LUND UNIVERSITY

Sparse Modeling Heuristics for Parameter Estimation - Applications in Statistical Signal Processing

Adalbjörnsson, Stefan Ingi

2014

Link to publication

Citation for published version (APA):
Adalbjörnsson, S. I. (2014). *Sparse Modeling Heuristics for Parameter Estimation - Applications in Statistical Signal Processing*. [Doctoral Thesis (compilation), Mathematical Statistics]. Centre for Mathematical Sciences, Lund University.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Sparse modeling heuristics for parameter estimation

Applications in statistical signal processing

Stefan Ingi Adalbjörnsson

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118
SE-221 00 Lund
Sweden

http://www.maths.lth.se/

Doctoral Theses in Mathematical Sciences 2014:6
ISSN 1404-0034

ISBN 978-91-7623-106-7
LUTFMS-1042-2014

© Stefan Ingi Adalbjörnsson, 2014

Printed in Sweden by KFS AB, Lund 2014

Contents

Acknowledgements **iii**

List of papers **v**

Introduction **1**
1 Background 1
2 Sparse modeling and estimation 3
3 Sparse recovery 5
4 Convex optimization 8
5 Recovery guarantees 11
6 Outline of the papers 13
7 Topics for future research 17

A Multi-pitch estimation exploiting block sparsity **25**
1 Introduction 26
2 Block sparse signal model 28
3 Pitch estimation using block sparsity 29
4 An efficient ADMM implementation 32
5 Numerical results 40
6 Conclusions 51
7 Appendix 53

B Sparse localization of harmonic audio sources **65**
1 Introduction 66
2 Signal model 67
3 Joint pitch and localization estimation 72
4 An efficient ADMM implementation 79
5 Numerical comparisons 83
6 Conclusions 90
7 Appendix 92

C Estimating periodicities in symbolic sequences using sparse modeling **101**
1 Introduction 102
2 Probabilistic model for symbolic sequences 103
3 Relaxation of the cardinality constraint 109
4 An efficient implementation 113
5 Numerical results 116
6 Conclusion 121

D High resolution sparse estimation of exponentially decaying N-D signals **129**
1 Introduction 130
2 The N-D signal model 131
3 An efficient ADMM implementation 135
4 Sparse dictionary learning 138
5 Numerical examples 139
6 Conclusions 149

E Joint model-order and fundamental frequency estimation in the presence of inharmonicity **159**
1 Introduction 160
2 Signal model and other estimators 161
3 Proposed robust covariance-fitting pitch estimator 163
4 Model-order selection 167
5 Simulations and results 171
6 Conclusion 176

Acknowledgements

I would like to thank my supervisor, Prof. Andreas Jakobsson, for all his help and support in writing this thesis. Andreas has been a pillar of support to me during my time as a Ph.D. student, always warmly greeting me with his insights into spectral analysis, traveling tips, wine pairings, child rearing (although I have no children of my own), or anything else I might have needed. This thesis could not have been written without him, or without the team of co-workers around him, many of which are co-authors on the papers included in the thesis. Much of the work has been performed in close collaboration with Ted Kronvall and Johan Swärd, I owe them much. They shared a great deal of the burden in putting this thesis together, for which I will always be indebted to them. All the good times we spent together during these last years are my fondest memories from my time as a Ph.D. student. My other co-authors also deserve, at the very least, a big thank you. In particular, I am grateful to Prof. Jian Li, who invited me to work in Gainesville for two summers; not only did the methods I was exposed to there greatly shape the content of this thesis, these visits were also some of the most exciting times during my Ph.D. I would also like to thank all my co-workers at the department for the numerous coffee breaks, wine tastings, and the (surprisingly many?!) work parties. Mona, James, and Maria L., thank you for keeping all the practical things working, and for a (partial) success in keeping the topics discussed at our social events out of the realm of stochastic differential equations and sigma algebras.

Finally, I would like to express my thanks to my family for all the things that are impossible to put into words. To Helena, thank you for being my everything, I promise to make time for summer vacations in the future!

Lund, 2014 *Stefan Ingi Adalbjörnsson*

List of papers

This thesis is based on the following papers:

A Stefan Ingi Adalbjörnsson, Andreas Jakobsson, and Mads G. Christensen, "Multi-pitch estimation exploiting block sparsity".
To appear in *Elsevier Signal Processing*.

B Stefan Ingi Adalbjörnsson, Ted Kronvall, Simon Burgess, Kalle Åström, and Andreas Jakobsson, "Harmonic audio localization".
Submitted for possible publication in *IEEE Journal of Selected Topics in Signal Processing*.

C Stefan Ingi Adalbjörnsson, Johan Swärd, Jonas Wallin, and Andreas Jakobsson, "Estimating periodicities in symbolic sequences using sparse modeling".
Submitted for possible publication in *IEEE Transactions on Signal Processing*.

D Johan Swärd, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson, "High resolution sparse estimation of exponentially decaying N-D signals".
Submitted for possible publication in *IEEE Transactions on Signal Processing*.

E Naveed R. Butt, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson, "Joint model-order and fundamental-frequency estimation in the presence of inharmonicity".
To be submitted.

Additional papers not included in the thesis:

1. Stefan Ingi Adalbjörnsson, Johan Swärd, Ted Kronvall, Andreas Jakobsson, "A Sparse Approach for Estimation of Amplitude Modulated Sinusoids", to be presented at *The Asilomar Conference on Signals, Systems, and Computers, Asilomar*, USA, November 2-5, 2014.

2. Ted Kronvall, Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "Joint DOA and Multi-Pitch Estimation Via Block Sparse Dictionary Learning", *22nd European Signal Processing Conference*, Lisbon, Portugal, September 1-5, 2014.

3. Stefan Ingi Adalbjörnsson, Johan Swärd, Andreas Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Two-Dimensional Signals", *22nd European Signal Processing Conference*, Lisbon, Portugal, September 1-5, 2014.

4. Ted Kronvall, Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "Joint DOA and Multi-Pitch estimation using Block Sparsity", *39th International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 4-9, 2014.

5. Johan Swärd, Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Signals", *39th International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 4-9, 2014.

6. Stefan Ingi Adalbjörnsson, Johan Swärd, Andreas Jakobsson, "Likelihood-based Estimation of Perioditicies in Symbolic Sequences", *21st European Signal Processing Conference*, Marrakech, Morocco, September 9-13, 2013.

7. Tommy Nilsson, Stefan Ingi Adalbjörnsson, Naveed R. Butt, Andreas Jakobsson, Multi-Pitch Estimation of Inharmonic Signals", *21st European Signal Processing Conference*, Marrakech, Morocco, September 9-13, 2013.

8. Stefan Ingi Adalbjörnsson, Andreas Jakobsson, and Mads G. Christensen, "Estimating Multiple Pitches using Block Sparsity", *38th International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26-31, 2013.

9. Naveed R. Butt, Stefan Ingi Adalbjörnsson, Samuel Somasundaram, Andreas Jakobsson, "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities", *38th International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26-31, 2013.

10. Stefan Ingi Adalbjörnsson, George O. Glentis, Andreas Jakobsson, Efficient Block and Time-Recursive Estimation of Sparse Volterra Systems", *IEEE Signal Processing Society Workshop*, Ann Arbor, USA, August 5-8, 2012.

11. Stefan Ingi Adalbjörnsson, Andreas Jakobsson, Sparse Estimation Of Spectroscopic Signals", *19th European Signal Processing Conference*, Barcelona, Spain, August 29 - September 1, 2011.

12. Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "Relax-Based Estimation of Voigt Lineshapes", *18th European Signal Processing Conference*, Aalborg, Denmark, August 23-27, 2010.

13. Erik Lindström, Jonas Ströjby, Stefan Ingi Adalbjörnsson, "Non-Linear Portmanteau Tests", *15th IFAC Symposium on System Identification*, St. Malo, France, July 6-8, 2009.

Introduction

This thesis is concerned with applications of sparse and robust modeling of various parameter estimation problems in audio modeling, audio localizations, DNA sequencing, and spectroscopy. These problems share the common characteristics of being well modeled using a sparse model formulation, such that the main parameters of interest are linked to only a few components, out of a large set of possible candidates. By imposing sparse constraints on the signal models one thereby allows for efficient estimation algorithms. In this introduction, we introduce the methodology and some of the underlying theory used in the following papers, as well as give some background to the studied problems, emphasizing their connection to the applied methods, and present an overview of the contributions in the thesis.

1 Background

During the recent decades, there has been a growing interest in the use of sparse linear models, where one considers signals that may be well modeled as the linear combination of a few vectors, out of a large set of feasible candidates. Originating as heuristics for solving under-determined system of equations, sparse modeling has become a thoroughly developed field with a rigorous mathematical and statistical theory, as well as a widely used tool in applications. Such models occur in a surprising number of applications, with one of the earliest examples being from reflection seismology [1], where one measures the reflections, stemming from abrupt changes in the earth's subsurfaces, from a series of shocks (impulses) to the surface. Other notable examples include genomics, where one commonly tries to infer what combination of DNA symbols that may be linked to various kinds of the experimental data, as seen in, for example, motif regression and prediction of DNA splice sites [2]. Further examples include various engineering applications related to line spectral analysis [3–5], such as direction of arrival estimation [6], radar imaging [7], and spectroscopy [8], as well as, for example, in numerous imaging and machine learning applications (see [9] for further examples as well

as a general overview on sparse modeling). As an illustrative example, consider the modeling of voiced speech, which may be well modeled using a few sinusoidal components. Such a signal can be represented using a sparse model by considering the signal as being formed by a few Fourier vectors, with frequencies corresponding to each of the sinusoids, and with the sinusoidal amplitudes forming the sparse set of coefficients, wherein one views the contribution from all other frequencies as contributing with zero coefficients. Since one does not know beforehand the signals frequencies, one instead considers a large number of possible frequencies, each thus represented by a Fourier vector and a corresponding coefficient. Consider a signal consisting of n samples; This may then be expressed in matrix-vector notation as

$$\mathbf{y} = \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 \dots \mathbf{a}_p x_p \tag{1}$$

$$= \begin{bmatrix} \mathbf{a}_1 \dots \mathbf{a}_p \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \tag{2}$$

$$\triangleq \mathbf{A}\mathbf{x} \tag{3}$$

where x_k denotes element k of the vector $\mathbf{x} \in \mathbb{C}^p$, $\mathbf{y} \in \mathbb{C}^n$ is the observation vector, $\mathbf{A} \in \mathbb{C}^{n \times p}$ is termed a dictionary matrix, such that each column represents one particular Fourier vector and the corresponding coefficient (or amplitude) is thus an element in the vector $\mathbf{x}$. Here, one might, for example, choose the frequencies in the dictionary such that $\mathbf{a}_k$ is a Fourier vector with normalized frequency k/p. Thus, given an harmonic signal, such as voiced speech, the signal could be approximated by finding for each frequency component in the signal, a corresponding Fourier vector in the dictionary, and setting its coefficient equal to the amplitude of the sinusoid, with all the other coefficients being set to zero. Note that if we use the previously suggested choice of dictionary, the worst approximation error would be $\pm\frac{1}{2}k/p$. Thus, as the quality of the approximation becomes better as one considers more and more Fourier vectors, for this form of signal representation to be useful, the model necessarily exhibits the typically undesirable characteristic of containing more unknowns than measurement, yielding an under-determined system of equations. Since we can assume that the dictionary matrix has full rank, the resulting systems of equations will be difficult to work with, given that they offer an infinite number of feasible solutions, were any two solutions can have wildly different characteristics. However, with prior knowledge that the system of equations has a sparse solution, i.e., that the coefficient

vector should be sparse and contains mostly zeros, one is, perhaps somewhat surprisingly, able to formulate highly efficient algorithms that are actually able to accurately reconstruct the signal using only the non-zero elements, offering, with high probability, a unique solution. This reconstruction can be done in various ways; some common choices include greedy methods that build up a solution one vector at a time, Bayesian methods, that use various prior distributions to promote sparsity, and convex relaxation techniques, where a difficult problem is approximated with a convex problem. In this thesis, we will mainly examine the last of these approaches. This choice of methodology, i.e., by mainly relying on convex relaxation, is a pragmatic one, allowing for sufficient flexibility for our purposes, i.e., the models are sufficiently detailed to include the relevant characteristics of the signals in question, and since the resulting criteria are convex, the computational effort will be tractable using the well developed theory that exist for convex optimization. In the next sections, we will give a brief overview of when and how this is possible, as well as present some of the basic theory that is useful for the analysis of such problems.

2 Sparse modeling and estimation

We are in this work primarily interested in modeling and estimation for *separable models*, formed as a linear combination of K components $\mathbf{a}(\vartheta_k)$, each scaled with the coefficient x_k, such that

$$\mathbf{y} = \sum_{k=1}^{K} x_k \mathbf{a}(\vartheta_k) + \boldsymbol{\varepsilon} \tag{4}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the vector of observations, $\vartheta_k \in \Omega \subset \mathbb{R}^M$ is the parameter vector containing the M unknown parameters, $\mathbf{a}(\cdot)$ is function such that $\mathbf{a}(\cdot) : \mathbb{R}^M \to \mathbb{C}^n$, and $\boldsymbol{\varepsilon}$ is a noise vector which is here, for simplicity, assumed to be uncorrelated (circularly symmetric) Gaussian distributed random variables. As is common for this form of signal models, a straightforward least squares or maximum likelihood solution will yield a complicated multi-modal optimization problem, typically having far too many local maxima for a gradient based, or similar, non-linear optimization to be applicable (see also, e.g. [3]). Thus, the resulting optimization is commonly done by evaluating the likelihood on a grid of values which leads to a high computational cost, especially in the multidimensional case. Furthermore, since the number of components, i.e., the model order,

K, is in general not known, one often needs to resort to solving the optimization problem for a possibly large number of different model orders and combined such solutions with an appropriate model choice criteria before a final estimate can be produced (see, e.g., [10] for an overview of the model order selection problem). Both of these difficulties are addressed by the sparse modeling approach to parameter estimation. The central idea is to approximate the non-linear model with a linear model. This is accomplished by assuming that the signal can be well approximated as a linear combination of vectors, where each vector corresponds to a particular grid point, such that the grid covers the entire parameter space. As a result, given a large enough dictionary, each signal component may be well approximated by an element that lies soe close to the true value that the resulting approximation error is small. The resulting linear system can be written in matrix form as

$$\mathbf{y} = \sum_{k=1}^{p} x_k \mathbf{a}_k + \boldsymbol{\varepsilon} \tag{5}$$

$$\triangleq \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon} \tag{6}$$

where p is the total number of grid points considered, assumed to be far larger than the number of observations, and each $\mathbf{a}_k$ corresponds to the vector representing the contribution from a specific grid point ϑ_k. Clearly, given that the dictionary needs to be fine enough, the size of the overall dictionary matrix $\mathbf{A}$, will grow rapidly, especially for multidimensional data set sets, quickly making it an unmanageable representation, both in terms of complexity and in terms of the necessary memory requirements. In Paper D, we examine an example of this problem, wherein we treat N-dimensional spectroscopy, such that the parameters space contains 2 dimensions for each of the N-dimensions. Even for low dimensional problems, forming a fine dictionary over $2N$ dimensions quickly becomes unfeasible, necessitating alternative solutions. We will examine this aspect further later on in Paper D.

As compared to the direct maximum likelihood approach using a grid of values, the difference with using the dictionary model in (5) is that the latter does not require *a priori* knowledge of the model order, and is rather only assuming that most of the amplitudes x_k are zero. Clearly, in case $\mathbf{A}$ is full rank, there are infinitely many solutions to the system of equations. To avoid this difficulty, one

may then select the solution that only has K non-zero elements, such that

$$\underset{\mathbf{x}}{\text{minimize}} \ \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_{\ell_2}^2 + \lambda||\mathbf{x}||_{\ell_0} \tag{7}$$

where λ is a positive tuning parameter that weighs the importance of the model fit and the sparsity level, and $\|\mathbf{x}\|_{\ell_0} = \sum_{k=1}^{p} \mathbf{1}_{x_k \neq 0}$, i.e., the function that counts the number of non-zero entries in a vector, and the ℓ_q-norm defined as

$$\|\mathbf{x}\|_{\ell_q}^q = \sum_{k=1}^{p} |x_k|^q \tag{8}$$

Such a solution would nicely impose the assumed sparsity structure, although it would require knowledge of the model order, K. Unfortunately, this problem is usually impossible to solve as it requires solving a least squares problem for all combination of K vectors (see, e.g., the discussion in [11]). This form of combinatorial problems are well known to be so-called NP-hard, meaning that they are as difficult to solve as some other problems that have a computational cost that will grow exponentially with the problem size, making it a daunting task even for small problems. As we here consider problems where the fidelity of the solution depends on having a large, or even a very large, dictionary, it is unfeasible to form this kind of solution. However, as we present in the next section there exist relaxations of (7) that are both easy to compute as well as having recovery guarantees for certain problems, i.e., instances when the relaxation will with high probability yield the same solution as (7).

3 Sparse recovery

The most well studied relaxation of (7) is the convex relaxation obtained by replacing the ℓ_0 penalty with the ℓ_1 norm, i.e., the convex optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \ \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_{\ell_2}^2 + \lambda||\mathbf{x}||_{\ell_1} \tag{9}$$

which is commonly referred to as either the least absolute shrinkage and selection operator (LASSO) [12] or basis pursuit denoising (BPDN) [13]. Although (9) does in general not offer a closed form solution, it can be recast as a second order cone program, allowing it to be solved using well developed interior point

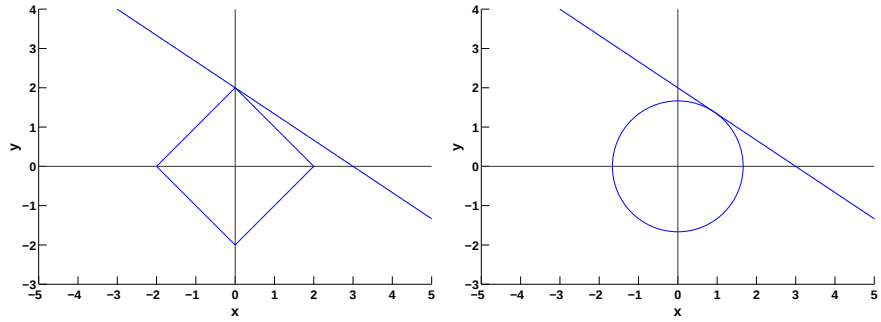

Figure 1: The straight line in both figures represent the solution set of an equation with two variables, on the left the intersection with the smallest ℓ_1 ball and on the right with the smallest ℓ_2 ball. As can be seen from the figures, the ℓ_1 solution has one of the coordinates as zero, whereas both are nonzero for the ℓ_2 solution. As a result, imposing a ℓ_1 norm constraint will favor sparser solution as compared to using an ℓ_2 criteria.

methods (see, e.g., [14]). Some intuition as to why the ℓ_1 penalty promotes sparse solutions can be gained by studying Figure 1, where a line is used to represent all solutions to an under-determined linear system with one equation and two variables. When comparing the minimum ℓ_1 solution with the minimum ℓ_2 solution, one can see that the ℓ_1 solution has one of the variables being exactly equal to zero, whereas both are non-zero for the ℓ_2 solution. Imposing the ℓ_1 norm as a penalty will thus favor a sparser solution as compared to the one found using an ℓ_2 criteria. This intuition can be made concrete by considering the first order Karush-Kuhn-Tucker (KKT) necessary condition for a solution to (9) to be optimal (see also ,e.g., [14, 15]). For many convex optimization problems, this condition is simply that if the gradient is equal to zero, one can be assured that the point is optimal. It may be noted that this implies that any locally optimal point is globally optimal, perhaps the most important attribute of convex optimization problems. However, since the here considered functions are not differentiable, the analysis needs to be performed using subdifferential calculus, where one similarly to the differential case may show that the necessary condition is that zero should be included in the subdifferential set (see, e.g., [15]). Thus, for the real valued version of (9), a necessary and sufficient condition for a minimizer $\mathbf{x}^*$ to

be optimal is that (see [2] for a more thorough treatment than presented here)

$$\mathbf{0} \in \mathbf{A}^T(\mathbf{A}\mathbf{x}^* - \mathbf{y}) + \lambda \mathbf{e} \tag{10}$$

where $\mathbf{e}$ is a vector such that the k:th element in the vector is either $e_k = \text{sign}(x_k^*)$, if x_k^* is non-zero, or $e_k \in [-1,\ 1]$, if $\mathbf{x}_k^*$ is zero. For the zero elements, this thus implies that if, say, variable $x_s^* = 0$, then

$$|\mathbf{a}_s^T(\mathbf{A}\mathbf{x}^* - \mathbf{y})| \leq \lambda \tag{11}$$

where $\mathbf{a}_s$ denotes column s of $\mathbf{A}$. Thus, now assuming $\mathbf{a}_s$ has unit norm for simplicity, if one were to solve

$$\underset{z}{\text{minimize}}\ \frac{1}{2}||\mathbf{A}\mathbf{x}^* - \mathbf{y} - \mathbf{a}_s z||_{\ell_2}^2 \tag{12}$$

this will yield an solution such that $|z^*| \leq \lambda$, with the intuitive interpretation being that if a least squares estimate using the residual leads to an estimated coefficient that is less than λ, then the coefficient is set to zero. For the non-zero variables, the KKT conditions are

$$\mathbf{0} = \tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{x}}^* - \mathbf{y}) + \lambda\, \text{sign}(\tilde{\mathbf{x}}^*) \tag{13}$$

where $\tilde{\mathbf{A}}$ is a matrix formed out of the columns of $\mathbf{A}$ that correspond to nonzero variables in $\mathbf{x}^*$, and $\tilde{\mathbf{x}}^*$ is the corresponding nonzero variables. When this is compared with the KKT conditions for the unpenalized least squares problem, it becomes clear that the estimated variables are shrunk by λ. As this shrinkage causes a bias towards zero, which can be troublesome in some applications, alternative penalty functions have been considered that minimize this effect, e.g., ℓ_q with $0 < q < 1$ [4, 7, 16] or the reweighted ℓ_1 [17], which is equivalent with a log penalty. In Figure 2, the comparison between the log penalty, the ℓ_1, and the ℓ_0 penalty is given. As can be seen the reweighted ℓ_1 penalizes larger amplitudes proportionally less than the ℓ_1 penalty, mimicking the ℓ_0 penalty more closely. This analysis framework is general enough to handle many other sparse criteria, e.g., in Paper A, we perform a similar analysis for the block sparse model (with sparsity within each block) making the intuitive connection between the tuning parameters and signal amplitudes concrete. Furthermore, in some cases the KKT conditions can be solved with a closed form expression, allowing for much improved computational complexity (see also Section 4.1).

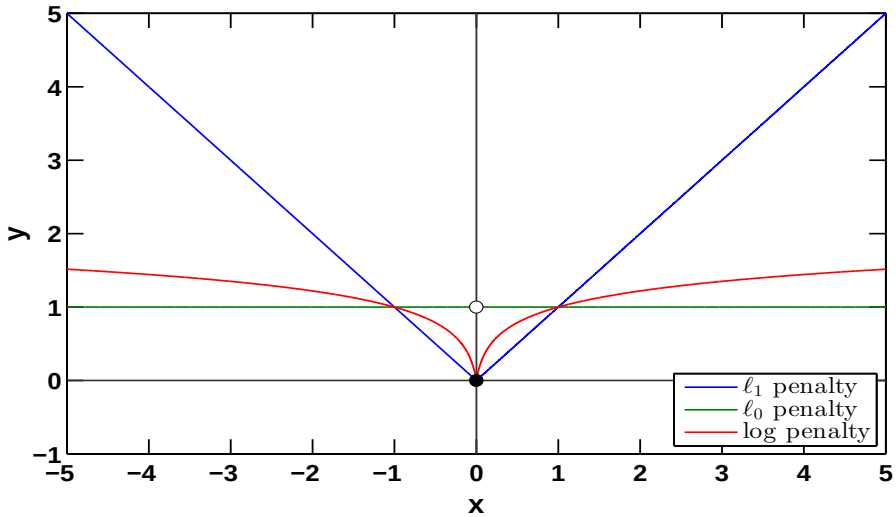

Figure 2: The log penalty is a better approximation for the ℓ_0 penalty than the convex ℓ_1 penalty.

4 Convex optimization

Most of the convex optimization problems considered in this thesis can be approached using the methodology of *disciplined convex programming*, a concept introduced in [18, 19] with a corresponding software package [20]. The methodology allows for transformation from a problem statement to a solvable form that may be performed automatically by a computer, a task that is far from trivial in many cases. This is done by formalizing how expert practitioners and theoreticians of convex optimization often approach mathematical modeling; a criterion is then formed by including function and restrictions that are known to be convex and manipulated in such ways that convexity is preserved. Once in standard form, the problem can be solved using interior point methods implemented in commonly available software packages such as SeDuMi [21] and SDPT3 [22]. However, convenient as it may be for prototyping new algorithms or methods, this approach applied to sparse modeling problems often leads to a prohibitive computational cost, which can give an overly pessimistic view of the feasibility of the approach. In the thesis, we consider two well studied approaches for solving the optimization problems encountered, namely, the alternating direction

Algorithm 1 The general ADMM algorithm

1: Initiate $\mathbf{z} = \mathbf{z}(0), \mathbf{u} = \mathbf{u}(0)$, and $\ell = 0$
2: **repeat**
3: $\quad \mathbf{z}(\ell+1) = \underset{\mathbf{z}}{\operatorname{argmin}} f_1(\mathbf{z}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z} - \mathbf{u}(\ell) - \mathbf{d}(\ell)||_2^2$
4: $\quad \mathbf{u}(\ell+1) = \underset{\mathbf{u}}{\operatorname{argmin}} f_2(\mathbf{u}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z}(\ell+1) - \mathbf{u} - \mathbf{d}(\ell)||_2^2$
5: $\quad \mathbf{d}(\ell+1) = \mathbf{d}(\ell) - (\mathbf{G}\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1))$
6: $\quad \ell \leftarrow \ell + 1$
7: **until** convergence

method of multipliers (ADMM) framework [23] and the cyclic coordinate descent (CCD) [24]. With these methods, the knowledge that a sparse solution is sought can be utilized in the calculations, resulting in a great increase in speed. For example, for (9), each step in such an implementation only requires a computational cost linear in the number of parameters, whereas each step of the interior point method requires a Newton step, and is thus approximately proportional to the number of parameters cubed.

4.1 Efficient implementation - the ADMM

We proceed by examining the two efficient optimization approaches used in the thesis, beginning with the ADMM. The ADMM formulation has been gaining notable attention in the recent literature as a method for solving distributed, large-scale, optimization problems (see, e.g., [23] for an overview of the technique). The framework is quite general and offers provable convergence with minimal assumptions. For example, the non-differentiable functions that commonly appear in sparse modeling applications are no problem. The way the ADMM works is by solving the considered optimization problem by increasing the number of variables so that the problem may be divided into smaller sub-problems, which are then coordinated to achieve a global optima that solves also the original problem. As it turns out, when introducing these variables for the problems involving sparsity promoting penalties, one can often find closed form solution for the KKT conditions of the sub-problems, thereby allowing for fast algorithms. More concretely, ADMM considers the convex optimization problem

$$\underset{\mathbf{z}}{\text{minimize}} \quad f_1(\mathbf{z}) + f_2(\mathbf{G}\mathbf{z}) \tag{14}$$

where $\mathbf{z} \in \mathbb{R}^p$ is the optimization variable, $f_1(\cdot)$ and $f_2(\cdot)$ are convex functions, and $\mathbf{G} \in \mathbb{R}^{N \times p}$ is a known matrix. If one introduces an auxiliary variable, $\mathbf{u}$, then (14) may be equivalently expressed as

$$\begin{aligned} \underset{\mathbf{z},\,\mathbf{u}}{\text{minimize}} \quad & f_1(\mathbf{z}) + f_2(\mathbf{u}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z} - \mathbf{u}||_2^2 \\ \text{subject to} \quad & \mathbf{G}\mathbf{z} - \mathbf{u} = \mathbf{0} \end{aligned} \tag{15}$$

Note from (15) that the constraint $\mathbf{G}\mathbf{z} - \mathbf{u} = \mathbf{0}$ ensures that the penalty function in the minimization will disappear for a feasible solution, ensuring that (14) and (15) actually solve the same problem. The ADMM solves the optimization problem in (15) via the dual function, defined as the infimum with respect to $\mathbf{u}$ and $\mathbf{z}$ of the augmented Lagrangian, i.e.,

$$L_\mu(\mathbf{z}, \mathbf{u}, \mathbf{d}) = f_1(\mathbf{z}) + f_2(\mathbf{u}) + \mathbf{d}^T(\mathbf{G}\mathbf{z} - \mathbf{u}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z} - \mathbf{u}||_2^2 \tag{16}$$

This is done in an iterative fashion, such that at step ℓ+1, one minimizes the Lagrangian for one of the variables, while holding the other one fixed at its most recent value, and then alternating, i.e.,

$$\mathbf{z}(\ell + 1) = \underset{\mathbf{z}}{\operatorname{argmin}}\, L_\mu(\mathbf{z}, \mathbf{u}(\ell), \mathbf{d}(\ell)) \tag{17}$$

$$\mathbf{u}(\ell + 1) = \underset{\mathbf{u}}{\operatorname{argmin}}\, L_\mu(\mathbf{z}(\ell + 1), \mathbf{u}, \mathbf{d}(\ell)) \tag{18}$$

where the notation $\mathbf{x}(\ell)$ denotes the vector $\mathbf{x}$ at iteration ℓ. Finally, one updates the dual variable by taking a gradient ascent step to maximize the dual function, resulting in

$$\tilde{\mathbf{d}}(\ell + 1) = \tilde{\mathbf{d}}(\ell) - \mu(\mathbf{G}\mathbf{z}(\ell + 1) - \mathbf{u}(\ell + 1)) \tag{19}$$

The general ADMM steps are outlined in Algorithm 1, using the scaled version of the dual variable $\mathbf{d}_k = \tilde{\mathbf{d}}_k/\mu$, which is more convenient for implementation. Clearly, the ADMM is only relevant when the optimizations in steps 3 and 4 in Algorithm 1 can be carried out easily as compared to the original problem. As it turns out, for many sparse recovery criteria, step 3 will involve solving a problem that is equivalent with a ridge regression least squares problem, solvable with a computational complexity that is approximately the square of the number of observations, but linear in the number of variables, while for step 4, one will

Algorithm 2 Cyclic coordinate descent for a general function f

1: Initiate $\mathbf{z} = \mathbf{z}(0)$, $\ell = 1$. z_k denotes coordinate k in the vector $\mathbf{z}$.
2: **repeat**
3: $\quad z_\ell \leftarrow \underset{z_\ell}{\operatorname{argmin}} f(z_1, \ldots, z_p)$
4: $\quad \ell \leftarrow \ell + 1 \mod p$
5: **until** convergence

often have a close formed solution that can be calculated with a computational complexity that is approximately linear in the number of parameters. In Paper A, we examine the ADMM implementation for the multi-pitch problem in further detail, also discussing how the general ADMM algorithm may be extended to more than two convex functions, as is required there.

4.2 Efficient implementation - the CCD

We proceed to examine the CCD, where the cost function is minimized by keeping all variables fixed except one, separating the optimization problem in a cyclic manner into one sub-problem per variable. In general, the CCD can fail to converge, or may converge very slowly. However, for many of the convex optimization problems commonly arising in sparse modeling, the situation is the opposite, and there even exists convergence proofs for these cases [2, 24]. In fact, in many applications, CCD implementations have empirically been shown to be the fastest algorithm available [25, 26]. The steps involved are outlined in Algorithm 2. Note that a significant performance increase is often possible, especially in batch applications, where a recursive algorithm is needed, by the so called *active set strategy*. The strategy simply involves not updating the parameters that are currently zero in every iteration, and perhaps only doing so once every tenth iteration or so. However, as compared to the ADMM approach, the CCD algorithm has a smaller scope of applicability.

5 Recovery guarantees

Substantial efforts have gone into determining recovery guarantees, statistical efficiency and uniqueness for sparse reconstruction and related problems, with notable contributions being made (primarily) by researchers in mathematics, statist-

ics, and signal processing. Here, we are mainly concerned with applying sparse modeling. As the provable theoretical results are quite pessimistic, posing much stronger restrictions on problem than has been empirically observed, we here only review some of the simpler existing results, as the conditions involved can be seen as giving some clues as to when sparse models can be applicable. The interested reader is referred to [2, 11] for further details. To begin with, we examine the noiseless scenario, i.e.,

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{x}\|_{\ell_0} \quad \text{subject to} \quad y = \mathbf{A}\mathbf{x} \tag{20}$$

where $\mathbf{A} \in \mathbb{R}^{n \times p}$ is assumed to be full rank with $p \gg n$. Clearly, a unique solution to (20) is not possible for every choice of matrix $\mathbf{A}$. As it turns out, the relevant property of $\mathbf{A}$ is how linearly dependent the columns are. We often use the notion of the spark of the matrix $\mathbf{A}$ to describe this.

Definition 1. The spark of a given matrix $\mathbf{A}$ is the smallest number of columns from $\mathbf{A}$ that are linearly dependent.

Thus, if $\mathbf{A}\mathbf{x} = 0$, it implies that $\|\mathbf{x}\|_{\ell_0} \geq \text{spark}(\mathbf{A})$, which coupled with the triangle inequality for the ℓ_0 penalty can be used to prove:

Theorem 1. If a system of linear equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ has a solution obeying $\|\mathbf{x}\|_{\ell_0} < \text{spark}(\mathbf{A})/2$, this solution is necessarily the sparsest possible.

Proof. Assume $\mathbf{x}$ and $\mathbf{y}$ both satisfy the linear system and the spark condition, then $\mathbf{A}(\mathbf{x} - \mathbf{y}) = 0$, and

$$\underbrace{\|\mathbf{x}\|_{\ell_0}}_{<\text{spark}(\mathbf{A})/2} + \underbrace{\|\mathbf{y}\|_{\ell_0}}_{<\text{spark}(\mathbf{A})/2} \geq \|\mathbf{x} - \mathbf{y}\|_{\ell_0} \geq \text{spark}(\mathbf{A}) \tag{21}$$

which is a contradiction, implying that any alternative solution has more than $\text{spark}(\mathbf{A})/2$ non-zero elements. See, e.g., [11] for further details. □

Despite its similarity to the rank of a matrix, the spark is unfortunately very difficult to calculate, in general requiring an infeasible combinatorial search. However, a simple bound of the spark is possible to obtain using the *mutual-coherence*.

Definition 2. The mutual-coherence of a given matrix $\mathbf{A}$ is the largest absolute normalized inner product between different columns from $\mathbf{A}$, where $\mathbf{a}_i$ denotes column i of $\mathbf{A}$

$$\mu(\mathbf{A}) = \underset{1 \leq i,j \leq p,\ i \neq j}{\text{maximize}} \frac{|\mathbf{a}_i^T \mathbf{a}_j|}{\|\mathbf{a}_i\|_{\ell_2} \|\mathbf{a}_i\|_{\ell_2}} \tag{22}$$

here it is possible to show that spark($\mathbf{A}$) $\geq 1 + 1/\mu(\mathbf{A})$. Thus, if one has a heuristic for finding a sparse solution to a linear system and it satisfies Theorem 1, with the spark replaced by this upper bound, one can claim that the found solution is indeed the sparsest one possible. Furthermore, it is possible to show that convex relaxation, using the ℓ_1 norm in (20), will always find any solution that satisfies $\|x\|_{\ell_0} < \frac{1}{2}(1 + 1/\mu(\mathbf{A})$. A similar analysis is possible for the noisy case, but the intuition is the same; the more linearly independent, or smaller the mutual-coherence is, the easier it is to find sparse solution. Similar results exist for the block sparse model, with a bound which depends on a generalization of the mutual-coherence that depends on both the coherence in each block as well as the coherence between different blocks (see, e.g., [27]).

6 Outline of the papers

This section gives an overview of the papers included in the thesis. Paper A through D compose the main contribution of the thesis, treating sparse modeling applied to different parameter estimation problems in audio modeling, audio localizations, DNA sequencing, and spectroscopy, whereas paper E treats robustness, and examines how one may extended the non-parametric idea of multidimensional covariance fitting presented in [28] to fundamental frequency estimation of inharmonic audio sources.

Paper A: Block sparse pitch estimation

In this paper, we consider the estimation of the fundamental frequency of a signal consisting of multiple pitches, i.e., a sum of components, where each component contains a sum of frequencies being integer multiples of the sought-after fundamental frequency. We model the signal using a block structure that groups together the variables that correspond to each possible fundamental frequency, then use sparse heuristics to obtain which groups and thus which fundamental frequencies that are present in the signal. We also present a novel idea to account for a possible ambiguity between the fundamental frequency and the half of that frequency. An efficient implementation is proposed using the ADMM methodology. The method does not need to assumed detailed prior knowledge about model orders, but nevertheless is shown in numerical simulations to attain similar or better results than previously proposed approaches which use such knowledge. The work in Paper A has been published in part as

Stefan Ingi Adalbjörnsson, Andreas Jakobsson, and Mads G. Christensen, "Estimating Multiple Pitches using Block Sparsity", *38th International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26-31, 2013.

and will appear as

Stefan Ingi Adalbjörnsson, Andreas Jakobsson, and Mads G. Christensen, "Multi-pitch estimation exploiting block sparsity", *Elsevier Signal Processing*.

Paper B: Harmonic Audio localization

In the second paper, we consider the problem of localizing multiple audio sources in a possibly reverberant environment using sound measurements obtained by an array of microphones with an arbitrary, but known, geometry. Only considering harmonic sources, the localization is accomplished using a two step procedure. In the first step, a generalization of Paper A to the case of multiple measurement vectors is used to estimate the phases and amplitudes of the pitches in the signal, whereafter a dictionary is created by mapping possible locations to amplitude attenuations and phase offsets for each pitch, using a similar variable grouping as in Paper A. The performance of the resulting algorithm is examined using simulated data, and is shown to attain a performance close to or even following the corresponding Cramér-Rao lower bound. The performance is further evaluated using real audio measurement, and is shown to yield accurate localization estimation also in such situations. The work in Paper B has been published in part as

Ted Kronvall, Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "Joint DOA and Multi-Pitch Estimation Via Block Sparse Dictionary Learning", *22nd European Signal Processing Conference*, Lisbon, Portugal, September 1-5, 2014.

Ted Kronvall, Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "Joint DOA and Multi-Pitch estimation using Block Sparsity", *39th International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 4-9, 2014.

and is submitted for possible publication as

Stefan Ingi Adalbjörnsson, Ted Kronvall, Simon Burgess, Kalle Åström, and Andreas Jakobsson, "Harmonic audio localization", submitted to: *IEEE Journal of Selected Topics in Signal Processing*.

Paper C: Estimation of periodicities in symbolic data

In the third paper, we consider the problem of finding hidden periodicities in symbolic data sequences. Commonly, this problem is approached by mapping the symbolic sequences to complex numbers, after which the analysis is done using various frequency estimation techniques. Our model instead uses a novel sparse logistic regression model to explicitly model the distribution of each symbol. The possible periodicities are thus accounted for by considering a possible change in distribution on each index sets that corresponds to a specific periodicity. Two algorithm are proposed for maximizing the resulting likelihood, the first being a greedy iterative approach that adds one index sets at a time, using an hypothesis testing framework as a stopping criterion, and the second an CCD algorithm that maximizes the penalized maximum likelihood. Using simulated data, the algorithms are shown to have superior performance as compared to previously published methods using simulated sequences. The work in Paper C has been published in part as

> Stefan Ingi Adalbjörnsson, Johan Swärd, Andreas Jakobsson, "Likelihood-based Estimation of Perioditicies in Symbolic Sequences", 21st European Signal Processing Conference, Marrakech, Morocco, September 9-13, 2013.

and is submitted for possible publication as

> Stefan Ingi Adalbjörnsson, Johan Swärd, Jonas Wallin, and Andreas Jakobsson, "Estimating periodicities in symbolic sequences using sparse modelling", submitted to: *IEEE Transactions on Signal Processing*.

Paper D: High resolution sparse estimation of exponentially decaying N-D signals

In the fourth paper, we examine the estimation of N-dimensional damped complex exponentials using sparse heuristics. We introduce the novel idea of using a dictionary learning to iterate between updating the frequency estimation using sparse heuristic, where the dictionary is composed of a grid of possible frequency components with fixed damping, and to updating the damping parameter for each mode using the residual and model fit from the sparse heuristics step. We also show how the model can be implemented using a dictionary composed of Kronecker products of smaller dictionary matrices, each containing the frequency

grid for one of the N-dimensions, resulting in a dramatic decrease in computational complexity. The method is shown to attain similar results as a statistically efficient parametric method, for medium to high signal to noise ratios, for well separated modes, as well as attaining a resolution superior to a zero-padded periodogram for closely space modes. The work in Paper D has been published in part as

> Johan Swärd, Stefan Ingi Adalbjörnsson, Andreas Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Signals", *39th International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 4-9, 2014.
>
> Stefan Ingi Adalbjörnsson, Johan Swärd, Andreas Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Two-Dimensional Signals", *22nd European Signal Processing Conference*, Lisbon, Portugal, September 1-5, 2014.

and is submitted for possible publication as

> Johan Swärd, Andreas Jakobsson, Stefan Ingi Adalbjörnsson, "High Resolution Sparse Estimation of Exponentially Decaying N-Dimensional Signals", submitted to: *IEEE Transactions on Signal Processing*.

Paper E: Joint Model-Order and Fundamental-Frequency Estimation in the Presence of Inharmonicity

In the final paper, we consider the robust estimation of a fundamental frequency for signals where the harmonics are allowed to deviate from being perfect multiples of the fundamental frequency. We approach the problem by adapting the non-parametric robust covariance fitting approach presented in [28] to the parameter uncertainty in the pitch model. Furthermore, we propose a scheme to estimate the commonly unknown number of harmonics. The resulting algorithm is shown to give good results as compared to a previously proposed method. The work in Paper E has been published in part as

> Tommy Nilsson, Stefan Ingi Adalbjörnsson, Naveed R. Butt, Andreas Jakobsson, "Multi-Pitch Estimation of Inharmonic Signals", *21st European Signal Processing Conference*, Marrakech, Morocco, September 9-13, 2013.

Naveed R. Butt, Stefan Ingi Adalbjörnsson, Samuel Somasundaram, Andreas Jakobsson: "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities", *38th International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26-31, 2013.

The extended version presented herein is planned to be submitted as a full paper.

7 Topics for future research

The papers presented in this thesis have been formulated focused on specific applications. However, they often share the common problem of formulating an efficient optimization, exploiting the inherent sparsity of the problem. This structure might of course be found in other problems, offering similar benefits. Thus, one might build on this thesis to form two main possible paths of future research; the first would be to build upon the methods and adapting them to variations on the presented problems, whereas the other would be to find problems that exhibit similar structure.

- As examples of the first path, Paper A has already been extended in [29] to the detection of tonals of rotating machinery. A possible extension of both these works would be a detection algorithm for audio chroma, i.e., the problem of detecting which of the 12 distinct semitones of the musical octave is present in an audio recording. This research direction might lead to possible applications such as automatic music transcription and cover song detection. However, the signal model might need to be modified to take into account some of the niceties of audio signals, such as timbre, amplitude modulation and, perhaps, inharmonicity, which is also an interesting research topic in itself.

- An example of the other path is the following: by interpreting the two stage procedure in paper D as a way of promoting sparsity, such that for each frequency there is only one damping, one allows for similar 2-dimensional parameter estimation problem to be approached in the same manner. Possible applications of this idea could include extending the algorithm in Paper B such that only one source is allowed from any one direction, or extending the work in Paper A, allowing each block of frequencies to depend on a parameter controlling inharmonicity, e.g., the one parameter string inharmonicity model briefly described in Paper E. Similarly, many commonly

used dictionaries, such as the Gabor dictionary, can be characterized in a similar manner, e.g., for the Gabor case, one might restrict every frequency to have only one width.

- An example that could be classified as belonging to both paths could be an extension of Paper A to harmonic audio sources where the fundamental frequency component varies over time. This could be done by using a chirp model for the fundamental frequencies, where the linear change in frequency over time would correspond to the damping parameter reminiscent of what is done in Paper D.

- In paper D, dimensionality reduction using compressive sampling would be a natural extension to the paper (see, e.g., [30, 31]). In broad terms, this approach relies on taking relatively few measurements of the signal using weighted linear combination of samples in a given basis, usually taken to consist of identically distributed random variables, and then performing the analysis on these samples.

- The robust estimator presented in paper E also has some interesting open questions: for example, a straightforward analysis reveals that if only one of the steering vectors/harmonics is allowed to vary, the non-convex criterion can be seen to be similar to a non-convex quadratic optimization problem, with two quadratic constraints, which is a problem recently shown not to have a duality gap [32], allowing for some interesting interpretations and perhaps optimization options.

- In Paper E, one might be able to decrease the high computational cost of the algorithm, by considering either an ADMM formulation or alternatively, interior point methods specifically geared toward convex optimization problems involving the log-determinant (see, e.g., [33]).

Clearly the field is abundant with open research problem, just waiting to be examined, offering a range of interesting and challenging topics to study. What are you waiting for? Lets get on with it! But first, lets proceed to examine the five papers constituting this thesis.

References

[1] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, 1979.

[2] P. G. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*, Springer Series in Statistics. Springer, 2011.

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

[4] I. F. Gorodnitsky and B. D. Rao, "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, March 1997.

[5] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul, jul 2008, pp. 10225–10229.

[6] D. Malioutov, M. Cetin, and A. S. Willsky, "A Sparse Signal Reconstruction Perspective for Source Localization With Sensor Arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, August 2005.

[7] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse Learning via Iterative Minimization With Application to MIMO Radar Imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1088–1101, March 2011.

[8] E. Gudmundson, Jun Ling, P. Stoica, Jian Li, and A. Jakobsson, "Spectral Estimation of Damped Sinusoids in the Case of Irregularly Sampled Data," in *Proceedings of the 9th International Symposium on Signals, Circuits and Systems (ISSCS 2009)*, Iasi, Romania, July 9-10 2009.

[9] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modelling of Signals and Images," *SIAM Review*, vol. 51, 2009.

[10] P. Stoica and Y. Selén, "Model-order Selection — A Review of Information Criterion Rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[11] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.

[12] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, pp. 129–159, 2001.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[15] D. Bertsekas, *Convex Optimization Theory*, Athena Scientific, 2009.

[16] R. Chartrand and B. Wohlberg, "A Nonconvex ADMM Algorithm for Group Sparsity with Sparse Groups," in *38th IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing*, 2013.

[17] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[18] M. Grant, *Disciplined Convex Programming*, Ph.D. thesis, Information Systems Laboratory, Department of Electrical Engineering, Stanford University, 2004.

[19] L. Liberti and N. Maculan, Eds., *Global Optimization: From Theory to Implementation*, Nonconvex Optimization and its Applications. Springer, 2006.

[20] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," `http://cvxr.com/cvx`, Sept. 2012.

[21] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, August 1999.

[22] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[23] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[24] P. Tseng, "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization," *Journal of Optimization Theory and Applications*, vol. 109, no. 3, pp. 475–494, 2001.

[25] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise Coordinate Optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.

[26] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[27] Y. V. Eldar, P. Kuppinger, and H. Bolcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3042–3054, 2010.

[28] M. Rübsamen and A. B. Gershman, "Robust Adaptive Beamforming Using Multidimensional Covariance Fitting," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 740–753, Feb. 2012.

[29] Lu. Wang, C. Wan, S. Li, and G. Bi, "Harmonic tonal detectors based on the BOGA," *Elsevier Signal Processing*, vol. 106, no. 0, pp. 215 – 230, 2015.

[30] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, March 2008.

[31] E. J. Candes, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[32] A. Beck and Y. C. Eldar, "Doubly Constrained Robust Capon Beamformer With Ellipsoidal Uncertainty Sets," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 753–758, Jan. 2007.

[33] L. Vandenberghe, S. Boyd, and S. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, pp. 499–533, 1998.

A

Paper A

Multi-pitch estimation exploiting block sparsity

Stefan Ingi Adalbjörnsson[1], Andreas Jakobsson[1], and Mads G. Christensen[2]

[1] *Centre for Mathematical Sciences, Lund University, Lund, Sweden*
[2] *Audio Analysis Lab, Dept. of Arch., Design & Media Technology, Aalborg University, Denmark*

Abstract

We study the problem of estimating the fundamental frequencies of a signal containing multiple harmonically related sinusoidal components using a novel block sparse signal representation. An efficient algorithm for solving the resulting optimization problem is devised exploiting a novel variable step-size alternating direction method of multipliers (ADMM). The resulting algorithm has guaranteed convergence and shows notable robustness to the f_0 vs $f_0/2$ ambiguity problem. The superiority of the proposed method, as compared to earlier presented estimation techniques, is demonstrated using both simulated and measured audio signals, clearly indicating the preferable performance of the proposed technique.

Key words: Pitch estimation, block sparsity, total variation, spectral smoothness, order estimation.

1 Introduction

Estimating the fundamental frequency of harmonically related signals form an integral part in a wide range of signal processing applications, and perhaps especially so in speech and audio processing. For example, the fundamental frequency, or *pitch*, is necessary when forming the long-term prediction used in linear prediction-based speech codecs [1], and is similarly the key component in music information retrieval applications, such as automatic music transcription, and in musical genre classification [2]. The fundamental frequency is also of notable importance in problem such as source separation, enhancement, compression, and classification (see, e.g., [3, 4] and the references therein), as well as in several biomedical, mechanical and acoustic applications, and the topic has for these reasons attracted a notable interest during the recent decades. Commonly, the pitch estimate is formed assuming a single source model, such that only a single fundamental frequency and its harmonics are assumed to be present in the signal, using different kinds of similarity measures, such as the cross-correlation, cepstrum, or the average squared difference function (see, e.g., [5–11]), although notable exceptions treating the multi-pitch problem can be found in, e.g., [3, 12–23]. Regrettably, the problem is hard, and most of these techniques will suffer from not yielding unique estimates even in the ideal case, even for a single source, and/or will typically also require perfect *a priori* knowledge of both the number of sources and the model order of each of these sources. Often, such limitations necessitate notable post-processing or correction steps in order to improve on an initially poor pitch estimate. In this work, we focus on improving the initial pitch estimate, proposing a novel multi-pitch estimation approach making no *a priori* model order assumptions. The method is based on a sparse signal recovery framework, wherein a signal is assumed to consist of only a small number of components from a large set of potential signal vectors. This approach has been found to yield high quality estimates in a wide variety of fields (see, e.g., [24–26]), and has also earlier been exploited in machine learning settings, where sparse modeling of pitch signals is accomplished by learning a dictionary of pitches from a training data set (see, e.g., [16, 21, 22]). For sinusoidal signals, it was early on shown that using a sparse representation technique allowed for high resolution frequency estimates; typical examples include [27, 28], wherein the sparse signal reconstruction from noisy observations was accomplished with the by now well-known sparse least squares (LS) technique. A similar approach may clearly also be applied to the pitch estimation problem, although one is then not fully exploiting the harmonic

signal structure. Herein, we instead propose a novel block sparse signal representation, such that each signal source is grouped in one data block for each pitch frequency. By then extending the representation to all considered pitch frequencies, reminiscent to the extended dictionaries used in, e.g., [13, 27, 29], the resulting model will be sparse in the sense that it will be formed from only a few of the possible blocks in the dictionary. Different from estimates such as the ones presented in [16, 21, 22], the presented method does not exploit any training data, with the method inferring the pitch parameters and the model orders from the spectral content of the signal. The proposed pitch estimation method instead exploits the group sparse structure, without requiring any prior knowledge of either the number of sources present, or their number of harmonics. The presented algorithm, in its presented form, does not take into account for any possible inharmonicity in the pitch structure, such that the higher order frequencies would not occur precisely as a multiple of the fundamental frequency. Such inharmonicities are common in audio signals, and should be taken into account for such signals. As we are here focusing on the general problem, occurring also for numerous other forms of signals, we have here opted to exclude the treatment of inharmonicity, although note that the algorithm may be extended to allow for this along the lines presented in [30, 31], or using a dictionary learning approach such as in [32, 33]. The theoretical study of block sparse signals was initially suggested in [34], where it is shown that including this structure in the estimation procedure has great practical consequences, improving both theoretical recovery limits and numerical results in many cases (see, e.g., [34–37]). Generally, this form of group sparse convex optimization problems are computationally cumbersome; for this reason, we also derive an efficient algorithm to form the estimate based on the alternating directions methods of multipliers (ADMM) (see, e.g., [38, 39]). The resulting algorithm will have a guaranteed convergence as well as exhibit a significant robustness to the common problem of the f_0 vs $f_0/2$ ambiguity, i.e., when a pitch candidate at half the nominal frequency fits the observed signal as well, or possibly even better, than the true pitch frequency. The remainder of this paper is organized as follows: in the next section, we briefly present the data model. Then, in Section 3, we introduce the proposed pitch estimation technique. Section 4 introduces the efficient ADMM-based implementation, and Section 5 includes numerical evaluations of the proposed method as compared to earlier techniques. Finally, Section 6 concludes on the work.

2 Block sparse signal model

Consider a complex-valued signal, $y(n)$, consisting of K harmonically related (signal) sources with fundamental frequencies f_k, for $k = 1, \ldots, K$, such that (see also [3])

$$y(n) = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} a_{k,\ell} e^{j2\pi f_k \ell n} + e(n) \tag{1}$$

for $n = 1, \ldots, N$, where $a_{k,\ell}$ and L_k denote the (complex-valued) amplitude of the ℓ:th harmonic of the k:th source, and the number of harmonically related sinusoids for the k:th source, respectively, and where $e(n)$ is an additive noise term, here assumed to be an identically distributed independent circularly symmetric complex Gaussian process with variance σ_e^2. It is worth noting that due to the restriction of the allowed frequency range, the number of harmonics are restricted as a function of the fundamental frequency, such that $L_k < \lfloor 1/f_k \rfloor$, $\forall k$, where $\lfloor \cdot \rfloor$ denotes the round-down to nearest integer operation. Let

$$\mathbf{y} = \begin{bmatrix} y(1) & \ldots & y(N) \end{bmatrix}^T \tag{2}$$

where $(\cdot)^T$ denotes the transpose. Then, (1) may be expressed succinctly as

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{V}_k \mathbf{a}_k + \mathbf{e} \stackrel{\Delta}{=} \mathbf{W}\mathbf{a} + \mathbf{e} \tag{3}$$

where $\mathbf{e}$ is a vector of noise terms constructed in the same manner as $\mathbf{y}$, and

$$\mathbf{W} = \begin{bmatrix} \mathbf{V}_1 & \ldots & \mathbf{V}_K \end{bmatrix} \tag{4}$$

$$\mathbf{V}_k = \begin{bmatrix} \mathbf{z}_k & \mathbf{z}_k^2 & \ldots & \mathbf{z}_k^{L_k} \end{bmatrix} \tag{5}$$

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1^T & \ldots & \mathbf{a}_K^T \end{bmatrix}^T \tag{6}$$

$$\mathbf{a}_k = \begin{bmatrix} a_{k,1} & \ldots & a_{k,L_k} \end{bmatrix}^T \tag{7}$$

with the vector powers, $\mathbf{z}_k^\ell$, being evaluated element-wise,

$$\mathbf{z}_k^\ell = \begin{bmatrix} e^{j2\pi f_k \ell} & \ldots & e^{j2\pi f_k N \ell} \end{bmatrix}^T \tag{8}$$

Reminiscent to the models considered for line spectra (see, e.g., [13, 27, 29]), the matrix $\mathbf{W}$ may be expanded to be formed instead over a (large) range of possible

fundamental frequencies, ν_ℓ, for $\ell = 1, \ldots, P$, where P denotes the total number of considered frequencies, such that the corresponding amplitude vector, $\mathbf{a}$, will have elements different from zero only for those frequencies actually coinciding with the frequencies in the signal. Thus, for the signal in (1), for each source in the signal, there will be a corresponding non-zero block in the amplitude vector, i.e., if the source has fundamental frequency ν_ℓ, the sub-block $\mathbf{a}_\ell$ will be non-zero. It should be noted that this formulation thus implicitly assumes that P is selected large enough so that the true pitch frequencies lie close to the used grid. Practical experience with similar methods, e.g., [27, 28], shows that they are quite robust to this approximation (see also the related discussions in [29, 40, 41]). Given the structure of (3), the resulting approximation of the signal is not only sparse, but thus also *block sparse*, since for each source present, several harmonics will be included in the signal.

3 Pitch estimation using block sparsity

Reminiscent of the block sparse formulations introduced in [34], one may thus form an estimate of the present sources as

$$\underset{\mathbf{a}}{\text{minimize}} \ \frac{1}{2}||\mathbf{y} - \mathbf{W}\mathbf{a}||_2^2 + \alpha \sum_{k=1}^{P} ||\mathbf{a}_k||_2 \tag{9}$$

where $||\cdot||_p$ denotes the ℓ_p norm, and with $\alpha > 0$ denoting a tuning parameter that controls the relative importance of the block sparsity promoting ℓ_2 norm and the squared ℓ_2 norm fitting term, discussed further below. It should be noted that the cost function is clearly convex as it is a sum of a norm and the composition of a norm and an affine function. The second term in (9) is included to promote a block sparse solution, i.e., a solution with the property that most blocks, $\mathbf{a}_i$, are zero (see also Appendix A). As noted, the number of harmonics of each source, L_k, is generally not known, and to be able to use the presented sparse approximation model, one needs to set some maximum allowed number of harmonics for all possible fundamental frequencies, say $L_{\max}$. This implies that the data blocks, $\mathbf{a}_k$, as given in (7), will typically contain some amplitudes that are close to zero, for those harmonics that are not present in the source signal. To allow for this, we introduce a further ℓ_1 penalty term, generally forcing small amplitudes to zero, resulting in the following sparse group lasso (see also [42] and the discussion in

Appendix B)

$$\underset{\mathbf{a}}{\text{minimize}}\ \frac{1}{2}||\mathbf{y}-\mathbf{W}\mathbf{a}||_2^2+\lambda||\mathbf{a}||_1+\alpha\sum_{k=1}^{P}||\mathbf{a}_k||_2 \tag{10}$$

where $\alpha > 0$ is a tuning parameter. Using the formulation in (10), this would imply that the (generic) f_0 harmonics will make up a subset of the block detailing the $f_0/2$ harmonics, i.e., the frequencies $\{f_0, 2f_0, 3f_0, ..., L_{f_0}f_0\}$ will be present in both blocks, and thus the minimization in (10) will then in all cases prefer the block corresponding the lower frequency. In order to partly resolve this problem, we introduce a further scaling of the norms in the minimization, such that the blocks are given comparable weights, instead forming the minimization as

$$\underset{\mathbf{a}}{\text{minimize}}\ \frac{1}{2}||\mathbf{y}-\mathbf{W}\mathbf{a}||_2^2+\lambda||\mathbf{a}||_1+\alpha\sum_{k=1}^{P}\sqrt{L_k}||\mathbf{a}_k||_2 \tag{11}$$

However, this does not completely remove the ambiguity from the model since one might well consider, in certain scenarios, restricting the maximum number of allowed harmonics such that the sub-vectors corresponding to some f_0 and $f_0/2$ could have the same number of elements. Thus, a signal composed of a fundamental frequency f_0, with L_{f_0} harmonics, can be written interchangeably using the first L_{f_0} elements of the sub-vector corresponding to the fundamental frequency f_0, or every other element of the first $2L_{f_0}$ elements of the sub-vector corresponding to $f_0/2$. By instead including a *total variation* penalty function

$$\text{Tv}(\mathbf{a}_k)=\sum_{i=1}^{L_{k-1}}\left|a_{k,i}-a_{k,i+1}\right|$$

in the cost function, blocks with constant amplitudes will not be penalized, whereas $f_0/2$ vectors, such as $\mathbf{a}_{f_0/2}$ mentioned above, will incur a large penalty. The resulting spectral smoothnes is similar to often imposed assumption in the modeling of audio signals, see e.g., [14]. Note that the total variation function is convex since it may be written as composition of an affine function, say $\mathbf{F}$, and the ℓ_1 norm, i.e.,

$$\sum_{k=1}^{P}\text{Tv}(\mathbf{a}_k)=||\mathbf{F}\mathbf{a}||_1 \tag{12}$$

where $\mathbf{F} \in \mathbb{R}^{\sum_{k=1}^{P} L_k \times \sum_{k=1}^{P} L_k}$ is created such that the rows corresponding to the first $L_k - 1$ elements of each block have a one on the diagonal and minus one on the first super-diagonal, and the row corresponding to element L_k is zero, or equivalently, a difference operator with rows $L_1, L_1 + L_2 \ldots, \sum_{k=1}^{P} L_k$ set to zero. Thus, we propose forming the pitch estimate via the minimization

$$\underset{\mathbf{a}}{\text{minimize}} \ \frac{1}{2}||\mathbf{y} - \mathbf{W}\mathbf{a}||_2^2 + \lambda||\mathbf{a}||_1 + \alpha \sum_{k=1}^{P} ||\mathbf{a}_k||_2 + \gamma \sum_{k=1}^{P} \text{Tv}(\mathbf{a}_k) \qquad (13)$$

where $\gamma > 0$ is a tuning parameter, which should be set small enough such that the total effect of adding the TV term is only to resolve the f_0 and $f_0/2$ ambiguity in a consistent and correct manner; in the numerical section it was set to 0.01 for all simulations. The tuning parameters, λ and α may, for instance, be estimated for example with a cross validation approach. However, in our experience, if the signal to noise ratio (SNR) is high enough, they may preferably be set by simply inspecting the amplitudes in the zero padded discrete Fourier transform, as is shown in Appendix B, i.e., by setting α as the smallest significant amplitude above the noise floor, and by setting λ similarly, but for each pitch. It is worth noting that an alternative formulation may be obtained by instead using a covariance fitting formulation; as recently shown in [43, 44], the sparse SPICE covariance fitting algorithm [45] may be equivalently expressed using an weighted penalized ℓ_1 formulation, for a particular choice of λ. One may similarly form a covariance fitting style minimization of the here proposed minimization by replacing the squared ℓ_2 fitting term in (11) or (13) with a corresponding ℓ_1 fitting term; we will below examine what such a choice would imply. Reminiscent of the work in [28, 46–48], another approach would be to instead consider other penalties, e.g., the ℓ_q penalties with $0 < q < 1$, or the reweighted ℓ_1, which would both lead to non-convex optimization problems, that can nevertheless often be efficiently solved with the benefit of, in many cases, sparser solutions, with less biased amplitude estimates, although with local minima being a recurring problem and without the global optimality conditions of convex optimization problems. Herein, given that our main objective is the estimation of the non-linear fundamental frequency parameters, we restrict our attention to convex criteria, but note that especially the reweighted ℓ_1 algorithm and the ℓ_q-like criteria suggested in [46] are easily adapted to the algorithm and the here presented criteria. Considering that the signals of interest are only approximately sparse in $\mathbf{W}$, and as two closely spaced fundamental frequencies will result in that the correspond-

ing matrices, $\mathbf{W}_s$ and $\mathbf{W}_r$, will be rather similar, one cannot expect the resulting (block) pseudo spectral solution, formed over the peaks of the 2-norm of the estimated amplitudes, $||\hat{\mathbf{a}}_k||_2$, to have exactly as many non-zero blocks as there are sources present in the signal. In order to determine the number of sources present, we therefore introduce a novel BIC-style criterion, such that the number of sources are selected as (cf. [29, 49])

$$\hat{K} = \underset{k \in [1, K_{\max}]}{\operatorname{argmin}} \, \mathrm{BIC}_k(\lambda, \alpha) \tag{14}$$

where $K_{\max}$ denotes the maximum number of considered sources, here selected as the number of peaks present in the initially obtained (block) pseudo-spectra, and where the (λ, α)-dependent BIC cost function is formed as

$$\mathrm{BIC}_k(\lambda, \alpha) = 2N \ln(\hat{\sigma}_k^2) + (2H_k + 1) \ln(N) \tag{15}$$

with $\hat{\sigma}_k^2$ denoting the variance of the estimation residual when modeling the pitch signal using

$$H_k = \sum_{\ell=1}^{k} \hat{L}_{k_\ell} \tag{16}$$

(dependent) sinusoidal components (see also [3]), where $\hat{L}_{k_\ell}$ is the number of frequencies corresponding to the non-zero elements of $\hat{\mathbf{a}}_{k_\ell}$. It should be stressed that the $\hat{L}_{k_\ell}$ considered harmonics are not necessarily consecutive, thereby allowing for the case of missing harmonics (including the possibility that the signal lacks the fundamental frequency component), which is a case commonly occurring in many form of acoustic signals.

4 An efficient ADMM implementation

As the minimizations in (11) and (13) are composed of simple convex functions, they may be solved using one of the freely available interior point based solvers, such as SeDuMi [50] and SDPT3 [51], although such solvers will scale badly both with increased data length and with the use of a finer grid size for the fundamental frequency. As a result, such a solution will in many cases be too computationally intensive to be practically useful. In order to form a more efficient implementation, we therefore reformulate the minimization in (11) using an ADMM formulation, which may be used to solve convex optimization problems which are

Algorithm 1 The general ADMM algorithm

1: Initiate $\mathbf{z} = \mathbf{z}(0), \mathbf{u} = \mathbf{u}(0)$, and $\ell = 0$
2: **repeat**
3: $\quad \mathbf{z}(\ell+1) = \underset{\mathbf{z}}{\operatorname{argmin}} f_1(\mathbf{z}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z} - \mathbf{u}(\ell) - \mathbf{d}(\ell)||_2^2$
4: $\quad \mathbf{u}(\ell+1) = \underset{\mathbf{u}}{\operatorname{argmin}} f_2(\mathbf{u}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z}(\ell+1) - \mathbf{u} - \mathbf{d}(\ell)||_2^2$
5: $\quad \mathbf{d}(\ell+1) = \mathbf{d}(\ell) - (\mathbf{G}\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1))$
6: $\quad \ell \leftarrow \ell + 1$
7: **until** convergence

the sum of two convex functions by decomposing the optimization into two simpler problems, which are then solved in an iterative fashion (see, e.g., [38]). For completeness and to introduce our notation, we here include a brief outline of the main steps involved.

Consider the convex optimization problem

$$\underset{\mathbf{z}}{\text{minimize}} \quad f_1(\mathbf{z}) + f_2(\mathbf{G}\mathbf{z}) \tag{17}$$

where $\mathbf{z} \in \mathbb{R}^p$ is the optimization variable, $f_1(\cdot)$ and $f_2(\cdot)$ are convex functions, and $\mathbf{G} \in \mathbb{R}^{N \times p}$ is a known matrix. If one introduces an auxiliary variable, $\mathbf{u}$, then (17) may be equivalently be expressed as

$$\begin{aligned} &\underset{\mathbf{z},\,\mathbf{u}}{\text{minimize}} \quad f_1(\mathbf{z}) + f_2(\mathbf{u}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z} - \mathbf{u}||_2^2 \\ &\text{subject to} \quad \mathbf{G}\mathbf{z} - \mathbf{u} = \mathbf{0} \end{aligned} \tag{18}$$

Under the assumption that there is no duality gap, which is true for all the optimization problems considered herein, one can solve the optimization problem via the dual function defined as the infimum with respect to $\mathbf{u}$ and $\mathbf{z}$ of the augmented Lagrangian [38]

$$L_\mu(\mathbf{z}, \mathbf{u}, \mathbf{d}) = f_1(\mathbf{z}) + f_2(\mathbf{u}) + \mathbf{d}^T(\mathbf{G}\mathbf{z} - \mathbf{u}) + \frac{\mu}{2}||\mathbf{G}\mathbf{z} - \mathbf{u}||_2^2 \tag{19}$$

which holds for all μ, since at any feasible point $||\mathbf{G}\mathbf{z} - \mathbf{u}||_2^2 = 0$. The ADMM does this by iteratively maximizing the dual function, such that at step ℓ+1, one minimizes the Lagrangian for one of the variables, while holding the other fixed

at its most recent value, i.e.,

$$\mathbf{z}(\ell+1) = \underset{\mathbf{z}}{\operatorname{argmin}}\, L_\mu(\mathbf{z}, \mathbf{u}(\ell), \mathbf{d}(\ell)) \tag{20}$$

$$\mathbf{u}(\ell+1) = \underset{\mathbf{u}}{\operatorname{argmin}}\, L_\mu(\mathbf{z}(\ell+1), \mathbf{u}, \mathbf{d}(\ell)) \tag{21}$$

where the notation $\mathbf{x}(\ell)$ denotes the vector $\mathbf{x}$ at iteration ℓ. Finally one updates the dual variable by taking a gradient ascent step to maximize the dual function, resulting in

$$\tilde{\mathbf{d}}(\ell+1) = \tilde{\mathbf{d}}(\ell) - \mu(\mathbf{G}\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1)) \tag{22}$$

from which the interpretation of μ as the dual variable step size may be seen (see also [38] for further details). The general ADMM steps are outlined in Algorithm 1, using the scaled version of the dual variable $\mathbf{d}_k = \tilde{\mathbf{d}}_k/\mu$, which is more convenient for implementation. As a stopping criterion, it is shown in [38] that by studying the necessary and sufficient conditions for the optimality of a solution, say $\mathbf{z}^*$, $\mathbf{u}^*$, and $\mathbf{d}^*$, of the minimization in (18), i.e., the primal feasibility

$$\mathbf{G}\mathbf{z}^* - \mathbf{u}^* = \mathbf{0} \tag{23}$$

and the dual feasibility

$$\mathbf{0} \in \partial f_1(\mathbf{z}^*) + \mathbf{G}^T\mathbf{d}^* \tag{24}$$

$$\mathbf{0} \in \partial f_2(\mathbf{d}^*) - \mathbf{d}^* \tag{25}$$

where ∂ is the sub-differential operator, imply that the so-called primal and dual residuals, which are defined as $\mathbf{r}_k = \mathbf{G}\mathbf{z}_k - \mathbf{u}_k$ and $\mathbf{s}_k = \mu\mathbf{G}^T(\mathbf{u}_k - \mathbf{u}_{k-1})$, respectively, will converge to zero. Thus, as a stopping criterion, one may use that the norm of the primal and dual residuals are small enough. Clearly, the ADMM is only relevant when the optimizations in steps 3 and 4 in Algorithm 1 can be carried out easily as compared to the original problem. We begin by examining the implementation of (11), and then proceed to extending this to form (13). One possibility to reformulate (11) in this fashion would be to choose $f_1(\cdot)$ as the 2-norm fitting term and $f_2(\cdot)$ as the sum of the sparse regularization term, i.e., with $\mathbf{G} = \mathbf{I}$ and

$$f_1(\mathbf{z}) = \frac{1}{2}||\mathbf{y} - \mathbf{W}\mathbf{z}||_2^2 \tag{26}$$

$$f_2(\mathbf{u}) = \lambda||\mathbf{u}||_1 + \alpha\sum_{k=1}^{P}\sqrt{\Delta_k}||\mathbf{u}_k||_2 \tag{27}$$

which yields

$$\mathbf{z}(\ell+1) = \underset{\mathbf{z}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{y} - \mathbf{W}\mathbf{z}||_2^2 + \frac{\mu}{2}||\mathbf{z} - \mathbf{u}(\ell) - \mathbf{d}(\ell)||_2^2 \tag{28}$$

$$= \left(\mathbf{W}^H\mathbf{W} + \mu\mathbf{I}\right)^{-1}\left(\mathbf{W}^H\mathbf{y} - \mathbf{u}(\ell) - \mathbf{d}(\ell)\right) \tag{29}$$

where $(\cdot)^H$ denotes the Hermitian (conjugate) transpose. It should be noted that the matrix inversion lemma can be used such that the solution can be calculated by solving an $N \times N$ system corresponding to the matrix $\mathbf{W}\mathbf{W}^H + \mathbf{I}/\mu$, i.e.,

$$\left(\mathbf{W}^H\mathbf{W} + \mu I\right)^{-1}\boldsymbol{\varkappa} = \frac{\mathbf{y}}{\mu} + 1/\mu\mathbf{W}^H\left(\mathbf{I}/\mu + \mathbf{W}\mathbf{W}^H\right)^{-1}\mathbf{W}\boldsymbol{\varkappa} \tag{30}$$

for some vector $\boldsymbol{\varkappa} \in \mathbb{C}^p$, thus transforming the $P \times P$ matrix inversion into that of an $N \times N$ matrix inversion. Moreover,

$$\mathbf{u}(\ell+1) = \underset{\mathbf{u}}{\operatorname{argmin}}\, \lambda||\mathbf{u}||_1 + \alpha\sum_{k=1}^{P}\sqrt{\Delta_k}||\mathbf{u}_k||_2 + \frac{\mu}{2}||\mathbf{z}(\ell+1) - \mathbf{u} - \mathbf{d}(\ell)||_2^2 \tag{31}$$

which decouples into P optimization problems as

$$\mathbf{u}_k(\ell+1) = \underset{\mathbf{u_k}}{\operatorname{argmin}}\, \lambda||\mathbf{u}_k||_1 + \alpha\sqrt{\Delta_k}||\mathbf{u}_k||_2 + \frac{\mu}{2}||\mathbf{z}_k(\ell+1) - \mathbf{u}_k - \mathbf{d}_k(\ell)||_2^2 \tag{32}$$

Here one can solve the sub-differential equations

$$\lambda\mathbf{r} + \alpha\sqrt{\Delta_k}\mathbf{s} + \mu(\tilde{\mathbf{z}}(\ell+1) - \tilde{\mathbf{u}}_k - \tilde{\mathbf{d}}(\ell)) = 0 \tag{33}$$

where the notation $\tilde{\mathbf{x}}$ denotes the real valued version of the complex vector $\mathbf{x}$, created as specified in Appendix A, and the vectors $\mathbf{s}$ and $\mathbf{r}$ are real-valued and are defined such that

$$\mathbf{s} = \begin{cases} \frac{\tilde{\mathbf{u}}_k}{||\tilde{\mathbf{u}}_k||_2} & \text{if } \tilde{\mathbf{u}}_k \neq 0 \\ \mathbf{v} & \text{otherwise} \end{cases} \tag{34}$$

with $||\mathbf{v}||_2 \leq 1$, and

$$\begin{bmatrix} r_i \\ r_{i+L_k} \end{bmatrix} = \begin{cases} \frac{\left[\tilde{\mathbf{u}}_{k,i},\, \tilde{\mathbf{u}}_{k,i+L_k}\right]^T}{\left\|\left[\tilde{\mathbf{u}}_{k,i},\, \tilde{\mathbf{u}}_{k,i+L_k}\right]\right\|_2} & \text{if } \left[\tilde{\mathbf{u}}_{k,i},\, \tilde{\mathbf{u}}_{k,i+L_k}\right]^T \neq 0 \\ \mathbf{p}_i & \text{otherwise} \end{cases} \tag{35}$$

Algorithm 2 PEBS$_2$ via ADMM

1: Initiate $\mathbf{z} = \mathbf{z}(0), \mathbf{u} = \mathbf{u}(0)$, and $\ell := 0$
2: **repeat**
3: $\quad \mathbf{z}(\ell+1) = \left(\mathbf{W}^H\mathbf{W} + \mu\mathbf{I}\right)^{-1}\left(\mathbf{W}^H\mathbf{y} - \mathbf{u}(\ell) - \mathbf{d}(\ell)\right)$
4: $\quad \mathbf{u}(\ell+1) = \bar{\Psi}\left(\Psi\left(\mathbf{z}(\ell+1) - \mathbf{d}(\ell+1), \frac{\lambda}{\mu}\right), \frac{\alpha\sqrt{\Delta_k}}{\mu}\right)$
5: $\quad \mathbf{d}(\ell+1) = \mathbf{d}(\ell) - (\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1))$
6: $\quad \ell \leftarrow \ell + 1$
7: **until** convergence

with $||\mathbf{p}_i||_2 \leq 1$, for $i = 1, \ldots, L_k$, where $\mathbf{a}_{i,j}$ denotes element j of sub-vector i and $[a,\ b]$ denoting a vector with two scalars a and b, and

$$\mathbf{r} = \begin{bmatrix} r_1 & \ldots & r_{2L_k} \end{bmatrix}^T \tag{36}$$

This leads to

$$\mathbf{u}(\ell+1) = \bar{\Psi}\left(\Psi\left(\mathbf{z}(\ell+1) - \mathbf{d}(\ell), \frac{\lambda}{\mu}\right), \frac{\alpha\sqrt{\Delta_k}}{\mu}\right) \tag{37}$$

where $\Psi(\cdot)$ is an element-wise shrinkage function, defined as

$$\Psi(\mathbf{a}, \gamma) = \frac{\max(|\mathbf{a}| - \gamma, 0)}{\max(|\mathbf{a}| - \gamma, 0) + \gamma} \odot \mathbf{a} \tag{38}$$

where the max function acts element-wise on the vector, and $\odot$ denotes the element-wise multiplication of two vectors. Similarly, $\bar{\Psi}(\cdot)$ is a vector shrinkage functions formed as

$$\bar{\Psi}(\mathbf{a}, \gamma) = \frac{\max(||\mathbf{a}||_2 - \gamma, 0)}{\max(||\mathbf{a}||_2 - \gamma, 0) + \gamma}\mathbf{a}$$

The resulting ADMM algorithm for (11), here termed the Pitch Estimation using ℓ_2 norm and Block Sparsity (PEBS$_2$), is summarized in Algorithm 2. For (13), one could similarly define $f_1(\cdot)$ as the sum of all the regularization terms. However, the subdifferential equations can then unfortunately not be solved as easily as before. Instead, we exploit the recent idea introduced in [39], where, by a clever choice of functions the $f_1(\cdot)$ and $f_2(\cdot)$, one may extend (17) to a minimization of a sum

of B convex functions, i.e.,

$$\underset{\mathbf{z}}{\text{minimize}} \quad \sum_{k=1}^{B} g_k(\mathbf{Hz}) \tag{39}$$

where $\mathbf{H}_k \in \mathbf{R}^{N\times p}$ are known matrices, and $g_k(\cdot)$ convex functions. This is accomplished by setting $f_1(\mathbf{z}) = 0$, and

$$f_2(\mathbf{Gu}) = \sum_{k=1}^{B} g_k(\mathbf{Gu}) = \sum_{k=1}^{B} g_k(\mathbf{H}_k\mathbf{u}^{(k)}) \tag{40}$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{H}_1^T & \dots & \mathbf{H}_K^T \end{bmatrix}^T \tag{41}$$

$$\mathbf{u} = \begin{bmatrix} (\mathbf{u}^{(1)})^T & \dots & (\mathbf{u}^{(K)})^T \end{bmatrix}^T \tag{42}$$

Thereby step 4 in Algorithm 1 is allowed to be decomposed into B independent optimization problems. Rewriting (13) on the form in (39), noting that for this case, $B = 3$, and

$$f_2(\mathbf{Gu}) = \frac{1}{2}||\mathbf{u}^{(1)} - \mathbf{y}|| + \lambda||\mathbf{u}^{(2)}||_1 + \alpha\sum_{k=1}^{P}\sqrt{\Delta_k}||\mathbf{u}_k^{(2)}||_2 + \gamma||\mathbf{u}^{(3)}||_1 \tag{43}$$

where $\mathbf{G} = \begin{bmatrix} \mathbf{A}^T & \mathbf{I} & \mathbf{F}^T \end{bmatrix}^T$, and

$$\mathbf{u} = \begin{bmatrix} (\mathbf{u}^{(1)})^T & (\mathbf{u}^{(2)})^T & (\mathbf{u}^{(3)})^T \end{bmatrix}^T \tag{44}$$

This implies that step 3 in Algorithm 1 can be solved as

$$\mathbf{z}(\ell+1) = \underset{\mathbf{z}}{\text{argmin}}\, ||\mathbf{Gz} - \mathbf{u}(\ell) - \mathbf{d}(\ell)||_2^2 \tag{45}$$

$$= \left[\mathbf{A}^H\mathbf{A} + \mathbf{F}^H\mathbf{F} + \mathbf{I}\right]^{-1}\left(\mathbf{A}^H\boldsymbol{\xi}^{(1)}(\ell) + \mathbf{F}^H\boldsymbol{\xi}^{(2)}(\ell) + \boldsymbol{\xi}^{(3)}(\ell)\right) \tag{46}$$

where $\mathbf{d}$ is decomposed in the same manner as $\mathbf{u}$, and

$$\xi^{(m)}(\ell) \triangleq \mathbf{u}^{(m)}(\ell) + \mathbf{d}^{(m)}(\ell) \tag{47}$$

for $m = 1, 2, 3$. Here, we are mostly interested in situations where the number of parameters far outnumbers the number of measurements, i.e., $N \ll p$. Thus, since (45) needs to be solved at each iteration, one may solve it efficiently using the matrix inversion lemma, i.e.,

$$\mathbf{z}(\ell+1) = \chi(\ell) - \left(\mathbf{F}^H\mathbf{F} + \mathbf{I}\right)^{-1}\mathbf{A}^H\left(\mathbf{I} + \mathbf{A}\left(\mathbf{F}^H\mathbf{F} + \mathbf{I}\right)^{-1}\mathbf{A}^H\right)^{-1}\mathbf{A}\chi(\ell) \tag{48}$$

with

$$\chi(\ell) = \left(\mathbf{F}^H\mathbf{F} + \mathbf{I}\right)^{-1}\left(\mathbf{A}^H\xi^{(1)}(\ell) + \mathbf{F}^H\xi^{(2)}(\ell) + \xi^{(3)}(\ell)\right) \tag{49}$$

where we instead of solving one full $p \times p$ system of equations solve two tridiagonal systems of equations, which may be solved using $\mathcal{O}(p)$ operations [52, p. 153] and one $N \times N$ system of equations. Furthermore, since

$$\left(\mathbf{I} + \mathbf{A}\left(\mathbf{F}^H\mathbf{F} + \mathbf{I}\right)^{-1}\mathbf{A}^H\right)^{-1}\mathbf{A}\chi_k \tag{50}$$

needs to be calculated at each step, the computational complexity can be decreased even further by calculating the Cholesky factor, and at each stop solving two triangular systems of equations. Thus, for a one time cost of $\mathcal{O}(N^3)$ operations, one can at each step solve two triangular systems of equations at cost of $\mathcal{O}(N^2)$ operations. Step 4 in Algorithm 1 thereby decomposes into three different and decoupled optimization problems; firstly, for the first block,

$$\begin{aligned}\mathbf{u}^{(1)}(\ell+1) &= \underset{\mathbf{u}}{\operatorname{argmin}}\,\frac{1}{2}\left\|\mathbf{u} - \mathbf{y}\right\|_2^2 + \frac{\mu}{2}\left\|\mathbf{A}\mathbf{z}(\ell+1) - \mathbf{u} - \mathbf{d}^{(1)}(\ell)\right\|_2^2 \\ &= \frac{\mathbf{y} - \mu\left(\mathbf{A}\mathbf{z}(\ell+1) - \mathbf{d}^{(1)}(\ell)\right)}{1+\mu}\end{aligned} \tag{51}$$

Secondly, the optimization problem for the second block is equivalent to (31),

leading again to

$$\mathbf{u}^{(2)}(\ell+1) = \underset{\mathbf{u}}{\operatorname{argmin}}\, \lambda||\mathbf{u}||_1 + \alpha \sum_{k=1}^{P} \sqrt{\Delta_k}||\mathbf{u}_k||_2 \tag{52}$$

$$+ \frac{\mu}{2}||\mathbf{z}(\ell+1) - \mathbf{u} - \mathbf{d}^{(2)}(\ell)||_2^2 \tag{53}$$

$$= \bar{\Psi}\left(\Psi\left(\mathbf{z}(\ell+1) - \mathbf{d}^{(2)}(\ell), \frac{\lambda}{\mu}\right), \frac{\alpha\sqrt{\Delta_k}}{\mu}\right) \tag{54}$$

Finally, the third block can be similarly updated to

$$\mathbf{u}^{(3)}(\ell+1) = \underset{\mathbf{u}}{\operatorname{argmin}}\, \gamma||\mathbf{u}||_1 + \frac{\mu}{2}||\mathbf{F}\mathbf{z}_{k+1} - \mathbf{u} - \mathbf{d}_k^{(3)}||_2^2 \tag{55}$$

$$= \Psi\left(\mathbf{F}\mathbf{z}(\ell+1) - \mathbf{d}^{(3)}(\ell), \frac{\gamma}{\mu}\right) \tag{56}$$

The resulting ADMM algorithm for the block sparse pitch estimation problem, including the TV penalty (PEBS$_2$TV), is summarized in Algorithm 3. Alternatively, if one wish to use an ℓ_1 norm for the model fit, as discussed above, one may simply change the appropriate step, i.e., the update for $\mathbf{u}_{k+1}^{(1)}$ in Algorithm 3 leads to

$$\begin{aligned}\mathbf{u}^{(1)}(\ell+1) &= \underset{\mathbf{u}}{\operatorname{argmin}}\, \frac{1}{2}\left\|\mathbf{u} - \mathbf{y}\right\|_1 + \frac{\mu}{2}\left\|\mathbf{A}\mathbf{z}(\ell+1) - \mathbf{u} - \mathbf{d}^{(1)}(\ell)\right\|_2^2 \\ &= \mathbf{y} + \Psi\left(\mathbf{A}\mathbf{z}^{(1)}(\ell+1) - \mathbf{d}^{(1)}(\ell), \frac{1}{\mu}\right)\end{aligned}$$

We denote the thus resulting estimators the PEBS$_1$ and PEBS$_1$TV, where the latter includes the TV penalty.

The computational cost of each iteration of Algorithm 2 and 3 is, for typical problem dimensions, dominated by calculating $\mathbf{A}\mathbf{x}$ and $\mathbf{A}^H\mathbf{y}$, for various vectors $\mathbf{x}$ and $\mathbf{y}$, and requires considerably less operations than the $\mathcal{O}(p^3)$ needed for the solvers mentioned earlier. It is worth noting that the cost of the PEBS algorithms may be significantly reduced for signals sampled at equidistant time-points by using fast Fourier transform (FFT) techniques. Further improvements are possible by addressing the choice of the dual variable step size, μ. Instead of tuning it for each problem depending on the typical sizes of the various inputs and outputs, an adaptive approach is possible using the following heuristic [38]: considering

Algorithm 3 PEBS$_2$TV via ADMM

1: Initiate $\mathbf{z} = \mathbf{z}(0)$, $\mathbf{u} = \mathbf{u}(0)$, and $\ell := 0$
2: **repeat**
3: $\quad \mathbf{z}(\ell) = \left[\mathbf{A}^H\mathbf{A} + \mathbf{F}^H\mathbf{F} + \mathbf{I}\right]^{-1}\left(\mathbf{A}^H\boldsymbol{\xi}^{(1)}(\ell) + \mathbf{F}^H\boldsymbol{\xi}^{(2)}(\ell) + \boldsymbol{\xi}^{(3)}(\ell)\right)$
4: $\quad \mathbf{u}^{(1)}(\ell+1) = \frac{\mathbf{y}-\mu\left(\mathbf{A}\mathbf{z}(\ell+1)-\mathbf{d}^{(1)}(\ell)\right)}{1+\mu}$
5: $\quad \mathbf{u}^{(2)}(\ell+1) = \bar{\Psi}\left(\Psi\left(\mathbf{z}(\ell+1) - \mathbf{d}^{(2)}(\ell), \frac{\lambda}{\mu}\right), \frac{\alpha\sqrt{\Delta_k}}{\mu}\right)$
6: $\quad \mathbf{u}^{(3)}(\ell+1) = \Psi\left(\mathbf{F}\mathbf{z}(\ell+1) - \mathbf{d}^{(3)}(\ell), \frac{\gamma}{\mu}\right)$
7: $\quad \mathbf{d}(\ell+1) = \mathbf{d}(\ell) - (\mathbf{G}\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1))$
8: $\quad \ell \leftarrow \ell + 1$
9: **until** convergence

the fact that μ can be seen as controlling the relative importance of the dual and primal feasibility condition suggests an adaptive choice by comparing the norms of the primal and dual residuals and adjusting μ appropriately, i.e., after step 9 in Algorithm 3, one may update μ according to

$$\mu(\ell+1) = \begin{cases} \mu(\ell)\tau & \text{if } ||\mathbf{r}(\ell)||_2 > \rho||\mathbf{s}(\ell)||_2 \\ \mu(\ell)/\tau & \text{if } ||\mathbf{s}(\ell)||_2 > \rho||\mathbf{r}(\ell)||_2 \\ \mu(\ell) & \text{otherwise} \end{cases} \tag{57}$$

where τ is the multiplicative change in the step size, and μ set such that the step size is changed to keep the ratio between the norms of the primal and dual residuals within a factor μ. In our experience, setting $\tau = 2$ and $\rho = 10$ results in about an order of magnitude fewer steps being needed. Note that changing μ here does not cause any additional computational cost in any of the above steps, except for the negligible cost of rescaling the dual variables, i.e., $\tilde{d}(\ell+1) = \mu_\ell/\mu_{\ell+1}d(\ell+1)$.

5 Numerical results

We proceed to examine the robustness and performance of the proposed estimators, using both simulated and real audio signals, comparing with the optimal filtering (Capon), approximative nonlinear least squares (ANLS), and multi-pitch

estimator based on subspace orthogonality (ORTH) algorithms [9, 53]. These estimators have in several studies been found to offer state-of-the-art performance, and have freely available implementations, allowing for easily reproducible comparisons in future studies. Initially, examining simulated signals, the performance of the estimates for the different algorithms are computed using 250 Monte-Carlo simulations and $N = 160$ samples, wherein the number of harmonics are selected uniformly over $\left[3, \min(\text{floor}(1/f), 10)\right]$ in each simulation, where f denotes the fundamental frequency, in order to ensure that all frequencies are below the Nyquist limit. Here, frequencies are given as normalized frequencies with unit cycles/sample, in the interval [0, 1], unless otherwise specified. The signal to noise ratio (SNR), defined as $10log_{10}(||\mathbf{y}||_2/||\mathbf{w}||_2)$, is set to 18 dB, unless otherwise stated. To ensure the best possible performance, the reference methods are allowed perfect *a priori* knowledge of both the number of present sources and their respective number of harmonics, whereas the proposed estimators are only given that the maximum number of harmonics for any present source is 10. All methods are given the same grid size, equivalent to 1000 equally spaced points in [0.025, 0.1]. We begin by examining the performance of the estimators in a case with one source when random harmonics are allowed to be missing. As shown in earlier studies (see, e.g., [9]), the reference methods are well able to estimate the pitch of a single source, but can be expected to suffer somewhat of a loss of performance when the number of assumed harmonics differ from the actual number present in the signal. To illustrate this, we simulate a signal with the fundamental frequency drawn uniformly on $[0.025, 0.05]$, with $L_1 = 10$ with $2 - 8$ harmonics missing at random, with all the amplitudes set to 1 with uniformly distributed phases. The results are shown in Figure 1, illustrating the ratio of estimates for which the estimated pitch is within ± 0.0002, i.e., approximately within two grid points from the true value, for a varying number of missing harmonics. As seen in the figure, it is clear that the PEBS estimators are performing as well as, or even better, than the reference methods. Of the methods, only ORTH is seen to suffer noticeably by the missing harmonics, which is natural due to the resulting loss of orthogonality between the subspaces. It is worth noting that the fundamental frequency is here allowed to be one of the randomly missing harmonics. We have here used $\alpha = c\chi$, $\lambda = (1 - c)\chi$, for $c = 0.5$ and $\chi = 0.2$. Next, we illustrate how the TV penalty influences the performance of the estimate. Figure 2 shows the results for a single pitch signal with fundamental frequency chosen uniformly in [0.04, 0.0625], with four harmonics, where, as before, all the amplitudes are

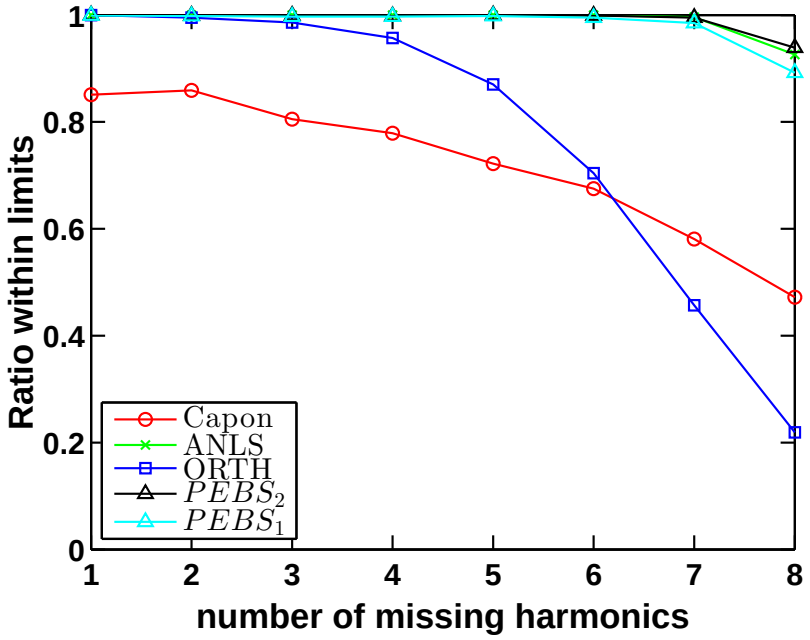

Figure 1: Ratio of estimated pitches where the fundamental frequency lies at most 0.0002 from the ground truth, plotted as a function of the number of harmonics that are missing for $\alpha = \lambda = 0.5\chi$ and $\chi = 0.2$. The fundamental frequency is uniformly distributed on $[0.025, 0.05]$.

set to 1 with random phases, and the dictionary for both methods is chosen such that a maximum of 8 harmonics are allowed for the frequency range [0.02, 0.1].

The result of this choice of signal and dictionary is that the cost function for PEBS_2 will not be able to distinguish between the block corresponding to f_0 and $f_0/2$ in a consistent manner. This is clearly visible in the figure, where one can see that the fundamental frequency is only correctly identified in roughly 60 % of the simulation for the PEBS_2 estimator, with noise in the spectrum basically deciding if f_0 or $f_0/2$ is chosen, whereas the PEBS_2TV estimate yields consistent performance for all SNRs. Here, and in all other simulations, γ was set to 0.01. We proceed with the more interesting case of more than one signal source, forming a signal consisting of two sources with the fundamental frequencies, f_k,

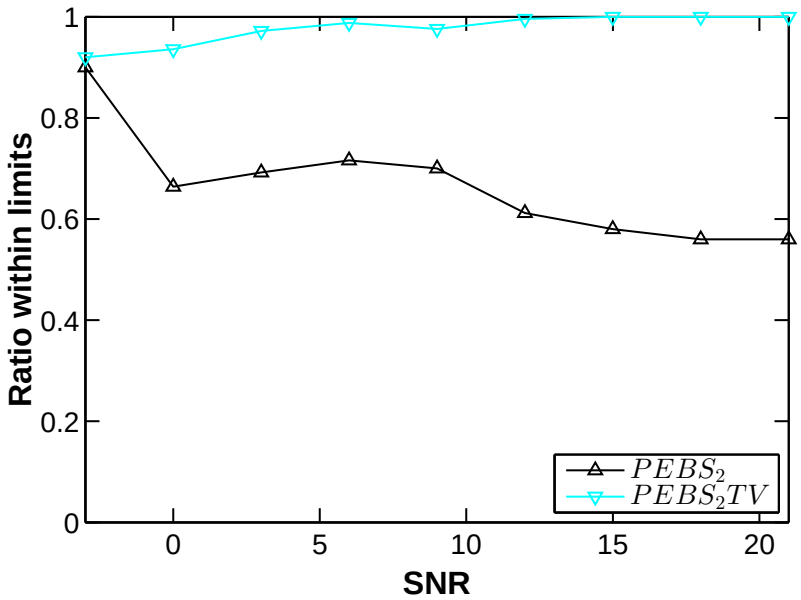

Figure 2: Ratio of estimated pitches where both fundamental frequencies lie at most two grid points from the ground truth, plotted as a function of SNR. The dictionary and signal are chosen such that there is ambiguity in the choice of f_0 vs $f_0/2$.

drawn uniformly on $[0.025, 0.1]$, where we have ensured that the minimum difference between the frequencies is at least 1/25 of the frequency range. To illustrate the effect of non-equal amplitudes, the amplitudes are here drawn such that both pitches have equal power, with $a_{i,k} \sim \mathbf{N}(1, 1)$, i.e., Gaussian with expected value one and variance one, with uniformly distributed phase, which also means that no harmonics will be missing, but some might have small amplitudes. Figure 3 shows the ratio of estimates where the estimated pitches are both within two grid points from the true value, for varying SNR, clearly showing the preferable performance of the proposed PEBS algorithms. As seen from the figure, the PEBS_2 estimates achieve almost perfect performance for SNRs greater than 5 dB, whereas the other examined estimators fail to do so, even for larger SNRs. The reference methods thus fail to properly identify the pitches for the two sources,

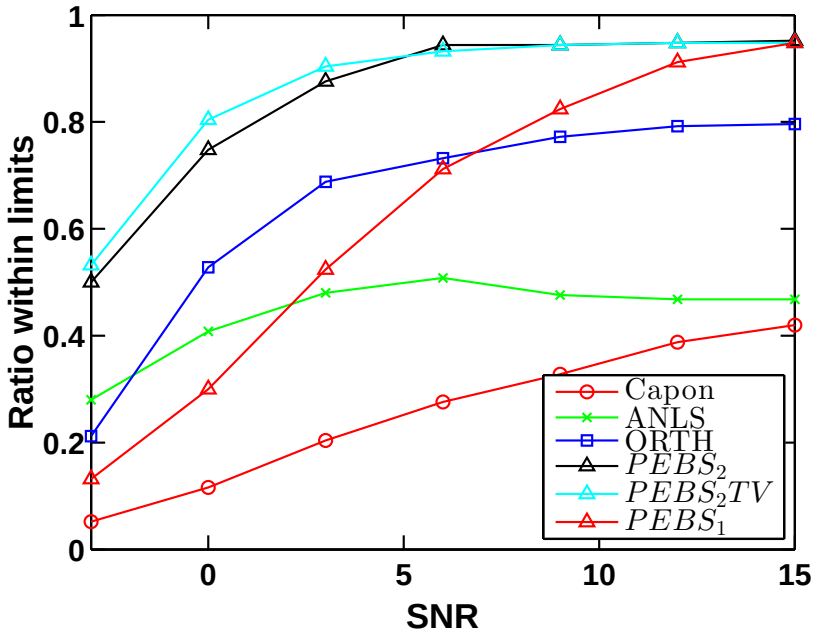

Figure 3: Ratio of estimated pitches where both fundamental frequencies lie at most two gridpoints from the ground truth, plotted as a function of SNR for $\alpha = \lambda = 0.5\chi$ and $\chi = 2.1\sigma_e$. The fundamental frequency is uniformly distributed on $[0.025, 0.1]$.

even though being provided perfect *a priori* information of the number of sources and harmonics. This can to some extent be explained by the fact that, being random variables, some of the amplitudes may well be quite small, mimicking the missing harmonics case previously studied. Also, as the fundamental frequency decreases, the harmonics become more closely spaced, implying a more difficult estimation problem. To examine the effects of closely spaced fundamental frequencies, we proceed to consider the pitches $f_1 = 0.02 + \xi$, where the random variable ξ, uniformly distributed on $[0,\ 0.00005]$ and redrawn for each Monte-Carlo simulation, is added to make sure that the signal is not lying exactly on the grid of proposed fundamental frequencies, and with $f_2 = f_1 + \Delta f$. Here, to clarify the effects of the source separation, $L_1 = 4$ and $L_2 = 4$, $\alpha_{k,l} = 1$,

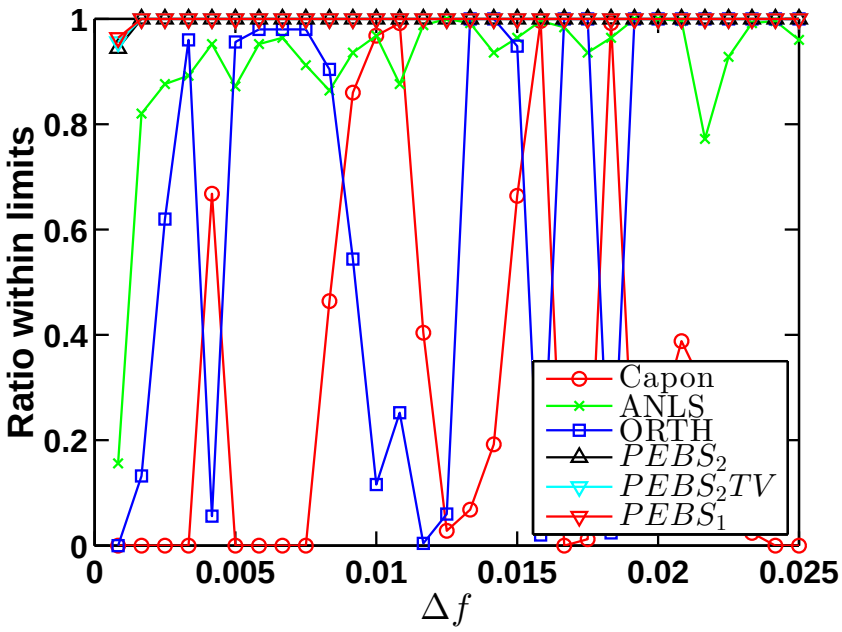

Figure 4: Ratio of estimated pitches where both fundamental frequencies lie at most two grid points from the ground truth, plotted as a function of Δf, for $f_0 = 0.025$, $\alpha = \lambda = 0.5\chi$, $L_1 = 7, L_2 = 5$ and $\chi = 0.2$.

$\forall k, l$, with the amplitudes having a uniformly distributed phase. Figure 4 shows the resulting performance as a function of Δf, again confirming the preferable performance of the proposed estimators. In particular, it is worth noting how the Capon and ORTH estimators suffers loss in performance as frequencies corresponding to the overtones of the fundamental frequencies. Here, the performance of the reference methods can be largely explained by the difficulty of estimating lower fundamental frequencies. To illustrate this, Figure 5 shows the ratio when selecting larger fundamental frequencies, $f_0 = 0.05$ instead of 0.025 in the previous example. As can be seen in the figure, the more well separated pitches are easier for the reference methods to resolve. As is clear from both figures, the proposed estimator does not suffer this shortcoming, and offer a uniformly preferable performance. We continue on to examine the robustness to the selection of the user parameters. Figure 6 illustrates the resulting performance as a function of χ

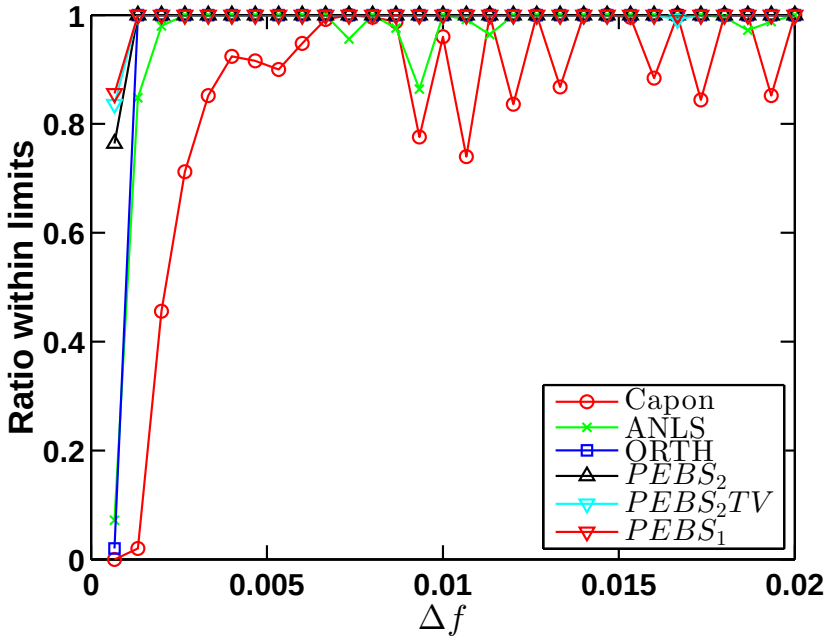

Figure 5: Ratio of estimated pitches where both fundamental frequencies lie at most two grid points from the ground truth, plotted as a function of Δf, for $f_0 = 0.05$, $\alpha = \lambda = 0.5\chi$, $L_1 = 7, L_2 = 5$ and $\chi = 0.2$.

for different values of c, for SNR=15 dB, while the other signal parameters are the same as for the signals used for Figure 3.To increase clarity, the results are here only compared to the ORTH estimator, which exhibited the best performance of the reference methods. As shown in the figure, the performance of the PEBS estimate is quite insensitive to the choice of the user parameters, although their relative ratio, typically estimated using a modified cross validation approach, where the prediction of the estimated model is done with a re-estimated LS solution using only the non-zero blocks chosen (see, e.g., [54]), does make some difference in performance. The figure illustrates that a better results was obtained by including the ℓ_1 penalty ($c \neq 0$), as compared to using only the block penalty ($c = 0$). Turning our attention to actual audio recordings, we consider a real audio signal[1]

[1]The authors are grateful to Mr Tommy Nilsson for this recording.

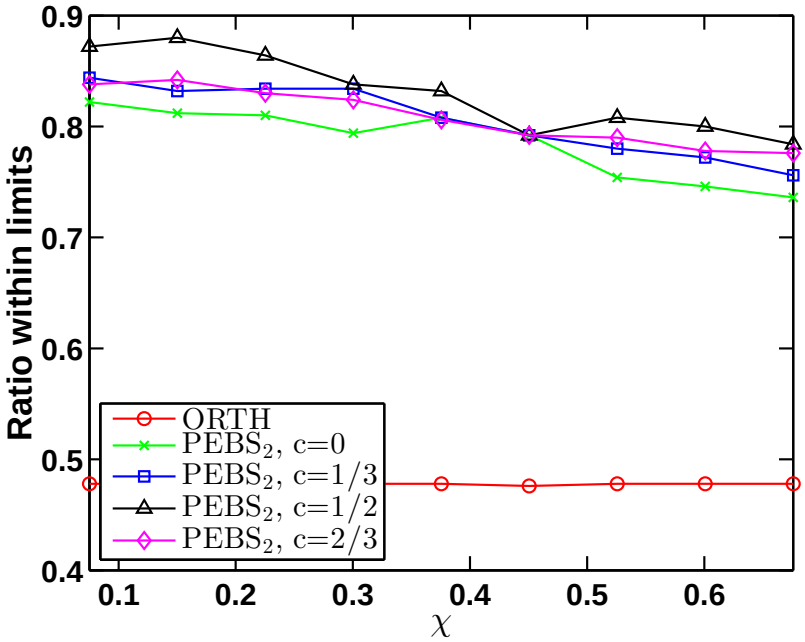

Figure 6: Ratio of estimated pitches where both fundamental frequencies lie at most two grid points from the ground truth, plotted as a function of χ for $\alpha = c\chi$, $\lambda = (1-c)\chi$, for $c \in \{0,\ 1/2,\ 1/3,\ 2/3\}$.

using a recorded guitar playing in succession three chords, first a single note, then a 2-note chord, and, finally, a 3-note chord. Figures 7-9 show the spectrogram of the recorded signal as well as the resulting PEBS_2TV and ORTH estimates, respectively. For this signal, where one may expect a fundamental frequency in the range 80 to 1600 Hz, and with varying number of pitches and harmonics, the f_0 vs $f_0/2$ ambiguity should be expected. As can be seen in the figures, the PEBS_2TV method estimates the fundamental frequencies consistently with the actual number of sources, as well as the fundamental frequencies of the underlying notes. Figure 8 also shows the (estimated) scaled standard deviation of the signal, clearly illustrating the initial uncertainty in the measurement when the chord is struck. The dictionary is chosen using the entire span of the fundamental frequency range of a guitar, and the number of harmonics is chosen to be

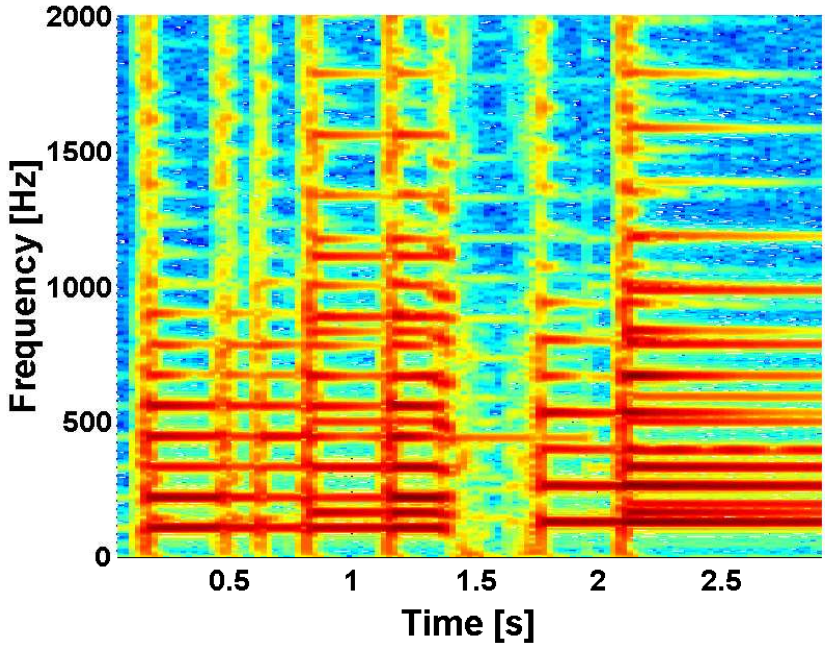

Figure 7: Spectrogram of recorded guitar sound.

at a maximum 8, c was set to 0.3 and χ was set to equal the standard deviation of the signal. Overall, PEBS$_2$TV manages to find the correct number of pitches and the true fundamental frequency. Since the estimator is not given the number of pitches, artificial fundamental frequency estimates appear when string is struck or damped. This shows the importance of better preprocessing or modeling for music signal applications. Furthermore, the frequency estimate at around 990 Hz might be due to the inharmonicity in the guitar (see, e.g., [55]). For comparison, we in Figure 9 show, the resulting estimates for the ORTH estimator, which was best performing of the reference methods for this signal. The model order was here set using oracle information of the number of pitches and manually tuning the number of pitches to give the best results. As can be seen, the ORTH estimator manages to do reasonably well, with the most troublesome region being between 1 and 1.5 seconds, where several cases of $f_0/2$ or $2f_0$ being chosen instead of the correct fundamental frequency.Finally, we examine a signal obtained by su-

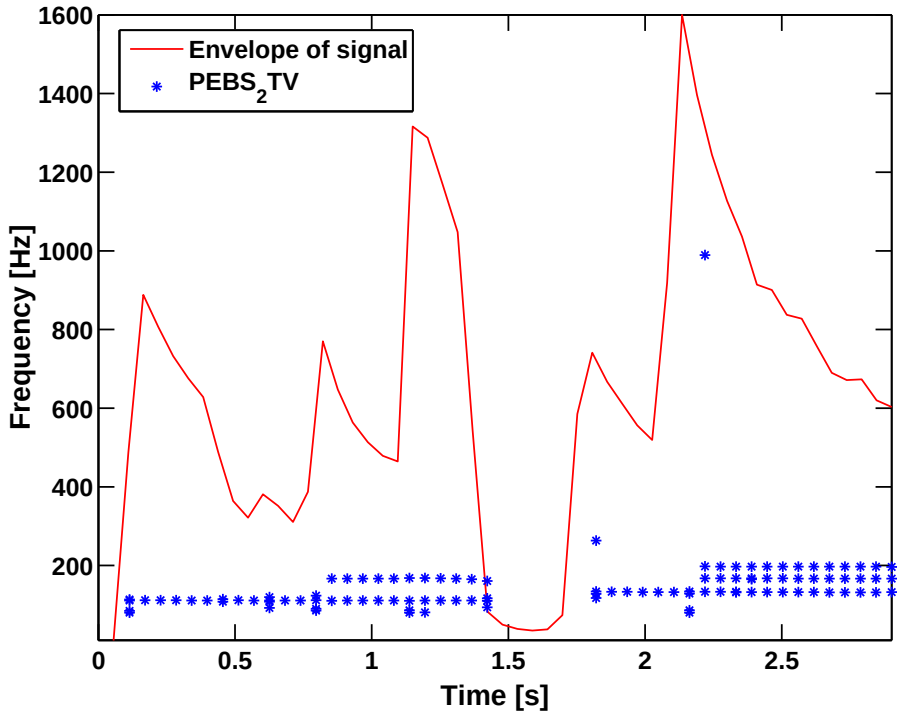

Figure 8: The PEBS estimate of the guitar recording, showing that the correct number of pitches and their corresponding frequencies are revealed. The scaled standard deviation of the signal is superimposed to illustrate at what time points the notes are struck or muted.

perimposing two recordings from the SQAM database [56], being a viola and the voice of a female speaker. The viola has a single fundamental frequency of about 131 Hz with roughly 15 overtones, although it may be noted that both the first and fifth harmonics are missing, and several other harmonics are quite small. For the speech signal, we have selected a part of the phrase "to administer", analyzing the two vowels "o" and "a", corresponding to the first third of the spectrogram in Figure 10. To allow the speech signal to be reasonably stationary, we use (non-overlapping and un-windowed) 20 ms time windows. During the examined time period, the voice varies considerable, and the number of harmonics can be seen to vary over the segments from one to eight with a fundamental frequency vary-

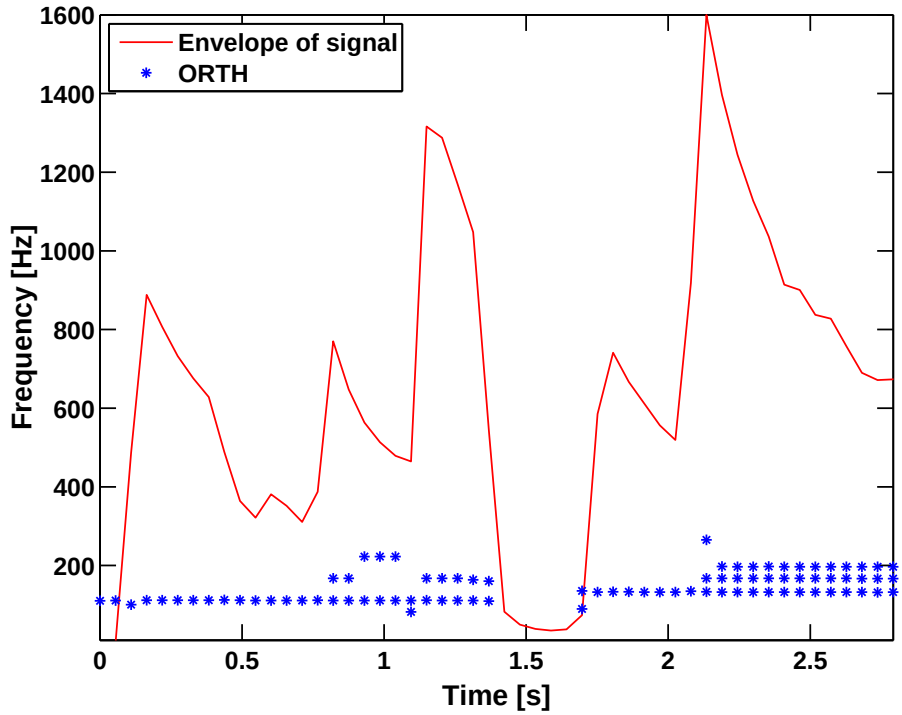

Figure 9: The ORTH estimate of the guitar recording, using oracle information of the model-orders. The scaled standard deviation of the signal is superimposed to illustrate at what time points the notes are struck or muted.

ing between 180 and 220 Hz. The spectrogram of the resulting signal is shown in Figure 10. To allow for the range of possible pitch frequencies a viola and a female voice may be expected to span, the dictionary was selected to cover the frequency range 130–1200 Hz, using 500 grid points, with the maximum number of harmonics set to $L_{\max} = 15$. Figures 11 and 12 show the resulting pitch estimates for PEBS$_2$TV and the ORTH estimator, respectively. Here, ORTH has been allowed oracle knowledge of the number of harmonics of each source, as well as the number of sources. As can be seen from the figures, the PEBS$_2$TV estimator is able to correctly identify the two pitch signals throughout, except in the transition period when the speech signal is too weak to be detected, whereas the ORTH estimate gives poor pitch estimates for the latter part of the signal, where

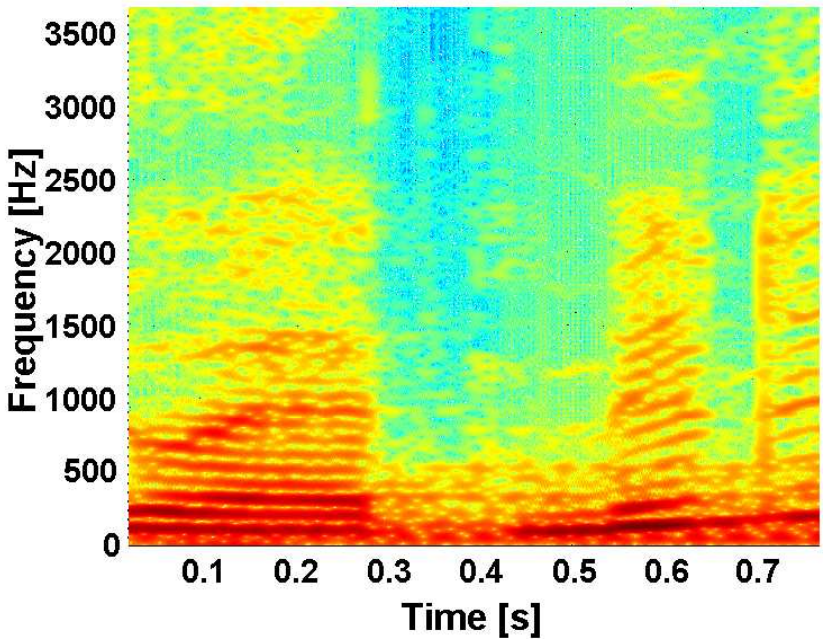

Figure 10: Spectrogram of recorded speech and viola.

it yields pitch estimates which are multiples of the correct pitch, corresponding to the higher order overtones. As the PEBS_2TV estimator does not assume prior knowledge of the number of sources, it may yield spurious pitches. This may be seen, for instance, at time 0.15 s, where a (weak) third pitch appears. By tuning the estimator better, or by allowing for information from previous frames, for instance via pitch tracking (see, e.g., [57]), this may easily be remedied.

6 Conclusions

In this work, we introduced the idea of using block sparsity in the estimation of the fundamental frequencies of a multi-pitch signal. Formulating the estimation as a sum of a fitting term and convex sparsity inducing norms, ensuring a block sparse solution, the proposed algorithm is shown to offer significantly improved performance as compared to a range of state-of-the-art multi-pitch estimators.

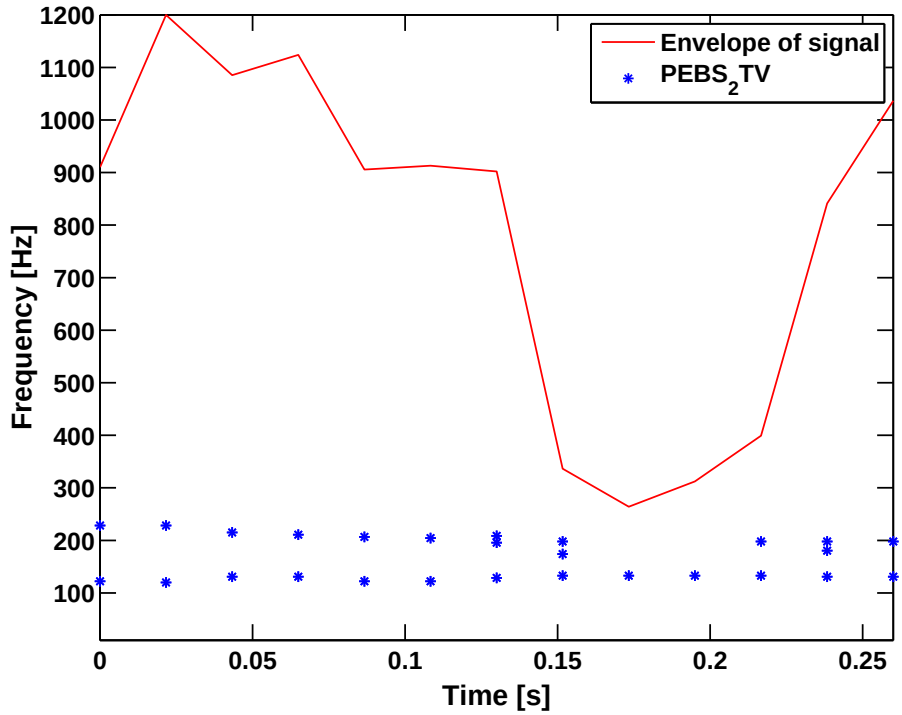

Figure 11: The PEBS_2TV estimate of the speech and viola recording. The scaled standard deviation of the signal is superimposed to illustrate at what time points the voice is silent.

Furthermore, by including a total variation penalty on each block, the algorithm avoids the f_0 vs $f_0/2$ ambiguity that many estimators suffer from. The algorithm is shown to be capable of handling issues such as missing harmonics as well as closely spaced fundamental frequencies. Furthermore, novel ADMM algorithms are devised for the entailing optimizations, resulting in a iterative dual ascent method, where each step has a simple closed form expression that scales well with the problem dimensions.

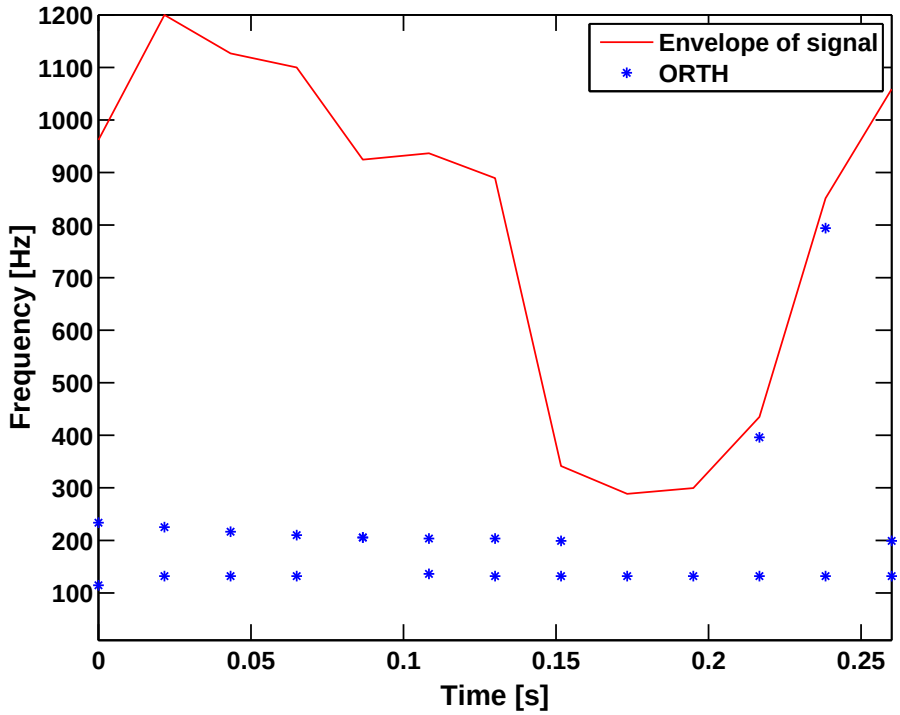

Figure 12: The ORTH estimate of the speech and viola recording, using oracle information of the model-orders. The scaled standard deviation of the signal is superimposed to illustrate at what time points the voice is silent.

7 Appendix

Appendix A

Insight into how the penalty term in (9) induces a block sparse solution can be gained by studying the sub-differential equations of the equivalent real-valued cost function (see also [42]), which may be expressed as

$$\tilde{\mathbf{W}}_\ell^T \left(\tilde{\mathbf{y}} - \sum_{k=1}^{P} \tilde{\mathbf{W}}_k \tilde{\mathbf{a}}_k \right) + \alpha \mathbf{s}_\ell = 0 \tag{58}$$

for $\ell = 1, 2, \ldots, P$, where $\mathbf{s}_\ell$ is either a vector such that $||\mathbf{s}_\ell||_2 \leq 1$, or equal to $\tilde{\mathbf{a}}_\ell / ||\tilde{\mathbf{a}}_\ell||$, depending on if $\tilde{\mathbf{a}}_\ell = 0$ or not, $\tilde{\mathbf{W}}$ is the real counterpart of $\mathbf{W}$, created such that

$$\tilde{\mathbf{W}}_\ell = \begin{bmatrix} \mathfrak{Re}\{\mathbf{W}_\ell\} & -\mathfrak{Im}\{\mathbf{W}_\ell\} \\ \mathfrak{Im}\{\mathbf{W}_\ell\} & \mathfrak{Re}\{\mathbf{W}_\ell\} \end{bmatrix}$$

where $\mathfrak{Re}\{\cdot\}$ and $\mathfrak{Im}\{\cdot\}$ denote the real and imaginary part of a matrix, and $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{a}}$ are formed similarly, i.e.,

$$\begin{aligned} \tilde{\mathbf{y}} &= \begin{bmatrix} \mathfrak{Re}\{\mathbf{y}\} & \mathfrak{Im}\{\mathbf{y}\} \end{bmatrix} \\ \tilde{\mathbf{a}}_\ell &= \begin{bmatrix} \mathfrak{Re}\{\mathbf{a}_\ell\} & \mathfrak{Im}\{\mathbf{a}_\ell\} \end{bmatrix} \end{aligned}$$

Thus, for any minimizing vector $\breve{\mathbf{a}}$, a necessary and sufficient condition for a sub-vector, or block, $\breve{\mathbf{a}}_\ell$, to be zero is that [42]

$$\left\| \tilde{\mathbf{W}}_\ell^T \left(\tilde{\mathbf{y}} - \sum_{k=1}^{P} \tilde{\mathbf{W}}_k \breve{\tilde{\mathbf{a}}}_k \right) \right\|_2 < \alpha \tag{59}$$

which shows the (block) sparsifying effect of the (block) 2-norm. Note further that if the inequality does not hold, $\tilde{\mathbf{a}}_\ell$ could have been found by solving

$$\tilde{\mathbf{a}}_\ell = \left(\tilde{\mathbf{W}}_\ell^T \tilde{\mathbf{W}}_\ell + \alpha / ||\tilde{\mathbf{a}}_\ell|| \right)^{-1} \tilde{\mathbf{W}}_\ell^T \left(y - \sum_{k \neq \ell} \tilde{\mathbf{W}}_k \tilde{\mathbf{a}}_k \right) \tag{60}$$

This can be recognized as being similar to the solution of a Tikhonov regularized LS, or ridge regression, solution which is known to lack a sparsifying effect. Thus, if the block is non-zero, one may expect each element in the block to be non-zero.

Appendix B

Similarily as in Appendix A, the sparsity of the solution of (11) may be understood by studying the subdifferential equations for the equivalent real-valued problem, which are given by

$$\tilde{\mathbf{W}}_\ell^T \left(\tilde{\mathbf{y}} - \sum_{k=1}^{P} \tilde{\mathbf{W}}_k \tilde{\mathbf{a}}_k \right) + \alpha \mathbf{s}_\ell + \lambda \mathbf{r}_\ell = 0 \tag{61}$$

for $\ell = 1, \ldots, P$, where $\mathbf{s}_\ell$ and $\mathbf{r}_\ell$ are real-valued vectors defined such that

$$\mathbf{s}_\ell = \begin{cases} \frac{\tilde{\mathbf{a}}_\ell}{||\tilde{\mathbf{a}}_\ell||_2} & \text{if } \tilde{\mathbf{a}}_\ell \neq 0 \\ \mathbf{v} & \text{otherwise} \end{cases} \tag{62}$$

where $||\mathbf{v}||_2 \leq 1$, and

$$\begin{bmatrix} r_{\ell,i} \\ r_{\ell,i+L_k} \end{bmatrix} = \begin{cases} \frac{\left[\tilde{\mathbf{a}}_{k,i}, \tilde{\mathbf{a}}_{k,i+L_k}\right]^T}{\left\|\left[\tilde{\mathbf{a}}_{k,i}, \tilde{\mathbf{a}}_{k,i+L_k}\right]\right\|_2} & \text{if } \left[\tilde{\mathbf{a}}_{k,i},\ \tilde{\mathbf{a}}_{k,i+L_k}\right]^T \neq 0 \\ \mathbf{p} & \text{otherwise} \end{cases} \tag{63}$$

with $||\mathbf{p}||_2 \leq 1$, for $i = 1, \ldots, L_k$, where $\mathbf{a}_{i,j}$ denotes element j of sub-vector i, $[a,\ b]$ a vector with two scalars a and b, and

$$\mathbf{r}_\ell = \begin{bmatrix} r_{\ell,1} & \ldots & r_{\ell,2L_k} \end{bmatrix}^T \tag{64}$$

This implies that for any minimizing vector $\breve{\mathbf{a}}$, it holds that $\breve{\mathbf{a}}_\ell = 0$ if

$$\left\| \tilde{\mathbf{W}}_\ell^T \left(\tilde{\mathbf{y}} - \sum_{k=1}^{P} \tilde{\mathbf{W}}_k \breve{\tilde{\mathbf{a}}}_k \right) - \lambda \mathbf{r} \right\|_2 \leq \alpha \tag{65}$$

or, equivalently, if

$$\sum_{k=1}^{L_\ell} \left\| \mathbf{z}_k (\|\mathbf{z}_k\|_2 - \lambda)^+ \right\|_2^2 \leq \alpha^2 \tag{66}$$

where $\mathbf{z}_k$ is a vector composed of the elements k and $k + L_\ell$ of the vector

$$\mathbf{z} = \tilde{\mathbf{W}}_\ell^T \left(\tilde{\mathbf{y}} - \sum_{k=1}^{P} \tilde{\mathbf{W}}_k \breve{\tilde{\mathbf{a}}}_k \right) \tag{67}$$

Interestingly, but perhaps not surprisingly, this is a similar solution as one would obtain from the analysis of the real-valued version of (10) analyzed in [42]. However, in this case, the analysis holds for any kind of non-overlapping sub-division of the sub-vectors, not only into the two variables corresponding to the same complex variables. This insight was used in [58] to generalize the above results to the case of multiple measurements vectors (array) case.

References

[1] P. Kroon and W. B. Kleijn, "Linear-prediction based analysis-by-synthesis coding," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds., chapter 3, pp. 79–119. Elsevier, Berlin, Germany, 1995.

[2] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 10, no. 5, pp. 293–302, July 2002.

[3] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, 2009.

[4] M. Müller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 6, pp. 1088–1110, 2011.

[5] W. Hess, *Pitch Determination of Speech Signals*, Springer, Berlin, 1983.

[6] H. Li, P. Stoica, and J. Li, "Computationally Efficient Parameter Estimation for Harmonic Sinusoidal Signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.

[7] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1917–1930, April 2002.

[8] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 609–612, 2004.

[9] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88, no. 4, pp. 972–983, April 2008.

[10] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "A Robust and Computationally Efficient Subspace-Based Fundamental Frequency Estimator," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 487–497, March 2010.

[11] Z. Zhou, H. C. So, and F. K. W. Chan, "Optimally Weighted Music Algorithm for Frequency Estimation of Real Harmonic Sinusoids," in *Proc. 37th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Kyoto, March 25-30 2012.

[12] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 8, no. 6, pp. 708–716, 2000.

[13] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 101–111, jan. 2003.

[14] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 11, no. 6, pp. 804–816, 2003.

[15] S. S. Abeysekera, "Multiple pitch estimation of poly-phonic audio signals in a frequency-lag domain using the bispectrum," in *Proc. IEEE International Symposium on Circuits and Systems*, 2004, vol. 14, pp. 469–472.

[16] M. D. Plumbley, S. A. Abdallah, T. Blumensath, and M. E. Davies, "Sparse representations of polyphonic music," *Signal Processing*, vol. 86, no. 3, pp. 417–431, March 2006.

[17] J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigne, and S. Sagayama, "Single and Multiple Contour Estimation Through Parametric Spectrogram Modeling of Speech in Noisy Environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1135–1145, May 2007.

[18] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1643–1654, Aug. 2010.

[19] E. Benetos and S. Dixon, "Joint Multi-Pitch Detection Using Harmonic Envelope Estimation for Polyphonic Music Transcription," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 6, pp. 1111–1123, Oct. 2011.

[20] A. Koretz and J. Tabrikian, "Maximum A Posteriori Probability Multiple-Pitch Tracking Using the Harmonic Model," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2210–2221, 2011.

[21] C. Lee, Y. Yang, and H. H. Chen, "Multipitch Estimation of Piano Music by Exemplar-Based Sparse Representation," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 608–618, 2012.

[22] M. Genussov and I. Cohen, "Multiple fundamental frequency estimation based on sparse representations in a structured dictionary," *Digit. Signal Process.*, vol. 23, no. 1, pp. 390–400, Jan. 2013.

[23] F. Huang and T. Lee, "Pitch Estimation in Noisy Speech Using Accumulated Peak Spectrum and Sparse Estimation Technique," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 1, pp. 99–109, Jan. 2013.

[24] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.

[25] D.L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289 –1306, 2006.

[26] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[27] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul, jul 2008, pp. 10225–10229.

[28] I. F. Gorodnitsky and B. D. Rao, "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, March 1997.

[29] P. Stoica and P. Babu, "Sparse Estimation of Spectral Lines: Grid Selection Problems and Their Solutions," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, Feb. 2012.

[30] T. Nilsson, S. I. Adalbjörnsson, N. R. Butt, and A. Jakobsson, "Multi-Pitch Estimation of Inharmonic Signals," in *European Signal Processing Conference*, Marrakech, Sept. 9-13, 2013.

[31] N. R. Butt, S. I. Adalbjörnsson, S. D. Somasundaram, and A. Jakobsson, "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities," in *Proc. 38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

[32] C. D. Austin, J. N. Ash, and R. L. Moses, "Dynamic Dictionary Algorithms for Model Order and Parameter Estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5117–5130, October 2013.

[33] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Signals," in *Proc. 39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Florence, Italy, May 4-9 2014.

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

[35] Y. V. Eldar, P. Kuppinger, and H. Bolcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3042–3054, 2010.

[36] X. Lv, G. Bi, and C. Wan, "The Group Lasso for Stable Recovery of Block-Sparse Signal Representations," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1371–1382, 2011.

[37] A. Juditsky, F. Karzan, A. Nemirovski, and B. Polyak, "Accuracy guaranties for ℓ_1 recovery of block-sparse signals," *Annals of Statistics*, vol. 40, pp. 3077–3107, 2012.

[38] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[39] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Algorithms for imaging inverse problems under sparsity regularization," in *Proc. 3rd Int. Workshop on Cognitive Information Processing*, May 2012, pp. 1–6.

[40] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182 –2195, May 2011.

[41] J. Fang, J. Li, Y. Shen, H. Li, and S. Li, "Super-resolution compressed sensing: An iterative reweighted algorithm for joint parameter learning and

sparse signal recovery," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 761–765, 2014.

[42] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A Sparse-Group Lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[43] P. Babu, *Spectral Analysis of Nonuniformly Sampled Data and Applications*, Ph.D. thesis, Uppsala University, 2012.

[44] C. R. Rojas, D. Katselis, and H. Hjalmarsson, "A Note on the SPICE Method," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4545–4551, Sept. 2013.

[45] P. Stoica, P. Babu, and J. Li, "SPICE : a novel covariance-based sparse estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629 –638, Feb. 2011.

[46] R. Chartrand and B. Wohlberg, "A Nonconvex ADMM Algorithm for Group Sparsity with Sparse Groups," in *38th IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing*, 2013.

[47] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[48] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse Learning via Iterative Minimization With Application to MIMO Radar Imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1088–1101, March 2011.

[49] P. Stoica and Y. Selén, "Model-order Selection — A Review of Information Criterion Rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[50] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, August 1999.

[51] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[52] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 3rd edition, 1996.

[53] M.G. Christensen, P. Stoica, A. Jakobsson, and S.H. Jensen, "The Multi-Pitch Estimation Problem: some New Solutions," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2007*, 15–20 April 2007, vol. 3, pp. III–1221–III–1224.

[54] Nicolai Meinshausen, "Relaxed lasso," *Computational Statistics and Data Analysis*, pp. 374–393, 2007.

[55] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, NY, 1988.

[56] "Sound Quality Assessment Materical Recodings for Subjective Tests," Tech. Rep., European Broadcasting Union, 1988.

[57] M. G. Christensen, "A Method for Low-Delay Pitch Tracking and Smoothing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 345–348.

[58] T. Kronvall, S. I. Adalbjörnsson, and A. Jakobsson, "Joint DOA and Multi-Pitch Estimation Using Block Sparsity," in *Proc. 39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014.

B

Paper B

Sparse localization of harmonic audio sources

Stefan Ingi Adalbjörnsson, Ted Kronvall, Simon Burgess, Kalle Åström, and Andreas Jakobsson

Centre for Mathematical Sciences, Lund University, Lund, Sweden

Abstract

In this paper, we propose a novel method for estimating the locations of near- and/or far-field harmonic audio sources impinging on an arbitrary, but calibrated, sensor array. Using a two-step procedure, we first estimate the fundamental frequencies and complex amplitudes under a sinusoidal model assumption, whereafter the location of each source is found by utilizing both the difference in phase and the relative attenuation of the amplitude estimates. As audio recordings often consist of multi-pitch signals exhibiting some degree of reverberation, where both the number of pitches and the source locations are unknown, we propose to use sparse heuristics to avoid the necessity of detailed a priori assumptions on the spectral and spatial model orders. The method's performance is evaluated using both simulated and measured audio data, with the former showing that the proposed method achieves near-optimal performance, whereas the latter confirms the method's feasibility when used with real recordings.

Key words: Multi-pitch estimation, near- and far-field localization, TDOA, block sparsity, convex optimization, ADMM, non-convex sparsity.

1 Introduction

Sound localization has been a topic of interest in a wide range of applications for centuries, and is well known to be a difficult problem, especially in a reverberating room environment (see, e.g., [1–7], and the references therein). Typically, localization estimates are formed by exploiting time of arrival (TOA), time difference of arrival (TDOA), and gain ratios of arrival as estimated over an array of sensors, often using cross-correlation or canonical correlation analysis (CCA) techniques, allowing the source positions to be determined using tri- och multilateration (see, e.g., [8], [9]). In cases when the sources are located far from the sensor array, so-called far-field sources, the range to the sources may not be determined due to the lack of curvature of the impinging sound pressure wavefront, which in this case is essentially planar, restricting the problem to that of determining the direction of arrival (DOA) to the source relative to the sensor array [10–12]. The problem of near-field source localization, and of far-field DOA estimation, has attracted substantial interest in the literature. Commonly, these problem are treated separately, such that the sources are either treated as being far- or near-field. In this work, we take on a different approach, considering both cases simultaneously, allowing for signals from both kind of sources, without requiring any a priori knowledge of either the number of sources, or if they are near- or far-field sources, or of detailed knowledge of the impinging signals. To allow for this very general problem formulation, we restrict our attention to harmonically related sound sources, such as voiced speech [13] and/or the many forms of harmonic audio sources, such as stringed, wind, and pitched percussion instruments [14]. Such sources may be well modelled as a sum of sinusoidal components, with frequencies which are integer multiples (or closely so in case of inharmonicity) of some fundamental (pitch) frequency [15]. Due to the sinusoidal nature of the signals, the measured signals may be well modelled as scaled and phase shifted versions of the source signals, or, typically, as a sum of such signals if measured in a reverberating room environment. Exploiting this, the joint estimation of the DOA and the pitch frequency has been addressed in [16–18], wherein the authors consider the estimation of the DOA of a single harmonic sound source using a uniform linear array (ULA) of receiver sensor, typically assuming oracle knowledge of the number of harmonic signals in the sound source. Here, we extend on these works, allowing for an unknown number of sources, each having an unknown number of harmonics, impinging in a reverberant room environment from either (or both) near- and far-field sources. This is done by exploiting a sparse recovery framework, for the

number of pitches, and, for each pitch, the number of harmonics, as well as the number of sources. Sparse recovery frameworks have in earlier works been found to allow for high quality estimates; typical examples include [19–22], wherein the sparse signal reconstruction from noisy observations were accomplished with the by now well-known sparse least squares (LS) technique. More recently, the technique has been extended to the case of harmonically related audio signals [23, 24]. Using the techniques introduced there, we propose a two-stage procedure, first creating a dictionary of candidate pitches to model the harmonic components of the sources, without taking the locations of the sources into account, and then, in a second stage, a dictionary of possible locations, including simultaneously near- and far-field locations, to model the observed phase differences, as well as the relative attenuations, of the amplitudes of each sinusoidal component.

The remainder of this paper is organized as follows: in the next section, we present the assumed signal model and discuss the imposed restrictions on the sensor array. Then, in section 3, we present the proposed pitch and localization estimator. Section 4 introduces a computationally efficient implementation based on the alternating direction method of multipliers (ADMM), followed in section 5 with an evaluation of the presented technique using both simulated and measured audio signals. Finally, we conclude on our work in section 6.

2 Signal model

In this work, we restrict our attention to complex-valued[1] harmonically related audio signals, formed from K separate audio sources, $x_k(t)$, for $k = 1, \ldots, K$, each consisting of L_k harmonically related sinusoids, such that (see also [15])

$$x_k(t) = \sum_{\ell=1}^{L_k} a_{k,\ell} e^{j\omega_k \ell t} \tag{1}$$

where $\omega_k = 2\pi f_k / f_s$ are the normalized fundamental frequencies, with sampling frequency f_s, and $a_{k,l}$ the complex amplitude of each harmonic. The resulting

[1] Clearly, the measured audio sources will be real-valued, but to simplify notation and in order to reduce complexity, we will here initially compute the discrete-time analytic signal versions of the measured signals, whereafter all processing is done on these signals (see also [15, 25]).

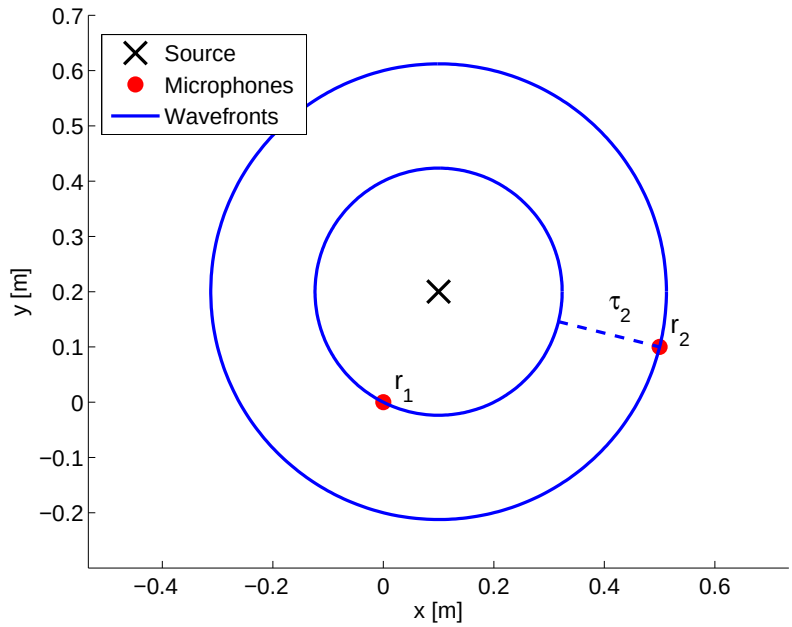

Figure 1: Illustration of a two sensor scenario, with spherical wavefronts propagating from the source. The dashed line shows the scaled TDOA of the second sensor with respect to the first sensor, i.e., τ_2.

multi-pitch signal

$$x(t) = \sum_{k=1}^{K} x_k(t) \tag{2}$$

may be the result of a combination of multi-pitch sources, for example, such as resulting from an instrument playing a musical chord, or from multiple speakers, or from combinations of such signals. When this form of signals impinge on a sensor array, the received k:th pitch at the m:th sensor may be expressed as

$$x_{k,m}(t) \triangleq \frac{1}{d_{k,m}} x_k(t - \tau_{k,m}) \tag{3}$$

for each pitch k over sensors $m = 1, \ldots, M$, with $\tau_{k,m}$ denoting the relative propagation delay, i.e., the TDOA, of the sound signal with respect to the time it

is measured by a selected reference sensor, so that $\tau_{k,1} \triangleq 0$, and where

$$d_{k,m} = \|\mathbf{s}_k - \mathbf{r}_m\|_2 \tag{4}$$

is the distance between the source having the k:th pitch and the m:th sensor, accounting for the attenuation of the signal when propagating in (3). Furthermore, $\mathbf{s}_k$ and $\mathbf{r}_m$ denote the coordinates of the k:th sound source and the m:th sensor, respectively, and with $\|\cdot\|_2$ denoting the Euclidean norm. An illustration of this is shown in Figure 1, for the case of a single source and two sensors. As is clear from the figure, the relative time delay between the first and mth sensors will be

$$\tau_{k,m} = \frac{d_{k,m} - d_{k,1}}{c} \tag{5}$$

where c is the propagation velocity. The impinging signal at sensor m may thus be expressed as

$$y_m(t) = \sum_{k=1}^{K} x_{k,m}(t) + e_m(t) \tag{6}$$

$$= \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} a_{k,\ell} d_{k,m}^{-1} e^{j\omega_k \ell (t - \tau_{k,m})} + e_m(t) \tag{7}$$

$$= \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} b_{k,\ell,m} e^{j\omega_k \ell t} + e_m(t) \tag{8}$$

where the TDOA phase information of the k:th pitch, for overtone ℓ and sensor m, is gathered in the complex amplitude of the signal, $b_{k,\ell,m}$, i.e.,

$$b_{k,\ell,m} \triangleq a_{k,\ell} d_{k,m}^{-1} e^{-j\omega_k \ell \tau_{k,m}} \tag{9}$$

and with $e_m(t)$ denoting an additive noise, which is here assumed to be circularly symmetric Gaussian distributed. For reverberating environments, or for other cases of coherent signals sharing the same fundamental frequency, each such contribution may be modelled as a separate source, increasing K accordingly. Here, we have selected to instead allow each source to have S_k coherent reflections (or, equivalently, allowing for S_k sources with the same fundamental frequency), extending the expected TDOA phase information accordingly, such that

$$b_{k,\ell,m} = \sum_{s=1}^{S_k} a_{k,\ell,s} d_{k,m,s}^{-1} e^{-j\omega_k \ell \tau_{k,m,s}} \tag{10}$$

where $a_{k,\ell,s}$, $d_{k,m,s}$, and $\tau_{k,m,s}$ denote the amplitude, the distance to the mth sensor, and the TDOA for the sth reflection, respectively. To simplify notation, and without loss of generality, we will here restrict our attention to the case when all sources and signals are restricted to a 2-D plane. The results generalize to 3-D by extending the parameter sets accordingly. Furthermore, we here assume a calibrated, although arbitrary, sensor array, only being chosen with enough sensors to avoid ambiguity. Such ambiguities arise as the phase difference between signals may, in general, map to several feasible source locations. To see this, consider a nominal complex amplitude from a single (near-field) sinusoidal signal, b, such that

$$b_m = \frac{a}{\|\mathbf{s} - \mathbf{r}_m\|_2} e^{i\omega\tau} = \frac{a}{\|\mathbf{s} - \mathbf{r}_m\|_2} e^{i\omega\tau + k2\pi} \tag{11}$$

is ambiguous for any $k \in \mathbb{Z}$. The reverse triangle inequality implies that

$$\Big| \|\mathbf{s} - \mathbf{r}_m\|_2 - \|\mathbf{s} - \mathbf{r}_1\|_2 \Big| \leq \|\mathbf{r}_m - \mathbf{r}_1\|_2 \tag{12}$$

and, given (5), that the TDOA for such a feasible source location must fulfill

$$\tau c \in \Big[-\|\mathbf{r}_m - \mathbf{r}_1\|_2, \|\mathbf{r}_m - \mathbf{r}_1\|_2 \Big] \tag{13}$$

limiting the TDOA to an interval that depends on the distance between the sensors, where the endpoints of the interval corresponds to source positions that are on either side of the sensors, positioned exactly on a line running through both. Thus, using (11), one may note that the same phase information is obtained for any TDOA such that

$$\tau_k c = \frac{\lambda \arg b}{2\pi} + \lambda k \tag{14}$$

where $k \in \mathbb{Z}$, $\arg b \in [-\pi, \pi]$, and $\lambda = 2\pi c/\omega$ is the wavelength of the signal, which, given (13), implies that only such TDOAs that satisfy

$$\tau_k c = \frac{\lambda \arg b}{2\pi} + \lambda k \in \Big[-\|\mathbf{r}_m - \mathbf{r}_1\|_2, \|\mathbf{r}_m - \mathbf{r}_1\|_2 \Big] \tag{15}$$

are feasible solutions. Therefore, if sensors are distanced by less than $\lambda/2$, the feasible τ is unique, and there is no ambiguity in the resulting estimates. In such cases, the TDOA for each sensor pair will form a single half of a hyperbola, as given by

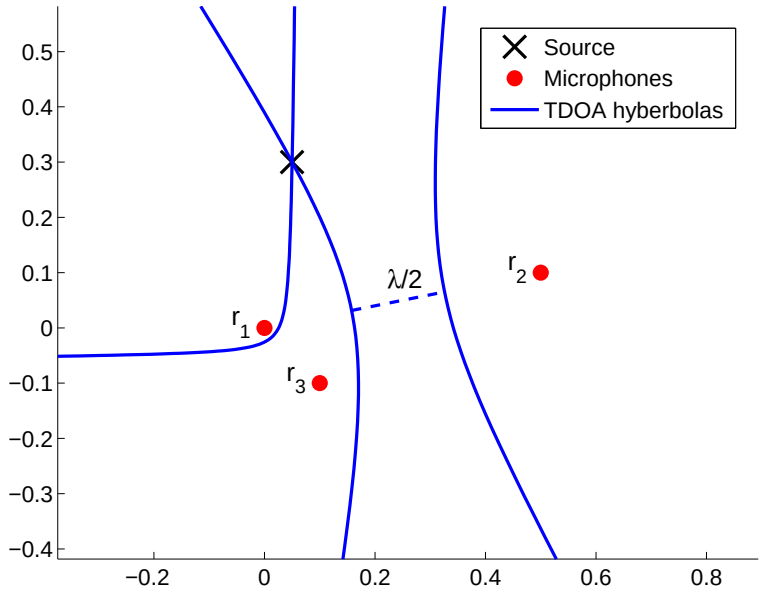

Figure 2: The three hyperbolas represent the possible locations of a source when the TDOA is estimated from the phase of the amplitude of a sinusoid with wavelength λ. As the distance between sensor 1 and 2 is greater than $\lambda/2$, two possible TDOAs and thus two hyperbolas, spaced $\lambda/2$ apart, are in the solution set for that sensor pair. However, the only feasible solution is the one marked with an 'x', as it is the only point which exists in the solution sets from both pairs of sensors.

(5), indicating feasible source locations. If instead some sensors are spaced further apart than $\lambda/2$, then, for all such sensor pairs, there will be more than one feasible TDOA, thereby yielding multiple hyperbolas indicating feasible source locations, with a minimum distance of $\lambda/2$ apart. When using multiple sensors, the feasible source locations are restricted to the intersection of hyperbolas for all sensor pairs. This is illustrated in Figure 2, where a single source emits a 1000 Hz signal, which is recorded by three sensors. As shown in the figure, between sensors one and three, which are less than $\lambda/2$ apart, the source gives a single TDOA and a

corresponding hyperbola, where the source may be located. Between sensors one and two, which are spaced by more than $\lambda/2$ apart, a second TDOA is feasible, λ/c apart from the true one, which yields the same same phase in the complex plane. However, as shown in the figure, the combined hyperbolas coincide in only a single feasible location, thus still allowing for an unambiguous estimate of the source location. For harmonic signals, consisting of multiple sinusoidal signals, each overtone will yield a separate set of hyperbolas, thus also expanding the range of possible locations. However, as we consider finding the location using all harmonics simultaneously, adding a harmonic does not increase the set of possible locations (as we only consider the intersection of all the harmonic's solution sets). Furthermore, using the amplitude attenuation information between sensors, as given by (9), the measured amplitudes may be expressed as

$$|b_m| = \frac{|a|}{\|\mathbf{s} - \mathbf{r}_m\|_2} \tag{16}$$

For each pair of microphones, these equations limit $\mathbf{s}$ to be on a circle. As each harmonic follows the same path loss model, each harmonic yields the same circle as the pitch, and thus does not add any information to the question of uniqueness. Instead, in the noisy case, adding more harmonics only adds to the precision of the location, as the signal-to-noise ratio (SNR) increases. Finally, as more sensors are added to the array, the set of possible locations quickly becomes small, and a unique solution generally exists. We thus deem that the imposed restriction on the array's geometry is mild.

3 Joint pitch and localization estimation

We proceed to detail the proposed two-step procedure to form reliable estimates of both the pitches and locations of the sources impinging on the array, without assuming detailed model knowledge of either the number of sources, K, the number of overtones for each source, L_k, the number of reflections experienced due to a possibly reverberant environment, S_k, or requiring knowledge about if sources are far- or near-field. In the first step, the amplitudes, phases, fundamental frequencies, and model orders of the present pitches are estimated, whereas, in the second step, the phase estimates are used to find the locations of these sources.

Let

$$\mathbf{\Phi} = \left\{ \left\{ b_{k,\ell,m} \right\}_{\substack{\ell=1,\ldots,L_k \\ m=1,\ldots,M}}, \omega_k, L_k \right\}_{k=1,\ldots,K} \tag{17}$$

denote the set of unknown parameters to be determined in the first step. Minimizing the squared model residual in (8), an estimate of $\mathbf{\Phi}$ may thus be formed as

$$\hat{\mathbf{\Phi}} = \arg\min_{\mathbf{\Phi}} \sum_{t=1}^{N} \sum_{m=1}^{M} \left| y_m(t) - \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} b_{k,\ell,m} e^{j\omega_k \ell t} \right|^2 \tag{18}$$

Clearly, given the dimensionality of the problem, and the required model order estimation steps in order to determine K and L_k, this is a non-trivial problem, and needs to be modified to allow for an efficient solution, as is detailed below. In the second step, the found amplitude and phase estimates, $\hat{b}_{k,l,m}$, are then exploited to form estimates of the source locations. Let

$$\mathbf{\Psi}_k = \left\{ \left\{ a_{k,\ell,s} \right\}_{\ell=1,\ldots,L_k}, \mathbf{s}_s \right\}_{s=1,\ldots,S_k} \tag{19}$$

Then, the locations may be determined by minimizing the squared model residual in (10), i.e.,

$$\hat{\mathbf{\Psi}}_k = \arg\min_{\mathbf{\Psi}_k} \sum_{\ell=1}^{\hat{L}_k} \sum_{m=1}^{M} \left| \hat{b}_{k,\ell,m} - \sum_{s=1}^{S_k} a_{k,\ell,s} d_{k,m,s}^{-1} e^{-j\omega_k \ell \tau_{k,m,s}} \right|^2 \tag{20}$$

where $\tau_{k,m,s}$ and $d_{k,m,s}$ are functions of the location $\mathbf{s}_s$, as defined in (4) and (5). As before, this minimization is also non-trivial, requiring an estimate of S_k, and also needs to be modified to allow for a reasonably efficient solution. In the following, we will elaborate on the proposed modifications of the above minimizations. In order to do so, we first extend the sparse pitch estimation algorithm presented in [23,24] to allow for multiple measurement vectors. For the second minimization, we then introduce a similar sparsity pattern to solve the localization problem. We begin by examining the extended pitch estimation algorithm.

3.1 Sparse pitch estimation

Define the measurement matrix

$$\mathbf{Y} = \left[\begin{array}{ccc} \mathbf{y}(1) & \ldots & \mathbf{y}(N) \end{array} \right]^T \tag{21}$$

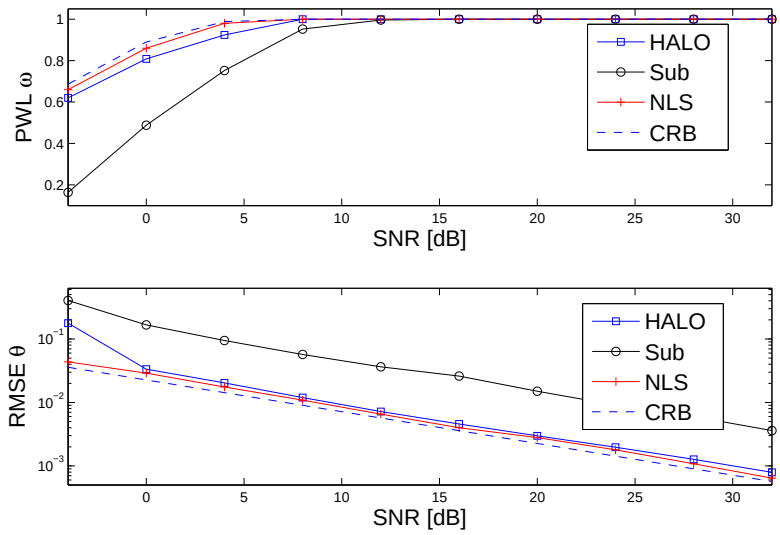

Figure 3: The PWL and RMSE for a single-pitch signal as compared with the optimal performance of an estimator reaching the CRB.

where

$$\mathbf{y}(t) = \left[\begin{array}{ccc} y_0(t) & \dots & y_{M-1}(t) \end{array}\right]^T \tag{22}$$

denotes a sensor snapshot for each time point $t = 1, \dots, N$, with $(\cdot)^T$ being the transpose. The measurements may then be concisely expressed as

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{W}_k \mathbf{B}_k + \mathbf{E} \tag{23}$$

where $\mathbf{E}$ denotes the combined noise term constructed similar to $\mathbf{Y}$, and

$$\mathbf{W}_k = \left[\begin{array}{ccc} \mathbf{w}_k^1 & \dots & \mathbf{w}_k^{L_k} \end{array}\right] \tag{24}$$

$$\mathbf{w}_k = \left[\begin{array}{ccc} e^{j\omega_k} & \dots & e^{j\omega_k N} \end{array}\right]^T \tag{25}$$

$$\mathbf{B}_k = \left[\begin{array}{ccc} \mathbf{b}_{k,1} & \dots & \mathbf{b}_{k,L_k} \end{array}\right]^T \tag{26}$$

$$\mathbf{b}_{k,\ell} = \left[\begin{array}{ccc} b_{k,\ell,1} & \dots & b_{k,\ell,M} \end{array}\right]^T \tag{27}$$

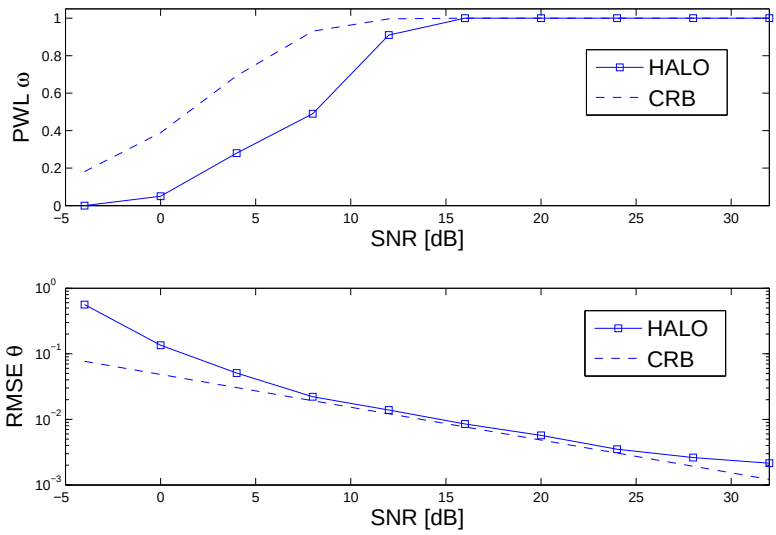

Figure 4: The PWL and RMSE for a multi-pitch signal with two pitches, as compared to the corresponding CRB.

Reminiscent to the sparse estimation framework proposed in [19], we form an extended dictionary of feasible fundamental frequencies, $\omega_1, \ldots, \omega_P$, where $P \gg K$, being chosen so large that K of these will reasonably well coincide with the true pitches in the signal. In the same manner, the number of harmonics of each pitch is extended to an arbitrary upper level, say $L_{\max}$, for all dictionary elements. The signal model may thus be expressed as

$$\mathbf{Y} = \sum_{p=1}^{P} \mathbf{W}_p \mathbf{B}_k + \mathbf{E} = \mathcal{W}\mathcal{B} + \mathbf{E} \tag{28}$$

where the block dictionary matrices are formed by stacking the matrices such that

$$\mathcal{W} = \begin{bmatrix} \mathbf{W}_1 & \ldots & \mathbf{W}_P \end{bmatrix} \tag{29}$$

$$\mathcal{B} = \begin{bmatrix} \mathbf{B}_1^T & \ldots & \mathbf{B}_P^T \end{bmatrix}^T \tag{30}$$

Note from (28) that if the element (ℓ, r) of the matrix $\mathbf{B}_k$ is non-zero, the frequency $\ell\omega_k$ is present in the signal at sensor r. Furthermore, since we assume all sensors to receive essentially the same signal, although time-delayed, we may assume that for a harmonic signal, the rows off a non-zero $\mathbf{B}_k$ will either be non-zero, implying that the harmonic ℓ is present in the pitch, or zero, if the harmonic is missing. An appropriate criterion, that promotes a combination of model to data fit and the sparsity pattern just described, may thus be formed as

$$\underset{\mathcal{B}}{\text{minimize}} \frac{1}{2}\|\mathbf{Y} - \mathcal{W}\mathcal{B}\|_{\mathcal{F}}^2 + \lambda \sum_{p=1}^{P} \sum_{\ell=1}^{L_p} \left\|\mathbf{b}_{p,\ell}\right\|_2 + \sum_{p=1}^{P} \gamma_p \left\|\mathbf{B}_p\right\|_{\mathcal{F}} \tag{31}$$

where two different kinds of group sparsities are imposed, and with $\|\cdot\|_{\mathcal{F}}$ denoting the Frobenius norm. This can be seen to be a generalization of the sparse group lasso to the multiple measurement case (see also [24, 26]). Here, the double sum of 2-norms, which is in the second entry of the minimization, should enforce sparsity in the solution in the rows of $\mathcal{B}$, and ideally only have as many non-zero rows as there are sinusoids in the signal. The third entry makes the solution (matrix) block sparse over the candidate pitches, penalizing the number of pitches with non-zero magnitude in the signal, ideally making them as many as there are pitches in the signal, i.e., K. Given an optimal point, $\hat{\mathcal{B}}$, the number of pitches is thus estimated as the number of non-zero matrices $\hat{\mathbf{B}}_k$, and, for each pitch, the number of harmonics, L_k, is estimated as the number of non-zero rows. The user parameters $\lambda, \gamma_p \in \mathbb{R}_+$ weighs the fit of the solution to its vector and matrix sparsity, respectively.

It is well known (see, e.g., [27]) that the amplitudes in the sparse estimate will be increasingly biased towards zero as sparse regularizers are increased. As we here intend to use both the estimated phases and the amplitudes, we propose to refine the amplitude estimates using a reweighting scheme similar to the one presented in [28]. This is accomplished by iteratively solving (31), such that at iteration $j+1$, one updates

$$\gamma_p^{(j+1)} = \frac{\gamma_p^{(0)}}{\left\|\hat{\mathbf{B}}_p^{(j)}\right\|_{\mathcal{F}} + \varepsilon} \tag{32}$$

where $\hat{\mathbf{B}}_p^{(j)}$ is block p of the optimal point for iteration j, and all $\gamma_p^{(0)}$ are set to be equal in the first iteration. As a result, the block matrices, $\hat{\mathbf{B}}_p^{(j)}$, which have

a small Frobenius norm at iteration j will be penalized harder in the next step, whereas the ones that have a larger Frobenius norm will be penalized less, and as a result reducing the bias. The resulting algorithm can be seen as a sequence of iterative convex programs to approximate the concave $\log(\sum_{p=1}^{P} \gamma_p^{(0)} \|\mathbf{B}_p\|_{\mathcal{F}} + \varepsilon)$ penalty function [29], where ε is chosen as a small number to avoid numerical difficulties. The introduction of the reweighting yields sparser estimates due to the introduction of the log penalty [28, 30], and the resulting technique may be viewed as an alternative to using an information criterion (as was done in [24], to avoid spurious peaks caused by the signal model and data miss-match).

It is worth noting that as we are here focusing on localization, we have selected to use a somewhat simplistic audio model that ignores several important features in harmonic audio signals, such as issues of inharmonicities, pitch halvings and doublings, and the commonly occurring forms of amplitude modulation exhibited by most audio sources (see also [15]). Clearly, the used model could be refined reminiscent to models such as the one used in [24,31], introducing a total variation penalty to each column of $\mathcal{B}$, and/or using an uncertainty volume to allow for inharmonicity. However, for localization purposes, these issues are of less concern, as halvings/doublings and/or amplitude modulations will not affect the below localization procedure more than marginally. Inharmonicity is more pressing, but we have in our numerical studies found that given the size of the calibration errors, the inharmonicity is not affecting the solution significantly, and in the interest of reducing the complexity, we have opted to exclude this aspect from the estimator.

As for the selection of the tuning parameters, one may use, for example, cross validation techniques, although it may be noted that, in high SNR cases, one can often get good results by simply inspecting the periodogram and by then setting the tuning parameters appropriately (see also [24] for a further discussion on this issue). Furthermore, we note that in the case of different noise variances at each sensor in the array, the Frobenius norm in the first entry of the minimization criterion may be replaced with a weighed Frobenius norm. Finally, we note that non-Gaussian noise distributions can also be used as long as the negative log-likelihood is convex.

3.2 Sparse phase- and attenuation- based localization

As the phase estimates in $\hat{\mathcal{B}}$ will inherently contain estimates of the TDOAs, this enables a range of post-processing steps to, for instance, estimate positions,

track, and/or calibrate the sensors. Here, we limit our attention to estimating the source positions. Let $\hat{\mathcal{B}}$ denote the solution obtained from minimizing (31), and consider a scenario where the sources are well separated in their pitch frequencies, and, initially, suffering from negligible reverberation, implying that $S_1 = \ldots = S_P = 1$. Then, the minimization in (20) may be seen as a generalization of the time-varying amplitude modulation problem examined in [32] (see also [12]) to the case of several realizations of the same signal, sampled at irregular time points, and with a different initial phase for each realization. Reminiscent to the solution presented in [12, p. 186], one may thus find the source locations, for far-field signals, for every pitch p with non-zero amplitudes in $\mathbf{B}_p$, as

$$\hat{\mathbf{s}}_p = \arg\max_{\mathbf{s}_p} \sum_{\ell=1}^{L_p} \left| \sum_{m=1}^{M} \hat{b}^2_{p,\ell,m} e^{-j2\omega_p \ell \tau_{p,\ell,m}} \right|^2 \tag{33}$$

where the TDOAs $\tau_{p,\ell,m}$ are found as a function of the source location $\mathbf{s}_p$, using (5). This minimization may be well approximated by 1-D searches over range and DOA (or over azimuth and elevation in the 3-D case). Considering also reverberating room environments, wherein each of the pitches may appear as originating from many different locations, the minimization needs to be extended to allow for varying number of reflections, S_k. To allow for such reflections, we proceed to model every non-zero amplitude block from the pitch estimation step as

$$\mathbf{B}_k = \sum_{s=1}^{S_k} \operatorname{diag}\left(\mathbf{a}_{k,s}\right) \mathbf{U}_{k,s} + \boldsymbol{\mathcal{E}}_k \tag{34}$$

with diag($\mathbf{x}$) denoting a diagonal matrix with the vector $\mathbf{x}$ along its diagonal, $\boldsymbol{\mathcal{E}}_k$ the combined noise term constructed in the same manner as $\mathbf{B}_k$, and

$$\mathbf{U}_{k,s} = \begin{bmatrix} \mathbf{u}^1_{k,s} & \ldots & \mathbf{u}^{\hat{L}_k}_{k,s} \end{bmatrix} \tag{35}$$

$$\mathbf{u}_{k,s} = \begin{bmatrix} \frac{e^{j\omega_k \tau_{k,1,s}}}{d_{k,1,s}} & \ldots & \frac{e^{j\omega_k \tau_{k,M,s}}}{d_{k,M,s}} \end{bmatrix}^T \tag{36}$$

$$\mathbf{a}_{k,s} = \begin{bmatrix} \mathbf{a}_{k,1,s} & \ldots & \mathbf{a}_{k,\hat{L}_k,s} \end{bmatrix}^T \tag{37}$$

where $\tau_{k,m,s}$ and $d_{k,m,s}$ are related to the source location as given by (4) and (5), respecively. Analogously to the above procedure for the pitch estimation, we then extend the dictionary of feasible source locations for the kth source, $\mathbf{s}_1, \ldots, \mathbf{s}_{S_k}$,

onto a grid of $Q \gg S_k$ candidate locations $\mathbf{s}_q$, for $q = 1, \ldots, Q$, with Q chosen large enough to allow some of the introduced dictionary elements to coincide, or closely so, with the true source locations in the signal. Clearly, this may force Q to be very large. Striving to keep the size of the dictionary as small as possible, we consider grid points in polar coordinates, with equal resolution for all considered DOAs, and linearly spaced grid points over the distance in each DOA. Thus, we get a denser grid in the close proximity to the sensor array, where the resolution capacity is highest, and then a less and less dense grid for sources further away from the array. Finally, to also allow for far-field sources, we can include one dictionary element for each direction at an infinite range, for which all the relative attenuations, $d_{k,l,s}$, are set to be equal to 1. Thus, we may estimate the source locations for the k:th pitch using a sparse modeling framework as

$$\underset{\mathbf{a}_{k,1},\ldots,\mathbf{a}_{k,Q}}{\text{minimize}} \frac{1}{2} \left\| \mathbf{B}_k - \sum_{q=1}^{Q} \operatorname{diag} \mathbf{a}_{k,q} \mathbf{U}_{k,q} \right\|_{\mathcal{F}}^2 + \sum_{q=1}^{Q} \varkappa_q \left\| \mathbf{a}_{k,q} \right\|_2 + \rho \sum_{q=1}^{Q} \left\| \mathbf{a}_{k,q} \right\|_1 \tag{38}$$

where, again, two types of sparsity is imposed on the solution. The 2-norm penalty term imposes sparsity to the blocks $\mathbf{a}_{k,q}$, i.e., penalizing the number of source locations present in the signal. Furthermore, the 1-norm term penalizes the number of harmonics, to allow for cases when some sources may have missing harmonics. Thus, here the number of sources is estimated as the number of nonzero blocks in an optimal point and any zero elements within a block corresponding to a missing harmonic. Here, $\varkappa_q, \rho \in \mathbb{R}_+$ are tuning parameters, controlling the amount of sparsity and the weight between sparsity in pitches and in harmonics, respectively, whereas the factor ρ is only used if two sources share the same fundamental frequency but differ in which harmonics are present. Finally, $\varkappa_q$ may be updated in the same manner as described in section III.A. As shown in the following section, the optimization problem in (31) and (38) are equivalent, so these tuning parameters may be set in a similar fashion.

4 An efficient ADMM implementation

It is worth noting that both the minimization in (31) and (38) are convex, as the tuning parameters are non-negative and all the functions are convex. Their solutions may thus be found using standard convex minimization techniques, e.g.,

Algorithm 1 The ADMM algorithm

1: Initiate $\mathbf{z} = \mathbf{z}_0$, $\mathbf{u} = \mathbf{u}_0$, and $k = 0$
2: **repeat**
3: $\quad \mathbf{z}_{k+1} = \underset{\mathbf{z}}{\operatorname{argmin}} f(\mathbf{z}) + \frac{\mu}{2}||\mathbf{z} - \mathbf{u}_k - \mathbf{d}_k||_2^2$
4: $\quad \mathbf{u}_{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}}\, g(\mathbf{u}) + \frac{\mu}{2}||\mathbf{z_{k+1}} - \mathbf{u} - \mathbf{d}_k||_2^2$
5: $\quad \mathbf{d}_{k+1} = \mathbf{d}_k - (\mathbf{z_{k+1}} - \mathbf{u_{k+1}})$
6: $\quad k \leftarrow k + 1$
7: **until** convergence

using CVX [33, 34], SeDuMi [35], or SDPT3 [36]. Regrettably, such solvers will scale poorly both with increasing data length, the use of a finer grid for the fundamental frequencies, and with the number of sensors. Furthermore, such implementations are unable to utilize the full structure of the minimization, and may, as a result, be computationally cumbersome in practical situations. To alleviate this, we proceed to introduce a novel ADMM re-formulation of the minimizations, offering efficient and fast implementations of both minimizations. For completeness and to introduce our notation, we briefly review the main steps involved in an ADMM (we refer the reader to [37, 38] for further details on the ADMM).

Considering the convex optimization problem

$$\underset{\mathbf{z}}{\text{minimize}}\, f(\mathbf{z}) + g(\mathbf{z}) \tag{39}$$

where $\mathbf{z} \in \mathbb{R}^p$ is the optimization variable, with $f(\cdot)$ and $g(\cdot)$ being convex functions. Introducing the auxiliary variable, $\mathbf{u}$ (39) may be equivalently be expressed as

$$\underset{\mathbf{z},\mathbf{u}}{\text{minimize}}\, f(\mathbf{z}) + g(\mathbf{u}), \quad \text{subject to}\, \mathbf{z} - \mathbf{u} = \mathbf{0} \tag{40}$$

since at any feasible point $\mathbf{z} = \mathbf{u}$. Under the assumption that there is no duality gap, which is true for the here considered minimizations, one may solve the optimization problem via the dual function defined as the infimum of the augmented Lagrangian, with respect to $\mathbf{x}$ and $\mathbf{z}$, i.e., (see also [37])

$$L_\mu(\mathbf{z}, \mathbf{u}, \mathbf{d}) = f(\mathbf{z}) + g(\mathbf{u}) + \mathbf{d}^T(\mathbf{z} - \mathbf{u}) + \frac{\mu}{2}||\mathbf{z} - \mathbf{u}||_2^2$$

The ADMM does this by iteratively maximizing the dual function such that at step $k+1$, one minimizes the Lagrangian for one of the variables, while holding the other fixed at its most recent value, i.e.,

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} L_{\mu}\left(\mathbf{z}, \mathbf{u}_k, \mathbf{d}_k\right) \tag{41}$$

$$\mathbf{u}_{k+1} = \arg\min_{\mathbf{u}} L_{\mu}\left(\mathbf{z}_{k+1}, \mathbf{u}_k, \mathbf{d}_k\right) \tag{42}$$

Finally, one updates the dual variable by taking a gradient ascent step to maximize the dual function, resulting in

$$\tilde{\mathbf{d}}_{k+1} = \tilde{\mathbf{d}}_k - \mu\left(\mathbf{z}_{k+1} - \tilde{\mathbf{d}}_{k+1}\right) \tag{43}$$

where μ is the dual variable step size. The general ADMM steps are summarized in Algorithm 1, using the scaled version of the dual variable $\mathbf{d}_k = \tilde{\mathbf{d}}/\mu$, which is more convenient for implementation. Thus, in cases when steps 3 and 4 of Algorithm 1 may be carried out more efficiently than for the original problem, the ADMM may be useful to form an efficient implementation of the considered minimization.

It may be noted that the minimizations in (31) and (38) are rather similar, both containing an affine function in a Frobenius norm, as well as a sum of the norm of different subset of the variable. In fact, by using the vec operation, both minimizations may be shown to be equivalent with the problem

$$\underset{\mathbf{z}}{\text{minimize}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 + \gamma \sum_{k=1}^{P} \|\mathbf{z}_k\|_2 + \delta \sum_{k=1}^{P} \sum_{g=1}^{G_k} \left\|\mathbf{z}_{k,g}\right\|_2$$

where the complex variable $\mathbf{z}$ is given as

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1^T & \dots & \mathbf{z}_P^T \end{bmatrix}^T \tag{44}$$

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{z}_{k,1}^T & \dots & \mathbf{z}_{k,G_k}^T \end{bmatrix}^T \tag{45}$$

where each $\mathbf{z}_k$ and $\mathbf{z}_{k,g}$ denote complex vectors with G_k and O elements, respectively. For the minimization in (31), this implies that

$$\mathbf{y} = \text{vec}(\mathbf{Y}) \tag{46}$$

$$\mathbf{z} = \text{vec}(\mathcal{B}) \tag{47}$$

$$\mathbf{A} = \mathbf{I} \otimes \mathcal{W} \tag{48}$$

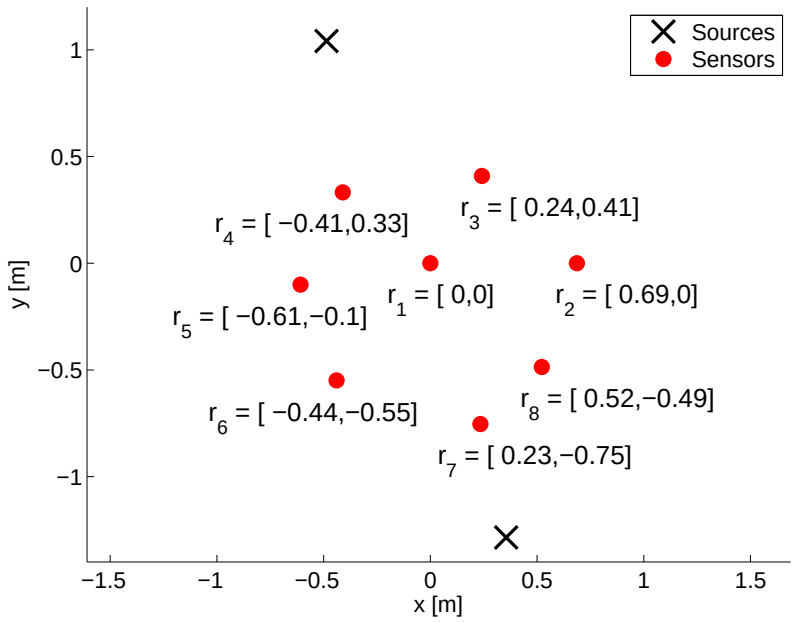

Figure 5: The two-source and eight-sensor layout in 2D. The 2D position of each sensor, shown in the plot with Cartesian coordinates as $r_m = [x, y]$, was obtained in an *a priori* calibration step.

where $\otimes$ and $\mathbf{I}$ denote the Kronecker product and an M-dimensional identity matrix, respectively, with G_k being equal to the number of harmonics, L_k, and O equals the number of sensors, M. Similarly, for the minimization in (38),

$$\mathbf{y} = \text{vec}(\mathbf{B}_p) \tag{49}$$

$$\mathbf{z} = \mathbf{a}_k \tag{50}$$

$$\mathbf{A} = \tilde{\mathbf{V}}_k \tag{51}$$

where

$$\mathbf{a}_k = \begin{bmatrix} \mathbf{a}_{k,1}^T & \dots & \mathbf{a}_{k,Q}^T \end{bmatrix}^T \tag{52}$$

$$\tilde{\mathbf{V}}_k = \begin{bmatrix} \tilde{\mathbf{V}}_{k,1} & \dots & \tilde{\mathbf{V}}_{k,Q} \end{bmatrix} \tag{53}$$

and $\mathbf{V}_{k,q} = \mathbf{U}_{k,q} \otimes \mathbf{I}$, with $\tilde{\mathbf{V}}_{k,q}$ being formed by removing all columns from $\mathbf{V}_{k,q}$ that correspond to zeros in the vector vec(diag($\mathbf{a}_{k,q}$)), and G_k being equal to L_k and O equals 1. Thus, we can formulate and ADMM solution for (44) that solves both problem (31) and (38). To that end, defining

$$f(\mathbf{z}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \tag{54}$$

$$g(\mathbf{u}) = \gamma \sum_{k=1}^{P} \|\mathbf{u}_k\|_2 + \delta \sum_{k=1}^{P} \sum_{g=1}^{Q_k} \left\|\mathbf{u}_{k,g}\right\|_2 \tag{55}$$

yields a quadratic problem in step 3 in Algorithm 1, with a closed form solution given by

$$\mathbf{z}_{k+1} = \left(\mu I + \mathbf{A}^H \mathbf{A}\right)^{-1} \left(\mu\left(\mathbf{u}_k - \mathbf{d}_k\right) + \mathbf{A}^H \mathbf{y}\right)$$

with $(\cdot)^H$ denoting the Hermitian transpose, whereas in step 4, by solving the sub-differential equations (see [24] for further details), one obtains

$$\mathbf{u}_{k+1} = \mathcal{S}^o\left(\mathcal{S}^i\left(\mathbf{z}_k - \mathbf{d}_k, \varkappa/\mu\right), \delta/\mu\right) \tag{56}$$

where the shrinkage operators $\mathcal{S}^o$ and $\mathcal{S}^i$ are defined using the vector shrinkage operator $\mathcal{S}$, defined for any vector $\mathbf{v}$ and positive scalar ξ such that

$$\mathcal{S}(\mathbf{v}, \xi) = \mathbf{v}\left(1 - \xi/||\mathbf{v}||_2\right)^+ \tag{57}$$

where $(\cdot)^+$ is the positive part of the scalar, and

$$\mathcal{S}(\mathbf{z}, \xi)^o = \left[\ \mathcal{S}^T(\mathbf{z}_1, \xi) \quad \ldots \quad \mathcal{S}^T(\mathbf{z}_P, \xi)\ \right]^T \tag{58}$$

$$\begin{aligned}\mathcal{S}(\mathbf{z}, \xi)^i = \Big[\ & \mathcal{S}^T(\mathbf{z}_{1,1}, \xi) \quad \ldots \quad \mathcal{S}^T(\mathbf{z}_{1,G_1}, \xi) \quad \ldots \\ & \mathcal{S}^T(\mathbf{z}_{P,1}, \xi) \quad \ldots \quad \mathcal{S}^T(\mathbf{z}_{P,G_P}, \xi)\ \Big]^T\end{aligned} \tag{59}$$

The resulting algorithm is here termed the Harmonic Audio LOcalization using block sparsity (HALO) estimator.

5 Numerical comparisons

We proceed to examine the performance of the proposed estimator using both synthetic and measured audio signals, initially examining the performance using

simulated audio signals. In the first examples, we limit ourselves to the case of letting a far-field signal impinge on a uniform linear array (ULA). Figure 3 shows the percentage within limits (PWL), defined as the ratio of pitch estimates within a limit of ± 0.1 Hz from the true pitch, and the root mean square error (RMSE) of the DOA, defined as

$$\mathrm{RMSE}_{\vartheta} = \sqrt{\frac{1}{nK} \sum_{k=1}^{K} \sum_{i=1}^{n} \left(\hat{\vartheta}_{k,i} - \vartheta_k \right)^2} \tag{60}$$

where n denotes the number of Monte Carlo (MC) simulation estimates, and K the number of pitches in the signal, for the resulting estimates. For comparison, we use the Cramér-Rao lower bound (CRB), the NLS estimator, and the Sub approach (see [16] for further details on these methods and for the corresponding CRB). These results have been obtained using $n = 250$ MC simulations of a single pitch signal, with $\omega_1 = 220$ Hz and $L_1 = 7$ harmonics, impinging from $\vartheta_1 = -30°$, where both the NLS and the Sub estimators have been allowed perfect a priori knowledge of both the number of sources and their number of harmonics, whereas the proposed method is allowed no such knowledge. As is clear from the figures, the HALO method offers a preferable performance as compared to the Sub estimator, and only marginally worse than the NLS estimator, in spite of both the latter being allowed perfect model orders information. Here, the number of sensors in the array was $M = 5$ and we used 20 ms of data sampled at $f_s = 8820$ Hz, i.e., $N = 176$ samples. Furthermore, $c = 343$ m/s and $d = c/f_s \approx 0.0389$ m. We proceed to consider the case of multi-pitch signals impinging on the array. Measuring as in the single-pitch case, we now form a multi-pitch signal with two pitches and fundamental frequencies $\{150,\ 220\}$ Hz containing $\{6,\ 7\}$ harmonics, coming from $\vartheta_1 = -30°$. Figure 4 shows the RMSE and PWL estimates, as obtained using 250 Monte Carlo simulations, clearly showing that the HALO estimator is able to reach close to optimal performance also in this case. Here, no comparison is made with the NLS and Sub estimators of [16] as these are restricted to the single-pitch case. Throughout these evaluations, we have used $L_{\max} = 15$. Also, as the resulting estimates were found to be appropriately sparse when using only the convex penalties, and no reweighing steps were used.

We next proceed to examine real measured signals. The measurements were made in an anechoic chamber, approximately $4 \times 4 \times 3$ meters in size, with the sensors and speakers located as shown in Figures 5 and 7. Two speakers were

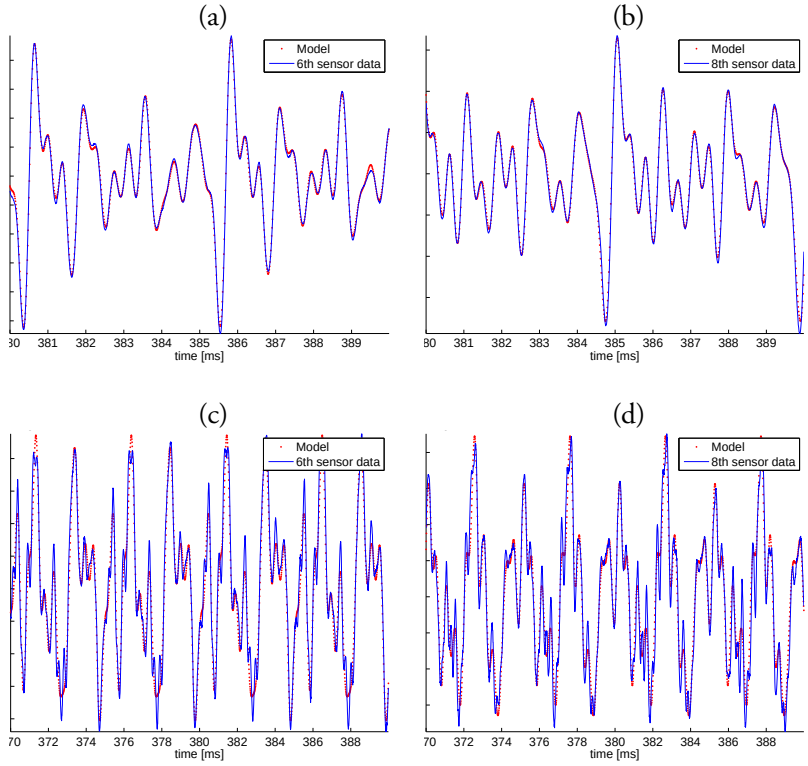

Figure 6: Time-domain data at a sensor (lined), overlaid with the signal model reconstruction (dotted). Panels (a) and (b) correspond to a speech recording, while Panels (c) and (d) correspond to a violin recording, in both cases, at sensor 6 and 8, respectively.

placed at locations (in polar coordinates) $\mathbf{s}_1 = [\vartheta_1, R_1] = [115.03°, 1.15\,\text{m}]$ and $\mathbf{s}_2 = [\vartheta_2, R_2] = [-74.53°, 1.33\,\text{m}]$, with respect to the central microphone, respectively. The positions of the sensors were determined by placing them together with the sources, using the acoustic method detailed in [39]. This is done by calibrating the sensors with a single moving source, using a correlation-based methodology. The positions were also confirmed via a computer vision approach were the positions were found by taking several photos and reconstructing the environment. The maximum deviation in position between these methods was less than 1 mm, which was considered to be precise enough. As the spatial impulse responses of the microphones were deemed to be reasonably omni-directional, as well as roughly the same for all the microphones, no further calibration of the sensor gains were performed. The positions were then projected onto a 2-D plane using principal component analysis. In order to illustrate the HALO estimator's ability to handle an environment with the same pitch signal originating from different sources, as in a reverberating room environment, we examine a case with two sources playing the same signal content. Both sources plays a (TIMIT) recording of a female voice saying 'Why were you away a year, Roy?', timing the source's playback so that the recording at each microphone sounds slightly echoic. The eight microphones all record at a sample rate of $f_s = 96$ kHz. The data is then divided into time frames of 10 ms, i.e., $N = 960$ samples, which allow each frame to be well modeled as being stationary. Examining a part of the speech that is voiced, arbitrarily selected as the frame starting 380 ms into the recording, about when the voice is saying the voiced phonetic sound 'a' in 'why', Panel (a) and (b) in Figure 6 show the signal measured at the 6th and 8th microphone, respectively, together with the reconstructed signal obtained from the pitch estimation step in HALO, obtained as

$$\hat{\mathbf{Y}} = \mathcal{W}\hat{\mathcal{B}} \tag{61}$$

using the resulting model orders and estimates. The estimator indicate that the signal contains a single pitch at $\hat{\omega}/2\pi = 193.5$ Hz, having $\hat{L} = 12$ overtones. As is clear from the figures, the estimator is well able to model the measured signal in spite of the presence of the reverberation. Comparing the figures, one may also note the time shift between the sensors, due to the additional time-delay for the wavefront traveling between them, corresponding to a linear combination of the two sources, each with their particular TDOA and attenuation. It should also be noted that the signals are not simply time-shifted versions of each other due

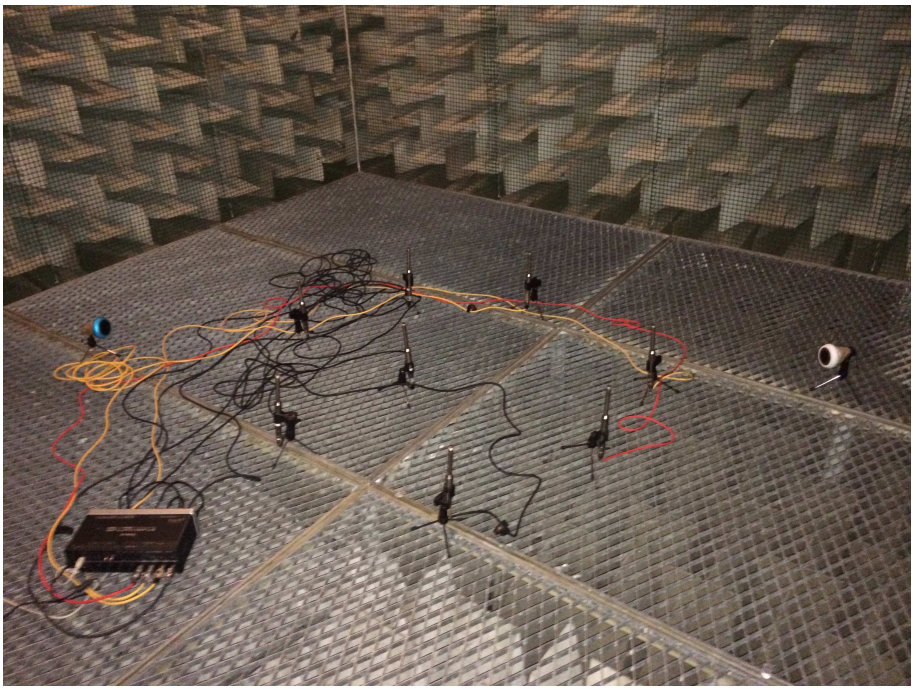

Figure 7: A photo showing the experimental setup in the anechoic chamber.

to the room environment and the attenuation of the signal when propagating in space (which would thus create problems for an estimator based on the cross-correlation between the sensors). The same situation is illustrated in Panel (c) and (d) in Figure 6 showing the results when the signal source is replaced with that of a part of a (SQAM) violin signal. Again, the estimator can be seen to be able to well model the impinging signals, which is estimated as being a single pitch with the fundamental frequency $\hat{\omega}/2\pi = 198.0$ Hz, containing $\hat{L} = 14$ harmonics. In order to examine the location estimation, we construct a 2-D grid of feasible locations, chosen such that the space is discretized into 1008 points, consisting of 72 directions between $[-180°, 180°)$, spaced every $5°$, where each direction allows for ranges $R \in [0.7, 2]$ m, spaced 10 cm apart. The resulting grid is shown in Figure 8, which is roughly covering the entirety of the anechoic chamber. To also allow for far-field sources, a range of $R = \infty$ is also added to the grid for each direction, which we have chosen to illustrate by the outer circle in Figure 8. For

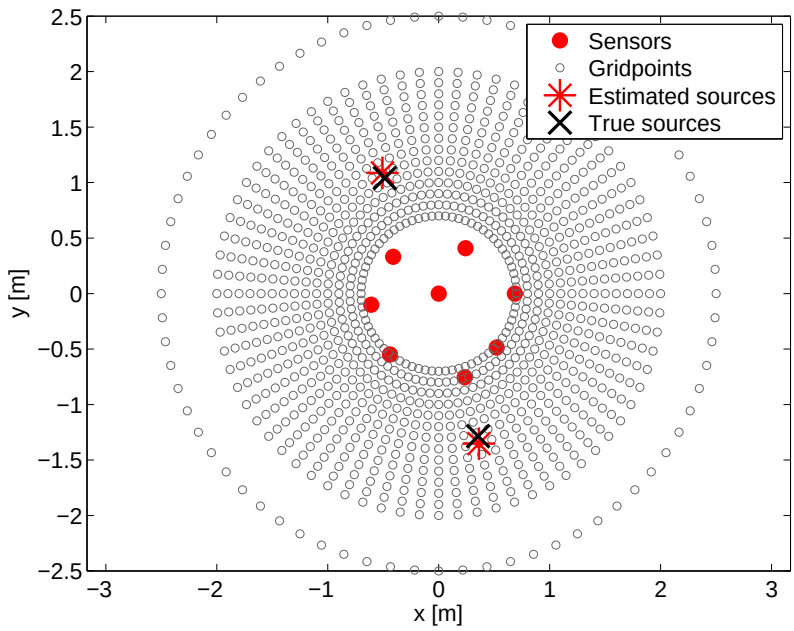

Figure 8: The experimental setup in the anechoic chamber, showing the sensor and loudspeaker locations, the considered dictionary grid, as well as the resulting estimated as obtained by the proposed algorithm.

these far-field grid points, the time-delays are instead computed as (see also [10])

$$\tau_m = \frac{\min_{\mathbf{z}} \|\mathbf{r}_m - \boldsymbol{\ell}(\mathbf{z})\|_2}{c} \tag{62}$$

for a location $\mathbf{z}$ on the line $\boldsymbol{\ell}(\cdot)$, which is perpendicular to the DOA and goes through $\mathbf{r}_1$. The figure also shows the locations for the sensors and the sound sources, as well as the estimated locations, as obtained by the second step of the HALO estimator (the estimated locations were identical for both audio recordings). The errors in position were 5 cm in range for each source, where a bias, overestimating the range, accounts for almost all of the error. On the other hand, as shown in the figure, the angles of the sources ϑ were accurately estimated. The overestimation of the range may to a large extent likely be explained by poor

scaling when calibrating the array.

Finally, we illustrate the algorithm's performance using Monte Carlo (MC) simulations, using simulated sources, one near- and one far-field source, detailed with $\omega = [200, 270]$ Hz, $L = [15, 14]$ harmonics, impinging from $\vartheta = [110°, -70°]$ at $R = [1.3, \infty]$ m, respectively. The sensors are placed as a uniform circular array, with 7 sensor placed evenly at a 0.5 m radius, together with a sensor being placed in the center of the array. First, we examine the position estimates using a coarse spacing for the possible sources, spaced by 11 cm in angle for all angles $\vartheta \in [-180°, 180°)$, and spaced by 10 cm in range, at $R \in [0.7, 3]$ m. In each MC simulation, the true location of each source was offset by a (uniformly distributed) range offset of plus minus one half the grid spacing. In all simulations, we ensured that neither of the sources were placed on a dictionary grid point. Figure 9 shows the PWL for the angle and range estimates, where the limit is chosen to be the same as the grid spacing, i.e., the ratio of estimates that are within ± 1 dm in range, and $\pm 5°$ in angle. As seen from the figure, both the range and the DOA of the sources are well determined, indicating that even with the use of a coarse grid, one is able to obtain reliable estimates. Proceeding to instead using a fine grid, the coarse estimates may then be refined by zooming in the grid over the found locations. Using a dictionary of the same size as the coarse grid, although centered around the found estimates, yields a resolution of ± 5 mm in range and $\pm 0.25°$ in angle. Figure 10 shows the resulting RMSE for the angle and pitch estimates on the finer grid, as compared to the CRB (given in the Appendix). As can be seen from the figure, the RMSE (and the corresponding CRB) of the far-field source is somewhat lower than the near-field source, although both sources are well estimated, yielding a performance close to being optimal. The slight offset from the CRB is deemed to be largely due to a small bias in the final estimates, resulting from the smoothness of the approximative cost function resulting from the additive convex constraints. As is clear from the above presentation, the HALO estimate exploits the harmonic structure in the received audio signals to position the sources, using the pitch estimates to form a sparse estimate over a wide range of feasible positions. Obviously, most audio signals are not harmonic at all times, and the estimator should thus be used in combination with a tracking technique, possibly using a methodology reminiscent to the one presented in [40, 41]. In such a tracking scheme, the estimated pitch amplitudes should be used as an indicator for the reliability of the obtained positioning, yielding poor or maybe even erroneous positioning for unvoiced or

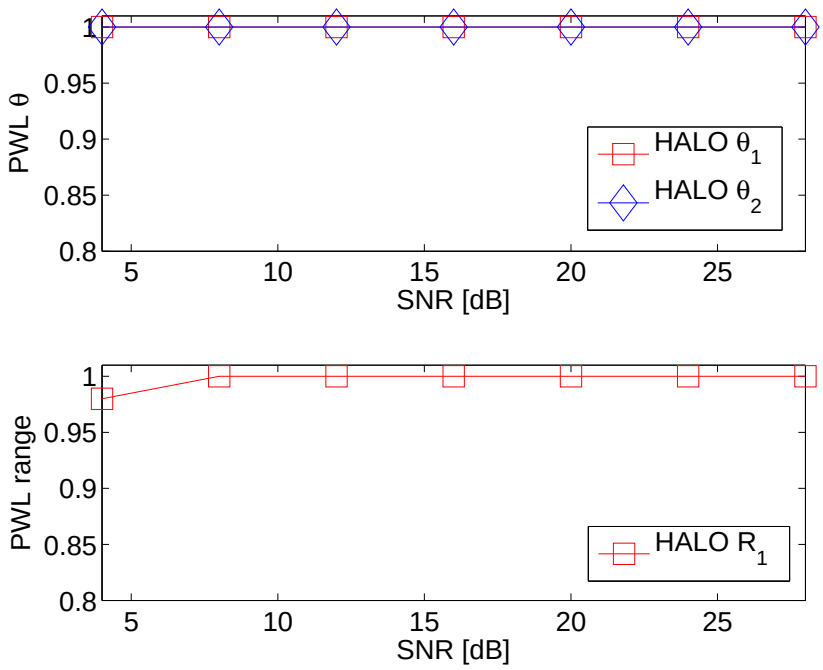

Figure 9: The PWL ratio for the angle and range estimates when using a coarsely spaced grid, indicating the ratio of estimates that are within ± 10 cm in range, and $\pm 5°$ in angle.

non-harmonic audio signals, whereas reasonably accurate positions may be expected for more harmonic signals.

6 Conclusions

In this paper, we have presented an efficient sparse modeling approach for localizing harmonic audio sources using a calibrated sensor array. Assuming that each harmonic components in each pitch can only come from one source, the localization estimate is based on the phase and attenuation information for all of the harmonics jointly. The resulting model phases and attenuation will then

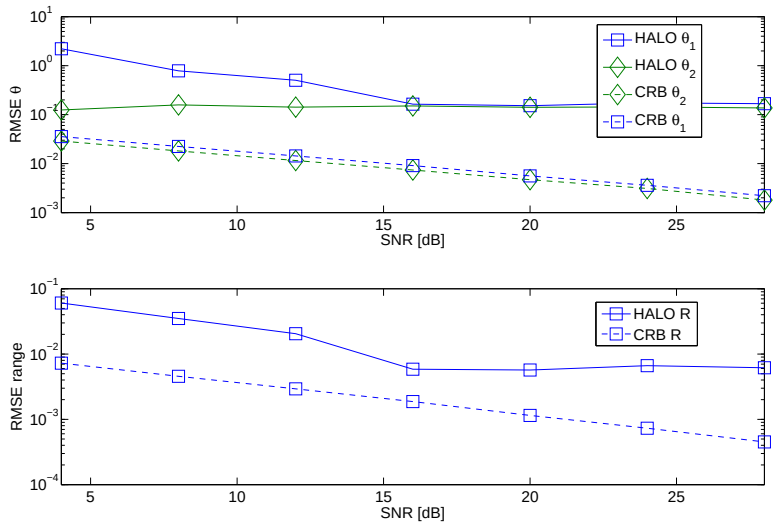

Figure 10: The RMSE for the angle and range estimates when using a finely spaced grid, approximately ±5 mm in range and ±0.25° in angle.

depend on the source location. By using sparse modeling, the method inherently estimates both the number of sources, the number of harmonics in each source, as well as the extent of a possibly occurring reverberation. The effectiveness of the resulting algorithm is shown using both simulated and measured audio sources.

Acknowledgements

The authors wish to express their gratitude to the Signal Processing Group at Electrical and Information Technology, Lund University, for allowing use of their experimental facilities, as well as to the authors of [16] for sharing their Matlab implementations. This work was supported in part by the Swedish Research Council and Carl Trygger's foundation. This work has been presented in part at the ICASSP 2014 conference [42].

7 Appendix

In this appendix, we briefly summarize the Cramér-Rao lower bound (CRB) for the examined localization problem. As is well known, under the assumption of complex circularly symmetric Gaussian distributed noise, the Slepian-Bangs formula yields [12, p. 382]

$$\left[P_{cr}^{-1}\right]_{ij} = \text{trace}\left[\mathbf{\Gamma}^{-1}\mathbf{\Gamma}_i'\mathbf{\Gamma}^{-1}\mathbf{\Gamma}_j'\right] + 2\mathcal{R}\left[\boldsymbol{\mu}_i'^H\mathbf{\Gamma}^{-1}\boldsymbol{\mu}_j'\right] \tag{63}$$

where $\mathcal{R}$ denotes the real part of a complex scalar, $\mathbf{\Gamma}$ the covariance matrix of the noise process, and $\boldsymbol{\mu}$ is the deterministic signal component, with $\mathbf{\Gamma}_i'$ and $\boldsymbol{\mu}_i'$ denoting the derivative of $\mathbf{\Gamma}$ and $\boldsymbol{\mu}$ with respect to element i of the parameter vector, respectively. For the case of uncorrelated noise with a known variance σ^2, this simplifies to

$$\left[P_{cr}^{-1}\right]_{ij} = 2\mathcal{R}\left[\boldsymbol{\mu}_i'^H\boldsymbol{\mu}_j'\right]/\sigma^2 \tag{64}$$

Using the assumed signal model as measured at sensor m, stacking the observations as in (21), and then using the vec operator on the resulting matrix results, one obtains the $\boldsymbol{\mu}$ function needed for the CRB calculations. Here, the parameters to be estimated are

$$\mathbf{\Delta} = \left\{\left\{a_{k,\ell},\, \varphi_{k,\ell}\right\}_{\ell=1,\ldots,L_k},\, \omega_k,\, \vartheta_{s,k},\, R_{s,k}\right\}_{\substack{s=1,\ldots,S\\ k=1,\ldots,K}} \tag{65}$$

Clearly, the resulting function may easily be derivated with respect to the amplitude, frequency and phase parameters. However, since the location parameter, $\vartheta_{s,k}$ and $R_{s,k}$, enter into the expression in a complicated manner depending on the sensor geometry, the corresponding derivatives are not straight forward for an arbitrary array. For this reason, for the considered array geometries, we here simply approximate the resulting expressions using numerically differentiated expressions.

References

[1] B. Champagne, S. Bedard, and A. Stephenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 2, pp. 148–152, Mar 1996.

[2] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberent rooms," in *Microphone Arrays: Techniques and Applications*, M. Brandstein and D. Ward, Eds., pp. 157–180. Springer-Verlag, New York, 2001.

[3] T. Gustafsson, B. D. Rao, and M. Trivedi, "Source localization in reverberant environments: modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Nov 2003.

[4] E. Kidron, Y. Y. Schechner, and M. Elad, "Cross-modal localization via sparsity," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1390–1404, April 2007.

[5] M. D. Gillette and H. F. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Processing Letters*, vol. 15, pp. 1–4, 2008.

[6] K. C. Ho and M. Sun, "Passive source localization using time differences of arrival and gain ratios of arrival," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 464–477, Feb 2008.

[7] X. Alameda-Pineda and R. Horaud, "A geometric approach to sound source localization from time-delay estimates," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 6, pp. 1082–1095, June 2014.

[8] H. F. Silverman and S. E. Kirtman, "A two-stage algorithm for determining talker location from linear microphone array data," *Computer Speech & Language*, vol. 6, no. 2, pp. 129 – 152, 1992.

[9] G. Shen, R. Zetik, and R.S. Thoma, "Performance comparison of toa and tdoa based location estimation algorithms in los environment," in *Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on*, March 2008, pp. 71–78.

[10] H. Krim and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Magazine*, pp. 67–94, July 1996.

[11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*, John Wiley and Sons, Inc., 2002.

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

[13] J. Benesty, M. Sondhi, M. Mohan, and Y. Huang, *Springer handbook of speech processing*, Springer, 2008.

[14] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, NY, 1988.

[15] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, 2009.

[16] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear Least Squares Methods for Joint DOA and Pitch Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 21, no. 5, pp. 923–933, 2013.

[17] S. Gerlach, S. Goetze, J. Bitzer, and S. Doclo, "Evaluation of joint position-pitch estimation algorithm for localising multiple speakers in adverse acoustical environments," in *Proc. German Annual Conference on Acoustics (DAGA)*, Düsseldorf, Germany, 2011, vol. Mar. 2011, pp. 633–634.

[18] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "Joint DOA and Multi-pitch Estimation Based on Subspace Techniques," *EURASIP J. on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–11, 2012.

[19] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul, jul 2008, pp. 10225–10229.

[20] I. F. Gorodnitsky and B. D. Rao, "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, March 1997.

[21] M. D. Plumbley, S. A. Abdallah, T. Blumensath, and M. E. Davies, "Sparse representations of polyphonic music," *Signal Processing*, vol. 86, no. 3, pp. 417–431, March 2006.

[22] M. Genussov and I. Cohen, "Multiple fundamental frequency estimation based on sparse representations in a structured dictionary," *Digit. Signal Process.*, vol. 23, no. 1, pp. 390–400, Jan. 2013.

[23] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Estimating Multiple Pitches Using Block Sparsity," in *Proc. 38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

[24] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Multi-Pitch Estimation Exploiting Block Sparsity," to appear in Elsevier Signal Processing.

[25] S. L. Marple, "Computing the discrete-time "analytic" signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, September 1999.

[26] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A Sparse-Group Lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[27] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.

[28] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[29] L. Qing, Z. Wen, and W. Yin, "Decentralized jointly sparse optimization by reweighted ell-q minimization," *Signal Processing, IEEE Transactions on*, vol. 61, no. 5, pp. 1165–1170, March 2013.

[30] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, 2010.

[31] N. R. Butt, S. I. Adalbjörnsson, S. D. Somasundaram, and A. Jakobsson, "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities," in *Proc. 38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

[32] O. Besson and P. Stoica, "Exponential signals with time-varying amplitude: parameter estimation via polar decomposition," *Signal Processing*, vol. 66, pp. 27–43, 1998.

[33] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," `http://cvxr.com/cvx`, Sept. 2012.

[34] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited, 2008, http://stanford.edu/∼boyd/graph_dcp.html.

[35] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, August 1999.

[36] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[38] N. Parikh and S. Boyd, "Proximal Algorithms," *Found. Trends Optim.*, vol. 1, pp. 127–239, 2014.

[39] Z. Simayijiang, F. Andersson, Y. Kuang, and K. Åström, "An automatic system for microphone self-localization using ambient sound," in *European Signal Processing Conference (Eusipco 2014)*, 2014.

[40] I. Potamitis, H. Chen, and G. Tremoulis, "Tracking of multiple moving speakers with multiple microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 520–529, Sept 2004.

[41] D. Gatica-Perez, G. Lathoud, J. Odobez, and I. McCowan, "Audiovisual Probabilistic Tracking of Multiple Speakers in Meetings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 601–616, Feb 2007.

[42] T. Kronvall, S. I. Adalbjörnsson, and A. Jakobsson, "Joint DOA and Multi-Pitch Estimation Using Block Sparsity," in *Proc. 39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014.

C

Paper C

Estimating periodicities in symbolic sequences using sparse modeling

Stefan Ingi Adalbjörnsson[1], Johan Swärd[1], Jonas Wallin[2], and Andreas Jakobsson[1]

[1] *Centre for Mathematical Sciences, Lund University, Lund, Sweden*
[2] *Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden*

Abstract

In this work, we propose a method for estimating statistical periodicities in symbolic sequences. Different from other common approaches used for the estimation of periodicities of sequences of arbitrary, finite, symbol sets, that often map the symbolic sequence to a numerical representation, we here exploit a likelihood-based formulation in a sparse modeling framework to represent the periodic behavior of the sequence. The resulting criterion includes a restriction on the cardinality of the solution; two approximate solutions are suggested, one greedy and one using an iterative convex relaxation strategy to ease the cardinality restriction. The performance of the proposed methods are illustrated using both simulated and real DNA data, showing a notable performance gain as compared to other common estimators.

Key words: Periodicity, symbolic sequences, spectral estimation, data analysis, DNA

1 Introduction

Sequences formed from a finite set of symbols, or *alphabet*, occur in a variety of fields, such as, for instance, in genomics, semantic analysis, and categorical time series [1, 2]. Frequently, there is an interest in determining reoccurring patterns, periodicities, in such sequences. For instance, in DNA analysis, the latent periodicities in DNA sequences have been found to be correlated with various forms of functional roles of importance [3–10]. Traditional spectral estimation techniques are not suitable for this problem as symbolic sequences lack the required algebraic structures. For DNA analysis, there is no natural ordering among the four occurring symbols, A, C, G, and T. In earlier literature, several authors have addressed the problem of estimating symbolic periodicity using heuristic mappings from the symbol set to sets of complex numbers. After the transformation the periodicities are estimated through standard estimation methods like, for instance, the periodogram. However, such estimates will suffer from the well-known high variability and/or poor resolution inherent to the periodogram [11]. Other examples of methods that use a mapping to transform the symbolic data include PAM- or QPSK-based mappings, minimum entropy mapping, mapping equivalences, or other transformations [4–7, 9, 10]. Generally, these mappings are computationally intensive, and/or suffer from difficulties expanding to a larger symbol sets, and often inadvertently impose a non-existing structure on the symbols. In this work, we instead use a probabilistic approach, modeling the symbolic sequences using a categorical distribution for each observation and try to infer not only the unknown probabilities but also the unknown indices where the distribution differs, resulting in a likelihood ratio test, which, for a given index set, is equivalent with the well studied problem of testing for independence in $2 \times J$ contingency tables, where J denotes the number of categories, see, e.g., [2]. However, if more than one statistical periodicity is considered at the same time, the number of possible combinations of index sets grows rapidly and an exact test will in many cases be computationally infeasible. By formulating the estimation of the unknown index sets, and the unknown probabilities, as a sparse logistic regression problem, we device two approximate solutions to the combinatorial problem using sparse heuristics. Namely, one greedy approach which builds up the solution by adding the sets in a sequential manner, and one using a convex relaxation of the cardinality constraint, resulting in the well-known (reweighted) LASSO problem. The resulting methods are firmly based in statistical theory, and also easily generalized to any finite symbol set.

The remainder of the paper is organized as follows: in the next section, we introduce the considered data model and show how the problem of choosing which indices that show a periodic change in the distribution can be interpreted as a sparse estimation problem. Then, in section III, we introduce a greedy algorithm that approximately solves the sparse problem, as well as a convex relaxation of the original problem, which may be efficiently solved using convex optimization algorithms. Then, in section IV, we outline some implementation issues, including a cyclic coordinate descent algorithm for solving the resulting convex relaxation problem. In section V, we examine the performance of the discussed estimators, showing the benefits of the proposed approach as compared to previously published methods. Finally, we conclude on the work in section VI.

2 Probabilistic model for symbolic sequences

Consider a symbolic sequence, $\{s_k\}_{k=1}^N$, where each symbol, s_k, is a stochastic variable drawn from a finite set, $\mathcal{A} = \{\alpha_1, \ldots, \alpha_B\}$, where B denotes the size of the alphabet. Assume that the symbols in the sequence are independent and identically distributed, such that

$$p_j \triangleq \text{Prob}(s_k = \alpha_j) \tag{1}$$

Then, if gathering a sequence of observations, $x_1, \ldots, x_N$, into the vector $\mathbf{x}$, the probability mass function (PMF) of $\mathbf{x}$ is given as

$$\begin{aligned} q_0(\mathbf{x}|\mathbf{p}) &\triangleq \text{Prob}(\mathbf{s} = \mathbf{x}) && (2)\\ &= \prod_{j=1}^{N}\prod_{\ell=1}^{B} p_\ell^{[x_j=\alpha_\ell]} = \prod_{\ell=1}^{B} p_\ell^{G_\ell} && (3) \end{aligned}$$

where $[\cdot]$ denotes the Iverson's bracket, which equals one if the statement inside the brackets is true, and zero otherwise, with each of the symbols appearing G_k times, and where $\mathbf{p}$ and $\mathbf{s}$ denote the vector of probabilities and the sequence of random variables, respectively, i.e.,

$$\begin{aligned} \mathbf{p} &= \begin{bmatrix} p_1 & \ldots & p_B \end{bmatrix}^T && (4)\\ \mathbf{s} &= \begin{bmatrix} s_1 & \ldots & s_N \end{bmatrix}^T && (5) \end{aligned}$$

with $(\cdot)^T$ denoting the transpose. As a result, the PMF is a function depending only on the number of times each symbol appears, and on the probability given

to each symbol. In general, the probabilities, p_k, are unknown and need to be estimated from the observed sequence. This can be done using the maximum likelihood (ML) estimate, formed as

$$\hat{p}_j = \frac{G_j}{N} \tag{6}$$

for $j = 1, \ldots, B$, which is an unbiased and asymptotically efficient estimate (see, e.g., [12, p. 475]). Furthermore, note that a symbol $\alpha \in \mathcal{A}$, occurring with periodicity m, i.e., with the symbol appearing at every mth index in the sequence, implies that all elements of the sequence should be equal to the symbol α in one of the m possible (disjoint) index sets

$$I(m, \ell) = \left\{ \ell, \ell + m, \ldots, \ell + \left\lfloor \frac{N - \ell}{m} \right\rfloor m \right\} \tag{7}$$

for all offsets $\ell \in \{1, \ldots, m\}$, where $\lfloor \cdot \rfloor$ denotes the rounding down operation. This means that if a periodicity m is present in a sequence, the sequence is clearly also periodic on the subharmonics i.e., for every mr:th symbol, for all natural numbers r [8]. To avoid ambiguity, we here refer to the period as the lowest possible such periodicity. Considering a sequence, $\mathbf{s}$, with a periodicity m in the symbol α, with offset n, this implies that all the symbols in the sequence at index k, will equal α, for $k \in I(m, n)$. Thus, it is a deterministic and not a statistical problem to determine if such a (deterministic) periodicity is present. However, of more interest are typically the statistical periodicities that occur in many forms of symbolic sequences, such as, e.g., DNA sequences. These are characterized by certain index sets having different distributions, such that the sequence may contain the periodicity over only a limited interval, and/or with some of the periodically occurring symbols occasionally being replaced by some other symbol, which may occur, for example, due to the presence of measurement noise, coding errors, or some, perhaps unknown, functional equivalence between symbols [3]. In such cases, the PMF for a symbolic sequence might instead be formed from two distribution, one for the indices, say I_1, corresponding to some unknown periodic index set $I(m, l)$, and another distribution for the complement index set, here denoted I_0. In this case, the PMF is

$$q_1(\mathbf{x}|\mathbf{p}_0, \mathbf{p}_1) \triangleq \prod_{j=1}^{N} \prod_{\ell=1}^{B} p_{0,\ell}^{[x_j = \mathcal{A}_\ell][j \in I_0]} p_{1,\ell}^{[x_j = \mathcal{A}_\ell][j \in I_1]}$$

$$= \prod_{\ell=1}^{B} p_{0,\ell}^{G_{0,\ell}} p_{1,\ell}^{G_{1,\ell}} \tag{8}$$

where $\mathbf{p}_0$, and similarly for $\mathbf{p}_1$, is a parameter vector containing the probabilities $p_{0,k}$, denoting the probability of a symbol, α_k, occurring in the index set I_0, and with $G_{0,k}$ and $G_{1,k}$ denoting the number of times the symbol α_k occurs in the set $I(m, n)$ and in its complement, respectively. The corresponding ML estimates are found as

$$\hat{p}_{0,j} = \frac{G_{0,j}}{|I_0|} \tag{9}$$

$$\hat{p}_{1,j} = \frac{G_{1,j}}{|I_1|} \tag{10}$$

for $j = 1, \ldots, B$, where $|S|$ denotes the cardinality of a set S, i.e., the number of elements in S. In a similar fashion, the addition of more than one periodicity can be accomplished by defining the distribution on more index sets, e.g. if one considers M disjoint index sets, $I_0, \ldots, I_{M-1}$, so that their union corresponds to the entire sequence, the PMF is

$$q_1(\mathbf{x}|\mathbf{p}_0, \ldots, \mathbf{p}_{M-1}) \triangleq \prod_{m=0}^{M-1} \prod_{k=1}^{B} p_{m,k}^{G_{m,k}} \tag{11}$$

where $G_{m,k}$ denotes the number of times the symbol α_k occurs in the set I_m. A similar model was considered in [8], although there they defined a statistical periodicity, say k, to be present when all index set $I(k, \ell)$, for $\ell = 1 \ldots, k$, have different distributions, and then set out to find the periodicity, k, by maximizing the log-likelihood using an information criteria penalty term to select the correct periodicity. If doing so, and the signal has a periodicity of k, then each index set corresponding to a different offset also has a unique distribution, implying a subdivision of the data into $\lfloor N/k \rfloor$ disjoint data sets, resulting in less data to be used to estimate these probabilities. For multiple periodicities, i.e., several index sets with different distributions, this results in a necessity to consider the overall periodicity of the sequence, i.e., if periods l and k are present, then the sequence will have a periodicity of lk, resulting in the need for substantially more data to achieve a similar performance as if only a single periodicity was present, as well as the need to perform on additional analysis to identify the factors constituting

lk. Furthermore, in the case when the sequence contains more than two periodicities, the problem quickly becomes infeasible. We instead want to find the index sets where the distributions differ as much as possible from the rest of the sequence. To that end, we recast the estimation problem in a sparse modeling framework. To do so, we note that one can interpret (11) as a multi-response logistic regression problem, which, as we will show, will be particularly useful for the case of several simultaneous periodicities. Furthermore, this mapping allows us to consider sequences one symbol at a time, which is particularly useful when the periodicity in a certain symbol is sought, or if the distribution of a particular symbol deviates especially much on a given index set. This, when applicable, decreases the variance of the estimated probabilities, thus improving the detection of periodicities only occurring in one symbol, or one subset of symbols. Rewriting (11) using logistic regression is accomplished by modeling the probability of each observation separately using a logistic function to map a linear model to the interval $[0,\ 1]$. To clarify the exposition, we first consider the case of a binary symbol set, a special case which will be shown to be particularly useful. Thus, consider a binary sequence which has a statistical periodicity on the indices I_1, and some other distribution on the indices I_0, so that the PMF may be expressed as

$$q_1(\mathbf{x}|\boldsymbol{\gamma}(\mathbf{c})) \triangleq \prod_{k=1}^{N} \gamma_k(\mathbf{c})^{x_k}(1-\gamma_k(\mathbf{c}))^{1-x_k} \tag{12}$$

where $\boldsymbol{\gamma}(\mathbf{c}) \in \mathbf{R}^N$ is a vector of probabilities, such that

$$Pr(s_k = 1) = \gamma_k(\mathbf{c}) \tag{13}$$

and the vector $\mathbf{c} \in \mathbf{R}^2$ models the probabilities for the index sets I_1 and its complement, I_0, such that

$$\boldsymbol{\gamma}(\mathbf{c}) = \begin{bmatrix} \gamma_1(\mathbf{c}) & \dots & \gamma_N(\mathbf{c}) \end{bmatrix}^T \tag{14}$$

$$\gamma_k(\mathbf{c}) = \frac{e^{\mathbf{h}_k^T\mathbf{c}}}{1+e^{\mathbf{h}_k^T\mathbf{c}}} \tag{15}$$

where

$$\mathbf{h}_k = \begin{cases} \begin{bmatrix} 1 & 1 \end{bmatrix}^T & \text{if } k \in I_1 \\ \begin{bmatrix} 1 & 0 \end{bmatrix}^T & \text{if } k \notin I_1 \end{cases} \tag{16}$$

Thus, there is a simple relationship between the parameters $p_{0,1}$ and $p_{1,1}$ in the original model in (8), i.e.,

$$P(s_k = 1) = p_{0,1} \quad \text{for } k \in I_0 \tag{17}$$
$$P(s_k = 1) = p_{1,1} \quad \text{for } k \in I_0 \tag{18}$$

and the parameter vector, $\mathbf{c}$, introduced in (12), i.e.,

$$\log\left(\frac{p_{0,1}}{1-p_{0,1}}\right) = \begin{bmatrix} 1 & 0 \end{bmatrix}^T \mathbf{c} \tag{19}$$
$$\log\left(\frac{p_{1,1}}{1-p_{1,1}}\right) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T \mathbf{c} \tag{20}$$

It should be noted that (19) implies that the probability of a symbol appearing in the set I_0 is given by the first element of the vector $\mathbf{c}$, and, similarly, one may by substituting (19) into (20) and simplifying, note that

$$\log\left(\frac{p_{1,1}}{1-p_{1,1}}\right) - \log\left(\frac{p_{0,1}}{1-p_{0,1}}\right) = \begin{bmatrix} 0 & 1 \end{bmatrix}^T \mathbf{c} \tag{21}$$

Thus, the second element in $\mathbf{h}_k$ control the change in probability on the index set, I_1, as compared to the indices in the set, I_0, e.g., if the second element is zero, then the probabilities are the same for both sets, whereas a positive or negative second element implies higher or lower probabilities on the set I_1, respectively. Extending the model to allow for the possibility of several periodicities using the logistic regression parameterization can be achieved by adding elements to the $\mathbf{c}$ vector such that each new element adjusts the probability for an additional index set. To that end, consider the case with M index sets, I_j, for $j = 1, \ldots, M$, corresponding to some specific periodicities with their different offsets, then $\mathbf{c} \in \mathbf{R}^M$ and every element of $\mathbf{h}_k^T \in \mathbf{R}^M$ is zero except the elements where k is in the corresponding index set, i.e.,

$$h_{k,j} = \begin{cases} 1 & k \in I_j \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

for $j = 1, \ldots, M$, and $d_{k,j}$ denotes element j of the vector $\mathbf{d_k}$. The resulting model can then be seen as the solution of the following optimization criterion

$$\begin{aligned} \underset{\mathbf{c}}{\text{maximize}} \quad & \prod_{k=1}^{N} \gamma_k(\mathbf{c})^{x_k}(1-\gamma_k(\mathbf{c}))^{1-x_k} \\ \text{subject to} \quad & \begin{cases} ||\mathbf{c}||_0 \leq L \\ \gamma_k(\mathbf{c}) = \frac{e^{\mathbf{h}_k^T \mathbf{c}}}{1+e^{\mathbf{h}^T \mathbf{c}}} \end{cases} \end{aligned} \tag{23}$$

where $||\cdot||_0$ denotes the ℓ_0 (pseudo) norm, which counts the number of nonzero elements of a vector, and L is the maximum number of periodicities that will be included in the model. It is worth noting that the expression for $\gamma_k(\mathbf{c})$ does not pose a restriction to the minimization, but has been included to emphasize that the probabilities for each observation are being modeled explicitly. Solving (23) for a given L, i.e., finding the maximum allowed number of simultaneous periodic sets, can be accomplished using an exhaustive search, since for each fixed k there are $(M)!/(M-k)!$ index sets. For each such set, the ML estimates may then be found using (6). However, the dimension of the parameter vector will grow quadratically with the maximum periodicity considered, since

$$M = \sum_{k=1}^{m_{max}} k = \frac{m_{max}(m_{max}+1)}{2} \tag{24}$$

where m_{max} is the maximum allowed periodicity, since each period k has k corresponding index sets, one for each possible offset. Thus, to evaluate the likelihood for all combinations of index sets will soon lead to a computationally infeasible problem. Generalization to larger symbol sets may be carried out in a similar manner, leading to the multi-response logistic regression model (see, e.g., [2] for a further discussion on multi-response logistic regression). The corresponding optimization problem is therefore given as the maximum of the log-likelihood with a cardinality constraint [13]

$$\begin{aligned} \underset{\mathbf{c}_1,\ldots,\mathbf{c}_B}{\text{maximize}} \quad & \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{\ell=1}^{B} x_{i\ell}(\mathbf{h}_i^T\mathbf{c}_\ell) - \log\left(\sum_{\ell=1}^{B} e^{\mathbf{h}_i^T\mathbf{c}_\ell}\right)\right] \\ \text{subject to} \quad & ||\mathbf{C}_{k\cdot}||_0 \leq L, \quad \text{for } k = 1, \ldots, R \end{aligned} \tag{25}$$

where $\mathbf{C}$ is a matrix constructed such that its k:th column is formed by the vector $\mathbf{c}_k$, and R is the number of considered index sets, with $\mathbf{C}_{k\cdot}$ denoting the restriction

that $||\mathbf{C}_{k\cdot}||_0$ forces the solution to adjust the B parameters corresponding to every index set simultaneously. Thus, the distributions can be changed on at most L index sets. As a result, the framework allows for flexibility in what is deemed a periodicity, e.g., one might test for a high probability of a certain symbol appearing, or even for if some symbols appear with low probability. Both of these ideas will be explored further in the following, where we outline a couple of possible algorithms for estimating periodicities for some commonly occurring situations, namely, estimation of an unknown periodicity, detection of an unknown periodicity, and, finally, estimation of multiple periodicities.

3 Relaxation of the cardinality constraint

For cardinality constrained, or sparse, least squares problems, there are a wide range of tools for forming approximate solutions, with many methods falling into two broad categories, namely greedy methods that build up a solution one variable at a time until either fitting criterion is satisfied, or the number of variables reaches the constraint, or methods that replace the cardinality constraint with a penalty function that promotes solutions that have few non-zero variables [14]. This implies that the optimization can be carried out without the combinatorial computation complexity inherent in cardinality constrained optimization problems. Typically, the penalty function is selected as the ℓ_1 norm, leading to a simple convex optimization problem. In the following two subsections, we propose both kinds of algorithms, first a greedy approach and then an iterative convex relaxation.

3.1 Greedy approach

In order to form a greedy estimate of the minimization in (25), one may note the analogy between this formulation and that of simple hypothesis test for testing if a distribution is different on some index sets (see also [3]). Thus, one may form a test to determine the hypothesis that a given sequence has a different distribution for the indices corresponding to $I(m, \ell)$, i.e., that the PMF is formed using (8), against the null hypothesis that the entire sequence has the same categorical distribution, such that the PMF instead follows (3), i.e.,

$$\mathrm{H}_0 : \mathbf{p}_0 = \mathbf{p}_1 \tag{26}$$

$$\mathrm{H}_1 : \mathbf{p}_0 \neq \mathbf{p}_1 \tag{27}$$

Such a test may be formed as a likelihood ratio (LR) test (see, e.g., [15, p. 375])

$$\lambda_{m,\ell}(\mathbf{x}_N) = \frac{q_0(\mathbf{x}_N|\mathbf{p}_0, \mathrm{H}_0)}{q_1(\mathbf{x}|\mathbf{p}_0, \mathbf{p}_1, \mathrm{H}_1)} \tag{28}$$

where the probabilities are determined using (6) under H_0, and using (9) and (10) under H_1. Thus, if one only seek to find a single index set, a suitable choice would be the one maximizing the LR, i.e.,

$$\arg\max_{m,\ell,i} \quad \lambda_{m,\ell}(f_i(\mathbf{x}_N)) \tag{29}$$

If the number of periodicities is unknown, i.e., the problem is one of detection and not estimation, one can allow for the possibility of no set being added by considering that if H_0 is true, it holds asymptotically that [15, p. 489]

$$-2\log(\lambda_{m,\ell}(\mathbf{x}_N)) \xrightarrow{d} \chi^2_{B-1} \tag{30}$$

where $\xrightarrow{d}$ denotes convergence in distribution and χ^2_k denotes the chi-squared distribution with k degrees of freedom. Thus, if no periodicity is present, a critical value, denoted T_α, for the likelihood ratio, below which no periodicity is deemed to be present, can be constructed for the likelihood ratio for each of the tests. Since M tests are formed in order to compute (29), and if assuming that these are independent, the critical value may be well approximated using extreme value theory as a quantile of the random variable

$$\psi = max\,(z_1, \ldots, z_M) \tag{31}$$

where each z_k is χ^2 distributed, implying that ψ will follow a Gumbel distribution (see, e.g., [16, p. 156]). In the case when multiple periodicities may be present, one can extend this procedure using a step-wise approach. To do so, first define I_1 as the index set containing all the indices in the sequence. Then, the initial step is performed by using the above algorithm to determine an index set $I_2 = I_{m_1,\ell_1}$, where m_1 and ℓ_1 denote the initially estimated periodicity and offset, respectively, found in the maximization of (29). In order to determine the next periodicity, the H_0 distribution is formed from (11), using one distribution for the found index set I_2 and one for all the other indices, $I_1 \setminus I_2$, where $\setminus$ denotes set subtraction operation. The second phase, m_2, and periodicity, ℓ_2, may be determined using

(29). This procedure can then be repeated until the zero hypothesis can not be rejected using a suitable quantile of (31), i.e., at iteration s the corresponding likelihood ratio test may be formed as

$$\lambda^{(s)}_{m,\ell}(\mathbf{x}_N) = \frac{q_0(\mathbf{x}_N|\mathbf{p}_0, \ldots, \mathbf{p}_{s-1}, \mathrm{H}_0)}{q_1(\mathbf{x}|\mathbf{p}_0, \ldots, \mathbf{p}_s, \mathrm{H}_1)} \tag{32}$$

Note that this assumes that the sets I_k being added to the zero hypothesis are disjoint, otherwise the likelihood would include some data points more than once. To ensure this we propose to only consider the indices that have not all ready been added to H_0 when evaluating $q_1(\mathbf{x}|\mathbf{p}_0, \mathbf{p}_1, \mathrm{H}_1)$ in (28), i.e., at iteration k the sets $I(m, \ell)$ are replaced with $I(m, l) \leftarrow I(m, l) \setminus I_{k-1}$, for all m and ℓ, where $\leftarrow$ denotes that the quantity on the left is replaced with the one on the right. The resulting greedy algorithm, here termed the greedy *P*eriodicity *E*stimation of *C*ategorical *S*equences (PECS$_G$) estimator, is outlined in Algorithm 1 below, with each step in the iteration requiring about $m_{max}N$ operations.

3.2 Iterative convex relaxation

It is worth noting that the optimization criterion in (25) is not convex as it restricts the parameter space to lie in a non-convex set. A commonly used relaxation for problems of this kind is to replace the ℓ_0 restriction with the convex ℓ_1 ball, which by taking the negative logarithm and using the Lagrange duality, results in the relaxed convex optimization criterion

$$\underset{\mathbf{c}}{\text{minimize}} \quad \sum_{k=1}^{N} -x_k \mathbf{h}_k^T \mathbf{c} + \log(1 + e^{\mathbf{h}_k^T \mathbf{c}}) + \lambda ||\mathbf{c}||_1 \tag{33}$$

where we have exploited the equality constraint for $p_k(\mathbf{c})$. Some adjustments may be done to this criterion; firstly, the penalty on $\mathbf{c}$ includes the first element. This is not appropriate since the first element controls the probability for all observations, and we have no reason to want to bias that probability towards $1/2$. This is easily accomplished by only penalizing the other elements of the vector, i.e., replacing $||\mathbf{c}||_1$ with $||\underline{\mathbf{c}}||_1$, where $\underline{\mathbf{c}}$ denotes the resulting vector once the first element of $\mathbf{c}$ is removed. However, the resulting expression will also have an undesirable ambiguity due to the lack of distinction being made between if the probability is higher or lower on the periodic indices. For instance, consider a case when every third index starting with 1 has the probability 0.1 of being 1, and all other indices have

Algorithm 1 The PECS_G estimator

```
Given a categorical sequence, x of length N
I_0 = {1, ..., N}
for s = 1 to max_iteration do
    {m_s, ℓ_s} = arg max_{m,ℓ} λ_{m,ℓ}(x_N)
    if λ_{m,ℓ}(x_N) > C_α then
        I_s = I_{m_s,ℓ_s}
    else
        break
    end if
    I(m, l) ← I(m, l) \ I_s for all m and l
    I_0 ← I_0 \ I_s
    H_0 distribution is replaced with (11) using I_0, ..., I_s
end for
```

probability 0.9 of being 1. Should this be considered two periodicities of 3 with probability 0.9, or one periodicity of 3 with probability 0.1? Such a distinction is of course not a problem specific for this model. However, since one is commonly interested in finding periodic indices where the probability is either higher or lower, such an ambiguous result would result in a non-consistent interpretation of the estimates. Fortunately, this can be easily handled by adding a constraint on $\underline{\mathbf{c}}$, ensuring that only periodicities with greater probability of a symbol appearing are considered, i.e., $c_k > 0$, for $k = 2, \ldots, M$, where c_i is the i:th element of the vector $\mathbf{c}$. This yields

$$\begin{aligned} \underset{\mathbf{c}}{\text{minimize}} \quad & \sum_{k=1}^{N} -x_k \mathbf{h}_k^T \mathbf{c} + \log(1 + e^{\mathbf{h}_k^T \mathbf{c}}) + \lambda ||\underline{\mathbf{c}}||_1 \\ \text{subject to} \quad & c_k \geq 0 \quad \text{for } k = 2, \ldots, M \end{aligned} \tag{34}$$

The resulting optimization is thus a sum of an affine function and the logarithm of a sum of exponential functions, and is thus a convex function. (see, e.g., [17, p. 93]). Thus, since the constraints can be seen as inequalities involving inner products with the Cartesian coordinate basis vectors, they are affine, and therefore convex functions, and the criterion is as a result a convex optimization problem in the standard form, as defined in [17, p. 136]. However, the cri-

terion in (34) will not yield sufficiently sparse estimates, as a result of the rather coarse approximation of the ℓ_1 norm to the desired ℓ_0 norm. Recently, interest in non-convex penalties that are closer, in some sense, to the ℓ_0 norm have been suggested, such as the use of the ℓ_q norm, for $0 < q < 1$ (see e.g., [18, 19]). Herein, we consider an alternative approach where the ℓ_1 penalty is replaced with the concave $\log(\cdot)$ penalty. The resulting optimization is then solved with an iteratively reweighted ℓ_1 minimization, using a technique suggested in [20]. The resulting algorithm thus solves, at iteration $j+1$, the minimization

$$\begin{aligned} \min_{\mathbf{c}} \quad & \sum_{k=1}^{N} -x_k \mathbf{h}_k^T \mathbf{c} + x_k \log(1 + e^{\mathbf{h}_k^T \mathbf{c}}) + \lambda \sum_{k=1}^{M} \frac{|c_k|}{|\hat{c}_k^{(j)}| + \varepsilon} \\ \text{s. t.} \quad & c_k \geq 0 \quad \text{for } k = 2, \ldots, M \end{aligned} \tag{35}$$

where $\hat{c}_k^{(j)}$ is the k:th element of the $\mathbf{c}$ estimate resulting from the j:th iteration, and ε is chosen as a small number to avoid numerical problems as well as to enable zero valued elements of $\mathbf{c}$ to transition from zero to non-zero values (see also [20]). The resulting sequence of convex minimizations yields a sufficiently sparse estimate of the periodicities (although at a high a computational complexity if implemented directly using a standard interior point-based solver). The resulting estimator is in the following referred to as the *P*eriodicity *E*stimation of *C*ategorical *S*equences using Logistic regression, PECS_L.

Comparing the two methods, PECS_G offers a faster solution, whereas PECS_L yields better results in the case of multiple periodicities. This is due to the fact that the iterative greedy procedure in PECS_G does not take into account the overlap between the two index sets, e.g., the index sets $I(k,1) \cap I(l,1) = I(kl,1)$, whereas, the logistic regression approach also takes the overlap into account in the estimation procedure.

4 An efficient implementation

In order to form an efficient solver for the minimization in (35), we proceed to develop a cyclic coordinate descent (CCD) algorithm. The CCD algorithm minimize the cost function in (35) one variable at a time, in a cyclical fashion, holding the other variables fixed at their most recent estimates. This will thus transform the $M-$dimensional optimization problem into a scheme where one instead repeatedly solves simpler one-dimensional problems.

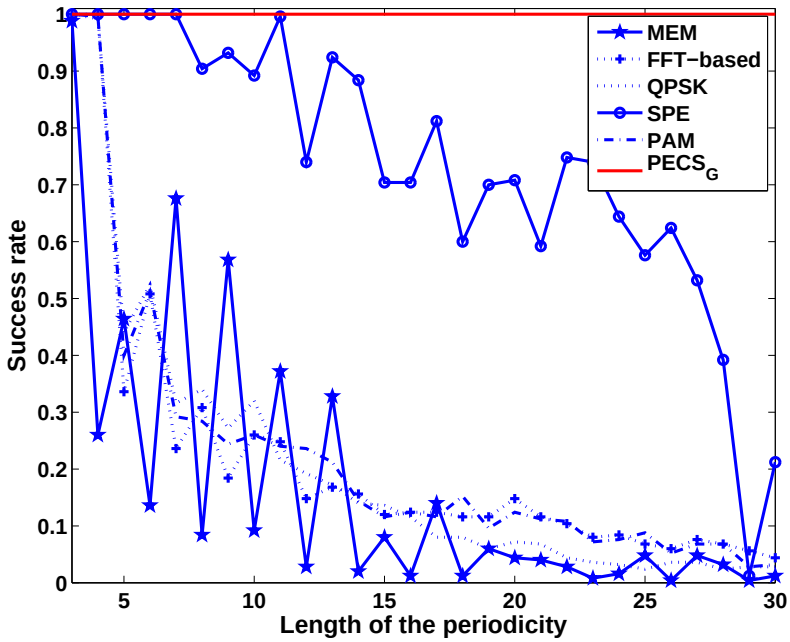

Figure 1: Rate of success in estimating deterministic periods.

It should be noted that such an approach is, in general, converging notoriously slowly, or in some cases, not at all. However, for the optimization problems often ecountered in sparse modeling, this does no longer hold, as in fact, convergence proofs exist [21, 22] and in many applications, CCD implementations have empirically been shown to be the fastest algorithm available [13, 23]. Below, we outline the steps involved in a CCD algorithm for the case of $c_k \geq 0$, with the other case being handled in a similar manner. Thus, consider $c_i^{(r)}$ as the r:th estimate of element i of the vector $\mathbf{c}$, then, for $i > 2$,

$$c_i^{(r+1)} = \arg\min_{c_i} \sum_{k=1}^{N} -x_k \mathbf{h}_k^T \mathbf{c} + \log(1 + e^{\mathbf{h}_k^T \mathbf{c}}) + \lambda ||\underline{\mathbf{c}}||_1$$

Algorithm 2 The PECS$_L$ estimator

```
Initiate c = c_0
for r = 1, ... do
    for i = 1, ..., M do
        if maximum of (41) ≥ 0 then
            c_i^(r) = 0
        else
            Update c_i^(r) according to (36)
        end if
    end for
end for
```

$$= \arg\min_{c_i} -\mathbf{x}^T \mathbf{H}_{(\cdot,i)} c_i + \lambda |c_i| + \sum_{k=1}^{N} \log(1 + a_{k,i} e^{h_{k,i} c_i}) \tag{36}$$

The notation $\mathbf{H}_{(\cdot,i)}$ denotes the i:th column of the matrix $\mathbf{H}$, $h_{k,i}$ the i:th element of the vector $\mathbf{h}_k$, and

$$\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix} \tag{37}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \dots & \mathbf{h}_N \end{bmatrix}^T \tag{38}$$

$$\mathbf{c} = \begin{bmatrix} c_1^{(r+1)} & \dots & c_{(i-1)}^{(r+1)} & c_i^{(r)} \dots & c_N^{(r+1)} \end{bmatrix}^T \tag{39}$$

$$a_{k,i} = \exp\left(\sum_{j, j \neq i} h_{k,j} c_j \right) \tag{40}$$

If the maximum value of the subdifferential set

$$\partial f_0 = -\mathbf{x}^T \mathbf{H}_{(\cdot,i)} + \lambda w + \sum_{k=1}^{N} \frac{a_{k,i} h_{k,i} e^{h_{k,i} c_i}}{1 + a_{k,i} e^{h_{k,i} c_i}} \tag{41}$$

with $c_i = 0$ is positive and $\{w \in [-1, 1]\}$, then the optimum is attained at $c_i = 0$ for the constrained optimization problem. On the other hand, if the maximum is negative, the stationary point may be found using a gradient approach (since the cost function is differentiable for all positive c_i). Note that this analysis gives

insight into both the sparsity promoting effect of the ℓ_1 norm as well as the role of the tuning parameter λ, in fact, rewriting (41) as

$$\partial f_0 = -\mathbf{x}^T \mathbf{H}_{(\cdot,i)} + \lambda w + \mathbf{r}_i^T \mathbf{H}_{(\cdot,i)} \tag{42}$$

where $\mathbf{r}_i = \begin{bmatrix} \frac{a_{1,i}}{1+a_{1,i}} & \cdots & \frac{a_{N,i}}{1+a_{N,i}} \end{bmatrix}$ can be interpreted as probabilities for each index. Furthermore, $\mathbf{r}_i^T \mathbf{H}_{(\cdot,i)}$ is the expected number of symbols on the periodicity corresponding to i and $\mathbf{x}^T \mathbf{H}_{(\cdot,i)}$ is the observed number of symbols on that periodicity, thus if

$$|\mathbf{r}_i^T \mathbf{H}_{(\cdot,i)} - \mathbf{x}^T \mathbf{H}_{(\cdot,i)}| < \lambda \tag{43}$$

implying that, if the expectation for the model with $c_i = 0$ is closer than λ to the observed number in the data, then set $c_i^{(i+1)=0}$. The resulting CCD algorithm is outlined in Algorithm 2.

5 Numerical results

We proceed to examine the performance of the proposed likelihood-based estimators using simulated DNA sequences, binary sequences, and measured DNA data. For DNA sequences, only $B = 4$ different symbols are present, namely A, C, G, and T. Initially, we examine a simulated DNA sequence containing one deterministic periodicity. Figure 1 illustrates the rate of successfully determining this periodicity as a function of the length of the periodicity, comparing the proposed PECS_G estimator with the MEM [10], PAM [7], QSPK [5], and SPE [24] estimators, as well as with a Fourier-based estimator detailed in [24]. Here, and in the following, the success rate has been determined using 250 Monte-Carlo simulations using $N = 1000$ equiprobable symbols, with the sought periodicity being inserted appropriately. As is clear from Figure 1, the proposed estimator succeeds in successfully determining all the considered periodicities, whereas all the other methods lose performance as the length of the periodicity grows. Of the other examined estimators, the SPE estimator seems to offer the second best performance, and we will for this reason only show the results for this estimator in the following comparisons, noting that all the other discussed estimators exhibits a notably worse performance than the SPE estimator in all the considered cases (see also [1]). Proceeding to examine also statistical periodicities, we vary p_1 for the index set corresponding to the generated periodicity, with $p_0 = 1/4$

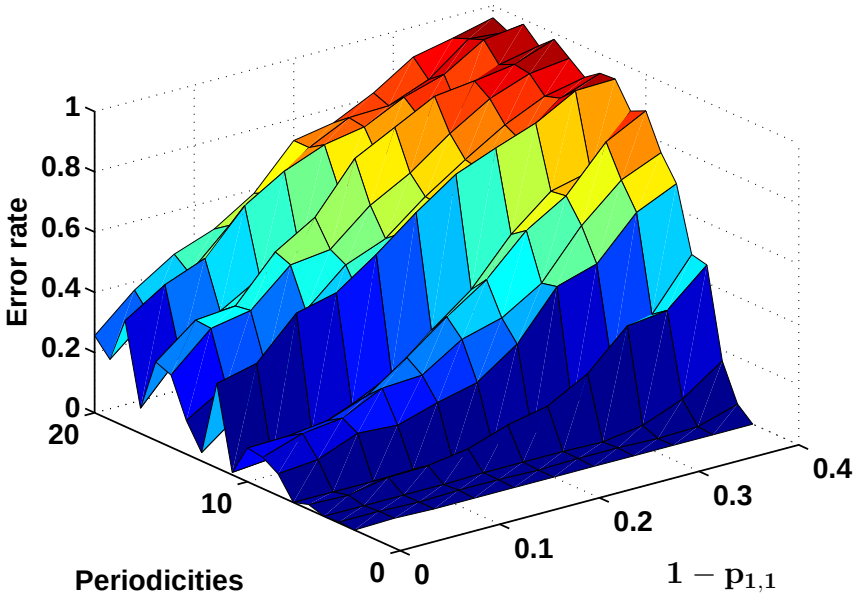

Figure 2: The error rate of finding the periodicity as a function of the negative probability, $1 - p_{1,1} =$, and the periodicity for the SPE algorithm.

on the complement set. Figures 2 and 3 show the resulting success rate for the SPE and PECS$_G$ estimators as a function of the periodicity and the probability p_1, again clearly illustrating how PECS$_G$ outperform SPE (and thus also all the other mentioned estimators) for all periodicities and p_1.

Next, we investigate how well PECS$_G$ and PECS$_L$ are able to resolves two periodicities in a binary sequence. In this case, some care needs to be taken when setting up the simulations, as when generating two periodicities, these may overlap or combine to create a new periodicity, e.g., if generating two periodicities of period six, these may be placed such that they instead form just a single periodicity with period three. Similarly, two periodicities with period four and twelve may cause the resulting sequence to have only a single periodicity of four. In order to avoid such ambiguities in the resulting performance measure, the test data has

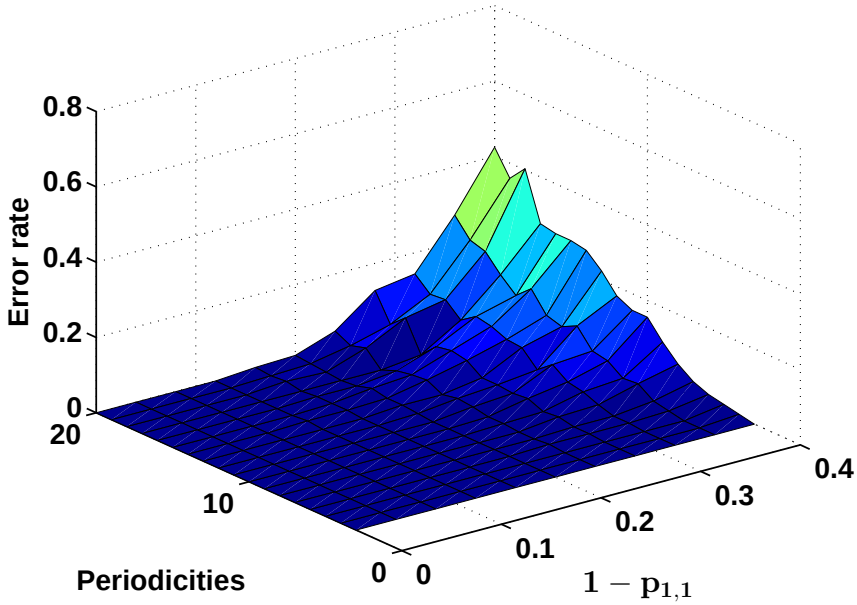

Figure 3: The error rate of finding the periodicity as a function of $1 - p_{1,1}$, and the periodicity for the proposed PECS$_G$ method.

been generated such that it avoids this form of ambiguities. Figure 4 illustrates the success rate of determining both periodicities correctly, as a function of the length of the two periodicities, with $N = 500$ and again using $p_1 = 3/4$ and $p_0 = 1/4$. Each point on the x-axis should be interpreted as the average error for all combinations of periodicities within the brackets, i.e., for instance $(14, 14 - 17)$ denotes all combinations $(14, 14), (14, 15), (14, 16)$ and $(14, 17)$. As may be seen from the figure, even when the sequence contains two periodicities of lengths up to 12, when most of the other discussed estimators completely fail to find even a single perfect periodicity, both PECS algorithms have a very low proportion of errors. From the figure, one can also observe that, as expected, the PECS$_L$ outperforms the PECS$_G$ when there is more than one periodicity present in the sequence. For the last simulated data experiment, we recreate a simulation experiment similar to

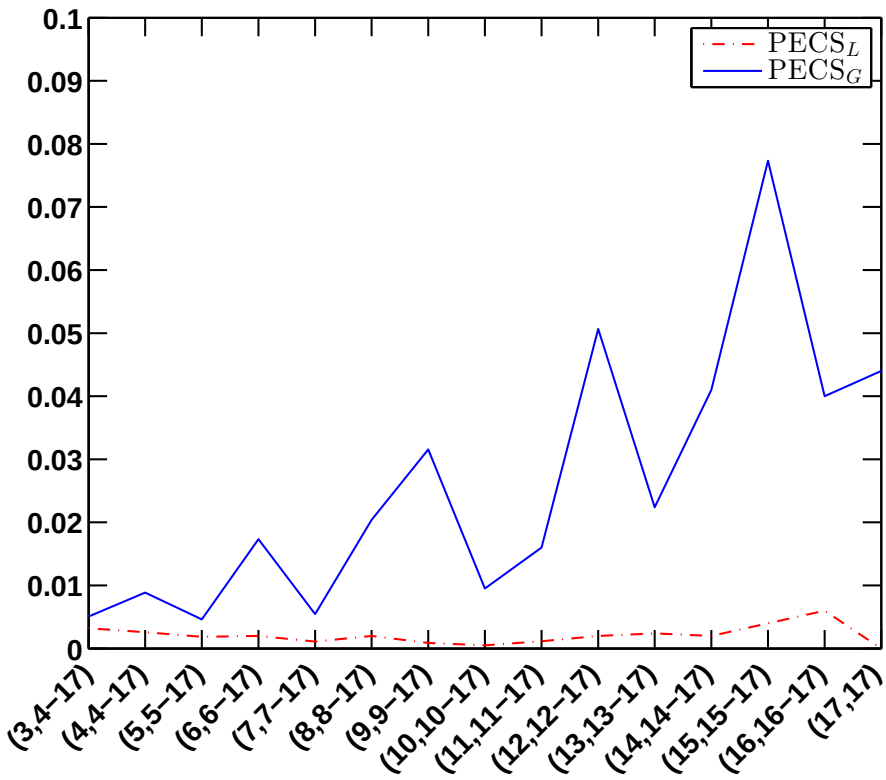

Figure 4: The proportion of incorrect estimations of two periodicities for the PECS algorithms. Each point on the x-axis represent average error for all combination of that point and smaller (or equal) periodicities.

the one that was used in [8], where a deterministic periodicity of 11 and 31 are present simultaineously in a signal generated from a 4 element set being uniformly distributed on the other indices. As can be seen in Figure 5, the PECS$_G$ estimator achieves almost 100 % success rate even before the method presented in [8] can start to be used, since it requires a minimum of $11 \times 31 = 341$ data points. Finally, we examine the performance of the PECS$_G$ estimator on measured genomic data, in the form of the gene C. elegans F56F11.4 [25]. Since genomic data is generally not stationary, the estimate has been formed using a sliding window

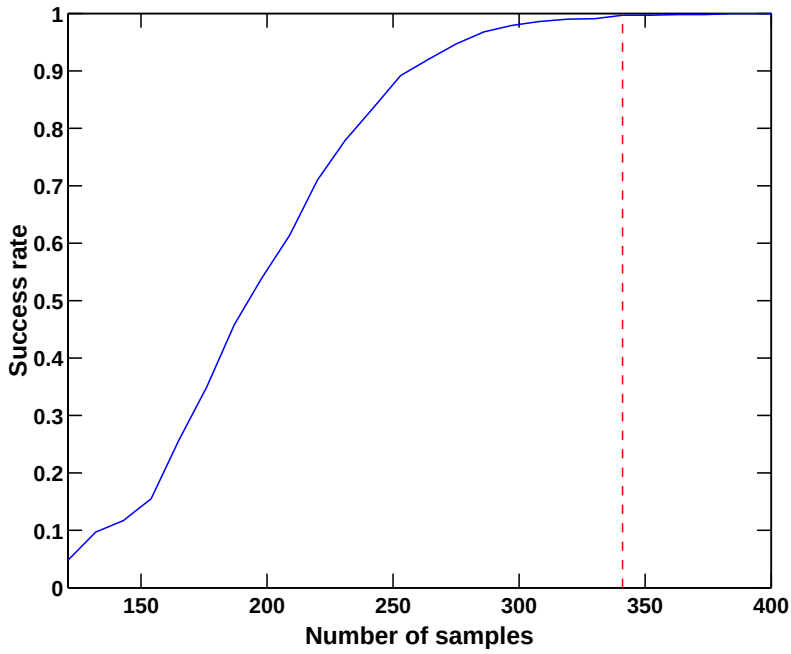

Figure 5: Rate of success for PECS$_G$ in estimating the periodicites of a signal with periodicites at 11 and 31, as a function of signal length. The dashed line denotes the minimum data needed for using [8].

with length $N = 360$. The results obtained by PECS$_G$ are shown in Figure 6, where the periodicities with a likelihood ratio greater than the 95% quantile of the maximum of $M = 465$ χ^2 distributed random variables are shown for each symbol. In earlier work, such as [10] and [24], a period of three was found at around index 7000. This period was also found when using PECS$_G$, but when looking at the corresponding $\tilde{p}$, one may note that this periodicity is actually constituted by the lack of the symbol C, i.e., this period is detected since the symbols A, G, and T are alternating in a non-periodic fashion, and since C is always absent at these indices, this apparently causes the Fourier based methods to indicate a periodicity of three. If one is not interested in finding these sorts of periodicities, one

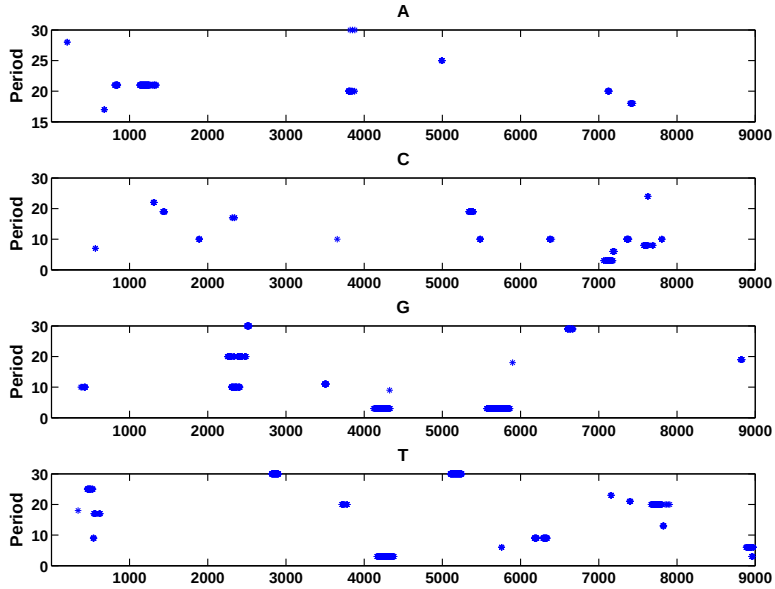

Figure 6: The periodicities of each symbol in the gene C.elegans F56F11.4 computed using a sliding window.

may restrict p_1 to be in $[1/2, 1]$, in the same manner as mentioned above. This will ensure that PECS_G only finds periodicities that are made up by an increased probability in the presence of a symbol.

6 Conclusion

In this work, we have presented a likelihood-based approach for modeling periodicities in symbolic sequences. Modeling the observations using a categorical distribution with periodic indices, possibly having a different distribution, leads to a difficult combinatorial problem. Here, we have proposed two algorithms to relax the problem using sparse heuristics: namely, one fast greedy approach which builds up the solution set in an iterative fashion, and one based on convex relaxa-

tion ideas, which has the benefit of a more efficient usage of the data. Finally, we show the benefits of the proposed algorithms as compared to previously published methods using simulation experiments as well as with real DNA data examples.

Acknowledgement

The authors would like to thank Prof. Lorenzo Galleani and Dr. Roberto Garello at Politecnico di Torino, Italy, for providing us with the their implementation of MEM-algorithm detailed in [10].

References

[1] S. I. Adalbjörnsson, J. Swärd, and A. Jakobsson, "'Likelihood-based Estimation of Periodicities in Symbolic Sequences," in *Proceedings of the 21th European Signal Processing Conference*, Marrakesh, 2013.

[2] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, second edition, 2007.

[3] M. B. Chaley, E. V. Korotkov, and K. G. Skryabin, "Method Revealing Latent Periodicity of the Nucleotide Sequences Modified for a Case of Small Samples," *DNA Res.*, vol. 6, no. 3, pp. 153–163, 1999.

[4] E. Korotkov and N. Kudryaschov, "Latent periodicity of many genes," *Genome Informatics*, vol. 12, pp. 437–439, 2001.

[5] D. Anastassiou, "Genomic Signal Processing," *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 8–20, July 2001.

[6] W. Wang and D. H. Johnson, "Computing linear transforms of symbolic signals," vol. 50, no. 3, pp. 628–634, March 2002.

[7] G. L. Rosen, *Signal Processing for Biologically-Inspired Gradient Source Localization and DNA Sequence Analysis,*, Ph.D. thesis, Georgia Institute of Technology, 2006.

[8] R. Arora, W. A. Sethares, and J. A. Bucklew, "Latent Periodicities in Genome Sequences," *IEEE J. Sel. Topics in Signal Processing*, vol. 2, no. 3, pp. 332–342, June 2008.

[9] L. Wang and D. Schonfeld, "Mapping Equivalence for Symbolic Sequences: Theory and Applications," vol. 57, no. 12, pp. 4895–4905, Dec. 2009.

[10] L. Galleani and R. Garello, "The Minimum Entropy Mapping Spectrum of a DNA Sequence," vol. 56, no. 2, pp. 771–783, Feb. 2010.

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

[12] E. L. Lehmann and G. Casella, *Theory of Point Estimation (Springer Texts in Statistics)*, Springer, 2nd edition, 1998.

[13] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[14] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.

[15] G. Casella and R. Berger, *Statistical Inference*, Duxbury, 2nd edition, 2002.

[16] P. Emberchts, C. Klüppelberg, and T. Mikosch, "Fluctuations of Maxima," in *Modelling Extremal Events*, vol. 33 of *Applications of Mathematics*, pp. 113–179. Springer Berlin Heidelberg, 1997.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[18] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," vol. 14, no. 10, pp. 707–710, Oct. 2007.

[19] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, 2010.

[20] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[21] P. Tseng, "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization," *Journal of Optimization Theory and Applications*, vol. 109, no. 3, pp. 475–494, 2001.

[22] P. G. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*, Springer Series in Statistics. Springer, 2011.

[23] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise Coordinate Optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.

[24] J. Swärd and A. Jakobsson, "Subspace-based estimation of symbolic periodicities," in *Intern. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26-31 2013.

[25] National Center for Biotechnology Information, "Genome sequence of the nematode C. elegans: a platform for investigating biology," http://www.ncbi.nlm.nih.gov/nuccore/FO081497.1.

D

Paper D

High resolution sparse estimation of exponentially decaying N-D signals

Johan Swärd, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson

Centre for Mathematical Sciences, Lund University, Lund, Sweden

Abstract

In this work, we consider the problem of high-resolution estimation of the parameters detailing an N-dimensional (N-D) signal consisting of an unknown number of exponentially decaying sinusoidal components. Since such signals are not sparse in an oversampled Fourier matrix, earlier approaches typically exploit large dictionary matrices that include not only a finely spaced frequency grid, but also a grid over the considered damping factors. Even in the 2-D case, the resulting dictionary is typically very large, resulting in a computationally cumbersome optimization problem. Here, we introduce a sparse modeling framework for N-dimensional exponentially damped sinusoids using the Kronecker structure inherit in the model, as well as introduce a novel dictionary learning approach that iteratively refines the estimate of the candidate frequency and damping coefficients for each component, thus allowing for smaller dictionaries, and for frequency and damping parameter that are not restricted to a grid. The performance of the proposed method is illustrated using simulated data, clearly showing the improved performance as compared to previous techniques.

Key words: Sparse signal modeling, Spectral analysis, sparse reconstruction, parameter estimation, dictionary learning, damped sinusoids.

1 Introduction

High-dimensional decaying sinusoidal signals occur in a wide variety of fields, such as spectroscopy, geology, sonar, and radar, and given the importance of such signals in a variety of applications, the topic has attracted notable attention in the recent literature (see, e.g. [1–11]). Common solutions include subspace-based algorithms [3–8], which are typically making relatively strong model assumptions, or the use of high-dimensional representations necessitating an iterative zooming procedure over multiple dimensions, such as the technique introduced in [11]. These kind of approaches often suffer from high complexity and sub-optimal performance, typically requiring an accurate initialization or model order information to yield reliable results, information which is commonly not available in many of the discussed applications. Often, the measurements are also assumed to be uniformly sampled, which may well be undesired in applications such as, for instance, spectroscopy. Furthermore, the number of modes present in the signal is generally unknown, or may vary over time, typically necessitating some form of model order selection decision. Given such difficulties, it is often of interest to formulate non-parametric or semi-parametric modeling techniques, imposing only mild assumptions of the *a priori* knowledge of the signal structure. Popular solutions include the so-called dCapon, dAPES, and dIAA spectral estimators, which all form generalized spectral estimates of the signal, constructing spectral representations over both the frequency and damping dimensions [12, 13] (see also [14, 15]). Although this form of techniques are robust to the made model order assumptions, they suffer difficulties in separating closely spaced modes from each other, and typically require notable computational efforts if not implemented carefully [15]. As an alternative, one may use sparse modeling of the signal, forming a large dictionary of all potential frequencies and damping candidates, thus generally having vastly more columns than rows. For a given signal and the resulting dictionary matrix, one thus wishes to find the sparsest solution to the resulting linear set of equations, mapping the signal to a linear combination of a few of the columns of the dictionary. Such techniques have successfully been applied to line spectral data, and the topic has attracted notable attention in the recent literature (see, e.g., [16–22]). Although these algorithms appear quite different from each other, they share the property that the considered dictionary grid should be selected sufficiently fine to allow for a sparse signal representation (see also [23, 24]), which, if extended to also consider damped modes, necessitates a large dictionary matrix containing elements with a sufficiently fine grid over the range of both

the potential frequencies and damping candidates (see, e.g., [13, 25, 26]); this will be particularly noticeable if treating large data sets, or data sets with multiple measurement dimensions. In order to mitigate this problem, we here introduce a tensor representation of the signal model, allowing us to exploit the resulting inherent Kronecker structure, which may be exploited to significantly reduce the required complexity as compared to a naive implementation of the sparse modeling framework. Furthermore, we propose a novel dictionary learning approach, wherein one iteratively decomposes the signal with a fixed small dictionary, adaptively learning the dictionary elements best suited to enhance sparsity. To this effect, we initially form a coarsely spaced dictionary with undamped modes over the range of considered frequency candidates, iteratively adapting both the frequency and damping settings for the dictionary elements, thereby also allowing for both a reduction and an expansion of the number of dictionary elements considered in each iteration of the optimization. In order to further reduce complexity, we propose a computationally efficient implementation based on the concept of the alternating direction method of multipliers (ADMM) (see, e.g., [27]), where the Kronecker structure of the resulting dictionary matrices may be exploited to dramatically decrease the cost of each iteration.

The remainder of the paper is organized as follows: in the next section, we introduce the considered data model. Then, in Section 3, we introduce the idea behind decoupling the search dimensions. Section 4 introduces the ADMM formulation of the estimator, and Section 5 illustrates the performance of the proposed estimator using simulated data. Finally, Section 6 contains our conclusions.

In the remainder of the paper, we use the following notation: scalars are represented using lower case letters, whereas vectors are represented with lower case bold-face letters. Matrices are represented with capital bold-face letters, tensors with capital bold Euler script letter, $(\cdot)^T$ denotes the transpose, and $(\cdot)^H$ the conjugate transpose.

2 The N-D signal model

Consider an N-dimensional signal consisting of a sum of K modes, i.e., K N-dimensional damped sinusoids such that observation $x_{\boldsymbol{\tau}}$ at a sampling point τ, where

$$\boldsymbol{\tau} = \begin{bmatrix} t_{i_1}^{(1)} & t_{i_2}^{(2)} & \dots & t_{i_N}^{(N)} \end{bmatrix}^T \tag{1}$$

and $t_{i_\ell}^{(\ell)}$ denotes the i_ℓ:th sampling point in dimension ℓ, may be well modeled as

$$x_{\boldsymbol{\tau}} = \sum_{k=1}^{K} g_k \prod_{\ell=1}^{N} \xi_{k,\ell}^{t_{i_\ell}^{(\ell)}} + \varepsilon_{\boldsymbol{\tau}} \tag{2}$$

where

$$\xi_{k,\ell} = e^{j\omega_k^{(\ell)} - \beta_k^{(\ell)}} \tag{3}$$

and with g_k denoting the complex amplitude of mode k, and $\varepsilon_{\boldsymbol{\tau}}$ is an additive noise term, here for simplicity assumed to be an independent identically distributed circularly symmetric Gaussian random variable. Assuming the signal is observed over $t_{i_n}^{(n)}$, for $i_n = 1, \ldots, I_n$, and $n = 1, \ldots, N$, the entire sequence may be stored in an N-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. It is worth noting that this formulation makes no restriction on any of the dimensions to have a sampling scheme that is equidistant, thus encompassing both missing data scenarios as well as irregular sampling. The entire model may thus be written in tensor format as the sum of K rank one tensors, such that

$$\boldsymbol{\mathcal{X}} = \sum_{k=1}^{K} g_k \tilde{\mathbf{a}}^{(1)}(k) \circ \tilde{\mathbf{a}}^{(2)}(k) \cdots \circ \tilde{\mathbf{a}}^{(N)}(k) + \boldsymbol{\mathcal{E}} \tag{4}$$

where $\circ$ denotes the outer product, $\boldsymbol{\mathcal{E}}$ is the tensor containing the noise terms, and

$$\tilde{\mathbf{a}}^{(n)}(k) = \begin{bmatrix} \xi_{k,n}^{t_1^{(n)}} & \ldots & \xi_{k,n}^{t_{I_n}^{(n)}} \end{bmatrix}^T \tag{5}$$

For an overview of tensor algebra sufficient for the here discussed results see, e.g., [28], which also use a notation consistent with the one used in this article. The model thus contain $(2N+1)K + 1$ unknown parameters, namely

$$\boldsymbol{\vartheta} \triangleq \begin{bmatrix} \{\{\omega_k^{(n)}, \beta_k^{(n)}\}_{n=1}^{N}, g_k\}_{k=1}^{K}, & K \end{bmatrix}^T \tag{6}$$

of which $2NK$ are non-linear parameters. Clearly, one could, in theory, form a non-linear least squares (LS) minimization over these parameters, as well as form a model order estimate from the resulting model order residuals for varying possible candidate model sizes. However, such a solution would in most practical

situations be computationally unfeasible, even for low dimensional data sets, especially as the optimization is well known to have numerous local minima [29]. To avoid this, we introduce a sparse modeling heuristic to approximate the model. This can be done by creating a large dictionary of candidate parameters, selected from a grid fine enough such that each true parameter lies sufficiently close to some grid point. For instance, if, to simplify our notation, one fixates all but the first frequency and damping coefficients, one may approximate (4) using a dictionary containing P_1 and J_1 candidate elements along the (first) frequency and damping dimension, respectively, such as

$$\mathcal{X} \approx \sum_{p=1}^{P_1} \sum_{j=1}^{J_1} g_{p,j} \mathbf{a}_{\omega_p}^{(1)}(\beta_j) \circ \mathbf{a}_{\omega_2}^{(2)}(\beta_2) \circ \cdots \circ \mathbf{a}_{\omega_N}^{(N)}(\beta_N) \tag{7}$$

where $\omega_2, \ldots, \omega_N$ and $\beta_2, \ldots, \beta_N$ denote the (for simplicity) fixed frequency and damping coefficients along the 2nd to N:th dimensions,

$$\mathbf{a}_{\omega}^{(n)}(\beta) = \begin{bmatrix} \xi_n^{t_1^{(n)}} & \ldots & \xi_n^{t_{I_n}^{(n)}} \end{bmatrix}^T$$

where

$$\xi_n = e^{j\omega^{(n)} - \beta^{(n)}} \tag{8}$$

and $g_{k,\ell}$ denotes the contribution of each of these dictionary elements in the approximation. Thus, as long as P_1 and J_1 are selected sufficiently large to allow for a grid of dictionary elements such that the true frequency and damping coefficients lie close to one of the grid points, only one $g_{p,j}$ should be non-zero for each of the K modes. By similarly extending the dictionary for each of the frequency and damping dimensions, such that $g_{p_1,\ldots,p_N,j_1,\ldots,j_N}$ denotes the contribution of the corresponding dictionary elements for the p_k:th and j_r:th frequency and damping dictionary elements, where $k, r \in \{1, \ldots, N\}$, the resulting (very large) dictionary would allow for a sparse approximative solution of the unknown parameters, such that most of the dictionary elements would not contribute to the approximation. Given such an approximative solution, the number of modes, K, may be estimated as the number of elements with non-zero contribution to the approximation. The non-linear parameters may then be estimated correspondingly, such that for any non-zero variables, e.g., $g_{h_1,\ldots,h_N,i_1,\ldots,i_N}$, the non-linear parameters are estimated as the frequency and damping coefficient that correspond to the found

coefficients. Such a solution may be obtained by reformulating the problem using the *vec* operator, defined here for tensors such that it is the usual *vec* operation on the mode-1 matricization, or unfolding (see also [28]), of a given tensor, i.e.,

$$vec\left(\boldsymbol{\mathcal{X}}\right) \triangleq vec\left(\mathbf{X}_{(1)}\right) \tag{9}$$

This allows for a sparse LS solution to be found by solving

$$\min_{\tilde{\mathbf{g}}} \left\| vec\left(\boldsymbol{\mathcal{X}}\right) - \tilde{\mathbf{A}}\tilde{\mathbf{g}} \right\|_2^2 + \rho(\tilde{\mathbf{g}}) \tag{10}$$

where $\tilde{\mathbf{g}} = vec\left(\boldsymbol{\mathcal{G}}\right)$, with $\boldsymbol{\mathcal{G}} \in \mathbb{C}^{P_1 \times \cdots \times J_N}$ denoting the tensor formed from the amplitudes of all of the dictionary elements, and the i:th column of $\tilde{\mathbf{A}}$ is formed as

$$\tilde{\mathbf{A}}_{:i} = vec\left(\mathbf{a}^{(1)}_{\omega_{k_1}}(\beta_{j_1}) \circ \mathbf{a}^{(2)}_{\omega_{k_2}}(\beta_{j_2}) \cdots \circ \mathbf{a}^{(N)}_{\omega_{k_N}}(\beta_{j_N})\right) \tag{11}$$

where the notation $\mathbf{A}_{:i}$ denotes the ith column of the matrix $\mathbf{A}$. The penalty term $\rho(\cdot)$ is added in (10) as the grid is typically chosen such that the number of elements in $vec(\boldsymbol{\mathcal{X}})$ is smaller than the number of columns in $\tilde{\mathbf{A}}$; thus, if assuming that $\tilde{\mathbf{A}}$ is of full rank, the system of equations is under-determined, with infinitely many solution, out of which one is interested in finding one that appropriately weighs sparsity and model fit. Ideally, $\rho(\cdot)$ could be chosen as a function counting the number of non-zero elements. However, the resulting optimization problem is well known to be combinatorial in nature and will be unfeasible to solve even for moderate problem sizes. Common approximative choices include the scaled ℓ_1 norm [17, 30], ℓ_q penalties [16, 31], and the reweighted ℓ_1 approach, which may be seen to correspond to the log penalty [32]. Herein, we consider the ℓ_1 and the log penalty. It is worth noting that the above sparsity restrictions allow for solutions having multiple damping coefficients for a given frequency. Such solutions imply that the component is not an exponentially damped sinusoid; as this is not relevant for the here considered application, we proceed to refine the constraint such that it will only yield unique frequency-damping pairs for each component. To this end, we propose an iterative dictionary learning approach such that the damping parameters for each sinusoidal component is held fixed during the sparse LS step, after which the damping parameters are found using the residual from the sparse LS step, one mode at the time, thus allowing for damping and frequency

estimation to be performed with a non-linear optimization algorithm, e.g., Newton's method. Thus, we initially fix all damping parameters to zero, modifying (7) such that the dictionary is only formed over the unknown frequencies, i.e.,

$$\boldsymbol{\mathcal{X}} \approx \sum_{p_1=1}^{P_1} \cdots \sum_{p_N=1}^{P_N} g_{p_1,\ldots,p_N} \mathbf{a}^{(1)}_{\omega_{p_1}}(\beta_{p_1}) \circ \cdots \circ \mathbf{a}^{(N)}_{\omega_{p_N}}(\beta_{p_N}) \tag{12}$$

The resulting minimization with respect to the K unknown frequencies, which may then be used to estimate the damping components, iteratively finding each of the set of estimates. To allow for a computationally efficient solution, the considered frequency and damping grids, respectively, are updated in each iteration, such that the dictionary is refined in each step of the iteration. However, even with such a reduction in complexity, the iterative optimization problems are clearly daunting, being formed over $J_1 \times \cdots \times J_N$ and $P_1 \times \cdots \times P_N$ dimensions, respectively. In the next two sections, we therefore proceed to examine how these minimizations may be performed in an efficient manner utilizing the Kronecker structure of the dictionary matrices for the sparse LS step, and by solving the non-linear damping parameter estimation one mode at a time.

3 An efficient ADMM implementation

The minimization problem considered in (10) may be solved using an approximation of the form

$$\min_{\tilde{\mathbf{g}}} \left\| vec(\boldsymbol{\mathcal{X}}) - \tilde{\mathbf{A}}\tilde{\mathbf{g}} \right\|_2^2 + \sum_{k=1}^{P_1 \times \cdots \times J_N} \lambda_k |\tilde{g}_k| \tag{13}$$

where λ_k denotes a set of tuning parameters, for $k = 1, \ldots, P_1 \times \cdots \times J_N$. In case these tuning parameters are all selected equal and the penalty is included as an inequality constraint, the resulting minimization is equivalent with the regular ℓ_1 penalized LS problem, often called basis pursuit denoising [33], or the LASSO [30]. For highly correlated dictionary elements, as may be required for high resolution N-D spectra, one may obtain sparser solutions using a reweighted LASSO formulation [32], such that the λ_k:s are instead selected as

$$\lambda_k = \frac{\varphi}{|\tilde{g}_k(\ell)| + \varepsilon} \tag{14}$$

Algorithm 1 Sparse LS via ADMM

1: Initiate $\mathbf{z} = \mathbf{z}(0), \mathbf{u} = \mathbf{u}(0)$, and $\ell = 0$
2: **repeat**
3: $\quad \mathbf{z}(\ell+1) = \left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \mu\mathbf{I}\right)^{-1}\left(\tilde{\mathbf{A}}^H\mathbf{y} - \mathbf{u}(\ell) - \mathbf{d}(\ell)\right)$
4: $\quad \mathbf{u}(\ell+1) = \Psi\left(\mathbf{z}(\ell+1) - \mathbf{d}(\ell+1), \frac{\lambda}{\mu}\right)$
5: $\quad \mathbf{d}(\ell+1) = \mathbf{d}(\ell) - (\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1))$
6: $\quad \ell \leftarrow \ell + 1$
7: **until** convergence

where the constant ε is included to avoid numerical problems when $g_k(\ell)$ is close to zero. Here, $\tilde{g}_k(\ell)$ denotes the value of g_k at iteration ℓ, and with $\varphi > 0$ denoting a tuning parameter controlling the sparsity at the solution. A general efficient iterative algorithm for solving problems such as (10), using an ADMM implementation was proposed in [27], and may be easily adapted to the here considered reweighted scenario. The steps involved are summarized in Algorithm 1, where the Ψ operator is a shrinkage operator, defined as

$$\Psi(\mathbf{x}, \gamma) = \mathbf{x}(1 - \gamma/|\mathbf{x}|)^+ \tag{15}$$

where $(\cdot)^+$ denotes the positive part of a scalar. The complexity of each iteration in the resulting algorithm is approximately $\mathcal{O}(n^2p)$, where p and n denote the columns and rows of $\mathbf{a}$, respectively. This is about the same as the computational cost for many LASSO solvers (see e.g. [34]). In the N-dimensional case, the overall computational complexity is about $\mathcal{O}(\prod_{n=1}^{N} J_n P_n \prod_{n=1}^{N} I_n^2)$, implying that even a 3-dimensional problem with 100 grid points in each dimension would result in a cost of approximately $100^{12} I_1 I_2$ operations, in each step, where I_n denotes the number of samples in dimension n. Fortunately, this complexity may be significantly reduced by exploiting the inherent Kronecker structure of the model. In order to do so, we rewrite (4) using tensor products as

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} + \boldsymbol{\mathcal{E}} \tag{16}$$

where the operator $\times_n$ represents the n-mode product of a tensor with a matrix, and the dictionary matrix for dimension n is given as

$$\mathbf{A}^{(n)} \triangleq \left[\ \mathbf{a}^{(n)}_{\omega_{k_1}}(\beta_{k_1}) \quad \ldots \quad \mathbf{a}^{(n)}_{\omega_{K_1}}(\beta_{K_1})\ \right] \tag{17}$$

Algorithm 2 Mode estimation

1: Use (10) to form initial estimates $\{g_{\boldsymbol{\gamma}_k}\}_{k=1}^{\hat{K}}$
2: Compute the residual according to (28)
3: **for** $k = 1, \ldots, \hat{K}$ **do**
4: Add the current mode to the residual:
$\boldsymbol{\mathcal{Y}}_k = \boldsymbol{\mathcal{R}}_k + g_{\boldsymbol{\gamma}_k} \mathbf{a}_k^{(1)} \circ \cdots \circ \mathbf{a}_k^{(N)}$
5: Estimate the frequencies and the dampings for the mode
6: Remove the current mode:
$\boldsymbol{\mathcal{R}}_k = \boldsymbol{\mathcal{Y}}_k - g_{\boldsymbol{\gamma}_k} \mathbf{a}_k^{(1)} \circ \cdots \circ \mathbf{a}_k^{(N)}$
7: **end for**

Expressed in this form, one may note that the matricization may be accomplished via Kronecker products instead (see, e.g., [28], [35]), yielding

$$\mathbf{X}_{(1)} = \mathbf{A}^{(1)} \mathbf{G}_{(1)} \left(\mathbf{A}^{(N)} \otimes \mathbf{A}^{(N-1)} \otimes \cdots \otimes \mathbf{A}^{(2)} \right)^T \tag{18}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{X}_{(1)} \in \mathbb{C}^{I_1 \times \prod_{n=2}^{N} I_n}$ is obtained by stacking all the mode-1 slices of $\boldsymbol{\mathcal{X}}$, and with $\mathbf{G}_{(1)}$ defined similarly. Vectorizing the resulting mode-1 slices yields (see, e.g., [36]),

$$\text{vec}\left(\mathbf{X}_{(1)}\right) = \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(2)} \otimes \mathbf{A}^{(1)} \right) \text{vec}\left(\mathbf{G}_{(1)}\right) \tag{19}$$

allowing us to express the parameters in (10) as

$$\tilde{\mathbf{g}} \triangleq \text{vec}\left(\mathbf{G}_{(1)}\right) \in \mathbb{C}^{\tilde{K} \times 1} \tag{20}$$

$$\tilde{\mathbf{A}} \triangleq \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(2)} \otimes \mathbf{A}^{(1)} \right) \in \mathbb{C}^{\tilde{I} \times \tilde{K}} \tag{21}$$

As a result, the full $\tilde{\mathbf{A}}$ matrix does not need to be formed, and vector multiplication of the form $\tilde{\mathbf{A}}\mathbf{x}$ and $\tilde{\mathbf{A}}^H\mathbf{y}$, for any appropriately sized vector $\mathbf{x}$ and $\mathbf{y}$, may be computed iteratively by each sub-matrix $\mathbf{A}^{(n)}$, and by then reshaping the resulting elements (see, e.g., [37, p. 28] for further details). This allows for a dramatic complexity reduction. To illustrate this, consider the case where each $\mathbf{A}^{(\ell)}$ matrix is $n \times n$. Then, the operation $\tilde{\mathbf{A}}\mathbf{x}$, which would require about $\mathcal{O}(n^{2N})$ multiplications if first forming $\tilde{\mathbf{A}}$ and then computing the inner-product using this matrix, may instead be formed using only $\mathcal{O}(Nn^{N+1})$ operations (see, e.g., [38]).

Furthermore, the LS step in the ADMM algorithm for solving (10) may also be computed significantly cheaper by utilizing its Kronecker structure, simply by calculating the singular value decomposition of each sub-matrix $\mathbf{A}^{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^H$, and then utilizing that the singular value decomposition of $\tilde{\mathbf{A}}$ is given by (see, e.g., [36, p. 246])

$$\tilde{\mathbf{A}} = \mathbf{U}_{\tilde{\mathbf{A}}} \mathbf{\Sigma}_{\tilde{\mathbf{A}}} \mathbf{V}_{\tilde{\mathbf{A}}}^H \tag{22}$$

where

$$\mathbf{U}_{\tilde{\mathbf{A}}} = \mathbf{U}_1 \otimes \cdots \otimes \mathbf{U}_N \tag{23}$$
$$\mathbf{\Sigma}_{\tilde{\mathbf{A}}} = \mathbf{\Sigma}_1 \otimes \cdots \otimes \mathbf{\Sigma}_N \tag{24}$$
$$\mathbf{V}_{\tilde{\mathbf{A}}}^H = \mathbf{V}_1^H \otimes \cdots \otimes \mathbf{V}_N^H \tag{25}$$

As a result, one may solve step 3 in Algorithm 1 by solving the equivalent LS problem

$$\min_{\tilde{\mathbf{z}}} \left\| \begin{bmatrix} \mathbf{U}_{\tilde{\mathbf{A}}}^H \mathbf{y} \\ \mathbf{V}_{\tilde{\mathbf{A}}}^H \boldsymbol{\xi} \end{bmatrix} - \begin{bmatrix} \mathbf{\Sigma}_{\tilde{\mathbf{A}}} \\ \mu \mathbf{I} \end{bmatrix} \tilde{\mathbf{z}} \right\| \tag{26}$$

where

$$\tilde{\mathbf{z}} = \left(\mathbf{\Sigma}_{\tilde{\mathbf{A}}}^2 + \mu^2 \mathbf{I} \right)^{-1} \left(\mathbf{\Sigma}_{\tilde{\mathbf{A}}} \mathbf{U}_{\tilde{\mathbf{A}}}^H \mathbf{y} + \mu^2 \mathbf{V}_{\tilde{\mathbf{A}}}^H \boldsymbol{\xi} \right) \tag{27}$$

with $\tilde{\mathbf{z}} = \mathbf{V}_{\tilde{\mathbf{A}}}^H \mathbf{z}$ and $\boldsymbol{\xi} = \tilde{\mathbf{A}}^H \mathbf{y} - \mathbf{u}(\ell) - \mathbf{d}(\ell)$. Thus, the LS step can be solved by three matrix vector multiplications, two Hadamard products between vectors, one scalar multiplication of a vector, and a vector-vector addition, which may all be calculated using their inherent Kronecker structure, significantly reducing the computational cost. For example if each $\mathbf{A}^{(\ell)}$ is $n \times n$, the cost for our approach is about $\mathcal{O}(3Nn^{N+1})$ versus $\mathcal{O}(n^{3N})$ for a solution that does not use the inherent structure of the problem.

4 Sparse dictionary learning

As noted above, the considered grid over the candidate frequency and damping coefficients are updated in alternating fashion. Let $\hat{K}$ denote the number of non-zero amplitudes after the sparse LS step. Then, the dictionary learning may be

done by forming the residual[1]

$$\mathcal{R} = \mathcal{X} - \mathcal{G}^* \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} \tag{28}$$

Using a relaxation-based procedure (see also [39]), one then iteratively adds back one mode at a time to the residual in (28), and form an estimate of the frequency and damping of this mode using an N-dimensional single mode solver, such as, for instance, the PUMA estimator [40]. Using the refined parameter estimates, the mode is then subtracted again, and the next mode is refined similarly. The procedure is summarized in Algorithm 2. Using the refined modes, the dictionary is then updated, such that it is separated into N dictionaries, one over each dimension, with each dictionary being centered in a fine grid around each of the found frequencies. As a result, the unused dictionary elements, having zero-amplitudes, are excluded from the updated dictionary (unless being close to one of the found modes). This also implies that closely spaced modes may yield overlapping dictionary elements; such duplicated dictionary elements are removed to avoid collinearity in the dictionary. For each grid point, the dictionary element is scaled according to the found damping coefficient of the corresponding mode, to ensure that all dictionary elements have the same norm, thus refining the dictionary iteratively over both frequencies and damping coefficients. We coin the resulting method the Sparse Exponential Mode Analysis (SEMA) algorithm.

5 Numerical examples

We proceed by examine the performance of the proposed method using simulated data. To simplify the presentation, we focus on the 1-D and 2-D cases, since problems of these dimensions offer more intuitive results that are also easier to analyze. Considering first the 1-D case, we illustrate the performance of the proposed method using simulated data. We initially consider a data vector containing $N = 128$ samples of a three mode signal, where the frequencies and damping parameters are chosen uniformly over $[0,\ 1]$ and $[0,\ 0.025]$, respectively. We note that we here use normalized frequencies, lying in the interval $[0,\ 1]$, denoted by the letter f. For now, we ensure that no modes are closer in frequency than $1/N$. Figures 1 and 2 depict the resulting performance of the SEMA algorithm, as compared to the non-parametric damped-Capon (dCapon) estimate [12,15], as a

[1] To simplify our notation, we have here suppressed the dependencies on the frequency ω and the damping β.

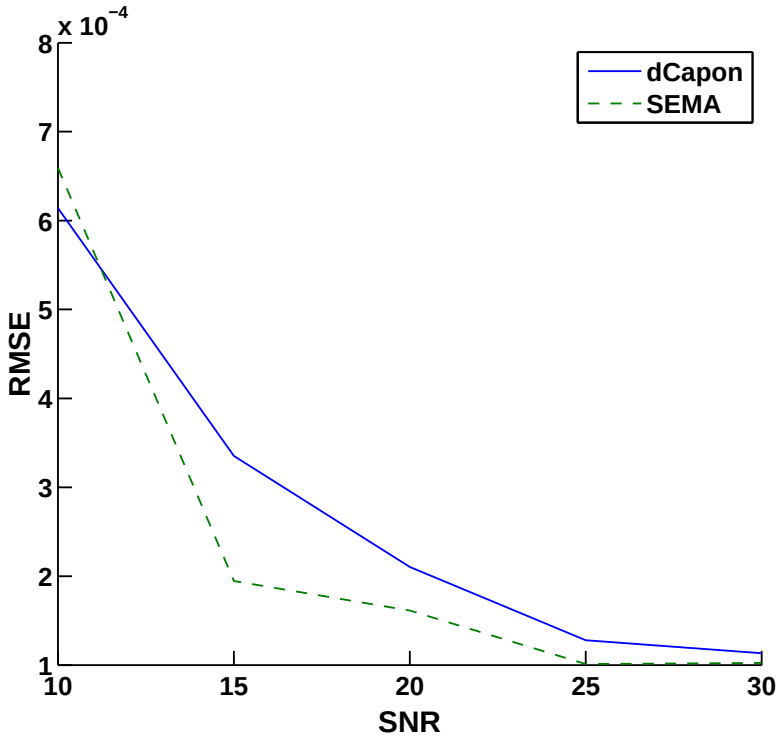

Figure 1: The RMSE of the frequency estimation as a function of SNR.

function of the signal-to-noise-ratio (SNR), defined as $\log_{10}(||\mathbf{y}||_2^2/N\sigma^2)$, where σ^2 denotes the variance of the noise. The two figures show the root mean squared error (RMSE) of the frequency and damping estimates, defined as

$$\text{RMSE} = \sqrt{\frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K}\left(\vartheta_{m,k} - \hat{\vartheta}_{m,k}\right)^2} \tag{29}$$

where $\vartheta_{m,k}$ denotes the estimate of either the frequency or the damping of mode k for Monte-Carlo simulation m, M is the total number of Monte-Carlo simulations, and K the number of modes. These results have been obtained using $M = 175$ Monte-Carlo simulations. In this example, dCapon have a frequency grid that is selected to be 6000×6000, uniformly covering frequencies and damp-

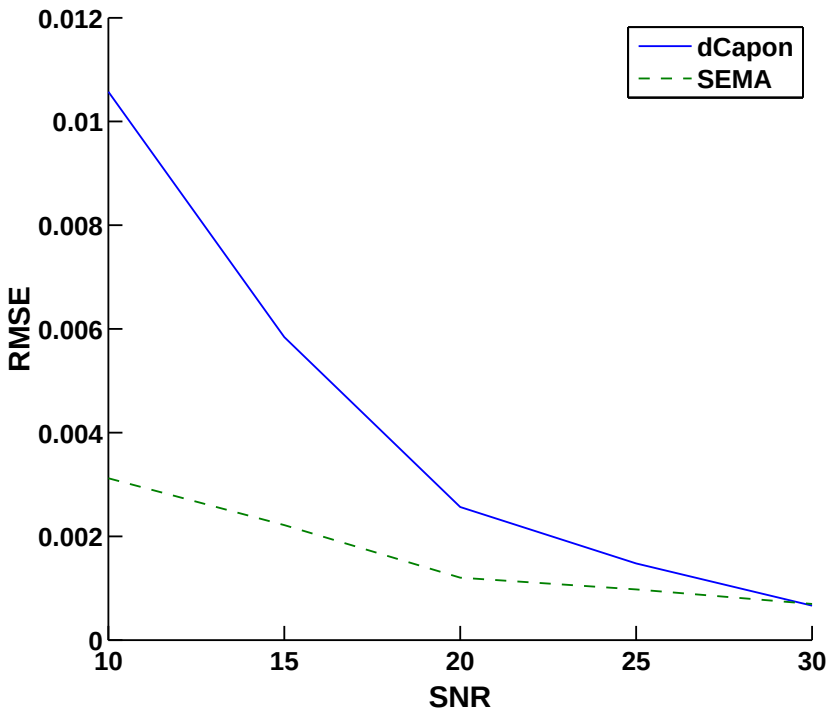

Figure 2: The RMSE of the damping estimation as a function of SNR.

ing factors in $[0, 1]$ and $[0, 0.025]$, respectively, and where the recommended filter length of $N/4$ is used. The SEMA algorithm on the other hand uses a dictionary containing only 128 elements in the first iteration, and, thereafter, uses only 40 grid points for each found mode when updating the dictionary in each subsequent iteration. As can be seen from the figures, the proposed SEMA algorithm yields notably better estimates than the dCapon estimator, without requiring a large dictionary grid over both dimensions, thereby allow for a substantially faster implementation. It is also worth noting that the dCapon estimation errors are here larger than the smallest possible error that is attainable given the current grid size, implying that the grid size does not in itself limit the quality of the estimates.

Next, we examine the ability of the methods to resolve two closely spaced spectral lines. In this case, we consider a signal containing two sinusoidal compon-

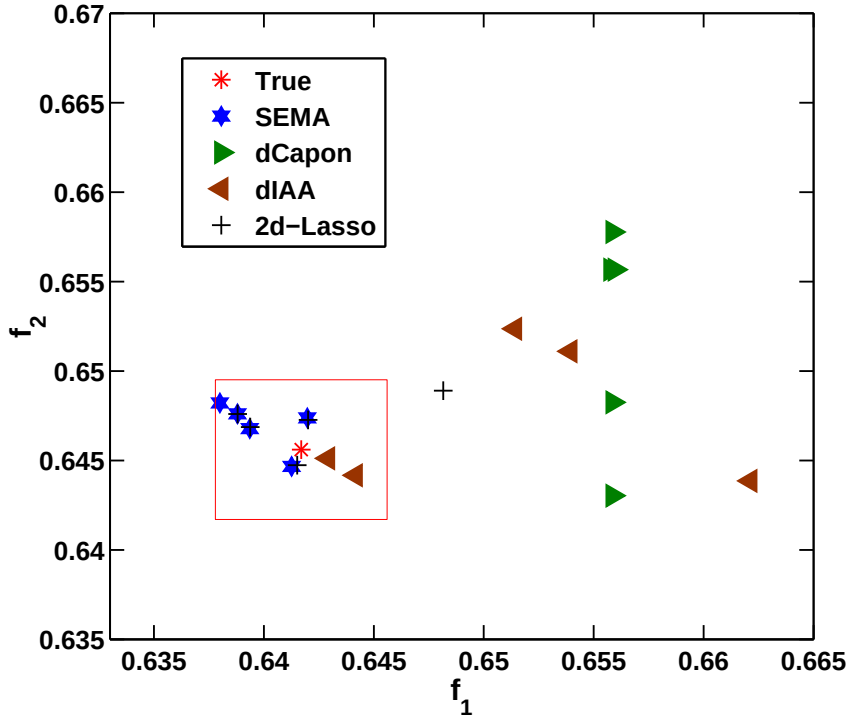

Figure 3: The result of resolving two closely spaced spectral peaks. The (red) square indicates the distance $1/(2N)$ from the true frequencies.

ents with frequencies, $f_1 = 0.6417$ and $f_2 = 0.6456$, i.e., separated by $0.5/N$, with random damping constants, being drawn uniformly from $[0, 0.025]$. Figure 3 illustrates the resulting frequency estimates as obtained from 5 Monte-Carlo simulations, and SNR $= 20$ dB. For comparison, the figure also shows the estimates obtained using 1-D SEMA, dCapon, dIAA [41], and for a Lasso method with a dictionary containing both frequencies and damping factors, and exploiting a zooming similar to the one used in SEMA. Here, to speed-up the computations, the frequency grid for dCapon and dIAA have been selected to only be formed on $[0.63, 0.67]$, allowing the methods notable *a priori* information on the frequency region of interest. The damping grid ranges over $[0, 0.025]$ and has size 500 for all methods, except for the used Lasso method, where, due to complexity reasons,

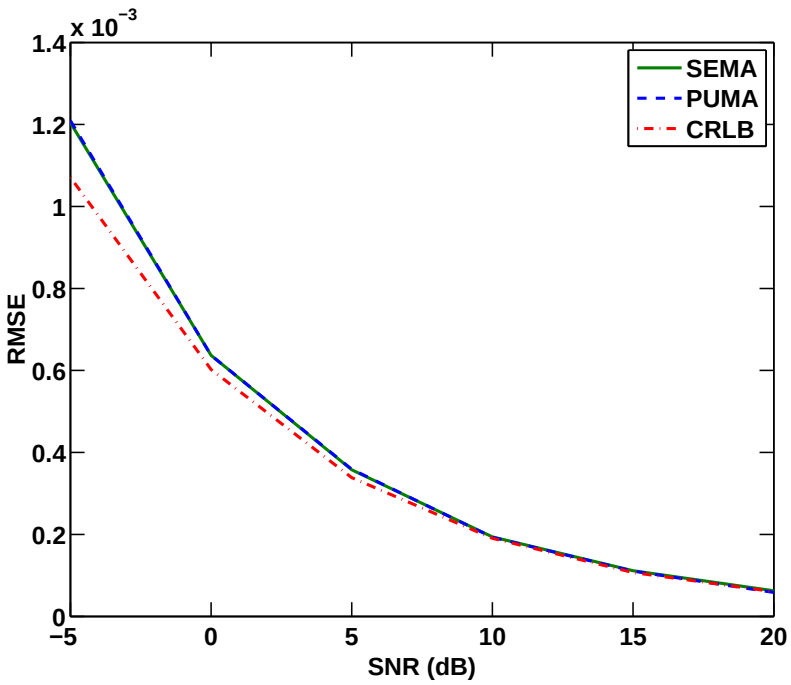

Figure 4: The average RMSE of $f_1^{(1)}$ and $f_2^{(1)}$ as a function of SNR.

it is set to 10. As seen in the figure, both the proposed method and the Lasso method clearly manage to resolve the two peaks, whereas dCapon and dIAA, in most cases, are unable to find the correct peaks. In the figure, the (red) square indicates the region $1/(2N)$ around the true frequencies.

We proceed to examine the performance of the SEMA algorithm for 2-D simulated data, examining the RMSE of two well separated peaks, showing that the proposed method has similar performance to the statistically efficient PUMA method [7], using simulated data mimicking a 2-D NMR signal, containing two damped sinusoids and having 33×31 samples. Figures 4-7 illustrates the performance of the SEMA estimator as compared to the parametric PUMA estimator and the corresponding Cramér-Rao lower bound (CRLB) [42]. The frequencies where randomly selected in the interval from 0 to 1 in normalized frequencies,

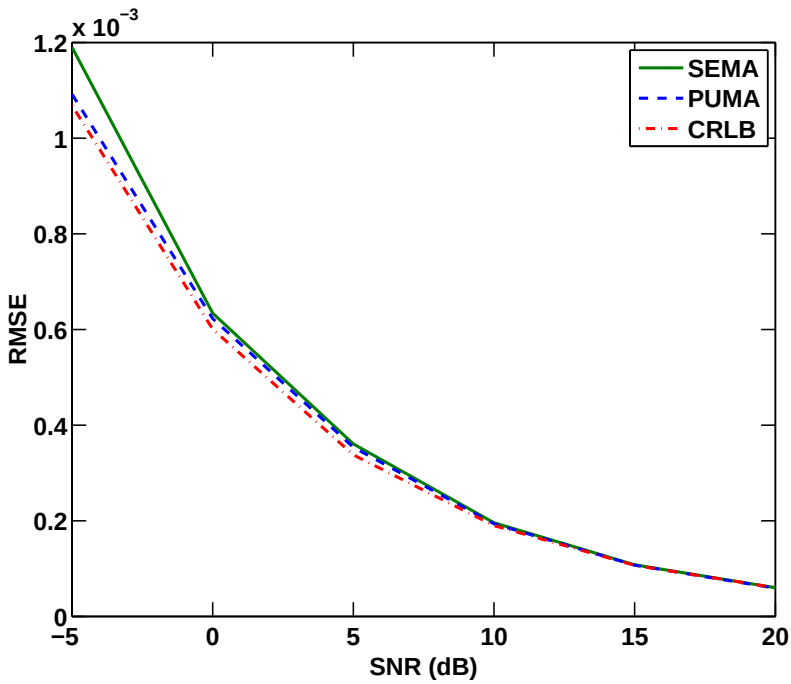

Figure 5: The average RMSE of $f_1^{(2)}$ and $f_2^{(2)}$ as a function of SNR.

and selected such that components where separated by at least $3/N$ in each dimension. If the spacing between the peaks is smaller, the estimation will degenerate for all methods. The damping parameters were set to $\beta_1 = (0.05 \quad 0.02)$ and $\beta_2 = (0.01 \quad 0.04)$ for all simulations. Each node was normalized in amplitude, thus making sure that both peaks were equally dominant. The PUMA algorithm was, as for all examples, allowed 100 iterations, as well as oracle model order information, and the initial grid for the proposed 2-D method was, as for the following examples, set to 100. The proposed method was allowed two iterations and used 33 grid points to zoom in on each found mode. The choice of λ governs the number of peaks that may be found. If set too high, peaks with low amplitude will be suppressed, and if set too low, peaks that originate from the noise will not be suppressed. However, due to the reweighting step, a too small λ will be compensated for, and therefore the algorithm is relatively robust

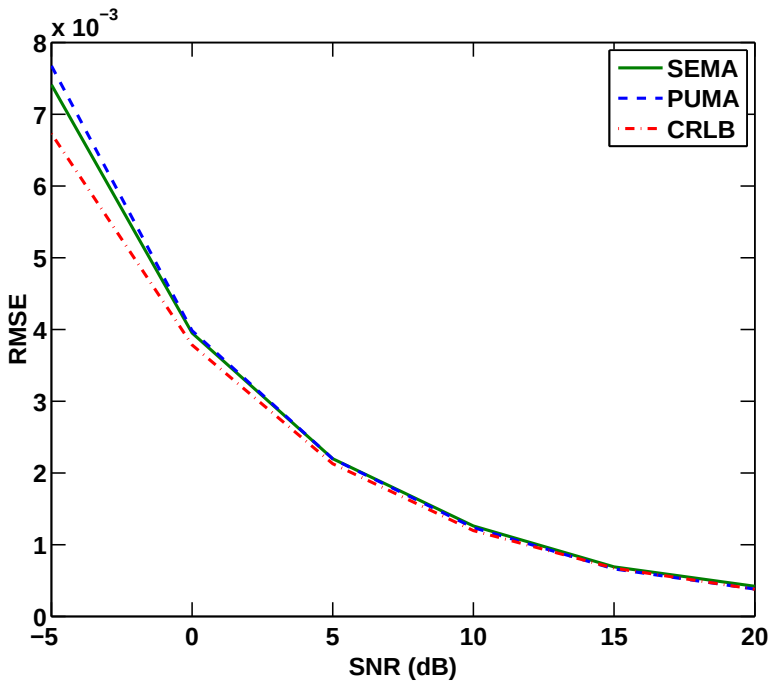

Figure 6: The average RMSE of $\beta_1^{(1)}$ and $\beta_2^{(1)}$ as a function of SNR.

to the choice of λ, as long as it is not set too large. Therefore, it is preferable to set λ to a small value. In these examples, we set λ equal to the tenth largest peak found in the periodogram. One could argue that we thereby limit the number of peaks that may be found, but that is easily avoided. If λ were set to equal the amplitude of the r:th largest peak and, when using the method, we found r peaks, one would run the algorithm a second time but with a somewhat smaller λ value. In this way, we make sure that we do not in fact limit the algorithm to a specified number of peaks. The test was performed using 250 Monte-Carlo simulations, for each value of the considered SNR. Figures 4-7 illustrate the total RMSE of all the unknown parameters. As can be seen from the figure, both the parametric PUMA, which has been allowed oracle model order information, and the proposed semi-parametric SEMA algorithms yield statistically efficient parameter estimates especially for larger larger SNR. Here, if the proposed algorithm

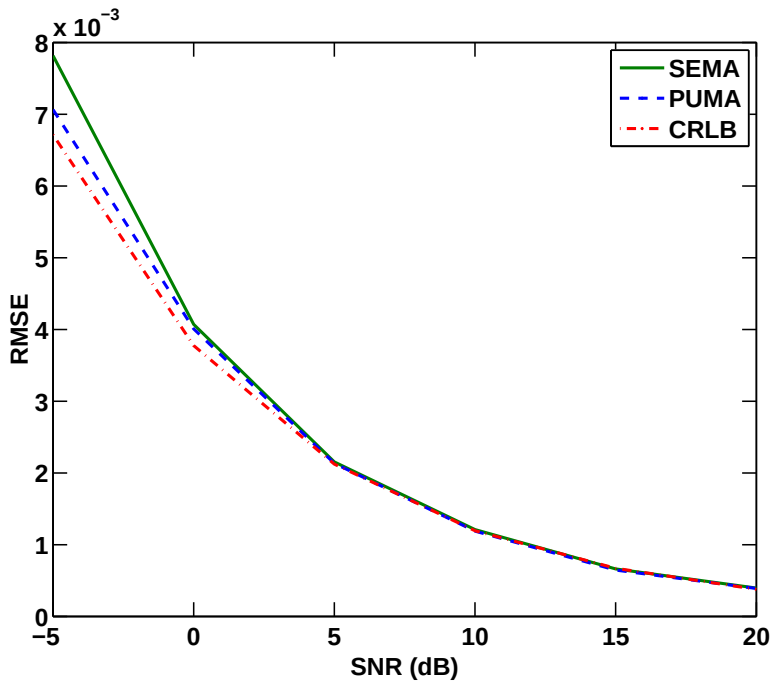

Figure 7: The average RMSE of $\beta_1^{(2)}$ and $\beta_2^{(2)}$ as a function of SNR.

did not manage to estimate the number of modes correctly, that estimate was then removed from the RMSE calculations for all methods. This happened two times out of 1500 Monte-Carlo simulations.

We proceed to examine the methods ability to resolve two closely spaced peaks. This was done by fixing the first mode at frequency $f_1 = (0.4, \quad 0.6)$, and letting the second mode gradually approach the first. The modes were initially separated by $1/N_1$ and $1/N_2$ in each frequency dimension, and the test was stopped when the modes were separated by $0.1/N_1$ and $0.1/N_2$. The data size for this example was again 33×31. The same SEMA settings as above were used. We also compare the estimates to that of a zero-padded 2-D periodogram, where 2^{13} zeros were padded in each dimension, but zoomed in on the correct frequencies (± 0.1 in each frequency). The damping parameters where fixed to 0.02 for all modes and dimensions, and the SNR was set to 10 dB. Furthermore, PUMA was

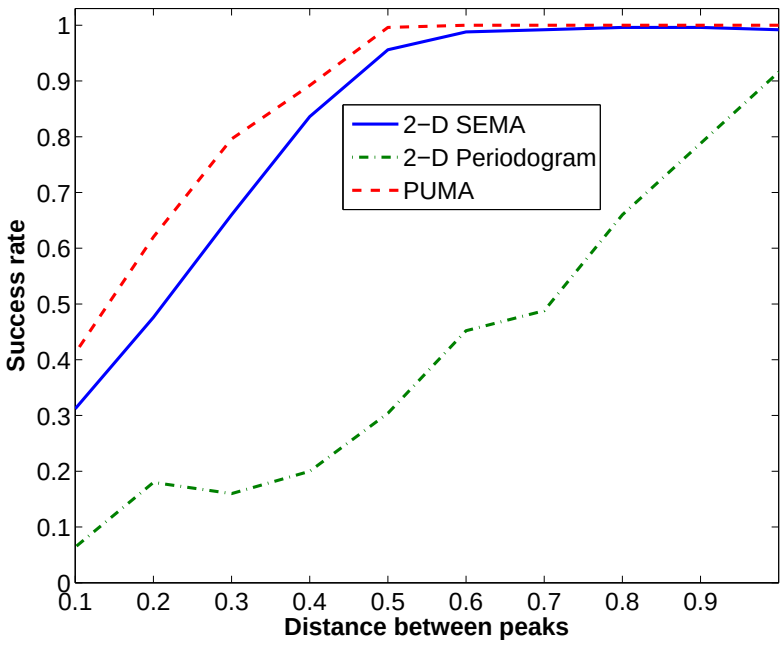

Figure 8: Ability to resolve two peaks as a function of the peak separation.

again allowed complete knowledge of the number of peaks. To determine whether or not two peak were resolved, we ensured that the method fulfilled at least two separation criteria: First, the peaks that were found had to be at least within a rectangle of size $1/N_1 \times 1/N_2$ from the correct frequencies; Secondly, the power of the valley between the peaks where allowed to be at most 90% of the average power of the peaks. If these two criteria were met, the modes were deemed to be resolved. The results are shown in Figure 8, where the x-axis should be interpreted as the distance divided by N_1, i.e., 0.1 means that the distance between the modes is $0.1/N_1$. As may be seen from the figure, the periodogram's ability to distinguish the two modes drastically decreases as the modes become closer. As may be expected. the PUMA method on the other hand manages to separate the modes very well until they are about 0.3 apart from each other. As can be seen from the figure, the SEMA method achieves about the same performance as

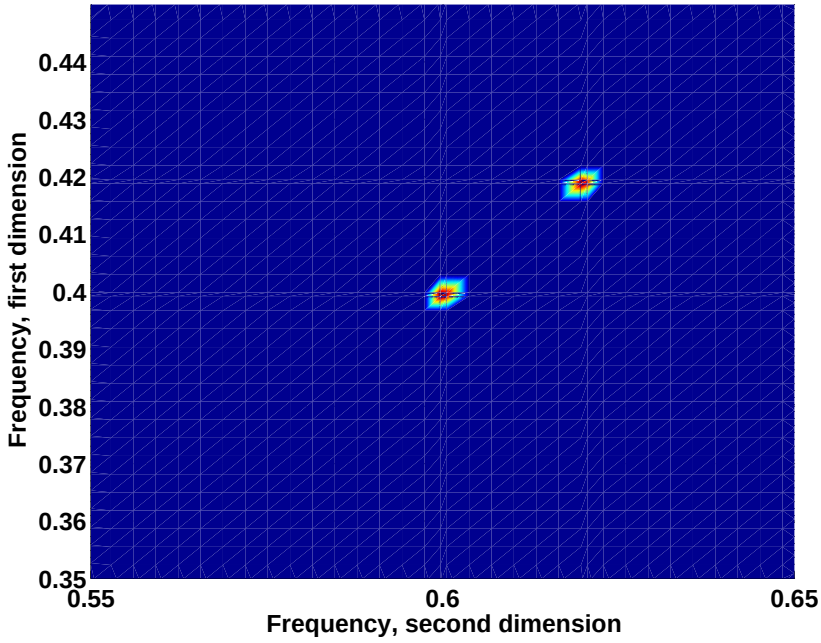

Figure 9: Resulting estimates using 2-D SEMA on two closely spaced modes.

PUMA until the distance is less than 0.4. It should be stressed that the PUMA estimator is given perfect prior knowledge about the number of modes, whereas the 2-D SEMA has no such prior information. As is clear from the figure, the SEMA estimate seems to be able to separate closely spaced modes almost as well as the parametric and statistically efficient PUMA estimator, without imposing any a priori model order information, as well as yielding far better performance than the periodogram estimate. A typical result is shown in Figures 9 and 10, where the peaks are separated by $0.5/N1$. It clearly shows how SEMA manages to separate the two peaks, whereas the periodogram only shows one peak.

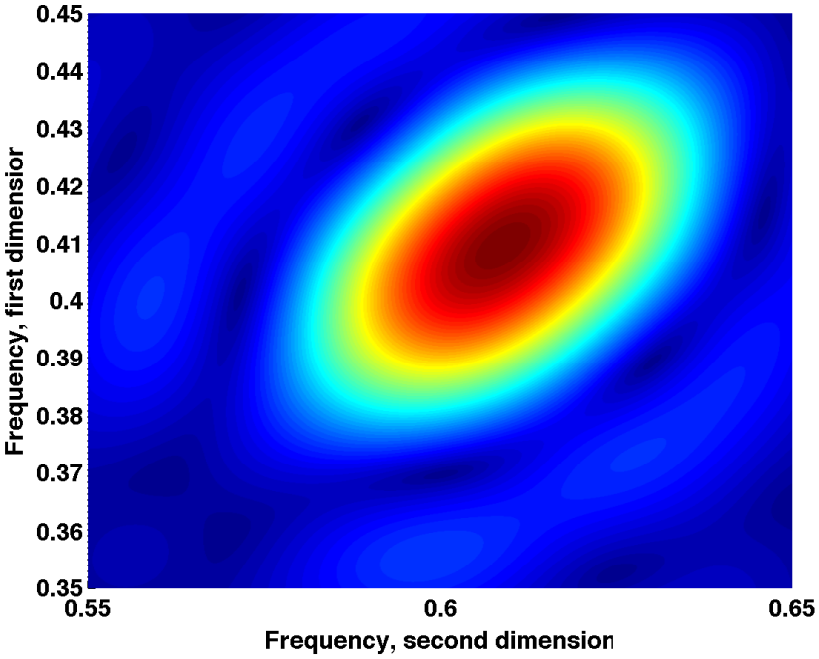

Figure 10: Resulting estimates using two dimensional periodogram on two closely spaced modes.

6 Conclusions

In this work, we have introduced a semi-parametric separable sparse model for (possibly non-uniformly sampled) N-dimensional damped sinusoidal signal components, forming a computationally efficient implementation exploiting the inherent structure of the resulting tensors. The proposed SEMA algorithms is found to yield highly accurate estimates of the frequency and damping coefficients of the signal modes, without imposing strong *a priori* knowledge on the number of modes present in the signal. The performance of the method is illustrated using 1- and 2-D simulated data as compared to the (parametric) PUMA estimator, the Cramér-Rao lower bound, and a zero-padded periodogram estimate, as well as the corresponding non-parametric Capon- and IAA-based estimators, and a LASSO-based estimator, clearly illustrating the achievable performance gain.

Acknowledgment

The authors would like to thank the authors of [7] and [40] for providing their implementation of the PUMA algorithm. This work was supported in part by the Swedish Research Council, and Crafoord's and Carl Trygger's foundations. This work has been presented in part at the ICASSP and EUSIPCO conferences [1, 2].

References

[1] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Signals," in *Proc. 39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Florence, Italy, May 4-9 2014.

[2] S. I. Adalbjörnsson, J. Swärd, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying Two-dimensionalimensional Signals," in *22nd European Signal Processing Conference*, Lisbon, Portugal, 2014.

[3] J. Liu and X. Liu, "An Eigenvector-Based Approach for Multidimensional Frequency Estimation With Improved Identifiability," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4543–4556, 2006.

[4] Y. Hua, "Estimating Two-Dimensional Frequencies by Matrix Enhancement and Matrix Pencil," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2267–2280, September 1992.

[5] J. Sacchini, W. Steedly, and R. Moses, "Two-dimensional Prony modeling and parameter estimation," *IEEE Transactions on Signal Processing*, vol. 41, no. 11, pp. 3127–3137, November 1993.

[6] S. Rouquette and M. Najim, "Estimation of Frequencies and Damping Factors by Two-Dimensional ESPRIT Type Methods," *IEEE Transactions on Signal Processing*, vol. 49, no. 49, pp. 237–245, January 2001.

[7] F. K. W. Chan, H. C. So, and W. Sun, "Subspace approach for two-dimensional parameter estimation of multiple damped sinusoids," *Signal Process.*, vol. 92, pp. 2172 – 2179, 2012.

[8] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-Order SVD-Based Subspace Estimation to Improve the Parameter Estimation Accuracy in Multidimensional Harmonic Retrieval Problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, July 2008.

[9] Y. Li, J. Razavilar, and K. J. R. Liu, "A High-Resolution Technique for Multidimensional NMR Spectroscopy," vol. 45, no. 1, pp. 78–86, 1998.

[10] W. Sun and H. C. So, "Accurate and Computationally Efficient Tensor-Based Subspace Approach for Multidimensional Harmonic Retrieval," vol. 60, no. 10, pp. 5077–5088, Oct. 2012.

[11] S. Sahnoun, E. H. Djermoune, and D. Brie, "Sparse Modal Estimation of 2-D NMR Signals," in *38th IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 26-31 2013.

[12] P. Stoica and T. Sundin, "Nonparametric NMR Spectroscopy," *J. Magn. Reson.*, vol. 152, pp. 57–69, 2001.

[13] E. Gudmundson, P. Stoica, J. Li, A. Jakobsson, M. D. Rowe, J. A. S. Smith, and J. Ling, "Spectral Estimation of Irregularly Sampled Exponentially Decaying Signals with Applications to RF Spectroscopy," *J. Magn. Reson.*, vol. 203, no. 1, pp. 167–176, March 2010.

[14] F. J. Frigo, J. A. Heinen, J. A. Hopkins, T. Niendorf, and B. J. Mock, "Using Peak-Enhanced 2D-Capon Analysis with Single Voxel Proton Magnetic Resonance Spectroscopy to Estimate T2* for Metabolites," in *Proc. of ISMRM*, 2004, vol. 12, p. 2437.

[15] G. O. Glentis and A. Jakobsson, "Computationally efficient damped Capon and APES spectral estimation," in *21st European Signal Processing Conference*, Marrakech, Morocco, Sept. 9-13 2013.

[16] I. F. Gorodnitsky and B. D. Rao, "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm," vol. 45, no. 3, pp. 600–616, March 1997.

[17] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul, jul 2008, pp. 10225–10229.

[18] P. Stoica, Jian Li, and Hao He, "Spectral Analysis of Nonuniformly Sampled Data: A New Approach Versus the Periodogram," vol. 57, no. 3, pp. 843–858, March 2009.

[19] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source Localization and Sensing: A Nonparametric Iterative Approach Based on Weighted Least Squares," vol. 46, no. 1, pp. 425–443, January 2010.

[20] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse Learning via Iterative Minimization With Application to MIMO Radar Imaging," vol. 59, no. 3, pp. 1088–1101, March 2011.

[21] P. Stoica, P. Babu, and J. Li, "SPICE : a novel covariance-based sparse estimation method for array processing," vol. 59, no. 2, pp. 629 –638, Feb. 2011.

[22] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, July 2012.

[23] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," vol. 59, no. 5, pp. 2182 –2195, May 2011.

[24] P. Stoica and P. Babu, "Sparse Estimation of Spectral Lines: Grid Selection Problems and Their Solutions," vol. 60, no. 2, pp. 962–967, Feb. 2012.

[25] S. I. Adalbjörnsson and A. Jakobsson, "Sparse Estimation of Spectroscopic Signals," in *19th European Signal Processing Conference, EUSIPCO 2011*, Barcelona, Spain, 2011.

[26] S. Sahnoun, E. Djermoune, C. Soussen, and D. Brie, "Sparse multidimensional modal analysis using a multigrid dictionary refinement," *EURASIP J. Applied SP*, vol. 60, pp. 1–10, 2012.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[28] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

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

[30] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[31] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," vol. 14, no. 10, pp. 707–710, Oct. 2007.

[32] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[33] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, pp. 129–159, 2001.

[34] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, April 2004.

[35] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds*, Dover Publications, Inc., New York, 1968.

[36] R. A. Horn and C. A. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, England, 1991.

[37] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 4th edition, 2013.

[38] C. Tadonki and B. Philippe, "Parallel numerical linear algebra," chapter Parallel Multiplication of a Vector by a Kronecker Product of Matrices, pp. 71–89. Nova Science Publishers, Inc., Commack, NY, USA, 2001.

[39] J. Li and P. Stoica, "Efficient Mixed-Spectrum Estimation with Applications to Target Feature Extraction," vol. 44, no. 2, pp. 281–295, February 1996.

[40] H. C. So, F. Chanand W. H. Lau, and C. Chan, "An efficient approach for two-dimensional parameter estimation of a single-tone," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 1999–2009, April 2010.

[41] E. Gudmundson, Jun Ling, P. Stoica, Jian Li, and A. Jakobsson, "Spectral Estimation of Damped Sinusoids in the Case of Irregularly Sampled Data," in *Proceedings of the 9th International Symposium on Signals, Circuits and Systems (ISSCS 2009)*, Iasi, Romania, July 9-10 2009.

[42] A. Månsson, A. Jakobsson, and M. Akke, "Multidimensional Cramer-Rao Lower Bound for Non-Uniformly Sampled NMR Signals," in *22nd European Signal Processing Conference*, Lisbon, Sept. 1-5 2014.

E

Paper E

Joint model-order and fundamental frequency estimation in the presence of inharmonicity

Naveed R. Butt[1], Stefan Ingi Adalbjörnsson[1],
Samuel D. Somasundaram[2], and Andreas Jakobsson[1]

[1] *Centre for Mathematical Sciences, Lund University, Lund, Sweden*
[2] *General Sonar Studies Group, Thales UK, Stockport, Cheshire, SK3 0XB, U.K.*

Abstract

Estimation of the fundamental frequency of a set of harmonically related sinusoids is an integral part of many signal processing algorithms with as diverse application as speech and audio signal processing and electrocardiography. Often, the harmonic structure may deviate from being exact multiples of the fundamental frequency, a phenomenon called *inharmonicity*, which if not properly accounted for will degrade the estimation performance. To address this problem, we develop a general robust fundamental frequency estimator that allows for a larger class of inharmonicities in the observed signal. We also propose a scheme to include the estimation of the often unknown number of harmonics in the signal. To this end, we incorporate the recently developed multi-dimensional covariance fitting approach by allowing the Fourier vector corresponding to each perturbed harmonic to lie within a small uncertainty hypersphere centered around its strictly harmonic counterpart. Within these hyperspheres, we find the best perturbed vectors fitting the covariance of the observed data. The proposed approach provides the estimate of the fundamental frequency in two steps, and, unlike other recent methods, involves only a single 1-D search over a range of candidate fundamental frequencies.

Key words: Fundamental frequency, inharmonicity, robust estimator, model-order estimation, multi-dimensional covariance fitting.

1 Introduction

The estimation of the fundamental frequency, or *pitch*, of a set of harmonically related sinusoids is an integral part of many signal processing algorithms. While these algorithms most commonly find application in speech and audio signal processing, they can, in principle, be applied to harmonically related signals appearing in other fields, such as electrocardiography (ECG) [1]. Most developed estimators assume that the harmonics are exact integer multiples of the fundamental frequency (see, e.g., [1–3] and references therein). However, this is not always the case, and the deviation of the higher frequencies from exact integer multiples of the fundamental frequency, a phenomenon called *inharmonicity*, is often observed in real-world signals. For instance, it is well known that inharmonicity arises in piano tones due to the stiffness in the piano strings [4]. Inharmonicity has also been considered in the modeling and coding of speech signals, and several different models of inharmonicity have been developed [5, 6], as, if not properly compensated for, the frequency deviations will lead to poor amplitude and pitch estimates [7]. To alleviate this problem, several *robust* fundamental frequency estimation algorithms have been proposed in the recent literature, allowing for inharmonicity in the observed signal. Most of these algorithms consider the scenario of stiff-stringed instruments where deviations from exact integer multiples of the fundamental frequency depend functionally on a single unknown stiffness parameter [8–11]. However, as discussed in [1, 7], and also elaborated upon below, a more general model that allows for random perturbations in the harmonics would lead to an estimator that covers a wider range of problems. Existing solutions, such as the maximum a posteriori (MAP) and subspace estimators presented in [1,7], suffer from requiring exhaustive grid searches, such that the estimates are formed based on searches close to the expected unperturbed harmonics. Clearly, such combinatorial grid search approaches would increasingly become computationally inefficient with increasing number of harmonics, or for signals containing multiple sources. An additional challenge in the fundamental frequency estimation problem is that the number of harmonics in the observed signal, or the model order, is not known a priori. To address these limitations, the main objective of this work is to develop a general robust fundamental-frequency and model-order estimator that does not require searches over individual perturbed harmonics. In this regard, we incorporate the recently developed multi-dimensional covariance-fitting (MDCF) approach from the beamforming literature [12] into the robust pitch estimation problem by allowing the Fourier vector corresponding to each

perturbed harmonic to lie within a small uncertainty hypersphere centered around its strictly harmonic counterpart. Within these hyperspheres, we find the best perturbed vectors fitting the covariance of the observed data. The proposed approach is more general than other recent robust methods such as [8–11] that deal only with simple parametric inharmonicity of the form in [4], and it avoids the exhaustive search approach of [1, 7]. We also note that the proposed approach is different from several other robust pitch estimators [13–16] that are robust to different kinds of noise or to missing data. In contrast, our work focuses on robustness to inharmonicity. Finally, we remark that the single-pitch approach developed here may be extended to include multi-pitch data, along the group-sparsity and the iterative relaxation-based ideas in [17] and [18].

The rest of the paper has been organized as follows. A review of the inharmonic signal model and some existing robust estimators is provided in Section 2. Following this review, a detailed derivation of the proposed robust covariance-fitting pitch estimator is given in Section 3. Section 4 covers the proposed scheme for the estimation of the model-order, and Section 5 gives the results of the numerical evaluation of the proposed estimator compared to the existing approaches. Finally, we provide our conclusions in Section 6.

2 Signal model and other estimators

Consider a harmonic signal with the fundamental frequency $\omega_0 > 0$, corrupted by an additive noise [1]

$$x(n) = \sum_{l=1}^{L} \alpha_l e^{in\omega_l} + e(n) \tag{1}$$

where $n = 0, \ldots, N-1$, L represents the number of harmonics, $\alpha_l = |\alpha_l| e^{i\angle\alpha_l}$ denotes the complex amplitude of the lth harmonic, and $e(n)$ is a zero-mean white complex circularly symmetric Gaussian noise process with unknown variance σ_e^2. The harmonic frequencies, ω_l, are often formed as $\omega_l = \omega_0 l$, where ω_0 denotes the fundamental frequency. As an alternative, the harmonic frequencies for e.g. a piano have been modelled as [4]

$$\omega_l(\omega_0, B) = l\omega_0\sqrt{1 + l^2 B} \tag{2}$$

where $B \ll 1$ is an unknown positive string stiffness parameter. The main problem with such parametric models is that they are instrument dependent and one

may have to consider many such models to develop an estimator that can be applicable to a wide range of pitch estimations problems. Additionally, in many audio signal processing problems, the inharmonicities may not be so well-behaved. To avoid such limitations, we will here consider the more general model used in [1], extending (1) to allow small independent perturbations in the harmonics, such that

$$x(n) = \sum_{l=1}^{L} \alpha_l e^{i\omega_l(\omega_0, \Delta_l)n} + e(n) \tag{3}$$

where $\omega_l(\omega_0, \Delta_l) = \omega_0 l + \Delta_l$, with Δ_l representing a perturbation of the l-th harmonic. Different from earlier works, we will herein not assume a priori knowledge of the number of harmonics, L. It is often a difficult problem to form reliable model-order estimates, and many methods suffer noticeable performance degradation in case of inaccurate knowledge of an assumed model order. We assume, without loss of generality, that the perturbations are normally distributed zero-mean random variables with unknown but small variances, $\sigma^2_{\Delta_l}$. Among the pitch estimation algorithms available in literature, the maximum likelihood (ML) estimator offers a very powerful tool for estimating the fundamental frequency of a perfectly harmonic signal. It is known to be computationally efficient, and reduces to the optimal nonlinear least squares (NLS) estimator in case of white noise [1]. A robust version of the ML estimator, that allows for parametric inharmonicity of the form (2) has been presented in [1]. The algorithm is, however, computationally inefficient as it requires a 2-D search over ω_0 and B. Two of the relatively recent approaches that cover the general inharmonicity model in (3) are the MAP method of [1] and the subspace-based method of [7]. The MAP approach estimates the fundamental frequency and the perturbations by maximizing the posterior likelihood of observing the measured data under an assumed prior on the distribution of the perturbations. The subspace-based method [7], on the other hand, exploits a MUSIC-like approach to estimate the perturbed frequencies. However, both methods form the estimates based on searches over the parameters $(\omega_0, \{\Delta_l\})$, and require reliable a priori information of the model order, L.

3 Proposed robust covariance-fitting pitch estimator

In this section, we present a detailed derivation of the proposed robust estimator. For the sake of simplicity of presentation of the main idea, and without loss of generality, we initially derive the proposed estimator for a known model order. A scheme to include model-order selection along with pitch estimation will then be discussed in the following section.

We begin by defining

$$\mathbf{x}(n) = \left[\begin{array}{cccc} x(n) & x(n-1) & \dots & x(n-M+1) \end{array}\right]^T \tag{4}$$

$$\mathbf{A}_\Delta = \left[\begin{array}{ccc} \mathbf{a}_M(\omega_0+\Delta_1) & \dots & \mathbf{a}_M(\omega_0 L+\Delta_L) \end{array}\right] \tag{5}$$

where $(\cdot)^T$ denotes the transpose, for $M < N$, with

$$\mathbf{a}_M(\omega) = \left[\begin{array}{cccc} 1 & e^{-i\omega} & \dots & e^{-i\omega(M-1)} \end{array}\right]^T \tag{6}$$

Note that $\mathbf{A}_\Delta$ is full-rank if $\omega_0 l + \Delta_l \neq \omega_0 m + \Delta_m$, $\forall\ l \neq m$. The covariance matrix of (3) can then be written as

$$\mathbf{R} = \mathrm{E}\{\mathbf{x}(n)\mathbf{x}^*(n)\} = \mathbf{A}_\Delta \mathbf{P} \mathbf{A}_\Delta^* + \sigma_e^2 \mathbf{I} \tag{7}$$

where $(\cdot)^*$ represents the Hermitian transpose, and

$$\mathbf{P} = \mathrm{diag}\{\left[\begin{array}{ccc} |\alpha_1|^2 & \dots & |\alpha_L|^2 \end{array}\right]\} \tag{8}$$

In order to utilize the powerful optimal filtering methods discussed in [1], the performance of which critically depends on knowing the correct frequency of each Fourier vector, we here propose to allow each perturbed Fourier vector $\mathbf{a}_M(\omega_0 l + \Delta_l)$ to lie within a small uncertainty hypersphere centered around its strictly harmonic counterpart $\mathbf{a}_M(\omega_0 l)$. Note that this relaxation of the model, from the parametric uncertainty to an uncertainty set for the entire vector, is made to avoid the search over all the perturbations $\{\Delta_l\}$ simultaneously, which would require an exponentially increasing number of grid points as the number of harmonics increases. Defining the nominal Fourier matrix

$$\mathbf{A} = \left[\begin{array}{ccc} \mathbf{a}_M(\omega_0) & \dots & \mathbf{a}_M(\omega_0 L) \end{array}\right] \tag{9}$$

the set of constraints on the L Fourier vectors may be written compactly as

$$\|(\mathbf{A}_\Delta - \mathbf{A})\mathbf{e}_l\|_2 \leq \varepsilon_l, \quad l = 1, \dots, L \tag{10}$$

where the radius, ε_l, of the lth uncertainty hypersphere is a user parameter reflecting on the expected level of inharmonicity, and where $\mathbf{e}_l$ is the l-th column vector of an $L \times L$ identity matrix. We thus seek a perturbation of all the Fourier vectors, each found within the given sphere, such that the resulting optimal filtering method will have a better performance than if one simply used the assumption of harmonicity. This is similar to the robust adaptive beamforming problem in array signal processing, where one seeks to create a beamformer (filter) for the case when one has uncertain knowledge on the signals impinging on the array (see, e.g., [19] for a study examining a variation of our problem, although with only uncertainty in one vector). In [12], it is shown that if there is uncertainty in several of vectors it will have a detrimental effect on the beamformer, if not properly accounted for. In the work, the authors also introduce the idea of the MDCF to counteract this problem. Building on this idea, we employ the MDCF concept to formulate the problem of simultaneously finding the perturbed Fourier vectors as the solution of the optimization problem

$$\begin{aligned} \max_{\mathbf{A}_\Delta, \mathbf{P}, \sigma_e^2 \geq 0} \quad & \log\det(\mathbf{A}_\Delta \mathbf{P} \mathbf{A}_\Delta^* + \sigma_e^2 \mathbf{I}) \\ \text{s.t.} \quad & \mathbf{A}_\Delta \mathbf{P} \mathbf{A}_\Delta^* + \sigma_e^2 \mathbf{I} \preceq \hat{\mathbf{R}} \\ & \|(\mathbf{A}_\Delta - \mathbf{A})\mathbf{e}_l\|_2 \leq \varepsilon_l, \quad l = 1, \ldots, L \\ & \mathbf{P} = \mathbf{P} \odot \mathbf{I}_L \succeq \mathbf{0} \end{aligned} \tag{11}$$

where $\mathbf{A} \preceq \mathbf{B}$ denotes that $\mathbf{B} - \mathbf{A}$ is positive semidefinite, $\odot$ is the element-wise matrix product, and the last constraint ensures that, in accordance with the definition in (8), $\mathbf{P}$ is positive semidefinite and diagonal, and $\hat{\mathbf{R}}$ is the sample covariance matrix, given as

$$\hat{\mathbf{R}} = \frac{1}{N - M + 1} \sum_{n=0}^{N-M} \mathbf{x}(n)\mathbf{x}^*(n) \tag{12}$$

Additionally, we assume that the frequency vector uncertainty sets are sufficiently separated from each other to ensure pairwise linear independence between the columns of $\mathbf{A}_\Delta$. As shown in [12], (11) may not be amenable to a standard numerical solution, and one may instead use semidefinite programming (SDP) to solve a local convex approximation of (11) as

$$\max_{\breve{\mathbf{A}}_\Delta, \sigma_e^2 \geq 0} \quad 2\mathcal{R}\left\{ \mathrm{tr}\{\breve{\mathbf{A}}^* \mathbf{R}_0^{-1} \breve{\mathbf{A}}_\Delta\} \right\} + \mathrm{tr}\{\mathbf{R}_0^{-1}\}\sigma_e^2 \tag{13}$$

$$\begin{aligned} \text{s.t.} \quad & \breve{\mathbf{A}}_{\Delta}\breve{\mathbf{A}}_{\Delta}^{*} + \sigma_e^2\mathbf{I} \preceq \hat{\mathbf{R}} \\ & \mathcal{R}\{(\mathbf{A}\mathbf{e}_l)^{*}\breve{\mathbf{A}}_{\Delta}\mathbf{e}_l\} \geq \nu_l\|\breve{\mathbf{A}}\mathbf{e}_l\|_2; \;\; l = 1,\ldots,L \\ & \mathcal{I}\{(\mathbf{A}\mathbf{e}_l)^{*}\breve{\mathbf{A}}_{\Delta}\mathbf{e}_l\} = 0; \;\; l = 1,\ldots,L \end{aligned}$$

where $\mathcal{R}\{\cdot\}$ and $\mathcal{I}\{\cdot\}$ denote the real and imaginary parts of a complex number, respectively,

$$\nu_l = \sqrt{\|\mathbf{A}\mathbf{e}_l\|_2^2 - \varepsilon_l^2} \tag{14}$$

$$\mathbf{R}_0 = \breve{\mathbf{A}}\breve{\mathbf{A}}^{*} + \sigma_0^2\mathbf{I} \tag{15}$$

with

$$\breve{\mathbf{A}} = \mathbf{A}\mathbf{P}_0^{\frac{1}{2}} \tag{16}$$

where $\mathbf{P}_0$ denotes an initial estimate of $\mathbf{P}$ obtained through any suitable spectral estimator, and σ_0^2 is formed by averaging the $M-L$ smallest eigenvalues of $\hat{\mathbf{R}}$. There are several reasons why this formulation cannot be directly used to estimate the perturbed frequencies $\{\omega_l\}$. Firstly, as can be seen from (6), the true Fourier vectors must satisfy

$$\overline{\mathbf{a}}_M(\omega)e^{-i\omega} = \underline{\mathbf{a}}_M(\omega) \tag{17}$$

where $\overline{\mathbf{a}}_M(\omega)$ and $\underline{\mathbf{a}}_M(\omega)$ are formed by taking, respectively, the first $M-1$ and the last $M-1$ elements of the vector $\mathbf{a}_M(\omega)$. However, the formulation in (13) imposes no such constraint on the structure of $\breve{\mathbf{A}}_{\Delta}$. Secondly, we note that, by virtue of (16), an estimated $\breve{\mathbf{A}}_{\Delta}$ would include estimates of the amplitudes of the harmonics. Thus, the cost function of (13) is not suitable for a grid search over the fundamental frequency as it may wrongly compensate for the frequency perturbations by adjusting the estimates of the amplitudes and the noise variance, σ_e^2. To address these issues for the robust pitch estimation problem, we propose the following two-step approach that can be applied over a very coarse grid of fundamental frequencies. We term the proposed two-step approach the robust covariance-fitting pitch (RCP) estimator.

3.1 Step one: coarse estimates

The main objective of the first step is to obtain an initial estimate of the perturbed matrix, $\mathbf{A}_{\Delta}$. This estimate will then be used as the *assumed* matrix in the second

step, and is formed using a single 1-D grid search over a range of fundamental frequencies. It is worth noting both that the search grid can be chosen to be rather coarse, and that the estimate may be formed without any search over the individual perturbations. The estimate is formed as:

(i) Form a grid of appropriate size, say K, over the expected range of fundamental frequencies, and choose a frequency point from the grid, say ω_0^k, and, assuming this to be the fundamental frequency, form the matrix $\mathbf{A}$ using (9), and $\mathbf{P}_0$ by computing the periodogram estimates at ω_0^k and its perfect harmonics. Using the evaluated $\mathbf{A}$ and $\mathbf{P}_0$, solve the SDP in (13) to get an initial estimate of $\breve{\mathbf{A}}_\Delta$.

(ii) In line with the discussion under (16), the perturbed harmonics are extracted from the estimated $\breve{\mathbf{A}}_\Delta$ by imposing the suggested structural constraint on its columns. More specifically, denoting the l-th column of the estimated $\breve{\mathbf{A}}_\Delta$ as $\mathbf{b}_l$, and noting that, to be a true Fourier vector for the l-th harmonic, it must satisfy $\overline{\mathbf{b}}_l \gamma_l = \underline{\mathbf{b}}_l$, where $\gamma_l = e^{-i\omega_l}$, and where $\overline{\mathbf{b}}_l$ and $\underline{\mathbf{b}}_l$ are defined similar to $\overline{\mathbf{a}}_M(\omega)$ and $\underline{\mathbf{a}}_M(\omega)$, respectively, form an estimate of the l-th harmonic frequency as $\hat{\omega}_l = -\mathcal{I}\{\ln(\hat{\gamma}_l)\}$, with

$$\hat{\gamma}_l = \frac{\overline{\mathbf{b}}_l^* \underline{\mathbf{b}}_l}{\|\overline{\mathbf{b}}_l\|_2^2} \tag{18}$$

(iii) Form an improved estimate of $\mathbf{A}_\Delta$, say $\hat{\mathbf{A}}_\Delta$, by substituting the estimates $\{\hat{\omega}_l\}$ in (5). With the estimate $\hat{\mathbf{A}}_\Delta$ now available, the problem reduces to a standard pitch estimation problem. Therefore, we propose to utilize the cost function

$$g_k \triangleq \operatorname{tr}\left[\left(\hat{\mathbf{A}}_\Delta^* \hat{\mathbf{R}}^{-1} \hat{\mathbf{A}}_\Delta\right)^{-1}\right] \tag{19}$$

which represents the total output power of a set of L Capon filters, and is maximized at the true perturbed frequencies (for details, see, e.g., [1, 20]).

(iv) Repeat (i)-(iii) for the K points in the grid, and choose $\{\hat{\omega}_l^{max}\}$ as the L estimates where $\{g_k\}_{k=1}^K$ is maximized.

3.2 Step two: refined estimates

While it is possible to use $\{\hat{\omega}_l^{max}\}$, obtained in the previous step, one may refine the estimates of the perturbed frequencies further by solving (13) with the following improved initializations. Firstly, in place of $\mathbf{A}$, $\hat{\mathbf{A}}_{\Delta}^{max}$ is used as the *assumed* Fourier matrix, where $\hat{\mathbf{A}}_{\Delta}^{max}$ is formed by substituting $\{\hat{\omega}_l^{max}\}$ in (5). Secondly, to give better initial estimates of the amplitudes of the harmonics, $\mathbf{P}_0$ should be formed by computing the periodogram amplitudes at $\{\hat{\omega}_l^{max}\}$. These two modifications together assure a better initialization for the SDP problem in (13), leading therefore to more accurate frequency estimates, which can be formed as in operation (ii) of the first step. For later use, we represent the final refined frequency estimates as $\{\hat{\hat{\omega}}_l^{max}\}$. We also note that in our numerical studies, further iterations did not yield any improvements, leading us to conclude that only a single refinement step was sufficient.

3.3 Selection of ε_l

Noting that the left side of (10) may be written as the summation

$$\sum_{m=1}^{M} \sqrt{2(1-\cos(\Delta_l m))} \tag{20}$$

one may give a rough range for the selection of ε_l, such that it does not violate (10), as $0 \leq \varepsilon_l \leq 2\sqrt{M}$. Practical experience shows that in order to restrict Δ_l to be very small (which is typically the case), one should choose $\varepsilon_l \leq \sqrt{M}/3$. Secondly, one should use a smaller ε_l in the second step of RCP as compared to the value used in the first step. This is because $\hat{\mathbf{A}}_{\Delta}^{max}$ is expected to be closer to the true value of $\mathbf{A}_{\Delta}$, as compared to $\mathbf{A}$ (which is used as the *assumed* matrix in the first step).

4 Model-order selection

In the previous section, we have considered the number of harmonics, L, to be exactly known. It is, however, possible that the model-order is known to lie within a small range. In such situations, it would be beneficial to include the model-order estimation within the fundamental-frequency estimation problem. We now discuss this aspect in context of the proposed robust pitch estimator. Define a set

of candidate models $\mathcal{M}_m$, $m \in \mathsf{Z}_q$, where

$$\mathsf{Z}_q = \{0, 1, \ldots, q-1\} \tag{21}$$

is the set of candidate model indices. One possible approach may then be to choose the model that maximizes the *a posteriori* probability of the model given the observation $\mathbf{x}$, i.e., we choose

$$\hat{\mathcal{M}} = \arg \max_{\mathcal{M}_m, m \in \mathsf{Z}_q} f(\mathcal{M}_m|\mathbf{x}) \tag{22}$$

$$= \arg \max_{\mathcal{M}_m, m \in \mathsf{Z}_q} \frac{f(\mathbf{x}|\mathcal{M}_m) f(\mathcal{M}_m)}{f(\mathbf{x})} \tag{23}$$

Assigning equal probabilities to all the models in the set, i.e., setting

$$f(\mathcal{M}_m) = \frac{1}{q}, \quad m \in \mathsf{Z}_q \tag{24}$$

leads to

$$\hat{\mathcal{M}} = \arg \max_{\mathcal{M}_m, m \in \mathsf{Z}_q} f(\mathbf{x}|\mathcal{M}_m) \tag{25}$$

Further, incorporating the dependency of the models on unknown parameters such as amplitudes, frequencies, and phases, one may rewrite (23) as

$$\hat{\mathcal{M}} = \int_{\boldsymbol{\Theta}} \arg \max_{\mathcal{M}_m, m \in \mathsf{Z}_q} f(\mathbf{x}|\boldsymbol{\vartheta}, \mathcal{M}_m) f(\boldsymbol{\vartheta}|\mathcal{M}_m) d\boldsymbol{\vartheta} \tag{26}$$

where all the unknown parameters have been gathered in the vector $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. It is not generally possible to obtain any simple analytic expression for the integral in (26). However, an asymptotic solution (for large N) has been developed in the literature, leading to the solution (interested readers are referred to [21] for details)

$$\hat{\mathcal{M}} = \arg \min_{\mathcal{M}_m, m \in \mathsf{Z}_q} -2 \ln f(\mathbf{x}|\hat{\boldsymbol{\vartheta}}, \mathcal{M}_m) + (5L+1) \ln N \tag{27}$$

where $\hat{\boldsymbol{\vartheta}}$ is an estimate of the unknown parameters. For the case of additive white complex Gaussian noise with variance σ_e^2, the first term in (27), i.e., the log-likelihood function, is equal to $N \ln \sigma_e^2$. Replacing σ_e^2 by an estimate of the noise

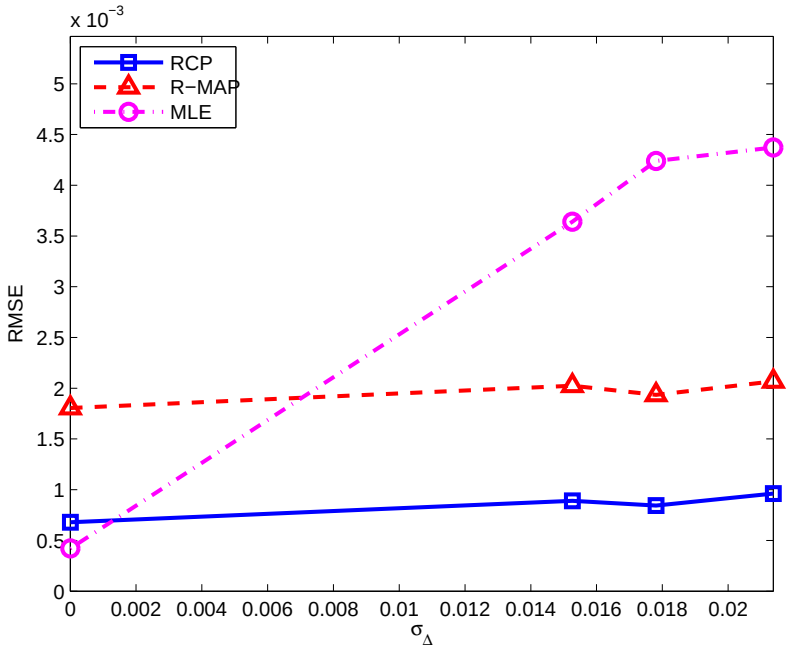

Figure 1: RMSE of the fundamental frequency estimates against the level of inharmonicity, for $\omega_0 = 0.2137$, at SNR level of 5 dB.

power for each candidate model order L, denoted $\hat{\sigma}_e^2(L)$, we may thus write an expression for the model-order estimate as

$$\hat{L} = \arg\min_L 2N \ln \hat{\sigma}_e^2(L) + (5L+1)\ln N \tag{28}$$

It remains now to obtain $\hat{\sigma}_e^2(L)$. For any method that estimates a noise-free signal-of-interest (SOI) for a pre-selected order, L, the noise power estimate may be written as the difference between the total measured power and power of the estimated SOI. For the proposed RCP estimator, the estimated power for a model-order L may therefore be given as (see discussion under (19))

$$\mathrm{tr}\left[\left((\hat{\hat{\mathbf{A}}}_{\Delta}^{max})^* \hat{\mathbf{R}}^{-1} \hat{\hat{\mathbf{A}}}_{\Delta}^{max}\right)^{-1}\right] \tag{29}$$

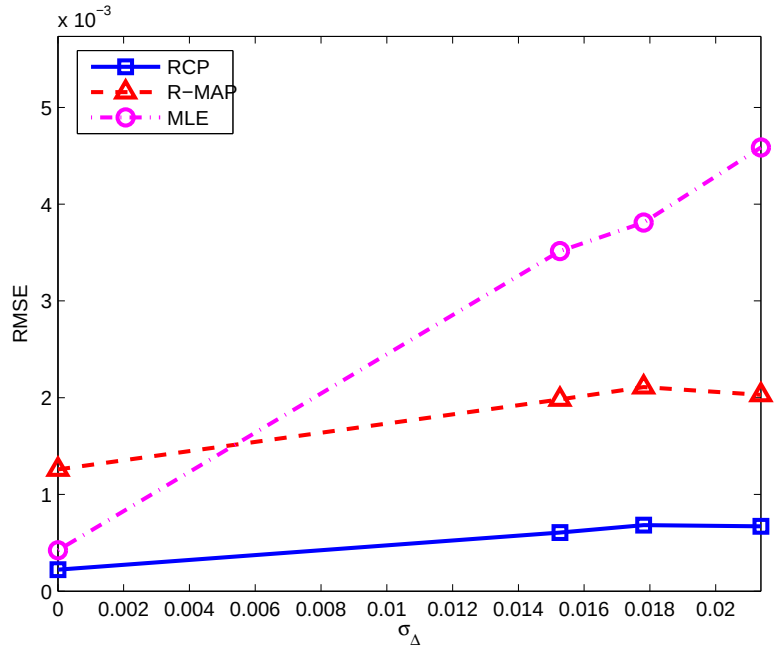

Figure 2: RMSE of the fundamental frequency estimates against the level of inharmonicity, for $\omega_0 = 0.2137$, at SNR level of 30 dB.

where $\hat{\hat{\mathbf{A}}}_{\Delta}^{max}$ is formed by substituting $\{\hat{\hat{\omega}}_l^{max}\}$ obtained in Section 3 into (5). Since an estimate of the power of the measurement $\mathbf{x}(n)$ may be given as $\frac{1}{M}\mathrm{tr}(\hat{\mathbf{R}})$, we can get an expression for $\hat{\sigma}_e^2(L)$ for the proposed approach as

$$\hat{\sigma}_e^2(L) = \frac{1}{M}\mathrm{tr}(\hat{\mathbf{R}}) - \frac{1}{M}\mathrm{tr}\left[\left((\hat{\hat{\mathbf{A}}}_{\Delta}^{max})^*\hat{\mathbf{R}}^{-1}\hat{\hat{\mathbf{A}}}_{\Delta}^{max}\right)^{-1}\right] \tag{30}$$

The estimation of the model order may therefore be included in the proposed RCP estimator as follows.

(i) Choose a set of candidate model orders.

(ii) For each model-order candidate, say L, estimate the noise power $\hat{\sigma}_e^2(L)$ using (30).

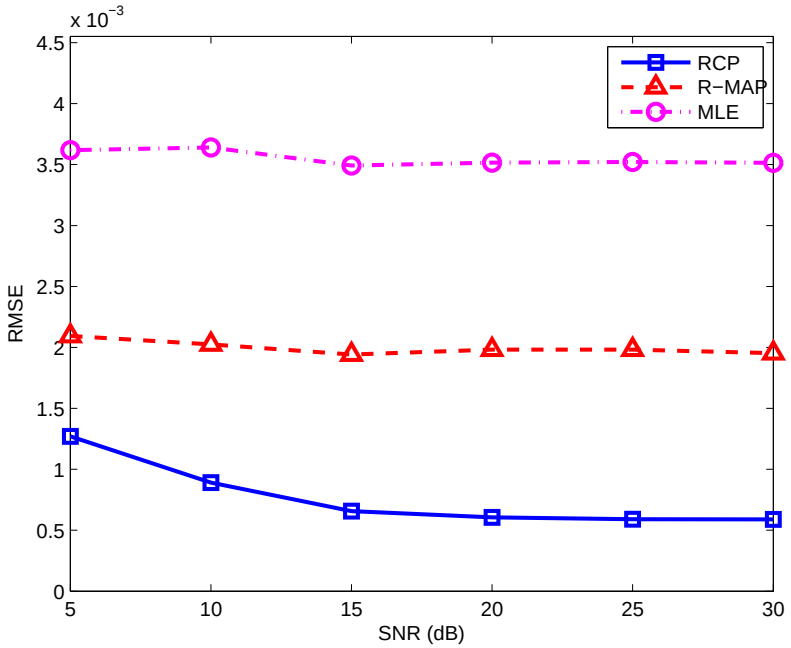

Figure 3: RMSE of the fundamental frequency estimates against SNR, for $\omega_0 = 0.2137$ and $\sigma_\Delta = \omega_0/14$.

(iii) Use $\hat{\sigma}_e^2(L)$ in (28) to choose the optimal model order among the candidates.

(iv) Choose the RCP frequency estimates corresponding to the optimal model order as the final estimates.

5 Simulations and results

We proceed to numerically evaluate the performance of the proposed RCP estimator, comparing to the MLE [1] and the robust MAP (R-MAP) [7] estimators. The results are obtained through a number of experiments based on Monte Carlo simulations using synthetic signals. In each case, the synthetic signal was generated using (3), with $L = 4$ harmonics having unit amplitudes and uniformly distributed phases that are randomized in each Monte Carlo run. The experi-

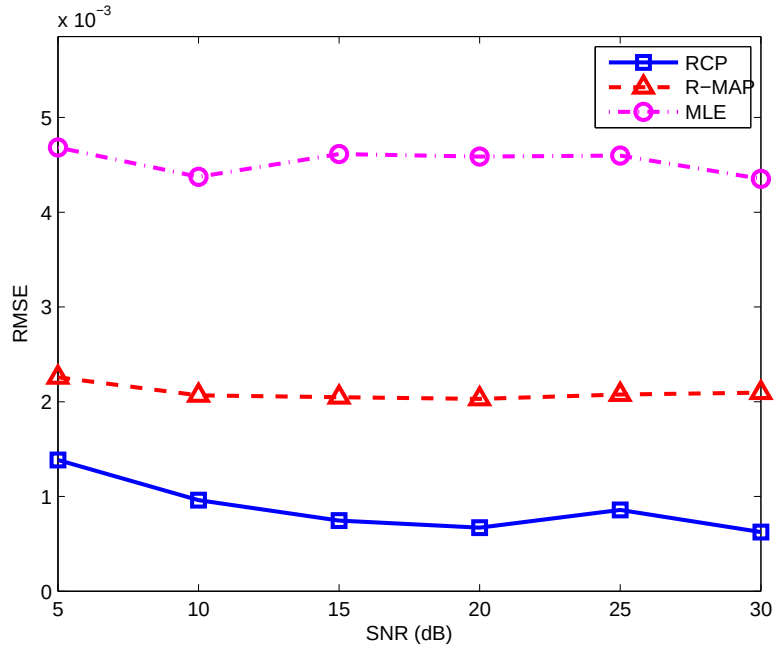

Figure 4: RMSE of the fundamental frequency estimates against SNR, for $\omega_0 = 0.2137$ and $\sigma_\Delta = \omega_0/10$.

ments were repeated for several different fundamental frequencies, and for five different signal-to-noise ratio (SNR) levels from $5 - 30$ dB, where the SNR is defined as $10\log_{10}(\mathrm{tr}(\mathbf{P})/\sigma_e^2)$. All algorithms were tested at different levels of inharmonicity by increasing the standard deviation of the perturbations, σ_Δ, from 0 to $\omega_0/10$, where a variance of 0 indicates a perfectly harmonic signal. A total of $J = 150$ Monte Carlo simulations were used in each experiment to evaluate the root mean square error (RMSE), defined for the fundamental frequency estimates as

$$RMSE = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\hat{\omega}_{0,j} - \omega_0)^2} \tag{31}$$

where ω_0 and $\hat{\omega}_{0,j}$ represent the true fundamental frequency and the estimated fundamental frequency in the j-th Monte Carlo run, respectively. A data length

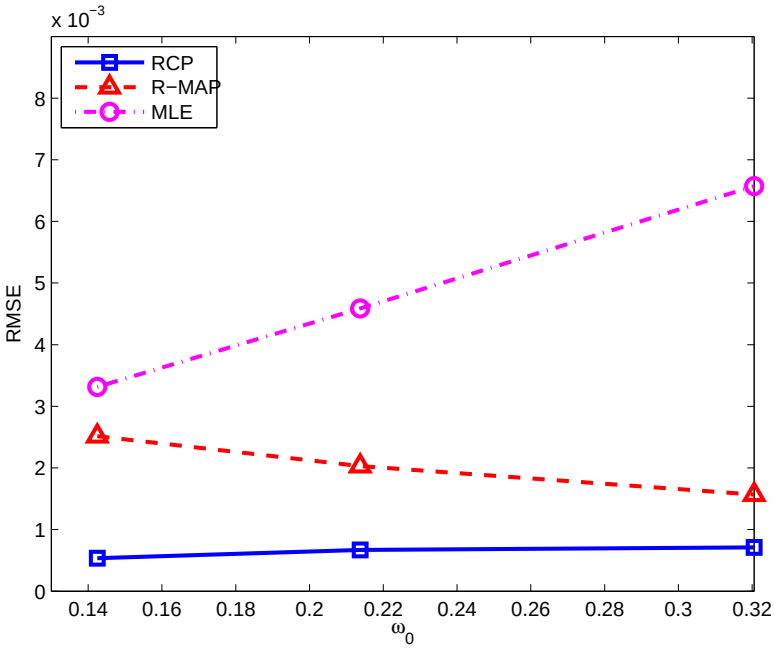

Figure 5: RMSE at $\omega_0 = 0.1425, 0.2137, 0.3206$, for SNR = 20 dB and $\sigma_\Delta = \omega_0/10$.

of $N = 200$ samples was used, while the sub-vector length for RCP was set to $M = 50$, which is in accordance with the limit, $M \leq N/2$, suggested in filtering literature (see, e.g., [1] and [20]). For this set of tests, the model order was assumed to be known. Typical results, comparing the proposed RCP estimator to the standard MLE and the R-MAP estimators, are shown in Figures 1-6. Following the guidelines in Section 3.3, all the results were obtained with the uncertainty parameter ε_l set to 4 for the first step and to 2 for the second step of RCP. The fundamental frequency search grid for MLE and R-MAP consisted of 300 equally-spaced points in the range [0.05, 0.5], whereas for the proposed RCP estimator, the grid consisted of only 30 equally-spaced points in the same range. To make the comparison fair, we here allow all the estimators knowledge of the true model order, L. Figures 1 and 4 show the RMSE of the fundamental frequency estimates against the level of inharmonicity for $\omega_0 = 0.2137$ at SNR levels of

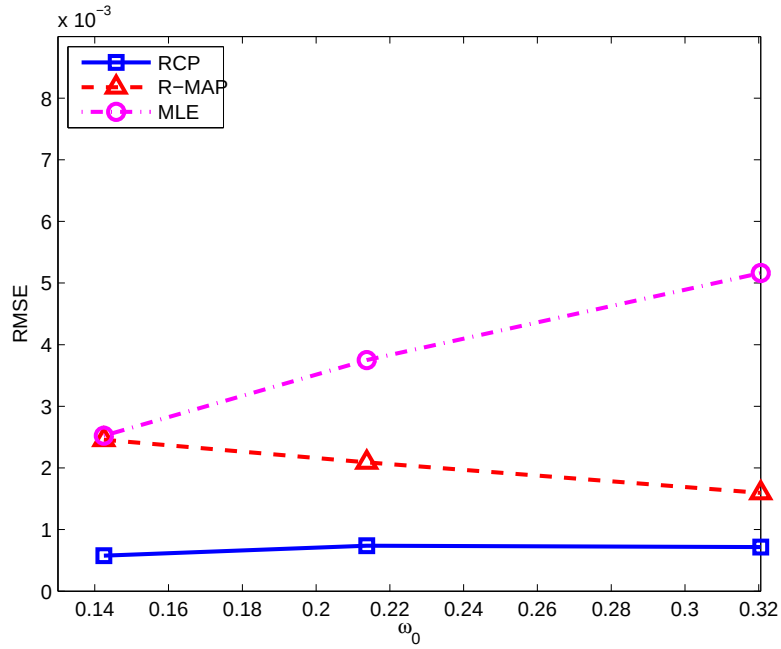

Figure 6: RMSE at $\omega_0 = 0.1425, 0.2137, 0.3206$, for SNR = 15 dB and $\sigma_\Delta = \omega_0/12$.

5 dB and 30 dB, respectively. As is clear from the figures, the proposed RCP estimator performs better at both low SNR and high SNR levels. As expected, the MLE method, not allowing for inharmonicity, suffers heavily with increase in inharmonicity. Figures 3 and 4 show the RMSE against SNR for $\omega_0 = 0.2137$ for σ_Δ equal to $\omega_0/14$ and $\omega_0/10$, respectively. We see that while the performance of all the estimators degrades slightly at lower SNRs, the RCP estimator provides more accurate estimates at all levels. Figures 5 and 6 show the RMSE at three different fundamental frequencies $\omega_0 = 0.1425, 0.2137$, and 0.3206, at SNR = 20 dB, $\sigma_\Delta = \omega_0/10$ and SNR =15 dB, $\sigma_\Delta = \omega_0/12$, respectively. We remark that the increase in the RMSE of MLE at the higher frequencies is because of a higher simulated inharmonicity at these frequencies. While both R-MAP and RCP show robustness to the inharmonicity, the proposed approach clearly provides more accurate estimates.

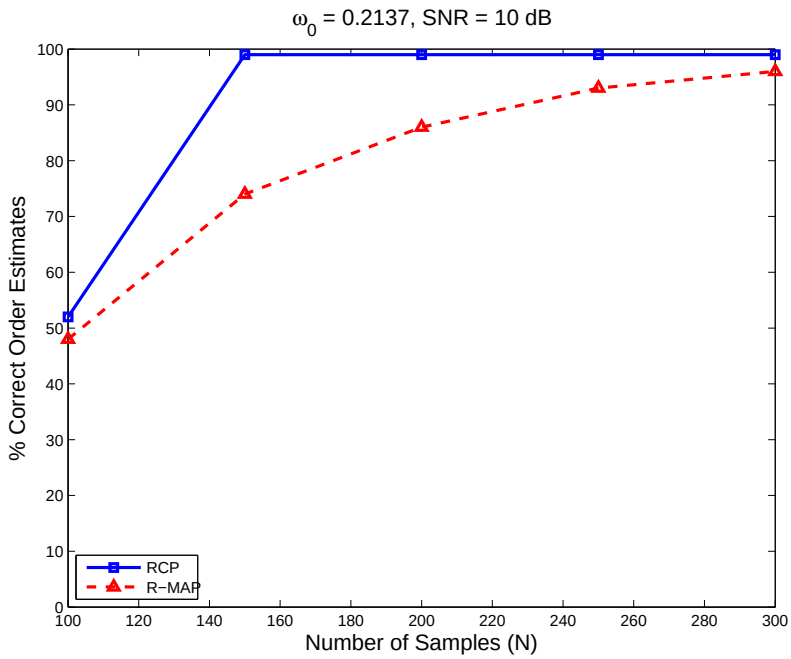

Figure 7: Percentage of correctly estimated model orders as a function of N with $\sigma_\Delta = \omega_0/10$.

In accordance with the discussion in Section 4, the ability of RCP to estimate the model order was also numerically studied. In these simulations, the candidate model orders were provided as $\{2 - 8\}$, and the results were compared to the R-MAP model-order estimates. Figure 7 shows the percentage of correctly estimated model orders as a function of the number of samples N. As expected, the model-order estimates improve with the increase in the number of available samples. Figure 8 shows percentage of correctly estimated model orders against a varying level of inharmonicity. While both RCP and R-MAP show similar and expected trends, the proposed RCP estimator clearly outperforms R-MAP in estimating the model order correctly.

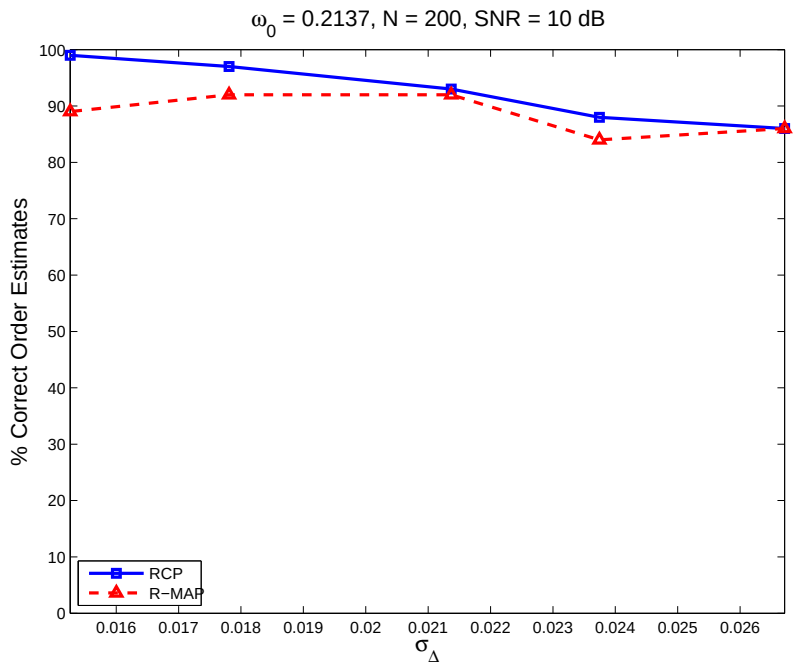

Figure 8: Percentage of correctly estimated model orders as a function of σ_Δ.

6 Conclusion

We have proposed a general robust fundamental-frequency and model-order estimator that allows for non-parametric inharmonicity in the observed signal. The proposed approach allows the Fourier vector corresponding to each perturbed harmonic to lie within a small uncertainty hypersphere centered around its strictly harmonic counterpart. Within these hyperspheres, we find the best perturbed vectors fitting the covariance of the observed data. The proposed approach provides the estimate of the fundamental frequency in two steps, and, unlike other recent methods, involves only a single 1-D search over a range of candidate fundamental frequencies. It is numerically shown to provide better fundamental frequency estimates than the MLE and R-MAP approaches under a variety of practical conditions covering various degrees of inharmonicity and SNR levels. The ability of the estimator to select the correct model-orders is also numerically shown.

References

[1] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, 2009.

[2] H. Kameoka, *Statistical Approach to Multipitch Analysis*, Ph.D. thesis, University of Tokyo, 2007.

[3] W. Hess, "Pitch and Voicing Determination," *Advances in Speech Signal Processing*, pp. 3–48, 1992.

[4] H. Fletcher, "Normal vibration frequencies of stiff piano string," *Journal of the Acoustical Society of America*, vol. 36, no. 1, 1962.

[5] T. D. Rossing, *The Science of Sound*, Addison-Wesley Publishing Co., 2 edition, 1990.

[6] E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," vol. 5, no. 5, pp. 389–406, Sep 1997.

[7] M. G. Christensen, P. Vera-Candeas, S. D. Somasundaram, and A. Jakobsson, "Robust Subspace-based Fundamental Frequency Estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, March 30-April 4, 2008.

[8] I. Barbancho, L. J. Tardon, S. Sammartino, and A. M. Barbancho, "Inharmonicity-Based Method for the Automatic Generation of Guitar Tablature," vol. 20, no. 6, pp. 1857–1868, Aug. 2012.

[9] E. Benetos and S. Dixon, "Joint Multi-Pitch Detection Using Harmonic Envelope Estimation for Polyphonic Music Transcription," vol. 5, no. 6, pp. 1111–1123, Oct. 2011.

[10] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "A Robust and Computationally Efficient Subspace-Based Fundamental Frequency Estimator," vol. 18, no. 3, pp. 487–497, March 2010.

[11] V. Emiya, R. Badeau, and B. David, "Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle," vol. 18, no. 6, pp. 1643–1654, August 2010.

[12] M. Rübsamen and A. B. Gershman, "Robust Adaptive Beamforming Using Multidimensional Covariance Fitting," vol. 60, no. 2, pp. 740–753, Feb. 2012.

[13] F. Huang and T. Lee, "Pitch Estimation in Noisy Speech Using Accumulated Peak Spectrum and Sparse Estimation Technique," vol. 21, no. 1, pp. 99–109, Jan. 2013.

[14] S. Gonzalez and M. Brookes, "A Pitch Estimation filter robust to high levels of noise (PEFAC)," in *19th European Signal Processing Conference (EUSIPCO)*, Barcelona, August 29 - September 2, 2011.

[15] J. A. Morales-Cordovilla, Ning Ma, V. Sanchez, J. L. Carmona, A. M. Peinado, and J. Barker, "A pitch based noise estimation technique for robust speech recognition with Missing Data," in *36th IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, May 22–27, 2011, pp. 4808–4811.

[16] J. O. Hong and P. J. Wolfe, "Robust and efficient pitch estimation using an iterative ARMA technique," in *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*, Makuhari, Japan, September 26–30, 2010.

[17] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Estimating Multiple Pitches Using Block Sparsity," in *Proc. 38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

[18] T. Nilsson, S. I. Adalbjörnsson, N. R. Butt, and A. Jakobsson, "Multi-Pitch Estimation of Inharmonic Signals," in *European Signal Processing Conference*, Marrakech, Sept. 9-13, 2013.

[19] J. Li, P. Stoica, and Z. Wang, "On Robust Capon Beamforming and Diagonal Loading," vol. 51, no. 7, pp. 1702–1715, July 2003.

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

[21] P. Stoica and Y. Selén, "Model-order Selection — A Review of Information Criterion Rules," vol. 21, no. 4, pp. 36–47, July 2004.