Estimation and Classification of Non-Stationary Processes

Applications in Time-Frequency Analysis

Brynolfsson, Johan

2019

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Total number of authors:
1

ESTIMATION AND CLASSIFICATION OF NON-STATIONARY PROCESSES

- APPLICATIONS IN TIME-FREQUENCY ANALYSIS -

JOHAN BRYNOLFSSON

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Contents

Contents

List of papers

This thesis is based on the following papers:

A Johan Brynolfsson and Maria Sandsten, "Parameter Estimation of Oscillating Gaussian Functions Using the Scaled Reassigned Spectrogram". *Elsevier Signal Processing*, vol. 150, pp. 20-32, September 2018

B Johan Brynolfsson, Isabella Reinhold, and Maria Sandsten, "A Time-Frequency-Shift Invariant Parameter Estimator for Oscillating Transient Functions Using the Matched Window Reassignment". To be submitted.

C Johan Brynolfsson, Johan Swärd, Andreas Jakobsson, and Maria Sandsten "Least Squares Estimation of Smooth and Mixed Spectra". To be submitted.

D Johan Swärd, Johan Brynolfsson, Andreas Jakobsson, and Maria Hansson-Sandsten, "Sparse Semi-Parametric Estimation of Harmonic Chirp Signals". *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1798-1807, 2016

E Johan Brynolfsson and Maria Sandsten, "Classification of One-Dimensional Non-Stationary Signals Using the Wigner-Ville Distribution in Convolutional Neural Networks", *25th European Signal Processing Conference*, Kos, Greece, 2017, August 28 - September 2, 2017

Additional papers not included in the thesis:

1. Johan Brynolfsson, Isabella Reinhold, Josefin Starkhammar, and Maria Sandsten, "The Matched Reassignment Applied to Echolocation Data", *44th International Conference on Acoustics, Speech, and Signal Processing*, Brighton, UK, May 12-17, 2019.

2. Johan Brynolfsson, Johan Swärd, Andreas Jakobsson, and Maria Sandsten, "Least Squares and Maximum Likelihood Estimation of Mixed Spectra", *26th European Signal Processing Conference*, Rome, Italy, September 3-7, 2018

3. Maria Sandsten, Johan Brynolfsson, and Isabella Reinhold, "The Matched Window Reassignment", *26th European Signal Processing Conference*, Rome, Italy, September 3-7, 2018

4. Maria Sandsten and Johan Brynolfsson, "Classification of Bird Song Syllables Using Wigner-Ville Ambiguity Function Cross-Terms", *25th European Signal Processing Conference*, Kos, Greece, 2017, August 28 - September 2, 2017

5. Christian Gustafsson, Fredrik Nordström, Emilia Persson, Johan Brynolfsson, and Lars E. Olsson, "Assessment of Dosimetric Impact of System Specific Geometry Distortion in an MRI Only Based Radiotherapy Workflow for Prostate", *Physics in Medicine and Biology*, vol. 62, no. 8, pp. 2976-2989, 2017

6. Johan Brynolfsson and Maria Hansson-Sandsten, "Parameter Estimation of Gaussian Functions using the Scaled Reassigned Spectrogram", *23rd European Signal Processing Conference*, Nice, France, August 31 - September 4, 2015

7. Maria Hansson-Sandsten and Johan Brynolfsson "The Scaled Reassigned Spectrogram with Perfect Localization for Estimation of Gaussian Functions", *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 100-104, 2015.

8. Johan Swärd, Johan Brynolfsson, Andreas Jakobsson, and Maria Hansson-Sandsten "Sparse Semi-Parametric Chirp Estimation", *The Asilomar Conference on Signals, Systems, and Computers*, Asilomar, USA, November 2-5, 2014.

9. Johan Brynolfsson and Maria Hansson-Sandsten "Multitaper Estimation of the Coherence Spectrum in Low SNR", *22nd European Signal Processing Conference*, Lisbon, Portugal, September 1-5, 2014.

10. Johan Swärd, Johan Brynolfsson, Andreas Jakobsson, and Maria Hansson-Sandsten "Smooth 2-D Frequency Estimation using Covariance Fitting", *22nd European Signal Processing Conference*, Lisbon, Portugal, September 1-5, 2014.

11. Johan Brynolfsson, Johan Swärd, Andreas Jakobsson, and Maria Hansson-Sandsten, "Smooth Time-Frequency Estimation Using Covariance Fitting". *39th International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 4-9, 2014.

12. Johan Brynolfsson and Maria Hansson-Sandsten, "Optimal Time Frequency Analysis of Multiple Time-Translated Locally Stationary Processes", *21st European Signal Processing Conference*, Marrakech, Morocco, September 9-13, 2013.

13. Johan Brynolfsson, Maria Hansson-Sandsten, Mikael Johansson, and Gerd Waldhauser, "Optimal Time-Frequency Analysis of EEG", *Medicinteknik-dagarna 2012*, Lund, Sweden, October 2-3, 2012.

Acknowledgements

The writing of this thesis marks the end of a seven-year-long, on and off journey, that has been truly fun, interesting and rewarding. When working on my Licentiate thesis, half-way through my doctoral studies, I found a job opportunity I felt I couldn't pass down. It was a difficult decision, but I believe it was the right one. After finishing my Licentiate thesis, I spent a year full-time at Spectronic Medical, before an agreement was made, letting me to continue my research part-time; allowing me to eat the cake and have it too. I would like to thank my supervisor, Prof. Maria Sandsten, for the great guidance, help, and encouragement in the work leading up to this thesis. I am very grateful for both allowing me to do my research part-time and giving me the opportunity to come back and finish my PhD. I am in large gratitude to my coauthor and co-supervisor Johan Swärd, working long hours on two interesting projects, leading to multiple conference papers, and two full papers. It has been fun from beginning to end! Thanks to Andreas Jakobsson, for your insights and amazing writing skills, and also thanks to my co-author Isabella Reinhold. This papers herein would truly have been lesser without you. I wish to thank my colleague Stefan Adalbjörnsson both in Lund and in Helsingborg, and my fellow PhD-students Filip Elvander, Maria Juhlin, and Ted Kronvall for many coffee and lunch discussions, and some very interesting conference trips; not only academically but also culturally and culinarily. Thank you also Carl Siversson at Spectronic Medical, for both freeing up time, allowing me to do research towards my PhD, and also for the interesting research I'm doing at Spectronic. Thank you Maria Lövgren and James Hakim for help with all matters practical. On a personal level, I would like to thank my parents and brother and sister. You are the best family one could ever have and I always look forward to our gatherings. Finally, I would like to express my love and gratitude to you, Emma! For your support and for always being so warm. You are the perfect partner in life and a great mother to our children.

Helsingborg, 2019   *Johan Brynolfsson*

Abstract

This thesis deals with estimation and classification problems of non-stationary processes in a few special cases. In paper A and paper D we make strong assumptions about the observed signal, where a specific model is assumed and the parameters of the model are estimated. In Paper B, Paper C, and Paper E more general assumptions about the structure of the observed processes are made, and the methods in these papers may be applied to a wider range of parameter estimation and classification scenarios. All papers handle non-stationary signals where the spectral power distribution may change with respect to time. Here, we are interested in finding time-frequency representations (TFR) of the signal which can depict how the frequencies and corresponding amplitudes change.

In Paper A, we consider the estimation of the shape parameter detailing time- and frequency translated Gaussian bell functions. The algorithm is based on the scaled reassigned spectrogram, where the spectrogram is calculated using a unit norm Gaussian window. The spectrogram is then reassigned using a large set of candidate scaling factors. For the correct scaling factor, with regards to the shape parameter, the reassigned spectrogram of a Gaussian function will be perfectly localized into one single point.

In Paper B, we expand on the concept in Paper A, and allow it to be applied to any twice differentiable transient function in any dimension. Given that the matched window function is used when calculating the spectrogram, we prove that all energy is reassigned to one single point in the time-frequency domain if scaled reassignment is applied. Given a parametric model of an observed signal, one may tune the parameter(s) to minimize the entropy of the matched reassigned spectrogram. We also present a classification scheme, where one may apply multiple different parametric models and evaluate which one of the models that best fit the data.

In Paper C, we consider the problem of estimating the spectral content of signals where the spectrum is assumed to have a smooth structure. By dividing the spectral representation into a coarse grid and assuming that the spectrum within each segment may be well approximated as linear, a smooth version of the Fourier transform is derived. Using this, we minimize the least squares norm of

the difference between the sample covariance matrix of an observed signal and any covariance matrix belonging to a piece-wise linear spectrum. Additionally, we allow for adding constraints that make the solution obey common assumptions of spectral representations. We apply the algorithm to stationary signals in one and two dimensions, as well as to one-dimensional non-stationary processes.

In Paper D we consider the problem of estimating the parameters of a multi-component chirp signal, where a harmonic structure may be imposed. The algorithm is based on a group sparsity with sparse groups framework where a large dictionary of candidate parameters is constructed. An optimization scheme is formulated such as to find harmonic groups of chirps that also punish the number of harmonics within each group. Additionally, we form a non-linear least squares step to avoid the bias which is introduced by the spacing of the dictionary.

In Paper E we propose that the Wigner-Ville distribution should be used as input to convolutional neural networks, as opposed to the often used spectrogram. As the spectrogram may be expressed as a convolution between a kernel function and the Wigner-Ville distribution, we argue that the kernel function should not be chosen manually. Instead, said convolutional kernel should be optimized together with the rest of the kernels that make up the neural network.

Populärvetenskaplig sammanfattning på svenska

I denna avhandling undersöker vi så kallade icke-stationära signaler. Olika signaler förekommer överallt omkring oss och innefattar allt som kan representeras i form av en serie av tal. Exempel på signaler är all sorts inspelade ljud där musik, eller mänskligt tal ingår men även djurläten så som fågelsång, delfinklick eller fladdermusklick som används för ekolokalisering. Hur vi människor uppfattar dessa ljud beror på frekvensen som de svänger med. En ljudsignal med mycket hög frekvens uppfattar vi som ljus och gäll, och en signal med låg frekvens uppfattas som basal och mörk. En annan viktig aspekt av signalen är vilken amplitud den har, det som i ljud-signaler motsvaras av volymen. Om en signal ändrar frekvens och/eller amplitud med tiden, vilket de flesta signaler gör, kallar vi den för icke-stationär. Ofta vill man analysera signaler genom att beräkna frekvensen som den svänger med, och dess amplitud, och i det icke-stationära fallet vill man alltså då mäta hur frekvenserna och amplituderna i en signal ändras med tiden. Detta fält av forskning inom signalbehandling kallas för tids-frekvens-analys. Detta är ett svårare problem än man kan tro, och det finns inte en universallösning som kan beräkna frekvens och amplitud för alla sorters icke-stationära signaler. För att försvåra saken ytterligare innehåller en signal ofta inte bara en komponent, utan är en sammansättning av flera signaler. Signalerna är heller nästan aldrig rena utan störs av vad som kallas brus. Bruset i ljudinspelning kan till exempel bestå av bakgrundsljud eller bara mätfel i mikrofonen som spelat in ljudet vilket kan ge ett konstant sprakande bakgrundsljud.

I denna avhandling tittar vi på några utvalda specialfall och tar fram metoder som är skräddarsydda för de olika fallen. Ett sätt att angripa problemen på är att ställa upp en matematisk modell som man tror ska passa för signalen som man vill undersöka. I den matematiska modellen finns parametrar som man kan vrida på tills modellen passar den uppmätta signalen så bra som möjligt.

I papper A presenteras en metod där amplituden i signalen kan modelleras av en så kallad Gaussklocka; en signalmodell av stor signifikans inom matematik, statistik och signalbehandling. Det unika i metoden vi presenterar är att den är

oberoende av vid vilken tidpunkt i signalen som Gaussklockan befinner sig och vilken frekvens signalen har.

I papper B generaliseras metoden som presenteras i papper A, och tillåter den att modellera, eller skatta parametrarna, i vilken transient modell som helst.

I papper C presenteras en metod för att skatta frekvenserna i en signal där man inte antar att det finns bara några få frekvenser, något som många metod gör. Istället antas att det är en fördelning av alla möjliga frekvenser.

I papper D presenteras en metod för att skatta linjär chirpar. Chirpar är signaler där frekvensen ändras kontinuerligt med tiden.

Slutligen presenteras, i papper E, en argumentation om hur tids-frekvens-representationer av signaler kan användas i neurala nätverk; ett verktyg där datorer tränas att klassificera signaler eller bilder utifrån många träningsexempel.

Introduction

In this thesis, methods in time-frequency analysis are investigated and applied in a selected set of special case scenarios. Time-frequency analysis is a useful tool when faced with observations of a non-stationary process, i.e., a process where the frequency content, or the power of the frequencies, is not constant for all eternity (or more realistically, for the duration of the observation). Rather, the spectral power distribution of the signal changes with respect to time. Throughout this thesis, it is assumed that the reader is familiar with methods in stationary spectral analysis. The introduction presents background information building up to the five papers presented in this thesis. For a deeper dive into time-frequency analysis I would heartily recommend *Time-Frequency Analysis, Concepts and Methods*, edited by Franz Hlawatsch and François Auger [1].

In this introduction t will denote time, and ω frequency. A noise free signal will be denoted $y(t)$ and an observation of $y(t)$ with additive, independently, and identically distributed (i.i.d.) noise will be denoted $x(t)$. The Fourier transform will, in this introduction, be expressed in angular frequency, and is defined as

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt \tag{1}$$

The inverse Fourier transform is defined as

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega \tag{2}$$

1 Background

Time-frequency estimation has long been a useful tool in the analysis and classification of both naturally occurring signals, such as human speech, bird song, and seismic signals etc., as well as artificial signals such as music, FM/AM radio, and radar signals to give a few examples. By estimating how the frequency content in a signal or process changes over time, one gets an important insight into the characteristics and behaviors, and help in classification, of the observed signal. The time-frequency estimators are in this introduction divided into two main

1

groups; parametric and non-parametric estimators. When using a non-parametric estimator, few assumptions, besides that the observation fulfill the Nyquist theorem, have to be made of the observed signal. For the parametric estimators, on the other hand, one instead has to make quite clear assumptions of the signal in question. By setting up a model of the signal, one may estimate the parameters detailing the model, via optimization of some appropriately chosen cost function. An estimate of the true time-frequency representation (TFR) can then be found by optimizing the cost function, and calculating the theoretic time-frequency content given the model and estimated parameters. The model assumed in a parametric estimator has to be exact in terms of model-order. If one also wants to estimate the model-order, i.e. the number of components in the signal, one may turn to the sub-class of semi-parametric estimators, where a model assumption must still be made, but where the number of components is not fixed.

1.1 A Chirp Signal Example

Throughout this introduction, a chirp signal will be used as an example to show the performance of the TFR estimators that are presented. A chirp is an oscillating signal where the frequency changes in a continuous fashion over time.

Example 1. *In this toy example the signal is a two-component linear chirp signal i.e., on the form*

$$x(t) = \sum_{k=1}^{2} a_k e^{i\left(\omega_k t + \frac{r_k}{2}t^2\right)} + n(t) \tag{3}$$

where both components have unit amplitude, $a_1 = a_2 = 1$, and where the starting frequencies are set to $\omega_1 = 0.5$ and $\omega_2 = 2$ rad/s and the rate of changes are set to $r_1 = 0.006$ and $r_2 = 0.002$ rad/s^2. Further the signal is corrupted by additive i.i.d. Gaussian white noise, $n(t)$, with variance 0.1. The signal is 128 samples long and the real part of a realization can be seen in Figure 1 and the ground truth TFR is seen in Figure 2.

2 Non-Parametric Methods

The class of non-parametric methods is a good choice when little is known about the observed signal or when an initial estimate is necessary. The estimates are
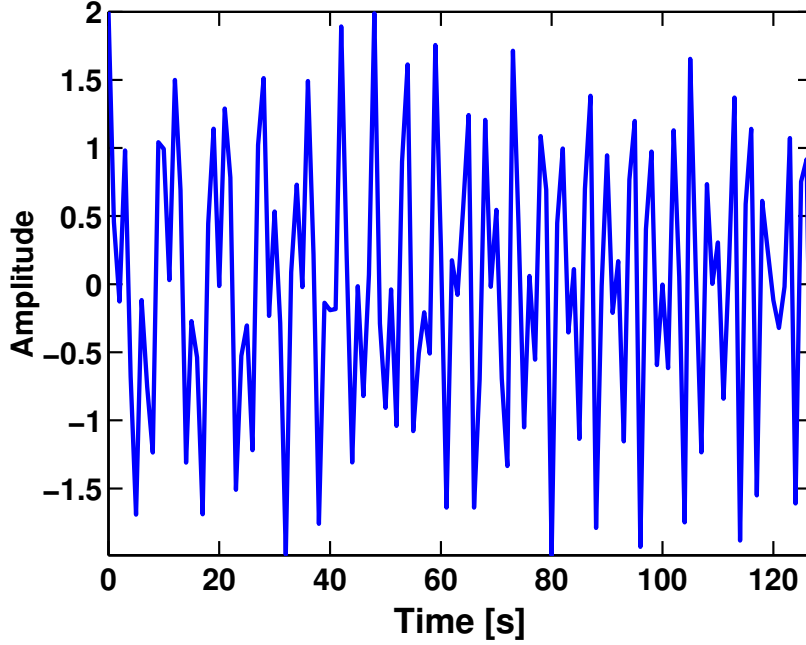
Figure 1: Real part of a realization of the chirp signal in Example 1.

generally quite robust to prior assumptions, and if they are smoothed, also robust in the terms of noise disturbances. They do however suffer from high variance and low resolution. When investigating the properties of a time-frequency estimator, there are a few properties that are of interest.

- **Positivity**
  As the TFR is a distribution of the signal energy into the time-frequency plane, an expected property of the TFR estimate is that it should be positive everywhere, such that

$$TFR(t, \omega) \geq 0 \quad \forall (t, \omega) \in \mathbb{R}^2 \tag{4}$$

- **Energy conservation**
  As the TFR distributes the signal energy into the time-frequency plane, the total energy of the observed signal should be same as the total energy of

3
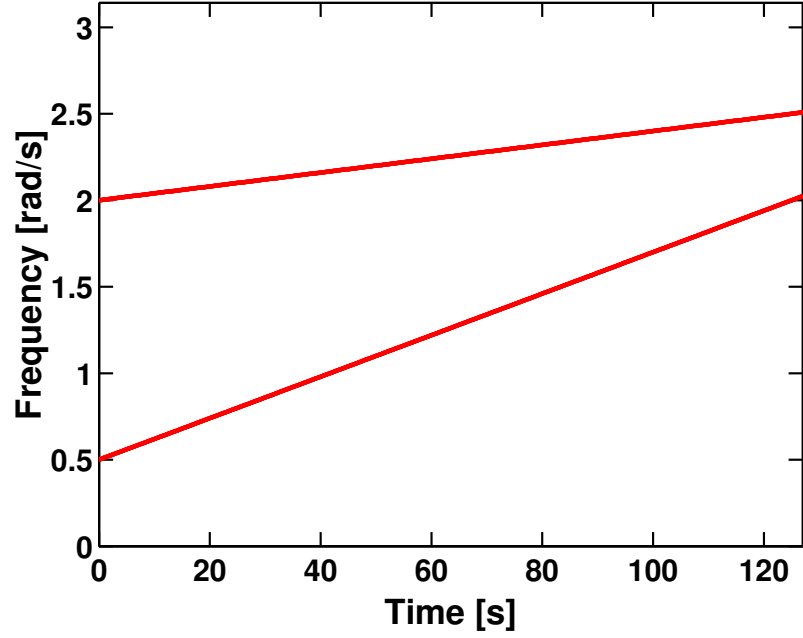
Figure 2: The ground-truth TFR of the two-component chirp signal in Example 1.

TFR estimate, i.e.,

$$\frac{1}{2\pi} \iint_{\mathbb{R}^2} \text{TFR}(t, \omega) dt d\omega = \int_{-\infty}^{\infty} |x(t)|^2 dt \tag{5}$$

- **Marginals**
  A stricter version of the energy conservation constraint is to enforce that the energy distributed at each instantaneous time or frequency must equal the energy of the signal at that exact time and frequency, respectively, expressed as

$$\int_{-\infty}^{\infty} \text{TFR}(t, \omega) dt = |X(\omega)|^2 \quad \forall \, \omega \in \mathbb{R} \tag{6}$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \text{TFR}(t, \omega) d\omega = |x(t)|^2 \quad \forall \, t \in \mathbb{R} \tag{7}$$

4

2.1 The Spectrogram

The spectrogram is one of the first methods used to estimate the TFR. By letting a suitable time-window, $h(t)$, move along the observed signal, $x(t)$, and estimating the frequency content within each time frame one gets an estimate of how the frequency changes over time, named the short time Fourier transform (STFT)

$$F_x^h(t, \omega) = \int_{-\infty}^{\infty} x(s)h^*(t - s)e^{-i\omega s}ds \tag{8}$$

where the spectrogram is then found as the squared absolute value of the STFT

$$S_x(t, \omega) = \left| F_x^h(t, \omega) \right|^2 \tag{9}$$

The time-window may be chosen to be a rectangular window, but more preferably it should also apply smoothing of the TFR estimate, analogous to the stationary spectral estimation case, to decrease side lobes and lower the variance. There are a wide range of window functions suggested where the Hann, Hamming, Bartlett and Blackman windows are a few common choices [2]. The choice of spectrogram analysis of signals is common in many fields, in part due to its simplicity and intuitiveness and also due to that it fulfills nice properties such as that it is real- and positive valued [3]. If a unit energy window function is used, it also fulfills the energy conservation property. One drawback of this method however is the resolution trade off between time and frequency. By applying a wide time-window one will get a more accurate frequency estimate at the cost of poor time resolution. Vice versa, a short time-window renders good resolution in time but poor frequency estimates. Another drawback is the high variance of the discrete Fourier transform.

In Figure 3, the spectrogram is shown for the two component chirp signal in Example 1, where one may especially note the wideness of the two chirp components. A unit norm Gaussian window function on the form

$$h(t) = \frac{1}{\pi^{1/4}\sqrt{\lambda}} \exp\left\{ -\frac{t^2}{2\lambda^2} \right\} \tag{10}$$

with shape parameter $\lambda = 10$, was used when calculating the spectrogram.

2.2 The Quadratic Class

A class of more recent methods of estimating the time-frequency distribution of a non-stationary signal is the quadratic class of estimators, also known as Cohen's
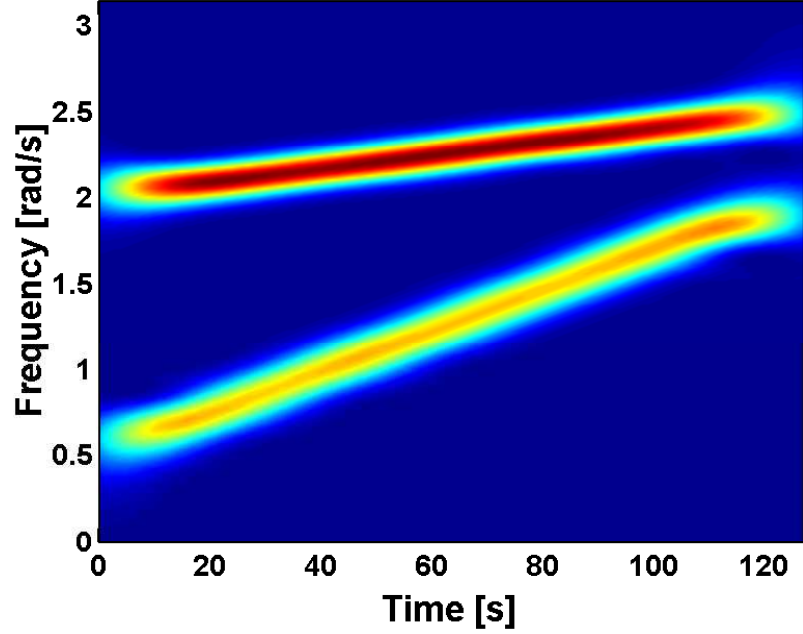
Figure 3: Spectrogram of the two-chirp example computed with a unit energy Gaussian window. Note the wideness of the two chirp components.

class, where the Wigner-Ville distribution (WVD) [4, 5] may be seen as the base distribution, and is defined as

$$W_x(t, \omega) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i\omega\tau} d\tau \tag{11}$$

Here, $x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right)$ is the instantaneous covariance function, making the distribution an intuitive transition from the stationary case. Benefits of the WVD include that an instantaneous frequency (IF) will be well estimated for a single component, but also that it fulfills a range of desired properties such as energy conservation, that the marginal distributions in time and frequency are fulfilled, and possible transition between signal representation and the TFR to name a few [3]. The WVD is unfortunately not guaranteed to be positive. However, the main drawback of the WVD is the cross-components that arises when the

6

analyzed signal is comprised of multiple components, making it difficult, if not impossible, to interpret the resulting TFR estimate. Assume that an observed signal consists of two components, i.e., $x(t) = x_1(t) + x_2(t)$. The WVD is then found as

$$W_x(t, \omega) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i\omega\tau} d\tau$$

$$= \int_{-\infty}^{\infty} \left(\left(x_1\left(t + \frac{\tau}{2}\right) + x_2\left(t + \frac{\tau}{2}\right)\right) \right) \tag{12}$$

$$\left(x_1\left(t - \frac{\tau}{2}\right) + x_2\left(t - \frac{\tau}{2}\right)\right)^* \right) e^{-i\omega\tau} d\tau \tag{13}$$

$$= W_{x_1}(t, \omega) + W_{x_2}(t, \omega) + 2\,\mathfrak{Re}\left\{W_{x_1 x_2}(t, \omega)\right\} \tag{14}$$

The cross-components will be oscillatory and situated at the midpoint between every pairwise combination of components existing in the signal.

The WVD of the chirp signal example is shown in Figure 4. One may note how the IFs of the two components are almost perfectly localized, but also that the oscillating cross-component between the two chirps makes the results harder to interpret. This would be even harder if there also was a third component in between the two, which essentially would be hidden by the cross-components.

The drawbacks of the WVD may be remedied, at the cost of lowered concentration, using some 2D smoothing kernel, analogous with the spectrogram. This gives rise to the quadratic class of TFRs [6]

$$C_x(t, \omega) = \iint_{\mathbb{R}^2} \varphi(s, \xi) W_x(t - s, \omega - \xi) ds \frac{d\xi}{2\pi} \tag{15}$$

where $\varphi(t, \omega)$ is some chosen smoothing kernel function. Note that the spectrogram is actually a member of the quadratic class with smoothing kernel $\varphi(t, \omega) = W_h(t, \omega)$, i.e., the WVD of the window function.

Another way to investigate the cross-components, and the smoothing kernel, is to transform the TFR into the ambiguity domain. The base distribution is then

$$A_x(\nu, \tau) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i\nu t} dt \tag{16}$$

This may also be written as a transform of the WVD

$$A_x(\nu, \tau) = \frac{1}{2\pi} \iint_{\mathbb{R}^2} W_x(t, \omega) e^{-it\nu} e^{i\tau\omega} d\nu d\omega \tag{17}$$

7
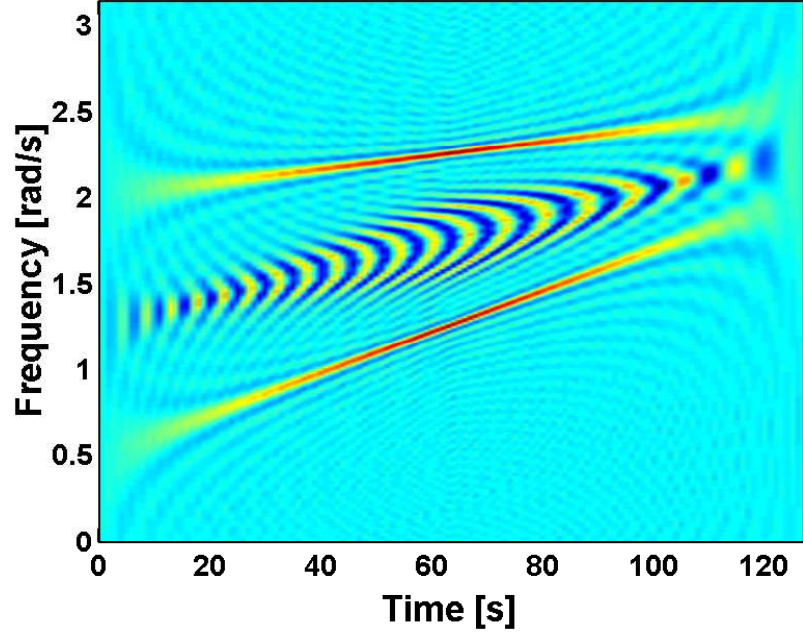
Figure 4: Wigner-Ville distribution of the chirp signal example. Note that the two components may be identified and the oscillating cross-component is clearly visible between them.

The ambiguity function is a function of time-lag, τ, and frequency-lag ν. Any true component in the WVD will naturally have no time- or frequency-lag between itself and itself, and will hence be situated in the origin of the ambiguity function. Cross-components arise when there is a time- and/or frequency-lag, as there is between two different components, and the cross-terms will then be situated away from the origin. A convolutional smoothing kernel in the time-frequency domain will, due to the convolution theorem, be multiplicative in the ambiguity domain. This gives the convolutional kernel a weighted masking interpretation in the ambiguity domain, where we want to keep terms situated close the origin unaltered, and mask out terms away from the origin. The kernel function is therefore commonly defined in the ambiguity domain.

An often mentioned kernel is the Choi-Williams (CW) kernel [7]. The kernel

is defined in the ambiguity domain as

$$\varphi_{CW}(\nu, \tau) = \exp\left(-\alpha(\nu\tau)^2\right) \tag{18}$$

Another common kernel is the Born-Jordan (BJ) kernel [8]

$$\varphi_{BJ}(\nu, \tau) = 2\frac{\sin\left(\frac{\nu\tau}{2}\right)}{\nu\tau} \tag{19}$$

Both these kernel result in TFRs that fulfill the marginal property. For this property to hold, the kernel must fulfill [3]

$$\varphi(0, \tau) = 1, \quad \forall\ \tau \in \mathbb{R} \tag{20}$$
$$\varphi(\nu, 0) = 1, \quad \forall\ \nu \in \mathbb{R} \tag{21}$$

This also means that all cross-components situated between terms that are located in parallel with the time- or frequency- axes, i.e., with time-lag $\tau = 0$ or frequency-lag $\nu = 0$, will be kept intact. As the example used in this introduction consists of two linear chirps, there will be cross-components parallel to the frequency axis at every time instance, and the resulting TFR using CW or BJ will still retain cross-components. The CW distribution of the chirp example is shown in Figure 5, where one may note the retained cross-components between the two chirps.

If more is known about the observation, a signal specific optimal kernel may be used. An example is [9], where the mean-squared-error optimal kernel for a stochastic processes with finite fourth order moments is derived. This is set up as

$$\varphi_{opt}(\nu, \tau) = \arg\inf_{\varphi} \iint_{\mathbb{R}^2} \mathbb{E}\left\{\left|A_x(\nu, \tau)\varphi(\nu, \tau) - \mathbb{E}\left\{A_y(\nu, \tau)\right\}\right|^2\right\} \frac{d\nu}{2\pi}d\tau \tag{22}$$

where $\mathbb{E}\{\cdot\}$ denotes expected value, and $y(t)$ is the noise free version of $x(t)$. The solution is

$$\varphi_{opt}(\nu, \tau) = \begin{cases} \frac{\mathbb{E}\{A_y(\nu, \tau)\}\mathbb{E}\{A_x^*(\nu, \tau)\}}{\mathbb{E}\{|A_x(\nu, \tau)|^2\}} & \text{if } (\nu, \tau) \in \mathbb{S}_{yx} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where

$$\mathbb{S}_{yx} = \left\{(\nu, \tau) \in \mathbb{R}^2\ :\ \left|\mathbb{E}\left\{A_y(\nu, \tau)\right\}\mathbb{E}\left\{A_x^*(\nu, \tau)\right\}\right| > 0\right\} \tag{24}$$

9
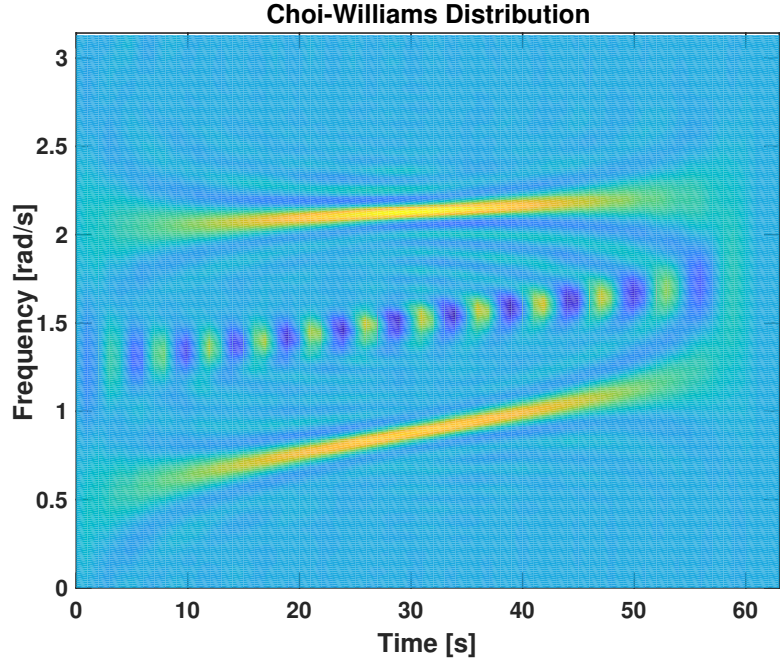
Figure 5: Choi-Willims distribution of the chirp signal example. The CW kernel removes a large part of the cross-terms that are not parallel with time- or frequency-axes. Compare with the WVD (Figure 4) and note that mostly cross-components parallel to the frequency axis is retained.

If there exists a $(\nu, \tau) \in S_{yx}$ such that $\mathbb{E}\left\{|A_x(\nu, \tau)|^2\right\} = 0$, the optimal kernel does not exist, but may be approximated within the subset

$$\mathbb{S}_\varepsilon = \left\{(\nu, \tau) \in \mathbb{S}_{yx} \ : \ \mathbb{E}\left\{|A_x(\nu, \tau)|^2\right\} > \varepsilon \left|\mathbb{E}\left\{A_y(\nu, \tau)\right\} \mathbb{E}\left\{A_x^*(\nu, \tau)\right\}\right|\right\} \quad (25)$$

Another approach is taken in [10], where kernels are optimized to separate different classes in an existing training set. There are also adaptive kernels proposed, where more general assumptions are made, see e.g., [11].

10

2.3 The Reassignment Method

As mentioned in Section 2.1, a main problem with the spectrogram is the resolution trade-off. One solution to this problem is the so called reassignment method, first suggested by [12], but was largely ignored at the time. However, as the method was reintroduced in [13], with a more thorough investigation, it received more attention. By reassigning every point of a TFR to its local center of gravity, a sparse estimate of the TFR is obtained, making for a better estimate if the spectral density is indeed sparse. The distance to the gravity center, for every point (t, ω) in the time-frequency plane may, for any member of the quadratic class, be calculated as

$$
\tilde{t}_x(t, \omega) = t - \frac{\iint_{\mathbb{R}^2} s\varphi(s, \xi) W_x(t - s, \omega - \xi) ds \frac{d\xi}{2\pi}}{\iint_{\mathbb{R}^2} \varphi(s, \xi) W_x(t - s, \omega - \xi) ds \frac{d\xi}{2\pi}} \tag{26}
$$

$$
\tilde{\omega}_x(t, \omega) = \omega - \frac{\iint_{\mathbb{R}^2} \xi\varphi(s, \xi) W_x(t - s, \omega - \xi) ds \frac{d\xi}{2\pi}}{\iint_{\mathbb{R}^2} \varphi(s, \xi) W_x(t - s, \omega - \xi) ds \frac{d\xi}{2\pi}} \tag{27}
$$

where $\varphi(t, \omega)$ is a chosen kernel function. Using this, the reassigned TFR may be expressed as

$$
R_x(t, \omega) = \iint_{\mathbb{R}^2} C_x(s, \xi) \delta\left(t - \tilde{t}_x(s, \xi), \omega - \tilde{\omega}_x(s, \xi)\right) ds \frac{d\xi}{2\pi} \tag{28}
$$

where $C_x(t, \omega)$ is the member of the quadratic class calculated with $\varphi(t, \omega)$ as kernel function and $\delta(t, \omega)$ is the 2-dimensional Dirac function. This renders a method giving sparse TFRs for sinusoids, impulses and chirp signals, whose IF estimates will be perfectly localized. The spectrogram has become a common choice to use for reassignment, again in part due to its simplicity to calculate and in part due to its positive valued nature. Further, for the spectrogram, it was shown in [13] that the center of gravity could be expressed compactly as

$$
\tilde{t}_x(t, \omega) = t + \Re\mathfrak{e}\left\{\frac{F_x^{th}(t, \omega)}{F_x^{h}(t, \omega)}\right\} \tag{29}
$$

$$
\tilde{\omega}_x(t, \omega) = \omega - \Im\mathfrak{m}\left\{\frac{F_x^{dh/dt}(t, \omega)}{F_x^{h}(t, \omega)}\right\} \tag{30}
$$

where F_x^{h} denotes the STFT of $x(t)$ using $h(t)$, F^{th} using $t \cdot h(t)$, and $F^{dh/dt}$ using $dh(t)/dt$ as their respective window functions. Further, $\Re\mathfrak{e}\{\cdot\}$ and $\Im\mathfrak{m}\{\cdot\}$

denotes the real and imaginary parts of a complex valued number, respectively. This renders a more computationally efficient method making it more viable to use.

The reassigned spectrogram of the two-chirp example signal is depicted in Figure 6. Note that the chirps are almost perfectly concentrated, apart from the edges where leakage occur.

The expression of the reassignment may be compactly formulated in a complex valued manner using the normalized reassignment vector [14]

$$r_x(t, \omega) = \frac{\tilde{t}_x(t, \omega) - t}{\Delta t_h} + i \frac{\tilde{\omega}_x(t, \omega) - \omega}{\Delta \omega_h} \tag{31}$$

where Δt_h and $\Delta \omega_h$ are the time- and frequency duration of the window function, defined as

$$\Delta t_h = \sqrt{\int_{-\infty}^{\infty} t^2 |h(t)|^2 dt} \tag{32}$$

$$\Delta \omega_h = \sqrt{\int_{-\infty}^{\infty} \omega^2 |H(\omega)|^2 d\omega} \tag{33}$$

A statistical analysis of the reassignment was made in [15, 16] where the probability density function of the normalized reassignment vectors in every point for a signal observed in white, circularly symmetric, Gaussian noise was derived. If the unit norm Gaussian window, defined in (10), is used when calculating the spectrogram the probability density function may be expressed as

$$f(r_x(t, \omega)) = \frac{1}{\pi \left(1 + |r_x(t, \omega)|^2|\right)} \left[1 + \frac{S_y(t, \omega)}{\sigma^2} \frac{|1 + r_x(t, \omega) r_y(t, \omega)^*|^2}{1 + |r_x(t, \omega)|^2}\right]$$
$$\times \exp \left\{ -\frac{S_y(t, \omega)}{\sigma^2} \frac{|r_x(t, \omega) - r_y(t, \omega)|^2}{1 + |r_x(t, \omega)|^2} \right\} \tag{34}$$

where σ^2 is the variance of the additive noise, and $S_y(t, \omega)$ and $r_y(t, \omega)$ are the spectrogram and the normalized reassignment vector of the noise-free signal, respectively. This gives a insight into the sensitivity the reassignment, with regards to noise disturbances. In paper A, we construct a parameter estimator utilizing the reassignment method. Using the above distribution function we can investigate how the estimator behaves for different levels of additive noise disturbances.
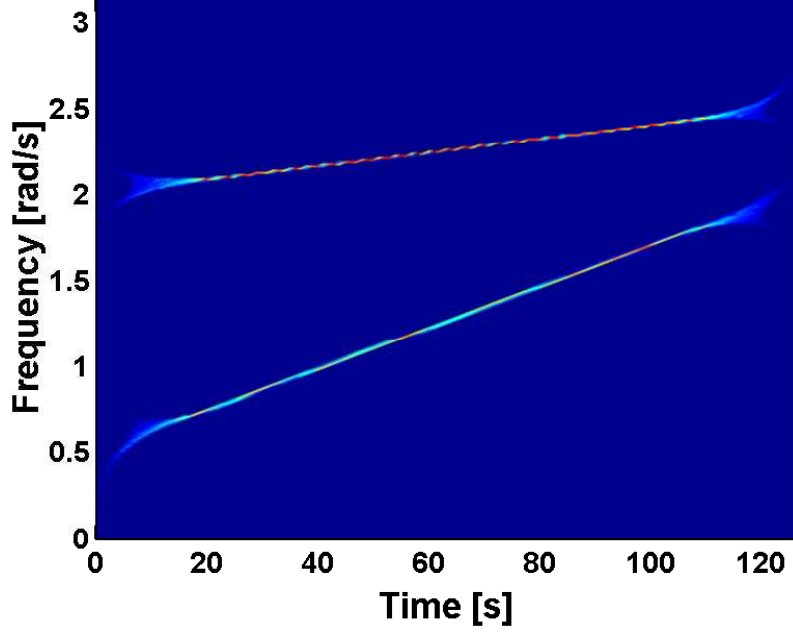
Figure 6: The reassigned version of the spectrogram in Figure 3 for the two chirp example, where one may note that the localization is almost perfect for the two chirps.

In [17] they further develop the reassignment method by introducing a parameter, making it adjustable and more versatile. There, the normal reassignment was interpreted as one step in an iterative scheme to move energy towards a fixed point. By introducing the non-normalized vector

$$R(t, \omega) = \begin{pmatrix} \frac{\iint_{\mathbb{R}^2} s\varphi(s,\xi)W_x(t-s,\omega-\xi)ds\frac{d\xi}{2\pi}}{\iint_{\mathbb{R}^2} \varphi(s,\xi)W_x(t-s,\omega-\xi)ds\frac{d\xi}{2\pi}} \\ \frac{\iint_{\mathbb{R}^2} \xi\varphi(s,\xi)W_x(t-s,\omega-\xi)ds\frac{d\xi}{2\pi}}{\iint_{\mathbb{R}^2} \varphi(s,\xi)W_x(t-s,\omega-\xi)ds\frac{d\xi}{2\pi}} \end{pmatrix} \quad (35)$$

the common reassignment method can be expressed as

$$\begin{bmatrix} \tilde{t}_x(t, \omega) \\ \tilde{\omega}_x(t, \omega) \end{bmatrix} = \begin{bmatrix} t \\ \omega \end{bmatrix} - R(t, \omega) \quad (36)$$

13

They then present a method based on the Levenberg-Marquardt algorithm, where the step taken depends on a ratio between the gradient of the center of mass plus a constant, and the center of gravity itself, such that

$$
\begin{bmatrix} \tilde{t}_x(t, \omega) \\ \tilde{\omega}_x(t, \omega) \end{bmatrix} = \begin{bmatrix} t \\ \omega \end{bmatrix} - \left(\nabla R(t, \omega) + cI \right)^{-1} R(t, \omega) \tag{37}
$$

where c is a parameter that may be tuned, and I is the identity matrix. By tuning the parameter c, perfect concentration may be achieved after reassignment for a larger set of signals than just chirps and impulses. In [18] we employ the interpretation presented in [17], but keep the common reassignment vector. Instead, we introduce separate parameters for the two dimensions, altering the step length taken in each dimension, such that

$$
\begin{bmatrix} \tilde{t}_x(t, \omega) \\ \tilde{\omega}_x(t, \omega) \end{bmatrix} = \begin{bmatrix} t \\ \omega \end{bmatrix} - \begin{bmatrix} c_t & 0 \\ 0 & c_\omega \end{bmatrix} R(t, \omega) \tag{38}
$$

We show that by scaling the reassignment vector with a factor of two, in both time and frequency, a Gaussian function is perfectly localized after reassignment, if a matched Gaussian window is used. We further present how it may be used to estimate the shape parameter detailing the function. This method is further investigated in paper A where we show that a Gaussian function can be reassigned into one single point, given that a Gaussian window is used, not necessarily the matched one, and if the correct scaling factors are applied. In Paper B we show that any transient and twice differentiable function can be reassigned into one single point, if the matched window function and a scaling factor of two is used.

3   Parametric and Semi-Parametric Methods

In the recent decade, there has been a surge of parametric methods, where a model of the observed signals is suggested and the parameters detailing the model are estimated. Given a good model of the observed process, the rendered estimates are generally quite accurate and robust in terms of noise disturbances and outliers [2]. However, if the model assumption is wrong, or even if only the model order assumption is wrong, the accuracy deteriorates.

Assume that the observed signals may be modeled as

$$
x(t) = f(t; \psi) + e(t) \tag{39}
$$

where $f(\cdot)$ is some model function and ψ is a vector of parameters detailing the model. Further, denote the observation vector as

$$\mathbf{x} = [x_1, \ldots, x_N]^T \tag{40}$$

where the j:th element in the vector, x_j is observed at time t_j, i.e. $x_j = x(t_j)$, and the observation times are collected as

$$\mathbf{t} = [t_1, \ldots, t_N]^T \tag{41}$$

One common choice is then to minimize the distance between the observations and the parametric model using some appropriate norm (commonly the ℓ_2-norm giving the least squares (LS) estimate). The estimate may then be formed as

$$\hat{\psi} = \underset{\psi}{\operatorname{argmin}} \, ||\mathbf{x} - f(\mathbf{t}; \psi)|| \tag{42}$$

where $|| \cdot ||$ is a chosen norm. Another choice is to use the Maximum Likelihood method, i.e. make a distribution assumption on the noise of the signal and maximize the likelihood function over the parameter(s) given the observation

$$\hat{\psi} = \underset{\psi}{\operatorname{argmax}} \, \mathcal{L}(\psi|\mathbf{x}) \tag{43}$$

where $\mathcal{L}(\psi|\mathbf{x})$ is the likelihood function of the parameters given the observations. The TFR estimate for the signal may then be obtained from the assumed model using the estimated parameters.

In Paper C, we present a spectral estimator for both stationary and non-stationary signals, where the spectral representation of the signal is assumed to be smooth. The estimator may be interpreted both as a non-parametric method as well as a parametrization of the spectrum of the process. In the non-stationary case, the smoothness in the spectrum is accounted for by dividing the time-frequency plane into a coarse grid structure where the spectrum within each grid segment is assumed to be well approximated by a plane. Using this formulation, an optimization function is set up where the distance between a covariance estimate and any covariance belonging to a smooth spectrum is minimized.

3.1 Sparse Semi-Parametric Estimation

If one has a good idea of the model of the observed signal, but does not know the model order, e.g., the number of components in a linear combination of some

model function, one may want to turn to the set of semi-parametric estimators. Assume that we have an observation $x(t)$, on the form

$$x(t) = \sum_{k=1}^{K} \alpha_k f(t; \psi_k) + e(t) \tag{44}$$

where K is the unknown number of components, α_k is the amplitude, and $e(t)$ additive noise corruption, often assumed to be circularly symmetric Gaussian white noise. One may then construct a large set of P candidate parameters

$$\mathbf{\Psi} = [\Psi_1, \ldots, \Psi_P] \tag{45}$$

where $P \gg K$, and define a realization of the model given

$$\mathbf{d}_j = f(\mathbf{t}; \Psi_j) \quad \forall j = 1...P \tag{46}$$

By collecting all the candidate realizations into a matrix

$$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_P] \tag{47}$$

an initial guess might be to estimate the amplitude of each candidate realization using LS

$$\hat{\mathbf{z}}_{LS} = \operatorname*{argmin}_{\mathbf{z}} \frac{1}{2} ||\mathbf{x} - \mathbf{D}\mathbf{z}||_2^2 \tag{48}$$

where $\mathbf{z}$ is the $P \times 1$ vector activating the candidate realizations by assigning them amplitude

$$\mathbf{z} = [z_1, z_2, \ldots, z_P]^T \tag{49}$$

However, even if the signal is observed noise free and all parameter values exist with perfect accuracy in the dictionary, the solutions may be a combination of all candidate realizations as the system is under-determined. If, additionally, the signal is observed with noise corruption, the solution will most likely be a weighted combination of all dictionary elements. An example of this may be seen in Figure 7, where this is applied to Example 1. A dictionary is set up using all possible combinations between 500 starting frequencies and 100 frequency rates, i.e., 50000 dictionary entries in total. The absolute value of the solution to (48)

16
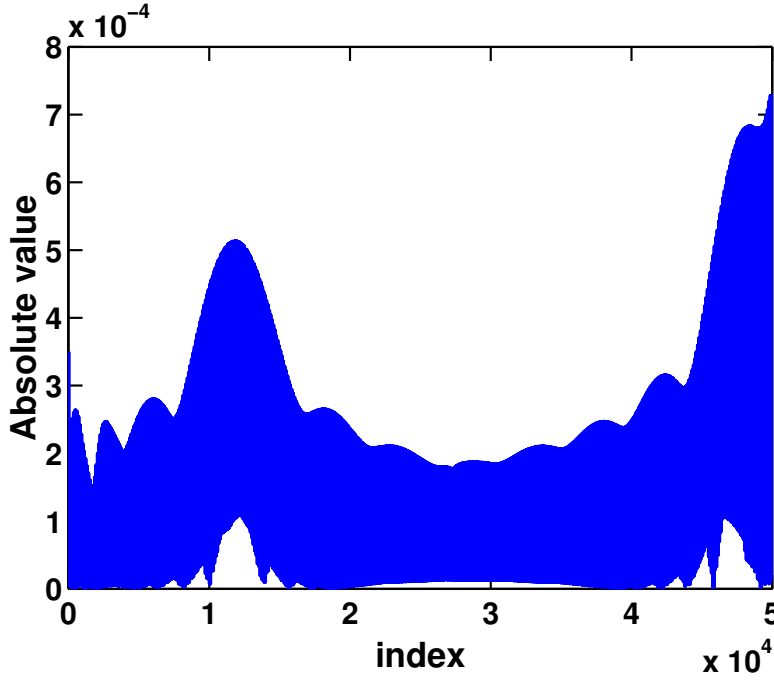
Figure 7: The solution to (48) applied to the two-component chirp signal showing the estimated amplitude for each of the 50000 dictionary elements. Note how all dictionary components are activated.

is shown in Figure 7 where one can see that all dictionary entries are assigned a non-zero amplitude.

To achieve the sought sparse solution it was suggested in [19] that an ℓ_1-norm penalty factor could be used

$$\hat{\mathbf{z}}_{LASSO} = \underset{z}{\operatorname{argmin}} \frac{1}{2}||\mathbf{x} - \mathbf{D}\mathbf{z}||_2^2 + \rho||\mathbf{z}||_1 \tag{50}$$

where the ℓ_1-norm is defined as

$$||\mathbf{z}||_1 = \sum_{j=1}^{P} |z_j| \tag{51}$$

In (50), ρ is a weighting factor, which is tuned in respect to how sparse one desires the solution to be. This method, denoted the LASSO-method (Least Absolute
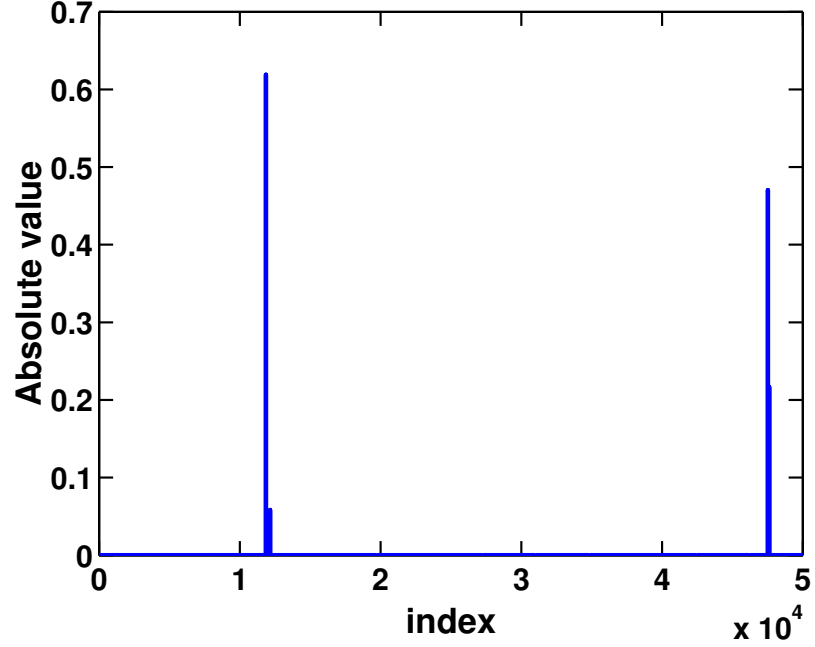
Figure 8: The solution to the (50) applied to the two-component chirp signal. Note that four dictionary components are activated, i.e., the two pairs of non-zero amplitudes, where as the signal only contains two components.

Shrinkage and Selection Operator), gained much popularity the recent decade. The LASSO estimate is biased for two reasons. The first one is caused by the term enforcing sparsity, which causes the amplitude of the activated components to be erroneously estimated. This may be remedied by applying, e.g., a LS step with only the activated dictionary elements. The second bias is introduced by the spacing of the dictionary. The true parameter value(s) are probably not included when creating the dictionary and hence only the closest one or two existing neighbors are activated. In paper D we remedy this by including a non-linear least squares step to estimate the parameter values outside the dictionary. Another issue with the LASSO method is the choice of ρ, which gives weight to how sparse one expects the observed signal to be. One may however easily confirm the resulting model order by reconstructing the signal using the given parameter estimates and

18

computing the residual. If the residual is too large, ρ may be decreased, allowing for more dictionary elements to be activated. If the reconstruction is "too good", and/or is not sparse enough, i.e., we have over-fitting, one may want to increase ρ. The LASSO approach is applied to the two-component chirp example and the result is seen in Figure 8, where the $\hat{\mathbf{z}}_{LASSO}$ is plotted. Note that more than two values in $\hat{\mathbf{z}}_{LASSO}$ are non-zero. As can be seen, two pairs of dictionary realizations are activated, which suggests that the true parameter values lies in between those existing in the dictionary. This causes the model order to be erroneously estimated. As suggested in [20] this may be remedied by instead adding weighting factors for each value of $\mathbf{z}$ using a diagonal weighting matrix and iteratively update it using previous estimates of $\hat{\mathbf{z}}$

$$\hat{\mathbf{z}}^{(b)} = \underset{\mathbf{z}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{x} - \mathbf{Dz}||_2^2 + \rho||\mathbf{W}^{(b-1)}\mathbf{z}||_1 \tag{52}$$

$$\mathbf{W}_{j,k}^{(b)} = \begin{cases} \frac{1}{|\hat{z}_j^{b-1}|+\varepsilon} & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} \tag{53}$$

where ε is a small factor to avoid zero-divisions and where $\mathbf{W}^{(0)}$ is set to be the identity matrix. The algorithm is generally quite robust to the choice of ε but it is suggested to be set slightly smaller than the smallest expected component magnitude in accordance to empirical evidence [20]. In the same paper, it is also shown that this iterative weighting converges to a log-norm penalty function, making it a better sparsifying penalty function than the ℓ_1-norm. Ideally, one would like to penalize the number of activated candidate realizations in the dictionary, i.e., use the ℓ_0 norm which, by assuming $0^0 = 0$, may be defined as

$$||\mathbf{z}||_0 = \sum_j |z_j|^0 \tag{54}$$

This is however neither a true norm, nor a convex function, making it difficult, and computationally costly to minimize. Here, the log-norm penalty function is a better approximation than the ℓ_1-norm. A simple way to convince oneself that this is a better approximation is

$$||\mathbf{W}^{(b-1)}\mathbf{z}^{(b)}||_1 = \sum_j \frac{|z_j^{(b)}|}{|z_j^{(b-1)} + \varepsilon|} \tag{55}$$

19
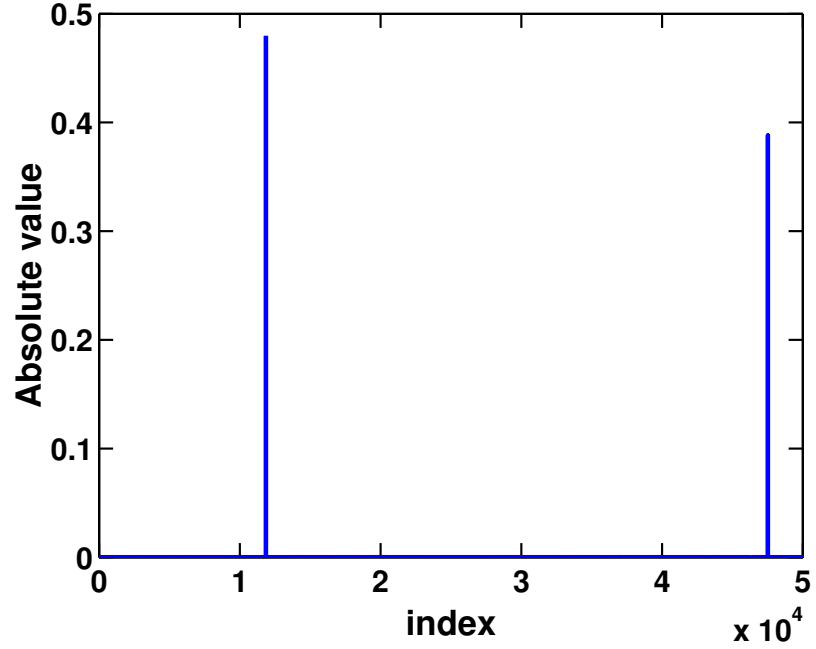
Figure 9: The solution to (52) applied to the two-component chirp signal. Note that two dictionary components are correctly activated, unlike the pure LASSO solution in Figure 8.

where, if $z_j^{(b)} \approx z_j^{(b-1)}$

$$\frac{|z_j^{(b)}|}{|z_j^{(b-1)} + \varepsilon|} \approx \begin{cases} 1 & z^{(b)} \neq 0 \\ 0 & z^{(b)} = 0 \end{cases} \tag{56}$$

i.e. the ℓ_0-norm.

The results of the iterative weighted LASSO applied to the chirp signal in Example 1 is seen in Figure 9 where two dictionary realizations are correctly activated.

20

3.2 Group Sparsity

A further development of this method was introduced in [21], where group sparsity is sought; i.e., from a set of candidate parameters we want to find a few number of groups of parameters that are, in some pre-defined way, connected. To generalize further, the number of elements in each group is also assumed to be unknown and the penalty term for sparsity is hence still included. The group sparsity with sparse groups estimate is then stated as

$$\hat{\mathbf{z}}_{GS} = \underset{z}{\arg\min} \frac{1}{2}||\mathbf{x} - \mathbf{Dz}||_2^2 + \rho||\mathbf{z}||_1 + \gamma \sum_{q=1}^{Q} ||\mathbf{z}[q]||_2 \tag{57}$$

where $\mathbf{z}[q]$ denotes the q:th group of corresponding candidate realizations in $\mathbf{D}$, Q is the number of candidate blocks in the dictionary, and ρ and γ are weighting factors dictating the overall sparsity and the number of groups. This is utilized in paper D where we look at linear chirp signals, which may be harmonically related, i.e., there are overtones that are integer multiplications of some fundamental frequency, or chirp as in this case. Each group in the dictionary is then a fundamental chirp and its overtones.

4 A Very Brief Introduction to Neural Networks

In paper E, we take a slight detour and look into applications of TFR as input to neural networks (NN). A quick overview of NNs will therefore be given here.

Prior to the explosion of NNs in image analysis, image processing and image classification required a great deal of craftsmanship. Features where pre-defined, followed by investigation of how well these features could be used as an identifier of a certain class of images as compared to other classes. By instead constructing a network containing a multitude of parameters, and optimizing the parameters using a training data set, there is no need to set any features by hand. Instead, the features are found in the optimization scheme, and are "chosen" by the network to separate different classes. The term neural network comes from the observation that the structure may be an interpretation of how the brain is built up, where similarly billions of neurons are connected to classify, interpret, and interact with the world.

In a neural network, a neuron is defined as a node taking multiple input values

outputting the weighted sum plus a bias term, i.e.,

$$h_j^{(k)} = b_j^{(k)} + \sum_{i=1}^{N_{k-1}} w_{j,i}^{(k)} h_i^{(k-1)} \tag{58}$$

where $h_j^{(k)}$ is the jth neuron in the kth layer of the network. A layer can then be more conveniently expressed in matrix form as

$$\mathbf{h}^{(k)} = W^{(k)} \mathbf{h}^{(k-1)} + \mathbf{b}^{(k)} \tag{59}$$

where, if there are N_k neurons in the kth layer, the output is

$$\mathbf{h}^{(k)} = \left[h_1^{(k)}, \ \ldots, \ h_{N_k}^{(k)} \right]^T \tag{60}$$

and (j, i)th entry in the matrix $W^{(k)}$ is $w_{j,i}^{(k)}$. If the output of a layer of neurons are sent directly to the next layer, the operation could be simplified to one single layer

$$\mathbf{h}^{(k+1)} = W^{(k+1)} \left(W^{(k)} \mathbf{h}^{(k-1)} + \mathbf{b}^{(k)} \right) + \mathbf{b}^{(k+1)} \tag{61}$$

$$= \left(W^{(k+1)} W^{(k)} \right) \mathbf{h}^{(k-1)} + \left(W^{(k+1)} \mathbf{b}^{(k)} + \mathbf{b}^{(k+1)} \right) \tag{62}$$

There is therefore a need to add a non-linear function. Two of the commonest choices of activation functions are the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{63}$$

and the Rectified Linear Unit (ReLu) function

$$\text{ReLu}(x) = \max(x, 0) \tag{64}$$

More choices are however of course available [22]. It is shown in [23], that any continuous function may be arbitrarily well approximated using a finite number of neurons with the sigmoid function used as activation functions. In [24] this is shown to hold for any non-linear activation function, given enough layers in the NN. Each layer in the NN can now be expressed as

$$\mathbf{h}^{(k)} = f(W^{(k)} \mathbf{h}^{(k-1)} + \mathbf{b}^{(k)}) \tag{65}$$
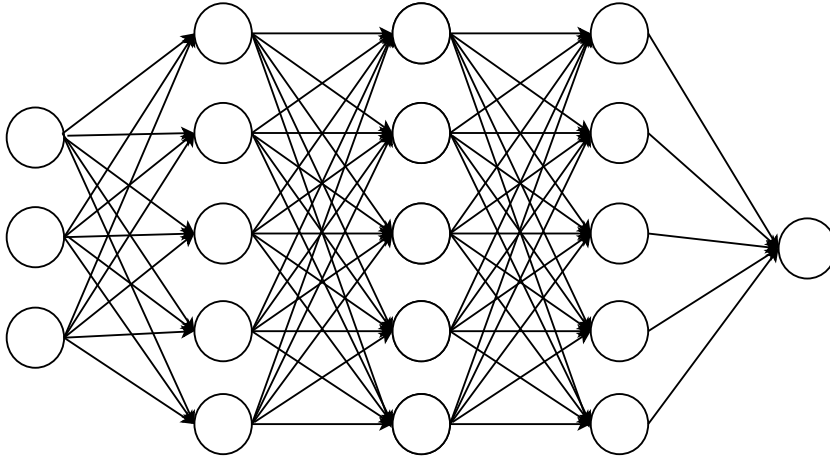
Figure 10: A simple example of neural network structure.

for some chosen activation function $f(x)$. Multiple layers may then be connected, where each the result from one is passed through an activation function and on to the next layer. An example of a NN structure is depicted in Figure 10.

If the matrices defining each layer are optimized without any restrictions (called a fully connected NN), there is a high risk of over-fitting. The parameters in the NN might be drawn to identify features specific to the examples in the training dataset, rather than a more general predictor. A common restriction to apply to the matrices is to enforce them to be Toeplitz. This is a matrix formulation of convolution, which also makes for a good interpretation of image feature extraction.

The last layer consists of a well chosen cost function, where the structure of the cost function depends on what problem the NN is set up to solve. If the NN is used to solve a regression problem, the network may simply output the number of values one wants to estimate. A distance measure may be applied to each of the values against the known true values. If a classification problem is considered, the output is commonly a vector with estimated probabilities of that an observation belongs to each of the possible classes. A chosen distance measure is then applied between the vector of probabilities and vector with zeros on all positions except for the correct class which contains a one.

One can now pass a set of observations through the network, and evaluate the result using the chosen measure. By calculating the gradient of the measure

and exploiting the chain rule, one can now estimate how each parameter in the network should be updated in each layer. Commonly, the gradients are estimated using a smaller set of observations, also called a batch, and a small gradient step is taken. A new batch is then evaluated, and parameters are updated again, and this procedure is continued until convergence. There are multiple schemes for choosing step-sizes when updating the parameters. An early method was stochastic gradient descent where a constant stepsize is chosen, and a small step is simply taken in the gradient direction, as estimated by the latest batch. Newer versions often include a momentum term, where the so-called Adam-scheme has grown in popularity as it is fast and consistently achieves good results [25].

5 Outline of the Papers

This section gives an overview of the papers included in the thesis. Paper A, Paper D, and Paper E are published; the versions herein have editorial changes and minor mistakes are fixed. Paper B and Paper C are to be submitted.

Paper A: Parameter Estimation of Oscillating Gaussian Functions Using the Scaled Reassigned Spectrogram

In this paper we apply the scaled reassignment method to Gaussian bell functions. If a Gaussian window function is used when calculating the spectrogram, perfect concentration of the reassigned spectrogram can be achieved, with all mass in one single point, if the correct scaling factor is used. The correct scaling factor is a function of the shape parameters of both the Gaussian window and the observed Gaussian function. By measuring the entropy of the reassigned spectrogram, when tuning the scaling factor, we can find a parameter estimate of the observed function, when the entropy is minimized and (close to) perfect concentration is achieved.

The work in paper A is published in part as

> Maria Hansson-Sandsten and Johan Brynolfsson "The Scaled Reassigned Spectrogram with Perfect Localization for Estimation of Gaussian Functions", *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 100-104, 2015.

> Johan Brynolfsson and Maria Hansson-Sandsten, "Parameter Estimation of Gaussian Functions using the Scaled Reassigned Spectrogram", *23rd European Signal Processing Conference*, Nice, France, Aug 31 - Sept 4, 2015

and has been published in full as

> Johan Brynolfsson and Maria Sandsten, "Parameter Estimation of Oscillating Gaussian Functions Using the Scaled Reassigned Spectrogram". *Elsevier Signal Processing*, vol. 150, pp. 20-32, Sept 2018

My Contributions: MS had the main idea for the project. I proposed the optimization scheme and derived and investigated the statistical properties of the reassignment. I implemented the algorithm and did the simulations. The paper was written jointly.

25

Paper B: A Time-Frequency-Shift Invariant Parameter Estimator for Oscillating Transient Functions Using the Matched Window Reassignment

In the second paper we generalize the estimator in paper A to be applied to any twice differentiable, transient and band-limited function. We show that the spectrogram of any function holding the stated properties may be reassigned into one single point, if the window function used to calculate the spectrogram is a time-reversed version of the envelope of the observed signal. We also show that it holds for signals in any number of dimensions. Based on this we further generalize the parameter estimator and also present a classification scheme based on the matched reassignment.

The work in paper B is published in part as

> Maria Sandsten, Johan Brynolfsson, and Isabella Reinhold, "The Matched Window Reassignment", *26th European Signal Processing Conference*, Rome, Italy, Sept 3-7, 2018

> Johan Brynolfsson, Isabella Reinhold, Josefin Starkhammar and Maria Sandsten, "The Matched Reassignment applied to Echolocation Data", *44th International Conference on Acoustics, Speech, and Signal Processing*, Brighton, UK, May 12-17, 2019.

and the full paper is to be submitted.

My Contributions: MS had the main idea for the project. I derived the method for complex, multidimensional signals, and proposed the parameter estimator. I implemented the algorithm and did the simulations. IR constructed the classifier and did the corresponding simulations. The paper was written jointly.

Paper C: Least Squares Estimation of Smooth and Mixed Spectra

In the third paper, we consider the problem of estimating spectral representations which are assumed to be smooth. We consider three different cases; one-dimensional stationary signals, one-dimensional non-stationary signals, and two-dimensional stationary signals. Here we divide the spectral representation into segments, and assume the the spectral representation within each segment may be approximated as changing linearly. We then construct a fast least-squares based estimator as well as an optimizing scheme where we may update the location of the separate segments, also allowing for mixed spectra.

The work in Paper C is published in part as

> Johan Brynolfsson, Johan Swärd, Andreas Jakobsson, and Maria Hansson-Sandsten 'Smooth Time-Frequency Estimation Using Covariance Fitting", presented at: *39th International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 4-9, 2014.

> Johan Swärd, Johan Brynolfsson, Andreas Jakobsson, and Maria Hansson-Sandsten "Smooth 2-D Frequency Estimation using Covariance Fitting", *22nd European Signal Processing Conference*, Lisbon, Portugal, Sept 1-5, 2014.

> Johan Brynolfsson, Johan Swärd, Andreas Jakobsson, and Maria Sandsten, "Least Squares and Maximum Likelihood Estimation of Mixed Spectra", *26th European Signal Processing Conference*, Rome, Italy, Sept 3-7, 2018

and the full paper is to be submitted.

My Contributions: The project was a close collaboration between JS and me. The idea was conceived jointly. Derivation and implementations where done jointly for the one-dimensional case and I extended it to the two-dimensional and time-frequency cases. The paper was written jointly.

Paper D: Sparse Semi-Parametric Estimation of Harmonic Chirp Signals

In the fourth paper, we consider the problem of estimating possibly harmonic chirp signals. A harmonic sinusoidal model is often used for many naturally occurring signals, such as speech, but where inharmonicities are often seen. A harmonic chirp model has been suggested as a more fitting model. Previously proposed algorithms estimating harmonic chirp structures require both the number of fundamental chirps as well as the number of overtones. Here we assume that both are unknown and propose a semi-parametric, dictionary based algorithm. To avoid the bias introduced by the dictionary approach an additional non-linear least squares step is proposed.

The work in Paper D has been published in part as

> Johan Swärd, Johan Brynolfsson, Andreas Jakobsson, and Maria Hansson-Sandsten "Sparse Semi-Parametric Chirp Estimation", *The Asilomar Conference on Signals, Systems, and Computers*, Asilomar, USA, Nov 2-5, 2014.

and has been published in full as

> Johan Swärd, Johan Brynolfsson, Andreas Jakobsson, and Maria Hansson-Sandsten, "Sparse Semi-Parametric Estimation of Harmonic Chirp Signals". *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1798-1807, 2016.

My Contributions: The project was a close collaboration between JS and me. The idea was conceived jointly. Derivation and implementations where done jointly. JS had previously worked with the ADMM optimization scheme and implemented it for this paper too. The paper was written jointly.

Paper E: Classification of One-Dimensional Non-Stationary Signals Using the Wigner-Ville Distribution in Convolutional Neural Networks

In the fifth paper, we argue that when a time-frequency representation is used as input to a neural network in a classification setting, the Wigner-Ville representation could be used for better accuracy, as compared to a manually chosen spectrogram. By allowing the neural network optimization to tune a convolutional kernel function, the network may choose any member of the quadratic class.

The work in Paper E has been published as

> Johan Brynolfsson and Maria Sandsten "Classification of One-Dimensional Non-Stationary Signals Using the Wigner-Ville Distribution in Convolutional Neural Networks", *25th European Signal Processing Conference*, Kos, Greece, Aug 28-Sept 2, 2017.

My Contributions: I had the idea for the project and did the implementation and simulations. The paper was written jointly.

References

[1] Franz Hlawatsch and F. Auger, Eds., *Time-Frequency Analysis*, ISTE Ltd, 2008.

[2] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, N.J., 2005.

[3] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, 2005.

[4] E. Wigner, "On the Quantum Correction for Thermodynamic Equilibrium," *Physical Review*, vol. 40, pp. 749–759, 1932.

[5] J. Ville, "Théorie et Applications de la Notion de Signal Analytique," *Câbles et Transmissions*, vol. 2A, pp. 66–74, 1948.

[6] L. Cohen, "Generalized Phase-Space Distribution Functions," *Journal of Mathematical Physics*, vol. 7, no. 5, pp. 781–786, 1966.

[7] H. Choi and W.J. Williams, "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 6, pp. 862–871, 1989.

[8] M. Born and P. Jordan, "Zur Quantenmechanik," *Zeits. Physik*, vol. 34, pp. 858–888, 1925.

[9] A. M. Sayeed and D. L. Jones, "Optimal Kernels for Nonstationary Spectral Estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 478–491, Feb 1995.

[10] B.W. Gillespie and L.E. Atlas, "Optimizing Time-Frequency Kernels for Classification," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 485–496, March 2001.

[11] Z. Hong and B. Zheng, "An Adaptive-Kernel Design Method Based on Ambiguity Domain," in *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1998.

[12] K. Kodera, R. Gendrin, and C. Villedary, "Analysis of Time-Varying Signals with Small BT Values," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-26, no. 1, pp. 64–76, Feb 1978.

[13] F. Auger and P. Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[14] E. Chassande-Mottin, F. Auger, and P. Flandrin, *Time-Frequency Analysis*, chapter Reassignment, pp. 249–277, U.K.: ISTE, 2008.

[15] E. Chassande-Mottin, F. Auger, and P. Flandrin, "Statistque des Vecteurs de Reallocation du Spectrogramme," *Rapport interne no. 96-01, Laboratoire Physique, ENS-Lyon (URA 1325 CNRS), Lyon (France)*, 1996.

[16] E. Chassande-Mottin, P. Flandrin, and F. Auger, "On the Statistics of Spectrogram Reassignment Vectors," *Multimedia Systems and Signal Processing*, vol. 9, no. 4, pp. 355–362, 1998.

[17] F. Auger, E. Chassande-Mottin, and P. Flandrin, "Making Reassignment Adjustable: The Levenberg-Marquardt Approach," in *37th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Kyoto, Japan, 25-30 Mar. 2012.

[18] M. Hansson-Sandsten and J. Brynolfsson, "The Scaled Reassigned Spectrogram with Perfect Localization for Estimation of Gaussian Functions," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 100–104, January 2015.

[19] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[20] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l_1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[21] M. Yuan and Y. Lin, "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[23] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function," *Math. Control Signal Systems*, vol. 2, pp. 303–314, 1989.

[24] K. Hornik, "Approximation Capabilities of Mutlilayer Feedforward Networks," *Neural Networks*, 1991.

[25] D.P. Kingma and J.L. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations*, 2015.