$

(b) $f_{50000,4}$
$\|\cdot\|^2_{\mathcal{L}_2} = 0.00825$

(c) $f_{40,200}$
$\|\cdot\|^2_{\mathcal{L}_2} = 0.4269$

Figure 2: Signals (upper row) and spectrogram (lower row) of our test signals



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

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

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

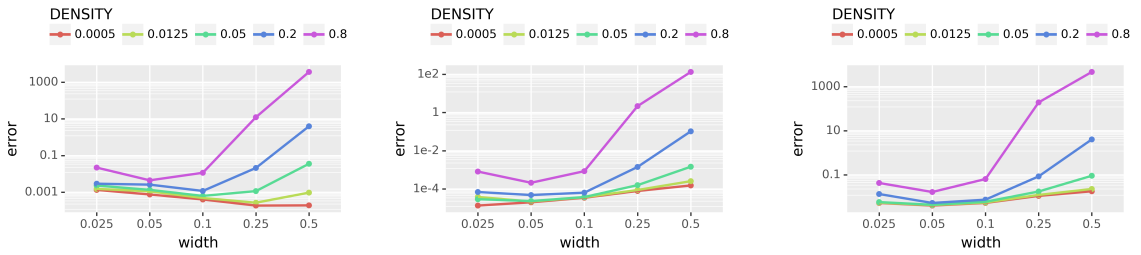
Figure 3: Noisy signals with SNR = 10

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 2: Number of frequency and time bands and the grid density

Effects of window length and grid density

We start by investigating the effects of the window width $w = \frac{W}{n} \in (0, 1]$, where W is the number of observations within the support of the window function, as well as the effect of the grid density $\alpha \cdot \beta$ of the frame in (3). 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 $M, N \in \{1, \ldots, n\}$. 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 Gröchenig (2013)[Theorem 7.5.3]. Here we report results for the case $M = N$, additional results for distinct time - and frequency resolutions can be found in the supplement, Section 9.3. Figure 4 shows the effect of different window lengths and grid densities according to Table 2 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.



(a) $f_{50,2}$ (b) $f_{50000,4}$ (c) $f_{40,200}$

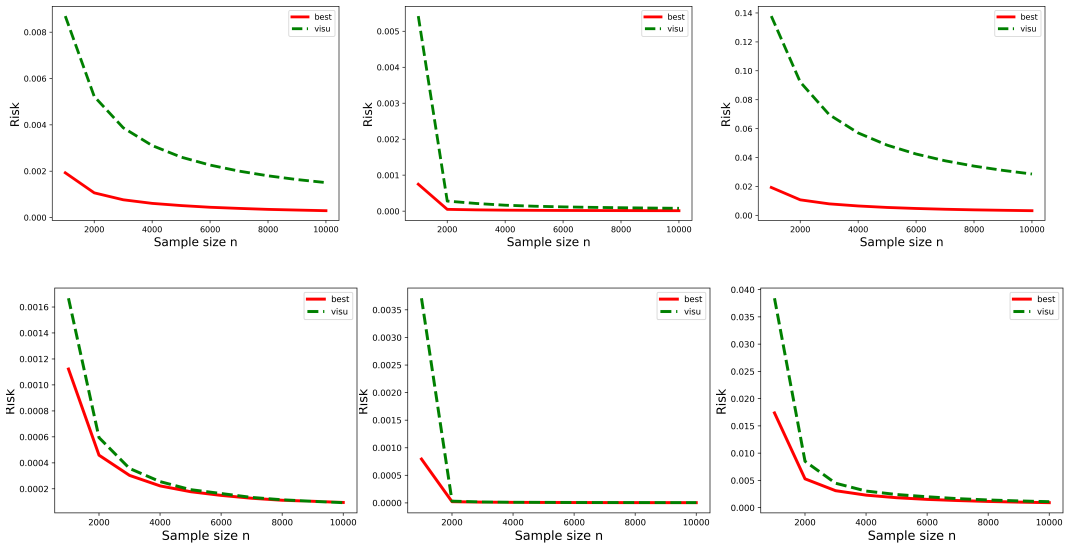
Figure 4: Effect of the window width w and the grid density with $M = N$ for the three signals for samples of sizes $n = 2000$, using hard thresholding with the optimal threshold.

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 $f_{50000,4}$ 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. 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

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 5, 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.



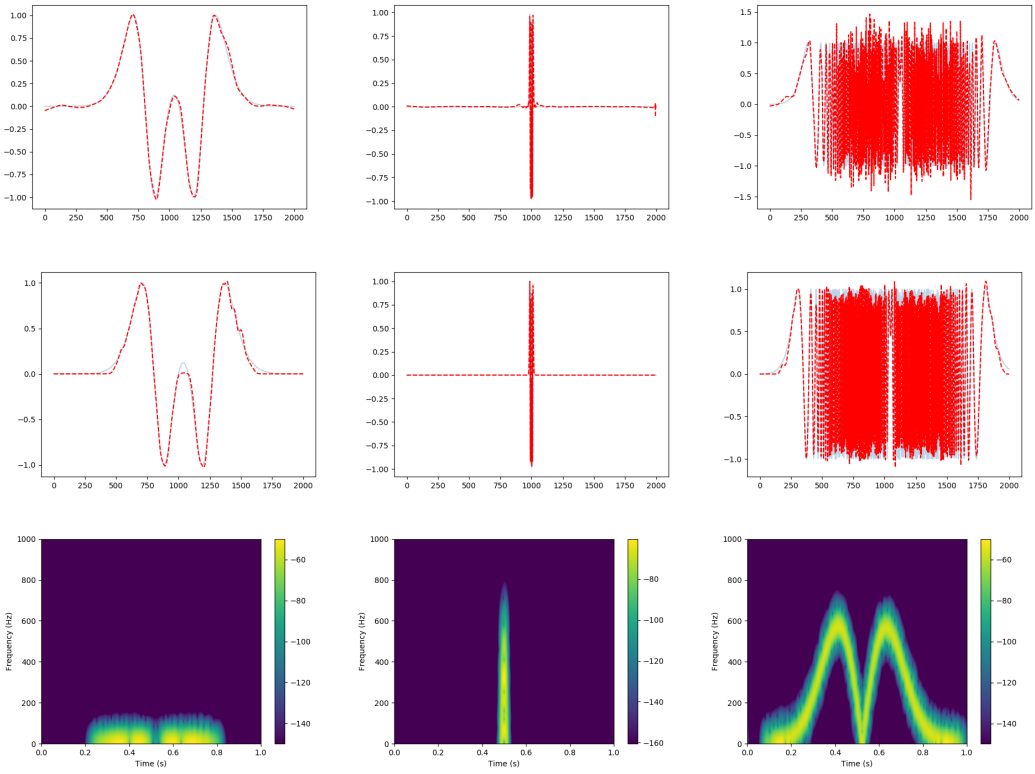
(a) $f_{50,2}$ (b) $f_{50000,4}$ (c) $f_{40,200}$

Figure 5: 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.

Comparison of Gabor-frame and wavelet-based methods

Next we investigate the reconstruction performance of our thresholding estimators in (9), 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

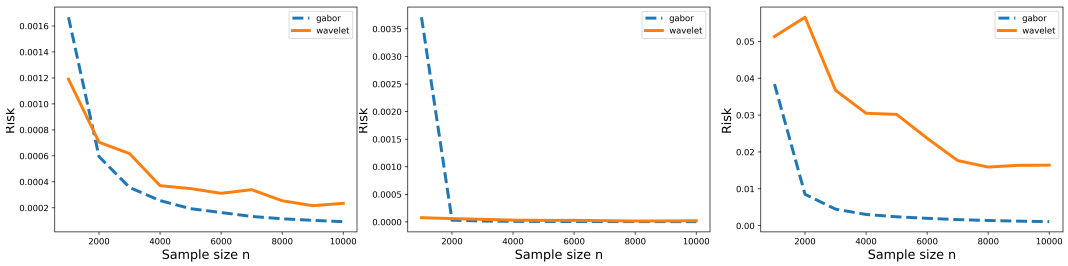
reconstructions of the signals in Figure 6, where the figures in the first row are from wavelet shrinkage, the figures in the second row from the Gabor-frame based method, and the last row has the spectrograms of the Gabor-frame based estimates. In particular for the signal $f_{40,200}$ the Gabor-frame 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



(a) $f_{50,2}$ (b) $f_{50000,4}$ (c) $f_{40,200}$

Figure 6: 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.

repeated simulation for various sample sizes. The results are given in Figure 7. 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.



(a) $f_{50,2}$ (b) $f_{50000,4}$ (c) $f_{40,200}$

Figure 7: MSE of the competing methods using universal hard thresholding. −−: Gabor-frame based method, — wavelet-shrinkage.

6.1 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, (14) can be written as

$$\log(\|f - f_\mu\|_{\mathcal{L}^2}^2) \leq \log(\tilde{C}_2^2 \, \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}^2) + \log(N_\mu) \cdot (1 - 2/p). \tag{18}$$

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 GitHub[2]. First, we analyse the Gaussian function

$$f(x) = \frac{1}{0.3 \cdot \pi} \exp\left(-\frac{1}{2}\left(\frac{x - 0.5}{0.15}\right)^2\right), \; x \in [0, 1],$$

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 $\mathcal{M}^p([0,1])$, $p > 0$. 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 (18) or the good time-frequency localisation of the signal and only focus on the "main slope" in the results. 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. Figure 8 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 (18) 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

[2]https://github.com/Heuersv/Compression-Simulations

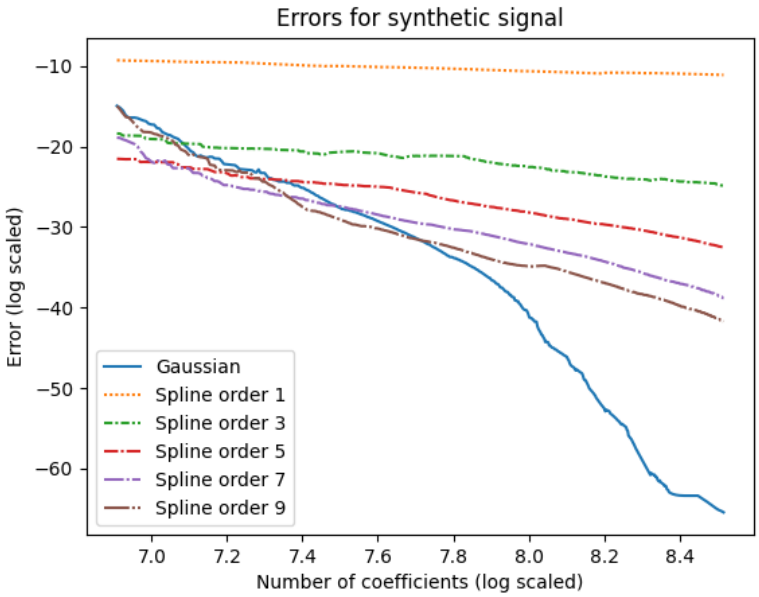


Figure 8: Errors for best N-term approximation of synthetic signal with different window functions.

3 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

Window function	estimated p	$1/k$, k spline order
Spline order 1	0.9318	1
Spline order 2	0.4538	0.5
Spline order 3	0.4170	0.3333
Spline order 4	0.3068	0.25
Spline order 5	0.2540	0.2
Spline order 6	0.2066	0.1667
Spline order 7	0.1715	0.1429
Spline order 8	0.1366	0.125
Spline order 9	0.1340	0.1111
Spline order 10	0.1147	0.1
Gaussian	0.0569	≈ 0

Table 3: Values of p from compressing a synthetic signal

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 advantage by using the compactly supported spline windows, but at the expense of worse compression rates.

We now turn to numerical illustrations using the recording of a common blackbird from Section 2. We use the same procedure as for the synthetic signal and only change the parameters to step sizes of 130 samples 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 9. Here, we can

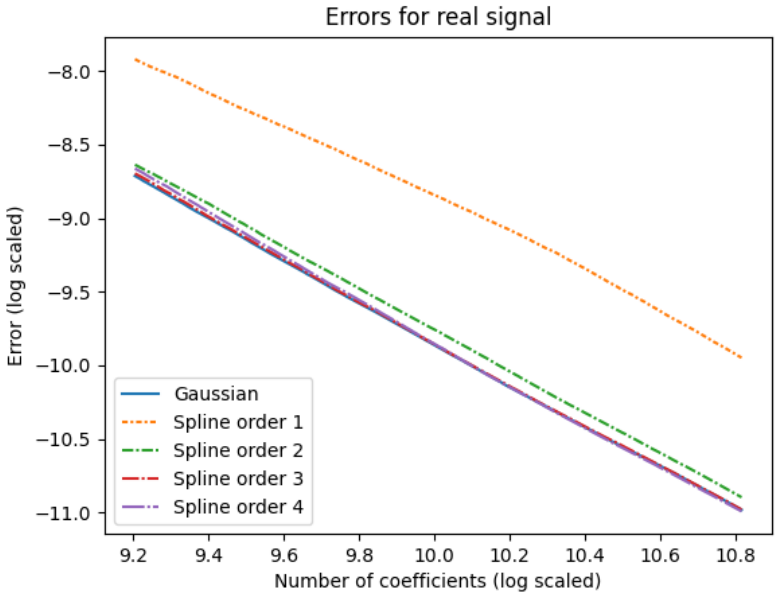


Figure 9: Errors for best N-term approximation of blackbird recording with different window functions.

see the linear dependency in (18) 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 4. 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.

Window function	estimated p from compression rate
Spline order 1	0.8844
Spline order 2	0.8322
Spline order 3	0.8317
Spline order 4	0.8207
Gaussian	0.8342

Table 4: Values of p from compressing the blackbird signal

7 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 particular 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 (Heuer et al., 2019).
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 (Dahlke et al., 2008) 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.

8 Proofs

8.1 Proofs of Proposition 2 and Theorem 3

Proof of Proposition 2. For (8), we show that for $\mu > 0$,

$$\mathbb{E}\Big[\sum_{\lambda\in\Lambda_0}\big|t_{\mathrm{s}}(Y(\tilde{h}_\lambda);\mu)-\vartheta_\lambda\big|^2\Big] \le \sum_{\lambda\in\Lambda_0}\min\big(|\vartheta_\lambda|^2,\varepsilon^2\|\tilde{h}\|_{\mathcal{L}^2}^2+2\,\mu^2\big)+\big(\#\Lambda_0\big)\,\frac{\varepsilon^3\,\|\tilde{h}\|_{\mathcal{L}^2}^3}{\mu}\,\exp\Big(-\frac{\mu^2}{2\,\varepsilon^2\,\|\tilde{h}\|_{\mathcal{L}^2}^2}\Big). \tag{19}$$

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 $\mathrm{j}\in\{\mathrm{s},\mathrm{h}\}$ that

$$\begin{aligned}&\sum_{\lambda\in\Lambda_0}\big|t_{\mathrm{j}}(Y(\tilde{h}_\lambda);\mu)-\vartheta_\lambda\big|^2\\ &=\sum_{\lambda\in\Lambda_0}\Big(\big(t_{\mathrm{j}}(\Re(Y(\tilde{h}_\lambda)))-\langle f,\Re(\tilde{h}_\lambda)\rangle\big)^2+\big(t_{\mathrm{j}}(\Im(Y(\tilde{h}_\lambda)))-\langle f,\Im(\tilde{h}_\lambda)\rangle\big)^2\Big).\end{aligned} \tag{20}$$

To bound the expected value of (20), we use the following fact, see Donoho and Johnstone (1994) or Candes (2006, Proof of Theorem 5.1) that for $W\sim\mathcal{N}(a,\sigma_0^2)$ we have that

$$\mathbb{E}\big[(t_{\mathrm{s}}(W;\mu)-a)^2\big] \le \min(a^2,\sigma_0^2+\mu^2)+2\,\frac{\sigma_0^3}{\mu}\,\varphi\Big(\frac{\mu}{\sigma_0}\Big), \tag{21}$$

where φ is the density of the standard normal distribution. Applying this to (20) we obtain

$$\begin{aligned}&\mathbb{E}\Big[\sum_{\lambda\in\Lambda_0}\Big(\big(t_{\mathrm{s}}(\Re(Y(\tilde{h}_\lambda));\mu)-\langle f,\Re(\tilde{h}_\lambda)\rangle\big)^2+\big(t_{\mathrm{s}}(\Im(Y(\tilde{h}_\lambda));\mu)-\langle f,\Im(\tilde{h}_\lambda)\rangle\big)^2\Big)\Big]\\
&=\sum_{\lambda\in\Lambda_0}\mathbb{E}\Big[\big(t_{\mathrm{s}}(\Re(Y(\tilde{h}_\lambda));\mu)-\langle f,\Re(\tilde{h}_\lambda)\rangle\big)^2\Big]+\sum_{\lambda\in\Lambda_0}\mathbb{E}\Big[\big(t_{\mathrm{s}}(\Im(Y(\tilde{h}_\lambda));\mu)-\langle f,\Im(\tilde{h}_\lambda)\rangle\big)^2\Big]\\
&\le\sum_{\lambda\in\Lambda_0}\Big(\min\big((\Re(\vartheta_\lambda))^2,\varepsilon^2\|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2+\mu^2\big)+\frac{\varepsilon^3\,\|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^3}{\mu}\,\exp\Big(-\frac{\mu^2}{2\,\varepsilon^2\,\|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2}\Big)\Big)\\
&+\sum_{\lambda\in\Lambda_0}\Big(\min\big((\Im(\vartheta_\lambda))^2,\varepsilon^2\|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2+\mu^2\big)+\frac{\varepsilon^3\,\|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^3}{\mu}\,\exp\Big(-\frac{\mu^2}{2\,\varepsilon^2\,\|\Im(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2}\Big)\Big)\\
&\le\sum_{\lambda\in\Lambda_0}\min\big(|\vartheta_\lambda|^2,\varepsilon^2\|\tilde{h}\|_{\mathcal{L}^2}^2+2\,\mu^2\big)+\big(\#\Lambda_0\big)\,\frac{\varepsilon^3\,\|\tilde{h}\|_{\mathcal{L}^2}^3}{\mu}\,\exp\Big(-\frac{\mu^2}{2\,\varepsilon^2\,\|\tilde{h}\|_{\mathcal{L}^2}^2}\Big),\end{aligned}$$

which is (19).

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 (Johnstone, 2013, Proposition 8.1)

$$\mathbb{E}\big[(t_{\mathrm{h}}(W;\mu)-a)^2\big] \le C\,\min(a^2,\mu^2)+C\,\sigma_0\,\mu\,\varphi(\mu/\sigma_0-1). \tag{22}$$

Proceeding from (20), since $\Re(Y(\tilde{h}_\lambda))\sim\mathcal{N}\big(\Re(\tilde{\vartheta}_\lambda),\epsilon^2\,\|\Re(\tilde{h}_\lambda)\|_{\mathcal{L}^2}^2\big)$ we have that

$$\begin{aligned}&\mathbb{E}\Big[\sum_{\lambda\in\Lambda_0}\big|t_{\mathrm{h}}(Y(\tilde{h}_\lambda);\mu)-\vartheta_\lambda\big|^2\Big]\\
&\le C\sum_{\lambda\in\Lambda_0}\Big(\min\big((\Re(\vartheta_\lambda))^2,\mu^2\big)+\min\big((\Im(\vartheta_\lambda))^2,\mu^2\big)\Big)\end{aligned}$$

$$+ C\,(\#\Lambda_0)\,\varepsilon\,\mu\cdot\Big(\|\Re(\tilde h_\lambda)\|_{\mathcal{L}^2}\,\varphi\Big(\frac{\mu}{\varepsilon\,\|\Re(\tilde h_\lambda)\|_{\mathcal{L}^2}}-1\Big)+\|\Im(\tilde h_\lambda)\|_{\mathcal{L}^2}\,\varphi\Big(\frac{\mu}{\varepsilon\,\|\Im(\tilde h_\lambda)\|_{\mathcal{L}^2}}-1\Big)\Big)$$
$$\le C\sum_{\lambda\in\Lambda_0}\min\big(|\vartheta_\lambda|^2,2\,\mu^2\big)+2\;C\,\varepsilon\,(\#\Lambda_0)\,\|\tilde h\|_{\mathcal{L}^2}\,\mu\,\varphi\Big(\frac{\mu}{\varepsilon\,\|\tilde h\|_{\mathcal{L}^2}}-1\Big).$$

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

Proof of Theorem 3. We may bound the second term in the oracle inequality (8) for functions in the modulation space associated to the weight function in (11) as follows.

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

$$\sum_{\|\lambda\|_2\le K}\min(\varepsilon^2\,\|\tilde h\|^2_{\mathcal{L}^2},|\vartheta_\lambda|^2)\le\ const.\cdot\|f\|^{\frac{2\,d\,p}{2\,d+s\,p}}_{\mathcal{M}^p_{m_s}}\,\varepsilon^{\frac{2\,d\,(2-p)+2sp}{2\,d+s\,p}}.$$

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

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

Note that

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

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

$$\begin{aligned}\sum_{\|\lambda\|_2\le K}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)&=\sum_{\|\lambda\|_2\le K_s}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)+\sum_{K_s<\|\lambda\|_2\le K}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)\\&\le\varepsilon^2\,C\,K_s^{2d}+\varepsilon^{2-p}\sum_{K_s<\|\lambda\|_2}|\vartheta_\lambda|^p\\&\le\varepsilon^2\,C\,K_s^{2d}+\varepsilon^{2-p}\,(1+K_s^2)^{-sp/2}\sum_{K_s<\|\lambda\|_2}m_s(\lambda)^p\,|\vartheta_\lambda|^p\\&\le\varepsilon^2\,C\,K_s^{2d}+\varepsilon^{2-p}\,K_s^{-sp}\,\tilde C_2^p\,\|f\|^p_{\mathcal{M}^p_{m_s}}\end{aligned}\tag{23}$$

by using $m_s(\lambda)^p>(1+K_s^2)^{sp/2}$ for $K_s<\|\lambda\|_2$ in the second last step, and (6) in the last step. Balancing both terms using

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

we get the result. □

Now let us estimate the truncation term in (10).

Lemma 9. *For $f\in\mathcal{M}^p_{m_s}(\mathbb{R}^d)$, $p\in(0,2]$ we have that* $\displaystyle\sum_{K<\|\lambda\|_2}|\vartheta_\lambda|^2\le K^{-2\,s}\,\tilde C^2\,\|f\|^2_{\mathcal{M}^p_{m_s}}$.

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

$$\begin{aligned}\sum_{K<\|\lambda\|_2}|\vartheta_\lambda|^2&\le\Big(\sum_{K<\|\lambda\|_2}|\vartheta_\lambda|^p\Big)^{2/p}\\&\le\Big(\big(1+K^2\big)^{-s\,p/2}\sum_{K<\|\lambda\|_2}m(\lambda)^p\,|\vartheta_\lambda|^p\Big)^{2/p}.\end{aligned}$$

Applying (6), we obtain the claim. □

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

8.2 Proof of Theorem 4

Proof of Theorem 4. The proof relies on the lower bound derived from Fano's lemma, see Tsybakov (2009, 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 $f_j \in \mathcal{M}^p_{s,C}$, where $M \asymp \exp(c\,m^{2d})$ for some $c > 0$, such that

$$\|f_j - f_k\|^2_{\mathcal{L}^2} \asymp \varepsilon^2\, m^{2d}, \qquad j \neq k. \tag{24}$$

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 $\mathrm{KL}(Y^{(j)}, Y^{(k)})$ we have that

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

so that by applying Tsybakov (2009, Theorem 2.5) yields the lower bound of order $\varepsilon^2\, m^{2d} = \varepsilon^{\frac{2\,d\,(2-p)+2sp}{2\,d+s\,p}}$, which is the statement of the theorem. Note that we need the upper bound in (24) for bounding the Kullback-Leibler divergence, and the lower bound for obtaining the rate.
To construct the test functions, we consider the Gaussian function

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

Given $m \in \mathbb{N}$, using the Varshamov-Gilbert bound (Tsybakov, 2009, Lemma 2.9) we may choose $M = \exp(c_1\, m^{2d})$, $c_1 = (\log 2)/8 > 0$ indices

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

of Hamming - distance $\|\iota - \tilde{\iota}\|_2^2 \geq m^{2d}/8$. Note that for vectors $\iota, \tilde{\iota} \in \{0,\}^{(m^{2d})}$, 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

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

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

$$\begin{aligned} f_\iota &= c\cdot\varepsilon \sum_{k_1,\ldots,k_{2d}=1}^{m} \iota_{(k_1,\ldots,k_{2d})^\top}\, T_{r\,(k_1,\ldots,k_d)^\top}\, M_{r\,(k_{d+1},\ldots,k_{2d})^\top}\, \varphi \\ &= c\cdot\varepsilon \sum_{k',k''}^{m} \iota_{k',k''}\, T_{r\,k'}\, M_{r\,k''}\, \varphi, \end{aligned} \tag{25}$$

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

Lemma 10. *We have that*

$$\begin{aligned}\big(V_\varphi f_\iota\big)\,(x,\omega) = c\cdot\varepsilon\sum_{k',k''}\iota_{k',k''}\,\exp\big(-2\,\pi i\,\langle\omega,rk'\rangle\big)\,2^{-d/2}\,\exp\big(\pi i\,\langle x-rk',\omega-rk''\rangle\big)\\ \cdot\exp\big(-\pi\|x-r\,k'\|_2^2/2\big)\,\exp\big(-\pi\|\omega-r\,k''\|_2^2/2\big).\end{aligned}\tag{26}$$

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

$$V_\varphi(\varphi)(x,\omega) = 2^{-d/2}\,\exp\big(-\pi\|x\|_2^2/2\big)\,\exp\big(\pi i\,\langle x,\omega\rangle\big)\,\exp\big(-\pi\|\omega\|_2^2/2\big).$$

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

$$\langle T_xf,g\rangle_{\mathcal{L}^2} = \langle f,T_{-x}\,g\rangle_{\mathcal{L}^2},\qquad \langle M_\omega f,g\rangle_{\mathcal{L}^2} = \langle f,M_{-\omega}\,g\rangle_{\mathcal{L}^2}$$

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

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

Using these formulas and the linearity of the STFT gives (26). □

Bounding the $\mathcal{L}_2$-distance

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

$$\|f_\iota - f_{\iota'}\|_{\mathcal{L}^2}^2 = 2^{-d/2}\,\|V_\varphi\,f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2,\qquad \iota,\iota'\in\{0,1\}^{(m^{2d})}.\tag{27}$$

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*

$$\|V_\varphi\,f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2 \geq \tilde c_1\,\varepsilon^2\,\|\iota-\iota'\|_2^2 \geq \tilde c_1\,\varepsilon^2\,m^{2d}/8,\tag{28}$$

$$\|V_\varphi\,f_\iota - V_\varphi f_{\iota'}\|_{\mathcal{L}^2}^2 \leq \tilde c_2\,\varepsilon^2\,\|\iota-\iota'\|_2^2 \leq \tilde c_2\,\varepsilon^2\,m^{2d}.\tag{29}$$

Proof of Lemma 11. First consider the lower bound (28). We have that

$$\begin{aligned}\big|\big(V_\varphi f_\iota\big) - \big(V_\varphi f_{\iota'}\big)\big|^2\,(x,\omega) \geq c^2\,\varepsilon^2 2^{-d}\Big(\sum_{k',k''}^{m}(\iota_{k',k''}-\iota'_{k',k''})^2\cdot\exp\big(-\pi\|x-r\,k'\|_2^2\big)\,\exp\big(-\pi\|\omega-r\,k''\|_2^2\big)\\ -\sum_{(k',k'')\neq(k'_1,k''_1)}^{m}|\iota_{k',k''}-\iota'_{k',k''}|\,|\iota_{k'_1,k''_1}-\iota'_{k'_1,k''_1}|\cdot\exp\big(-\frac{\pi}{2}\,(\|x-r\,k'\|_2^2+\|x-r\,k'_1\|_2^2)\big)\\ \cdot\exp\big(-\frac{\pi}{2}\,(\|\omega-r\,k''\|_2^2+\|\omega-r\,k''_1\|_2^2)\big)\Big).\end{aligned}\tag{30}$$

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

$$\sum_{(k',k'')\neq(0,0)}\exp\big(-a\,(\|x\|_2^2+\|x-r\,k'\|_2^2)\big)\,\exp\big(-a\,(\|\omega\|_2^2+\|\omega-r\,k''\|_2^2)\big)\leq 4\,\exp\big(-a\frac{r^2}{8}\big)\,\exp\big(-a\frac{\|x\|_2^2+\|\omega\|_2^2}{2}\big).\tag{31}$$

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

$$e^{-(\|x\|_2^2+\|\omega\|_2^2)} \sum_{(k',k'')\neq(0,0)} \exp\big(-\|x - r\,k'\|_2^2\big)\,\exp\big(-\|\omega - r\,k''\|_2^2\big). \tag{32}$$

Of course,

$$\sum_{(k',k'')\neq(0,0)} \exp\big(-\|x - r\,k'\|_2^2\big)\,\exp\big(-\|\omega - r\,k''\|_2^2\big) \le \sum_{(k',k'')} \exp\big(-\|x - r\,k'\|_2^2\big)\,\exp\big(-\|\omega - r\,k''\|_2^2\big),$$

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 $[-r/2, r/2]^{2\,d}$, and we can bound

$$\sum_{(k',k'')} \exp\big(-\|x - r\,k'\|_2^2\big)\,\exp\big(-\|\omega - r\,k''\|_2^2\big) \le \sum_{(k',k'')} \exp\big(-r^2\,\|k'\|_2^2/4\big)\,\exp\big(-r^2\,\|k''\|_2^2/4\big) \le 2, \tag{33}$$

for sufficiently large r.

Now, concerning (32) if $\|x\|_2 \ge r/2$ then $\|x\|_2^2 \ge \|x\|_2^2/2 + r^2/8$ and hence $e^{-(\|x\|_2^2+\|\omega\|_2^2)} \le e^{-r^2/8}\,e^{-(\|x\|_2^2+\|\omega\|_2^2)/2}$, which together with (33) implies (31). The case $|\omega| \ge r/2$ is similar.

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

$$\|x - r\,k'\|_2^2 \ge (r\,\|k'\|_2 - \|x\|_2) \ge r^2(\|k'\|_2 - 1/2)^2 \ge r^2\,(\|k'\|_2^2/2 - 1/4),$$

and hence in view of (33), and similarly for $k'' \neq 0$, and hence we can bound

$$\sum_{(k',k'')\neq(0,0)} \exp\big(-\|x - r\,k'\|_2^2\big)\,\exp\big(-\|\omega - r\,k''\|_2^2\big) \le 4\,\exp(-r^2/4).$$

□

We resume the proof of Lemma 11. Using (31) we may upper bound the second term in the bracket in (30) (which is subtracted) by

$$\begin{aligned}\sum_{(k',k'')\neq(k_1',k_1'')}^{m} &|\iota_{k',k''} - \iota'_{k',k''}|\,|\iota_{k_1',k_1''} - \iota'_{k_1',k_1''}| \cdot \exp\big(-\frac{\pi}{2}\,(\|x - r\,k'\|_2^2 + \|x - r\,k_1'\|_2^2)\big)\\ &\cdot \exp\big(-\frac{\pi}{2}\,(\|\omega - r\,k''\|_2^2 + \|\omega - r\,k_1''\|_2^2)\big)\Big)\\ \le 4\,\exp(-\pi r^2/16) &\sum_{(k',k'')} (\iota_{k',k''} - \iota'_{k',k''})^2 \cdot \exp\big(-\frac{\pi}{4}\,\|x - r\,k'\|_2^2\big)\,\exp\big(-\frac{\pi}{2}\,\|\omega - r\,k''\|_2^2\big).\end{aligned}$$

Thus, we may lower bound the difference in brackets in (30) by

$$\begin{aligned}\sum_{(k',k'')} (\iota_{k',k''} - \iota'_{k',k''})^2 \cdot \Big(&\exp\big(-\pi\,\|x - r\,k'\|_2^2\big)\,\exp\big(-\pi\,\|\omega - r\,k''\|_2^2\big)\\ &- 4\,\exp(-\pi r^2/16)\,\exp\big(-\frac{\pi}{4}\,\|x - r\,k'\|_2^2\big)\,\exp\big(-\frac{\pi}{2}\,\|\omega - r\,k''\|_2^2\big)\Big).\end{aligned}$$

Now, the integrals of $\exp\big(-\pi\,\|x - r\,k'\|_2^2\big)$, $\exp\big(-\pi\,\|\omega - r\,k''\|_2^2\big)$ and $\exp\big(-\frac{\pi}{4}\,\|x - r\,k'\|_2^2\big)$, $\exp\big(-\frac{\pi}{2}\,\|\omega - r\,k''\|_2^2\big)$ are positive and do not depend on r. Thus, by choosing r large enough we obtain the desired first inequality in (28), while the second follows from the assumption on ι and ι'.

For the upper bound, we obtain similarly that

$$\sum_{(k',k'')} (\iota_{k',k''} - \iota'_{k',k''})^2 \cdot \Big(\exp\big(-\pi \|x - r\,k'\|_2^2\big)\, \exp\big(-\pi \|\omega - r\,k''\|_2^2\big) + 4\, \exp(-\pi r^2/16)\, \exp\big(-\frac{\pi}{4} \|x - r\,k'\|_2^2\big)\, \exp\big(-\frac{\pi}{2} \|\omega - r\,k''\|_2^2\big)\Big),$$

which yields the first inequality in (29), while the second follows from $\|\iota - \iota'\|_2^2 \leq m^{2d}$. □

Bounding the modulation norm

Lemma 13. *For the functions defined in* (25)*, we have that*

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

where the constant can be made small by decreasing the constant c in (25)*.*

Proof. We have that

$$\begin{aligned} \|f_\iota\|_{\mathcal{M}^p_{m_s}}^p &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_\varphi f_\iota(x,\omega)|^p\, m_s(x,\omega)^p \,\mathrm{d}x\,\mathrm{d}\omega \\ &\leq c^p\, \varepsilon^p \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Big(\sum_{k',k''}^{m} \exp\big(-\pi\|x - r\,k'\|_2^2/2\big) \\ &\qquad \cdot \exp\big(-\pi\|\omega - r\,k''\|_2^2/2\big)\Big)^p \big(1 + \|x\|_2^2 + \|\omega\|_2^2\big)^{ps/2} \,\mathrm{d}x\,\mathrm{d}\omega. \end{aligned} \tag{34}$$

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

$$\begin{aligned} \sum_{j=1}^{m} \exp\big(-a\,(t - r\,j)^2\big) \leq 4\Big(&\sum_{j=2}^{m-1} \exp\big(-a\,(t - r\,j)^2\big)\, \mathbf{1}_{|t - r\,j| \leq r/2} \\ &+ \exp\big(-a\,(t-r)^2\big)\, \mathbf{1}_{t \leq 3r/2} + \exp\big(-a\,(t - rm)^2\big)\, \mathbf{1}_{t \geq (m - 1/2)\,r}\Big). \end{aligned} \tag{35}$$

This follows from

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

for $a^{1/2} \cdot r$ sufficiently large, which is immediate from the geometric series. The functions on the right of (35) have disjoint support, so that

$$\begin{aligned} \Big(\sum_{j=1}^{m} \exp\big(-a\,(t - r\,j)^2\big)\Big)^p &\leq 4^p \Big(\sum_{j=2}^{m-1} \exp\big(-a\,p\,(t - r\,j)^2\big)\, \mathbf{1}_{|t - r\,j| \leq r/2} \\ &\qquad + \exp\big(-a\,p\,(t-r)^2\big)\, \mathbf{1}_{t \leq 3r/2} + \exp\big(-a\,p\,(t - rm)^2\big)\, \mathbf{1}_{t \geq (m-1/2)\,r}\Big) \\ &\leq 4^p \sum_{j=1}^{m} \exp\big(-a\,p\,(t - r\,j)^2\big), \end{aligned}$$

and inserting this bound in each of the sums in (34) we bound the integral by

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big(\sum_{k',k''}^{m}\exp\big(-\pi\|x-r\,k'\|_2^2/2\big)\exp\big(-\pi\|\omega-r\,k''\|_2^2/2\big)\Big)^p\big(1+\|x\|_2^2+\|\omega\|_2^2\big)^{ps/2}\,\mathrm{d}x\,\mathrm{d}\omega$$

$$\leq \text{const.}\cdot\sum_{k',k''}^{m}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\exp\big(-p\,\pi\|x-r\,k'\|_2^2/2\big)\exp\big(-p\,\pi\|\omega-r\,k''\|_2^2/2\big)\big(1+|x_1^{ps}|+\ldots+|x_d|^{ps}$$

$$+|\omega_1^{ps}|+\ldots+|\omega_d|^{ps}\big)\,\mathrm{d}x\,\mathrm{d}\omega.$$

$$\leq \text{const.}\cdot\sum_{k',k''}^{m}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\exp\big(-p\,\pi\|x-r\,k'\|_2^2/2\big)\exp\big(-p\,\pi\|\omega-r\,k''\|_2^2/2\big)\big(1+\|x\|_2^2+\|\omega\|_2^2\big)^{ps/2}\,\mathrm{d}x\,\mathrm{d}\omega \tag{36}$$

Since for $b\geq 0$,

$$\int_{\mathbb{R}}\exp\big(-p\,\pi t^2\big)\,|t+b|^{ps}\,\mathrm{d}x\leq\text{const.}\cdot|b|^{ps},$$

we may further bound (36) by

$$\text{const.}\cdot\sum_{k_1,\ldots,k_{2d}=1}^{m}\big(1+k_1^{ps}+\ldots+k_{2d}^{ps}\big)\leq\text{const.}\cdot m^{2d-1}\cdot m^{ps+1}$$

$$\leq\text{const.}\cdot m^{ps+2d},$$

which together with (34) proves the lemma. □

Conclusion of the proof of Theorem 4. From the above paragraph we obtain as condition for $f_\iota\in\mathcal{M}^p_{s,C}$ in $d=1$ that $m^{ps+2\,d}\varepsilon^p\lesssim 1$, which leads to $m\asymp\varepsilon^{-\frac{p}{ps+2d}}$. Now (28) implies (24), and the conclusion follows as in the first paragraph of the proof. □

8.3 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 $f\in\mathcal{M}^p_{m_{u,v}}(\mathbb{R}^d)$, $p\in(0,2]$ we have the bound*

$$\sum_{\|\lambda\|_2\leq K}\min(\varepsilon^2\,\|\tilde{h}\|_{\mathcal{L}^2}^2,|\vartheta_\lambda|^2)\leq\text{const.}\cdot\|f\|_{\mathcal{M}^p_{m_{u,v}}}^{\frac{p\,d\,(v+u)}{d(v+u)+pvu}}\,\varepsilon^{\frac{(2-p)\,d\,(v+u)+2pvu}{d(v+u)+pvu}}. \tag{37}$$

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 (23) we have that

$$\sum_{\|\lambda\|_2\leq K}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)\leq\sum_{\|x\|_2\leq K_u,\|\omega\|_2\leq K_v}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)+\sum_{\|x\|_2>K_u}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)+\sum_{\|\omega\|_2>K_v}\min(\varepsilon^2\,,|\vartheta_\lambda|^2)$$

$$\leq\varepsilon^2\,C\,K_v^d\,K_u^d+\tilde{C}_2^p\,\|f\|^p_{\mathcal{M}^p_{m_{u,v}}}\,\varepsilon^{2-p}\,\big(K_u^{-up}+K_v^{-vp}\big)$$

Balancing the terms with

$$K_v=\text{const.}\cdot\|f\|_{\mathcal{M}^p_{m_{u,v}}}^{\frac{pu}{d(v+u)+pvu}}\,\varepsilon^{-\frac{pu}{d(v+u)+pvu}},$$

$$K_u=\text{const.}\cdot\|f\|_{\mathcal{M}^p_{m_{u,v}}}^{\frac{pr}{d(v+u)+pvu}}\,\varepsilon^{-\frac{pv}{d(v+u)+pvu}},$$

we get the result. □

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

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

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 $f_j \in \mathcal{M}^p_{u,v,C}$, where $M \asymp \exp(c\,m_1^d\,m_2^d)$ for some $c > 0$, such that

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

The conclusion then follows as in the proof of Theorem 4.
Again we consider the Gaussian function $\varphi(x) = \exp\big(-\pi\,\|x\|_2^2\big)$, and given $m_1, m_2 \in \mathbb{N}$, using the Varshamov-Gilbert bound (Tsybakov, 2009, Lemma 2.9) we may choose $M = \exp(c_1\,m_1^d\,m_2^d)$, $c_1 = (\log 2)/8 > 0$ indices

$$\iota = \big(\iota_{(k_1,\ldots,k_{2d})}\big)_{k_1,\ldots,k_d=1,\ldots,m_1,k_{d+1},\ldots,k_{2d}=1,\ldots m_2} \in \{0,1\}^{(m_1^d\,m_2^d)}$$

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

$$\begin{aligned} f_\iota &= c\cdot\varepsilon \sum_{k_1,\ldots,k_2=1}^{m_1} \sum_{k_{d+1},\ldots,k_{2d}=1}^{m_2} \iota_{(k_1,\ldots,k_{2d})^\top}\,T_{r\,(k_1,\ldots,k_d)^\top}\,M_{r\,(k_{d+1},\ldots,k_{2d})^\top}\,\varphi \\ &= c\cdot\varepsilon \sum_{k',k''} \iota_{k',k''}\,T_{r\,k'}\,M_{r\,k''}\,\varphi, \end{aligned} \tag{38}$$

where we denote $k' = (k_1,\ldots,k_d)^\top$ and $k'' = (k_{d+1},\ldots,k_{2d})^\top$.
Bounding the $\mathcal{L}_2$-distance

Analoguously to Lemma 11, we have that

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

$$\begin{aligned} &\|V_\varphi\,f_\iota - V_\varphi f_{\iota'}\|^2_{\mathcal{L}^2} \geq \tilde{c}_1\,\varepsilon^2\,\|\iota - \iota'\|_2^2 \geq \tilde{c}_1\,\varepsilon^2\,m_1^d\,m_2^d/8, \\ &\|V_\varphi\,f_\iota - V_\varphi f_{\iota'}\|^2_{\mathcal{L}^2} \leq \tilde{c}_2\,\varepsilon^2\,\|\iota - \iota'\|_2^2 \leq \tilde{c}_2\,\varepsilon^2\,m_1^d\,m_2^d. \end{aligned} \tag{39}$$

Bounding the modulation norm

Lemma 17. *For the functions defined in* (38)*, we have that*

$$\|f_\iota\|^p_{\mathcal{M}^p_{m_{u,v}}} \leq \textit{const.}\cdot\,\varepsilon^p\,(m_1^d\,m_2^d)\,(m_1^{pu} + m_2^{pv})\,,$$

where the constant can be made small by decreasing the constant c in (38)*.*

For the choices m_1 and m_2 above, the upper bound in Lemma 17 remains bounded, and inserting these choices in (39) gives the rate. □

8.4 Proofs of Section 5

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

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

where we used (6) in the last step. Hence

$$\mu \le N_\mu^{-1/p}\, \tilde{C}_2\, \|f\|_{\mathcal{M}^p(\mathbb{R}^d)}. \tag{40}$$

Further, for $\lambda \in I_\mu^{\complement}$ we have that

$$|\langle f, \tilde{h}_\lambda\rangle|^2 \le \mu^{2-p}\, |\langle f, \tilde{h}_\lambda\rangle|^p \tag{41}$$

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

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

Using (6) we obtain

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

and inserting the bound (40) yields the result. □

Proof of Theorem 7. From (4) we have the bound

$$\|g\|^p_{\mathcal{M}^p_m(\mathbb{R}^d)} \le \bar{C} \cdot \sum_{\lambda\in\Lambda} |c_\lambda|^p\, m(\lambda)^p.$$

Further we may write

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

where D_h is the Gabor synthesis operator associated with $\{h_\lambda \mid \lambda \in \Lambda\}$. From Gröchenig (2013, Prop. 5.1.1 (b)) we obtain the estimate

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

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

9 Proofs

9.1 Proof of Theorem 1

Proof. Following Galperin and Samarah (2004, Theorem 3.1), it is sufficient to show

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

with

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

We therefore need $V_h h \in \mathcal{L}^{p^*}_{w_{r,s}}$. Since (see, e.g., Gröchenig, 2013, Lemma 3.1.1)

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

we have

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

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

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

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 $\hat{f_\omega}$ 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

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

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

$$\|V_h h\|^{p^*}_{\mathcal{L}^{p^*}_{w_{r,s}}} \lesssim \int_{\mathbb{R}} (1+|\omega|)^{sp^*} \|f_\omega\|^{p^*}_{\mathcal{L}^q} \mathrm{d}\omega. \tag{42}$$

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

$$N_\omega = \left\{ t \in \mathbb{R} \;\middle|\; \left| t - \frac{\omega}{2} \right| \le \frac{|\omega|}{4} \right\}.$$

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

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

where we used the integrability of $\left|\frac{\sin(x)}{x}\right|^{\alpha}$ for $\alpha > 1$ in the last step.
Inserting this into (42), we get

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

In the second summand we integrate a continous function over a compact set, this integral is always finite. Therefore, we have $\|V_h h\|_{\mathcal{L}^{p^*}_{w_{r,s}}} < \infty$, 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 (6) for all B-spline window functions of order $k > 1/p$. □

References

Benedek, A. and R. Panzone (1961). The space l_p with mixed norm. *Duke Mathematical Journal 28*(3), 301–324.

Candes, E. J. (2006). Modern statistical estimation via oracle inequalities. *Acta Numerica 15*, 257–325.

Chambolle, A., R. A. De Vore, 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.

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.

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.

Daubechies, I., M. Defrise, and C. Mol (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constrains. *Communications on Pure and Applied Mathematics 57*.

DeVore, R. and B. Lucier (1992). Fast wavelet techniques for near-optimal image processing. In *MILCOM*, Volume 92, pp. 11291135. IEEE.

DeVore, R. A., B. Jawerth, and V. Popov (1992). Compression of wavelet decompositions. *American Journal of Mathematics 114*(4), 737–785.

Doerfler, M. (2001). Time-frequency analysis for music signals: A mathematical approach. *Journal of New Music Research 30*, 3–12.

Donoho, D. L. and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association 90*(432), 1200–1224.

Donoho, D. L. and J. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*(3), 425–455.

Engl, H., M. Hanke, and A. Neubauer (2000). *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands.

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.

Gröchenig, K. (2013). *Foundations of Time-Frequency Analysis*. Springer Science & Business Media.

Gröchenig, K. and S. Samarah (2000). Nonlinear approximation with local fourier bases. *Constructive Approximation 16*(3), 317–331.

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*.

Johnstone, I. M. (2013). Gaussian estimation: Sequence and wavelet models. *Unpublished manuscript*.

Kerkyacharian, G., O. Lepski, and D. Picard (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probability Theory and Related Fields 121*(2), 137–170.

Kerkyacharian, G., O. Lepski, and D. Picard (2008). Nonlinear estimation in anisotropic multi-index denoising. sparse case. *Theory of Probability & Its Applications 52*(1), 58–77.

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.

Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *The Annals of Statistics 36*(4), 1957–1982.

Samarah, S. and S. Al-Sa'di (2006). Characterization of modulation spaces by nonlinear approximation. *arXiv preprint math/0602250*.

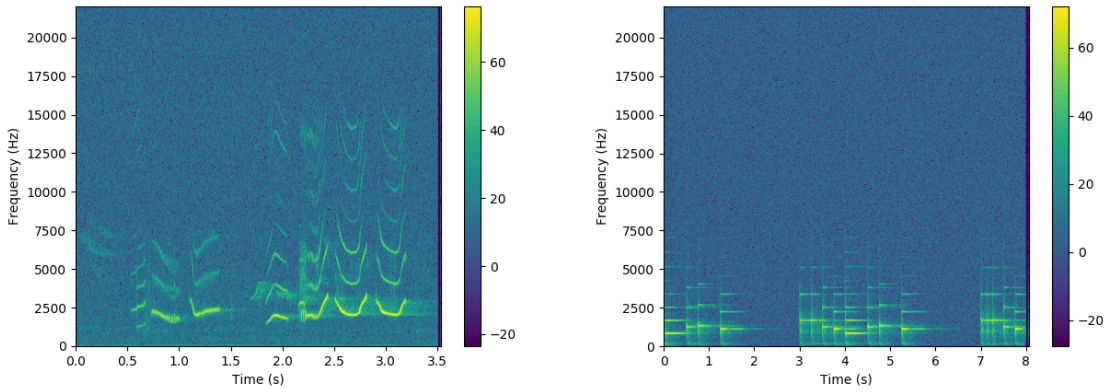
Strohmer, T. (2001). Approximation of dual gabor frames, window decay, and wireless communications. *Applied and Computational Harmonic Analysis 11*(2), 243 – 262.

Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*, Volume 606. Springer series in statistics.

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.

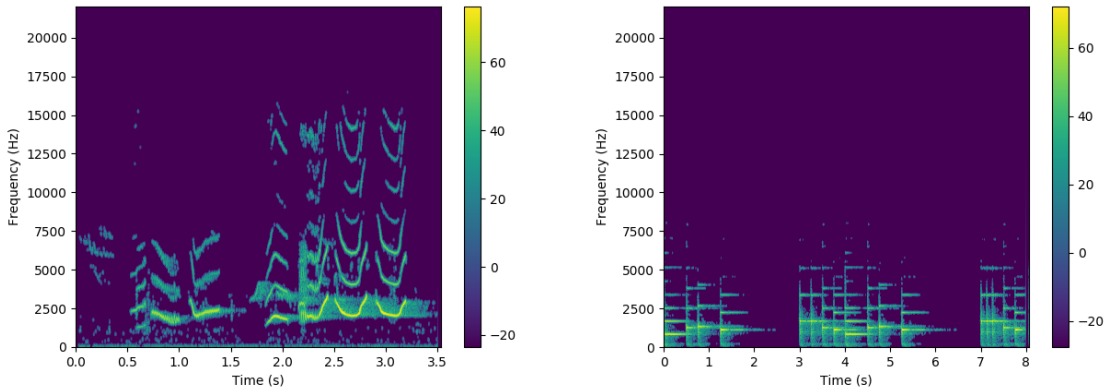
Supplement: For review only

9.2 Figures for the blackbird and piano recording



(a) blackbird, noisy recording

(b) Melody, noisy recording



(c) blackbird, reconstruction

(d) Melody, reconstruction

Figure 10: Spectrograms of blackbirds and piano recording with additive noise and SNR = 30 dB in (a) and (b), and after performing a denoising based on hard-thresholding in (c) and (d).

9.3 Additional simulations

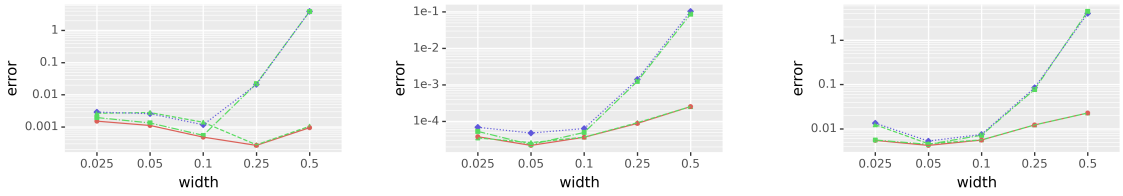
Effects of window length and grid density

We also analyze the error from an irregular grid, i e. unequal amount of frequency and time bands, $M \neq N$. Giving the characteristics of some functions with signals being located in a single time or frequency fragment, it might be useful to consider an amplification of the

time-frequency representation in either time or frequency plane. Figure 11 shows the MSE for four different scenarios according to table 5.

Scenario	S1	S2	S3	S4
N	400	400	100	100
M	400	100	400	100
$\alpha \cdot \beta$	0.0125	0.05	0.05	0.3125

Table 5: Gabor grid density



(a) $f_{50,2}$ (b) $f_{50000,4}$ (c) $f_{40,200}$

Figure 11: Effect of the window width w and the grid density. For the three signals for samples of sizes $n = 2000$, using hard thresholding with the optimal threshold. —S1, – –S2, –·–S3, $\cdots$ S4

For all signals we observe that S2 performs better than S3 for a long window and S3 performs better than S2 for a short window. An increased discretization of the time domain is required as the window grows to make up for the loss in time resolution. Similarly, an increased discretization of the frequency domain is required as the window decreased to make up for the loss in the frequency resolution. For $f_{50,2}$, scenario S3 perform as good S1, which means that the lost of local information in the time domain in S3 does not affect the denoising method since the signal is located only in a small frequency domain. Scenario S2 and S4 also perform equal. This means that a gain in local information in the time domain does not improve the denoising of $f_{50,2}$.

For $f_{50000,4}$ we observed opposite results. Scenario S2 performs as good as S1 and better than S3 which in turn perfoms almost as good as S4. This is not surprising given that the signal lies in a small time fragment. Therefore an increasing of the local information in the frequency domain does not lead to better results.