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.

Supplement

A. *Additional results for denoising the signals from Section V-C in the main paper*

We provide some additional simulation results for the setting from Section V-C in the main paper, in particular for SureShrink, BayesShrink and musical noise block soft thresholding.

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

B. *Additional simulations*

Effects of window length and grid density

In the setting of Section V-A in the main paper, 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 1 shows the MSE for four different scenarios according to table III.

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

audio	SNR	SureShrink			BayesShrink		
		STFT$_{rect}$	STFT$_{gauss}$	wavelet$_{db4}$	STFT$_{rect}$	STFT$_{gauss}$	wavelet$_{db4}$
blackbird	-10	8.339	**9.128**	2.9	7.921	**8.394**	2.836
	1	14.918	**16.3**	10.018	13.588	**14.112**	9.506
	10	20.356	**21.973**	16.559	19.199	**19.77**	15.902
	30	32.968	**34.11**	31.956	33.406	**34.112**	31.586
melody	-10	8.196	**8.35**	3.938	8.041	**8.043**	3.913
	1	14.911	**15.708**	11.346	13.694	**13.855**	11.096
	10	20.681	**22.21**	18.113	20.283	**20.801**	17.894
	30	34.174	**37.831**	34.57	34.543	**37.427**	34.436
ECG	-10	3.452	**3.694**	3.556	4.344	**4.486**	3.791
	1	11.551	**11.685**	11.25	11.283	**11.644**	11.036
	10	17.005	**18.44**	17.838	16.926	**18.286**	17.647
	30	30.057	30.074	**30.953**	30.358	30.786	**31.008**
guitar	-10	4.171	**4.212**	3.81	**4.907**	4.768	3.784
	1	11.548	**11.829**	9.546	11.218	**11.274**	9.541
	10	17.77	**18.184**	16.443	17.966	**18.106**	16.357
	30	33.061	**35.98**	33.323	33.621	**35.832**	33.242
human speech	-10	**6.713**	6.576	4.096	**6.925**	6.862	4.119
	1	13.773	**14.281**	11.195	13.151	**13.263**	10.958
	10	20.05	**20.818**	18.046	19.887	**20.121**	17.702
	30	33.834	**37.344**	34.42	34.524	**37.221**	34.134

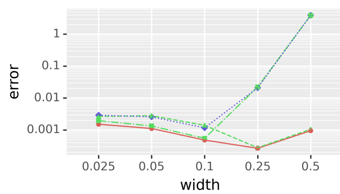
TABLE I: SureShrink and BayesShrink. All values in decibel (10 $\log_{10} x$, dB).

audio	SNR	SureShrink			BayesShrink		
		STFT$_{rect}$	STFT$_{gauss}$	wavelet$_{db4}$	STFT$_{rect}$	STFT$_{gauss}$	wavelet$_{db4}$
blackbird	-10	**5.665**	4.901	1.459	7.253	**7.318**	1.81
	1	**13.536**	12.565	7.667	14.487	**14.774**	8.481
	10	**19.632**	19.274	13.897	20.295	**20.854**	14.578
	30	32.68	**33.448**	30.239	33.53	**34.136**	30.939
melody	-10	5.251	**5.545**	2.578	7.202	**7.48**	2.762
	1	13.462	**13.505**	10.112	14.434	**14.814**	10.547
	10	20.15	**20.374**	16.829	20.812	**21.415**	17.191
	30	32.773	**37.029**	33.739	34.887	**37.728**	33.979
ECG	-10	0.989	1.122	**2.35**	2.484	2.618	**2.784**
	1	9.876	**10.047**	9.225	10.482	**10.715**	10.125
	10	16.633	**17.61**	15.639	16.644	**17.91**	16.643
	30	30.018	30.074	**30.949**	30.158	30.546	**31.023**
guitar	-10	3.516	**3.842**	3.536	4.301	**4.418**	3.557
	1	10.573	**10.58**	8.688	11.297	**11.366**	8.565
	10	17.827	**17.896**	15.893	18.22	**18.332**	16.024
	30	31.369	**34.708**	32.907	33.896	**35.872**	33.032
human speech	-10	5.324	**5.531**	3.446	6.376	**6.438**	3.55
	1	**12.835**	12.739	9.875	13.484	**13.578**	10.277
	10	**19.701**	19.663	16.364	20.254	**20.422**	16.957
	30	32.119	**36.801**	33.219	34.893	**37.56**	33.654

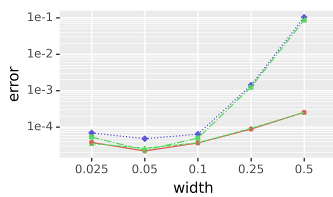
TABLE II: Musical noise block soft thresholding. All values in decibel (10 $\log_{10} x$, dB).

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

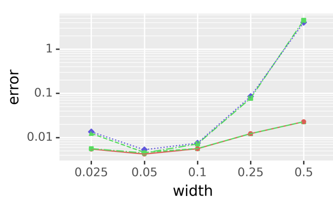
TABLE III: Gabor grid density



(a) $f_{50,2}$



(b) $f_{50000,4}$



(c) $f_{40,200}$

Fig. 1: 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

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.