PhD Thesis

Thesis · December 2013

1 author:

Sašo Muševič
University Pompeu Fabra
6 PUBLICATIONS **23** CITATIONS

Non-stationary sinusoidal analysis

Sašo Muševič

TESI DOCTORAL UPF / 2013

Directors de la tesi

Prof. Xavier Serra, Dr. Jordi Bonada
Department of Information and Communication Technologies

The court's PhD was appointed by the recto of the Universitat Pompeu Fabra on ....., 2010.

Chairman

Member

Member

Member

Secretary

The doctoral defense was held on ....., 2010,
at the Universitat Pompeu Fabra and scored as .....

PRESIDENT MEMBERS

SECRETARY

To Felix Arthur

Acknowledgements

In the three and a half PhD years, I have probably met more interesting people that I have in my entire life before that. It is impossible to thank everyone in particular, as another dissertation of the same size would be required. Thank you all - you know who you are!
There has been numerous exceptional people that I've had the privilege to meet and sometimes collaborate with. I would like to acknowledge the tremendous mark certain individuals have left on my personal life and academic career alike.
many thanks to Mark Plumbly and Axel Röbel who have showed great hospitality and interest in my work. The research stays at QMUL and IRCAM have made crucial advancements in my thesis.
I would also like to thank Wen Xue, QMUL for daily reviews of my scattered ideas.
A very warm and special thanks go to the people at MTG. I have enjoyed every single day in the office, the discussions, sparking ideas and clashes alike. Putting together such a vibrant and successful group of people requires an exceptional team leader with great perseverance. Xavier Serra has taken MTG from its beginnings through a quick expansion and through a financial crisis. Thank you for creating such a unique place.
What makes the biggest impact on ones life are not the flashy, one-off events, but people and places one sees every day. Big, big thanks go to the audio signal processing team: Ricard, Jordi J., Jordi B., Marti, Merlijn, Keita and Graham. Thank you for creating a fantastic academic and friendly atmosphere!
Very special thanks go to Jordi Bonada - you deserve the biggest accolade in

helping me get through my thesis! Your *crystal ball* ability to find solutions to my problems and digging up tiny bugs in thousand lines of my code on impossible deadlines has put the progress of my research back on track on quite a few occasions. I only wish to have your motivation and passion - this thesis would not have been finished without you.
On a personal note, I would like to mention a few very special people.
Leonardo Aldrey, you have been and still are the biggest unstoppable source of positive energy and inspiration I have ever seen. I can only hope that my advice to you has made a fraction of a positive impact that you presence has made on mine. It kept my head above water and gave me that much needed morning boost to face the day.
My mother and my sister you have wholeheartedly supported every move that I've ever made, including the move to Spain and eventually to Scotland, further away from you. I could not wish for a better mum and sis!
Last but not least, my biggest thanks go to my father. You have known what's best for me more than I did and you were there for me when I needed you the most, including helping me create a life on the other side of Europe. Thank you.

And finally:
Fiona, my best friend, my love, and mother of my child, you deserve this thesis as much as I do.
...and Felix: you haven't really done much apart from cry, sleep and bless us with a very occasional smile in the two weeks since you've been born, but I feel like I have to thank you already. Not for the sleepless nights, when trying to finish my PhD and looking after you has caused havoc, but for the great time we are going to have together.

Abstract

Many types of everyday signals fall into the *non-stationary sinusoids* category. A large family of such signals represent audio, including acoustic/electronic, pitched/transient instrument sounds, human speech/singing voice, and a mixture of all: music. Analysis of such signals has been in the focus of the research community for decades. The main reason for such intense focus is the wide applicability of the research achievements to medical, financial and optical applications, as well as radar/sonar signal processing and system analysis. Accurate estimation of sinusoidal parameters is one of the most common digital signal processing tasks and thus represents an indispensable building block of a wide variety of applications.

Classic time-frequency transformations are appropriate only for signals with slowly varying amplitude and frequency content - an assumption often violated in practice. In such cases, reduced readability and the presence of artefacts represent a significant problem. Time and frequency resolution cannot be increased arbitrarily due to the well known time-frequency resolution trade-off by Heisenberg.

The main objective of this thesis is to revise and improve existing methods, and to propose several new approaches for the analysis of non-stationary sinusoids. This dissertation substantially contributes to the existing sinusoidal analysis algorithms: a) it critically evaluates and disseminates in great detail current analysis methods, b) provides significant improvements for some of the most promising existing methods, c) proposes several new approaches for analysis of the existing sinusoidal models and d) proposes a very general and flexible sinusoidal model together with a fast, direct

estimator.

Resumen

Muchos tipos de señales que encontramos a diario pertenecen a la categoría de *sinusoides no estacionarias*. Una gran parte de esas señales son sonidos que presentan una gran variedad de características: acústicos/electrónicos, sonidos instrumentales harmónicos/impulsivos, habla/canto, y la mezcla de todos ellos que podemos encontrar en la música. Durante décadas la comunidad científica ha estudiado y analizado ese tipo de señales. El motivo principal es la gran utilidad de los avances científicos en una gran variedad de áreas, desde aplicaciones médicas, financiera y ópticas, a procesado de radares o sonar, y también a análisis de sistemas. La estimación precisa de los parámetros de sinusoides no estacionarias es una de las tareas más comunes en procesado digital de señales, y por lo tanto un elemento fundamental e indispensable para una gran variedad de aplicaciones.

Las transformaciones de tiempo y frecuencia clásicas son solamente apropiadas para señales con variación lenta de amplitud y frecuencia. Esta suposición no suele cumplirse en la prctica, lo que conlleva una degradación de calidad y la aparición de artefactos. Además, la resolución temporal y frecuencial no se puede incrementar arbitrariamente debido al conocido principio de incertidumbre de Heisenberg.

El principal objetivo de esta tesis es revisar y mejorar los métodos existentes para el análisis de sinusoides no estacionarias, y también proponer nuevas estrategias y aproximaciones. Esta disertación contribuye sustancialmente a los análisis sinusoidales existentes: a) realiza una evaluación crítica del estado del arte y describe con gran detalle los métodos de análisis existentes, b) aporta mejoras sustanciales a algunos de los métodos existentes más prometedores, c) propone varias aproximaciones nuevas para el análisis de los modelos sinusoidales existentes i d) propone un modelo sinusoidal muy general y flexible con un algoritmo de análisis directo y rápido.

Preface

A sinusoid, by far the most appropriate mathematical function to describe any form of vibration, occupies a rather important place in research concerning acoustics and sound. It should come as no surprise that human senses depend heavily on vibration.

Sound is a perfect example of information transmission in which sinusoids play a central role. It is widely believed that a mixture of sinusoids and noise (Serra, 1989) adequately describes sound signals as far as the human perception is concerned. This notion has been disputed and the idea of *transients* has been proposed as an integral part of sonically complete sound analysis (Verma and Meng, 1998, 2000). It is however also common knowledge that acoustic vibration in most cases comes with at least a pinch of non-linearity, causing the frequency and amplitude of the vibration to change rather fast. Membranes, strings, stiff objects as well as air pressure fluctuations exhibit what is often referred to as *stationary* vibration, only in very limited circumstances when the excitation force is almost negligible. In majority of real world cases however, the non-linear behaviour of vibrating objects produces signals that differ substantially from a stable, stationary sinusoid. At the same time it is interesting to observe how such discrepancies produce sounds that enjoy great interest of musicians and music lovers. It seems that inharmonicity and other forms of distortion caused by such non-linear behaviour stir human minds, making it more desirable where one would expect the opposite.

Surprisingly, a great deal of difficulties analysing acoustic sounds can be accounted to modulations sourcing from *artistic articulations*. Such articulations involve rapid changes of pitch and intensity. While the frequency of

the lower harmonic might not fluctuate significantly the frequency change rate of higher harmonics changes much quicker. For example, a voice vibrato can exhibit a frequency change rate of thousands of hertz per second for the 5th harmonic.

With the advent of widespread use of electronic instruments, be it analogue or digital, the range of sounds used in music production has extended enormously. Before the electronic music revolution, the amplitude and frequency modulations of instrument sounds were confined within the limits set by acoustical and physical properties of the instrument, as well as performer's skills. Later, analogue synthesisers were able to create sounds modulated beyond what was imagined before, and the arrival of purely digital sound sources and transformations gave musicians free hands on synthesizing virtually any sound imaginable. One can imagine the difficulties in facilitating robust and accurate analysis of such wide spectrum of sounds.

Probably the most difficult sounds to analyse are percussive ones. Such very short duration with typically very quickly changing amplitude and unstable/inharmonic frequency content have arisen doubts about adequacy of the aforementioned *sinusoidal + noise* model. The reason for that is the amplitude and frequency of such percussive sounds may change significantly in times shorter than one period of the signal. In such cases the signal does not resemble a sinusoid any more and is commonly referred to as *transient*. Dismissing the problem by labelling it as non-relevant for human perception has been largely criticised as percussive sounds - the most perceptually prominent class of transients, have always been an indispensable part of music in every culture around the world.

A rather unexpected family of non-stationary sinusoidal signal, commonly found in music arises from a pair of sinusoids with very similar frequencies. In such cases, the pair can be considered as a single sinusoid with rather specific amplitude and frequency modulation. Very little effort have been put into accurate resolution of such pairs of sinusoids, often called *overlapping*. In essence, it's an attempt to double the frequency resolution of the transform.

It is desirable that a framework for extracting parameters of sinusoids from a signal should be designed to deal with such *non-stationary* sinusoids. Methods based on the Fourier transform (FT) have been extensively used in such applications, but it's easy to recognise its intrinsic limits. An attempt to extend an analysis framework to include amplitude and frequency modulated sinusoids seems natural.

Numerous methods have been developed for the problem at hand. An attempt to compare and evaluate all of them would be a huge task, far beyond

the extent of a single PhD thesis. It is important to mention that a unified testing framework (as a set of synthetic modulated single sinusoids and absolute lower achievable error) has existed for a very long time (Rao, 1947), yet many *state-of-the-art* algorithms have not been tested and compared in such framework. An algorithm, performing well in such framework would be expected to perform well on real world examples. This mostly holds when analysing a recording of a single quasi-harmonic sound source of not very low frequency (relative to the length of the observation frame). Unfortunately, in the more complex cases, when more than one sound source is present, the *single isolated sinusoid* assumption can be violated to a large extent. An effort to reach an agreement on a unified sinusoidal modelling testing framework for such real world signals has not yet gained a widespread attention of the research community, however it seems crucial to take steps in the direction, as problem at hand.

In this thesis a specific family of non-stationary sinusoidal analysis methods will be analysed in depth. In particular, fast real-time algorithms are preferred above off-line CPU-intensive methods, even potentially sacrificing some accuracy. Generally, iterative improvement methods will be avoided when possible, due to difficulties of defining the convergence region when dealing with real world signals. Specifically, a number of state-of-the-art kernel based methods are described, evaluated and improved. In addition, a new family of estimators is proposed and evaluated.

This dissertation will cover two main sinusoidal models: a complex polynomial amplitude modulated complex sinusoid with exponential damping (cPACED) and the generalised sinusoid - a complex sinusoid with a log-amplitude and frequency modulation expressed as a linear combination of predefined model functions. Lastly, a hybrid sinusoidal model will be tackled by a subtle modification of existing analysis method. State-of-the-art methods were generalized and unified, extended for accurate analysis of transient-like signals. Next, a completely different approach of computing closed form expressions for the FT of cPACED using symbolic computing software is considered. The fact that multivariate polynomial systems can many times have very simple solutions is exploited to derive an estimator capable of accurately estimating parameters of sinusoids with extremely similar frequencies.

Contents

Abstract ix

Resumen x

Preface xi

List of Figures xviii

List of Tables xxiii

I Background 1

1 Problem definition 3
1.1 General Considerations 4
1.2 Short Time Fourier Transform 6
1.3 Cramer-Rao bounds 7

II State-of-the-art 11

2 Introduction 13
2.1 The energy reallocation 13
2.2 Least-Square estimators 14
2.3 High-resolution methods 18
2.4 Various 21

3 Generalised sinusoid analysis using derivatives 23
3.1 Reassignment and derivative method 23
3.2 Generalised reassignment 26
3.3 Distribution-derivative method 26
3.4 Estimating static parameters: amplitude and phase 28
3.5 Unified reassignment, derivative and distribution derivative algorithm 30

III Contribution 35

4 Practical and Theoretical Comparison of Generalised Reassignment, Derivative and Distribution Derivative methods 37
4.1 Theoretical equivalence of generalised reassignment and generalised derivative method 37
4.2 Per-Parameter Comparison 41
4.3 Signal-to-Residual-Ratio Comparison 56
4.4 Frequency dependent SRR Comparison 61
4.5 Low-frequency analysis with GRM and DDM 70
4.6 Conclusion 74

5 Reassignment with adaptive Fourier poly-phase kernel 75
5.1 GRM using a generic kernel 75
5.2 Polynomial-phase Fourier kernel 76
5.3 Tests and Results 78
5.4 Conclusion 81

6 cPACED analysis using Gamma function 83
6.1 Lower incomplete Gamma function and cPACED model . . . 84
6.2 Pole estimator using multiple windows 86
6.3 Complex polynomial amplitude estimator 89
6.4 Tests and Results 89
6.5 Conclusion 92

7 Non-stationary sinusoidal analysis using Chebyshev polynomials and Gröbner basis 95
7.1 Discrete Fourier-Chebyshev basis 99
7.2 System of multivariate polynomials 106
7.3 Solutions to various signal models 109

7.4 Tests and results . . . 116
7.5 Discussion . . . 119

8 Reassignment for cPACED 121
8.1 cPACED model . . . 121
8.2 Pole estimator using derivatives . . . 122
8.3 Complex polynomial amplitude estimator . . . 125
8.4 Tests and results . . . 125
8.5 Discussion . . . 127

9 DDM for a Hybrid Sinusoidal Model - Generalised Sinusoid with Complex Amplitude Modulation 129
9.1 Non-linear multivariate polynomial system of equations . . . 131
9.2 DDM for cPACED . . . 133
9.3 Tests and results . . . 134
9.4 Discussion . . . 136

10 Variable bandwidth DDM for bird song analysis 139

IV Conclusion 145

11 Conclusion 147
11.1 Contributions . . . 149
11.2 Future Work . . . 150

V Appendix 153

A Proofs of equations 155
A.1 Reassignment . . . 155
A.2 Derivative method . . . 159
A.3 Generalized reassignment . . . 162
A.4 Distribution derivative method . . . 168

B Some code 171

Bibliography 173

List of Figures

3.1 Maximal estimation error for all phases, when approximating the first derivative. Significant approximation accuracy is gained by doubling the derivation filter length 32

4.1 Reassignment method (reduced frequency range): log-amplitude estimation error 45
4.2 Reassignment method (reduced frequency range): linear log-AM estimation error 46
4.3 Reassignment method (reduced frequency range): phase estimation error 46
4.4 Reassignment method (reduced frequency range): frequency estimation error 47
4.5 Reassignment method (reduced frequency range): linear FM estimation error 47
4.6 Derivative method (reduced frequency range): log-amplitude estimation error 48
4.7 Derivative method (reduced frequency range): linear log-AM estimation error 48
4.8 Derivative method (reduced frequency range): phase estimation error 49
4.9 Derivative method (reduced frequency range): frequency estimation error 49
4.10 Derivative method: (reduced frequency range) linear FM estimation error 50

4.11 Generalized reassignment (reduced frequency range): log-amplitude estimation error . . . 50
4.12 Generalized reassignment (reduced frequency range): linear log-AM estimation error . . . 51
4.13 Generalized reassignment (reduced frequency range): phase estimation error . . . 51
4.14 Generalized reassignment (reduced frequency range): frequency estimation error . . . 52
4.15 Generalized reassignment (reduced frequency range): linear FM estimation error . . . 52
4.16 All methods (full frequency range): log-amplitude estimation error 53
4.17 All methods (full frequency range): linear log-AM estimation error 54
4.18 All methods (full frequency range): phase estimation error 54
4.19 All methods (full frequency range): frequency estimation error . 55
4.20 All methods (full frequency range): linear FM estimation error . 55
4.21 GRM: $Hann$ window, $dgr = 2$, SRR average and standard deviation for different zero-paddings . . . 57
4.22 GRM: $Hann^2$ window, $dgr = 2$, SRR average and standard deviation for different zero-paddings . . . 58
4.23 DDM: $Hann$ window, $dgr = 2, Q = 2$, SRR average and standard deviation for different zero-paddings . . . 59
4.24 DDM: $Hann$ window, $dgr = 2, Q = 5$, SRR average and standard deviation for different zero-paddings . . . 59
4.25 DDM: $Hann$ window, $dgr = 2, Q = 7$, SRR average and standard deviation for different zero-paddings . . . 60
4.26 CRB: SRR average and standard deviation for different zero-paddings . . . 60
4.27 DDM: $Hann$ window, $dgr = 2, Q = 2$, no zero-padding, SRR average for each tested frequency below 10kHz . . . 61
4.28 DDM: $Hann$ window, $dgr = 2, Q = 5$, no zero-padding, SRR average for each tested frequency below 10kHz . . . 62
4.29 DDM: $Hann$ window, $dgr = 2, Q = 7$, no zero-padding, SRR average for each tested frequency below 10kHz . . . 62
4.30 DDM: $Hann$ window, $dgr = 2, Q = 2$, no zero-padding, SRR average for each tested frequency above 10kHz . . . 63
4.31 DDM: $Hann$ window, $dgr = 2, Q = 5$, no zero-padding, SRR average for each tested frequency above 10kHz . . . 63
4.32 DDM: $Hann$ window, $dgr = 2, Q = 7$, no zero-padding, SRR average for each tested frequency above 10kHz . . . 64

4.33 DDM: $Hann$ window, $dgr = 2, Q = 2$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend) 65
4.34 DDM: $Hann$ window, $dgr = 2, Q = 5$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend) 65
4.35 DDM: $Hann$ window, $dgr = 2, Q = 7$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend) 66
4.36 GRM: $Hann^2$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency below 10kHz 67
4.37 GRM: $Hann^2$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency above 10kHz 67
4.38 GRM: $Hann$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend) 68
4.39 GRM: $Hann^2$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend) 69
4.40 Linear system conditioning: $dgr = 2$, no zero-padding, condition number average for each tested frequency 70
4.41 DDM - low frequency range (20-500Hz): $dgr = 2, Q = 2$, no zero-padding, $Hann$ window, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend 71
4.42 DDM - low frequency range (20-500Hz): $dgr = 2, Q = 5$, no zero-padding, $Hann$ window, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend 71
4.43 DDM - low frequency range (20-500Hz): $dgr = 2, Q = 2$, no zero-padding, $Hann^2$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend 72
4.44 DDM - low frequency range (20-500Hz): $dgr = 2, Q = 5$, no zero-padding, $Hann^2$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend 72
4.45 GRM - low frequency range (20-500Hz): $dgr = 2$, no zero-padding, $Hann$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend 73
4.46 GRM - low frequency range (20-500Hz): $dgr = 2$, no zero-padding, $Hann^2$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend 73

5.1 Group 1 (AM only) 79
5.2 Group 2 (FM only) 80
5.3 Group 3 (AM and FM) 80

6.1 GAMMA 3^{rd} degree cPACED - full frequency range (200-2000Hz): $Hann, Hann^2, Hann', (Hann^2)'$ window set, SRR average 91
6.2 GAMMA 3^{rd} degree cPACED - full frequency range (200-2000Hz): $Hann, Hann^2, Hann', (Hann^2)'$ window set, SRR average for each tested frequency 91
6.3 GAMMA 3^{rd} degree cPACED - low frequency range (20-500Hz): $Hann, Hann^2, Hann', (Hann^2)'$ window set, SRR average for each tested frequency 92

7.1 Above: product ($s = s_{sn}s_s$). Middle: stationary factor (s_s). Below: non-stationary factor (s_{sn}). 97
7.2 Comparison of the condition number an alternant matrix derived from different functions sets. Above: all function sets. Below: a zoom-in with only discrete Fourier and discrete Chebyshev polynomials constructed by the recursion on the degree for $k = 0$ and the recursion on the time variable for $k > 0$. 104
7.3 Approximation of the non-stationary sinusoid defined by 7.42 and 7.43 with Fourier-Chebyshev functions of ascending degree. . 105
7.4 L_1 norm of the signal in the transform domain. Left: log scale, first 100 coefficients. Right: linear scale, first 10 coefficients. . . . 106
7.5 Approximation of a function by polynomials with polynomials. From such polynomial approximation The signal and its approximations are plotted on the x-y plane and the same plane is used for a real-complex plots for the roots. This way the real roots will have no y-axis component and will closely match the zeros of the function. The real roots are marked as '+' and the complex ones with '$*$'. 109
7.6 Relative error of the estimators with respect to *non-stationarity*. The FM/log-AM is jointly increased with higher values on the x-axis. The dashed line corresponds to 1% accuracy. 111
7.7 Error of joint estimation of the frequencies of two complex stationary overlapping sinusoids of equal amplitude. The relative phase difference at time 0 remains constant. Above: the $\hat{q}$ was set to the magnitude peak of the spectrum. Bottom: the $\hat{q}$ was set to the average of the frequencies of the sinusoids. The dashed line marks the 1% accuracy limit. 113

7.8 Magnitude spectrum of the original Hanning window function, it's time derivative, and time ramped versions, all required by reassignment. . . . 115
7.9 Magnitude spectrum of the original Hanning window function multiplied by Chebyshev polynomials . . . 116
7.10 SRR for single sinusoid model analysis. . . . 118
7.11 SRR for twin sinusoid model. . . . 120

8.1 Amplitude polynomial and exponential damping estimates separately (above) and cumulative (below). . . . 123
8.2 SRR: Mean and variance . . . 127

9.1 SRR between two sinusoids with significantly different parameters 24dB. . . . 131
9.2 SRR: Mean and variance . . . 136

10.1 Birdchirp, approximately 100kHz/s . . . 140
10.2 Bird sound, high FM and stationary sections show . . . 141
10.3 Bird chirp: DDM, Q=2 . . . 142
10.4 Bird chirp: DDM, Q=8 . . . 142
10.5 Bird chirp: DDM, Q=16 . . . 143
10.6 Bird chirp: DDM, Q=32 . . . 143
10.7 Bird chirp: DDM, Q=2, sparse representation . . . 144
10.8 Bird chirp: DDM, Q=16, sparse representation . . . 144

List of Tables

2.1 Estimation results of chirp rate presented in (Pantazis et al., 2009a) 18

Part I

Background

Chapter 1

Problem definition

We have to remember that what we observe is not nature in itself but nature exposed to our method of questioning.

The reality we can put into words is never reality itself.

Physics and Philosophy,1958.
Werner Heisenberg

Last years have seen a dramatic increase in demand of high-accuracy sounds analysis (Hainsworth et al., 2001; Badeau et al., 2006; Bonada, 2008; Betser, 2009; Wen and Sandler, 2009). Many divergent fields have expressed a necessity for high-accuracy analysis of non-stationary sinusoids. Prosody modelling and realistic voice synthesis (King, 2010; Tokuda et al., 2013) rely heavily on robust analysis, most of bio-acoustic signals seem to be highly non-linear (Stowell et al., 2013), and other fields non-related to acoustics such as seismic signal analysis (Bjerhammar, 1951), magnetic resonance imaging (Pattichis et al., 2000), spectrometry (Hänsch et al., 1979) and synthetic aperture radar technology (Li and Stoica, 1996) are facing, in its core, exactly the same problems. It seems that regardless of the nature of oscilation, be it air pressure, electro-magnetic field or heart beat - the problem of quick frequency and amplitude changes will pose a significant obstacle in achieving acceptable analysis results.
Since the development of the first non-stationary sinusoidal analysis meth-

ods, the well-known time-frequency (TF) ambiguity gave birth to an unexpected dilemma. Observing a non-stationary phenomena, the same signal could be explained as a single highly-modulated or two moderately modulated sinusoids very close in frequency. Both explanations can be equally accurate resulting in an ambiguity that represents a problem even for methods with inherently high frequency resolution like Empirical Mode Decomposition (EMD) (Rilling and Flandrin, 2008; Borgnat et al., 2010; WU et al., 2011). The frequency and amplitude content of such *non-stationary* signals may change rapidly enough (Wen and Sandler, 2010) for classic time-frequency (TF) analysis methods to result in *coloured* distributions (Daubechies et al., 2011). It is important to distinguish sinusoidal parameter *estimation* from *representation*. For most analysis methods that utilise a test function, sometimes called a *kernel* or an *atom*, the difference between the estimation and representation can be substantially blurred. A small set of kernels is commonly required for a single estimation. Typically the kernels centred around certain frequency give very similar estimates - they *snap* towards a nearby sinusoid, a phenomena reminiscent of the gravitational pull. This way, the TF energy in vicinity of a sinusoid is *squeezed* closer to the sinusoids, effectively modifying the representation by increasing readability and clarity. The most common estimation method of this kind is the *reassignment* method (Kodera et al., 1976).

1.1 General Considerations

This section presents mathematical definitions for the problems at hand. Throughout this document, certain assumptions are made in order to avoid repetitions thus preserving compactness and clarity.

A *frame based parameter estimator* is any method capable of estimating any number of parameters from a single time-limited sampled sequence (frame) of the signal under study. It is important to stress that any information about the signal in times before or after the frame is assumed to be unknown. That may seem a rather *pessimistic* assumption, however in many practical applications such information is either unavailable or its accuracy cannot be determined. Rather than relying on potentially inaccurate assumptions, an attempt to construct a sufficiently good estimator that does not require assumptions seems reasonable.

A *non-stationary* sinusoid is a sinusoid with varying amplitude and/or frequency. It is very common and practical to express it in the complex domain:

$$s(t) = A(t)e^{j(\phi(t)+\phi_0)}, \tag{1.1}$$

where $A(t) \in \mathbb{R}$ is the *amplitude function*, $\phi(t) \in \mathbb{R}$ the *phase function*, ϕ_0 the *initial phase* and t the time. A very important is the notion of an instantaneous frequency defined as the time derivative of phase: $\omega(t) = f'(t)$. In cases where $A(t)$ or $\omega(t)$ are not constant with respect to time, the signal $s(t)$ is said to be a *non-stationary complex sinusoid*. A non-stationary sinusoid is said to be *amplitude modulated* (AM) if $A(t)$ is not constant and *frequency modulated* (FM) if $\omega(T)$ is not constant. Since human perception of sound intensity is known to be logarithmic, $A(t)$ can be conveniently defined as a real exponential function, producing a compact representation of $s(t)$ (Wen and Sandler, 2009):

$$s(t) = e^{p(t)+jq(t)} = e^{r(t)}, \tag{1.2}$$

where $q(t)$ is the *phase function*, $p(t)$ is commonly referred to as *log-amplitude* (log-AM) and $r(t)$ is a complex valued function of time, combining phase and amplitude trajectories.
Until present day, the only mathematically tractable models are one of the following:

- $e^{r(t)}, r(t) \in \mathbb{C}$ where $r(t)$ is a polynomial - *generalised sinusoid*
- $A(t)e^{\alpha+j\omega t}, A(t) = \sum_k a_k t^k, \omega, \alpha \in \mathbb{R}$ - *PACED* (Polynomial Amplitude-Complex Exponential with exponential Damping if $A(t) \in \mathbb{R}$), or *cPACED* if $A(t) \in \mathbb{C}$ and PACE/cPACE if $\alpha = 0$

More models exist, for instance a sinusoidal log-AM/FM (Wen and Sandler, 2009) by simply substituting the polynomials with trigonometric functions. Such model sadly isn't tractable without an a-priori estimate of the modulation frequency, reducing the algorithm into a grid-search type - an alternative was not yet proposed.
The generalised sinusoid and cPACED models encompass a wide variety of signal families, thus an additional categorisation is welcome. A complex sinusoid with a log-AM/FM function $r(t)$ of a n^{th} degree polynomial will be denoted as a n^{th} *degree generalised sinusoid*. Historically, the research has many times focused on cases where the degree of amplitude function is not the same as that of phase function: as it will be shown such models do not offer measurable advantage, but will be covered for the sake of completeness. In such cases the degree of amplitude polynomial is a degree lower than that of the phase polynomial, but equal to the degree of the instantaneous frequency polynomial.

Similarly, a n^{th} *degree cPACED* model has a n^{th} degree polynomial amplitude function $A(T)$. In the research community the amplitude and phase functions are erroneously referred to as AM/FM (or log-AM/FM) for brevity - the notation will be adopted in order to comply with the rest of publications.
Some low-degree cPACE methods (eg: up to second degree polynomial Pantazis et al. (2011)), require initial frequency estimates for all the sinusoids in the signal, which implies successful identification of all salient peaks. The benefit is a joint estimation in the least-square sense and enabling the use of a window function to minimise the inter-sinusoid interference. Alternatively, the methods based on rotational invariance require little extra parameters apart from the signal itself. The cPACED model was shown to be the most general one still tractable by methods based on rotational invariance Badeau et al. (2006). Many versions of such methods have been successfully used in various audio coding/analysis applications Jensen et al. (2004); Hermus et al. (2005), the main advantage is overcoming the time-frequency resolution trade-off, this is inherent in the Fourier Transform (FT). It will be shown however, that such methods bear significant computational burden. It is therefore desirable to construct an efficient method, able to estimate the parameters of a high degree cPACED model.

1.2 Short Time Fourier Transform

In order to obtain frequency domain information, the Time Fourier Transform (STFT) is defined as follows (Marchand and Depalle, 2008):

$$S_w(t,\omega) = \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} s(\tau)w(\tau - t)\exp\big(-i\omega(\tau - t)\big)d\tau, \tag{1.3}$$

where $w(t) \in \mathbb{R}$ is a time limited window function of length T: $w(t) = 0, |t| > \frac{T}{2}$. The above expression can be represented in the common *log-polar* form:

$$S_w(t,\omega) = \exp\big(\lambda(t,\omega) + j\phi(t,\omega)\big), \tag{1.4}$$

$\lambda(t,\omega)$ being the *spectral amplitude* and $\phi(t,\omega)$ is *spectral phase*, many times referred to as the *magnitude* and *phase* spectrum. It is important to note that $\lambda(t,\omega), \phi(t,\omega)$ are functions of frequency and time as well, where the time variable designates the center of the observation frame, rather than a time instant within the frame. Setting $t = 0$ (or equivalently, shifting the signal, so that it always holds $t = 0$) gives, without any loss of generality, a

more widely used representation:

$$S_w(0,\omega) = S_w(\omega) = \int_{-\frac{T}{2}}^{\frac{T}{2}} s(\tau)w(\tau)e^{-j\omega\tau}d\tau. \tag{1.5}$$

A generalisation of STFT can simply be achieved by realising equation 1.3 is in fact an *inner product* of the signal and a windowed stationary complex exponential:

$$S_w(t,\omega) = \langle s(t), w(t)e^{j\omega t}\rangle, \tag{1.6}$$

where the window function is assumed to be non-zero only for $t \in \left[-\frac{T}{2}, \frac{T}{2}\right]$ acting as the limits of integration in equation 1.3. It is now trivial to generalise the complex exponential to an arbitrary *kernel* $\Psi(t)$:

$$\langle s(t), w(t)\Psi(t)\rangle. \tag{1.7}$$

Adopting the use of an arbitrary kernel allows for a very important flexibility when the signal under study does not correlate strongly with a stationary complex exponential, inevitably leading to numerical problems.

1.3 Cramer-Rao bounds

Every real-world signal contains some degree of white additive Gaussian noise which limits the best achievable accuracy of estimators. The so called *Cramer-Rao bounds* (CRB) (Cramér, 1946; Rao, 1947; Kay, 1946) represent theoretically optimal accuracy achievable by any estimator. Such bounds depend on the signal model and naturally on the *signal-to-noise ratio* (SNR). Specifically the CRBs express *the lower bound on the variance of the estimates* which in case of unbiased estimator corresponds to an inverse of the *Fisher information matrix* (FIM) (Edgeworth, 1908). For model parameters θ_n the FIM is defined in terms of *expected values*:

$$\mathbf{F_{kl}} = -E\left[\frac{\partial^2\Lambda}{\partial\theta_k\theta_l}\right], \tag{1.8}$$

where Λ denotes a log-probability density function of the signal under study s and the assumed model x (containing the parameters θ_n), corrupted by white Gaussian noise (eg: the SNR) of variance σ:

$$\Lambda = -\frac{1}{\sigma^2} < s - x, s - x >, \tag{1.9}$$

As already mentioned earlier in this section, the signal model 1.1 can be divided in two significantly different families, for which the CRBs have been derived (Stoica and Arye, 1989; Friedlander and Francos, 1993a). It is important to distinguish CRBs for single sinusoid and two (or more) sinusoids. Essentially, the multi sinusoid CRBs converge to the single sinusoid ones when the frequency difference between the sinusoids gets large enough. However, analytic derivation of a multi sinusoid CRB is substantially more complex (Stoica and Arye, 1989), forcing researchers to use approximations (Swingler, 1993). Further, the multi sinusoid CRBs depend significantly on specific combinations of parameter values of the signal under study, not only on the SNR, making them rather impractical for evaluation.
Despite that, CRBs were used extensively for evaluating single sinusoid estimators. The main reason is the effect of model parameter values on CRBs is negligible compared to that of SNR (Guotong Zhou, 1996). CRBs of a traditionally very common non-stationary single sinusoidal signal model with first degree log-AM and second degree FM (Abe and Smith, 2005; Marchand and Depalle, 2008; Wen and Sandler, 2009; Muševič and Bonada, 2010b; Hamilton et al., 2009)

$$s(t) = \exp\left(a_0 + \mu_0 t + j(\phi_0 + \omega_0 t + \psi_0 \frac{t^2}{2})\right), \tag{1.10}$$

depend on expressions:

$$\epsilon_k(\underline{\mu}_0, N) = \sum_{n=0}^{N-1} \left(\frac{n-n_0}{N}\right)^k \exp\left(2\underline{\mu}_0 \frac{n-n_0}{n}\right) \tag{1.11}$$

$$D_1(\underline{\mu}_0, N) = 2(\epsilon_0\epsilon_2 - \epsilon_1^2) \tag{1.12}$$

$$D_2(\underline{\mu}_0, N) = 2(\epsilon_0\epsilon_2\epsilon_4 - \epsilon_1^2\epsilon_4 + 2\epsilon_1\epsilon_2\epsilon_3 - \epsilon_2^3), \tag{1.13}$$

where $\underline{\mu}_0$ is the linear log-AM normalized with sampling frequency $\underline{\mu}_0 = \frac{\mu_0}{fs}$ and n_0 is the sample at which estimation is made. The optimal choice is at the centre of frame (Djuric and Kay, 1990) which is a common choice for most estimators. The crucial piece of information is the value of ϵ_k depends on the normalized log-AM (rather than non-normalized one), which greatly reduces the dependence of ϵ_k on log-AM (non-normalized) as $2\underline{\mu}_0 \frac{n-n_0}{n} \approx 0$.

The approximate CRB expressions are:

$$CRB_{\underline{a}_0,N}(\sigma,\underline{\mu}) \approx \frac{\sigma^2\epsilon_2}{D_1} \tag{1.14}$$

$$CRB_{\underline{\mu}_0,N}(\sigma,a_0,\underline{\mu}) \approx \frac{\sigma^2\epsilon_0}{a_0^2N^2D_1} \tag{1.15}$$

$$CRB_{\underline{\phi}_0,N}(\sigma,a_0,\underline{\mu}) \approx \frac{\sigma^2(\epsilon_2\epsilon_4-\epsilon_3^2)}{a_0^2D_2} \tag{1.16}$$

$$CRB_{\underline{\omega}_0,N}(\sigma,a_0,\underline{\mu}) \approx \frac{\sigma^2(\epsilon_0\epsilon_4-\epsilon_2^2)}{a_0^2N^2D_2} \tag{1.17}$$

$$CRB_{\underline{\psi}_0,N}(\sigma,a_0,\underline{\mu}) \approx \frac{\sigma^2(\epsilon_0\epsilon_2-\epsilon_1^2)}{a_0^2N^4D_2} \tag{1.18}$$

Complete derivations of above expressions is of limited significance since they are merely an approximation. Some interesting conclusions can be drawn however, most notably that all CRBs (except the one for amplitude) depend on amplitude, linear log-AM and SNR and not of FM.
In practice it is possible to compute above values numerically for much wider family of signals via inversion of the matrix defined by the equation 1.8. Although such procedure is more CPU intensive it will be used throughout this document, as it is not a part of the estimators itself, can generally be precomputed and will accommodate large number of different models. Under more general conditions when $p(t), q(t)$ from 1.2 are arbitrary functions, the elements of FIM (equation 1.8) consist of:

$$\frac{\partial\Lambda}{\theta_k} = \frac{2}{\sigma^2}\Re\left[\left\langle s-x, \frac{\partial x}{\partial\theta_k}\right\rangle\right] \tag{1.19}$$

$$\frac{\partial^2\Lambda}{\theta_k\theta_l} = -\frac{2}{\sigma^2}\Re\left[\left\langle \frac{\partial x}{\partial\theta_k}, \frac{\partial x}{\partial\theta_l}\right\rangle - \left\langle s-x, \frac{\partial^2 x}{\partial\theta_k\theta_l}\right\rangle\right] \tag{1.20}$$

$$\tag{1.21}$$

Since $E[s-x]$=0, further simplification is possible:

$$\frac{\partial\Lambda}{\theta_k} = \frac{2}{\sigma^2}\Re\left[\left\langle \frac{\partial x}{\partial\theta_k}, \frac{\partial x}{\partial\theta_l}\right\rangle\right]. \tag{1.22}$$

In case that θ_k, θ_l are parameters of $p(t), q(t)$ respectively, then $\left\langle \frac{\partial x}{\partial\theta_k}, \frac{\partial x}{\partial\theta_l}\right\rangle$ is purely imaginary and the corresponding $F_{k,l} = 0$. The amplitude and phase Fischer information matrices are therefore independent and can be

computed separately:

$$F_{k,l} = \frac{2}{\sigma^2} \sum_t e^{2p(t)} \frac{\partial p}{\partial \theta_k} \frac{\partial p}{\partial \theta_l}, \tag{1.23}$$

where θ_k, θ_l are both parameters of $p(t)$ and:

$$F_{k,l} = \frac{2}{\sigma^2} \sum_t e^{2p(t)} \frac{\partial q}{\partial \theta_k} \frac{\partial q}{\partial \theta_l} \tag{1.24}$$

where θ_k, θ_l are both parameters of $q(t)$. Specifically, when $p(t), q(t)$ are polynomials the corresponding elements of the Fischer matrix are:

$$F_{k,l} = \frac{2}{\sigma^2} \sum_t t^{k+l} e^{2p(t)}. \tag{1.25}$$

Since all the parameters are normalized with the sampling frequency, the assumption $2p(t) \approx 0$ is not too far from the truth.
Under such assumptions a performance of an algorithm can be summarised in a single plot per estimation parameter.
CRBs for close frequency sinusoids are analytically far more complex even for constant amplitude chirps (Yau and Bresler, 1992; Badeau et al., 2008a), but still follow the same general expression 1.8.

Part II

State-of-the-art

Chapter 2

Introduction

2.1 The energy reallocation

Many different approaches have been taken in order to tackle the TF trade-off of such signal representations. A very old idea of the *reassignment* has recently seen numerous enhancements and generalisations (Auger et al., 2012; Auger and Flandrin, 1995a; Kodera et al., 1976; Chassande-Mottin et al., 2010; Nilsen, 2009). For the purpose of this document, a more general notion of *reallocation* will be used. Reallocating spectral energy of a 1 or 2 dimensional TF distributions is the main topic of this thesis. The purpose of such reallocation is twofold: estimation of parameters of underlying non-stationary sinusoids and, using the same procedure on the whole frequency range, an enhanced TF representation of the non-stationary sinusoids. By *enhanced representation of non-stationary sinusoids*, reduction of artefacts and compact energy representation is assumed. Different model assumptions will lead to different estimators and representations, sharing the same core idea of the energy reallocation.
The step from parameter estimation to representation deserves an explanation. The reallocation in TF plane designates a *change of the frequency and time variable* from one point to another. The reallocation of time variable is often neglected and considered of a lesser importance. Frequency reallocation is a *change of initial frequency to the estimated one*. This initial frequency is the only parameter of reallocation methods, and due to CPU friendly nature the whole frequency range can be reallocated utilising fast algorithms like FFT. In practice, the frequency variable is *binned* (discretized to a pre-defined value set) in order to use FFT, however reas-

signed frequency values aren't. Often, the reassigned frequency is artificially discretized to the same resolution as the initial spectrum, however higher resolutions with smoothing could potentially yield better results. If not stated otherwise, the reassigned frequency will be discretized to the same resolution as the initial FFT bins for the purpose of this thesis.
Blind assumption that certain TF region contains a sinusoid and performing the reallocation analysis regardless might result in erroneous estimates for non-stationary parameters and phase. The overall magnitude of such reallocation however will remain low - a very desirable feature, as bulk processing without any a-priori knowledge and therefore no input parameters will yield a *consistent* result, in a sense that no phantom energy will be created in the TF regions that do not contain any sinusoids. Most of the algorithms require a solution of a linear system of equations for each initial frequency. This might seem CPU intensive, however the derivation of these linear systems is such that matrix pivoting (used by the LU decomposition) does not affect numerical stability and is thus not required. In such case, block operations allow the solution of thousands of linear systems at once, heavily relying on underlying *single instruction multiple data* (SIMD) CPU architecture present in every modern processor.
The topology of reallocated distributions has only lately been explored in Auger et al. (2012). In some specific cases when analytical expressions for reallocated distributions exist, an interesting, but not surprising behaviour, similar to attraction regions in gravitational fields was observed. The energy in the vicinity (frequency wise) of a sinusoid is reallocated closer to the sinusoids, leaving some regions completely devoid of energy. Clearly, a sinusoid acts as a mass attracting other nearby mass.
The notion of *equal noise band-width* (ENBW) (Harris, 1978; Nuttall, 1981) plays an important role in practical considerations as it represents a main-lobe width and side attenuation jointly in a single measure. In all the presented methods more than 1 window function will be required. The *effective* ENBW will be defined as a ENBW of a sum of all kernels used (window function is assumed to be a part of the kernel function). Such measure will accommodate for the fact that in some cases the same window shifted in frequency is used.

2.2 Least-Square estimators

A comprehensive summary of LS-based amplitude estimators of stationary sinusoids in (Stoica et al., 2000) outlines a Matched Filter (MAFI) frame-

work which is shown to be more general form of LS and Weighted LS (WLS) methods (Stoica et al., 1998) and also entails the general spectral estimators such as Amplitude-Phase Estimator (APES) (Li and Stoica, 1996) and Capon (Capon, 1969), both originating from the radar imaging research community.
An LS estimator in its most direct form allows for joint estimation of stationary complex sinusoids:

$$\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-2] \\ s[N-1] \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\omega_1} & \cdots & e^{j\omega_K} \\ \vdots & \ddots & \vdots \\ e^{j(N-2)\omega_1} & \cdots & e^{j(N-2)\omega_K} \\ e^{j(N-1)\omega_1} & \cdots & e^{j(N-2)\omega_K} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{K-1} \\ \alpha_K \end{bmatrix} + \begin{bmatrix} v[0] \\ v[1] \\ \vdots \\ v[N-2] \\ v[N-1] \end{bmatrix} \tag{2.1}$$

where $s[n] \in \mathbb{C}$ are the sampled signal values, $v[n]$ residual samples, N the length of observation frame and the signal model consist of K exponentials:

$$s[n] = \sum_{k=1}^{K} \alpha_k e^{j\omega_k}. \tag{2.2}$$

The equation 2.1 can be compacted into a vector notation:

$$\mathbf{s} = \tilde{\mathbf{A}}\tilde{\alpha} + \mathbf{v}, \tag{2.3}$$

to allow a matrix notation for the LS estimator:

$$\tilde{\alpha} = (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H \mathbf{s}, \tag{2.4}$$

where noise vector was left out. A very attractive ability to estimate multiple sinusoids on LS basis sadly implies the restriction of stationary sinusoids. The discussion following outlines an attempt to overcome this restriction.
It is important to try at all cost to avoid approximations when designing an estimation algorithm. Such procedure inevitably leads to inaccuracy, however sometimes the desired model is simply not tractable in its original form. The Quasi-Harmonic Model (QHM) (Pantazis et al., 2008, 2009b,a, 2011; Kafentzis et al., 2012) is based on a truncated 2^{nd} degree Taylor series approximation of a modulation function of a complex signal model:

$$s(t) = A(1 + \gamma_1 t + \gamma_2 t^2) e^{j(\phi_0 + \omega_0 t + \phi_1 t + \phi_2 t^2)} \tag{2.5}$$

$$\approx A e^{j\phi_0} \left(1 + (\gamma_1 + j\phi_1)t + \left(\gamma_2 - \frac{\phi_1^2}{2} + j(\phi + \gamma_1\phi_1)\right) t^2\right) e^{j\omega_0 t}, \tag{2.6}$$

where all the parameters are real valued and ω_0 is a known rough estimate of a frequency (Kay, 1988; Stoica and Moses, 1997) and ϕ_1 is the mismatch that is to be estimated. Above equation can be rewritten:

$$\hat{s}(t) = (a + bt + ct^2)e^{j\omega_0 t}, \tag{2.7}$$

with parameters $a, b, c \in \mathbb{C}$ - a well known cPACED model is recognised. Clearly, the initial model 2.5 drastically differs from the final approximation above, resulting in the need to iteratively re-estimate the parameter estimations. Generally, there is no guarantee that such iterative procedure would indeed converge to the correct values, especially when multicomponent signals are analysed (Pantazis et al., 2011). By far the most important is the frequency estimate, as the amplitude parameters can be estimated with an arbitrary accuracy as long as the frequency estimate is sufficiently good (Kay, 1993).
The parameters can now be estimated via LS by formulating a matrix equation:

$$s(t) = \overbrace{[E_1 E_2 E_3]}^{E} \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \tag{2.8}$$

where

$$E_k = \left[t^k e^{j\omega_0 t}\right], t \in [-\frac{T}{2}, \frac{T}{2}]. \tag{2.9}$$

It is common to apply a window function with the desired properties:

$$w(t)s(t) = w(t)\left(E \begin{bmatrix} a \\ b \\ c \end{bmatrix}\right). \tag{2.10}$$

The LS solution for a, b, c can now be obtained by matrix Moore-Penrose pseudoinverse (Moore, 1920; Bjerhammar, 1951; Penrose, 1955):

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = ((wE)^H wE)^{-1}(wE)^H ws. \tag{2.11}$$

Once a, b, c are estimated the corresponding estimates of the initial model 2.5 can easily be calculated. Importantly, the frequency mismatch ϕ_1 and frequency change rate ϕ_2 can be used to construct an updated matrix E from 2.9:

$$E_k = \left[t^k e^{j(\omega_0+\phi_t)t+j\phi_2 t^2}\right], t \in [-\frac{T}{2}, \frac{T}{2}]. \tag{2.12}$$

this effectively removes any existing frequency modulation and refines the frequency estimate which is crucial for eventually estimating of the amplitude parameters (Kay, 1993).
A major advantage of the simplified model 2.7 and the LS estimate 2.11 is a straightforward generalisation to joint estimation of multiple sinusoids. Such joint estimation requires a multi-component signal model, for example a harmonic one:

$$s(t) = \sum_{k=-K}^{K} (a_k + b_k t + c_k t^2) e^{jk\omega_0 t}. \tag{2.13}$$

Joint multi-component estimator can be derived by replacing window function w with a diagonal matrix W with diagonal elements of w (Stylianou, 1996) and replacing the single component parameters a, b, c with $\{a_k, b_k, c_k\}, k = -K..K$ yielding:

$$\begin{bmatrix} \vec{a} \\ \vec{b} \\ \vec{c} \end{bmatrix} = ((WE)^H WE)^{-1} (WE)^H W s. \tag{2.14}$$

Note that 2.14 includes negative frequencies to accommodate for real signals. Since a_k, b_k, c_k are complex, the model deviates from a pure harmonic one and it was therefore named Quasi-Harmonic. In fact, if the model 2.13 would have been modified to allow non-harmonic components the solution 2.14 expression would have not changed.
The method achieves satisfactory results analysing a synthetic harmonic signal. The first iteration misses the target by around 50% for values bellow 1000Hz/s, but the accuracy quickly drops even lower when chirp rate climbs towards 2000Hz/s. Iteration does however converge the estimate towards the correct value.
It is important to note that the above example is a noiseless harmonic signal with the sinusoids spaced quite far away from each other in the frequency domain, thus convergence is rather unsurprising. As pointed out in (Pantazis et al., 2009a) higher harmonics can potentially reach very high values (above 10.000Hz/s, a standard limit for most of spectral energy reallocation algorithms (Wen and Sandler, 2009)), clearly invalidating the assumption made in 2.5 and thus possibly breaking the convergence. In (Pantazis et al., 2009b, 2011) a first degree truncated Taylor approximation (as opposed to the second one described above) is considered in a similar iterative process and used for speech analysis.
The model 2.5 is a very general one, for which no direct estimation method

harmonic	chirp rate	1 iter	10 iter
1	200	55	200
2	400	194	400
3	600	353	599
4	800	457	800
5	1000	515	1000
6	1200	525	1199
7	1400	519	1400
8	1600	490	1600
9	1800	456	1799
10	2000	441	2000

Table 2.1: Estimation results of chirp rate presented in (Pantazis et al., 2009a)

exists to date. An attempt to handle such model was considered in (Friedlander and Francos, 1993b) via minimisation of the log-likelihood function over amplitude and phase parameters, achieving near-CRB accuracy for a single component case. It will however be shown in chapter 9 that a multivariate non-linear system for the model 2.5 can be constructed and solved analytically without approximations of any kind.

2.3 High-resolution methods

The methods were developed in XVIII. century to study gas expansion behaviour is *Prony's method* (de Prony, 1795; Hildebrand, 1956), later adapted for audio analysis (Tufts and Kumaresan, 1980a,b; Kumaresan and Tufts, 1980, 1982; Laroche, 1989). The main idea is to model the signal as a sum of L damped sinusoids

$$s(t) = \sum_{l}^{l=L} e^{\alpha t} \cos(\omega_l t + \phi_l), \alpha \in \mathbb{R} \tag{2.15}$$

by defining the following system:

$$A = \begin{bmatrix} s[K] & s[K-1] & \cdots & s[0] \\ s[K+1] & s[K] & \cdots & s[1] \\ \vdots & \vdots & \ddots & \vdots \\ s[N] & s[N-1] & \cdots & s[N-K] \end{bmatrix}, \tag{2.16}$$

where $s[k] = s(k\triangle t)$, N is the number of observed samples and $\triangle t = \frac{1}{fs}$ is the length of 1 sample. System A has the following two significant properties: it is singular $|A| = 0, N - K > 2L, K > 2L$ and its null-space (kernel) is of dimension $\dim(\ker(A)) = p + 1 - 2L$. Equivalently, for any vector $x[k] \in \ker(A)$ the following holds:

$$\sum_{k=0}^{K} s[m-k]x[k] = 0. \tag{2.17}$$

Further, it can be shown (Henderson, 1981) that the K complex roots of the polynomial

$$P(z) = \sum_{k=0}^{K} x[k]z^k \tag{2.18}$$

include $2L$ conjugate roots z_l that correspond to ω_l frequency and damping α parameters from the model 2.15. The general outline of algorithms based on Prony's method is the following:

1. select N and K and derive matrix A
2. compute $\ker(A)$ via SVD or via eigenvalues/eigenvectors of $A^H A$
3. from $x \in \ker(A)$ form complex polynomial $P(z)$
4. compute roots of $P(z)$

Many audio signals violate the assumptions of being a mixture of damped sinusoids at least to some extent. Frequency modulated sinusoid is a common example clearly breaking the assumption. In such cases it will be said that the signal contains certain number of *pseudo-sinusoids*. Various criteria commonly depending on the ratio of singular values (Kumaresan and Tufts, 1982) can be employed to determine the number of sinusoids of interest.
A later development spawned by the works of Pisarenko (Pisarenko, 1973) and Berni (Berni, 1975) successfully exploits underlying model by assuming a mixture of damped sinusoids and Gaussian noise. On the other hand, modern high-resolution methods build on the idea of *signal subspace* (van der Veen et al., 1993) and rotational invariance (Roy et al., 1986) were shown to result in even higher resolution (Kumaresan and Tufts, 1982), especially the generalised MUSIC algorithm (Schmidt, 1981, 1986), its variant root-MUSIC (Barabell, 1983) and the Toeplitz Approximation Method (TAM) (kun, 1983). It was shown in (Stoica and Nehorai, 1988; Kot et al., 1987)

that the error variance of the original methods by Prony and Pisarenko are not close to the CRB, thus further advancement was sought.
The latest development in the field exploit a specific property of the underlying Hankel and Pascal-Vandermonde matrices (Badeau et al., 2006) resulting in *Estimation of Signal Parameters via Rotational Invariance Technique* algorithm (ESPRIT) (Roy et al., 1986; Roy and Kailath, 1987) and its variation TLS-ESPRIT (Roy and Kailath, 1987). It was shown in Porat and Friedlander (1987); Stoica and Soderstrom (1991); Eriksson et al. (1993); Hua and Zhang (1991) that MUSIC,ESPRIT and Matrix Pencil (MP) have an asymptotic efficiency close to 1, with ESPRIT and MP slightly outperforming the MUSIC algorithm.
Generally, above mentioned methods only worked for a mixture of stationary, rather than cPACED sinusoids (sometime referred to as *quasi-polynomials* (V. Slivinskas, 1992; Badeau et al., 2008a)). In (Badeau et al., 2006, 2008b) the currently most sophisticated high-resolution method is defined, achieving a multi-component cPACED model by identifying multiple poles corresponding to a single non-stationary sinusoid (rather than multiple stationary sinusoids) and grouping them together to derive a single, non-stationary sinusoid.
A major advantage of the methods mentioned in this subsection is a very high frequency resolution when compared to FT based methods. It is important to reiterate that in its core, these methods were designed to analyse a number of stationary sinusoids. The extension to cPACED is not inherent to the method (a LS approach is used in (Roy et al., 1986)) and can lead to ambiguity when 2 partials are sufficiently close in frequency, as a tight group of poles must artificially be grouped into 1, 2 or more sinusoids. The symmetric structure of poles exhibited by a non-stationary sinusoid ease the decision on many occasions (Badeau et al., 2006), but does not remove inherent ambiguity. Some real-world examples include a critically damped harmonic oscillator and involves a double pole (Serway et al., 2010), Laguerre functions (the exponentials modulated by Laguerre polynomials) are often used in the estimations of time delay (Sabatini, 1997; Fischer and Medvedev, 1999) and biomedical applications for modelling the florescence decay (Olga et al., 2005). Multi-pole signals also appear as solutions to Schrödinger equation for hydrogen-like atoms (Hänsch et al., 1979), laser physics, as transverse laser modes (Milonni and Eberly, 1988), and in finance, for modeling the evolution of interest rates (Filipovic, 2000).
An important drawback is extremely high CPU complexity when compared to spectral energy re-allocation methods, especially when only a small number of sinusoids is of interest. Analysing a single sinusoid would result

in complexity about 200 times higher than a similar energy reallocation method with comparable accuracy (Muševič and Bonada, 2013).
As already mentioned, the absence of any kind of transform increases the accuracy, but operating completely in the time domain, the methods become very sensitive to the model mismatch. It is important to stress that failure to estimate the number of prominent sinusoids contained in the signal (Kumaresan and Verma, 1987) affects the accuracy gravely. Interestingly, this phenomena is a by-product of the ability to jointly estimate multiple components, which can now be viewed as a advantage and a drawback, depending on the specific application at hand.

2.4 Various

Percussion sounds can be successfully modelled using a sinusoid with Gammatone amplitude envelope Scholler and Purwins (2011) - a desirable model for such signals has to include an exponential damping parameter Christensen and van de Par (2006) to express the exponential energy loss of a vibrating object without continuous energy supply. Lastly, the main benefit of the complex polynomial amplitude compared to the real one is the ability to encode frequency modulations to some extent, enabling some desirable audio coding properties Bartkowiak (2007); Pantazis et al. (2011).

Chapter 3

Generalised sinusoid analysis using derivatives

One of the most widespread sinusoidal models is a so called *generalised sinusoid*. Assuming the polynomial log-AM/FM yields:

$$s(t) = e^{r(t)} = \exp\left(\sum_{k=0}^{K} r_k t^k\right) \tag{3.1}$$

$$= e^{p(t)+jq(t)} = \tag{3.2}$$

$$= \exp\left(\sum_{k=0}^{K} (p_k + jq_k) t^k\right), \tag{3.3}$$

where r_k are complex parameters composed of log-amplitude parameters p_k and phase parameters q_k and p_0, q_0 are the static log-amplitude and static phase respectively.

3.1 Reassignment and derivative method

Reassignment method was first proposed in (Kodera et al., 1978, 1976) and generalized for any time-frequency distribution (including the STFT) in (Auger and Flandrin, 1995b). Its main idea is to *reassign* spectrum energy from its current time/frequency location to a new one, located in the time/frequency gravity of the energy. This fairly simple concept has proven extremely powerful in removing spectral energy from the time frequency regions without any sinusoidal content. Despite numerous differences, a

general notion of *energy reallocation* can be attributed to all the methods described in this thesis. It could be viewed as a sidelobe supression method, especially when non-stationary frequency parameter estimates are used to *demodulate* the signal. Such FM removal further improves energy concetration and centres the spectral energy at the frequency of the underlying sinusoid. This phenomena was observed and mathematicall justified in Abe and Smith (2005) QIFFT algorithm.
The assumed model is linear log-AM/FM and historically, such model bears the following notation:

$$s(t) = \exp\left(a_0 + \mu_0 t + j(\phi_0 + \omega_0 t + \psi_0 t^2)\right), \tag{3.4}$$

where parameters $a_0, \mu_0, \phi_0, \omega_0, \psi_0$ correspond to log-amplitude, log-amplitude change rate, phase, frequency and frequency change rate, respectively. Note that the last term will later be changed to $\frac{t^2}{2}$ rather then t^2 to accommodate for the time derivation. By using the Gaussian window function:

$$w(t) = \frac{1}{\sqrt{2\pi}}\sigma e^{-\frac{1}{2\sigma}t^2} = \sqrt{\frac{p}{\pi}}e^{-pt^2}, \tag{3.5}$$

where σ is the standard deviation of the Gaussian and $p = 1/2\sigma^2$, an analytic expression for bounded FT can be obtained Abe and Smith (2005)

$$\langle s, we^{j\omega}\rangle = e^{u(\omega)+jv(\omega)} \tag{3.6}$$

where

$$u(\omega) = a_0 + \frac{\mu_0^2}{4p} - \frac{1}{4}\log\left(1 + \left(\frac{\psi_0}{p}\right)^2\right) - \frac{p}{4(p^2-\psi_0^2)}\left(\omega - \omega_0 - \frac{\mu_0\psi_0}{p}\right)^2 \tag{3.7}$$

$$v(\omega) = \psi_0 + \frac{\mu_0^2}{4\psi_0} + \frac{1}{2}\arctan\left(\frac{\psi_0}{p}\right) - \frac{\psi_0}{4(p^2-\psi_0^2)}\left(\omega - \omega_0 + \frac{p\mu_0}{\psi_0}\right)^2. \tag{3.8}$$

Above expression reaches a global maximum at:

$$argmax|e^{u(\omega)+jv(\omega)}| = \omega_0 + \frac{\mu_0\psi_0}{p}. \tag{3.9}$$

Clearly, a non-zero μ_0 and ψ_0 introduce a bias to the frequency estimate, if it were to be estimated by identifying a maximum in the magnitude spectrum. Interestingly, removing either AM or FM would remove such

bias. There are no analytical expressions for more general family of window functions, for example the popular *raised cosine* windows (Nuttall, 1981; Harris, 1978).Such non-linear behaviour drastically reduces the accuracy of this simple frequency estimator. Reassignment was designed to surpass such bias as long as the sinusoid under study matches the assumed model. The family of tractable models for reassignment has been extended in the last years. Most of additional flexibility stems from the simplification of the definition of the method itself Marchand (2008). The expressions however remain identical to classical reassignment as long as linear chirps are concerned (the original model on which reassignment is based).
Detailed and compact derivation of the estimator expressions can be found in Appendix A.1. The estimates (excluding amplitude and phase) are as follows:

$$\hat{\omega}(t,\omega) = \omega - \Im\left(\frac{S_{w'}}{S_w}\right) \tag{3.10}$$

$$\hat{\mu}(t,\omega) = -\Re\left(\frac{S_{w'}}{S_w}\right) \tag{3.11}$$

$$\hat{\psi}(t,\omega) = \frac{\Im\left(\frac{S_w S_{w''} - (S_{w'})^2}{(S_w)^2}\right)}{\Re\left(\frac{S_{tw'} S_w - S_{tw} S_w}{(S_w)^2}\right)}, \tag{3.12}$$

where $\hat{\omega}, \hat{\mu}, \hat{\psi}$ stand for the frequency, log-amplitude change rate and frequency change rate estimates. Generally, the *hat* symbol will designate an expression for estimate.
The derivative method was first designed as a high accuracy stationary estimator (Marchand, 1998; Desainte-Catherine and Marchand, 1998) and later generalized to a non-stationary one (Marchand and Depalle, 2008; Hamilton et al., 2009). Summary of the derivations of the estimators (again, excluding amplitude and phase) can be found in Appendix A.2:

$$\hat{\omega}(t,\omega) = \Im\left(\frac{S'_w}{S_w}\right) \tag{3.13}$$

$$\hat{\mu}(t,\omega) = \Re\left(\frac{S'_w}{S_w}\right) \tag{3.14}$$

$$\hat{\psi}(t,\omega) = \frac{\Im\left(\frac{S''_w}{S_w}\right) - 2\hat{\mu}(t,\omega)\hat{\omega}(t,\omega)}{1 + \Re\left(\frac{S'_{tw} S_w - S_{tw} S'_w}{(S_w)^2}\right)}, \tag{3.15}$$

which represent analogous expressions to 3.10-3.12. Importantly, the equivalence of the derivative and reassignment method has been shown in (Muševič

and Bonada, 2010b; Hamilton et al., 2009).
The complete derivations of both methods are included in Appendixes A.1 and A.2 in order to prepare necessary mathematical background for generalization to an arbitrary modulation degree and proof of equivalence of both algorithms in such general case.

3.2 Generalised reassignment

The reassignment method has been updated to work with an arbitrary order polynomial log-AM/FM (Wen and Sandler, 2009), that is an arbitrary K for the model in equation 3.1. The estimates can generally be defined as solutions to a specific linear system of equations, however for the sake of completeness, the solutions for the 2nd degree model from 3.1 which roughly corresponds to 3.4 follows:

$$r_2 = p_2 + jq_2 = \gamma_0 + j\psi_0 = \frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}} \Rightarrow \tag{3.16}$$

$$p_2 = \gamma_0 = \Re\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right) \tag{3.17}$$

$$q_2 = \psi_0 = \Im\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right), \tag{3.18}$$

where γ_0 is the second order log-AM parameter. The parameter does not exist in model 3.4, however if the signal does not exhibit such second order log-AM then the expression 3.17 should be nil anyway. Estimates for first order parameters follow:

$$q_1 = \omega_0 = \omega - \Im\left(\frac{S_{w'}}{S_w}\right) - \psi_0 \Re\left(\frac{S_{tw}}{S_w}\right) \tag{3.19}$$

$$p_1 = \mu_0 = -\Re\left(\frac{S_{w'}}{S_w}\right) + \psi_0 \Im\left(\frac{S_{tw}}{S_w}\right) \tag{3.20}$$

3.3 Distribution-derivative method

The distribution derivative method (Betser, 2009) belongs to the same family of methods as the reassignment and the derivative method as it reallocates spectrum energy to construct a more compact and readable representation. Its design allows departure from the Fourier transform, but before exploring various other alternative *kernels*, analogous expressions to the

others already presented in this chapter follow (full derivation in appendix A.4):

$$r_2 = p_2 + jq_2 = \gamma_0 + j\psi_0 = \frac{S_{w'}(\omega_2)S_w(\omega_1) - S_{w'}(\omega_1)S_w(\omega_2)}{S_{tw}(\omega_2)S_w(\omega_1) - S_{tw}(\omega_1)S_w(\omega_2)}, \tag{3.21}$$

where ω_1, ω_2 are two frequencies, in practice chosen to be as close to the actual frequency ω_0 as possible. The estimator for 1st order parameter:

$$r_1 = p_1 + jq_1 = \mu_0 + j\omega_0 = j\omega - \frac{S_{w'}}{S_w} - r_2\frac{S_{tw}}{S_w}. \tag{3.22}$$

Taking the real and imaginary part will give linear log-AM/frequency estimates identical to that of the reassignment and generalized reassignment 3.10,3.19. Note that the frequency variable is omitted as it can now be arbitrary chosen (eg: either ω_1 or ω_2 from 3.21 or any other appropriate value). A notable difference compared to the reassignment and derivative method is the necessity to use more than a single (in above case 2) frequency value. This may seem impractical, but as will be shown offers more flexibility and robustness.

The reassignment and derivative method can easily be redefined to avoid using the Fourier Transform if desired. There is however little need to do so, as far as modulations are below some reasonable bounds, keeping the correlation between the signal and the kernel high enough to avoid the numerical problems. A simple *polynomial phase kernel* could be used (Muševič and Bonada, 2011) for the analysis of highly modulated sinusoids. Wavelets and Gabor functions (Daubechies, 1990; Feichtinger and Strohmer, 1998) make a very attractive family of kernels since much more flexible time-frequency tiling is possible.

A very important property of the methods presented in this section is the absence of restrictions on the kernel functions used. In the latter case of the distribution derivative method with the Fourier kernel, the 2 frequencies can be arbitrarily close, either by calculating the Discrete Time Fourier Transform (DTFT) at arbitrary frequencies or taking bins of significantly zero-padded Fast Fourier Transform (FFT).

Interestingly, there is no assumption in the design of any of the methods that would require the kernels to be orthogonal or even belong to a tight frame (Christensen, 2003) as long as they are not too similar, in which case numerical problems would ensue.

3.4 Estimating static parameters: amplitude and phase

Estimation of the static amplitude and phase is not directly possible with the mentioned methods, as computing first or higher order time derivatives removes constants. A straight-forward, but CPU intensive way to estimate the static parameters is based on assumption that non-static parameters can be estimated with high enough accuracy.
The static parameter can be separated from the signal:

$$s(t) = e^{r_0}\gamma(t). \tag{3.23}$$

Estimates of the non-static parameters can be used to compute an estimate $\hat{\gamma}(t)$, which can in turn be used to estimate the static parameter:

$$\langle s, \hat{\gamma}\rangle = e^{r_0}\langle \gamma, \hat{\gamma}\rangle \Rightarrow \tag{3.24}$$

$$r_0 = \log(\langle s, \hat{\gamma}\rangle) - \log(\langle \gamma, \hat{\gamma}\rangle). \tag{3.25}$$

A great difficulty with the above approach is the CPU heavy calculation of $\hat{\gamma}$ from the parameter estimates, when it must be repeated for every frequency bin. It is possible to sacrifice some accuracy by simply using an already calculated inner product, for example a bin from FFT or an average of bins around the maxima - any kernel with large correlation with the signal will suffice Betser (2009).
The calculation of $\hat{\gamma}$ could be done via Sub-band Sinusoidal Synthesis method (Wen and Sandler, 2013) however the accuracy deteriorates in high frequency band.
Importantly, the phase and amplitude estimates are not required if only reallocation of the original representation is desired. However, a much sharper reallocated representation is obtained by swapping the original amplitude with the one estimated via 3.24.
In practice the inner product is replaced with a simple sum:

$$\langle x(t), y(t)\rangle = \int_{-\infty}^{\infty} x(t)\bar{y}(t)dt = \sum_{k=-\infty}^{\infty} x(kT)\bar{y}(kT), \tag{3.26}$$

as many times an analytical expression for above integral cannot be expressed with elementary arithmetic functions. Except for the infinite Gauss window 3.5 when the spectral amplitude and phase indeed have analytically

tractable form 3.7 for log-AM/FM, the more general family of raised cosine windows:

$$w(t) = \sum_{k=0}^{K} c_k \cos(2\pi k t) \tag{3.27}$$

and substituting $\omega_\Delta = \omega - \omega_0$ results in the following expression:

$$\begin{aligned}
&S_w(\omega) = \Gamma_w(\omega_\Delta, \mu_0, \psi_0) = \\
&\sum_{k=0}^{K} \frac{1}{\sqrt{\psi_0}} \left(\frac{1}{4} + \frac{j}{4}\right) c_k e^{\frac{i(2jk\pi + T(\mu_0 + j\omega_\Delta))^2}{2T^2\psi_0}} \sqrt{\pi} \\
&\left(\operatorname{Erf}\left[\frac{\left(\frac{1}{4} - \frac{j}{4}\right)(A)}{T\sqrt{\psi_0}}\right] - \operatorname{Erf}\left[-\frac{\left(\frac{1}{4} - \frac{j}{4}\right)(B)}{T\sqrt{\psi_0}}\right]\right. \\
&\left. + e^{\frac{4k\pi(\mu_0 + j\omega_\Delta)}{T\psi_0}} \left(\operatorname{Erf}\left[\frac{\left(\frac{1}{4} - \frac{j}{4}\right)(C)}{T\sqrt{\psi_0}}\right] - \operatorname{Erf}\left[-\frac{\left(\frac{1}{4} - \frac{j}{4}\right)(D)}{T\sqrt{\psi_0}}\right]\right)\right)
\end{aligned} \tag{3.28}$$

$$\begin{aligned}
A &= 4k\pi + T(-2j\mu_0 + T\psi_0 + 2\omega_\Delta) \\
B &= -4k\pi + T(2j\mu_0 + T\psi_0 - 2\omega_\Delta) \\
C &= 4k\pi + T(2j\mu_0 + T\psi_0 - 2\omega_\Delta) \\
D &= -4k\pi + T(-2j\mu_0 + T\psi_0 + 2\omega_\Delta)
\end{aligned}$$

Above expression has several problematic properties: with $\psi_0 = 0$ it results in multiplying 0 and ∞ therefore another, different expression should be derived for $\Gamma_w(\omega_\Delta, \mu_0, 0)$:

$$\Gamma_w(\omega_\Delta, \mu_0, 0) = \sum_{k=0}^{K} \frac{c_k e^{-\frac{1}{2}T(\mu_0 + j\omega_\Delta)} T^2 \left(e^{T(\mu_0 + j\omega_\Delta)} - 1\right)(\mu_0 + j\omega_\Delta)\cos(k\pi)}{4k^2\pi^2 + T^2(\mu_0 + j\omega_\Delta)^2} \tag{3.29}$$

Again, above term is not defined when $k = 0, \omega_\Delta = 0, \mu_0 = 0$ therefore we require another integral for this case:

$$\Gamma_w(0, 0, 0) = c_0 T \tag{3.30}$$

In practice, numerical errors appear when mentioned values are small enough, so we are forced to choose some threshold in each case. Even in such cases, certain parameter combinations result in multiplying a very large and a very

small number, leading to numerical errors.
Computing above expressions is clearly intractable and inefficient, but can lead to more accurate estimation compared to just using equation 3.26. The effects of noise will almost always overshadow the intricate details described in above derivations.

3.5 Unified reassignment, derivative and distribution derivative algorithm

In Hamilton and Depalle (2012b) it was shown that the reassignment, derivative and distribution method can be fused into a more general method, utilising either different kernel and signal or kernel derivatives. Interestingly, the last step of the method is still solution of a linear system of equations, however this equations now come from either reassignment, derivative or a combination.
The unified view on the three methods has greatly relaxed the restrictions on the contents of the linear system. Unfortunately, computing signal derivatives is a very erroneous procedure, as it will be shown later in this chapter. Further, the condition number of the matrix naturally plays a big role. The amplitude of a higher degree time derivative of the kernel function (required for generalised reassignment) is factor $\frac{fs}{N}$ bigger than the previous one, causing the inner products to raise exponentially - the condition number drops significantly. Further, kernel derivatives inherently include window function derivatives which must be zero at the start and the end of the time frame. This poses severe restrictions on window function, forcing a suboptimal choice in terms of TF properties. Distribution derivative method only requires first derivative of the kernel, allowing the use of almost arbitrary window function such as Hann window with very good TF properties. The requirement to use more than one frequency may seem restrictive, but the freedom to choose them very close together further extends flexibility to control TF properties of the method.
In this section a general algorithm based on the reassignment, derivative and distribution derivative will be outlined. The *hybrid* algorithm steps beyond the historical framework of non-stationary sinusoidal analysis, in the sense that it allows for any order of modulation with an arbitrary modulation function set and any reasonable kernel transform. In the case of polynomial modulation function it will by shown that high order models quickly become numerically unstable and limit the maximum modulation degree that can be used in practice.

The reallocation estimators use the time derivative of the signal to generate the systems of equations. An arbitrary generalised sinusoid model:

$$s(t) = e^{r(t)}, r(t) = \sum_{k=0}^{K} r_k m_k(t), \tag{3.31}$$

with $m_k(t)$ an arbitrary modulation function. To arrive to a linear system in respect to r_k a time derivative must be taken:

$$s'(t) = r'(t)s(t) = \sum_{k=0}^{K} r_k m'_k(t)s(t). \tag{3.32}$$

Above equality represents the core of all reallocation algorithms, as the signal derivative is expressed by the linear combination of terms involving known model functions and the original signal - the right hand side of the equation consists of known quantities and can in practice easily be computed. The left hand side requires more attention, as it is generally not available in practice and must thus be computed. In (Desainte-Catherine and Marchand, 2000) a simple approach by applying the general definition of a function derivative:

$$s'(t) = \lim_{\epsilon \to 0} \frac{s(t+\epsilon) - s(t)}{\epsilon}, \tag{3.33}$$

directly to the discrete time sampled signal:

$$s[n] = s[n+1] - s[n], \tag{3.34}$$

yields limited quality results. The error can be minimised by up-sampling (Raspaud and Marchand, 2007) effectively sending ϵ close to zero. A simple differentiator filter can be derived from sampling theorem definition (Marchand and Depalle, 2008):

$$s(t) = \sum_{n=-\infty}^{+\infty} s[n] sinc(\overbrace{f_s - m}^{u(t)}) \Rightarrow \tag{3.35}$$

$$s'(t) = \sum_{n=-\infty}^{+\infty} s[m] \left(\frac{\cos(\pi u(t))}{u(t)} - \frac{\sin(\pi u(t))}{\pi u(t)^2} \right) \Rightarrow \tag{3.36}$$

$$s'[n] = s'(n/f_s) = f_s \sum_{m \neq n} s[m] \left(\frac{(-1)^{n-m}}{(n-m)} \right). \tag{3.37}$$

The discrete derivative s' can thus be obtained convolving the discrete signal s by the following filter:

$$h[n] = f_s \frac{(-1)^n}{n} \text{ for } n \neq 0, \text{ and } h(0) = 0. \tag{3.38}$$

Evidently, the above filter has infinite time support. In practice, an additional finite time support window function is used. Figure 3.1 shows the accuracy of the mentioned filter at various frequencies.

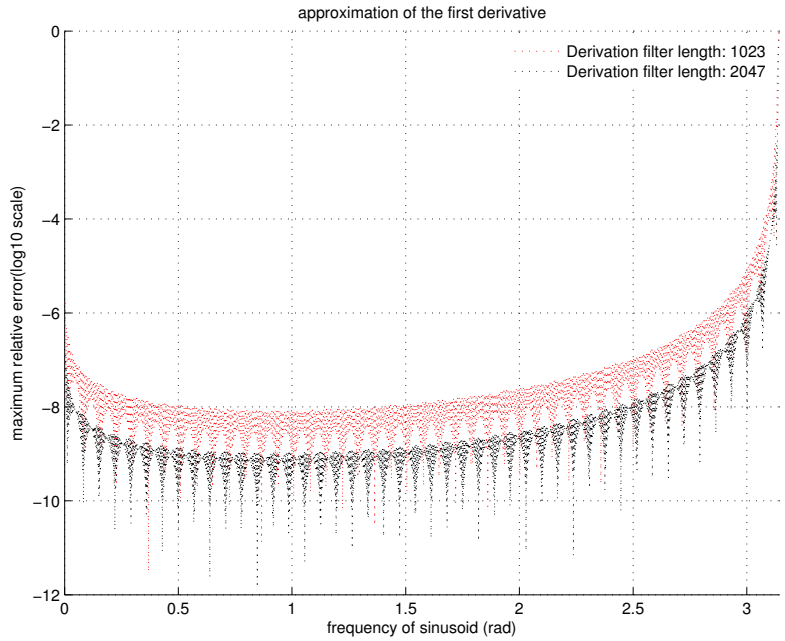


Figure 3.1: Maximal estimation error for all phases, when approximating the first derivative. Significant approximation accuracy is gained by doubling the derivation filter length

Computing signal derivatives is a CPU intensive and error prone - it should be avoided if possible. Luckily, using *integration per-partes* it is easy to derive the so called *distribution derivative* rule:

$$\langle x', y \rangle = x\bar{y}\big|_{-\infty}^{\infty} - \langle x, y' \rangle, \tag{3.39}$$

and assuming y approaches 0 when $t \to \pm\infty$, the distribution derivative rule is obtained:

$$\langle x', y \rangle = -\langle x, y' \rangle. \tag{3.40}$$

The assumption $\lim_{t\to\pm\infty} y^{(n)} = 0, n = 0 \cdots k-l$ provides a simple *chain* distribution derivative rule:

$$\langle x^{(k)}, y\rangle = (-1)^{k-l}\langle x^{(l)}, y^{(k-l))}\rangle. \tag{3.41}$$

An arbitrary choice of $k < l$ gives birth to as many equations as required. It is now easy to imagine x being the signal under study and y an arbitrary kernel including the window function. It is now assumed that any degree time derivative of the signal can be computed with a reasonable accuracy. The kernel function can be well-defined so any order derivative exists - the right hand side of the equation 3.41 thus can be computed in practice.
The derivatives of the generalised sinusoid are particularly easy to express recursively:

$$s^{(d)} = \sum_{k=0}^{K} r_k (m_k' s)^{(d-1)}. \tag{3.42}$$

Using an arbitrary kernel Ψ_ω (centred around frequency ω) and a limited time support window function w yields the final equality:

$$\langle s^{(m-b)}, (w\Psi_\omega)^{(b)}\rangle = (-1)^{d+b}\sum_{k=0}^{K} r_k \langle (m_k' s)^{(m-d-1)}, (w\Psi_\omega)^{(d)}\rangle. \tag{3.43}$$

Since model functions m_k are also known a priori the inner products can easily be computed, yielding a linear equation with respect to r_k. Naturally, at least K equations are required to estimate all r_k. However, it is common to set $m_0 = 1$ for the model to include static phase and amplitude. In such case $m_0' = 0$ and thus even the first derivative does not include r_0 any more:

$$s' = r's = \sum_{k=1}^{K} r_k m_k' s. \tag{3.44}$$

The expression 3.43 can be used to derive various linear systems of equations for the estimation of the parameters of interest. A rigorous and detailed tests would be timely, as there are numerous ways to construct a linear system. Naturally, one would strive to use a certain construction that would perform best in all cases, rather than adapt the process to the signal understudy. It will be shown in chapter 4 that indeed the DDM performs equally or superior to GRM in all cases, with superior numerical stability and an option to arbitrarily over determine the linear system. The latter will greatly improve the analysis of sinusoids exhibiting extreme frequency

modulation (see chapter 10). This desired property however, diminishes the practical relevance of the unified approach described here. However a great contribution of alleviating the assumptions that DDM, GRM and DAM are separate, independent methods is tremendous. Algorithms outlined in chapters 5, 6, 7, 8 and 9 were all greatly inspired by the ideas outlined in this chapter.

Part III

Contribution

Chapter 4

Practical and Theoretical Comparison of Generalised Reassignment, Derivative and Distribution Derivative methods

4.1 Theoretical equivalence of generalised reassignment and generalised derivative method

In this section it will be show that the reassignment and derivative method are theoretically identical for any modulation degree (Muševič and Bonada, 2010b) as opposed to *frequency slope* (2nd degree FM) in (Hamilton et al., 2009). The estimators of both methods are in fact a two special cases of the same mathematical derivation. It will be immediately clear that computing the signal derivative should be avoided, thus the reassignment method is preferred, which is confirmed by the tests. The difference is can only be detected in practice, since the mathematical derivations assume the derivative can be computed with an arbitrary accuracy.
Despite it is straightforward to see that derivative method cannot achieve the accuracy of reassignment in practice, the mathematical derivations will consolidate and unify the theory behind the methods, revealing details po-

tentially important for the design of more advanced methods.
The definitions of spectral amplitude and phase from section 1.1 equation 1.4:

$$S_w(t,\omega) = \exp\big(\lambda(t,\omega) + j\phi(t,\omega)\big). \tag{4.1}$$

Reassigned frequency and time can be expressed (for details see appendix A.3) for a general case as:

$$\hat{\omega}(t,\omega) = \frac{\partial}{\partial t}\phi(t,\omega) = \Im\left(\frac{\frac{\partial S_w}{\partial t}}{S_w}\right) \tag{4.2}$$

$$\hat{t}(t,\omega) = t - \frac{\partial}{\partial \omega}\phi(t,\omega) = t - \Im\left(\frac{\frac{\partial S_w}{\partial \omega}}{S_w}\right). \tag{4.3}$$

And common notation for the linear-AM/FM generalised sinusoid:

$$s(t) = \exp\big(\lambda_0 + \mu_0 t + j(\phi_0 + \omega_0 t + \frac{\psi_0}{2}t^2)\big). \tag{4.4}$$

As pointed out in Marchand and Depalle (2008) and Röbel (2002), general log-AM and FM expressions can be written as:

$$\hat{\mu}(t,\omega) = \frac{\partial}{\partial t}a(t,\omega) = \Re\left(\frac{\frac{\partial S_w}{\partial t}}{S_w}\right) \tag{4.5}$$

$$\hat{\psi}(t,\omega) = \frac{\partial \hat{\omega}}{\partial \hat{t}} = \frac{\partial \hat{\omega}}{\partial t} \Big/ \frac{\partial \hat{t}}{\partial t} \tag{4.6}$$

$$\frac{\partial \hat{\omega}}{\partial t} = \Im\left(\frac{\frac{\partial^2 S_w}{\partial t^2} S_w - \left(\frac{\partial S_w}{\partial t}\right)^2}{\left(S_w\right)^2}\right) \tag{4.7}$$

$$\frac{\partial \hat{t}}{\partial t} = 1 - \Im\left(\frac{\frac{\partial^2 S_w}{\partial \omega \partial t} S_w - \frac{\partial S_w}{\partial \omega}\frac{\partial S_w}{\partial t}}{\left(S_w\right)^2}\right). \tag{4.8}$$

The above equations provide estimate expressions independent of the method used and thus hold for both reassignment and the derivative method. For reassignment, the following expressions with some restrictions apply:

$$\frac{\partial}{\partial t}S_w = -S_{w'} + j\omega S_w \tag{4.9}$$

$$\frac{\partial}{\partial \omega}S_w = -jS_{tw} \tag{4.10}$$

$$\frac{\partial^2}{\partial \omega \partial t}S_w = jS_{tw'} + jS_w + \omega S_{tw} \tag{4.11}$$

$$\frac{\partial^2}{\partial t^2}S_w = S_{w''} - 2j\omega S_{w'} - \omega^2 S_w. \tag{4.12}$$

For the derivative method, slightly simpler expressions hold:

$$\frac{\partial}{\partial t} S_w = S'_w \tag{4.13}$$

$$\frac{\partial}{\partial \omega} S_w = -j S_{tw} \tag{4.14}$$

$$\frac{\partial^2}{\partial \omega \partial t} S_w = -j S'_{tw} \tag{4.15}$$

$$\frac{\partial^2}{\partial t^2} S_w = S''_w. \tag{4.16}$$

Substituting reassignment STFT expressions 4.9-4.12 into general equations for parameter estimations 4.2-4.8 yields:

$${}^{\mathrm{R}}\hat{\omega}(t,\omega) = \omega - \Im\left(\frac{S_{w'}}{S_w}\right) \tag{4.17}$$

$${}^{\mathrm{R}}\hat{\mu}(t,\omega) = -\Re\left(\frac{S_{w'}}{S_w}\right) \tag{4.18}$$

$${}^{\mathrm{R}}\hat{\psi}(t,\omega) = \frac{\Im\left(\frac{S_w S_{w''} - (S_{w'})^2}{(S_w)^2}\right)}{\Re\left(\frac{S_{tw'} S_w - S_{tw} S_w}{(S_w)^2}\right)}, \tag{4.19}$$

which are well known reassignment expressions for estimating parameters of log-AM/FM sinusoids. Analogously, substituting derivative method STFT expressions 4.13-4.16 into same equations results in:

$${}^{\mathrm{D}}\hat{\omega}(t,\omega) = \Im\left(\frac{S'_w}{S_w}\right) \tag{4.20}$$

$${}^{\mathrm{D}}\hat{\mu}(t,\omega) = \Re\left(\frac{S'_w}{S_w}\right) \tag{4.21}$$

$${}^{\mathrm{D}}\hat{\psi}(t,\omega) = \frac{\Im\left(\frac{S''_w}{S_w}\right) - 2\,{}^{\mathrm{D}}\hat{\mu}(t,\omega)\,{}^{\mathrm{D}}\hat{\omega}(t,\omega)}{1 + \Re\left(\frac{S'_{tw} S_w - S_{tw} S'_w}{(S_w)^2}\right)}, \tag{4.22}$$

which are the derivative method expressions as given in Marchand and Depalle (2008) and Hamilton et al. (2009).

This section has clearly demonstrated that reassignment and the derivative method are in fact analogous methods, derived from the same general linear log-AM/linear FM equations. The only difference is the definition of STFT, which results in quite different expressions for parameter estimates. Mathematically identical proof was already given in (Marchand and Depalle,

2008) and (Hamilton et al., 2009), however it was given for each parameter of linear log-AM/linear FM sinusoids separately and thus did not prove the equivalence of the two methods for arbitrarily modulated sinusoids. Both methods should theoretically be generalised to an arbitrary modulation degree and their equivalence in such a generalised case should been shown. In order to show that, an arbitrary order time derivatives of general linear FM parameter expressions (equation 4.6) should be considered: $\frac{\partial^n \hat{\omega}}{\partial \hat{t}^n} = \frac{\partial^n \hat{\omega}}{\partial t^n} / \frac{\partial^n \hat{t}}{\partial t^n}$. Such expressions would contain STFTs of the form $\frac{\partial^{k+l} S_w}{\partial t^k \partial \omega^l}$. It is possible to transform the general expressions into reassignment ones, containing STFTs of the form $S_{w^{(k)}t^l}$ and analogously into the derivative method ones, containing STFTs of the form $S_{t^l}^{(k)}$. It is straightforward that reassignment and corresponding derivative method expressions are identical for all modulation degrees. The same procedure can be performed for log-AM, concluding the proof of equality of the two methods for an arbitrary modulated sinusoid.

The derivative method requires computation of signal time-derivatives, as opposed to reassignment, which requires computation of the window time-derivatives. In practice, it is impossible to avoid errors computing time derivative of the signal in time domain. For that purpose, a derivation filter is used, however unacceptable errors occur at high frequencies (Marchand and Depalle, 2008). Further, using such filter increases the frame length requirements of the STFT and raises computational complexity. When performing the STFT analytical expression for window function is known in most cases, therefore exact analytical expression for its time derivatives can generally be computed before performing the STFT, which does not add any computational complexity. It can be concluded that lower computational complexity and higher accuracy is expected from the reassignment estimates compared to the derivative method ones. However, tests have shown that in the reduced frequency range (up to 3/4 Nyquist) methods perform comparably (Marchand and Depalle, 2008).

A recent development has provided the unification of the derivative, generalised reassignment and the distribution derivative method into one single hybrid method (Hamilton and Depalle, 2012b). Despite its high flexibility the hybrid method suffers from the same problems as each method separately - the unification does not provide any added value over the original methods.

4.2 Per-Parameter Comparison

In this section a classic comparison tests for non-stationary methods is presented. Thousands of individual tests with various test signal parameters and white noise levels are conducted. Such tests provide a rigorous assessment of accuracy, by comparing estimations to the CRBs to have a complete picture of how close to perfection the algorithms are. It will be shown that some of the presented algorithms indeed come very close to optimal for most of the mid and high noise range. Since very low noise cases (high SNR) are not very common in practice, the presented algorithms will be deemed optimal for all practical purposes.

Generalised reassignment and distribution derivative methods will be shown to approach the CRB quality even for very high SNR. An interesting phenomena exhibited by all the algorithms is reaching a certain *plateau* of quality, when higher SNR does not improve the quality any more. It is unreasonable to expect algorithms would be *infinitely* accurate, simply due to finite representation of the numbers in computers.

Possibly more interesting is the dependence of accuracy with respect to the actual signal parameters. The derivative method, for instance, is expected to deteriorate at higher frequencies due to the inaccurate derivative computation, which is even more evident with the second derivative. This will limit the derivative method to about 3/4 of the Nyquist frequency range and linear-AM/FM.

Further, the choice of window function will become increasingly important in case of the generalised reassignment. The restriction $w^{(k)}(-\frac{T}{2}) = w^{(k)}(\frac{T}{2}) = 0, k = 0...$ pose the main obstacle for very high accuracy 3rd degree log-AM/FM model due to suboptimal TF properties of such window functions.

The distribution derivative does not suffer from any of the aforementioned shortcomings. Only the first window derivative w' is required even for 3rd degree modulation, allowing for use of the Hann window, drastically reducing the effective bandwidth of the method.

All tests are conducted in the presence of log-AM and FM. A single complex sinusoid is analyzed, $fs = 44100Hz, H = 511$, Hanning window is used and signal to noise ratio (SNR) expressed in $10\log_{10}(\sigma^2/a^2)$ ranges from -20dB to + 120dB. Frequency ranges linearly from 0 to 3/8Fs (100 different values), phase linearly from -0.8π to 0.8π (9 different values), linear log-AM linearly from -100 to 100 (5 different values, including 0), linear FM linearly from -10000 to 10000 (5 different values, including 0). In (Marchand and Depalle, 2008) the error variance is considered in its strict definition,

that is: $var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$. The values of X are normalized, i.e.: the square of the mean is subtracted. Variance of a consistently biased estimator could be very low, as the square of the mean value of the error (the *bias*) is subtracted. This makes a biased estimator appear *better*, than it is. For this reason, two different measures will be computed: $E[X^2] - E[X]^2$ denoted as *normalized* and $E[X^2]$ as *non-normalized* variance. Any difference in these two measure signals a biased estimator.
In practice the magnitude spectrum computed from the FFT is used to find a peak. That is, a frequency FFT bin at which magnitude spectrum reaches a maximum is looked for. Such procedure results in a significant error in frequency estimate, even with zero-padding. Since all methods need an initial estimate of frequency and magnitude spectrum peak is the only available estimate, robustness to such error is very important. Denoting $\tilde{\omega}$ as the initial frequency estimate made from the FFT magnitude spectrum, the frequency and the linear log-AM/FM estimates can be written as follows:

$$\tilde{\omega}_0 = \hat{\omega}(t, \tilde{\omega}) \tag{4.23}$$
$$\tilde{\mu}_0 = \hat{\mu}(t, \tilde{\omega}) \tag{4.24}$$
$$\tilde{\psi}_0 = \hat{\psi}(t, \tilde{\omega}) \tag{4.25}$$

Since the values of spectrum at $\tilde{\omega}$ were already computed, above estimates take almost no computational overhead and provide a fast way to get relatively good estimates of parameters. However, new frequency estimate $\tilde{\omega}_0$ provides a better estimate than $\tilde{\omega}$, so it can used to get new, better set of estimates:

$$\tilde{\omega}_1 = \hat{\omega}(t, \tilde{\omega_0}) \tag{4.26}$$
$$\tilde{\mu}_1 = \hat{\mu}(t, \tilde{\omega_0}) \tag{4.27}$$
$$\tilde{\psi}_1 = \hat{\psi}(t, \tilde{\omega_0}) \tag{4.28}$$

Note that ω_0 is generally not equal to any FFT bin frequency, therefore above procedure requires additional computations of DTFT at ω_0. It is natural to generalise this recursive run of the algorithm to any degree:

$$\tilde{\omega}_k = \hat{\omega}(t, \tilde{\omega}_{k-1}) \tag{4.29}$$
$$\tilde{\mu}_k = \hat{\mu}(t, \tilde{\omega}_{k-1}) \tag{4.30}$$
$$\tilde{\psi}_k = \hat{\psi}(t, \tilde{\omega}_{k-1}) \tag{4.31}$$

It's important to note, that to achieve above estimates the most efficient way would be to first compute ω_{k-1} and only then $\tilde{\omega}_k, \tilde{\mu}_k, \tilde{\psi}_k$. Finally, phase and amplitude estimates can be computed as proposed in section 3.4:

$$\tilde{a}_k = \left| \frac{S_w(\tilde{\omega}_{k-1})}{\langle \exp(j(\tilde{\omega}_k t + \tilde{\psi}_k t^2)), e^{j\tilde{\omega}_{k-1}t} \rangle} \right| \tag{4.32}$$

$$\tilde{\phi}_k = \angle \left(\frac{S_w(\tilde{\omega}_{k-1})}{\langle \exp(j(\tilde{\omega}_k t + \tilde{\psi}_k t^2)), e^{j\tilde{\omega}_{k-1}t} \rangle} \right). \tag{4.33}$$

The continuous FT of a time-limited signal:

$$X(\omega) = \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j\omega t} dt, \tag{4.34}$$

must in practice be approximated by a DTFT:

$$X[\omega_l] = \sum_{k=-\frac{T}{2}}^{\frac{T}{2}} x[k] e^{-j\omega_l k}. \tag{4.35}$$

The DTFT spectrum is evaluated only on selected *bin* frequencies ω_l. The error of such approximation

$$X(\omega_l) - X[\omega_l] = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} X(\omega_k + 2\pi k). \tag{4.36}$$

Clearly the $X(\omega_k \pm 2\pi k)$ must decay quickly with k to minimise this error. For slowly varying sinusoids and a reasonable window function this automatically holds. The signal x can however represent a derivative of a non-stationary sinusoids which can depart from the slowly varying modulation assumption and thus cause performance degradation.

In following plots, the errors are shown for 3 different runs of each algorithm: 'no DTFT' corresponds to parameters with index 0 ($\tilde{a}_0, \tilde{\mu}_0, \tilde{\phi}_0, \tilde{\omega}_0, \tilde{\psi}_0$), 'DTFT 1' corresponds to parameters with index 1 (i.e.: DTFT is used once) and 'DTFT 2' corresponds to parameters with index 2 (i.e.: DTFT is used twice recursively). In (Marchand and Depalle, 2008), the 'no DTFT' variant of reassignment and 'DTFT 1' variant of derivative was compared.

In the case of the derivative method, a derivation filter of length 1023 was used. Further, derivatives of the signal are estimated using a larger time frame to avoid derivation filter edge effect. Time frame was enhanced by two

lengths of derivation filter, but only the accurate middle part was kept after convolution with the filter. If a higher order derivative would be required an even further frame extension would be required. Therefore derivative method has some additional requirements, which could be used by other two compared algorithms in some way to improve its estimates. Therefore, it can be argued, that the comparison is not entirely fair, as reassignment and generalized method were given a smaller time frame.

The generalized algorithm was tested using a polynomial model. To simplify the implementation of the algorithm, the degree of amplitude and phase polynomials are required to be set to the same value. In the tests, the 2^{nd} degree model was used. This means, that log-amplitude, linear log-AM and second order log-AM were calculated (although the latest was not used in comparison). As described in the appendix A.3, the corrections arising from the time reassignment will be accounted for. Since second order log-AM is null for all analysed signals, our log-amplitude model is of 1 degree higher order then necessary. In (Wen and Sandler, 2009) it was shown that setting the polynomial order too high introduces some additional error due to over-fitting. Therefore, we can expect higher error in log-amplitude and log-AM estimates. It's interesting to speculate, that the error is caused by afore mentioned *corrections*. The error in the second order log-AM estimate, which should be null, is propagated to lower level estimates as a *correction* term.

In the tests conducted, Hanning window function was used, although its second derivative doesn't satisfy the assumption made in A.57. This could cause an error in some estimates of the generalized method, however its results improve in most cases.

In the first series of tests a *reduced frequency range* of up to 16kHz is used (about 3/4 of the Nyquist frequency), in order to show comparable performance of the derivative method. The results of these extensive tests outline several interesting properties. A Firstly, the only estimates exhibiting a bias (i.e., normalised and non-normalised variance differ significantly) are the amplitude estimates for RM and DAM. Since GRM does not exhibit such bias the reason should be sought in the method differences. The reassigned-time corrections (for details see A.1,A.2 and A.3) might explain such phenomena. The bias is significant however not particularly large and since GRM does not suffer from it, no further attention will be given to it. Importantly, the accuracy for all the estimators reaches a *plateau* at some high SNR. That is: the accuracy does not continue to improve after certain very low-noise level is reached. Clearly, one of the reasons is the limitation outlined in equation 4.36 and another the already mentioned time-

corrections. Importantly, no other approximations were made during the derivation of the algorithms, thus the reasons for any potential inaccuracies must arrise from practical and implementation issues, which are commonly harder to detect.
The DAM requires an accurate computation of the signal derivative, which is either time consuming or erroneous (see figure 3.1, section 3.5 and (Marchand and Depalle, 2008)). However, it was shown in (Marchand and Depalle, 2008) that in the reduced frequency range, substituting the computed derivative with the precomputed analytical derivative of the test signal did not affect the accuracy significantly. The tests conducted correspond to the AM/FM case in (Marchand and Depalle, 2008). Amplitude estimation of both DAM and RM profit greatly from re-estimation in mid-SNR region. Clearly, the GRM is an optimal algorithm for all practical purposes. It reaches the plateau at about 90dB SNR - well below noise levels encountered in real-world recordings. Further, re-estimation has no observable effect on accuracy, a very desirable and expected property as the accuracy is very close to CRB anyway.

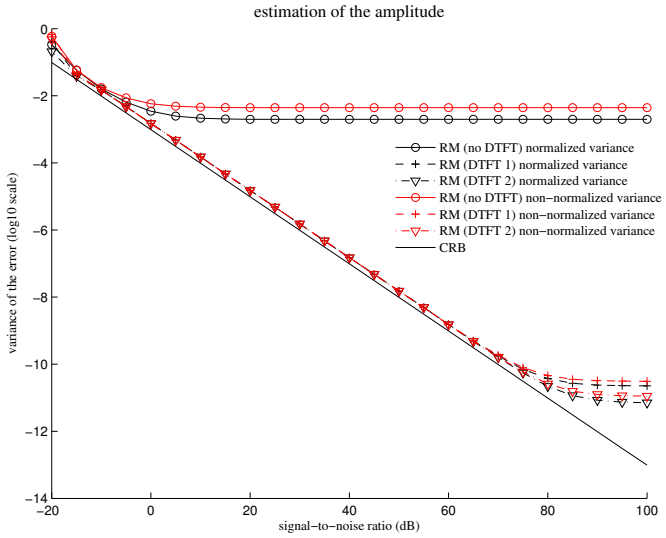


Figure 4.1: Reassignment method (reduced frequency range): log-amplitude estimation error

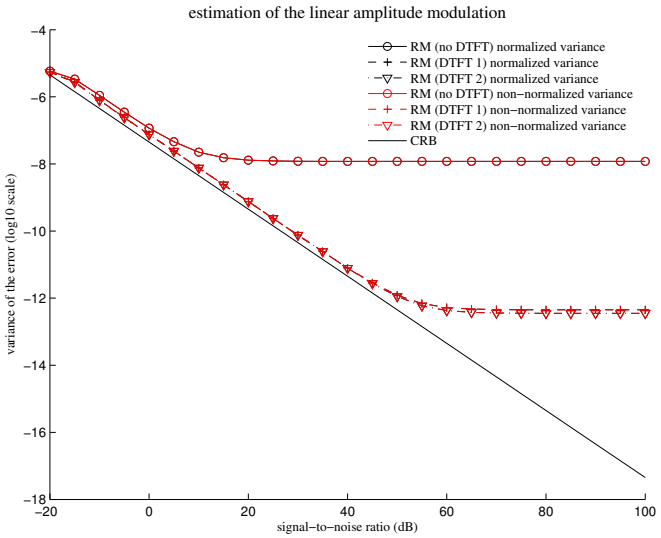


Figure 4.2: Reassignment method (reduced frequency range): linear log-AM estimation error

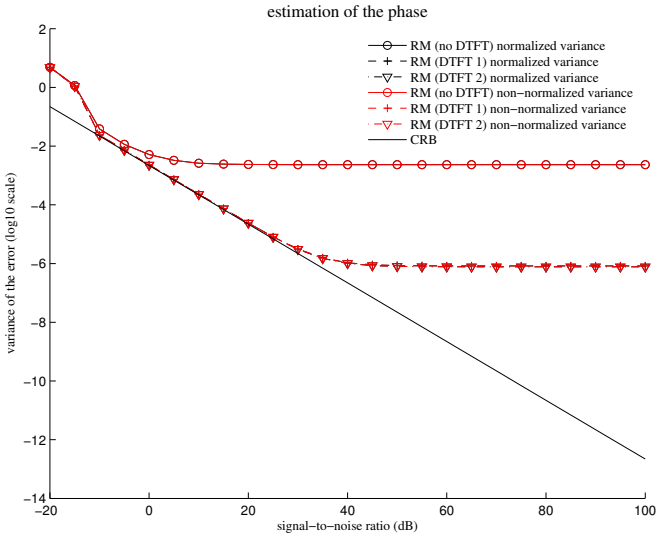


Figure 4.3: Reassignment method (reduced frequency range): phase estimation error

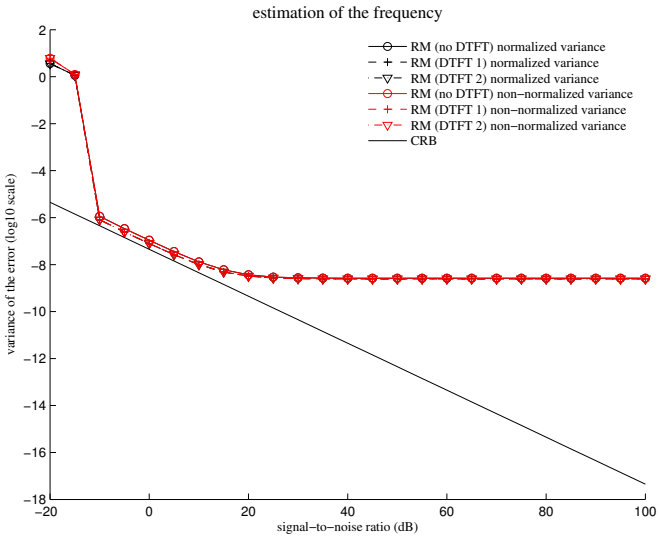


Figure 4.4: Reassignment method (reduced frequency range): frequency estimation error

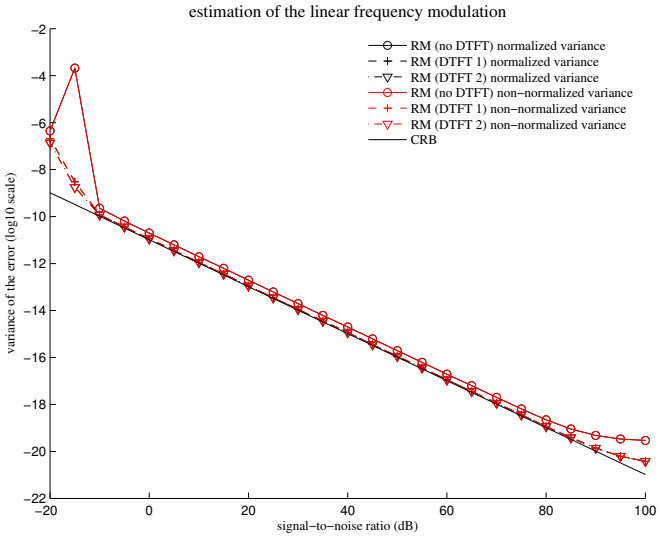


Figure 4.5: Reassignment method (reduced frequency range): linear FM estimation error

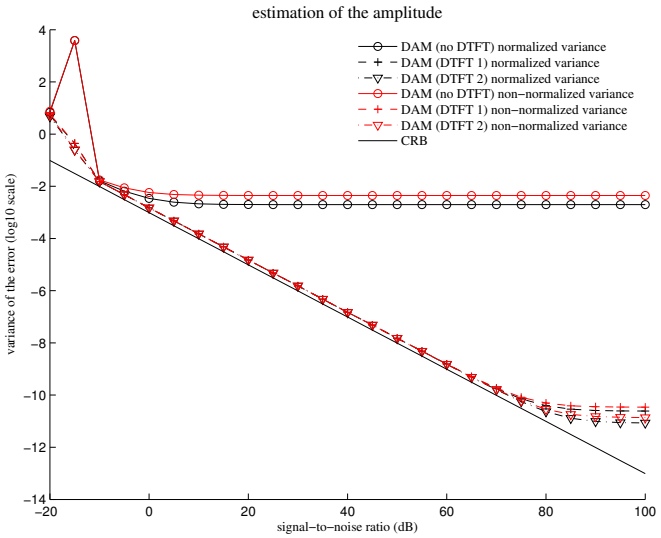


Figure 4.6: Derivative method (reduced frequency range): log-amplitude estimation error

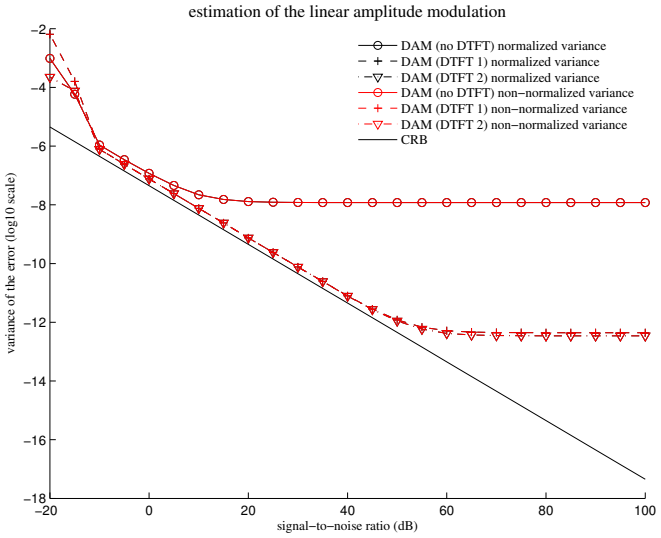


Figure 4.7: Derivative method (reduced frequency range): linear log-AM estimation error

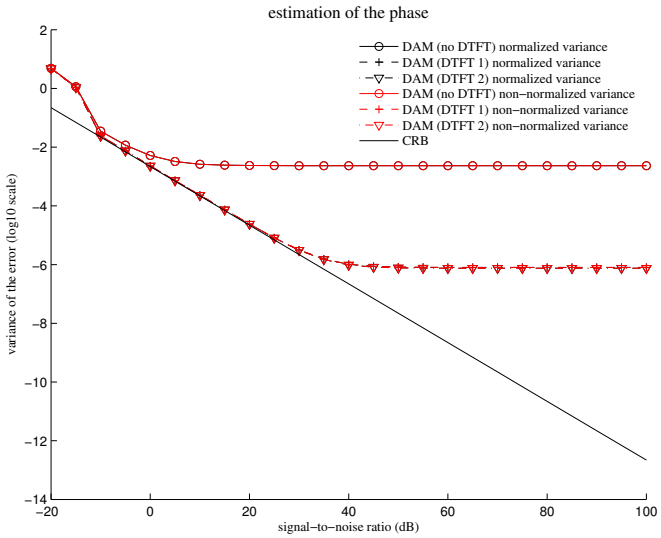


Figure 4.8: Derivative method (reduced frequency range): phase estimation error

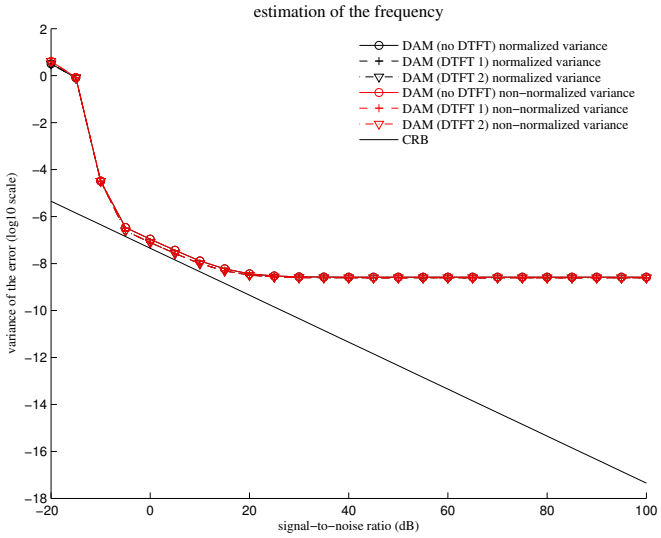


Figure 4.9: Derivative method (reduced frequency range): frequency estimation error

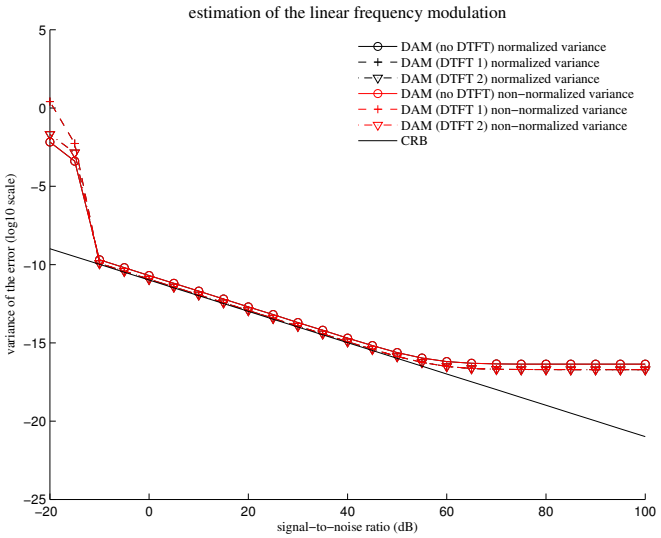


Figure 4.10: Derivative method: (reduced frequency range) linear FM estimation error

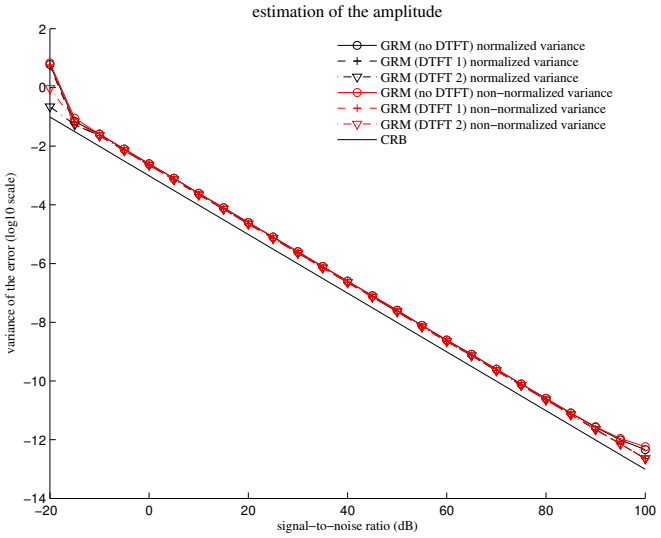


Figure 4.11: Generalized reassignment (reduced frequency range): log-amplitude estimation error

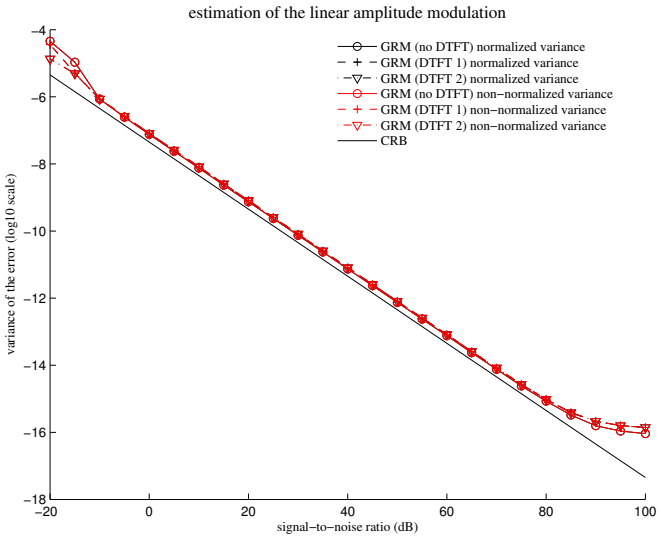


Figure 4.12: Generalized reassignment (reduced frequency range): linear log-AM estimation error

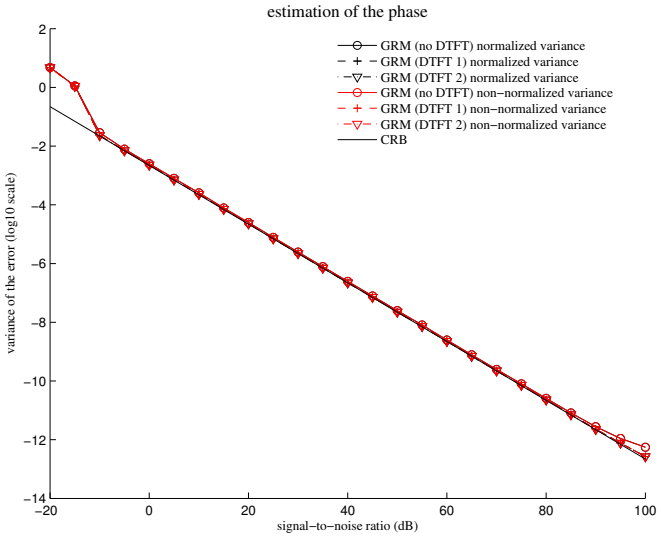


Figure 4.13: Generalized reassignment (reduced frequency range): phase estimation error

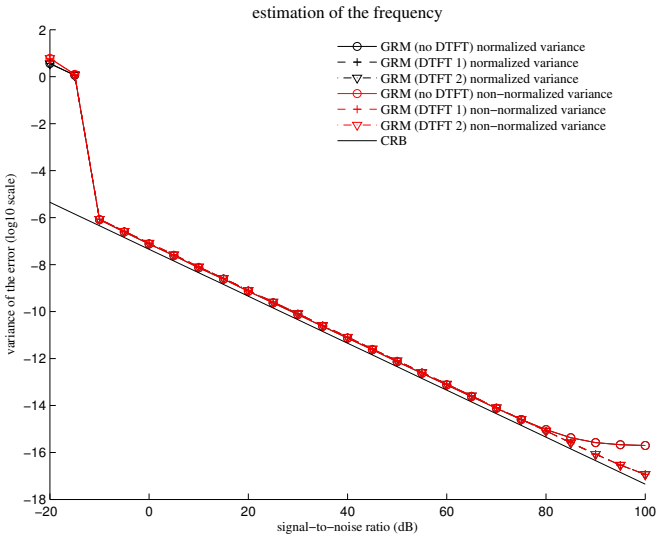


Figure 4.14: Generalized reassignment (reduced frequency range): frequency estimation error

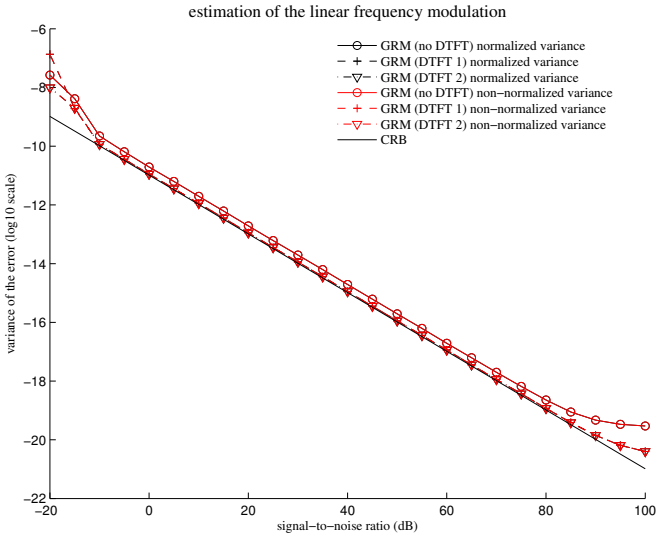


Figure 4.15: Generalized reassignment (reduced frequency range): linear FM estimation error

In the second suite of tests only the non-normalised variance was calculated

(so any existing estimation bias is not removed), for DAM 2 different differentiation filter sizes were used (the regular 1023 and 2047 samples), single re-estimation was performed to get best possible accuracy for all methods and the full frequency range up to the Nyquist frequency was used.
The GRM and RM behave similarly as in the previous tests (i.e.: the frequency range does not affect the accuracy), DAM on the other hand suffers from significant derivative filter accuracy degradation in high frequency range, even longer derivative filter does not alleviate the problem significantly. Although larger filters might affect the accuracy beneficially, however the computational cost exceeds that of RM and GRM with the original 1023 sample length filter already and any effort to improve the performance of DAM may rightfully be deemed in vain. The GRM clearly outperforms RM as it correctly applies the time-corrections arising from higher order modulations to lower order parameters. Most importantly, the model assumed matches the analysed signal exactly and thus GRM represents a perfect single sinusoid estimator

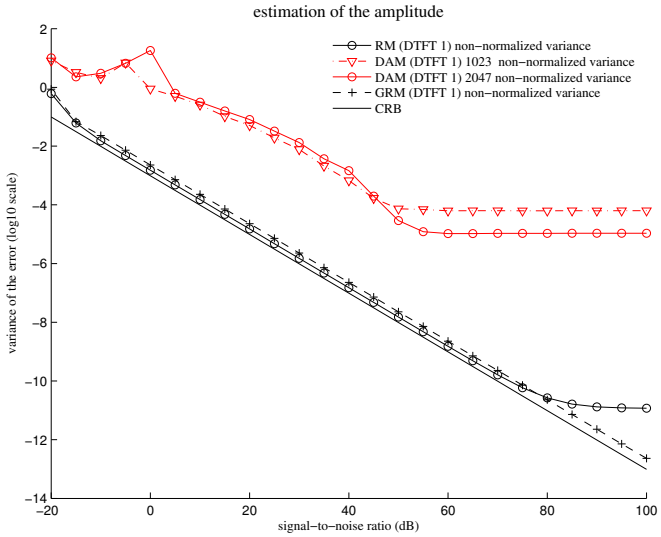


Figure 4.16: All methods (full frequency range): log-amplitude estimation error

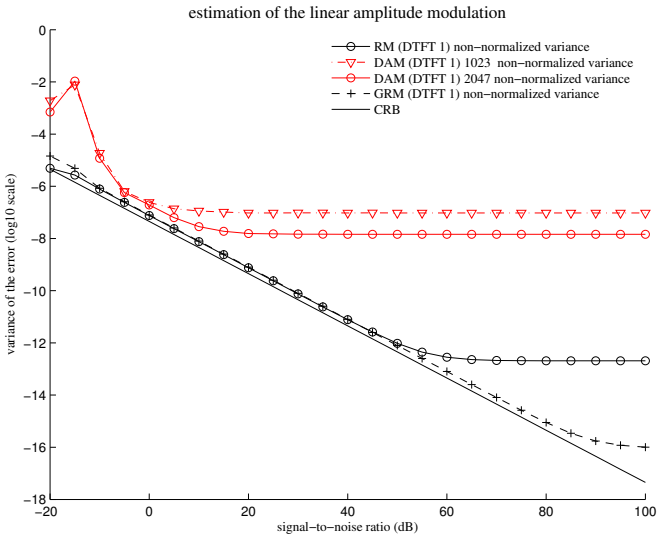


Figure 4.17: All methods (full frequency range): linear log-AM estimation error

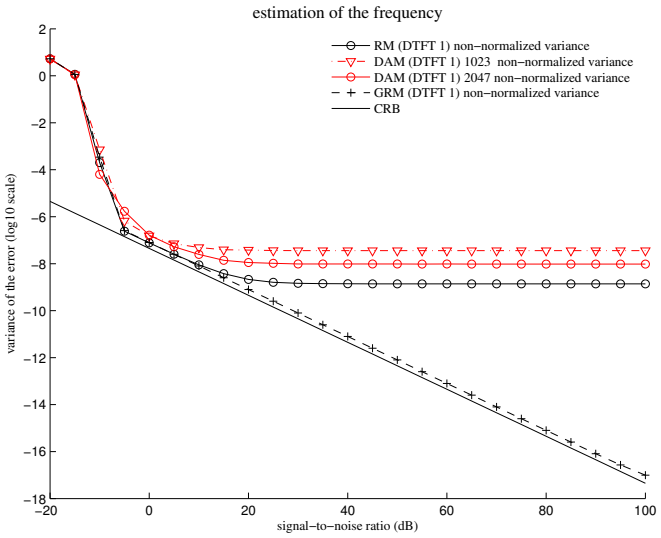


Figure 4.18: All methods (full frequency range): phase estimation error

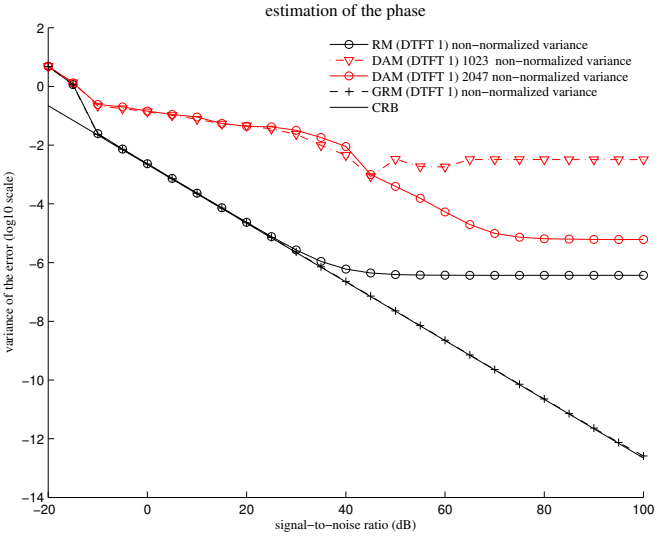


Figure 4.19: All methods (full frequency range): frequency estimation error

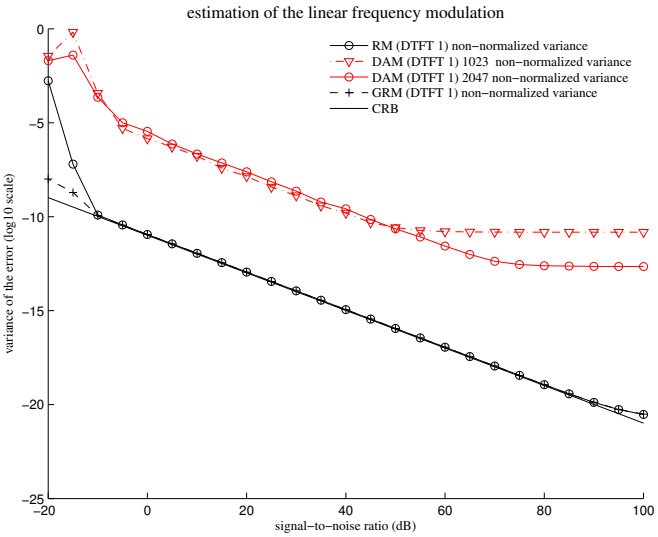


Figure 4.20: All methods (full frequency range): linear FM estimation error

4.3 Signal-to-Residual-Ratio Comparison

In the preceding section 4.2 an extensive comparison of DAM, RM and GRM algorithms was presented. The Distribution Derivative Method (Betser, 2009) falls into the same category as the aforementioned algorithms, an extensive *per-parameter* comparison of DAM, RM, GRM and DAM was conducted in (Hamilton and Depalle, 2012a), suggesting the GRM and DDM achieve similar results for the second degree polynomial model. In this section, a more elaborate comparison of GRM and DDM with respect to algorithm parameters was conducted. The DDM allows for an arbitrarily overdetermined linear system, giving an extra dimension of freedom over GRM. When considering only the bins of FFT it is expected that zero-padding will have a more pronounced effect on the DDM than on GRM since required frequency bins might end up very close together by the zero-padding process, causing an ill-conditioned linear system, whereas for GRM only 1 frequency bin is required. Naturally, it is possible to avoid such ill-conditioning by altering the bin selection strategy, however it will be assumed that choosing the frequency bins as close to each other as possible is desired, in order to minimise the interference of any neighbouring sinusoids. The results are based on exact same test set as the previous section 4.2 (full frequency range). The algorithm parameters, i.e. the zero-padding and the number of bins used (DDM), are varied to study the overall quality of the algorithms.

To meaningfully compare the overall algorithm accuracy (in SRR) to the CRBs, a *maximum* achievable SRR corresponding to CRBs for the specific test case should computed:

$$\max SRR(\hat{s}(r_2 \pm \sqrt{\epsilon_{r_2}}, r_1 \pm \sqrt{\epsilon_{r_1}}, r_0 \pm \sqrt{\epsilon_{r_0}}), \tag{4.37}$$

where ϵ_{r_k} is the CRB for parameter r_k and $\hat{s}$ is the generalised sinusoid with the specified parameters. For the following tests the equation 4.37 was evaluated in a *brute force* manner for each test case in the set. The figures below depict the average and standard deviation over all the test cases to accommodate for variations of CRBs and algorithm accuracies across the test cases.

For GRM the *Hann* window cannot be used due to specific requirements (see section A.3), however the tests show only marginal accuracy degradation (compared to CRB) of about 5dB in the low to mid-SNR region (see figures 4.21,4.22). In the high-SNR region (above 30dB) the effect of discontinuous second window derivative significantly reduces the accuracy as the maximum SRR of about 40dB is not surpassed. It could be argued that

such high-SNR cases aren't very common in most real-world applications and desirable properties of the *Hann* window might be of higher priority. Using $Hann^2$ window (figure 4.22) clearly renders GRM near-optimal in almost all SNR cases, with less than 5dB below CRB. Zero-padding does not significantly affect the GRM accuracy as the only benefit is that the initial frequency estimate is generally closer to the actual value. This is in accordance with figures 4.12- 4.15 in section 4.2, where it was shown that the effect of re-estimation (i.e.: using the frequency estimate in the next algorithm run) does not yield any observable improvement.

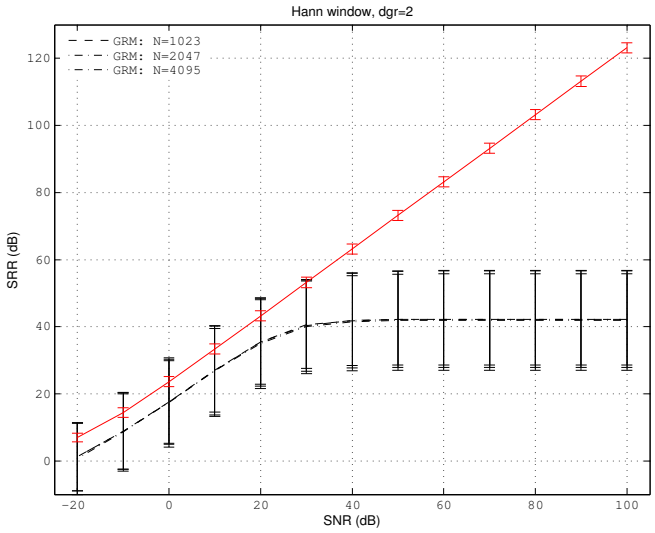


Figure 4.21: GRM: *Hann* window, $dgr = 2$, SRR average and standard deviation for different zero-paddings

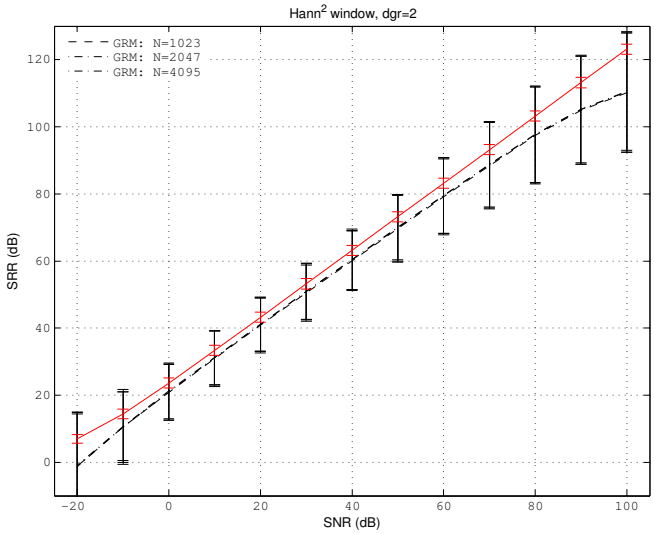


Figure 4.22: GRM: $Hann^2$ window, $dgr = 2$, SRR average and standard deviation for different zero-paddings

For DDM the $Hann$ window is used due to no specific requirements for its desirable time-frequency properties. The terms $S_w(\omega)$ and $S_{t^k w}(\omega)$ were computed via FFT for all the frequency bins (rather than much slower DTFT), however only 2, 5 and 7 bins were used, thus the CPU consumption shown on the figures might depict a higher value than the absolute minimum required. Zero-padding does have a measurable effect on the estimator when the system is overdetermined ($dgr = 2, Q = 5$ and $dgr = 2, Q = 7$). Interestingly, 4 times zero-padding reduces the accuracy of up to 10dB, yielding an increased computation time (shown next to the legend) and lower quality. For the current test set and frame length, 5 frequency bins and no zero-padding practically overlaps with the maximum SRR for CRBs computed for $N = 1023$. It thus seems desirable to keep the frame size at minimum to increase the accuracy of DDM, contrary to a general consensus that applies for most TF analysis methods.
From the figures and discussion it is reasonable to deduce that overdetermined ($Q = 5$) DDM with $Hann$ window, no zero-padding and GRM with $Hann^2$ window, no zero-padding represent near-optimal algorithms for the 3^{rd} degree polynomial generalised sinusoid. It is important to note that DDM $Q = 5$ uses 3 more frequency bins than the $Q = 2$ and is thus expected to be more susceptible to inter-partial interference.

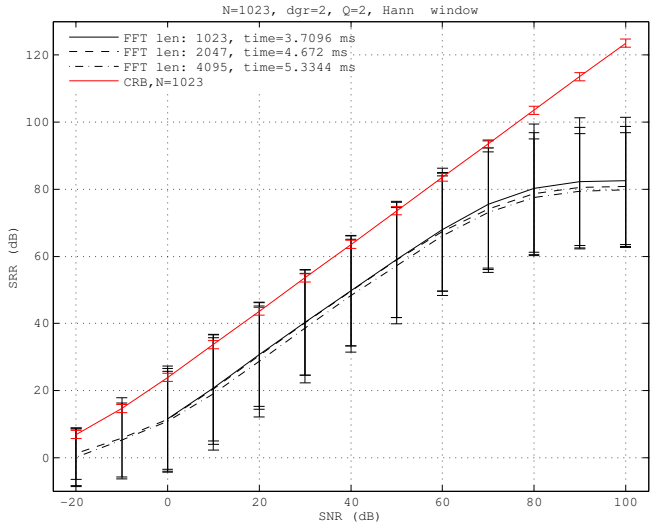


Figure 4.23: DDM: *Hann* window, $dgr = 2, Q = 2$, SRR average and standard deviation for different zero-paddings

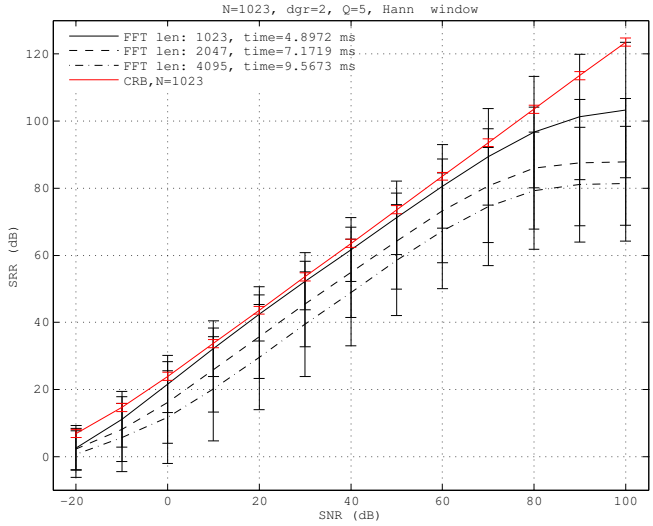


Figure 4.24: DDM: *Hann* window, $dgr = 2, Q = 5$, SRR average and standard deviation for different zero-paddings

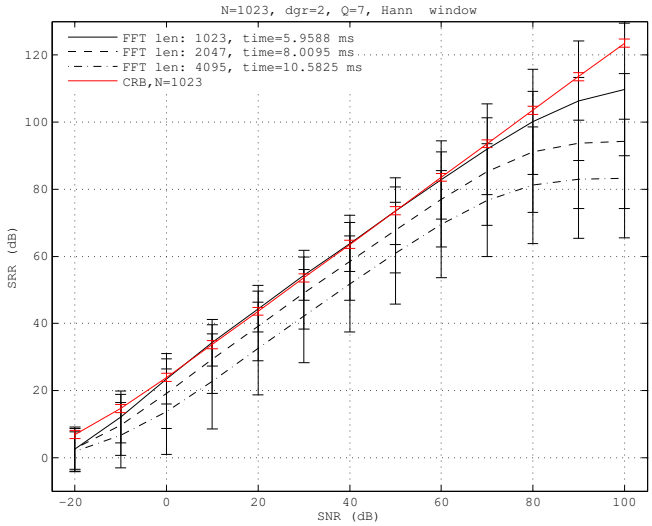


Figure 4.25: DDM: *Hann* window, $dgr = 2, Q = 7$, SRR average and standard deviation for different zero-paddings

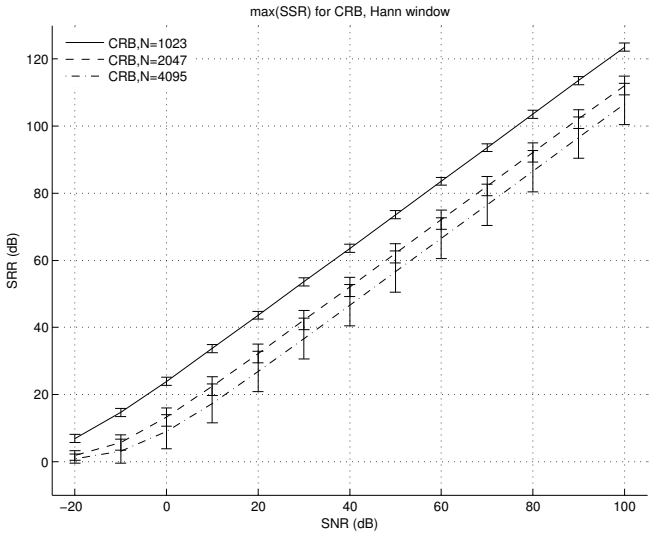


Figure 4.26: CRB: SRR average and standard deviation for different zero-paddings

4.4 Frequency dependent SRR Comparison

Naturally, lowering the frequency or shortening the analysis frame eventually decreases the accuracy. To observe the dependency of accuracy on the frequency, the test results described recently will be averaged on all the parameters except frequency. Figures 4.27,4.28 and 4.29 show SRR for each tested frequency below 10kHz.
Importantly, in all the cases the frequency accuracy curves largely overlap for majority of the SNR. As already observed in section 4.2 the accuracy reaches a plateau at some SNR, after which no further improvement can be observed. This cut-off SNR is lower with frequencies approaching 0 and frequencies approaching $f_s/2$ (figures 4.30,4.31 and 4.32. Frequencies above 20kHz were not tested and since $f_s/2 = 22050Hz$ the self-interference is not as pronounced (2000Hz difference) as in the lower frequency region (only 200Hz difference).
For the lowest frequency tested (200Hz) there is a significant drop for low-noise region in all cases. The cut-off SNR and plateau level stay roughly the same for $Q = 5$ and $Q = 7$, however the mid-noise region suffers a bigger drop for $Q = 7$, signalling the effect of self-interference is starting to dominate as the 2 extra frequency bins boost the interference slightly.

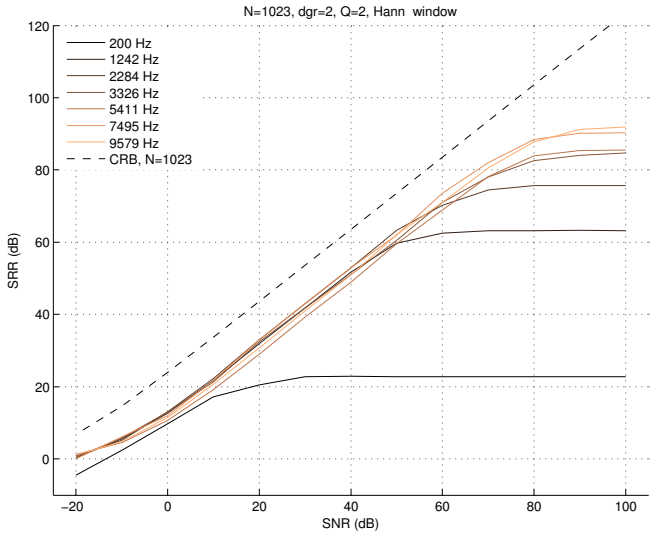


Figure 4.27: DDM: $Hann$ window, $dgr = 2, Q = 2$, no zero-padding, SRR average for each tested frequency below 10kHz

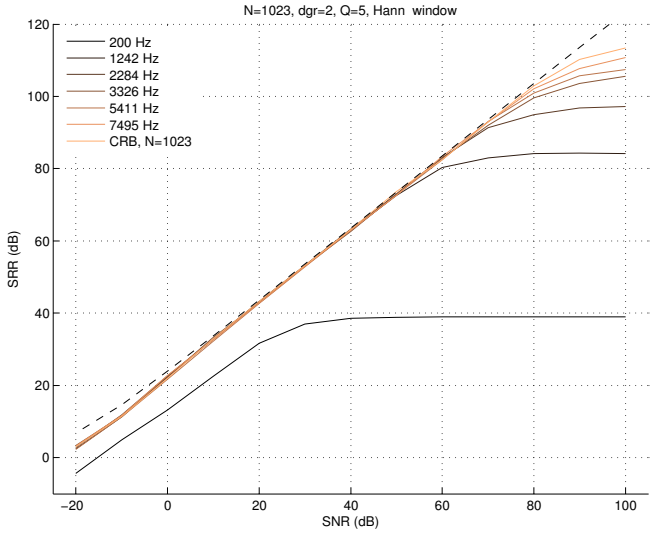


Figure 4.28: DDM: *Hann* window, $dgr = 2, Q = 5$, no zero-padding, SRR average for each tested frequency below 10kHz

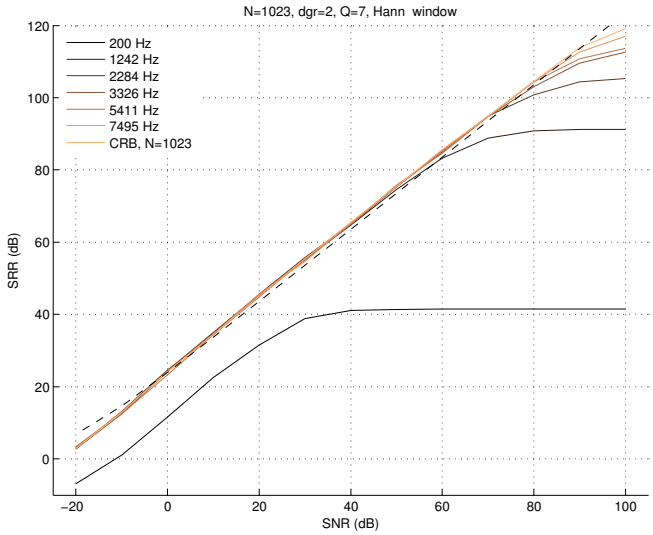


Figure 4.29: DDM: *Hann* window, $dgr = 2, Q = 7$, no zero-padding, SRR average for each tested frequency below 10kHz

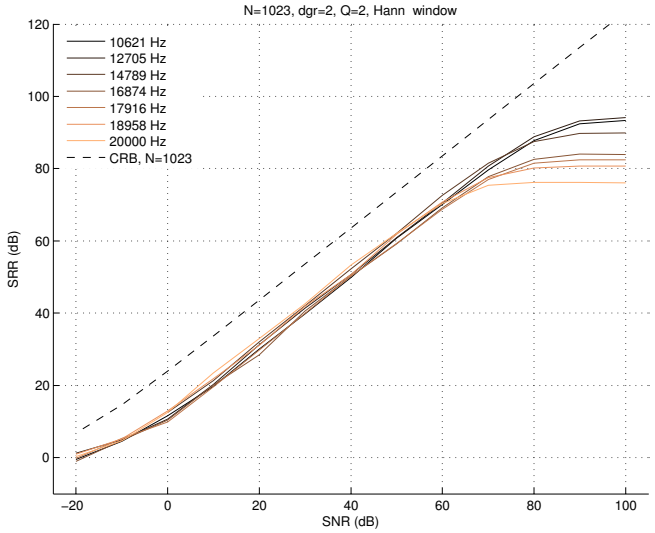


Figure 4.30: DDM: *Hann* window, $dgr = 2, Q = 2$, no zero-padding, SRR average for each tested frequency above 10kHz

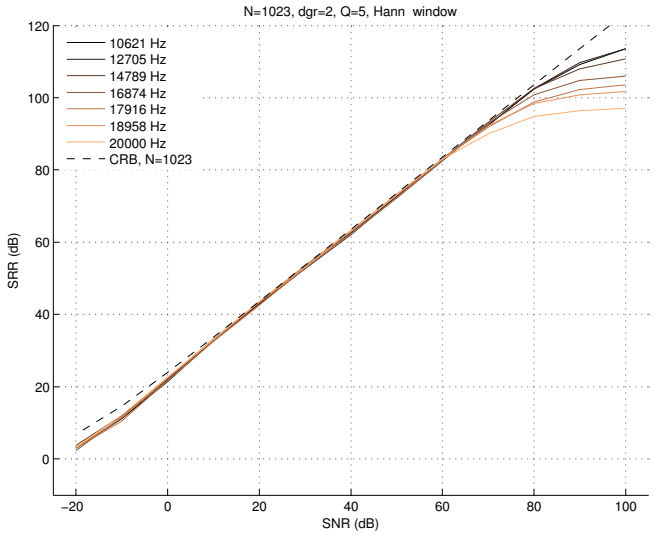


Figure 4.31: DDM: *Hann* window, $dgr = 2, Q = 5$, no zero-padding, SRR average for each tested frequency above 10kHz

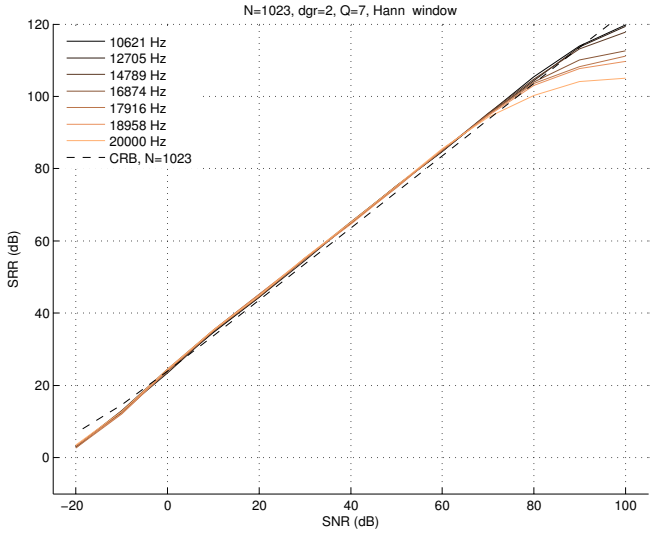


Figure 4.32: DDM: *Hann* window, $dgr = 2, Q = 7$, no zero-padding, SRR average for each tested frequency above 10kHz

Figures 4.33, 4.34 and 4.35 show a more compact representation of frequency dependant accuracy plots. Again, the slightly overdetermined version of DDM, $Q = 5$ is proven to be a good balance between low-bandwidth requirement and high accuracy. The low frequency range is affected significantly for all cases, however comparing the $Q = 2$ and $Q = 5$ versions, the lower bandwidth requirement of the $Q = 2$ does not boost the performance noticeably, at least at the current frequency test set. A more detailed tests focused only on low-frequency range would be required. The symmetrical, bell shaped accuracy curve for the analysis of low-noise signals (SNR ¿ $80dB$) can be attributed to spectral leakage (self-interference), which is completely dominated by the effect of noise at mid and high-noise levels at the whole frequency range excluding very low frequencies.

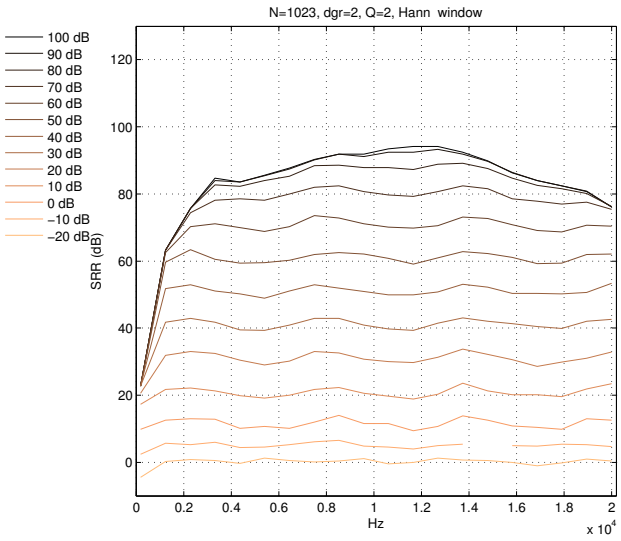


Figure 4.33: DDM: $Hann$ window, $dgr = 2, Q = 2$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend)

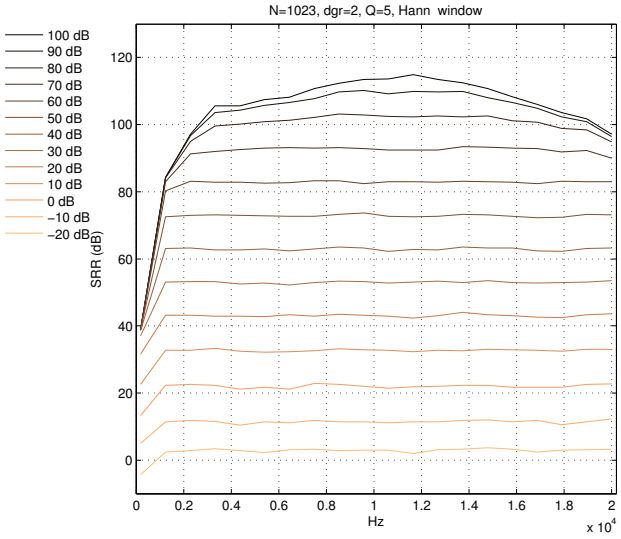


Figure 4.34: DDM: $Hann$ window, $dgr = 2, Q = 5$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend)

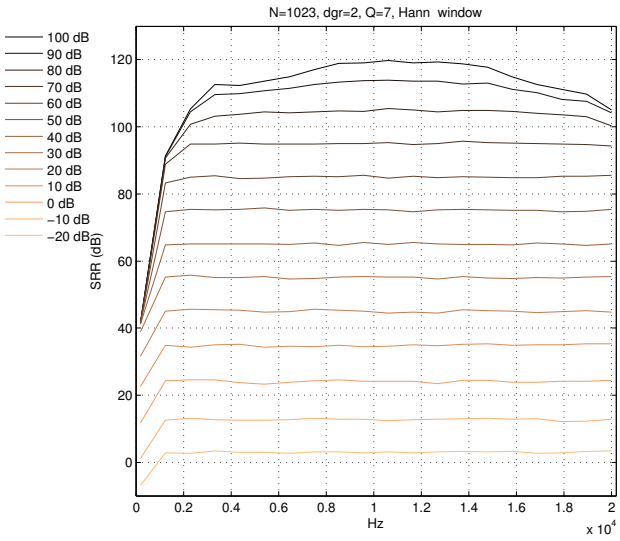


Figure 4.35: DDM: *Hann* window, $dgr = 2, Q = 7$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend)

GRM exhibits similar yet not completely identical behaviour to the DDM. Accuracy at high-frequencies (see figure 4.37) resembles very much the corresponding figures 4.28 and 4.31. In contrast, the low-frequency range accuracy does not exhibit a significant drop in the low to mid noise range as it is the case with DDM (figures 4.28 and 4.31).

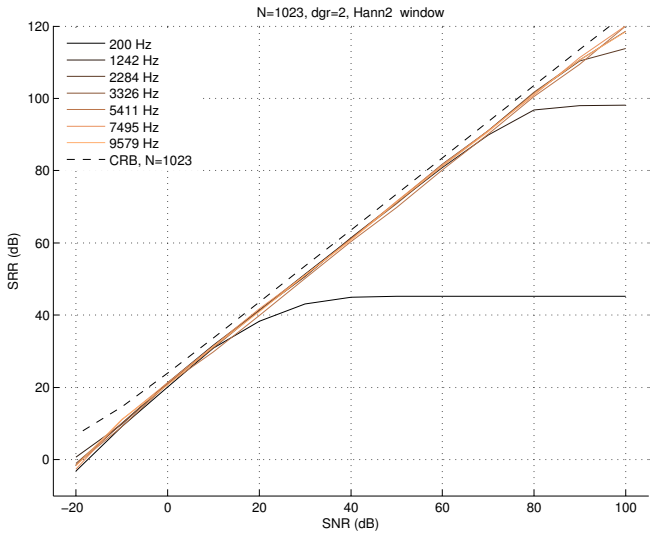


Figure 4.36: GRM: $Hann^2$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency below 10kHz

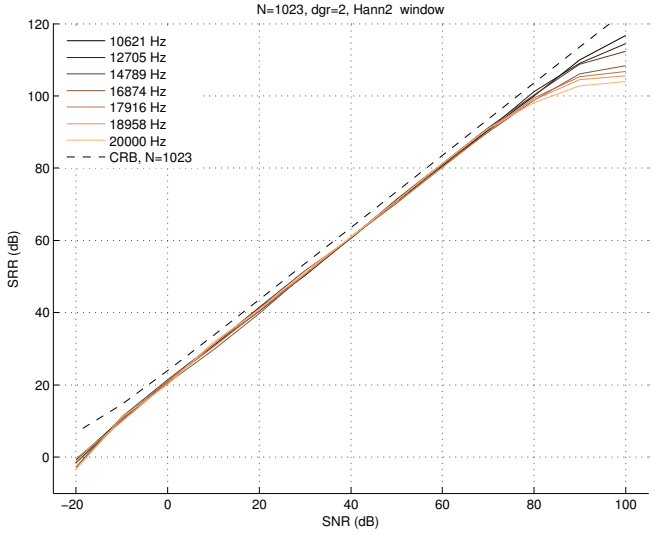


Figure 4.37: GRM: $Hann^2$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency above 10kHz

Figure 4.38 reveals rather poor performance due to the unsuitable use of

the $Hann$ window while figure 4.39 ($Hann^2$ window) exhibits similar trend as that of DDM $Q = 5, Q = 7$. In very low-noise cases, (SNR=100dB) a gradual drop with increasing frequency can be observed and since the drop starts at about 5000Hz it cannot be attributed to the spectral leakage (self-interference).

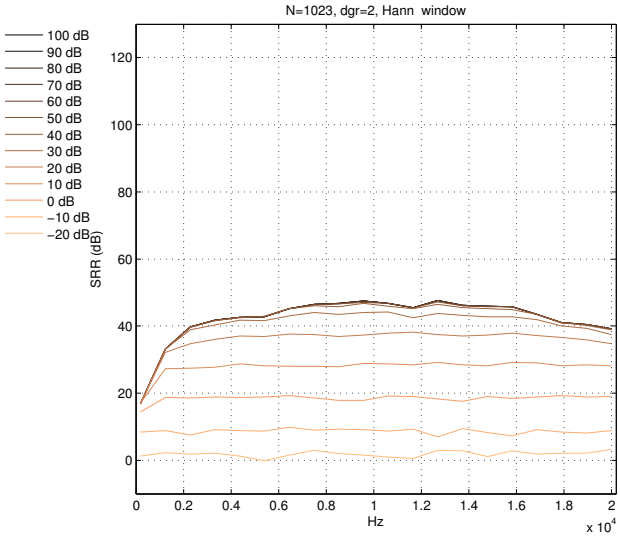


Figure 4.38: GRM: $Hann$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend)

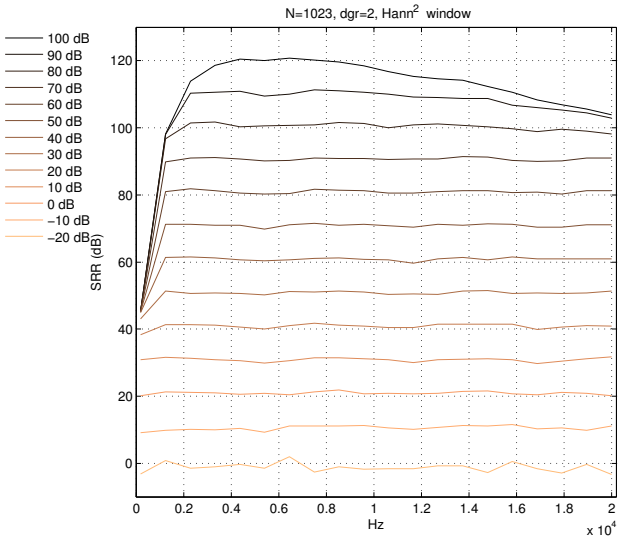


Figure 4.39: GRM: $Hann^2$ window, $dgr = 2$, no zero-padding, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend)

From the figures and discussion in this section it is clear that for estimation of non-stationary sinusoidal parameters zero-padding is not required. Further, for the second degree polynomial generalised sinusoid a slightly overdetermined DDM with $Hann$ and GRM with $Hann^2$ window perform comparably in most cases.
Since both methods derive a linear system of equations, from which the estimates are calculated, the condition number represents an important measure for numerical stability. Figure 4.40 depicts the $2 - norm$ condition number:

$$cond(A) = \frac{\max SVD(A)}{\min SVD(A)}. \tag{4.38}$$

Above expression can be calculated for square or rectangular matrix alike and large numbers mean badly conditioned system. An important trend difference is observed: the system conditioning for DDM does not depend on frequency and remains constant above 10^{-3}. GRM on the other hand, suffers from much more ill-conditioned system which progressively worsens with increasing frequency. Such trend explains the asymmetric accuracy curve observed in figure 4.38, where accuracy unexpectedly drops with increasing frequency, dominating the effect of self-interference not exhibited by the DDM.

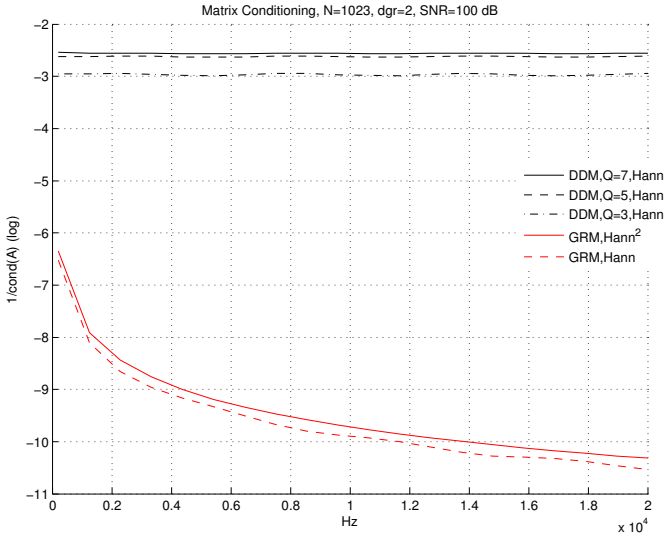


Figure 4.40: Linear system conditioning: $dgr = 2$, no zero-padding, condition number average for each tested frequency

4.5 Low-frequency analysis with GRM and DDM

In the preceding section 4.4 an accuracy overview for full frequency range has been presented. As expected, the spectral analysis algorithms behave substantially different at frequencies close to 0 and $f_s/2$, when self-interference becomes the strongest contributing factor causing reduced quality estimations.
A frequency range between 20 and 500Hz is put into focus in this section. At such low frequencies, the window function TF properties become crucial as can be deduced from the following figures.
The DDM with $Hann^2$ (figures 4.43 and 4.44) window yields superior accuracy throughout the specified frequency range for the SNR range above 0dB, whereas the ultra-low SNR range (below 0dB) is not affected by the choice of the window. The GRM with $Hann^2$ window performs similarly to the DDM with $Hann$ with the exception of high-noise region (below 0dB) where GRM performs better with frequencies down to 125Hz.

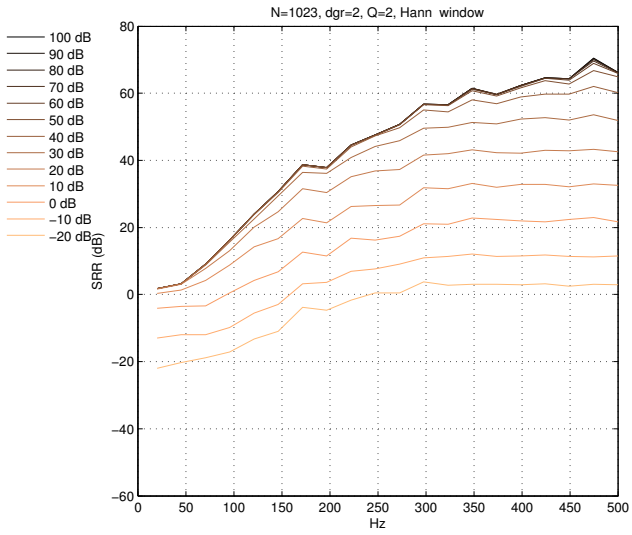


Figure 4.41: DDM - low frequency range (20-500Hz): $dgr = 2, Q = 2$, no zero-padding, $Hann$ window, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend

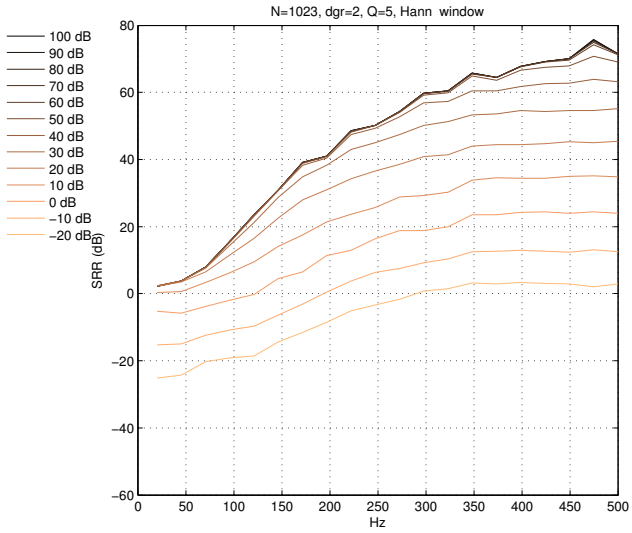


Figure 4.42: DDM - low frequency range (20-500Hz): $dgr = 2, Q = 5$, no zero-padding, $Hann$ window, SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend

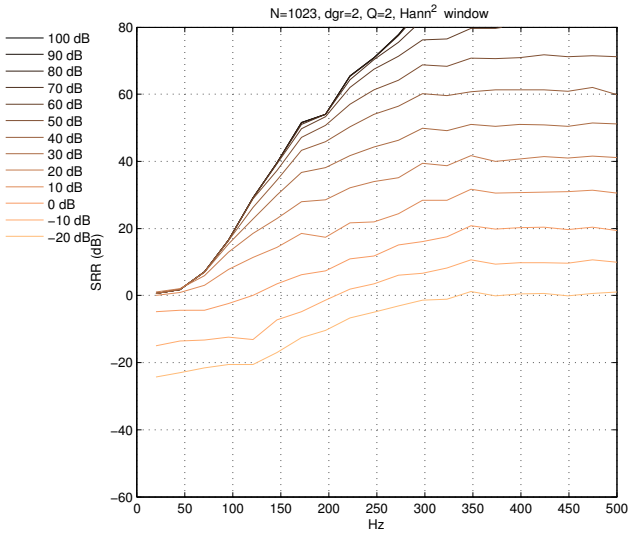


Figure 4.43: DDM - low frequency range (20-500Hz): $dgr = 2, Q = 2$, no zero-padding, $Hann^2$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend

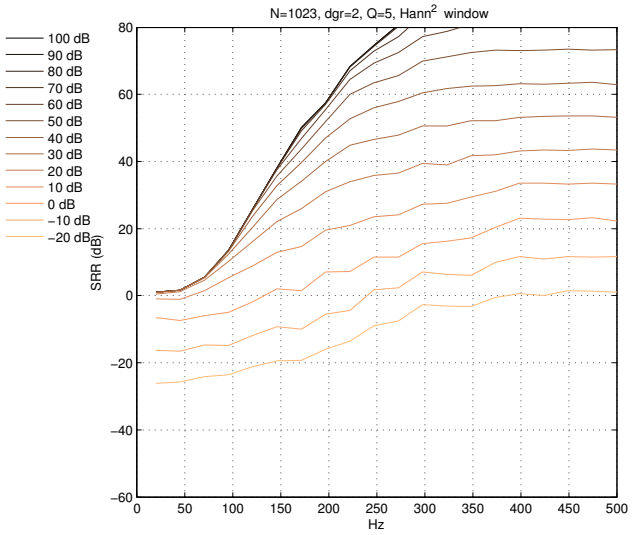


Figure 4.44: DDM - low frequency range (20-500Hz): $dgr = 2, Q = 5$, no zero-padding, $Hann^2$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend

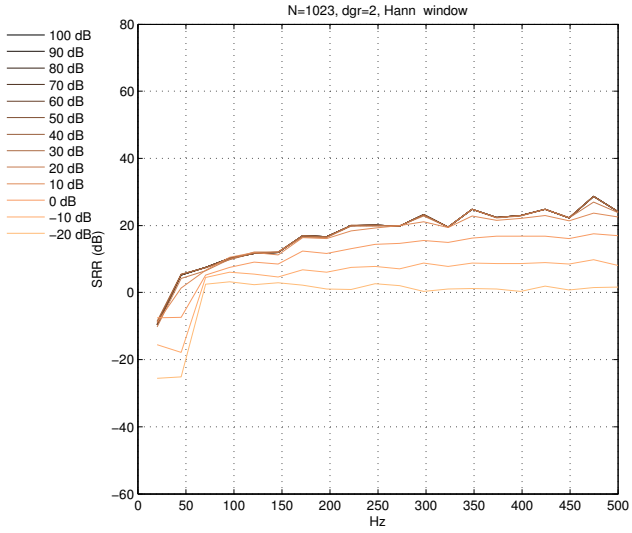


Figure 4.45: GRM - low frequency range (20-500Hz): $dgr = 2$, no zero-padding, $Hann$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend

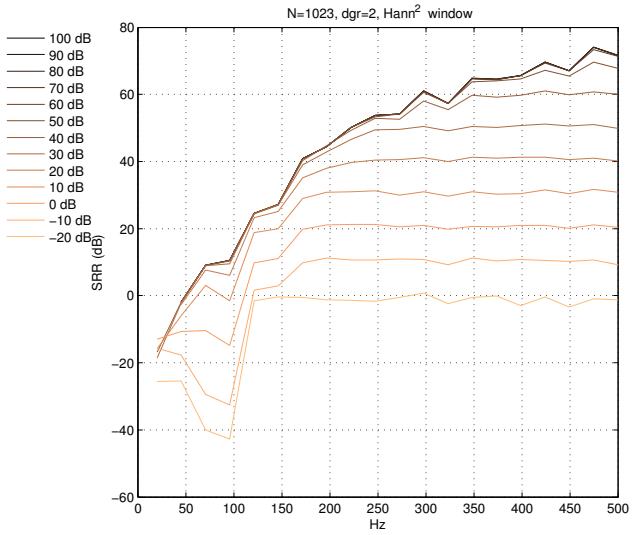


Figure 4.46: GRM - low frequency range (20-500Hz): $dgr = 2$, no zero-padding, $Hann^2$ SRR average for each tested frequency, -20dB to 100dB SNR (designated on the legend

4.6 Conclusion

In the preceding section 4.5 a detailed comparison of the second degree polynomial DDM and GRM in variety of situations was tested and compared. It is easy to conclude that DDM is the most accurate and flexible method. It is attractive to consider a 3^{rd} order modulation model and re-evaluate the methods the GRM might outperform DDM. It would greatly time consuming to run complete tests extended to 3^{rd} order, however initial tests for a single frequency reveal enough information to conclude the GRM vs DDM comparison. For 3^{rd} degree GRM the condition number drops below 10^{-18}, which is well below the boundary at which the results can be considered appropriate by the Matlab programming language. Indeed, initial test confirm the estimate error rises significantly. The condition number for the DDM however drops to about 10^{-6}, still considered within bounds. Full test suite was not performed as it is clear that DDM is significantly more numerically stable than GRM. The GRM requires higher window derivatives with each degree. It is trivial to see that each window derivative is of order f_s/N higher than the previous one, thus some of the entries to the linear system matrix increase exponentially. Secondly, the higher degree polynomial terms decrease exponentially with the same factor, forcing some of the entries to a severely low values. Combination of both phenomena causes this rapid drop in system condition number for GRM. The DDM on the other hand only suffers from the second phenomena. It is easy to imagine a partial solution to this problem, by simply normalising the model functions. Still, DDM will exhibit higher numerical stability and it can thus be considered of superior accuracy. In addition, the ability to arbitrarily over-determine the linear system without side effects will prove to be of tremendous advantage when analysing highly modulated signals, see chapter 10.

Chapter 5

Reassignment with adaptive Fourier poly-phase kernel

It has been shown in (Betser, 2009) that extreme modulation cause reduced estimation accuracy for the distribution derivative method due to reduced signal to kernel correlation for FT. This makes it impossible to analyse highly modulated sinusoids even with presented non-stationary methods if they rely on FT as the underlying transform. As it will be shown, any number of arbitrary kernels can be used as there is absolutely no restrictions. An appropriate selection of the set of the kernel functions is a very different subject and depends on the signals under study.
It is highly desirable that the transform correlates well with the signal under study - very high amounts of FM clearly lower the correlation between the signal and the stationary complex sinusoid, the Fourier kernel. In this section an adaptive method based on reassignment for the analysis of the generalised sinusoid model $exp(r(t)), r(t) \in \mathbb{C}$ (see section 1.1) is presented. The adaptive kernel proposed is a polynomial-phase complex sinusoid $\exp(j(\omega t + \psi t^2))$.

5.1 GRM using a generic kernel

Using the integration *per partes*, Leibniz integration rule and the restriction $w(-\frac{T}{2}) = w(\frac{T}{2}) = 0$ (required by the generalized reassignment), the following useful equality can be produced (for complete derivation see (Betser,

2009; Wen and Sandler, 2009; Muševič and Bonada, 2010a)):

$$\frac{\partial}{\partial t}\langle s, w\Psi\rangle = -(\langle s, w\Psi'\rangle + \langle s, w'\Psi\rangle) \tag{5.1}$$

The above equality is more commonly referred to as the *distribution derivative* rule $\langle x', y\rangle = -\langle x, y'\rangle$. Left hand side can also be expressed as:

$$\begin{aligned} &\frac{\partial}{\partial t}\langle s, w\Psi\rangle = \langle \frac{\partial}{\partial t} s, w\Psi\rangle = \\ &\langle r's, w\Psi\rangle = \sum_{k=1}^{K} r_k \langle m_k' s, w\Psi\rangle \Rightarrow \end{aligned} \tag{5.2}$$

$$\sum_{k=1}^{K} r_k \langle m_k' s, w\Psi\rangle = -(\langle s, w\Psi'\rangle + \langle s, w'\Psi\rangle). \tag{5.3}$$

To compute $K-1$ non-stationary parameters, another $K-2$ time derivatives are required. Its computation can efficiently be performed by the following *pyramid-like* scheme:

$$\begin{array}{ccccc} & & \langle sh, \Psi_g w\rangle & & \\ & \swarrow & & \searrow & \\ & -\langle sh, \Psi_g' w\rangle & & -\langle sh, \Psi_g w'\rangle & \\ \swarrow & & \searrow & & \swarrow & & \searrow \\ \langle sh, \Psi_g'' w\rangle & & +2\langle sh, \Psi_g' w'\rangle & & + \langle sh, \Psi_g w''\rangle, & & \\ & \vdots & & & & & \end{array} \tag{5.4}$$

where $h(t)$ stands either for $h(t) = 1$ to calculate right hand side or $h(t) = h_k'(t), k = 1 : K-1$ to calculate the left hand side of the equation 5.3. Importantly, the above derivation holds for any kernel Ψ and window function w as long as $w^{(k-1)}\left(-\frac{T}{2}\right) = w^{(k-1)}\left(\frac{T}{2}\right) = 0$ for the desired k.

5.2 Polynomial-phase Fourier kernel

In Betser (2009) it was demonstrated that the estimation accuracy is inversely proportional to the kernel-to-signal correlation. Therefore maximising the correlation should improve the accuracy and since the signal is

modelled as a non-stationary sinusoid, a natural choice for kernel function would be the same as the model. The proposed kernel function follows:

$$\Psi_g(t) = e^{jg(t)}, \tag{5.5}$$

where $g(t)$ is a real polynomial of order M: $g(t) = \sum_{k=1}^{K} g_m t^k$. Note that $g_0 = 0$, as any non-zero value would introduce bias in the phase estimation. From scheme 5.4 it is clear that a $(K-1)$-th degree time derivative of the kernel function is required. The specific case of the polynomial-phase Fourier kernel the following scheme similar to 5.4 can be used in order to calculate the kernel function time derivatives:

$$\begin{array}{ccccccc} & & \Psi'_g = g'\Psi_g & & & & \\ & & \swarrow & & \searrow & & \\ & \Psi''_g = g''\Psi_g & & + & & g'\Psi'_g & \\ & \swarrow & & \searrow\swarrow & & \searrow & \\ \Psi'''_g = g'''\Psi_g & & + & & 2g''\Psi'_g & + & g'\Psi''_g, \\ & & \vdots & & & & \end{array} \tag{5.6}$$

The main advantage of such algorithm is a less restricted kernel, thus the selection of $m_k(t)$ functions can therefore be matched with appropriate kernel functions to maximize correlation and avoid accuracy deterioration in the case of extreme parameter values.

The algorithm should initially be invoked with $g(t) = j\hat{\omega}t$, where $\hat{\omega}$ is a frequency of the magnitude spectrum peak. This yields an initial estimate of the polynomial $r(t)$: $\hat{r}(t) = \sum_{k=1}^{K} \hat{r}_k t^k$. This initial run of the algorithm is identical to generalized reassignment as described in (Wen and Sandler, 2009). In the second iteration the kernel function can be adapted to the signal by setting $g(t) = j\Im(\hat{r}(t)) = j\sum_{k=1}^{K} \hat{q}_k t^k$.

From equations 5.3, 5.4 and 5.6 the following linear system can be directly deduced:

$$A = \begin{array}{lll} \langle s, \Psi_g w\rangle & \langle s, \Psi'_g w\rangle + \langle s, \Psi_g w'\rangle & \dots \\ \langle \mathbf{st}, \mathbf{\Psi_g w}\rangle & \langle st, \Psi'_g w\rangle + \langle st, \Psi_g w'\rangle & \dots \\ \langle st^2, \Psi_g w\rangle & \langle st^2, \Psi'_g w\rangle + \langle st^2, \Psi_g w'\rangle & \dots \\ \vdots & \vdots & \ddots \end{array} \tag{5.7}$$

Of a particular interest is the term written in bold, $\langle st, \Psi_g w\rangle$. When the kernel $\Psi_g(t)$ closely matches the target signal $s(t)$ then the product $\bar{\Psi}_g(t)s(t) \approx 1$ and the following can be deduced:

$$\langle st, \Psi_g w\rangle = \int ts(t)\bar{\Psi}_g(t)w(t)dt \approx \int tw(t)dt. \tag{5.8}$$

For any symmetric window function $w(t)$ and $t \in [-\frac{T}{2}, \frac{T}{2}]$ the above expression is very close to 0. Such cases occur when the signal exhibits low or no amplitude modulation causing the linear system of equations close to singular, rendering the algorithm essentially useless as the condition number of the matrix approaches 0. Such a drawback can simply be avoided by artificially inducing some amplitude modulation into the signal and then subtracting it from the estimate obtained. A very small amount of the amplitude modulation of magnitude around 10^{-10} is sufficient to stabilize the system and significantly improve the estimates. Importantly, it is expected that in practice the noise level will dominate the effect of the aforementioned artificial AM anyway.

5.3 Tests and Results

The tests conducted were identical to those in Wen and Sandler (2009). The metric used was the signal to residual ratio (SRR):

$$SRR = \frac{\sum_{i=1}^{N} h_i s_i^2}{\sum_{i=1}^{N} h_i (s_i - \hat{s}_i)^2}, \tag{5.9}$$

where $s_i, i = 1..N$ are samples of the original signal $s(t)$ (without noise), $\hat{s}_i, i = 1..N$ are the samples of the estimated signal and $h_i, i = 1..N$ are samples of the weighting function - Hanning window. A model degree of 3 was chosen and the Hann^2 function of length 1024 was used as the window function. The test signals analyzed were real sinusoids sampled at 44100Hz. The parameters of the test sinusoids were varied in the following way: 10 phase values in the [0,0.45]π interval, 10 linear log-amplitude modulation values in the [0,0.0045] /frame interval (roughly corresponds to the [0,200] /s interval), 10 frequency values in the [255,255.9] bins interval (roughly corresponds to the [10.982, 11.021]Hz) and 10 linear frequency modulation values in the [0,27] bins/frame interval (roughly corresponds to the [0,16.000] Hz/s). The tests were conducted in 3 separate groups for the original reassignment (labeled GEN RM) and the one using the polynomial-phase kernel (labeled GEN RM PPT). In group 1 (figure 5.1), the linear frequency modulation was set to 0 while the log-amplitude modulation was varied (x-axis) in the mentioned range. In group 2 (figure 5.2) the log-amplitude modulation was set to 0 while the linear frequency modulation was varied (x-axis) in the mentioned range. In group 3 (figure 5.3), both the FM and log-AM were jointly varied (x-axis) in double the range compared

to the groups 1 and 2. In the first part (labeled SNR: Inf dB in the plots) no noise was added to the signal and in the second part (labeled SNR: 0dB in the plots) a Gaussian white noise of the energy equal to that of the clean signal was added. The range of the log-AM/FM for group 3 was doubled intentionally to examine properties of both algorithms in highly modulated cases.

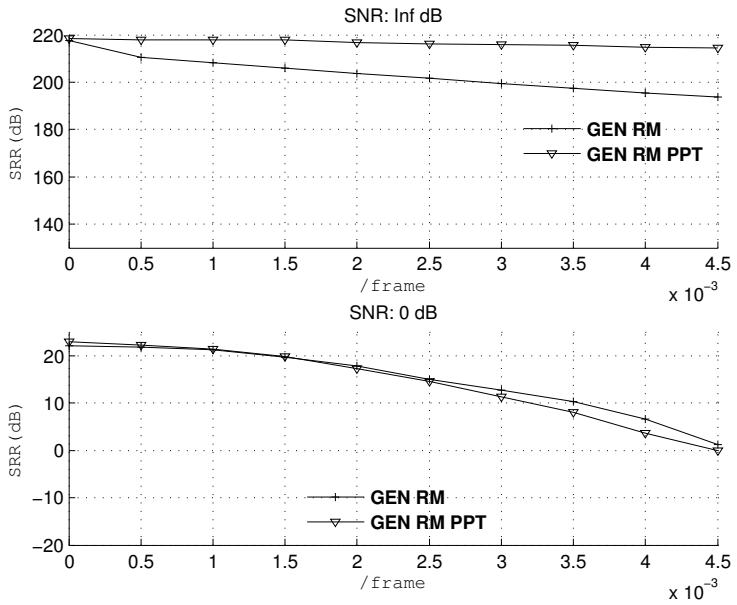


Figure 5.1: Group 1 (AM only)

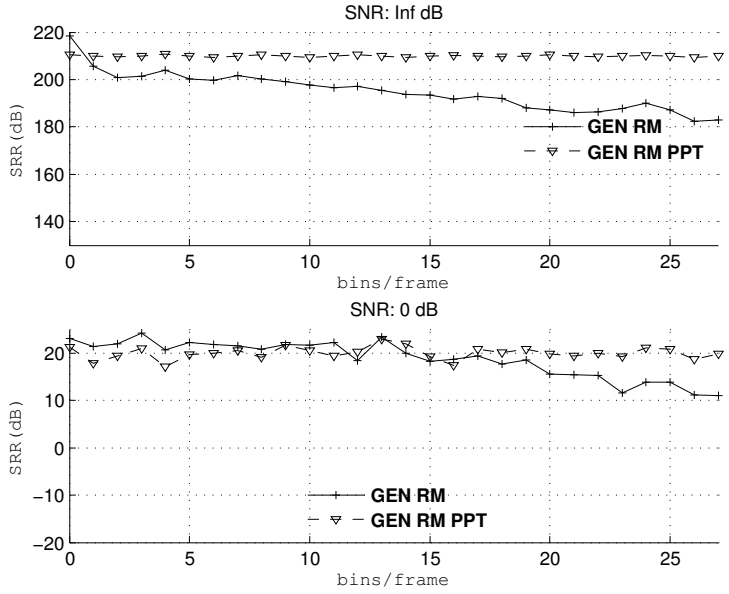


Figure 5.2: Group 2 (FM only)

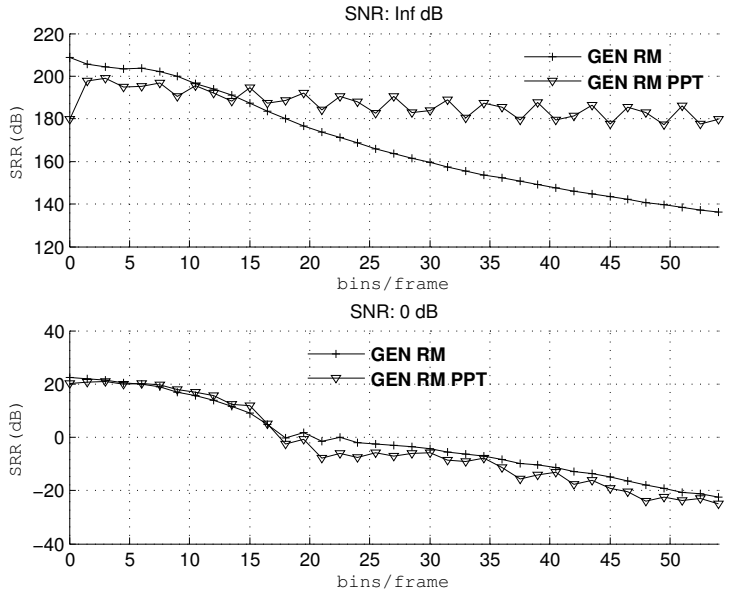


Figure 5.3: Group 3 (AM and FM)

5.4 Conclusion

In this chapter, an improvement of the generalised reassignment method was described. The main idea of the improvement is the use of an adaptive polynomial-phase Fourier kernel in conjunction with the general reassignment algorithm. The algorithm exhibits a significant improvement in accuracy compared to the original method in the case of clean signal, as the effect of frequency modulation is minimised by the adaptive kernel. For a stationary sinusoid, the accuracy is comparable to the original method, however an increase in accuracy is observed in the case of non-stationary ones, reaching almost 50dB in the most modulated case (group 3). The method does not improve the analysis of the original algorithm if 0 dB Gaussian white noise added. The reason for this is the kernel adaptation works in the opposite way to which is desired. This is because it uses the estimate of the original method, which is not precise enough at such a high noise level, therefore the error in the input parameters corrupts the final estimate.
In group 3, the most modulated case corresponds to 32.000Hz/s change. This may seem excessive for analysing real world music related signals. However, a higher order modulation polynomials could exhibit even larger linear FM values, as its contribution can be cancelled or balanced out by the second or higher order terms. So as the kernel is adapted to the sinusoid in question, the energy concentration of its representation in the transform domain is increased: the bandwidth of the non-stationary sinusoid is reduced. This is a desirable property in the case of multicomponent signals, where side-lobes of a sinusoid cause significant interference to the neighbouring partials.

Chapter 6

cPACED analysis using Gamma function

cPACED model is unique in a way that it is the only one with analytically computable FT for wide variety of window functions. In this chapter an algorithm based on analytical FT of the cPACED model will be described and evaluated.
The Gamma function (Bateman Manuscript Project et al., 2006c) is the extension of the factorial function (with its argument shifted down by 1) to real and complex numbers. For a positive integer n, the Gamma function is defined as:

$$\Gamma(n) = (n-1)!. \tag{6.1}$$

The Gamma function is defined for all complex except negative integers and zero. For complex numbers with positive real part, it is defined as an improper integral:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \tag{6.2}$$

One can quickly recognise similarity of the integrand with the cPACED model (see definition in section 1.1):

$$s(t) = a(t)e^{\beta_0 t} = \sum_k^K a_k t^k e^{(\alpha + j\omega_0)t}. \tag{6.3}$$

6.1 Lower incomplete Gamma function and cPACED model

Since the analysis frame is always time limited in practice the upper bound in equation 6.3 is inappropriate. The *lower incomplete* Gamma function is conveniently defined as:

$$\gamma(k,x) = \int_0^x t^{k-1} e^{-t} dt. \tag{6.4}$$

Each term of the STFT of the signal in equation 6.3 can now be expressed in terms of equation 6.4 (assuming rectangular window):

$$\begin{aligned} &\int_{-\frac{T}{2}}^{\frac{T}{2}} t^k e^{(\alpha+j\omega_0)t} e^{-j\omega t} dt = \\ &\frac{(-1)^k}{(\alpha+j(\omega_0-\omega))^k} \int_{\frac{T}{2}(\alpha+j(\omega_0-\omega))}^{-\frac{T}{2}(\alpha+j(\omega_0-\omega))} \tau^k e^{-\tau} d\tau = \\ &\frac{(-1)^k}{(\alpha+j(\omega_0-\omega))^k} \left(\gamma\Big(k+1, -\frac{T}{2}(\alpha+j\,(\omega_0-\omega)\,)\Big) - \gamma\Big(k+1, \frac{T}{2}(\alpha+j(\omega_0-\omega))\Big)\right) \end{aligned} \tag{6.5}$$

Integration *per-partes* allows for the following expression:

$$\gamma(k,x) = (k-1)\gamma(k-1,x) - x^{k-1}e^{-x}. \tag{6.6}$$

Together with $\gamma(1,x) = 1 - e^{-x}$ and assuming k to be a non-negative integer as required by 6.3 the following closed form expression for each term of the STFT can be obtained:

$$\gamma(k,x) = (k-1)! \left(1 - e^{-x} \sum_{n=0}^{k-1} \frac{x^k}{n!}\right). \tag{6.7}$$

Substituting $\beta_0 = \alpha + j(\omega_0 - \omega)$, the FT of signal 6.3 can be expressed:

$$\begin{aligned} \langle s, e^{j\omega} \rangle &= \sum_{k=0}^{K} a_k \langle t^k e^{(\alpha+j\omega_0)t}, e^{j\omega} \rangle = \\ \sum_{k=0}^{K} a_k \frac{(-1)^k}{\beta_0^k} &\left(\gamma\Big(k+1, -\frac{T}{2}\beta_0\Big) - \gamma\Big(k+1, \frac{T}{2}\beta_0\Big)\right). \end{aligned} \tag{6.8}$$

The left hand side of the above equation can be computed from the signal and can be considered a constant. A quick glance over the above expressions

it is easy to see that equation 6.8 contains a linear combination of terms each consisting of two polynomials (in terms of $\frac{T}{2}\beta_0$), multiplied by $\exp(\frac{T}{2}\beta_0)$ and $\exp(-\frac{T}{2}\beta_0)$ respectively. If more equations similar to 6.8 are available, the exponential factors could be eliminated and the resulting expression would be a multivariate polynomial system of β_0 and a_k, which would hopefully be simple enough to solve. As will be shown an arbitrary number of equations very similar to 6.8 can be obtained by using different window functions.
In order to do such computations, the above equation would have to be manually manipulated, which is a very complex task even for small K. Further, the expression corresponds to a square-window. Considering a large family of *raised cosine* window functions adds even more complexity to the equation 6.8:

$$\langle s(t), \cos(l\pi t)e^{j\omega}\rangle = \frac{1}{2}\langle s(t), e^{j(\omega - l\pi)t}\rangle + \frac{1}{2}\langle s(t), e^{j(\omega + l\pi)t}e^{j\omega}\rangle. \tag{6.9}$$

Using the above expression the commonly used $Hann, Hamming$ as well as $Blackman, Nutall, Blackman - Nutall, Blackman - Harris, Flat - top$ and $Cosine/Power - of - Cosine$ windows can be expressed as a linear combination of expressions 6.9 and 6.8. The equalities

$$S_{w_{Hann}}(\omega) = \frac{1}{2}S(\omega) + \frac{1}{4}\left(S(\omega + 2\pi) + S(\omega - 2\pi)\right) \tag{6.10}$$

$$S_{w_{Hamming}}(\omega) = 0.54\, S(\omega) + 0.46\left(S(\omega + 2\pi) + S(\omega - 2\pi)\right) \tag{6.11}$$

hold for $Hann$ and $Hamming$ window respectively. Symbolic computing software like (sag; Wolfram Research) can be used to derive complete closed-form expressions and potentially simplify them yielding a simple and computationally cheap implementation.
For the purpose of this chapter a 2^{nd} degree cPACED model $(a_2t^2 + a_1t + a_0)e^{\alpha t}$ is considered. For the 4 parameters to be estimated, at least 4 equations are required. All the previously described methods utilised either signal derivatives, kernel derivatives or different kernel frequencies (see section 3.5) to deduce enough linear equations to construct a linear system with respect to the signal model parameters. Since the window function can justifiably be considered a part of the kernel, another way to produce equations would be to simply use a different window function.

6.2 Pole estimator using multiple windows

The algorithm presented in this section constructs an arbitrary number of equations containing the model parameters by using different kernels, similar to the DDM and GRM. However, rather than different kernel frequencies (Betser, 2009) or different kernel derivatives (Wen and Sandler, 2009) a different window functions are used. This may offer some advantage since the only restriction is the window functions have to be different enough not to cause numerical problems. Further, the algorithm will use analytical closed form expressions for the FT of a cPACED sinusoid which generally doesn't result in linear system of equations. It will be shown however, that only a single non-linear term is produced, which can easily be eliminated by utilising a symbolic computing software. Such algorithms suffer from very high CPU complexity - the Büchberger algorithm (Cox et al., 2007) for example, commonly used for solving multivariate polynomial systems exhibits the complexity of $(n!)^3$ where n is the number of symbolic variables. Clearly, it is crucial to keep the number of symbolic variables at a minimum. For that reason the following set of window functions yields a sufficiently simple problem, so that the closed form expression for the pole in terms of FT using the different windows can be be computed:

$$w_1(t) = c_{10} + c_{11}\cos(2\pi t/T) \tag{6.12}$$

$$w_1'(t) = -c_{11}\frac{2\pi}{T}\sin(2\pi t/T) \tag{6.13}$$

$$w_2(t) = c_{20} + c_{21}\cos(2\pi t/T) + c_{22}\cos(4\pi t/T) \tag{6.14}$$

$$w_2'(t) = -c_{21}\frac{2\pi}{T}\sin(2\pi t/T) - c_{22}2\pi\sin(4\pi t/T) \tag{6.15}$$

The above system only uses 5 parameters, since second and fourth window are derived from the first and second respectively, thus avoiding the introduction of new symbolic variables. It is crucial to understand that the analytic solution can be computed in *offline* fashion - the resulting equations are then evaluated in real-time. This way, the vast majority of computational load is completed before actual run of the algorithm.
By means of symbolic computing algorithms implementation (Wolfram Research) the terms containing $a_0, a_1, a_2, \exp(\alpha t - \omega t)$ were eliminated from 6.12, resulting in a polynomial expression $\sum_{k=0}^{6} \mu_k \beta_0^k = 0$ where coefficients

μ_k:

$$\mu_0 = \pi^6(-16(-c_{21}+c_{22})(4c_{11}c_{20}+c_{10}(-4c_{21}+c_{22}))S_{w1'} - 16c_{11}(4c_{11}c_{20}+ c_{10}(-4c_{21}+c_{22}))S_{w2'}) \tag{6.16}$$

$$\mu_1 = \pi^6(-144ic_{11}c_{20}c_{22}S_{w1} + 144ic_{10}c_{11}c_{22}S_{w2}) \tag{6.17}$$

$$\mu_2 = \pi^4T^2(-12(4c_{11}c_{20}(c_{21}+2c_{22}) - c_{10}(4c_{21}^2+c_{21}c_{22}+4c_{22}^2))S_{w1'} + 12c_{11}(4c_{11}c_{20}+ \tag{6.18}$$

$$\mu_3 = \pi^4T^2(-12ic_{11}c_{22}(5c_{20}-4c_{21}+c_{22})S_{w1} - 12ic_{11}(-5c_{10}+4c_{11})c_{22}S_{w2}) \tag{6.19}$$

$$\mu_4 = \pi^2T^4(12(c_{11}(c_{20}(c_{21}-7c_{22})+3(c_{21}-c_{22})c_{22}) - c_{10}(c_{21}^2-5c_{21}c_{22}+4c_{22}^2))S_{w1'} - 12c_{11}(-c_{10}(c_{21}+2c_{22})+c_{11}(c_{20}+3c_{22}))S_{w2'}) \tag{6.20}$$

$$\mu_5 = -36ic_{11}c_{22}\pi^2T^4((c_{20}-c_{21}+c_{22})S_{w1} + (-c_{10}+c_{11})S_{w2}) \tag{6.21}$$

$$\mu_6 = T^6((-c_{10}(c_{21}-4c_{22})+c_{11}(c_{20}-3c_{22}))(-c_{21}+4c_{22})S_{w1'} + c_{11}(-c_{10}(c_{21}-4c_{22})+ c_{11}(c_{20}-3c_{22}))S_{w2'}) \tag{6.22}$$

Since c_{kl} are predefined constants 6 estimates $\hat{\beta}_0$ can be obtained. A lower degree polynomial can obtained by setting a common constraint $w(-\frac{T}{2}) = w(\frac{T}{2}) = 0$ which in turn forces:

$$c_{10} = c_{11} \tag{6.23}$$

$$c_{21} = c_{20} + c_{22}, \tag{6.24}$$

and the equation set 6.16-6.22 is simplified to:

$$\mu_0 = 48c_{11}c_{22}\pi^6(c_{11}S_{w2'} - c_{20}S_{w1'}) \tag{6.25}$$

$$\mu_1 = 144jc_{11}c_{22}\pi^6(c_{11}S_{w2} - c_{20}S_{w1}) \tag{6.26}$$

$$\mu_2 = 36c_{11}c_{22}\pi^4T^2((3c_{22}-c_{20})S_{w1'} + c_{11}S_{w2'}) \tag{6.27}$$

$$\mu_3 = 12jc_{11}c_{22}\pi^4T^2((3c_{22}-c_{20})S_{w1} + c_{11}S_{w2}) \tag{6.28}$$

$$\mu_4 = 0 \tag{6.29}$$

$$\mu_5 = 0 \tag{6.30}$$

$$\mu_6 = 0, \tag{6.31}$$

and the polynomial is now reduced to the 3^{rd} degree:

$$\begin{aligned} jT^2((3c_{22}-c_{20})S_{w1} + c_{11}S_{w2})\beta_0^3 + \\ 3T^2((3c_{22}-c_{20})S_{w1'} + c_{11}S_{w2'})\beta_0^2 + \\ 12j\pi^2(c_{11}S_{w2} - c_{20}S_{w1})\beta_0 + \\ 4\pi^2(c_{11}S_{w2'} - c_{20}S_{w1'}) = 0. \end{aligned} \tag{6.32}$$

If $Hann^2$

$$(\frac{1}{2}(1+\cos(2\pi t/T)))^2 = \frac{1}{8}(3+4\cos(2\pi t/T)+\cos(4\pi t/T)), \tag{6.33}$$

and $Hann$ window are selected as $w_1(t), w_2(t)$ respectively the polynomial simplifies even further to:

$$j2T^2S_{w2}\beta_0^3 + 6T^2S_{w2'}\beta_0^2 + 6j\pi^2(4S_{w2} - 3S_{w1})\beta_0 + 2\pi^2(4S_{w2'} - 3S_{w1'}) = 0. \quad (6.34)$$

Above equation allows for estimation of complex poles of a cPACED sinusoid, but not the modulation polynomial coefficients. Importantly, the choice of window function is in fact fairly arbitrary. For instance, the following window function set:

$$w_1(t) = c_{10} + c_{11}\cos(2\pi t/T) + c_{12}\cos(4\pi t/T) \quad (6.35)$$
$$w_2(t) = c_{20} + c_{21}\cos(2\pi t/T) + c_{22}\cos(4\pi t/T) \quad (6.36)$$
$$w_3(t) = c_{30} + c_{31}\cos(2\pi t/T) + c_{32}\cos(4\pi t/T) \quad (6.37)$$
$$w_4(t) = c_{40} + c_{41}\cos(2\pi t/T) + c_{42}\cos(4\pi t/T) \quad (6.38)$$

can be used and a substantially more complex polynomial than 6.34 would be obtained. The computational complexity of finding the solution using the above system is substantially higher than the one of 6.12-6.15. For the window configuration in 6.12-6.15 only 5 window parameters were used whereas the function set in 6.35-6.38 uses 12, some of which might be 0 (if one of the windows is chosen to be $Hann$ for example). The solution for the above system has failed to be computed on a regular computer available at the time of writing in a reasonable time.

Generally, as far as mathematical derivations in this section are concerned, the window functions need not exhibit any of the restrictions applicable to DDM and GRM. The generally desired bell-shaped window is merely a sensible choice, so the samples close the middle of the time frame are preferred over the ones close to the boundaries. Many state-of-the-art analysis methods require the use of window derivatives as well, and by definition, at least one, the original window or its derivative, is not bell-shaped. Since such algorithms can achieve very high accuracy (see chapter 4) despite such seemingly inappropriate windows, its use will be considered in this chapter as well.

It is tempting to imagine an even simpler window function system with only 4 parameters:

$$w_1(t) = c_{10} + c_{11}\cos(2\pi t/T) \quad (6.39)$$
$$w_2(t) = w_1'(t) = -c_{11}2\pi/T\sin(2\pi t/T) \quad (6.40)$$
$$w_3(t) = c_{20} + c_{21}\cos(2\pi t/T) \quad (6.41)$$
$$w_4(t) = w_3'(t) = -c_{21}2\pi/T\sin(2\pi t/T), \quad (6.42)$$

however the functions w_2 and w_4 only differ by a factor, thus effectively removing 1 equation and rendering the system under-defined. It is now evident that the 5 parameter window function set in 6.12-6.15 if only the raised cosine window function family is considered.

6.3 Complex polynomial amplitude estimator

The expression 6.34 will be used to estimate the pole of a cPACED sinusoid. Naturally the next task at hand is the estimation of the complex amplitude polynomial coefficients. As already mentioned in (Badeau et al., 2006, 2008b) the pole estimation is essential for accurate polynomial amplitude estimation. Many LS based approaches could be employed (Stoica et al., 2000), however a simple analytical expression can also be obtained by assuming the pole estimate is very close to the actual value $\hat{\beta}_0 \approx \beta_0$:

$$\langle s, w_1\overline{e^{-\hat{\beta}_0}}\rangle \approx \langle a, w_1\rangle = a_0 T + \frac{(\pi^2-1)a_2T^2}{\pi^2} \tag{6.43}$$

$$\langle s, w_2\overline{e^{-\hat{\beta}_0}}\rangle \approx \langle a, w_2\rangle = a_1 T \tag{6.44}$$

$$\langle s, w_3\overline{e^{-\hat{\beta}_0}}\rangle \approx \langle a, w_3\rangle = \frac{1}{8}T(6a_0 + \frac{(2\pi^2-15)a_2T^2}{\pi^2}, \tag{6.45}$$

from which the following estimators can be deduced:

$$a_0 = \frac{(45-6\pi^2)\langle s, w_1\overline{e^{-\hat{\beta}_0}}\rangle + 8(\pi^2-6)\langle s, w_3\overline{e^{-\hat{\beta}_0}}\rangle}{9T} \tag{6.46}$$

$$a_1 = \frac{\langle s, w_2\overline{e^{-\hat{\beta}_0}}\rangle}{T} \tag{6.47}$$

$$a_2 = \frac{2\pi(3\langle s, w_1\overline{e^{-\hat{\beta}_0}}\rangle - 4\langle s, w_3\overline{e^{-\hat{\beta}_0}}\rangle)}{3\pi T}. \tag{6.48}$$

As already mentioned, numerous methods for estimating the coefficients of the complex polynomial amplitude exist, once the pole has been estimated. In chapter 8, a simple and straightforward method, inspired by DDM is employed. The above estimators follow the ideas presented in this chapter by using the analytical closed form expressions for FT rather than constructing a linear system of equations like DDM and GRM.

6.4 Tests and Results

The tests were performed on the 2^{nd} degree cPACED model $(a_2t^2 + a_1t + a_0)e^{\alpha+j\omega_0}$ with the $[a_2, a_1, a_0] = [p_2 + jq_2, p_1 + jq_1, p_0 + jq_0]$. The test values

were chosen so all the terms of the amplitude polynomial have equal impact on the amplitude/phase trajectory:

$$q_2, p_2 \in \left[-\left(\frac{fs}{2T}\right)^2, \left(\frac{fs}{2T}\right)^2 \right] \tag{6.49}$$

$$\alpha, q_1, p_1 \in \left[-\frac{fs}{2T}, \frac{fs}{2T} \right], \tag{6.50}$$

while $p_0 + jq_0$ was randomised for each test case so the following holds:

$$p_0 + jq_0 = e^{j\phi}, \phi \in [0, 2\pi]. \tag{6.51}$$

Essentially, the overall amplitude at $t = 0$ is forced to 1 and the phase is random. A window size of 1023 samples and sampling frequency of 44100 was used.

The SRR corresponding to CRB was computed by taking the maximum SRR achieved for a combination of specific per-parameter CRBs. Figure 6.1 depicts the overall SRR-SNR dependency averaged over the full frequency range. The algorithm achieves the quality of approximately 20dB below the CRB. Interestingly, the accuracy is not correlated to frequency, as depicted on figure 6.2. The absence of self-interference can be explained by the use of window functions with desirable TF properties. Compared to GRM or DDM, only the first window derivative is used, rather than higher order derivatives (GRM) or time-ramped windows $w(t)t^k$ (DDM). A more detailed inspection of low-frequency behaviour 6.3 indeed reveals a favourable behaviour, as much of the accuracy is retained even below 200Hz (about 4 wavelengths).

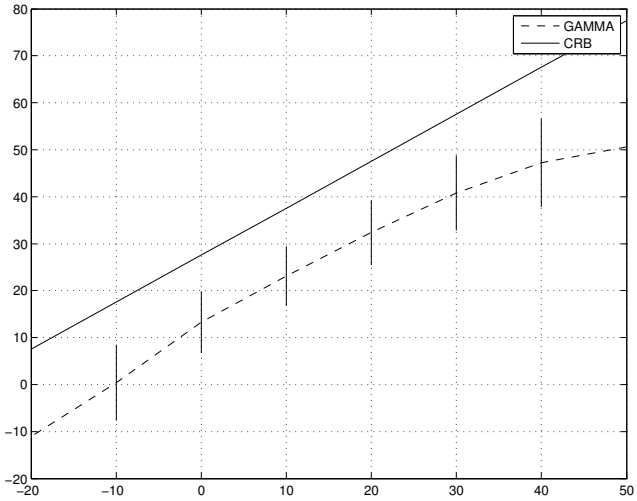


Figure 6.1: GAMMA 3^{rd} degree cPACED - full frequency range (200-2000Hz): $Hann, Hann^2, Hann', (Hann^2)'$ window set, SRR average

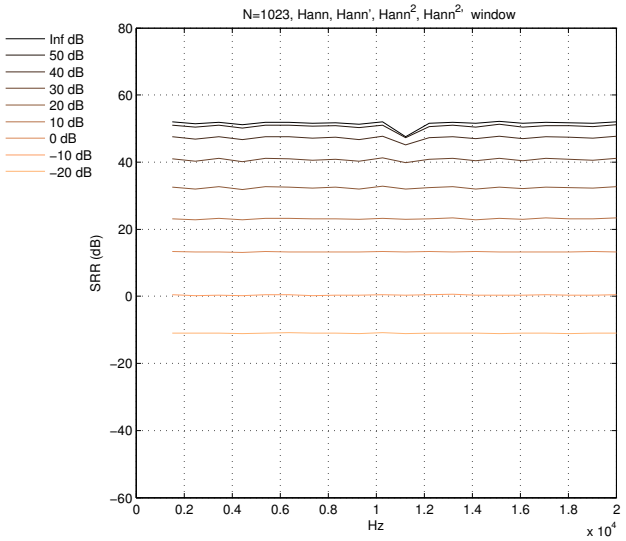


Figure 6.2: GAMMA 3^{rd} degree cPACED - full frequency range (200-2000Hz): $Hann, Hann^2, Hann', (Hann^2)'$ window set, SRR average for each tested frequency

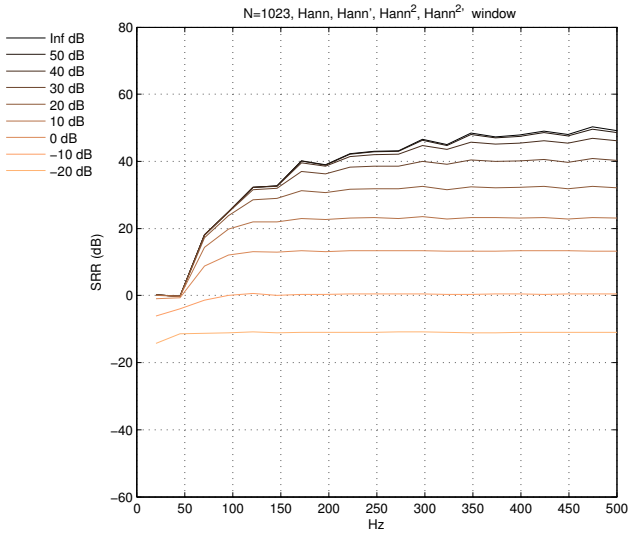


Figure 6.3: GAMMA 3^{rd} degree cPACED - low frequency range (20-500Hz): $Hann, Hann^2, Hann', (Hann^2)'$ window set, SRR average for each tested frequency

6.5 Conclusion

In this section a novel algorithm for estimation of the parameters of a cPACED sinusoid has been presented. Specifically, a 2^{nd} degree cPACED model was tested and compared to the CRB. The algorithm exhibits low bandwidth requirement resulting in relatively high accuracy in low frequency regions, while general accuracy remains approximately 20dB bellow the CRB.

Proposed method can be categorised as an energy reallocation method, as the input frequency is used in the pole estimator, from which resulting in a new frequency estimate, allowing for the spectral energy reallocation. The downside of the algorithm is the estimator expressions depend on the window function set used. It was shown that a fairly exhaustive family of raised cosine window functions yield a desirable analytic solution. However, the use of a different window function set would require a complete recalculation of the estimator expression. The procedure however should be straightforward and should follow the same general guideline as the one for the proposed system 6.12-6.15. Since the system 6.35-6.38 is too complex to be solved on an average computer at the time of writing, a similar, simpler

system might be imagined:

$$w_1(t) = \frac{1}{2} + \frac{1}{2}\cos(2\pi t/T) \tag{6.52}$$

$$w_2(t) = a_1 + \frac{1}{2}\cos(2\pi t/T) + (\frac{1}{2} - a_1)\cos(4\pi t/T) \tag{6.53}$$

$$w_3(t) = a_2 + \frac{1}{2}\cos(2\pi t/T) + (\frac{1}{2} - a_2)\cos(4\pi t/T) \tag{6.54}$$

$$w_4(t) = a_3 + \frac{1}{2}\cos(2\pi t/T) + (\frac{1}{2} - a_3)\cos(4\pi t/T). \tag{6.55}$$

The windows w_2, w_3, w_4 are clearly inspired by the Hamming window and w_1 is the Hann window. Unfortunately, utilising the exact same procedure for solving the system, all the parameters cancel out, revealing that the above system is in fact ill-conditioned, even for distinct a_1, a_2, a_3. The difficulty of using an automated problem solver is apparent, as most of the calculations are hidden from the user and thus hard to track, even harder to modify.

It is possible to imagine a modification of the proposed algorithm by replacing the window function set with the evaluation of the FT 6.8 at different frequencies. Again, the procedure of eliminating the variables reveals the system is ill-conditioned and a closed form solution thus cannot be determined.

Chapter 7

Non-stationary sinusoidal analysis using Chebyshev polynomials and Gröbner basis

In this chapter some ideas from

The complex polynomial amplitude model (cPACE) has received much attention (Pantazis et al., 2009b, 2011; Kafentzis et al., 2012) due to its simple extension to the LS-based joint estimation. It is imperative to understand that cPACE model describes FM via the imaginary part of the complex polynomial amplitude. Such procedure cannot accurately describe a linear chirp (i.e.: a generalised sinusoid) for example as the transition from cPACE to generalized sinusoid model forces the FM function to be rather complex. A high degree polynomial amplitude can approximate a linear FM, however it seems futile to estimate numerous parameters just to transform them into a single one, while clearly making a very crude approximation. In practice however it is questionable which model better describes the signal under study, the cPACE or generalised sinusoid? The cPACE model has been successfully used in speech/voice analysis (Pantazis et al., 2011) therefore no clear answer can be expected. Electronic sound sources might produce more *mathematical* modulations suggesting the generalized sinusoid could be a better fit, but no such claims can be made for voice/speech for example.
Derivations that follow essentially extend the cPACE degree much higher than in (Pantazis et al., 2009b; Badeau et al., 2006). A transition from high-degree cPACE to generalised sinusoid model is performed via analyt-

ical solutions of a multivariate polynomial system, arising from truncated Taylor series. As a consequence a pair of stationary sinusoids with very close frequencies (below 1 FFT bin) can be resolve to some extent. An important notion is that many non-linear polynomial systems are in fact easily solvable via Gröbner basis (using the Buchberger algorithm (Cox et al., 2007)). Severe CPU complexity of such symbolic computing algorithms can be avoided by computing analytical solutions in an off-line manner. These solutions come in form of univariate polynomials (Cox et al., 2007) and are therefore easily solvable in real-time.

For purpose of this section a complex sinusoid is defined as in (Wen and Sandler, 2009):

$$s(t) = e^{R(t)}, R(t) = \sum_{m=0}^{M-1} p_m \alpha_m(t) + j \sum_{m=0}^{M-1} q_m \psi_m(t), \tag{7.1}$$

where $R(t)$ is a complex function, a linear combination of 2M real functions $\alpha_m(t), \psi_m(t)$, weighted with real parameters p_m, q_m respectively. The function pairs $\alpha_m(t)$ and $\psi_m(t)$ can theoretically be different, however the most straight-forward choice for both is monomials: $\alpha_m(t) = \psi_m(t) = t^m$. Other choices can be motivated by specific *a priori* knowledge about the signal under study. For example, analysing *vibrato* sounds (e.g.: assumed sinusoidal modulation with frequency ω_m), $\psi_0(t) = 1, \psi_1(t) = \cos(\omega_m t), \psi_2(t) = \sin(\omega_m t)$ was proposed in (Wen and Sandler, 2009). In the monomial case, (e.g.: $\alpha_m(t) = \psi_m(t) = t^m$), p_0 corresponds to the stationary log-amplitude while $p_i, i > 0$ corresponds to the i-th order log-amplitude modulation. Analogously, q_0 corresponds to the stationary phase, q_1 to the stationary frequency and parameters $q_i, i > 1$ to the $(i-1)$-th degree frequency modulation. However the only necessary condition to separate this representation into 2 factors (e.g. a stationary and non-stationary one) is given by the generally agreed-on definition of a stationary complex sinusoid $e^{p_0 + j(q_0 + q_1 t)}$ which follows: $\alpha_0(t) = \psi_0(t) = 1$ and $\psi_1(t) = t$. In such case the following separation can be obtained:

$$s(t) = s_s(t) s_{ns}(t) = e^{p_0 + j(q_0 + q_1 t)} e^{R_{ns}(t)},$$
$$R_{ns}(t) = \sum_{m=2}^{M-1} (j q_m \psi_m(t) + p_m \alpha_m(t)) + p_1 \alpha_1(t), \tag{7.2}$$

where s_s, s_{ns} are the stationary and non-stationary factors respectively. The crucial observation is that the signal s_{ns} is *much less* oscillatory compared to s_s. An example of a decomposition of a non-stationary sinusoid is

depicted in Fig. 7.1. Most of the oscillation is captured by the stationary factor leaving the non-stationary part easy to approximate with a polynomial basis. It is expected the non-stationary parameters can be deduced from the coefficients of the approximated polynomial. The factor s_{ns} is of

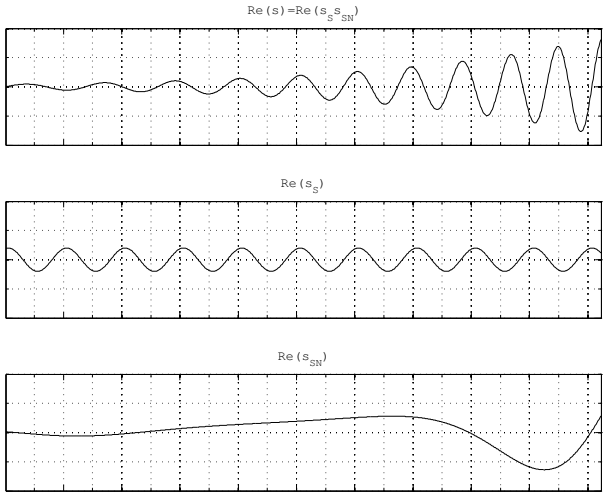


Figure 7.1: Above: product ($s = s_{sn}s_s$). Middle: stationary factor (s_s). Below: non-stationary factor (s_{sn}).

the form $e^{f(t)}$ and the time derivatives needed for Taylor series expansion can be computed with the following recursive formula:

$$\left(e^{f(t)}\right)^{(k)} = \left(f'(t)e^{f(t)}\right)^{(k-1)}. \tag{7.3}$$

A convenient choice of the expansion point $t = 0$ yields the following Taylor

series for $e^{f(t)}$:

$$\begin{aligned} & e^{f(0)} \\ & +e^{f(0)}f'(0)x \\ & +e^{f(0)}(f'(0)^2+f''(0))\frac{x^2}{2!} \\ & +e^{f(0)}(f'(0)^3+3f'(0)f''(0)+f^{(3)}(0))\frac{x^3}{3!} \\ & +e^{f(0)}(f'(0)^4+f^{(4)}(0)+3f''(0)^2+ \\ & \quad 4f^{(3)}(0)f'(0)+6f'(0)^2f''(0))\frac{t^4}{4!} \\ & \quad \vdots \end{aligned} \tag{7.4}$$

Let's now assume $\alpha_m(t) = \psi_m(t) = h_m(t) = t^m$. This assumption vastly simplifies the expansion in equation 7.4 as most of the terms simplify to a complex constant:

$$\begin{aligned} f(0) &= 0, \\ f'(0) &= p_1, \\ f''(0) &= 2!(p_2+jq_2), \\ f^{(3)}(0) &= 3!(p_3+jq_3), \\ &\vdots \\ f^{(l)}(0) &= l!(p_l+jq_l). \end{aligned} \tag{7.5}$$

For sake of generality and compactness the equation 7.4 is rewritten as:

$$e^{f(x)} = e^{f(0)}\sum_{l=0}^{l=\infty}\gamma[l]\frac{x^l}{l!}, \tag{7.6}$$

where $\gamma[l]$ stand for the coefficients composed of parameters of the function f. If the modulation functions are indeed polynomials (eg: monomials), than the coefficients $\gamma[l]$ are multivariate polynomials of the parameters p_k, q_k. If the $\gamma[l]$ can be estimated from signal then a system of multivariate polynomials can be constructed and the parameter estimation problem is solved by solving this system.

7.1 Discrete Fourier-Chebyshev basis

This section constructs a function set necessary for accurate estimation of the coefficients $\gamma[l]$ from the signal observations. The signal under study will be conveniently defined by combining equation 7.7 and 7.6:

$$s(t) = e^{p_0+jq_0}e^{jq_1t}\sum_{l=0}^{l=\infty}\gamma[l]\frac{t^l}{l!}. \tag{7.7}$$

Common definitions of the *dot* product in the space of continuous functions and its analogous definition in discrete time are assumed:

$$\langle p, q\rangle = \int_{-\infty}^{\infty} p(t)\bar{q}(t)dt \tag{7.8}$$

$$\langle p, q\rangle = \sum_{l=-\infty}^{\infty} p[l]\bar{q}[l]. \tag{7.9}$$

A peak in Fourier magnitude spectrum gives a good approximation $\hat{q}_1$ of the frequency parameter q_1. With definition of Fourier kernel $\Psi_\omega(t) = e^{j\omega t}$ the following approximation can be obtained:

$$\langle s, \Psi_\omega\rangle \approx e^{p_0+jq_0}\sum_{l=0}^{\infty}\frac{\gamma[l]}{l!}\langle e^{jq_1t}t^l, \Psi_{\hat{q}_1}\rangle. \tag{7.10}$$

It can now be recognized the Fourier basis must be extended to match the terms for $l > 0$. It is crucial the extended function set remains orthogonal. For a fixed frequency ω_0 the following equality gives a good clue:

$$\langle \Psi_{\omega_0}t^n, \Psi_{\omega_0}t^l\rangle = \int_{-\infty}^{\infty}\Psi_{\omega_0}\bar{\Psi}_{\omega_0}t^{n+l}dt = \langle t^n, t^l\rangle. \tag{7.11}$$

Orthogonalizing the function set $\Psi_{\omega_0}t^n$ is indeed identical to that of orthogonalizing monomials. There are numerous polynomial orthonormal basis (Legendre, Chebyshev, Laguerre, Jacobi, Hermite...) of which the Legendre polynomials (Bateman Manuscript Project et al., 2006b) are the most straightforward as they can be derived from from monomials via the Gram-Schmidt orthogonalization process. Of all the mentioned polynomial basis only Chebyshev (Bateman Manuscript Project et al., 2006a) and Legendre are orthogonal with the respect to the inner product defined in equation 7.8 and are therefore the only candidates. A little less obvious issue is this polynomial sets are all defined in continuous time and do not

assure orthogonality when discrete version of the inner product (equation 7.9) is used. Using original polynomials simply sampled at equidistantly spaced times causes unacceptable errors in practice due to *reduced* orthogonality (Mukundan et al., 2001). The solution however is straightforward as discrete time versions of Legendre/Chebyshev polynomials were discovered (Morrison, 1969)(Clenshaw and Mühlig, 1963) as well as numerically stable methods for its computation (Aburdene and Dorband, 1996)(Mukundan, 2004) extensively used in image coding applications (Liao and Pawlak, 1996)(Mukundan et al., 2000). A recursive construction over degree of the polynomials and its time variable (Mukundan, 2004) offers sufficient stability even for higher order polynomials. To author's knowledge such construction is only known for discrete Chebyshev polynomials. A report on superior image coding ability of the Chebyshev over Legendre basis (Mukundan et al., 2000), followed by the analysis of the condition number of the alternant matrix (Aitken, 1956) similar to that in (Li and Wen, 2010) favoured recursive construction of Chebyshev polynomials on degree and time variable over sampled continuous Chebyshev and Legendre polynomials, due to superior numerical stability.

Alternant matrix condition number analysis

There has been a report on superior image coding ability of the Chebyshev polynomials over the Legendre ones (Mukundan et al., 2000). However images are of entirely different nature as sound signals therefore a further investigation is required. The analysis of the condition number of the alternant matrix similar to that in (Li and Wen, 2010) was conducted. The alternant matrix of a function set $f_l, l = 1..n$ is defined as in (Aitken, 1956):

$$A = \begin{bmatrix} f_1(t_1) & f_2(t_1) & \dots & f_n(t_1) \\ f_1(t_2) & f_2(t_2) & \dots & f_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(t_m) & f_2(t_2) & \dots & f_n(t_m) \end{bmatrix}. \tag{7.12}$$

The condition number of an alternant matrix A gives a bound on inaccuracy of the solution x of a system $Ax = b$. The condition number is defined as follows (Cheney and Kincaid, 2007)

$$\kappa(A) = \|A\|\|A^{-1}\|. \tag{7.13}$$

An extended comparison is depicted in figure 7.2. To compare relevant function sets the following were tested:

- time sampled continuous Legendre polynomials (with recursion on degree)
- discrete Legendre polynomials
  - direct formula
  - with recursion on degree
- time sampled continuous Chebyshev polynomials
- discrete Chebyshev polynomials
  - with recursion on degree
  - with recursion on degree and time variable
- monomials
- discrete Fourier basis

The most readable and compact form of defining polynomial sets is by a recursive construction formula. Explicit formulas are complex and offer little insight, and will thus be omitted if recursive formula is known. Continuous Legendre polynomials can be defined with recursive formula on its degree (Bateman Manuscript Project et al., 2006a):

$$(n+1)l_{n+1}(t) = (2n+1)tl_n(t) - nl_{n-1}(t), \tag{7.14}$$
$$l_0(t) = 1, \tag{7.15}$$
$$l_1(t) = t, \tag{7.16}$$

Explicit formula for its discrete version (Morrison, 1969):

$$L_n[k] = \sum_{l=0}^{n} (-1)^l \binom{n}{l} \binom{n+l}{l} \frac{k^{(l)}}{N^{(l)}}, \tag{7.17}$$

where $k^{(l)}$ is the *backward factorial* defined as:

$$x^{(l)} = x(x-1)(x-2)...(x-l+1). \tag{7.18}$$

A recursive formula is also known (Neuman and Schonbach, 1974):

$$(n+1)(N-n)L_{n+1}[k] = (2n+1)(N-2k)L_n[k] - n(n+N+1)L_{n-1}[k] \tag{7.19}$$
$$L_0[k] = 1, \tag{7.20}$$
$$L_1[k] = \frac{(N-2k)}{N} \tag{7.21}$$

Similar definitions hold for continuous Chebyshev polynomials (Bateman Manuscript Project et al., 2006a) :

$$c_{n+1}(t) = 2tc_n(t) - c_{n-1}(t), \tag{7.22}$$
$$c_0(t) = 1, \tag{7.23}$$
$$c_1(t) = t. \tag{7.24}$$

A recursive formula is known for its discrete analogy (Clenshaw and Mühlig, 1963):

$$C_n[k] = \frac{1}{n}(2n-1)C_1[k]C_{n-1}[k] - \frac{1}{n}(n-1)\left(1 - \frac{(n-1)^2}{N^2}\right)C_{n-2}[k], \tag{7.25}$$
$$C_0[k] = 1, \tag{7.26}$$
$$C_1[k] = \frac{2k+1-N}{N}. \tag{7.27}$$

Estimation of the discrete Legendre polynomials was constructed using the following equation (Aburdene and Dorband, 1996):

$$P = \left(I^-((B'B)\cdot B)'\right)\left(Diag(I'B'B)\right)^{-1}B, \tag{7.28}$$
$$[B]_{i,j} = \binom{i}{j}, \tag{7.29}$$
$$[I^-]_{i,i} = (-1)^i, \tag{7.30}$$
$$i = 0..N-1, \tag{7.31}$$
$$j = 0..N-1, \tag{7.32}$$
$$A\cdot B = C \Rightarrow [C]_{i,j} = [A]_{i,j}[B]_{i,j}. \tag{7.33}$$

The n-th row of the matrix C contains the values of the discrete Legendre polynomial of n-th degree. Above procedure is an implementation of evaluation of the explicit formula 7.17 and is not considered to be numerically stable for degrees above 8 ((Press, 1992) chapter *6.8: Spherical Harmonics*).
The most numerically stable construction of discrete Chebyshev polynomials can be achieved by using the recursion on the degree only for the $k = 0$ and the recursion on the time variable for $k > 1$ using the recursion on the time variable. Including normalization terms the recursions are as follows

(Mukundan, 2004):

$$\widehat{C}_n[1] = \left(1 + \frac{n(1+n)}{1-N}\right)\widehat{C}_n[0] \tag{7.34}$$

$$\widehat{C}_n[k] = \beta_1\widehat{C}_n[k-1] + \beta_2\widehat{C}_n[k-2], \tag{7.35}$$

$$n = 1, 2, ..., N-1; \quad k = 2, 3, ..., \frac{N}{2} \tag{7.36}$$

where

$$\beta_1 = \frac{-n(n-1) - (2k-1)(k-N-1) - k}{k(N-k)} \tag{7.37}$$

$$\beta_2 = \frac{(k-1)(k-N-1)}{k(N-k)}. \tag{7.38}$$

The recursion can be terminated at $N/2$ as the symmetry can be used to calculate the rest:

$$\widehat{C}_n[N-1-k] = (-1)^n\widehat{C}_n[k]. \tag{7.39}$$

It is apparent from the figure 7.2 that discrete Chebyshev polynomial basis constructed by the recursion on the degree for $k = 0$ and the recursion on the time variable for $k > 0$ is numerically the most stable choice. Surprisingly it seems to be marginally more stable than the discrete Fourier basis, however there might exist a better construction of Fourier basis providing even higher numerical stability. Please note that discrete Legendre polynomial does not achieve the same performance when compared to (Li and Wen, 2010). To authors knowledge the results presented here are correct. In (Li and Wen, 2010) much better results for the discrete Legendre polynomials were reported, however the authors were unavailable for comments and could not provide the Matlab code mentioned to achieve reported results. In any case, the discrete Chebyshev polynomials achieve even lower condition number that the one reported in (Li and Wen, 2010) for the discrete Legendre polynomials.

The signal model in equation 7.1 is defined in continuous time, but the discrete Chebyshev basis can be thought of as set of continuous polynomials, sampled at equidistantly spaced times. Note that these are not equal to continuous Chebyshev polynomials and their coefficients change with the number of the time samples , eg: length of observation frame and degree. It is thus safe to conclude that discrete Chebyshev decomposition can be matched to the model defined in continuous time which concludes the justification of its use.

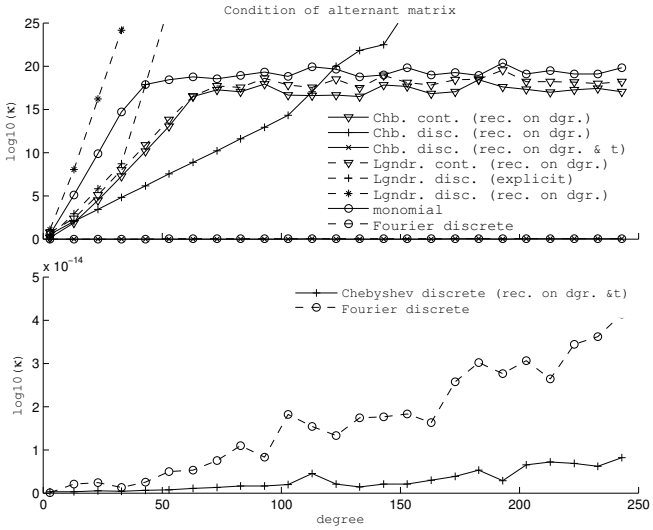


Figure 7.2: Comparison of the condition number an alternant matrix derived from different functions sets. Above: all function sets. Below: a zoom-in with only discrete Fourier and discrete Chebyshev polynomials constructed by the recursion on the degree for $k = 0$ and the recursion on the time variable for $k > 0$.

Decomposition of non-stationary sinusoid with discrete Fourier-Chebyshev functions

For the purpose of decomposition motivated *discrete Fourier-Chebyshev function of order* l can be defined :

$$\Upsilon_l^w[n] = e^{j2\pi\frac{n}{N}w} C_l[n], w = 0...N-1, n = 0...N-1, \tag{7.40}$$

where $C_l[n]$ is the discrete Chebyshev polynomial of order l, w is the frequency bin index, n the sample index and N the buffer length in samples. It is now trivial to see that $< \Upsilon_{l_1}^w, \Upsilon_{l_2}^w >= \delta_{l_1 l_2}$, where $\delta_{l_1 l_2}$ is the Kronecker delta function and consequently the set $\Upsilon_l^{w_0}, l = 0...N-1$ is an orthogonal basis for some fixed w_0 and normalizing the C_l yields an orthonormal basis. A non-stationary sinusoid can therefore be represented with the following decomposition once the nearest frequency bin $\hat{q}_1$ is identified:

$$\mu_{\hat{q}_1}[l] =< s, \Upsilon_l^{\hat{q}_1} >, l => 0. \tag{7.41}$$

An example of a non-stationary sinusoid similar to the one in Fig. 7.1 is considered. The modulation functions are monomials of 2^{nd} degree for the

phase parameters and 1^{st} degree for the log-amplitude. The specific parameter values are 1^{st} order (linear) log-amplitude change rate $150/s$, static log-amplitude 0, 1^{st} order (linear) frequency rate $5000Hz/s$, frequency $440Hz$, phase 0π, sample rate 44100 and buffer length 511 samples. Parameter values can be written compactly as the log-amplitude/phase parameter vectors respectively:

$$\vec{p} = [0, 230] \tag{7.42}$$

$$\vec{q} = [0.35\pi, 440, 10000]. \tag{7.43}$$

Fig. 7.3 shows consecutive approximations of the sinusoid with the Υ_l^w with degree $l = 0, 1, 2, 3$. Analysis of the L_1 norm of the representation

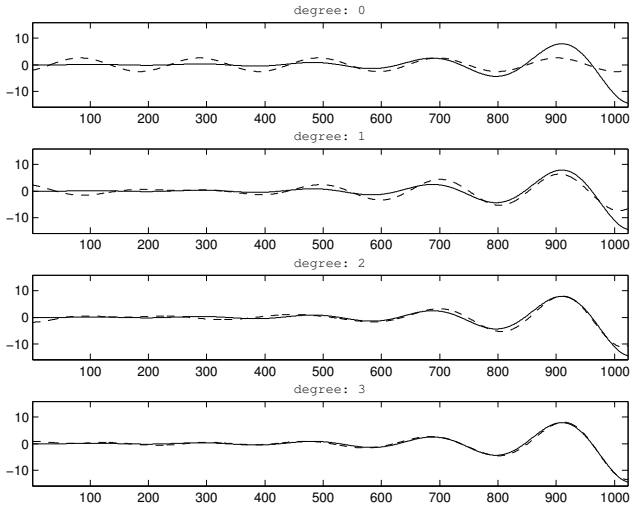


Figure 7.3: Approximation of the non-stationary sinusoid defined by 7.42 and 7.43 with Fourier-Chebyshev functions of ascending degree.

vector $\mu_w[l]$ in respect to the degree l reveals how non-stationarity of the sinusoid is encoded in the transform domain. From figure 7.4 it is easy to conclude that there is no information included in coefficients higher than about 25 as it reaches a plateau governed by noise. The plot in linear scale suggest even more compressed representation as coefficients reach near zero values at a degree below 10. More extreme modulations naturally require more coefficients however modulations present in the example from Fig. 7.3 and 7.4 are above the limits of standard tests in context of non-stationary sinusoidal analysis (Marchand and Depalle, 2008) and correspond to modulations found in real audio signals.

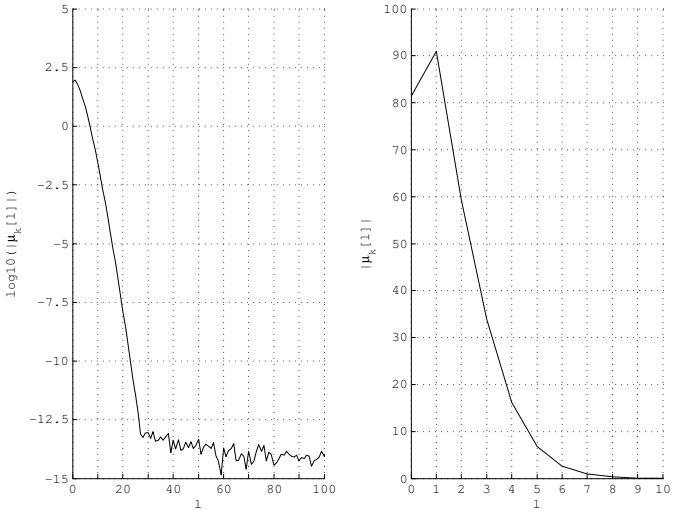


Figure 7.4: L_1 norm of the signal in the transform domain. Left: log scale, first 100 coefficients. Right: linear scale, first 10 coefficients.

7.2 System of multivariate polynomials

Combining the equations of Taylor expansion of non-stationary part 7.6 with rest of the signal yields:

$$s(t) = e^{p_0 + j(q_0 + q_1 t)} \sum_{l=0}^{l=\infty} \gamma[l] \frac{t^l}{l!}, \tag{7.44}$$

where $\gamma[l]$ is multivariate polynomial of the non-stationary parameters $p_1, p_2, ..., q_2, q_3, ...$ and can be expressed via equations 7.4 and 7.5. Estimation of $\gamma[l]$ can be achieved using the discrete Fourier-Chebyshev basis decomposition described in section 7.1. The discrete Chebyshev polynomials can be though of as a set of continuous polynomials sampled at equidistant times. It is therefore possible to represent each discrete Chebyshev polynomial as a sum of monomials and thus transform discrete Chebyshev moments into discrete monomial moments. Such transition depends on the buffer length N and can be performed with a matrix multiplication:

$$\mu_w T_N = \gamma_w, \tag{7.45}$$

where the T_N matrix is composed of coefficients of the discrete Chebyshev polynomials of degree/length N. Each row of the matrix contains coefficients of one polynomial, each column represents coefficient of a certain

degree. Therefore an entry at row n, column m corresponds to a $m-order$ coefficient of the discrete Chebyshev polynomial of $n-th$ degree.
To derive a simple multivariate system, the equation 7.44 can be rewritten using the frequency q_1:

$$s(t) = e^{p_0+j(q_0+q_1 t)} \sum_{l=0}^{l=\infty} \gamma_{q_1}[l] \frac{t^l}{l!}. \tag{7.46}$$

Inserting the terms from 7.5 (which assumes polynomial modulation functions for log-amplitude and phase) into equation 7.4, matching corresponding coefficients with 7.6, the following equalities are obtained:

$$\begin{aligned} e^{p_0+jq_0} &= \gamma_{q_1}[0] \\ e^{p_0+jq_0} p_1 &= \gamma_{q_1}[1] \\ e^{p_0+jq_0}(p_1^2 + 2jq_2) &= \gamma_{q_1}[2] \\ e^{p_0+jq_0}(p_1^3 + 6jp_1q_2) &= \gamma_{q_1}[3] \\ e^{p_0+jq_0}(p_1^4 + 12jp_1^2q_2 - 12q_2^2) &= \gamma_{q_1}[4], \\ &\vdots \end{aligned} \tag{7.47}$$

which is recognized as a system of multivariate polynomials of 4 parameters (p_0, p_1, q_0, q_2). The frequency parameter q_1 does not appear in the system as it was *cancelled out*. In practice however, an estimate of the frequency rather than the exact value is available yielding a slightly different system (see subsection 7.3). The solutions of a multivariate polynomial system 7.47 are estimates of the sinusoidal parameters.

Gröbner basis and Buchberger algorithm

This subsection briefly outlines the most important aspects of the Gröbner basis and Buchberger algorithm (Cox et al., 2007). For the sake of brevity many concepts are simplified, and sometimes renamed to avoid lengthy definitions, proofs and theorems.
A specific multiplication/addition sequence, called the *Buchberger algorithm*, can be applied to each multivariate system, in order to transform it into a special set of polynomials, the so called *Gröbner basis*. The main idea of this iterative process is to gradually eliminate variables from the system by iteratively combining 2 polynomials and cancel out certain terms.
A crucial requirement of this process is an ability to *order* multivariate monomials, the so called *monomial ordering* must be defined. The *degree*

ordering of terms (denoted by $\prec$) of an univariate monomial arises naturally: $t \prec t^2 \prec t^3 \prec t^4$. However, in case of a multivariate monomial, additional information is required. For example, the two terms x^2y, xy^2 cannot be *degree ordered*. A *lexicographic ordering* requires the parameters to be ordered in a sequence, also denoted by $\prec$. Assuming a parameter ordering $x \prec y$, it is possible to order the two terms in question:

$$x^2y \prec xy^2, \tag{7.48}$$

or more generally:

$$x^{k_1}y^{l_1} \prec x^{k_2}y^{l_2} \Leftrightarrow k_1 > k_2 \text{ or } (l_1 > l_2, \text{ if } k_1 = k_2). \tag{7.49}$$

Since $x \prec y$, the power of x *dominates* the power of y. Above *lexicographic ordering* rule can be easily extended to an arbitrary number of parameters. The most important property of the Gröbner basis is that the first polynomial is an univariate polynomial of the first parameter (according to the defined parameter ordering), from which solutions for it can be directly computed. The second polynomial contains only the first two parameters and solutions for the first one can be directly inserted since they were computed in the previous step, making the second polynomial univariate as well. Generally, a chain of univariate polynomials can be obtained by substituting solutions from previous polynomials into the next one (Buchberger, 1976). This unique property of the Gröbner basis, computed with the Buchberger algorithm using lexicographic ordering generally allows computation of all existing solutions to a complex valued system - the existence and finiteness of the Gröbner basis is guaranteed (Buchberger, 1976). All the computations presented in the following subsection were done with the Buchberger implementation from *SAGE open-source mathematics software* (sag). As an example we approximate an unknown non-stationary sinusoid with polynomials of increasing degrees. The roots of the polynomial without the imaginary component will closely match zeros of the signal as degree is increased. Some number of complex roots, which do not correspond to the actual zeros of the underlying function might occur. In figure 7.5 the lower degree approximations exhibit only real roots (top and middle), and higher degree approximation (below) produces conjugate pairs of complex root (marked with '$*$') which do not correspond to an any of the zeros of the function. According to this example it is important to conclude that as an effect of approximation some solutions must be identified as spurious. If the model assumes real solutions the complex solutions must be discarded. If complex solutions are allowed, as is often the case, there is no way to distinguish

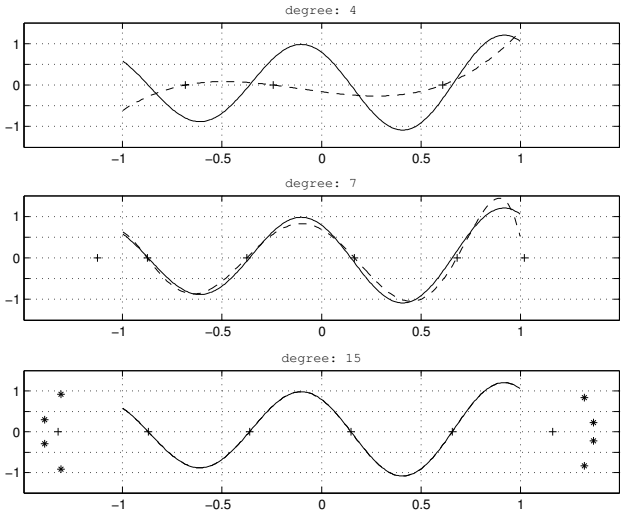


Figure 7.5: Approximation of a function by polynomials with polynomials. From such polynomial approximation The signal and its approximations are plotted on the x-y plane and the same plane is used for a real-complex plots for the roots. This way the real roots will have no y-axis component and will closely match the zeros of the function. The real roots are marked as '+' and the complex ones with '∗'.

acceptable solutions from spurious ones. It will be shown however that certain polynomials from Gröbner basis always give the *right* solutions and the others the spurious ones. A more detailed explanation of the theory of Gröbner basis and Buchberger algorithm is out of scope of this document: a discussion on spurious complex roots and a short,condensed introduction in the topic of computational commutative algebra can be found in (Cox et al., 2007).

7.3 Solutions to various signal models

In this section the Gröbner basis of systems defined by various signal models will be used to estimate the parameters of a non-stationary sinusoid. The steps of the Buchberger algorithm will be omitted for the sake of compactness. In all cases, the peak location in magnitude spectrum $\hat{q}_1$ was used as a initial estimation of frequency. To get a better estimate of frequency, a slightly different model separation to stationary - non-stationary parts than

in 7.2 will be used:

$$s(t) = s_s(t)s_{ns}(t) = e^{p_0+j(q_0+\hat{q}_1 t)}e^{R_{ns}(t)},$$
$$R_{ns}(t) = p_1 t + j\overbrace{(q_1-\hat{q}_1)}^{\Delta q_1} t + \sum_{m=2}^{M-1}(jq_m + p_m)t^m. \quad (7.50)$$

This way the frequency offset Δq_1 can be estimated in the same way as the rest of the non-stationary parameters.

Linear FM, first order AM complex sinusoid

The first model considered is the most common one in the context of non-stationary sinusoidal modelling:

$$\begin{aligned} s(t) =& e^{p_0+p_1 t+j(q_0+q_1 t+q_2 t^2)} \\ =& e^{p_0+j(q_0+\hat{q}_1 t)}e^{p_1 t+j(\Delta q_1 t+q_2 t^2)} \\ =& e^{j\hat{q}_1 t}e^{r_0+r_1 t+r_2 t^2}, \\ r_0 =& p_0 + jq_0, \\ r_1 =& p_1 + j\Delta q_1, \\ r_2 =& jq_2, \end{aligned} \quad (7.51)$$

The corresponding multivariate system:

$$e^{r_0} - \gamma_{\hat{q}_1}[0] = 0 \quad (7.52)$$
$$e^{r_0}r_1 - \gamma_{\hat{q}_1}[1] = 0 \quad (7.53)$$
$$e^{r_0}\left(r_1^2 + 2r_2\right) - 2\gamma_{\hat{q}_1}[2] = 0 \quad (7.54)$$
$$e^{r_0}\left(r_1^3 + 6r_1 r_2\right) - 6\gamma_{\hat{q}_1}[3] = 0 \quad (7.55)$$
$$e^{r_0}\left(r_1^4 + 2r_1^2 r_2 + 2r_2^2\right) - 24\gamma_{\hat{q}_1}[4] = 0 \quad (7.56)$$
$$\vdots$$

The set of univariate polynomials that give direct means of solving above system depend on the degree and can be obtained using the Buchberger algorithm. The minimum number of polynomials in the original set must naturally be equal or greater to the number of unknown variables, in this particular case 3. The Buchberger algorithm returns the following solutions

for r_2, r_1 for system consisting of equations 7.52,7.53 and 7.54:

$$2r_2(\gamma_{\hat{q}_1}[0])^2 + (\gamma_{\hat{q}_1}[1])^2 - 2\gamma_{\hat{q}_1}[0]\gamma_{\hat{q}_1}[2] = 0 \tag{7.57}$$

$$r_1\gamma_{\hat{q}_1}[0] - \gamma_{\hat{q}_1}[1] = 0. \tag{7.58}$$

Note that in this simple example, the solution equations can be obtained by a few algebraic operations on the system, eg: the Buchberger algorithm can be performed by hand. As the modulation degree increases, human capacity of solving the system is quickly surpassed. Note that above equations also hold for the case of $r_2 = p_2 + jq_2$, e.g.: log-AM and FM are polynomials of the second order. The degree of the model can be arbitrarily increased and

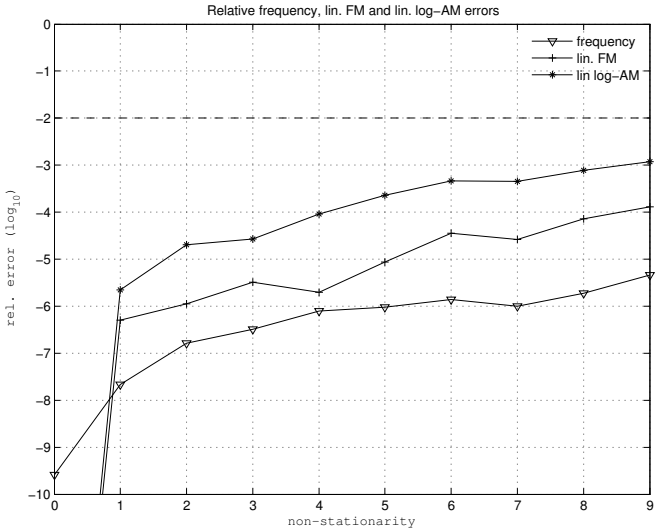


Figure 7.6: Relative error of the estimators with respect to *non-stationarity*. The FM/log-AM is jointly increased with higher values on the x-axis. The dashed line corresponds to 1% accuracy.

the Gröbner basis will provide the solutions for higher order parameters as well. For instance, the relevant univariate polynomial for 3^{rd} degree log-AM/FM parameter r_3:

$$\begin{aligned} &3r_3(\gamma_{\hat{q}_1}[0])^3 - (\gamma_{\hat{q}_1}[1])^3 + \\ &\quad +3\gamma_{\hat{q}_1}[0]\gamma_{\hat{q}_1}[1]\gamma_{\hat{q}_1}[2] - 3(\gamma_{\hat{q}_1}[0])^2\gamma_{\hat{q}_1}[3] = 0. \end{aligned} \tag{7.59}$$

Deriving above solution by hand would be time consuming and error prone procedure, whereas the Buchberger algorithm derives it in fraction of a second.

Two complex sinusoids, no FM, first order AM

The following *twin sinusoid* model is considered:

$$\begin{aligned} s(t) =& e^{p_{10}+p_{11}t+j(q_{10}+q_{11}t)} + e^{p_{20}+p_{21}t+j(q_{20}+q_{21}t)} \\ =& e^{p_{10}+j(q_{10}+\hat{q}t)}e^{p_{11}t+j\Delta q_{11}t} + e^{p_{20}+j(q_{20}+\hat{q}t)}e^{p_{21}t+j\Delta q_{21}t} \\ =& e^{j\hat{q}t}(e^{r_{10}+r_{11}t} + e^{r_{20}+r_{21}t}), \end{aligned} \tag{7.60}$$

$$\begin{aligned} r_{10} =& p_{10} + jq_{10}, \\ r_{11} =& p_{11} + j\Delta q_{11}, \\ r_{20} =& p_{20} + jq_{20}, \\ r_{21} =& p_{21} + j\Delta q_{21}, \\ \Delta q_{11} =& q_{11} - \hat{q}, \\ \Delta q_{21} =& q_{21} - \hat{q}, \end{aligned}$$

where $\hat{q}$ is an estimate of frequency close to both q_{11}, q_{21} preferably between these values. The resulting multivariate system:

$$e^{r_{10}} + e^{r_{20}} - \gamma_{\hat{q}}[0] = 0 \tag{7.61}$$
$$e^{r_{10}}r_{11} + e^{r_{20}}r_{21} - \gamma_{\hat{q}}[1] = 0 \tag{7.62}$$
$$e^{r_{10}}r_{11}^2 + e^{r_{20}}r_{21}^2 - 2\gamma_{\hat{q}}[2] = 0 \tag{7.63}$$
$$e^{r_{10}}r_{11}^3 + e^{r_{20}}r_{21}^3 - 6\gamma_{\hat{q}}[3] = 0. \tag{7.64}$$

$$\vdots$$

The quadratic equation from its Gröbner basis, contaning only parameter r_{21}:

$$r_{21}^2 a_2 + r_{21}a_1 + a_0 = 0, \tag{7.65}$$

where

$$\begin{aligned} a_2 =& \gamma_{\hat{q}}[1]^2 - 2\gamma_{\hat{q}}[0]\gamma_{\hat{q}}[2] \\ a_1 =& 6\gamma_{\hat{q}}[0]\gamma_{\hat{q}}[3] - 2\gamma_{\hat{q}}[1]\gamma_{\hat{q}}[2] \\ a_0 =& 4\gamma_{\hat{q}}[2]^2 - 6\gamma_{\hat{q}}[1]\gamma_{\hat{q}}[3]. \end{aligned} \tag{7.66}$$

A closer look at the equations 7.61-7.64 reveals the following *symmetry*: if the parameters $r_{11} \leftrightarrow r_{21}$ and $r_{10} \leftrightarrow r_{20}$ are swapped, the system remains the same: the system is *invariant* to the mentioned variable swap. By swapping $r_{11} \leftrightarrow r_{21}$ in equation 7.65, the quadratic equation for r_{11} can be derived. The solutions of r_{11} are therefore identical to those of r_{21}, which

appears to be counter intuitive. In practice however it has been observed, that one of the two solutions to equation 7.65 corresponds to r_{21} and another to r_{11}. Similar holds for the r_{10},r_{20} pair, thus it is reasonable to rewrite 7.65 by swapping r_{21} with a new variable r_1, leaving out the notion whether the parameter belongs to the first or the second sinusoid:

$$\begin{aligned} r_1^2 a_2 + r_1 a_1 + a_0 &= 0 \\ a_2 &= \gamma_{\hat{q}}[1]^2 - 2\gamma_{\hat{q}}[0]\gamma_{\hat{q}}[2] \\ a_1 &= 6\gamma_{\hat{q}}[0]\gamma_{\hat{q}}[3] - 2\gamma_{\hat{q}}[1]\gamma_{\hat{q}}[2] \\ a_0 &= 4\gamma_{\hat{q}}[2]^2 - 6\gamma_{\hat{q}}[1]\gamma_{\hat{q}}[3]. \end{aligned} \tag{7.67}$$

The example in Fig. 7.7 shows frequency error for two complex stationary sinusoids with frequencies close to 1760Hz. In practice, the magnitude

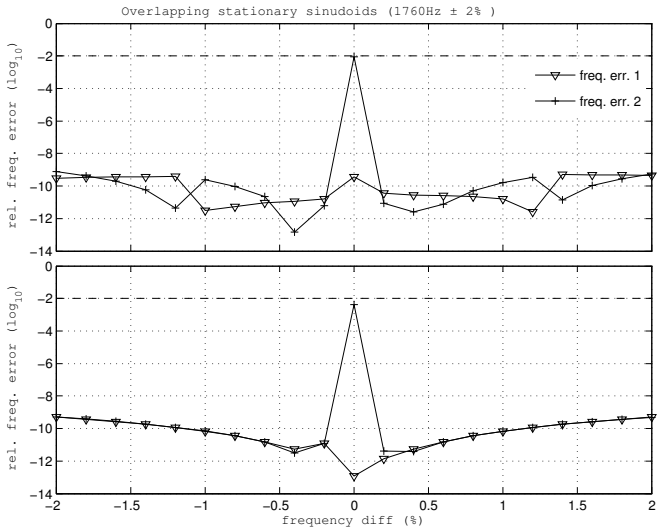


Figure 7.7: Error of joint estimation of the frequencies of two complex stationary overlapping sinusoids of equal amplitude. The relative phase difference at time 0 remains constant. Above: the $\hat{q}$ was set to the magnitude peak of the spectrum. Bottom: the $\hat{q}$ was set to the average of the frequencies of the sinusoids. The dashed line marks the 1% accuracy limit.

peak of the spectrum would be chosen for the required frequency estimate $\hat{q}$. Severe zero-padding would eventually place the spectrum peak very close to middle frequency of the overlapping partials, assuming the amplitudes are almost equal. It is also informative to see how the estimator behaves

when frequencies other than the FFT bins are used, therefore an average frequency $\hat{q} = (q_1 + q_2)/2$ was tested. Both plots in Fig. 7.7 exhibit roughly the same accuracy, rendering the use of the FFT bin value nearest to the spectrum peak appropriate for use in practice.
Above promising example serves merely as a *proof-of-concept*, more rigorous tests and comparison is required to establish the overall accuracy and usefulness in practice.

Practical considerations

In practical audio applications, a minimal accuracy requirement of a sinusoidal parameter estimator would generally correspond to the loudness and frequency resolution of human perception. If the analysis stage is followed by transformation/resynthesis, the accuracy requirement might rise well beyond human hearing abilities, as transformations might amplify the error.
Model based estimators suffer accuracy degradation for two main reasons: presence of Gaussian noise and discrepancy between the assumed model and reality. The single sinusoid estimators fail to take into account a neighboring sinusoid, however estimators normally implicitly limit this effect. Any FFT bin value arising from a time limited signal (which is always the case, since time infinite signals cannot be analyzed in practice) is largely affected by a bandwidth corresponding to the main lobe of the window function used, assuming the effect of the side lobes is negligible. Generally, the inter-sinusoid interference largely depends on the window function used. A sinusoid pair with frequencies sufficiently close would produce a single peak in magnitude spectrum for an arbitrary window function, suggesting high inter-sinusoid interference. A model based estimator, assuming 2 (or more) sinusoids inside the main lobe bandwidth would greatly improve accuracy in such cases.
It is interesting to note that non-stationary parameter estimators incur higher bandwidth. For example the *Quadratically Interpolated FFT* (QIFFT) (Abe and Smith, 2005) method requires 3 FFT bin values rather than just one, effectively increasing its bandwidth and making it more sensitive for interference. Similar holds for the distribution derivative method (Betser, 2009), whereas generalized reassignment (Wen and Sandler, 2009) requires a closer look. The use of various window functions $(w(t), tw(t), \frac{\partial w(t)}{\partial t})$ at the same frequency seems to avoid increasing the bandwidth. A close inspection reveals that is not the case. Figure 7.8 shows a magnitude spectrum of all the window functions required by reassignment, assuming the original one is Hanning window. Clearly, the side lobe attenuation is largely reduced in

all cases. It is also evident that energy concentration around the base frequency is reduced. In case of $w'(t), w''(t)$ most of the energy is confined in the region bounded by the Hanning window main lobe (± 2 frequency bins), whereas the time ramped window functions $w'(t)t, w(t)t$ cross that bound slightly. Evidently, time ramping increases the main lobe energy scatter, whereas time derivation decreases the side lobe attenuation. Simple mathematical deduction would generalize this observation to an arbitrary window function, however it would be out of scope of this chapter. In the definition

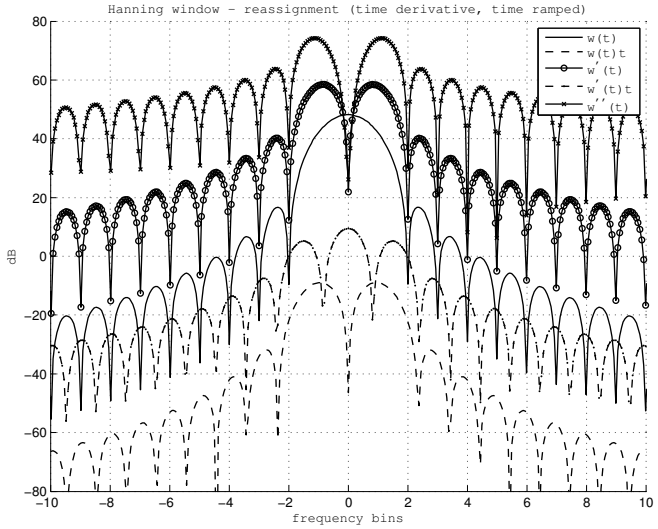


Figure 7.8: Magnitude spectrum of the original Hanning window function, it's time derivative, and time ramped versions, all required by reassignment.

of the proposed method a simple square window was implicitly used. It is however trivial to introduce an arbitrary window function by substituting $s(t)$ with $s(t)w(t)$. Such substitution introduces a slight change in the system and its solutions, however the same methods for solving it apply. Using again the Hanning window as an example, figure 7.9 shows magnitude spectrums of the original window and its multiplications with the Chebyshev polynomials of various degrees. Similarly as in the case of reassignment, energy concentration and side lobe attenuation is observed for derived window functions. The higher the Chebyshev polynomial degree, the better approximation and accuracy, therefore a tradeoff is required to balance effect of inter-sinusoidal interference. Comparing figures 7.8 and 7.9 suggest the required bandwidth is lower in case of reassignment, making it more

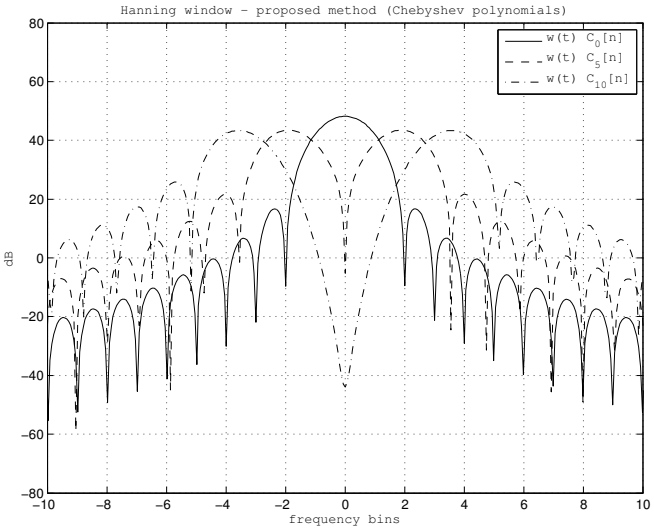


Figure 7.9: Magnitude spectrum of the original Hanning window function multiplied by Chebyshev polynomials

suitable for a single sinusoid analysis. However, the proposed method can resolve overlapping sinusoids which is inherently impossible with reassignment. Nevertheless, reducing bandwidth requirement by using a specially designed window function would be of great benefit.

7.4 Tests and results

Single sinusoid

Single sinusoid tests conducted were identical to those in (Wen and Sandler, 2009),(Muševič and Bonada, 2011) with an exception of a modified Signal-to-Noise Ratio (SNR) range. The metric used was Signal-to-Residual Ratio (SRR) defined by:

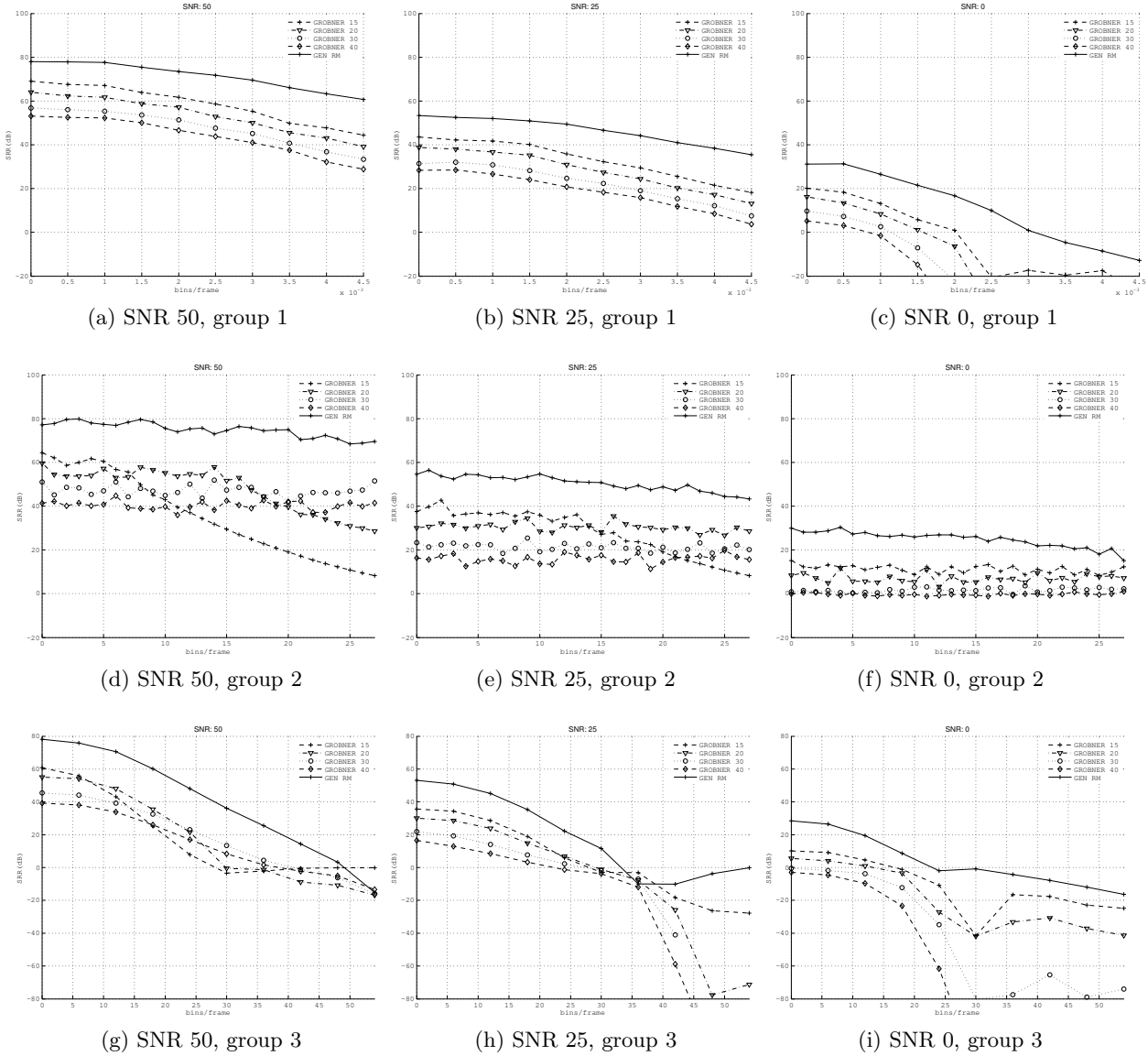
$$SRR = \frac{\sum_i h_i s_i^2}{\sum_i h_i (s_i - \hat{s}_i)^2}, \tag{7.68}$$

where $s_i, \hat{s}_i$ are discrete time samples of the original (excluding noise) and estimated signal respectively, while h_i are samples of the weighting function - in this case the Hanning window. The test signals analysed were real sinusoids sampled at 44100 Hz, window length of 1023 samples. The parameters of the test sinusoids were varied in the following way: 10 phase values

in the [0,0.45]π interval, 10 linear log-amplitude modulation values in the [0,0.0045] /frame interval (roughly corresponds to the [0,200] /s interval), 10 frequency values in the [255,255.9] bins interval (roughly corresponds to the [10.982, 11.021]Hz) and 10 linear frequency modulation values in the [0,27] bins/frame interval (roughly corresponds to the [0,16.000] Hz/s). The tests were conducted in 3 separate groups. In group 1 (subfigures 7.10a,7.10b and 7.10c), the linear frequency modulation was set to 0 while the log-amplitude modulation was varied (x- axis) in the mentioned range. In group 2 (subfigures 7.10d,7.10e and 7.10f) the log-amplitude modulation was set to 0 while the linear frequency modulation was varied (x-axis) in the mentioned range. In group 3 (subfigures 7.10g,7.10h and 7.10i), both the FM and log-AM were jointly varied (x-axis) in double the range compared to the groups 1 and 2. The SNR levels of 50dB, 25dB and 0dB were considered. The frequency range was selected around half of Nyquist frequency in order to avoid self-interference.
It is evident from results on Fig. 7.11 that the reassignment is more suitable for single sinusoid analysis. There is approximately $15dB$ SRR drop of for proposed method when compared to reassignment. Additionally, for groups 2 (FM only) and 3 (AM and FM), $50dB$ SNR the approximation degree of 15 suffers a severe drop of accuracy due to inability of Chebyshev polynomials of this degree to capture the modulations. A similar decrease of quality is also observed for the approximation degree of 20 for some higher modulation levels. This results are in accordance with the example from 7.1, where Fig. 7.4 demonstrates that the proposed representation reaches noise level for a highly modulated sinusoid at a degree around 25. More importantly, the same accuracy drop is not observed in group 1 (AM only), suggesting that frequency modulation is a much more difficult to model as amplitude modulation. A perhaps surprising fact is that lower degree approximations generally perform better than higher ones, which can be explained with the effect of noise which is amplified with each added degree of approximation. Results not presented suggest that accuracy analysing clean signal is exactly the same for all approximation degrees above 25 (sufficient to code the non-stationarities considered).

Two overlapping sinusoids

To author's knowledge a standardized test for evaluating joint estimations have not yet been developed. Tests were performed on a pair of synthesized stationary sinusoids with a difference between frequencies varying from 0 to 2.5 bins (26 values), relative phase difference varying from 0 to π (11 val-



(a) SNR 50, group 1 (b) SNR 25, group 1 (c) SNR 0, group 1

(d) SNR 50, group 2 (e) SNR 25, group 2 (f) SNR 0, group 2

(g) SNR 50, group 3 (h) SNR 25, group 3 (i) SNR 0, group 3

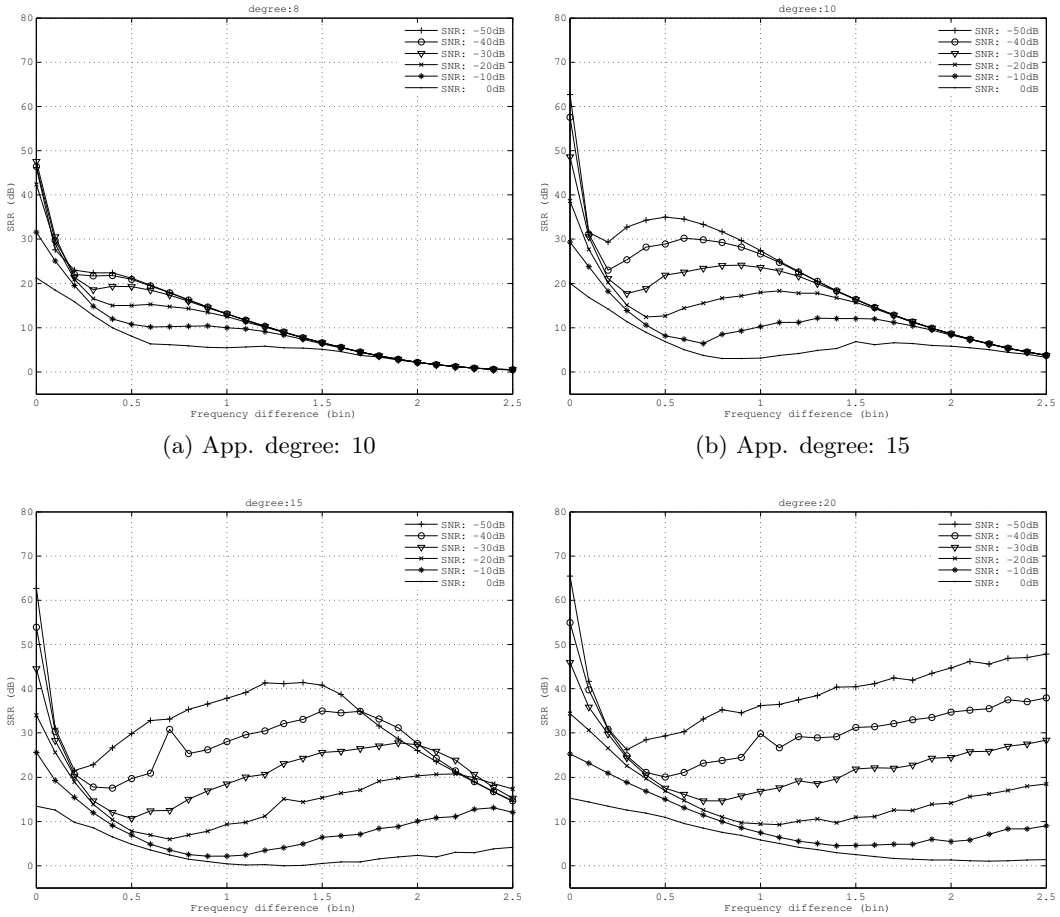
Figure 7.10: SRR for single sinusoid model analysis.

ues), frequency range from 200 Hz to 22000 Hz, while amplitudes were kept equal. Various window lengths and approximation degrees were studied. During initial testing it was discovered that approximation degree affects

accuracy to much bigger extent than the length of the window, therefore only results for $N = 1023$ are presented. Tests were performed in presence of Gaussian noise, various SNRs were considered, analytic signal was computed via discrete Hilbert transform (Marple, 1999). Fig. 7.11a, 7.11b, 7.11c and 7.11d reveal that increasing approximation degree improves results for higher frequency differences and deteriorates those for very low frequency differences. A combination of frequency difference (a property of the signal) and the approximation degree (a parameter of the algorithm) significantly affects the accuracy, which is generally considered undesirable and poses a drawback in this case.

7.5 Discussion

In practice, very little computational overhead is needed to derive lower approximation degree results from the higher ones, therefore an algorithm based on iteratively lowering the approximation degree should be efficient, would improve the overall accuracy and at the same time remove the need to predefine the approximation degree parameter. The presented method seems promising for the analysis of overlapping sinusoids and should be evaluated using real world musical signals. One drawback of the presented models is the assumption of the availability of the analytical signal, which were computed via discrete Hilbert transform (Marple, 1999). To avoid this problem a change of the analytical signal model to non-analytical one is required. Experimenting with such non-analytical models proved to be significantly harder, as the Gröbner basis for single non-stationary sinusoid failed to be found with the original version of the Buchberger algorithm in a reasonable time on an average desktop computer. Importantly, the same goes for joint estimation of two complex non-stationary sinusoids. Further, the symbolic computing algorithms for solving multivariate systems fail to exploit the inherent repetitive structure of the systems, caused by their intricate connection to Taylor series. Further, the property of *invariance* to variable swap in case of twin sinusoid model is not preserved, which is believed to increase the total computational cost enormously. The exact effect of breaking the invariance depends heavily on the structure of a specific system and is hard to predict. It is crucial to note that Gröbner basis of all mentioned models do exist and, if known, would provide means to accurate joint multiple-sinusoid non-stationary overlapping sinusoid analysis. Importantly, the Gröbner basis need not be computed in *real-time* - once known, they can be hard-coded in the algorithm and optimized for



(a) App. degree: 10

(b) App. degree: 15

(c) App. degree: 10

(d) App. degree: 15

Figure 7.11: SRR for twin sinusoid model.

maximum performance.

Chapter 8

Reassignment for cPACED

This chapter presents a new method for the analysis of cPACED sinusoids using high order derivatives. The polynomial amplitude is cancelled out to derive a closed form expression for the pole estimation. Since a high degree derivative is used it bears resemblance to the GRM, however the estimator is defined in the form of roots of a polynomial, more reminiscent to the high-resolution method (Badeau et al., 2006). It is of great importance that the accuracy indeed reaches that of the high-resolution methods, but exhibits superior performance. It is by no means an attempt to supersede the frequency resolution of the high-resolution methods, which remains by far the highest at 1 FFT bin.

8.1 cPACED model

The cPACED model is defined as follows:

$$s(t) = a(t)e^{(\mu_0 + j\omega_0)t}, \tag{8.1}$$

$$a(t) = p(t) + jq(t) = \sum_{k=0}(p_k + jq_k)t^k, \tag{8.2}$$

where $a(t)$ is the complex polynomial amplitude with real polynomials $p(t), q(t)$ and its real coefficients p_k, q_k respectively, while μ_0, ω_0 are the exponential damping and frequency parameter respectively, referred to as *pole* when combined into a complex number $\mu_0 + j\omega_0$. Such model covers all the parameters of a Gamma-tone envelope except the exact transient time inside the observed time frame as it is described in Christensen and van de

Par (2006).
Since the polynomial coefficients are complex, they affect AM as well as FM. Transforming the polynomial into the polar form yields the exponential AM and FM separately:

$$s(t) = \sqrt{p(t)^2 + q(t)^2} \exp\left((\mu_0 + j\omega_0)t + j \arctan \frac{q(t)}{p(t)}\right), \tag{8.3}$$

$$= \exp\left((\mu_0 t + \alpha(t)) + j(\omega_0 t + \phi(t))\right), \text{where} \tag{8.4}$$

$$\alpha(t) = \frac{1}{2}\log(p(t)^2 + q(t)^2), \tag{8.5}$$

$$\phi(t) = \arctan \frac{q(t)}{p(t)}. \tag{8.6}$$

The Taylor series of $\phi(t), \alpha(t)$ suggest a certain degree of ambiguity is expected. By denoting the phase and log-amplitude power series respectively

$$M_\alpha(t) = \sum_{k=0}^{\infty} \frac{\alpha^{(l)}(0)}{l!} t^l, \tag{8.7}$$

$$M_\phi(t) = \sum_{k=0}^{\infty} \frac{\phi^{(l)}(0)}{l!} t^l, \tag{8.8}$$

the actual linear phase parameter (i.e. frequency) is a sum of ω_0 and $\phi'(0)$ and the actual exponential damping parameter is a sum of μ_0 and $\alpha'(0)$. It is crucial to recognise this duality when assessing the accuracy of the algorithm. Figure 8.1 shows an example of such duality. Evaluating estimations of amplitude polynomial and exponential damping separately shows significant discrepancies, however the cumulative amplitude envelope is much more accurate.

8.2 Pole estimator using derivatives

A Fourier Transform of signal $s(t)$ at frequency ω, using a window function $w(t)$ is defined as an inner product:

$$S_w(\omega) = \langle s, w\Psi_{j\omega} \rangle, \tag{8.9}$$

where $\Psi_x = \exp(xt)$. The FT of signal time derivative is designated as

$$S'_w(\omega) = \left\langle s', w\Psi_{j\omega} \right\rangle = \tag{8.10}$$

$$- \left\langle s, w'\Psi_{j\omega} \right\rangle + j\omega \langle s, w\Psi_{j\omega} \rangle, \tag{8.11}$$

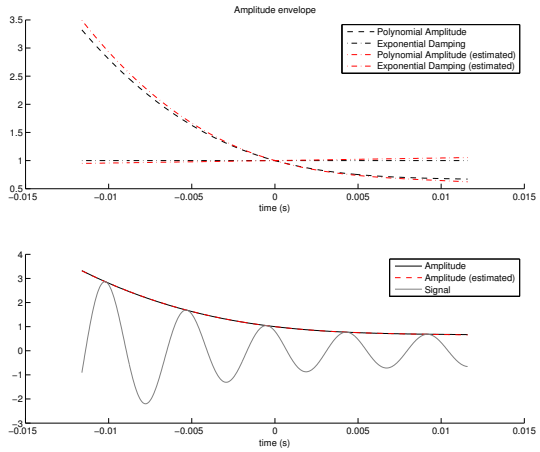


Figure 8.1: Amplitude polynomial and exponential damping estimates separately (above) and cumulative (below).

where the second equality follows from the distribution derivative rule (Betser, 2009) and implies $w(-\frac{T}{2}) = w(\frac{T}{2}) = 0$, where T is the length of the observation window. Higher signal derivatives can be easily derived using higher window derivatives, and further restrictions on window function apply: for $l-th$ signal derivative $w^{(l-1)}(-\frac{T}{2}) = w^{(l-1)}(\frac{T}{2}) = 0$. The FT of the derivative of cPACED model follows:

$$S'_w(\omega) = \frac{\partial}{\partial t}\langle a\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = \tag{8.12}$$

$$\langle a'\Psi_{\beta_0}, w\Psi_{j\omega}\rangle + \beta_0 S_w(\omega) \Rightarrow \tag{8.13}$$

$$\langle a'\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = S'_w(\omega) - \beta_0 S_w(\omega), \tag{8.14}$$

where the pole is designated as $\beta_0 = \mu_0 + j\omega_0$ for compactness. Time derivatives of both hand sides of 8.14 yield:

$$\langle a''\Psi_{\beta_0}, w\Psi_{j\omega}\rangle + \beta_0\langle a'\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = S''_w(\omega) - \beta_0 S'_w(\omega) \Rightarrow \tag{8.15}$$

$$\langle a''\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = S''_w(\omega) - 2\beta_0 S'_w(\omega) + \beta_0^2 S_w(\omega). \tag{8.16}$$

A general expression can easily be proven:

$$\langle a^{(k)}\Psi\beta_0, w\Psi_{j\omega}\rangle = \sum_{l=0}^{k} S_w^{(l)} \binom{k}{l} (-\beta_0)^{k-l}. \tag{8.17}$$

Proof by induction: for $k = 1$ the above expression simplifies to 8.14. Assuming 8.17 for some k, its derivative is:

$$\langle a^{(k+1)}\Psi_{\beta_0}, w\Psi_{j\omega}\rangle + \beta_0\langle a^{(k)}\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = \sum_{l=0}^{k} S_w^{(l+1)}\binom{k}{l}(-\beta_0)^{k-l}. \quad (8.18)$$

Inserting the induction assumption and rearranging the indexes yields:

$$\begin{aligned}\langle a^{(k+1)}\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = &\sum_{l=0}^{k} S_w^{(l+1)}\binom{k}{l}(-\beta_0)^{k-l} - \sum_{l=0}^{k} S_w^{(l)}\binom{k}{l}(-\beta_0)^{k+1-l} = \\ &\sum_{l=0}^{k+1} S_w^{(l)}(-\beta_0)^{k-l}\left(\binom{k-1}{l} + \binom{k-1}{l-1}\right) = \\ &\sum_{l=0}^{k+1} S_w^{(l)}(-\beta_0)^{k-l}\binom{k}{l}, \quad (8.19)\end{aligned}$$

concluding the proof. Assuming the degree of the amplitude polynomial is K and using the fact that for polynomial $p(x)$ of degree D, $\frac{\partial^{D+1}p(x)}{\partial x^{D+1}} = 0$, the following equation can be obtained:

$$\langle a^{(K+1)}\Psi_{\beta_0}, w\Psi_{j\omega}\rangle = \sum_{l=0}^{K+1} S_w^{(l)}(-\beta_0)^{k-l}\binom{k}{l} = 0, \quad (8.20)$$

which is a $(K+1)^{th}$ degree polynomial in respect to β_0 and it's K roots $\hat{\beta}_{0,k}, k = 1..K$ are the estimates for the pole β_0. Note that both α_0 and ω_0 are estimated jointly. The K estimates will in general not be equal even in a noise-less case as already outlined in Badeau et al. (2006). All the poles should in theory give the *correct* result. The pole estimate is used to estimate the coefficients of the polynomial. In such setting even a inaccurate pole estimation leads to acceptable SRR, since the polynomial amplitude coefficients *make up* for a inaccurate pole estimate. Nonetheless, the *best* estimate can be chosen by $\max_k \left|\langle s, e^{j\Im(\hat{\beta}_{0,k})}\rangle\right|$.

8.3 Complex polynomial amplitude estimator

Pole estimates can be used to construct a simple linear system $Ax = b$:

$$A = \begin{bmatrix} \langle t^K \Psi_{\hat{\beta}_0}, w\Psi_{j\omega_1} \rangle & \cdots & \langle t\Psi_{\hat{\beta}_0}, w\Psi_{j\omega_1} \rangle & \langle \Psi_{\hat{\beta}_0}, w\Psi_{j\omega_1} \rangle \\ \langle t^K \Psi_{\hat{\beta}_0}, w\Psi_{j\omega_2} \rangle & \cdots & \langle t\Psi_{\hat{\beta}_0}, w\Psi_{j\omega_2} \rangle & \langle \Psi_{\hat{\beta}_0}, w\Psi_{j\omega_2} \rangle \\ \vdots & & \vdots & \vdots \\ \langle t^K \Psi_{\hat{\beta}_0}, w\Psi_{j\omega_R} \rangle & \cdots & \langle t\Psi_{\hat{\beta}_0}, w\Psi_{j\omega_R} \rangle & \langle \Psi_{\hat{\beta}_0}, w\Psi_{j\omega_R} \rangle \end{bmatrix} \tag{8.21}$$

$$x = \begin{bmatrix} a_K \\ a_{K-1} \\ \vdots \\ a_1 \\ a_0 \end{bmatrix} \quad b = \begin{bmatrix} S_w(\omega_1) \\ S_w(\omega_2) \\ \vdots \\ S_w(\omega_{R-1}) \\ S_w(\omega_R) \end{bmatrix}, \tag{8.22}$$

where $\hat{\beta}_0 = \hat{\alpha}_0 + j\hat{\omega}_0$ is the pole estimation acquired as described in section 8.2. Solutions of above linear system give estimates for coefficients of the complex amplitude polynomial. The window function does not have to be the same as the one used for pole estimation - restriction on edge values does not apply. Each row of matrix A corresponds to an arbitrary frequency, the most reasonable choice being the ones carrying most of the energy of the sinusoid in question, eg: as close to magnitude peak frequency as possible. An efficient algorithm implementation can utilise FFT bin values and zero-padding to avoid costly computation of DTFT at specific frequencies and to adjust inter-bin frequency difference. Matrix A need not be square, many times an overdetermined system is desired: the number of estimation frequencies R can be larger than the number of unknowns K. Such systems can solved on least-square basis via Moore-Penrose matrix pseudo-inverse A^+:

$$A^+ = (A^*A)^+A^* = (A^*A)^{-1}A^*, \tag{8.23}$$

where A^* designates a conjugate transpose of matrix A and $A^{-1} = A^+$ if A is square.

8.4 Tests and results

A polynomial amplitude of degree 3 was studied and the polynomial denoted as: $[a_3, a_2, a_1, a_0] = [p_3 + jq_3, p_2 + jq_2, p_1 + jq_1, p_0 + jq_0]$. The test values for p_3, p_2, p_1 were chosen so all the terms of the amplitude polynomial have

equal impact on the final value:

$$p_3 \in \left[-\left(\frac{fs}{2T}\right)^3, \left(\frac{fs}{2T}\right)^3 \right] \tag{8.24}$$

$$p_2 \in \left[-\left(\frac{fs}{2T}\right)^2, \left(\frac{fs}{2T}\right)^2 \right] \tag{8.25}$$

$$p_1 \in \left[-\frac{fs}{2T}, \frac{fs}{2T} \right]. \tag{8.26}$$

The exact same value sets were used for the imaginary part of the polynomial $q(t)$. A Hann^3 window function of length 512 samples was used for pole estimation and Hann window for the complex polynomial coefficients estimation. Damping factor was varied in the bounds [-100,100] and only one frequency of 1000Hz was considered. For each parameter except frequency, only the 2 extreme values and 0 have been tested in order to keep computational time reasonable. The comparison to a 3^{th} degree (i.e. 4 poles and amplitudes) simple high-resolution method implementation from DESAM Toolbox (Lagrange, 2010) (section 5.1.2.) without whitening was conducted. The signal tested is the real part of the complex cPACED signal, reflecting the real world scenario when analytical signal isn't available. To measure accuracy, the commonly used Signal-to-Residual-Ratio (SRR) metric was used:

$$SRR = \frac{\langle s, ws \rangle}{\langle s - \hat{s}, w(s - \hat{s}) \rangle}, \tag{8.27}$$

where $s, \hat{s}$ are the original signal (without noise) and the estimated signal respectively, and w the Hann window. The Signal-to-Noise-Ratio (SNR) range from [50,-20] was studied. The total computation times for both methods follow:

Proposed method	28s
High-resolution	5400s

(8.28)

Since HRM involves singular value decomposition of correlation matrix of size $N/2 \times N/2$ the computation cost is significantly higher as the proposed method only requires $K - 1$ FFTs for the pole and K DTFTs for the complex polynomial estimates.

The classic Cramer-Rao bounds (CRBs) parameter-by-parameter comparison would total to 10 plots (8 for the real/imaginary polynomial coefficients and 2 for the pole), overcomplicating the results and obscuring the overall accuracy. A more intuitive approach would involve only one SRR/SNR plot,

thus a different upper accuracy bound is required. A *near perfect* estimator can be constructed by substituting the pole estimates with actual poles and solving the linear system 8.21. The mean and variance of the SRR, computed with the aforementioned estimator represents a good upper SRR bound. Figure 8.2 depicts the mean and variance of the baseline estimator, proposed method and HRM. At low SNR the methods perform roughly the same, HRM reaching the upper bound while proposed method performing 2dB below. For high SNRs both methods reach a plateau, however the proposed method's plateau is about 10dB higher.

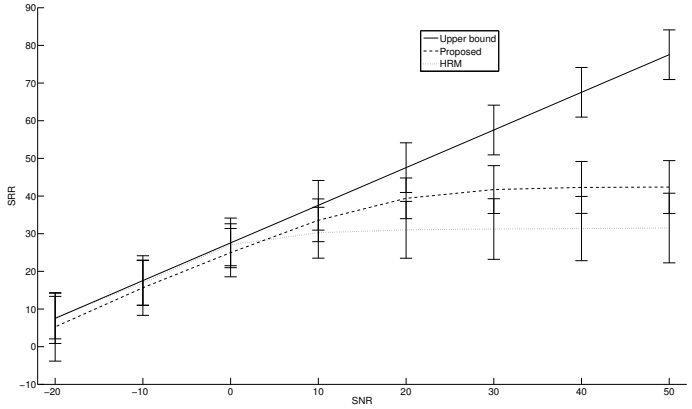


Figure 8.2: SRR: Mean and variance

8.5 Discussion

In this chapter a novel method for analysing cPACED signals was represented, tested and compared to the HRM. In conducted tests the HRM performs marginally better in high noise cases, while the proposed method performs significantly better in low-noise cases. More rigorous testing, involving multicomponent signals is required, but is out of scope of this document.

While HRM is a computationally intense method and is designed to jointly estimate parameters of multiple cPACED sinusoids in the entire frequency range, the proposed method focuses on a narrow frequency range to estimate a single cPACED sinusoid. The flexibility of HRM is a huge overkill for the tests conducted, which is reflected in substantially higher (about 2 degrees of magnitude) computational costs. However the proposed method could be

invoked on different frequency ranges, effectively covering the whole spectrum. The advantage of the proposed method in this case is the ability to process only certain frequency regions of interest, reducing the final computational cost.

The pole estimator is inspired by the generalised reassignment method Wen and Sandler (2009), imposing significant requirement on the number and type of the window function. Such window set usually exhibits suboptimal time-frequency properties and quickly reduces the condition number of the resulting linear system to a value too low to handle even using very high precision computation. The distribution derivative method Betser (2009) circumvents this problem by constructing the linear system using FT values at different frequencies, rather than using a single FT value, but using different windows. A version of the proposed method that does not impose the window requirement would enable an extremely accurate, very high degree cPACED analysis and would, analogously, correspond to the distribution derivative method. cPACE model is well suited for analysis of close frequency non-stationary sinusoids, as the amplitude beating function can be much better approximated with a polynomial than a generalised sinusoid, thus coding such signals is expected to be of higher accuracy with the proposed method when compared to the generalised reassignment.

Chapter 9

DDM for a Hybrid Sinusoidal Model - Generalised Sinusoid with Complex Amplitude Modulation

The hybrid sinusoidal model will be defined as follows:

$$s(t) = a(t)e^{r(t)}, \tag{9.1}$$

$$a(t) = \sum_{k=0} a_k m_k(t), a_k \in \mathbb{C} \tag{9.2}$$

$$r(t) = \sum_{l=0} r_l n_l(t), r_l \in \mathbb{C} \tag{9.3}$$

where m_k, n_k are the complex amplitude and log-AM/FM model functions respectively. To accommodate for the static amplitude and phase, $m_0 = n_0 = 1$ is assumed. A most common, but by no means mandatory selection for the model functions are polynomials: $m_k = n_k = t^k$.
The above model 9.1 is ambiguous with respect to parameters r_0 and a_0. To show this, the following derivation is considered:

$$s(t) = a(t)e^{r(t)} = a_0\tilde{a}(t)e^{r_0+\tilde{r}(t)} \tag{9.4}$$

$$= \tilde{a}(t)\exp(\overbrace{\log(|a_0|) + j\angle(a_0) + r_0}^{\tilde{r}_0} + \tilde{r}(t)), \tag{9.5}$$

Clearly a_0 and r_0 are in fact the same parameter in either Cartesian or polar coordinates. The decision seems irrelevant, however as will be shown in section 9.1, using the Cartesian form would result in a rank-deficient system, therefore the model will be constrained to $a_0 = 1$.
It is important to note that since modulation functions are complex they both contribute to overall AM/FM. If the same model functions are used ($m_k = n_k$) that can lead to some ambiguity, especially when the energy of m_k, n_k declines fast with k. Such ambiguity can be demonstrated when using polynomials for the modulation functions $m_k = n_k = t^k$:

$$a(t)e^{r(t)} = \exp(\log(a(t)) + r(t)) \tag{9.6}$$
$$\approx \exp(a_1 t + (2a_2 - {a_1}^2)t^2 + r(t)), \tag{9.7}$$

using the 2^{nd} degree truncated Taylor expansion. It is expected that an estimator for the model in 9.1 could be inaccurate when separate parameter estimates are considered, but generally much more flexible due to twin AM/FM functions.
An example is shown in figure 9.1, where 2 hybrid model sinusoids with significantly different a_1, r_1 reach SRR of $24dB$. To reach higher SRR the parameters would have to eventually match exactly, however in noisy conditions a relatively high SRR is achievable with substantial error in per-parameter estimates. From figure 9.1 is also evident that the proposed model includes sinusoids with *negative* amplitude, suggesting good coding abilities for sinusoid pairs with close frequencies (i.e.: beating partials). The negative amplitude can only occur when $\Im[a(t)] = 0$, since overall amplitude corresponds to: $\sqrt{a(t)\bar{a}(t)}e^{\Re[r(t)]}$. The notion of negative amplitude is purely artificial ($\sqrt{a(t)\bar{a}(t)}e^{\Re[r(t)]}$ cannot be negative), as it does not have a natural, physical meaning, however it comes handy as a mathematical generalisation. If negative amplitude is not allowed and $\Im[a(t)] = 0$, the derivative of $\sqrt{a(t)\bar{a}(t)}e^{\Re[r(t)]} = |a(t)|e^{\Re[r(t)]}$ is not continuous for all roots of $a(t)$. In such cases the absolute value can be easily dropped and negative amplitude introduced, leading to mathematically sound model.

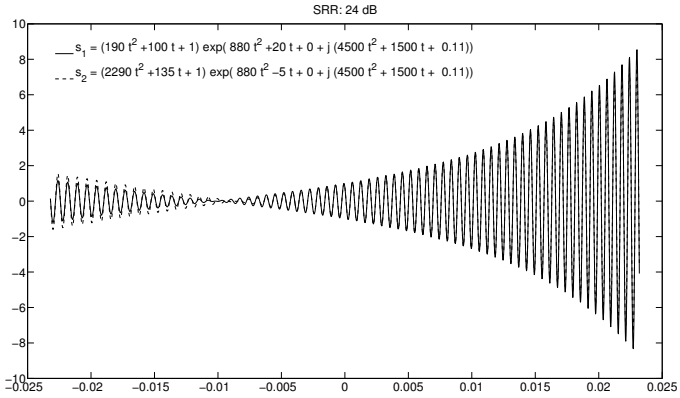


Figure 9.1: SRR between two sinusoids with significantly different parameters 24dB.

9.1 Non-linear multivariate polynomial system of equations

The non-linear system can be derived by considering the signal time derivative must be manipulated in the following way:

$$s'(t) = a'(t)e^{r(t)} + r'(t)a(t)e^{r(t)}, \Rightarrow \tag{9.8}$$

$$a(t)s'(t) = a'(t)s(t) + a(t)r'(t)s(t). \tag{9.9}$$

The last row can be rewritten in a more verbose form that reveals the non-linearity of the system:

$$m_0 s' + \sum_{k=1}^{K} a_k m_k s' = \sum_{k=1}^{K} a_k m_k' s + \left(m_0 + \sum_{k=1}^{K} a_k m_k \right) \sum_{l=1}^{L} r_l n_l' s, \tag{9.10}$$

where time variable t was omitted for compactness. The only non-linear terms arise from the last - double sum expression. Multiplying both sides with a window function $w(t)$ and taking a Fourier Transform (FT) at frequency ω yields:

$$S'_{wm_0}(\omega) + \sum_{k=1}^{K} a_k S'_{wm_k}(\omega) = \sum_{k=1}^{K} a_k S_{wm_k'}(\omega) + \sum_{l=1}^{L} r_l S_{wm_0 n_l'}(\omega) + \sum_{k=1}^{K} \sum_{l=1}^{L} a_k r_l S_{wm_k n_l'}(\omega), \tag{9.11}$$

where $S_f(\omega) = \langle s(t)f(t), e^{j\omega t}\rangle$, is FT of the signal s multiplied by function f and $S'_g(\omega) = \langle s'(t)g(t), e^{j\omega t}\rangle$ is FT of the signal derivative multiplied by function g at frequency ω. Note that $n'_0 = m'_0 = 0$ and thus the sums on the right-hand side start at index 1 rather than 0. Above equation can be viewed as a (non-linear) multivariate polynomial with respect to parameters $a_k : k = 1 \dots K, r_l : l = 1 \dots L$. The expressions S'_f, S_g can be considered constants for any f, g as they can be directly computed from the signal. To calculate S'_f accurately, sample difference in time domain should be avoided Marchand and Depalle (2008). A common approach is the use of distribution derivative rule $\langle x', y\rangle = -\langle x, y'\rangle$ and a real window function w as a part of the kernel y:

$$\begin{aligned} S'_{gw}(\omega) =& \langle s'g, w\psi_\omega\rangle = \langle s', \bar{g}w\psi_\omega\rangle && (9.12)\\ =& -\langle s, \bar{g}'w\psi_\omega\rangle - \langle s, \bar{g}w'\psi_\omega\rangle + j\omega\langle s, \bar{g}w\psi_\omega\rangle && (9.13)\\ =& -\langle sg', w\psi_\omega\rangle - \langle sg, w'\psi_\omega\rangle + j\omega\langle sg, w\psi_\omega\rangle && (9.14)\\ =& -S_{g'w}(\omega) - S_{gw'}(\omega) + j\omega S_{gw}(\omega), && (9.15)\end{aligned}$$

where ψ_ω is generally a kernel function with FT centred around frequency ω, for the last equality to hold the kernel is set simply to the Fourier kernel: $\psi_\omega(t) = e^{j\omega t}$. Higher time derivatives can accurately be computed by chaining the above expression. Rearranging the equation and collecting together the model parameters yields:

$$\begin{aligned} S'_{wm_0}(\omega) = \sum_{k=1}^{K} a_k(S_{wm'_k}(\omega) - S'_{wm_k}(\omega)) + \\ \sum_{l=1}^{L} r_l S_{wm_0 n'_l}(\omega) + \sum_{k=1}^{K}\sum_{l=1}^{L} a_k r_l S_{wm_k n'_l}(\omega). \quad (9.16)\end{aligned}$$

Taking the FT at different frequencies close to the peak provides as many equations as necessary. Considering a range of frequency values $\omega_1, \dots, \omega_{(K-1)(L-1)}$, the following system of equations can be deduced:

$$A_a(\omega) = (S_{wm'_1}(\omega) - S'_{wm_1}(\omega) \quad S_{wm'_2}(\omega) - S'_{wm_2}(\omega) \dots \quad S_{wm'_K}(\omega) - S'_{wm_K}(\omega)) \quad (9.17)$$

$$A_r(\omega) = (S_{wm_0 n'_1}(\omega) \quad S_{wm_0 n'_2}(\omega) \dots \quad S_{wm_0 n'_L}(\omega)) \quad (9.18)$$

$$A_{ar}(\omega) = (S_{wm_1 n'_1}(\omega) \dots S_{wm_K n'_1}(\omega) S_{wm_1 n'_2}(\omega) \dots S_{wm_K n'_2}(\omega) \dots S_{wm_K n'_L}(\omega)) \quad (9.19)$$

$$A = \begin{pmatrix} A_a(\omega_1) & A_r(\omega_1) & A_{ar}(\omega_1) \\ A_a(\omega_2) & A_r(\omega_2) & A_{ar}(\omega_2) \\ \vdots & \vdots & \vdots \\ A_a(\omega_{(K-1)(L-1)}) & A_r(\omega_{(K-1)(L-1)}) & A_{ar}(\omega_{(K-1)(L-1)}) \end{pmatrix} \quad (9.20)$$

$$x = (a_1, a_2, \dots, a_K, r_1, r_2, \dots, r_L, a_1 r_1, \dots, a_1 r_L, a_2 r_1, \dots, a_2 r_L, \dots, a_K r_1, \dots, a_K r_L) \quad (9.21)$$

$$b = (S'_{wm_0}(\omega_1), S'_{wm_0}(\omega_2), \dots, S'_{wm_0}(\omega_{(K-1)(L-1)})), \quad (9.22)$$

thus solving $Ax = b$ estimates the parameters of the model and the cross-products. The above system is general enough to be applied to a variety of model functions. If $K = 0$ the model corresponds to the generalised sinusoid and the algorithm is identical to the classical DDM. Analogously, if $L = 1$, the model corresponds to the cPACED and the algorithm the DDM version of the method described in section 8. GRM and DDM, originally designed for the analysis of generalised sinusoid model have been modified to work with cPACED models. Moreover, the DDM was adapted for the hybrid sinusoidal model, combining the cPACED and generalised sinusoid model.

9.2 DDM for cPACED

To compare the proposed method to the current state-of-the-art methods, the scope should be limited to either generalised sinusoid model or cPACED model. The accuracy for the generalised model has been studied in depth. This section will compare the accuracy of a single sinusoid of the adapted DDM method described in the preceding section with the HRM and GRM for cPACED (see section 8). The model functions in equation 9.16 must be replaced with polynomials:

$$S'_w(\omega) + \sum_{l=1}^{L} a_l S'_{t^l w}(\omega) = \sum_{l=1}^{L-1} l a_l S_{t^{l-1} w}(\omega) + \sum_{l=1}^{L} a_l \sum_{k=0}^{K-1} (k+1) r_{k+1} S_{t^{l+k}}(\omega) + \sum_{k=0}^{K-1} (k+1) r_{k+1} S_{t^k}(\omega) \quad (9.23)$$

For a cPACED sinusoids with polynomial amplitude of degree 3 (ie: K=1, L=3) the following special case can be deduced:

$$S'_w(\omega) + a_1 S'_{tw}(\omega) + a_2 S'_{t^2 w}(\omega) + a_3 S'_{t^3 w}(\omega) = a_1 S_w(\omega) + 2a_2 S_{tw}(\omega) + 3a_3 S_{t^2 w}(\omega) + a_1 r_1 S_{tw}(\omega) + a_2 r_1 S_{t^2 w}(\omega) + a_3 r_1 S_{t^3 w}(\omega) + r_1 S_w(\omega) \quad (9.24)$$

Grouping the linear and non-linear terms in respect to a_l, r_1:

$$S'_w(\omega) = a_1(S_w(\omega) - S'_{tw}(\omega)) + a_2(2S_{tw}(\omega) - S'_{t^2 w}(\omega)) + a_3(3S_{t^2 w}(\omega) - S'_{t^3 w}(\omega)) + r_1 S_w(\omega) + a_1 r_1 S_{tw}(\omega) + a_2 r_1 S_{t^2 w}(\omega) + a_3 r_1 S_{t^3 w}(\omega). \quad (9.25)$$

The distribution derivative rule can be applied to the S' terms:

$$S'_{t^l w}(\omega) = -lS_{t^{l-1}w}(\omega) - S_{t^l w'}(\omega) + j\omega S_{t^l w}(\omega), \quad \text{for } l > 0 \quad (9.26)$$

$$S'_w(\omega) = -S_{w'}(\omega) + j\omega S_w(\omega), \quad \text{for } l = 0 \quad (9.27)$$

Following the approach of the distribution derivative method and considering FT at 7 different frequencies:

$$A = \begin{pmatrix} S_w(\omega_1) - S'_{tw}(\omega_1) & \cdots & S_w(\omega_7) - S'_{tw}(\omega_7) \\ 2S_{tw}(\omega_1) - S'_{t^2w}(\omega_1) & \cdots & 2S_{tw}(\omega_7) - S'_{t^2w}(\omega_7) \\ 3S_{t^2w}(\omega_1) - S'_{t^3w}(\omega_1) & \cdots & 3S_{t^2w}(\omega_7) - S'_{t^3w}(\omega_7) \\ S_w(\omega_1) & \cdots & S_w(\omega_7) \\ S_{tw}(\omega_1) & \cdots & S_{tw}(\omega_7) \\ S_{t^2w}(\omega_1) & \cdots & S_{t^2w}(\omega_7) \\ S_{t^3w}(\omega_1) & \cdots & S_{t^3w}(\omega_7) \end{pmatrix}'$$

,

$$x = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ r_1 \\ r_1 a_1 \\ r_1 a_2 \\ r_1 a_3 \end{pmatrix}, b = \begin{pmatrix} S'_w(\omega_1) \\ S'_w(\omega_2) \\ S'_w(\omega_3) \\ S'_w(\omega_4) \\ S'_w(\omega_5) \\ S'_w(\omega_6) \\ S'_w(\omega_7) \end{pmatrix}.$$

Solving the linear system $Ax = b$ is the estimator for the model parameters. For high parameter values, the frequency spread of the signal might be large - a small number of frequency bins (in above case 7) might not suffice to cover all the information in the Fourier domain. In such cases more frequency bins can be considered, forcing the matrix A to be overdetermined, resulting in a least-squares solution via Moore-Penrose pseudoinverse: $x = (A^H A)^{-1} A^H b$.

9.3 Tests and results

A polynomial amplitude of degree 3 was studied and the polynomial denoted as: $[a_3, a_2, a_1, 1] = [p_3 + jq_3, p_2 + jq_2, p_1 + jq_1, 1]$. The test values for p_3, p_2, p_1

were chosen so all the terms of the amplitude polynomial have equal impact on the final value:

$$p_3 \in \left[-\left(\frac{fs}{2T}\right)^3, \left(\frac{fs}{2T}\right)^3 \right] \tag{9.28}$$

$$p_2 \in \left[-\left(\frac{fs}{2T}\right)^2, \left(\frac{fs}{2T}\right)^2 \right] \tag{9.29}$$

$$p_1 \in \left[-\frac{fs}{2T}, \frac{fs}{2T} \right]. \tag{9.30}$$

The exact same value sets were used for the imaginary part of the polynomial $q(t)$. A Hann window function of length 511 samples was used for pole estimation and Hann window for the complex polynomial coefficients estimation. Damping factor was varied in the bounds [-100,100] and only one frequency of 1000Hz was considered. For each parameter except frequency, only the 2 extreme values and 0 have been tested in order to keep computational time reasonable. The comparison to a 3^{th} degree (i.e. 4 poles and amplitudes) simple high-resolution method (HRM) implementation from DESAM Toolbox Lagrange (2010) (section 5.1.2.) without whitening and the cPACED reassignment method (cPACED-RM) Muševič and Bonada (2013) was conducted. The signal tested is the real part of the complex cPACED signal, reflecting the real world scenario when analytical signal isn't available.
To measure accuracy, the commonly used Signal-to-Residual-Ratio (SRR) metric was used:

$$SRR = \frac{\langle s, ws \rangle}{\langle s - \hat{s}, w(s - \hat{s}) \rangle}, \tag{9.31}$$

where $s, \hat{s}$ are the original signal (without noise) and the estimated signal respectively, and w the Hann window. The Signal-to-Noise-Ratio (SNR) range from [50,-20] was studied. The total computation times for both methods follow:

cPACED-RM	15s
cPACED-DDM	26s
High-resolution	5400s

(9.32)

Since HRM involves singular value decomposition of correlation matrix of size $N/2 \times N/2$ the computation cost is significantly higher as the proposed method only requires $K - 1$ FFTs for the pole and K DTFTs for the complex polynomial estimates.

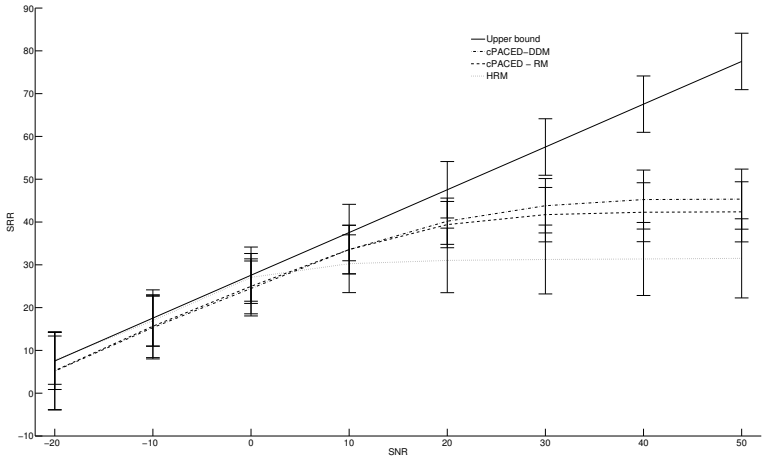


Figure 9.2: SRR: Mean and variance

The classic Cramer-Rao bounds (CRBs) parameter-by-parameter comparison would total to 10 plots, overcomplicating the results and obscuring the overall accuracy and . A more intuitive approach would involve only one SRR/SNR plot, thus a different upper accuracy bound is required. For each test case the CRBs for each parameter were computed. Denoting a CRB for parameter a_0 as ϵ_{a_0}, the minimum SRR for the specific CRB set can be defined:

$$\min SRR(\hat{s}(a_3 \pm \epsilon_{a_3}, a_2 \pm \epsilon_{a_2}, a_1 \pm \epsilon_{a_1}, r_0 \pm \epsilon_{r_0}, r_1 \pm \epsilon_{r_1}). \tag{9.33}$$

The mean and variance of the minimum SRR represents a good upper SRR bound. Figure 9.2 depicts the mean and variance the upper SRR bound, proposed method (cPACED-DDM), cPACED-RM and HRM. At low SNR the methods perform roughly the same, HRM reaching the upper bound while cPACED-DDM and cPACED-RM performing 2dB below. For high SNRs the methods reach a plateau, however the plateau for cPACED-DDM and cPACED-RM is about 10dB higher.

9.4 Discussion

In this section currently the most flexible sinusoidal method for TF energy re-allocation analysis has been described. The concept used in the distribution derivative method is used to solve a non-linear multivariate system of polynomial obtained by the first signal derivative. It is important to note

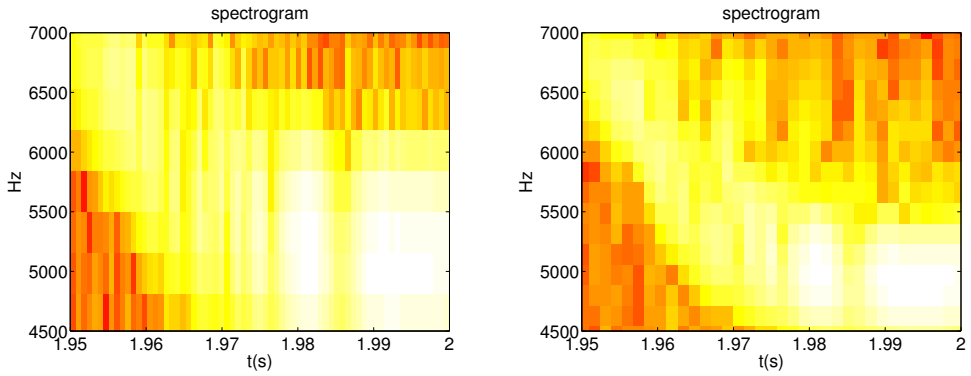
that higher signal derivatives would provide enough equations for a solution to exist, however a significantly more complex system would be obtained. Even if solution could eventually be obtained, it is desirable to avoid higher signal derivatives due to ill conditioning.
The method was compared favourably to the high-resolution method for the cPACED signal model, however a much more flexible models are possible with the proposed method. Further, the proposed sinusoidal model seems promising for the analysis of overlapping partials, as the beating function corresponds to real value amplitude/frequency modulated sinusoid - a subfamily of signals described by the proposed model.
On the other hand high-resolution methods' intrinsic frequency resolution of 1 frequency bin Porat and Friedlander (1987); Hua and Zhang (1988) for damped sinusoids has not been surpassed, as common window function mainlobe width (several bins) and significant sidelobe amplitude both reduce the frequency resolution.

Chapter 10

Variable bandwidth DDM for bird song analysis

In this chapter the non-stationary sinusoidal analysis methods will be tested on a real world signals. Analysing bird chirp sounds represents a significant challenge as frequency change can reach 100kHz/s (Stowell and Plumbley, 2012). Successful detection of the frequency slope is crucial for high accuracy analysis necessary for machine recognition of bird sounds, a very important tool that can facilitate unattended monitoring and other applications (Walters et al., 2012) in the field of bioacoustics and ecology. For birdsongs, important tasks include recognition of species and individuals (Cheng et al., 2012). A particularly important is successful application of current methods to songbird mixtures, rather than monophonic cases as singing often occurs within flocks or dawn choruses.
The methods described in previous chapter were not explicitly designed for the analysis of sound mixtures, however the use of window functions with desirable TF properties allow for high accuracy analysis of sinusoids with reasonably different frequencies. Importantly, high AM/FM drastically increase the bandwidth a sinusoid occupies, making resolution of sinusoids close in frequency even harder. Figure 10.1 depicts 2 spectrograms using different window length of a relevant part of a bird chirp. Clearly, none of the TF resolution combinations yield satisfactory results for an accurate chirp rate estimation. Further, it is unlikely that any TF resolution combination would result in a readable spectrogram.

A peak-tracking algorithm (T.F.Quateri, 1993) would have experienced severe problems avoiding the spurious peaks generated by the extreme FM



(a) Win. length: 64 samples (b) Win. length:128 samples

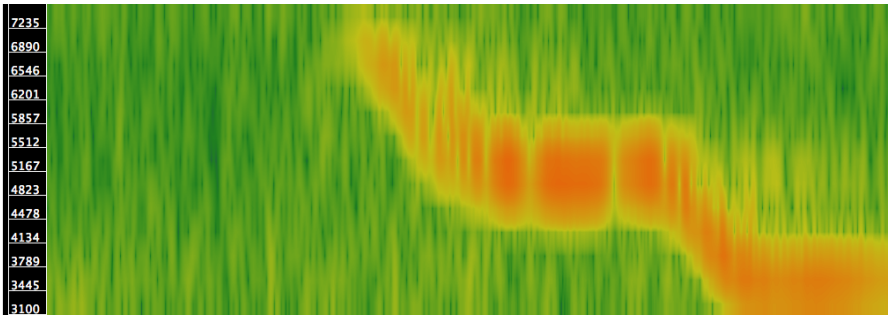
Figure 10.1: Birdchirp, approximately 100kHz/s

(AM), but it is safe to assume that with some modifications specifically tailored for this particular case an acceptable result could be obtained. An example can be seen on figure 10.2. However, such specialisation is undesired as it tends to spawn numerous sub-algorithms limited to a very specific task only. In the present case, a high-level algorithm is employed to *fix* the problem of a low-level one, i.e.: the inherently low resolution of FT in cases of highly modulates sinusoids. Methods described in the previous chapters have been designed for cases like the bird chirp sounds. It has been shown 5 that extreme AM/FM cases can be tackled with GRM by an adaptive kernel and re-estimation. However, the current upper tested FM value was about 35kHz/s, while the bird chirps under study exhibit modulations up to roughly 3 times higher.

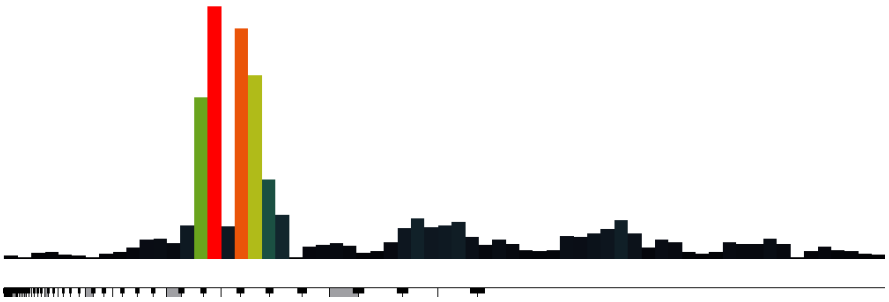
Rather than using a GRM based approach employed in chapter 5 a similar, more straightforward approach using DDM can be envisaged. The re-estimation procedure using the polynomial-phase kernel requires computation of the kernel, a costly operation that should be avoided if possible.

The key to successful estimation of signals occupying an extended amount of bandwidth is to force the set of kernels used by the estimator to cover a significant amount of the signal's bandwidth. Using the polynomial-phase Fourier kernel is one way to modify a single kernel to cover more bandwidth - the use such kernel with GRM is straightforward. Another simple way to achieve the same goal is to use a number of Fourier kernels or simply multiple FFT bin values - a technique inherent to DDM.

A simple practical example demonstrates the usefulness of the ability to



(a) Spectrogram



(b) Mag. spectrum located at the middle of the above spectrogram

Figure 10.2: Bird sound, high FM and stationary sections show

construct and arbitrarily overdetermined linear system for DDM (for details see appendix A.4). Rather than DDM spectrogram (short for DDM reassigned spectrogram), where all the non-stationary parameter estimates are discarded, the frequency trajectory estimates are plotted for each bin. To asses the effect of the afore mentioned procedure, the amplitude information was discarded and the non-stationary parameters are superimposed on the spectrogram, the results are shown on figures 10.3-10.6. Each successive figure depicts an increasing number of FFT bins used for DDM. As expected, the TF regions without any sinusoidal content exhibit random frequency trajectory estimations, whereas in the parts containing the bird chirp the trajectories align and are clearly correlated. Such representations are well suited for regression shrinkage algorithms to further enforce sparsity (Siedenburg and Drfler, 2011). Using only 2 frequency bins causes inconclusive results for the initial part with extreme FM, while using 32 bins largely solves the problem. The figures 10.3-10.6 have proved that using a larger number of FFT bins for DDM successfully resolves all the parts of

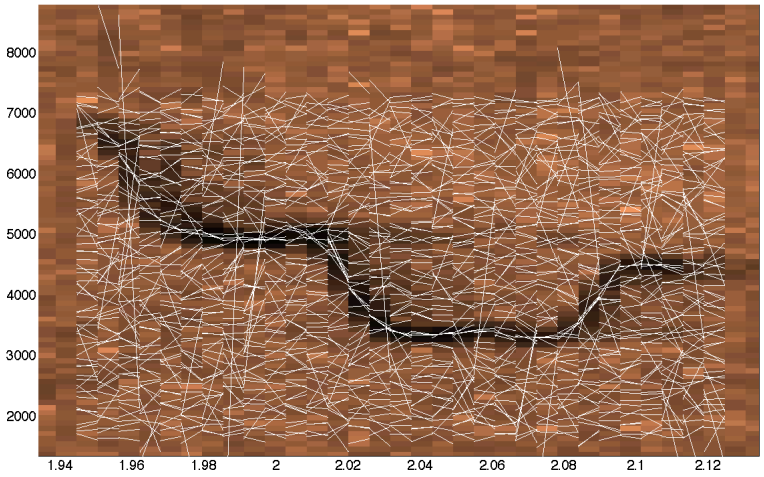


Figure 10.3: Bird chirp: DDM, Q=2

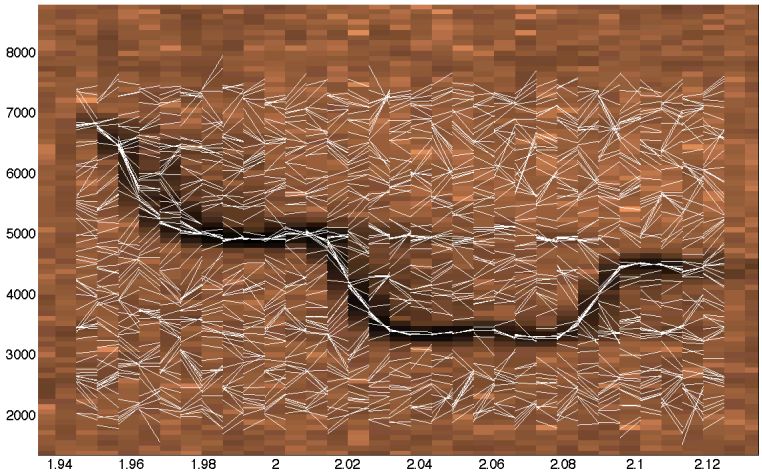


Figure 10.4: Bird chirp: DDM, Q=8

selected bird chirp sound. For the plots the amplitude estimate has been discarded to asses the structure of the frequency trajectory field. Introducing the amplitude estimate results in, what is many times referred to as a *sparse representation*, higher amplitude trajectories are plotted with darker colours. Figures 10.7 and 10.8 clearly show how using 16 FFT bins DDM significantly improves the sparsity and reduces ambiguity. The discussion in this chapter has been focused on a single bird syllabus, comprised of

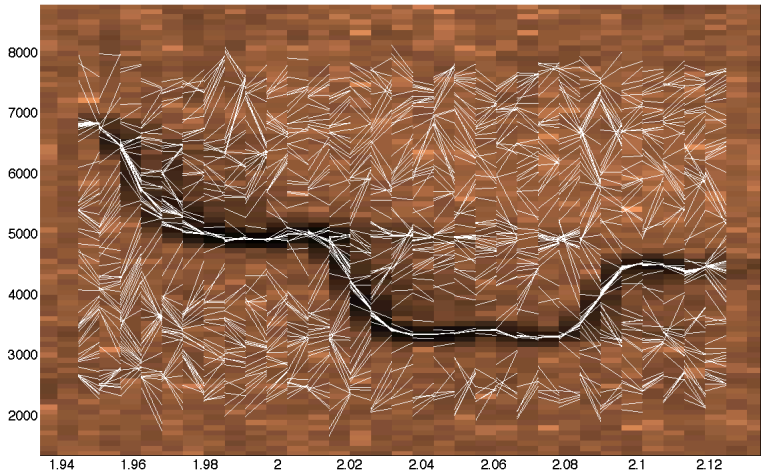


Figure 10.5: Bird chirp: DDM, Q=16

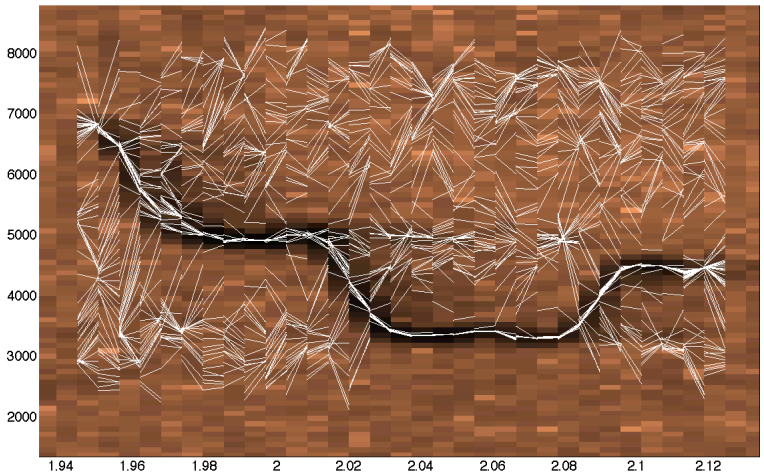


Figure 10.6: Bird chirp: DDM, Q=32

a single modulated sinusoid. The technique proposed has been, however, successfully applied to improve multiple birdsong tracking using Markov renewal process clustering to separate syllabus and segregate event streams (Stowell et al., 2013).

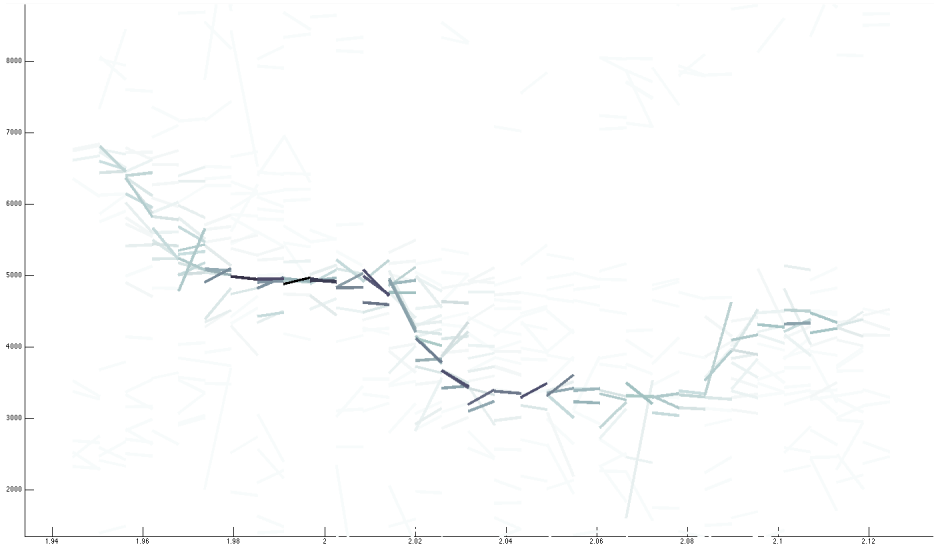


Figure 10.7: Bird chirp: DDM, Q=2, sparse representation

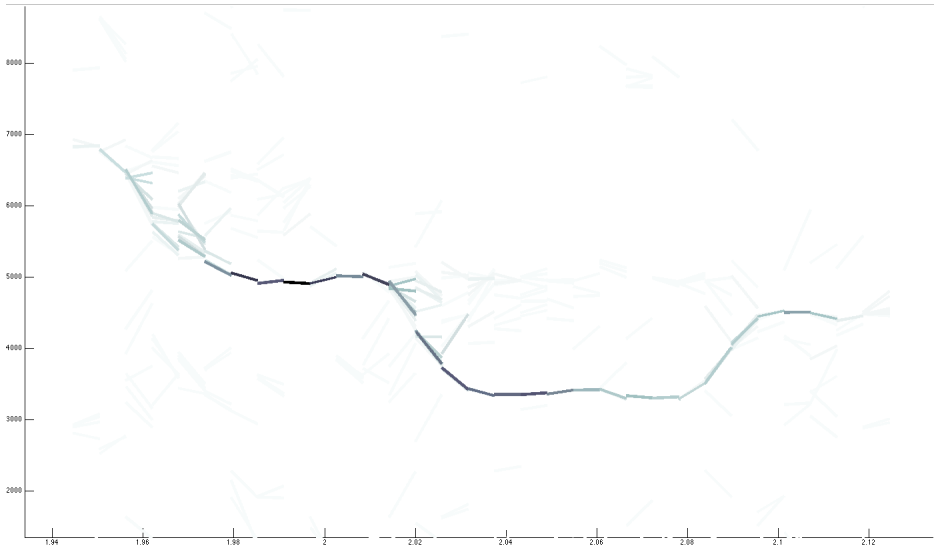


Figure 10.8: Bird chirp: DDM, Q=16, sparse representation

Part IV

Conclusion

Chapter 11

Conclusion

All the algorithms considered in this thesis make use of a kernel function and derive estimates using the inner product of the kernel with the signal under study. Such procedure inherently limits the frequency resolution, especially when non-stationary signals are studied, thus the non-stationary analysis methods play an important role in at least partly alleviating this restriction.
All but one algorithm described in this thesis use the signal time derivative, since it can be conveniently expressed as a linear combination of the modulation functions and the original signal in the case of generalised sinusoid. For the polynomial amplitude, the high order derivatives cancel out the effect of AM, and a simple pole estimator can be obtained. It is safe to say that the use of the signal derivative has had a major impact on non-stationary sinusoidal analysis. In the current context, the only way to avoid using the signal derivative is by considering the direct analytical expression for FT, which only exists for the polynomial amplitude model. Please note that the use of the inner product of a kernel function with the signal is not avoided even in this case, thus the underlying resolution is only increased for the non-stationary sinusoids, reasonably isolated in the TF domain.
To reach beyond the resolution limitation set by the inner product, one has to abandon it, forcing the calculations to be done in the time domain. The so called high-resolution methods, based on the matrix pencil and rotational invariance methods offer one such alternative (section 2.3). Unfortunately, a significant increase in the CPU complexity compared to the kernel based methods (up to 2 degrees magnitude for a single sinusoid, see section 8.4). Importantly, these methods were not designed for the analysis

of non-stationary sinusoids, but rather a mixture of stationary ones. Fortunately, the extension is trivial, a simple consequence to a classic result about algebraic functions, which can however, cause certain ambiguity.
It unlikely that kernel based sinusoidal analysis methods would ever reach the resolution of the high-resolution methods. Certain techniques allow for the estimation of stationary sinusoids with very similar frequencies (see chapter 7), effectively increasing the resolution, however the parameter estimation of a non-stationary sinusoid pair, be it generalised or polynomial amplitude sinusoid, remains an elusive problem. If achieved, it would effectively double the resolution of non-stationary analysis methods greatly diminishing the advantage of high-resolution methods.
The first sinusoidal models were assuming two significant parts: a sinusoidal and noise part, sometimes referred to as the deterministic and stochastic part. It has soon become clear that certain sonically extremely relevant classes of signals cannot be easily classified as either. This new class of signals was named *transients*, signifying a usually short lived transitional state from stochastic to deterministic or vice versa. Such phenomena is easily observed the very early moments of a vibrating object, to which a sudden force has been applied. Such signals represent a very important class of sounds called *percussive* sounds - great amount of effort has been dedicated to its analysis. Apart from adaptive TF kernel based analysis, the transient analysis has largely been tackled by brute-force type approaches, resulting in substantially higher CPU requirements. With non-stationary analysis methods developing rapidly in the last years, signals with very rapidly changing amplitude can accurately be resolved (chapter 5), arriving very close to the generally loosely defined notion of transients. Naturally, an interesting question of the difference between transients and highly modulated sinusoids arises. In practice, it is fairly easy to adequately approximate a transient signal with a DDM or GRM, however a small but significant detailed remains unsolved. Both models, the generalised and polynomial amplitude sinusoid can describe a very significant drops of amplitude to near zero values, but cannot describe a signal that is exactly zero for some extended amount of time - which is possibly the only widely accepted and well specified feature of a transient sound. An amplitude modulation involving the step function cannot be modelled by any of the methods presented in this thesis directly. This comes as no surprise, as the step function isn't a very descriptive by itself. In addition, it exhibits a discontinuity, while it's derivative reaches infinitum at the same time instant, causing all sorts of difficulties in mathematical derivations. However, the recent advancement that introduced the hybrid model (see chapter 9) has severely widened the

class of signals that can be efficiently analysed, opening a possibility to successfully analyse even an amplitude modulation involving a step function.

11.1 Contributions

This dissertation significantly contributes to the field of non-stationary kernel based sinusoidal modelling. Several algorithms for non-stationary sinusoidal analysis have been unified, thoroughly tested, compared, improved and advanced - new relevant algorithms have been proposed.
In chapter 3 the state-of-the-art algorithms (GRM, DDM, DAM) using signal derivatives are presented, the section 4.1 unifies the GRM and DAM methods for any modulation degree. This crucial step has resulted in eventual unification with DDM (see section 3.5) to construct a GRM-DDM-DAM hybrid method. As shown in chapter 4, the DDM exhibits advantageous properties, thus the use of the hybrid method does not offer any practical advantage. Nonetheless, the idea of arbitrarily constructed linear system relaxes previously assumed restrictions and allows for new, more flexible algorithms.
Chapter 4 presents extensive comparison between GRM, DDM and DAM. It is shown that DDM clearly outperforms the rest in all the cases. Such decision can be drawn since DDM exhibits superior numerical stability and flexibility while slightly outperforming in parameter estimate accuracy.
An improvement of the GRM for analysis of extremely modulated, transient like sinusoids is proposed in chapter 5. The technique employed is shown to remove the unwanted effect of AM/FM and successfully estimate even the highest AM/FM sinusoids.
A novel estimator for the polynomial amplitude model is presented in chapter 6. Complex analytical expressions for FT of the model are manipulated using symbolic computing software. In contrast with DDM and GRM, the algorithm proposed uses multiple window functions, allowing for very low required bandwidth. Despite highly complex FT expressions, the pole estimates are shown to be roots of a sufficiently low degree polynomial for the case when a specific set of window functions is used.
Chapter 7 exploits the theory of computational algebraic geometry as a tool for solving a multivariate polynomial systems arising when converting a polynomial amplitude to the generalised sinusoid model. Various nonlinear systems can be constructed and solved, including stationary sinusoids with very close frequencies. The ability to resolve stationary sinusoids with

frequency difference bellow 1 FFT bin has not been observed even in high-resolution methods.
A high order derivative of the signal can be used to eliminate the effect of any existing polynomial amplitude and thus can be used as a pole estimator. In chapter 8, the 3^{rd} order of aforementioned method, reminiscent of GRM is compared to the high-resolution method, exhibiting very comparable results, however for analysis of a single sinusoid the proposed method is 2 magnitudes faster.
An efficient, novel algorithm for a very flexible sinusoidal model, a hybrid between generalised and polynomial amplitude sinusoid, is presented in chapter 9. Such models have only been tackled using approximate, iterative improvement type algorithms and a fast, direct algorithms is designed using techniques derived from the ideas used in solving multivariate polynomial systems, largely used in chapter 7.
Finally, an arbitrarily overdetermined DDM is successfully used in a practical application for the analysis of a birdsong. Modulations, significantly exceeding even those tackled in chapter 5 are quite often found in bird chirps. Chapter 10 briefly demonstrates the flexibility of DDM can be used to construct a favourable TF representation, improving the performance of high-level machine learning algorithms for bird chirp clustering and identification.

11.2 Future Work

The future of non-stationary sinusoidal analysis represents some significant challenges. Importantly more research is required to achieve significant benefits of non-stationary methods in practice. The following topics should receive a great deal of attention:

- joint estimation
- transient analysis
- overlapping non-stationary sinusiods

Joint estimation of non-stationary sinusoids has been researched (Friedlander and Francos, 1993a; Pantazis et al., 2011) , however algorithms less reliant on statistical modelling and iterative improvement would be greatly appreciated. A direct, possible LS based algorithm for the joint estimation of a mixture of generalised sinusoids has not yet been proposed.s

Further, a robust and fast algorithm for the analysis of transients has also not yet been proposed. Iterative approaches (Christensen and van de Par, 2006) have been proposed, however in chapter 9 a big step towards a fast FFT based algorithm has been made.
Last but not least, the ever elusive problem of overlapping non-stationary sinusoids could be tackled by multivariate polynomial system solving algorithms. An appropriate representation of the problem as a multivariate system would enable similar procedure used in chapter 9 to derive a closed form estimator using only signal derivatives.

Part V

Appendix

Appendix A

Proofs of equations

A.1 Reassignment

In this appendix a formal proofs of the integral equations used by the reassignment method are presented. Please note that the following proofs differ from the original derivations in (Kodera et al., 1978) as they are, by author's opinion, a simpler and more compact representations (Hainsworth et al., 2001; Hamilton and Depalle, 2012a,b; Muševič and Bonada, 2010b; Marchand, 2008). Expression for frequency reassignment can be obtained directly from 1.4:

$$\hat{\omega}(t,\omega) = \frac{\partial}{\partial t}\phi(t,\omega) = \Im\left(\frac{\partial}{\partial t}\log\left(S_w(t,\omega)\right)\right) \tag{A.1}$$

Polynomial model is implicitly assumed: *initial phase* is the *static* phase parameter and frequency is the *first order* phase parameter. It may seem counter-intuitive to think of a frequency as function of time and frequency itself, but note, that $\hat{\omega}$ represents the *reassigned* frequency. The same procedure can be used on spectral amplitude function to derive the expression for linear log-AM:

$$\hat{\mu}(t,\omega) = \frac{\partial}{\partial t}\lambda(t,\omega) = \Re\left(\frac{\partial}{\partial t}\log\left(S_w(t,\omega)\right)\right) \tag{A.2}$$

Again, polynomial model is assumed. Luckily, above time derivatives are easily computable:

$$\frac{\partial}{\partial t}\log S_w(t,\omega) = \frac{\frac{\partial}{\partial t}S_w(t,\omega)}{S_w(t,\omega)} \tag{A.3}$$

$$\frac{\partial}{\partial t}S_w(t,\omega) = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} s(\tau)w(\tau-t)\exp(-j\omega(\tau-t))d\tau = \tag{A.4}$$

$$\int_{-\infty}^{\infty} s(\tau)\overbrace{\frac{\partial}{\partial t}\left(w(\tau-t)\right)}^{w'(\tau-t)}e^{-j\omega(\tau-t)}d\tau + \int_{-\infty}^{\infty} s(\tau)w(\tau-t)\overbrace{\frac{\partial}{\partial t}\left(e^{-j\omega(\tau-t)}\right)}^{-j\omega e^{-j\omega(\tau-t)}}d\tau = \tag{A.5}$$

$$-S_{w'}(t,\omega) + j\omega S_w(t,\omega) \Rightarrow \tag{A.6}$$

$$\frac{\partial}{\partial t}\log S_w(t,\omega) = j\omega - \frac{S_{w'}(t,\omega)}{S_w(t,\omega)} \Rightarrow \tag{A.7}$$

$$\hat{\omega}(t,\omega) = \omega - \Im\left(\frac{S_{w'}(t,\omega)}{S_w(t,\omega)}\right) \tag{A.8}$$

$$\hat{\mu}(t,\omega) = -\Re\left(\frac{S_{w'}(t,\omega)}{S_w(t,\omega)}\right), \tag{A.9}$$

where $S_{w'}$ is STFT using time derivative of the window function, rather than the original window. Without going into further details of (Kodera et al., 1978) and (Auger and Flandrin, 1995b) we state, that above procedure reassigns our parameter estimations from time instant t to a new, reassigned

time instant $\hat{t}$:

$$\hat{t}(t,\omega) = t - \frac{\partial}{\partial\omega}\phi(t,\omega) \tag{A.10}$$

$$= t - \Im(\frac{\partial}{\partial\omega}\log S_w(t,\omega)) = t - \Im\left(\frac{\frac{\partial}{\partial\omega}S_w(t,\omega)}{S_w(t,\omega)}\right) \tag{A.11}$$

$$\frac{\partial}{\partial\omega}S_w(t,\omega) = \frac{\partial}{\partial\omega}\int_{-\infty}^{\infty} s(\tau)w(\tau-t)\exp(-j\omega(\tau-t))d\tau \tag{A.12}$$

$$= \int_{-\infty}^{\infty} s(\tau)w(\tau-t)\overbrace{\frac{\partial}{\partial\omega}\left(e^{-j\omega(\tau-t)}\right)}^{-j(\tau-t)e^{-j\omega(\tau-t)}}d\tau \tag{A.13}$$

$$= -jS_{tw}(t,\omega) \Rightarrow \tag{A.14}$$

$$\Im\left(\frac{\frac{\partial}{\partial\omega}S_w(t,\omega)}{S_w(t,\omega)}\right) = -\Re\left(\frac{S_{tw}(t,\omega)}{S_w(t,\omega)}\right) \Rightarrow \tag{A.15}$$

$$\hat{t}(t,\omega) = t - \Re\left(\frac{S_{tw}(t,\omega)}{S_w(t,\omega)}\right), \tag{A.16}$$

where S_{tw} is STFT computed using *time ramped* window function rather than original one. Therefore in practice, the parameters are estimated not at desired time, but reassigned time. Some additional steps are needed to bring them back to desired time instant. That is, the time difference $t - \hat{t}$ has to be accounted for, but only for parameters which are considered to change with time. In above case, log-AM/FM do not change in time, but log-amplitude, initial phase and frequency do. Because the polynomial model was implied and the time difference is known to be $\Delta t = t - \hat{t}$, the following corrections can be made:

$$\hat{\mu}_0^R = \hat{\mu} \tag{A.17}$$

$$\hat{a}_0^R = \hat{a} + \hat{\mu}\Delta t \tag{A.18}$$

$$\hat{\psi}_0^R = \hat{\psi} \tag{A.19}$$

$$\hat{\omega}_0^R = \omega + \hat{\psi}\Delta t \tag{A.20}$$

$$\hat{\phi}_0^R = \hat{\phi} + \omega\Delta t + \frac{\hat{\psi}}{2}\Delta t^2 \tag{A.21}$$

$$\text{where } \Delta t = t - \hat{t} = \Re\left(\frac{S_{tw}(t,\omega)}{S_w(t,\omega)}\right) \tag{A.22}$$

Above is only possible, when all estimates are made, however as Marchand states in (Marchand, 2008) that time difference Δt is generally smaller than

sampling rate except for very low SNRs. In such cases, noise causes estimation errors much larger than those caused if we simply ignore the time difference. Therefore, the above corrections were not covered in the tests. In order to express linear frequency modulation in terms of STFT of the signal, it's *reassigned* time derivative has to be derived:

$$\frac{\partial\hat{\omega}}{\partial\hat{t}} = \frac{\frac{\partial\hat{\omega}}{\partial t}}{\frac{\partial\hat{t}}{\partial t}} \tag{A.23}$$

$$\frac{\partial\hat{\omega}}{\partial t} = \overbrace{\frac{\partial\omega}{\partial t}}^{0} - \Im\left(\frac{\frac{\partial}{\partial t}S_{w'}S_w - S_{w'}\frac{\partial}{\partial t}S_w}{S_w^2}\right) \tag{A.24}$$

$$\frac{\partial\hat{t}}{\partial t} = \overbrace{\frac{\partial t}{\partial t}}^{1} + \Re\left(\frac{\frac{\partial}{\partial t}S_{tw}S_w - S_{tw}\frac{\partial}{\partial t}S_w}{S_{w^2}}\right) \tag{A.25}$$

To complete above equation, time derivative of STFT using time ramped window (S_{tw}) has to be computed:

$$\frac{\partial}{\partial t}S_{wt} = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} s(\tau)(\tau-t)w(\tau-t)e^{-j\omega(\tau-t)}d\tau \tag{A.26}$$

$$= \int_{-\infty}^{\infty} s(\tau)\frac{\partial}{\partial t}\left((\tau-t)w(\tau-t)e^{-j\omega(\tau-t)}\right)d\tau \tag{A.27}$$

$$= \int_{-\infty}^{\infty} s(\tau)\frac{\partial}{\partial t}\left((\tau-t)w(\tau-t)\right)e^{-j\omega(\tau-t)}d\tau + \tag{A.28}$$

$$\int_{-\infty}^{\infty} s(\tau)(\tau-t)w(\tau-t)\overbrace{\frac{\partial}{\partial t}\left(e^{-j\omega(\tau-t)}\right)}^{j\omega e^{-j\omega(\tau-t)}}d\tau \tag{A.29}$$

$$= \int_{-\infty}^{\infty} s(\tau)\overbrace{\frac{\partial}{\partial t}\left[(\tau-t)\right]}^{-1}w(\tau-t)e^{-j\omega(\tau-t)}d\tau + \tag{A.30}$$

$$\int_{-\infty}^{\infty} s(\tau)(\tau-t)\overbrace{\frac{\partial}{\partial t}\left[w(\tau-t)\right]}^{-w'(\tau-t)}e^{-j\omega(\tau-t)}d\tau + j\omega S_{tw} \tag{A.31}$$

$$= -S_w - S_{tw'} + j\omega S_{tw} \tag{A.32}$$

It is easy to generalize A.32 and A.16 to get higher order time and frequency derivatives of STFT using window derivatives, multiplied by t^l:

$$\frac{\partial}{\partial t} S_{t^l w^{(k)}} = -l S_{t^{l-1} w^{(k)}} - S_{t^l w^{(k+1)}} + j\omega S_{t^l w^{(k)}} \tag{A.33}$$

$$\frac{\partial}{\partial \omega} S_{t^l w^{(k)}} = -j S_{t^{l+1} w^{(k)}} \tag{A.34}$$

Terms $S_{t^l w^{(k)}}$ involved in above equation should not be hard to compute, as window function is known in advance and can be chosen in such a way it exhibits desired properties despite high order derivatives and multiplication with t^l. However, overall effect of window function properties and properties of it's derivatives on parameter estimation might be questionable.
We can use A.33 and A.34 to finalize A.24 and A.25 to finally get expression for linear frequency modulation:

$$\frac{\partial \hat{\omega}}{\partial t} = \Im\left(\frac{S_{w''}}{S_w}\right) - \Im\left(\left(\frac{S_{w'}}{S_w}\right)^2\right) \tag{A.35}$$

$$\frac{\partial \hat{t}}{\partial t} = \Re\left(\frac{S_{w'} S_{tw}}{S_w^2}\right) - \Re\left(\frac{S_{tw'}}{S_w}\right) \tag{A.36}$$

$$\frac{\partial \hat{\omega}}{\partial \hat{t}} = \frac{\frac{\partial \hat{\omega}}{\partial t}}{\frac{\partial \hat{t}}{\partial t}} = \frac{\Im\left(\frac{S_{w''}}{S_w}\right) - \Im\left(\left(\frac{S_{w'}}{S_w}\right)^2\right)}{\Re\left(\frac{S_{w'} S_{tw}}{S_w^2}\right) - \Re\left(\frac{S_{tw'}}{S_w}\right)} \tag{A.37}$$

Above expressions give estimations for linear log-AM and FM parameters. However, with help of A.34 and A.33 the expressions for $\frac{\partial^k \phi}{\partial t^k}$ and $\frac{\partial^k \lambda}{\partial t^k}$ for arbitrary k should eventually be derived. Such procedure might result in very long and complicated expressions, but could generally be done in practice with use of some symbolic programming techniques.

A.2 Derivative method

Presented derivation of algorithm is slightly different than in Marchand and Depalle (2008), but it arrives at the same expressions for parameter. The algorithm exploits the fact, that a time derivative of complex exponential is still a complex exponential with the same frequency. Consider the generally

modulated signal from 3.1:

$$s'(t) = (p'(t) + jq'(t))s(t) \Rightarrow \tag{A.38}$$

$$\Im\left(\frac{s'(t)}{s(t)}\right) = q'(t) \tag{A.39}$$

$$\Re\left(\frac{s'(t)}{s(t)}\right) = p'(t) \tag{A.40}$$

Taking STFT of both sides results in:

$$\begin{aligned} S'_w(t,\omega) &= \int_{-\infty}^{\infty} w(\tau - t)(a'(\tau) + j\varphi'(\tau))s(\tau)e^{-j\omega(\tau - t)}d\tau \\ &= \mathfrak{F}\left[a'(t) + j\varphi'(t)\right] * S_w(t,\omega), \end{aligned} \tag{A.41}$$

where S'_w signifies a STFT of a time derivative of the signal, rather than original signal. In case when $a(t)$ an $i\varphi(t)$ are polynomials, above expression changes to:

$$\begin{aligned} S'_w(t,\omega) &= \sum_{k=1}^{K-1} k(p_k + jq_k)\mathfrak{F}\left[t^{k-1}\right] * S_w(t,\omega) \\ &= \sum_{k=1}^{K-1} k(p_k + jq_k)(j)^k\delta^{(k-1)}(\omega) * S_w(t,\omega) \\ &= \sum_{k=1}^{K-1} k(p_k + jq_k)(j)^k\frac{\partial^{k-1}}{\partial\omega^{k-1}}S_w(t,\omega) \\ &= \sum_{k=1}^{K-1} k(p_k + jq_k)\overbrace{(j)^{k-1}(-j)^{k-1}}^{1}S_{wt^{k-1}}(t,\omega) \\ &= \sum_{k=1}^{K-1} k(p_k + jq_k)S_{wt^{k-1}}(t,\omega) \end{aligned} \tag{A.42}$$

Evidently, the STFT of time derivative of the non-stationary sinusoid is a linear combination of frequency derivatives of STFT of original signal. In case of linear log-AM/FM model 3.4, a specific case of A.42 for $K = 3$ and $p_2 = 0$:

$$S'_w(t,\omega) = \big(\mu_0 S_w(t,\omega) + j\omega_0 S_w(t,\omega)\big) + 2j\psi_0 S_{wt}(t,\omega) \tag{A.43}$$

In practice, first approximation of the frequency is made by locating a peak in magnitude spectrum, that is an extrema in $\|S_w\| = \sqrt{S_w S_w^*}$. This extrema intuitively corresponds to centre of energy in frequency domain. Further, derivative of magnitude spectrum in respect to frequency is nil at the peak location. It is crucial to investigate behaviour of the frequency derivatives of real/imaginary parts of the magnitude spectrum separately:

$$\|S_w\| = \sqrt{S_w S_w^*} = \sqrt{\Re(S_w)^2 + \Im(S_w)^2} \Rightarrow \tag{A.44}$$

$$\frac{\partial \|S_w\|}{\partial \omega} = \frac{1}{\sqrt{\Re(S_w)^2 + \Im(S_w)^2}} \left(2\Re(S_w) \frac{\partial}{\partial \omega} \Re(S_w) + 2\Im(S_w) \frac{\partial}{\partial \omega} \Im(S_w) \right) \tag{A.45}$$

It would be very difficult to derive an analytic expression for above terms even in case of linear log-AM/FM. However, some helpful assumptions about behaviour of some terms in above expression can be made. Magnitude spectrum reaches very large values around sinusoid frequency ω_0, thus $\sqrt{\Re(S_w)^2 + \Im(S_w)^2}$ will not be zero around ω_0. Further, it is safe to assume that generally both $\Re(S_w)$ and $\Im(S_w)$ will not reach zero at ω_0. In practice it can be observed, that ratio between values of real/imaginary part depends on the instantaneous phase. In fact, the commonly used expression for instantaneous phase $\angle(S_w(t,\omega)) = arctan\left(\frac{\Im(S_w(t,\omega))}{\Re(S_w(t,\omega))}\right)$ confirms this assumption, e.g.: it is a function of ratio between imaginary and real parts of FT spectrum. Thus it can be written:

$$\frac{\partial \|S_w(t,\omega_0)\|}{\partial \omega} = 0 \Leftrightarrow \tag{A.46}$$

$$\frac{\partial}{\partial \omega} \Re(S_w(t,\omega_0)) = \frac{\partial}{\partial \omega} \Im(S_w(t,\omega_0)) = 0 \tag{A.47}$$

This assumption should be taken with precaution: at some very specific phase values either of $\Re(S_w(t,\omega_0)), \Im(S_w(t,\omega_0))$ could be 0, which in turn would not make previous conclusion necessary. It will however be assumed, that even in such cases A.47 holds. This fact has an useful effect:

$$\frac{\partial}{\partial \omega} \Re(S_w(t,\omega_0)) = \frac{\partial}{\partial \omega} \Im(S_w(t,\omega_0)) = 0 \Rightarrow \tag{A.48}$$

$$\frac{\partial S_w(t,\omega_0)}{\partial \omega} = \frac{\partial}{\partial \omega} \Re(S_w(t,\omega_0)) + j \frac{\partial}{\partial \omega} \Im(S_w(t,\omega_0)) = 0 \Rightarrow \tag{A.49}$$

$$\frac{\partial S_w(t,\omega_0)}{\partial \omega} = S_{wt}(t,\omega_0) = 0 \tag{A.50}$$

This finally gives the expression for linear AM and frequency estimates of sinusoid using A.43:

$$\hat{\omega}_0(t,\omega) = \Im\left(\frac{S'_w(t,\omega)}{S_w(t,\omega)}\right) \tag{A.51}$$

$$\hat{\mu}_0(t,\omega) = \Re\left(\frac{S'_w(t,\omega)}{S_w(t,\omega)}\right) \tag{A.52}$$

To derive linear FM estimation expression, 2^{nd} derivative of the signal needs to be computed:

$$s''(t) = (\mu_0^2 - \omega_0^2 - 2\omega_0\psi_0 t - \psi_0^2 t^2 + j(\psi_0 + 2\mu_0\omega_0 + 2\mu_0\psi_0 t))s(t) \tag{A.53}$$

Using same assumptions used to derive A.47, we can express linear frequency modulation as follows:

$$\hat{\psi}_0(t,\omega) = \Im\left(\frac{S''_w(t,\omega)}{S_w(t,\omega)}\right) - 2\hat{\omega}_0(t,\omega)\hat{\mu}_0(t,\omega) \tag{A.54}$$

Accuracy of estimates relies greatly on equation A.43, particularly on the last term $S_{wt}(t,\omega)$, which is assumed to be zero at the frequency of the sinusoid ω_0. This was in fact concluded in A.47. However $S_{wt}(t,\omega)$ depends on our initial frequency estimate $\tilde{\omega}$ therefore $S_{wt}(t,\tilde{\omega})$ does not reach exact 0 at least for this reason.

A.3 Generalized reassignment

Generalized method described in (Wen and Sandler, 2009) essentially uses signal derivatives to derive expressions for estimating signal parameters, thus it is derived from derivative method. However, *integration-by-parts* is exploited to represent STFT of signal time derivative with reassignment operators. This technique is used to Such procedure poses some restrictions to the window function, but avoids computing signal derivatives, which is indeed a computationally costly and erroneous operation. Further, the method can be used recursively to estimate higher order modulations. An arbitrary order time derivative of 3.1 can be expressed:

$$s^{(l)}(t) = (r'(t)s(t))^{(l-1)} = \sum_{k=1}^{K}\Big(s(t)\big(p_k + jq_k\big)kt^{k-1}\Big)^{(l-1)} \tag{A.55}$$

After applying STFT to both sides and assuming t=0 we get:

$$S_w^{(l)}(0,\omega) = \sum_{k=1}^{K} \int_{-\frac{T}{2}}^{\frac{T}{2}} w(\tau)\Big(\big(p_k + jq_k\big)k\tau^{k-1}s(\tau)\Big)^{(l-1)} e^{-j\omega\tau} d\tau \tag{A.56}$$

For $l = 1$ above expression is trivially computable. However, for $k > 1$ the procedure is not straight forward if computation of signal derivatives is to be avoided. Further, it will be shown that an arbitrary window derivative is needed for such recursive operation, therefore $S_w^{(k)}$ is swapped with $S_{w^{(l)}}^{(k)}$ in the left-hand side of A.56, to get more general expression. Using integration-by-parts on it gives:

$$\begin{aligned}
S_{w^{(k)}}^{(l)} &= \int_{-\frac{T}{2}}^{\frac{T}{2}} \overbrace{\frac{\partial}{\partial\tau}s^{(l-1)}(\tau)}^{\partial v}\overbrace{w^{(k)}(\tau)e^{-j\omega\tau}d\,\tau}^{u} \\
&= s^{(l-1)}(\tau)w^{(k)}(\tau)e^{-j\omega\tau}|_{-\frac{T}{2}}^{\frac{T}{2}} - \\
&\qquad \int_{-\frac{T}{2}}^{\frac{T}{2}} s^{(l-1)}(\tau)\big(w^{(k+1)}(\tau)e^{-j\omega\tau} - j\omega w^{(k)}(\tau)e^{-j\omega\tau}\big)d\tau \\
&= \overbrace{s^{(l-1)}(\tau)w^{(k)}(\tau)e^{-j\omega\tau}|_{-\frac{T}{2}}^{\frac{T}{2}}}^{=0 \text{ if } w^{(k)}(-\frac{T}{2})=w^{(k)}(\frac{T}{2})=0} - S_{w^{(k+1)}}^{(l-1)} + j\omega S_{w^{(k)}}^{(l-1)} \\
&= -S_{w^{(k+1)}}^{(l-1)} + j\omega S_{w^{(k)}}^{(l-1)}
\end{aligned} \tag{A.57}$$

The first term in A.57 must be eliminated, if signal derivatives are to be avoided. Luckily, the only restriction is to choose a window function, whose desired order time derivative reaches 0 at it's both borders, eg: $w^{(k)}(-\frac{T}{2}) = w^{(k)}(\frac{T}{2}) = 0$. Above rule can be generalized for any signal. For the sake of clarity it is rewritten in more generic form, for signal x(t):

$$X_{w^{(k)}}^{(l)} = -X_{w^{(k+1)}}^{(l-1)} + j\omega X_{w^{(k)}}^{(l-1)} \tag{A.58}$$

For left/right hand side of A.56 substituting $x(t) = s(t)$ and $x(t) = s(t)r'(t)$ will give a system of complex linear equations. A very efficient algorithm can be designed, based on equation A.58 using *pyramid* like data flow, as described in section 3.1 of Wen and Sandler (2009). From the complex equation A.56 with the use of A.58 it is possible to derive a system of equations, linear in respect to r_k by letting the degree of the derivative l go up to the desired modulation degree K, constructing $K - 1$ equations

from which all parameters except r_0 (log-amplitude and phase) which can be eventually estimated in some manner.
If the model of generalized method is assumed to be polynomial, the degree of log-amplitude polynomial is 1 and the degree of phase polynomial is 2, this exactly corresponds to the assumptions, implicitly taken in the original reassignment and derivative methods. For such low degrees, we can solve the system of linear equations defined by A.56:

$$r_2 = p_2 + jq_2 = \gamma_0 + j\psi_0 = \frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}} \Rightarrow \tag{A.59}$$

$$p_2 = \gamma_0 = \Re\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right) \tag{A.60}$$

$$q_2 = \psi_0 = \Im\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right), \tag{A.61}$$

where γ_0 denotes second order log-AM. There is a slight difference between classic reassignment and above expression, even if $p_2 = \gamma_0 = 0$ is assumed. Generally, any assumed model does not fit the signal under inspection exactly and since ψ_0 is assumed to be a real number, only the imaginary part of right hand side of A.59 is considered. If the signal fits well the assumed model, γ_0 will be nil anyway, but it does not affect the expression A.61. After evaluating linear FM parameter, the result can be used to solve the following complex equation in order to derive expressions for frequency ($q_1 = \omega_0$) and linear log-AM ($p_1 = \mu_0$):

$$p_1 + jq_1 = -jq_2\frac{S_{tw}}{S_w} + j\omega - \frac{S_{w'}}{S_w} \Rightarrow \tag{A.62}$$

$$q_1 = \omega_0 = \omega - \Im\left(\frac{S_{w'}}{S_w}\right) - q_2\Re\left(\frac{S_{tw}}{S_w}\right) \tag{A.63}$$

$$p_1 = \mu_0 = -\Re\left(\frac{S_{w'}}{S_w}\right) + q_2\Im\left(\frac{S_{tw}}{S_w}\right) \tag{A.64}$$

In order to compare above expressions to reassignment parameter estimates, they should be rewritten with additional notations: 'RM' for reassignment

method and 'GRM' for generalized method.

$$\psi_0^{GRM} = \Im\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right) \tag{A.65}$$

$$\psi_0^{RM} = \frac{\Im\left(\frac{S_{w''}}{S_w}\right) - \Im\left(\frac{S_{w'}^2}{S_w^2}\right)}{\Re\left(\frac{S_{tw} S_{w'}}{S_w^2}\right) - \Re\left(\frac{S_{tw'}}{S_w}\right)} \tag{A.66}$$

$$\omega_0^{RM} = \omega - \Im\left(\frac{S_{w'}}{\S w}\right) \tag{A.67}$$

$$\omega_0^{GRM} = \overbrace{\omega - \Im\left(\frac{S_{w'}}{S_w}\right)}^{\omega_0^{RM}} - \psi_0^{GRM} \overbrace{\Re\left(\frac{S_{tw}}{S_w}\right)}^{t_\Delta} = \omega_0^{RM} - \psi_0^{GRM} t_\Delta \tag{A.68}$$

$$\mu_0^{RM} = -\Re\left(\frac{S_{w'}}{S_w}\right) \tag{A.69}$$

$$\mu_0^{GRM} = \overbrace{-\Re\left(\frac{S_{w'}}{S_w}\right)}^{\mu_0^{RM}} + \psi_0^{GRM} \Im\left(\frac{S_{tw}}{S_w}\right) = \mu_0^{RM} + \psi_0^{GRM} \Im\left(\frac{S_{tw}}{S_w}\right) \tag{A.70}$$

$$\tag{A.71}$$

The time difference t_Δ was defined in A.22. The only difference in frequency estimates is the correction factor, which corresponds exactly to the correction defined by A.20. Linear log-AM is assumed to be constant with time, therefore no correction arising from t_Δ are applied as expected. However another *correction* term $\psi_0^{GRM} \Im\left(\frac{S_{tw}}{S_w}\right)$ is present. Interestingly, linear FM estimation is used in expression for linear log-AM estimation. The other factor $\Im\left(\frac{S_{tw}}{S_w}\right)$ is in fact equal to $\frac{\partial}{\partial\omega}\lambda(t,\omega)$, which can be quickly deduced from A.16.

It was shown, that one difference between generalized and reassignment method are the corrections, defined in equations A.17, A.18, A.19, A.20 and A.21. That is, solving the system of linear equations already implies the corrections. As already mentioned in section A.1, these corrections only boost the accuracy significantly at very high SNRs. Therefore, we expect the generalized method to behave similarly to the reassignment in low to mid SNRs. There is an additional *correction* term entering expression for linear log-AM, which depends on linear FM. It can be considered as a cancellation of the effect the second order FM has on first order log-AM estimate. It is interesting to investigate further how higher degree model estimator

depend on the lower degree estimates and how the lower degree estimates change when the model degree is increased. Since AM/FM effect on spectrum is highly non-linear in terms of energy distribution (at least peak frequency exhibits complex behaviour in respect to AM/FM, (Kodera et al., 1978; Abe and Smith, 2005)). Adding another degree of modulation is thus expected to *update* all lower degree estimates.
The following notations will come handy: the generalized method of degree 1 (stationary frequency, linear log-AM) will be notated as 'GM1' and the one of degree 2 (linear FM, 2^{nd} order log-AM) will be notated as 'GM2'. Generally, the generalized method of degree D will be labeled 'GMD'. First degree expressions will be studied first. From A.56 using the rule from A.58, the linear log-AM and frequency expressions are:

$$S_w = r_1^{GM1} S_{w'} = r_1^{GM1}(-S_{w'} + j\omega S_w) \Rightarrow \tag{A.72}$$

$$r_1^{GM1} = j\omega - \frac{S_{w'}}{S_w} \Rightarrow \tag{A.73}$$

$$p_1^{GM1} = \Re(r_1^{GM1}) = \mu_0^{GM1} = -\Re\left(\frac{S_w}{S_{w'}}\right) \tag{A.74}$$

$$q_1^{GM1} = \Im(r_1^{GM1}) = \omega_0^{GM1} = \omega - \Im\left(\frac{S_w}{S_{w'}}\right) \tag{A.75}$$

The estimates of second order generalized method can now be expressed with above expressions:

$$r_2^{GM2} = p_2^{GM2} + jq_2^{GM2} = \gamma_0^{GM2} + j\psi_0^{GM2} = \frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}} \Rightarrow \quad \text{(A.76)}$$

$$p_2^{GM2} = \gamma_0^{GM2} = \Re\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right) \quad \text{(A.77)}$$

$$q_2^{GM2} = \psi_0^{GM2} = \Im\left(\frac{S_w S_{w''} - S_{w'}^2}{S_{tw} S_{w'} - S_w S_{tw'}}\right) \quad \text{(A.78)}$$

$$r_1^{GM2} = p_1^{GM2} + jq_1^{GM2} = \mu_0^{GM2} + j\omega_0^{GM2} = \quad \text{(A.79)}$$

$$= \overbrace{j\omega - \frac{S_{w'}}{S_w}}^{r_1^{GM1}} - r_2^{GM2}\frac{S_{tw}}{S_w} = r_1^{GM1} - (p_2^{GM2} + jq_2^{GM2})\frac{S_{tw}}{S_w} = \quad \text{(A.80)}$$

$$= r_1^{GM1} - (\gamma_0^{GM2} + j\psi_0^{GM2})\frac{S_{tw}}{S_w} \Rightarrow \quad \text{(A.81)}$$

$$\omega_0^{GM2} = \omega_0^{GM1} - \psi_0^{GM2}\overbrace{\Re\left(\frac{S_{tw}}{S_w}\right)}^{t_\Delta} - \gamma_0^{GM2}\Im\left(\frac{S_{tw}}{S_w}\right) = \quad \text{(A.82)}$$

$$= \omega_0^{GM1} - \psi_0^{GM2} t_\Delta - \gamma_0^{GM2}\Im\left(\frac{S_{tw}}{S_w}\right)$$

$$\mu_0^{GM2} = \mu_0^{GM1} + \psi_0^{GM2}\Im\left(\frac{S_{tw}}{S_w}\right) - \gamma_0^{GM2}\overbrace{\Re\left(\frac{S_{tw}}{S_w}\right)}^{t_\Delta} = \quad \text{(A.83)}$$

$$= \mu_0^{GM1} - \gamma_0^{GM2} t_\Delta + \psi_0^{GM2}\Im\left(\frac{S_{tw}}{S_w}\right)$$

Above derivations suggest that lower degree parameters are indeed affected by higher degree estimates. A straightforward correction arises from the simple time shift (or reassigned time), where lower degree parameter μ_0 is adjusted by the corresponding higher one γ_0 multiplied by the time shift. Clearly, the corrections are more important when high order modulations are of bigger values. It is however important to realise that these corrections do not significantly affect the accuracy of generalised reassignment at mid to low SNR when modulation parameters are in some reasonable range, assuming the analysed signal fits well the underlying model.

A.4 Distribution derivative method

The distribution theory Schwartz (1966) defines the derivative of a distribution with the respect to a test function ψ (also referred to as atom):

$$< s', \psi >= - < s, \psi' >, \tag{A.84}$$

where s is the distribution under study and $< \_, \_ >$ is a L^2 inner product:

$$< x, y >= \int_{-\infty}^{\infty} x(t)\bar{y}(t)dt. \tag{A.85}$$

The time derivative of the model 3.1 can therefore be expressed with a combination of such inner products:

$$< s', \psi >= \sum_{k=1}^{K} r_k < st^{k-1}, \psi >= - < s, \psi' > . \tag{A.86}$$

It is interesting to compare above expression to equation A.55. A special case when $\psi(t) = w(t)\exp(-j\omega t)$ allows comparison of both expressions:

$$\begin{aligned} S'_w(\omega) &=< s'(t), w(t)\exp(-j\omega t) > & \text{(A.87)}\\ &= \sum_{k=1}^{K} kr_k < st^{k-1}, w(t)\exp(-j\omega t) > & \text{(A.88)}\\ &= \sum_{k=1}^{K} kr_k S_{t^{k-1}w}(\omega), & \text{(A.89)} \end{aligned}$$

which is exactly the right hand side of equation A.55 for $l = 1$. With $- < s, \psi' >= -S'_w + j\omega S_w$ a specific equation corresponding to model 3.4 can be obtained:

$$r_1 S_w(\omega) + 2r_2 S_{tw}(\omega) = -S'_w(\omega) + j\omega S_w(\omega). \tag{A.90}$$

It is now possible to use two different Fourier atoms at different frequencies yielding two equations with two unknowns:

$$\begin{aligned} r_1 S_w(\omega_1) + 2r_2 S_{tw}(\omega_1) &= -S'_w(\omega_1) + j\omega_1 S_w(\omega_1) & \text{(A.91)}\\ r_1 S_w(\omega_2) + 2r_2 S_{tw}(\omega_2) &= -S'_w(\omega_2) + j\omega_1 S_w(\omega_2). & \text{(A.92)} \end{aligned}$$

Above system is trivially solvable. The main difference compared to the (generalized) reassignment is that in order to obtain enough equations,

STFTs at different frequencies rather, than higher order derivatives are considered. Such mechanism keeps the equations simple, as well as doesn't pose additional constraints on the window function (as it is the case with the generalized reassignment). The frequencies ω_1, ω_2 can be arbitrarily close, but far apart to avoid numerical problems. A generic system of equations to solve any polynomial degree log-AM/FM is easily deduced from A.87:

$$\mathbf{Ar} = \mathbf{b}, \tag{A.93}$$

where

$$\mathbf{A} = \begin{bmatrix} < s, \psi_1 > & 2 < st, \psi_1 > & \cdots & K < st^{K-1}, \psi_1 > \\ < s, \psi_2 > & 2 < st, \psi_2 > & \cdots & K < st^{K-1}, \psi_2 > \\ \vdots & \vdots & \ddots & \vdots \\ < s, \psi_M > & 2 < st, \psi_M > & \cdots & K < st^{K-1}, \psi_M > \end{bmatrix} \tag{A.94}$$

$$\mathbf{r} = \begin{bmatrix} -S'_w + j\omega S_w \\ \vdots \\ r_{K-1} \end{bmatrix} \tag{A.95}$$

$$\mathbf{b} = \begin{bmatrix} - < s, \psi'_1 > \\ \vdots \\ - < s, \psi'_M > \end{bmatrix}. \tag{A.96}$$

Bibliography

Sage mathematical software. URL http://www.sagemath.org. 85, 108

State-space and singular value decomposition based approximation methods for harmonic retrieval problem. *J. of Opt. Soc. of America*, page 17991811, December 1983. 19

M. Abe and III Smith, J.O. Am/fm rate estimation for time-varying sinusoidal modeling. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, volume 3, pages 201–204, Philadelphia,PA, March 2005. doi: 10.1109/ICASSP.2005.1415681. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1415681. 8, 24, 114, 166

M.F. Aburdene and J.E. Dorband. Parallel computation of discrete legendre transforms. volume 6, pages 3225–3228, Atlanta, GA, May 1996. doi: 10.1109/ICASSP.1996.550563. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=550563. 100, 102

A. C. Aitken. *Determinants and Matrices*. Oliver and Boyd, Edinburgh, UK, 1 edition, 1956. 100

F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. Acoust., Speech, Signal Process.*, 43(5):1068–1089, 1995a. ISSN 1053-587X. doi: 10.1109/78.382394. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=382394. 13

F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *Signal Processing, IEEE Transactions on*, 43(5):1068–1089, 1995b. ISSN 1053-587X. doi: 10.1109/78.382394. 23, 156

F. Auger, E. Chassande-Mottin, and P. Flandrin. On phase-magnitude relationships in the short-time fourier transform. *Signal Processing Letters, IEEE*, 19(5):267–270, 2012. ISSN 1070-9908. doi: 10.1109/LSP.2012.2190279. 13, 14

R. Badeau. Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (hr-nmf). In *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, pages 253–256, 2011. doi: 10.1109/ASPAA.2011.6082264.

R. Badeau, B. David, and G. Richard. High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials. *Signal Processing, IEEE Transactions on*, 54(4):1341–1350, 2006. ISSN 1053-587X. doi: 10.1109/TSP.2006.870556. 3, 6, 20, 89, 95, 121, 124

R. Badeau, B. David, and G. Richard. Cramer-rao bounds for multiple poles and coefficients of quasi-polynomials in colored noise. *Signal Processing, IEEE Transactions on*, 56(8):3458–3467, 2008a. ISSN 1053-587X. doi: 10.1109/TSP.2008.921719. 10, 20

R. Badeau, G. Richard, and B. David. Performance of esprit for estimating mixtures of complex exponentials modulated by polynomials. *Signal Processing, IEEE Transactions on*, 56(2):492–504, 2008b. ISSN 1053-587X. doi: 10.1109/TSP.2007.906744. 20, 89

Roland Badeau and Drémeau Angélique. Variational bayesian em algorithm for modeling mixtures of non-stationary signals in the time-frequency domain (hr-nmf). In *Acoustics, Speech and Signal Processing, 2013. Proceedings of the 2013 IEEE International Conference on*, 2013.

A. Barabell. Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '83.*, volume 8, pages 336–339, 1983. doi: 10.1109/ICASSP.1983.1172124. 19

Maciej Bartkowiak. A complex envelope sinusoidal model for audio coding. In *Proc. of the 10 th Int. Conference on Digital Audio Effects, DAFx-07*, 2007. 21

Bateman Manuscript Project, Harry Bateman, and Arthur Erdélyi. *Higher Transcendental Functions: Volume II*. Dover Publications, Mineola, NY, 2006a. ISBN 048644614X. 99, 101, 102

Bateman Manuscript Project, Harry Bateman, and Arthur Erdélyi. *Higher Transcendental Functions: Volume II*. Dover Publications, Mineola, NY, 2006b. ISBN 048644614X. 99

Bateman Manuscript Project, Harry Bateman, and Arthur ErdÍyi. *Higher*

Transcendental Functions: Volume I. Dover Publications, Mineola, NY, 2006c. ISBN 048644614X. 83

Albert J. Berni. Target identification by natural resonance estimation. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-11(2): 147–154, 1975. ISSN 0018-9251. doi: 10.1109/TAES.1975.308051. 19

M. Betser. Sinusoidal polynomial parameter estimation using the distribution derivative. *IEEE Trans. Signal Process.*, 57(12):4633–4645, December 2009. ISSN 1053-587X. doi: 10.1109/TSP.2009.2027401. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5164904. 3, 26, 28, 56, 75, 76, 86, 114, 123, 128

Arne Bjerhammar. Application of calculus of matrices to method of least squares; with special references to geodetic calculations. *Trans. Roy. Inst. Tech. Stockholm*, (49), 1951. 3, 16

J. Bonada. Wide-band harmonic sinusoidal modeling. In *Proc. 11th Int. Digital Audio Effects*, pages 265–272, Espoo, Finland, September 2008. ISBN 978-951-22-9517-3. URL http://www.acoustics.hut.fi/dafx08/papers/dafx08_46.pdf. 3

P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and Jun Xiao. Testing stationarity with surrogates: A time-frequency approach. *Signal Processing, IEEE Transactions on*, 58(7):3459–3470, 2010. ISSN 1053-587X. doi: 10.1109/TSP.2010.2043971. 4

B. Buchberger. A theoretical basis for the reduction of polynomials to canonical forms. *ACM SIGSAM Bulletin*, 10:19–29, August 1976. ISSN 0163-5824. doi: 10.1145/1088216.1088219. URL http://dl.acm.org/citation.cfm?id=1088216.1088219. 108

J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969. ISSN 0018-9219. doi: 10.1109/PROC.1969.7278. 15

ric Chassande-Mottin, Franois Auger, and Patrick Flandrin. *Reassignment*, pages 249–277. ISTE, 2010. ISBN 9780470611203. doi: 10.1002/9780470611203.ch9. URL http://dx.doi.org/10.1002/9780470611203.ch9. 13

E. W Cheney and David Kincaid. *Numerical Mathematics and Computing*. Brooks/Cole, Belmont, CA, 6 edition, 2007. ISBN 9780495114758. 100

Jinkui Cheng, Bengui Xie, Congtian Lin, and Liqiang Ji. A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features. *Bioacoustics*, 21(2):157–171, June 2012. ISSN 0952-4622, 2165-0586.

doi: 10.1080/09524622.2012.669664. URL http://www.tandfonline.com/doi/abs/10.1080/09524622.2012.669664. 139

M.G. Christensen and S. van de Par. Efficient parametric coding of transients. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(4):1340–1351, 2006. ISSN 1558-7916. doi: 10.1109/TSA.2005.858038. 21, 121, 151

O. Christensen. *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2003. ISBN 9780817642952. URL http://books.google.es/books?id=X92UUiIkvAEC. 27

C.W. Clenshaw and H. Mühlig. Mathematical tables, vol. 5: Chebyshev series for mathematical function. *J. Appl. Math. Mech.*, 43:147, 1963. ISSN 0163-5824. doi: 10.1002/zamm.19630430321. 100, 102

David A Cox, John B Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer NY, New York, NY, 2 edition, 2007. ISBN 9780387356501. 86, 96, 107, 109

Harald Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946. URL http://books.google.com/books?hl=fr&lr=&id=CRTKKaJ0ODYC&pgis=1. 7

I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on*, 36(5):961–1005, 1990. ISSN 0018-9448. doi: 10.1109/18.57199. 27

Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis*, 30(2):243 – 261, 2011. ISSN 1063-5203. doi: 10.1016/j.acha.2010.08.002. URL http://www.sciencedirect.com/science/article/pii/S1063520310001016. 4

G. M. Riche de Prony. Essai exprimental et analytique: sur les lois de la dilatabilit de fluides lastiques et sur celles de la force expansive de la vapeur de leau et de la vapeur de lalcool diffrentes tempratures. *Journal de lcole polytechnique*, 1(22):2476, 1795. 18

M. Desainte-Catherine and S. Marchand. High-precision fourier analysis of sounds using signal derivatives. *J. Audio Eng. Soc.*, 48(7/8):654–667, September 2000. URL http://www.aes.org/e-lib/browse.cfm?elib=12054. 31

Myriam Desainte-Catherine and Sylvain Marchand. High precision fourier analysis of sounds using signal derivatives. *Audio Engineering So-*

ciety, 1953 Journal of, 1998. URL http://citeseer.ist.psu.edu/desainte-catherine98high.html. 25

P.M. Djuric and S.M. Kay. Parameter estimation of chirp signals. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(12):2118–2126, 1990. ISSN 0096-3518. doi: 10.1109/29.61538. 8

F. Y. Edgeworth. On the probable errors of frequency-constants. *Journal of the Royal Statistical Society*, 71(3):pp. 499–512, 1908. ISSN 09528385. URL http://www.jstor.org/stable/2339293. 7

A. Eriksson, P. Stoica, and T. Soderstrom. Second-order properties of music and esprit estimates of sinusoidal frequencies in high snr scenarios. *Radar and Signal Processing, IEE Proceedings F*, 140(4):266–272, 1993. ISSN 0956-375X. 20

H.G. Feichtinger and T. Strohmer. *Gabor analysis and algorithms: theory and applications*. Applied and numerical harmonic analysis. Birkhèauser, 1998. ISBN 9780817639594. URL http://books.google.es/books?id=TF0D4Fyz_rAC. 27

D. Filipovic. Exponential-polynomial families and the term structure of interest rates. *Bernoulli*, 6(6):10811107, 2000. 20

B. R. Fischer and A. Medvedev. L2 time delay estimation by means of laguerre functions. In *American Control Conference, 1999. Proceedings of the 1999*, volume 1, pages 455–459 vol.1, 1999. doi: 10.1109/ACC.1999.782869. 20

B. Friedlander and J. Francos. Estimation of amplitude and phase of non-stationary signals. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 431 –435 vol.1, nov 1993a. doi: 10.1109/ACSSC.1993.342549. 8, 150

Benjamin Friedlander and J. Francos. Estimation of amplitude and phase of non-stationary signals. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 431–435 vol.1, 1993b. doi: 10.1109/ACSSC.1993.342549. 18

Swami A. Guotong Zhou, Giannakis G.B. On polynomial phase signals with time-varying amplitudes. *Signal Processing, IEEE Transactions on*, 44 (4):848 –861, Apr 1996. ISSN 1053-587X. doi: 10.1109/78.492538. 8

S.W. Hainsworth, M.D. Macleod, and P.J. Wolfe. Analysis of reassigned spectrograms for musical transcription. In *IEEE Workshop Applicat. of Signal Process. to Audio and Acoustics*, pages 23–26, New Paltz, NY, October 2001. doi: 10.1109/ASPAA.2001.969533. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=969533. 3, 155

B. Hamilton and P. Depalle. Comparisons of parameter estimation methods for an exponential polynomial sound signal model. In *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio*, 3 2012a. URL http://www.aes.org/e-lib/browse.cfm?elib=16180. 56, 155

B. Hamilton and P. Depalle. A unified view of non-stationary sinusoidal parameter estimation methods using signal derivatives. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Kyoto, Japan, 2012b. 30, 40, 155

B. Hamilton, P. Depalle, and S. Marchand. Theoretical and practical comparisons of the reassignment method and the derivative method for the estimation of the frequency slope. In *IEEE Workshop Applicat. of Signal Process. to Audio and Acoustics*, pages 345 –348, oct. 2009. doi: 10.1109/ASPAA.2009.5346513. 8, 25, 26, 37, 39, 40

T. W. Hänsch, A. L. Schawlow, and G. W. Series. The spectrum of atomic hydrogen. *Sci. Amer.*, 240(94):531544, March 1979. 3, 20

F.J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978. ISSN 0018-9219. doi: 10.1109/PROC.1978.10837. 14, 25

T. Henderson. Geometric methods for determining system poles from transient response. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(5):982–988, 1981. ISSN 0096-3518. doi: 10.1109/TASSP.1981.1163675. 19

Kris Hermus, Werner Verhelst, Philippe Lemmerling, Patrick Wambacq, and Sabine Van Huffel. Perceptual audio modeling with exponentially damped sinusoids. *Signal Processing*, 85(1):163–176, 2005. 6

F. B. Hildebrand. *Introduction to Numerical Analysis*. New York: McGraw-Hill, 1956. 18

Y. Hua and Yu Zhang. Perturbation analysis of tk method for harmonic retrieval problems. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(2):228–240, 1988. ISSN 0096-3518. doi: 10.1109/29.1515. 137

Y. Hua and Yu Zhang. On svd for estimating generalized eigenvalues of singular matrix pencil in noise. In *Circuits and Systems, 1991., IEEE International Sympoisum on*, pages 2780–2783 vol.5, 1991. doi: 10.1109/ISCAS.1991.176121. 20

J Jensen, R Heusdens, and SH Jensen. A perceptual subspace approach for modeling of speech and audio signals with damped sinusoids. *IEEE*

Transactions on Speech and Audio Processing, 12:121–132, 03/2004 2004. 6

G.P. Kafentzis, Y. Pantazis, O. Rosec, and Y. Stylianou. An extension of the adaptive quasi-harmonic model. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 4605–4608, 2012. doi: 10.1109/ICASSP.2012.6288944. 15, 95

S. M. Kay. *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1 edition, 1988. 16

Steven M. Kay. *Mathematical Methods of Statistics*. Prentice Hall, 1946. URL http://books.google.com/books?hl=fr&lr=&id=CRTKKaJOODYC&pgis=1. 7

Steven M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory (v. 1)*. Prentice Hall, 1 edition, April 1993. ISBN 0133457117. URL http://www.worldcat.org/isbn/0133457117. 16, 17

Simon King. A tutorial on HMM speech synthesis (invited paper). In *Sadhana – Academy Proceedings in Engineering Sciences, Indian Institute of Sciences*, 2010. 3

K. Kodera, R. Gendrin, and C. Villedary. Analysis of time-varying signals with small BT values. *IEEE Trans. Acoust., Speech, Signal Process.*, 26(1):64–76, 1978. ISSN 0096-3518. doi: 10.1109/TASSP.1978.1163047. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1163047. 23, 155, 156, 166

Kunihiko Kodera, Claude De Villedary, and Roger Gendrin. A new method for the numerical analysis of non-stationary signals. *Physics of the Earth and Planetary Interiors*, 12(23):142 – 150, 1976. ISSN 0031-9201. doi: 10.1016/0031-9201(76)90044-3. URL http://www.sciencedirect.com/science/article/pii/0031920176900443. 4, 13, 23

A.C. Kot, S. Parthasarathy, D.W. Tufts, and R.J. Vaccaro. The statistical performance of state-variable balancing and prony's method in parameter estimation. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87.*, volume 12, pages 1549–1552, 1987. doi: 10.1109/ICASSP.1987.1169515. 19

R. Kumaresan and D.W. Tufts. Improved spectral resolution iii: Efficient realization. *Proceedings of the IEEE*, 68(10):1354–1355, 1980. ISSN 0018-9219. doi: 10.1109/PROC.1980.11867. 18

R. Kumaresan and D.W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 30(6):833–840, 1982. ISSN

0096-3518. doi: 10.1109/TASSP.1982.1163974. 18, 19

R. Kumaresan and S. Verma. On estimating the parameters of chirp signals using rank reduction techniques. In *Proc. 21st Asilomur Conf Signals, Sjst. Comput.*, pages 555–558, 1987. 21

Mathieu Lagrange. The desam toolbox: spectral analysis of musical audio. In *Proc. 13th Int. Digital Audio Effects*, Graz, Austria, September 2010. ISBN 978-3-200-01940-9. 126, 135

J. Laroche. A new analysis/synthesis system of musical signals using prony's method-application to heavily damped percussive sounds. In *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*, pages 2053–2056 vol.3, 1989. doi: 10.1109/ICASSP.1989.266864. 18

Guoqi Li and Changyun Wen. Legendre polynomials in signal reconstruction and compression. In *Proc. IEEE Conf. Ind. Electronics and Applicat.*, pages 1636–1640, June 2010. doi: 10.1109/ICIEA.2010.5514776. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5514776. 100, 103

Jian Li and Petre Stoica. An adaptive filtering approach to spectral estimation and sar imaging. *Signal Processing, IEEE Transactions on*, 44(6): 1469–1484, 1996. ISSN 1053-587X. doi: 10.1109/78.506612. 3, 15

S.X. Liao and M. Pawlak. On image analysis by moments. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(3):254–266, March 1996. ISSN 0162-8828. doi: 10.1109/34.485554. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=485554. 100

S. Marchand. Improving spectral analysis precision with an enhanced phase vocoder using signal derivatives. In *Proc. 1st Int. Digital Audio Effects Workshop*, pages 114–118, Barcelona, Spain, November 1998. doi: 10.1.1.41.2547. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.2547. 25

S. Marchand and P. Depalle. Generalization of the derivative analysis method to Non-Stationary sinusoidal modeling. In *Proc. 11th Int. Digital Audio Effects*, pages 281–288, Espoo, Finland, March 2008. URL http://hal.archives-ouvertes.fr/hal-00351950/en/. 6, 8, 25, 31, 38, 39, 40, 41, 43, 45, 105, 132, 159

Sylvain Marchand. *Advances in Spectral Modeling of Musical Sound*. N/A, 2008. URL http://docs.google.com/viewer?url=http://dept-info.labri.u-bordeaux.fr/~sm/Habilitation/habilitation.pdf. 25, 155, 157

Jr. Marple, L. Computing the discrete-time analytic signal via FFT. *IEEE Trans. Signal Process.*, 47(9):2600–2603, September 1999. ISSN 1053-587X. doi: 10.1109/78.782222. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=782222. 119

P. W. Milonni and J. H. Eberly. *Lasers*. Wiley Series in Pure and Applied Optics. New York: Wiley-Interscience, 1988. 20

E. H. Moore. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26(9):394395, 1920. 16

Norman Morrison. *Introduction to Sequential Smoothing and Prediction*. McGraw-Hill, New York, NY, 1 edition, 1969. ISBN 0754321069. 100, 101

R. Mukundan. Some computational aspects of discrete orthonormal moments. *IEEE Trans. Image Process.*, 13(8):1055–059, August 2004. ISSN 1057-7149. doi: 10.1109/TIP.2004.828430. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1315694. 100, 103

R. Mukundan, S. H. Ong, and P. A. Lee. Discrete orthogonal moment features using chebyshev polynomials. In *Proc. Int. Conf. Image and Vision Computing New Zealand*, pages 20–25, Hamilton, New Zealand, November 2000. URL ir.canterbury.ac.nz/handle/10092/446. 100

R. Mukundan, S.H. Ong, and P.A. Lee. Discrete vs. continuous orthogonal moments for image analysis. In *Proc. Int. Conf. Imaging Systems, Sci. and Technology*, pages 23–29, Las Vegas, NV, July 2001. URL http://hdl.handle.net/10092/470. 100

S. Muševič and J. Bonada. Comparison of non-stationary sinsoid estimation methods using reassignment and derivatives. In *Proc. 7th Sound and Music Computing Conf.*, Barcelona, Spain, July 2010a. URL http://smcnetwork.org/files/proceedings/2010/14.pdf. 76

S. Muševič and J. Bonada. Comparison of non-stationary sinusoid estimation methods using reassignment and derivatives. In *Proc. 7th Sound and Music Computing Conf.*, Barcelona, Spain, July 2010b. URL http://smcnetwork.org/files/proceedings/2010/14.pdf. 8, 25, 37, 155

Sašo Muševič and J. Bonada. Generalized reassignment with an adaptive polynomial-phase fourier kernel for the estimation of non-stationary sinusoidal parameters. In *Proc. 14th Int. Digital Audio Effects*, pages 371–374, Paris, France, September 2011. URL http://recherche.ircam.fr/pub/dafx11/Papers/45_e.pdf. 27, 116

Sašo Muševič and Jordi Bonada. Derivative analysis of complex polynomial

amplitude, complex exponential with exponential damping. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013. 21, 135

C. P Neuman and D. I Schonbach. Discrete (Legendre) orthogonal polynomials-a survey. *Int. J. Numer. Methods Eng.*, 8(4): 743–770, January 1974. ISSN 1097-0207. doi: 10.1002/nme.1620080406. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.1620080406/abstract. 101

G.K. Nilsen. Recursive time-frequency reassignment. *Signal Processing, IEEE Transactions on*, 57(8):3283–3287, 2009. ISSN 1053-587X. doi: 10.1109/TSP.2009.2020355. 13

Albert H. Nuttall. Some windows with very good sidelobe behavior. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(1):84–91, 1981. ISSN 0096-3518. doi: 10.1109/TASSP.1981.1163506. 14, 25

V.Ivanova Olga, Laura Marcu, and C.K.Khoo Michael. A nonparametric method for analysis of fluorescence emission in combined time and wavelength dimensions. *Annals of Biomedical Engineering*, 33(4):531–544, 2005. ISSN 0090-6964. doi: 10.1007/s10439-005-2512-5. URL http://dx.doi.org/10.1007/s10439-005-2512-5. 20

Y. Pantazis, O. Rosec, and Y. Stylianou. Chirp rate estimation of speech based on a time-varying quasi-harmonic model. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3985–3988, 2009a. doi: 10.1109/ICASSP.2009.4960501. xxiii, 15, 17, 18

Y. Pantazis, O. Rosec, and Y. Stylianou. Adaptive am/fm signal decomposition with application to speech analysis. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(2):290–300, 2011. ISSN 1558-7916. doi: 10.1109/TASL.2010.2047682. 6, 15, 16, 17, 21, 95, 150

Yannis Pantazis, Olivier Rosec, and Yannis Stylianou. On the properties of a time-varying quasi-harmonic model of speech. In *INTERSPEECH*, pages 1044–1047, 2008. 15

Yannis Pantazis, Olivier Rosec, and Yannis Stylianou. Am-fm estimation for speech based on a time-varying sinusoidal model. In *INTERSPEECH*, pages 104–107, 2009b. 15, 17, 95

M.S. Pattichis, C.S. Pattichis, M. Avraam, A. Bovik, and K. Kyriacou. Am-fm texture segmentation in electron microscopic muscle imaging. *Medical Imaging, IEEE Transactions on*, 19(12):1253–1257, 2000. ISSN 0278-0062. doi: 10.1109/42.897818. 3

Roger Penrose. A generalized inverse for matrices. In *Proceedings of the Cambridge Philosophical Society 51*, page 406413, 1955. doi: 10.1017/S0305004100030401. 16

V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysical Journal of the Royal Astronomical Society*, 33(3):347366, 1973. doi: 10.1111/j.1365-246X.1973.tb03424.x. 19

B. Porat and Benjamin Friedlander. On the accuracy of the kumaresan-tufts method for estimating complex damped exponentials. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(2):231–235, 1987. ISSN 0096-3518. doi: 10.1109/TASSP.1987.1165121. 20, 137

William H Press. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge Uni. Press, Cambridge, NY, 2 edition, 1992. ISBN 0521431085. 102

C. Radhakrishna Rao. Minimum variance and the estimation of several parameters. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43(02):280–283, 1947. doi: 10.1017/S0305004100023471. URL http://dx.doi.org/10.1017/S0305004100023471. xiii, 7

M. Raspaud and S. Marchand. Enhanced resampling for sinusoidal modeling parameters. In *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*, pages 327–330, 2007. doi: 10.1109/ASPAA.2007.4393020. 31

G. Rilling and P. Flandrin. One or two frequencies? the empirical mode decomposition answers. *Signal Processing, IEEE Transactions on*, 56(1):85–95, 2008. ISSN 1053-587X. doi: 10.1109/TSP.2007.906771. 4

Axel Röbel. Estimating partial frequency and frequency slope using reassignment operators. In *Proc. Int. Comput. Music Conf.*, pages 122–125, Gothenburg, Sweden, September 2002. doi: 10.1.1.8.7437. URL http://articles.ircam.fr/textes/Roebel02a/index.pdf. 38

R. Roy and T. Kailath. Total least squares esprit. In *21st Asilomar Conference on Signals, Systems, and Computers*, page 297301, November 1987. 20

R. Roy, A. Paulraj, and T. Kailath. Esprit-a subspace rotation approach to estimation of parameters of cisoids in noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(5):1340–1342, 1986. ISSN 0096-3518. doi: 10.1109/TASSP.1986.1164935. 19, 20

A.M. Sabatini. Correlation receivers using laguerre filter banks for modelling narrowband ultrasonic echoes and estimating their time-of-flights. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*,

44(6):1253–1263, 1997. ISSN 0885-3010. doi: 10.1109/58.656629. 20

R.O. Schmidt. *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. Stanford University, 1981. URL http://books.google.es/books?id=30qvmgEACAAJ. 19

R.O. Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276–280, 1986. ISSN 0018-926X. doi: 10.1109/TAP.1986.1143830. 19

Simon Scholler and P. Purwins. Sparse approximations for drum sound classification. *IEEE Journal of Selected Topics in Signal Processing*, 5: 933 – 940, 2011 2011. 21

L Schwartz. *Thorie des Distributions*. Hermann, 1966. 168

X. Serra. *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition*. PhD thesis, Stanford University, Stanford, CA, 1989. URL https://ccrma.stanford.edu/files/papers/stanm58.pdf. xi

R.A. Serway, J. John W. Jewett, and V. Peroomian. *Physics for Scientists and Engineers*. Number v. 1 in Physics for Scientists and Engineers. Brooks/Cole, Cengage Learning, 2010. ISBN 9781439048382. URL http://books.google.es/books?id=6upvonUt0O8C. 20

Kai Siedenburg and Monika Drfler. Structured sparsity for audio signals. In *Proc. 14th Int. Digital Audio Effects*, 2011. 141

P. Stoica and Nehorai Arye. Music, maximum likelihood, and cramer-rao bound. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(5):720 –741, may 1989. ISSN 0096-3518. doi: 10.1109/29.17564. 8

P. Stoica and A. Nehorai. Study of the statistical performance of the pisarenko harmonic decomposition method. *Radar and Signal Processing, IEE Proceedings F*, 135(2):161–168, 1988. ISSN 0956-375X. 19

Petre Stoica and Randolph L. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1 edition, February 1997. 16

Petre Stoica and T. Soderstrom. Statistical analysis of music and subspace rotation estimates of sinusoidal frequencies. *Signal Processing, IEEE Transactions on*, 39(8):1836–1847, 1991. ISSN 1053-587X. doi: 10.1109/78.91154. 20

Petre Stoica, Andreas Jakobsson, and Jian Li. Matched-filter bank interpretation of some spectral estimators. *Signal Process.*, 66(1):45–59, April 1998. ISSN 0165-1684. doi: 10.1016/S0165-1684(97)00239-9. URL http://dx.doi.org/10.1016/S0165-1684(97)00239-9. 15

Petre Stoica, Hongbin Li, and Jian Li. Amplitude estimation of sinusoidal signals: survey, new results, and an application. *Signal Processing, IEEE Transactions on*, 48(2):338–352, 2000. ISSN 1053-587X. doi: 10.1109/78.823962. 14, 89

D. Stowell and M.D. Plumbley. Framewise heterodyne chirp analysis of birdsong. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 2694–2698, 2012. 139

Dan Stowell, Saso Musevic, Jordi Bonada, and Mark D. Plumbley. Improved multiple birdsong tracking with distribution derivative method and markov renewal process clustering. *CoRR*, abs/1302.3462, 2013. 3, 143

Yannis Stylianou. *Harmonic Plus Noise Models for Speech, Combined with Statistical Methods, for Speech and Speaker Modification*. PhD thesis, 1996. 17

David N. Swingler. Frequency estimation for closely spaced sinsoids: Simple approximations to the cramer-rao lower bound. *IEEE Transactions on Signal Processing*, 41(1):489–494, 1993. 8

T.F.Quateri. *Discrete-Time Speech Signal Processing: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 2002, Signal Processing Series., 1993. 139

Keiichi Tokuda, Yoshihiko Nankaku, Tomoki Today, Heiga Zen, Junichi Yamagishi, and Keiichiro Oura. Speech synthesis based on hidden markov models. *Proceedings of the IEEE*, 101(6), June 2013. (in press). 3

D.W. Tufts and R. Kumaresan. Improved spectral resolution. *Proceedings of the IEEE*, 68(3):419–420, 1980a. ISSN 0018-9219. doi: 10.1109/PROC.1980.11657. 18

D.W. Tufts and R. Kumaresan. Improved spectral resolution ii. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '80.*, volume 5, pages 592–597, 1980b. doi: 10.1109/ICASSP.1980.1170988. 18

V. Simonyte V. Slivinskas, M. Radavicius. Cramer-rao bound for the estimates of frequencies and damping factors of quasipolynomials in noise. *Lithuanian Mathematical Journal*, 32:327–333, JulySeptember 1992. 20

A.-J. van der Veen, E.F. Deprettere, and A.L. Swindlehurst. Subspace-based signal analysis using singular value decomposition. *Proceedings of the IEEE*, 81(9):1277–1308, 1993. ISSN 0018-9219. doi: 10.1109/5.237536. 19

Tony S. Verma and Teresa H. Y. Meng. Extending spectral modeling syn-

thesis with transient modeling synthesis. *Comput. Music J.*, 24(2):47–59, July 2000. ISSN 0148-9267. doi: 10.1162/014892600559317. URL http://dx.doi.org/10.1162/014892600559317. xi

T.S. Verma and Teresa H.Y. Meng. An analysis/synthesis tool for transient signals that allows a flexible sines+transients+noise model for audio. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 6, pages 3573–3576 vol.6, 1998. doi: 10.1109/ICASSP.1998.679647. xi

Charlotte L. Walters, Robin Freeman, Alanna Collen, Christian Dietz, M. Brock Fenton, Gareth Jones, Martin K. Obrist, Sbastien J. Puechmaille, Thomas Sattler, Bjrn M. Siemers, Stuart Parsons, and Kate E. Jones. A continental-scale tool for acoustic identification of european bats. *Journal of Applied Ecology*, 49(5):1064–1074, 2012. ISSN 1365-2664. doi: 10.1111/j.1365-2664.2012.02182.x. URL http://dx.doi.org/10.1111/j.1365-2664.2012.02182.x. 139

X. Wen and M. Sandler. Notes on model-based non-stationary sinusoid estimation methods using derivative. In *Proc. 12th Int. Digital Audio Effects*, pages 113–120, Como, Italy, September 2009. URL http://dafx09.como.polimi.it/proceedings/papers/paper_27.pdf. 3, 5, 8, 17, 26, 44, 76, 77, 78, 86, 96, 114, 116, 128, 162, 163

X. Wen and M. Sandler. Fast Additive Sinusoidal Synthesis With a Subband Sinusoidal Method. *IEEE Signal Processing Letters*, 20:467–470, May 2013. doi: 10.1109/LSP.2013.2250958. 28

Xue Wen and Mark Sandler. On the characterization of slowly varying sinusoids. *EURASIP J. Audio Speech Music Process.*, 2010:7:1–7:12, January 2010. ISSN 1687-4714. doi: 10.1155/2010/941732. URL http://dx.doi.org/10.1155/2010/941732. 4

Inc. Wolfram Research. Mathematica. URL http://www.wolfram.com/mathematica/. 85, 86

HAU-TIENG WU, PATRICK FLANDRIN, and INGRID DAUBECHIES. One or two frequencies? the synchrosqueezing answers. *Advances in Adaptive Data Analysis*, 03(01n02):29–39, 2011. doi: 10.1142/S179353691100074X. URL http://www.worldscientific.com/doi/abs/10.1142/S179353691100074X. 4

S.F. Yau and Y. Bresler. A compact cramer-rao bound expression for parametric estimation of superimposed signals. *Signal Processing, IEEE Transactions on*, 40(5):1226 –1230, may 1992. ISSN 1053-587X. doi: 10.1109/78.134484. 10