The Affine Uncertainty Principle, Associated Frames and Applications in Signal Processing

Florian Lieb

The Affine Uncertainty Principle, Associated Frames and Applications in Signal Processing

Florian Lieb

Dissertation
zur Erlangung des Grades eines Doktors der Ingenieurwissenschaften
- Dr.-Ing. -

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im März 2018

Datum des Promotionskolloquiums: 25. Juli 2018

Gutachter:
Prof. Dr. Peter Maaß (Universität Bremen)
Prof. Dr. Hans-Georg Stark (Hochschule Aschaffenburg)

Für Jutta, Hanna & Moritz

Contents

Preface

This journey would not have been possible without the support and help of some people, who need to be mentioned here. First, I would like to thank the people from the BioMEMS lab for assistance with the processing of MEA data. Furthermore, I want to thank Daniel Lantzberg for always taking the time to discuss my problems and proofreading of this thesis.

I would like to thank Prof. Dr. Hans-Georg Stark for encouraging me to pursue my own ideas and for supervising my work in the past years. I also want to thank Prof. Dr. Peter Maaß for providing the opportunity of writing this thesis.

Most of all, I want to thank my family, especially my wife, for her endless support and for reminding me every once in a while what the really important things in my life are: *my family*.

Florian Lieb
Aschaffenburg, September 2018

Chapter 1

Introduction

1.1 Motivation

The succinct term *uncertainty principle* was first characterized by Werner Heisenberg in the early 20th century. It has become famous for being a characteristic feature of all quantum mechanical systems, limiting the ability to measure physical properties of subatomic particles without altering the system being measured. Heisenberg demonstrated, that position and velocity of such particles cannot be measured with absolute precision, or to quote Heisenberg:

> *Any use of the words "position" and "velocity" with an accuracy exceeding that given by [the uncertainty principle] is just as meaningless as the use of words whose sense is not defined.*
>
> — W. Heisenberg, *The Physical Principles of the Quantum Theory*, 1930

The link to signal processing was closely associated in Heisenberg's reasoning, since Fourier analysis was essential in his interpretation (Aitchison et al., 2004). In Fourier analysis the uncertainty relation states that a function and, simultaneously, its Fourier transform cannot be sharply concentrated. The similarity between Heisenberg's and this uncertainty relation is not coincidental, as position and momentum are related via the Fourier transform. Since then, various kinds of uncertainty inequalities have been derived based on Heisenberg's fundamental principle: the general Robertson-Schrödinger inequality, the Hirschman-Beckner entropic inequality, the Donoho-Stark inequalities or Lieb's inequality which all differ in their concept of measuring concentration or spread of a function, to name only a few examples mentioned by Ricaud and Torrésani (2014).

Current developments in signal processing still consider uncertainty principles at the core of recent research interests. With ever growing amounts of data the efficiency of sophisticated signal processing algorithms requires optimal localization properties of underlying function systems. Finding appropriate localization measures in order to obtain functions with such properties and deriving discretization schemes based on these functions has been one of the main challenges of the research project *UNLocX - Uncertainty principles versus localization properties, function systems for efficient coding schemes* funded by the European Commission (2010 - 2013). A major part of the theoretical foundation of this project was based on the observation that in particular situations the uncertainty principle can be circumvented (Maaß et al., 2010). This does not imply, however, that meaningless words have been used. On the contrary, it had led to a more detailed analysis of uncertainty principles, most importantly the affine uncertainty principle associated to the wavelet transform, which is where the violation was first observed.

Since one of the first publications by Dahlke and Maaß (1995), various affine uncertainty principles have been introduced (Flandrin, 2001). Each of them results in a different minimizing waveform for the one-dimensional case: a general solution is given in (Dahlke and Maaß, 1995, Thm. 3.1) from which the Klauder wavelet is derived by Flandrin (2001, Prop. 10.2.1) and using a slightly different localization measure leads to the Altes wavelet (Flandrin, 2001, Prop. 10.2.3). Based on the shortcomings revealed by Maaß et al. (2010), Levie et al. (2014) derived a general concept of the affine uncertainty principle. It resulted in yet another, not minimizing, but equalizing waveform. As one of the latest results emerging from the UNLocX project, specific discretization schemes based on this equalizing waveform are still inadequately covered. In particular, the construction of wavelet frames, even tight wavelet frames, is either restricted to compactly supported functions or lacking proper dual frames which are wavelet frames themselves (Feichtinger et al., 2012).

Such frame discretizations, on the other hand, are needed in sophisticated signal processing applications. For example, audio processing algorithms might benefit from better localized function systems as important features can be represented more sparsely. This leads to an increased robustness with respect to noise or missing audio samples. Applications in life sciences may also benefit from frame constructions with optimal localization properties. Imaging mass spectrometry provides a thorough analysis of metabolic patterns essential for pharmaceutical, medical (especially pathological) and industrial research. Observing neuronal activity using microelectrode arrays gives insights to intercellular communication of neuronal cells from which many biomedical and pharmaceutical applications benefit. Although these two applications are by far not the only ones, they may benefit from sparse and localized function systems alleviating

the distinction between important signal characteristics and noise. Sparsely localized signal representations can improve separation, in particular if temporal signal features and noise characteristics are similar in shape and amplitude. As the two life science approaches are essential to this thesis, they are briefly introduced next.

1.2 Microelectrode Arrays

Electrophysiology is the part of neuroscience which analyzes electrochemical processes linking neurons and muscle tissue. In the human body, for example, stimulating muscles, physical sensations and even thoughts are based primarily on electrical signals called action potentials (Aach et al., 2014). With clinically well established methods like electrocardiography (ECG/EKG), electroencephalography (EEG) or electromyography (EMG), such signals are already used for diagnostic purposes of heart, brain or muscle. However, not all electrophysiological issues can be investigated on the human body, for example in the context of pharmaceutical and toxicological research. Instead, in vitro approaches are increasingly utilized to gain insights into intracellular communication as well as cellular responses to chemical or physical stimuli. They offer an alternative methodology to animal testing, which might be faster, more efficient and morally acceptable (Daus et al., 2012).

An elaborated non-invasive method for cellular electrophysiological characterization of in vitro models are microelectrode arrays (MEAs). MEAs are essentially just planar electrodes arranged into a matrix with distances between electrodes in the sub-millimeter range onto which immobilized cellular tissue such as neural networks are coupled. This allows a long-term and spatially resolved analysis of neuronal activity by detecting voltage differences between cell and electrode resulting from neuronal ion currents. Intensities of these action potentials are inversely related to the gap between neurons and corresponding electrode (Liu and Wang, 2009). The shape is further affected by the culture medium between electrode and cell so that transduced voltages do not reflect transmembrane potentials anymore (Daus et al., 2012). Additionally, recorded signals are superimposed by noise from biochemical processes and thermal or extrinsic noise, which complicates exact localization of action potentials. These localizations, however, are fundamental for subsequent spike and burst analysis. Bursts and burst rates are a common feature of neuronal networks indicating synchronization or carrying information (Obien et al., 2015). These can be triggered intrinsically by the network itself, or extrinsically provoked by chemical stimulus. Hence, burst patterns might be directly related to treatments with specific drugs (Chiappalone et al., 2003; El Hady et al., 2013; Martinoia et al., 2005).

With advances in lithography the increasing number of electrodes per mm^2 facilitates a more detailed electrophysiological analysis (Lonardoni et al., 2015; Wark et al., 2013). Unfortunately, these high-density MEAs also increases the noise contribution and reliably separating action potentials from background noise becomes more and more challenging requiring efficient algorithms based on localized function systems.

1.3 MALDI Imaging

Imaging mass spectrometry enables detailed analysis of tissue sections exposing DNA, proteins, peptides, metabolites or lipids in a spatial and morphological context. The fields of application is manifold: biomarker analysis, for example, to evaluate drug responses in pharmacological drug development, tumor classification in pathological applications in order to identify cancer subtypes or the origin of tumorous cells, identification of metabolites in forensic toxicology, or even characterization of proteins in wine to analyze polyphenols responsible for taste and color (Gross and Beifuss, 2012). One of the most used imaging mass spectrometry techniques is MALDI imaging: Matrix-Assisted Laser Desorption/Ionization. It allows analyzing molecular compositions of tissue while retaining spatial distributions.

Before a tissue section can be analyzed through MALDI imaging, it has to be prepared such that analytes can be desorbed and ionized. This is done by applying a matrix solution onto the tissue, which absorbs the main energy of laser pulses and protects the tissue from photolytic decomposition while at the same time transferring necessary energy for ionization onto tissue molecules (Karas and Hillenkamp, 1988). After ionization, individual singly charged molecules are accelerated by a constant electrical field. At the end of the mass analyzer a detector counts arriving molecules, which, depending on their masses, need a different amount of time to cover the flight path. Instead of specifying the actual time of flight, however, the mass-to-charge ratio m/z is used as it simplifies the interpretation of the measured data. This method is called MALDI-TOF (time-of-flight). Currently, there are two different MALDI-TOF approaches available, differing mainly in the flight path. In linear MALDI-TOF the flight path is straight: after acceleration molecules drift to the detector on a straight trajectory. Ionized molecules in MALDI-TOF instruments in reflection mode are at least once reflected on their way to the detector, increasing mass resolution (Flensburg et al., 2004; Goodwin et al., 2008).

With a current minimum lateral resolution of 10 µm, laser pulses generate a single spectrum for each spatial tissue spot (Ogrinc Potočnik et al., 2015). This procedure is visualized in Figure 1.1. The hereby obtained number of spots is in the magnitude of 10^4 in the 2D case

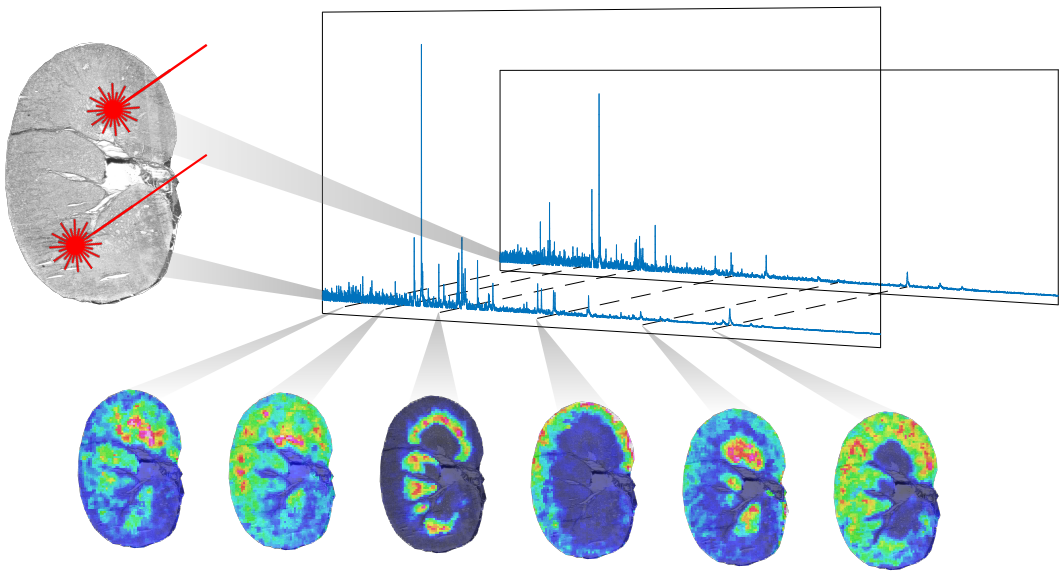


Figure 1.1: Basic MALDI imaging principle displaying spatially resolved mass spectra of a rat kidney (SCiLS GmbH, Bremen).

and 10^6 in 3D cases. In addition, each spectrum contains 10^3 - 10^5 data points depending whether linear or reflector mode is used (Kobarg, 2014). Not only does this lead to considerable amounts of MALDI data, but also visualization and analysis is challenging and requires efficient algorithms extracting only the most important information. Spatial segmentation, for example, can be used for unsupervised mining of large MALDI imaging data sets in order to reveal prominent features. Hereby, spectra are grouped into different clusters based on hierarchical and statistical similarities (Alexandrov and Kobarg, 2011; Alexandrov et al., 2010; Deininger et al., 2008). These clusters can be visualized as a segmentation map with different colors for different clusters and regions of interest can be easily associated with corresponding m/z values.

Another example requiring efficient algorithms is the co-localization of MALDI imaging data with annotated regions. The distinction of different tissue types is of great interest in histopathological research. Finding m/z-markers with high intensities in specific annotated regions and low intensities in other regions (co-localization) helps discriminating healthy and tumorous tissue sections (Alexandrov, 2012). Regardless whether MALDI imaging data is spatially segmented or co-localized regions are estimated, a reliable separation of important features and noise is of fundamental importance.

1.4 Publications and Roadmap of this Thesis

During the course of this thesis, some parts presented in the following have already been published or are submitted for publication:

- Levie, R., Stark, H.-G., Lieb, F. and Sochen, N. (2014). "Adjoint translation, adjoint observable and uncertainty principles". *Advances in Computational Mathematics*, 40(3):609-627.

- Lieb, F. (2015). "Audio inpainting using M-frames". In: *Current Trends in Analysis and its Applications: Proceedings of the 9th ISAAC Congress, Kraków 2013*, Springer International Publishing, 705-713.

- Lieb, F., Stark, H.-G. and Thielemann, C. (2017). "A stationary wavelet transform and a time-frequency based spike detection algorithm for extracellular recorded data". *Journal of Neural Engineering*, 14(3):36013.

- Mayer, M., Arrizabalaga, O., Lieb, F., Ciba, M., Ritter, S. and Thielemann, C. (2018). "Electrophysiological investigation of human embryonic stem cell derived neurospheres using a novel spike detection algorithm". *Biosensors and Bioelectronics*, 100(Supplement C):462-468.

- Lieb, F. and Stark, H.-G. (2018). "Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches". *Signal Processing*, 153:291 – 299.

Additionally, the following European patent is pending:

- Lieb, F., Hochschule Aschaffenburg (2015). "Verfahren zur Analyse eines Datensatzes einer Flugzeit-Massenspektrometrie-Messung", Deutsches Patentamt, 10 2015 010 602.3.

The thesis can be structured into the following parts. First, a brief preliminary introduction summarizing the basic mathematical concepts which are used throughout the thesis is given. Subsequently, a construction scheme of approximately tight wavelet frames based on non-compactly supported windows is described in Chapter 3. This scheme relies on an extension of nonstationary Gabor frames to frame systems without compact support. Sufficient conditions are derived for which corresponding approximate dual frames lead to reconstruction errors within machine precision. Moreover, fast algorithms for analysis and synthesis based on such wavelet frames are introduced.

The following parts consider the application of such wavelet frame constructions to diverse signal processing approaches. In Chapter 4, such frames are used to show that missing samples,

either randomly or consecutively distributed, can be more efficiently recovered using these frame constructions than with current time-frequency representations. This is numerically justified by comparing various algorithms as well as certain sparse representations.

Chapter 5 is focused on spike detection in noisy extracellular recorded data. Current state-of-the-art algorithms are compared with two newly proposed approaches based on different simulated data. Furthermore, a novel spike detection approach is introduced which enables the usage of the proposed algorithms on real MEA recordings.

In Chapter 6, a novel peak picking algorithm for noisy MALDI-TOF data is presented. It is based on the sparse approximation of frame multipliers, with the additional option to include spatial information in the peak picking process. Performance of this method is evaluated on simulated as well as real data sets.

Finally, the last chapter concludes with summarizing the main results, illustrating the common thread among all parts of this thesis and discussing open issues and possible future directions.

Chapter 2

Preliminaries

In this chapter important mathematical notations used throughout this thesis are recalled. Whenever possible a continuous setting is used in order to avoid the heavy notation of a finite dimensional setting. In such continuous settings functions are generally assumed to be square integrable, i.e., $f \in L^2(\mathbb{R})$ with inner product

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x)\overline{g}(x)dx. \tag{2.0.1}$$

The Fourier transform of a square integrable function is defined as follows.

Definition 2.0.1 (Fourier Transform)**.** The Fourier transform on $L^2(\mathbb{R})$ is defined by

$$(\mathcal{F}f)(\nu) = \hat{f}(\nu) = \int_{\mathbb{R}} f(x)e^{-2\pi i \nu x}dx, \tag{2.0.2}$$

where $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ denotes the Fourier operator. The operator $\mathcal{F}^{-1}$ denotes the inverse Fourier transform defined as

$$\left(\mathcal{F}^{-1}\hat{f}\right)(x) = \int_{\mathbb{R}} \hat{f}(\nu)e^{2\pi i \nu x}d\nu. \tag{2.0.3}$$

A fundamental property of the Fourier transform is given by Plancherel's theorem, showing that the Fourier operator $\mathcal{F}$ is unitary:

$$\langle f, g \rangle = \left\langle \hat{f}, \hat{g} \right\rangle, \forall f, g \in L^2(\mathbb{R}). \tag{2.0.4}$$

Consequently, the Fourier transform preserves energy since $\|f\|_2 = \|\mathcal{F}f\|_2$.

Apart from the Fourier operator, the following three linear operators are repeatedly used throughout this thesis:

- The translation or time shift operator $T_a : L^2(\mathbb{R}) \to L^2(\mathbb{R})$

$$T_a f(x) = f\,(x-a)\,, \tag{2.0.5}$$

  for some time shift parameter $a \in \mathbb{R}$.

- The modulation or frequency shift operator $M_b : L^2(\mathbb{R}) \to L^2(\mathbb{R})$

$$M_b f(x) = e^{2\pi i b x} f\,(x)\,, \tag{2.0.6}$$

  for some frequency shift parameter $b \in \mathbb{R}$.

- The dilation or scaling operator $D_d : L^2(\mathbb{R}) \to L^2(\mathbb{R})$

$$D_d \psi(x) = \frac{1}{\sqrt{d}}\,\psi\left(\frac{x}{d}\right), \tag{2.0.7}$$

  for some scaling parameter $d > 1$.

These operators satisfy the following relations (Christensen, 2016, Eq. 2.29):

$$\mathcal{F}T_a = M_{-a}\mathcal{F}, \quad \mathcal{F}M_b = T_b\mathcal{F}, \quad \mathcal{F}D_d = D_{\frac{1}{d}}\mathcal{F}. \tag{2.0.8}$$

Another central concept in this thesis are frames, a generalization to orthonormal bases defined as follows.

Definition 2.0.2 (Frames in Hilbert Spaces)**.** A sequence $\{g_k\}_{k\in\mathbb{Z}}$ of elements in a Hilbert space $\mathcal{H}$ is a frame for $\mathcal{H}$, if there exists constants $A, B > 0$ such that

$$A\,\|f\|_2^2 \leq \sum_{k\in\mathbb{Z}} |\langle f, g_k\rangle|^2 \leq B\,\|f\|_2^2\,, \tag{2.0.9}$$

for every $f \in \mathcal{H}$. A frame is said to be tight whenever the frame bounds are equal, that is $A = B$. A tight frame with $A = B = 1$ is called a Parseval frame.

Closely associated to frames are corresponding analysis, synthesis and frame operators (Christensen, 2016; Gröchenig, 2001) defined by

- The *analysis* operator $\Phi : \mathcal{H} \to \ell^2(\mathbb{Z})$ reads

$$\Phi f = \{\langle f, g_k \rangle\}_{k \in \mathbb{Z}}. \tag{2.0.10}$$

- The *synthesis* operator $\Phi^* : \ell^2(\mathbb{Z}) \to \mathcal{H}$ is defined by

$$\Phi^* \{c_k\}_{k \in \mathbb{Z}} = \sum_{k \in \mathbb{Z}} c_k g_k. \tag{2.0.11}$$

- The composition of both operators results in the *frame* operator $S : \mathcal{H} \to \mathcal{H}$

$$Sf = \Phi^* \Phi f = \sum_{k \in \mathbb{Z}} \langle f, g_k \rangle \, g_k. \tag{2.0.12}$$

  The notation $S_{g,\gamma} = \sum_{k \in \mathbb{Z}} \langle f, g_k \rangle \, \gamma_k$ is used whenever the frame operator results from analysis and synthesis operators with frames $\{g_k\}_{k \in \mathbb{Z}}$ and $\{\gamma_k\}_{k \in \mathbb{Z}}$.

The concept of frames also plays a significant role in time-frequency representations (Gröchenig, 2001, Ch. 5) as well as wavelet theory (Daubechies, 1992) when discretizing the short-time Fourier and wavelet transform defined as follows.

Definition 2.0.3 (Short-Time Fourier Transform)**.** The short-time Fourier transform (STFT) of a function $f \in L^2(\mathbb{R})$ with respect to some window function $g \in L^2(\mathbb{R})$ is defined by

$$V_g f(a, b) = \langle f, T_a M_b g \rangle \tag{2.0.13}$$

$$= \int_{\mathbb{R}} f(x) \overline{g}\,(x - a)\, e^{-2\pi i b x} dx, \tag{2.0.14}$$

for $a, b \in \mathbb{R}$.

Definition 2.0.4 (Continuous Wavelet Transform)**.** The continuous wavelet transform (CWT) of a function $f \in L^2(\mathbb{R})$ with respect to some admissible function $\psi \in L^2(\mathbb{R})$ is defined by

$$W_\psi f(a, d) = \langle f, T_a D_d \psi \rangle \tag{2.0.15}$$

$$= \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{d}} \overline{\psi} \left(\frac{x - a}{d} \right) dx, \tag{2.0.16}$$

for $a, d \in \mathbb{R}$ and $d > 1$. Admissibility is given whenever

$$C_\psi = \int_{\mathbb{R}} \left|\hat{\psi}(\nu)\right|^2 \frac{d\nu}{|\nu|} < \infty. \tag{2.0.17}$$

Although basic mathematical principles in this thesis are described in a continuous setting, the proposed algorithms and applications require a finite dimensional setting. When dealing with such signals of finite length, difficulties may emerge at the boundaries. Strang and Nguyen (1996) considered various possibilities how a signal might be extended beyond its boundary. Feichtinger and Strohmer (1998, Ch. 8) argued that in most cases circular extensions are advantageous, implying that a finite sequence $f \in \mathbb{C}^L$ is extended to an infinite one by periodization: $f[n + mL] = f[n]$ for $n = 0, 1, \ldots, L-1$ and $m \in \mathbb{Z}$. This requires, however, that time and frequency shift parameters $a, b \in \mathbb{N}$ of the discrete translation and modulation operator, $T_{ak} f[n] = f[n - ak]$ and $M_{bl} f[n] = e^{2\pi i bln/L} f[n]$ for $f \in \mathbb{C}^L$ and $k, l \in \mathbb{Z}$, are integer factors of L. These periodic signal extensions induce circular convolutions which can be efficiently evaluated using a fast Fourier transform (fft) algorithm (Aldroubi and Unser, 1996, Ch. 2.2.3), with the discrete Fourier transform defined as follows.

Definition 2.0.5 (Discrete Fourier Transform)**.** Let $f \in \mathbb{C}^L$. The L-dimensional discrete Fourier transform (DFT) $\mathcal{F}_L : \mathbb{C}^L \to \mathbb{C}^L$, is defined as

$$(\mathcal{F}_L f)[n] = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} f[l] \omega_L^{nl}, \tag{2.0.18}$$

where $\omega_L = e^{-2\pi i/L}$ is the L-th primitive root of unity.

Chapter 3

Construction Schemes for Approximately Tight Wavelet Frames

3.1 Introduction

In signal processing, uncertainty principles emanate from the need to determine certain features of signals, e.g., position or frequency, with reasonable accuracy. With regard to position and frequency, the corresponding uncertainty principle states that a signal can not be well concentrated in time- and frequency domain simultaneously. Not only position and frequency are desirable function properties, but also scale or orientation. By minimizing each of the resulting uncertainty principles, explicit functions can be derived which are supposed to be optimally localized in corresponding transform domains (Dahlke and Maaß, 1995; Dahlke et al., 2008; Sagiv et al., 2006). Only recently, however, it has been shown that some of the considered uncertainty principles might not attain a lower bound (Maaß et al., 2010, 2014), misleading the concept of uncertainty minimizers. Therefore, Levie et al. (2014) introduced a more general uncertainty principle remedying these flaws. Their results have lead to a new uncertainty principle for the wavelet transform with a strict lower bound. This concept is concisely summarized as follows.

A commonly used measure of spread or concentration is the variance (Ricaud and Torrésani, 2014). With expectation value $e_f(\mathcal{T}) = \langle f, \mathcal{T} f \rangle$ for a self-adjoint operator $\mathcal{T} \in \mathcal{H}$ acting on some state $f \in \mathcal{H}$, the variance $v_f(\mathcal{T})$ is defined by $v_f(\mathcal{T}) = e_f\left(\mathcal{T}^2\right) - e_f(\mathcal{T})^2$. Known from quantum mechanics, the uncertainty principle for two self-adjoint operators $\mathcal{T}_1$ and $\mathcal{T}_2$ is

then defined by the Robertson-Schrödinger inequality (Folland and Sitaram, 1997)

$$v_f(\mathcal{T}_1)v_f(\mathcal{T}_2) \geq \frac{1}{4}\left|\langle f, [\mathcal{T}_1, \mathcal{T}_2]f\rangle\right|^2, \tag{3.1.1}$$

for all suitably chosen f. The commutator $[\mathcal{T}_1, \mathcal{T}_2]$ is given by $[\mathcal{T}_1, \mathcal{T}_2] = \mathcal{T}_1\mathcal{T}_2 - \mathcal{T}_2\mathcal{T}_1$. For more information on the theory of commutators and involved operators, e.g., domain, range, boundedness, it is referred to Putnam (1967).

Grossmann et al. (1985) associated the wavelet transform with the one-dimensional affine group. Generators of this group are given by the infinitesimal operators of translation and dilation $\mathcal{T}_a = i\frac{d}{dx}$ and $\mathcal{T}_d = i\left(\frac{1}{2} + x\frac{d}{dx}\right)$ and lead to a minimizer for the uncertainty principle in (3.1.1) derived by Dahlke and Maaß (1995). Levie et al. (2014) argued, however, that such an approach is ambiguous since the generators would not measure position and scale. The approach proposed instead, is based on so called adjoint translations (Lantzberg et al., 2012; Levie and Sochen, 2017a,b).

Adjoint translation operators $\breve{\mathcal{T}}$ of some operator $\mathcal{T}$ satisfy the canonical commutation relation

$$\left[\mathcal{T}, \breve{\mathcal{T}}\right] = i1, \tag{3.1.2}$$

where 1 denotes the identity element. According to Levie et al. (2014, Prop. 11), a solution to (3.1.2) is guaranteed, albeit it might not necessarily be unique: the sum of $\breve{T}$ and any self-adjoint operator which commutes with T also satisfies (3.1.2). The adjoint translation operators for translation and dilation $\breve{\mathcal{T}}_a$ and $\breve{\mathcal{T}}_d$ are easily computed by $\breve{\mathcal{T}}_a f(x) = xf(x)$ and $\mathcal{F}\breve{\mathcal{T}}_d\mathcal{F}^* \hat{f}(\nu) = -\ln\nu\,\hat{f}(\nu)$, with ln denoting the natural logarithm. For a more rigorous discussion about corresponding domains, existence and uniqueness of such adjoint translation operators in general it is referred to (Levie et al., 2014).

In order to summarize the main result from Levie et al. (2014), let the space $H^2_+(\mathbb{R})$ be defined by $H^2_+(\mathbb{R}) = \left\{f \in L^2(\mathbb{R}) \;\middle|\; \operatorname{supp}\left(\hat{f}(\nu)\right) \subseteq \mathbb{R}^+\right\}$. Then, Proposition 22 by Levie et al. (2014) states that the uncertainty principle for the wavelet transform reads

$$v_{\hat{f}}\left(-i\frac{d}{d\nu}\right) v_{\hat{f}}(\ln\nu) \geq \frac{1}{4}\left|C_f\right|^2, \tag{3.1.3}$$

for some signal $f \in H^2_+(\mathbb{R})$ and admissibility constant C_f. It has been further shown that equality of (3.1.3) can not be achieved. The resulting waveform which would satisfy (3.1.3) is

explicitly given by

$$\hat{f}(\nu) = Ce^{\nu((i\alpha+\mu(\beta+1-\ln\nu))}, \tag{3.1.4}$$

for some $C, \alpha, \beta \in \mathbb{R}$ and $\mu > 0$. This function, however, only satisfies the admissibility condition for wavelets asymptotically as $\mu \to \infty$, (Levie et al., 2014, Cor. 24). Despite this shortcoming, with sufficiently large μ (e.g. $\mu = 25$) this equalizing function can be used in a finite dimensional setting for wavelet analysis as $\hat{f}(0)$ can be made sufficiently small. In the following definition of the Equalizer as well as all subsequent evaluations, the parameters α and β are set to 0.

Definition 3.1.1 (Wavelet Equalizer)**.** For some $\mu > 0$ the uncertainty Equalizer for the wavelet transform is defined in the frequency domain by

$$\Psi_{\mathrm{Eq}}(\nu) = Ce^{\mu\nu(1-\ln\nu)}, \tag{3.1.5}$$

for all positive frequencies $\nu \in \mathbb{R}^+$ and normalization constant $C \in \mathbb{R}$. Possible choices for C may be such that $\left\|\Psi_{\mathrm{Eq}}\right\|_2 = 1$ or $C = e^{-\mu}$ leading to a maximum value $\Psi_{\mathrm{Eq}}(1) = 1$.

With this definition the question now arises if it is possible to construct a discretized version of the continuous wavelet transform based on the Equalizer. Additionally, perfect reconstruction is desirable. Construction schemes for wavelet frames as well as dual frames for perfect reconstruction are presented by Bayram and Selesnick (2009) and Balazs et al. (2011) and are implemented in the LTFAT toolbox (Průša et al., 2014). However, both approaches are based on band-limited wavelets. Unfortunately, the Equalizer is neither band- nor time limited. The existence of wavelet frames with non-compactly supported functions is shown by Christensen (2016, Prop. 15.2.6). In a more general setting a similar result is proven by Dörfler and Matusiak (2014) for so called nonstationary Gabor frames, a generalization of Gabor and wavelet frames to irregular, and hence almost arbitrary, sampling schemes.

Although the existence of such frames has been investigated in the literature quite recently, the analysis of dual frames for which perfect reconstruction is possible is rarely addressed. Feichtinger et al. (2014) introduced approximate dual frames based on compactly supported wavelets in order to reduce the complexity of computing the canonical dual frame elements. A detailed analysis of dual wavelet frame pairs has been introduced by Chui and Shi (2000, Thm. 1), where necessary conditions are only satisfied for band-limited wavelets (Christensen, 2016, p. 403). A more general result has been introduced by Liu and Sun (2009) and Li and Sun (2012). They have proven that in the continuous case the wavelet frame operator

with two arbitrary wavelets converges to the identity, whenever dilation parameters tend to 1 and simultaneously shift parameters to 0, provided both wavelets satisfy proper decay and smoothness conditions. In practical applications, however, such theoretical results are not feasible. Dörfler and Matusiak (2015) have introduced approximate dual windows for nonstationary Gabor frames which are based on non-compactly supported window functions. With the results presented therein, it is possible to construct corresponding dual frames, but only in an approximate manner such that perfect reconstruction is numerically not achievable.

In the following chapter, the diagonality of frame operator is characterized first with respect to equidistantly spaced time- and frequency sampling points. The result is then generalized with respect to a non-equidistantly spaced time- or frequency sampling. In contrast to Dörfler and Matusiak's (2015) results, these frame constructions and proposed dual frames lead to a frame operator which is, at least within numerical precision, the identity operator in a finite dimensional setting. Additionally, tight frames are characterized and fast algorithms for synthesis and analysis are proposed. Finally, the results are applied to construct and evaluate wavelet frames with non-compactly supported functions.

3.2 Gabor Frames

3.2.1 Preliminaries

Discretization of the short-time Fourier transform in (2.0.14) leads to the so called Gabor transform. Instead of using all possible translations and modulations, it is sufficient to use only a countable subset of time-frequency shifts. Thus, the Gabor expansion of a signal f can be written as

$$f = \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} \langle f, T_{ak} M_{bl} g \rangle \, T_{ak} M_{bl} \gamma, \tag{3.2.1}$$

for suitable windows $g, \gamma \in L^2(\mathbb{R})$ and time and frequency shift parameters $a, b > 0$. Fundamental results and definitions concerning such expansions are summarized in the following from Feichtinger and Strohmer (1998, 2003) and Gröchenig (2001), starting with the definition of a Gabor frame itself.

Definition 3.2.1 (Gabor Frame)**.** Let $g \in L^2(\mathbb{R})$ be a window function. For some real numbers $a, b > 0$ the collection

$$\mathcal{G}(g, a, b) = \{T_{ak} M_{bl} g : k, l \in \mathbb{Z}\} \tag{3.2.2}$$

with time and frequency shift parameters a and b is called a Gabor system. Whenever this system satisfies the frame condition in (2.0.9), it is referred to as a Gabor frame for $L^2(\mathbb{R})$.

The frame operator associated with such Gabor frames is defined as follows:

Definition 3.2.2 (Gabor Frame Operator)**.** For any $f \in L^2(\mathbb{R})$ and $g, \gamma \in L^2(\mathbb{R})$, the Gabor frame operator $S : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, associated to Gabor frames $\mathcal{G}(g, a, b)$ and $\mathcal{G}(\gamma, a, b)$, has the form

$$S_{g,\gamma} f = \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} \langle f, T_{ak} M_{bl} g \rangle \, T_{ak} M_{bl} \gamma. \tag{3.2.3}$$

For the special case when $\gamma = g$, the Gabor frame operator $S_{g,g}$ is simply denoted by S.

The Gabor frame operator commutes with time-frequency shifts, implying that $S T_{ak} M_{bl} = T_{ak} M_{bl} S$ as shown by Christensen (2016, Lem. 12.3.1). Further, the Gabor frame operator S is bounded (cf. Gröchenig (2001, Cor. 6.2.3)) whenever the window function g is in the Wiener space $W(\mathbb{R})$ defined as follows.

Definition 3.2.3 (Wiener Space)**.** A function $g \in L^\infty(\mathbb{R})$ satisfying

$$\|g\|_W = \sum_{k\in\mathbb{Z}} \operatorname*{ess\,sup}_{x\in[0,1]} |g(x+k)| < \infty, \tag{3.2.4}$$

belongs to the Wiener space $W(\mathbb{R})$.

Walnut (1992, Prop. 2.4.) showed that the Gabor frame operator maps a function onto a sum of weighted and translated copies of itself. It states that for $g, \gamma \in W(\mathbb{R})$ and $a, b > 0$, the Gabor frame operator $S_{g,\gamma}$ in (3.2.3) can be written as

$$S_{g,\gamma} f = \frac{1}{b} \sum_{l\in\mathbb{Z}} G_l^{g,\gamma} T_{\frac{l}{b}} f, \tag{3.2.5}$$

where the correlation $G_l^{g,\gamma}$ is defined as

$$G_l^{g,\gamma}(x) = \sum_{k\in\mathbb{Z}} \overline{g}\left(x - \frac{l}{b} - ak\right) \gamma(x - ak), \quad l \in \mathbb{Z}. \tag{3.2.6}$$

A formal proof of Walnut's representation theorem is found in (Gröchenig, 2001, Thm. 6.3.2). For $l = 0$ the term $G_l^{g,\gamma}$ represents the diagonal part of the Gabor frame operator and $G_l^{g,\gamma}$ with $l \neq 0$ the side diagonal part.

The Gabor frame operator is diagonal whenever the side diagonals vanish, i.e., $G_l^{g,\gamma} = 0$ for all $l \neq 0$. Daubechies et al. (1986) introduced the following theorem, which characterizes the diagonality of the Gabor frame operator for compactly supported functions g.

Theorem 3.2.4 (Painless Non-Orthogonal Expansions (Gröchenig, 2001, Thm. 6.4.1))**.** *Let $g \in W(\mathbb{R})$ be compactly supported on the interval $[0, N]$ and choose time and frequency shift steps a and b such that $a \leq N$ and $b \leq \frac{1}{N}$. The frame operator $S_{g,g}$ turns into a (pointwise) multiplication operator*

$$S_{g,g} f(x) = \frac{1}{b} G_0^{g,g}(x) f(x) = \frac{1}{b} \sum_{k \in \mathbb{Z}} |g(x - ak)|^2 \ f(x). \tag{3.2.7}$$

Proof. The proof analogously follows Gröchenig's (2001, Thm. 6.4.1) argumentation. Consider the corresponding correlation function from (3.2.6)

$$G_l^{g,g}(x) = \sum_{k \in \mathbb{Z}} \overline{g}\left(x - \frac{l}{b} - ak\right) g(x - ak), \quad l \in \mathbb{Z}. \tag{3.2.8}$$

Assume $l \neq 0$. The intersection of the supports of $\overline{g}(x - b^{-1}l - ak)$ and $g(x - ak)$ is either empty or a set of measure zero iff $b = N^{-1}$ for all $k \in \mathbb{Z}$. This implies that $G_l^{g,g} = 0$ whenever $l \neq 0$. For $l = 0$, $G_0^{g,g}(x) = \sum_{k \in \mathbb{Z}} \overline{g}(x - ak) g(x - ak)$ and (3.2.7) follows from Walnuts representation of the frame operator in (3.2.5). □

Hence, whenever a window g has compact support and the frequency shift parameter b is chosen appropriately, the resulting Gabor frame operator is diagonal. If the time shift parameter a is chosen such that $G_0^{g,g}$ is bounded from below and from above, the resulting collection of time- and frequency shifted functions constitutes a Gabor frame. The canonical dual window γ° is then easily computed by inverting the diagonal frame operator

$$\gamma^\circ = S^{-1} g = b \left(G_0^{g,g}\right)^{-1} g. \tag{3.2.9}$$

Further, if $G_0^{g,g}(x) \propto 1$ for all $x \in \mathbb{R}$ the corresponding frame is tight and a Parseval frame whenever $G_0^{g,g} = 1$.

3.2.2 Diagonality of the Gabor Frame Operator

Unfortunately, Daubechies non-orthogonal painless expansions only work for compactly supported functions. In general the frame operator is not diagonal and the estimation of canonical

dual windows depends on the frame operator and its inverse. Computational aspects of estimating canonical duals have been treated by Li (1995) and Janssen and Søndergaard (2007) and non-canonical dual windows by Werther et al. (2005). In all those cases, a fixed Gabor frame $\mathcal{G}(g, a, b)$ leads to a dual window γ such that the resulting frame operator $S_{g,\gamma}$ is the identity.

A different approach would be to fix a window g and an appropriate choice of γ and evaluate for which parameters a and b the frame operator is diagonal. Obviously, the existence of these parameters depends on the chosen γ. Sun (2010) proved, that for $\gamma = g$ the frame operator $S_{g,\gamma}$ converges to the identity, whenever the sampling parameters a and b tend to zero. With a more suitable choice for γ, Sun's result can be relaxed such that $S_{g,\gamma}$ converges to the identity for all suitable $a > 0$ if b tends to zero. The following theorem shows an upper bound for the diagonality of the Gabor frame operator with respect to sampling parameters a and b.

Theorem 3.2.5 (Diagonality of the Gabor Frame Operator)**.** *Let $g \in W(\mathbb{R})$. Let the time sampling parameter $a > 0$ such that there exists positive constants A, B satisfying*

$$0 < A \leq G_0^{g,g} \leq B < \infty \quad (a.e.). \tag{3.2.10}$$

Define the dual window by

$$\gamma = \gamma^\circ = b \left(G_0^{g,g}\right)^{-1} g, \tag{3.2.11}$$

where $b > 0$ is the frequency shift parameter. With I being the identity operator, the reconstruction error can now be bounded by

$$\left\| I - S_{g,\gamma} \right\|_2 \leq \frac{\sum\limits_{l \in \mathbb{Z}\setminus\{0\}} \left\| G_l^{g,g} \right\|_\infty}{\operatorname{ess\,inf} G_0^{g,g}}. \tag{3.2.12}$$

Furthermore,

$$\lim_{b \to 0} \left\| I - S_{g,\gamma}^{(a,b)} \right\|_2 = 0 \tag{3.2.13}$$

where the notation of the frame operator is slightly changed to indicate its dependence on the parameters a and b.

Proof. Due to (3.2.10) the dual window γ is well-defined. To show that (3.2.12) holds, Walnuts

representation of the Gabor frame operator gives

$$\left\| \left(I - S_{g,\gamma}^{(a,b)} \right) f \right\|_2 = \left\| f - \frac{1}{b} \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \overline{g} \left(\cdot - \frac{l}{b} - ak \right) \frac{b}{G_0^{g,g}} g(\cdot - ak) f \left(\cdot - \frac{l}{b} \right) \right\|_2 \tag{3.2.14}$$

$$= \left\| f - \frac{1}{G_0^{g,g}} \left[\sum_{k \in \mathbb{Z}} |g(\cdot - ak)|^2 f + \sum_{l \in \mathbb{Z} \setminus \{0\}} \sum_{k \in \mathbb{Z}} \overline{g} \left(\cdot - \frac{l}{b} - ak \right) g(\cdot - ak) f \left(\cdot - \frac{l}{b} \right) \right] \right\|_2 \tag{3.2.15}$$

$$= \left\| \frac{1}{G_0^{g,g}} \sum_{l \in \mathbb{Z} \setminus \{0\}} \sum_{k \in \mathbb{Z}} \overline{g} \left(\cdot - \frac{l}{b} - ak \right) g(\cdot - ak) f \left(\cdot - \frac{l}{b} \right) \right\|_2 . \tag{3.2.16}$$

This expression can be bounded by

$$\left\| \left(I - S_{g,\gamma}^{(a,b)} \right) f \right\|_2 \leq (\operatorname{ess\,inf} G_0)^{-1} \left\| \sum_{l \in \mathbb{Z} \setminus \{0\}} G_l^{g,g} f \left(\cdot - \frac{l}{b} \right) \right\|_2 \tag{3.2.17}$$

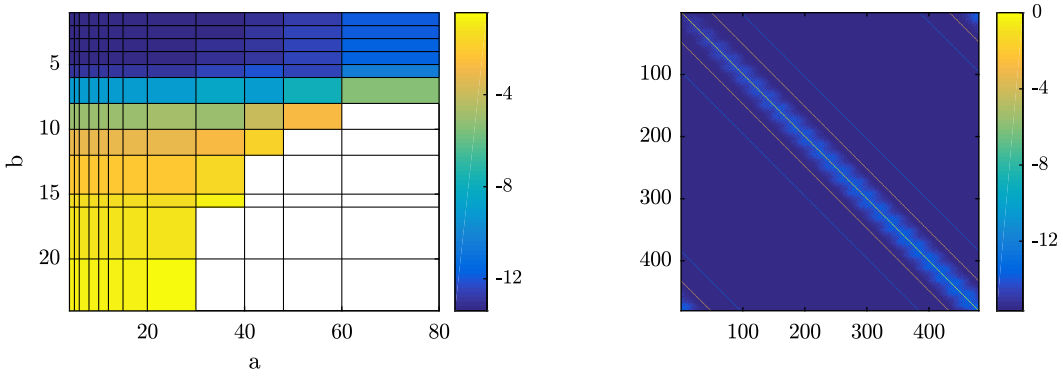
$$\leq (\operatorname{ess\,inf} G_0)^{-1} \sum_{l \in \mathbb{Z} \setminus \{0\}} \left\| G_l^{g,g} \right\|_\infty \| f \|_2 . \tag{3.2.18}$$

Equation (3.2.13) follows immediately from (3.2.12) and (Gröchenig, 2001, Lemma 6.5.2), which states that

$$\lim_{b \to 0} \sum_{l \in \mathbb{Z} \setminus \{0\}} \left\| G_l^{g,g} \right\|_\infty = 0, \tag{3.2.19}$$

for any $a > 0$. □

Theorem 3.2.5 states that the frame operator with g and dual window γ according to (3.2.11) is diagonal whenever the following two conditions hold: First, the time sampling parameter a is chosen such that (3.2.10) holds, meaning sufficient overlap of neighboring copies of g is required. Secondly, the frequency sampling parameter b has to be chosen such that the side diagonals of the Gabor frame operator vanish. For compactly supported functions g the corresponding parameter choice is explicitly given by Daubechies painless non-orthogonal expansions. For more general windows this is only true whenever b tends to 0.



(a) Reconstruction error $\left\| I - S_{g,\gamma}^{(a,b)} \right\|_2$

(b) Gabor frame operator $S^{(a,b)}$ for $a = 20$ and $b = 10$

Figure 3.1: Gabor frame operator for a Gaussian window in dependence of time and frequency sampling parameter a and b. Both color bars are logarithmic.

In a finite dimensional setting, however, it is sufficient that the diagonality of the Gabor frame operator is within machine precision. This mainly depends on the frequency sampling parameter b and the decay behavior of windows g. Hence, the limiting process of (3.2.19) can be stopped for b-values where the reconstruction error is within machine precision. In summary, the Gabor frame operator is diagonal in a finite dimensional setting if g is decaying fast enough and b is sufficiently small. This is illustrated in the following discrete example.

Example. Let g be the canonical Gaussian, which is invariant under Fourier transform, with length $L = 480$. With this choice only integer divisors of L can be used for time and frequency shift parameters a and b. The first 19 of these divisors are 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 20, 24, 30, 32, 40, 48, 60 and 80 which can be easily verified. The Gabor frame operator $S_{g,\gamma}^{(a,b)}$ based on windows g and γ as defined in (3.2.11) can then be explicitly computed, e.g., using (Qiu and Feichtinger, 1995, Eq. (6)). According to the Balian-Low theorem, $\mathcal{G}(g, a, b)$ is a frame only if $ab < L$ (Gröchenig, 2001, Thm. 7.5.3 and Cor. 8.4.4).

Figure 3.1a shows the reconstruction error in logarithmic representation of the Gabor frame operator in dependence of parameters a and b. Parameter combinations which do not satisfy the Balian-Low Theorem are excluded and corresponding rectangles remain white. According to the figure, the reconstruction error is within machine precision for small values of b. The time sampling parameter a does not directly influence the error. Large values, however, cause insufficient overlap of Gaussian windows, resulting in large values of dual windows, since the

lower bound A in (3.2.10) is close to zero. The Gabor frame with $a = 30$ and $b = 4$ results in a frame of redundancy 4 and a reconstruction error of $\left\| I - S^{30,4} \right\|_2 = 7.3\text{e}{-14}$.

Figure 3.1b illustrates the Gabor frame operator exemplarily for $a = 20$ and $b = 10$. Clearly visible are the side diagonals responsible for a reconstruction error of 0.0008. Further, numerical noise can be observed around the main diagonal, which arises from computing the frame operator explicitly by $\Phi^*\Phi$.

The aforementioned example shows, that it is possible to construct frames with predefined dual windows, such that the approximation error $\left\| I - S_{g,\gamma} \right\|_2 < \varepsilon$ of the Gabor frame operator is within machine precision by choosing a suitably small frequency sampling parameter b. Since the approximation depends on the decay properties of g, and hence convergence is nonuniform, it is difficult to derive a general rule which assigns for each ε a corresponding b.

In the following section this extension of Daubechies non-orthogonal expansions is applied to a generalization of Gabor frames.

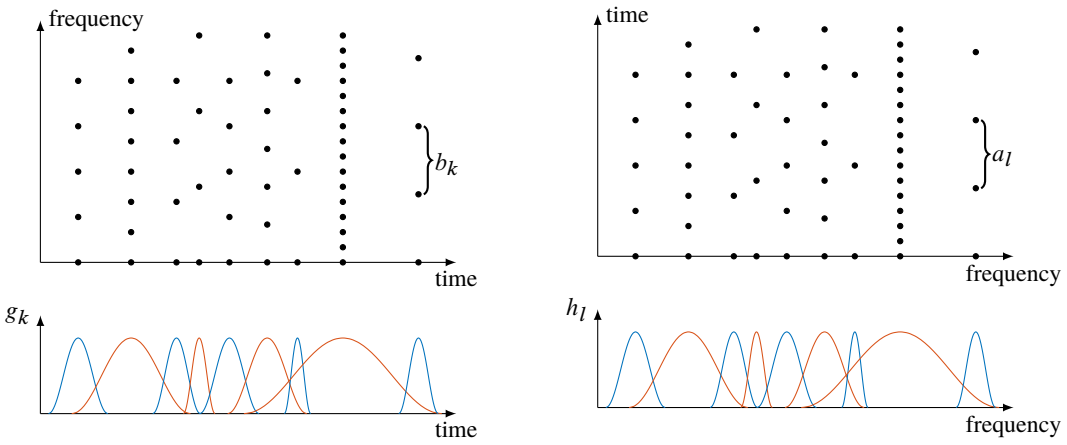
3.3 Nonstationary Gabor Frames

3.3.1 Preliminaries

One of the main disadvantages of the Gabor transform is its fixed time-frequency resolution resulting from inflexible time and frequency shift parameters a and b. Balazs et al. (2011) proposed the concept of nonstationary Gabor frames, a generalization of Gabor frames and non-orthogonal painless expansions, overcoming such rigid time-frequency discretizations. It has been widely used for adaptive signal analysis in acoustics (Liuni et al., 2013; Ricaud et al., 2014; Schörkhuber et al., 2013), in nonlinear approximation theory (Ottosen and Nielsen, 2017a,b) and even data mining (Sattar et al., 2016). Formally, nonstationary Gabor frames are defined as follows.

Definition 3.3.1 (Nonstationary Gabor Frames)**.** Let the set $\mathcal{G}(g_k, b_k) = \{M_{b_k l} g_k : k, l \in \mathbb{Z}\}$ of window functions $g_k \in W(\mathbb{R})$ with corresponding frequency shift parameter b_k be denoted a nonstationary Gabor system. Whenever this set satisfies the frame condition in (2.0.9) it is called a nonstationary Gabor frame.

Instead of fixed uniform translations a, windows g_k can be non-uniformly distributed on the time-axis. These windows can vary in shape or width, allowing adaptive and irregular sampling schemes. Figure 3.2a schematically shows an example of such an irregular sampling of the



(a) Resolution changing over time

(b) Resolution changing over frequency

Figure 3.2: Irregular sampling scheme resulting from nonstationary Gabor frames.

time-frequency plane with windows g_k of varying width. Clearly, whenever $g_k = g(\cdot - ak)$ for some time shift parameter $a > 0$ and $b_k = b$ for all $k \in \mathbb{Z}$ the resulting frame is a regular Gabor frame.

Any function $f \in L^2(\mathbb{R})$ can be decomposed into analysis coefficients by taking inner products of f with nonstationary Gabor frame elements, i.e.,

$$c_{k,l} = \left\langle f, M_{b_k l} g_k \right\rangle \quad k, l \in \mathbb{Z}. \tag{3.3.1}$$

The corresponding nonstationary Gabor frame operator for frames $\mathcal{G}(g_k, b_k)$ and $\mathcal{G}(\gamma_k, b_k)$ is given by

$$S_{g,\gamma} f = \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \left\langle f, M_{b_k l} g_k \right\rangle M_{b_k l} \gamma_k. \tag{3.3.2}$$

Its Walnut representation is due to Dörfler and Matusiak (2014, Prop. 3.3) and reads

$$S_{g,\gamma} f = \sum_{l \in \mathbb{Z}} G_l^{g,\gamma} T_{\frac{l}{b_k}} f, \tag{3.3.3}$$

with

$$G_l^{g,\gamma} = \sum_{k \in \mathbb{Z}} \frac{1}{b_k} \overline{g_k} \left(\cdot - \frac{l}{b_k} \right) \gamma_k(\cdot). \tag{3.3.4}$$

for $g_k, \gamma_k \in W(\mathbb{R})$.

Balazs et al. (2011) introduced a generalization of Daubechies painless non-orthogonal expansions. The following theorem is one of their main results, albeit with an alternative version of the proof:

Theorem 3.3.2 (Painless Nonstationary Expansions (Balazs et al., 2011, Thm. 1))**.** *Let $g_k \in W(\mathbb{R})$ be compactly supported on $[p_k, q_k]$ for all $k \in \mathbb{Z}$. Let the frequency shift parameter $b_k > 0$ for each corresponding g_k be such that $b_k \leq \frac{1}{q_k - p_k}$. If there exists positive constants A, B such that*

$$0 < A \leq \sum_{k \in \mathbb{Z}} \frac{1}{b_k} |g_k|^2 \leq B < \infty \quad (a.e.), \tag{3.3.5}$$

the collection $\mathcal{G}(g_k, b_k)$ is a frame for $L^2(\mathbb{R})$. Further, the frame operator $S_{g,g}$ in Equation (3.3.2) is the multiplication operator

$$S_{g,g} f = \left(\sum_{k \in \mathbb{Z}} \frac{1}{b_k} |g_k|^2 \right) f. \tag{3.3.6}$$

Proof. With Walnuts representation of the nonstationary Gabor frame operator, the proof follows Daubechies painless non-orthogonal expansions analogously. First, assume that $l \neq 0$ and consider the correlation $G_l^{g,g}$ in (3.3.4). For each k, intersecting the supports of $\overline{g_k}\left(\cdot - b_k^{-1} l\right)$ and $g_k(\cdot)$ results in either the empty set or a set of measure zero and hence $G_{l \neq 0}^{g,g} = 0$. Equation (3.3.6) then follows immediately for $l = 0$ and Walnuts representation of the nonstationary Gabor frame operator. □

Corollary 3.3.3. *Canonical dual windows γ_k° are given by*

$$\gamma_k^\circ = \frac{1}{\sum_{k \in \mathbb{Z}} \frac{1}{b_k} |g_k|^2} g_k, \tag{3.3.7}$$

for every $k \in \mathbb{Z}$.

An equivalent result to Theorem 3.3.2 can be formulated for compactly supported functions in the Fourier domain, allowing adaptivity in the frequency domain. Figure 3.2b exemplarily illustrates such an irregular sampling scheme where the resolution changes over frequency, as windows can be placed at arbitrary locations on the frequency axis. The following Corollary is an immediate result from Theorem 3.3.2.

Corollary 3.3.4 (Painless Nonstationary Expansions in Frequency Domain (Balazs et al., 2011, Cor. 2))**.** *Let $h_l \in W(\mathbb{R})$ be such that $\hat{h}_l$ is bandlimited on the interval $[p_l, q_l]$ for all $l \in \mathbb{Z}$. The corresponding nonstationary Gabor system is the set $\mathcal{G}(h_l, a_l) = \{T_{a_l k} h_l : k, l \in \mathbb{Z}\}$. If time shift parameters a_l are chosen such that $a_l \leq \frac{1}{q_l - p_l}$, the corresponding nonstationary Gabor frame operator $S_{h,h}$ is the convolution operator*

$$S_{h,h} f = \mathcal{F}^{-1}\left(\sum_{l \in \mathbb{Z}} \frac{1}{a_l} \left|\hat{h}_l\right|^2 \cdot \hat{f} \right), \quad f \in L^2(\mathbb{R}). \tag{3.3.8}$$

Thus, $\mathcal{G}(h_l, a_l)$ is frame for $L^2(\mathbb{R})$ if and only if there exist constants $A, B > 0$ satisfying the inequality

$$0 < A \leq \sum_{l \in \mathbb{Z}} \frac{1}{a_l} \left|\hat{h}_l\right|^2 \leq B < \infty, \tag{3.3.9}$$

almost everywhere.

Similar to Corollary 3.3.3, inverting the frame operator gives canonical dual windows $\hat{\gamma}_l^\circ$ in frequency domain

$$\hat{\gamma}_l^\circ = \frac{\hat{h}_l}{\sum_{l \in \mathbb{Z}} \frac{1}{a_l} \left|\hat{h}_l\right|^2}, \quad \forall l \in \mathbb{Z}. \tag{3.3.10}$$

Corresponding analysis coefficients are then defined by

$$c_{k,l} = \left\langle f, T_{a_l k} h_l \right\rangle = \left\langle \hat{f}, M_{-a_l k} \hat{h}_l \right\rangle \quad k, l \in \mathbb{Z}, \tag{3.3.11}$$

showing that the results from Theorem 3.3.2 and above Corollary 3.3.4 are essentially the same up to a Fourier transform. Hence, $S_{h,h}$ also admits a Walnut representation.

Obviously, such painless constructions only work for windows with compact support, either time- or bandlimited. More general windows which are neither time- nor bandlimited are considered in the following.

3.3.2 Diagonality of the Nonstationary Gabor Frame Operator with Non-Compactly Supported Functions

Dörfler and Matusiak (2014, Thm. 3.4 and Cor. 3.5) have proven the existence of nonstationary Gabor frames with windows which are neither compactly supported nor bandlimited. In order to summarize their results, it is necessary to assume that windows g_k are localized around

time-sampling points a_k.

Definition 3.3.5 (δ-separated Set)**.** A set of sampling points $\{a_k \in \mathbb{R} : k \in \mathbb{Z}\}$ is δ-separated if there exists $\delta > 0$, such that

$$|a_k - a_m| > \delta, \quad \forall k \neq m. \tag{3.3.12}$$

Dörfler and Matusiak then show that the nonstationary Gabor frame operator $S_{g,g}$ is bounded from above and below if for every k, $g_k \in W(\mathbb{R})$ and the set of windows $\{g_k\}_{k\in\mathbb{Z}}$ satisfies the following two conditions:

- for two constants $A, B > 0$

$$0 < A \leq \sum_{k\in\mathbb{Z}} |g_k(x)|^2 \leq B < \infty, \tag{3.3.13}$$

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

- for constants $C_k > 0$, windows g_k have polynomial decay around a δ-separated set $\{a_k \in \mathbb{R} : k \in \mathbb{Z}\}$ such that

$$|g_k(x)| \leq C_k \left(1 + |x - a_k|\right)^{-\rho_k}, \quad \forall x \in \mathbb{R}, \tag{3.3.14}$$

  where $\rho_k > 2$ for all $k \in \mathbb{Z}$.

With these two conditions, Theorem 3.4 by Dörfler and Matusiak (2014) can be concisely summarized as follows. If windows g_k have polynomial decay and sufficient overlap, a sequence $\{b_k^0\}_{k\in\mathbb{Z}}$ exists such that for $b_k \leq b_k^0$ for all $k \in \mathbb{Z}$, the system $\mathcal{G}(g_k, b_k)$ constitutes a frame for $L^2(\mathbb{R})$. Thus, nonstationary Gabor frames can always be constructed for functions with sufficient decay properties by choosing sufficiently dense frequency sampling parameters b_k. For the sake of completeness, an equivalent result holds for windows h_l which decay polynomially in frequency domain: with sufficient overlap of windows $\hat{h}_l$, there exists a sequence $\{a_l^0\}_{l\in\mathbb{Z}}$ such that for all $a_l \leq a_l^0$ the system $\mathcal{G}(h_l, a_l)$ forms a frame for $L^2(\mathbb{R})$.

In the regular Gabor case the frame operator commutes with time-frequency shifts. For nonstationary Gabor frames the frame operator might not commute with modulations, i.e., $S^{-1}(M_{b_k l} g_k) \neq M_{b_k l} S^{-1} g_k$. Holighaus (2014, Thm. 3) proved, that under certain conditions the inverse frame operator of nonstationary Gabor frames possesses a similar structure as the original frame operator. Further, these conditions also guarantee that the canonical dual frame of $\mathcal{G}(g_k, b_k)$ is again a nonstationary Gabor frame with the same modulation parameters b_k

(Holighaus, 2014, Cor. 5). Unfortunately, this only holds for compactly supported windows. For functions which are neither time- nor bandlimited Dörfler and Matusiak (2015) proposed construction schemes for approximately dual frames, where approximate dual frames are characterized by the following definition adapted from Christensen and Laugesen (2010).

Definition 3.3.6 (Approximate Dual Frame)**.** Two frames $\mathcal{G}(g_k, b_k)$ and $\mathcal{G}(\gamma_k, b_k)$ are said to be approximate dual frames, whenever

$$\left\| I - S_{g,\gamma} \right\|_2 < 1, \tag{3.3.15}$$

for any frame operator S with windows g and γ.

As stated by Christensen (2016, Ch. 6.5), an upper bound of 1 might not be sufficient to guarantee that a function f is approximately close to $S_{g,\gamma} f$. In a finite dimensional setting this would contradict the perfect reconstruction constraint. A more reasonable approach is given by

$$\left\| I - S_{g,\gamma} \right\|_2 \leq \varepsilon_r \tag{3.3.16}$$

for some positive $\varepsilon_r \ll 1$. With the following lemmata it can be shown, that under certain assumptions, this approximation can be made arbitrarily small, similar to Theorem 3.2.5 for the regular Gabor case. This extends the results by Dörfler and Matusiak (2015) as they only considered approximate dual frames as defined in (3.3.15). Starting point is a brief lemma about an upper bound involving δ-separated sets, taken from Dörfler and Matusiak (2014).

Lemma 3.3.7 (Dörfler and Matusiak (2014, Lem. 2.2b))**.** *Let $\{a_k \in \mathbb{R} : k \in \mathbb{Z}\}$ be a δ-separated set. For $\rho > 1$ the following inequality holds*

$$\operatorname*{ess\,sup}_{x \in \mathbb{R}} \sum_{k \in \mathbb{Z}} (1 + |x - a_k|)^{-\rho} \leq 2 \left(1 + (1+\delta)^{-\rho} (\delta^{-1} + \rho)(\rho - 1)^{-1}\right). \tag{3.3.17}$$

The corresponding proof can be found in (Dörfler and Matusiak, 2014). The following lemma characterizes the side diagonals of the nonstationary Gabor frame operator with respect to frequency sampling parameters b_k.

Lemma 3.3.8. *Let g_k have polynomial decay around a δ-separated set $\{a_k \in \mathbb{R} : k \in \mathbb{Z}\}$ for all $k \in \mathbb{Z}$, i.e., there exist constants $C_k \in [C_L, C_U]$ and $\rho_k \in [\rho_L, \rho_U]$, where the sets $[C_L, C_U]$ and $[\rho_L, \rho_U]$ are positive and $\rho_L > 2$, such that*

$$|g_k(x)| \leq C_k \, (1 + |x - a_k|)^{-\rho_k}, \quad \forall x \in \mathbb{R}, \tag{3.3.18}$$

for all windows g_k.

Then

$$\lim_{b_k \to 0} \sum_{l \in \mathbb{Z} \setminus \{0\}} \left\| G_l^{g,g} \right\|_\infty = 0, \quad \forall k \in \mathbb{Z}, \tag{3.3.19}$$

where $G_l^{g,g}$ *is defined according to* (3.3.4).

Proof. The general idea of this proof is based on the proof of Theorem 3.4 by Dörfler and Matusiak (2014) with minor modifications. First, let $0 < \mu < \rho_L - 2$. Using the polynomial decay of g_k gives

$$\left| \frac{1}{b_k} \overline{g_k} \left(x - \frac{l}{b_k} \right) g_k(x) \right| = \frac{1}{b_k} \left| g_k \left(x - \frac{l}{b_k} \right) \right| |g_k(x)| \tag{3.3.20}$$

$$\leq \frac{C_k^2}{b_k} \left(1 + \left| x - a_k - \frac{l}{b_k} \right| \right)^{-\rho_k} (1 + |x - a_k|)^{-\rho_k} \tag{3.3.21}$$

$$\leq \frac{C_k^2}{b_k} \left(1 + \left| x - a_k - \frac{l}{b_k} \right| \right)^{-\rho_k + (1+\mu)} (1 + |x - a_k|)^{-\rho_k} \tag{3.3.22}$$

$$\leq \frac{C_k^2}{b_k} (1 + |x - a_k|)^{-(1+\mu)} \left(1 + \left| \frac{l}{b_k} \right| \right)^{-\rho_k + (1+\mu)} \tag{3.3.23}$$

$$\leq C_k^2 (1 + |x - a_k|)^{-(1+\mu)} |l|^{-\rho_k + (1+\mu)} b_k^{\rho_k - (2+\mu)}, \tag{3.3.24}$$

where the estimate from (3.3.22) to (3.3.23) results from the inequality

$$(1 + |x + y|)^{-\rho} \leq (1 + |x|)^{\rho} (1 + |y|)^{-\rho}, \tag{3.3.25}$$

for $x, y \in \mathbb{R}$ and $\rho \geq 0$ (Dörfler and Matusiak, 2014). Now, choose $\varepsilon < C_L$ and set $b_k = \left(\frac{\varepsilon}{C_k} \right)^{1/\rho_k}$. Thus,

$$\left\| G_l^{g,g} \right\|_\infty = \operatorname*{ess\,sup}_{x \in \mathbb{R}} \left| \sum_{k \in \mathbb{Z}} \frac{1}{b_k} \overline{g_k} \left(x - \frac{l}{b_k} \right) g_k(x) \right| \tag{3.3.26}$$

$$\leq \operatorname*{ess\,sup}_{x \in \mathbb{R}} \sum_{k \in \mathbb{Z}} C_k^{1 + (2+\mu)/\rho_k} |l|^{-\rho_k + 1 + \mu} \varepsilon^{1 - (2+\mu)/\rho_k} (1 + |x - a_k|)^{-(1+\mu)}$$

$$\leq \max_{k \in \mathbb{Z}} R_k |l|^{-\rho_L + 1 + \mu} \varepsilon^{1 - (2+\mu)/\rho_L} \operatorname*{ess\,sup}_{x \in \mathbb{R}} \sum_{k \in \mathbb{Z}} (1 + |x - a_k|)^{-(1+\mu)}, \tag{3.3.27}$$

where $R_k = C_k^{1+(2+\mu)/\rho_k}$. Using Lemma 3.3.7, the essential supremum on the right hand side of (3.3.27) can then be bounded by

$$\left\| G_l^{g,g} \right\|_\infty \leq \max_{k\in\mathbb{Z}} R_k \, |l|^{-(\rho_L-1-\mu)} \, \varepsilon^{1-(2+\mu)/\rho_L} 2\left(1 + (1+\delta)^{-(1+\mu)}(\delta^{-1}+1+\mu)\mu^{-1}\right). \tag{3.3.28}$$

Hence, with $\Lambda = 2\left(1 + (1+\delta)^{-(1+\mu)}(\delta^{-1}+1+\mu)\mu^{-1}\right)$

$$\sum_{l\in\mathbb{Z}\setminus\{0\}} \left\| G_l^{g,g} \right\|_\infty \leq \Lambda \max_{k\in\mathbb{Z}} R_k \varepsilon^{1-(2+\mu)/\rho_L} \sum_{l\in\mathbb{Z}\setminus\{0\}} |l|^{-(\rho_L-1-\mu)} \tag{3.3.29}$$

The summation of (3.3.29) over all $l \in \mathbb{Z}\setminus\{0\}$ reduces to the Riemann zeta function and is convergent as per definition $\rho_L - 1 - \mu > 1$. Further, Equation (3.3.29) tends to 0 whenever $\varepsilon \to 0$ since the exponent of ε satisfies $1 - (2+\mu)/\rho_L > 0$ and the claim in (3.3.19) follows readily. □

Now, analogously to Theorem 3.2.5 for Gabor frames, a similar result can be derived for nonstationary Gabor frames. It is an adaption of Proposition 4.1 by Dörfler and Matusiak (2015).

Theorem 3.3.9 (Diagonality of the Nonstationary Gabor Frame Operator)**.** *For every $k \in \mathbb{Z}$, let g_k have polynomial decay around a δ-separated set and let there exist positive constants A, B such that*

$$0 < A \leq G_0^{g,g} = \sum_{k\in\mathbb{Z}} \frac{1}{b_k} |g_k(x)|^2 \leq B < \infty, \quad \forall x \in \mathbb{R}. \tag{3.3.30}$$

Define dual windows by

$$\gamma_k = \left(G_0^{g,g}\right)^{-1} g_k, \tag{3.3.31}$$

for all $k \in \mathbb{Z}$. With identity operator I, the nonstationary Gabor frame operator satisfies

$$\left\| I - S_{g,\gamma} \right\|_2 \leq \frac{\sum_{l\in\mathbb{Z}\setminus\{0\}} \left\| G_l^{g,g} \right\|_\infty}{\operatorname{ess\,inf} G_0^{g,g}}. \tag{3.3.32}$$

Furthermore, if $b_k \to 0$ for every $k \in \mathbb{Z}$,

$$\lim_{b_k\to 0} \left\| I - S_{g,\gamma} \right\|_2 = 0, \tag{3.3.33}$$

where $S_{g,\gamma}$ depends on b_k, see (3.3.3)*.*

Proof. The proof for estimating the upper bound in (3.3.32) follows similar arguments as in Equations (3.2.14) to (3.2.18). Equation (3.3.33) follows then directly from Lemma 3.3.8 and the lower bound of $G_0^{g,g}$. □

This result shows that, quite similar to the Gabor case in a finite dimensional setting, the nonstationary Gabor frame operator can be made diagonal within machine precision for sufficiently small frequency shift parameters b_k. Similarly, by duality of (3.3.11) and (3.3.1), all of the above assumptions are also valid for functions $\hat{h}_l$ with sufficient decay properties in frequency domain. This can be summarized in the following Corollary, showing that the nonstationary Gabor frame operator tends to the identity, whenever time sampling parameters a_l tend pointwise to zero.

Corollary 3.3.10. *For every $l \in \mathbb{Z}$, let h_l be such that $\hat{h}_l$ has polynomial decay around a separated set and let there exist positive constants A, B such that*

$$0 < A \leq G_0^{h,h} = \sum_{l \in \mathbb{Z}} \frac{1}{a_l} \left| \hat{h}_l \right|^2 \leq B < \infty. \tag{3.3.34}$$

With identity operator I, the nonstationary Gabor frame operator satisfies

$$\left\| I - \mathcal{F}S_{h,\gamma} \right\|_2 \leq \frac{\sum\limits_{k \in \mathbb{Z} \setminus \{0\}} \left\| G_k^{h,h} \right\|_\infty}{\operatorname{ess\,inf} G_0^{h,h}}, \tag{3.3.35}$$

where dual windows are defined by

$$\hat{\gamma}_l = \left(G_0^{h,h} \right)^{-1} \hat{h}_l \tag{3.3.36}$$

for all $l \in \mathbb{Z}$. Furthermore, if $a_l \to 0$ for every $l \in \mathbb{Z}$,

$$\lim_{a_l \to 0} \left\| I - \mathcal{F}S_{h,\gamma} \right\|_2 = 0, \tag{3.3.37}$$

where $\mathcal{F}S_{h,\gamma}$ depends on a_l.

In order to put the results from Theorem 3.3.9 and Corollary 3.3.10 into more practical terms, consider the following definition, similar to the ε-concentration proposed by Donoho and Stark (1989):

Definition 3.3.11 (Essential Support)**.** Let $f \in L^2(\mathbb{R})$ be a function with $\|f\|_2 = 1$. For any

ε, with $0 < \varepsilon < 1$, there exists an interval $[N_1, N_2] \subset \mathbb{R}$ with $N_1 < N_2$, such that

$$\left\| f - f\chi_{[N_1,N_2]} \right\|_2^2 = \int_{\mathbb{R}} \left| f(x) - f(x)\chi_{[N_1,N_2]} \right|^2 dx < \varepsilon, \tag{3.3.38}$$

where χ is the characteristic function of the specified set. The essential support of f is then denoted by the interval $U^\varepsilon = [N_1, N_2]$.

With this definition, so called almost painless nonstationary Gabor frames can be constructed, similar to (Dörfler and Matusiak, 2015, Def. 4). For some ε, let

$$z_k = g_k \chi_{U_k^\varepsilon}, \tag{3.3.39}$$

be the painless approximation of windows g_k without compact support. Whenever frequency-shift parameters b_k are chosen such that $\mathcal{G}(z_k, b_k)$ satisfies Theorem 3.3.2, $\mathcal{G}(g_k, b_k)$ is called an almost painless nonstationary Gabor frame. Obviously, the smaller ε the smaller the difference between painless and almost painless nonstationary Gabor frame. This implies the following discrete construction scheme for nonstationary Gabor frames $\mathcal{G}(g_k, b_k)$ with windows g_k without compact support:

1. Choose arbitrary g_k with sufficient decay properties (e.g., Gaussians with different variances) localized at different time locations such that Equations (3.3.13) and (3.3.14) are satisfied.

2. Choose b_k by considering the corresponding painless approximation frame $z_k = g_k \chi_{U_k^\varepsilon}$ for a specific ε, such that $\mathcal{G}(z_k, b_k)$ is a frame according to Theorem 3.3.2.

The resulting nonstationary Gabor frame $\mathcal{G}(g_k, b_k)$ with non-compactly supported g_k can then be related to Theorem 3.3.9 as follows. With decreasing ε the frequency shift parameter b_k will also decrease according to Theorem 3.3.2 and therefore, the limiting case of $\varepsilon \to 0$ implies $b_k \to 0$ for all $k \in \mathbb{Z}$. Theorem 3.3.9 then states that the nonstationary Gabor frame operator is the identity, if dual frames are chosen according to (3.3.31). In a finite dimensional setting the resulting frame operator is diagonal within machine precision for sufficiently small choices of ε. Hence, no computational expensive inverting of the frame operator is required, resulting in fast algorithms to compute dual frames.

3.3.3 Partitions of Unity

Whenever the smallest and largest eigenvalue of the frame operator are equal, the frame is said to be tight. In the previous cases of painless nonstationary Gabor frames as well as nonstationary Gabor frames based on non-compactly supported functions g_k which lead to a diagonal frame operator this implies

$$\sum_{k\in\mathbb{Z}} |g_k|^2 = \text{const.} \tag{3.3.40}$$

Clearly, whenever two adjacent windows g_k and g_{k+1} are not related anymore, e.g., by a simple shift, it is quite difficult to derive conditions such that (3.3.40) holds. For the construction of approximately tight wavelet frames, however, it is sufficient to consider the following problem

$$\sum_{j\in\mathbb{Z}} f(x - j\Delta T) = \text{const}, \quad \forall x \in \mathbb{R}. \tag{3.3.41}$$

If there exists an f and a corresponding ΔT such that (3.3.41) holds, then f is said to form a partition of unity. The simplification is justified by assuming that all windows g_k result from a single positive function which is appropriately shifted. The following Proposition characterizes the conditions for ΔT such that f forms a partition of unity. The result is similar to the general version of the Poisson summation formula (Benedetto and Zimmermann, 1997; Castaneda et al., 2011).

Proposition 3.3.12 (Characterization of Partitions of Unity)**.** *Let $f \in L^2(\mathbb{R})$ be a real-valued and positive function. Further, let ΔT be a positive constant. Then*

$$\sum_{j\in\mathbb{Z}} f(x - j\Delta T) = \frac{1}{\Delta T}\left(\hat{f}(0) + \sum_{j\in\mathbb{N}_+} \hat{f}\left(\frac{j}{\Delta T}\right)\cos\left(2\pi x\frac{j}{\Delta T}\right)\right), \tag{3.3.42}$$

is constant for all $x \in \mathbb{R}$ whenever the Fourier transform $\hat{f}\left(\frac{j}{\Delta T}\right) = 0$ for all $j \in \mathbb{N}_+$.

Proof. Rewrite the left hand side of (3.3.42) as the convolution of f with the Dirac comb $d(x) = \sum_{j\in\mathbb{Z}} \delta(x - j\Delta T)$,

$$\sum_{j\in\mathbb{Z}} f(x - j\Delta T) = f(x) * d(x). \tag{3.3.43}$$

The Fourier transform of the Dirac comb is again a Dirac comb (Mallat, 2008, Thm. 2.4)

$$\mathcal{F}d = \hat{d}(\nu) = \frac{1}{\Delta T}\sum_{j\in\mathbb{Z}}\delta\left(\nu - \frac{j}{\Delta T}\right). \tag{3.3.44}$$

This gives

$$\begin{aligned}
\sum_{j\in\mathbb{Z}} f(x - j\Delta T) &= \mathcal{F}^{-1}(\hat{f}\cdot\hat{d}) \\
&= \frac{1}{\Delta T}\left(\ldots + \int_{\mathbb{R}}\hat{f}(\nu)\delta\left(\nu + \frac{1}{\Delta T}\right)e^{2\pi i\nu x}d\nu + \int_{\mathbb{R}}\hat{f}(\nu)\delta(\nu)\,e^{2\pi i\nu x}d\nu\right. \\
&\qquad\left. + \int_{\mathbb{R}}\hat{f}(\nu)\delta\left(\nu - \frac{1}{\Delta T}\right)e^{2\pi i\nu x}d\nu + \ldots\right) \\
&= \frac{1}{\Delta T}\left(\ldots + \hat{f}\left(-\frac{1}{\Delta T}\right)e^{-2\pi i\frac{1\cdot x}{\Delta T}} + \hat{f}(0) + \hat{f}\left(\frac{1}{\Delta T}\right)e^{2\pi i\frac{1\cdot x}{\Delta T}} + \ldots\right) \\
&= \frac{1}{\Delta T}\left(\sum_{j\in\mathbb{Z}}\hat{f}\left(\frac{j}{\Delta T}\right)e^{2\pi i\frac{j\cdot x}{\Delta T}}\right).
\end{aligned} \tag{3.3.45}$$

Equation (3.3.42) then follows from simplifying the last expression using Euler's formula and the fact that $\hat{f}$ is an even function. □

Proposition 3.3.12 can be illustrated by considering the following finite dimensional examples, which are also useful when constructing tight wavelet frames later on: the Hann and the Gaussian window as well as the Equalizer.

Example (Hann Window). Let g be the Hann window defined by

$$g(x) = \begin{cases} \frac{1}{2} + \frac{1}{2}\cos(\pi x) & -1 \le x \le 1 \\ 0 & \text{else} \end{cases}, \tag{3.3.46}$$

with the Fourier transform of the squared Hann window $f(x) = (g(x))^2$

$$\hat{f}(\nu) = \frac{3\sin(2\pi\nu)}{8\pi\left(4\nu^5 - 5\nu^3 + \nu\right)}. \tag{3.3.47}$$

Since $g(x)$, and thus $(g(x))^2$, is compactly supported, its Fourier transform is continuous.

Proposition 3.3.12 states that f forms a partition of unity whenever ΔT is chosen such that $\sin(2\pi j/\Delta T) = 0$ for all $j \in \mathbb{N}^+$. This leads to

$$\Delta T = \frac{2}{k} \tag{3.3.48}$$

for $k \in \mathbb{Z}\backslash\{0\}$. For a discrete Hann window of length L

$$g[n] = \frac{1}{2} - \frac{1}{2}\cos\left(2\pi\frac{n}{L}\right), \quad 0 \leq n \leq L-1, \tag{3.3.49}$$

this result implies that $g[n]$ forms a partition of unity if ΔT is an integer divisor of L.

Example (Gaussian Window). For a positive parameter α, let g be the Gaussian such that

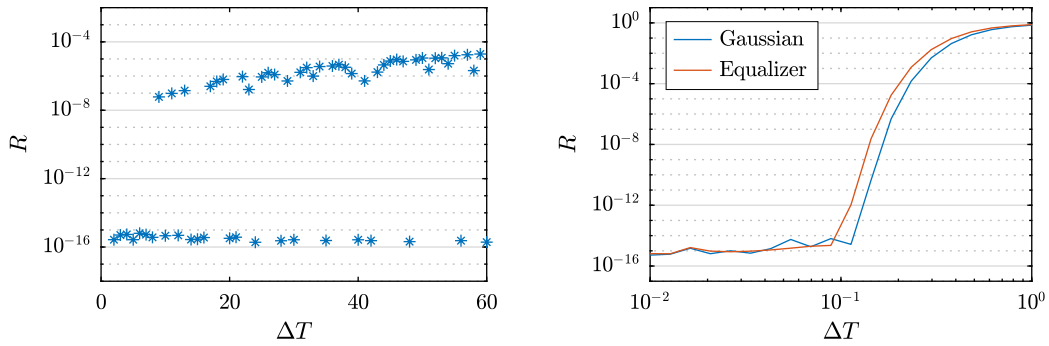
$$\hat{g}(\nu) = \mathcal{F}(g(x)) = \mathcal{F}\left(\sqrt{\frac{\pi}{\alpha}}e^{-\frac{(\pi x)^2}{\alpha}}\right) = e^{-\alpha\nu^2}, \tag{3.3.50}$$

where the Fourier transform of the Gaussian distribution is explicitly given by Abramowitz and Stegun (2012, p. 302, Eq. 7.4.6). Unfortunately, $\hat{g}(\nu) \neq 0$ for all $\nu \in \mathbb{R}$ and hence, a Gaussian cannot form an exact partition of unity according to Proposition 3.3.12. With sufficiently small ΔT, however, $\hat{g}(1/\Delta T)$ can be made arbitrarily small, which leads to a partition of unity in a finite dimensional setting. The summation properties for the squared Gaussian $(g(\nu))^2$ are also easily derived by substituting α appropriately, simply resulting in smaller ΔT values in order to form a partition of unity.

Example (Uncertainty Equalizer). From its definition in (3.1.5) and the corresponding derivation, the uncertainty Equalizer is square integrable. It can be used to construct partitions of unity whenever Proposition 3.3.12 is satisfied. In frequency domain the Equalizer is positive and real-valued, however, a closed form of its Fourier transform does not exist. However, the Fourier transform of Ψ_{Eq} can be characterized as follows. It can be verified that $\Psi_{\text{Eq}} \in L^1(\mathbb{R})$, e.g. for $\mu = 25$

$$\int_0^\infty \frac{1}{e^\mu} e^{\mu\nu(1-\ln\nu)} d\nu = 0.5005, \tag{3.3.51}$$

using numerical methods (Shampine, 2008). As a consequence of the Riemann-Lebesgue Lemma it follows that the Fourier transform satisfies $\hat{\Psi}_{\text{Eq}}(\xi) \to 0$ as $|\xi| \to \infty$ (Rudin, 1987, p. 103). This shows that, similar to the Gaussian case, the Equalizer forms a partition of unity for sufficiently small ΔT in a finite dimensional setting. Similarly, squaring the Equalizer $(\Psi_{\text{Eq}})^2$



(a) squared Hann window

(b) squared Gaussian and squared Equalizer

Figure 3.3: Summation property with respect to the shift parameter ΔT for the three example functions.

just results in doubling the parameter μ and the same arguments as above apply.

Figure 3.3 illustrates the deviation R from the constant $\frac{\hat{g}(0)}{\Delta T}$ defined by

$$R = \left\| \frac{\hat{g}(0)}{\Delta T} - \sum_{j \in \mathbb{Z}} g(\cdot - j\Delta T) \right\|_2 = \left\| \frac{1}{\Delta T} \sum_{j \in \mathbb{N}_+} \hat{g}\left(\frac{j}{\Delta T}\right) \cos\left(2\pi \frac{j}{\Delta T} \cdot\right) \right\|_2, \tag{3.3.52}$$

for all three examples in finite dimensions. The squared Hann window of length $L = 1680$ forms a partition of unity whenever ΔT is an integer divisor of L as can be verified in Fig. 3.3a. The first 24 integer divisors of 1680 are 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 15, 16, 20, 21, 24, 28, 30, 35 40, 42, 48, 56 and 60. The squared Gaussian window in Figure 3.3b is based on (3.3.50) with $\alpha = 1$ and the squared Equalizer as defined in (3.1.5) with $\mu = 25$. It can further be deduced, that the Gaussian decays slightly faster to zero than the Equalizer, since the error R is less for the same ΔT value. The resulting plot shows, that the Gaussian as well as the Equalizer forms a partition of unity within machine precision, whenever ΔT is sufficiently small.

3.3.4 Algorithms for Analysis and Synthesis

For subsequent usages, it is advantageous to consider nonstationary Gabor frames with adaptivity in frequency domain. The following algorithms are therefore based on computational aspects of (3.3.11) but can be easily transferred to (3.3.1). Starting from (3.3.11), analysis coefficients

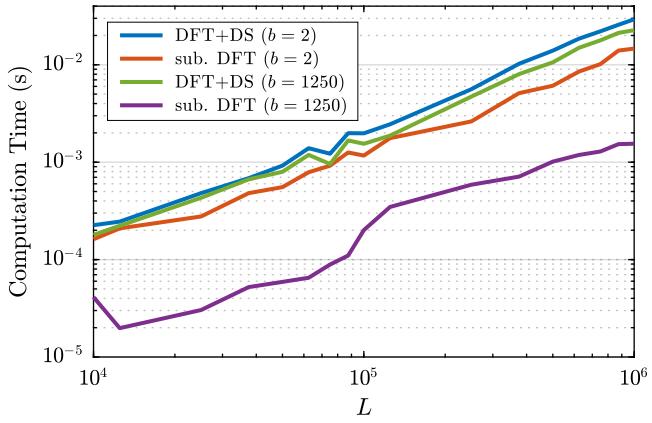


Figure 3.4: Computation time for the subsampled DFT (sub. DFT) and the full DFT with subsequent downsampling (DFT+DS) for $b = 2$ and $b = 1250$.

$c_{k,l}$ can be efficiently computed by considering

$$c_{k,l} = \langle f, T_{a_l k} h_l \rangle \tag{3.3.53}$$

$$= \left\langle \hat{f}, M_{-a_l k} \hat{h}_l \right\rangle \tag{3.3.54}$$

$$= \mathcal{F}^{-1}\left(\hat{f} \cdot \overline{\hat{h}_l} \right)(a_l k), \tag{3.3.55}$$

where the last equation is derived by writing the inner product and modulation in (3.3.54) as an inverse Fourier transform. Based on (3.3.55), the algorithm for computing time-frequency coefficients in a finite dimensional setting is straight forward. However, it might be favorable to further improve the computation of the finite dimensional inverse Fourier transform. Depending on a_l, only a subset of the inverse Fourier transform output is needed. For large a_l only a few coefficients remain after downsampling, leading to the question if such a subsampled Fourier transform can be more efficiently computed. In order to answer this, the following operator is considered:

Definition 3.3.13 (Discrete Projection Operator)**.** For $N < L$ being an integer divisor of L, let the discrete projection operator $\mathcal{Q}_N : \mathbb{C}^L \to \mathbb{C}^N$ be defined by

$$(\mathcal{Q}_N y)[n] = \sum_{j=0}^{L/N-1} y[n + jN], \tag{3.3.56}$$

where $y \in \mathbb{C}^L$, $n = 0, 1, 2, \ldots, N-1$.

Based on this projection operator, the succeeding theorem states an equivalent computation for a subsampled discrete Fourier transform, which is quite similar to (Søndergaard, 2007, Eq. (4.8)).

Proposition 3.3.14 (Subsampled DFT)**.** *Let $y \in \mathbb{C}^L$ and let $b, N \in \mathbb{N}$ such that $Nb = L$. Then the following holds*

$$(\mathcal{F}_L y)[bk] = \frac{1}{\sqrt{b}} \mathcal{F}_N \left(\mathcal{Q}_N y\right)[k], \tag{3.3.57}$$

where $k = 0, 1, 2, \ldots, N-1$.

Proof. The above statement can be proved by a simple rearrangement of the summation order in the definition of the DFT in (2.0.18):

$$(\mathcal{F}_L y)[bk] = \frac{1}{\sqrt{L}} \sum_{l=0}^{L-1} y[l] \omega_L^{kbl} \tag{3.3.58}$$

$$= \frac{1}{\sqrt{L}} \sum_{\beta=0}^{N-1} \sum_{\alpha=0}^{b-1} y[\beta + \alpha N] \omega_L^{(\beta+\alpha N)kb} \tag{3.3.59}$$

$$= \frac{1}{\sqrt{L}} \sum_{\beta=0}^{N-1} \sum_{\alpha=0}^{b-1} y[\beta + \alpha N] \omega_N^{\beta k} \tag{3.3.60}$$

$$= \frac{\sqrt{N}}{\sqrt{L}} \mathcal{F}_N \left(\sum_{\alpha=0}^{b-1} y[\beta + \alpha N] \right)[k] \tag{3.3.61}$$

$$= \frac{1}{\sqrt{b}} \mathcal{F}_N \left(\mathcal{Q}_n y\right)[k]. \tag{3.3.62}$$

□

An equivalent calculation leads to a corresponding result for the inverse discrete Fourier transform. Figure 3.4 illustrates that Proposition 3.3.14 is useful even for small downsampling factors. It compares computation times using the subsampled DFT (r.h.s. of (3.3.57)) and a DFT with subsequent downsampling (l.h.s. of (3.3.57)) for signals with increasing length L. The figure illustrates that even for the smallest downsampling factor $b = 2$, the subsampled

Algorithm 1: Nonstationary Gabor Analysis

Input : f - function of length L
L - signal length
$\hat{h}_l$ - set of K window functions in frequency domain
a_l - set of K modulation parameters

Output: c_l - analysis coefficients

```
F ⟵ F_L f
for l ← 1 to K do
    T ⟵ Q_{L/a_l} (F · ĥ_l)
    c_l ⟵ F^{-1}_{L/a_l} T
end
```

DFT yields a gain of factor 2. With larger downsampling factors this increases to a speedup of factor 10. The resulting algorithm for computing nonstationary Gabor coefficients for a signal $f \in \mathbb{C}^L$ of length L can now be summarized in Algorithm 1.

Synthesis follows by reversing the steps of analysis in (3.3.55), provided the dual windows γ_l are given. This is best formulated in the continuous case as

$$Sf = \mathcal{F}^{-1}\mathcal{F}\left(\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} c_{k,l}\, T_{a_l k}\gamma_l\right) \tag{3.3.63}$$

$$= \mathcal{F}^{-1}\left(\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} c_{k,l}\, M_{-a_l k}\hat{\gamma}_l\right) \tag{3.3.64}$$

$$= \mathcal{F}^{-1}\left(\sum_{l\in\mathbb{Z}} \mathcal{F}(c_l)\,\hat{\gamma}_l\right) \tag{3.3.65}$$

Before the synthesis algorithm in a finite dimensional setting can be stated, however, the following definition is needed.

Definition 3.3.15 (Discrete Periodization Operator)**.** Let $\mathcal{P}_N : \mathbb{C}^N \to \mathbb{C}^L$, with $N < L$ and $L/N \in \mathbb{Z}$, be defined as

$$(\mathcal{P}_N y)[j] = y[j \bmod N], \tag{3.3.66}$$

where $y \in \mathbb{C}^N$ and $j = 0, 1, 2, \ldots, L-1$.

Algorithm 2: Nonstationary Gabor Synthesis

Input : c_l - wavelet coefficients
L - length of transform
γ_l - set of K dual window functions in frequency domain
a_l - set of K modulations

Output: f - reconstructed signal

1 $f \longleftarrow 0$
2 **for** $l \leftarrow 1$ **to** K **do**
3 $\quad T \longleftarrow \mathcal{P}_{\frac{L}{a_l}} \left(\mathcal{F}_{\frac{L}{a_l}} c_l \right)$
4 $\quad f \longleftarrow f + \gamma_l \cdot T$
5 **end**
6 $f \longleftarrow \mathcal{F}_L^{-1} f$

With this definition, Algorithm 2 summarizes the computation of the inverse nonstationary Gabor transform based on (3.3.65). The purpose of using the periodization operator $\mathcal{P}_N$ in line 3 of Algorithm 2 needs some justification. When reversing the analysis steps of Algorithm 1 an inverse of the projection operator $\mathcal{Q}_N$ is needed to recover y from $\mathcal{Q}_N y$. In general this is not possible, since $\mathcal{Q}_N$ is not invertible. If, however, the essential support of y is smaller than N, $\mathcal{P}_N \mathcal{Q}_N y$ leads to a periodized version of $\mathcal{Q}_N y$ with period N. Since y is assumed to be concentrated in an interval smaller than N only one of these periods coincides with y. The subsequent pointwise multiplication of $\mathcal{P}_N \mathcal{Q}_N y$ with the dual window γ_l weighs the periodized signal such that the desired period remains, provided that the window $\hat{h}_l$ and its dual γ_l have the same localization properties.

It is easily verified that both assumptions are satisfied for painless constructions. The parameter a_l is always chosen such that the support of corresponding windows $\hat{h}_l$ is smaller than $N = \frac{L}{a_l}$. Furthermore, the corresponding dual windows γ_l have the same compact support as $\hat{h}_l$ since the frame operator is diagonal. As a result, Algorithm 1 and 2 can be simplified in the painless case by utilizing the exact positions where windows $\hat{h}_l$ and γ_l are localized.

In the case of non-compactly supported window functions $\hat{h}_l$ both assumptions are still valid, albeit only approximately. By choosing an appropriate ε, the corresponding a_l-values are small enough such that $\hat{h}_l$ is sufficiently small outside its essential support of size $N = \frac{L}{a_l}$. Further,

with sufficient overlap, lower and upper bound in (3.3.13) are reasonably close to each other such that dual windows have approximately the same localization properties as the original windows.

Analysis and synthesis procedures are already implemented in the LTFAT toolbox for painless nonstationary Gabor frames (Průša et al., 2014). The algorithms above extend these procedures to non-compactly supported window functions.

3.4 Wavelet Frames Based on Nonstationary Gabor Frames

3.4.1 Preliminaries

In the following the connection of wavelet and nonstationary Gabor frames is shown. The discretization of the dilation operator in (2.0.7) will be given for $l \in \mathbb{Z}$ and $d > 1$ by

$$D_{d^l}\psi(x) = \sqrt{d^l}\,\psi\left(d^l x\right). \tag{3.4.1}$$

The resulting discretized wavelet transform of a function f with wavelet ψ reads

$$W_\psi^{a,d} f(k,l) = \langle f, D_{d^l} T_{ak}\psi\rangle = \langle f, T_{ad^l k} D_{d^l}\psi\rangle = \left\langle \hat{f}, M_{-ad^l k} D_{d^{-l}}\Psi\right\rangle, \tag{3.4.2}$$

where $\Psi = \hat{\psi}$ denotes the Fourier transform of ψ in the following.

Setting $h_l = D_{d^l}\psi$ and $a_l = ad^l$ the collection $\mathcal{G}(h_l, a_l)$ is a nonstationary Gabor system. It is a nonstationary Gabor frame, whenever the conditions of Theorem 3.3.2 are satisfied for band limited wavelets Ψ. Similarly, whenever non band limited wavelets Ψ have polynomial decay around δ-separated sets in frequency domain, the system $\mathcal{G}(h_l, a_l)$ is also a frame, provided that corresponding time sampling is sufficiently dense. In fact, all of the previously made assumptions on nonstationary Gabor frames as well as Theorem 3.3.9 also apply to the wavelet case. This implies that the wavelet frame operator with wavelets h_l and corresponding dual wavelets γ_l° computed according to (3.3.10) converges to the identity whenever shift parameters a_l tend to zero for all $l \in \mathbb{Z}$. Additionally, the wavelet ψ has to satisfy proper decay conditions and the dilation parameter d is chosen such that (3.3.9) holds.

The dual wavelets defined by (3.3.10) are scaled versions of ψ itself, whenever corresponding frames are tight. Therefore, dilates of a function Ψ have to satisfy

$$\sum_{l\in\mathbb{Z}} |D_{d^{-l}}\Psi|^2 = \text{const.} \tag{3.4.3}$$

Daubechies et al. (1986, Sec. C) introduced logarithmic warping in order to obtain tight wavelet frames generated from a system of translates instead of dilates. Such logarithmic warping is also used in a group theoretical approach about the uncertainty principle of the wavelet transform (Levie et al., 2014). Therefore, two important properties resulting from logarithmic warping can be summarized in the following Proposition.

Proposition 3.4.1. *Let $\mathcal{W} : \mathbb{R}^+ \to \mathbb{R}$ be the warping function $\mathcal{W} = \log_d$ for some $d > 0$. Denote by $\Psi(\nu)$ the function composition $\Psi(\nu) = (\phi \circ W)(\nu)$ for any function ϕ defined in the warped frequency domain. Then, the following statements hold:*

i) If $\phi \in L^2(\mathbb{R})$, then Ψ is an admissible wavelet.

ii) Dilates of the function Ψ turn into translates of ϕ.

Proof. The first statement follows from

$$C_\Psi = \int_0^\infty \frac{|\Psi(\nu)|^2}{|\nu|} d\nu = \int_0^\infty \frac{|\phi(\log_d \nu)|^2}{\nu} d\nu = \int_{\mathbb{R}} \ln d \, |\phi(\zeta)|^2 \, d\zeta = \ln d \, \|\phi\|_2^2 \,, \tag{3.4.4}$$

the second from

$$D_{d^{-l}} \Psi(\nu) = \frac{1}{\sqrt{d^l}} \Psi\left(d^{-l}\nu\right) = \frac{1}{\sqrt{d^l}} \phi\left(\log_d \nu - l\right). \tag{3.4.5}$$

□

Equation (3.4.5) shows how dilates can form a partition of unity. Assuming

$$\sum_{l\in\mathbb{Z}} \left|\hat{h}_l\right|^2 = \sum_{l\in\mathbb{Z}} |D_{d^{-l}}\Psi(\nu)|^2 = \sum_{l\in\mathbb{Z}} \left|\frac{1}{\sqrt{d^l}} \phi\left(\log_d \nu - l\right)\right|^2 = \text{const}, \tag{3.4.6}$$

partitions of unity can be computed according the results from the previous section. Nevertheless, there are a few aspects in the finite dimensional setting which require careful considerations. For example, dilates with small l will result in an infinite number of window functions in the neighborhood of the zero frequency. Furthermore, since the whole frequency axis needs to be covered, some functions have to be placed on the negative frequency axis as well. Before these aspects are addressed in the next section, general parameters for the construction of wavelet frames are introduced. These parameters are chosen to be consistent with the painless

nonstationary Gabor construction in the LTFAT toolbox (Průša et al., 2014). The discrete and finite set of dilations is mainly characterized by the following parameters:

- $f_{\min}$: the location in Hz of the first window on the frequency axis,
- *bw*: the bandwidth in Hz of the first window on the frequency axis,
- *bins*: the number of windows per frequency bin, i.e., $d^{\frac{l}{bins}}$ with $l \in \mathbb{Z}$.

The construction schemes presented in the following section extends the functionality of the methods already implemented in the LTFAT toolbox, including non-compactly supported functions on one hand, and a more flexible construction of tight wavelet frames for compactly supported functions on the other hand.

3.4.2 Tight Wavelet Frames with Compactly Supported Functions

As a representative for all compactly supported functions, the construction of wavelet frames $\mathcal{G}(h_l, a_l)$ is described with Hann windows. It is possible, however, to generalize the results of this section to any compactly supported window in the finite dimensional case.

In the following let ϕ be the Hann window defined in (3.3.46). The warping function is given by $\mathcal{W} = \tau \log_d$, where $\tau > 0$ is a fixed constant. The resulting warped Hann window Ψ is then defined as

$$\Psi_{d,l,\tau}(\nu) = D_{d^{-l}}\Psi(\nu) = \begin{cases} \frac{1}{\sqrt{d^l}}\left(\frac{1}{2} + \frac{1}{2}\cos\left(\pi\tau(\log_d \nu - l)\right)\right) & d^{l-\frac{1}{\tau}} \leq \nu \leq d^{l+\frac{1}{\tau}} \\ 0 & \text{else} \end{cases}, \tag{3.4.7}$$

for $\nu, d \in \mathbb{R}^+$ and $l \in [l_1, l_2 = l_1 + \Delta l, l_3 = l_1 + 2\Delta l, \ldots, l_n = l_1 + (n-1)\Delta l]$. Hence, the parameter τ serves for adjusting the bandwidth *bw* of the first warped Hann window.

If the first wavelet is supposed to be specified at $f_{\min}$ it follows by a simple computation that

$$l_1 = \log_d f_{\min}. \tag{3.4.8}$$

For a given bandwidth *bw* the parameter τ is computed by

$$\tau = \frac{\ln(d)}{\ln\left(\frac{1}{2}\left(bw \cdot d^{-l_1} + \sqrt{\left(bw \cdot d^{-l_1}\right)^2 + 4}\right)\right)}, \tag{3.4.9}$$

Algorithm 3: Tight Wavelet Frames with Hann Windows

Input : L - signal length
sf - sampling frequency
d - scaling base
$f_{\min}$ - location of first window
bw - bandwidth of first window
$bins$ - number of frequency bins

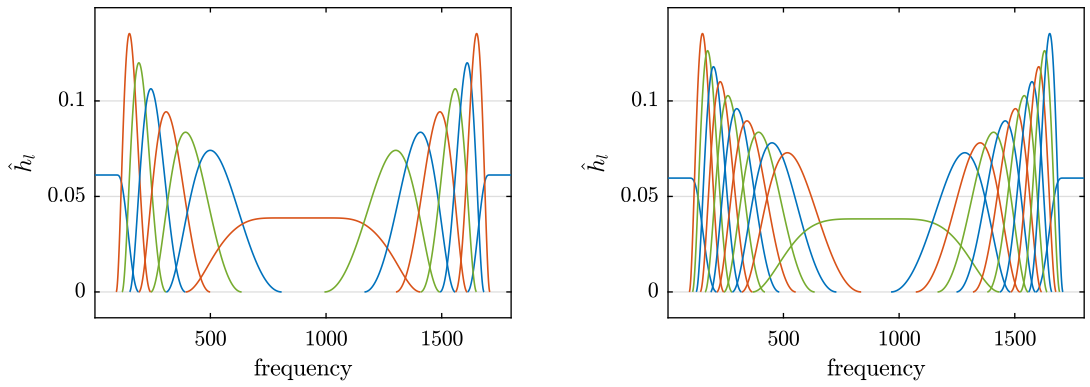
Output: $\hat{h}_l$ - set of window functions
a_l - set of corresponding time sampling steps

compute initial parameters l_1, τ and Δl according to Equations (3.4.8) - (3.4.10)
$l \leftarrow l_1$
while $d^{l+\frac{1}{\tau}} < \frac{sf}{2}$ **do**
  $\hat{h}_l \leftarrow \Psi_{d,l,\tau}(\nu)$
  $\hat{h}_{-l} \leftarrow \Psi_{d,l,\tau}(-\nu)$
  $a_{-l,l} \leftarrow L\left(d^{l+\frac{1}{\tau}} - d^{l-\frac{1}{\tau}}\right)^{-1}$
  $l \leftarrow l + \Delta l$
end
add two cover functions at zero and Nyquist frequency

which results from setting $bw = d^{l_1+\frac{1}{\tau}} - d^{l_1-\frac{1}{\tau}}$ and solving for τ. If the difference between two consecutive l's satisfies

$$\Delta l = \frac{2}{\tau \cdot bins}, \tag{3.4.10}$$

with $bins > 2$ being an integer, the translated versions of the warped Hann windows will sum up to a constant, cf. (3.3.48). This procedure can be repeated, progressively covering the positive frequency axis until the support of a scaled window reaches the Nyquist frequency. The time sampling parameter a_l for each corresponding window $\hat{h}_l$ is chosen according to Corollary 3.3.4. Since it is assumed that all input signals f are real valued, it is sufficient to cover negative frequencies by mirroring corresponding wavelets, i.e., $\Psi_{d,l,\tau}(-\nu)$. Furthermore, two additional windows have to be added covering the zero and Nyquist frequency. Usually, wavelet coefficients resulting from these two windows do not contain any useful information and can therefore be neglected (Balazs et al., 2011). For frames, however, the windows are needed



(a) *bins* $= 4$ **(b)** *bins* $= 7$

Figure 3.5: Two examples of window functions with warped Hann windows, each constituting a tight wavelet frame with corresponding a_l, differing only in the number of bins.

to preserve the structure of the partition of unity. The construction scheme is summarized in Algorithm 3.

Two examples of tight wavelet frames based on the warped Hann window can be seen in Figure 3.5. Signal length is $L = 1800$ at a sampling frequency of $sf = 1800\,\text{Hz}$. Furthermore, $d = 2, f_{\text{min}} = 150\,\text{Hz}$ and $bw = 150\,\text{Hz}$. Figure 3.5a and 3.5b show windows $\hat{h}_l$ for *bins* $= 4$ and *bins* $= 7$, resulting in tight wavelet frames $\mathcal{G}(h_l, a_l)$ with bounds $A = 1.5$ and $A = 2.625$ and redundancies 2.76 and 4.18, respectively. For convenient plotting, each window is normed such that $\left\|\hat{h}_l\right\|_2 = 1$. The frequency axis in Fig. 3.5 is adapted to match the frequency domain representation of MATLAB 's fft-routine, where the second half with frequencies larger than the Nyquist frequency corresponds to negative frequencies.

3.4.3 Approximately Tight Wavelet Frames with Non-Compactly Supported Functions

General Construction

In this section a general framework to construct wavelet frames from non-compactly supported mother wavelets is presented. In order to be consistent with the construction of frames with compactly supported functions, the same characteristic parameters f_{min}, *bw* and *bins* are used. The position of the first wavelet f_{min} and the number of bins can be adapted without further consideration. The bandwidth *bw* of the first wavelet, however, needs to be defined for functions

without compact support. A common approach would be to define the bandwidth as the width where the function drops below a certain point, e.g. the 3 dB bandwidth. Similar to this approach, the bandwidth *bw* can be defined by utilizing the essential support in Definition 3.3.11. Hence,

$$bw = N_2^{\varepsilon_{bw}} - N_1^{\varepsilon_{bw}}, \tag{3.4.11}$$

where the bandwidth depends on the chosen ε_{bw}.

As a starting point of the construction scheme, let ϕ be a continuous square integrable function, which attains its maximum at zero. Similar to the painless construction in the previous section, the warping function is given by $\mathcal{W} = \tau \log_d$ for some parameter $\tau > 0$. The scaled versions of ϕ can be written as

$$\Psi_{d,l,\tau} = D_{d^{-l}} \Psi(\nu) = \phi(\tau \log_d \nu - \tau l), \tag{3.4.12}$$

where $d, \nu > 0$ and $l \in [l_1, l_2 = l_1 + \Delta l, l_3 = l_1 + 2\Delta l, \ldots, l_n = l_1 + (n-1)\Delta l]$.

Initial parameters l_1, τ and Δl can then be computed in the following manner. The position of the first wavelet immediately yields

$$l_1 = \log_d f_{\min}, \tag{3.4.13}$$

since the maximum value of ϕ is located at zero. The parameter τ can be computed by considering the bandwidth of the first wavelet

$$\tau \log_d \nu - \tau l_1 \in U^{\varepsilon_{bw}} \tag{3.4.14}$$

$$\nu \in \left[d^{l_1 + \frac{N_1^{\varepsilon_{bw}}}{\tau}}, d^{l_1 + \frac{N_2^{\varepsilon_{bw}}}{\tau}} \right], \tag{3.4.15}$$

for some ε_{bw}. Hence

$$bw = d^{l_1 + \frac{N_2^{\varepsilon_{bw}}}{\tau}} - d^{l_1 + \frac{N_1^{\varepsilon_{bw}}}{\tau}}. \tag{3.4.16}$$

The nonlinear expression in (3.4.16) is guaranteed to have a solution for $\tau \in (0, \infty)$, which can be seen from the following Lemma.

Lemma 3.4.2. *Let $bw, l_1 > 0$ and $d > 1$, be positive real numbers. Further let, $N_1, N_2 \in \mathbb{R}^+$ such that $N_2 > N_1$. Then, the nonlinear expression $\Xi(\tau)$*

$$\Xi(\tau) = d^{\frac{N_2}{\tau}} - d^{\frac{N_1}{\tau}} - d^{-l_1} bw, \tag{3.4.17}$$

has exactly one zero in the interval $(0, \infty)$.

Proof. Obviously, $\lim_{\tau\to\infty} \Xi(\tau) = -d^{-l_1} bw < 0$. On the other hand, $d^{N_2} > d^{N_1}$ since $N_2 > N_1$ and therefore $\Xi(\tau)$ is strictly monotonic decreasing with $\lim_{\tau\to 0} \Xi(\tau) = +\infty > 0$. □

The parameter τ can then be estimated using numerical methods for finding real roots, e.g., as proposed by Forsythe et al. (1977, p. 161). After computing the parameters for the initial wavelet, dilates can be constructed with

$$\Delta l = \frac{1}{\tau \cdot bins}. \tag{3.4.18}$$

By subsequently increasing l with Δl the functions $\Psi_{d,l,\tau}$ cover the positive frequency axis up to the Nyquist frequency. Similar to the painless construction, negative frequencies are covered by using mirrored versions of the corresponding windows. Again, two additional windows are needed to cover zero and Nyquist frequency explicitly. The time sampling parameter a_l is computed as described in Section 3.3.2: for each scaled wavelet $\hat{h}_l$ the time sampling step a_l is chosen such that

$$a_l \leq L \left(\left(N_2^{\varepsilon_a}\right)_l - \left(N_1^{\varepsilon_a}\right)_l \right)^{-1}, \tag{3.4.19}$$

where $\left(N_{1/2}^{\varepsilon_a}\right)_l$ denotes the bounds of the essential support of corresponding wavelets $\hat{h}_l$ based on parameter ε_a. Note the difference between ε_a and ε_{bw}: the former is a measure responsible for the redundancy of the frame, the latter only affecting the bandwidth of the first wavelet. According to the results in Section 3.3.2, the corresponding wavelet frame operator is numerically diagonal for suitable small ε_a and the canonical dual frame of the painless approximation. The general construction scheme is summarized in Algorithm 4.

Examples of Tight Wavelet Frames

In the following finite dimensional examples let $L = 1800$ and the sampling frequency $sf = 1800\,\text{Hz}$ be fixed. Further, the first wavelet should be located at $f_{\min} = 150\,\text{Hz}$ with a bandwidth $bw = 150\,\text{Hz}$ for $\varepsilon_{bw} = 1\text{e}{-6}$. Different generating functions ϕ might lead to different total numbers of windows for the same *bins* parameter, depending on the decay behavior. Thus, the number of bins is adjusted for each function such that the total number of wavelets $K = 38$ is constant. In the case of tight frame constructions the function

$$Z = \sum_{l=1}^{K} \left| \hat{h}_l \right|^2, \tag{3.4.20}$$

Algorithm 4: Wavelet Frames with Non-Compactly Supported Functions

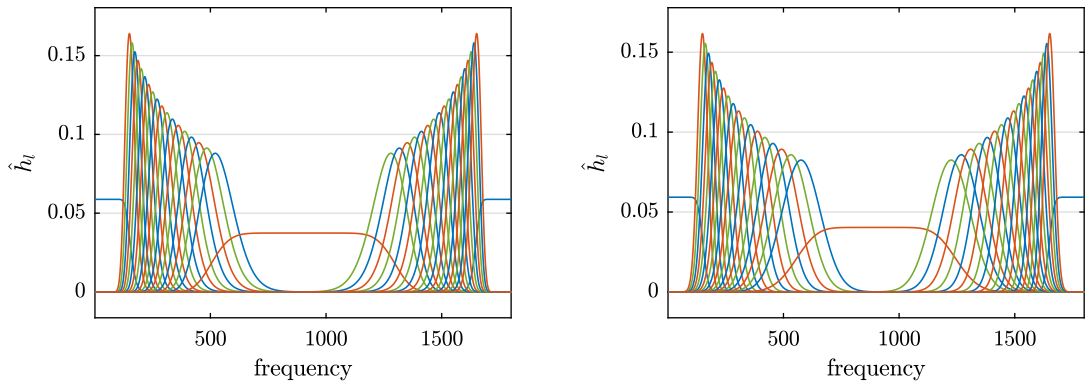
Input : ϕ - generating wavelet
L - signal length
sf - sampling frequency
d - scaling base
$f_{\min}$ - location of first window
bw - bandwidth of first window
ε_{bw} - bandwidth parameter for essential support
$bins$ - number of frequency bins
ε_a - parameter for choosing time sampling steps

Output: $\hat{h}_l$ - set of window functions
a_l - set of corresponding time sampling steps

compute initial parameters l_1, τ (based on ε_{bw}) and Δl according to Equations (3.4.8), (3.4.16) and (3.4.18)
$l \leftarrow l_1$
while $d^{l+\frac{N_2}{\tau}} < \frac{sf}{2}$ **do**
  $\hat{h}_l \leftarrow \Psi_{d,l,\tau}$
  $\hat{h}_{-l} \leftarrow$ mirrored version of $\Psi_{d,l,\tau}$
  $a_{-l,l} \leftarrow L\left(N_2^{\varepsilon_a} - N_1^{\varepsilon_a}\right)^{-1}$
  $l \longleftarrow l + \Delta l$
end
add the two cover functions at zero and Nyquist frequency

should be, at least within numerical precision, a constant. As shown in Section 3.3.3, this depends on the chosen number of bins.

First, consider ϕ to be the Gaussian. The resulting set of warped Gaussian wavelets $\hat{h}_l$ is illustrated in Figure 3.6a. Figure 3.6b illustrates the set of wavelets, where ϕ is chosen such that Ψ is the Equalizer with $\mu = 25$, i.e., $\phi(\nu) = \Psi_{\mathrm{Eq}}(d^\nu)$ and $\phi = 0$ for negative frequencies. This substitution is justified since the domain of the Equalizer is restricted to $\mathbb{R}^+$. The singularity at $\nu = 0$ introduced by setting function values to zero for negative frequencies is in a finite dimensional setting close to machine precision. Corresponding values Z for the partition of unity are given by $Z = 3.18$ with absolute numerical error $\max(Z) - \min(Z) = 3.6\mathrm{e}{-15}$ for



(a) Warped Gaussian

(b) Equalizer

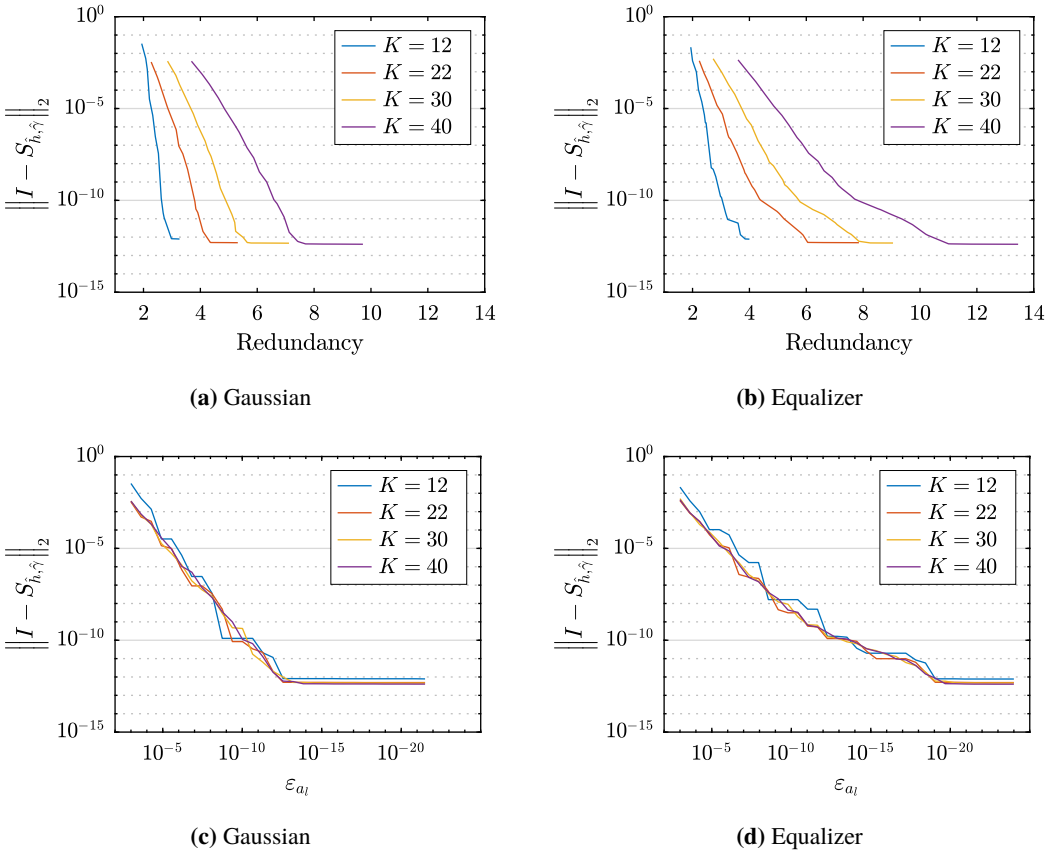
Figure 3.6: Examples of non-compactly supported wavelet functions forming a partition of unity.

the warped Gaussian and $Z = 3.25$ with $\max(Z) - \min(Z) = 3.1\text{e}{-15}$ for the Equalizer. Each set of wavelet functions would constitute a tight frame, at least within numerical precision, if parameters ε_a are chosen sufficiently small. A more detailed evaluation follows in the subsequent section.

3.4.4 Evaluation of Wavelet Frames

After presenting the construction schemes for wavelet frames in the preceding section, in the following these frames are evaluated more closely. The main result in Section 3.3.2 states that the wavelet frame operator is the identity for sufficiently dense time-sampling parameters a_l, given that consecutive window functions have adequate overlap. In the case of non-compactly supported and strictly positive generating functions ϕ, e.g., Gaussians, the last condition is always satisfied in a finite dimensional setting. As a consequence, the diagonality of the frame operator remains unaffected by the total number K of chosen wavelets. This result is also reflected by Corollary 3.3.10, indicating that primarily the choice of sufficiently small time-sampling parameters a_l ensures that the frame operator is approximating the identity operator.

Now, the question might be asked how diagonality of the wavelet frame operator affects redundancy, and hence density of time sampling parameters, of the frame. Obviously, if the required redundancy is too large such frame constructions might not be a great advantage from a computational perspective. There is no trivial way to directly connect redundancy and



(a) Gaussian

(b) Equalizer

(c) Gaussian

(d) Equalizer

Figure 3.7: Approximation error $\left\|I - S_{\hat{h},\hat{\gamma}}\right\|_2$ in dependence of the frame redundancy and the parameter ε_{a_l} for the Gaussian and the Equalizer.

diagonality of the frame operator though. This is best motivated by the following example.

For computational convenience, fix $L = 480$, $sf = 480\,\text{Hz}$, $f_{\text{min}} = 40\,\text{Hz}$, $d = 2$ and a bandwidth $bw = 20\,\text{Hz}$ for $\varepsilon_{bw} = 1\text{e}{-6}$. For each ϕ adjust the number of bins such that the total number of wavelets K is identical. Corresponding dual windows $\hat{\gamma}$ are computed according to (3.3.36). With a fixed K the approximation error $\left\|I - S_{\hat{h},\hat{\gamma}}\right\|_2$ can then be estimated depending on the parameter ε_{a_l}, which is directly related to the redundancy of the frame. The smaller ε_{a_l} is chosen, the larger the essential support of corresponding $\hat{h}_l$. Consequently, the density of the time sampling increases according to (3.4.19). Figure 3.7 illustrates the approximation error for the Gaussian and the Equalizer for different K in dependance of the redundancy and the parameter ε_{a_l}. Figures 3.7a and 3.7b show the approximation error as the redundancy

increases for different total numbers of wavelets. It can be seen, that once a certain redundancy is reached, the frame operator is close to the identity operator (within machine precision) and the approximation error stagnates with further increasing redundancy. Furthermore, since the Equalizer does not decay as fast as the Gaussian, the redundancy is larger before the error reaches its saturation value. In terms of the parameter ε_{a_l}, Figures 3.7c and 3.7d illustrate nicely that the total number of wavelets K is irrelevant concerning the diagonality of the wavelet frame operator. The value of ε_{a_l} for which the approximation error reaches its saturation depends, as expected, on the decay behavior of the generating wavelet. In the Gaussian case it is around 1e−14 and for the Equalizer around 1e−19. In both cases numerical experiments have shown, that this value is independent of the chosen $f_{\min}$ and bw.

Comparing the above results with the ones from Feichtinger et al. (2014) shows the flexibility of the proposed approach. Feichtinger et al. approximated canonical dual frame elements by inverting the wavelet frame operator using its thresholded pseudoinverse. They show that this is, however, only possible for certain dilation parameters d, else the frame operator would not commute with time shifts anymore, which is also discussed by Chui and Shi (2000). Furthermore, in their numerical experiments the constructed frame has redundancy between 1.5 and 2, and dual windows can be approximated within an error of 0.007 compared to the original canonical dual window (Feichtinger et al., 2014, Fig. 2). This leads to reconstruction errors of the same magnitude when doing synthesis based on these approximated dual frames.

3.5 Conclusion

Based on the results of Dörfler and Matusiak (2014, 2015) explicit construction schemes for approximately tight wavelet frames with non-compactly supported functions have been introduced in this section. Sufficient conditions are derived, for which a wavelet frame $\mathcal{G}(h_l, a_l)$ leads to a diagonal frame operator. This allows the numerical construction of wavelet transforms with perfect reconstruction. Wavelets can be based on non-compactly supported functions in frequency domain, extending the constructions presented by Bayram and Selesnick (2009, Eq. (42) on p. 2964) and Balazs et al. (2011). While at first glance obviously not very convenient, the usage of non-compactly supported functions in frequency domain allows wavelet transforms with compactly supported wavelets in time domain. Thus, real-time implementations of the continuous wavelet transform are possible, similar to the proposed approach by Holighaus et al. (2013). Furthermore, the algorithm presented by Balazs et al. (2011) is extended by a more flexible construction scheme for tight wavelet frames with compactly supported wavelets.

The proposed construction is based on computing dual frame elements without inverting the actual frame operator. Since canonical dual wavelet frames may not be wavelet frames itself, see for example (Bownik and Lemvig, 2007; Christensen, 2006; Daubechies and Han, 2002), the structure and localization properties of canonical dual atoms might be lost. Defining dual windows as proposed in (3.3.10) for compactly supported functions, the structure can be controlled fairly easily by assuming adequate overlap of $\hat{h}_l$. Additionally, with sufficient overlap the dual windows in a finite dimensional setting satisfy $\hat{\gamma}_l = c\hat{h}_l$ up to machine precision for some constant $c > 0$. Hence, with certain redundancy of the frame it can be guaranteed that the approximate dual frame is again a wavelet frame.

The increased redundancy, however, might cause higher computational complexities if, for example, analysis and synthesis algorithms are invoked repeatedly. In this case it may be beneficial to choose a single time sampling parameter $a_l = a$ for all l. The frame operator then commutes with time shifts and canonical dual frames can be computed using the inverse of the frame operator. This way, the redundancy can be reduced by choosing a larger time sampling parameter, while still retaining perfect reconstruction. Existing algorithms like (Janssen and Søndergaard, 2007) can then be used to efficiently invert the frame operator. Obviously, such an approach increases computation time of the dual frame elements, but might reduce complexity of computationally demanding applications, where the dual frame is frequently needed.

Chapter 4

Audio Inpainting: Evaluation of Time-Frequency Representations and Structured Sparsity Approaches

This chapter is based on (Lieb, 2015) and is a slightly amended version of the publication (Lieb and Stark, 2018); therefore, most passages have been quoted verbatim.

4.1 Introduction

The term inpainting originally refers to filling in blanks or gaps in images in an indistinguishable way. It is a well known problem, resulting form restoring damaged paintings and artwork (Bertalmío et al., 2000). Whenever spots of paint are missing, they have to be filled in a meaningful way, depending on the neighborhood of the spot. Therefore, surrounding textures are used to estimate the original color of the spot. Clearly, this leaves a lot of room for interpretation and inpainting becomes highly subjective, as there might be more than one plausible solution to how lost information can be restored. In modern ages inpainting images is still important as digital images can be distorted by missing pixels, overlaid text or unwanted structures.

Besides images, missing information can also be a problem in audio signals. It can arise from impulsive noise like clicks, from clipping where the signal is truncated at a certain threshold or simply from lost bits in signal transmission (Godsill and Rayner, 1998). Adler et al. (2012) denoted the problem of filling in missing audio data as audio inpainting. Similar to image inpainting, many different approaches for inpainting audio signals have been proposed. There

are algorithms based on interpolation and autoregressive models (Etter, 1996; Janssen et al., 1986; Lagrange et al., 2005), extrapolation (Maher, 1993), imputation (Gemmeke et al., 2010; Smaragdis et al., 2009), concealment (Ofir et al., 2007) or sparsity based approaches (Adler et al., 2012; Perraudin et al., 2014; Siedenburg et al., 2014). When the number of missing samples is large, in particular approaches based on sparse signal representations recently have been successfully employed (Mach, 2016; Mach and Ozdobinski, 2013).

Sparse representations of audio signals are even more important in the context of audio coding (Plumbley et al., 2010). Usually, sound or audio signals exhibit only few important frequencies. Assuming that missing information in time domain will not drastically affect this property, it is plausible that sparsity in frequency domain is a useful guide to reconstruct the original. In fact, sparsity in time-frequency representations has been successfully applied to denoise (Siedenburg and Dörfler, 2012, 2013; Yu et al., 2008) as well as to inpaint audio data (Mach, 2016).

Mathematically, the audio inpainting problem is formulated as proposed by Adler et al. (2012). Let $s \in \mathbb{R}^L$ be an observed audio signal with partially missing information and let these positions in the signal be known. Consider a masking operator $\mathcal{M} : \mathbb{R}^L \to \mathbb{R}^M$ which selects the reliable part of s. In other words, $\mathcal{M}$ is a rectangular matrix whose elements are obtained by taking the rows of an $L \times L$ identity matrix which correspond to the non-missing samples of s. In the following let $y = \mathcal{M}s$. Further, denote by $\Phi^* : \mathbb{C}^N \to \mathbb{R}^L$ the synthesis operator of some sparsity inducing linear operator $\Phi : \mathbb{R}^L \to \mathbb{C}^N$ such that $\Phi^*\Phi = I$. Examples are the Gabor transform or more adaptive time-frequency representations like wavelets or ERBlets (Necciari et al., 2013). Audio inpainting then describes the problem of finding missing segments of a signal such that the time-frequency representation is sparse, while at the same time staying close to the original samples. This is reflected by the following minimization problem

$$\min_{x \in \mathbb{C}^N} \|x\|_0 \quad \text{subject to} \quad y = \mathcal{M}\Phi^* x. \tag{4.1.1}$$

However, the above equation is NP-hard. In signal processing a convex relaxation has been introduced by Chen et al. (1998) as Basis Pursuit Denoising and independently by Tibshirani (1996) which he denoted LASSO (Least Absolute Shrinkage and Selection Operator). In both cases the ℓ_0-norm is replaced by a ℓ_1-norm. It is obvious that (4.1.1) depends crucially on the choice of Φ. Adler et al. (2012) have used a regular Gabor transform and in the following the potential of more flexible representations, in particular wavelets and ERBlets, is explored.

Another point to consider is the structure of time-frequency coefficients of audio signals. It is known that audio signals consist mostly of tonal and transient parts (Daudet et al., 2001). Thus,

a single large coefficient among a neighborhood of small values might be discarded, whereas, on the other hand, a single small value in a neighborhood of large values should be kept. It has been shown that this concept, denoted social or structured sparsity, improves denoising and declipping of audio signals in similar inverse problems (Dörfler et al., 2013; Kowalski et al., 2013; Siedenburg et al., 2014).

Recent audio inpainting results published are based almost exclusively on declipping, where the proposed algorithms exploit the fact that missing information is assumed to attain values larger than or equal to the clipping threshold (Bilen et al., 2015a,b; Kitić et al., 2013, 2015; Siedenburg et al., 2014). In more general inpainting problems, as considered in the following, such assumptions do not apply and related algorithms have to be adapted. Steps in this direction have been undertaken by Adler et al. (2012) and Kereliuk and Pasquier (2013) for gap lengths of up to 10 ms. The combination of suitable sparse signal representations with structured sparsity approaches, however, has not been investigated for such general inpainting problems.

With the tools mentioned above, inpainting procedures are introduced working well for increased gap lengths of up to 30 ms. Numerical results indicate that the proposed methods favor non-stationary signal transforms as compared to the Gabor transform and clearly outperform the current state-of-the-art approach used by Adler et al. (2012) and Kereliuk and Pasquier (2013). It can be further shown that exploiting time-frequency structures is beneficial for inpainting missing audio data, in terms of signal-to-noise ratio and, even more importantly, in terms of perceived audio quality. This provides a useful complement to the literature and extends a previously described approach applying social sparsity operators to declipping (Siedenburg et al., 2014) to much more general inpainting problems.

4.2 Mathematical Framework

4.2.1 Inverse Problems and Convex Optimization Algorithms

A general convex relaxation of (4.1.1) can be written in Lagrangian form by

$$\min_x \frac{1}{2} \|Fx - y\|_2^2 + \Gamma(Ax), \tag{4.2.1}$$

for suitable linear operators F and A. In fact, (4.1.1) may be relaxed to (4.2.1) when choosing $F = \mathcal{M}\Phi^*$ and replacing the ℓ_0-norm with $\Gamma(Ax) = \lambda \|x\|_1$ for some $\lambda > 0$. In this case the operator A is the identity operator. In general, the convex constraint Γ can be any sparsity enforcing penalty.

Algorithm 5: Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)

Input : $x^0 \in \mathbb{C}^N$, $z^0 = x^0$, $\gamma > 0$, $\tau^0 = 1$

Output: x^{k+1}

for $k = 0, 1, 2, \ldots$ **do**

$z^{k+1} \leftarrow \text{prox}_{\gamma\Gamma}\left(x^k - \gamma F^*\left(Fx^k - y\right)\right)$

$\tau^{k+1} \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + 4\left(\tau^k\right)^2}\right)$

$x^{k+1} \leftarrow z^{k+1} + \frac{\tau^k - 1}{\tau^{k+1}}\left(z^{k+1} - z^k\right)$

end

Before algorithms solving (4.2.1) can be summarized, however, some terminology of convex optimization needs to be introduced. Moreau (1965) introduced so called proximity operators, extending projection operators onto closed convex sets by substituting indicator functions with arbitrary non-smooth functions. Such proximity operators are defined following the notation by Combettes and Pesquet (2011).

Definition 4.2.1 (Proximity Operator)**.** Let $\Gamma : \mathbb{C}^N \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous, convex function. The proximity operator $\text{prox}_\Gamma : \mathbb{C}^N \to \mathbb{C}^N$ is uniquely defined by

$$\text{prox}_\Gamma(x) = \underset{z \in \mathbb{C}^N}{\arg\min}\, \Gamma(z) + \frac{1}{2}\|x - z\|_2^2, \tag{4.2.2}$$

for all $x \in \mathbb{C}^N$.

For $\Gamma(Ax) = \lambda \|x\|_1$ with $\lambda \in \mathbb{R}^+$ as before, the resulting proximity operator is the LASSO soft-thresholding operator $\mathbb{S}^{\text{L}}$ defined by

$$\mathbb{S}^{\text{L}}_\lambda(x_j) = \text{prox}_{\lambda\|\cdot\|_1}(x_j) = x_j\left(1 - \frac{\lambda}{|x_j|}\right)^+, \tag{4.2.3}$$

where $x \in \mathbb{C}^N$ is indexed by $j = 0, 1, \ldots, N-1$ and the notation $(\cdot)^+ = \max(\cdot, 0)$ denotes the pointwise maximum with zero (Tibshirani, 1996). This can be shown by a straight forward calculation using subgradients (Hastie et al., 2015, Ch. 2). Proximity operators can be combined with bounded linear operators as follows. Let A be a bounded linear operator such that $AA^* =$

Algorithm 6: Douglas-Rachford (DR) Algorithm

Input : $x^0 \in \mathbb{C}^N$, $\gamma > 0$ and $\tau^k \in]0, 2[$

Output: x^{k+1}

for $k = 0, 1, 2, \ldots$ **do**
  $z^k \leftarrow \text{prox}_{\gamma h}\left(x^k\right)$
  $x^{k+1} \leftarrow x^k + \tau^k \left(\text{prox}_{\gamma \Gamma}\left(2z^k - x^k\right) - z^k\right)$
end

γI. The corresponding proximity operator $\text{prox}_{\Gamma \circ A}$ is given by Fadili and Starck (2009, Lemma 2) as

$$\text{prox}_{\Gamma \circ A}(x) = x + \frac{1}{\gamma} A^* \left(\text{prox}_{\gamma \Gamma}\left(Ax\right) - Ax\right), \tag{4.2.4}$$

where $\circ$ denotes composition. Whenever A is a frame, then $\Gamma \circ A$ is always lower semi-continuous and convex (Fadili and Starck, 2009).

The audio inpainting problem of the form (4.2.1) can then be solved using proximal splitting algorithms summarized by Combettes and Pesquet (2011). One of such algorithms is the accelerated forward-backward algorithm (Starck et al., 2015, Eq. (7.40)) also known as the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) by Beck and Teboulle (2009) outlined in Algorithm 5. Convergence of the iterates in the accelerated forward-backward algorithm has been recently proved by Chambolle and Dossal (2015). FISTA has been successfully used for declipping (Siedenburg et al., 2014) and denoising (Siedenburg and Dörfler, 2013) audio signals as well as decomposing them into transients and tonals (Bayram and Akyıldız, 2014).

Apart from FISTA, the other proximal splitting algorithm which recently has received considerable attention (see, for example, (Combettes and Pesquet, 2011) and references therein) is the Douglas-Rachford (DR) splitting algorithm first introduced by Combettes and Pesquet (2007). Whereas FISTA requires one of the two summands to be Lipschitz-differentiable, this assumption is relaxed in the Douglas-Rachford algorithm at the cost of an additional proximal step. Although this might increase the computational complexity, Combettes and Pesquet (2011) stated, however, that in general it is not clear a priori which algorithm may be more efficient. Denoting the data fidelity term in (4.2.1) by

$$h(x) = \frac{1}{2} \left\| Fx - y \right\|_2^2, \tag{4.2.5}$$

the DR algorithm is summarized in Algorithm 6 (Combettes and Pesquet, 2011, Alg. 4.2).

Depending on the choice of F and A in (4.2.1) the audio inpainting problem can be divided into synthesis and analysis approach (Elad et al., 2007). First, with $F = \mathcal{M}\Phi^*$ and $A = I$ the *synthesis* approach is given by

$$\min_{x\in\mathbb{C}^N} \frac{1}{2} \left\|\mathcal{M}\Phi^* x - y\right\|_2^2 + \Gamma(x). \tag{4.2.6}$$

The *analysis* approach, on the other hand, defines the inpainting problem by assuming $F = \mathcal{M}$ and $A = \Phi$, i.e.,

$$\min_{x\in\mathbb{C}^L} \frac{1}{2} \|\mathcal{M}x - y\|_2^2 + \Gamma(\Phi x). \tag{4.2.7}$$

Both approaches differ from each other by formulating the reconstruction problem in the signal or the transform domain. Consequently, solutions in the synthesis approach have to be in the range of Φ, whereas the solution can be any vector in $\mathbb{R}^L$ in the analysis case (Starck et al., 2010, Ch. 7.5.1). Synthesis and analysis approach are equivalent whenever the operator Φ is orthonormal (Elad et al., 2007, Thm. 1). There has been an increased interest in the analysis approach recently, reporting results where it outperforms the synthesis approach (Almeida et al., 2016; Cleju et al., 2012; Selesnick and Figueiredo, 2009). However, it is still not clear a priori which of the two approaches gives better results (Elad et al., 2007; Starck et al., 2010).

The synthesis as well as the analysis inpainting formulation in (4.2.6) and (4.2.7) can be solved using the splitting algorithms introduced above, provided the proximity operator of h is given for the DR-algorithm. This operator can be derived as follows. Justified by assuming a noiseless approximation of x at the non-missing coefficients of s, the proximity operator of h prox_h can be defined as the projection onto the set $Fx = y$. The projection of a vector x onto the affine subspace $C = \{x|Fx = y\}$ is given by

$$P_C x = x - F^* \left(FF^*\right)^{-1} (Fx - y), \tag{4.2.8}$$

assuming that FF^* is invertible (Bauschke and Combettes, 2017, Ex. 28.14 (iii)). The resulting proximity operator in the synthesis approach can be obtained by

$$\begin{aligned} \text{prox}_h(x) &= P_C x = x - \Phi\mathcal{M}^* \left(\mathcal{M}\Phi^* x - y\right) && (4.2.9)\\ &= \Phi\Phi^* x - \Phi\mathcal{M}^*\mathcal{M}\Phi^* x + \Phi\mathcal{M}^* y && (4.2.10)\\ &= \Phi \left(\left(I - \mathcal{M}^*\mathcal{M}\right)\Phi^* x + \mathcal{M}^*\mathcal{M}s\right), && (4.2.11) \end{aligned}$$

with I denoting the $L \times L$ identity matrix in (4.2.11). This follows readily since $FF^* = \mathcal{M}\Phi^*\Phi\mathcal{M}^* = I$, where in this case I denotes the $M \times M$ identity matrix. Note that $\mathcal{M}^*\mathcal{M}$ is a $L \times L$ matrix whose elements are the Kronecker delta at the non-missing parts of s and zero otherwise. The proximity operator of h in the analysis case can be derived similarly and is given by $\text{prox}_h(x) = (I - \mathcal{M}^*\mathcal{M})\, x + \mathcal{M}^*\mathcal{M}s$. Corresponding proximity operators for Γ in the analysis case can be constructed using (4.2.4).

4.2.2 Structured Sparsity

A main disadvantage of the soft-thresholding operator in (4.2.3) is, that there is a global threshold for each coefficient value. As mentioned before, audio signals possess a certain structure in the time-frequency domain which is not utilized in this approach. Harmonics, for example, extend over a certain period of time, and hence consecutive coefficients in time direction should have higher energy. Noise, on the other hand, resulting from missing samples in time domain introduces a persistent spread of large energy coefficients over a large frequency range at a given time instance. Thus, the idea pursued here is to enforce persistence in time direction by local thresholding in order to improve reconstruction quality in inpainting problems. The concept of structured or social sparsity, which has been elaborated extensively (Dörfler et al., 2013; Kowalski, 2009; Kowalski and Torrésani, 2009a,b; Kowalski et al., 2013; Siedenburg and Dörfler, 2011; Siedenburg and Dörfler, 2012, 2013; Siedenburg et al., 2014) is useful in this context.

Kowalski (2009) considered inverse problems of the form (4.2.6) with Γ being a mixed norm. Such a norm is defined as follows.

Definition 4.2.2 (Weighted Mixed Norm)**.** Let $x \in \mathbb{C}^N$ be indexed by disjoint groups $g \in [1, 2, 3, \ldots, K]$ and members inside groups $m_g \in [1, 2, 3, \ldots, K_g]$ such that $N = \sum_{g=1}^{K} K_g$. Further, let $w_{g,m_g} \in \mathbb{R}_+^N$ be a positive weights and $p, q \geq 1$. Then, the weighted mixed norm of x is defined as

$$\|x\|_{w;p,q} = \left(\sum_{g=1}^{K} \left(\sum_{m_g=1}^{K_g} w_{g,m_g} \left| x_{g,m_g} \right|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}. \tag{4.2.12}$$

For $p = \infty$ or $q = \infty$ the corresponding sums are replaced by the supremum.

The combination of intra-group-norm ℓ_p and inter-group-norm ℓ_q immediately leads to the behavior for the case $p = 2$ and $q = 1$. Sparsity is enforced between groups, meaning only

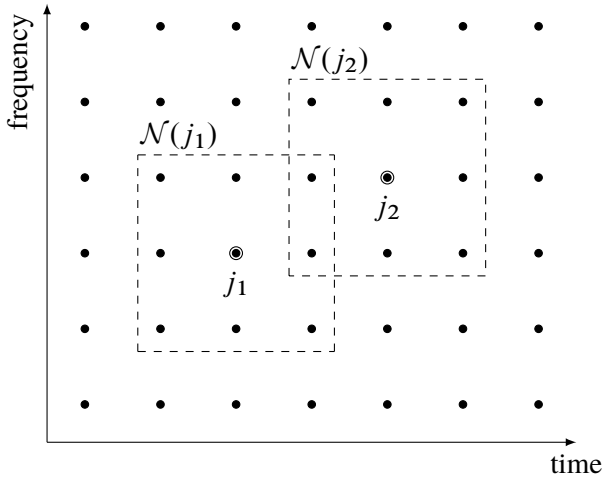


Figure 4.1: Overlapping group structure in time-frequency domain (adapted from Kowalski and Torrésani, 2009b, Fig. 1).

the most energetic groups (in the ℓ_2-sense) remain. The proximity operator of the ℓ_{21}-norm is known as Group LASSO and is given by Kowalski (2009, Prop. 2.2) as

$$\mathbb{S}_{\lambda}^{\mathrm{GL}}(x_g) = \mathrm{prox}_{\lambda\|\cdot\|_{w;2,1}} x_g = x_g \left(1 - \frac{\lambda\sqrt{w_g}}{\|x_g\|_2}\right)^{+}, \tag{4.2.13}$$

where weights w_{g,m_g} are chosen such that $w_{g,m_g} = w_g$ for all g, m_g. It is a generalization of the soft-threshold operator in (4.2.3) and enforces sparsity between disjoint groups of time-frequency coefficients. As there is no dependence on the member index anymore, entire groups are either retained or discarded. This might be too restrictive, for example, whenever groups are large. Additionally, the choice of groups is quite difficult if the structure of time-frequency representations is not known a priori. Therefore, a neighborhood based selection of coefficients is proposed by introducing overlapping groups (Kowalski and Torrésani, 2009b). Each coefficient is evaluated based on the magnitude of its surrounding coefficients: a single large coefficient surrounded by only small coefficients can be discarded, whereas a small coefficient among a neighborhood of large coefficients is retained.

This thresholding operator is defined following the notation of Kowalski et al. (2013). Let x be indexed by $j \in \mathbb{N}$ and denote by $\mathcal{N}(j)$ a set of indices as the weighted neighborhood. For each $k \in \mathcal{N}(j)$ the weights $w_k > 0$ satisfy $\sum_{k\in\mathcal{N}(j)} w_k^2 = 1$. The Windowed Group LASSO

(WGL) is defined by

$$\mathbb{S}_{\lambda}^{\mathrm{WGL}}(x_j) = x_j \left(1 - \frac{\lambda}{\sqrt{\sum_{k \in \mathcal{N}(j)} w_k \left| x_k \right|^2}} \right)^+ , \tag{4.2.14}$$

as a generalization of the Group LASSO: whenever neighborhoods are disjoint the WGL reduces to the Group LASSO. Figure 4.1 shows such overlapping group structures for time-frequency sampling points j_1 and j_2 and corresponding 3×3 neighborhoods $\mathcal{N}(j_1)$ and $\mathcal{N}(j_2)$. As a consequence of the desired overlap of groups, the thresholding operator does not directly correspond to a variational problem anymore (Kowalski and Torrésani, 2009a). Nonetheless, Kowalski et al. (2013) derived a proximity operator for the WGL thresholding operator. Although the convergence of proximal splitting algorithms with such an proximity operator has not been proven (Kowalski et al., 2013, Conjecture 1), several numerical experiments indicate convergence (Kowalski et al., 2013; Siedenburg and Dörfler, 2013; Siedenburg et al., 2014).

Siedenburg et al. (2014) successfully used empirical Wiener thresholding operators inside a proximal splitting algorithm to inpaint missing audio data resulting from clipped samples. The Empirical Wiener (EW) thresholding is closely related to the soft-thresholding operator by squaring the weighted threshold. It is also known as the Non-Negative Garrotte introduced by Antoniadis (2007) and is defined as

$$\mathbb{S}_{\lambda}^{\mathrm{EW}}(x_j) = x_j \left(1 - \frac{\lambda^2}{\left| x_j \right|^2} \right)^+ . \tag{4.2.15}$$

Such a construction exists similarly for the WGL threshold operator. Siedenburg (2012) introduced the Persistent Empirical Wiener (PEW) thresholding operator as

$$\mathbb{S}_{\lambda}^{\mathrm{PEW}}(x_j) = x_j \left(1 - \frac{\lambda^2}{\sum_{k \in \mathcal{N}(j)} w_k \left| x_k \right|^2} \right)^+ . \tag{4.2.16}$$

The main difference of the empirical Wiener approaches compared to LASSO and WGL is the reduced energy loss of remaining coefficients.

In summary, four thresholding operators are introduced: LASSO, WGL, EW and PEW, which might be used inside proximal splitting algorithms. Convergence has been proved only for the LASSO, the remaining generalized thresholding procedures empirically converge in the experiments below, yet a convergence proof is still lacking.

4.2.3 Time-Frequency Representations

Gabor Transform

As stated in (3.2.1), the Gabor transform, denoted in the following by Φ_{GAB}, admits signal expansions by

$$f = \Phi_{\mathrm{GAB}}^{*} \Phi_{\mathrm{GAB}} f = \sum_{k,l \in \mathbb{Z}} \langle f, T_{ak} M_{bl} g \rangle \, T_{ak} M_{bl} g, \tag{4.2.17}$$

for suitable choices of g and time and frequency sampling parameters a and b. In the following it is assumed that the corresponding Gabor frame $\mathcal{G}(g, a, b)$ is a Parseval frame. Therefore, for a given window function g sampling parameters a and b are chosen according to Theorem 3.2.4 and such that $G_0^{g,g} = \frac{1}{b} \sum_{k \in \mathbb{Z}} |g(x - ak)|^2 = 1$ holds for all $x \in \mathbb{R}$ (cf. Proposition 3.3.12 and the following examples). The choice of the window function g also affects the resolution in the time-frequency domain. A spatially concentrated window corresponds to a good time-resolution at a less localized frequency resolution and vice versa for concentrated windows in the frequency domain. Efficient algorithms to compute the Gabor transform based on Gabor frames can be found in the LTFAT toolbox (Průša et al., 2014).

Wavelet Transform

Tight wavelet frames based on nonstationary Gabor frames have been covered in Section 3.4. Signals f can be expanded with such tight frames by

$$f = \Phi_{\mathrm{WAV}}^{*} \Phi_{\mathrm{WAV}} f = \sum_{k,l \in \mathbb{Z}} \langle f, D_{d^l} T_{ak} \psi \rangle \, D_{d^l} T_{ak} \psi = \sum_{k,l \in \mathbb{Z}} \left\langle f, T_{a_l k} h_l \right\rangle T_{a_l k} h_l, \tag{4.2.18}$$

for suitable choices of wavelets ψ. Analysis and synthesis algorithms are summarized in Algorithms 1 and 2, respectively. Construction schemes for compactly supported wavelets are outlined in Algorithm 3. Tight wavelet frames based on non-compactly supported functions as outlined in Algorithm 4 are not used in the context of audio inpainting as computation time of wavelet coefficients increases significantly. Considered wavelet frames are characterized by the parameters f_{min}, *bw* and *bins* as described in Section 3.4.1.

ERBlet Transform

The ERB (Equivalent Rectangular Bandwidth) frequency scale is adapted to human auditory perception and is introduced by Glasberg and Moore (1990). It can be seen as a hybrid model

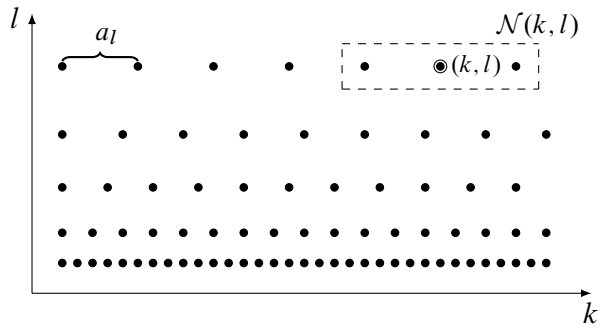


Figure 4.2: Irregular sampling scheme resulting from nonstationary Gabor frames. Neighborhoods $\mathcal{N}$ can only be defined in time (k) direction.

of the Gabor and wavelet discretization scheme: the frequency sampling is approximately linear for low frequencies and logarithmic for large frequencies. According to Glasberg and Moore (1990), the bandwidth *bw* in Hertz of the filter centered at frequency ν satisfies

$$bw(\nu) = 24.7 + \frac{\nu}{9.265}. \tag{4.2.19}$$

Necciari et al. (2013) introduced the ERBlet transform with perfect reconstruction based on the concept of nonstationary Gabor frames. Again, signals can be expanded by

$$f = \Phi^*_{\mathrm{ERB}} \Phi_{\mathrm{ERB}} f = \sum_{k,l \in \mathbb{Z}} \left\langle f, T_{a_l k} g_l \right\rangle T_{a_l k} g_l, \tag{4.2.20}$$

where for each window function g_l centered at corresponding frequency ν_l Equation (4.2.19) holds. A fast implementation of the ERBlet transform can also be found in the LTFAT toolbox. The parametrization of the ERBlet transform consists of the following two parameters: *bins* the number of bins per frequency channel and *qvar* the bandwidth variation factor, which accounts for the resolution of the time-frequency representation. Larger values result in better time-resolution and smaller values in better frequency-resolution. For more detailed information on computational aspects it is referred to (Necciari et al., 2013).

Structured Sparsity for ERBlets and Wavelets

Since both, ERBlet and wavelet transform in (4.2.18) and (4.2.20) are derived from nonstationary Gabor frames, each window g_l or, respectively, h_l has a specific time-shift parameter a_l. An example of such an irregular sampling scheme can be seen in Figure 4.2. This irregularity

allows to define social sparsity with horizontal neighborhoods $\mathcal{N}$, enforcing persistence in time direction only. Instead of defining an absolute number of neighbor coefficients, it is more suitable to define persistence in seconds or milliseconds. Then, for smaller values l the neighborhood consists of more coefficients and decreases with increasing scale (see Fig. 4.2). If for some l the chosen persistence is smaller than the time-shift parameter a_l, the neighborhood consists only of the point itself.

4.3 Experimental Setup

4.3.1 Performance Measures

A widely used measure to evaluate overall quality of reconstructed audio signals is the signal-to-noise ratio (SNR) defined on the missing samples

$$\mathrm{SNR}_{\boldsymbol{M}}(s,\tilde{s}) = 20\log_{10}\frac{\sigma\left(s(I_{\mathcal{M}})\right)}{\sigma\left(s(I_{\mathcal{M}})-\tilde{s}(I_{\mathcal{M}})\right)}, \tag{4.3.1}$$

where s is the original signal, $\tilde{s}$ the reconstructed signal, σ the standard deviation and $I_{\mathcal{M}}$ the set of indices marking the missing samples (Perraudin et al., 2014).

By definition the SNR can only measure energy differences, hence favoring low energy residuals. Such residuals, even though their energy is low, might be audible and affect perceived audio quality (You et al., 2010). A standardized perceptual quality measure for audio data is the widely used PEAQ measure (Perceived Audio Quality). However, it has been shown that it can lead to unreliable results (Creusere et al., 2008; Siedenburg and Dörfler, 2013; You et al., 2010). Therefore, the evaluation of inpainted audio signals will be focused on SNR based methods only. For subjectively evaluating perceived audio quality, however, reconstructed signals can be found in the GitHub repository *https://github.com/flieb/AudioInpainting*.

4.3.2 Test Signals

Siedenburg and Dörfler (2013) provided a set of test signals with different tonal content for denoising audio signals based on the same structured sparsity thresholding operators as introduced in Sec. 4.2.2. What makes the usage of these test signals appealing is the fact, that Siedenburg and Dörfler provided an extensive comparison of different neighborhood choices $\mathcal{N}$ for their signals. Their results can be used similarly in the proposed case of audio inpainting. It can be argued, that inpainting of missing audio data performs best if the neighborhood is adapted to

Table 4.1: Properties of the four audio signals used for inpainting, summarized from Siedenburg and Dörfler (2013, Table 1 and 2).

Index	Description	sf (kHz)	Duration (s)	Orientation of $\mathcal{N}$	Extension of $\mathcal{N}$
1	Strings	44.1	2.3	horizontal	72 ms
2	Piano	44.1	2.3	horizontal	48 ms
3	Percussion	44.1	2.3	rectangular	24 ms and 172 Hz
4	Jazz-Quintet	44.1	2.3	horizontal	48 ms

the signals time-frequency characteristic. According to Siedenburg and Dörfler (2013), a neighborhood $\mathcal{N}$ is characterized by its orientation and extension. Horizontal orientation induces persistence in time direction, vertical in frequency direction and rectangular in both directions. The extension specifies the corresponding size of the neighborhood. The parameters for these signals are summarized in Table 4.1 according to Siedenburg and Dörfler (2013, Table 1 and 2), however, in milliseconds and Hertz rather than number of coefficients.

In case of the percussion signal a rectangular neighborhood choice is not possible for nonstationary Gabor frame approaches due to irregular sampling in frequency direction (cf. Fig. 4.2). Neighborhoods are therefore chosen to be horizontal for wavelet and ERBlet time-frequency representations.

4.3.3 Parameter Settings

Time-Frequency Representation Parameters

In the following numerical evaluations the Gabor transform Φ_{GAB} is used with a Hann window of length 23 ms and time sampling parameter $a = 3.6$ ms. Parameters for the ERBlet transform Φ_{ERB} are $qvar = 0.08$ and $bins = 18$. The window is chosen to be a compactly supported Hann window in frequency domain. The wavelet transform Φ_{WAV} is based on a warped Hann window and parameters $f_{\mathrm{min}} = 100$ Hz, $bw = 3$ Hz and $bins = 120$.

Spatial Structure of Missing Samples

Generally, there are two choices for the mask $\mathcal{M}$. First, the mask $\mathcal{M}$ is chosen such that the retaining samples are uniformly distributed. In the following experiments 80% of the total

number of samples are deleted. On the other hand, an alternative choice of $\mathcal{M}$ results in consecutive missing samples. Obviously, position and length of such gaps play a crucial role for inpainting missing data. A gap in a tonal section of the signal might be easier to reconstruct than during a transient part. Therefore, gaps of a certain length are generated periodically every 300 ms, as proposed by Adler et al. (2012). The gap length is varied between 5 and 30 ms, which corresponds to 221 up to 1323 data points at a sampling rate of 44.1 kHz, in contrast to a maximum proposed gap length of 10 ms by Adler et al. (2012) and Kereliuk and Pasquier (2013).

Algorithm Specific Parameters

The FISTA as well as the Douglas-Rachford algorithm have a maximum iteration number of 200. However, if the relative norm $\|x\|_{\text{rel}}$ between two solutions x^k and x^{k-1} from consecutive iterations drops below a tolerance ε_{rel} the iteration stops before reaching the maximum iteration number. This relative norm is defined by

$$\|x\|_{\text{rel}} = \frac{\left\| x^k - x^{k-1} \right\|_2}{\left\| x^k \right\|_2}. \tag{4.3.2}$$

For all subsequent experiments the tolerance is set to $\varepsilon_{\text{rel}} = 1\text{e}{-4}$. Since all corresponding frames are tight with bound 1, the operator norm of Φ is given by $\gamma = \|\Phi^*\Phi\| = 1$ and for the DR-algorithm $\tau^k = 1$ for all k as proposed by Perraudin et al. (2014).

4.3.4 Comparison Algorithm

The inpainting results from Adler et al. (2012, Fig. 2) clearly show, that Janssen et al.'s (1986) approach is superior to their orthogonal matching pursuit. A recent contribution for inpainting missing audio samples by Oudre (2015) is also based on Janssen's method and demonstrates that it is still considered a state-of-the-art algorithm for interpolating audio data. A MATLAB implementation is provided by Adler et al. (2012), and is used as a comparison approach in the following numerical evaluations with parameters as suggested by Adler et al. (2012) and Oudre (2015). Unfortunately, whenever the binary mask $\mathcal{M}$ is chosen such that the missing samples are randomly distributed, Janssen's approach can not be applied. With 80% missing samples, the algorithm requires about 60 GB of memory.

Table 4.2: SNR (dB) for the two proposed algorithms and analysis and synthesis approach. Test signals are reconstructed from 20% remaining samples based on the soft-thresholding operator $\mathbb{S}^{\mathrm{L}}$ and the proposed time-frequency representations.

		FISTA			DR-Algorithm		
Test Signal		Φ_{GAB}	Φ_{WAV}	Φ_{ERB}	Φ_{GAB}	Φ_{WAV}	Φ_{ERB}
1	Synthesis	15.5	25.5	25.9	18.7	26.0	26.4
	Analysis	18.6	25.2	25.6	16.9	25.9	26.3
2	Synthesis	16.8	25.1	25.2	20.1	25.9	25.9
	Analysis	19.7	25.1	25.2	18.3	25.7	25.6
3	Synthesis	17.4	18.9	19.1	18.6	19.2	19.3
	Analysis	18.5	19.2	19.2	17.9	19.2	19.3
4	Synthesis	13.6	19.3	20.1	16.2	19.8	20.4
	Analysis	16.1	19.7	20.4	15.1	19.7	20.4

4.4 Results

The results presented in this section are based on a parametrization of the regularization parameter λ. The values for λ are taken from the interval $[1\mathrm{e}{-5}, 10]$, for each thresholding operator. For small values, the thresholding operator removes only very few time-frequency coefficients in each iteration which in return leads to small SNR values. Large λ values, on the other hand, ensures that all coefficients are removed and, hence, again results in a small SNR. For values in between, the signal-to-noise ratio attains a maximum for a specific λ value. This parameter sweep of λ is repeated for each time-frequency representation and each thresholding operator and the maximum SNR value is selected in the following evaluations.

4.4.1 Analysis versus Synthesis and FISTA versus DR-Algorithm

As mentioned earlier, the synthesis approach in (4.2.6) is to be compared to the analysis one (4.2.7) and, moreover, FISTA to the DR-algorithm. SNR-values of the reconstructed audio signals are shown in Table 4.2 based on depleted audio signals with only 20% remaining samples. The time-frequency representations described in Sec. 4.2.3 and the LASSO thresholding

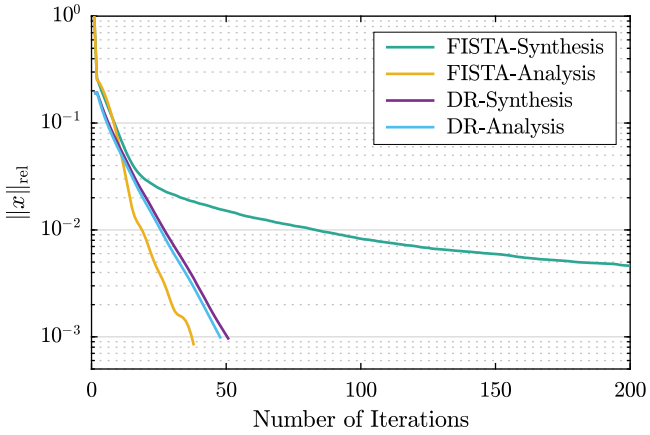


Figure 4.3: Convergence rate for the four approaches based on a Gabor transform with LASSO thresholding operator and regularization parameter $\lambda = 0.05$.

operator are used. Only the splitting algorithms based on the Gabor transform show significant differences. FISTA favors the analysis approach, whereas, the DR-algorithm yields better results for the synthesis method.

In terms of convergence rates, it is known from Chambolle and Dossal (2015) that FISTA converges with $O\left(1/k^2\right)$ and the Douglas-Rachford algorithm converges generally with $O\left(1/k\right)$ (He and Yuan, 2012). Therefore, it could be assumed, that FISTA would be the better choice. Numerical results for synthesis and analysis inpainting cases, however, show that the convergence rate for FISTA with the analysis approach and the DR-algorithm are quite similar. FISTA based on the synthesis approach shows the slowest convergence. This is visualized in Fig. 4.3 for test signal 4 (jazz-quintet) and the LASSO approach based on the Gabor transform but similar results can be observed for the other test signals as well.

Since the DR-algorithm based on the synthesis model results in slightly better SNR-values compared to FISTA and convergence is only slightly slower, it is chosen for all subsequent numerical evaluations.

4.4.2 Results for Randomly Distributed Missing Samples

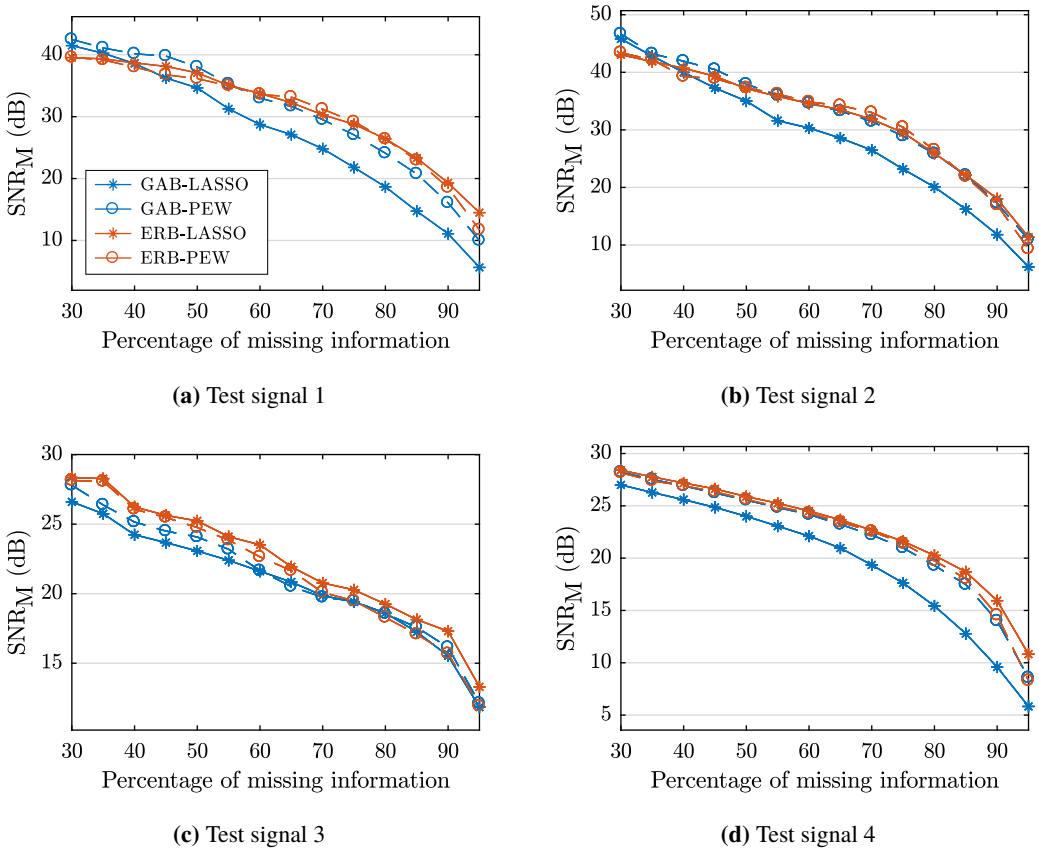
The mask $\mathcal{M}$ is fixed to retain 20% of the audio samples. For each regularization parameter λ the DR-algorithm approximates a corresponding solution for the inpainting problem. Of all solutions, the ones with the largest SNR are summarized in Table 4.3.

SNR values associated with the Gabor transform confirm the assumption made in the intro-

Table 4.3: Signal-to-noise ratios in dB for the four thresholding operators. The test signals are reconstructed from 20% remaining samples based on the proposed time-frequency (TF) representations.

		Test Signal			
TF Repr.	Threshold Op.	1	2	3	4
Φ_{GAB}	$\mathbb{S}^{\text{L}}$	18.7	20.1	18.6	16.2
	$\mathbb{S}^{\text{EW}}$	22.6	24.4	18.1	18.5
	$\mathbb{S}^{\text{WGL}}$	20.6	21.5	18.6	17.6
	$\mathbb{S}^{\text{PEW}}$	24.1	25.9	18.6	19.6
Φ_{WAV}	$\mathbb{S}^{\text{L}}$	26.0	25.9	19.2	19.8
	$\mathbb{S}^{\text{EW}}$	25.2	25.6	17.8	18.5
	$\mathbb{S}^{\text{WGL}}$	25.9	26.0	19.4	19.7
	$\mathbb{S}^{\text{PEW}}$	25.4	26.9	18.5	19.5
Φ_{ERB}	$\mathbb{S}^{\text{L}}$	26.4	25.9	19.3	20.4
	$\mathbb{S}^{\text{EW}}$	25.3	25.5	17.5	18.8
	$\mathbb{S}^{\text{WGL}}$	26.4	26.0	19.4	20.4
	$\mathbb{S}^{\text{PEW}}$	26.4	26.5	18.3	19.8

duction: the PEW operator improves inpainting of audio data significantly in most cases compared to LASSO. This is in accordance with the results from Siedenburg et al. (2014, Fig. 4) as well as the results from Siedenburg and Dörfler (2013, Fig. 5) where the PEW operator performed best when declipping and denoising audio data. Wavelets and ERBlets, on the other hand, improve the Gabor-PEW based signal-to-noise ratios even further. Although SNR values for the jazz quintet signal (test signal 4) are close for the GAB-PEW, ERB-LASSO and ERB-PEW approaches, the perceptual quality of the reconstructions differs significantly. The neighborhood approach eliminates many audible artifacts when listening to the reconstructions of GAB-LASSO and GAB-PEW approaches. The ERBlet approach, on the other hand, further reduces the musical noise introduced by Gabor representations. This can be verified by listening to the reconstructions in the GitHub repository *https://github.com/flieb/AudioInpainting/tree/master/Results*.



(a) Test signal 1

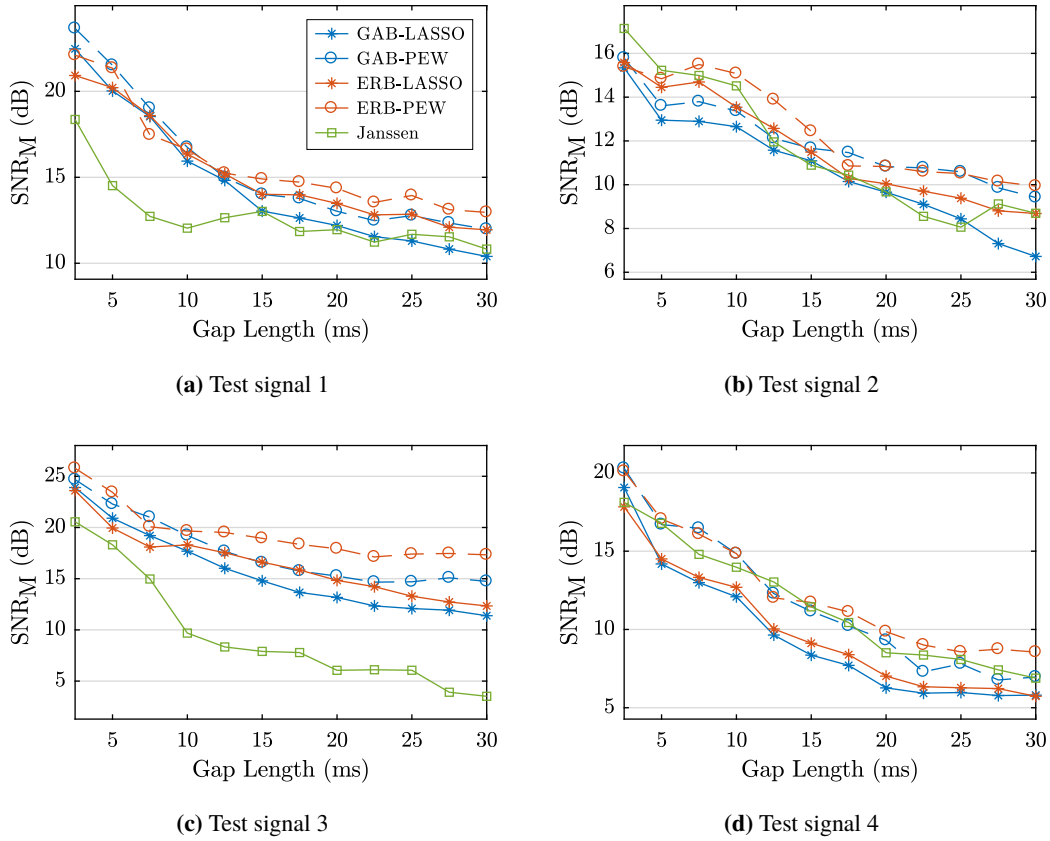
(b) Test signal 2

(c) Test signal 3

(d) Test signal 4

Figure 4.4: SNR with increasing number of missing samples for all four test signals.

In Figure 4.4 it can be observed how performance evolves in the range of 30 up to 95% of missing samples for all four test signals. Only Gabor and ERBlet time-frequency representations with LASSO and PEW thresholding operators are illustrated such that the figures remain discernible. Except for low percentages of missing audio samples in the first two test signals, the ERBlet based time-frequency representation gives an overall better signal-to-noise ratio. Whereas there are significant differences between the LASSO and PEW thresholding for Gabor time-frequency representations, the differences for ERBlets are only marginal. Figure 4.4 hence reflects the results observed in Table 4.3 for large percentages of missing samples. For smaller percentages missing samples are much more scattered throughout the data, since they are uniformly distributed. Consequently, perceptual quality between different thresholding operators and time-frequency representations does not differ much.



(a) Test signal 1

(b) Test signal 2

(c) Test signal 3

(d) Test signal 4

Figure 4.5: Signal-to-noise ratios as a function of gap size for all four test signals.

4.4.3 Results for Consecutive Missing Samples

Here, $\mathcal{M}$ is chosen such that a gap with specified length is repeated every 300 ms. In favor of illustrational convenience, the succeeding results also feature only Gabor and ERBlet transforms with thresholding operators LASSO and PEW. In order to avoid that the DR-algorithm amplifies high-frequency noise at the edge of gaps, instead of reducing it, minor parameter settings need to be adapted. Firstly, increasing the width of the Gabor window improves the inpainting performance, since wider windows fill gaps with information from surrounding samples. A Hann window of length 150 ms leads to reasonable results and is therefore used for subsequent evaluations. Additionally, for test signal 3 the *qvar* parameter of the ERBlet transform is changed to $qvar = 0.2$ with $bins = 8$, increasing the window width in order to obtain reasonable results. The remaining transform parameters remain as before. Since the structure of gaps generates

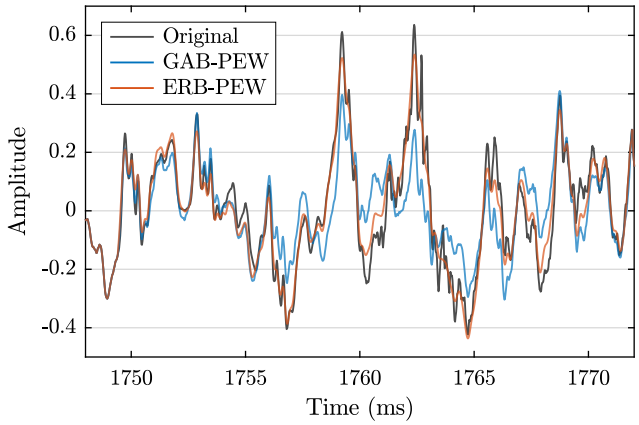


Figure 4.6: Inpainting solutions for the jazz quintet signal with gap length 22.5 ms.

noise with persistence in frequency direction, it is disadvantageous to consider rectangular neighborhoods since this would enforce noise coefficients. Instead, only horizontal neighborhoods of length 48 ms are used in this approach.

Figure 4.5 shows maximum signal-to-noise ratios (where the maximum is taken over all λ values) of the four reconstructed test signals as well as Janssen's algorithm with increasing gap length. In nearly all cases, Janssen's approach as well as the Gabor-LASSO results in the lowest SNR compared to the remaining methods. For large gap sizes the PEW approaches show the best performance for ERBlets as compared to the Gabor transform.

Figure 4.6 illustrates the inpainted solutions for the jazz quintet signal (test signal 4) with gap size 22.5 ms based on PEW threshold operators for Gabor and ERBlet transform. It clearly shows that the ERB-PEW approach provides more accurate results, resembling features of the original signal more closely than the GAB-PEW approach.

4.5 Conclusion

Several sparsity constraints for the problem in (4.2.6) are considered for inpainting missing audio samples based on a set of time-frequency representations. These constraints promote sparsity similar to the well known soft-thresholding by either reducing the loss of energy for remaining coefficients (EW), introducing neighborhoods (WGL) or the combination of both (PEW).

The results for randomly distributed missing samples clearly favor ERBlet and wavelet time-frequency representations. SNR values indicate, that the reconstruction is quite good, despite

the loss of 80% of the original data. This might improve reconstruction of packets lost during audio streaming. If the audio data to be streamed is interleaved in different packets as suggested by Ciaramella et al. (2016, Fig. 1 and Fig. 2), missing samples are randomly distributed. Hence, the results from Section 4.4.2 can be directly applied. Since timing is crucial in streaming applications, a more detailed convergence analysis as well as fast implementations are required to ensure a quick reconstruction of missing audio samples. Furthermore, acceleration methods as proposed by Rajmic et al. (2015) can further speed up the reconstruction process.

The performance for consecutive missing samples is also dominated by usage of the ERBlet transform. The PEW threshold operator has shown to approximate the original signal better than the Gabor transform based PEW, especially for large gaps. For Gabor based time-frequency representations, it obviously is advantageous to use window functions with a longer duration than 20 ms as proposed in (Kowalski and Torrésani, 2009a; Siedenburg, 2012; Siedenburg and Dörfler, 2013; Siedenburg et al., 2014). They extent signal information from the borders of a gap into the gap itself, improving inpainting substantially. In case of the proposed non-stationary time-frequency representations the window length is not fixed, instead it adaptively scales with respect to frequency. Further, Janssen's algorithm which has shown good performance in (Adler et al., 2012) for gap sizes up to 10 ms, mostly can not keep up with the performance of PEW threshold operators. Additionally, for large numbers of missing samples Janssen's algorithm is computationally not feasible.

Clearly, the SNR results favor ERB- or wavelet approaches, yet the perceptual quality of the reconstructions is most important. Example files for the jazz quintet signal can be found in the GitHub repository. The listener will notice that both LASSO approaches (Gabor as well as ERBlet transform) feature clearly audible artifacts, whereas PEW thresholding operators perform much better in this respect. In summary, based on subjective observations the ERB-PEW approach leads to a better perceptual quality with less prominent artifacts compared to the Gabor-PEW approach.

Comparing analysis and synthesis approaches, it can moreover be concluded that in combination with a Gabor transform, the analysis approach yields an improvement for iterative thresholding type algorithms like FISTA. Interestingly, the synthesis approach performs better while used in combination with the Douglas-Rachford algorithm. Whereas this leads to significantly different signal-to-noise ratios for the Gabor transform, differences are only minor for ERB- and wavelet like time-frequency representations. The denoising results from Siedenburg and Dörfler (2013) as well as the declipping results from Siedenburg et al. (2014) might be further improved by using analysis models, instead of the proposed synthesis ones. In addition, the

declipping results in (Siedenburg et al., 2014) also benefit from using longer window functions for the Gabor transform.

The neighborhood choice for numerical evaluations has been based on results from Siedenburg and Dörfler (2013) where the underlying time-frequency representation is the Gabor transform. As can be seen for test signal 3 in Table 4.3 this might not be the best choice for inpainting problems. Hence, a more detailed analysis of neighborhood structures is required. Neighborhoods which are better adapted to the problem of inpainting as well as better adapted to the time-frequency representation might further improve reconstruction results. Especially for consecutive missing samples, neighborhoods can easily be constructed such that they match the surrounding time-frequency coefficients more adaptively.

So far only audio signals have been used for inpainting. Whether similar conclusions can be transferred to speech signals has not been investigated and leaves room for future contributions.

Chapter 5

Spike Detection Algorithms for Extracellular Recorded Data

This chapter is an amended version of the results published by Lieb et al. (2017), with Section 5.5 focusing on the novel spike detection algorithm introduced by Mayer et al. (2018).

5.1 Introduction

A detailed analysis of intercellular communication, either spontaneous or stimulated, is facilitated by cultivating neuronal cells on microelectrode arrays (MEAs) and recording the electrophysiological activity of resulting neuronal networks. In recent years this in vitro method has emerged as an effective technique for neurotoxicological screenings, see for example (Hondebrink et al., 2016; Johnstone et al., 2010; McConnell et al., 2012; Tukker et al., 2016). Extracellular recordings are based on rapid changes of the cell membrane potential caused by biological ion displacement. The amplitude of the resulting voltage difference strongly depends on the coupling quality and the distance between cell and electrode (Daus et al., 2012; Obeid and Wolf, 2004). The larger the distance between firing neuron and recording electrode, the smaller the recorded amplitude. This might lead to recordings where spike amplitudes are not necessarily larger than noise. In general, this noise is characterized by colored 1/f noise as a superposition of electrochemically induced noise, thermal noise, noise induced by stochastic fluctuation of electrolytic and conducting resistors as well as extrinsic noise like 50 Hz net frequency or biological background noise (Obien et al., 2015).

In the past, the majority of neurons cultivated on MEAs are based on two-dimensional neural networks (Frega et al., 2014). Only recently, Smith et al. (2015) succeeded in cultivating differentiated human embryonic stem cells (hESCs) on MEAs to form functional three-dimensional networks allowing more detailed electrophysiological investigations. However, 3D cell models result in smaller spike amplitudes and an even worse signal-to-noise ratio. Unfortunately, this turns detection of spikes with small amplitudes in large background noise into a nontrivial task. Separating spikes reliably from noise, however, is the basis for all subsequent processing steps like clustering and analyzing spike trains (Rossant et al., 2016), making highly accurate spike detection algorithms indispensable.

Several algorithms for spike detection have already been published (Azami and Escudero, 2016; Azami and Sanei, 2014; Azami et al., 2015; Choi and Kim, 2002; Choi et al., 2006; Kim and McNames, 2007; Liu et al., 2012; Maccione et al., 2009; Nabar and Rajgopal, 2009; Natora et al., 2010; Nenadic and Burdick, 2005; Obeid and Wolf, 2004; Quiroga et al., 2004; Salmasi et al., 2016; Shahid et al., 2010; Shalchyan et al., 2012). Based on their methodological concept most of them can be divided into three categories, as suggested in (Obeid and Wolf, 2004):

A. Simple Thresholding
   Spike detection algorithms of the first category rely on the most prominent feature of the spike shape, its amplitude. It is assumed that the amplitude of a spike is larger or has a larger peak-to-peak value than noise and can hence be detected by a simple threshold. Related algorithms are published, e.g., by Maccione et al. (2009); Obeid and Wolf (2004) or Quiroga et al. (2004). Clearly, whenever the amplitude of the temporal noise is larger than the spike's amplitude this approach might fail and is easily susceptible to errors.

B. Template Based Correlation
   The second category is based on actual spike shapes, whereby signal segments are correlated with template waveforms. Whenever the resulting similarity is larger than a predefined threshold the corresponding segment is considered a spike. Obviously, the main challenge is to find reasonable templates in the first place. This either requires knowledge of the spike's waveform or generic approaches to find suitable templates. Such algorithms are described by Kim and McNames (2007); Liu et al. (2012); Natora et al. (2010) and Shahid et al. (2010).

C. Transient Energy
   Spike detection algorithms of the third category exploit the transient behavior of spikes, i.e., abrupt amplitude changes. Such transients introduce a frequency pattern with distinct

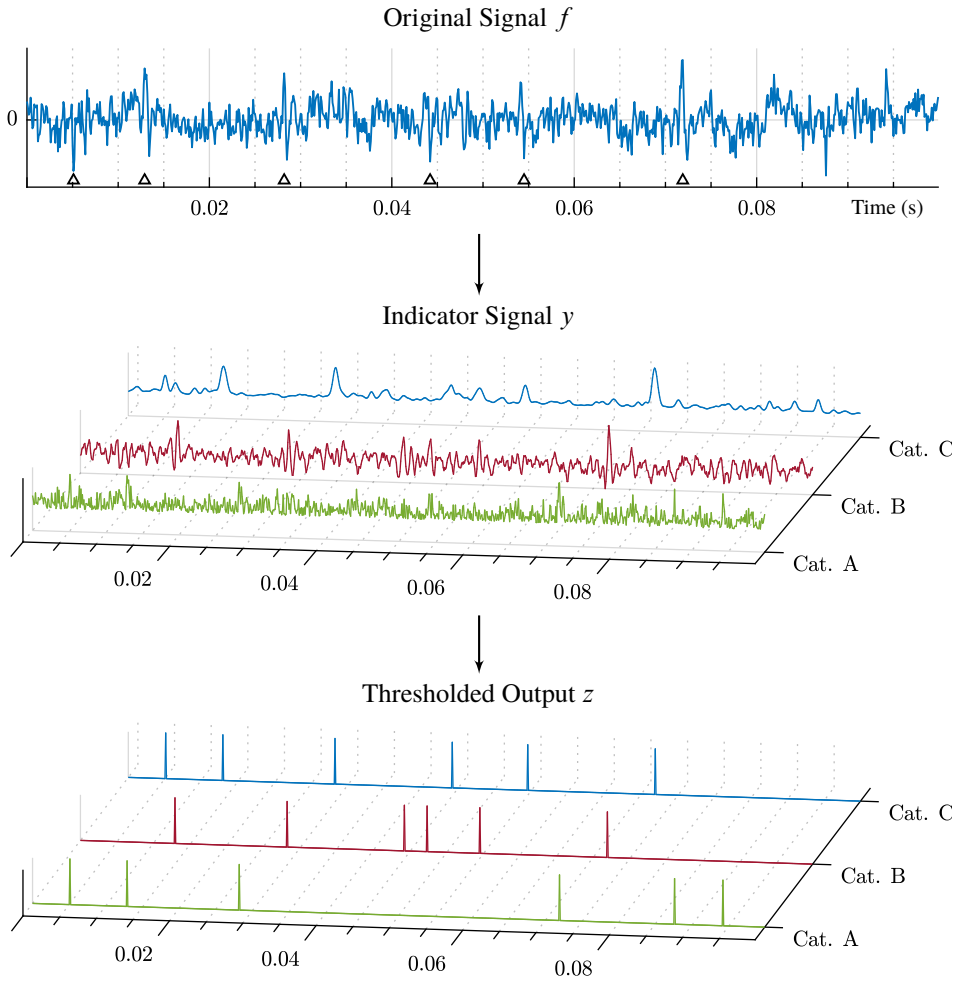


Figure 5.1: Basic scheme of spike detection algorithms based on algorithms from categories A, B and C. The simulated test signal has noise level $\kappa = 5$, see (5.3.2).

characteristics, different from noise. For example, the methods based on nonlinear Teager energy operators introduced by Azami and Sanei (2014); Choi and Kim (2002); Choi et al. (2006) and Nabar and Rajgopal (2009) fall into this category, as well as the wavelet based approaches from Nenadic and Burdick (2005) and Shalchyan et al. (2012).

More recent algorithms which are entropy based (Azami and Escudero, 2016) or based on fractal dimensions (Azami et al., 2015; Salmasi et al., 2016) cannot be classified to any of these categories. Despite this diversity, almost all spike detection algorithms mentioned above follow

a two-step-procedure. First, from the raw input signal f an indicator signal y is computed. This computation is specific for the respective algorithm and should not be sensitive to noise. Subsequently, y is thresholded in order to localize spike occurrences and associate corresponding time stamps. For example, putting $y = |f|$ and thresholding this indicator signal results in one of the simplest representative of a spike detection algorithm belonging to category A.

Figure 5.1 on the previous page illustrates this two-step-procedure with one representative algorithm per Category A, B or C respectively. The input signal f is constructed from known spikes, whose locations are marked in Fig. 5.1 with black triangles. This signal is superimposed with noise reflecting the characteristics of MEA recordings. On this basis the indicator signal y is computed based on an representative algorithm from each category. Category A is represented by the absolute value procedure (ABS) described above. Categories B and C are represented by the HBBSD-algorithm (Natora et al., 2010) and the SWTTEO algorithm proposed later on. From the resulting spike locations in the output signal z it can be deduced that there are differences in the precision of the algorithms and apparently, the performance of the algorithms is based on the accuracy of the indicator signal y.

Despite the diversity of already proposed spike detection algorithms, the performance whenever signal-to-noise ratios are low still leaves room for further improvement. In the following, two new spike detection algorithms are introduced. The first one is based on the characteristic time-frequency pattern of spikes, the other relies on energy distributions of stationary wavelet coefficients. The performance of both algorithms is first compared with current state-of-the-art algorithms based on two different simulated data sets before a novel spike detection approach for real MEA data is introduced.

5.2 New Spike Detection Algorithms

5.2.1 Preliminaries

Modified Gabor Transform

Fast algorithms for computing the Gabor transform are publicly available in the LTFAT toolbox (Průša et al., 2014). However, these algorithms lack fast implementations whenever only a small subset of the full frequency range is required for further analysis. Instead of computing a full Gabor transform with a subsequent selection of desired frequency bins, the following approach introduces a modified Gabor transform which only computes time-frequency coefficients in a

Algorithm 7: Discrete Gabor Transform with Specified Frequency Range (DGTSF)

Input : $f \in \mathbb{R}^N$ - input signal
g - window function in frequency domain
$[f_{\min}, f_{\max}]$ - frequency interval
M - number of frequency bins
a - time shift parameter

Output: $c \in \mathbb{C}^{N/a \times M}$ - time-frequency coefficients

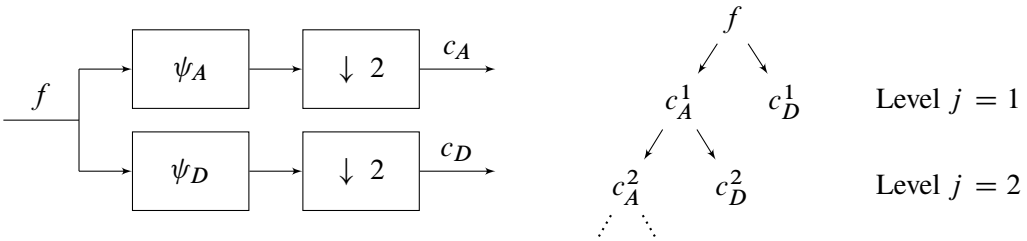
$F \leftarrow \mathcal{F}_N f$
$g \leftarrow T_{f_{\min}} g$
$b \leftarrow (f_{\max} - f_{\min})/(M-1)$
for $l = 0, 1, 2, \ldots, M-1$ **do**
  $c_l \leftarrow T_{bl} g \cdot F$
end
$c_{k,l} \leftarrow e^{-2\pi i nakbl/N} \mathcal{F}^{-1}_{\frac{N}{a}} \mathcal{Q}_{\frac{N}{a}} c_l$ // see Prop. 3.3.14 for the subsampled DFT

specified frequency range.

In a finite dimensional setting, the Gabor transform as defined in a continuous setting in (4.2.17), decomposes every signal $f \in \mathbb{R}^N$ into its time-frequency coefficients $c[k,l]$ by

$$c[k,l] = \sum_{n=0}^{N-1} f[n]\overline{g}[n-ak]e^{-2\pi i nbl/N}, \tag{5.2.1}$$

for some window function g and $k \in [0, 1, \ldots, K-1]$, $l \in [0, 1, \ldots, L-1]$. The frequency and time shift parameters a and b satisfy $aK = bL = N$. With some modifications it is possible to specify only a certain frequency range. In general, the frequency index l takes values from the integer interval $[0, L-1]$ corresponding to L frequency bins, ranging from zero to twice the Nyquist frequency (Průša et al., 2014). Considering for l a suitable subset of $[0, L-1]$ unwanted frequencies can be omitted. For example, if $l \in [\lfloor (L-1)/8 \rfloor, \lfloor (L-1)/4 \rfloor]$, only frequencies from a quarter of the Nyquist frequency up to half the Nyquist frequency are regarded. Hence, time-frequency coefficients can be computed for specific frequency ranges and undesirable frequency contributions, e.g., in lower frequency ranges resulting from 1/f noise, are disregarded.



(a) Single DWT decomposition step

(b) Recursive tree structure for a 2-level DWT

Figure 5.2: Block diagram and tree structure of the DWT.

As indicated by Søndergaard (2012, Eq. (4.2)), Equation (5.2.1) can be rewritten as

$$c[k,l] = \langle f, M_{bl} T_{ak} g \rangle \tag{5.2.2}$$

$$= \langle \mathcal{F}_N f, T_{bl} M_{-ak} \mathcal{F}_N g \rangle \tag{5.2.3}$$

$$= e^{-2\pi i nakbl/N} \mathcal{F}_N^{-1} \left(\mathcal{F}_N f \cdot T_{bl} \overline{\mathcal{F}_N g} \right)[ak]. \tag{5.2.4}$$

The phase factor can be ignored in cases where only the magnitude of time-frequency coefficients is relevant. By shifting the analyzing window g only to specified frequencies in the interval $[f_{\min}, f_{\max}]$ the restriction of the frequency index l to a subset of $[0, L-1]$ in (5.2.1) can be efficiently implemented. The algorithm for computing the *discrete Gabor transform with specified frequency range* (DGTSF) is summarized in Algorithm 7. Note that the subsampled N-point inverse DFT in (5.2.4) can be efficiently computed by utilizing the result from Proposition 3.3.14.

Discrete and Stationary Wavelet Transform

Introduced by Mallat (1989), the fast algorithm for the *discrete wavelet transform* (DWT) is nowadays a common tool in signal processing. The DWT basically consists of high- and low-pass filtering with filters ψ_D and ψ_A and a subsequent dyadic down sampling resulting in detail and approximation coefficients, c_D and c_A, respectively. A schematic block diagram of a single level wavelet decomposition is depicted in Figure 5.2a. This procedure may be recursively iterated j times, by replacing the input f with the approximation coefficients c_A of the previous decomposition. The resulting decomposition is called a j-level wavelet decomposition and is visualized for two levels in Figure 5.2b.

One main disadvantage of the discrete wavelet transform is its lack of time-invariance. This means that a shifted version of any input signal does not necessarily result in a shifted version of wavelet coefficients, due to downsampling after applying the wavelet filter. The *stationary wavelet transform* (SWT) introduced by Nason and Silverman (1995) consists only of filtering with no subsequent downsampling. Instead, the filter coefficients are dyadically upsampled by padding them with zeroes at even- or odd-indexed positions. In contrast to the DWT, where the number of DWT coefficients is equal to the number of data points of f, this results in a redundant signal representation.

Teager Energy Operator

The fundamentals of the Teager energy operator (TEO) were originally derived by Kaiser (1993). In the continuous case the operator reads

$$\mathcal{B}_{\text{TEO}}(f(t)) = \left(\frac{d}{dt}f(t)\right)^2 - f(t)\frac{d^2}{dt^2}f(t). \tag{5.2.5}$$

The term 'energy operator' is motivated by an analogy to a classical harmonic oscillator of mass m. In this case $f(t) = A\cos(\omega t + \phi)$ and the total energy reads $E = \frac{1}{2}m\omega^2 A^2$, (Tipler and Mosca, 2015, p. 421). It is then easy to verify that $\mathcal{B}_{\text{TEO}}(f(t)) = \omega^2 A^2$. Thus, in fact $\mathcal{B}_{\text{TEO}}(f(t))$ is proportional to the energy associated with oscillatory motion.

For discrete signals the Teager energy operator is defined as

$$\mathcal{B}_{\text{TEO}}(f[n]) = f^2[n] - f[n-1]f[n+1]. \tag{5.2.6}$$

Again, a similar calculation shows that $\mathcal{B}_{\text{TEO}}(A\cos[\Omega n + \phi]) = A^2\sin^2(\Omega)$. Here, Ω denotes the discrete frequency $\Omega = 2\pi\omega/sf$, with sampling frequency sf. For small Ω it holds that $\sin(\Omega) \approx \Omega$ and the Teager energy operator is approximately proportional to the systems energy as in the continuous case.

5.2.2 Time-Frequency based Convolution Spike Detection Algorithm (TIFCO)

The time-frequency based convolution spike detection algorithm, denoted TIFCO in the following, assumes that action potentials exhibit a certain time-frequency behavior which can be detected even in noisy measurements. For example, they exhibit prominent features in the range

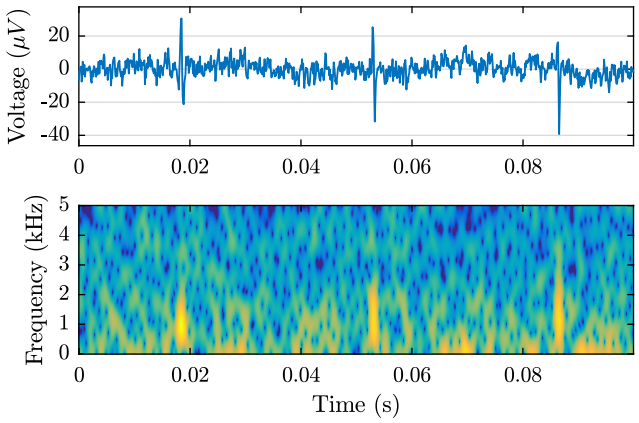


Figure 5.3: Characteristic time-frequency pattern of spikes embedded in colored noise. The colorscale is logarithmic with yellow indicating large and blue small coefficients.

of 500 Hz up to 3500 Hz, as visualized in Figure 5.3. These patterns of large coefficients spread over several frequency bins at the occurrence of a spike and distinguish spikes from 1/f noise. Therefore, the DGTSF with an appropriate choice of the frequency range can be used to compute corresponding time-frequency coefficients $c[k,l]$. Subsequently, the time-frequency coefficients are filtered in order to emphasize spike specific patterns in the time-frequency plane. Generally, such a filter can be defined as follows. Letting $\eta \in \mathbb{R}^{K' \times L'}$ be the kernel of a two-dimensional convolution, the filtered time-frequency coefficients $v[k,l]$ can be defined by

$$v[k,l] = \left(|c|^2 * \eta\right)[k,l] \tag{5.2.7}$$

$$= \sum_{k'} \sum_{l'} \left|c[k-k', l-l']\right|^2 \eta\left[k', l'\right], \tag{5.2.8}$$

where k', l' are the corresponding indices for the filter kernel η. The choice of the filter should be motivated by enforcing spike specific structures. Gaussian filters or more advanced filters from image processing, for example, are conceivable. Alternatively, a moving average filter of size $K' \times L'$ introduces persistence in the desired frequency and time range, while simultaneously lessens the influence of large isolated noise coefficients.

Finally, a decision variable y is computed by summing over all frequency bins, i.e.,

$$y[k] = \sum_{l=0}^{L-1} v[k,l]. \tag{5.2.9}$$

Algorithm 8: Time-Frequency based Convolution Spike Detection Algorithm (TIFCO)

input : $f \in \mathbb{R}^N$ - input signal
g - window function for DGTSF
$[f_{\min}, f_{\max}]$ - frequency interval
M - number of frequency bins
a - time shift parameter
$\eta \in \mathbb{R}^{K' \times L'}$ - 2D filter kernel

output: $y \in \mathbb{R}^{N/a}$ - indicator signal

1 $c \leftarrow \text{DGTSF}\,(f, g, [f_{\min}, f_{\max}], M, a)$
2 $c \leftarrow |c|^2 * \eta$
3 $y \leftarrow \sum_l c[k, l]$

Thus, whenever a spike is present at a time instance k, the decision variable y yields a larger value compared to a time instance where only noise is present. This decision variable y can then be thresholded in order to extract exact spike locations as indicated in Fig. 5.1. The TIFCO algorithm is summarized in Algorithm 8.

5.2.3 Stationary Wavelet Transform based TEO Spike Detection Algorithm (SWTTEO)

This newly proposed algorithm is based on a further development of the wavelet TEO algorithm (WTEO) from Nabar and Rajgopal (2009). Their spike detection algorithm is based on a low-pass filter using the first and second level approximation coefficients of the DWT. The Teager energy operator is then applied to each sub-band with a subsequent thresholding, followed by expanding the corresponding decimated coefficients to the length of the original data.

The SWT based TEO spike detection algorithm (SWTTEO) adopts the following changes to improve the WTEO algorithm. The discrete wavelet transform used by Nabar and Rajgopal (2009) is substituted by the stationary wavelet transform. Hence, no down sampling of wavelet approximation coefficients is applied, superseding the need to expand decimated coefficients. For each low-pass sub-band the TEO is computed and subsequently convolved with a smoothing window w. The indicator signal y is then derived by adding these smoothed TEO sub-bands. Since it is assumed that each sub-band contains spike information the summation of smoothed

Algorithm 9: Stationary Wavelet based TEO Spike Detection Algorithm (SWTTEO)

input : $f \in \mathbb{R}^N$ - input signal
ψ - discrete wavelet
j - decomposition level of SWT
w - smoothing window

output: $y \in \mathbb{R}^N$ - indicator signal

1 $c_j \leftarrow j$-level SWT of f with wavelet ψ
2 $y \leftarrow \sum_j \left(\mathcal{B}_{\text{TEO}}(c_j) * w\right)$

TEO outputs is preferable to the maximum filter suggested by Nabar and Rajgopal (2009). Finally, the thresholding is applied to the indicator signal y instead of thresholding individual sub-bands proposed by Nabar and Rajgopal. The complete algorithm for the SWTTEO spike detection is outlined in Algorithm 9.

The number of levels for the wavelet decomposition depends on the frequency content of the spike shapes itself. Figure 5.3, for example, shows spike specific frequencies in the range from 500 up to approximately 3500 Hz for a signal sampled at 10 kHz. Hence a decomposition level of 2 or 3 would be reasonable for this example. The first level approximation coefficients would then cover frequencies from zero up to half the sampling frequency, the second level would contain frequencies from zero to a quarter of the sampling frequency and so on. Choosing a decomposition level larger than 3 would only cover lower frequency bands, where no useful spike information is present. This coincides with the results from Nabar and Rajgopal where the best wavelet decomposition level is also found to be 2 for signals sampled at 10 kHz. However, for some signals it might be useful to consider only a subset of the decomposition level. For example, if noise is predominately associated with high frequencies, the performance of the algorithm may improve by ignoring the first or even the first two levels of approximation coefficients. This restrains the influence of high frequency induced errors.

Furthermore, Choi et al. (2006) have shown that especially when signal-to-noise ratios are low, peaks resulting from noise can be efficiently removed by smoothing the TEO output with a Hamming window. Typically, this window is chosen such that its length corresponds to the range of a spike duration. As shown later, smoothing the TEO output also reduces the chance of falsely detected spikes if the input signal consists of stationary wavelet coefficients.

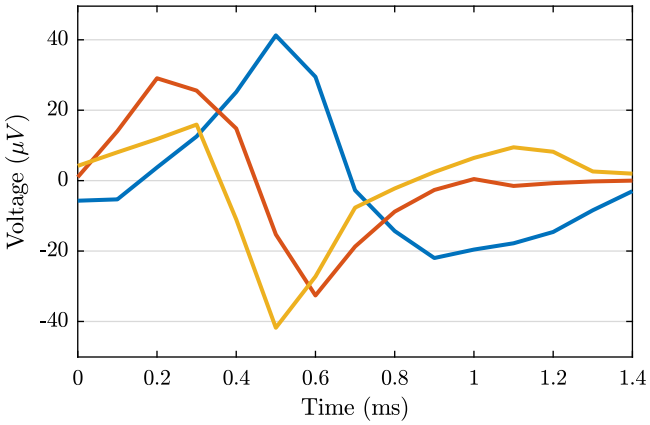


Figure 5.4: Spike templates extracted from real MEA recordings (sampled at 10 kHz), serving as basis spikes for generating test signals.

5.3 Experimental Setup

5.3.1 Simulated Data Set (HAB Data Set)

The evaluation of spike detection algorithms is biased towards the data sets used. The usage of real MEA recordings is not practicable due to the lack of reliably annotated spike locations. On the other hand, generating signals with prior information on the location of action potentials should also closely resemble original recorded MEA data. A frequently used model describes the signal as a linear combination of template spikes and modeled noise (Choi et al., 2006; Nabar and Rajgopal, 2009; Natora et al., 2010; Shalchyan et al., 2012). The model introduced in the following additionally includes the reciprocal relationship between the distance of the neuron itself and the electrode as stated by Martinez et al. (2009). This implies that neurons residing in close proximity of the electrode have a higher amplitude than more distant neurons.

In order to get realistic template spike shapes, three action potentials from a MEA recording with sampling frequency of 10 kHz are extracted. The used cells are human embryonic stem cells (hESC, line WA09) which have been differentiated into three dimensional neurospheres. The recordings result from measuring electrical activity of these neurospheres, 24 hours after plating on MEA chips. The three spikes are visualized in Figure 5.4 and differ in shape and polarity. In the following, these three spike templates ψ_i are labeled by the index $i \in \{1, 2, 3\}$. Maccione et al. (2009) assume that there are only a few neurons in the local proximity of a microelectrode. Spikes from neurons further away are not recognizable in noise anymore and

are neglected in this model. Therefore, each of the three spike templates corresponds to a single neuron and is scaled with a fixed parameter $\xi_i < 1$. This parameter reflects the distance from the neuron to the electrode. The simulated data is now constructed by randomly picking one of the three scaled action potentials $\xi_i \psi_i$ and positioning it at a random chosen time τ in the test signal such that there are no overlapping spikes (overlap is considered in the following data set). This procedure is repeated until the desired number of spikes N_{sp} is reached. The resulting signal s_0 can be written as

$$s_0[n] = \sum_{\beta=1}^{N_{\text{sp}}} \xi_i \psi_{i^{(\beta)}} [n - \tau[\beta]] , \tag{5.3.1}$$

where $\tau[\beta]$ contains the time stamps for N_{sp} spikes and $i^{(\beta)}$ is the index of the β-th spike form. Additive background noise can be modeled by an autoregressive model as described by Hayes (1996, Ch. 8.5.1). The model parameters are chosen such that the characteristic noise of actual MEA recordings is approximated best. Using Gaussian white noise as input to the model, noise vectors $p[n]$ are generated and corresponding amplitudes are modeled by a noise level parameter $\kappa > 0$. The resulting data model then reads

$$s[n] = s_0 + \kappa p[n] = \sum_{\beta=1}^{N_{\text{sp}}} \xi_i \psi_{i\beta} [n - \tau[\beta]] + \kappa p[n]. \tag{5.3.2}$$

As the maximum amplitude of the spike templates are constant, the parameter κ is directly related to the signal-to-noise ratio of generated test signals. For large values of κ the spikes are buried in noise, whereas for small values spikes are clearly discernible. By keeping the amount of noise κ as well as the number of spikes N_{sp} fixed, uncorrelated test signals with the same noise amplitude can be generated. Each signal then differs in the positions of spikes and noise characteristics p which are both based on normally distributed pseudorandom numbers. This allows an unbiased performance evaluation independent of absolute spike positions for each noise amplitude.

5.3.2 University of Leicester Data Set (UL Data Set)

In addition to the data set proposed above, the performance of the algorithms are also evaluated on a publicly available data set. This data set is provided by the University of Leicester (Quiroga, 2017) and the corresponding data model described by Quiroga et al. (2004) is closely related to (5.3.2). MEA recordings are similarly simulated by superimposing modeled background

noise with randomly placed spikes. These spikes have distinct amplitudes, shapes and durations (Quiroga et al., 2004, Fig. 4a). Additionally, no spike of the same class can occur within a refractory period of 2 ms, whereas overlap of spikes from different classes is possible (Quiroga et al., 2004, Fig. 4c). The data is simulated at a sampling rate of 96 kHz and subsequently down-sampled to 24 kHz to account for quantization effects resulting from analog-to-digital conversions. The data set comprises 8 simulated signals with differing noise level, ranging from 0.05 up to 0.4, where this number denotes the standard deviation of the noise relative to the spike's amplitude (Quiroga et al., 2004, Ch. 4). Each test signal has a duration of 60 seconds and contains approximately 3400 spikes.

5.3.3 Performance Measures

The most common measure to verify the performance of spike detection algorithms are Receiver Operating Characteristic (ROC) graphs (Fawcett, 2006). These ROC curves plot the True Positive Rate (TPR) against the False Positive Rate (FPR), defined as

$$\mathrm{TPR} = \frac{\text{Number of True Positives}}{\text{Number of inserted Spikes}}, \qquad \mathrm{FPR} = \frac{\text{Number of False Positives}}{\text{Number of detected Spikes}}. \tag{5.3.3}$$

The ROC graphs show the relative tradeoff between the number of true positives and the number of false positives. The closer a graph approximates the optimal classification value (zero false positives and 100 percent true positives) the better the performance of the algorithm. The ROC curves are obtained as follows: Each algorithm leads to an indicator signal y, which is thresholded in order to indicate peak locations. The threshold is varied such that desired FPRs are obtained and corresponding TPRs are recorded.

A different frequently used performance measure is the Detection Rate (DR), defined as the percentage of correctly identified spikes, i.e., the true positive rate (TPR). In order to obtain the DR, the threshold for the indicator signal y must be chosen such that the number of detected spikes equals the number of inserted spikes in the test signal. Thus, the performance of all algorithms can be analyzed with respect to the noise level parameter κ of the test signal. Averaging over several detection rates for one specific κ leads to more reliable results whenever test signals are independently generated. Taking the mean value of the DR (mDR) over all noise levels gives a reasonable number on the performance of a spike detection algorithm,

$$\mathrm{mDR} = \frac{1}{R}\sum_{r=1}^{R} \mathrm{DR}_r. \tag{5.3.4}$$

Here r enumerates R different noise levels.

Due to noise, a tolerance of ± 0.2 ms is introduced, wherein a spike was classified as correctly detected. At a sampling rate of 10 kHz this equals ± 2 samples.

5.3.4 Current State-of-the-Art Methods

In this section common state-of-the-art-algorithms are briefly summarized, which serve as comparison algorithms when benchmarking both proposed methods. Not all algorithms mentioned in Sec. 5.1 are considered in this context. For some algorithms it has already been shown that they are inferior to other algorithms. Thus, it is sufficient to compare the proposed algorithms to the superior ones only. The fractal detector from Salmasi et al. (2016) is not easily applicable for signals with a sampling rate of 10 kHz, due to missing data points for the boxcounting method as investigated by Pissulla (2016). The implementation of Azami et al.'s (2015) algorithm shows poor performance in the simulated data set and has also been excluded (see supplemental material on the GitHub repository *https://github.com/flieb/SpikeDetection-Toolbox/*). Since the performance of the algorithm described by Azami and Escudero (2016, Table 2 and 3) is quite similar to (Azami et al., 2015) this algorithm is also excluded. The algorithm proposed by Liu et al. (2012) uses the same public data set and hence results can be easily compared. Finally, the HBBSD algorithm outperforms Shahid et al.'s (2010) cepstrum of bispectrum approach (Natora et al., 2010, Fig. 5). Categorized into the three groups from Section 5.1 this leaves the following state-of-the-art algorithms to be considered:

Absolute Thresholding Spike Detection Algorithm (ABS)

Spikes are detected by checking whether the input signal is above or below a certain threshold. In the case of symmetric thresholds, the resulting indicator signal is given by $y = |f|$ (cf. Section 5.1). The ABS method is simple and easy to implement, but very sensitive to noise. This algorithm from Category A is included since it is well-known and still used extensively for spike detection, see for example (Daus et al., 2011; Karkare et al., 2013; Takekawa et al., 2010; Wallace et al., 2008).

Precision Timing Spike Detection Algorithm (PTSD)

A more advanced algorithm of Category A is described by Maccione et al. (2009). The indicator signal y is based on absolute voltage differences between local minima and maxima within a

predefined time-period denoted peak lifetime period. Whenever such a difference is greater than the threshold, the corresponding signal segment is classified as a spike.

Hybrid Blind Beamforming Spike Detection Algorithm (HBBSD)

The HBB spike detection algorithm is the only representative from Category B and is described by Natora et al. (2010). First, template spike forms are computed by utilizing higher order statistics. Second, the indicator signal y results from convolving these templates with the original signal. No prior knowledge of spike forms are needed and the algorithm also takes into account that there can be more than just one spike template.

Multiresolution TEO Spike Detection Algorithm (MTEO)

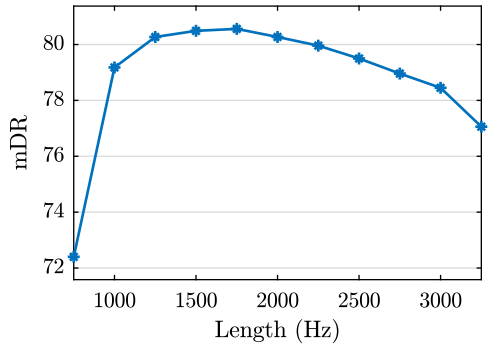
The multiresolution Teager energy operator (MTEO) is defined by Choi and Kim (2002) as

$$\mathcal{B}^{q}_{\text{TEO}}(f[n]) = f^2[n] - f[n-q]\,f[n+q], \tag{5.3.5}$$

where $q \in \mathbb{N}^+$ denotes a resolution parameter. Since it is easy to show that $\mathcal{B}_{\text{TEO}}(A\cos(\Omega n + \phi)) \simeq A^2\sin^2(q\Omega)$ (cf. Eq. (5.2.6)), the parameter q is related to the frequency of action potentials. As the optimal value of q depends on the width of action potentials, the simultaneous use of different values is advantageous (Choi and Kim, 2002). Choi et al. (2006) have shown, that values 1, 3 and 5 lead to reasonable results in the context of spike detection. Additionally, Hamming windows of length $4q + 1$ are normalized by the noise power for each value q (Choi et al., 2006, Eq. (13)). Thus, each q-channel contributes with an equal weight to the indicator signal y, which is composed by a maximum filter picking for every time instance the maximum from the three smoothed TEO outputs.

Stationary Wavelet Transform based Spike Detection Algorithm (SWT)

This spike detection algorithm is based on the stationary wavelet transform and is introduced by Shalchyan et al. (2012). The input signal is decomposed into five levels of a stationary wavelet transform, in the following denoted with W_j, with j denoting the corresponding decomposition level. These detail coefficients are then thresholded by the following level dependent threshold $\lambda_j = 0.8\sqrt{2\log(N)}\text{MAD}(W_j)/0.6745$, where N is the signal length and MAD denotes the median absolute deviation operator (Shalchyan et al., 2012, Eq. (5)). This gives W_{λ_j} and the energy of each thresholded scale j is computed according to $EW_j = \sum_n (W_{\lambda_j}[n] - \overline{W_{\lambda_j}}[n])^2$,



	$K' = 1$	$K' = 5$	$K' = 9$
$L' = 1$	80.6	80.6	80.4
$L' = 33$	81.0	80.9	80.6
$L' = 67$	81.4	81.4	81.2
$L' = 101$	81.9	81.8	81.5

Figure 5.5 & Table 5.1: Performance of the TIFCO algorithm with Hann windows of various lengths and with averaging kernels η of different size.

where the overline denotes the mean. The absolute values of the three scales with the largest energy are summed up and the resulting function is smoothed by a Bartlett window. Why only 80% of the threshold is used is not justified and might be the reason why this algorithm does not yield good performance in the simulated data sets proposed here. Note that the SWT algorithm is the only algorithm which does not follow the basic scheme in Figure 5.1. Individual sub-bands of the SWT are thresholded before a decision variable is accumulated.

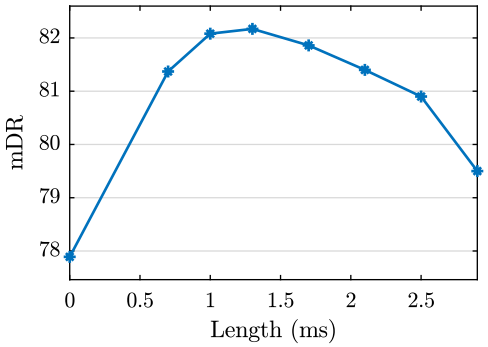
5.3.5 Parameter Evaluation

The following parameter evaluation for both proposed algorithms are based on the HAB data set with noise level $\kappa \in [1, 2, \ldots, 11]$. For each noise level 20 independent test signals are averaged, before the mean value of all noise levels leads to the mDR as defined in (5.3.4).

Parameter Choice for the TIFCO Algorithm

The time sampling parameter a and the number of frequency bins M do not directly influence the performance, but rather the run time of the algorithm. Hence, these parameters are fixed to $a = 1$ and $M = 101$ in the following. Further, the frequency interval is adapted to spike specific time-frequency patterns as depicted in Figure 5.3, i.e., $f_{\text{min}} = 500\,\text{Hz}$ and $f_{\text{max}} = 3500\,\text{Hz}$. This leaves the choice of the window function g and the filter kernel η.

Firstly, windows g are chosen to be compactly supported in frequency domain such that the width can be properly specified. Figure 5.5 shows the performance of the TIFCO algorithm based on Hann windows with increasing length. The best performance is found for a length of



Wavelet	mDR	Wavelet	mDR
db1	80.8	sym7	80.9
db4	79.5	coif1	81.7
sym4	81.9	bior3.9	81.6
sym5	82.1	rbio3.9	80.8

Figure 5.6 & Table 5.2: Performance of the SWTTEO algorithm using smoothing windows of various lengths and different wavelets.

1750 Hz, corresponding approximately to a window size of 2.3 ms, which seems a reasonable result. The window width is larger than a spike duration, but not too large in order to avoid multiple spikes in a single data segment.

Secondly, the choice of the 2-D filter kernel η is motivated by the characteristic time-frequency pattern of spikes. This implies a kernel which puts emphasis on a few consecutive large coefficients in time direction as well as sequential large coefficients over several frequencies. In order to emphasize this spike pattern a moving average kernel is chosen, but other choices like Gaussian or sharpening filters as well as nonlinear approaches such as median filters are possible. The average kernel $\eta \in \mathbb{R}^{K' \times L'}$ is defined by

$$\eta[k', l'] = \frac{1}{K' \cdot L'}, \tag{5.3.6}$$

for all k' and l'. If $K' = L' = 1$ no filtering is done. The results in Table 5.1 show that averaging over all frequency coefficients yields the best performance rate as the influence of 1/f noise is attenuated.

Parameter Choice for the SWTTEO Algorithm

It can be empirically verified that a decomposition level of 2 for the SWT leads to the best performance, which also confirms the results of Nabar and Rajgopal (2009). Further, the performance of the SWTTEO algorithm based on a selection of different wavelets can be seen in Table 5.2, where the naming convention of wavelets follows the MATLAB notation (see the MATLAB routine `wfilters` and references therein). With a decomposition level of 2 and a

smoothing window of length 1.3 ms, the sym5-wavelet gives the best mDR.

Figure 5.6 illustrates the performance of the SWTTEO algorithm with respect to smoothing windows of increasing lengths. Obviously, smoothing the output of the energy operator indeed increases the accuracy of the SWTTEO algorithm (a window length of 0 ms in Fig. 5.6 indicates no smoothing). The best performance is achieved for window lengths which are close to the spike duration. Further, the results in Figure 5.6 indicate that the performance is robust even if spike durations are not known explicitly a priori, as within the range of 1 to 2 ms the performance decrease is only marginal.

Parameter Choice for Current State-of-the-Art Algorithms

The choice of parameters for the comparison algorithms described in Section 5.3.4 is crucial for a fair comparison. Therefore, recommended values according the respective literature are empirically adjusted for optimal performance based on the HAB and the UL data set. This results in the following parameters.

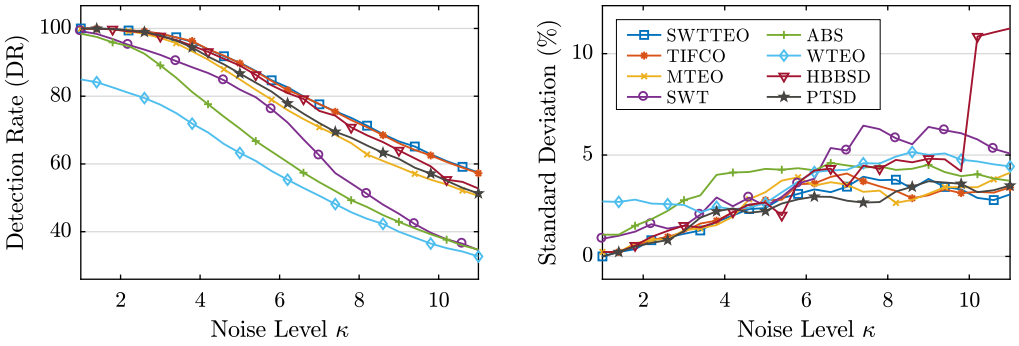
The WTEO algorithm is based on the sym5 wavelet and a decomposition level of 2. Originally, the db4 wavelet is suggested by Nabar and Rajgopal (2009), however, the sym5 wavelet leads to a better performance. The SWT algorithm uses a decomposition level of 5 with the same sym5 wavelet. For the HBBSD algorithm the filter length is set to 1.4 ms and the remaining parameters are chosen as proposed by Natora et al. (2010, Sec. III.C). The PTSD algorithm requires a peak lifetime period and a refractory period, which were set to 0.5 ms and 1.2 ms respectively. The MTEO as well as the ABS algorithm do not have additional parameters.

5.4 Results

5.4.1 HAB Data Set

In the following numerical evaluations, test signals of length 6 s are generated with 100 randomly inserted spikes at a sampling frequency of 10 kHz. The noise level is varied by choosing κ from the interval $[1, 2, \ldots, 11]$ and for each noise level the resulting performance is averaged over 20 independent test signals.

First, the performance of all considered algorithms is evaluated based on the detection rate (DR) illustrated in Figure 5.7. The resulting DR and corresponding standard deviation are shown in Figures 5.7a and 5.7b. It shows that both proposed algorithms achieve a higher detection rate compared to all current state-of-the-art algorithms, regardless of the noise level. The



(a) Detection rate

(b) Standard deviation

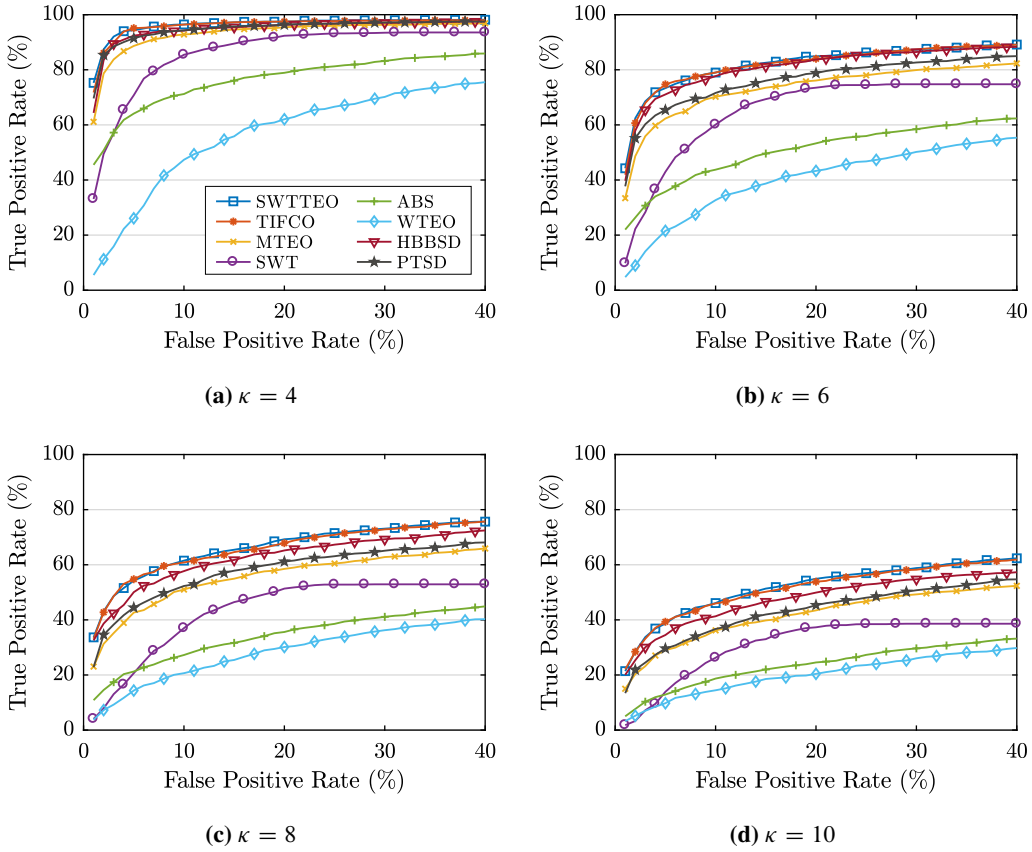
Figure 5.7: Detection rate and corresponding standard deviation depending on different signal-to-noise ratios for the HAB data set.

mean detection rate (mDR) for the SWTTEO and the TIFCO methods are 82.285 and 82.094 respectively, followed by the HBBSD algorithm with 80.503, the PTSD with 78.76 and the MTEO with 77.48. The poor performance of the WTEO algorithm results from not smoothing the TEO output as well as expanding decimated wavelet coefficients. Both accentuate noisy spikes.

Second, Figure 5.8 shows ROC curves for noise levels $\kappa \in [4, 6, 8, 10]$. For each algorithm, the TPR is computed by thresholding the indicator signal y to yield the desired false positive rate (FPR). Apart from the WTEO, SWT and ABS algorithm, the majority of the algorithms perform quite similar for small noise levels κ. The performance gap between both proposed algorithms and current state-of-the-art algorithms emerges with increasing κ and emphasizes the superiority of the TIFCO and SWTTEO algorithm.

5.4.2 UL Data Set

The data set from the University of Leicester has been simulated at a sampling rate of 24 kHz, requiring adjustments for minimum and maximum frequency of the TIFCO algorithm. Therefore, the parameters f_{min} and f_{max} are changed to 1 kHz and 8 kHz respectively. All other parameters remain as previously described, as empirically adjusting these parameters does not lead to significant performance improvements. Further, in order to separate overlapping spikes the analyzing window function for the Gabor transform needs to be more narrow. But, this short window may cause spikes with longer durations to be detected twice: the positive and negative



(a) $\kappa = 4$ **(b)** $\kappa = 6$

(c) $\kappa = 8$ **(d)** $\kappa = 10$

Figure 5.8: ROC curves for all tested algorithms with four different noise levels κ based on the HAB data set.

part separately. Choosing a wider window in contrast reduces the chance to detect overlapping spikes. A reasonable tradeoff can be achieved by setting the window width for the DGTSF to 3000 Hz.

Figure 5.9 shows for each algorithm the percentage of correctly identified spikes depending on the eight different noise levels of the UL data set. The HBBSD algorithm is not included since reasonable spike templates could not be found. For small as well as large noise levels both proposed methods outperform all other spike detection algorithms. At the largest noise level, both proposed methods still achieve a detection rate of approximately 80% in contrast to only 44% correctly detected spikes by the ABS algorithm. The tradeoff of the window length is the main reason why the TIFCO algorithm shows a slightly inferior performance to the SWTTEO

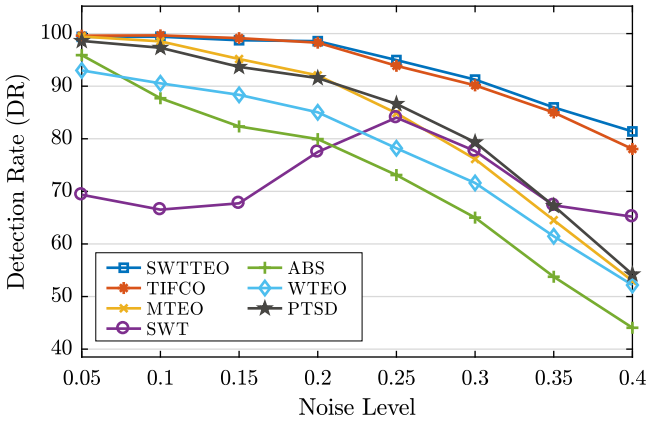


Figure 5.9: Detection rate of spike detection algorithms in dependance on the noise level based on the UL data set.

algorithm. Here, the length of the smoothing window for the SWTTEO algorithm does not influence the detection of overlapping spikes. As mentioned before, the results in Fig. 5.9 can be directly compared with the piecewise optimal morphological filter approach by Liu et al. (2012, Table 1). Liu et al. based their evaluations on the same data set, but only up to a noise level of 0.2. Their algorithm achieved a corresponding detection rate of 95.18%, whereas the SWTTEO approach obtains 98.56% and the TIFCO 98.27% for the same noise level (0.2). Unfortunately, there are no results for larger noise levels, but based on the detection rates it can be concluded that the performance will be inferior to both proposed approaches.

5.4.3 Runtime

The future development of microelectrode arrays with higher density require fast algorithms for spike detection. Therefore, a basic run time evaluation of both proposed spike detection algorithms is visualized in Figure 5.10. The run time is compared with the ABS algorithm which has the lowest computational complexity. For each signal length, Figure 5.10 shows the mean and standard deviation of 10 independent runs. It is not surprising that the TIFCO approach performs worst, as it has the most computational expensive steps. The SWTTEO algorithm is in the best case one order of magnitude slower than the ABS algorithm. Nonetheless, it is still feasible to use the SWTTEO algorithm whenever online spike detection (during acquisition) is required (Franke et al., 2012).

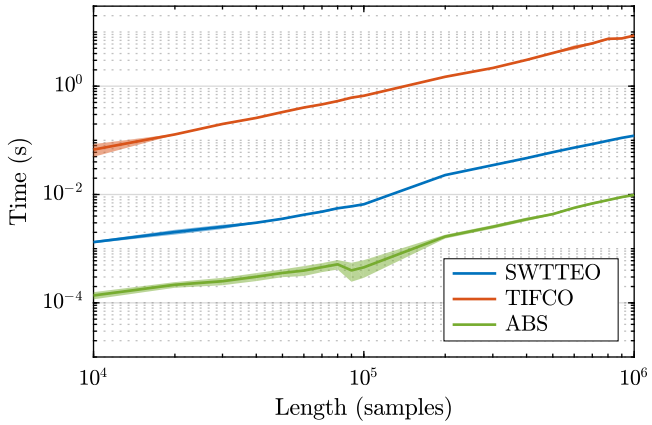
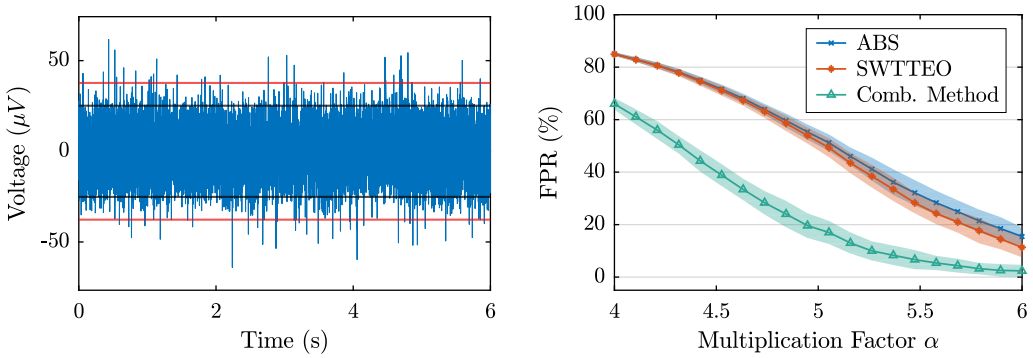


Figure 5.10: Run time comparison of both proposed algorithms and the ABS algorithm.

5.5 Application to MEA Recordings

In real MEA recordings the number of spikes present is in general unknown. This requires an appropriate choice of the threshold to extract spike locations from the indicator signal y. A widely used method is based on choosing the threshold to be a multiple of the noise level in the recording. This noise level can be estimated either statically or dynamically as the standard deviation of the signal's noise (Daus et al., 2012). In combination with the simple thresholding algorithm (ABS), multiplication factors α typically range between 4 and 6 (Karkare et al., 2013; Nick et al., 2014; Regalia et al., 2016). Such approaches, however, are not easily applicable to the proposed algorithms SWTTEO and TIFCO, since a reasonable choice of multiplication factors is difficult. The range of the indicator signal y does not depend on the noise level anymore and might differ for different electrodes. Therefore, a novel approach has been introduced by Mayer et al. (2018), which combines the well established simple thresholding with the more reliable SWTTEO approach.

The proposed spike detection algorithm for MEA recordings can be described as follows. First, the noise level σ_{est} in a recording is estimated using methods described in (Biffi et al., 2010; Nick et al., 2014). Both methods are essentially based on windowing the recording with a subsequent estimation of either the root-mean-square or some percentiles of the voltage distribution. The threshold λ_{est} then results by multiplying the estimated noise level with some fixed factor α. All positions where the recording exceeds $\lambda_{\text{est}} = \alpha\sigma_{\text{est}}$ are considered as potential spike candidates. The resulting number of potentially detected spikes is then used to detect the same number of candidate spikes using the SWTTEO algorithm. This results in two independent



(a) Test signal with indicated thresholds for multiplication factors $\alpha = 4$ (black) and $\alpha = 6$ (red).

(b) False positive rate with standard deviation

Figure 5.11: False positive rate (FPR) for the proposed algorithm based on different multiplication factors α.

spike candidate lists of equal length. In the last step both lists are intersected and only those spikes which have been detected by both algorithms remain. This obviously reduces the number of detected spikes, but it also reduces the false positive rate (FPR).

Results based on the simulated HAB data set can be seen in Figure 5.11. On the left side Fig. 5.11a shows a simulated MEA signal with $\kappa = 5$, where 100 spikes are present. The threshold thr_{est} is indicated by black and red horizontal lines for multiplication factors $\alpha = 4$ and $\alpha = 6$. The proposed algorithm then gives the false positive rate illustrated in Fig. 5.11b with respect to different values for α. For comparison the FPR for the ABS algorithm as well as the SWTTEO algorithm are included. For each multiplication factor α the FPR is averaged over 20 independent simulations and the mean and standard deviation are visualized.

The simulated results justify the combination of the two spike detection algorithms for a more sensitive analysis of electrophysiological processes of human embryonic stem cells. At the same time, the approach can be based on well known concepts for unsupervised spike detection. This will be useful in order to determine the influence of antiepileptic drugs and ionized radiation on electrophysiological properties of 3D neuronal networks (Mayer et al., 2018).

5.6 Conclusion

Two novel spike detection algorithms have been introduced, one exploiting spike specific frequency patterns and the other analyzing energy of stationary wavelet coefficients. The per-

formance of both algorithms is compared to current state-of-the-art methods based on two simulated data sets. Regardless of the used data set, both proposed approaches show superior performance. This is also visualized in Fig. 5.1, where only the SWTTEO algorithms detects all spikes correctly. Furthermore, a straightforward application to real MEA recordings has been introduced, combining a well-established method for threshold selection with the sensitivity of the SWTTEO approach.

The marginal superiority of SWTTEO over TIFCO might lead to the question what the advantage of the TIFCO algorithm is. Other than the SWTTEO, the TIFCO algorithm is a basic framework with a simple moving average filter to emphasize spike features. More sophisticated algorithms originating from image processing can be considered to further accentuate spike-specific frequency patterns. Further improvement may also be drawn from evaluating the phase of complex time-frequency coefficients. So far, only the magnitude is utilized to extract spike features. More advanced algorithms from the field of audio processing like reassigning the time-frequency plane (Auger et al., 2013) are based on the phase and may improve the performance of the TIFCO algorithm. Additionally, it is noted that using a time-frequency representation based on wavelet frames as introduced in Section 3.4 does not lead to an improved performance. This is not surprising, since spike durations are usually within a certain range, which makes fixed length windows more appropriate rather than scaled ones.

Both proposed algorithms perform reasonably well in the case of overlapping spikes without any further adjustments. However, for the TIFCO approach the length of the window function directly influences the separability of overlapping spikes, leaving room for further improvement. For instance, the DGTSF can be computed with two windows of different length. A short one to distinguish overlapping spikes and a longer one to reduce the chance of detecting the same spike twice. The SWTTEO algorithm also does not explicitly handle overlapping spikes. Here, an additional feature extraction of corresponding wavelet coefficients as proposed by Song and Li (2015) can enhance the detection of overlapping spikes even further.

In MEA recordings the absolute number of spikes present is usually unknown. This complicates the sole usage of one of the proposed methods, since an appropriate threshold needs to be estimated. This threshold can be based on the amount of noise present in the original signal f, but can also depend on noise present in the decision variable y. Further, approaches independent of noise levels, like evaluating the energy of spike indicators in the decision variable y might lead to better threshold estimates. Only if these threshold estimates are reliable, the spike detection algorithms TIFCO and SWTTEO can be evaluated on real data-sets.

Chapter 6

Peak Detection for MALDI Mass Spectrometry Using Frame Multipliers

6.1 Introduction

Matrix assisted laser desorption ionization (MALDI) is a widely used form of imaging mass spectrometry, analyzing molecular compositions of tissue sections as described in more detail in Section 1.3. Following these introductory remarks, the processing of MALDI imaging data can be structured as shown in the schematic diagram in Figure 6.1. This pipeline is proposed by Alexandrov et al. (2010) and has been modified to include all preprocessing steps.

First, initial preprocessing includes baseline removal, spectra smoothing and normalization. Obtained spectra exhibit an intensity offset for lower m/z values which originate from resulting clusters of the applied matrix during ionization (Sun and Markey, 2011). This offset is called baseline and might differ for each spectrum. Sophisticated algorithms for baseline removal are based on asymmetric least squares (de Noo et al., 2006), wavelets (Sun and Markey, 2011) or top hat filters (Sauve and Speed, 2004). After removing the baseline, Deininger et al. (2011) illustrated the advantages of normalizing MALDI imaging data spectra wise. A frequently used normalization procedure is the total ion count (TIC) normalization, which is shown to reduce artifacts resulting from matrix inhomogeneities. Spectra smoothing is also a quite commonly used preprocessing approach (Bauer et al., 2011; Du et al., 2006; Shin et al., 2010; Sun and Markey, 2011; Wijetunge et al., 2015; Yang et al., 2009). Such methods include wavelet based denoising (Coombes et al., 2005b; Kwon et al., 2008; Shin et al., 2010) and Savitzky-Golay or Gaussian filters (Yang et al., 2009).

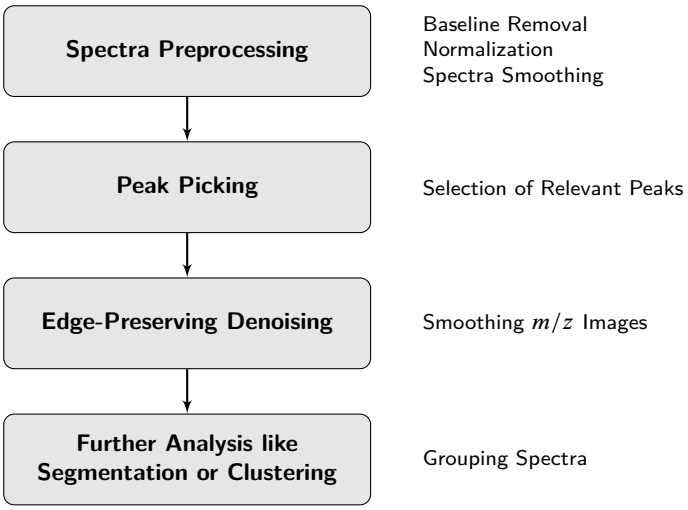


Figure 6.1: MALDI processing pipeline (adapted from Alexandrov et al., 2010, Scheme 1).

After preprocessing peak picking algorithms detect prominent peaks in mass spectra, separating peaks corresponding to molecules from noise. This noise might arise from varying thickness of the applied matrix, artifacts resulting from ion suppression or from electronic noise (Deininger et al., 2011). Shin et al. (2010) showed that the variance of the noise is larger in lower mass regions and decreases with increasing m/z values. This makes accurate peak picking quite challenging, since any further analysis like spatial segmentation or clustering groups of similar spectra are based upon the detected peaks. After selecting prominent peaks, Alexandrov et al. (2010) demonstrated that smoothing m/z images greatly enhances segmentation results. This is also confirmed and further improved by Kobarg (2014).

The number of proposed peak picking algorithms is quite large with highly diverse approaches. Early MALDI peak picking approaches are based on simple local maxima (Breen et al., 2000) or fitting Gaussian distributions to mass peaks (Kempka et al., 2004). More advanced algorithms make use of the continuous wavelet transform (CWT) (Antoniadis et al., 2010; Du et al., 2006; Lange et al., 2006) or the discrete wavelet transform (Alexandrov et al., 2009; Coombes et al., 2005b; Kwon et al., 2008). Especially the CWT approach based on ridges and zero-crossings of wavelet coefficients have gained recent popularity (Antoniadis et al., 2010; Du et al., 2006; Zhang et al., 2015). Two independent comparisons of MALDI peak picking algorithms favor wavelet approaches over other conventional methods (Bauer et al., 2011; Yang et al., 2009).

Wijetunge et al. (2015) recently introduced a new peak detection algorithm. Their approach has shown superior performance compared to the ridge line wavelet approach from Du et al.

(2006), the discrete wavelet transform based algorithm from Coombes et al. (2005b) and a Bayesian approach based on adaptive regression kernels (House et al., 2011). Simulations were performed on a publicly available data set, mimicking the features of real MALDI-TOF data. This makes the Wijetunge approach an ideal candidate to verify performance of the newly proposed peak picking algorithm.

The peak picking algorithms described above are all based on spectra wise peak picking. Spatial information, however, might improve the peak picking process. Large peaks which are spatially surrounded by small peaks might be more likely to be ignored. In contrast, a relative small peak in a neighborhood of larger peaks might be relevant. Alexandrov and Bartels (2013) introduced an algorithm which detects peaks not spectra wise, but in corresponding m/z images. This is shown to evidently improve the sensitivity of peak picking algorithms compared to other classical spectra wise methods (Alexandrov and Bartels, 2013, Fig. 5).

The algorithm proposed in the following is also based on spectra wise peak detection, but can be modified to include spatial information. The combination of spectra wise peak picking and spatial awareness reduces the second and third steps in Figure 6.1 into a single step. This implies, that the proposed algorithm edge-preservingly smooths m/z images while detecting peaks. Furthermore, the algorithm is designed such that preprocessing spectra becomes obsolete. By dividing spectra into small sections baseline effects, for example, can be sufficiently suppressed by the algorithm. Essentially, the proposed algorithm inherently combines the first three pipeline steps in Fig. 6.1 into a single step, reducing computational complexity significantly.

6.2 Algorithm Description

The peak picking algorithm introduced in the following can be split into two variants. The basic approach is a spectra wise algorithm, whereas the modified version also utilizes neighboring spectra. Both methods are described in detail in the following sections, after defining some required mathematical concepts first.

6.2.1 Frame Multiplier

The proposed approach is based on an adapted and further improved audio processing algorithm tracing sound objects in nonstationary background noise based on Gabor multipliers (Dörfler and Matusiak, 2013, 2014). Especially the time-varying noise behavior justifies such an approach for detecting peaks in MALDI-TOF data, since the noise level of spectra decreases with increasing

mass-to-charge ratio (Kwon et al., 2008; Shin et al., 2010; Wu et al., 2010). A key concept for the proposed peak picking algorithm are frame multipliers (Balazs, 2007), where the involved frame elements depend on two parameters, e.g., time and frequency.

Definition 6.2.1 (Frame Multiplier)**.** Let $g_{k,l}$ and $\gamma_{k,l}$ be a frame and its dual and denote by Φ_g and Φ_γ^* the analysis and synthesis operators of the corresponding frames. Further, let I_m be a diagonal operator in the corresponding transform domain. Thus, for any suitable $f \in L^2(\mathbb{R})$ the operator $\mathcal{M}_m : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ defined by

$$\mathcal{M}_m f = \Phi_\gamma^* I_m \Phi_g f = \sum_{k,l \in \mathbb{Z}} m_{k,l} \left\langle f, g_{k,l} \right\rangle \gamma_{k,l}, \tag{6.2.1}$$

is called a frame multiplier with frame mask $m_{k,l} \in L^\infty\left(\mathbb{R}^2\right)$.

Essentially, the frame multiplier acts in the transform domain of frame g by a pointwise multiplication with mask m and a subsequent inverse transform based on γ. Balazs (2007) showed that the frame multiplier is bounded whenever $m \in L^\infty(\mathbb{R})$. The easiest example of a frame multiplier is given, with a slight abuse of notation, by setting $g_{k,l}$ and $\gamma_{k,l}$ to be the complex exponential $g_{k,l}(\cdot) = e^{2\pi i k \cdot}$ and its complex conjugate: a convolution operator. Whenever $g_{k,l}$ and $\gamma_{k,l}$ are Gabor frames as defined in Definition 3.2.1, the multiplier $\mathcal{M}_m$ is known as a Gabor multiplier first introduced by Feichtinger and Nowak (2003). Such Gabor multipliers are usually used with a known mask m to filter time-frequency coefficients by enforcing certain structures and suppressing unwanted time-frequency components (Balazs et al., 2010). The mask m, however, can also be used to measure similarity between two signals f_1 and f_2 by considering the following equation

$$f_2 = \sum_{k,l \in \mathbb{Z}} m_{k,l} \left\langle f_1, g_{k,l} \right\rangle \gamma_{k,l}. \tag{6.2.2}$$

Assuming that f_1 has non-vanishing coefficients $\left\langle f_1, g_{k,l} \right\rangle$, the frame mask m describes the transition of f_1 to f_2 in the corresponding domain. Depalle et al. (2007), Kronland-Martinet et al. (2010) and Olivero et al. (2013) utilized this concept in order to morph one sound signal into another. For example, Kronland-Martinet et al. (2010) morphed a clarinet tone into a saxophone tone by estimating masks m with specific characteristics. On the other hand, this approach can also be used to identify differences between two signals. If $f_1 = f_2$ then the mask is $m = 1$ for all $k, l \in \mathbb{Z}$. Otherwise, the mask m indicates how much, and, in particular, at which spatial locations both signals differ.

6.2.2 Multiplier Estimation with Sparsity Constraints

Given two signals f_1 and f_2, similarity between these two signals can be measured by estimating the mask m in (6.2.2). Obviously, a simple approach would be to set

$$m_{k,l} = \frac{\langle f_2, g_{k,l} \rangle}{\langle f_1, g_{k,l} \rangle}, \tag{6.2.3}$$

which might be unbounded in general. Using some appropriate regularization term $d(m)$ depending on the mask m and a regularization parameter $\lambda > 0$, similarity can be estimated by the following minimization problem

$$\min_m \frac{1}{2} \| f_2 - \mathcal{M}_m f_1 \|_2^2 + \lambda d(m). \tag{6.2.4}$$

A reasonable choice for the regularization term $d(m)$ is given by $d(m) = \| |m| - 1 \|_1$, introducing sparse frame masks (Dörfler and Matusiak, 2014; Kronland-Martinet et al., 2010). It is further shown that considering only the magnitude of masks reduces spurious noise artifacts whenever transform coefficients are complex valued, e.g., for Gabor frames (Kronland-Martinet et al., 2010).

The sensitivity can be adjusted with the regularization parameter λ. For relatively small values any difference between the transform coefficients of f_1 and f_2 is captured by the mask. With increasing λ small deviations between the transform coefficients are ignored and the resulting mask is set to 1. This way, only the most prominent differences between f_1 and f_2 can be detected.

Unfortunately, the inverse problem in (6.2.4) does not admit a closed form solution. Rewriting this equation gives

$$\min_m \frac{1}{2} \left\| \Phi_\gamma^* \left(\Phi_g f_2 - m \cdot \Phi_g f_1 \right) \right\|_2^2 + \lambda d(m). \tag{6.2.5}$$

Minimizing this expression is not easy, since the operator Φ_γ^* is not injective. As suggested by Dörfler and Matusiak (2014) and Olivero et al. (2013) it is sufficient, however, to formulate the problem in the transform domain. For notational convenience let the transform coefficients be denoted by

$$c_1 = \Phi_g f_1 = \langle f_1, g_{k,l} \rangle, \qquad c_2 = \Phi_g f_2 = \langle f_2, g_{k,l} \rangle, \tag{6.2.6}$$

where the dependence on k and l is implicit from now on. The resulting inverse problem

$$m = \arg\min_m \frac{1}{2} \|c_2 - mc_1\|_2^2 + \lambda \| |m| - 1 \|_1 , \tag{6.2.7}$$

then admits a closed form solution. For the purpose of detecting peaks this expression can be further simplified. Whenever the coefficients c_1 and c_2 are complex valued, e.g., for Gabor frames only the magnitude contains valuable information. Replacing c_1 and c_2 with its magnitude, the absolute value of the mask in the regularization term can be omitted. The following Theorem 6.2.2 derives the closed form solution of the simplified inverse problem stated in (6.2.7).

Theorem 6.2.2. *Let c_1 and c_2 be the transform coefficients of f_1 and f_2 with corresponding frame $g_{k,l}$ as defined in* (6.2.6)*. The simplified minimization problem of* (6.2.7)

$$m = \arg\min_m \frac{1}{2} \| |c_2| - m\,|c_1| \|_2^2 + \lambda \|m - 1\|_1 , \tag{6.2.8}$$

has the following closed form solution

$$m = \left(\frac{|c_2|}{|c_1|} - 1 \right) \left(1 - \frac{\lambda}{|c_1|^2 \left| \frac{|c_2|}{|c_1|} - 1 \right|} \right)^+ + 1, \tag{6.2.9}$$

where $(\cdot)^+ = \max(0, \cdot)$ denotes the maximum with zero.

Proof. Obviously, $m = 1$ whenever $|c_1|$ is trivial. Assuming that $|c_1|$ is non-trivial leads to

$$\begin{aligned} & 0 \in \nabla_m \left(\frac{1}{2} \| |c_2| - m\,|c_1| \|_2^2 \right) + \lambda \partial \left(\|m - 1\|_1 \right) \\ \Leftrightarrow \quad & 0 \in m\,|c_1|^2 - |c_2|\,|c_1| + \lambda \partial \left(\|m - 1\|_1 \right) \\ \Leftrightarrow \quad & m \in \frac{|c_2|}{|c_1|} - \frac{\lambda}{|c_1|^2} \partial \left(\|m - 1\|_1 \right) . \end{aligned} \tag{6.2.10}$$

The subdifferential of the ℓ_1-norm consists of the following subgradients, which can be evaluated for each coefficient separately

$$\partial\,|m - 1| = \begin{cases} \{1\} & \text{if } m > 1 \\ \{-1\} & \text{if } m < 1 \\ [-1, 1] & \text{if } m = 1 \end{cases} . \tag{6.2.11}$$

Considering all three cases in (6.2.10) leads to the closed form solution

$$m = \begin{cases} \frac{|c_2||c_1|-\lambda}{|c_1|^2} & \text{if } \frac{|c_2|}{|c_1|} - 1 > \frac{\lambda}{|c_1|^2} \\ \frac{|c_2||c_1|+\lambda}{|c_1|^2} & \text{if } \frac{|c_2|}{|c_1|} - 1 < -\frac{\lambda}{|c_1|^2} \\ 1 & \text{if } -\frac{\lambda}{|c_1|^2} \leq \frac{|c_2|}{|c_1|} - 1 \leq \frac{\lambda}{|c_1|^2} \end{cases} . \tag{6.2.12}$$

With $y = |c_2|\,|c_1|^{-1}$, the equivalence of (6.2.12) and (6.2.9) follows directly: assuming $1 - \lambda\,|c_1|^{-2}\,|y-1|^{-1} \leq 0$ in (6.2.9) is equivalent to the third row in (6.2.12). On the other hand, $1 - \lambda\,|c_1|^{-2}\,|y-1|^{-1} > 0$ leads to the first and second row of (6.2.12) by considering the cases $y - 1 > 0$ and $y - 1 < 0$, respectively.

□

6.2.3 Peak Picking Algorithm

Basic Algorithm

Based on Theorem 6.2.2 the peak picking algorithm can now be defined in a finite dimensional setting as follows. Let $f \in \mathbb{R}^L$ be a single raw MALDI spectrum of length L. This signal is divided into overlapping slices f_i of length Δ and overlap $o \in (0,1)$. An overlap between slices is required, otherwise peaks might be unintentionally separated into two consecutive slices. For both slices the transform coefficients based on the given frame $g_{k,l}$ can be computed in the next step. Frames, which might be interesting to consider are Gabor and wavelet frames, resulting in either time-frequency or time-scale coefficients. From these coefficients the mask m can now be estimated using (6.2.9) for a given regularization parameter λ.

Once the mask is estimated, it indicates at the most prominent differences between two consecutive and overlapping transform coefficients c_i and c_{i+1} for corresponding f_i and f_{i+1}. If both slices are similar with respect to λ, the mask is constant and supposedly no peak is present. Otherwise, if the mask takes values different from 1, a closer inspection might indicate possible peak locations. Generally, there are three possible cases to consider:

- There is a peak present in f_i and no peak at the same position in f_{i+1}, resulting in values smaller than one in the mask m_i, in order to lessen the influence of large coefficients in c_i compared to c_{i+1}.
- A peak in f_{i+1} and no peak at the same location in f_i yields a mask with coefficients larger than one.

- There are peaks in f_i and f_{i+1} at the same location, meaning that these two peaks are located exactly $(1-o)\Delta$ apart in the original spectrum f. Both transform coefficients might not differ significantly enough for the mask m_i to deviate from a constant one.

From the first two cases, peak locations can be easily estimated by considering the values of $m_i - 1$. Resulting negative values, denoted in the following by $(\cdot)_{\text{neg}}$, are accumulated directly by summing the absolute mask coefficients over all frequency/scale indices,

$$z_k^i = \sum_l \left| \left(m_{k,l}^i - 1 \right)_{\text{neg}} \right|. \tag{6.2.13}$$

With every slice f_i the signal z_k^i then accumulates to an indicator signal for peaks. On the other hand, positive entries in $m_i - 1$, indicating at a peak in f_{i+1}, can be used in the subsequent step when the difference between f_{i+1} and f_{i+2} is estimated as an additional source of information. Obviously, when analyzing m_{i+1} negative coefficients are expected at the positions where in the preceding step positive coefficients occurred. If this is the case, it can be proceeded with (6.2.13). If, however, $m - 1$ is zero, where negative coefficients are expected it can only be due to the third case mentioned above. In such cases, the information of preceding as well as subsequent masks can be used to circumvent that peaks are getting ignored, whenever peaks are present in two subsequent slices f_i and f_{i+1} at the same relative location. The mask for f_{i-1} and f_i should lead to positive values and negative values for f_{i+1} and f_{i+2}, provided all peaks in the spectrum are not periodically spaced. In MALDI data such regular patterns can be neglected though.

In general, Equation (6.2.9) tends to be more sensitive for peaks in f_i instead of f_{i+1}, since the regularization parameter λ gets weighted with $|c_i|^2$. Consequently, the first case of the three mentioned above is usually the one occurring most. A summary of the proposed peak picking approach can be found in Algorithm 10. Note that the inner products in line 3 and 4 are finite dimensional with respect to the frame length Δ. Additionally, the computation for N spectra can be done using an additional loop or, as implicitly indicated in Algorithm 10, a vectorized approach.

Slicing spectra into smaller parts has several advantages. First, a raw spectrum could be analyzed without preprocessing the baseline. If the slice length Δ is chosen small enough, the influence of baseline effects of two consecutive slices can be made negligibly small. A second advantage is that the algorithm's sensitivity can be adjusted to the noise level. Based on time-frequency or time-scale coefficients for both slices, the amount of noise present in these

Algorithm 10: MALDI Peak Picking Algorithm

Input : $f \in \mathbb{R}^{N\times L}$ - Raw MALDI data,
$\lambda \in \mathbb{R}^{+}$ - Regularization parameter,
Δ - Length of slices,
o - Overlap between slices,
$g_{k,l}$ - Frame of length Δ,

Output: $z \in \mathbb{R}^{N\times L}$

$M \leftarrow$ Total number of slices with length Δ and overlap o dividing L
for $i = 0, 1, 2, \ldots, M-1$ **do**
  $c_1 \leftarrow \langle f_i, g_{k,l} \rangle$
  $c_2 \leftarrow \langle f_{i+1}, g_{k,l} \rangle$
  $y \leftarrow |c_2|\,|c_1|^{-1}$
  $m_i \leftarrow (y-1)\left(1 - \frac{\lambda}{|c_1|^2|y-1|}\right)^{+} + 1,$
  $z_i \leftarrow \sum_l \left|\left(m^i_{k,l} - 1\right)_{\text{neg}}\right|$
end
$z \leftarrow \sum_i z_i$

slices can be estimated, e.g., using the noise variance estimation proposed by Mallat (2008, Eq. (11.85) on p. 565). The regularization parameter λ can then be weighted according to the noise level. Hence, the sensitivity increases whenever the noise variance decreases within a single spectrum.

Modified Algorithm with Spatial Awareness

The algorithm described in the previous section can be modified to allow spatial awareness. In the context of MALDI, this means the peak picking process depends on neighboring spectra. The possibility of a peak being detected is larger if the spectra of neighboring spots also contain peaks at approximately the same m/z ratio. Whereas a peak surrounded by noise in the spatial neighborhood might be more likely to be ignored.

The spatial awareness can be included in the closed form solution of the mask m similar to the Windowed Group LASSO approach in (4.2.14), with the remaining algorithm as described

in Algorithm 10. To formulate the spatial awareness mathematically, let for every spectrum the set $\mathcal{N}$ be its neighboring spectra, including the actual spectrum itself. Further, denote by w_j, $j \in \mathcal{N}$, a weight corresponding to each neighbor such that $\sum_{j\in\mathcal{N}} w_j = 1$. By defining

$$\tilde{y} = \sum_{j\in\mathcal{N}} w_j \frac{|c_{2,j}|}{|c_{1,j}|}. \tag{6.2.14}$$

and

$$\tilde{c}_1 = \sum_{j\in\mathcal{N}} w_j \, c_{1,j}, \tag{6.2.15}$$

based on transform coefficients c_1 and c_2 of neighboring spectra, the estimation of the mask m can be formulated as

$$m = \left(\frac{|c_2|}{|c_1|} - 1\right)\left(1 - \frac{\lambda}{|\tilde{c}_1|^2\,|\tilde{y} - 1|}\right)^+ + 1. \tag{6.2.16}$$

This scales the regularization parameter λ for each coefficient depending on the characteristics of neighboring spectra. The weights w can, for example, be a simple average kernel, where each element is defined by $w_j = \frac{1}{|\mathcal{N}|} \forall\, j \in \mathcal{N}$ and $|\mathcal{N}|$ denotes the cardinality of $\mathcal{N}$. Other choices of weights could include Gaussian kernels with different variances or circular average filters. Non-linear approaches such as median filters (Lim, 1990, Ch. 8.2.2) or edge-preserving Kuwahara filters (Bartyzel, 2016; Kuwahara et al., 1976) are also possible.

6.3 Results Simulated Data

6.3.1 Data Set

The performance of the proposed peak picking algorithm is evaluated based on the same simulated data set, which is used by Wijetunge et al. (2015). This data set is introduced by Coombes et al. (2005a) and is based on the physical principles of time-of-flight mass spectrometry, emulating characteristics of real MALDI-TOF data. In total, the data set consists of 2500 individual spectra with annotated peak locations each having a length of 15,000 up to 30,000 samples. Each spectrum is independent, which implies that the spatial awareness approach can not be utilized in the simulated data set. For more detailed information on the model and computation of the simulated data set it is referred to (Coombes et al., 2005a; Morris et al., 2005).

6.3.2 Performance Measures

The following performance measures for evaluating the proposed algorithm are essentially identical with the ones in (5.3.3). However, the naming convention is adapted to be consistent with the measures mentioned by Wijetunge et al. (2015) and Yang et al. (2009), allowing a direct comparability of results. For a single spectrum, the Sensitivity is defined as

$$\text{Sensitivity} = \frac{\text{Number of correctly identified peaks}}{\text{Number of reference peaks}}. \tag{6.3.1}$$

The False Discovery Rate (FDR) is defined as

$$\text{FDR} = \frac{\text{Number of falsely identified peaks}}{\text{Number of total peaks detected}}. \tag{6.3.2}$$

Ideally, an optimal peak picking algorithm has a Sensitivity of 1 and a FDR of 0. The larger the Sensitivity and at the same time the lower the FDR, the better the performance of the algorithm. Both values can be combined into a single performance measure, denoted by F1-score (Wijetunge et al., 2015; Yang et al., 2009), defined by

$$\text{F1-score} = \frac{2 \cdot (1 - \text{FDR}) \cdot \text{Sensitivity}}{1 - \text{FDR} + \text{Sensitivity}}. \tag{6.3.3}$$

This gives a single performance value, taking the Sensitivity as well as the FDR into account. In the following numerical evaluation, a peak is classified as correctly identified, if it is within 1% of the expected m/z value as proposed by Wijetunge et al. (2015) and Yang et al. (2009).

6.3.3 Parameter Settings

The following parameters for the peak picking algorithm are proposed. The length of slices is set to 60 samples with an overlap of 0.5. The frame is set to be either a Gabor or a wavelet frame as defined in Chapter 3, resulting in two distinct algorithms. The Gabor frame is based on a Hann window with a width of 20 samples and a time- and frequency-sampling step size of 1 each, i.e., $a = 1$ and $b = 1$. The wavelet frame is based on the concept introduced in Section 3.4. The frame parameters have been chosen, such that the lower frequency range is oppressed in order to reduce baseline effects. The parameters $f_{\min}$ and bw are set to 1000 Hz each and the number of bins is set to 30. The generating waveform is the Equalizer defined in (3.1.5). The regularization parameter λ controls the number of detected peaks and is chosen such that the number of detected spikes equals the number of inserted spikes.

Table 6.1: Performance of the proposed peak picking algorithms, showing mean and standard error over all 2500 spectra.

	Sensitivity (%)	FDR (%)	F1-score (%)
Wijetunge et al. (2015)	86.86 ± 0.18	28.89 ± 0.22	77.29 ± 0.14
Proposed algorithm (Gabor)	87.89 ± 0.08	20.70 ± 0.12	83.23 ± 0.09
- with baseline removed	94.26 ± 0.07	12.20 ± 0.14	90.76 ± 0.09
Proposed algorithm (Wavelet)	92.13 ± 0.07	19.81 ± 0.14	85.59 ± 0.10
- with baseline removed	92.14 ± 0.07	19.65 ± 0.14	85.70 ± 0.10

6.3.4 Results

The performance of the newly proposed algorithm is compared to the recently introduced peak picking algorithm by Wijetunge et al. (2015). The algorithms are applied to all 2500 spectra and the resulting mean Sensitivity, FDR and F1-score can be seen in Table 6.1. The proposed algorithm is once evaluated without baseline correction and once with baseline correction using the MATLAB routine *msbackadj* (Andrade and Manolakos, 2003). The results presented by Wijetunge et al. (2015, Table 3) are reproducible, but are outperformed by the proposed peak picking algorithm even with no baseline removed: the sensitivity is higher with a lower false discovery rate. Using wavelet frames, the influence of the baseline can be completely disregarded by an appropriate frame parametrization. The sensitivity of the proposed algorithm yields the best performance at the lowest FDR for Gabor frames and a prior baseline removal.

The run time of Wijetunge's algorithm is approximately 60 seconds for a single spectrum on a 2.9 GHz i7 QuadCore processor, resulting in a computation time of more than 41 hours for all 2500 spectra. The proposed algorithm, however, takes roughly 5 minutes to process all 2500 spectra in a vectorized implementation of Algorithm 10.

6.4 Results Real Data

It is almost impossible to verify the performance of the proposed peak picking algorithm on real MALDI data, since peak locations are generally not known. Nonetheless, it is possible to apply the peak picking approach to MALDI data sets resulting in two possible applications: peak picking and denoising. Treating the output z primarily as an indicator variable for possible

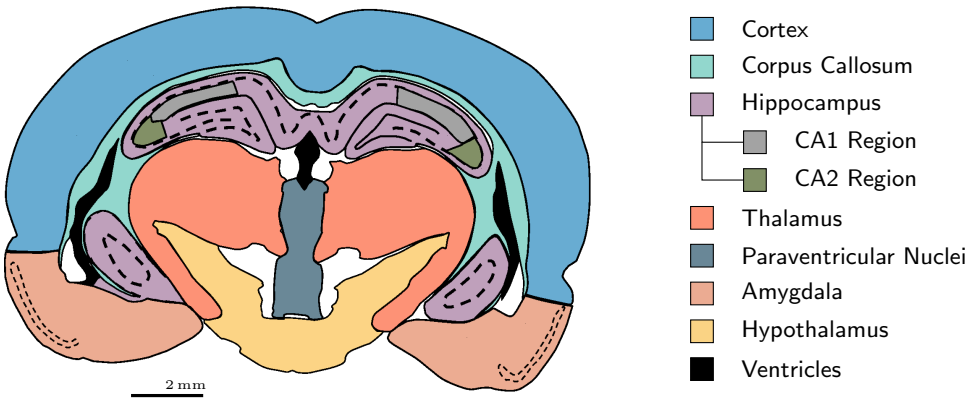


Figure 6.2: Basic anatomical annotation of the rat brain data set (adapted from Alexandrov et al., 2010, Fig. 5C).

peak locations leads to a classical peak picking application (similar as described in Chapter 5). Considering the output z, on the other hand, as actual 'spectra' itself results in a strongly denoised MALDI spectrum. It is important to note, that in this case, peak intensities in z do not have any relation to peak intensities in the original spectrum and the term 'spectra' is only used to indicate the intended usage. Despite this drawback, it might be beneficial for revealing certain structures in m/z images, which might be challenging to detect otherwise.

6.4.1 Data Sets

Linear TOF - Coronal Rat Brain Data Set

The rat brain data set has been used in quite a few publications (Alexandrov, 2012; Alexandrov and Bartels, 2013; Alexandrov and Kobarg, 2011; Alexandrov et al., 2010; Jones et al., 2012). A 10 µm tissue section of a rat brain was prepared for MALDI imaging using sinapinic acid as a matrix. With a lateral resolution of 80 µm 20,185 spectra were acquired, each containing 6,618 m/z bins. The spectra are preprocessed by removing the baseline using a top hat filter and TIC normalization before applying the peak picking algorithm. No spectra smoothing has been applied. Figure 6.2 shows the schematically annotated rat brain atlas.

Reflector TOF - FFPE Lung Data Set

MALDI imaging based on formalin-fixed and paraffin-embedded (FFPE) tissue samples is gaining increased interest in pathological applications (Aichler and Walch, 2015; Buck et al.,

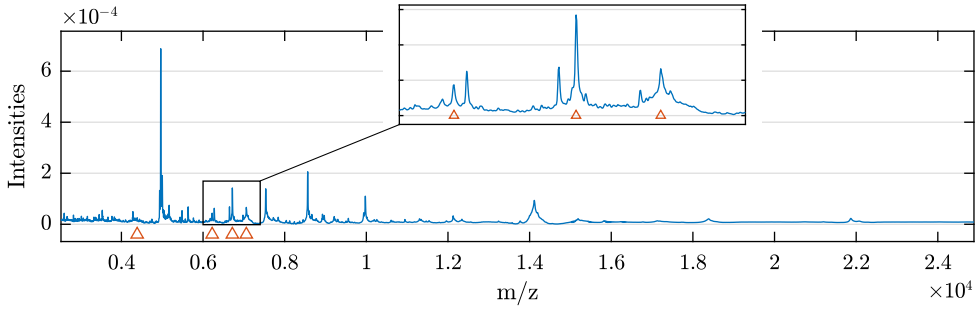
2015; Ly et al., 2016; Norris and Caprioli, 2013). Sample preparation and acquisition of the following FFPE data set is similar as described in more detail by Boskamp et al. (2017). Briefly summarized, MALDI imaging data of a human lung FFPE tissue sample was obtained in positive ion reflector mode on a MALDI-TOF instrument by Bruker Daltonics (Autoflex Speed). The data set consists of 3567 spectra, each containing 20992 samples in the mass range of 700 - 4000 m/z. The baseline has been removed prior to TIC normalization. Unfortunately, the lung data set is not annotated.

6.4.2 Results Coronal Rat Brain Data Set

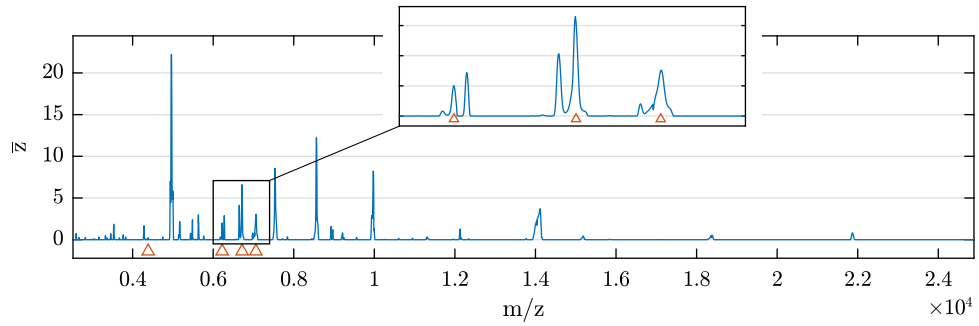
The peak picking approach based on Algorithm 10 is applied to the entire rat brain data set with the following input parameters: the slice length and overlap remain as previously defined ($\Delta = 60$ and $o = 0.5$. The Gabor frame is based on a Hann window of width 15 samples. The regularization parameter is fixed to $\lambda = 1.5\text{e}{-3}$.

The resulting sparse output z then contains only nonzero values where a peak is detected. The mean spectrum $\overline{z}$ over all spots is depicted in Fig. 6.3. Hereby, Figure 6.3a shows the mean spectrum of the original rat brain data set. In comparison, the mean spectrum after peak picking is visualized below in Fig. 6.3b. It shows the same prominent features as the original data set. However, only 48% of all m/z images contain non-trivial coefficients. This means, that for the remaining 52% of m/z values no peak is detected in any of the 20,185 spots. Clearly, this procedure is sensitive to the choice of the regularization parameter λ. Smaller values increase the sensitivity, resulting in more detected peaks per single spectrum. Larger values, on the other hand, increase the denoising effect by choosing less peaks. Recall from (6.2.13) that z, and hence the mean spectrum of z, does not reflect original intensities anymore, but rather significant changes between Gabor coefficients of two consecutive slices. This means the larger a peak in z, the larger the difference between two slices. Despite this drawback, it might still be useful to analyze m/z images after peak picking.

In the following, let the original approach as summarized in Algorithm 10 be denoted as the *basic* approach. The modified approach proposed in (6.2.16) is based on an average filter for a 3×3 neighborhood $\mathcal{N}$. Hence, the filter coefficients are $w_j = \frac{1}{9}$ for all $j \in \mathcal{N}$. The *basic* as well as the *average* approach are applied to the rat brain data set with the same parameter settings as previously defined. Four selected m/z images are shown in Figure 6.4, where corresponding m/z-values are indicated with a red triangle in the mean spectrum of Fig. 6.3. The denoising effect of the proposed algorithm is clearly visible when comparing raw data with



(a) Original mean spectrum



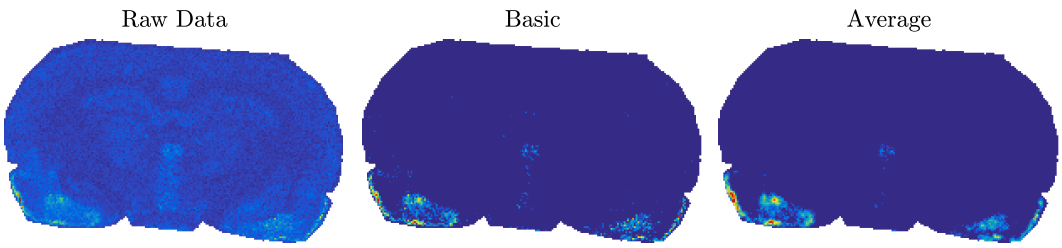
(b) Mean spectrum after peak picking

Figure 6.3: Mean spectra of the original rat brain data set and after peak picking. Red triangles indicate m/z locations for Fig. 6.4

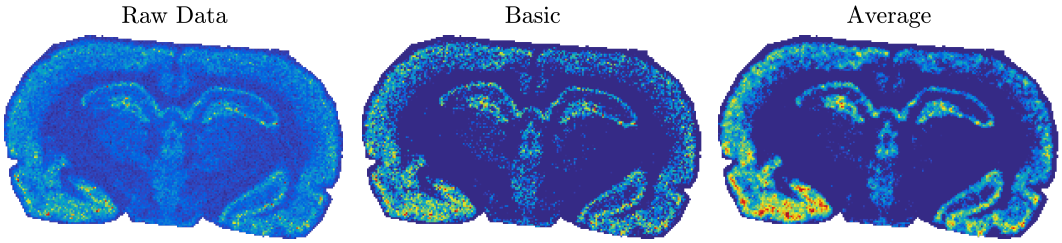
the basic or average approach. Additionally, the neighborhood based approach smooths peak areas in m/z images, while preserving edges. This can be nicely seen, for example, in Figures 6.4c and 6.4d. The sensitivity of the proposed peak picking approach is large enough to also detect low intensity peaks. Alexandrov and Bartels (2013) showed that the low intensity peak at $m/z = 4385.9$, depicted in Fig. 6.4a, is not detected by other spectrum-wise peak picking approaches. Nonetheless, detecting peaks which are present in a small number of spectra is substantial in order to distinguish between matrix and actual protein peaks (Alexandrov and Bartels, 2013, Fig. 5).

Detection of Overlapping Peaks

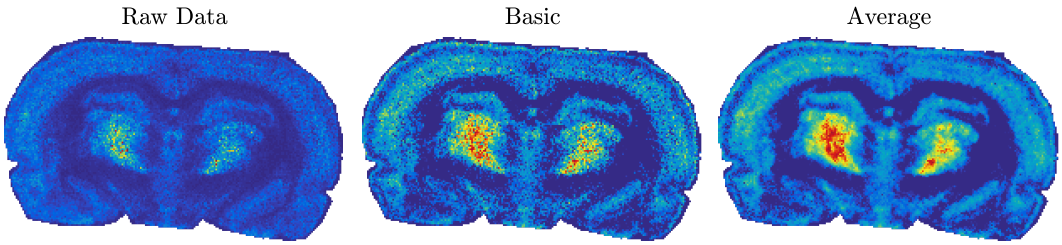
The detection of overlapping peaks is crucial whenever the sampling rate is low and isotopes are not clearly separated. Figure 6.5 shows part of a rat brain spectrum with overlapping peaks and



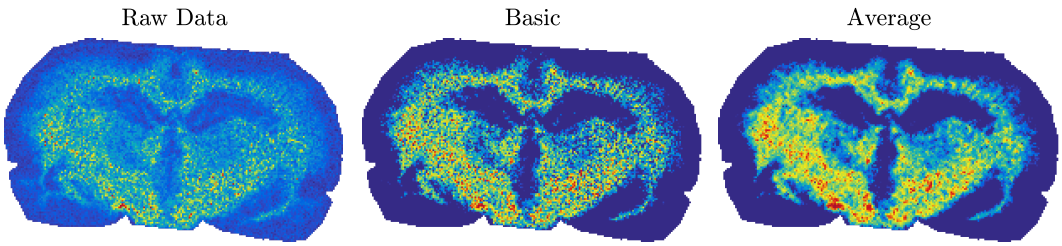
(a) $m/z = 4385.8525$



(b) $m/z = 6223.2627$



(c) $m/z = 6717.3877$



(d) $m/z = 7060.0215$

Figure 6.4: Four m/z images of the rat brain data set showing the raw data, data after basic peak picking and data after spatially aware peak picking using an average filter.

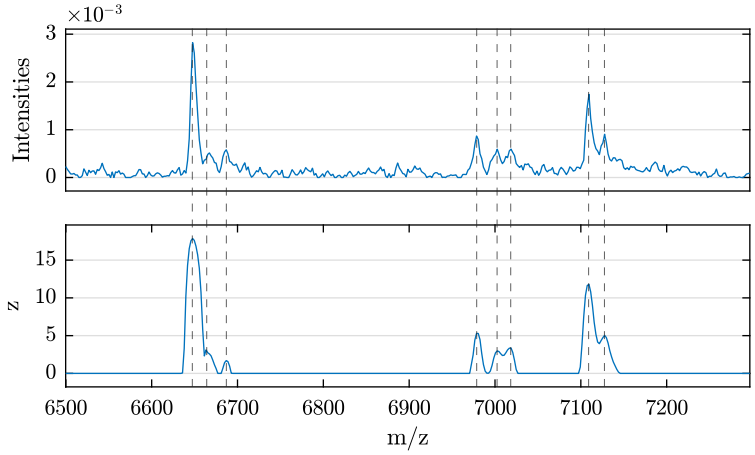


Figure 6.5: Detection of overlapping peaks in the rat brain data set. Local maxima in z are indicated by dashed lines.

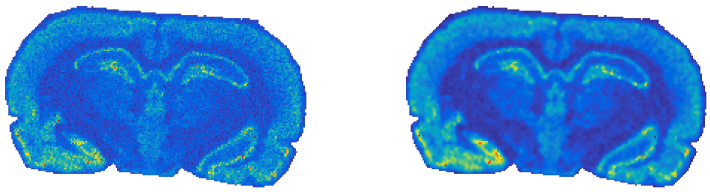
the output of the proposed peak picking algorithm z. Local maxima in z indicate the correct positions of overlapping peaks in the original spectrum.

Whenever the proposed algorithm is based on Gabor frames, the capability of separating overlapping peaks strongly depends on the chosen window width of the frame itself. If the window function is too wide, two overlapping peaks might be detected as a single peak. On the other hand, narrower windows might fail to correctly detect wider peaks in the spectrum. One solution to overcome this shortcoming is to use so called multi-window Gabor frames (Zeevi et al., 1998). Such frames consist of multiple single frames, where the window width varies from frame to frame. This allows the combination of estimated masks from different frame multipliers in order to improve the detection of overlapping peaks.

Using wavelet frames, on the other hand, the scaling of the window function simplifies the detection of overlapping peaks. Additional steps, however, have to be taken in order to correctly identify overlapping spikes from noise. If a peak is detected, wavelet coefficients corresponding to scaling factors resulting in narrow wavelets reveal information whether this peak is a single peak or consists of overlapping peaks. This is similar as proposed in the previous Chapter (Song and Li, 2015).

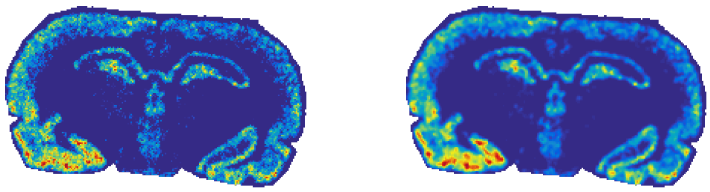
Comparison with Edge-Preserving Denoising

Kobarg (2014) showed that smoothing m/z images enhances segmentation maps. The hereby used bilateral filter preserves edges and local structures in m/z images. This bilateral smoothing



(a) Original

(b) Bilateral filtered



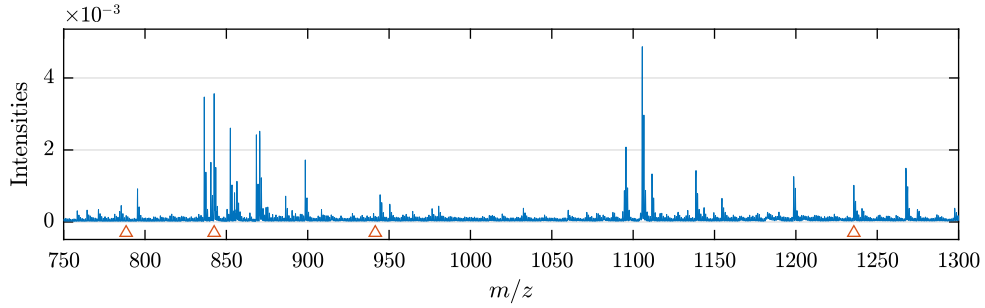
(c) Proposed algorithm

(d) Proposed algorithm bilateral filtered

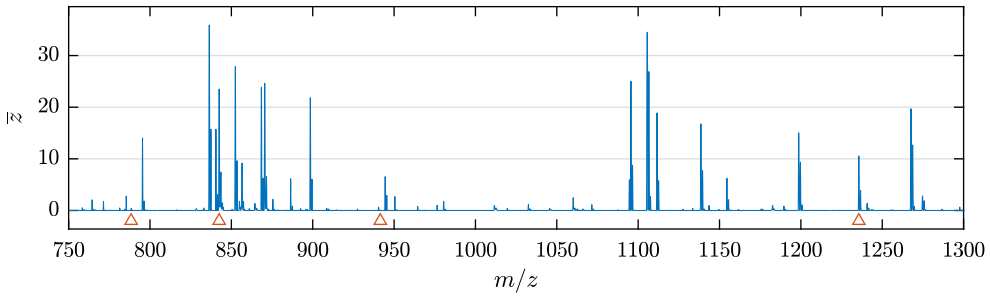
Figure 6.6: Comparison of bilateral filtering and the proposed algorithm based on the spatially aware approach with an average filter ($m/z = 6223.2627$).

filter has been introduced by Tomasi and Manduchi (1998) and can be summarized as follows. A Gaussian low-pass filter with standard deviation σ_b smooths pixels in a $w \times w$ neighborhood. Additionally, intensity differences of pixels inside this neighborhood are also weighted with a separate Gaussian filter with standard deviation τ_b. Hence, sharp transitions between intensities inside a neighborhood remain, whereas similar intensities are smoothed.

Figure 6.6 demonstrates the effects of a bilateral smoothing filter on the same data as shown by Kobarg (2014, Fig. 5.2) with similar parameters ($w = 3$, $\sigma_b = 1.5$ and $\tau_b = 5/16$). Compared to the original image, bilateral filtering spatially smooths the image while preserving local structures as illustrated in Figures 6.6a and 6.6b. Kobarg (2014) improved segmentation maps by clustering spectra with K-means based on the bilateral filtered rat brain data set. It can be assumed that using m/z images based on the proposed spatially aware peak picking method might further improve segmentation results. Comparing Figures 6.6b and 6.6c demonstrates the advantages of the proposed algorithm. Noise is removed, while preserving edges and local structures. The bilateral filtered image after spatially aware peak picking is shown in Figure 6.6d for illustrational purpose only.



(a) Original mean spectrum



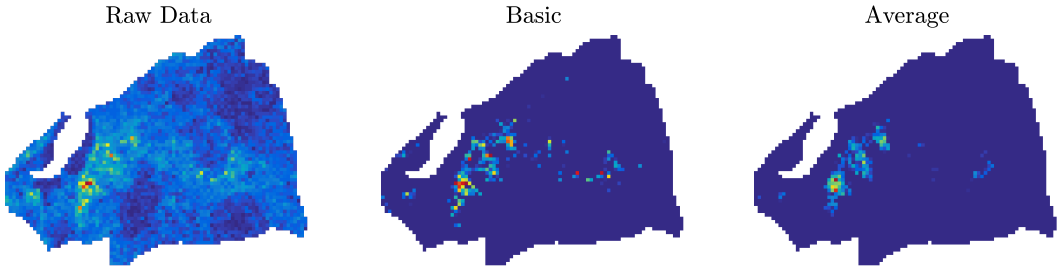
(b) Mean spectrum after peak picking.

Figure 6.7: Part of the mean spectrum of the lung data set before and after peak picking. Red triangles mark positions of m/z values in Figure 6.8.

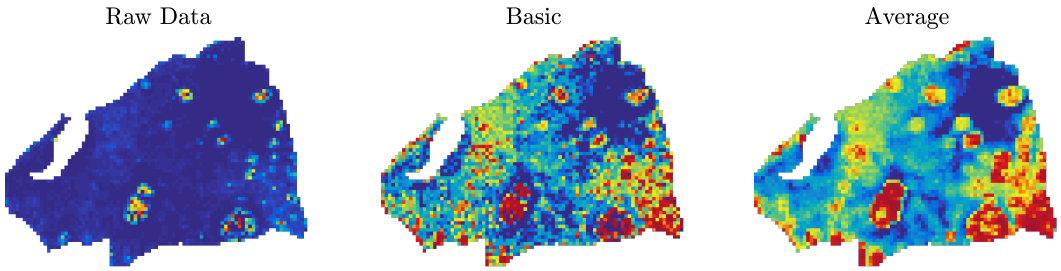
6.4.3 Results FFPE Lung Data Set

The proposed peak picking approach is applied to the lung data set with similar parameter settings as previously used: $\Delta = 60$, $o = 0.5$, a Gabor frame of length Δ based on a Hann window of length 8 and a regularization parameter $\lambda = 3\text{e}{-3}$. The mean spectrum of the resulting denoised data z retains only 15% all 20992 m/z images with non-zero information. A small section of the resulting mean spectrum is depicted in Figure 6.7, showing the similarity of the data set after peak picking with the original one. Again, note that peak intensities of both spectra are not correlated anymore. Nonetheless, considering m/z images based on denoised data z may reveal structures which would remain hidden in the original data set.

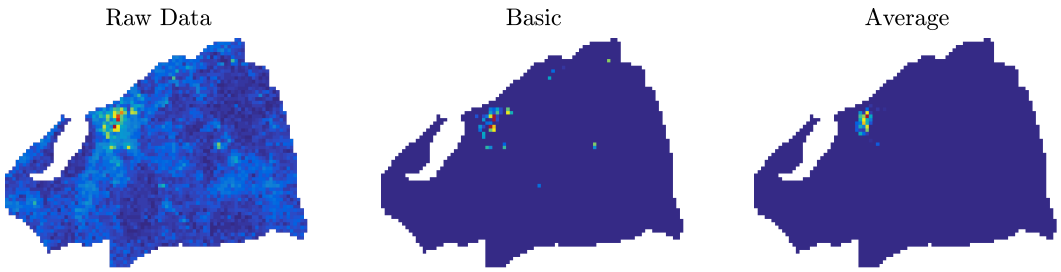
In order to show this, m/z images after basic and average peak picking are visualized in Figure 6.8. Corresponding mass-to-charge values are indicated in the mean spectrum in Fig. 6.7 by red triangles. Figures 6.8a and 6.8c show m/z images corresponding to small intensities in the mean spectrum, which are sparsely localized. On the other hand, Figures 6.8b and 6.8d



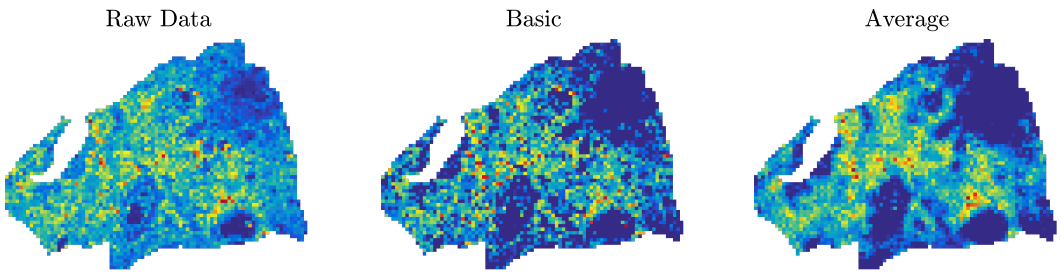
(a) $m/z = 788.3945$



(b) $m/z = 842.4932$



(c) $m/z = 941.5156$



(d) $m/z = 1235.6282$

Figure 6.8: Four m/z images of the lung data set showing the raw data, data after basic peak picking and data after spatially aware peak picking using an average filter.

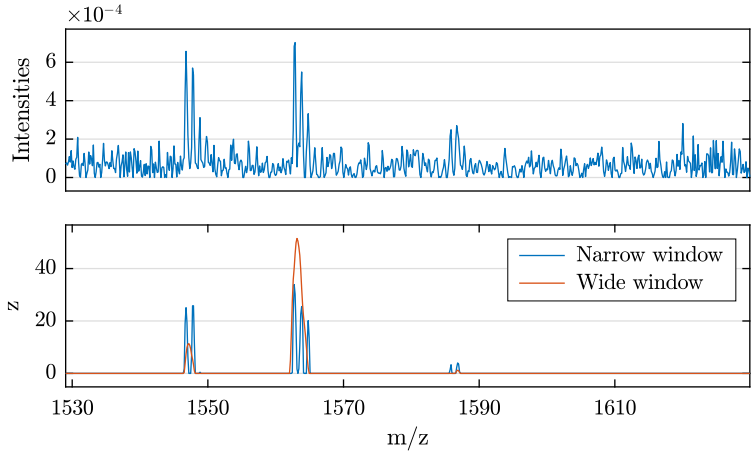


Figure 6.9: Detection of isotopic distributions or envelopes based on Gabor frames with different sized window functions.

reveal certain structures which are not readily visible in the original data, even with hotspot removal applied. So-called hotspots are removed by setting a certain percentage of the largest peak intensities to the lowest intensity among them. Especially the band structure of localized dots on the left hand side of the lung tissue in Fig. 6.8b becomes visible only when using the average based neighborhood approach.

Detection of Isotope Patterns

With the increased mass resolution of the reflector mode TOF, isotope patterns are well separated. The proposed approach based on Gabor frames with short windows is capable of detecting each isotopic peak separately. This is crucial for accurate protein identification (Spraggins et al., 2015). On the other hand, wider window functions can be used to find entire isotope patterns without resolving isotopic distributions, which may help to detect isotopic envelopes (Robinson et al., 2006). Both approaches are illustrated in Fig. 6.9 by showing part of an original spectrum and the same spectrum after applying the peak picking algorithm with a narrow and a wide window. It can be seen that even the isotope pattern which is buried in noise ($m/z \approx 1586$) can be reliably detected.

6.5 Conclusion

A novel spatially aware peak picking algorithm for MALDI imaging data has been introduced. It is based on sparse estimations of frame multiplier masks, detecting peaks spectra wise by measuring similarity between overlapping parts of a spectrum. A slight modification of the algorithm also allows for incorporating spatial information into the peak picking process. This combines the usual three preprocessing steps in Figure 6.1 into a single step, reducing computational complexity and simplifying parameter choices.

On simulated data the accuracy of the peak picking algorithm shows a significant increase, while at the same time reducing false discovery rates by more than 50% compared to the recently introduced algorithm by Wijetunge et al. (2015). Furthermore, the proposed algorithm is applied to two real MALDI-TOF data sets in order to highlight the advantages of including spatial information in the peak picking process. Although the denoised data does not correspond to original intensities anymore, the visualization can be advantageous in order to detect spatial patterns which would remain unnoticed otherwise. Hence, the algorithm can either be used as an actual peak picking approach indicating at peak locations or as a denoising approach exposing hidden peptide structures. This might be advantageous as a preprocessing step prior to segmentation or clustering algorithms (Kobarg, 2014).

Results from simulated data show, that baseline effects are successfully suppressed using the wavelet frame construction from Chapter 3. For Gabor frames this has not been observed. Using Gabor frames based on the DGTSF described in Algorithm 7 in the previous chapter might also lead to negligible baseline affects by suppressing lower frequency contributions. Additionally, emphasizing characteristic peak features in estimated masks as proposed in the TIFCO algorithm (Algorithm 8) might also further improve peak detection. Furthermore, estimating peak parameters such as peak area or peak width has not been addressed so far. Wijetunge et al. (2015) stated that the peak area is more important than actual peak intensity when estimating molecular abundances. As demonstrated by Zhang et al. (2015) these parameters can be easily extracted from corresponding wavelet coefficients. For Gabor frames such a quantization of peaks is still an open topic and leaves room for further improvement.

As the sampling of the m/z axis is not equidistant, the peak width changes with increasing mass-to-charge ratio. To overcome this problem, the proposed peak picking approach can be utilized with Gabor frames where the window size increases with corresponding m/z sampling distance, e.g., the multi window Gabor frames (Zeevi et al., 1998). For wavelet frames, a similar behavior can be realized by adapting discrete scales of the frame to corresponding m/z values.

With increasing mass resolution, for example using MALDI-Fourier transform ion cyclotron resonance (FT-ICR), the reliability to discriminate metabolites significantly improves to the range of millidaltons (Palmer et al., 2017). However, FT-ICR spectroscopy with high spectral resolution also results in larger noise (Liu and Xiao, 2014). Additionally, the data set size can reach up to 500 GB (Buck et al., 2015). The proposed algorithm is capable of reliably detecting spectral patterns in noisy data while at the same time reducing the size of large data sets, making it also a good approach for analyzing MALDI-FT-ICR data. To further compress data sets it is possible to retain coefficients only when the approximated mask differs from 1. For example, using $z^i = \left\| m^i - 1 \right\|_2$ instead of (6.2.13) results in a reduction of spectral resolution depending on the size of the slice length Δ. Instead of extracting the exact location of a spike in the mask, z_i then just gives a non-trivial value whenever a spike is present in the current mask m^i. The compressed data set z^i allows an initial examination of main global features of the data set, while keeping details hidden in the mask of the frame multiplier for later usages.

Chapter 7

Conclusion

The main part of this thesis describes sophisticated signal processing applications in audio processing and life science which are based on sparse signal representations. These representations are mostly determined by nonstationary Gabor frames, a generalization of Gabor and wavelet frames to irregular and arbitrary sampling schemes. The theory of nonstationary Gabor frames is well established for compactly supported functions (Balazs et al., 2011; Holighaus, 2014), but lacks explicit construction schemes whenever underlying functions do not have compact support. Furthermore, the construction of tight frames with compactly supported functions was unflexible and confined to certain parameter settings. So initially the question was addressed how a non-compactly supported function, such as the minimizing function for the wavelet uncertainty principle in (3.1.5), can be used to construct frames and possibly dual frames.

In the first part of the thesis, conditions motivated by results from Dörfler and Matusiak (2014, 2015) have been derived for which the nonstationary Gabor frame operator tends towards the identity operator. This has lead to a characterization of nonstationary Gabor frames with non-compactly supported functions and corresponding dual frames. The main result can be summarized as follows: for non-compactly supported window functions with proper decay the nonstationary Gabor frame operator converges to the identity whenever corresponding frequency and time sampling is sufficiently dense. It has been further shown, that in a finite dimensional setting this leads to frames inducing reasonable redundancies and a diagonal frame operator. Furthermore, corresponding dual frames are well defined, although in general the dual frame of a nonstationary Gabor frame is not necessarily a nonstationary Gabor frame (Dörfler and Matusiak, 2015). The algorithms derived in this part allow the construction of approximately tight wavelet frames based on non-compactly supported functions in frequency domain such as

the Equalizer in (3.1.5). In the remaining part of this thesis these frame constructions have been exploited in signal processing applications.

In Chapter 4 different audio inpainting approaches have been evaluated based on various time-frequency representations in order to reconstruct missing audio samples. It has been shown, that reconstruction results favor adaptive nonstationary over conventional Gabor frame approaches, not only quantitatively but also audibly. Including additional information from surrounding time-frequency neighborhoods further reduce audible artifacts. The usage of wavelet frames with non-compactly supported functions such as the minimizing waveform, however, does not yield significantly better reconstructions. It just increases computational complexity, since analysis and synthesis operator are repeatedly applied in corresponding proximal splitting algorithms.

Two spike detection algorithms for extracellular recorded data have been introduced in Chapter 5 which are capable of detecting spikes even in prominent noise. Both algorithms are based on sparse representations of spike features and have been demonstrated to perform superior with respect to all current state-of-the art methods on simulated data. On real data, a novel approach has been proposed which links a well-established but error-prone spike detection algorithm with the sensitivity of one of the proposed algorithms. The sparse representation did not favor the wavelet frame construction described in Chapter 3, instead Gabor frames with fixed window sizes have lead to more reasonable results.

In Chapter 6 a different problem in life sciences applications is addressed: peak detection in MALDI imaging. A novel algorithm has been proposed based on Gabor or wavelet frame multipliers which leads to a better detection rate at a lower false discovery rate with respect to a recently introduced peak detection algorithm. The wavelet frame construction using the minimizing function can be directly applied and leads to negligible baseline effects. Additionally, including spatial information has shown to smooth corresponding m/z images edge-preservingly and might lead to a better detection of biomarker structures.

Retrospectively it appears, at least for audio inpainting and MALDI peak detection, that the performance is not dominated by localization properties of underlying function systems. This means, whether a compactly supported function or the uncertainty minimizer is used for constructing wavelet frames only leads to marginal differences. The same can be observed for Gabor frames: whether Gaussian windows, which minimize the corresponding uncertainty principle, significantly change the performance of algorithms based on sparse representations just because underlying function systems are optimally localized is not confirmed. Using Gaussians is supposed to ease the readability of spectrograms as side lobes disappear (Pfander, 2013). Hence, the choice of a window function is more biased towards computational complexity,

where compactly supported functions have a major advantage.

Whether other symmetries between Gabor and wavelet frames can be found is still an open issue. For example the estimation of optimal window functions by Feichtinger et al. (2012) is based on Gabor frame multipliers and corresponding eigenvectors. With the construction schemes derived in Chapter 3 and wavelet frame multipliers from Chapter 6, a similar approach for wavelet frames may lead to similar sets of eigenvectors as indicated by Lieb (2012).

The results presented in this thesis might also be applicable in other areas where signal information needs to be separated from noise. Such areas may include speech enhancement and recognition (Mak and Yu, 2014), anomaly detection in data mining approaches (Agrawal and Agrawal, 2015), peak detection in social media, e.g., Twitter data (Ranco et al., 2015) or wavelet analysis of big financial data (Sun et al., 2015). The MALDI frame multiplier approach might be suitable for all signals with time-varying noise characteristics. The framework presented for spike detection in MEA data, on the other hand, might be better suited for detecting anomalies in large static background noise.

Going from one-dimensional signal expansions to higher dimensional transforms, similar uncertainty principles also lead to minimizing waveforms, for example the minimizer corresponding to the shearlet transform (Levie et al., 2014, Fig. 4). Uncertainty principles of other higher dimensional signal transforms are stated by Stark and Sochen (2011) and might also lead to minimizing functions. Explicit frame constructions based on these functions, however, are still an open topic.

Bibliography

Aach, T., Flohr, T., Mertelmeier, T., Kötz, K., Beer, S., Botterweck, H., Ermert, H., Schaeffter, T., Gleich, B., Borgert, J., et al. (2014). *Medizinische Bildgebung*. De Gruyter.

Abramowitz, M. and Stegun, I. (2012). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Books on Mathematics. Dover Publications.

Adler, A., Emiya, V., Jafari, M., Elad, M., Gribonval, R., and Plumbley, M. D. (2012). "Audio inpainting". *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(3):922–932.

Agrawal, S. and Agrawal, J. (2015). "Survey on anomaly detection using data mining techniques". *Procedia Computer Science*, 60(Supplement C):708–713.

Aichler, M. and Walch, A. (2015). "MALDI imaging mass spectrometry: Current frontiers and perspectives in pathology research and practice". *Laboratory Investigation*, 95(4):422–431.

Aitchison, I. J. R., MacManus, D. A., and Snyder, T. M. (2004). "Understanding Heisenberg's 'magical' paper of july 1925: A new look at the calculational details". *American Journal of Physics*, 72(11):1370–1379.

Aldroubi, A. and Unser, M. (1996). *Wavelets in Medicine and Biology*. Taylor & Francis.

Alexandrov, T. (2012). "MALDI imaging mass spectrometry: Statistical data analysis and current computational challenges". *BMC Bioinformatics*, 13(16):S11.

Alexandrov, T. and Bartels, A. (2013). "Testing for presence of known and unknown molecules in imaging mass spectrometry". *Bioinformatics*, 29(18):2335–2342.

Alexandrov, T. and Kobarg, J. (2011). "Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering". *Bioinformatics*, 27(13):i230–i238.

Alexandrov, T., Decker, J., Mertens, B., Deelder, A. M., Tollenaar, R. A. E. M., Maaß, P., and Thiele, H. (2009). "Biomarker discovery in MALDI-TOF serum protein profiles using discrete wavelet transformation". *Bioinformatics*, 25(5):643–649.

Alexandrov, T., Becker, M., Deininger, S., Ernst, G., Wehder, L., Grasmair, M., von Eggeling, F., Thiele, H., and Maaß, P. (2010). "Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering". *Journal of Proteome Research*, 9(12):6535–6546.

Almeida, L. A. D., Wakin, M., and Sava, P. (2016). "Data denoising and interpolation using synthesis and analysis sparse regularization". In *SEG Technical Program Expanded Abstracts 2016*, pages 4112–4117. Society of Exploration Geophysicists.

Andrade, L. and Manolakos, E. S. (2003). "Signal background estimation and baseline correction algorithms for accurate DNA sequencing". *Journal of VLSI signal processing systems for signal, image and video technology*, 35(3):229–243.

Antoniadis, A. (2007). "Wavelet methods in statistics: Some recent developments and their applications". *Statistics Surveys*, 1:16–55.

Antoniadis, A., Bigot, J., and Lambert-Lacroix, S. (2010). "Peaks detection and alignment for mass spectrometry data". *Journal de la Société Française de Statistique*, 151(1):17–37.

Auger, F., Flandrin, P., Lin, Y.-T., McLaughlin, S., Meignen, S., Oberlin, T., and Wu, H.-T. (2013). "Time-frequency reassignment and synchrosqueezing: An overview". *IEEE Signal Processing Magazine*, 30(6):32–41.

Azami, H. and Escudero, J. (2016). "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation". *Computer Methods and Programs in Biomedicine*, 128: 40–51.

Azami, H. and Sanei, S. (2014). "Spike detection approaches for noisy neuronal data: Assessment and comparison". *Neurocomputing*, 133:491–506.

Azami, H., Escudero, J., Darzi, A., and Sanei, S. (2015). "Extracellular spike detection from multiple electrode array using novel intelligent filter and ensemble fuzzy decision making". *Journal of Neuroscience Methods*, 239:129–138.

Balazs, P. (2007). "Basic definition and properties of bessel multipliers". *Journal of Mathematical Analysis and Applications*, 325(1):571–585.

Balazs, P., Laback, B., Eckel, G., and Deutsch, W. A. (2010). "Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(1):34–49.

Balazs, P., Dörfler, M., Jaillet, F., Holighaus, N., and Velasco, G. (2011). "Theory, implementation and applications of nonstationary Gabor frames". *Journal of Computational and Applied Mathematics*, 236(6):1481–1496.

Bartyzel, K. (2016). "Adaptive Kuwahara filter". *Signal, Image and Video Processing*, 10(4): 663–670.

Bauer, C., Cramer, R., and Schuchhardt, J. (2011). "Evaluation of peak-picking algorithms for protein mass spectrometry". In *Data Mining in Proteomics: From Standards to Applications*, pages 341–352. Humana Press, Totowa, NJ.

Bauschke, H. H. and Combettes, P. L. (2017). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer International Publishing.

Bayram, İ. and Akyıldız, Ö. D. (2014). "Primal-dual algorithms for audio decomposition using mixed norms". *Signal, Image and Video Processing*, 8(1):95–110.

Bayram, I. and Selesnick, I. W. (2009). "Frequency-domain design of overcomplete rational-dilation wavelet transforms". *IEEE Transactions on Signal Processing*, 57(8):2957–2972.

Beck, A. and Teboulle, M. (2009). "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". *SIAM Journal on Imaging Sciences*, 2(1):183–202.

Benedetto, J. J. and Zimmermann, G. (1997). "Sampling multipliers and the Poisson summation formula". *Journal of Fourier Analysis and Applications*, 3(5):505–523.

Bertalmío, M., Sapiro, G., Caselles, V., and Ballester, C. (2000). "Image inpainting". In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000, New Orleans, LA, USA, July 23-28, 2000*, pages 417–424.

Biffi, E., Ghezzi, D., Pedrocchi, A., and Ferrigno, G. (2010). "Development and validation of a spike detection and classification algorithm aimed at implementation on hardware devices". *Computational Intelligence and Neuroscience*, 2010:8.

Bilen, Ç., Ozerov, A., and Pérez, P. (2015a). "Joint audio inpainting and source separation". In *Latent Variable Analysis and Signal Separation*, pages 251–258. Springer International Publishing.

Bilen, Ç., Ozerov, A., and Pérez, P. (2015b). "Audio declipping via nonnegative matrix factorization". In *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–5.

Boskamp, T., Lachmund, D., Oetjen, J., Hernandez, Y. C., Trede, D., Maaß, P., Casadonte, R., Kriegsmann, J., Warth, A., Dienemann, H., Weichert, W., and Kriegsmann, M. (2017). "A new classification method for MALDI imaging mass spectrometry data acquired on formalin-fixed paraffin-embedded tissue samples". *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1865(7):916–926.

Bownik, M. and Lemvig, J. (2007). "The canonical and alternate duals of a wavelet frame". *Applied and Computational Harmonic Analysis*, 23(2):263–272.

Breen, E. J., Hopwood, F. G., Williams, K. L., and Wilkins, M. R. (2000). "Automatic Poisson peak harvesting for high throughput protein identification". *Electrophoresis*, 21(11):2243–2251.

Buck, A., Ly, A., Balluff, B., Sun, N., Gorzolka, K., Feuchtinger, A., Janssen, K.-P., Kuppen, P. J., van de Velde, C. J., Weirich, G., Erlmeier, F., Langer, R., Aubele, M., Zitzelsberger, H., Aichler, M., and Walch, A. (2015). "High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples". *The Journal of Pathology*, 237(1):123–132.

Castaneda, J., Heusdens, R., and Hendriks, R. (2011). "A generalized Poisson summation formula and its application to fast linear convolution.". *IEEE Signal Processing Letters*, 18 (9):501–504.

Chambolle, A. and Dossal, C. (2015). "On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm"". *Journal of Optimization Theory and Applications*, 166 (3):968–982.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). "Atomic decomposition by basis pursuit". *SIAM Journal on Scientific Computing*, 20(1):33–61.

Chiappalone, M., Vato, A., Tedesco, M., Marcoli, M., Davide, F., and Martinoia, S. (2003). "Networks of neurons coupled to microelectrode arrays: A neuronal sensory system for pharmacological applications". *Biosensors and Bioelectronics*, 18(5):627–634. Selected papers from the Seventh World Congress on Biosensors Kyoto, Japan 15-17 May 2002.

Choi, J. H. and Kim, T. (2002). "Neural action potential detector using multi-resolution teo". *Electronics Letters*, 38(12):541–543.

Choi, J. H., Jung, H. K., and Kim, T. (2006). "A new action potential detector using the mteo and its effects on spike sorting systems at low signal-to-noise ratios". *IEEE Transactions on Biomedical Engineering*, 53(4):738–746.

Christensen, O. (2006). "Pairs of dual Gabor frame generators with compact support and desired frequency localization". *Applied and Computational Harmonic Analysis*, 20(3):403–410.

Christensen, O. (2016). *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Springer International Publishing.

Christensen, O. and Laugesen, R. S. (2010). "Approximately dual frames in Hilbert spaces and applications to Gabor frames". *Sampling Theory in Signal and Image Processing*, 9(1-3): 77–89.

Chui, C. K. and Shi, X. (2000). "Orthonormal wavelets and tight frames with arbitrary real dilations". *Applied and Computational Harmonic Analysis*, 9(3):243–264.

Ciaramella, A., Gianfico, M., and Giunta, G. (2016). "Compressive sampling and adaptive dictionary learning for the packet loss recovery in audio multimedia streaming". *Multimedia Tools and Applications*, 75(24):17375–17392.

Cleju, N., Jafari, M. G., and Plumbley, M. D. (2012). "Choosing analysis or synthesis recovery for sparse reconstruction". In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 869–873.

Combettes, P. L. and Pesquet, J. C. (2007). "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery". *IEEE Journal of Selected Topics in Signal Processing*, 1 (4):564–574.

Combettes, P. L. and Pesquet, J.-C. (2011). "Proximal splitting methods in signal processing". In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer New York, New York, NY.

Coombes, K. R., Koomen, J. M., Baggerly, K. A., Morris, J. S., and Kobayashi, R. (2005a). "Understanding the characteristics of mass spectrometry data through the use of simulation". *Cancer Informatics*, 1(1):41–52.

Coombes, K. R., Tsavachidis, S., Morris, J. S., Baggerly, K. A., Hung, M. C., and Kuerer, H. M. (2005b). "Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform". *Proteomics*, 5:4107–4117.

Creusere, C. D., Kallakuri, K. D., and Vanam, R. (2008). "An objective metric of human subjective audio quality optimized for a wide range of audio fidelities". *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1):129–136.

Dahlke, S. and Maaß, P. (1995). "The affine uncertainty principle in one and two dimensions". *Computers & Mathematics with Applications*, 30(3):293–305.

Dahlke, S., Kutyniok, G., Maaß, P., Sagiv, C., Stark, H.-G., and Teschke, G. (2008). "The uncertainty principle associated with the continuous shearlet transform". *International Journal of Wavelets, Multiresolution and Information Processing*, 06(02):157–181.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Daubechies, I. and Han, B. (2002). "The canonical dual frame of a wavelet frame". *Applied and Computational Harmonic Analysis*, 12(3):269–285.

Daubechies, I., Grossmann, A., and Meyer, Y. (1986). "Painless nonorthogonal expansions". *Journal of Mathematical Physics*, 27(5):1271–1283.

Daudet, L., Molla, S., and Torrésani, B. (2001). "Transient detection and encoding using wavelet coefficient trees". In *18 Colloque sur le traitement du signal et des images, FRA, 2001*, pages 321–324. GRETSI, Groupe d´Etudes du Traitement du Signal et des Images.

Daus, A. W., Goldhammer, M., Layer, P. G., and Thielemann, C. (2011). "Electromagnetic exposure of scaffold-free three-dimensional cell culture systems.". *Bioelectromagnetics*, 32 (5):351–359.

Daus, A. W., Layer, P. G., and Thielemann, C. (2012). "A spheroid-based biosensor for the label-free detection of drug-induced field potential alterations". *Sensors and Actuators B: Chemical*, 165(1):53–58.

de Noo, M. E., Mertens, B. J., Ozalp, A., Bladergroen, M. R., van der Werff, M. P., van de Velde, C. J., Deelder, A. M., and Tollenaar, R. A. (2006). "Detection of colorectal cancer using MALDI-TOF serum protein profiling". *European Journal of Cancer*, 42(8):1068–1076.

Deininger, S.-O., Ebert, M. P., Fütterer, A., Gerhard, M., and Röcken, C. (2008). "MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers". *Journal of Proteome Research*, 7(12):5230–5236.

Deininger, S.-O., Cornett, D. S., Paape, R., Becker, M., Pineau, C., Rauser, S., Walch, A., and Wolski, E. (2011). "Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations". *Analytical and Bioanalytical Chemistry*, 401(1):167–181.

Depalle, P., Kronland-Martinet, R., and Torrésani, B. (2007). "Time-Frequency multipliers for sound synthesis". In SPIE, editor, *SPIE annual Symposium Wavelet XII*, volume 6701 of *Proceedings of the SPIE*, pages 670118–1–670118–15, San Diego, United States.

Donoho, D. L. and Stark, P. B. (1989). "Uncertainty principles and signal recovery". *SIAM Journal on Applied Mathematics*, 49(3):906–931.

Dörfler, M. and Matusiak, E. (2013). "Tracing sound objects in audio textures". In *10th international conference on Sampling Theory and Applications (SampTA 2013)*, pages 408–411, Bremen, Germany.

Dörfler, M. and Matusiak, E. (2014). "Sparse Gabor multiplier estimation for identification of sound objects in texture sound". In *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers*, pages 443–462. Springer International Publishing, Cham.

Dörfler, M. and Matusiak, E. (2014). "Nonstationary Gabor frames - existence and construction". *International Journal of Wavelets, Multiresolution and Information Processing*, 12(3):1450032.

Dörfler, M. and Matusiak, E. (2015). "Nonstationary Gabor frames - approximately dual frames and reconstruction errors". *Advances in Computational Mathematics*, 41(2):293–316.

Dörfler, M., Kowalski, M., and Siedenburg, K. (2013). "Audio inpainting with social sparsity". In *Proceedings of Spars2013, Lausanne, Switzerland*.

Du, P., Kibbe, W. A., and Lin, S. M. (2006). "Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching". *Bioinformatics*, 22 (17):2059–2065.

El Hady, A., Afshar, G., Broeking, K., Schlüter, O., Geisel, T., Stühmer, W., and Wolf, F. (2013). "Optogenetic stimulation effectively enhances intrinsically generated network synchrony". *Frontiers in Neural Circuits*, 7:167.

Elad, M., Milanfar, P., and Rubinstein, R. (2007). "Analysis versus synthesis in signal priors". *Inverse Problems*, 23(3):947.

Etter, W. (1996). "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters". *IEEE Transactions on Signal Processing*, 44(5):1124–1135.

Fadili, M. J. and Starck, J. L. (2009). "Monotone operator splitting for optimization problems in sparse recovery". In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1461–1464.

Fawcett, T. (2006). "An introduction to ROC analysis". *Pattern Recognition Letters*, 27(8): 861–874.

Feichtinger, H. G. and Nowak, K. (2003). "A first survey of Gabor multipliers". In *Advances in Gabor Analysis*, pages 99–128. Birkhäuser Boston, Boston, MA.

Feichtinger, H. G. and Strohmer, T. (1998). *Gabor Analysis and Algorithms. Theory and Applications.* Birkhäuser, Boston.

Feichtinger, H. G. and Strohmer, T. (2003). *Advances in Gabor Analysis*. Birkhäuser, Basel.

Feichtinger, H. G., Onchis-Moaca, D., Ricaud, B., Torrésani, B., and Wiesmeyr, C. (2012). "A method for optimizing the ambiguity function concentration". In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 804–808.

Feichtinger, H. G., Onchis, D. M., and Wiesmeyr, C. (2014). "Construction of approximate dual wavelet frames". *Advances in Computational Mathematics*, 40(1):273–282.

Flandrin, P. (2001). "Inequalities in Mellin-Fourier signal analysis". In *Wavelet Transforms and Time-Frequency Signal Analysis*, pages 289–319. Birkhäuser Boston, Boston, MA.

Flensburg, J., Haid, D., Blomberg, J., Bielawski, J., and Ivansson, D. (2004). "Applications and performance of a MALDI-ToF mass spectrometer with quadratic field reflectron technology". *Journal of Biochemical and Biophysical Methods*, 60(3):319–334.

Folland, G. B. and Sitaram, A. (1997). "The uncertainty principle: A mathematical survey". *Journal of Fourier Analysis and Applications*, 3(3):207–238.

Forsythe, G., Malcolm, M., and Moler, C. (1977). *Computer methods for mathematical computations*. Prentice-Hall series in automatic computation. Prentice-Hall.

Franke, F., Jäckel, D., Dragas, J., Müller, J., Radivojevic, M., Bakkum, D., and Hierlemann, A. (2012). "High-density microelectrode array recordings and real-time spike sorting for closed-loop experiments: an emerging technology to study neural plasticity". *Frontiers in Neural Circuits*, 6:105.

Frega, M., Tedesco, M., Massobrio, P., Pesce, M., and Martinoia, S. (2014). "Network dynamics of 3D engineered neuronal cultures: A new experimental model for in-vitro electrophysiology". *Scientific Reports*, 4:5489.

Gemmeke, J., Hamme, H. V., Cranen, B., and Boves, L. (2010). "Compressive sensing for missing data imputation in noise robust speech recognition". *IEEE Journal of Selected Topics in Signal Processing*, 4(2):272–287.

Glasberg, B. R. and Moore, B. (1990). "Derivation of auditory filter shapes from notched-noise data". *Hearing Research*, 47:103–138.

Godsill, S. J. and Rayner, P. J. W. (1998). *Digital Audio Restoration.* Springer, Berlin.

Goodwin, R. J. A., Pennington, S. R., and Pitt, A. R. (2008). "Protein and peptides in pictures: Imaging with MALDI mass spectrometry". *Proteomics*, 8(18):3785–3800.

Gröchenig, K. (2001). *Foundations of Time-Frequency Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston.

Gross, J. and Beifuss, K. (2012). *Massenspektrometrie: Ein Lehrbuch*. Springer Berlin Heidelberg.

Grossmann, A., Morlet, J., and Paul, T. (1985). "Transforms associated to square integrable group representations. I. General results". *Journal of Mathematical Physics*, 26(10):2473–2479.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

Hayes, M. (1996). *Statistical digital signal processing and modeling*. John Wiley & Sons.

He, B. and Yuan, X. (2012). "On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method". *SIAM Journal on Numerical Analysis*, 50(2):700–709.

Holighaus, N. (2014). "Structure of nonstationary Gabor frames and their dual systems". *Applied and Computational Harmonic Analysis*, 37(3):442–463.

Holighaus, N., Dörfler, M., Velasco, G. A., and Grill, T. (2013). "A framework for invertible, real-time constant-Q transforms". *IEEE Transactions on Audio, Speech, and Language Processing*, 21(4):775–785.

Hondebrink, L., Verboven, A., Drega, W., Schmeink, S., de Groot, M., van Kleef, R., Wijnolts, F., de Groot, A., Meulenbelt, J., and Westerink, R. (2016). "Neurotoxicity screening of (illicit) drugs using novel methods for analysis of microelectrode array (mea) recordings". *NeuroToxicology*, 55:1–9.

House, L. L., Clyde, M. A., and Wolpert, R. L. (2011). "Bayesian nonparametric models for peak identification in MALDI-TOF mass spectroscopy". *The Annals of Applied Statistics*, 5 (2B):1488–1511.

Janssen, A. and Søndergaard, P. L. (2007). "Iterative algorithms to approximate canonical Gabor windows: Computational aspects". *Journal of Fourier Analysis and Applications*, 13 (2):211–241.

Janssen, A., Veldhuis, R., and Vries, L. (1986). "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(2):317–330.

Johnstone, A. F., Gross, G. W., Weiss, D. G., Schroeder, O. H.-U., Gramowski, A., and Shafer, T. J. (2010). "Microelectrode arrays: A physiologically based neurotoxicity testing platform for the 21st century". *NeuroToxicology*, 31(4):331–350.

Jones, E. A., Deininger, S.-O., Hogendoorn, P. C., Deelder, A. M., and McDonnell, L. A. (2012). "Imaging mass spectrometry statistical analysis". *Journal of Proteomics*, 75(16):4962–4989.

Kaiser, J. F. (1993). "Some useful properties of teager's energy operators". In *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 149–152.

Karas, M. and Hillenkamp, F. (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons". *Analytical Chemistry*, 60(20):2299–2301.

Karkare, V., Gibson, S., and Marković, D. (2013). "A 75-µW, 16-channel neural spike-sorting processor with unsupervised clustering". *IEEE Journal of Solid-State Circuits*, 48(9):2230–2238.

Kempka, M., Sjödahl, J., Björk, A., and Roeraade, J. (2004). "Improved method for peak picking in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry". *Rapid Communications in Mass Spectrometry*, 18(11):1208–1212.

Kereliuk, D. and Pasquier, P. (2013). "Audio interpolation and morphing via structured-sparse linear regression". In *Proceedings of the International Conference on Sound and Music Computing*, pages 546–552, Stockholm, Sweden.

Kim, S. and McNames, J. (2007). "Automatic spike detection based on adaptive template matching for extracellular neural recordings". *Journal of Neuroscience Methods*, 165(2): 165–174.

Kitić, S., Jacques, L., Madhu, N., Hopwood, M. P., Spriet, A., and Vleeschouwer, C. D. (2013). "Consistent iterative hard thresholding for signal declipping". In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5939–5943.

Kitić, S., Bertin, N., and Gribonval, R. (2015). "Sparsity and cosparsity for audio declipping: A flexible non-convex approach". In Vincent, E., Yeredor, A., Koldovský, Z., and Tichavský, P., editors, *Latent Variable Analysis and Signal Separation*, pages 243–250, Cham. Springer International Publishing.

Kobarg, J. H. (2014). "*Signal and image processing methods for imaging mass spectrometry data*". PhD thesis, University of Bremen.

Kowalski, M. (2009). "Sparse regression using mixed norms". *Applied and Computational Harmonic Analysis*, 27(3):303–324.

Kowalski, M. and Torrésani, B. (2009a). "Sparsity and persistence: Mixed norms provide simple signal models with dependent coefficients". *Signal, Image and Video Processing*, 3 (3):251–264.

Kowalski, M. and Torrésani, B. (2009b). "Structured sparsity: From mixed norms to structured shrinkage". In *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*.

Kowalski, M., Siedenburg, K., and Dörfler, M. (2013). "Social sparsity! Neighborhood systems enrich structured shrinkage operators". *IEEE Transactions on Signal Processing*, 61(10): 2498–2511.

Kronland-Martinet, R., Torrésani, B., and Olivero, A. (2010). "A new method for Gabor multiplier estimation: Application to sound morphing". *Proceedings of EUSIPCO 2010*.

Kuwahara, M., Hachimura, K., Eiho, S., and Kinoshita, M. (1976). "Processing of RI-Angiocardiographic images". In *Digital Processing of Biomedical Images*, pages 187–202. Springer US, Boston, MA.

Kwon, D., Vannucci, M., Song, J. J., Jeong, J., and Pfeiffer, R. M. (2008). "A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise". *Proteomics*, 8(15):3019–3029.

Lagrange, M., Marchand, S., and Rault, J.-B. (2005). "Long interpolation of audio signals using linear prediction in sinusoidal modeling". *Journal of the Audio Engineering Society*, 53:891–905.

Lange, E., Gropl, C., Reinert, K., Kohlbacher, O., and Hildebrandt, A. (2006). "High-accuracy peak picking of proteomics data using wavelet techniques". *Pacific Symposium Biocomputing*, pages 243–254.

Lantzberg, D., Lieb, F., Stark, H. G., Levie, R., and Sochen, N. (2012). "Uncertainty principles, minimum uncertainty samplings and translations". In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 799–803.

Levie, R. and Sochen, N. (2017a). "A wavelet plancherel theory with application to sparse continuous wavelet transform". *arXiv:1712.02770v3*.

Levie, R. and Sochen, N. (2017b). "Uncertainty principles and optimally sparse wavelet transforms". *arXiv:1707.04863v1*.

Levie, R., Stark, H.-G., Lieb, F., and Sochen, N. (2014). "Adjoint translation, adjoint observable and uncertainty principles". *Advances in Computational Mathematics*, 40(3):609–627.

Li, K. and Sun, W. (2012). "Convergence of wavelet frame operators as the sampling density tends to infinity". *Applied and Computational Harmonic Analysis*, 33(1):140–147.

Li, S. (1995). "On general frame decompositions". *Numerical Functional Analysis and Optimization*, 16(9-10):1181–1191.

Lieb, F. (2012). "Uncertainty principles and construction of minimizers in signal processing". Master's thesis, University of Applied Sciences Aschaffenburg.

Lieb, F. (2015). "Audio inpainting using m-frames". In *Current Trends in Analysis and Its Applications: Proceedings of the 9th ISAAC Congress, Kraków 2013*, pages 705–713. Springer International Publishing, Cham.

Lieb, F. and Stark, H.-G. (2018). "Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches". *Signal Processing*, 153:291 – 299.

Lieb, F., Stark, H.-G., and Thielemann, C. (2017). "A stationary wavelet transform and a time-frequency based spike detection algorithm for extracellular recorded data". *Journal of Neural Engineering*, 14(3):036013.

Lim, J. (1990). *Two-dimensional Signal and Image Processing*. Prentice-Hall international editions. Prentice Hall.

Liu, B. and Sun, W. (2009). "Inversion of the wavelet transform using Riemannian sums". *Applied and Computational Harmonic Analysis*, 27(3):289–302.

Liu, Q. and Wang, P. (2009). *Cell-based Biosensors: Principles and Applications*. Artech House Engineering in Medicine & Biology Series. Artech House.

Liu, Q. and Xiao, S. (2014). "Effects of spectral resolution and signal-to-noise ratio of hyperspectral sensors on retrieving atmospheric parameters". *Optics Letters*, 39(1):60–63.

Liu, X., Yang, X., and Zheng, N. (2012). "Automatic extracellular spike detection with piecewise optimal morphological filter". *Neurocomputing*, 79:132–139.

Liuni, M., Robel, A., Matusiak, E., Romito, M., and Rodet, X. (2013). "Automatic adaptation of the time-frequency resolution for sound analysis and re-synthesis". *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5):959–970.

Lonardoni, D., Marco, S. D., Amin, H., Maccione, A., Berdondini, L., and Nieus, T. (2015). "High-density MEA recordings unveil the dynamics of bursting events in cell cultures". In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3763–3766. IEEE.

Ly, A., Buck, A., Balluff, B., Sun, N., Gorzolka, K., Feuchtinger, A., Janssen, K.-P., Kuppen, P. J. K., van de Velde, C. J. H., Weirich, G., Erlmeier, F., Langer, R., Aubele, M., Zitzelsberger, H., McDonnell, L., Aichler, M., and Walch, A. (2016). "High-mass-resolution MALDI mass spectrometry imaging of metabolites from formalin-fixed paraffin-embedded tissue". *Nature Protocols*, 11(8):1428–1443.

Maaß, P., Sagiv, C., Sochen, N., and Stark, H.-G. (2010). "Do uncertainty minimizers attain minimal uncertainty?". *Journal of Fourier Analysis and Applications*, 16(3):448–469.

Maaß, P., Sagiv, C., Stark, H.-G., and Torrésani, B. (2014). "Signal representation, uncertainty principles and localization measures". *Advances in Computational Mathematics*, 40(3):597–607.

Maccione, A., Gandolfo, M., Massobrio, P., Novellino, A., Martinoia, S., and Chiappalone, M. (2009). "A novel algorithm for precise identification of spikes in extracellularly recorded neuronal signals". *Journal of Neuroscience Methods*, 177(1):241–249.

Mach, V. (2016). "*Inpainting of missing audio signal samples*". PhD thesis, Brno University of Technology.

Mach, V. and Ozdobinski, R. (2013). "Optimizing dictionary learning parameters for solving audio inpainting problem". *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, 2(1):39–44.

Maher, R. C. (1993). "A method for extrapolation of missing digital audio data". In *Audio Engineering Society Convention 95*.

Mak, M.-W. and Yu, H.-B. (2014). "A study of voice activity detection techniques for NIST speaker recognition evaluations". *Computer Speech & Language*, 28(1):295–313.

Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd edition.

Mallat, S. G. (1989). "A theory for multiresolution signal decomposition: The wavelet representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Martinez, J., Pedreira, C., Ison, M. J., and Quiroga, R. Q. (2009). "Realistic simulation of extracellular recordings". *Journal of Neuroscience Methods*, 184(2):285–293.

Martinoia, S., Bonzano, L., Chiappalone, M., Tedesco, M., Marcoli, M., and Maura, G. (2005). "In vitro cortical neuronal networks as a new high-sensitive system for biosensing applications". *Biosensors and Bioelectronics*, 20(10):2071–2078.

Mayer, M., Arrizabalaga, O., Lieb, F., Ciba, M., Ritter, S., and Thielemann, C. (2018). "Electrophysiological investigation of human embryonic stem cell derived neurospheres using a novel spike detection algorithm". *Biosensors and Bioelectronics*, 100(Supplement C):462–468.

McConnell, E. R., McClain, M. A., Ross, J., LeFew, W. R., and Shafer, T. J. (2012). "Evaluation of multi-well microelectrode arrays for neurotoxicity screening using a chemical training set". *NeuroToxicology*, 33(5):1048–1057.

Moreau, J. (1965). "Proximité et dualité dans un espace hilbertien". *Bulletin de la Société Mathématique de France*, 93:273–299.

Morris, J. S., Coombes, K. R., Koomen, J., Baggerly, K. A., and Kobayashi, R. (2005). "Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum". *Bioinformatics*, 21(9):1764–1775.

Nabar, N. and Rajgopal, K. (2009). "A wavelet based teager energy operator for spike detection in microelectrode array recordings". In *TENCON 2009 - 2009 IEEE Region 10 Conference*, pages 1–6.

Nason, G. P. and Silverman, B. W. (1995). "The stationary wavelet transform and some statistical applications". In *Wavelets & Statistics: Lecture Notes in Statistics*, pages 281–300. Springer-Verlag.

Natora, M., Franke, F., and Obermayer, K. (2010). "Spike detection in extracellular recordings by hybrid blind beamforming". In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4636–4641.

Necciari, T., Balazs, P., Holighaus, N., and Søndergaard, P. (2013). "The ERBlet transform: An auditory-based time-frequency representation with perfect reconstruction". In *Proceedings*

of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), pages 498–502.

Nenadic, Z. and Burdick, J. W. (2005). "Spike detection using the continuous wavelet transform". *IEEE Transactions on Biomedical Engineering*, 52(1):74–87.

Nick, C., Daus, A., Bestel, R., Goldhammer, M., Steger, F., and Thielemann, C. (2014). "Drcell - a software tool for the analysis of cell signals recorded with extracellular microelectrodes". *Signal processing: an international journal (SPIJ)*, 7(2):96–109.

Norris, J. L. and Caprioli, R. M. (2013). "Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research". *Chemical Reviews*, 113(4):2309–2342.

Obeid, I. and Wolf, P. D. (2004). "Evaluation of spike-detection algorithms for a brain-machine interface application". *IEEE Transactions on Biomedical Engineering*, 51(6):905–911.

Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J., and Frey, U. (2015). "Revealing neuronal function through microelectrode array recordings". *Frontiers in Neuroscience*, 8: 423.

Ofir, H., Malah, D., and Cohen, I. (2007). "Audio packet loss concealment in a combined MDCT-MDST domain". *IEEE Signal Processing Letters*, 14(12):1032–1035.

Ogrinc Potočnik, N., Porta, T., Becker, M., Heeren, R. M. A., and Ellis, S. R. (2015). "Use of advantageous, volatile matrices enabled by next-generation high-speed matrix-assisted laser desorption/ionization time-of-flight imaging employing a scanning laser beam". *Rapid Communications in Mass Spectrometry*, 29(23):2195–2203.

Olivero, A., Torrésani, B., and Kronland-Martinet, R. (2013). "A class of algorithms for time-frequency multiplier estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 21(8):1550–1559.

Ottosen, E. S. and Nielsen, M. (2017a). "A characterization of sparse nonstationary Gabor expansions". *Journal of Fourier Analysis and Applications*, pages 1–24.

Ottosen, E. S. and Nielsen, M. (2017b). "Nonlinear approximation with nonstationary Gabor frames". *arXiv:1704.07176*.

Oudre, L. (2015). "Automatic detection and removal of impulsive noise in audio signals". *Image Processing On Line*, 5:267–281.

Palmer, A., Phapale, P., Chernyavsky, I., Lavigne, R., Fay, D., Tarasov, A., Kovalev, V., Fuchser, J., Nikolenko, S., Pineau, C., Becker, M., and Alexandrov, T. (2017). "FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry". *Nature Methods*, 14 (1):57–60.

Perraudin, N., Shuman, D. I., Puy, G., and Vandergheynst, P. (2014). "UNLocBoX A matlab convex optimization toolbox using proximal splitting methods". *CoRR*, abs/1402.0779.

Pfander, G. E. (2013). "Gabor frames in finite dimensions". In Casazza, P. G. and Kutyniok, G., editors, *Finite Frames: Theory and Applications*, pages 193–239. Birkhäuser Boston, Boston.

Pissulla, M. (2016). "Weiterentwicklung einer toolbox zur detektion von spikes in MEA datensätzen". Bachelor's Thesis, University of Applied Sciences Aschaffenburg.

Plumbley, M. D., Blumensath, T., Daudet, L., Gribonval, R., and Daviesouou, M. E. (2010). "Sparse representations in audio and music: From coding to source separation". *Proceedings of the IEEE*, 98(6):995–1005.

Průša, Z., Søndergaard, P. L., Holighaus, N., Wiesmeyr, C., and Balazs, P. (2014). "The Large Time-Frequency Analysis Toolbox 2.0". In *Sound, Music, and Motion*, Lecture Notes in Computer Science, pages 419–442. Springer International Publishing.

Putnam, C. R. (1967). *Commutation Properties of Hilbert Space Operators and Related Topics*. Springer Berlin Heidelberg.

Qiu, S. and Feichtinger, H. G. (1995). "Discrete Gabor structures and optimal representation". *IEEE Transactions on Signal Processing*, 43(10):2258–2268.

Quiroga, R. Q. (2017). Homepage University of Leicester. *https://www2.le.ac.uk/centres/csn*. Accessed: 2017-08-21.

Quiroga, R. Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering". *Neural Computing*, 16(8):1661–1687.

Rajmic, P., Bartlová, H., Průša, Z., and Holighaus, N. (2015). "Acceleration of audio inpainting by support restriction". In *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 325–329.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). "The effects of twitter sentiment on stock price returns". *PLOS ONE*, 10(9):1–21.

Regalia, G., Coelli, S., Biffi, E., Ferrigno, G., and Pedrocchi, A. (2016). "A framework for the comparative assessment of neuronal spike sorting algorithms towards more accurate off-line and on-line microelectrode arrays data analysis". *Computational Intelligence and Neuroscience*, 2016:19.

Ricaud, B. and Torrésani, B. (2014). "A survey of uncertainty principles and some signal processing applications". *Advances in Computational Mathematics*, 40(3):629–650.

Ricaud, B., Stempfel, G., Torrésani, B., Wiesmeyr, C., Lachambre, H., and Onchis, D. (2014). "An optimally concentrated Gabor transform for localized time-frequency components". *Advances in Computational Mathematics*, 40(3):683–702.

Robinson, N. E., Zabrouskov, V., Zhang, J., Lampi, K. J., and Robinson, A. B. (2006). "Measurement of deamidation of intact proteins by isotopic envelope and mass defect with ion cyclotron resonance fourier transform mass spectrometry". *Rapid Communications in Mass Spectrometry*, 20(23):3535–3541.

Rossant, C., Kadir, S. N., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., Tolias, A. S., Solomon, S., Buzsaki, G., Carandini, M., and Harris, K. D. (2016). "Spike sorting for large, dense electrode arrays". *Nature Neuroscience*, 19(4):634–641.

Rudin, W. (1987). *Real and Complex Analysis*. Mathematics series. McGraw-Hill.

Sagiv, C., Sochen, N. A., and Zeevi, Y. Y. (2006). "The uncertainty principle: Group theoretic approach, possible minimizers and scale-space properties". *Journal of Mathematical Imaging and Vision*, 26(1):149–166.

Salmasi, M., Büttner, U., and Glasauer, S. (2016). "Fractal dimension analysis for spike detection in low SNR extracellular signals". *Journal of Neural Engineering*, 13(3).

Sattar, F., Cullis-Suzuki, S., and Jin, F. (2016). "The use of adaptive Erblet transform with application to ecological data". In *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 317–322.

Sauve, A. C. and Speed, T. P. (2004). "Normalization, baseline correction and alignment of high-throughput mass spectrometry data". *Proceedings Gensips*.

Schörkhuber, C., Klapuri, A., and Sontacchi, A. (2013). "Audio pitch shifting using the constant-Q transform". *Journal of the Audio Engineering Society*, 61(7/8):562–572.

Selesnick, I. W. and Figueiredo, M. A. T. (2009). "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors". In *Proceedings SPIE*, volume 7446, pages 74460D–74460D15.

Shahid, S., Walker, J., and Smith, L. S. (2010). "A new spike detection algorithm for extracellular neural recordings". *IEEE Transactions on Biomedical Engineering*, 57(4):853–866.

Shalchyan, V., Jensen, W., and Farina, D. (2012). "Spike detection and clustering with unsupervised wavelet optimization in extracellular neural recordings". *IEEE Transactions on Biomedical Engineering*, 59(9):2576–2585.

Shampine, L. (2008). "Vectorized adaptive quadrature in matlab". *Journal of Computational and Applied Mathematics*, 211(2):131–140.

Shin, H., Sampat, M. P., Koomen, J. M., and Markey, M. K. (2010). "Wavelet-based adaptive denoising and baseline correction for MALDI TOF MS". *OMICS: A Journal of Integrative Biology*, 14(3):283–295.

Siedenburg, K. (2012). "Persistent empirical Wiener estimation with adaptive threshold selection for audio denoising". In *Proceedings of the 9th Sound and Music Computing Conference, Copenhagen, Denmark*, pages 426–433.

Siedenburg, K. and Dörfler, M. (2011). "Structured sparsity for audio signals". In *Proceedings of DAFx11*, pages 23–28.

Siedenburg, K. and Dörfler, M. (2012). "Audio denoising by generalized time-frequency thresholding". In *Audio Engineering Society Conference: Applications of Time-Frequency Processing in Audio*.

Siedenburg, K. and Dörfler, M. (2013). "Persistent time-frequency shrinkage for audio denoising". *Journal of the Audio Engineering Society*, 61(1/2):29–38.

Siedenburg, K., Kowalski, M., and Dörfler, M. (2014). "Audio declipping with social sparsity". In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1577–1581.

Smaragdis, P., Raj, B., and Shashanka, M. (2009). "Missing data imputation for spectral audio signals". In *2009 IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6.

Smith, I., Silveirinha, V., Stein, J. L., de la Torre-Ubieta, L., Farrimond, J. A., Williamson, E. M., and Whalley, B. J. (2015). "Human neural stem cell-derived cultures in three-dimensional substrates form spontaneously functional neuronal networks". *Journal of Tissue Engineering and Regenerative Medicine*, 11(4):1022–1033.

Søndergaard, P. L. (2007). "*Finite discrete Gabor analysis*". PhD thesis, Technical Univerisity of Denmark.

Søndergaard, P. L. (2012). "Efficient algorithms for the discrete gabor transform with a long fir window". *Journal of Fourier Analysis and Applications*, 18(3):456–470.

Song, R.-Q. and Li, H.-G. (2015). "Overlapping spikes sorting using feature fusion". In *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pages 391–394.

Spraggins, J. M., Rizzo, D. G., Moore, J. L., Rose, K. L., Hammer, N. D., Skaar, E. P., and Caprioli, R. M. (2015). "MALDI FTICR IMS of intact proteins: Using mass accuracy to link protein images with proteomics data". *Journal of The American Society for Mass Spectrometry*, 26(6):974–985.

Starck, J., Murtagh, F., and Fadili, J. (2010). *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press.

Starck, J., Murtagh, F., and Fadili, J. (2015). *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambridge University Press.

Stark, H.-G. and Sochen, N. (2011). "Square integrable group representations and the uncertainty principle". *Journal of Fourier Analysis and Applications*, 17(5):916–931.

Strang, G. and Nguyen, T. (1996). *Wavelets and Filter Banks*. Wellesley-Cambridge Press.

Sun, C. S. and Markey, M. K. (2011). "Recent advances in computational analysis of mass spectrometry for proteomic profiling". *Journal of Mass Spectrometry*, 46(5):443–456.

Sun, E. W., Chen, Y.-T., and Yu, M.-T. (2015). "Generalized optimal wavelet decomposing algorithm for big financial data". *International Journal of Production Economics*, 165(Supplement C):194–214.

Sun, W. (2010). "Asymptotic properties of Gabor frame operators as sampling density tends to infinity". *Journal of Functional Analysis*, 258(3):913–932.

Takekawa, T., Isomura, Y., and Fukai, T. (2010). "Accurate spike sorting for multi-unit recordings". *European Journal of Neuroscience*, 31(2):263–272.

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

Tipler, P. A. and Mosca, G. (2015). *Physik für Wissenschaftler und Ingenieure*. Springer Spektrum, 7 edition.

Tomasi, C. and Manduchi, R. (1998). "Bilateral filtering for gray and color images". In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846.

Tukker, A., de Groot, M., Wijnolts, F., Kasteel, E., Hondebrink, L., and Westerink, R. (2016). "Is the time right for in vitro neurotoxicity testing using human iPSC-derived neurons?". *ALTEX - Alternatives to animal experimentation*, 33(3):261–271.

Wallace, D. J., zum Alten Borgloh, S. M., Astori, S., Yang, Y., Bausen, M., Kugler, S., Palmer, A. E., Tsien, R. Y., Sprengel, R., Kerr, J. N. D., Denk, W., and Hasan, M. T. (2008). "Single-spike detection in vitro and in vivo with a genetic Ca2+ sensor". *Nature Methods*, 5(9): 797–804.

Walnut, D. F. (1992). "Continuity properties of the Gabor frame operator". *Journal of Mathematical Analysis and Applications*, 165(2):479–504.

Wark, H. A. C., Sharma, R., Mathews, K. S., Fernandez, E., Yoo, J., Christensen, B., Tresco, P., Rieth, L., Solzbacher, F., Normann, R. A., and Tathireddy, P. (2013). "A new high-density (25 electrodes/mm2) penetrating microelectrode array for recording and stimulating sub-millimeter neuroanatomical structures". *Journal of Neural Engineering*, 10(4):045003.

Werther, T., Eldar, Y. C., and Subbanna, N. K. (2005). "Dual Gabor frames: Theory and computational aspects". *IEEE Transactions on Signal Processing*, 53(11):4147–4158.

Wijetunge, C. D., Saeed, I., Boughton, B. A., Roessner, U., and Halgamuge, S. K. (2015). "A new peak detection algorithm for MALDI mass spectrometry data based on a modified asymmetric pseudo-voigt model". *BMC Genomics*, 16(12):S12.

Wu, L.-C., Chen, H.-H., Horng, J.-T., Lin, C., Huang, N. E., Cheng, Y.-C., and Cheng, K.-F. (2010). "A novel preprocessing method using hilbert huang transform for MALDI-TOF and SELDI-TOF mass spectrometry data". *PLOS ONE*, 5(8):1–15.

Yang, C., He, Z., and Yu, W. (2009). "Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis". *BMC Bioinformatics*, 10.

You, J., Reiter, U., Hannuksela, M. M., Gabbouj, M., and Perkis, A. (2010). "Perceptual-based quality assessment for audio-visual services: A survey". *Signal Processing: Image Communication*, 25(7):482–501.

Yu, G., Mallat, S., and Bacry, E. (2008). "Audio denoising by time-frequency block thresholding". *IEEE Transactions on Signal Processing*, 56(5):1830–1839.

Zeevi, Y. Y., Zibulski, M., and Porat, M. (1998). "Multi-window Gabor schemes in signal and image representations". In *Gabor Analysis and Algorithms: Theory and Applications*, pages 381–407. Birkhäuser Boston, Boston, MA.

Zhang, Z.-M., Tong, X., Peng, Y., Ma, P., Zhang, M.-J., Lu, H.-M., Chen, X.-Q., and Liang, Y.-Z. (2015). "Multiscale peak detection in wavelet space". *Analyst*, 140:7955–7964.

List of Figures

List of Algorithms

List of Tables