Parameter Estimation by Sparse Reconstruction with Wideband Dictionaries

MAKSIM BUTSENKO

TTÜ PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies
Thomas Johann Seebeck Department of Electronics
The dissertation was accepted for the defence of the degree of Doctor of Philosophy
in Electronics and Telecommunication on 21 November 2018

Supervisor: Prof. Olev Märtens,
                    Thomas Johann Seebeck Department of Electronics,
                    Tallinn University of Technology,
                    Tallinn, Estonia

Co-supervisor: Prof. Yannick Le Moullec,
                    Thomas Johann Seebeck Department of Electronics,
                    Tallinn University of Technology,
                    Tallinn, Estonia

Opponents: Prof. Andi Kivinukk,
                    School of Digital Technologies,
                    Tallinn University,
                    Tallinn, Estonia

                    Associate Prof. Thomas Arildsen,
                    Department of Electronic Systems,
                    Aalborg University,
                    Aalborg, Denmark

Defence of the thesis: 17 December 2018, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any academic degree elsewhere.

Maksim Butsenko   _____
                                    signature

Signaali parameetrite hindamine kasutades hajusat taastamist laiaribaliste sõnastikega

MAKSIM BUTSENKO

To my parents:
Nadežda and Viktor

Contents

List of publications

The work of this thesis is based on the following publications:

A   Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. Estimating Sparse Signals Using Integrated Wide-band Dictionaries in *Proc. of 42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, pp. 4426-4430, 5-9 Mar. 2017. [ETIS 3.1]

B   Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. Estimating Sparse Signals Using Integrated Wide-band Dictionaries in *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4170-4181, 2018. [ETIS 1.1]

C   Maksim Butsenko, Olev Märtens, Andrei Krivošei and Yannick Le Moullec. Sparse Reconstruction Method for Separating Cardiac and Respiratory Components from Electrical Bioimpedance Measurements in *Elektronika ir Elektrotechnika*, vol. 24, no. 5, pp. 57-61, 2018. [ETIS 1.1]

D   Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. The Zoomed Iterative Adaptive Approach. Accepted and to be presented at *2018 International Symposium on Intelligent Signal Processing and Communication Systems*. [ETIS 3.1]

In addition, the following publications where co-authored by the author of the thesis:

E   Maksim Butsenko and Tõnu Trump. An Affine Combination Of Adaptive Filters For Sparse Impulse Response Identification in *Proc. of 23rd Telecommunications Forum Telfor (TELFOR)*, Belgrade, Serbia, pp. 396−399, 24-26 Nov. 2015. [ETIS 3.1]

F   Maksim Butsenko and Tõnu Trump. An Affine Combination of Adaptive Filters for Channels with Different Sparsity Levels in *Telfor Journal*, vol. 8, no. 1, pp. 32−37, 2016. [ETIS 1.2]

G   Andrei Krivošei, Mart Min, Paul Annus and Maksim Butsenko. Decomposition of the EBI Signal into Components using two Channel Cross-Compensating Singular Spectrum Analysis in *13th Annual IEEE International Symposium on Medical Measurements & Applications*, Rome, Italy, 2018. [ETIS 3.1]

Author's contributions to the publications

The author's contribution to the publications in this thesis are:

A   In Publication A, a technique for reducing the size of the dictionary in sparse signal reconstruction by using novel idea of wideband dictionary is presented. The proposed approach is described and evaluated on one-dimensional data. The contribution of the author of thesis lies in the initial discussion over the idea, formulation and implementation of the proposed algorithm in MATLAB environment, designing and conducting the numerical experiments, evaluating the results and preparing the publication.

B   In Publication B, the discussion on the idea of wideband dictionary presented in a Publication A is expanded and analyzed further. The proposed approach is generalized for the multi-dimensional case and evaluated on two-dimensional data. The contribution of the author of thesis lies in the initial discussion over the idea, formulation and implementation of the proposed algorithm in MATLAB environment, designing and conducting the numerical experiments, evaluating the results and preparing the publication.

C   Publication C applies the methods described in Publications A and B to a different problem. Here a sparse reconstruction framework was used as a basis for blind source separation of the respiratory and cardiac components from electrical bioimpedance measurements. The author of this thesis formulated the problem and the idea for the solution, designed, conducted the experiments on the data and prepared the publication.

D   Publication D presents the idea of incorporating an iterative zooming procedure for the iterative adaptive approach (IAA) spectral estimator. It is usually complicated either by the computational

burden related to the use of large dictionaries or to the risk of poor conditioning. The proposed method allows for computationally efficient calculation of IAA spectra using wideband dictionary described in the Publications A and B. The contribution of the author of thesis lies in the initial discussion over the idea, formulation and implementation of the proposed algorithm in MATLAB environment, designing and conducting the numerical experiments, evaluating the results and preparing the publication.

List of abbreviations

ADMM   Alternative Direction Method of Multipliers
AR   Autoregressive
ARMA   Autoregressive-Moving Average
BIC   Bayesian Information Criterion
CS   Compressive Sensing
DFT   Discrete Fourier Transform
DPSS   Discrete Prolate Spheroid Sequences
EBI   Electrical Bio-Impedance
ECG   Electrocardiography
ESPRIT   Estimation of Signal Parameters by Rotational Invariance Techniques
FFT   Fast Fourier Transform
GPU   Graphics Processing Unit
IAA   Iterative Adaptive Approach
LASSO   Least Absolute Shrinkage and Selection Operator
LDA   Levinson-Durbin Algorithm
LS   Least-Squares
MA   Moving Average
MSE   Mean Squared Error
MUSIC   Multiple Signal Classification
NLS   Nonlinear Least Squares
NMR   Nuclear Magnetic Resonance
NUFFT   Non-Uniform Fast Fourier Transform
PSD   Power Spectral Density
RIP   Restricted Isometry Property
RMSE   Root Mean Squared Error
SNR   Signal-to-Noise Ratio
SPICE   Sparse Covariance-Based Estimation Method
WLS   Weighted Least Squares

List of notations

$\mathbf{a}, \mathbf{b}, ...$	boldface lowercase letters denote vectors
$\mathbf{A}, \mathbf{B}, ...$	boldface upper case letters denote matrices
$\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{I}}, ...$	calligraphic boldface upper case letters denote parameter sets
$A, a, ...$	nonbold letters denote scalars
$(\cdot)^T$	transpose
$(\cdot)^H$	Hermitian transpose
$\hat{(\cdot)}$	estimated parameter
$\{\cdot\}$	set of elements
$\lvert \cdot \rvert$	magnitude
$\lVert \cdot \rVert_0$	ℓ_0-"pseudonorm"
$\lVert \cdot \rVert_1$	ℓ_1-norm (Manhattan distance)
$\lVert \cdot \rVert_2$	ℓ_2-norm (Euclidean distance)
$\lVert \cdot \rVert_\infty$	ℓ_∞-norm (infinity norm)
$E[\cdot]$	expected value of a random variable
$\approx$	approximately equal to
$\in$	belongs to (a set)
$\odot$	element-wise multiplication
$\circ$	outer product
$\otimes$	Kronecker product
$\mathrm{vec}(\cdot)$	vectorization of a matrix or a tensor
$\mathrm{diag}(\cdot)$	diagonal matrix with specified diagonal elements
$\mathrm{argmin}(\cdot)$	argument that minimizes
$\mathbb{Z}$	the set of integers

Introduction

1. Motivation

This thesis concerns the field of sparse signal processing and algorithms which efficiently exploit sparsity to find a solution to the optimization problems which are not attainable otherwise. In this work, this concept is applied to the well-known problem of parameter estimation - recovering parameters of interest from limited measurements of an actual signal. Probably the most widely known method in this area is the computation of the signal periodogram through Fast Fourier Transform (FFT) in order to estimate the spectral density of the signal or the frequency content of the signal. In this case, the parameter of interest is the frequency. Although the problem and many solutions are well-known for decades, this area of research is still fairly active and many contributions to the field are being made each year. So far, the research community have not find a single estimator which can cover all the scenarios efficiently and the choice of the estimator is usually a trade-off between different properties of the estimator and constraints required by the application. For a wide range of applications the FFT is a suitable method; however, it is not considered to be a high-resolution method due to its resolution limit [1] and it also requires re-formulation of its fast computational algorithm if the application considers a non-uniformly sampled data, which is often the case in many different fields and applications; consider for example radar imaging [2], [3], astronomy [4], seismology [5] or genetics [6]. Although the re-formulation of the fast computation algorithm is certainly possible and is usually done through the interpolation of the oversampled FFT and referred to as non-uniform FFT (NUFFT) [7], [8], it does add a computational cost and still does not improve resolution limits. In addition to the applications mentioned before, one application that requires high-resolution, fast computational time and also capability of working with non-uniformly sampled data is multi-dimensional nuclear magnetic resonance (NMR) spectroscopy [9], [10] - a technique for determination of chemical structure and molecular interactions [11]. Considering the requirements that are posed by applications mentioned above, this thesis introduces a new method for high-resolution parameter estimation, which often shows considerable speed-up compared to the state of the art methods, can work with the non-uniform data directly and also mitigates problem of grid-mismatch inherent to many similar methods. Although this new estimator still can not solve
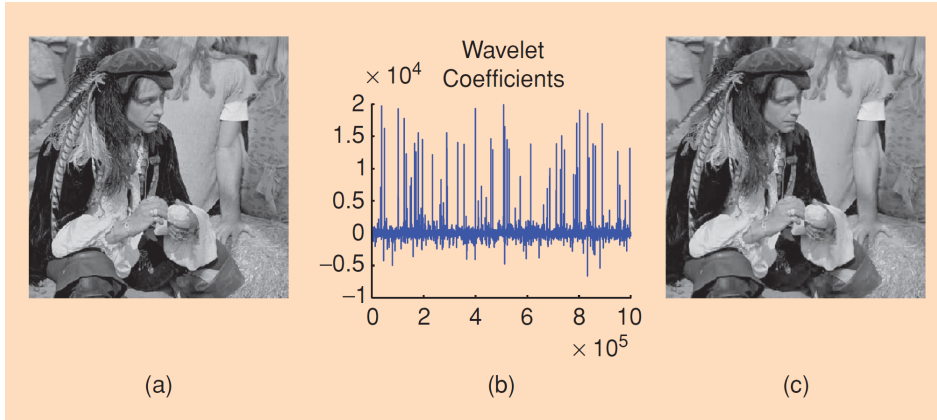
Figure 1 Original image (a), wavelet transform coefficients (b) and resulted reconstruction image (c) after removing 99% of the wavelet transform coefficients [12] ©2008 IEEE.

all the problems that arise in a reality, the author of this thesis hopes that this proposed tool might be useful for those working in the area of signal estimation, not only due to its decent and computationally fast performance, but also through its intuitive implementation and adaptability for the different estimation methods.

In the modern days it is evident that the amount of captured and stored data grows every day and there is an obvious need for the ability to transfer this data more quickly, store it in a more compact way or to process it faster. The concept of sparsity forms a base from which many algorithms and methods for tackling those problems can be proposed. Sparsity is a property of a signal where the investigated model has only a small number of nonzero components or where a signal of interest consists of samples which are mostly zeros or near-zeros. Although most of the signals are not directly sparse, there exist many useful transforms which can represent signal in a sparse way. Consider a simple example: a sinusoidal signal is not sparse in the time-domain as it requires many samples to be represented explicitly, but Fourier Transform helps to represent it in a sparse way in the frequency-domain, where such signal has only one relevant peak. Consider another example of an image which is transformed using Haar wavelet [13]. In the wavelet domain we see that only a very limited number of coefficients carry most of the energy and therefore the information in the image. Therefore, it can be said that the image is sparse in the wavelet domain. One can remove most of the smaller coefficients, apply Inverse Wavelet Transform and get a reconstructed version of the original image. As can be seen from a Figure 1, the difference between the original and the reconstructed images are hardly noticeable. As shown, sparsity is usually present in the signal; however, it may require additional transformation into the domain which represents the signal in a sparse way, be it a Fourier Transform, Wavelet Transform or any other kind of suitable mathematical representation.

2.   State of the art

In the recent years, the research community in the field of signal processing have been heavily involved with sparsity by applying methods to different problems, such as but not limited to sparse array processing [14]–[18], signal denoising [19]–[21], image processing [22]–[24], channel estimation [25], [26] and deriving mathematical proofs and theoretical limits for corresponding algorithms [12], [27]–[31]. However, even before the signal processing community fully embraced the concept of sparsity, statisticians were already aware of its benefits when Hastie et al [32] came up with the informal "bet on sparsity" principle, which states:

> "Use a procedure that does well in sparse problems, since no procedure does well in dense problems."

Such procedures in the signal processing world often take form of finding sparse representation of the signal at hand [21], [28], [31], [33]–[35]. Exploiting sparse represenation of the signal is often useful as many of the applications mentioned above result in discrete and peaky spectra. This is the perfect setup for the methods considering sparsity [36], [37] as resulted spectra is well described by just a few elements in appropriately chosen dictionary. However, if spectral resolution required by the application is high then the dictionary becomes restrictively large and employing such methods requires dealing with increased computational complexity. Therefore, the need for efficient computational methods for such problems arises and one can consider exploiting numerical structures of resulting dictionary matrices and transforms [38], [39]. Another possible direction looks at employing various screening methods for deciding which dictionary elements were present in the signal [40], [41] and therefore reducing considerably the size of the problem to solve. However, by employing dictionary-based methods one has to take into account the problem of off-grid estimation [42], [43] - estimated signal falling between two dictionary elements (off-grid). Few methods of dealing with such issue have been proposed, such as adaptive grid or atomic-norm minimization [44]–[46]. However, they still come with the additional trade-off of not having convex problem formulation for former and the requirement of solving computationally demanding optimization problem for latter method respectively.

This work considers discrete-time signals and assumes that they were appropriately sampled beforehand. A signal is presented as a complex-valued signal, which usually results in easier mathematical formulation and is generally preferred in the signal-processing literature. The purpose of the next sections in the Introduction is to present an overview of preliminaries and theoretical background necessary for understanding of the topic, also to introduce existing methods and their drawbacks. The following work is organized as follows. Section 3 gives an overview of two groups of parameter estimation algorithms with

description of some widely used algorithms. Section 4 introduces terminology and the main process of sparse reconstruction framework, which is the basis of the methods described in the publications. Section 5 introduces one particular algorithm of solving convex optimization problems, which is used widely in this work. Introduction is followed by main body of this work which consists of four publications. The main findings are then summarized in the Conclusion.

3. Parameter estimation

In signal processing as well as in other scientific fields we are often interested in estimating the value of an unknown parameter from a limited set of observations. As any estimate will depend on the observations we make, then the estimate itself is considered to be a random variable. Therefore, in order to evaluate the efficiency of the estimator, it is necessary to establish its statistical properties. Mainly we are interested in bias and variance of the estimator. Consider estimating the value of an unknown parameter θ from a sequence x_n, for $n = 1, 2, ..., N$. We are interested in acquiring the estimate, $\hat{\theta}$, which should be equal to the actual value of the parameter, at least in the average sense. The difference between the expected value of our estimate and the actual value θ is called the bias

$$Bias = \theta - E[\hat{\theta}] \tag{1}$$

If the bias is zero, then the expected value and true value are equal

$$E[\hat{\theta}] = \theta \tag{2}$$

and estimator is considered to be unbiased. If the bias is not zero, then the estimator is said to be biased. The estimator is considered to be asymptotically unbiased if the bias approaches zero, when N, number of observations, tends towards infinity

$$\lim_{N \to \infty} E[\hat{\theta}] = \theta. \tag{3}$$

Bias is an objective parameter of the estimator. In general, an unbiased or asymptotically unbiased estimator is preferred, however if other properties of the estimator are more important, such as low variance for example, one might prefer an estimator with small bias to the unbiased estimator. One has also consider the variance of the estimator, as it is a measure of dispersion of a random variable around its mean. Low variance means that observations are more closely concentrated around the mean of the parameter and therefore the estimate is more precise (in the case of unbiased estimator). A popular graphical representation of low/high bias and variance is showed in Figure 2. In order for the estimate to converge to the mean of the parameter, the variance of the unbiased estimator has to go to zero as the number of observations tends towards infinity

$$\lim_{N \to \infty} Var[\hat{\theta}] = \lim_{N \to \infty} E\left\{|\theta - E[\hat{\theta}]|^2\right\} = 0 \tag{4}$$
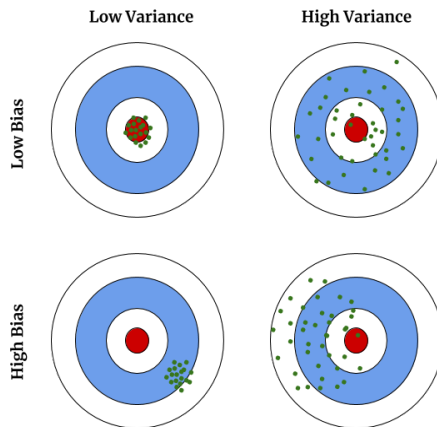
Figure 2 Graphical illustration of low-high bias and variance [49].

Another important property of an estimator is its robustness, which shows how well an estimator can maintain its estimation quality in a varying environment, i.e. when assumptions on the distribution changes. There are other properties (such as consistency, effeciency etc.) of the estimator that can be considered in the estimation theory; however, we refer interested readers to the classical books by Stoica [1], Kay [47] and Hayes [48] for more details.

3.1. Non-parametric estimators

Non-parametric estimators do not pose any prior assumption on the signal and therefore are quite robust tools which work reasonably well in many circumstances. The periodogram is arguably one of the most widely known and used non-parameteric methods for spectral estimation; it is based on power spectral density (PSD) of the signal. The term was proposed by Schuster [50] and can be written as

$$\phi(\omega) = \lim_{N \to \infty} E\left\{\frac{1}{N}\Big|\sum_{t=1}^{N} y(t)e^{-i\omega t}\Big|^2\right\} \tag{5}$$

where $y(t)$ is a discrete-time data sequence, N denotes the number of samples and $\omega = 2\pi f$, where f is the frequency. A periodogram is essentially an estimate of a PSD from a signal with a limited number of samples and therefore we omit limit and expectation operators from the PSD definition

$$\hat{\phi}_p(\omega) = \frac{1}{N}\Big|\sum_{t=1}^{N} y(t)e^{-i\omega t}\Big|^2 \tag{6}$$

From the practical considerations it is impossible to evaluate $\hat{\phi}_p(\omega)$ over continuous frequencies and therefore the frequency space is sampled in order to estimate the approximation of the actual PSD of the signal. The estimate is then calculated by the means of the Discrete Fourier Transform (DFT)

$$Y(k) = \sum_{t=1}^{N} y(t)e^{-i\frac{2\pi}{N}tk}, \qquad k = 0, ..., N-1 \tag{7}$$

where the contributions of each frequency are then forming the spectrum of the signal. Computation of the DFT is rarely done directly by (7) as more efficient algorithms exist. Regular DFT results in complexity of $\mathcal{O}(N^2)$ and algorithms which can perform the computation with a lower number of operations are called Fast Fourier Transforms (FFT). By employing FFT, one can compute the periodogram with computational complexity as low as $\mathcal{O}(N \log N)$. Arguably the most widely used modern implementation of the FFT algorithm was proposed by Cooley and Tukey [51] in 1965; however, already in 1805 Carl Friedrich Gauss described an algorithm similar to the FFT for the computations of coefficients of a finite Fourier series [52]. The Cooley and Tukey version of FFT known as radix-2 FFT algorithm is not the most efficient by modern standards; however, its simplicity and ease of implementation keep it popular. An example of a periodogram estimate is shown in Figure 3 for a signal consisting of three sinusoids with frequencies $f_1 = 0.2$, $f_2 = 0.3$ and $f_3 = 0.8$, and with additive White Gaussian noise with signal-to-noise ratio (SNR) of 0 dB. Figure 3 and following figures in this chapter were generated in MATLAB environment.

Although the periodogram is a simple and computationally efficient method for estimating signal spectra, there exist several severe drawbacks (otherwise spectral analysis would not be such an active research field). Due to smearing effect, the resolution limit for classical periodogram is $1/N$, meaning that the method is not able to reliably resolve details in the signal spectrum which are separated by less than $1/N$ cycles per sampling interval. Another effect is known as spectral leakage - power from frequency bands with high power "leaks" to adjacent bands which contain less power. Smearing and spectral leakage are more critical for spectra with large deviations in amplitude, so-called peaky spectra. For smooth spectra, these effects are less serious. And although bias of the periodogram might be considerable in the case of peaky spectra, it is not the main limitation of the periodogram. This comes from the fact that in the case of large N, the periodogram is an unbiased estimator

$$\lim_{N \to \infty} E\{\hat{\phi}(\omega)\} = \phi(\omega) \tag{8}$$

meaning that if the only problem of this estimator would lie in its bias, it could be simply solved by increasing the number of collected samples N (considering it is possible). However, the main problem of the periodogram lies in its large variance. The periodogram is an inconsistent spectral estimator [1] and its
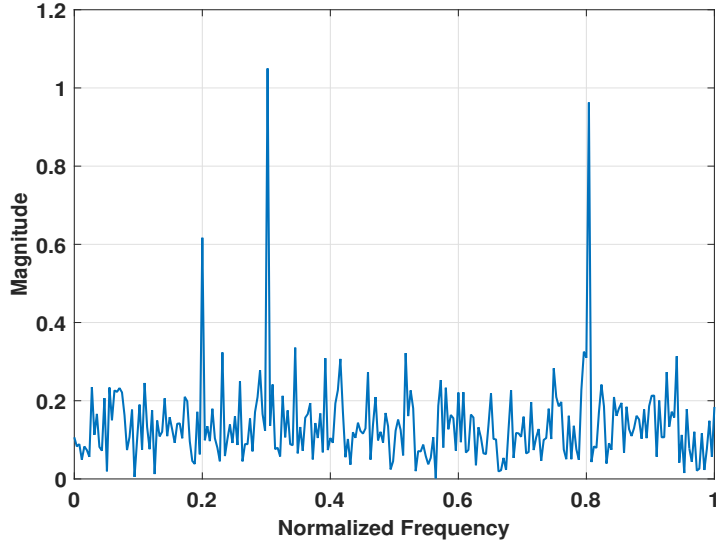
Figure 3 Periodogram estimate of a signal with three sinusoids ($f_1 = 0.2$, $f_2 = 0.3$ and $f_3 = 0.8$).

variance does not go to zero even if the number of samples goes to infinity. This limitation is known and several improved methods were proposed to decrease the variance; however, this comes at the cost of increasing the bias of the estimate. For example, one such method is Blackman-Tukey spectral estimator, which considers adding windowing (or weighting) to the initially estimated spectra. The method considers correlogram, which is an equivalent estimator to the periodogram and can be represented as follows

$$\hat{\phi}_c(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k)e^{-i\omega k} \tag{9}$$

where $\hat{r}$ denotes an estimate of the covariance lag $r(k)$ from the data sequence $\{y(n), ..., y(N)\}$. The Blackman-Tukey modification of the initial correlogram is given by

$$\hat{\phi}_{BT}(\omega) = \sum_{k=-(M-1)}^{M-1} w(k)\hat{r}(k)e^{-i\omega k} \tag{10}$$

where $w(k)$ is a weighting coefficient and $M \leq N$. This approach can be seen as equivalent to the "locally" weighted average of the periodogram [1] and results in smoothed version of the spectra estimated by the periodogram and hence reducing fluctuations and decreasing variance. However, this approach also introduces the undesirable effect of reducing the resolution and therefore choosing the appropriate window to use is a trade-off between statistical variance

23

and spectral resolution. This is controlled by the window length. Another trade-off that needs to be considered is the one between smearing and leakage effects, which are controlled by the shape of the window (weighting coefficients) [1]. This trade-off usually considers the application at hand, which dictates the most appropriate window shape to use. Different types of windows have been proposed, each one optimizing a specific property of the estimate. For example: Rectangular, Bartlett, Hanning, Hamming, Blackman etc.

Bartlett method [53], [54] splits available observations N into $L = N/M$ data sequences of M samples each. The periodogram is then calculated for each L data sequence and the results are averaged. It is easy to conclude that by using the Barlett method we reduce the spectral resolution by a factor of L compared to the periodogram directly calculated from N observations. The resulting gain in variance reduction can also be shown to have the same factor L [1].

Another popular method for estimating the PSD is the Welch estimator [55]. It can be seen as extension to the Bartlett method. The data sequences are allowed to overlap in the Welch method, which is not the case with the Bartlett method. This results in more periodograms for averaging and decreases the variance of estimate of PSD. In addition, each sequence is windowed before computing the periodogram, which gives control over the resolution trade-off.

To sum up, the non-parametric estimators are reliable tools for the task of estimating the signal's PSD. They are fast, robust, have been well studied and provide reasonable estimation performance. Their biggest limitation is their inferior resolution and variance.

3.2. Parametric estimators

Parametric estimators seek to establish the data model which describes the signal under the consideration and use this model to estimate the spectrum. Therefore the task itself is to estimate the parameters of the model. When the signal indeed confirms to the model, the parametric estimators often result in more accurate estimates as compared to non-parametric methods. On the other hand, when assumptions made by the data model do not hold for the signal under evaluation, then parametric methods perform worse than non-parametric ones, as latter are more robust to this sort of problems since they do not make any assumption about the data. Discussion on parametric methods can be divided into methods for the rational spectra (which forms a dense set on continuous spectra) and the methods for discrete spectra (sinusoids in the presence of noise).

A rational PSD can be described as a rational function of $e^{-i\omega}$

$$\phi(\omega) = \frac{\sum_{k=-m}^{m} \gamma_k e^{-i\omega k}}{\sum_{k=-n}^{n} \rho_k e^{-i\omega k}} \qquad (11)$$
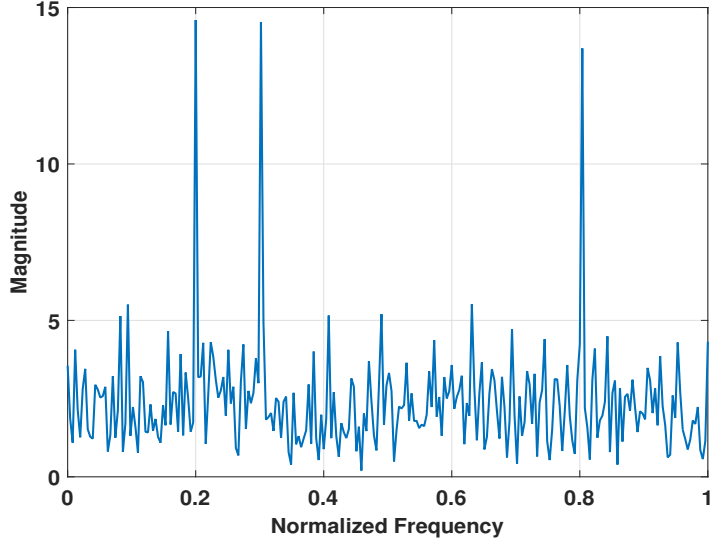
Figure 4 Least squares estimate for a signal with three sinusoids ($f_1 = 0.2$, $f_2 = 0.3$ and $f_3 = 0.8$).

and then represented [47] in the following form

$$\phi(\omega) = \left| \frac{B(\omega)}{A(\omega)} \right|^2 \sigma^2 \tag{12}$$

where σ^2 is a positive scalar and $A(\omega)$ and $B(\omega)$ are polynomials

$$\begin{aligned} A(\omega) &= 1 + \alpha_1 e^{-i\omega} + ... + \alpha_n e^{-in\omega} \\ B(\omega) &= 1 + \beta_1 e^{-i\omega} + ... + \beta_m e^{-im\omega} \end{aligned} \tag{13}$$

Arbitrary rational PSD in (12) corresponds to white noise with the power σ^2 filtered with transfer function [1], [47]

$$H(\omega) = B(\omega)/A(\omega) \tag{14}$$

which can be represented in the time domain through Z-transform

$$y(t) = \frac{B(z)}{A(z)} e(t) \tag{15}$$

where $e(t)$ is white noise of variance equal to σ^2. This approach allows us to represent the spectral estimation as a signal modeling problem. For a more thorough discussion we refer interested readers to [1], [47], [48]. Next, we consider several methods for acquiring signal model parameters.

A signal $y(t)$, that satisfies (15) is called an autoregressive moving average (ARMA, ARMA(n, m)) signal [1] and forms a class of ARMA signals.

$$A(z)y(t) = B(z)e(t) \tag{16}$$

In the case of $m = 0$, ARMA signal turns into an autoregressive (AR) signal, which can be considered a sub-class of ARMA signals.

$$A(z)y(t) = e(t) \tag{17}$$

AR type of signals are frequently used in applications as the AR model can represent spectra with narrow peaks, which are quite prevalent in practice. In addition, the AR model is a topic well researched, different methods for estimating the model parameters exist and the stability of the estimate can be guaranteed. For example, Yule-Walker method constructs system of linear equations (also known as Yule-Walker equations) and uses the relationship between the covariances and the AR parameters to obtain the solution. Another method for obtaining the AR model parameters is a recursive method known as Levinson-Durbin algorithm (LDA), which utilizes the structure of Yule-Walker system of equations to find the solution. The structure is Hermitian and Toeplitz and it allows LDA to be computationally more efficient than the standard Yule-Walker method. The description of aforementioned methods is not presented as it described extensively in the literature, for example [48]. We will, however, consider the least squares (LS) method for AR estimate as it coincides well with the mathematical presentation of the following topics. The LS method results in the similar formulation as the one used for finding the best linear predictor and therefore AR modeling is often referred to as linear predictive modeling [1]. Considering the output signal $y(t)$ and the input signal $\mathbf{x}(t)$ one seek to find set of parameters $\boldsymbol{\beta}$ for linear approximation

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} \sum_{t=0}^{N-1} |y(t) - \boldsymbol{\beta}\mathbf{x}(t)| \tag{18}$$

Solution for the least squares estimator can be find analytically as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{19}$$

where $\mathbf{X} = [\mathbf{x}(0) \ldots \mathbf{x}(N)]$ and $\mathbf{y} = [y(0) \ldots y(N)]^T$. Example of LS estimate for the same signal consisting of three sinusoids and SNR of 0 dB is shown in Figure 4.

In the case of $n = 0$ ARMA signal turns into a moving average (MA) signal; however, the MA signal cannot model the narrowband spectra[1].

$$y(t) = B(z)e(t) \tag{20}$$

---

[1]In theory it is possible, however only if order of the model is very large, which is highly impractical.

Instead, the MA model can provide good approximation for the spectra with wider peaks. As such spectra are less common in practice, the MA model does not find such widespread use (and corresponding interest from scientific community) as the AR model. In addition, the MA model is more difficult to solve than the AR model and it is usually reasonable to solve the ARMA instead, as it provides similar computational difficulties, but results in more general solution.

Considering methods for discrete spectra, which is common in many different applications, such as communications, radar, sonar, geophysical seismology etc., a signal can be well described by the sinusoidal model as follows

$$y(t) = x(t) + e(t) \tag{21}$$

$$x(t) = \sum_{k=1}^{N} \alpha_k e^{i(\omega_k t + \rho_k)} \tag{22}$$

where $x(t)$ denotes the signal consisting of the sum of complex-valued sinusoids with α_k, ω_k and ρ_k corresponding to k-th amplitude, frequency and phase, respectively. The noise $e(t)$ is usually assumed to be complex-valued Gaussian white noise. As parameters relate to the signal in nonlinear way, one can use the nonlinear least square (NLS) method

$$f(\omega, \alpha, \rho) = \sum_{t=1}^{N} \left| y(t) - \sum_{k=1}^{n} \alpha_k e^{i(\omega_k t + \rho_k)} \right|^2 \tag{23}$$

Subspace-based methods such as MUltiple SIgnal Classification (MUSIC) [56] and Pisarenko's method [57] (which is a special case of MUSIC, when $M = p+1$) employ eigen-decomposition to separate signal subspace and noise subspace of the considered spectra (hence the name). Although mathematical description of subspace methods is not in the scope of this thesis, it is important to note that the aforementioned methods can be also considered high-resolution methods and exhibit satisfactory statistical properties. ESPRIT [58] (Estimation of Signal Parameters by Rotational Invariance Techniques) is an additional method, which displays similar and often even better statistical accuracy than MUSIC.

3.3. Semi-parametric estimators

Non-parametric methods provide reasonable estimates to the PSD of the signal without making any assumption on the signal content and also usually have lower computational complexity. Parametric methods use underlying information about the signal, which often can result in better statistical performance and higher resolution than non-parametric methods. However, they are not that robust as they depend on the signal model, which should correspond well to the measured signal and fail if it is not the case. They also require a fair amount of additional

fine-tuning by the algorithm designer in order to choose the best parameters for the model. The research community has been interested in methods which produce high-resolution estimates, but with robustness of non-parametric methods and without inherent design complexity of parametric methods. For certain situations such methods exist and we will refer to them as semi-parametric methods. In this work the assumed constraint that allows for such methods to perform as well as they do, is the assumption of sparsity of the signal.

4. Sparsity and sparse reconstruction

The idea behind sparse reconstruction or sparse estimation methods (also sparse approximation, sparse representation) is to establish underdetermined system of equations and in order for it to be solvable enforce the solution to be sparse, as in only few of the candidates to combine a solution are selected. By candidates we consider here columns in a matrix $\mathbf{D}$, which is referred to as the dictionary. In general we use the term dictionary to describe a matrix (or a tensor) which consists of basic elements, which are also referred to as candidates or as atoms (meaning smallest building blocks of the solution). The goal is to find a sparse representation of the signal of interest as a linear combination of as few atoms as possible. Often the problem is expanded to search also for the basic elements or candidates themselves. Here, however, we consider the pre-defined dictionary. Let us consider some dictionary $\mathbf{D}$ consisting of signal candidates and solution vector $\mathbf{x}$ which forms a combination of columns from $\mathbf{D}$ to reconstruct the signal $\mathbf{y}$.

$$\mathbf{y} = \mathbf{D}\mathbf{x} \tag{24}$$

In real-life applications we have to consider added noise, therefore the reconstruction takes form

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e} \tag{25}$$

with $\mathbf{e}$ corresponding to white Gaussian noise vector, which is the usual assumption due to its favorable mathematical representation and often good approximation of the actual systems. A common solution to this sort of a problem would be minimizing squared ℓ_2-norm of residuals, or finding shortest Euclidean distance between signal vector $\mathbf{y}$ and its approximation, which is reconstructed from combination of columns from dictionary $\mathbf{D}$ described by solution vector $\hat{\mathbf{x}}$

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \, \frac{1}{2} ||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 \tag{26}$$

Solution to this optimization problem can be obtained from

$$\hat{\mathbf{x}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{y} \tag{27}$$

which is the already known LS estimator. However, although this estimator minimizes residual sum of squares, the optimal solution will be some combination

of most of the (or all) columns in the dictionary $\mathbf{D}$ and therefore not sparse. So, even if we end up with an optimal estimator in some statistical sense, it is not providing us any meaningful information about the signal vector $\mathbf{y}$. This form of solution can be considered as regression problem; however, for regression the usual assumption is that $M < N$ and in our case it does not hold. The description of the dictionary $\mathbf{D}$ will be given later, but it is usually constructed to be overcomplete, as in having $M >> N$, considerably larger number of columns than number of rows in the dictionary. Therefore we end up with an undetermined system of equations without a unique solution. In order to proceed further, we have to put additional constraints on the problem. This constraint would be the notion of sparsity - we assume that $\mathbf{y}$ is itself sparse or sparse in some domain (recur beginning of our discussion on wavelet and Fourier Transform). Therefore the solution to the optimization problem $\hat{\mathbf{x}}$ should be sparse as well. We should be able to represent $\mathbf{y}$ with only some limited number of columns from the dictionary $\mathbf{D}$. We will come back to this later, drawing parallels with the framework of compressed sensing. As we are assuming that the solution should be sparse, we need to impose an additional penalty on our estimator to guarantee that the solution will indeed be sparse.

4.1. Penalties, regularization

The intuitive way for describing signal sparsity is the ℓ_0-"pseudonorm" [2] - number of non-zeros elements in the vector, which results in the following optimization problem

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{Dx}||_2^2 + \lambda||\mathbf{x}||_0 \tag{28}$$

where λ is a tuning parameter controlling sparsity of the solution. This problem, however, is known to be NP-hard and is unfeasible for large $\mathbf{x}$ as it results in the extensive combinatorial search for the solution. The problem is therefore often relaxed and ℓ_0-penalty is replaced with ℓ_1-norm [59] which is defined as

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i| \tag{29}$$

and results in the following optimization problem

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{Dx}||_2^2 + \lambda||\mathbf{x}||_1 \tag{30}$$

with the very favourable property of being a convex problem and therefore local minimum is also a global minimum of the solution. This is the form of the widely

---

[2]Not to be confused with the actual pseudonorm definition. Note the absence of quotation marks. ℓ_0-"pseudonorm" is not a norm, because it does not fulfill one of the definitions of the norm - it is not homogeneous.

popular LASSO estimator [60] (least absolute shrinkage and selection operator). An estimate corresponding to previously considered signal with sinusoids is shown in Figure 5 and Figure 6 compares resolution limits of periodogram and LASSO estimators. From Figure 5 the sparsity of the solution can be clearly seen. Only the sinusoids are recovered as including other parts of the spectra does not result in the sparse solution which describes the measured signal in an optimal way. Another possible convex relaxation is ℓ_2-norm penalty

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \ \frac{1}{2}||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_2 \tag{31}$$

In statistics and machine learning literature this is also known as ridge regression or Tikhonov regularization. One can also consider convex combination of ℓ_1 and ℓ_2 penalties which results in so-called elastic-net regularization [61].

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \ \frac{1}{2}||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 + \lambda_2||\mathbf{x}||_2^2 + \lambda_1||\mathbf{x}||_1 \tag{32}$$

Sparse reconstruction can be seen as somewhat hybrid of parametric and non-parametric estimators [62]. It does use some information about the signal - assumption is made that the signal should be sparse and can be represented with a small number of significant coefficients. However, no explicit assumption on the number of required coefficients is made. It is left to the estimator to find the best model order. The result of this is the group of methods which can be considered high-resolution estimators, as they often outperform periodogram, but at the same time are considerably more robust to the model assumption in comparison to parametric methods.

4.2. Dictionary construction

Next, let us consider dictionary construction for the problem of estimating the frequencies f_k, for $k = 1, \dots, K$, of a measured signal y_n, with

$$y_n = \sum_{k=1}^{K} \beta_k e^{2i\pi f_k t_n} + \epsilon_n \tag{33}$$

for $n = 1, \dots, N$, and with K denoting the (unknown) number of sinusoids in the considered signal. Let β_k and f_k denote the complex amplitude and frequency of the kth frequency, respectively, t_n the nth sample time, and ϵ_n the additive noise at time t_n. The original sparse formulation of this estimation problem presented in [37] considers the LASSO minimization discussed above

$$\underset{\mathbf{x}}{\operatorname{argmin}} \ \frac{1}{2}||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1 \tag{34}$$
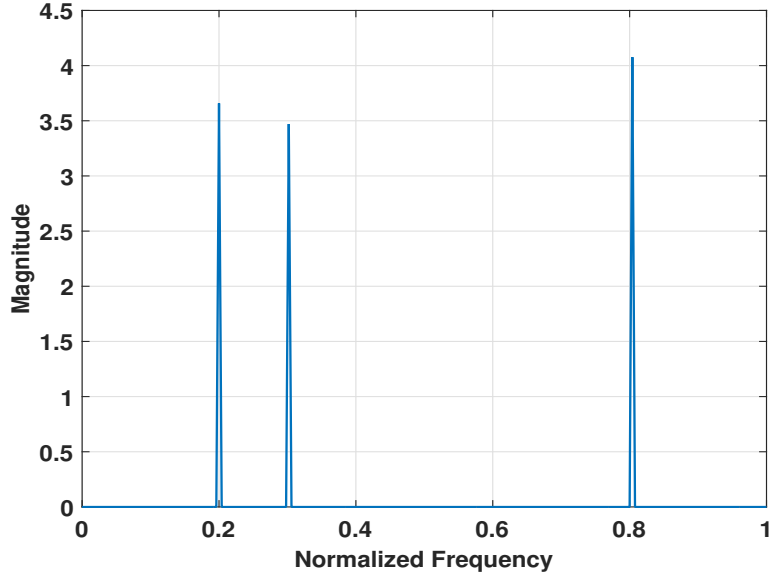
Figure 5 LASSO estimate for a signal with three sinusoids ($f_1 = 0.2$, $f_2 = 0.3$ and $f_3 = 0.8$).

with

$$\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_N \end{bmatrix}^T \tag{35}$$

$$\mathbf{D} = \begin{bmatrix} \boldsymbol{d}_1 & \cdots & \boldsymbol{d}_L \end{bmatrix} \tag{36}$$

$$\boldsymbol{d}_\ell = \begin{bmatrix} e^{2i\pi \hat{f}_\ell t_1} & \cdots & e^{2i\pi \hat{f}_\ell t_N} \end{bmatrix}^T \tag{37}$$

where $\hat{f}_\ell$ for $\ell = 1, \ldots, L$ denotes the $L \gg K$ candidate frequencies in the dictionary, $\mathbf{D}$, which are typically selected to be closely spaced to allow for minimal grid mismatch, and $(\cdot)^T$ the transpose. The desired frequencies and the model order are then found as the non-zero elements in $\hat{\mathbf{x}}$.

The above example uses a simple case of a sinusoid dictionary element. Regardless of that, this approach is very effective when dealing with narrowband signals that are often modeled as a sum of sinusoids. However, the choice of the proper dictionary element should depend on the task to be solved and indeed one can use a whole multitude of different functions, for example wavelet-, curvelet- or bandelet-based dictionary elements for image processing [63]. The field of online dictionary learning goes even further than that, without implicitly constructing the dictionary beforehand, but rather coming up with the methods which the learn properties of the analyzed signal and construct the dictionary according to that [64], [65]. An obvious resemblance can be found in the methods of supervised learning from the field of machine learning. Without any explicit mathematical description of the data structure, the model is built based
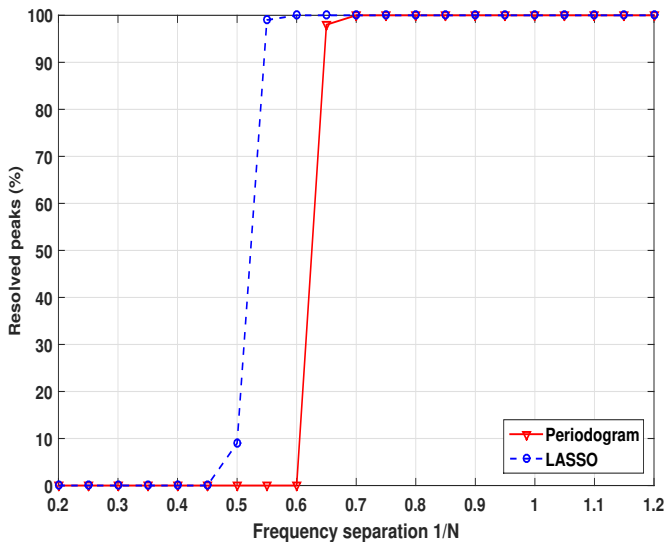
31

Figure 6 Resolution limits for periodogram and LASSO. Resolution limit is defined through the peak resolving ability, which is described in numerical section of Publication D.

on learning example of the considered data, which results in finer adaptation to the task at hand [63].

4.3. Parameter selection

Recalling the previous discussion on parametric estimators, we made a remark that the considered sparse reconstruction methods do not require any choice of the model parameters. Still, the considered approach requires the choice of the regularization parameter λ. It is usually chosen through some heuristic based on the data or through the procedure of cross-validation [66], [67] where data is divided into two parts (usually denoted training and test part), then fitted on the training part and validated on the test part, which helps to pick the best model parameters while avoiding "overfitting" [67]. In this thesis we approach the choice of the parameter by selecting $\lambda = \alpha \|\mathbf{D}^H\mathbf{y}\|_\infty$, where $\alpha \in [0, 1]$ is user-selectable parameter which sets the ratio for largest inner-product of the dictionary $\mathbf{D}$ and the data [62]. For $\alpha = 1$ it means λ would be the smallest tuning parameter value for which all the coefficients in the solution are zero [68]. The choice of α then is done by employing grid-search, where we evaluate different values of α and pick the one based on the results. The impact of λ on the estimate is shown in Figure 7. With decreasing λ we relax the sparsity constraint and allow more peaks to be considered as a part of the solution.
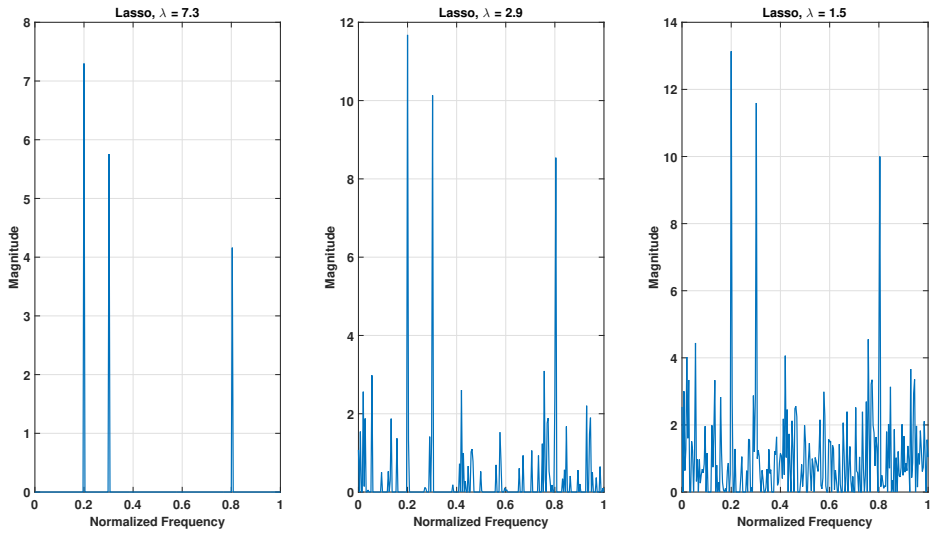
Figure 7 LASSO solution for different λ values.
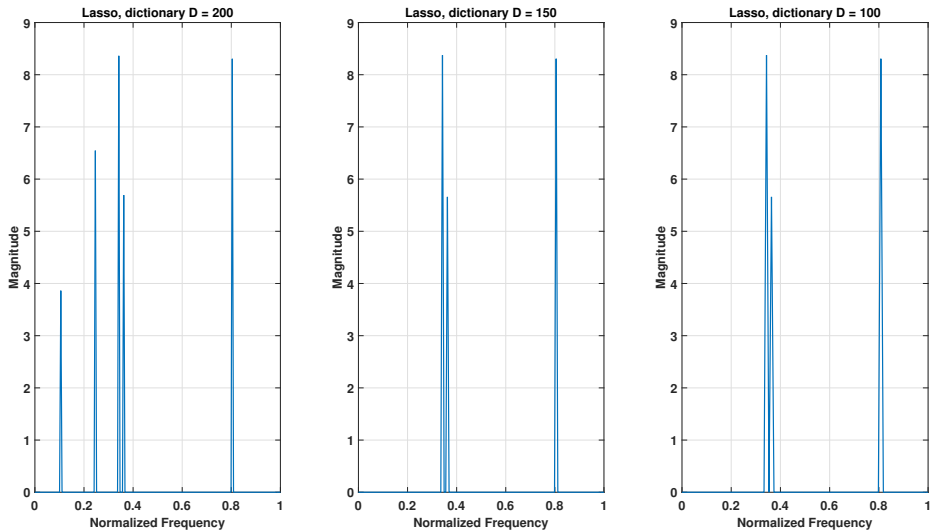
Figure 8 LASSO solution for different number of elements in dictionary **D**.

4.4. Grid selection

Previously we have described the construction of the signal candidate dictionary **D**, however we did not comment on number of elements in the dictionary. Ideal dictionary would have elements at the same frequencies as the actual frequencies in the estimated spectra. This, however, is an unrealistic assumption, as it would

require infinitely small gaps between the dictionary elements. Instead of relying on dictionary elements which perfectly coincide with the estimated sinusoids in the spectra, the construction of the dictionary is usually done with the assumption that the actual frequencies lie quite close to the ones the dictionary consists of. Therefore, we end up with the small "rounding" error.

It is evident that by increasing number of elements in the dictionary we are getting more precise estimate and vice versa. A finer dictionary, however, requires more computations in order to find the solution and therefore selecting the grid for the dictionary can be seen as a trade-off between required resolution and computational complexity. It is also important to note that by making dictionary too coarse, the risk is that the estimator might miss the frequency located in between two dictionary elements [36], [37], as depicted by Figure 8, where a signal consisting of five sinusoids is considered and dictionaries with smaller number of elements are recovering only three sinusoids. This effect is known as the grid-mismatch problem or off-grid estimation [42], [43]. In the past years there have been several ideas on how to solve this. One approach that has been studied is an adaptive grid selection, where the dictionary grid is also a part of optimization problem, meaning that we also search for the optimal frequency grid and not only for the best set of elements in the dictionary to solve the problem. This might seem like a hard problem to solve as one has to find the best set of elements and then to tune the grid in the dictionary, which itself will affect how the best set of elements should be chosen. However, it is often possible to separate this sort of problem into separate problems, first solving for x and then updating the grid and then repeat the process iteratively until convergence. Although, this sort of approach is easy to use, one still has to be careful in choosing the initial grid, to avoid poor estimates and what is more important, the problem formulation is no longer convex in this case and therefore one cannot guarantee convergence to global optima. Another interesting approach proposed lately is atomic-norm minimization [44], [45]. Here, instead of constructing discrete grid on which signal is evaluated, the authors employ atomic norm penalty, as introduced in [46] and formulate problem using infinite grid (continuous dictionary). This can be seen as generalization of the LASSO problem and allows for a way to determine the most suitable convex penalty for signal recovery. This approach often results in an accurate signal reconstruction, however at the cost of requiring solving computationally demanding optimization problems, which might limit the size of the considered problem.

4.5. Relation to a compressed sensing

In the last decade, compressed sensing (CS) has attracted considerable interest in the research community in areas of mathematics, electrical engineering and signal processing. So far the classical approach was lead by the Nyquist-Shannon sampling theory which requires a certain minimum of samples to be collected

in order to perfectly reconstruct a bandlimited signal. However, we often apply compression methods afterwards to reduce the amount of captured information (MP3 or JPEG compression for correspodingly audio and image data for example). The fundamental change that compressed sensing brings is that rather than sampling at a higher frequency and then compressing the sampled signal, CS allows to collect signal in an already compressed form at a lower sampling rate if the signal can be represented with a small number of significant coefficients, $k \ll N$, i.e. the signal is sparse in some domain. The field of CS has gained considerable interest with the research community after the theoretical works of Donaho [27] and Candes [12], [29] showed that a sparse signal can be recovered exactly from a small set of measurements. Different applications already leverage the idea, for example applications in medical imaging applications [69], [70] where speedups of several times were noted. To briefly present the main idea behind compressed sensing we start with the description of its process as

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{38}$$

where $\mathbf{x}$ is a signal and $\mathbf{y}$ its sampled form we wish to capture, A is $M \times N$ and is referred to as the sensing matrix and it has to satisfy the restricted isometry property (RIP) [71]. One class of matrices which are useful and satisfy RIP with high-probability are random matrices, where the matrix elements are chosen according to any sub-gaussian distribution. The original signal might not be sparse and as compressed sensing requires sparsity for its formulation, it is often required to add a transform to the suitable basis $\mathbf{\Phi}$ where $\mathbf{x}$ is actually sparse.

$$\mathbf{y} = \mathbf{\Phi}\mathbf{A}\mathbf{x} \tag{39}$$

Here lies the additional benefit of constructing the sensing matrix $\mathbf{A}$ to be random, as when $\mathbf{\Phi}$ is an orthonormal basis, $\mathbf{A}\mathbf{\Phi}$ will satisfy RIP. Recovery can be done by the already familiar ℓ_1-minimzation

$$\hat{\mathbf{x}} = \underset{\mathbf{z}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{A}\mathbf{z} - \mathbf{y}||_2^2 + \lambda||\mathbf{z}||_1 \tag{40}$$

or by greedy algorithms including different flavours of basis pursuit [72], [73] or iterative thresholding methods [74]. Therefore, clear similarity can be noted between CS framework and LASSO as both consider sparse signals and can be solved by using similar ℓ_1-minimization problem. However, one can also argue that CS is a more general approach for sampling, whereas LASSO is just one particular method for reconstruction. In addition, with LASSO one is usually not concerned with converting signal into sparse form as it is implied that signal is sparse, on the other hand for the CS the choice of proper transform matrix $\mathbf{\Phi}$ is usually part of the problem.

5.   Convex optimization and ADMM

As was discussed before, convex problems have the advantageous property of solution local minima being also a global minima. Mathematics of convex optimization has been known for a while; however, modern day interest re-ignited when interior-point methods developed in the 1980s were found suitable for solving convex problems. Then, since the 1990s many applications of convex optimization have been discovered in areas such as automatic control systems, estimation and signal processing, communications and networks, electronic circuit design, data analysis and modeling, statistics, finance and many more [75]. In addition, general-purpose software for nonlinear convex optimization solvers such as MOSEK, Sedumi, SDPT3 [76]–[78] are readily available for the public. This stimulated additional interest in the research community. Modelling packages such as CVX, Yalmip (Matlab), CVXMOD (Python) [75], [79] facilitate further development and prototyping of new optimization problems and regularizers. Although such frameworks are great for experimentation, they are generally too complex and computationally expensive. Therefore, it is often required to implement faster solver directly for the considered problem. Next, we will present one such method, which is also used to solve various problems presented further in this thesis.

5.1.   The alternative direction method of multipliers

The Alternative Direction Method of Multipliers (ADMM) is a Lagrangian-based method developed in the 1970s, which gained new attention in the 21st century due to the rise of Machine Learning and Big-Data and its requirements of scalable distributed solution methods [80]. ADMM works by decomposing a large problem into smaller ones, then iteratively solving these small subproblems while coordinating to find a solution to the original large problem. This approach helps to distribute solving subproblems to different computing nodes or/and leverage modern hardware for parallel computing (graphics processing units (GPU)) [81]. Next, we will proceed with a brief description of the basic ADMM formulation; however, the interested reader is referred to a more in-depth overview of the method combined by Boyd et al. [80]. ADMM solves following convex optimization problem

$$\text{minimize} \quad f_1(\boldsymbol{z}) + f_2(\mathbf{G}\boldsymbol{z}), \tag{41}$$

where $\boldsymbol{z}$ is the optimization variable, $f_1(\cdot)$ and $f_2(\cdot)$ are convex functions, and $\mathbf{G}$ is a known matrix. If we introduce an auxiliary variable $\boldsymbol{u} = \mathbf{G}\boldsymbol{z}$, then we can rewrite accordingly

$$\text{minimize} \quad f_1(\boldsymbol{z}) + f_2(\boldsymbol{u}) + \frac{\rho}{2}\|\mathbf{G}\boldsymbol{z} - \boldsymbol{u}\|_2^2 \tag{42}$$

$$\text{subject to} \quad \mathbf{G}\boldsymbol{z} - \boldsymbol{u} = 0, \tag{43}$$

The augmented Lagrangian for the scaled form of ADMM is formed then as

$$L_\rho(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{d}) = f_1(\boldsymbol{z}) + f_2(\boldsymbol{u}) + \frac{\rho}{2}\|\mathbf{G}\boldsymbol{z} - \boldsymbol{u} + \boldsymbol{d}\|_2^2, \qquad (44)$$

where $\boldsymbol{d}$ denotes the scaled dual variable. The problem is solved by iterating between solving for $\boldsymbol{z}$, while keeping $\boldsymbol{u}$ fixed at the value of the previous iteration and vice versa. Solution holds for all ρ, since at any feasible point $\|\mathbf{G}\boldsymbol{z} - \boldsymbol{u}\|_2^2 = 0$. At iteration (j+1) solutions are obtained by solving

$$\boldsymbol{z}^{(j+1)} = \underset{\mathbf{x}}{\mathrm{argmin}} L_\rho(\boldsymbol{z}, \boldsymbol{u}^{(j)}, \boldsymbol{d}^{(j)}) \qquad (45)$$

$$\boldsymbol{u}^{(j+1)} = \underset{\mathbf{x}}{\mathrm{argmin}} L_\rho(\boldsymbol{z}^{(j+1)}, \boldsymbol{u}, \boldsymbol{d}^{(j)}) \qquad (46)$$

and updating the scaled dual variable as

$$\boldsymbol{d}^{(j+1)} = \boldsymbol{d}^{(j)} - \rho(\mathbf{G}\boldsymbol{z}^{(j+1)} - \boldsymbol{u}^{(j+1)}) \qquad (47)$$

ADMM optimization is a great mathematical tool when solutions for (45) and (46) can be found easier than the original problem at hand.

5.2. ADMM for LASSO

Let us consider the formulation of ADMM for LASSO (3) by splitting the initial variable $\mathbf{x}$ into two variables, $\mathbf{x}$ and $\boldsymbol{z}$, with LASSO taking the following form

$$\underset{\mathbf{x}, \boldsymbol{z}}{\mathrm{minimize}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\boldsymbol{z}\|_1 \qquad (48)$$

The Augmented Lagrangian is

$$L_\rho(\mathbf{x}, \boldsymbol{z}, \boldsymbol{u}) = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\boldsymbol{z}\|_1 + \frac{\rho}{2}\|\mathbf{x} - \boldsymbol{z} + \boldsymbol{u}\|_2^2 \qquad (49)$$

To minimize (49) with respect to variables $\mathbf{x}$ and $\boldsymbol{u}$, we need to differentiate the Lagrangian, set the resulting derivative to zero and solve for the corresponding variable at step $k + 1$. For $\mathbf{x}$ this approach yields an expression similar to ridge regression

$$\mathbf{x}^{k+1} = \left(\mathbf{D}^T\mathbf{D} + \rho\mathbf{I}\right)^{-1}\left(\mathbf{D}^T\mathbf{y} + \rho(\boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)})\right) \qquad (50)$$

and ADMM can be interpreted as a method for solving the lasso problem by iteratively carrying out ridge regression [80]. Due to ℓ_1-penalty, $\boldsymbol{z}$ is non-differentiable directly, but a closed-form solution can be found by using subdifferential calculus and represented in the form of *soft-thresholding* operator

$$\boldsymbol{z}^{k+1} = \mathcal{S}(\mathbf{x}^{(k+1)} + \boldsymbol{u}^{(k)}, \lambda/\rho) \qquad (51)$$

Finally, the last step of ADMM is to update dual variable u according to (47). The principal computational cost is the inversion in (50), which requires $\mathcal{O}(M^3)$ [82]. However, it can be pre-computed offline and thus LASSO through ADDM requires at most $\mathcal{O}(M^2)$ operations, which can possibly be further reduced by utilizing efficiently the structure of the dictionary matrix through factorization similar to approaches presented in [83], [84].

6.   Problem statement and research questions

To conclude on the previous discussion, it is clear that semi-parametric estimators can provide a viable alternative to parametric and non-parametric estimators and result in high-resolution estimates. In some conditions they can outperform non-parametric estimators in terms of frequency resolution without introducing model sensitivity as parametric estimators do. However, we need to consider that semi-parametric estimators might still require considerable amount of computational resources to reach the optimal solution. Therefore, it is of interest to formulate efficient algorithms for solving the optimization problem. State-of-the-art estimators mostly utilize methods of sparse reconstruction and dictionary-based approach. However, for dictionary-based estimators the problem of grid-selection arises and although different attempts at mitigating it exists in the literature, they all have their drawbacks and therefore new methods are still needed. This is true for parameter estimation in general as well. Although this area of research has been very active and have seen many brilliant contributions through the last 70 years, it is still not solved and many new methods are being proposed each year. To summarize, the general requirements for the ideal estimator are:

- Reliable, unbiased estimate with low variance

- Robust

- Able to provide high-resolution estimates

- Computationally efficient

- Simple problem formulation and not limited by the problem size

- Does not suffer from off-grid problem

- Is able to work with non-uniform data without requiring reformulation of estimator

- Provides reliable estimate with small number of samples

Research questions:

1. Can wideband dictionary formulation help to mitigate off-grid estimation issue for sparse signal reconstruction?

2. If wideband based dictionary is suitable, then what sort of computational complexity reduction can one expect from iterative formulation of the estimation problem?

3. How suitable are different types of wideband elements?

4. What are the possible drawbacks of wideband dictionaries compared to classical narrowband dictionaries?

5. How well does synthetic tests correspond to actual real-life data?

6. What other estimators can benefit, and in what way, from dictionaries constructed in a wideband manner?

7. Contributions of the thesis

The main focus of this thesis is on methods of sparse signal estimation. The main contributions of this thesis for that field can be summarized as follows:

1. A novel method of constructing the dictionary for sparse signal estimation. The proposed method of wideband dictionaries decreases probability of missing off-grid components and can make reliable estimations in situations where the number of samples is considerably less than the number of dictionary elements. The percentage of correct model order estimation in this scenario is $40 - 50\%$ higher than conventional method (90-100% vs 50-60%). For the same resolution, computational complexity of the proposed method can be $20 - 30$ times lower, which results in a considerable reduction in the time required to make an estimation [Publications A and B].

2. Application of the method described in Publications A and B to actual real-life signals in the domain of electrical bio-impedance. Describing method of sparse reconstruction for the separation of cardiac and respiratory signal components from electrical bio-impedance measurements [Publication C].

3. A novel procedure for iterative zooming for the IAA algorithm. The proposed approach allows for higher precision estimates on relevant areas of the spectrum without the need for using a large dictionary of finely spaced elements. Therefore, the proposed approach reduces the computational complexity required for the estimation. The proposed method coupled with the wideband dictionary results in increased estimation performance in terms of frequency estimation as well as peak resolving ability. Average runtime of the proposed method is up to $4.5\textbf{x}$ faster [Publication D].

This thesis is based on four Publications (A to D). The outline of the Publications is presented below.

Publication A: Estimating Sparse Signals Using Integrated Wide-band Dictionaries

In this paper, we present a technique for reducing the size of the dictionary in sparse signal reconstruction by formulating an initial dictionary containing elements that span bands of the considered parameter space. We allow for the use of this banded dictionary in a first-stage estimation procedure, in which large parts of the parameter space is discarded for further analysis, thereby reducing the overall computationally complexity required to allow for a reliable signal reconstruction. We illustrate the presented principle on the problem of estimating sinusoidal components corrupted by white noise. The work in Publication A has been published as

Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. "Estimating Sparse Signals Using Integrated Wide-band Dictionaries" in *Proc. of 42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, pp. 4426-4430, 5-9 Mar. 2017.

Publication B: Estimating Sparse Signals Using Integrated Wide-band Dictionaries

In this paper, we introduce a wideband dictionary framework for estimating sparse signals. By formulating integrated dictionary elements spanning bands of the considered parameter space, one may efficiently find and discard large parts of the parameter space not active in the signal. After each iteration, the zero-valued parts of the dictionary may be discarded to allow a refined dictionary to be formed around the active elements, resulting in a zoomed dictionary to be used in the following iterations. Implementing this scheme allows for more accurate estimates, at a much lower computational cost, as compared to directly forming a larger dictionary spanning the whole parameter space or performing a zooming procedure using standard dictionary elements. Different from traditional dictionaries, the wideband dictionary allows for the use of dictionaries with fewer elements than the number of available samples without loss of resolution. The technique may be used on both one- and multi-dimensional signals, and may be exploited to refine several traditional sparse estimators, here illustrated with the LASSO and the SPICE estimators. Numerical examples illustrate the improved performance. The work in Publication B has been be published as

Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. "Estimating Sparse Signals Using Integrated Wide-band Dictionaries" in *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4170-4181, 2018.

Publication C: Sparse Reconstruction Method for Separating Cardiac and Respiratory Components from Electrical Bioimpedance Measurements

In this work, we investigate the possibility of employing a sparse reconstruction framework for the separation of cardiac and respiratory signal components from the bioimpedance measurements. The signal decomposition is complicated by the nonstationarity of the signal and overlapping of their spectra. The signal has a harmonic structure which is sparse in the spectral domain. We approach the problem by considering a dictionary with integrated wideband elements describing spectral components of the considered signal. The parameter estimation task is solved through the means of sparse reconstruction where solving the optimization problem returns a sparse vector of relevant dictionary atoms. The work in Publication C has been published as

Maksim Butsenko, Olev Märtens, Andrei Krivošei and Yannick Le Moullec. "Sparse Reconstruction Method for Separating Cardiac and Respiratory Components from Electrical Bioimpedance Measurements" in *Elektronika ir Elektrotechnika*, vol. 24, no. 5, pp. 57-61, 2018.

Publication D: The Zoomed Iterative Adaptive Approach

In this work, we investigate the possibility of incorporating a zooming procedure for the iterative adaptive approach (IAA), and thereby allow for higher precision on relevant areas of the spectrum. These kinds of zooming schemes have been used successfully together with several other methods, and have in many cases shown dramatical decrease in the computational cost. It has earlier been noted that the IAA method does not easily allow for these kind of zooming approaches, as the covariance formulation dictates that the resolution must be the same over all regions of the spectrum. In this paper, we present an iterative zooming procedure which allows for an efficient local estimation of IAA spectrum. Numerical examples illustrate the improved performance as compared to the classical IAA estimate. The work in Publication D has been accepted and will be published as

Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. "The Zoomed Iterative Adaptive Approach" at *2018 International Symposium on Intelligent Signal Processing and Communication Systems*.

References

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[2] E. G. Larsson, P. Stoica, and J. Li, "Sar image construction from gapped phase-history data", in *Image Processing, 2001. Proceedings. 2001 International Conference on*, IEEE, vol. 3, 2001, pp. 608–611.

[3] H. Choi and D. C. Munson Jr, "Direct-fourier reconstruction in tomography and synthetic aperture radar", *International journal of imaging systems and technology*, vol. 9, no. 1, pp. 1–13, 1998.

[4] J. D. Scargle, "Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data", *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.

[5] S. Baisch and G. H. Bokelmann, "Spectral analysis with incomplete time series: An example from seismology", *Computers & Geosciences*, vol. 25, no. 7, pp. 739–750, 1999.

[6] A. W.-C. Liew, J. Xian, S. Wu, D. Smith, and H. Yan, "Spectral estimation in unevenly sampled space of periodically expressed microarray time series data", *BMC bioinformatics*, vol. 8, no. 1, p. 137, 2007.

[7] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast fourier transform", *SIAM review*, vol. 46, no. 3, pp. 443–454, 2004.

[8] J. A. Fessler and B. P. Sutton, "Nonuniform fast fourier transforms using min-max interpolation", *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 560–574, 2003.

[9] E.-H. Djermoune, G. Kasalica, and D. Brie, "Estimation of the parameters of two-dimensional nmr spectroscopy signals using an adapted subband decomposition", in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, IEEE, 2008, pp. 3641–3644.

[10] E. H. D. S. Sahnoun and D. Brie, "Sparse Modal Estimation of 2-D NMR Signals", in *38th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[11] J. C. Hoch and A. S. Stern, *NMR data processing*. Wiley, 1996.

[12] E. J. Candès and M.B. Wakin, "An introduction to compressive sampling", *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[13] C. Mulcahy, "Image compression using the haar wavelet transform", *Spelman Science and Mathematics Journal*, vol. 1, no. 1, pp. 22–31, 1997.

[14] Z. Shi, C. Zhou, Y. Gu, N. A. Goodman, and F. Qu, "Source estimation using coprime array: A sparse reconstruction perspective", *IEEE Sensors Journal*, vol. 17, no. 3, pp. 755–765, 2017.

[15] 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.

[16] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays", *IEEE transactions on signal processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[17] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing", in *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*, IEEE, 2009, pp. 626–629.

[18] P. Stoica, D. Zachariah, and L. Li, "Weighted SPICE: A Unified Approach for Hyperparameter-Free Sparse Estimation", *Digit. Signal Process.*, vol. 33, pp. 1–12, Oct. 2014.

[19] L. Guo, H. Gao, J. Li, H. Huang, and X. Zhang, "Machinery vibration signal denoising based on learned dictionary and sparse representation", in *Journal of Physics: Conference Series*, IOP Publishing, vol. 628, 2015, p. 012124.

[20] Y. Ding and I. W. Selesnick, "Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization", *IEEE signal processing letters*, vol. 22, no. 9, pp. 1364–1368, 2015.

[21] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries", *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[22] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation", *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering", *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[24] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution", *IEEE transactions on image processing*, vol. 23, no. 6, pp. 2569–2582, 2014.

[25] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing", *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1708–1721, 2010.

[26] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization", *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.

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

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

[29] E. J. Candès, 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.

[30] V. Papyan, Y. Romano, and M. Elad, "Convolutional neural networks analyzed via convolutional sparse coding", *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2887–2938, 2017.

[31] T. Peleg, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery", *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2286–2303, 2012.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning 2nd edition*. New York: Springer, 2009.

[33] Y. T. Nguyen, M. G. Amin, M. Ghogho, and D. McLernon, "Local sparse reconstructions of doppler frequency using chirp atoms", in *Radar Conference (RadarCon), 2015 IEEE*, IEEE, 2015, pp. 1280–1284.

[34] D. D. Ariananda, H. Jamali-Rad, Z. Tang, G. Leus, and X. Campman, "Deterministic fourier-based dictionary design for sparse reconstruction", in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2016 IEEE*, IEEE, 2016, pp. 1–5.

[35] Y. T. Nguyen, D. McLernon, M. Ghogho, and S. A. R. Zaidi, "Sparse reconstruction of time-frequency representation using the fractional fourier transform", in *Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), International Conference on*, IEEE, 2017, pp. 16–20.

[36] 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.

[37] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra", in *17th World Congress IFAC*, Seoul, Jul. 2008, pp. 10 225–10 229.

[38] G. O. Glentis, K. Zhao, A. Jakobsson, H. Abeida, and J. Li, "SAR Imaging via Efficient Implementations of Sparse ML Approaches", *Elsevier Signal Processing*, vol. 95, pp. 15–26, 2014.

[39] J. R. Jensen, G. O. Glentis, M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Computationally Efficient IAA-based Estimation of the Fundamental Frequency", in *European Signal Processing Conference*, Bucharest, Romania, Aug. 2012.

[40] L. E. Ghaoui, V. Viallon, and T. Rabbani, "Safe Feature Elimination for the LASSO and Sparse Supervised Learning Problems", *Pacific Journal of Optimization*, vol. 8, no. 4, pp. 667–698, Jan. 2012.

[41] Z. J. Xiang, Y. Wang, and P. J. Ramadge, "Screening Tests for Lasso Problems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1008–1027, May 2016.

[42] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing", *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[43] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, "On the relation between sparse reconstruction and parameter estimation with model order selection", *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 560–570, 2010.

[44] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation", *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[45] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid", *IEEE transactions on information theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[46] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems", *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[47] S. M. Kay, *Fundamentals of statistical signal processing. Vol 1, Estimation theory*. Englewood Cliffs, NJ: Prentice-Hall PTR, 1993.

[48] M. H. Hayes, *Statistical digital signal processing and modeling*. John Wiley & Sons, 2009.

[49] Scott Fortmann-Roe. (2012). Understanding the bias-variance tradeoff. [Accessed: 12.06.2018], [Online]. Available: `http://scott.fortmann-roe.com/docs/BiasVariance.html`.

[50] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena", *Journal of Geophysical Research*, vol. 3, no. 1, pp. 13–41, 1898.

[51] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series", *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[52] M. Heideman, D. Johnson, and C. Burrus, "Gauss and the history of the fast fourier transform", *IEEE ASSP Magazine*, vol. 1, no. 4, pp. 14–21, 1984.

[53] M. Bartlett, "Smoothing periodograms from time-series with continuous spectra", *Nature*, vol. 161, no. 4096, p. 686, 1948.

[54] M. S. Bartlett, "Periodogram analysis and continuous spectra", *Biometrika*, vol. 37, no. 1/2, pp. 1–16, 1950.

[55] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms", *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[56] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[57] V. F. Pisarenko, "The retrieval of harmonics from a covariance function", *Geophysical Journal International*, vol. 33, no. 3, pp. 347–366, 1973.

[58] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques", *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[59] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ 1 minimization", *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

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

[61] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 301–320, 2005.

[62] J. Swärd, *PhD Thesis. Parameter Estimation - in sparsity we trust*. Mathematical Statistics, Centre for Mathematical Sciences, Lund University, 2017.

[63] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling", *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[64] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding", *Journal of Machine Learning Research*, vol. 11, no. Jan, pp. 19–60, 2010.

[65] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding", in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Omnipress, 2010, pp. 399–406.

[66] Ö. Batu and M. Çetin, "Hyper-parameter selection in non-quadratic regularization-based radar image formation", in *Algorithms for Synthetic Aperture Radar Imagery XV*, International Society for Optics and Photonics, vol. 6970, 2008, p. 697 009.

[67] P. Boufounos, M. F. Duarte, and R. G. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation", in *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*, IEEE, 2007, pp. 299–303.

[68] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, Simon, J. Taylor, and R. J. Tibshirani, "Strong rules for discarding predictors in lasso-type problems", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 2, pp. 245–266, 2012.

[69] J. Sun, H. Li, Z. Xu, *et al.*, "Deep admm-net for compressive sensing mri", in *Advances in Neural Information Processing Systems*, 2016, pp. 10–18.

[70] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing mri", *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[71] E. J. Candes, "The restricted isometry property and its implications for compressed sensing", *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[72] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit", *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[73] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[74] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance", *IEEE Journal of selected topics in signal processing*, vol. 4, no. 2, pp. 298–309, 2010.

[75] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[76] Mosek, APS. (2010). The mosek optimization software. [Accessed: 21.05.2018], [Online]. Available: `http://www.mosek.com`.

[77] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones", *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[78] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "Sdpt3—a matlab software package for semidefinite programming, version 1.3", *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[79] M. Grant, S. Boyd, and Y. Ye, *Cvx: Matlab software for disciplined convex programming*, 2008.

[80] 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.

[81] Nvidia Blog. (2013). Cuda spotlight: Gpu-accelerated neuroscience. [Accessed: 12.06.2018], [Online]. Available: `https : / / devblogs . nvidia.com/cuda-spotlight-gpu-accelerated-neuroscience/`.

[82] T. Kronvall, *PhD Thesis. Group-Sparse Regression: With Applications in Spectral Analysis and Audio Signal Processing*. Mathematical Statistics, Centre for Mathematical Sciences, Lund University, 2017.

[83] G.-O. Glentis and A. Jakobsson, "Efficient implementation of iterative adaptive approach spectral estimation techniques", *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4154–4167, 2011.

[84] M. Xue, L. Xu, and J. Li, "Iaa spectral estimation: Fast implementation using the gohberg–semencul factorization", *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3251–3261, 2011.

A.   Estimating Sparse Signals Using Integrated Wide-band Dictionaries

Maksim Butsenko[*], Johan Swärd[†], and Andreas Jakobsson[†]

[*]*Dept. of Radio and Comm. Eng., Tallinn University of Technology, Estonia*
[†]*Dept. of Mathematical Statistics, Lund University, Sweden*

Abstract

In this paper, we present a technique for reducing the size of the dictionary in sparse signal reconstruction by formulating an initial dictionary containing elements that spans bands of the considered parameter space. We allow for the use of this banded dictionary in a first-stage estimation procedure, in which large parts of the parameter space is discarded for further analysis, thereby reducing the overall computationally complexity required to allow for a reliable signal reconstruction. We illustrate the presented principle on the problem of estimating sinusoidal components corrupted by white noise.

Keywords: sparse signal reconstruction, dictionary learning, convex optimization

1. Introduction

A wide range of applications yields signals that may be well approximated using a sparse reconstruction framework, and the area has attracted dramatic interest in the recent literature (see, e.g., [1]–[3] and the references therein). Much of this work has focused on formulating convex algorithms that exploit different sparsity inducing penalties, thereby encouraging solutions that are well represented using just a few elements from some known dictionary matrix, $\mathbf{D}$. If the dictionary is appropriately chosen, even very limited measurements can be shown to allow for an accurate signal reconstruction [4], [5]. Recently, increasing attention has been given to signals that are best represented using a continuous parameter space. In such cases, the discretization of the parameter space that is typically used to approximate the true parameters will not represent the noise-free signal exactly, resulting in solutions that are less sparse than desired. This problem has been examined in, e.g., [6]–[8], wherein discretization recommendations and new bounds of the reconstruction guarantees were presented, taking the grid mismatch into consideration. Typically, this results in the use of large and over-complete dictionaries, which, although quite efficient, often violate the assumptions required to allow for a perfect recovery guaranty.

As an alternative, one may formulate the reconstruction problem using a continuous dictionary, such as in, e.g., [9]–[11]. Such formulations typically use an atomic norm penalty, as introduced in [12], which allows for a way to determine the most suitable convex penalty to recover the signal, even over a continuous parameter space. Such a solution often offers an accurate signal reconstruction, but typically requires one to solve large and rather complicated optimization problems, thereby limiting the size of the considered problem.

In this work, we examine an alternative way of approaching the problem, proposing the use of wide-band dictionary elements, such that the dictionary is formed over B subsets of the continuous parameter space. In the estimation procedure, the activated subsets are retained and refined, whereas non-activated sets are discarded from the further optimization. Without loss of generality, the proposed principle is here illustrated on the problem of estimating the frequencies of K complex-valued sinusoid corrupted by white circularly symmetric Gaussian noise. This is a classical estimation problem, originally expressed using a sparse reconstruction framework in [13], and having since attracting notable attention (see, e.g., [14]–[17]). Here, using the classical formulation, the resulting sinusoidal dictionary will allow for a K-sparse representation of frequencies on the grid, whereas the grid mismatch of any off-grid components will typically yield solutions with more than K components. Extending the dictionary to use a finely spaced dictionary, as suggested in, e.g., [8], will yield the desired solution, although at the cost of an increased complexity. In this work, we instead proceed to divide the spectrum into B (continuous) frequency bands, each band possibly containing multiple spectral lines. This allows for an initial coarse estimation of

the signal frequencies, without (significantly) increasing the risk of missing any off-grid components.

The proposed principle may also be used when solving the reconstruction problem using gridless methods, such as the methods in [9]–[11]. It has been shown that if the reconstruction problem allows for any prior knowledge about the location of the frequencies, e.g., the frequencies are located within a certain region of the spectrum, one may use this information to improve the estimates[18]. The proposed method may then be used for attaining such prior information, and thus improving the overall estimates as a result.

2. Problem statement

Consider the problem of estimating the frequencies f_k, for $k = 1, \ldots, K$, of a measured signal y_n, with

$$y_n = \sum_{k=1}^{K} \beta_k e^{2i\pi f_k t_n} + \epsilon_n \tag{1}$$

for $n = 1, \ldots, N$, and where K denotes the (unknown) number of sinusoids in the signal. Furthermore, let β_k and f_k denote the complex amplitude and frequency of the kth frequency, respectively, t_n the nth sample time, and ϵ_n the additive noise at time t_n. The classical sparse formulation of this estimation problem, as presented in [13], considers the LASSO minimization (see also [19])

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{Dx}||_2^2 + \lambda||\mathbf{x}||_1 \tag{2}$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_N \end{bmatrix}^T \tag{3}$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_1 & \cdots & \mathbf{d}_L \end{bmatrix} \tag{4}$$

$$\mathbf{d}_\ell = \begin{bmatrix} e^{2i\pi \hat{f}_\ell t_1} & \cdots & e^{2i\pi \hat{f}_\ell t_N} \end{bmatrix}^T \tag{5}$$

where $\hat{f}_\ell$ for $\ell = 1, \ldots, L$ denotes the $L \gg K$ candidate frequencies in the dictionary, $\mathbf{D}$, typically selected to be closely spaced to allow for minimal grid mismatch, and $(\cdot)^T$ the transpose. The penalty on the 1-norm of $\mathbf{x}$ will ensure that the found solution, $\hat{\mathbf{x}}$, will be sparse, with λ denoting a user parameter governing the desired sparsity level of the solution. The desired frequencies, as well as their order, are then found as the non-zero elements in $\hat{\mathbf{x}}$. As shown in [8], the number of dictionary elements, L, typically has to be large to allow for reliable high-resolution frequency estimates.

As an alternative, one may use a zooming procedure, where one first employ an initial coarse frequency dictionary, $\mathbf{D}_1$, and then employ a fine dictionary,
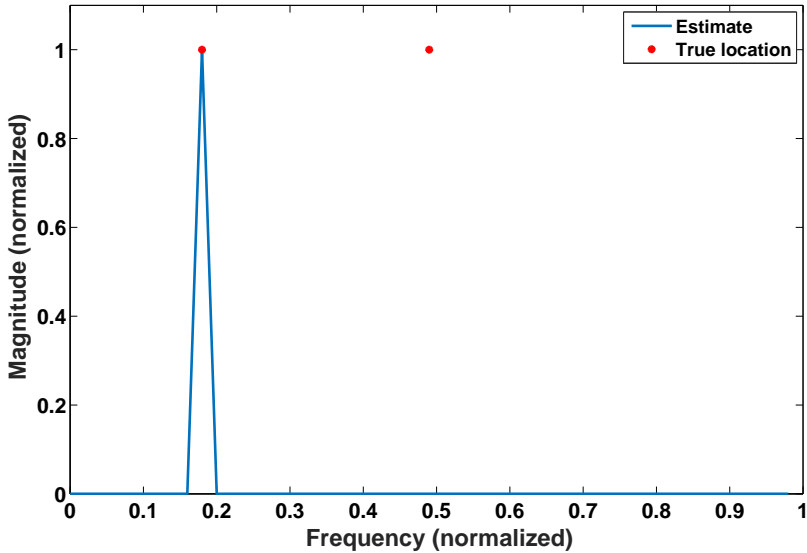
Figure 1 The inner-product of a dictionary containing $L = 50$ (narrowband) candidate frequency elements and the noise-free signal, with $N = 100$.

$\mathbf{D}_2$, centered around the initially found frequency estimates (see, e.g., [20], [21] for similar approaches). This allows for computationally efficient solution of the optimization problem in (3), but suffers from the problem of possibly missing off-grid components far from the initial coarse frequency grid. This is illustrated in Figure 2, where the inner-product between the dictionary and the signal is depicted together with the location of the true peaks. In this noise-free example, we used $N = 100$ samples and $L = 50$ dictionary elements, with one of the frequencies being situated in between two adjacent grid points in the dictionary. As seen in the figure, the coarse initial estimate fails to detect the presence of the second sinusoid, which is thereby discarded as a possibility in the following refined estimate. Increasing the number of candidate frequencies will result in that the side-lobes of the more finely spaced frequencies will lessen the gap between the frequency grid points, making the inner-product between the dictionary and the signal larger for sinusoidal components that lies between two candidate frequencies. However, doing so will increase computational complexity correspondingly, begging the question if one may retain a low number of candidate frequencies, while reducing the likelihood of missing any off-grid components. This is the problem we examine in the following.

3. Integrated Wide-band dictionaries

To allow for off-grid components, we here instead propose forming a wide-band dictionary over B frequency bands, with each integrated wide-band dictionary
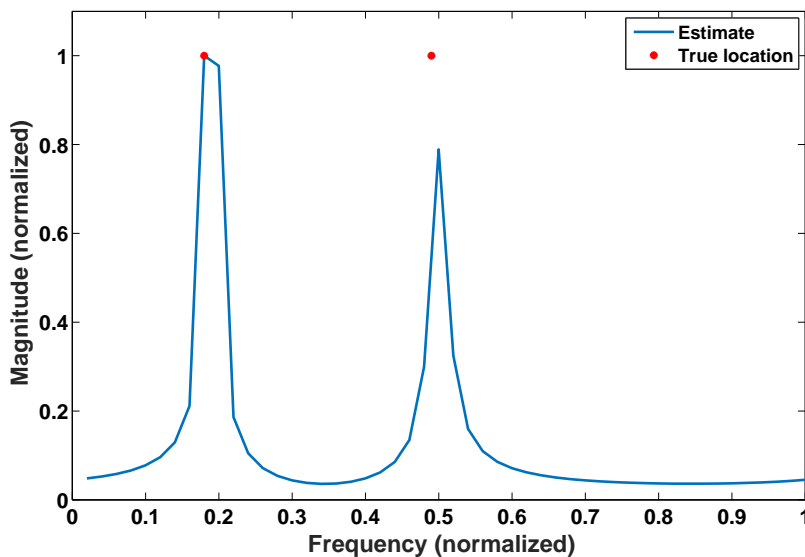
Figure 2 The inner-product of a dictionary containing $B = 50$ (wide-band) candidate frequency elements and the noise-free signal, with $N = 100$.

element being formed as

$$\mathbf{a}_b = \int_{f_b}^{f_{b+1}} e^{2i\pi ft} df \tag{6}$$

where f_b and f_{b+1} are the two frequencies bounding the frequency band, for $b = 1, \ldots, B$. The resulting elements are then gathered into the dictionary, $\mathbf{A}$, formed as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_B \end{bmatrix} \tag{7}$$

with the bth dictionary element at time t_n being formed as

$$a_{b,n} = \frac{e^{2i\pi f_{b+1}t_n} - e^{2i\pi f_b t_n}}{2i\pi t_n} \tag{8}$$

where $a_{b,n}$ denotes the nth element in column b of $\mathbf{A}$. The inner-product between the proposed dictionary, $\mathbf{A}$, and the earlier signal is shown in Figure 4, using the same number of dictionary elements as in that case, i.e., with $B = 50$, clearly indicating that the proposed dictionary is able to locate the off-grid frequency. This is due to the wide-band nature of the proposed dictionary, which thus has less power concentrated at the grid points, but covers a wider range of frequencies, not reducing to zero, or close to zero, anywhere within the band (as is the case for the narrowband dictionary elements). As a result, using the wide-band dictionary

57

elements, it is possible to use a smaller initial dictionary, thereby reducing the computational complexity, without increasing the risk of missing components in the signal.

4.   Efficient implementation

To form a computationally efficient solution of the problem and to showcase the complexity reduction provided by the method proposed in this paper, we proceed to solve (3) using the popular ADMM algorithm [22]. In order to do so, the variable $\mathbf{x}$ is split into two variables, here denoted $\mathbf{x}$ and $\mathbf{z}$, after which the (scaled) augmented Lagrangian may be formulated as

$$L_{\mathbf{x},\mathbf{z},\boldsymbol{u}} = ||\mathbf{y} - \mathbf{Ax}||_2^2 + \lambda ||\mathbf{z}||_1 + \rho ||\mathbf{x} - \mathbf{z} + \boldsymbol{u}||_2^2 \tag{9}$$

where $\boldsymbol{u}$ is the scaled dual variable and ρ is the step length (see [22] for a detailed discussion). The minimization is thus formed by iteratively solving (29) for $\mathbf{x}$ and $\mathbf{z}$, as well as updating the scaled dual variable $\boldsymbol{u}$. This is done by finding the (sub-)gradient for $\mathbf{x}$ and $\mathbf{z}$ of the augmented Lagrangian, and setting it to zero, fixing the other variables to their latest values. The steps for the jth iteration are thus

$$\mathbf{x}^{(j+1)} = \left(\mathbf{A}^H \mathbf{A} + \rho \mathbf{I} \right)^{-1} \left(\mathbf{A}^H \mathbf{y} + \mathbf{z}^{(j)} - \boldsymbol{u}^{(j)} \right) \tag{10}$$

$$\mathbf{z}^{(j+1)} = S(\mathbf{x}^{(j+1)} + \boldsymbol{u}^{(j)}, \lambda/\rho) \tag{11}$$

$$\boldsymbol{u}^{(j+1)} = \boldsymbol{u}^{(j)} + \mathbf{x}^{(j+1)} - \mathbf{z}^{(j+1)} \tag{12}$$

where $(\cdot)^H$ denotes the Hermitian transpose, $(\cdot)^{(j)}$ the jth iteration, and $S(\mathbf{x}, \kappa)$ the soft threshold operator, defined as

$$S(\boldsymbol{v}, \kappa) = \frac{\max\left(|\boldsymbol{v}| - \kappa, 0 \right)}{\max\left(|\boldsymbol{v}| - \kappa, 0 \right) + \kappa} \odot \boldsymbol{v} \tag{13}$$

where $\kappa \odot \boldsymbol{v}$ denotes the element-wise multiplication for any vector $\boldsymbol{v}$ and scalar κ. The computationally most demanding part of the resulting ADMM implementation is to form the inverse in (30) and to calculate $\mathbf{A}^H \mathbf{y}$. These steps are often done by QR factorizing the inverse prior to the iteration, so that this part is only calculated once, and then using the factors when forming the inner product. The total computational cost for the step in (30) depends on the size of the matrix $\mathbf{A}$ (or, correspondingly, $\mathbf{D}$, if using the narrowband dictionary). If $\mathbf{A}$ is an $N \times L$ matrix, and if $L < N$, computing the inverse will cost approximately L^3 operations, plus an additional $L^2 N$ operations to form the Gram-matrix $\mathbf{A}^H \mathbf{A}$. Furthermore, to compute $\mathbf{A}^H \mathbf{y}$ requires LN operations, and the final step to compute $\mathbf{x}$ costs L^2 operations. If instead $L > N$, one may make use of the Woodbury matrix identity [23], allowing the inverse to be formed using $N^3 + 3LN^2$ operations, whereafter one has to compute $\mathbf{A}^H \mathbf{y}$ and

Settings	Complexity ratio	Grid distance (10^{-3})
D1000	1	0.50
B20 Q25	31	1.0
B20 Q40	7	0.63
B40 Q25	26	0.50
B40 Q40	7	0.31
B75 Q25	16	0.27
B75 Q40	6	0.17
B75 Q323	1	0.02

Table 1 Complexity reduction compared to using the full dictionary and the distance between the final grid for different settings. Here, D1000 indicates the one-stage narrowband dictionary using a dictionary with $L = 1000$ elements, whereas B20 Q25 indicates the two-stage dictionary using $B = 20$ wide-band elements, followed by $Q = 25$ narrowband elements in the second-stage dictionary.

the final matrix-vector multiplication, together costing $LN + L^2$ operations. In total, the x-step will have the cost of roughly $L^3 + (N + 1)L^2 + NL$, if $L < N$, or $N^3 + 3LN^2 + LN + L^2$, if $N < L$.

Since using the banded dictionary allows for a smaller dictionary, one may calculate the computational benefit of using the integrated dictionary as compared to just using an ordinary dictionary with large L. Consider using only a single-stage narrowband dictionary, $\mathbf{D}_1$, with $L > N$ dictionary elements. This requires $C_1 = N^3 + 3LN^2 + L^2 + LN$ operations if using the above ADMM solution, with the dictionary $\mathbf{D}_1$ in place of $\mathbf{A}$ in (30)-(32). If, on the other hand, one uses a two-stage wide-band dictionary with N dictionary elements in the initial coarse dictionary, $\mathbf{A}_1$ (which is more than required, but simplifies the calculations), the cost of forming the first stage (coarse) minimization is $C_2 = 2(N^3 + N^2)$. By taking the difference, i.e., forming $R = C_1 - C_2 = N^3 + 3LN^2 + L^2 + LN - 2(N^3 + N^2)$, one obtains the available computational resources, R, that are left for a second stage dictionary, $\mathbf{A}_2$, without increasing the overall computational cost above that of the narrowband dictionary solution. Assuming that the $\mathbf{A}_2$ dictionary has $Z > N$ grid points available, one may deduce the grid size by solving $R = N^3 + 3N^2Z + Z^2 + ZN$, yielding that one is able to use a fine grid of $Z = (-3N^2 + \sqrt{9N^4 + 2N^3 + N^2 + 4R} - N)/2$ candidates in a secondary refinement step, without increasing the total computational complexity, as compared to using the single stage narrowband dictionary. To illustrate the resulting difference, consider the following settings: $L = 1000$ and $N = 100$, yielding $Z \approx 936$ grid points to be distributed over the activated bands. If the number of activated bands are three in the settings above, that would yield a grid separation of $1.6 \cdot 10^{-5}$, which should be compared to the ordinary dictionary having a grid separation of $5 \cdot 10^{-4}$; a difference of roughly a factor 31.
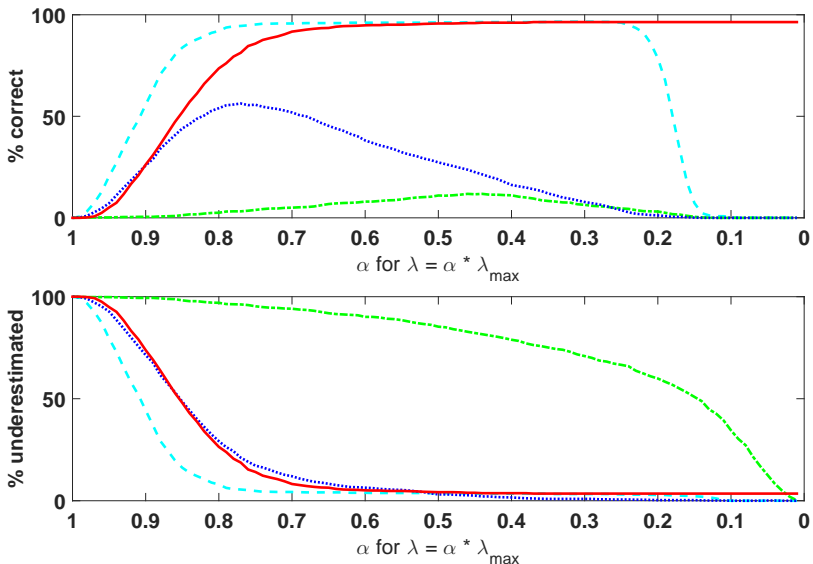
Figure 3 The probability of (top) correctly estimating and (bottom) underestimating the number of spectral lines, for the (single-stage) narrowband dictionary, using $L = 1000$ elements (cyan, dashed) and $L = 75$ elements (green, dot-dashed), and for the wide-band dictionary, using $B = 75$ elements (blue, dotted), and the (two-stage) wide-band dictionary, using $B = 75$ elements, together with $Q = 25$ elements per activated bands in the refining dictionary (red, solid).

5.   Numerical examples

In this section, we proceed to examine the performance of proposed method, initially illustrating that the proposed (two-stage) wide-band estimator has the same estimation quality as when using the ordinary (one-stage) Lasso estimator. We considered a signal consisting of $N = 75$ samples containing $K = 3$ (complex-valued) sinusoids corrupted by a zero-mean white Gaussian noise with signal-to-noise ratio (SNR) of SNR= 10dB. In each simulation, the sinusoidal frequencies are drawn from a uniform distribution, over $[0, 1)$, and all the amplitudes have magnitude 1 and phase drawn from a uniform distribution, over $[0, 2\pi)$. The performance is then computed using three different dictionaries, namely the (ordinary) narrowband dictionary, $\mathbf{D}$, with $L = 1000$ and $L = 75$ elements, respectively, and the proposed wide-band dictionary, $\mathbf{A}$, using $B = 75$ elements, followed by a second-stage narrowband dictionary using $Q = 25$ elements per active band. For each dictionary, we evaluate the performance for varying values of the user parameter α using $\lambda = \alpha\lambda_{max}$, where $\lambda_{max} = \max_i |\mathbf{x}_i^H \mathbf{y}_i|$ is the smallest tuning parameter value for which all coefficients in the solution are zero [24]. Each estimated result is then compared to the ground truth, counting the number of correct and underestimated
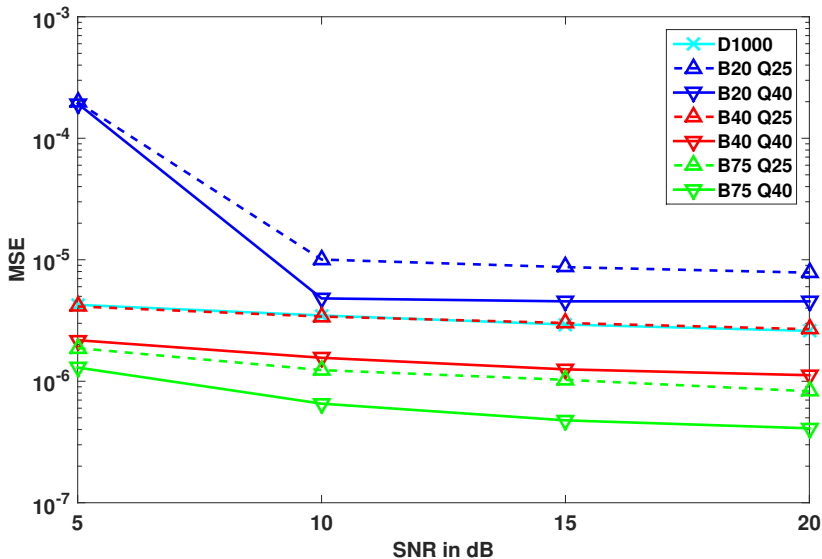
Figure 4 Mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $L = 1000$, as compared to the two-stage dictionary, using B integrated wide-band elements in the first stage, followed by Q narrowband elements in the second stage.

model order estimates. The result is shown in Figure 10. As can be seen from the figure, the best results are achieved when $\alpha \leq 0.65$, in which case the proposed wide-band dictionary, using $B = 75$ bands, followed by a second stage narrowband dictionary, with $Q = 25$ per activated band, have similar performance to the narrowband dictionary using $L = 1000$ dictionary elements. Proceeding, we asses the mean-square error (MSE) for different settings of the two-stage dictionary, showing the MSE as a function of SNR for various sizes of the first-stage wide-band dictionary (B) and second-stage narrowband refining dictionary (Q). Figure 2 shows the resulting MSE, for the estimates with correctly estimated model order; Table 1 shows the corresponding complexity cost and the final grid distance of the second-stage dictionary. As can be seen from the figure, the two-stage dictionary using a wide-band dictionary, with $B = 40$ bands, followed by a refining dictionary using $Q = 25$ narrowband elements, achieves the same performance as the single-stage narrowband dictionary using $L = 1000$ elements, although the latter requires about 26 times fewer operations. Furthermore, it may be noted that using the same overall complexity, as resulting from using $B = 75$ and $Q = 323$, we achieve 25 times higher resolution as compared to the single-stage dictionary. All results are computed using 1000 Monte-Carlo simulations.

References

[1] M.Unser and P.Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2013.

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

[3] E. J. Candès and M.B. Wakin, "An introduction to compressive sampling", *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[4] E. J. Candès, 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.

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

[6] M. A. Herman and T. Strohmer, "Genral Deviants: An Analysis of Perturbations in Compressed Sensing", *IEEE J. Sel. Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[7] 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.

[8] 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.

[9] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed Sensing Off the Grid", *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[10] Y. Chi and Y. Chen, "Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization", *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[11] Z. Yang and L. Xie, "Enhancing Sparsity and Resolution via Reweighted Atomic Norm Minimization", *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb. 2016.

[12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems", *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[13] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra", in *17th World Congress IFAC*, Seoul, Jul. 2008, pp. 10 225–10 229.

[14] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data", *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

[15] 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, Jul. 2012.

[16] 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, Mar. 1997.

[17] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Multi-Pitch Estimation Exploiting Block Sparsity", *Elsevier Signal Processing*, vol. 109, pp. 236–247, Apr. 2015.

[18] Z. Yang and L. Xie, "Frequency-Selective Vandermonde Decomposition of Toeplitz Matrices With Applications", *Publication: eprint arXiv:1605.02431*, 2016.

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

[20] E. H. D. S. Sahnoun and D. Brie, "Sparse Modal Estimation of 2-D NMR Signals", in *38th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[21] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying N-dimensional Signals", *Elsevier Signal Processing*, vol. 128, pp. 309–317, Nov. 2016.

[22] 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.

[23] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th. The John Hopkins University Press, 2013.

[24] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, Simon, J. Taylor, and R. J. Tibshirani, "Strong rules for discarding predictors in lasso-type problems", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 2, pp. 245–266, 2012.

B

B. Estimating Sparse Signals Using Integrated Wide-band Dictionaries

Maksim Butsenko[*], Johan Swärd[†], and Andreas Jakobsson[†]

[*]*Thomas Johann Seebeck Dept. of Electronics, Tallinn University of Technology, Estonia*
[†]*Dept. of Mathematical Statistics, Lund University, Sweden*

Abstract

In this paper, we introduce a wideband dictionary framework for estimating sparse signals. By formulating integrated dictionary elements spanning bands of the considered parameter space, one may efficiently find and discard large parts of the parameter space not active in the signal. After each iteration, the zero-valued parts of the dictionary may be discarded to allow a refined dictionary to be formed around the active elements, resulting in a zoomed dictionary to be used in the following iterations. Implementing this scheme allows for more accurate estimates, at a much lower computational cost, as compared to directly forming a larger dictionary spanning the whole parameter space or performing a zooming procedure using standard dictionary elements. Different from traditional dictionaries, the wideband dictionary allows for the use of dictionaries with fewer elements than the number of available samples without loss of resolution. The technique may be used on both one- and multi-dimensional signals, and may be exploited to refine several traditional sparse estimators, here illustrated with the LASSO and the SPICE estimators. Numerical examples illustrate the improved performance.

Keywords: sparse signal reconstruction, dictionary learning, convex optimization

1.   Introduction

A wide range of common applications yield signals that may be well approximated using a sparse reconstruction framework, and the area has as a result attracted notable interest in the recent literature (see, e.g., [1]–[3] and the references therein). Much of this work has focused on formulating convex algorithms that exploit different sparsity inducing penalties, thereby encouraging solutions that are well represented using only a few elements from some (typically known) dictionary matrix, $\mathbf{D}$. If the dictionary is appropriately chosen, even very limited measurements can be shown to allow for an accurate signal reconstruction [4], [5]. Recently, increasing attention has been given to signals that are best represented using a continuous parameter space. In such cases, the discretization of the parameter space that is typically used to approximate the true parameters will not represent the noise-free signal exactly, resulting in solutions that are less sparse than desired. This problem has been examined in, e.g., [6]–[8], wherein discretization recommendations and new bounds of the reconstruction guarantees were presented, taking possible grid mismatches into consideration. Typically, this results in the use of large and over-complete dictionaries, which, although quite efficient, often violate the assumptions required to allow for a perfect recovery guarantee.

As an alternative, one may formulate the reconstruction problem using a continuous dictionary, such as in, e.g., [9]–[11]. This kind of formulations typically use an atomic norm penalty, as introduced in [12], which allows for a way to determine the most suitable convex penalty to recover the signal, even over a continuous parameter space. These solutions often offer an accurate signal reconstruction, but also require the solving of large and computationally rather cumbersome optimization problems, thereby limiting the size of the considered problems.

In this work, we examine an alternative way of approaching the problem, proposing the use of wideband dictionary elements, such that the dictionary is formed over B subsets of the continuous parameter space. In the estimation procedure, the activated subsets are retained and refined, whereas non-activated sets are discarded from the further optimization. This screening procedure may be broken down into two steps. The first step is to remove the parts of the parameter space not active in the signal, whereafter, in the second step, a smaller dictionary is formed covering only the parts of the parameter space that were active in the first step. This smaller dictionary may then again be expanded with candidates close to the activated elements, thereby yielding a zoomed dictionary in these regions. The process may then be repeated to further refine the estimates as desired. Without loss of generality, the proposed principle is here illustrated on the problem of estimating the frequencies of K complex-valued M-dimensional sinusoid corrupted by white circularly symmetric Gaussian noise. The one-dimensional case of this is a classical estimation problem, originally expressed using a sparse

reconstruction framework in [13], and having since attracting notable attention (see, e.g., [14]–[17]). Here, using the classical formulation, the resulting sinusoidal dictionary will allow for a K-sparse representation of frequencies on the grid, whereas the grid mismatch of any off-grid components will typically yield solutions with more than K components. Extending the dictionary to use a finely spaced dictionary, as suggested in, e.g., [8], will yield the desired solution, although at the cost of an increased complexity. In this work, we instead proceed to divide the spectrum into B (continuous) frequency bands, each band possibly containing multiple spectral lines. This allows for an initial coarse estimation of the signal frequencies, without the risk of missing any off-grid components. Due to the iterative refining of the dictionary, closely spaced components are successfully separated as the dictionary is refined; as the wideband elements span the full band, no power is off-grid, avoiding the problem of a non-sparse solution due to dictionary mismatch.

Other screening methods that decrease the dictionary size have been proposed. For instance, in [18]–[23], methods for finding the elements in the dictionary that corresponds to zero-valued elements in the sparse vector were proposed. Based on the inner product between the large dictionary and the signal, a rule was formed for deeming whether or not a dictionary element was present in the signal or not. Although these methods show a substantial decrease in computational complexity, one still has to form the inner product between the likely large dictionary and the signal. To alleviate this, one may instead use the here proposed wideband dictionary elements, thereby discarding large parts of the parameter space. Since the wideband dictionary is magnitudes smaller than the full dictionary required to achieve the reconstruction, the computational complexity is significantly reduced.

The proposed principle is not limited to methods that use discretization of the parameter space; it may also be used when solving the reconstruction problem using gridless methods, such as the methods in [9]–[11]. It has been shown that if the reconstruction problem allows for any prior knowledge about the location of the frequencies, e.g., the frequencies are located within a certain region of the spectrum, one may use this information to improve the estimates[24]. The proposed method may also be used to attain such prior information, and thus improving the overall estimates as a result.

To illustrate the performance of the proposed dictionary, we make use of two different sinusoidal estimators, namely the LASSO [25] and the SPICE estimators [26], [27]; the first finding the estimate by solving a penalized regression problem, whereas the latter instead solves a covariance fitting problem.

The remainder of this paper is organized as follows: in the next section, the problem of estimating an M-dimensional sinusoidal signal is introduced, followed, in Section III, by the introduction of the proposed wideband dictionary. In Section IV, a discussion about the computational complexity reduction allowed by the proposed wideband dictionary is given, and, in Section V, the performance

of the proposed wideband dictionary is illustrated by numerical examples. Finally, in Section VI, we conclude on our work.

2. Problem statement

To illustrate the wideband dictionary framework consider the problem of estimating the K frequencies $f_k^{(m)}$, for $k = 1, \ldots, K$ and $m = 1, \ldots M$, of an M-dimensional signal $y_{n_1, \ldots, n_M}$, with

$$y_{n_1, \ldots, n_M} = \sum_{k=1}^{K} \beta_k e^{2i\pi f_k^{(1)} t_{n_1}^{(1)} + \cdots + 2i\pi f_k^{(M)} t_{n_M}^{(M)}} + \epsilon_{n_1, \ldots, n_M} \tag{1}$$

for $n_m = 1, \ldots, N_m$, and where K denotes the (unknown) number of sinusoids in the signal. Furthermore, let β_k and $f_k^{(m)}$ denote the complex amplitude and frequency of the kth frequency and mth dimension, respectively, $t_{n_m}^{(m)}$ the n_mth sample time in the mth dimension, and $\epsilon_{n_1, \ldots, n_M}$ an additive noise observed at time $t_{n_1}, \ldots, t_{n_M}$. The signal model in (1) may be equivalently described by an M-dimensional (M-D) tensor

$$\mathcal{Y} = \sum_{k=1}^{K} \beta_k \tilde{d}_{(k)}^{(1)} \circ \tilde{d}_{(k)}^{(2)} \cdots \circ \tilde{d}_{(k)}^{(M)} + \mathcal{E} \tag{2}$$

where $\circ$ denotes the outer product, and

$$\tilde{d}_{(k)}^{(m)} = \left[\begin{array}{ccc} e^{2i\pi f_k^{(m)} t_1^{(m)}} & \ldots & e^{2i\pi f_k^{(m)} t_{Nm}^{(m)}} \end{array} \right]^T \tag{3}$$

To determine the parameters of the model in (1) or (2), as well as the model order, we proceed by creating a dictionary containing a set of signal candidates, each representing a sinusoid with a unique frequency. By measuring the distance between the signal candidates and the measured signal, and by promoting a sparse solution, one may find a small set of candidates that best approximates the signal. To this end, we form a dictionary on the form

$$\mathbf{D}^{(m)} = \left[\begin{array}{ccc} d_1^{(m)} & \ldots & d_{P_m}^{(m)} \end{array} \right] \tag{4}$$

$$d_{(p)}^{(m)} = \left[\begin{array}{ccc} e^{2i\pi f_p^{(m)} t_1^{(m)}} & \ldots & e^{2i\pi f_p^{(m)} t_{Nm}^{(m)}} \end{array} \right]^T \tag{5}$$

for $m = 1, \ldots, M$ and $p = 1, \ldots, P_m$, where $P_m \gg K$ denotes the number of candidates in dimension m. Here, the dictionary is assumed to be fine enough so that the unknown sinusoidal component will (reasonably well) coincide with K dictionary elements[1]. Often, it is more convenient to work with a vectorized

---

[1] As noted in [7], [8], the dictionary generally needs to be selected sufficiently fine to allow for a reconstruction of the signal, whereas increasing the size of the dictionary will also increase the computational complexity of the estimate. As shown in the following, the discussed method relaxes this requirement by instead defining a dictionary covering bands of potential candidates, rather than a set of individual dictionary candidates.
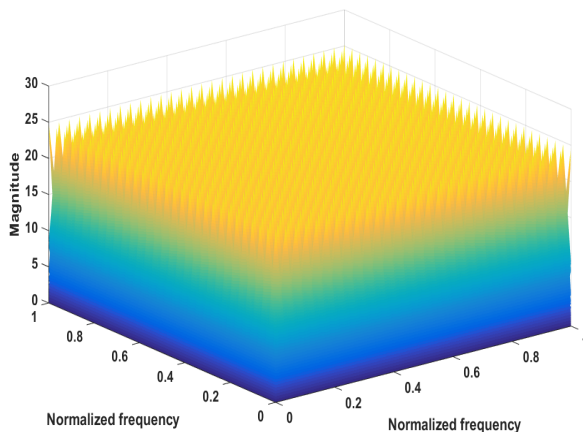
Figure 1 Fine-grid dictionary for two-dimensional signal estimation with $N_1 = 30$, $N_2 = 30$, and $P = 60$ elements per dimension.

version of the tensor. Let $\mathbf{y} = \text{vec}(\mathcal{Y})$, where $\text{vec}(\cdot)$ stacks the tensor into a vector. One may then re-write (2) as

$$\mathbf{y} = \left(\mathbf{D}^{(M)} \otimes \mathbf{D}^{(M-1)} \otimes \cdots \otimes \mathbf{D}^{(1)} \right) \boldsymbol{\beta} \tag{6}$$

where $\otimes$ denotes the Kronecker product, suggesting that one may find both the unknown parameters and the model order by forming the LASSO problem (see, e.g., [13], [25])

$$\min_{\boldsymbol{\beta}} ||\mathbf{y} - \mathbf{D}\boldsymbol{\beta}||_2^2 + \lambda ||\boldsymbol{\beta}||_1 \tag{7}$$

where $\mathbf{D} = \left(\mathbf{D}^{(M)} \otimes \mathbf{D}^{(M-1)} \otimes \cdots \otimes \mathbf{D}^{(1)} \right)$ and $||\cdot||_q$ denotes the q-norm. A visual representation of such dictionary is shown in Figure 1 for the 2-D case. The penalty on the 1-norm of $\boldsymbol{\beta}$ will ensure that the found solution will be sparse, with λ denoting a user parameter governing the desired sparsity level of the solution. The frequencies, as well as their order, are then found as the non-zero elements in $\boldsymbol{\beta}$.

As shown in [8], the number of dictionary elements, P, typically has to be large to allow for an accurate determination of the correct parameters. This means that for multi-dimensional signals, the dictionary quickly becomes inhibitory large. Thus, it is often not feasible in practice to directly compute the solution of (7) using a dictionary constructed from such finely space candidates. As an alternative, one may use a zooming procedure, where one first employs an initial coarse dictionary, $\mathbf{D}_1$, to determine the parameter regions of interest, and then employ a fine dictionary, $\mathbf{D}_2$, centered around the initially found candidates (see, e.g., [28], [29] for similar approaches). This allows for a computationally
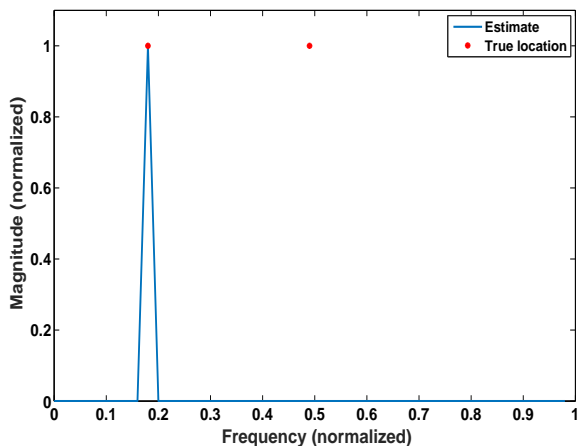
Figure 2 The inner-product of a dictionary containing $P = 50$ (narrowband) candidate frequency elements and the noise-free signal, with $N = 100$.

efficient solution of the optimization problem in (7), but suffers from the problem of possibly missing off-grid components far from the initial coarse frequency grid. This is illustrated in Figure 2 for a 1-D signal, where the inner-product between the dictionary and the signal is depicted together with the location of the true peaks. In this noise-free example, we used $N = 100$ samples and $P = 50$ dictionary elements, with one of the frequencies being situated in between two adjacent grid points in the dictionary. As seen in the figure, the coarse initial estimate fails to detect the presence of the second signal component, which is thereby discarded as a possibility in the following refined estimate. Increasing the number of candidate frequencies will result in the side-lobes of the dictionary elements decreasing the gap between the frequency grid points, making the inner-product between the dictionary and the signal larger for components that lie in between two candidate frequencies. However, doing so will increase the computational complexity correspondingly, begging the question if one may retain a low number of candidate frequencies, while still reducing the likelihood of missing any off-grid components. This is the problem we shall examine in the following.

3.   Integrated Wideband dictionaries

We note that the above problem results from the dictionary being formed over a set of single-component candidates, thereby increasing the risk of neglecting the off-grid components. In order to avoid this, we here propose a wideband dictionary framework, such that each of the dictionary elements is instead formed over a range of such single-component candidates. This is done by letting
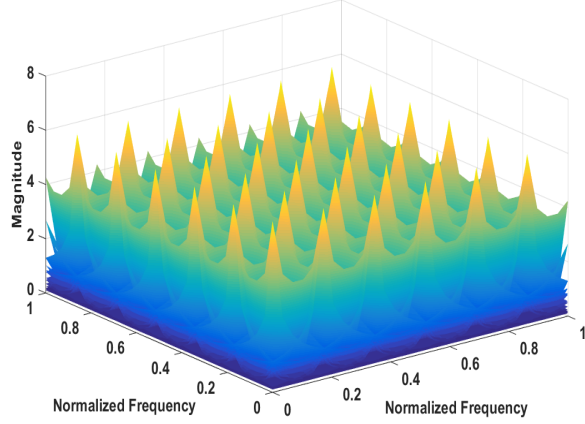
Figure 3 Wideband dictionary with integrated sinusoids elements formed with $N_1 = 30$, $N_2 = 30$, and $B = 6$ bands per dimension.

the dictionary elements be formed over an integrated range of the parameter(s) of interest, in this case being the frequencies of the candidate sinusoids. For a multi-dimensional sinusoidal dictionary, the resulting B integrated wideband elements should thus be formed as

$$
a_{b^{(1)},\ldots,b^{(M)}}\left(t^{(1)},\ldots,t^{(M)}\right) =
$$
$$
\int_{f_{b^{(1)}}}^{f_{b^{(1)}+1}} \ldots \int_{f_{b^{(M)}}}^{f_{b^{(M)}+1}}
$$
$$
e^{2i\pi\left(f^{(1)}t^{(1)}+\cdots+f^{(M)}t^{(M)}\right)}df^{(1)}\ldots df^{(M)} \tag{8}
$$

for $t^{(m)} = 1,\ldots,N_m$ for all $m = 1,\ldots,M$, where $f_{b^{(m)}}$ and $f_{b^{(m)}+1}$ are the two frequencies bounding the frequency band, for $b = 1,\ldots,B$, for the mth dimension. The resulting elements are then gathered into the dictionary, $\mathbf{B}$, where each column contains a specific wideband of the M-D parameter space for all time samples, where each element is formed as the solution from (8), such that, in this case,

$$
a_{b^{(1)},\ldots,b^{(M)}}\left(t^{(1)},\ldots,t^{(M)}\right) =
$$
$$
\prod_{m=1}^{M} \frac{e^{2i\pi f_{b^{(m)}+1}t^{(m)}} - e^{2i\pi f_{b^{(m)}}t^{(m)}}}{2i\pi t^{(m)}} \tag{9}
$$

Note that (9) corresponds to the M-D inverse Fourier transform of 1, i.e., it is the M-D inverse Fourier transform of an M-D section in the frequency domain
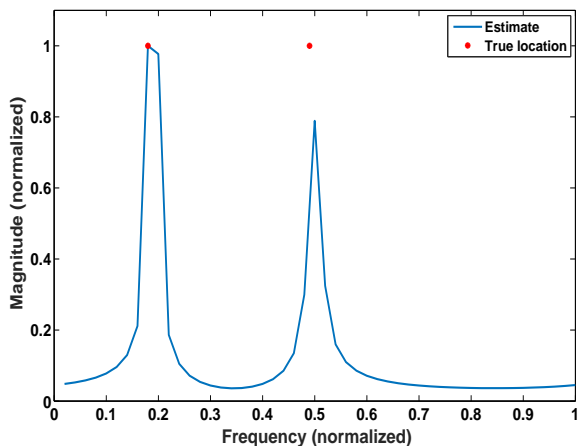
73

Figure 4 The inner-product of a dictionary containing $B = 50$ (wideband) candidate frequency elements and the noise-free signal, with $N = 100$.

with unit amplitude. For the 1-D case, this simplifies to

$$
\begin{cases} 1, & \text{for } f_a \leq f \leq f_b \\ 0 \end{cases} \quad \xrightarrow{\mathcal{F}^{-1}} \quad \frac{e^{2i\pi f_b t} - e^{2i\pi f_a t}}{2i\pi t} \tag{10}
$$

Algorithm 1 summarizes the usage of the wideband dictionary in a sparse reconstruction framework. In Figure 3, we show a visual representation of the resulting wideband dictionary for $M = 2$ dimensions. The inner-product between the proposed dictionary, **B**, and the earlier 1-D signal is shown in Figure 4, using the same number of dictionary elements as in Figure 1, clearly indicating that the proposed dictionary is able to locate the off-grid frequency. This is due the wideband nature of the proposed dictionary, which thus has less power concentrated at the grid points, but covers a wider range of frequencies, not reducing to zero, or close to zero, anywhere within the band (as is the case for the narrowband dictionary elements). As a result, using the wideband dictionary elements, it is possible to use a smaller dictionary, thereby reducing the computational complexity, without increasing the risk of missing components in the signal. To further show this, 1000 Monte-Carlo simulations were conducted for each considered signal to noise ratio (SNR), here defined as

$$
\text{SNR} = 10\log_{10}\left(\frac{P_y}{\sigma^2}\right) \tag{11}
$$

where P_y is the power of signal, and σ^2 the variance of the noise. In each simulation, we considered a signal containing two sinusoids, where the frequencies were randomly selected on $(0, 1]$ with a spacing of at least $2/N$, with $N = 100$ denoting the signal length. The sinusoids had the magnitudes 4 and 5,

Algorithm 1 Sparse reconstruction with LASSO using the wideband dictionary for the 1-D case

1: choose the number of zooming steps, I_{zoom}
2: choose the number of bands, B_1
3: set the frequency bin $\Delta_1 = \frac{1}{B_1}$
4: $\mathcal{F}_1 = \{f_k : f_k = k\Delta_1, \text{for } k = 1, \ldots, B_1\}$
5: form the dictionary $\mathbf{B}_1$ according to (9)
6: solve $\min_{\boldsymbol{\beta_1}} ||\mathbf{y} - \mathbf{B_1}\boldsymbol{\beta_1}||_2^2 + \lambda||\boldsymbol{\beta_1}||_1$
7: $\mathcal{I}_1 = \{i : \beta_1(i) > 0, \text{for } i = 1, \ldots, B_1\}$
8: $\mathcal{F}_{active} = \{f_k \in \mathcal{F}_1 : k \in \mathcal{I}_1\}$
9: **for** $z = 2$ to I_{zoom} **do** {zooming procedure}
10:    choose the number of bands, B_z
11:    select the frequency bin $\Delta_z = \frac{\Delta_{z-1}}{B_z}$
12:    $\mathcal{F}_z = \{\mathbf{f}_k : \mathbf{f}_k = [f_k + \Delta_z, f_k + 2\Delta_z, \ldots, f_k + B_z\Delta_z]^T, f_k \in \mathcal{F}_{active}\}$
13:    form the dictionary $\mathbf{B}_z$ according to (9)
14:    solve $\min_{\boldsymbol{\beta_z}} ||\mathbf{y} - \mathbf{B}_z\boldsymbol{\beta_z}||_2^2 + \lambda||\boldsymbol{\beta_z}||_1$
15:    $\mathcal{I}_z = \{i : \beta_z(i) > 0, \text{for } i = 1, \ldots, \prod_1^z B_z\}$
16:    $\mathcal{F}_{active} = \{f_k \in \mathcal{F}_z : k \in \mathcal{I}_z\}$
17: **end for**

with a randomly selected phase between $(0, 2\pi]$. Two dictionaries were given, one containing ordinary sinusoids and one containing the proposed wideband components, both containing $P = B = 50$ elements. For each dictionary, the inner-products with the signal where computed, where the amplitudes were normalized so that the largest estimated peak had unit magnitude. Figure 5 shows the variance of the smallest peak for different SNR-levels. As is clear from the figure, the variance of the peaks are much lower for the banded case. The reason why the sinusoidal dictionary results in a larger variance is due to the fact that the main lobe is much thinner in this case than in the banded counterpart. This means that when the sinusoids happen to have frequencies that do not overlap with the main lobe of the dictionary, the power in the inner-product will be small. This will not only make such components harder to detect, but will also make it more difficult to determine a suitable hyperparameter, λ.

When P decreases below N, the gaps between the frequency candidates in the single-component dictionary become so large that if one of the sinusoids in the signal has its frequency values between two adjacent grid points, the likelihood that this sinusoid lie in the null-space of the dictionary increases. This problem is avoided with the wideband dictionary as it is more likely to eliminate any gaps.

This property is depicted in Figure 6, where the success rate of finding the true support is displayed as a function of the number of samples, N, and the number of bands in the dictionary, B, for different number of sinusoids in the
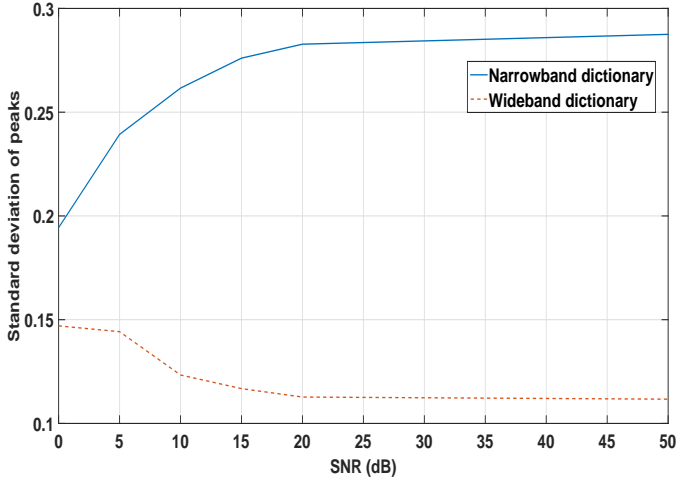
Figure 5 The standard deviation of the peaks as a function of SNR.

signal, K. The estimation was done for a noise-free signal by solving (7), using wideband dictionaries and letting

$$\lambda = 0.3 \max_{i=1,\dots,B} |\boldsymbol{d}_i^T \mathbf{y}| \tag{12}$$

where $\boldsymbol{d}_i$ denotes the ith column of $\mathbf{D}$ and the coefficient 0.3 is selected given the observations in Figure 10. For a more complete discussion on how one should select λ, we refer the reader to the original presentation of the LASSO [25]. In the top left figure, the signal contains three sinusoids, and it is clearly the case that the banded dictionary is able to retrieve the true support for all setting of N and B/N, except for the case when $N = 30$ and $B/N < 7$. In the top right and bottom figures, where $K = 7$ and $K = 11$, respectively, it is shown that when the number of sinusoids in the signal increases, a larger number of samples is needed to allow for a successful reconstruction, which is reasonable, as one needs more information to be able to correctly estimate more parameters. However, the banded dictionary is able to retrieve the true support as long as the number of samples is big enough and the ratio B/N is not too small. It is further clear from the figures, that the banded dictionary actually retrieves the true support even though $B < N$.

The proposed approach is not the only way to form a wideband dictionary. For example, one could populate the dictionary using discrete prolate spheroid sequences (DPSS) [30]. For an integer Q and with real-valued $0 < W < \frac{1}{2}$, the DPSS are a set of Q discrete-time sequences for which the amplitude spectrum is band-limited. The most interesting property of the DPSS for our discussion is the fact that the energy spectrum of the dictionary elements are highly concentrated in the range $[-W, W]$, suggesting that the DPSS could be a suitable basis for the candidates in a wideband dictionary, where the candidates are formed such that
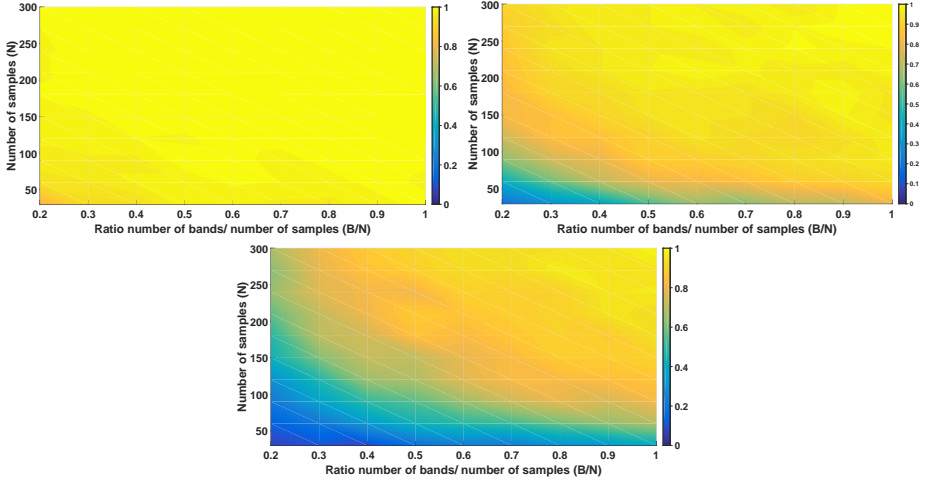
Figure 6 The success rate of finding the true support as a function of the number of samples (y-axis) and the ratio between the number of bands in the dictionary and the number of samples (x-axis), for different values of K. Top left corner, $K = 3$, top right corner, $K = 7$, and bottom, $K = 11$.

each covers a $1/B$-th part of the spectrum. In the numerical section below, we examine how the use of DPSS candidates compare to the integrated wideband candidates in (9). It is worth stressing that the wideband dictionary framework introduced here is not limited to the LASSO-style minimizations such as the one examined in (7). There are many other popular methods that could be implemented using this approach. As an example of how the wideband dictionary can be applied for other typical sparse estimation algorithms, consider the SPICE algorithm [14], [27], formed as the solution to

$$\underset{\tilde{\mathbf{p}} \geq 0}{\text{minimize}} \; \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} + ||\boldsymbol{W}_p \mathbf{p}||_1 + ||\boldsymbol{W}_\sigma \boldsymbol{\sigma}||_1, \tag{13}$$

where

$$\mathbf{R}(\tilde{\mathbf{p}}) = \mathbf{APA}^* \tag{14}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{I} \end{bmatrix} \tag{15}$$

$$\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_M \end{bmatrix}^T \tag{16}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 & \cdots & \sigma_N \end{bmatrix}^T \tag{17}$$

$$\tilde{\mathbf{p}} = \begin{bmatrix} \mathbf{p}^T & \boldsymbol{\sigma}^T \end{bmatrix}^T \tag{18}$$

$$\mathbf{P} = \text{diag}(\tilde{\mathbf{p}}) \tag{19}$$

and

$$W_p = \text{diag}\left(\begin{bmatrix} w_1 & \ldots & w_P \end{bmatrix}\right) \tag{20}$$

$$W_\sigma = \text{diag}\left(\begin{bmatrix} w_{P+1} & \ldots & w_{P+N} \end{bmatrix}\right) \tag{21}$$

$$w_k = ||\mathbf{a}_k||_2^2/||\mathbf{y}||_2^2, \text{ for } k = 1, \ldots, P + N \tag{22}$$

Alternatively, one may consider the more general $\{r, q\}$-SPICE formulation[2] [31]

$$\underset{\tilde{\mathbf{p}} \geq 0}{\text{minimize}}\ \mathbf{y}^*\mathbf{R}^{-1}\mathbf{y} + ||W_p\mathbf{p}||_r + ||W_\sigma\boldsymbol{\sigma}||_q \tag{23}$$

Using the wideband dictionary over $\mathbf{B}$ in (13) or (23) will allow for much smaller dictionaries as opposed to using ordinary sinusoidal dictionaries. Many other sparse reconstruction techniques may be extended similarly. Generally, the wideband dictionary may be used either as an energy detector which finds the parts of the spectrum that have most energy, or in a zooming procedure similar to the one described above.

4.   Parameter Selection

From our discussion on the integrated wideband dictionary and its use for sparse signal estimation, one may note that there are two parameters which should be chosen by the user, namely the number of used bands and the number of zooming steps. The choice of the number of the bands, B, will depend on the required resolution, whereas the number of zooming steps will decrease the computation complexity (for a fixed resolution) as with each zooming step inactive parts of the spectra are discarded from future computations. Therefore, the choice of the total number of bands one should use is dependent on the required resolution. Furthermore, for each zooming step, the distribution of these bands should be made such that the subsequent selection will guarantee a high likelihood of including the true support. This idea is illustrated in Figure 6, where the success rate of finding the true support is shown to depend on the number of bands, the number of samples, and the number of components in the data. As may be expected, the use of the wideband dictionary does not remove such user choices; in fact, the here proposed framework does not remove any of the usual user choices or limitations of a sparse reconstruction technique, be it the LASSO, SPICE, or any other dictionary based technique, and the same restrictions will apply that do so for the particular method if used with a narrowband dictionary. Rather, the wideband dictionary allows for an efficient refinement procedure speeding up the calculations required in forming the estimate.

---

[2]In this formulation, we assume that the columns of the dictionaries are normalized to have norm equal to $1/||\mathbf{y}||_2^2$.
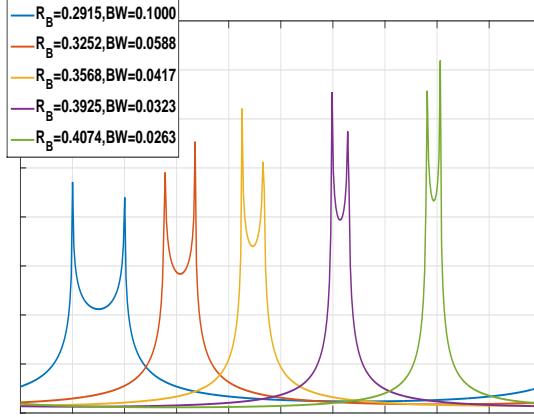
Figure 7 The figure shows wideband elements with varying bandwidths (BW), with the resulting values of R_B, plotted next to each other using the same scale to illustrate the difference in gain as the bandwidth of the band varies.

On the other hand, the use of a wideband dictionary does introduce the need to select the number of used bands, which directly relates to the width of the used bands as these bands are assumed to span the full parameter range. Due to the integration, the wideband elements will suffer from a reduced gain in the middle of the covered band; this will be negligible for bands of limited width, but will be pronounced, and will affect the estimation results, for wider bands. This is illustrated in Figure 7, showing a single wideband element for varying bandwidths. As shown in the figure, the ratio between the minimum and maximum gains of the wideband element, here denoted R_B, will depend on the bandwidth of the band, and is thus related to both the number of bands, B, and the number of samples in the signal, N.

To examine this aspect further, we proceed by formulating the ratio R_B for the case of one-dimensional dictionaries. Introducing $\Delta = f_b - f_a$, (10) may be expressed as

$$a(k) = e^{2i\pi f_a k} e^{i\pi \Delta k} \psi_k / \pi k, \quad \text{for } k \in \mathbb{Z}_+ \tag{24}$$

where $\psi_k = \sin(\pi \Delta k)$, implying that the discrete Fourier transform of $a(k)$ may be expressed as

$$G(f) = \Delta + \sum_{k=1}^{N-1} \frac{e^{i\pi \Delta k} \psi_k}{\pi k} e^{2i\pi(f_a - f)k} \tag{25}$$
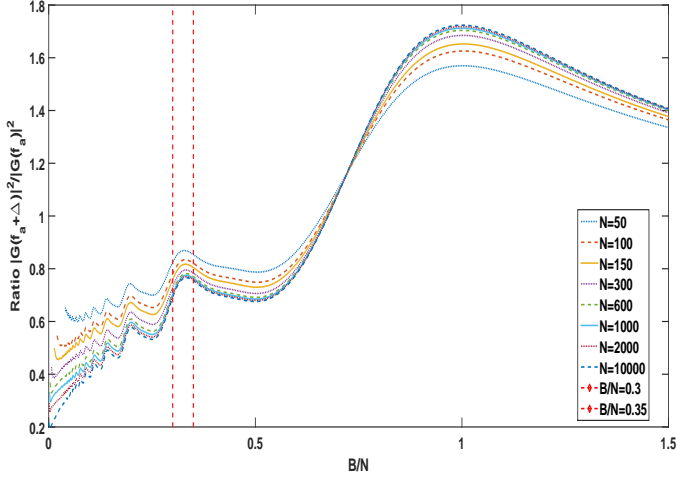
79

Figure 8 The ratio between the maximum and minimum gain of a band as a function of B/N. As shown in the figure, the ratio R_B will have a local peak around $B = N/3$, with $R_B \approx 0.8$. It is worth noting that, for $B > N$, the ratio will smoothly decrease to one, without any maxima after $B = N$.

for $f \in [0, 1)$. Thus,

$$
\begin{aligned}
|G(f)|^2 =& \Delta^2 + 2 \sum_{\ell=1}^{N-2} \sum_{k=\ell+1}^{N-1} \frac{\psi_k \psi_{k-\ell} \chi_\ell}{\pi^2 k(k-\ell)} + \\
& 2\Delta \sum_{k=1}^{N-1} \frac{\psi_k \chi_k}{\pi k} + \sum_{k=1}^{N-1} \frac{\psi_k^2}{\pi^2 k^2}
\end{aligned}
\tag{26}
$$

where $\chi_\ell = \cos\left(2\pi\ell(f_a - f + \frac{\Delta}{2})\right)$. The ratio between the minimum and maximal gain of a band may then be expressed as

$$
R_B = \frac{|G(f_a + \frac{\Delta}{2})|^2}{|G(f_a)|^2}
\tag{27}
$$

Figure 8 illustrates resulting ratios for varying number of samples, indicating that it is advantageous to select B to be roughly $N/3$, as this yields a good trade-off between the computational gain and the likelihood of accurately capturing the correct model order in few refinement steps, as shown in Figure 9, which illustrates the impact of R_B on the resulting estimates. Here, we have considered a signal consisting of $N = 50$ samples containing $K = 2$ (complex-valued) sinusoids corrupted by a zero-mean white Gaussian noise with SNR= 20dB. The figure shows the percentage of correctly estimated model orders for different ratios of B/N and the corresponding R_B ratio (here, to simplify the presentation, the shown second stage zooming used a constant $B_2 = 5$ elements), computed using 1000 Monte-Carlo simulations. In the figure, the resulting values of R_B
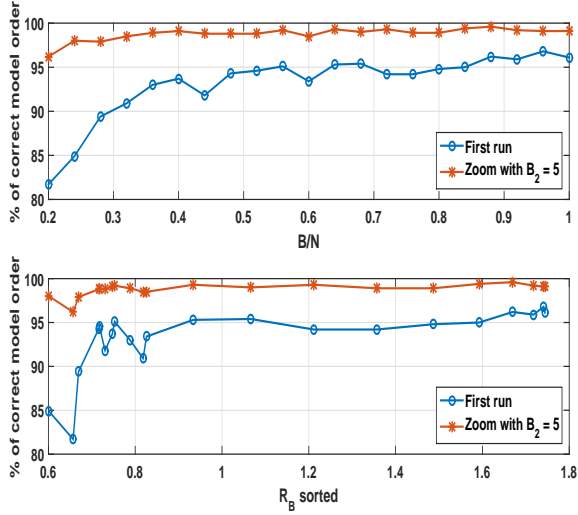
Figure 9 Percentage of finding correct model order as a function of ratio of bands to number of samples (top) and R_B (bottom).

and corresponding estimates have been sorted in ascending order (based on R_B value) as ratio R_B varies heavily in the range $B/N < 0.3$. Figure 9 confirms our suggestion of choosing the number of bands based on $B \approx N/3$ criteria as further increase of this ratio results only in marginal improvements in estimation performance.

5.   Complexity analysis

To illustrate the computational benefits of using the wideband dictionary as compared to forming the full dictionary, we proceed with our example of determining K M-D sinusoids by solving (7) using the popular ADMM algorithm [32]. In order to do so, we first transform the problem into a vector form reminiscent to (6), and split the variable β into two variables, here denoted $\mathbf{x}$ and z, after which the optimization problem may be reformulated as

$$\min_{\mathbf{x},z} \frac{1}{2}||\mathbf{y} - \mathbf{Ax}||_2^2 + \lambda||z||_1 \quad \text{subj. to} \quad \mathbf{x} = z \tag{28}$$

having the (scaled) augmented Lagrangian

$$\frac{1}{2}||\mathbf{y} - \mathbf{Ax}||_2^2 + \lambda||z||_1 + \frac{\rho}{2}||\mathbf{x} - z + u||_2^2 \tag{29}$$

where u is the scaled dual variable and ρ is the step length (see [32] for a detailed discussion on the ADMM). The minimization is thus formed by iteratively solving (29) for $\mathbf{x}$ and z, as well as updating the scaled dual variable u. This is done by

finding the (sub-)gradient for $\mathbf{x}$ and $\boldsymbol{z}$ of the augmented Lagrangian, and setting it to zero, fixing the other variables to their latest values. The steps for the jth iteration are thus

$$\mathbf{x}^{(j+1)} = \left(\mathbf{A}^H\mathbf{A} + \rho\mathbf{I}\right)^{-1} \left(\mathbf{A}^H\mathbf{y} + \boldsymbol{z}^{(j)} - \boldsymbol{u}^{(j)}\right) \tag{30}$$

$$\boldsymbol{z}^{(j+1)} = S(\mathbf{x}^{(j+1)} + \boldsymbol{u}^{(j)}, \lambda/\rho) \tag{31}$$

$$\boldsymbol{u}^{(j+1)} = \boldsymbol{u}^{(j)} + \mathbf{x}^{(j+1)} - \boldsymbol{z}^{(j+1)} \tag{32}$$

where $(\cdot)^H$ denotes the Hermitian transpose, $(\cdot)^{(j)}$ the jth iteration, and $S(\boldsymbol{v}, \kappa)$ is the soft threshold operator, defined as

$$S(\boldsymbol{v}, \kappa) = \frac{\max\left(|\boldsymbol{v}| - \kappa, 0\right)}{\max\left(|\boldsymbol{v}| - \kappa, 0\right) + \kappa} \odot \boldsymbol{v} \tag{33}$$

where $\odot$ denotes the element-wise multiplication for any vector $\boldsymbol{v}$ and scalar κ.

The computationally most demanding part of the resulting ADMM implementation is to form the inverse in (30) and to calculate $\mathbf{A}^H\mathbf{y}$. These steps are often done by QR factorizing the inverse in (30) prior to the iteration, so that this part is only calculated once. After this, the QR factors are used when forming the inner product. To give a simple example on the difference between the two types of dictionaries, we exclude any further computational speed-ups and show the difference on brute force computations of the above ADMM. This is done to give an idea on the effect $P < N$ has on the computational complexity. The total computational cost for the step in (30) depends on the size of the matrix $\mathbf{A}$. Let $N = \prod_{m=1}^{M} N_m$ and $P = \prod_{m=1}^{M} P_m$, then $\mathbf{A}$ is a $N \times P$ matrix. If $P < N$, computing the inverse will cost approximately P^3 operations, plus an additional P^2N operations to form the Gram matrix $\mathbf{A}^H\mathbf{A}$. Furthermore, to compute $\mathbf{A}^H\mathbf{y}$ requires PN operations, and the final step to compute $\mathbf{x}$ costs P^2 operations. If instead $P > N$, one may make use of the Woodbury matrix identity [33], allowing the inverse to be formed using $N^3 + 3PN^2$ operations, whereafter one has to compute $\mathbf{A}^H\mathbf{y}$ and the final matrix-vector multiplication, together costing $PN + P^2$ operations. In total, the x-step will have the cost of roughly $P^3 + (N+1)P^2 + NP$, if $P < N$, or $N^3 + 3PN^2 + PN + P^2$, if $N < P$.

Since using the banded dictionary allows for a smaller dictionary, one may calculate the computational benefit of using the integrated dictionary as compared to just using an ordinary dictionary with large P. Consider using only a single-stage narrowband dictionary, $\mathbf{D}_1$, with $P > N$ dictionary elements. This requires $C_1 = N^3 + 3PN^2 + P^2 + PN$ operations if using the above ADMM solution, with the dictionary $\mathbf{D}_1$ in the place of $\mathbf{A}$ in (30)-(32). If, on the other hand, one uses a multiple-stage wideband dictionary with N dictionary elements in the initial coarse dictionary, $\mathbf{B}_1$ (which is more than required, but simplifies the calculations), the cost of forming the first stage (coarse) minimization is

$C_2 = 2(N^3 + N^2)$. By taking the difference, i.e., forming

$$R = C_1 - C_2 = N^3 + 3PN^2 + P^2 +$$
$$+ PN - 2(N^3 + N^2)$$

one obtains the available computational resources, R, that are left for the dictionaries of the zoomed-in stages, without increasing the overall computational cost above that of the narrowband dictionary solution. Let $\mathbf{B}_z$ denote the zoomed-in dictionary with ηN number of bands, where $0 < \eta < 1$ denotes the ratio between the number of available bands in the dictionary and the number of samples. Then, one may deduce the grid size for each $\mathbf{B}_z$ that is allowed without increasing the overall computational complexity as compared to using the narrowband dictionary by solving

$$R = K I_z \left((\eta N)^3 + (N + 1)(\eta N)^2 + \eta N^2 \right)$$

where I_z denotes the number of zooming steps and K the number of sinusoids in the signal. To illustrate the resulting difference, consider the following settings: $P = 1000$, $N = 100$, $K = 5$, and $\eta = 2/3$. To only use half the resources that are needed to solve the full narrowband problem, one may, using the wideband dictionary, use 4 stages of zooming, resulting in a grid spacing of roughly 10^{-9}, as compared to 10^{-3} for the narrowband dictionary. One may of course also use a zooming procedure when using the narrowband dictionaries, although this would increase the risk of missing any off-grid component. This means that the smallest number of dictionary elements, for the narrowband dictionary to avoid missing any off-grid components, is $P = N$, and thus the wideband dictionary would need only at most η^2 of the computational resources needed for the ordinary dictionary, at each zooming stage.

6. Numerical examples

In this section, we proceed to examine the performance of the proposed method, initially illustrating that the use of a two-stage wideband estimator will have the same estimation quality as when using the ordinary (one-stage) narrowband LASSO estimator.

6.1. One-dimensional data

We initially considered a signal consisting of $N = 75$ samples containing $K = 3$ (complex-valued) sinusoids corrupted by a zero-mean white Gaussian noise with SNR$= 10$dB. In each simulation, the sinusoidal frequencies are drawn from a uniform distribution, over $[0, 1)$, with all amplitudes having unit magnitude and phases drawn from a uniform distribution over $[0, 2\pi)$. The performance is then
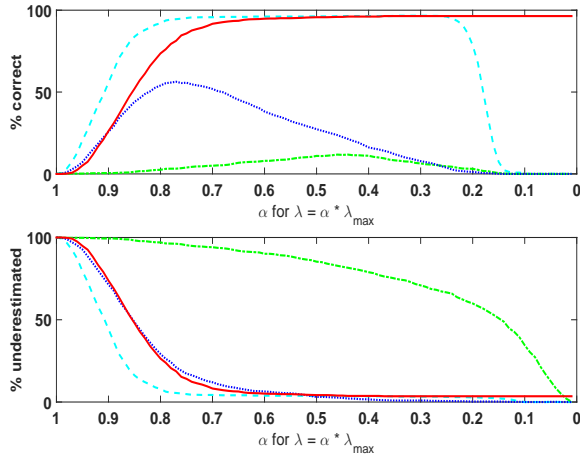
Figure 10 The probability of (top) correctly estimating and (bottom) underestimating the number of spectral lines, for the (single-stage) narrowband dictionary, using $P = 1000$ elements (cyan, dashed) and $P = 75$ elements (green, dot-dashed), and for the initial wideband dictionary, using $B_1 = 75$ elements (blue, dotted), and the (two-stage) wideband dictionary, using $B_1 = 75$ elements, together with $B_2 = 25$ elements per activated bands in the refining dictionary (red, solid).

computed using three different dictionaries, namely the (ordinary) narrowband dictionary, $\mathbf{D}$, with $P = 1000$ and $P = 75$ elements, respectively, and the proposed wideband dictionary, $\mathbf{B}$, using $B_1 = 75$ elements, followed by a second-stage narrowband dictionary using $B_2 = 25$ elements per active band. For each dictionary, we evaluate the performance for varying values of the user parameter α using $\lambda = \alpha\lambda_{max}$, where $\lambda_{max} = \max_i |\mathbf{x}_i^H \mathbf{y}|$ is the smallest tuning parameter value for which all coefficients in the solution are zero [19]. Here, $\mathbf{x}_i$ denotes either the ith column of the $\mathbf{D}$ dictionary or the ith column of the $\mathbf{B}$ dictionary. Each estimated result is then compared to the ground truth, counting the number of correctly estimated and underestimated model orders. The results are shown in Figure 10. As can be seen from the figure, the best results are achieved when $\alpha \leq 0.65$, in which case the proposed wideband dictionary, using $B_1 = 75$ bands, followed by a second stage narrowband dictionary, with $B_2 = 25$ for each activated band, have similar performance to the narrowband dictionary using $P = 1000$ dictionary elements.

Proceeding, we assess the mean-square error (MSE), defined as

$$\text{MSE} = \frac{1}{K} \sum_{k=1}^{K} (f_k - \hat{f}_k)^2 \tag{34}$$

where f_k and $\hat{f}_k$ denote the true and the estimated frequency, respectively, for the two-stage dictionary, showing the MSE as a function of SNR for the first-stage
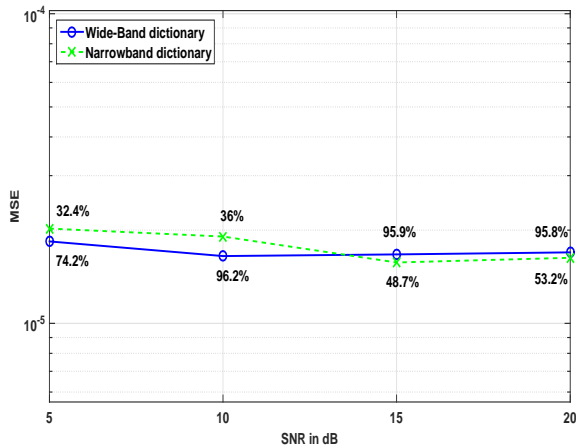
Figure 11 Mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $P = 100$, as compared to the two-stage dictionary, using $B_1 = 20$ integrated wideband elements in the first stage, followed by $B_2 = 5$ wideband elements in the second stage. The percentage of correct model order estimation (excluding outliers) is shown as a percentage on top of the corresponding MSE value.

wideband dictionary, $\mathbf{B}_1$, and second-stage wideband refining dictionary, $\mathbf{B}_2$. Here, and in the following, we consider situations where the number of elements in the dictionary is less than number of samples. As was described before, this is a situation where the performance of narrowband dictionaries can deteriorate seriously. For this experiment, we considered a signal with $N = 300$ samples containing $K = 2$ (complex-valued) sinusoids, being corrupted by different levels of zero-mean white Gaussian noise with SNR in the range $[5, 20]$ dB. Figure 2 shows the resulting MSE for the LASSO estimator for the estimates with correctly estimated model order; for runs with the correct model order estimation we also removed outliers from the final MSE calculation. We consider an estimate as an outlier if $|f - \hat{f}| > \Delta f$, where Δf was defined as two times the possible resolution, where possible resolution is defined as $1/P$ for the narrowband dictionary and $1/(B_1 \cdot B_2)$ for the wideband dictionary. Figure 12 shows the MSE for the same experiment done using the SPICE estimator. The number of outliers removed for the LASSO estimator was: 4, 0, 0, 0 for the wideband dictionary and 7, 16, 10 and 11 for the narrowband dictionary (corresponding to SNRs of 5, 10, 15, and 20 dB). The number of outliers removed for the SPICE estimator was; 17, 1, 1, 0 for the wideband dictionary and 52, 80, 117, and 103 for the narrowband dictionary. As can be seen from the figures, the two-stage dictionary using a wideband dictionary using $B_1 = 20$ bands, followed by a refining dictionary using $B_2 = 5$ wideband elements, achieves the same performance as the single-stage narrowband dictionary using $P = 100$ elements in terms of resolution. However, the narrow-band dictionary will for this case
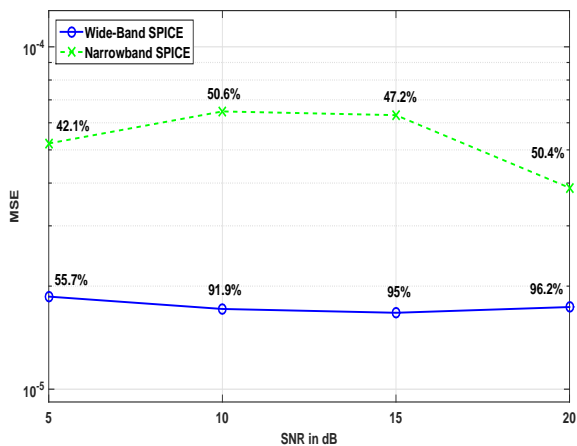
Figure 12 Mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $P = 100$, as compared to the two-stage dictionary, using $B_1 = 20$ integrated wideband elements in the first stage, followed by $B_2 = 5$ wideband elements in the second stage. The percentage of correct model order estimation (excluding outliers) is shown as a percentage on top of the corresponding MSE value.

fail to reliably restore the signal with reconstruction success rates of merely $30 - 50\%$.

Table 1 shows the relative complexity between using a full narrowband dictionary (using $P = 1000$, $N = 200$, and $K = 2$) and some different settings for the wideband dictionaries used in the numerical section. To simplify the comparison, the given complexity is the one of solving the ADMM without utilising any structures of the dictionary matrices. From the table, it is clear that it is more efficient to use the zooming procedure utilising the wideband dictionary as compared to solving the same problem using a full narrowband dictionary.

Next, we consider non-uniformly sampled data with $N = 400$ samples, for $K = 2$ sinusoids. For this experiment, we also added a third estimation step

Settings	Relative complexity
$P = 1000, N = 200, K = 2$	1
$B_1 = 20, B_2 = 5$	0.001
$B_1 = 20, B_2 = 40$	0.015
$B_1 = 10, B_2 = 10, B_3 = 5$	0.001

Table 1 Relative complexity between using the narrow- and wideband dictionaries. Here, P indicates the number of columns in the narrowband dictionary, whereas B_1 and B_2 indicate the number of wideband elements in the first and second stage of the zooming procedure, respectively. In the last row, a third stage has been added using B_3 wideband elements.
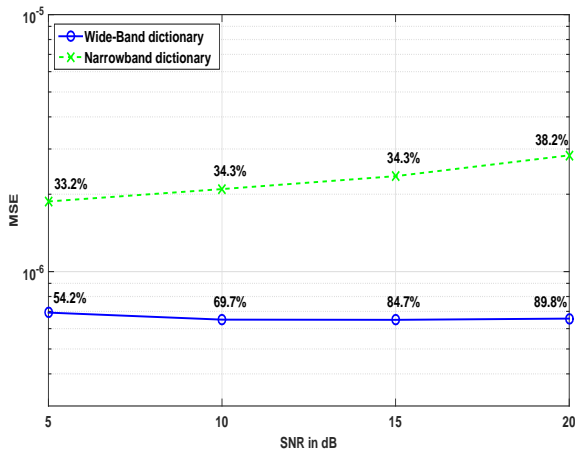
Figure 13 Signal estimation for non-uniform sampling: mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $P = 200$ elements, as compared to the three-stage dictionary, using $B_1 = 10$ integrated wideband elements in the first stage, followed by $B_2 = 10$ and $B_3 = 5$ wideband dictionaries in the second stage and third stage per active band detected in the previous stage. The correct model order estimations are shown in percentage above each point.

for the iterative wideband dictionary. After initial estimation with $B_1 = 10$ wideband dictionary elements, we zoom into the active bands with $B_2 = 10$ dictionary elements per active band, and then once again with $B_3 = 5$ dictionary elements. In spite of the three stage zooming, the method requires considerably less computational operations as compared to using a corresponding narrowband dictionary, but results in better performance both in terms of resolution and model-order accuracy. The resulting MSEs are shown in Figure 13. All results are computed using 1000 Monte-Carlo simulations.

6.2. Two-dimensional data

In this subsection, we present results on a 2-D data set. In this example, each dimension is sampled uniformly with $N = 100$ samples. We compare a narrowband dictionary with $P = 49$ elements per dimension with the wideband dictionary using $B_1 = 7$ bands per dimension in the first step and a wideband dictionary with $B_2 = 7$ elements per active band in a second (zooming) step. Here, we use two separate wideband dictionaries, the first, $\mathbf{B}$, using integrated dictionary elements as defined in (8), and the second, $\mathbf{B}_{DPSS}$, which contains elements based on DPSS. For the DPSS-based dictionary, we used a sequence length of $Q = 100$ and $W = 1/2.1$. Using $W < 1/2.1$ results in dictionary elements which concentrate energy in a more narrow band and are therefore not suitable for the dictionary with $B_1 = B_2 = 7$ elements. We considered a signal
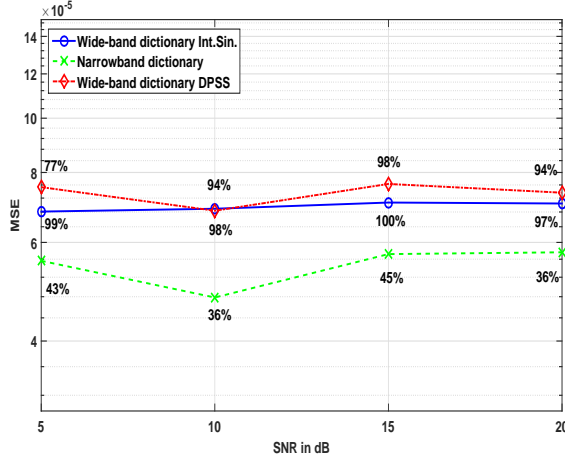
Figure 14 Signal estimation in two dimensions: mean-square error curves for different SNR levels for the single-stage narrowband dictionary $\mathbf{D}$*, using* $P = 49$ *per dimension, as compared to the two-stage dictionaries (DPSS based and integrated sinusoids based), using* $B_1 = 7$ *wideband elements in the first stage, followed by* $B_2 = 7$ *wideband elements in the second stage (per active band).*

containing $K = 2$ (complex-valued) sinusoids per dimension, with the signal being corrupted by a zero-mean white Gaussian noise. In each simulation, the sinusoidal frequencies are drawn from a uniform distribution, over $[0, 1)$, with all the amplitudes having unit magnitude. The two dictionaries are compared against each other based on the MSE performance in the same manner as in the previous subsection, with the MSE being calculated as the average value for both dimensions if the model order estimate for the iteration was correct. Outliers are removed before the MSE calculation. The number of outliers removed was: 6, 15, 25, 30 for the $\mathbf{B}_{DPSS}$ dictionary and 20, 22, 17 and 24 for the narrowband dictionary (corresponding to SNRs of 5, 10, 15, and 20 dB). The wideband dictionary $\mathbf{B}$ did not result in any outliers. The percentages of correct model order estimates are shown for each SNR value. Figure 14 shows the resulting MSE curves. It can be seen that the wideband dictionary with integrated sinusoids outperforms the DPSS-based wideband dictionary both in terms of MSE and model-order accuracy. Comparing to using the narrowband dictionary, it can be seen that both wideband dictionaries outperform it considerably in terms of model-order estimation, although the narrowband dictionary shows slightly better performance in terms of MSE. Also in this example, the wideband dictionaries provide a considerable reduction in computational complexity as well as a robustness in terms of estimating off-grid components. All results are computed using 100 Monte-Carlo simulations.

Using the same setup as described above we also evaluated the performance of the proposed approach when the number of sinusoids to detect is higher. Again, we
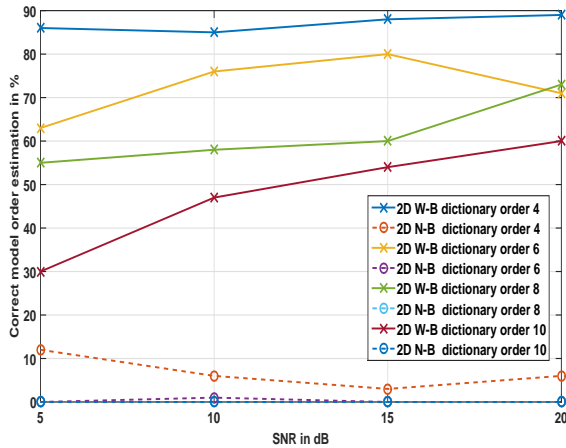
Figure 15 Percentage of correct model order esimations for different number of sinusoids and for different SNR levels for wideband dictionary (W-B) and narrowband dictionary (N-B).

considered the ordinary narrowband dictionary, $\mathbf{D}$, and the wideband dictionary, $\mathbf{B}$, from the previous experiment. We calculated the percentage of correct model order estimation for signals with $K = 4, 6, 8$, and 10 (complex-valued) sinusoids. The results were computed using 100 Monte-Carlo simulations; the correct model order estimation percentages for different SNR levels are shown in Figure 15. The best regularization parameters λ for solving the LASSO for each case were found beforehand with the grid-search method. For this, we selected the range of parameter $\alpha \in [0.7, 0.05]$ with the step-size 0.05 and ran 100 Monte-Carlo simulations for each model order and then picked the best parameter for the selected model order based on model order accuracy. For the two-step wideband dictionary, a grid-search was done for the set of α parameter for the both stages. It can be clearly seen that for situations where the number of elements in the dictionary is lower than the number of samples, the narrow-band dictionary fails to produce any meaningful results.

6.3. Measured data example

Finally, we examine the performance of the proposed wideband framework on measured nuclear magnetic resonance (NMR) data, again comparing with using the full narrowband dictionary. The measured data NMR measurement consist of $N = 256$ samples, and contains five damped sinusoidal signals. To make the comparison fair, we neglect the damping in the modelling (as the wideband dictionary will implicitly allow for the resulting wide peaks, whereas the narrowband dictionary will require an additional parameter to do so). This results in estimates containing clusters of peaks instead of individual component.
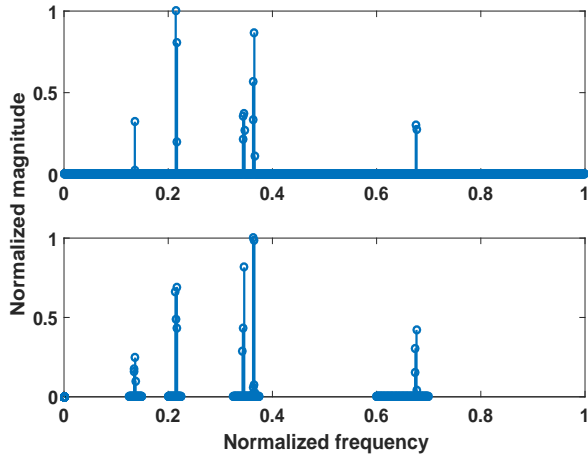
Figure 16 The resulting estimates using a dictionary with 2000 narrowband elements (top), and a two-stage zooming approach using wideband elements, using $B_1 = 40$ in the first stage and $B_2 = 50$ for each activated bands in a second stage. The signal is a measured NMR signal of length $N = 256$.

Figure 16 illustrates the resulting estimates, showing the result of using a narrowband dictionary with 2000 elements (top), as well as a two-stage wideband dictionary (using $B_1 = 40$ elements in the first stage and $B_2 = 50$ elements in the second). The resulting estimates will thus have the same final grid resolution. As can be expected, both estimators show similar results, having the same support and roughly the same relative amplitudes. Using the introduced ADMM implementation described in Section 5, the wideband estimate was formed in 0.315 seconds, which was 20 times faster than the narrowband estimate. Here, one may note that if an iterative narrowband zooming would be used, it would require at least 256 elements in the first stage to avoid losing any peaks; doing so would require more complexity than the two-stage wideband estimator.

7.   Conclusion

In this paper, we have introduced a wideband dictionary framework, allowing for a computationally efficient reconstruction of sparse signals. Wideband dictionary elements are formed as spanning bands of the considered parameter space. In the first stage, one may typically use a coarse grid using the integrated wideband dictionary locating the bands of interest, whereafter non-active parts of the parameter space are discarded. In the next stage, a refining dictionary can be used to more precisely determine the parameters of interest on the active bands from the previous step, allowing for an iterative zooming procedure. The technique is illustrated for the problem of estimating multidimensional sinusoids corrupted

by Gaussian noise, showing that the same accuracy can be achieved, although at a computationally substantially lower cost and with much less risk of missing any off-grid components. The proposed framework is here illustrated for the LASSO and SPICE estimators, but other sparse reconstruction techniques may be extended similarly.

References

[1] M.Unser and P.Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2013.

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

[3] E. J. Candès and M.B. Wakin, "An introduction to compressive sampling", *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[4] E. J. Candès, 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.

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

[6] M. A. Herman and T. Strohmer, "Genral Deviants: An Analysis of Perturbations in Compressed Sensing", *IEEE J. Sel. Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[7] 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.

[8] 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.

[9] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed Sensing Off the Grid", *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[10] Y. Chi and Y. Chen, "Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization", *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[11] Z. Yang and L. Xie, "Enhancing Sparsity and Resolution via Reweighted Atomic Norm Minimization", *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb. 2016.

[12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems", *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[13] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra", in *17th World Congress IFAC*, Seoul, Jul. 2008, pp. 10 225–10 229.

[14] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data", *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

[15] 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, Jul. 2012.

[16] 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, Mar. 1997.

[17] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Multi-Pitch Estimation Exploiting Block Sparsity", *Elsevier Signal Processing*, vol. 109, pp. 236–247, Apr. 2015.

[18] L. E. Ghaoui, V. Viallon, and T. Rabbani, "Safe Feature Elimination for the LASSO and Sparse Supervised Learning Problems", *Pacific Journal of Optimization*, vol. 8, no. 4, pp. 667–698, Jan. 2012.

[19] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, Simon, J. Taylor, and R. J. Tibshirani, "Strong rules for discarding predictors in lasso-type problems", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 2, pp. 245–266, 2012.

[20] Z. J. Xiang, Y. Wang, and P. J. Ramadge, "Screening Tests for Lasso Problems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1008–1027, May 2016.

[21] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "A Dynamic Screening Principle for the Lasso", in *Proceedings of the 22nd European Signal Processing Conference*, Lisbon, Portugal, Sep. 2014.

[22] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the Duality Gap: Safe Rules for the Lasso", *Publication: eprint arXiv:1505.03410*, 2015.

[23] J. Liu, Z. Zhao, J. Wang, and J. Ye, "Safe screening with variational inequalities and its applicaiton to lasso", in *International Conference on Machine Learning*, Beijing, China, Jun. 2014.

[24] Z. Yang and L. Xie, "Frequency-Selective Vandermonde Decomposition of Toeplitz Matrices With Applications", *Publication: eprint arXiv:1605.02431*, 2016.

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

[26] 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.

[27] P. Stoica, D. Zachariah, and L. Li, "Weighted SPICE: A Unified Approach for Hyperparameter-Free Sparse Estimation", *Digit. Signal Process.*, vol. 33, pp. 1–12, Oct. 2014.

[28] E. H. D. S. Sahnoun and D. Brie, "Sparse Modal Estimation of 2-D NMR Signals", in *38th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[29] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying N-dimensional Signals", *Elsevier Signal Processing*, vol. 128, pp. 309–317, Nov. 2016.

[30] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — V: The Discrete Case", *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May 1978.

[31] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "Generalized sparse covariance-based estimation", *Signal Processing*, vol. 143, pp. 311–319, 2018.

[32] 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.

[33] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th. The John Hopkins University Press, 2013.

C

C. Sparse Reconstruction Method for Separating Cardiac and Respiratory Components from Electrical Bioimpedance Measurements

Maksim Butsenko, Olev Märtens, Andrei Krivošei and Yannick Le Moullec
Thomas Johann Seebeck Dept. of Electronics, Tallinn University of Technology,

Estonia

Abstract

In this work, we investigate the possibility of employing sparse reconstruction framework for the separation of cardiac and respiratory signal components from the bioimpedance measurements. The signal decomposition is complicated by the nonstationarity of the signal and overlapping of their spectra. The signal has a harmonic structure which is sparse in the spectral domain. We approach the problem by considering the dictionary with integrated wideband elements describing spectral components of the considered signal. The parameter estimation task is solved through the means of sparse reconstruction where solving the optimization problem returns a sparse vector of relevant dictionary atoms.

Keywords: Electrical bioimpedance, parameter estimation, sparse reconstruction, cardiac and respiratory signals, wideband dictionary

1.   Introduction

Electrical bioimpedance (EBI) measurements based applications for medical signal monitoring can provide interesting alternative to the conventional approaches due to non-invasiveness and cost-effectiveness. EBI measurements can characterize different properties of the human tissues and structures as well as various physiological dynamic processes in the human body, e.g. respiration and cardiac activities [1]. Medical applications can utilize EBI signal to analyze cardiac activities from simple heart-rate monitors [2] up to more sophisticated estimation of the cardiac output [3], [4] or central aortic blood pressure waveform [5]. EBI-based, noninvasive, continuous cardiac output monitoring can have several clinical applications in cardiology, emergency care, anesthesiology [3].

Extraction and separation of the cardiac and/or respiratory signal(s) is challenging; as these two signal components can vary significantly in time and frequency domains, they can have overlapping spectra and can be accompanied by severe noise, disturbances and artefacts. Respiratory and motion-based signal components can be considered as noise signals if one is only interested in the cardiac component. However, it is more interesting to try to extract both the cardiac and respiratory signals. Different approaches to this problem have been proposed in the recent literature including adaptive filtering [6], adaptive phase-locked loop [7], method based on the signal shape-locked-loop decomposer solution [8], principal and independent component analysis [9] and artificial neural networks [10]. These proposed methods are quite sophisticated; however, they either require considerable amount of fine-tuning and computational resources (neural networks) or require some time for converging to an optimal solution and exhibit low robustness to changes in the parameters of the signal (adaptive filters). Considering the variety of the proposed solutions, it is interesting to note that there is still no clearly established method available for extraction and separation of cardiac and respiration signal components, partly due to the problems mentioned above, but also partly due to the variations in the anatomy and physiology of human beings and their behaviour at various times and in physical and mental situations.

In this work we are looking at employing a sparse signal processing technique for reconstructing respiratory and cardiac signals from EBI measurements (Fig. 1 shows an example of the EBI signal). The notion of sparsity usually implies that a signal is sparse in some domain (has only a few number of significant components). It has been noted that a large number of common applications results in signals that may be well approximated using a sparse reconstruction framework, and this area has seen increase in interest from the scientific community (see, e.g., [11], [12] and the references therein). A large part of these works has focused on convex algorithms that make use of different sparsity inducing penalties, which result in solutions that are well represented using only few elements from some (usually known) dictionary matrix, $\mathbf{D}$. If

the dictionary is appropriately chosen, it can be shown that even very limited measurements allow for an accurate signal reconstruction [13], [14]. We propose to reconstruct the relevant components of the EBI signal by considering wideband dictionary elements introduced in [15], [16]. These elements are formed as bands of frequencies covering the entire spectrum of interest. By refining the estimate through iteratively zooming into active parts of the spectra, the proposed algorithm is able to estimate signal parameters and reconstruct the signal without any prior knowledge.

2.   Sparse reconstruction

The signal model for the considered EBI measurements can be written as

$$y(t) = s_C(t) + s_R(t) + e(t) \tag{1}$$

where $s_C(t)$ is the cardiac component, $s_R(t)$ respiratory component and $e(t)$ additive noise. We start with the assumption that both the cardiac and the respiratory components in the EBI signal are well approximated by a sum of sinusoids. Considering discrete-time signal consisting of N samples, we can then reformulate (1) as

$$y_n = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \beta_{k,l} e^{2i\pi f_k l t_n} + \epsilon_n \tag{2}$$

for $n = 1, \ldots, N$, where K denotes the number of sources and L_k the number of sinusoids for the kth source. Furthermore, f_k denotes the kth fundamental frequency and $\beta_{k,l}$ the complex amplitude corresponding to the ℓth harmonic of the kth fundamental frequency, respectively, t_n the nth sample time, and ϵ_n the additive noise at time t_n. In the problem at hand K is known as we deal only with two sources. Hence, for reconstructing each source signal we can tackle the problem of estimating the frequencies f_k, for $k = 1, \ldots, L_k$, of a measured signal y_n. The classical sparse formulation of this estimation problem, as presented in [17], considers the LASSO minimization [18]

$$\min_{\mathbf{x}} \frac{1}{2} ||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 + \lambda ||\mathbf{x}||_1 \tag{3}$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_N \end{bmatrix}^T \tag{4}$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_1 & \cdots & \mathbf{d}_P \end{bmatrix} \tag{5}$$

$$\mathbf{d}_p = \begin{bmatrix} e^{2i\pi \hat{f}_p t_1} & \cdots & e^{2i\pi \hat{f}_p t_N} \end{bmatrix}^T \tag{6}$$

where $\hat{f}_p$ for $p = 1, \ldots, P$ denotes the $P \gg L_k$ candidate frequencies in the dictionary, $\mathbf{D}$, typically selected to be closely spaced to allow for minimal grid mismatch, and $(\cdot)^T$ the transpose. The penalty on the 1-norm of $\mathbf{x}$ will ensure that the found solution, $\hat{\mathbf{x}}$, will be sparse, with λ denoting a user parameter governing the desired sparsity level of the solution. The desired frequencies, as well as their order, are then found as the non-zero elements in $\hat{\mathbf{x}}$. In this work we are limiting our approach to the use of LASSO minimization; however, in the literature there exist several well-known methods of solving undetermined system of linear equations (e.g. basis pursuit [19], orthogonal matching pursuit [20], etc.).

Deviating from the classical estimation method where the dictionary consists of closely spaced sinusoids, in this work we propose exploiting integrated wideband dictionary elements [1] in $\mathbf{D}$ [15], [16], such that each element is formed over the band

$$\boldsymbol{d}_b(t_n) = \int_{f_b}^{f_{b+1}} e^{2i\pi f t_n} df = \frac{e^{2i\pi f_{b+1} t_n} - e^{2i\pi f_b t_n}}{2i\pi t_n} \tag{7}$$

As the resulting dictionary elements cover the band of frequencies from f_b to f_{b+1}, for $b = 1, \ldots, B_1$, the entire spectrum may be formed with $B_1 \ll P$, where B_1 denotes the (initial) number of wideband dictionary elements. This approach coupled with iterative zooming into the active part of the spectra results in decreased computational complexity and improved estimation performance. For the in-depth discussion on the use and benefits of wideband dictionary elements based dictionaries we refer the reader to the comprehensive analysis in [16].

The resulting solution to the minimization in (3), $\hat{\mathbf{x}}$, can be further used to reconstruct an approximation of the initial EBI waveform as

$$\hat{y} = \mathbf{D}_C \hat{\mathbf{x}}_C + \mathbf{D}_R \hat{\mathbf{x}}_R \tag{8}$$

where $\mathbf{D}_C$ and $\mathbf{D}_R$ denote dictionaries, while $\hat{\mathbf{x}}_C$ and $\hat{\mathbf{x}}_R$ denote solutions for the LASSO minimization for cardiac and respiratory components respectively.

3.   The proposed algorithm

We approach our analysis of the bioimpedance data by selecting two time windows t_c and t_r, corresponding to the cardiac and the respiratory component respectively. Ideally we would like the time window to be close to the period of the analyzed signal. A reasonable assumption for the rest rate of an adult is around 50-90 heartbeats per minute (bpm) for the cardiac component, which corresponds approximately to the heart rate with the frequency of $0.83...1.5$

---

[1]It should be noted that the proposed framework does not require the use of a wideband dictionary; it is here used for illustration as it has been found to often yield preferable performance [15], [16].
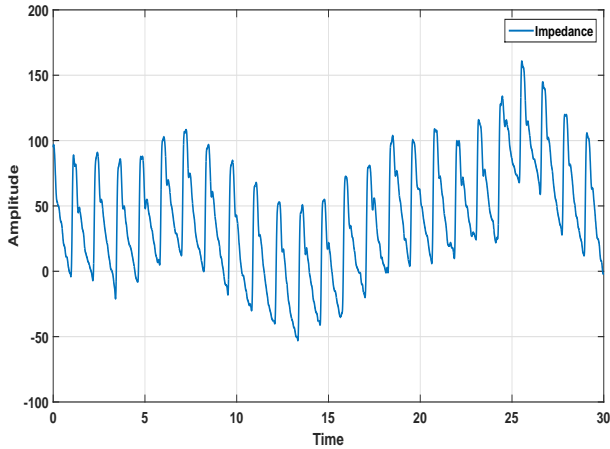
Figure 1 Example of a waveform of an electrical bioimpedance measurements.

Hz; and around 10-18 breaths per minute for the respiratory component, which corresponds approximately to the respiratory rate with the frequency of $0.16...0.30$ Hz. Correspondingly, we assign the time windows to be $t_c = 1.5$ seconds and $t_r = 6.25$ seconds. The respiratory component is reconstructed from the EBI data that is filtered through a 512-tap low-pass filter with cut-off frequency $f_{LP} = 0.4$ Hz and the cardiac component is reconstructed from the EBI data that is filtered through a 512-tap band-pass filter with $f_{BP_1} = 0.7$ Hz and $f_{BP_2} = 3$ Hz. For each component we continue in a similar manner by analyzing data for each time window. We solve the LASSO minimization in (3) using the initial dictionary $\mathbf{D}_1$ with B_1 number of elements constructed according to (7). Next, we discard the inactive bands of the spectra and solve the minimization task again now using the dictionary $\mathbf{D}_2$ with B_2 on parts of the spectra which were active in the first estimation step. By discarding inactive bands, we reduce the amount of computations required. In our experience, a two-step iterative zooming procedure (coupled with the proper choice of number of elements in the both dictionaries) is enough to achieve reasonable estimation precision; however, this choice is rather arbitrary. After the final estimation step, the algorithm returns a sparse vector which corresponds to the dictionary elements which fit the data in the best way. By extracting the fundamental frequency (first significant peak) from the vector, we get the corresponding rate of cardiac or respiratory part of the signal respectively. Then we move the window further on with an overlap of 50% with the previous data block. LASSO is solved through the alternating direction method of multipliers (ADMM) [21], derivation of the solution is presented in [15], [16], [22] and we refer the interested reader to description therein.

It has to be noted that using general LASSO formulation has a common weakness known as f_0 vs. $f_0/2$ ambiguity [22] (here f_0 denotes the fundamental
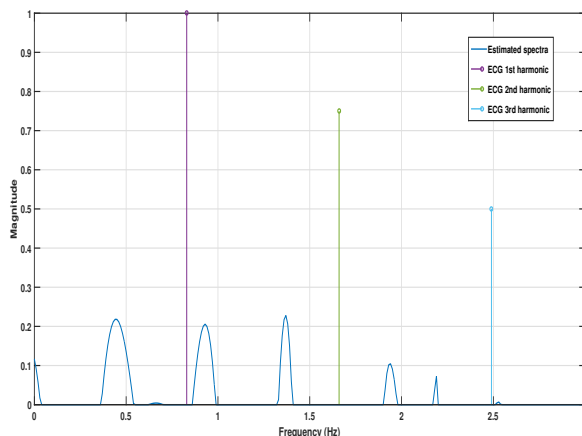
Figure 2 Half fundamental frequency problem (magnitudes of ECG harmonics were chosen arbitrarily to show the location of the harmonics and do not reflect the actual values of the signal).

frequency). In some cases, the optimal solution to the minimization problem is going to be described similarly well (or even in a better way) by dictionary elements with fundamental frequency at half the actual frequency. As every other harmonic component in this solution will coincide with the actual signal component, then the algorithm returns $f_0/2$ as fundamental frequency of the analyzed signal. The problem is illustrated in the Fig. 2. The actual fundamental frequency of the cardiac signal as represented by ECG spectra is located at 0.83 Hz. From the estimated signal spectra the first significant peak can be found at 0.44 Hz and second at 0.93 Hz. It is clear that in this particular case the algorithm returned the fit with harmonic structure with fundamental frequency at $f_0/2$ the actual one.

The currently proposed solution works reliably for the measurements where the cardiac component of the EBI signal lies in the frequency range of 0.8 to 1.4 Hz, corresponding to 48 to 84 heartbeats per minute. In this case $f_0/2$ frequency lies well below the normal heart-rate of the human beings and can therefore be detected as wrong estimate and corrected accordingly. However, with the heart-rate values above this range, $f_0/2$ frequency will yield an estimate which falls in the range which can be considered to lie in the normal heartbeat range and therefore requires additional processing in terms of signal-tracking, which will exclude abrupt changes in the continuous estimation process. Proper remedy to this sort of situation can be found in the form of solutions which promote block-sparsity [22], [23], where the dictionary has an internal block structure and the final solution has to be block-sparse (i.e. use only few blocks to describe the solution).
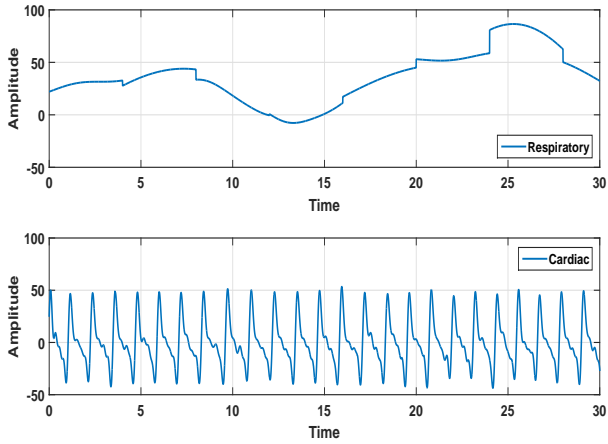
Figure 3 Separated respiratory (top) and cardiac (bottom) signal components

4.   Reconstructing the signal

In our experiments for testing the proposed method, we used the collection of EBI signals recorded in clinical conditions. The EBI measurements were taken by using a CircMon device (JR Medical Ltd, Estonia) operating at a frequency of 30 kHz and having a tetrapolar configuration of electrodes (together eight standard ECG electrodes on hands and feet). Initially the EBI signal is sampled at a rate of 1 kHz and is further decimated to 200 Hz using a 5-point averaging filter. All the simulations and analysis in this work were conducted in the MATLAB environment.

For the reconstruction, we have chosen to use a two-step estimation with $B_1 = 100$ dictionary elements in the first estimation run and $B_2 = 50$ dictionary elements in the second refining run. This choice results in the final resolution of 0.01 Hz. The component reconstruction is carried out by multiplying the solution $\hat{x}$ found by LASSO with the dictionary $\mathbf{D}$ for each time window t_c or t_r. We can interpret the reconstructed block as a weighted sum of relevant dictionary elements. Fig. 3 shows the separated respiratory and cardiac parts of the signal. As the reconstruction is done piecewise block by block, the breaks can be noted on reconstructed signal (especially on the respiratory component in Fig. 3). This can be mitigated by employing "smoothing" moving average filter for example. Fig. 4 shows the original EBI data and reconstructed EBI data which consists of reconstructed respiratory and cardiac components. In Fig. 5 we show the resulting fundamental frequency estimate of the cardiac signal component. We compare the result of the proposed method with the ground-truth, which is the frequency that is extracted from the ECG measurement done at the same time as EBI measurements. As we can see from the figure, the resulting estimate follows
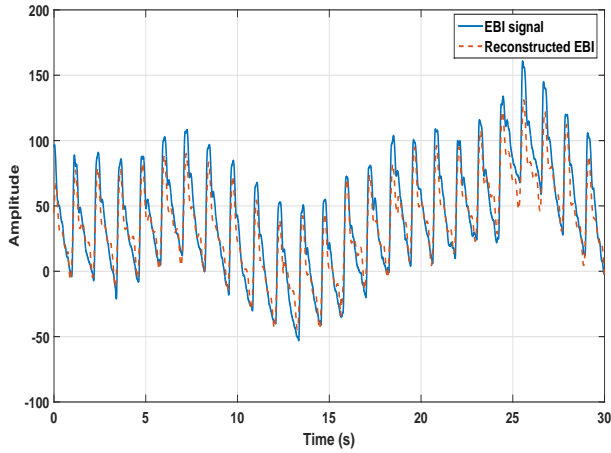
Figure 4 Impedance signal as reconstructed from the sum of the separated cardiac and respiratory parts, compared to the original impedance signal

the actual frequency of the cardiac component closely and mean-square error (MSE) as compared to ECG measurements is 0.0007.

Direct comparison with other algorithms is complicated by two factors. First, different authors use different metrics to evaluate efficacy of their approach. For example, classification metrics in [9], stroke volume estimation accuracy in [6] and in [10], visual comparison of original and reconstructed waveforms in [6] and [8]. Second and more important is the fact that different authors use different datasets (usually collected by their research group) and their measurements are not available for the wider research community, possibly due to privacy and ethical constraints related to the patient data. Authors tend to concentrate more on considering how their approach is able to capture the phenomena they are investigating. In our work the proposed algorithm is evaluated on 8 different EBI datasets measuring different levels of activity for different subjects. For each set we have evaluated the proposed algorithm based on two MSE-s. Reconstruction MSE denoted as MSE_R (9) is calculated by comparing original EBI waveform with the waveform that is reconstructed from two separated signal components according to (8).

$$\text{MSE}_R = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 \tag{9}$$

where y_n and $\hat{y}_n$ denote the nth sample of the EBI signal and reconstructed signal, respectively. Frequency estimation MSE denoted as MSE_F (10) is calculated by comparing frequency estimate of cardiac component with the ECG measurements

106
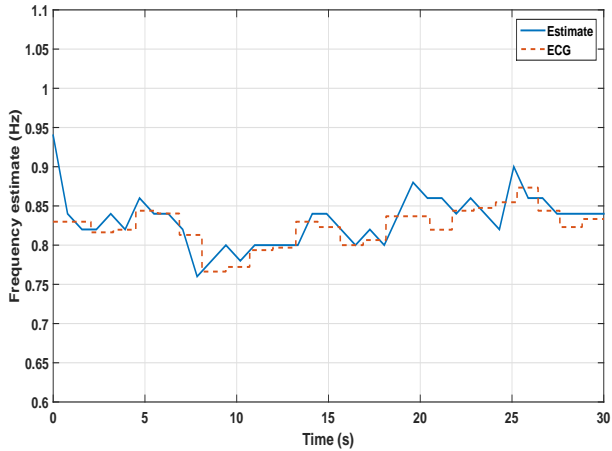
Figure 5 Cardiac component frequency estimate compared to the frequency estimate as measured from ECG signal

which were collected simultaneously with the corresponding EBI dataset.

$$\text{MSE}_F = \frac{1}{M} \sum_{m=1}^{P} (f_m - \hat{f}_m)^2 \qquad (10)$$

where f_m denotes ground truth frequency and $\hat{f}_m$ frequency estimated by the proposed method. M is the number of blocks for which the frequency estimate is calculated and depends on the length of the recorded signal, time window t_c and window overlapping factor. Here we considered ECG measurements as the ground truth for the comparison. Results are presented in Table 1.

5. Conclusion

In this work we have presented a novel method for separating cardiac and respiratory components from the electrical bioimpedance measurements. Signal components are recovered by the means of sparse reconstruction using the dictionary consisting of wideband elements. The proposed approach is able to reliably reconstruct both signal components and track fundamental frequency of cardiac component with high precision. Although the algorithm is not robust to the f_0 vs. $f_0/2$ ambiguity problem, relevant post-processing techniques provide a temporary remedy, while the planned extension to the proposed algorithm with the use of block-sparse dictionary structure should provide the necessary robustness in the future.

EBI dataset	Duration	MSE_R	MSE_F
02ecg	107 s	2.16e-06	0.00129
06ecg	37 s	5.09e-06	0.00120
07ecg	67 s	3.75e-06	0.00658
08ecg	23 s	2.97e-05	0.00335
12ecg	117 s	1.13e-06	0.00251
17ecg	24 s	3.88e-06	0.00752
101ecg	103 s	1.04e-05	0.00540
102ecg	82 s	3.90e-06	0.00060

Table 1 Mean square errors for reconstruction of EBI signals and frequency estimate of corresponding cardiac component

Acknowledgments

References

[1] S. Grimnes and O. Martinsen, *Bioimpedance and Bioelectricity Basics*. Academic Press, 2014.

[2] D. H. Díaz, Ó. Casas, and R. Pallas-Areny, "Heart rate detection from single-foot plantar bioimpedance measurements in a weighing scale", in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, IEEE, 2010, pp. 6489–6492.

[3] D. G. Jakovljevic, M. I. Trenell, and G. A. MacGowan, "Bioimpedance and bioreactance methods for monitoring cardiac output", *Best Practice & Research Clinical Anaesthesiology*, vol. 28, no. 4, pp. 381–394, 2014.

[4] B. Saugel, M. Cecconi, JY. Wagner, and DA. Reuter, "Noninvasive continuous cardiac output monitoring in perioperative and intensive care medicine", *British Journal of Anaesthesia*, vol. 114, no. 4, pp. 562–575, 2015.

[5] A. Krivoshei, H. Uuetoa, M. Min, P. Annus, T. Uuetoa, and J. Lamp, "Cap waveform estimation from the measured electrical bioimpedance values: Patient's heart rate variability analysis", in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 2788–2791.

[6] V. K. Pandey, P. C. Pandey, N. J. Burkule, and LR. Subramanyan, "Adaptive filtering for suppression of respiratory artifact in impedance cardiography", in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 7932–7936.

[7] A. Krivoshei, M. Min, T. Parve, and A. Ronk, "An adaptive filtering system for separation of cardiac and respiratory components of bioimpedance signal", in *Medical Measurement and Applications, 2006. MeMea 2006. IEEE International Workshop on*, 2006, pp. 10–15.

[8] A. Krivoshei, V. Kukk, and M. Min, "Decomposition method of an electrical bio-impedance signal into cardiac and respiratory component", *Physiological Measurement*, vol. 29, no. 6, pp. 15–25, 2008.

[9] I. Nejadgholi and M. Bolic, "A comparative study of pca, simca and cole model for classification of bioimpedance spectroscopy measurements", *Elsevier Computers in Biology and Medicine*, vol. 63, pp. 42–51, 2015.

[10] SMM. Naidu, U. R. Bagal, P. C. Pandey, S. Hardas, and N. D. Khambete, "Monitoring of stroke volume through impedance cardiography using an artificial neural network", in *Communications (NCC), 2015 Twenty First National Conference on*, IEEE, 2015, pp. 1–6.

[11] M.Unser and P.Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2013.

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

[13] E. J. Candès, 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.

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

[15] M. Butsenko, J. Swärd, and A. Jakobsson, "Estimating Sparse Signals Using Integrated Wide-band Dictionaries", in *42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, New Orleans, USA, Mar. 2017.

[16] ——, "Estimating Sparse Signals Using Integrated Wideband Dictionaries", *IEEE Trans. Signal Process.*, vol. 66, pp. 4170–4181, Aug. 2018.

[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. 10 225–10 229.

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

[19] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit", *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[20] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[21] 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.

[22] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Multi-pitch estimation exploiting block sparsity", *Elsevier Signal Processing*, vol. 109, pp. 236–247, 2015.

[23] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso", *Journal of Computational and Graphical Statistics*, vol. 22, pp. 231–245, 2011.

D

D. The Zoomed Iterative Adaptive Approach

Maksim Butsenko[*], Johan Swärd[†], and Andreas Jakobsson[†]

[*]*Thomas Johann Seebeck Dept. of Electronics, Tallinn University of Technology, Estonia*
[†]*Dept. of Mathematical Statistics, Lund University, Sweden*

Abstract

In this work, we investigate the possibility of incorporating a zooming procedure for the well known iterative adaptive approach (IAA), and thereby allow for higher precision on relevant areas of the spectrum. These kinds of zooming schemes have been used successfully together with several other methods, and have in many cases shown dramatical decrease in the computational cost. It has earlier been noted that the IAA method does not easily allow for these kind of zooming approaches, as the covariance formulation dictates that the resolution must be the same over all regions of the spectrum. In this paper, we present an iterative zooming procedure which allows for an efficient local estimation of IAA spectrum. The proposed algorithm is shown to allow for significant speed-ups without sacrificing estimation performance.

Keywords: iterative adaptive approach, IAA, wideband dictionary, iterative zooming

1. Introduction

High-resolution spectral estimation is a topic of notable relevance in a wide range of applications, and the area has as a result attracted significant interest during the recent decades. Early estimators were often formed as being either parametric or non-parametric, with the former approaches generally posing strong assumptions on knowledge of the signal of interest and the distribution of any corrupting noise processes, whereas the latter generally avoid exploiting any information of the signal of interest [1], [2]. Recent efforts have focused more on so-called semi-parametric estimators, allowing for partial knowledge of the signal of interest, but often avoid making assumptions on the model order or on the noise distribution. The iterative adaptive approach (IAA) [3], [4] is one such semi-parametric estimator that has been found to offer preferable performance in several applications (see, e.g.,[5]–[14]). The IAA estimate is formed over a grid of potential dictionary candidates, often being selected as sinusoidal components over the entire range of frequencies, without imposing assumptions of the signal being sparse, or using knowledge of the assumed model order. The estimate iteratively computes the contribution for each such component using a weighted least squares (WLS) formulation, wherein the WLS estimates are weighted using the covariance matrix formed from the contributions found in the earlier iteration.

Due to its iterative formulation, with each WLS estimate depending on the inverse covariance matrix formed from all components, the IAA estimate is computationally cumbersome, especially when being formed over a fine dictionary and/or multiple dimensions. To alleviate this drawback, significant efforts have been made to formulate computationally efficient implementations [15]–[18]. As the IAA estimate is formed using the inverse covariance matrix, the formulation requires the estimate to be computed using the same spectral resolution over all frequencies, without allowing for a refined estimation over regions of interest.

The here proposed zooming procedure for IAA (zIAA) is formed initially using a coarse dictionary. The contribution of each element is then refined, such that parts of the spectrum that are deemed to contain only (possibly colored) noise are formed using only one initial run of the refining algorithm, whereas regions of interests are refined further using a finer dictionary covering only the frequency ranges of interest. As the contribution from all frequencies are still formed, the covariance matrix will not become poorly conditioned, as would be the case if only parts of the spectrum were refined. This zooming procedure may then be repeated until the signal of interest has been estimated using the desired resolution. As the WLS estimates only need to be formed over the frequencies of interest, the resulting estimate require a notably lower computational complexity as compared to forming the classical IAA spectrum. Typically, further computational reductions may also be achieved using the inherent structure of the estimates, reminiscent to the approaches in [15]–[18].

In this work, exploiting the ideas introduced in [19], [20] for the LASSO, we also investigate a zoomed IAA estimator formed over wideband integrated dictionary elements. These elements are formed using bands of frequencies, together covering the entire frequency range, thereby capturing all the energy in the band and resulting in more precise estimates. By iteratively forming a refined dictionary over the bands of interest, the proposed algorithm allows for a zooming over frequency regions of interest, without requiring knowledge of the number of signal components within such regions. A related zooming procedure was proposed in [21] for spatial sources. Such a procedure is not applicable for IAA, as it results in different spectral resolutions for different part of the spectra, which will cause numerical instability in the IAA estimate. Furthermore, just as other similarly introduced zooming procedures, such an approach using a narrowband dictionary will increase the risk of missing off-grid elements in the initial coarse estimation steps and therefore exclude this part of the spectra from further zooming. On the other hand, the here proposed method makes use of uniform element spacing covering the full band, ensuring that no components are missed in the initial search step [20].

2. The zIAA algorithm

Without loss of generality, we proceed to outline the proposed algorithm using a dictionary consisting of sinusoidal components, although it should be noted that other alternative dictionary elements could also be used (see, e.g., [11], [15]). Consider N (possibly non-uniform) samples of a signal $\mathbf{y}$, consisting of K complex sinusoids measured in the presence of an additive noise, such that

$$\mathbf{y} = \tilde{\mathbf{A}}\boldsymbol{\alpha} + \mathbf{w}, \tag{1}$$

where

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_K \end{bmatrix}, \tag{2}$$

$$\mathbf{a}_k = \begin{bmatrix} e^{2i\pi f_k t_1} & \dots & e^{2i\pi f_k t_N} \end{bmatrix}^T, \tag{3}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \dots & \alpha_K \end{bmatrix}^T, \tag{4}$$

with $(\cdot)^T$ denoting the transpose, $\{t_n\}_{n=1}^N$ the sampling times, α_k and f_k the complex amplitude and the frequency of the kth component, respectively, and $\mathbf{w}$ is an additive noise vector. The IAA estimate is formed by expanding the dictionary over the full range of frequencies, such that the used dictionary, $\mathbf{A}$, contains $P \gg N$ dictionary elements. These elements are selected fine enough so that each of the frequencies of interest can be deemed to be (almost) coinciding with one of the candidates. The IAA estimate is then computed by iteratively

solving

$$s_k = \frac{\mathbf{a}_k^H \mathbf{R}^{-1} \mathbf{y}}{\mathbf{a}_k^H \mathbf{R}^{-1} \mathbf{a}_k}, \text{ for } k = 1, \dots, P, \tag{5}$$

$$\mathbf{P} = \text{diag}(s_1, \dots, s_P), \tag{6}$$

$$\mathbf{R} = \mathbf{A} \mathbf{P} \mathbf{A}^H, \tag{7}$$

until convergence (typically about 10-15 iterations are sufficient for convergence [3], [4]). Here, $(\cdot)^H$ denotes the conjugate transpose, and $\text{diag}(\cdot)$ a diagonal matrix formed with the specified vector along its diagonal. Upon convergence, the estimated complex amplitudes, s_k, coinciding with the components of interest will thus form estimates of the corresponding components in $\boldsymbol{\alpha}$.

As the IAA estimate forms a WLS estimate in (5), for each of the P frequencies in the dictionary, the inverse covariance matrix, $\mathbf{R}^{-1}$, needs to be recomputed in each iteration of the algorithm. Although this may be done computationally efficiently by exploiting the inherent structure of the dictionary and the Toeplitz structure of the covariance matrix [15], [16], the computational cost remains significant for large dictionaries and/or if the estimate is formed over several dimensions. Furthermore, to avoid poor conditioning, the inverse covariance matrix needs to be formed on a uniformly spaced frequency grid, necessitating that P is selected large enough to yield the desired resolution over the frequencies of interest.

In this work, we propose exploiting integrated wideband dictionary elements[1] in $\mathbf{A}$ [19], [20], such that each element is formed over the band

$$\mathbf{a}_b(t_n) = \int_{f_b}^{f_{b+1}} e^{2i\pi f t_n} df = \frac{e^{2i\pi f_{b+1} t_n} - e^{2i\pi f_b t_n}}{2i\pi t_n}. \tag{8}$$

As the resulting dictionary elements cover the band of frequencies from f_b to f_{b+1}, for $b = 1, \dots, B_1$, the entire spectrum may be formed with $B_1 \ll P$, where B_1 denotes the (initial) number of wideband dictionary elements.

The wideband zoomed IAA estimate is then formed similar to the IAA estimate, by iteratively forming (5)-(7), although with only B_1 amplitudes estimates in place of P. After convergence, the spectral regions of interest are selected, typically by either exploiting *a priori* information or by retaining only the spectral regions with significant power, say above some threshold, τ. The dictionary is then refined such that each selected wideband component is replaced with Z components (thus forming new expanded dictionary matrix $\mathbf{A}_2$ with size $N \times B_1 Z$), such that they together constitute the frequencies covered by the earlier wideband dictionary element. Now the new grid, with size $B_1 Z$, is

---

[1]It should be noted the proposed zooming framework does not require the use of a wideband dictionary; it is here used for illustration as it has been found to often yield preferable performance [19], [20].
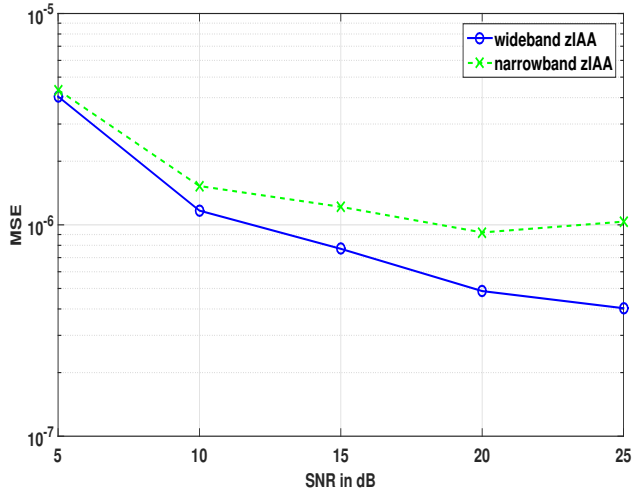
Figure 1 Frequency estimation mean-square error curves for the zIAA with wideband dictionary and narrowband dictionary.

divided into two disjoint sets, one containing the gridpoints corresponding to the noise part of the spectrum, and one containing the gridpoints corresponding to the signal of interest.

Using the estimated covariance matrix, $\mathbf{R}$, attained from the old grid of size B_1, the amplitudes for the new grid are updated using a single iteration of (5). From now on, the amplitudes corresponding to the noisy parts are fixed and only the amplitudes corresponding to the signal of interest are updated. These are updated by iterating (5)-(7) until convergence. This scheme is much faster than the original IAA version due to the iterative zooming procedure. The complexity will also be reduced since we only need to estimate the noise part of the spectrum in one pass, thus we only need to iterate (5)-(7) for the amplitudes corresponding to the signal of interest. In this way, the spectral estimate is refined also for these frequencies, although at a low computational cost, allowing the covariance matrix to be formed from the contributions from both forms of refined estimates.

It is worth noting that this implies that the resulting covariance matrix estimate, formed both using the refined wideband dictionary, reminiscent to (7), and from the corresponding noise floor estimate, will have an inherent structure that may be exploited when forming a computationally efficient implementation. However, even without exploiting such structure, the proposed algorithm yields a notable computational reduction as compared to the classical IAA formulation. The procedure may then be refined further after each iteration, until the estimate has been formed at the desired resolution.

To increase the speed-up further, one may refrain from updating the covariance matrix when performing the zooming procedure. This results in that the estimated

	10 dB	15 dB	20 dB	25 dB
IAA	54ms	52ms	56ms	52ms
wb zIAA	32ms	32ms	34ms	32ms
Times faster	**1.7x**	**1.6x**	**1.7x**	**1.6x**
wb zIAA$_R$	12ms	11ms	12ms	12ms
Times faster	**4.5x**	**4.6x**	**4.5x**	**4.5x**

Table 1 Average runtime in milliseconds for the IAA and zIAA estimates and speed-up comparing to ordinary IAA, using only a single refinement step.

covariance obtained from the first IAA solution is used throughout the algorithm. This will decrease the computational complexity but will also degrade the estimation precision. The pros and cons with this relaxation are investigated in the next section.

The work presented here is based on a simple thresholding idea where bands are deemed active if the energy in a band is higher than some predefined threshold τ. However, different model-order selection tools can be incorporated to the proposed framework. For instance, one may use the Bayesian information criterion (BIC) [22], as was done in the original IAA paper [4] (see also, e.g., [23], for alternative methods). Another idea is to choose the threshold τ as a fixed ratio to the largest peak found in the resulting spectrum, making the threshold data-dependent, such that $\tau = \alpha \max \{s_k\}$, where $\alpha \in (0, 1]$.

3. Numerical examples

We start by comparing the difference in performance from using a wideband dictionary to a narrowband dictionary. Both dictionaries use $B_1 = 200$ elements in the first stage and $Z = 5$ zooming elements for each active band in the following zooming step. We simulated a signal of length $N = 50$ containing $K = 3$ complex-valued sinusoids, corrupted by a zero-mean white (circularly symmetric) Gaussian noise. Frequencies were uniformly drawn from $[0, 1)$ and with random phase, all having unit magnitude. We are considering (the sum of) the mean-square error (MSE) for different signal-to-noise ratios (SNR), here defined as

$$\text{SNR} = 10\log_{10}\left(\frac{P_y}{\sigma^2}\right), \tag{9}$$

where P_y is the power of the signal and σ^2 is the variance of the Gaussian noise. The resulting MSE curves of the estimated frequencies are presented in Figure 1. As can be seen from the figure, the wideband dictionary clearly outperforms the narrowband dictionary due to its ability to capture all the energy in the band, also for off-grid components, and to produce more precise estimation of a spectra.

	10 dB	15 dB	20 dB	25 dB
IAA	5%	32%	64%	89%
wb zIAA	13%	66%	86%	92%
wb zIAA$_R$	12%	63%	82%	83%

Table 2 Percentage of correct model order estimations for different SNR levels.

We proceed with examining whether it is necessary to update the covariance matrix at all after the convergence of IAA with the initial coarse grid. To answer this question, we estimate the frequencies using three different versions of IAA using the same setup as described above, but increasing number of sinusoids to $K = 7$. The first one being the original version of IAA using $P = 1000$ sinusoids in its dictionary. The rest of the two versions are using a wideband dictionary with updated $\mathbf{R}$ as described originally, i.e., zIAA, and using wideband dictionary without updating $\mathbf{R}$, denoted zIAA$_R$. Both of these two versions had an initial grid size of $B_1 = 200$, and used $Z = 5$ for the zoom in. All the three estimators thus have a resulting dictionary containing 1000 elements, yielding the same resolution. Figure 2 shows the MSE of the estimated frequencies and Table 1 gives the corresponding runtimes for the estimates, indicating that even without using the inherent structure of the proposed estimator, the wideband zIAA is about 1.6 times faster than the classical IAA estimator, even when using only a single refinement step. For the wideband zIAA$_R$ case where covariance matrix $\mathbf{R}$ does not get updated after the initial IAA cycle, we can see that the proposed estimator is about 4.5 times faster than the classical IAA, and still is able to outperform it in the MSE sense and only showing marginally worse estimation performance in comparison with the wideband zIAA. Table 2 summarizes the percentage of correct model order estimations for each considered algorithm. Here, a method was considered returning the correct model order if its $K = 7$ highest peaks were within two grid-points of the actual frequency value. For lower SNR, the methods all show similar and poor performance, with close to 0% correct model order estimates and very noisy spectral estimates. As is clear from the table, the proposed refinement step actually improves the likelihood of a correct model order estimate, although it may be noted that the zIAA$_R$ version seems to be performing marginally worse than zIAA.

Next, we compare ordinary IAA with both narrow- and wideband dictionary based zoomed IAA estimators in terms of a peak resolution ability. Here, we use a signal of length $N = 50$ consisting of two sinusoids with varying distance and SNR = 20 dB. Ordinary IAA uses $P = 1000$ elements and both zoomed IAA dictionaries use $B_1 = 200$ elements in the first stage and $Z = 5$ zooming elements for each active band in the following zooming step, thus resulting in estimates with the same resolution. In order to determine the percentage of resolved peaks, we let $P(\omega)$ denote an amplitude estimate at a frequency ω,
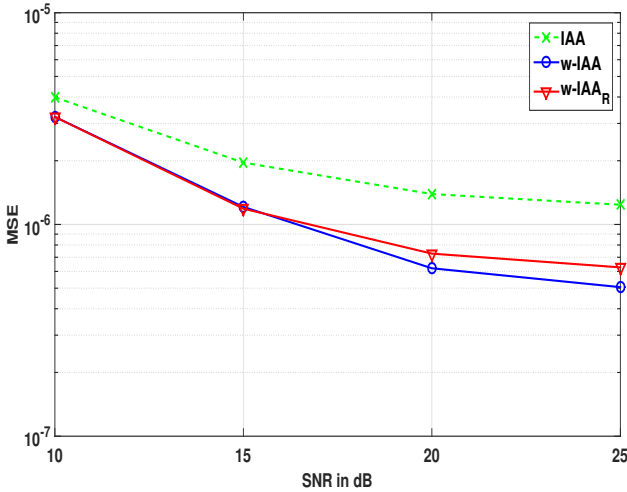
Figure 2 Frequency estimation mean-square error curves for the ordinary IAA and zIAA estimates. As can be seen from the figure, the introduced speed-up does not yield any loss of performance as compared to the ordinary IAA estimate, and is even, when using the wideband dictionary, yielding somewhat preferable performance.

forming the measure (see, e.g., [24], [25])

$$\gamma = 2P(\omega_3) - P(\omega_1) - P(\omega_2) < 0, \tag{10}$$

where ω_3 is the midway frequency between ω_1 and ω_2

$$\omega_3 = (\omega_1 + \omega_2)/2. \tag{11}$$

The sinusoids are then deemed resolvable if $\gamma < 0$ and irresolvable otherwise. Figure 3 shows the resolution ability of the estimators, again indicating that the proposed wideband zIAA algorithm yields preferable performance. All the results in this section were obtained from 100 Monte-Carlo runs.

4.   Conclusion

In this work, we have investigated how to allow for a larger grid for the IAA estimator by incorporating a zooming procedure. The IAA estimator is prone to experience numerical issues when the underlying grid structure is imbalanced. To avoid having to increase number of grid points such that the whole grid structure has the desired resolution, we instead propose an iterative formulation, where we use a wideband dictionary instead of the more traditional narrowband dictionary. The proposed estimator is shown to offer a significant speed-up as compared to the regular IAA formulation, without suffering from any decrease in performance.
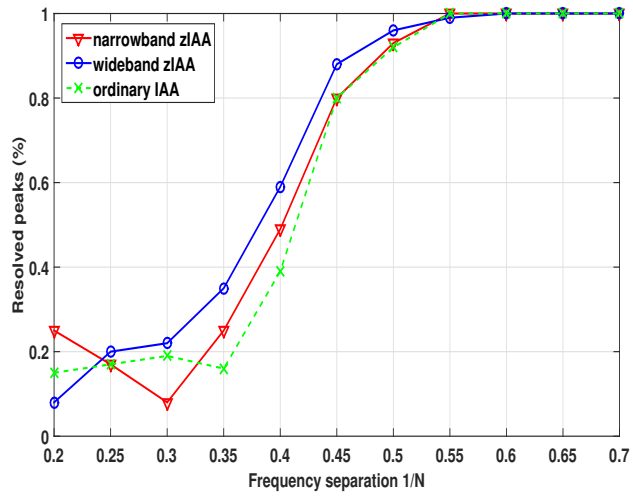
Figure 3 The peak resolving ability of the discussed estimators; one may again note that the proposed speed-up does not reduce the resolution of the resulting estimates. As expected, the use of a wideband dictionary is even yielding a somewhat improved performance.

References

[1]   P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[2]   D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*. McGraw-Hill Boston, 2000.

[3]   T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, 2010.

[4]   T. Yardibi, J. Li, and P. Stoica, "Nonparametric and sparse signal representations in array processing via iterative adaptive approaches", in *Asilomar*, Pacific Grove, USA, Oct. 2008.

[5]   P. Stoica, J. Li, and J. Ling, "Missing Data Recovery via a Nonparametric Iterative Adaptive Approach", *IEEE Signal Process. Lett.*, vol. 16, pp. 241–244, Apr. 2009.

[6]   L. Du, J. Li, P. Stoica, H. Ling, and S. S. Ram, "Doppler spectrogram analysis of human gait via iterative adaptive approach", *Electronics Letters*, vol. 45, no. 3, pp. 186–188, 2009.

[7]   Z. Chen, J. Li, P. Stoica, and K. W. Lo, "Iterative adaptive approach for wide-band active sonar array processing", in *OCEANS 2010 IEEE-Sydney*, IEEE, 2010, pp. 1–10.

[8]   W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. A. Sadjadi, "Iterative Adaptive Approaches to MIMO Radar Imaging", *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 5–20, Feb. 2010.

[9]   X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse learning via iterative minimization with application to mimo radar imaging", *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1088–1101, 2011.

[10] D. Vu, L. Xu, M. Xue, and J. Li, "Nonparametric missing sample spectral analysis and its applications to interrupted sar", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 1–14, 2012.

[11] J. R. Jensen, G. O. Glentis, M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Computationally Efficient IAA-based Estimation of the Fundamental Frequency", in *European Signal Processing Conference*, Bucharest, Romania, Aug. 2012.

[12] Z. Yang, H. W. X. Li, and W. Jiang, "Adaptive Clutter Suppression Based on Iterative Adaptive Approach For Airborne Radar", *Elsevier Signal Processing*, vol. 93, pp. 3567–3577, 2013.

[13] Y. Chen, H.-C. So, and W. Sun, "ℓp-norm based iterative adaptive approach for robust spectral analysis", *Signal Processing*, vol. 94, pp. 144–148, 2014.

[14] G. O. Glentis, K. Zhao, A. Jakobsson, H. Abeida, and J. Li, "SAR Imaging via Efficient Implementations of Sparse ML Approaches", *Elsevier Signal Processing*, vol. 95, pp. 15–26, 2014.

[15] G. O. Glentis and A. Jakobsson, "Efficient Implementation of Iterative Adaptive Approach Spectral Estimation Techniques", *IEEE Trans. Signal Process.*, vol. 59, pp. 4154–4167, Sep. 2011.

[16] M. Xue, L. Xu, and J. Li, "IAA Spectral Estimation: Fast Implementation using the Gohberg-Semencul Factorization", *IEEE Trans. Signal Process.*, vol. 59, pp. 3251–3261, Jul. 2011.

[17] G.-O. Glentis and A. Jakobsson, "Superfast Approximative Implementation of the IAA Spectral Estimate", *IEEE Trans. Signal Process.*, vol. 60, pp. 472–478, Jan. 2012.

[18] J. Karlsson, W. Rowe, L. Xu, G. Glentis, and J. Li, "Fast Missing-Data IAA by Low Rank Completion", in *38th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[19] M. Butsenko, J. Swärd, and A. Jakobsson, "Estimating Sparse Signals Using Integrated Wide-band Dictionaries", in *42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, New Orleans, USA, Mar. 2017.

[20] ——, "Estimating Sparse Signals Using Integrated Wideband Dictionaries", *IEEE Trans. Signal Process.*, vol. 66, pp. 4170–4181, Aug. 2018.

[21] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays", *IEEE transactions on signal processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[22] G. Schwarz, "Estimating the Dimension of a Model", *Ann. Stat.*, vol. 6, pp. 461–464, 1978.

[23] P. Stoica and Y. Selén, "Model-order Selection — A Review of Information Criterion Rules", *IEEE Signal Process. Mag.*, vol. 21, pp. 36–47, Jul. 2004.

[24] A. Jakobsson and P. Stoica, "Combining Capon and APES for Estimation of Spectral Lines", *Circuits, Systems, and Signal Processing*, vol. 19, no. 2, pp. 159–169, 2000.

[25] Q. T. Zhang, "A Statistical Resolution Theory of the AR Method of Spectral Analysis", *IEEE Trans. Signal Process.*, vol. 46, pp. 2757–2766, Oct. 1998.

Conclusions

This thesis presented a wideband dictionary framework for estimating sparse signals. This formulation allows for considerable reduction in computational complexity and decreases the probability of missing off-grid components. For situations where the number of samples is considerably smaller than the number of dictionary elements, the percentage of correct model order estimation for the proposed wideband dictionary is $40 - 50\%$ higher than for the conventional method ($90 - 100\%$ vs $50 - 60\%$). As most of the errors in this situation come from missing off-grid components, it can be stated that wideband elements provide reliable remedy for such cases, without increasing complexity of problem formulation and therefore limiting the size of the considered problems as atomic-norm minimization; or loosing the benefits of convexity of a solution as in the case of adaptive grid approach. By employing an iterative zooming procedure and decreasing risk of missing off-grid components we showed that it is possible to formulate the problem with a smaller initial dictionary and therefore reduce the amount of computation that are needed for the required resolution. For the same resolution, the computational complexity of the proposed method can be $20 - 30$ times lower, which results in a considerable reduction in time required to make an estimation.

This work examines more closely sparse reconstruction method based on solving LASSO minimization; however, the approach is not limited to only this method and the proposed procedure for constructing the dictionary is suitable for any reconstruction method where one generally considers dictionaries consisting of candidate elements. We also have showed that such a formulation results in similar improvements for SPICE and IAA estimators for example. For the latter, we proposed the novel approach which allows for higher precision estimates on relevant areas of the spectrum, which was complicated beforehand as IAA uses WLS formulation, which depends on inverse covariance matrix based on estimates from the earlier iteration. Such a formulation is cumbersome as one has to compute the estimate using the same spectral resolution over all frequencies, which requires large dictionary. The proposed algorithm results in increased estimation performance in terms of frequency estimation as well as peak resolving ability. The average runtime for the proposed method is up to 4.5 times faster.

For most part of the discussion we have considered so-called integrated wideband dictionaries; however, this choice was rather arbitrary for the initial

evaluation of the idea and we assume that different methods for forming wideband dictionary elements can be proposed. We have also evaluated the DPSS-based dictionary elements and confirmed that similar estimation performance can be expected. Also, most of the experiments were conducted on synthetic data were the signal was generated from the signal model. However, to evaluate our assumptions and performance of the proposed methods on the real-life signals we considered two examples of measured data. We have examined performance of the wideband framework on NMR data, where the wideband estimate is formed 20 times faster than the conventional narrowband estimate. We also have considered electrical bio-impedance signals. We have described the method of sparse reconstruction for the separation of cardiac and respiratory signal components from electrical bio-impedance measurements, which is able to reliably reconstruct both components from the signal. However, further work is needed to improve the robustness of the proposed method to the half the fundamental frequency problem.

When comparing the proposed wideband framework to the conventional narrowband one, it is important to note that additional design complexity arises by the addition of the user-defined parameters, which should be defined for choosing number of iteration steps and number of dictionary elements at each iteration. We have provided guidelines and theoretical formulation for the best approach for the parameter selection. However, it has to be noted that the huge benefit of wideband dictionary lies in its adaptability and intuitive implementation. One does not need to change the optimization algorithm, it is only required to change the dictionary formulation and add iterative procedure, which are both straightforward tasks. We hope that this encourages those working in the area of signal estimation to consider applying the proposed method for their problems.

To summarize the conclusion let us look at the answers that this thesis gives to the research questions that were posed in the beginning of this work.

Research questions:

1. Can wideband dictionary formulation help to mitigate off-grid estimation issue for sparse signal reconstruction?

   Answer: Yes. This work showed that wideband dictionaries can provide remedy for this problem. Wideband dictionary element span the whole sub-band and therefore reduce probability of missing certain parts of the signal.

2. If wideband based dictionary is suitable, then what sort of computational complexity reduction can one expect from iterative formulation of the estimation problem?

Answer: For the same resolution, the computational complexity can be up to 20-30 times lower.

3. How suitable are different types of wideband elements?

    Answer: This work investigated integrated sinusoid based dictionaries as well as DPSS-based dictionaries. Both dictionaries showed that they provide similar estimation performance in terms of MSE and model order estimation. There were not any substantial differences noted between two types of the wideband dictionaries.

4. What are the possible drawbacks of wideband dictionaries compared to classical narrowband dictionaries?

    Answer: Major drawback is the necessity of choosing additional user-defined parameters for the algorithm.

5. How well does synthetic tests correspond to actual real-life data?

    Answer: We have seen that by running estimators on real-life data, such as NMR spectroscopy and EBI signals, we can expect similar estimation performance as well as similar reduction in computational complexity as on the synthetic data that was generated in MATLAB during experiments.

6. What other estimators can benefit, and in what way, from dictionaries constructed in a wideband manner?

    Answer: Main investigation was conducted and main results formulated by using LASSO estimator. However we also showed that for example IAA and SPICE estimators showed similar improvement by employing wideband dictionaries. In general dictionary-based estimators can be expected to benefit from the usage of wideband dictionaries.

Acknowledgements

The decision to switch from an industry position and start PhD studies was quite spontaneous for me. For that I have to thank Dr. Tõnu Trump and his lectures in adaptive filtering and signal processing. They inspired me and showed that academia can be exciting and is full of interesting problems waiting to be solved. In addition to that, Dr. Trump was my first supervisor and I am very grateful for his support and guidance during the first half of my studies.

My gratitude goes to my supervisors Prof. Olev Märtens and Prof. Yannick Le Moullec for their great support during the final years of my studies. Also to Prof. Enn Velmre for his help in proof-reading and providing valuable feedback on the draft of this work. I am very thankful to Prof. Toomas Rang, the head of the Thomas Johann Seebeck Department of Electronics, for his support and motivation during the final stage of my studies.

To my friends and wonderful colleagues at Lund University - Prof. Andreas Jakobsson, Dr. Johan Swärd, Dr. Ted Kronvall and Filip Elvander. You have taught me a lot and changed the way I look at academia. Thank you for this valuable lesson and for your friendship.

To my friends and colleagues at Tallinn University of Technology with whom we shared day-to-day fun and labour of PhD studies: Sander Ulp, Egon Astra, Julia Berdnikova, Hip Kõiv, Dr. Ahti Ainomäe and Dr. Andrei Krivošei - thank you for all the great moments.

To my family, everything I accomplished is because of your dedication, eternal support and patience. I am forever grateful!

To my dear girlfriend Darja. Thank you for your love, support and all the happiness you brought to my life!

Tallinn, 2018   *Maksim Butsenko*

Abstract
Parameter Estimation by Sparse Reconstruction with Wide-band Dictionaries

Parameter estimation in general and spectral analysis in particular have been fruitful research areas for a long time and still rightfully remain so. Many algorithms and methods for parameter estimation were proposed during decades of research, but as technology evolves, so evolve our requirements to estimators. Non-parametric estimators were for a long time the most popular method of spectral analysis; however, their major drawback is their resolution limitation and high variance. Parametric estimators can provide high-resolution estimates, however this requires considered signal to correspond well to underlying signal model and they perform much worse than non-parametric estimators if this is not the case. Semi-parametric estimators can often provide high-resolution estimates without big dependency on the signal model as their only assumption is that the signal should be sparse. In fact, a wide range of common applications considers signals that are well approximated by sparse reconstruction framework, and this area has attracted noteworthy interest in the recent literature. Considerable number of these work focuses on formulating convex optimization algorithms that make use of different sparsity inducing penalties, thereby encouraging solutions that are well represented using only a few elements from some dictionary matrix. It can be shown that when the dictionary is chosen properly, even limited number of measurements allows for an accurate signal reconstruction.

In this work a novel procedure for constructing dictionaries for parameter estimation by sparse reconstruction methods is considered. Instead of forming the dictionary as a finite set of discrete narrowband components for evaluation of continuous parameter space, this work considers wideband dictionary elements, such that continuous parameter space is divided into B subsets. During the estimation procedure, the activated subsets are selected for further refinement and non-activated subsets are discarded from further optimization. Afterwards, a new dictionary is formed for each of the activated subsets, resulting in the zoomed dictionary for that particular region of considered parameter space. An iterative procedure may then be repeated further until required resolution is reached.

The initial problem statement and the plausibility of the approach are validated by showing that the method is suitable for one-dimensional data. The

formulation of the wideband framework for multi-dimensional data and validation on different estimators (LASSO, SPICE, IAA), sampling scenarios and data from different sources are considered. An efficient implementation of the algorithm by alternative direction method of multipliers is formulated and corresponding complexity reduction calculations considered. A comprehensive overview of the best approaches to selection of the parameters for the framework is provided. The proposed approach is tested mainly by using corresponding signal model and conducting series of numerical experiments by running multiple Monte-Carlo simulations. However, the proposed method is also tested on real-life signals by considering NMR data and by investigating the possibility of employing similar sparse reconstruction framework for the separation of cardiac and respiratory signal components from the electrical bioimpedance measurements.

The wideband framework allows for considerable reduction in computational complexity and decreases probability of missing off-grid components. For situations where the number of samples is considerably smaller than the number of dictionary elements, the percentage of correct model order estimation for the proposed wideband dictionary is $40 - 50\%$ higher than for the conventional method ($90 - 100\%$ vs $50 - 60\%$). Most of the errors in this situation come from missing off-grid components. Similar methods for grid-selection issues exist; however, they often impose increasing complexity of the problem formulation and therefore limiting size of considered problems as for example in the case of atomic-norm minimization; or loosing the benefits of convexity of a solution as in the case of adaptive grid approach. By employing iterative zooming procedure and decreasing risk of missing off-grid components we show that it is possible to formulate problem with smaller initial dictionary and therefore reduce amount of computations that are needed for the required resolution. For the same resolution, the computational complexity of the proposed method can be $20 - 30$ times lower, which results in considerable reduction in the time required to make an estimation.

The proposed wideband dictionary method has the additional benefit of its adaptability and intuitive implementation. This can help to encourage those working in the area of signal estimation to consider applying the proposed method for their problems as wideband dictionary successfully replaces the classical narrowband dictionary for considered problems by providing at least similar performance and often outperforming the classical framework.

Kokkuvõte
Signaali parameetrite hindamine kasutades hajusat taastamist laiaribaliste sõnastikega

Signaali parameetrite hindamine üldiselt ja spektraalanalüüs eriti on pikalt olnud väga tulemuslikud uurimissuunad ning on seda endiselt. Aastate jooksul on välja pakutud mitmeid algoritme ja meetodeid, kuid tehnoloogia arenedes on arenenud ka meie nõudmised signaali omadusi hindavatele algoritmidele. Mitte-parameetrilised meetodid on olnud pikka aega enamkasutatavate seas, kuid nende miinuseks võib lugeda signaali piiratud lahutusvõime ja suurt dispersiooni. Parameetrilised meetodid pakuvad kõrget lahutusvõimet, kuid selleks peab vaadeldav signaal hästi kokku langema olemasoleva signaali mudeliga. Vastasel juhul on parameetriliste meetodite tulemus halvem kui mitte-parameetrilistel. Semi-parameetrilised meetodid on võimelised pakkuma kõrget signaali lahutusvõimet ilma, et nad sõltuks oluliselt signaali mudelist, kuna nende ainuke eeldus on, et signaal peaks olema hõre (ingl.k. sparse) – seletatav vaid vähese arvu signaalikomponentide abil. Tõesti, suur osa rakendusi tegelebki signaalidega, mis võivad olla aproksimeeritud hõreda taastamise abil ning antud valdkond on pälvinud suurt tähelepanu viimaste aastate kirjanduses. Suur osa nendest töödest on pühendatud kumerate algoritmide (ingl.k. convex optimization) formuleerimisele, mis kasutavad erinevaid hõredust taotlevaid piiranguid ning sellega saavutavad lahendusi, mis on esitatavad vaid mõne elemendiga sõnastiku maatriksist. Kui sõnastik on sobivalt valitud, siis on võimalik näidata, et juba väga piiratud arv mõõtmisi võib viia signaali piisavalt täpse taastamiseni.

Selles töös tutvustatakse uut meetodit sõnastike moodustamiseks signaali parameetrite hindamiseks kasutades hõredat taastamist. Selle asemel, et moodustada sõnastik kui lõplik hulk diskreetseid kitsaribalisi komponente pideva parameetri ruumi hindamiseks, vaatleb antud töö laiaribalisi sõnastiku elemente, kus pidev ruum on jaotatud B alamhulgaks. Hindamisprotseduuri jooksul need alamhulgad, kus määratakse signaal olevat, on valitud täpsustamiseks ja alamhulgad, kus signaali olemasolu ei ole määratud, on välja arvatud edaspidisest optimeerimisest. Seejärel moodustatakse uus sõnastik iga aktiivse alamhulga jaoks, mille tulemusel saame täpsema sõnastiku selle konkreetse regiooni

kirjeldamiseks. Antud iteratiivset protseduuri võib seejärel korrata kuni soovitud lahutusvõime on saavutatud.

Algne probleemi püstitus ja kirjeldatud lähenemise valideerimine oli näidatud ühemõõtmeliste signaalide jaoks ning seejärel oli laiaribaliste sõnastike meetod formuleeritud mitmemõõtmeliste signaalide jaoks ja valideeritud erinevate algoritmide jaoks (LASSO, SPICE, IAA), ühtlase ja mitte-ühtlase sammuga diskreetimise meetodite ja erinevatest signaaliallikatest pärit signaalide jaoks. Töös on vaadeldud efektiivset algoritmi formuleerimist vahelduvate suundadega kordajate meetodiga (ingl.k alternating direction method of multipliers) ja välja toodud vastav algoritmilise keerukuse arvutus. Lisaks on toodud ulatuslik ülevaade parameetrite valimise parimatest viisidest. Meetod on valideeritud põhiliselt kasutades signaali mudelit ning eksperimenteerides Monte Carlo simulatsioonidega. Lisaks on meetod valideeritud reaalsete signaalidega, näiteks tuumamagnetresonantsspektroskoopia signaalide hindamisel ning elektrilise bioimpedantsi signaalide puhul. Laiaribaliste sõnastike kasutamine annab märkimisväärse vähenemise algoritmilises keerukuses ning vähendab ka tõenäosust nn. võrest väljaspool asuvaid komponente mitte märgata. Nendel juhtudel, kui diskreetide arv on väiksem kui sõnastikus olevate elementide arv, oli õige mudeli järgu tabamise protsent välja pakutud meetodi jaoks $40-50\%$ kõrgem kui harilike meetodite jaoks ($90-100\%$ vs $50-60\%$). Enamik vigu antud olukorras tuleb võrest väljaspool asuvate komponentide mitte tabamisest. Sarnaseid võre valimise meetodeid eksisteerib teisigi, kuigi tihti kaasneb nendega probleemi formuleerimise matemaatilise keerukuse kasv, mis seega piirab vaadeldava probleemi suurust nagu näiteks atomic-normi minimeerimise puhul; või kaotab nende lahendus kumeruse eeliseid nagu adaptiivse võre puhul. Välja pakutud iteraatiivse täpsustamise meetodi kasutamisel langeb oht võrest väljaspool asuvaid komponenti mitte tabada ning antud töös näidatakse, et on võimalik formuleerida probleemi ka väiksema algsõnastiku abil ning selle kaudu vähendada arvutusi, mis on vajalikud soovitud lahutusvõime saavutamiseks. Sama signaali lahutusvõime jaoks on välja pakutud meetodi algoritmiline keerukus kuni $20-30$ korda väiksem, mis märkimisväärselt vähendab aega, mis on vajalik signaali hindamiseks. Antud töös välja pakutud laiaribalise sõnastiku meetodi lisaeelised on ka tema adaptiivsus ning lihtne juurutamine. Antud asjaolud peaksid julgustama kolleege, kes töötavad signaali hindamise valdkonnas, et nad kaaluks antud meetodi rakendamist nende töös esinevatele probleemidele, kuna laiaribalise sõnastiku meetod asendab edukalt harilikku kitsaribalist sõnastiku pakkudes sarnast ning tihti ka tunduvalt paremat tulemust väiksema algoritmilise keerukusega.

Curriculum Vitae

Education

2013– ...	**Doctor of Philosophy**, *Information and Communication Technology*, Tallinn University of Technology.
2011–2013	**Master of Science**, *Telecommunications Engineering*, Tallinn University of Technology, *cum laude*.
2008–2011	**Bachelor of Science**, *Telecommunications Engineering*, Tallinn University of Technology, *cum laude*.
2007–2008	**Military Service**, Kuperjanov Infantry Battalion.
2004–2007	**Secondary Education**, Nõo Science Gymnasium.

Working Experience

02.2018– ...	**Data Scientist**, Taxify.
01.2017– ...	**Early Stage Researcher**, Tallinn University of Technology, Thomas Johann Seebeck Department of Electronics.
10.2016–02.2017	**Data Science Internship**, Mooncascade.
02.2016–05.2016	**Visiting Researcher**, Lund University, Department of Mathematical Statistics.
08.2013–12.2016	**Early Stage Researcher**, Tallinn University of Technology, Department of Radio and Communications Engineering.
04.2013–08.2013	**Technical Consultant**, Levira, Play-Out Center.
01.2011–03.2013	**Technical Manager**, Levira, Play-Out Center.
09.2010–12.2010	**Play-Out Engineer**, Levira, Play-Out Center.

Teaching Experience

University courses:
- Master's level:
  - Standardization in Telecommunication
  - Data Compression
  - Adaptive Signal Processing
- Bachelor's level:
  - Telecommunication
  - Signal Processing I
  - Computer Networks

Mektory School of Technology: *Multimedia and Telecommunication*

TTÜ continuing education courses: *Multimedia and Telecommunication*

TTÜ educational lectures in secondary schools around Estonia: *Multimedia*

Publications

2018   M. Butsenko, J. Swärd and A. Jakobsson, "Zoomed Iterative Adaptive Approach", has been accepted to *2018 International Symposium on Intelligent Signal Processing and Communication Systems*.

2018   A. Krivošei, M. Min, P. Annus and M. Butsenko, "Decomposition of the EBI Signal into Components using two Channel Cross-Compensating Singular Spectrum Analysis" in *13th Annual IEEE International Symposium on Medical Measurements & Applications*, Rome, Italy, 2018.

2018   M. Butsenko, O. Märtens, A. Krivošei and Y. Le Moullec, "Sparse Reconstruction Method for Separating Cardiac and Respiratory Components from Electrical Bioimpedance Measurements" in *Elektronika ir Elektrotechnika*, vol. 24, no. 5, pp. 57-61, 2018.

2018   Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. Estimating Sparse Signals Using Integrated Wide-band Dictionaries in *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4170-4181, 2018.

2017   M. Butsenko, J. Swärd and A. Jakobsson, "Estimating Sparse Signals Using Integrated Wide-Band Dictionaries", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, USA, March 5-9, 2017.

2016   M. Butsenko and T. Trump, "An Affine Combination of Adaptive Filters for Channels with Different Sparsity Levels", *Telfor Journal*, Vol. 8, No. 1, pp. 32-37, July 2016.

2015   M. Butsenko and T. Trump, "An Affine Combination of Adaptive Filters for Sparse Impulse Response Identification", *Telecommunications Forum Telfor (TELFOR)*, 2015 23rd, Belgrade, 2015, pp. 396-399.

Computer Skills

Basic	PHP, C, Hadoop
Intermediate	R, SQL
Advanced	MATLAB, Python

Languages

Russian	**Native**
Estonian	**Fluent**
English	**Fluent**
French	**Basic**
Swedish	**Basic**

Interests

- Music
- Backpacking
- Football
- Education
- Homebrewing
- Badminton

Elulookirjeldus

Hariduskäik

2013– ...	**Doktoriõpe**, *Info- ja kommunikatsioonitehnoloogia*, Tallinna Tehnikaülikool.
2011–2013	**Magistriõpe**, *Telekommunikatsioon*, Tallinna Tehnikaülikool, *cum laude.*
2008–2011	**Bakalaureuseõpe**, *Telekommunikatsioon*, Tallinna Tehnikaülikool, *cum laude*.
2007–2008	**Ajateenistus**, Kuperjanov Üksikjalaväepataljon.
2004–2007	**Keskkool**, Nõo Reaalgümnaasium.

Töökäik

02.2018– ...	**Andmeteadlane**, Taxify.
01.2017– ...	**Nooremteadur**, Tallinna Tehnikaülikool, Thomas Johann Seebecki elektroonikainstituut.
10.2016–02.2017	**Andmeteaduse intern**, Mooncascade.
02.2016–05.2016	**Külalisteadur**, Lundi Ülikool, Matemaatilise statistika instituut.
08.2013–12.2016	**Nooremteadur**, Tallinna Tehnikaülikool, Raadio- ja sidetehnika instituut.
04.2013–08.2013	**Tehniline konsultant**, Levira.
01.2011–03.2013	**Tehnikajuht**, Levira.
09.2010–12.2010	**Väljastuskeskuse insener**, Levira.

Õpetamiskogemus

Ülikoolikursused:
- Magistriõpe kursused:
  - Sidetehnilised standardid
  - Allikakodeerimine
  - Adaptiivne signaalitöötlus
- Bakalaureuseõpe kursused:
  - Side
  - Signaalitöötlus I
  - Arvutivõrgud

Mektory Tehnoloogiakool: *Multimeedia ja telekommunikatsioon*

TTÜ täienduskoolitused: *Multimeedia and telekommunikatsioon*

TTÜ populaarteaduslikud loengud keskkoolides: *Multimeedia*

Teadusartiklid

2018   M. Butsenko, J. Swärd and A. Jakobsson, "Zoomed Iterative Adaptive Approach", has been accepted to *2018 International Symposium on Intelligent Signal Processing and Communication Systems*.

2018   A. Krivošei, M. Min, P. Annus and M. Butsenko, "Decomposition of the EBI Signal into Components using two Channel Cross-Compensating Singular Spectrum Analysis" in *13th Annual IEEE International Symposium on Medical Measurements & Applications*, Rome, Italy, 2018.

2018   M. Butsenko, O. Märtens, A. Krivošei and Y. Le Moullec, "Sparse Reconstruction Method for Separating Cardiac and Respiratory Components from Electrical Bioimpedance Measurements" in *Elektronika ir Elektrotechnika*, vol. 24, no. 5, pp. 57-61, 2018.

2018   Maksim Butsenko, Johan Swärd, and Andreas Jakobsson. Estimating Sparse Signals Using Integrated Wide-band Dictionaries in *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4170-4181, 2018.

2017   M. Butsenko, J. Swärd and A. Jakobsson, "Estimating Sparse Signals Using Integrated Wide-Band Dictionaries", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, USA, March 5-9, 2017.

2016   M. Butsenko and T. Trump, "An Affine Combination of Adaptive Filters for Channels with Different Sparsity Levels", *Telfor Journal*, Vol. 8, No. 1, pp. 32-37, July 2016.

2015   M. Butsenko and T. Trump, "An Affine Combination of Adaptive Filters for Sparse Impulse Response Identification", *Telecommunications Forum Telfor (TELFOR)*, 2015 23rd, Belgrade, 2015, pp. 396-399.

Arvutioskused

Basic	PHP, C, HADOOP
Intermediate	R, SQL
Advanced	MATLAB, PYTHON

Keeled

Vene	**emakeel**
Eesti	**kõrgtase**
Inglise	**kõrgtase**
Prantsuse	**algtase**
Rootsi	**algtase**

Hobid

- Muusika
- Matkamine
- Jalgpall
- Haridus
- Pruulimine
- Sulgpall