Imperial College of Science, Technology and Medicine
Department of Electrical and Electronic Engineering

A total variation approach to sampling and sparse reconstruction from Fourier measurements

Maxime Ferreira Da Costa

Supervised by Dr.Wei Dai

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London and
the Diploma of Imperial College London, August 2018.

Abstract

Inferring the fine scale properties of a signal from its coarse measurements is a common signal processing problem that finds a myriad of applications in various areas of experimental sciences. Line spectral estimation is probably one of the most iconic instances of this category of problems and consists of recovering the locations of highly localized patterns, or spikes, in the spectrum of a time signal by observing a finite number of its uniform samples. Recent advances have shown that convex programming could be used to estimate the frequency components of a spectrally sparse signal. This thesis focuses on the total variation (TV) approach to perform this reconstruction.

It is conjectured that a phase transition on the success of the total variation regularization occurs whenever the distance between the spectral components of the signal to estimate crosses a critical threshold. We prove the necessity part of this conjecture by demonstrating that TV regularization can fail bellow this limit. In addition, we enrich the sufficiency side by proposing a novel construction of a dual certificate built on top of a so-called diagonalizing basis which can guarantee a prefect reconstruction of the spectrum up to near optimal regimes.

Moreover, we study the computational cost of the TV regularization, which remains the major bottleneck to its application to practical systems. A low-dimensional semidefinite program is formulated and its equivalence with the TV approach is ensured under the existence of a certain trigonometric certificate verifying the sparse Fejér-Riesz condition, leading to potential order of magnitude changes in the computational complexity of the algorithm.

This low dimensional algorithm is then applied in the context of multirate sampling systems in order to jointly estimate sparse spectra at the output of several samplers. We demonstrate the sub-Nyquist capabilities and the high computational efficiency of such systems.

Statement of originality

I declare that this thesis is the result of my own work.

Information and ideas derived from the work of others has been acknowledged in the text and a list of references is given in the bibliography.

The material of this thesis has not been and will not be submitted for another degree at any other university or institution.

Copyright declaration

Acknowledgements

This journey, as a PhD student, was filled with excitements, disappointments, hope, frustrations, successes and, most of all, filled with doubts and unanswered questions. I would hereby like to express my gratitude to all the people who made those past years a very special and unique experience in a lifetime.

First of all, I would to thank my supervisor, Dr. Wei Dai, for is content trust, and for the freedom he gave me on my research topics while always providing valuable advices. I am grateful to Dr. Moez Draief for advising the first months of my PhD program and for allowing me to visit Huawei Technologies research center in Paris in summer 2015. I would also like to acknowledge Prof. Antony G. Constantinides who advised my master's thesis and passed on his passion for research to me.

My research has been stimulated by several discussions I had with researchers of great talent. I would like to thank my colleagues at the Alain Turing Institute for allowing me to take part of their group seminar, and in particular Dr. Armin Eftekhari for his help and technical explanations. I am also grateful to Dr. Gongguo Tang and his student Qiuwei Li for the very stimulating and challenging conversations we had, and for welcoming me in their research group at Colorado School of Mines in autumn 2017.

I would also like to acknowledge all my colleagues and friends in the Communication and Signal Processing group who made my time at Imperial College unforgettable. Among all the brilliant people of our research group, I would particularly like to thank Thibaud Gabillard for his perpetual kindness and our discussions on orbital mechanics, Antonio Campello for our unrelated mathematical conversations and the whiskeys, Enrico Piovano for his unbounded entropy, and Hamdi Joudeh for his wise advices.

I am also indebted to my parents, my family, and my friends in France and all around the world for their great help in both good and bad times.

My last thoughts are for my wife, Jinwen, for her countless love, her unbreakable support, and her encouragements to pursue my dreams.

Maxime Ferreira Da Costa

London, August 2018.

In memory of my mother, Nelly Le Rolland.

Contents

Mathematical notations

Integers

$\llbracket s \rrbracket$	the integer sequence $[1, \cdots, s]$
$\llbracket a, b \rrbracket$	the integer sequence $[a, a+1, \cdots, b]$
$\gcd(\cdot, \cdot)$	the greatest common divisor of two integers

Number sets

$\mathbb{T}$	the unidimensional torus $\mathbb{T} \simeq \mathbb{R}/\mathbb{Z} \simeq [0, 1)$
$\mathbb{U}$	the set of unitary complex number $\mathbb{U} = \{z \in \mathbb{C} : \ \lvert z \rvert = 1\}$
$\mathcal{B}_{\mathbb{T}}(x, \varepsilon)$	the open ball over the torus of center $x \in \mathbb{T}$ and radius $\varepsilon \in \left[0, \frac{1}{2}\right)$

Function and measure spaces

$L_1(\mathbb{R}, \mathbb{C})$ — the space of absolutely integrable complex valued functions of the real variable

$\mathcal{C}(E, F)$ — the set of continuous maps between two spaces E and F

$\mathcal{M}(\mathbb{K})$ — the set of complex valued Radon measure over $\mathbb{K}$ ($\mathbb{K} = \mathbb{T}$ or $\mathbb{R}$)

$\operatorname{supp}(\cdot)$ — the support of a function or Radon measure defined over the domain $\mathcal{D}$

$$\forall \mu, \quad \operatorname{supp}(\mu) = \{x \in \mathcal{D} : \ \mu(x) \neq 0\}$$

$\mathcal{M}(X)$ — the set of complex valued Radon measure supported over the set X

$\hat{f}$ — the continuous time Fourier transform of a function $f \in L_1(\mathbb{R}, \mathbb{C})$

$$\forall \xi \in \mathbb{R}, \quad \hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi x\xi} \mathrm{d}x$$

$\mathcal{B}_\sigma$ — the Bernstein space of bandlimited functions for the band limit σ

$$\mathcal{B}_\sigma = \left\{ f \in \mathcal{C}(\mathbb{R}, \mathbb{C}) : \ \operatorname{supp}\left\{\hat{f}\right\} \subset [-\sigma, \sigma] \right\}$$

Linear algebra

$\bar{z}$	the complex conjugate of a complex number z
$\mathbf{X}, \mathbf{Y}, \mathbf{M}, \ldots$	a matrix
$\mathcal{L}, \mathcal{R}, \ldots$	a linear operator
$z^*, \mathbf{X}^*, \mathcal{L}^*, \ldots$	the adjunction of $z, \mathbf{X}, \mathcal{L}, \ldots$
$z^\mathsf{T}, \mathbf{X}^\mathsf{T}, \ldots$	the transposition of $z, \mathbf{X}, \ldots$
e_l	the l^{th} element of the canonical basis of $\mathbb{C}^d$ in dimension d
$\mathbf{C}_\mathcal{I}$	a selection matrix $\mathbf{C}_\mathcal{I} \in \{0,1\}^{r \times d}$ for a subset $\mathcal{I} \subseteq [\![d]\!]$ of cardinality r is the boolean matrix whose rows are equal to $\{e_l^*, l \in \mathcal{I}\}$
$\mathcal{H}_m$	the canonical Toeplitz Hermitian matrix generator in dimension m

$$\begin{aligned} \mathcal{H}_m : \mathbb{C}^{m+1} &\rightarrow \mathbb{C}^{(m+1)\times(m+1)} \\ u &\mapsto \mathcal{H}_m(u) = \begin{bmatrix} u_0 & u_1 & \dots & u_m \\ \overline{u_{-1}} & u_0 & \dots & u_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{u_m} & \overline{u_{m-1}} & \dots & u_0 \end{bmatrix} \end{aligned}$$

$\mathcal{T}_m$	the set of complex 1-periodic trigonometric polynomials of degree m

Miscellaneous

δ_x	the Kronecker symbol for the event x
$\operatorname{sign}(\cdot)$	the sign of a non-zero complex number, defined by

$$\forall z \in \mathbb{C}^*, \quad \operatorname{sign}(z) = \frac{z}{|z|}$$

Chapter 1

Introduction

1.1 Harmonic analysis and reconstruction

1.1.1 Time and frequency band limits

The continuous Fourier transform builds an isomorphism on the space of functions with finite energy. The knowledge to the temporal information of a signal is enough to completely characterize its spectrum, and reciprocally the knowledge of the spectrum determines a signal in the time domain. Furthermore, the continuous Fourier transform is a non-local operator, and temporal informations are spread in the whole spectrum and vice-versa. More importantly, the notion of time and frequency are dual to each other, and *exchangeable* within the definition of Fourier's basis

$$e^{i2\pi ft}.$$

However, this interplay becomes more subtle whenever one has only a partial knowledge of a signal. For instance, in practical scenarios, an experimenter will surely be limited to acquire time-limited observations. Moreover, her measurement device might be constrained by the law of physics to sense only a low-pass version, or a narrow-band version of the ground truth signal, or both. This is typically the case in electromagnetic systems where the transmission is limited in time and where the receiving antennas can only receive fragments that lies within the spectral bandwidth they have been conceived to work on. Those observation limits, and the non-locality properties of the Fourier transform induces blurring and distortion in the dual domain. This phenomena can be explained by the time-frequency analog of the well known Heisenberg uncertainty principle in quantum mechanics [37] stating that a temporal event of finite temporal duration Δt should have a frequency spread Δf verifying

$$\Delta t \Delta f \geq \frac{1}{4\pi}, \tag{1.1}$$

highlighting the existence of a limit between the maximal observability in one domain from limited temporal knowledge of its dual [41]. The Heisenberg's relation (1.1) also suggests that highly localized events in one domain might only be resolved by providing a nearly complete observation of the dual domain. Finally, most of the modern measurement systems are *digital*, and may only access punctual information of the signal they are monitoring, adding an extra layer of distortions and ambiguities between the sampled data and the ground truth.

Hence, measuring in one domain induces unavoidable uncertainties in the dual one, and unveiling the properties in time or frequency from measurements in its dual is surely a challenging task. This reconstruction problem is one of the core questions of *harmonic analysis*, which was qualified by Fourier in his scientific manifesto [34] as

> "[...] the ability of humankind to supplement the brevity of life and the imperfection of the senses."

Understanding the time-frequency relations and limits has become an even more important concern with the raise of bandlimited communications, where a transmitter is constrained to operate within a fixed spectral band of width Δf and aims to maximize the amount of information it can send within a finite time range Δt. Modern development of this theory had also crucial impacts in various areas of experimental and imaging sciences, such as astronomy, crystallography, medical and magnetic resonance imaging, and optics. Despite, its longevity, harmonic analysis remains, as of today, an active area of research filled with unknowns and new potential applications.

1.1.2 Structured reconstruction

If the uncertainty principle (1.1) highlights the existence of ambiguities and draws a statistical limit when trying to distinguish between any two bandlimited signals x_1 and x_2 from a time limited observation window, it doesn't necessarily implies that any kind of reconstruction is impossible. It is fairly well understood that the use of a *prior knowledge* on the structure of the signal to reconstruct can avoid any reconstruction ambiguity. The most common example consists in assuming that the continuous time signal $x(\cdot)$ is spectrally bandlimited within the frequency range $[-B, B]$, and the Nyquist-Shannon sampling theorem [51], [59] ensures that $x(\cdot)$ can be fully determined by the observation of a series of uniform temporal observations $\{y_k\}_{k\in\mathbb{Z}}$ acquired at a sampling frequency $F_s \geq 2B$ through the relation

$$\forall t \in \mathbb{R}, \quad x(t) = \sum_{k\in\mathbb{Z}} y_k \operatorname{sinc}\left(\pi\left(F_s t - k\right)\right).$$

If the above theorem still requires to acquire an unlimited number of measurements to guarantee a complete reconstruction, an other important line of work from Slepian, Pollak and Landau studied the properties of temporal analog functions maximizing the energy concentration within a fixed bandwidth [62]. This work led to the development of the discrete prolate spheroidal wave functions, which are optimal digital waveforms on N points maximizing the energy transmission within the reduced bandwidth W, and found great applications in digital communications and filter design [61]. Slepian's $2NW$-Theorem ensures that the set of discrete signals on N points guarantying zero spectral information leakage within a reduced bandwidth W tends to a linear space of dimension $2NW$ at a linear convergence rate.

Many others structured prior models have been assumed and studied, and a comprehensive tour of modern reconstruction problems from discrete samples can be found in [31], [73]. This thesis will focused on the reconstruction of signals that can be modeled in the time or frequency domain by a stream of punctual events of negligible width ($\Delta t = 0$ or $\Delta f = 0$) from observations in the dual domain. This model is often refereed as finite rate of innovation or spikes model. Many algorithms [47], [74] are known to be able to reconstruct the exact location of the events with infinite precision in absence of noise, although the model clearly fails to meet the Heisenberg's condition (1.1).

This work studies the reconstruction problem under the lens of convex optimization, and in particular under the total variation (TV) regularization paradigm. We start by discussing the *resolution limits* of this problem – the necessary minimal distance between *two events* to guarantee a robust recovery – and review recent results related to this limit. We later introduce the TV estimation, firstly proposed in [18], to tackle this problem. Novel approaches to study the achievable resolution limits of TV regularization are proposed in Chapter 2 and 3. The computational complexity of the TV regularization estimator is explored in Chapter 4, and an equivalent low-dimensional convex formulation is guaranteed provided the existence of a solution admitting a certain sparse sum-of-squares decomposition. Chapter 5 focuses on the joint spectral estimation problem from the output of synchronized samplers, and the TV regularization method is adapted to this extended problem. Finally, Chapter 6 draws a conclusion and proposes further research directions and extensions of the line spectral estimation problem.

1.2 Mathematical definitions and conventions

This section presents the mathematical conventions that will be used through this thesis.

Linear algebra Vector spaces of matrices are all endowed with the Frobenius inner product denoted $\langle \cdot, \cdot \rangle$ and defined by $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}\left(\mathbf{A}^* \mathbf{B}\right)$, where $\mathrm{tr}\left(\cdot\right)$ is the trace operator. $\left\|\cdot\right\|_\infty$ is used to denote the supremal norm in various normed spaces, while $\left\|\cdot\right\|_2$ denotes the usual Euclidean norm. The adjunction is taken respectively to the canonical inner product associated with the Euclidean space.

A vector y belonging to an Euclidean space $\mathbb{C}^d$ with explicitly an *odd* dimension $d = 2m + 1$ is indexed in $[\![-m, m]\!]$ so that $y = \left[y_{-m}, y_{-m+1}, \cdots, y_m\right]^\mathsf{T}$.

Trigonometric polynomials The space of trigonometric polynomials of order $m \in \mathbb{N}$ is denoted $\mathcal{T}_m$. Trigonometric polynomials are assumed to be $1-$periodic, so that any element $Q \in \mathcal{T}_m$ writes under the form

$$\forall \omega \in \mathbb{T}, \quad Q\left(\omega\right) = \sum_{k=-m}^{m} q_k e^{i2\pi k\omega}.$$

for some coefficients vector $q = [q_{-m}, q_{-m+1} \cdots, q_m]^\mathsf{T} \in \mathbb{C}^{2m+1}$ indexed in $[\![-m, m]\!]$.

Toeplitz matrices The adjoint $\mathcal{H}_m^*$ of the canonical Toeplitz Hermitian matrix generator in dimension m, denoted $\mathcal{H}_m$, is defined in Page xviii is characterized for every matrix $\mathbf{H} \in \mathbb{C}^{(m+1)\times(m+1)}$ by

$$\forall k \in [\![0, m]\!], \quad \mathcal{H}_m^* (\mathbf{H}) [k] = \langle \mathbf{\Theta}_k, \mathbf{H} \rangle = \operatorname{tr} (\mathbf{\Theta}_k^* \mathbf{H}),$$

whereby $\mathbf{\Theta}_k$ is the elementary Toeplitz matrix that is equal to 1 on the k^{th} upper diagonal and zero elsewhere, i.e.

$$\forall (i,j) \in [\![0, m]\!]^2, \quad \mathbf{\Theta}_k (i,j) = \begin{cases} 1 & \text{if } j - i = k \\ 0 & \text{otherwise.} \end{cases}$$

1.3 Line spectral estimation

1.3.1 Observation model and problem formulation

In its canonical formulation, the line spectral estimation problem aims to estimate the parameters of a sparse measure $\mu \in \mathcal{M}(\mathbb{T})$ modeled as a stream of Dirac pulses placed at unknown locations of the form

$$\forall \omega \in \mathbb{T}, \quad \mu(\omega) = \sum_{k=1}^{s} c_k \delta_{x_k}(\omega) \tag{1.2}$$

from its projection unto the first $2m+1$ complex trigonometric moments $y \in \mathbb{C}^{2m+1}$ taken by convention from the trigonometric order $-m$ to the trigonometric order m. Namely, the components of the observation vector y are given by

$$y_k = \left\langle e^{i2\pi k\omega}, \mu \right\rangle = \int_{\mathbb{T}} e^{-i2\pi k\omega} \mathrm{d}\mu(\omega)$$

for $|k| \leq m$. In the above, the finite subset $X = \{x_k\}_{k=1}^s \subset \mathbb{T}$ classically represents the support of the frequencies to estimate and the vector $c \in \mathbb{C}^s$ contains the associated complex amplitudes. The sparse measure μ is assumed to be unknown, meaning that both X, c, and s are unknown parameters to be estimated. Similarly, the observation vector $y = [y_{-m}, \cdots, y_m]^\mathsf{T} \in \mathbb{C}^{2m+1}$ can be expressed under the integral representation

$$y = \int_{\mathbb{T}} a_m(\omega) \, \mathrm{d}\mu(\omega), \tag{1.3}$$

whereby each *atom* $a_m(\cdot) \in \mathbb{C}^{2m+1}$ is the vector given by

$$\forall \omega \in \mathbb{T}, \quad a_m(\omega) = \left[e^{-i2\pi m\omega}, e^{-i2\pi(m-1)\omega}, \cdots, e^{i2\pi m\omega}\right]^\mathsf{T}. \tag{1.4}$$

Recovering μ from the sole knowledge of y is obviously an ill-posed problem, since the set of measures $\mu \in \mathcal{M}(\mathbb{T})$ leading to the same observation vector y under the consistency relation (1.3) forms an affine subspace of $\mathcal{M}(\mathbb{T})$ of uncountable dimension. In particular, the discrete Fourier transform $\tilde{y} \in \mathbb{C}^{2m+1}$ of the vector y can be interpreted as a sparse measure μ_{DFT} of the form

$$\forall \omega \in \mathbb{T}, \quad \mu_{\mathrm{DFT}}(\omega) = \sum_{k=-m}^{m} \tilde{y}_k \delta_{\frac{k}{2m+1}}(\omega)$$

that satisfies $y = \int_{\mathbb{T}} a_m(\omega) \, \mathrm{d}\mu_{\mathrm{DFT}}(\omega)$ and therefore is a solution of Equation (1.3). However, μ_{DFT} is unlikely to be an interesting representation on the studied reconstruction context, since its support $X_{\mathrm{DFT}} \subset \mathbb{T}$ has, in the general case, a fixed cardinality $2m+1$ no manner the number of spikes s involved in the original definition of the measure μ.

In the parsimonious reconstruction paradigm, the line spectral estimation problem aims to recover the *sparsest* measure $\mu_0 \in \mathcal{M}(\mathbb{T})$, supported on a set $X_0 \subset \mathbb{T}$ of minimal cardinality, that is consistent with the measurements y under the observation constraint described by Equation (1.3). Hence the *optimal estimator* for the line spectral estimation problem can be formulated as the output of the abstract optimization program over $\mathcal{M}(\mathbb{T})$

$$\mu_0 = \arg \min_{\mu \in \mathcal{M}(\mathbb{T})} \|\mu\|_0 \text{ subject to } y = \int_{\mathbb{T}} a(\omega) \, \mathrm{d}\mu(\omega), \tag{1.5}$$

whereby $\|\cdot\|_0$ denotes the pseudo-norm counting the potentially infinite cardinality of a complex Radon measure in $\mathcal{M}(\mathbb{T})$ given by

$$\begin{aligned} \|\cdot\|_0 : \mathcal{M}(\mathbb{T}) &\to \mathbb{R}^+ \cup \{+\infty\} \\ \mu \mapsto \|\mu\|_0 &= \begin{cases} \mathrm{card}(\mathrm{supp}(\mu)) & \text{if } \mathrm{supp}(\mu) \text{ is finite} \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{1.6}$$

One major particularity and difficulty of the line spectral estimation problem resides in the fact that one seeks for a *continuous* reconstruction of the spikes locations, which can be off the grid defined by the discrete Fourier transform of the samples. As a result, the object that one aims to recover is a sparse measure, and not a sparse vector, which draws a fundamental difference with the classic finite dimensional inverse problem framework. The abstract formulation (1.5) of the optimal estimator μ_0 has two drawbacks: The $\|\cdot\|_0$ pseudo-norm used in the cost function of Program (1.5) is not continuous, and the feasible set is, as discussed early, an affine subspace of $\mathcal{M}(\mathbb{T})$ of infinite and non-countable dimension. As a result, Program (1.5) cannot be tackled directly via the mean of classic descent algorithms and generic optimization solvers. Moreover, the abstract formulation (1.5) do not give, at this point, any insight on how to algorithmically compute the optimal estimator μ_0.

1.3.2 Application range

Although the sparsity model described in Subsection 1.3.1 seems idealistic since very few signals rigorously follow the spikes model (1.2), most of the natural signal can be assumed to be *compressible*. More precisely, a wide range of those can be represented as a sum of a few number of highly localized patterns, obtained by convolving a stream of Dirac spikes by a point spread function. When the point spread function is assumed to be known, the presented model is particularly fit to reconstruct signals falling in this category, since a simple prior equalization step is enough to transform the observation vector into a surrogate one following the sparsity model described by Equation (1.2).

On the practical side, the applications of line spectral estimation are many. Applications to the reconstruction of sparse spectra from time domain measurements includes radar signal processing [35], cognitive radio [3], and spectrography. On the other hand, sampling in frequency domain occurs in optical super-resolution [39], [48], magnetic resonance imaging, and crystallography.

1.4 The spectral resolution limit

1.4.1 The Rayleigh principle

Measurement devices, whenever they are electromagnetic or optical are limited by diffraction and aberrations resulting in the presence of distortions or blurring effects at their output. This phenomenon was principally highlighted by Rayleigh in the beginning of the past century by considering the diffraction patterns of two point sources passing though a circular aperture [55]: Fraunhofer's work on diffraction explains that the light intensity on the observer's screen can be modeled as the sum of the squares of two translated order-one Bessel functions of first kind. The nominal width of the Bessel spread function is inversely proportional to the width of the diffracting circular aperture. When the point sources are far apart, they are easily identifiable by targeting the center of each of the diffraction pattern. However, it becomes more complicated to distinguish the point sources when they get closer to each other.

The Rayleigh limit characterizes the *resolution* of a measurement device, defined by the minimum distance above which an observer is able to distinguish two objects of small size. The notion of resolution of a system is quite different form the notion of the *precision*, although often wrongly mingled, that refers to the estimation error of a measurement device. If Rayleigh's motivation were mostly empirical, recent lines of work in harmonic analysis aim to prove the existence of a statistical limit under which it is impossible to reconstruct of measure following a given observation model. In the case of the line spectral estimation problem described in Section 1.3, the corresponding diffracting point spread function can be assimilated to a Dirichlet kernel of order varying with the number of measurements, and will be defined later. The rest of this section covers the literature related to the resolution limits inherent to the line spectral estimation problem.

1.4.2 Coherence, Ingham inequalities, and minimal distance

There is a vast literature on the finite dimensional sparse inverse problem framework, which consists in recovering a sparse vector $x \in \mathbb{C}^d$ from noisy observations $y \in \mathbb{C}^r$, with $r \leq d$

acquired through a fat measurement operator $\mathbf{M} \in \mathbb{C}^{r \times d}$ of the form

$$y = \mathbf{M}x + \eta, \tag{1.7}$$

where $\eta \in \mathbb{C}^r$ is a noise vector. Generally, the robustness of the reconstruction of a sparse vector x from the knowledge of the observations y is related to the restricted isometry properties (RIP) [16], the randomness [20], or the incoherence properties [19] of the matrix $\mathbf{M}$. When letting by $(m_1, \cdots, m_n)$ the columns of the matrix $\mathbf{M}$, the coherence $\xi(\mathbf{M})$ is defined by

$$\xi(\mathbf{M}) \triangleq \max_{\substack{i,j \in [\![1,d]\!] \\ i \neq j}} \frac{\langle m_i, m_j \rangle}{\|m_i\|_2 \|m_j\|_2}.$$

It has been shown that the robustness of the sparse inversion (1.7) increases whenever the measurement operator $\mathbf{M}$ satisfies a tighter RIP property, or whenever the coherence $\xi(\mathbf{M})$ decreases.

By opposition to the "classic" and well-explored finite-dimensional inverse problem framework, one seeks, in the studied settings, to reconstruct *continuously* a measure $\mu \in \mathcal{M}(\mathbb{T})$, instead of assuming that μ is supported on prior discrete set $X \subset \mathbb{T}$ guiding the reconstruction. By a direct calculation, the coherence $\xi(\omega, \omega + \Delta\omega)$ between two atoms $a_m(\omega)$ and $a_m(\omega + \Delta\omega)$ defined in Equation (1.4) writes

$$\begin{aligned} \xi_m(\omega, \omega + \Delta\omega) &\triangleq \frac{\langle a_m(\omega), a_m(\omega + \Delta\omega) \rangle}{\|a_m(\omega)\|_2 \|a_m(\omega + \Delta\omega)\|_2} \\ &= \sum_{k=-m}^{m} e^{i2\pi k \Delta\omega} \\ &= D_m(\Delta\omega) \end{aligned}$$

whereby $D_m(\cdot)$ denotes the Dirichlet kernel of trigonometric degree m defined by

$$\begin{aligned} D_m : \mathbb{T} &\to \mathbb{R} \\ \omega &\mapsto \sum_{k=-m}^{m} e^{i2\pi k \omega} = \frac{\sin(2\pi(m+1)\omega)}{(m+1)\sin(2\pi\omega)}. \end{aligned} \tag{1.8}$$

Consequently, the coherence between two atoms tends to $D_m(0) = 1$ whenever $\Delta\omega \to 0$, and inferring on the joined presence of ω and $\omega + \Delta\omega$ in the support set X of the measure to reconstruct will become a harder and harder task [48]. As a result, the dictionary used for the inversion is always coherent, and one cannot directly rely on a notion of RIP or incoherence to study the performance of an off the grid reconstruction estimator for the line spectral estimation problem.

However, the atoms *involved* in the construction of the observations y can be incoherent if one assumes the existence of a minimal separation between them. By analogy with the finite dimensional case, one can conjecture that the reconstruction performances of the support set X are driven by its minimal warp-around distance over the torus, denoted

$\Delta_{\mathbb{T}}(X)$, and defined by

$$\forall X \subseteq \mathbb{T}, \quad \Delta_{\mathbb{T}}(X) \triangleq \inf_{\substack{x,x' \in X \\ x \neq x'}} \min_{p \in \mathbb{Z}} |x - x' + p| \,. \tag{1.9}$$

The necessity for such separation can be partially justified by the early work of Slepian [61] on discrete prolate spheroidal sequences proving that a discrete signal of length $2m+1$ cannot concentrate its energy in a bandwidth narrower than $o\left(\frac{1}{m}\right)$. Hence, a "reasonable" separation condition should be at least of $\Delta_{\mathbb{T}}(X) = \Omega\left(\frac{1}{m}\right)$ in the asymptotic regime. This first intuition can be strengthen through the lens of Ingham inequalities [40], [50]: A generalization of Perseval's theorem to non-harmonic series within a time limit of the form

$$C_1(T,\gamma)\,\|c\|_2^2 \leq \int_0^T \left| \sum_{k \in \mathbb{Z}} c_k e^{i2\pi\lambda_k t} \right|^2 \mathrm{d}t \leq C_2(T,\gamma)\,\|c\|_2^2 \,, \tag{1.10}$$

whereby $\gamma \in \mathbb{R}^+$ is the minimal successive gap between two successive harmonics $\gamma \leq \lambda_{k+1} - \lambda_k$. It is known that the lower inequality of (1.10) holds for a value of $C_1(T,\gamma)$ verifying

$$C_1(T,\gamma) = \frac{2T}{\pi}\left(1 - \frac{1}{T^2\gamma^2}\right),$$

provided that $\gamma > \frac{1}{T}$, and that the left-hand side inequality of (1.10) breaks whenever $\gamma < \frac{1}{T}$. Ingham's inequality (1.10) can be interpreted as an asymptotic for the minimal distance of the line spectral problem whenever $m \to +\infty$ by rescaling the parameters $T = 2m$, $x_k = \frac{\lambda_k}{T}$. It comes that no Riesz basis, framing the energies of the samples $y \in \mathbb{C}^{2m+1}$ and the norm of vector of the complex amplitudes $c \in \mathbb{C}^s$ can be found whenever the measure μ as a support verifying $\Delta_{\mathbb{T}}(X) < \frac{1}{2m}$ in the limit where $m \to \infty$.

A recent similar result, introduced and demonstrated in [49], states that the line spectral estimation problem is statistically intractable whenever

$$\Delta_{\mathbb{T}}(X) < \frac{1}{2m}, \tag{1.11}$$

in the limit where $m \to +\infty$, in the sense that one can always find another discrete support set $X' \subset \mathbb{T}$ that can explain the observations y within exponentially small noise levels with respect to the number of measurements m. Hence, under this critical *resolution limit*, X and X' are statistically indistinguishable in the asymptotic regime, no matter the chosen estimator. This result is explained by the presence of a phase transition on the behaviors of the extremal singular values of Vandermonde matrices with collapsing nodes around the unit circle, and is discussed in the next subsection.

1.4.3 Resolution and the stability of Vandermonde matrices

1.4.3.1 Relationship with Vandermonde matrices

Consider a system of noisy observations of the form

$$y = \int_{\mathbb{T}} a_m(\omega)\,\mathrm{d}\mu(\omega) + \eta \tag{1.12}$$

where $\eta \in \mathbb{C}^{2m+1}$ is a noise term. Denote by $X \subset \mathbb{T}$ the discrete support set of the measure $\mu \in \mathcal{M}(\mathbb{T})$ following the spikes model (1.2), and assume that X has cardinality s. The Vandermonde matrix $\mathbf{V}_m(X) \in \mathbb{C}^{(2m+1)\times s}$ associated with X is defined by

$$\forall X \subset \mathbb{T}, \quad \mathbf{V}_m(X) = \begin{bmatrix} e^{-i2\pi x_1 m} & e^{-i2\pi x_2 m} & \cdots & e^{-i2\pi x_s m} \\ e^{-i2\pi x_1 (m-1)} & e^{-i2\pi x_2 (m-1)} & \cdots & e^{-i2\pi x_s (m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{i2\pi x_1 m} & e^{i2\pi x_2 m} & \cdots & e^{i2\pi x_s m} \end{bmatrix}, \tag{1.13}$$

the columns for $\mathbf{V}_m(X)$ correspond to the atom $\{a_m(x_1), \cdots, a_m(x_s)\}$, therefore one can reduce Equation (1.12) to the bivariate problem

$$y = \mathbf{V}_m(X)\, c + \eta \tag{1.14}$$

whereby $c \in \mathbb{C}^s$ is the vector of the associated complex amplitudes of the measure μ.

1.4.3.2 Performance limits of an oracle estimator

Consider an idealistic scenario where an oracle gives the exact value of the number of spikes s and of the support set $X \subset \mathbb{T}$ containing the frequencies of the sparse measure μ to reconstruct. The amplitude vector $c \in \mathbb{C}^s$ is therefore the only parameter that remains to be estimated in order to fully characterize the measure μ. A possible and common approach consists in estimating c by a least-square estimator $\hat{c}_{\mathrm{LS}}$, which will also minimize the mean square error, provided that the noise term η has independent and with zero mean entries. The worst-case error in least-square estimation for a simple linear problem is highly related to the condition number of the observation matrix $\kappa(\mathbf{M}) \triangleq \frac{\sigma_{\max}(\mathbf{M})}{\sigma_{\min}(\mathbf{M})}$ whereby $\sigma_{\max}(\mathbf{M})$ and $\sigma_{\min}(\mathbf{M})$ respectively denotes the maximal and minimal singular values of $\mathbf{M}$. The following lemma provides a sharp bound on the error of the least square approach as a function of the signal-to-noise ratio and the spectral condition number of the observation matrix.

Lemma 1.1 (Worst-case bound for least-square estimation)**.** *Fix a full rank matrix* $\mathbf{M} \in \mathbb{C}^{r\times d}$ *with* $r \geq d$ *and consider the inverse problem*

$$y = \mathbf{M}x + \eta \tag{1.15}$$

where $\eta \in \mathbb{C}^r$ *is a perturbation. Define the signal-to noise ratio in the sample domain as* $\mathrm{SNR} = \frac{\|\mathbf{M}x\|_2}{\|\eta\|_2}$. *The least square estimator* $\hat{x}_{\mathrm{LS}}$ *of* x *given the observation* y *as in (1.15) verifies*

$$\|x - \hat{x}_{\mathrm{LS}}\|_2 \leq \frac{\kappa(\mathbf{M})\,\|x\|_2}{\mathrm{SNR}}, \tag{1.16}$$

whereby $\kappa(\mathbf{M})$ *is the spectral condition number of the matrix* $\mathbf{M}$. *Moreover, the bound (1.15) is sharp in the sense that there always exists a choice of* η *and* x *for which the* *equality case is achieved in (1.16).*

Proof. Denote by $\mathbf{M}^\dagger = (\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^*$ the left Moore-Penrose pseudo-inverse of $\mathbf{M}$, we have that

$$\begin{aligned}\hat{x}_{\text{LS}} &= \mathbf{M}^\dagger y = \mathbf{M}^\dagger \mathbf{M} x + \mathbf{M}^\dagger \eta \\ &= x + \mathbf{M}^\dagger \eta.\end{aligned}$$

Hence, it is possible to bound the quadratic error as follows

$$\begin{aligned}\|x - \hat{x}_{\text{LS}}\|_2 &= \left\|\mathbf{M}^\dagger \eta\right\|_2 \\ &\leq \frac{1}{\sigma_{\min}(\mathbf{M})}\|\eta\|_2 &(1.17)\\ &\leq \frac{\sigma_{\max}(\mathbf{M})\|x\|_2}{\|\mathbf{M}x\|_2} \times \frac{1}{\sigma_{\min}(\mathbf{M})}\|\eta\|_2 &(1.18)\\ &= \frac{\kappa(\mathbf{M})\|x\|_2}{\text{SNR}}.\end{aligned}$$

We conclude on the sharpness of the result by noticing that the equality case can be achieved in (1.17) for a choice of η in the left singular-space of $\mathbf{M}$ associated to the smallest singular value, while equality in (1.18) can also be achieved for a choice of x in the right singular-space of $\mathbf{M}$ associated to the maximal singular value. □

1.4.3.3 The condition number of Vandermonde matrices

Lemma 1.1 and the reformulation (1.13) of the line spectral estimation problem, suggests that the robustness of the observation model (1.3) is intimately linked to the behaviors of the condition number of Vandermonde matrices with collapsing nodes around the unit circle. The existence of a phase transition on the conditioning of such matrices is highlighted in [49] and sharper and extended bounds are proposed in [2] whenever the minimal separation of the support set X crosses the threshold $\Delta_{\mathbb{T}}(X) = \frac{1}{2m}$. The main bounds are recalled from [2], [49] in the following proposition.

Proposition 1.1 (Condition number of Vandermonde matrices)**.** *For any $X \subset \mathbb{T}$ and any $m \in \mathbb{N}$, if $\Delta_{\mathbb{T}}(X) > \frac{1}{2m}$ then the following inequalities hold*

$$\begin{aligned}\sigma_{\min}(\mathbf{V}_m(X)) &\geq \sqrt{2m - \Delta_{\mathbb{T}}(X)^{-1}} \\ \sigma_{\max}(\mathbf{V}_m(X)) &\leq \sqrt{2m + \Delta_{\mathbb{T}}(X)^{-1}}. &(1.19)\end{aligned}$$

Moreover, denote by $s = |X|$ the cardinality of X. If $s = \Omega(\ln m)$ then there exists a set $X \subset \mathbb{T}$ with minimal distance $\Delta_{\mathbb{T}}(X) < \frac{1-\delta}{2m}$ for some $\delta > 0$ for which the following inequalities hold

$$\begin{aligned}\sigma_{\min}(\mathbf{V}_m(X)) &\leq 2^{-\Omega(\delta s)} \\ \sigma_{\max}(\mathbf{V}_m(X)) &\geq \sqrt{2m}. &(1.20)\end{aligned}$$

The approach introduced in [49] to demonstrate the stability bounds (1.19) consists in relating the extremal singular values of $\mathbf{V}_m(X)$ with the existence of a *preconditioner*: A surrogate continuous function used to bound the Rayleigh quotient of a matrix. The Beurling-Selberg extremal functions, denoted c_T (resp. C_T) may serve as optimal preconditioner in the studied case. They correspond to the functions in the Bernstein space of bandlimited functions $\mathcal{B}_1$ realizing the best one-sided approximation from below (resp. from above) of the indicator function ξ_T of the interval $[-T,T] \subseteq \mathbb{R}$ for the L_1 distance. The Beurling-Selberg minorant can be defined as the solution of the convex program

$$\begin{aligned} c_T = &\arg\min_{f\in\mathcal{B}_1} \int_{\mathbb{R}} \left(\xi_T(u) - f(u)\right) \mathrm{d}u \\ \text{subjet to} \quad & f(t) \leq \xi_T(t), \quad \forall t \in \mathbb{R} \end{aligned}$$

while the Beurling-Selberg majorant can be view as a solution of

$$\begin{aligned} C_E = &\arg\min_{f\in\mathcal{B}_1} \int_{\mathbb{R}} \left(f(u) - \xi_T(u)\right) \mathrm{d}u \\ \text{subjet to} \quad & f(t) \geq \xi_T(t), \quad \forall t \in \mathbb{R} \end{aligned}$$

The extremal functions c_E and C_E were initially studied for their implication in number theory [45], and have well known series representations [1]. More importantly, Selberg demonstrated that c_T and C_T achieves the same minimal distance to the indicator function ξ_T in the L_1 sense, no matter the value of the time limit T, so that

$$\forall T \in \mathbb{R}, \quad \int_{\mathbb{R}} \left(\xi_T(u) - c_T(u)\right) \mathrm{d}u = \int_{\mathbb{R}} \left(C_T(u) - \xi_T(u)\right) \mathrm{d}u = 1. \tag{1.21}$$

This fundamental property (1.21) leads to a direct and elegant way to bound on the condition number of a Vandermonde matrix $\mathbf{V}_m(X)$ whenever $\Delta_{\mathbb{T}}(X) > \frac{1}{2m}$. A partial proof of Proposition 1.1 is recalled from [2], [49].

Partial proof of Proposition 1.1. Denote by $v \in \mathbb{C}^{2m+1}$ the vector with r^{th} entry $v_r = \sum_{j=1}^{s} c_k e^{i2\pi x_k r}$ for every $r \in [\![-m,m]\!]$. Let $T = m\Delta_{\mathbb{T}}(X)$, for all $c \in \mathbb{C}^s$ one has

$$\begin{aligned} \|Vc\|_2^2 &= \sum_{r=-m}^{m} \|v_r\|_2^2 = \sum_{r\in\mathbb{Z}} \xi_T\left(\Delta_{\mathbb{T}}(X)\, r\right) \|v_r\|_2^2 \\ &\leq \sum_{r\in\mathbb{Z}} C_T\left(\Delta_{\mathbb{T}}(X)\, r\right) \|v_r\|^2 \\ &= \sum_{r\in\mathbb{Z}} C_T\left(\Delta_{\mathbb{T}}(X)\, r\right) \sum_{k,k'=1}^{s} c_j \overline{c_{j'}} e^{i2\pi(x_k - x_{k'})r}. \\ &= \sum_{k,k'=1}^{s} c_k \overline{c_{k'}} \left(\sum_{r\in\mathbb{Z}} C_T\left(\Delta_{\mathbb{T}}(X)\, r\right) e^{i2\pi(x_k - x_{k'})r} \right). \end{aligned}$$

Since C_T is integrable over $\mathbb{R}$, applying the Poisson summation formula on the inner sum of the last equality yields

$$\sum_{r\in\mathbb{Z}} C_T\left(\Delta_{\mathbb{T}}\left(X\right)r\right)e^{i2\pi(x_k-x_{k'})r} = \Delta_{\mathbb{T}}\left(X\right)^{-1}\sum_{r\in\mathbb{Z}}\hat{C}_T\left(\Delta_{\mathbb{T}}\left(X\right)^{-1}\left(r-\left(x_k-x_{k'}\right)\right)\right)$$
$$= \Delta_{\mathbb{T}}\left(X\right)^{-1}\hat{C}_T\left(0\right)\delta_{k-k'},$$

since $\operatorname{supp}\left(\hat{C}_T\right) = [-1,1]$ by the assumption $C_T \in \mathcal{B}_1$, and since $\Delta_{\mathbb{T}}\left(X\right)^{-1}\left(x_k - x_{k'}\right) > 1$ whenever $k \neq k'$ by the minimal separation assumption on the support set X. Moreover, one has that

$$\forall T\in\mathbb{R},\quad \hat{C}_T\left(0\right) = \int_{-\infty}^{\infty} C_T\left(u\right)\mathrm{d}u = \int_{-\infty}^{\infty}\xi_T\left(u\right)\mathrm{d}u + 1$$
$$= 2m\Delta_{\mathbb{T}}\left(X\right) + 1.$$

It comes that

$$\forall c\in\mathbb{C}^s,\quad \left\|Vc\right\|_2^2 \leq \left(2m + \Delta_{\mathbb{T}}\left(X\right)^{-1}\right)\left\|c\right\|_2^2$$

and one concludes that $\sigma_{\max}\left(\mathbf{V}_m\left(X\right)\right) \leq \sqrt{2m + \Delta_{\mathbb{T}}\left(X\right)^{-1}}$.

A similar reasoning can be followed to derive the desired bound on $\sigma_{\min}\left(\mathbf{V}_m\left(X\right)\right)$ by introducing the Beurling-Selberg minorant c_T.

The instability bounds (1.20) are proven by a direct construction of an $X \subset \mathbb{T}$ and of a vector $c \in \mathbb{C}^s$ so that $\left\|Vc\right\|_2^2$ can be made exponentially small. □

It comes from Proposition 1.1 that if $\left\{X_m\right\}_{m\in\mathbb{N}}$ is a sequence of support with minimal distance verifying $\Delta_{\mathbb{T}}\left(X_m\right) \geq \frac{\alpha}{m}$ for $\alpha > \frac{1}{2}$, then the spectral condition number $\kappa\left(\mathbf{V}_m\left(X\right)\right)$ of the matrix $\mathbf{V}_m\left(X_m\right)$ will be bounded by

$$\kappa\left(\mathbf{V}_m\left(X\right)\right) \leq \sqrt{\frac{2\alpha+1}{2\alpha-1}}$$

which is independent of the value of m. Hence, from Lemma 1.1, the least square estimator on a support given by an oracle described in Subsection 1.4.3.1 will return an estimate $\hat{c}_{\mathrm{LS}}^{(m)}$ of the complex amplitudes verifying

$$\left\|c^{(m)} - \hat{c}_{\mathrm{LS}}^{(m)}\right\|_2 \leq \sqrt{\frac{2\alpha+1}{2\alpha-1}}\left\|c^{(m)}\right\|_2 \times \mathrm{SNR}^{-1},$$

for any sequence $\left\{c^{(m)}\right\}_{m\in\mathbb{N}}$ of complex amplitudes associated with the sequence of supports. On the other hand if $\left\{X_m\right\}_{m\in\mathbb{N}}$ is a sequence of support with poorly separated minimal distance verifying $\Delta_{\mathbb{T}}\left(X_m\right) < \frac{1-\delta}{2m}$ with $\left|X_m\right| = \Omega\left(\ln m\right)$, then by Lemma 1.1, one can have

$$\left\|c_m - \hat{c}_{\mathrm{LS}}\right\|_2 \geq 2^{\Omega(\delta s)}\left\|c_m\right\|_2\mathrm{SNR}^{-1}$$
$$\geq 2^{\Omega(\delta\ln(m))}\left\|c_m\right\|_2\mathrm{SNR}^{-1}$$

for some choices of complex amplitudes $c_m \in \mathbb{C}^s$. Therefore, even with the knowledge of the support set, the Euclidean distance between the least-square estimator and the coefficient vector can diverge when m grows large, even under exponentially small signal-to-noise ratios. One concludes on the *statistical intractability* of the line spectral estimation problem below the resolution limit $\frac{1}{2m}$ when m tends to infinity.

1.5 Estimating the line spectrum

There is a rich literature on spectral estimation. The continuous reconstruction imposed by the line spectral estimation framework requires to look for algorithms that are discretization free in order to avoid the inherent risk of basis mismatch and of discretization artifacts. A review of the main existing estimators is provided in this section. The existing techniques can mostly be classified into two mains categories. One the one hand, *subspace based methods* aim to build a matrix out of the observation vector y in such a way that the parameters of the measure $\mu \in \mathcal{M}(\mathbb{T})$ to reconstruct are related to the algebraic and spectral properties of this matrix. The spectral properties of the involved matrices are known to be robust up to certain levels of noise which varies with the separation of the measure [68], [77]. On the other hand, *optimization based methods* seek to propose convex alternatives to the optimal estimator defined by Program (1.5), either by relaxing the consistency constraints and the cost function, or by adding a data fidelity term to perform a reconstruction while denoising the signal.

1.5.1 Subspace based methods

Prony's methods [9], [53] also known as annihilating filter methods use the auto-regressive structure of the observation vector $y \in \mathbb{C}^{2m+1}$ in order to ensure the existence of a vector $h \in \mathbb{C}^{m+1}$ solution of the linear system

$$\mathcal{H}(y)\, h = y_{|0..m}. \tag{1.22}$$

The solution h can be read as an annihilating polynomial whose roots are on each of the elements of the support set X. A Cadzow's denoising step [13] can be added to ensure robustness in noisy environments. Prony's method can also be used to reconstruct sparse signals under the wider finite rate of innovation framework [26], [74].

Multiple subspace classification (MUSIC) is a popular algorithm coming in several variants. The single snapshot case [58], closer to the problem exposed in Section 1.3 aims to build a rectangular tap-delayed Hankel matrix $\mathbf{H} \in \mathbb{C}^{l \times (2m+1)}$ out of the observation vector y for a fix integer $l \in \mathbb{N}$ with $s \leq l \leq 2m+1$. The spectral properties of Hankel matrices [21] ensure the existence of a decomposition of the data matrix under the form

$$\mathbf{H} = \mathbf{V}_m(X)\, \mathbf{D} \mathbf{V}_m(X)^*, \tag{1.23}$$

whereby $\mathbf{V}_m(X)$ is the Vandermonde matrix defined in Equation (1.13). Computing this factorization leads directly to an estimate of the set X. The MUSIC algorithm is known to be robust to noise under the condition that $\Delta_{\mathbb{T}}(X) \geq \frac{1}{m}$ whenever the number of spikes is small [47].

Modified matrix pencil (MMP) is a reconstruction method proposed in [49] that relies on the algebraic properties of the solution of the generalized eigenvalue problem

$$\mathbf{T}_0 u = \lambda_k \mathbf{T}_1 u$$

whereby $(\mathbf{T}_0, \mathbf{T}_1)$ are two delayed version of the Toeplitz data structure $\mathcal{H}(y)$ used in Equation (1.22). It can be shown that the generalized eigenvalues λ_k are complex numbers whose arguments verify $\arg(\lambda_k) = -2\pi x_k$, leading again to an algorithm estimating the support set X. Robustness up to the resolution $\Delta_{\mathbb{T}}(X) \geq \frac{1}{m}$ is ensured [23]. This method can also be extended to reconstruct a stream of spikes affected by multiple known convolution kernels $\{g_l\}_{l=1}^{L}$ so that the signal to reconstruct may read

$$\forall t \in \mathbb{R}, \quad y(t) = \sum_{l=1}^{L} (g_l * \mu_l)(t).$$

1.5.2 Convex based methods

Low rank Hankel reconstruction is a method that can be seen as a convex lifting of the MUSIC algorithm, and consists in optimizing on the set of Hankel matrices with a rank constraint instead of building the matrix $\mathbf{H}$ in Equation (1.23) directly from the sampled data [14]. The rank constraint can classically be relaxed into a nuclear norm minimization program leading to

$$\mathbf{H}_\star = \arg\min \|\text{hankel}(u)\|_* \quad \text{s.t.} \quad \|u - y\| \leq \delta,$$

whereby $\|\cdot\|_*$ denotes the nuclear norm and whereby δ reflects the noise level. The parameters of the measure can be reconstructed from the output $\mathbf{H}_\star$ using the Vandermonde decomposition (1.23).

Total variation norm minimization (TV regularization) is another convex approach to the line spectral estimation introduced in [18] and will be the main focus of this thesis. A wider background on TV regularization is given in Section 1.6.

1.6 Background on the total variation regularization

1.6.1 Definitions and properties

In the recent years, a growing enthusiasm has been placed in tackling the line spectral estimation problem though the lens of *convex optimization* after the pioneer work [17], [18] demonstrated that convex programming could recover any sparse measure having a support with minimal distance verifying $\Delta_{\mathbb{T}}(X) \geq \frac{2}{m}$ in absence of noise, and for sufficiently large values of the trigonometric order m. By analogy with the successful

ℓ_1-minimization approach to solve finite dimensional linear inverse problems [15], [19], the authors' original idea consists in swapping the cardinality counting pseudo-norm in (1.5) by the total variation norm, also often referred as total mass of a measure, and denoted $|\cdot|(\mathbb{T})$. The total variation norm is formally defined as the dual norm associated to the infinite norm $\|\cdot\|_\infty$ for the weak-$*$ topology. Its rigorous definition is given as follows

$$\begin{aligned} |\cdot|(\mathbb{T}) : \mathcal{M}(\mathbb{T}) &\to \mathbb{R}^+ \\ \mu &\mapsto |\mu|(\mathbb{T}) = \sup_{\substack{f \in \mathcal{C}(\mathbb{T},\mathbb{C}) \\ \|f\|_\infty \leq 1}} \Re\left[\int_{\mathbb{T}} \overline{f(\omega)} \mathrm{d}\mu(\omega)\right]. \end{aligned} \tag{1.24}$$

Since a minimal separation of the form $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha}{m}$ of some $\alpha > 0$ will be assumed in this thesis on the measures $\mu \in \mathcal{M}(\mathbb{T})$ to reconstruct, the supremum in Expression (1.24) will always be reached for a function of $\mathcal{C}(\mathbb{T}, \mathbb{C})$ of bounded variation, and the definition of the total variation norm may always be reduced to

$$\forall \mu \in \mathcal{M}(\mathbb{T}), \quad |\mu|(\mathbb{T}) = \int_{\mathbb{T}} \mathrm{d}|\mu|, \tag{1.25}$$

which is lighter and easier to deal with. In particular, if $\mu \in \mathcal{M}(\mathbb{T})$ is a s-sparse measure of the form (1.2) for some complex vector $c \in \mathbb{C}^s$ and $X = \{x_1, \ldots, x_s\} \subset \mathbb{T}$, one has

$$|\mu|(\mathbb{T}) = \|c\|_1.$$

Consequently, it is easy to interpret the total variation norm as an extension of the finite dimensional ℓ_1-norm to the set of Radon measures. Moreover, it is interesting to notice that the total variation ball $\mathcal{B}_{\mathrm{TV}}(\mathbb{T})$, defined by

$$\begin{aligned} \mathcal{B}_{\mathrm{TV}}(\mathbb{T}) &= \{\mu \in \mathcal{M}(\mathbb{T}) : \ |\mu|(\mathbb{T}) = 1\}, \\ &= \mathrm{conv}\left(\{\delta_\omega : \ \omega \in \mathbb{T}\}\right) \end{aligned} \tag{1.26}$$

is the *smallest possible convex set* containing every possible 1-sparse measure of $\mathcal{M}(\mathbb{T})$. Hence, by analogy with ℓ_1-minimization, one can conjecture that any regularization based on the total variation norm might have a high sparsity promoting power [22]. Finally, the so-called *total variation regularization* of the combinatorial Program (1.5) is simply obtained by changing the original cost function for its convex surrogate

$$\mu_{\mathrm{TV}} = \arg\min_{\mu \in \mathcal{M}(\mathbb{T})} |\mu|(\mathbb{T}) \text{ subject to } y = \int_{\mathbb{T}} a(\omega)\, \mathrm{d}\mu(\omega), \tag{1.27}$$

which leads to a well-defined *convex program* over the space of measure $\mathcal{M}(\mathbb{T})$.

The total variation approach to line spectral estimation is a canonical example of the wider theory of convex regularization of linear inverse problems defined over the set of measures. More generic aspects and extension of this theory are discussed in [5], [24], [28].

1.6.2 Lagrangian duality and certifiability

1.6.2.1 Duality and polynomial representations

A classic argument in convex optimization consists in studying the associated Lagrangian problem in order to derive properties on the primal solution [12], [38]. In particular, it is shown in [22] that the Lagrangian dual of Program (1.27) writes

$$\begin{aligned} q_\star = \arg \max_{q \in \mathbb{C}^{2m+1}} & \Re(\langle q, y \rangle) \\ \text{subject to} \quad & |\langle a_m(\omega), q \rangle| \leq 1, \quad \forall \omega \in \mathbb{T}, \end{aligned} \tag{1.28}$$

which is a semi-infinite program [60]: An optimization program involving variable from finite dimensional spaces over a set defined by infinitely many constraints. It is worth noticing the constraints of Program (1.28) can be reduced to

$$\forall \omega \in \mathbb{T}, \quad \langle a_m(\omega), q \rangle = \sum_{k=-m}^{m} q_k e^{i2\pi k(-\omega)} = Q(-\omega),$$

whereby $Q \in \mathcal{T}_m$ is the trigonometric polynomial with coefficient vector $q \in \mathbb{C}^{2m+1}$. Hence, one might rewrite the semi-infinite program as

$$\begin{aligned} q_\star = \arg \max_{q \in \mathbb{C}^{2m+1}} & \Re(\langle q, y \rangle) \\ \text{subject to} \quad & \|Q\|_\infty \leq 1. \end{aligned} \tag{1.29}$$

1.6.2.2 Dual certifiability

One of the principal question arising from the convex reformulation of the problem is to understand *whether the TV regularization (1.27) is tight or not*. In other words, one need to establish under which conditions it is possible to guarantee that the output μ_{TV} of Program (1.27) is equal to the output μ_0 of the optimal estimator (1.5). There is an extensive literature on the certifiability of the TV regularization approach for a broad class of sparse linear inverse problems over the set Radon measures [4], [5], [28]. In particular, the success of an instance of TV regularization is known to be conditioned by the existence of a so called *dual certificate*: A continuous function representing the values of the optimal dual Lagrange variables of Program (1.27) and satisfying some extremal interpolation properties. The existence of such certificate is enough to guarantee both the *tightness* of the relaxation and the *uniqueness* of the output of the convex program. In our setting of interest, the specific geometry of the dual constraint of Program (1.29) enforces the dual certificate to be an element of $\mathcal{T}_m$. The following theorem, recalled from [18], provides the complete statement of the dual certifiability conditions. We will make an extensive use of this fundamental result in the rest of this thesis.

Theorem 1.1 (Dual certifiability)**.** *The output μ_{TV} of the convex optimization program (1.27) is equal to the ground truth measure $\mu_\star = \sum_{k=1}^{s} c_k \delta_{x_k}$ if there exists a complex* *trigonometric polynomial $Q \in \mathcal{T}_m$ satisfying*

$$\begin{cases} Q(x_k) = \operatorname{sign}(c_k), & \forall k \in [\![s]\!] \\ |Q(\omega)| < 1, & \forall \omega \notin X. \end{cases} \tag{1.30}$$

Conversely, if $\mu_{\mathrm{TV}} = \mu_\star$ then there exists a complex trigonometric polynomial $Q \in \mathcal{T}_m$ verifying

$$\begin{cases} Q(x_k) = \operatorname{sign}(c_k), & \forall k \in [\![s]\!] \\ |Q(\omega)| \leq 1, & \forall \omega \in \mathbb{T}. \end{cases} \tag{1.31}$$

Therefore, the previous theorem links to reconstruction problem under the TV norm constraint to an interpolation problem for the infinite norm on its dual space. Figure 1.1 provides a visual representation of Conditions (1.30) and (1.31). It is particularly interesting to study this theorem under the light of the early work of Landau in sampling theory on the critical density of balayage and interpolation by entire functions [42]. This work highlights similar relations linking the existence of balayages for the p-Schatten norm to the feasibility of an interpolation problem for the dual norm.

1.6.2.3 The spectral resolution conjecture

Although Theorem 1.1 reduces the reconstruction guarantees to an apparently simple interpolation problem over the set of trigonometric polynomials, it is highly non-trivial, and still an open research topic, to understand the exact conditions on the support $X \subset \mathbb{T}$ and the coefficients $c \in \mathbb{C}^s$ of a measure to guarantee the existence of an element $Q \in \mathcal{T}_m$ verifying the condition (1.30). Chapters 2 and 3 will seek to give necessary and sufficient conditions for the existence of such trigonometric polynomials.

Experimental results provided in [17], [32] suggest the existence of a phase transition on the capability of Program (1.27) to recover a sparse $\mu \in \mathcal{M}(\mathbb{T})$ in terms of the minimal separation of the elements of its discrete support X. The expected phase transition is formulated in the following conjecture.

Conjecture 1.1 (Spectral resolution limits via TV regularization)**.** *Let $X \subseteq \mathbb{T}$ be a subset of the torus. Denote by $y(\mu) \in \mathbb{C}^{2m+1}$ the vector of the first trigonometric moments of μ up to the order m obtained through the relation (1.3).*

The TV regularized program (1.27) with input $y(\mu_\star)$ has for unique output $\mu_\star$ for every measure $\mu_\star \in \mathcal{M}(X)$ if and only if

$$\Delta_{\mathbb{T}}(X) \geq \frac{1}{m} + o\left(\frac{1}{m}\right) \tag{1.32}$$

in the asymptotic regime where $m \to +\infty$.

1.6.3 The Fejér-Riesz theorem and semidefinite representations

Although the associated dual Program (1.28) is a convex program, its semi-infinite nature does not immediately guarantee that its output μ_{TV} can be computed in a finite time

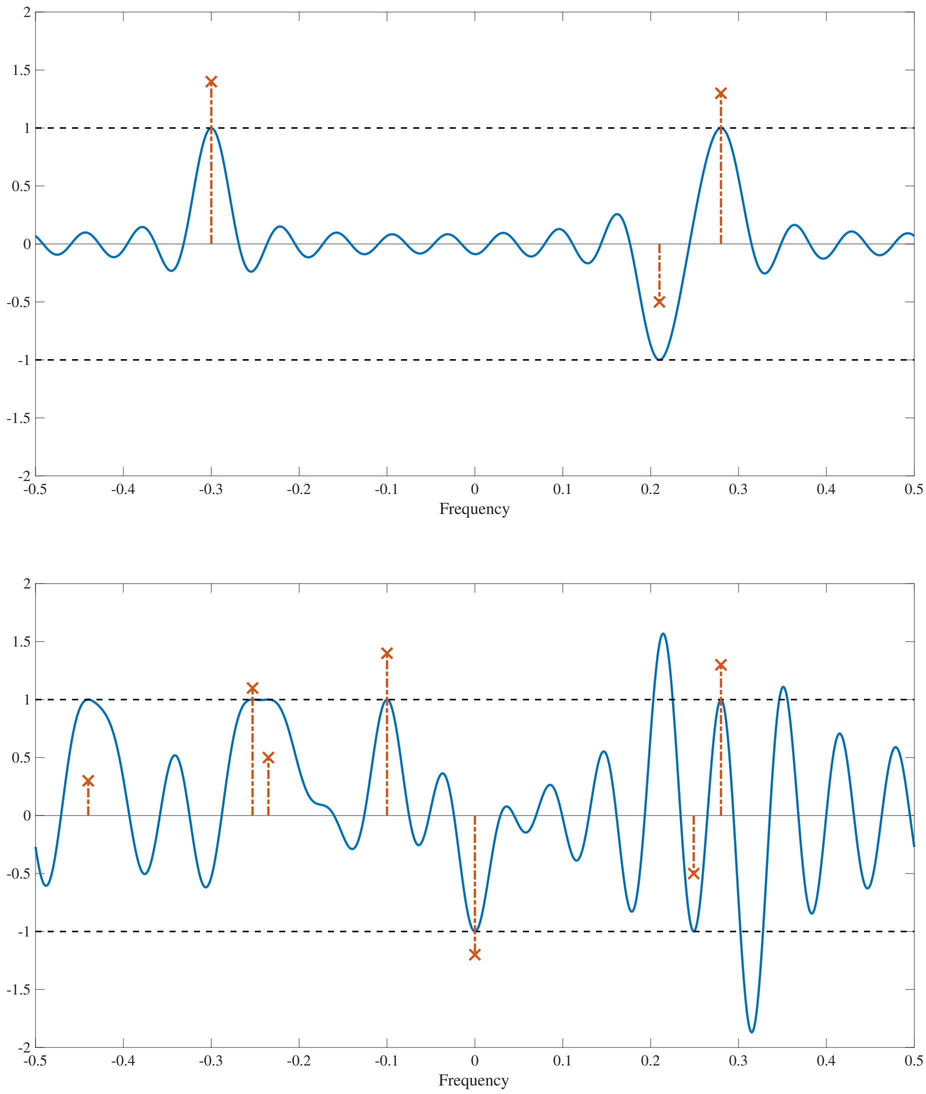

Figure 1.1: A representation of the dual certificate conditions.
Top: The candidate trigonometric polynomial (blue) interpolates the sign pattern of the the sparse measure (in red) while having a modulus bounded by 1 around the torus, and hence is a dual certificate for Theorem 1.1.
Bottom: Any trigonometric polynomial interpolating the sign pattern of the sparse measure fails to meet the extremal condition $|Q(\omega)| \leq 1$ for every $\omega \in \mathbb{T}$, hence TV regularization will fail in this setting.

by an algorithm. If gridded approximation of the dual TV problem have been studied in a wider framework to approach the optimal [10], [30], [66], they cannot handle the full complexity of the constraint, leading to provably unfeasible solutions.

Instead, one would like to determine whether the infinitely many dual constraints, compacted under the relation $\|Q\|_\infty \leq 1$, can be verified by a finite number of inequations. To this aim, one can make use of the *Gram parametrization theory* of trigonometric polynomials [27], [44], which aims to build relationships between trigonometric inequalities and the positivity of certain associated Hermitian matrices. This theory finds many useful applications in signal and data processing, including the design of finite and infinite response filters, and the sum-of-squares relaxation via the use of Lasserre hierarchies [43]. Most of the results within this line of work rely on a very fundamental factorization theorem from Fejér and Riesz [78], recalled in the following.

Theorem 1.2 (Fejér-Riesz theorem). *A trigonometric polynomial $R \in \mathcal{T}_m$ takes real and positive values on the torus, i.e.*

$$\forall \omega \in \mathbb{T}, \quad R(\omega) = \sum_{k=-m}^{m} r_k e^{i2\pi k\omega} \geq 0$$

for some coefficients vector $r \in \mathbb{C}^{2m+1}$, if and only if there exists a half-degree trigonometric polynomial $\tilde{P}$ of the form $\tilde{P}(\omega) = \sum_{k=0}^{m} \tilde{p}_k e^{i2\pi k\omega}$ such that

$$\forall \omega \in \mathbb{T}, \quad R(\omega) = \left|\tilde{P}(\omega)\right|^2.$$

We continue by introducing a brief review of the Gram parametrization theory of trigonometric polynomials. It is easy to verify that a trigonometric polynomial $R \in \mathcal{T}$ takes real values around the unit circle for all $\omega \in \mathbb{T}$, if and only if its coefficients vector $r \in \mathbb{C}^{2m+1}$ satisfies the Hermitian symmetry condition

$$\forall k \in [\![0, m]\!], \quad r_{-k} = \overline{r_k}. \tag{1.33}$$

Let by $a_{m,+}(\cdot) \in \mathbb{C}^{m+1}$ the *positive atomic vector* defined by $a_{m,+}(\omega) = \left[1, e^{i2\pi\omega}, \cdots, e^{i2\pi m\omega}\right]$ for all $\omega \in \mathbb{T}$. Every element $R \in \mathcal{T}_m$ can then be associated with a subset of matrices of $\mathbb{C}^{(m+1)\times(m+1)}$, called *Gram set* of R, as defined below.

Definition 1.1 (Gram set). A complex matrix $\mathbf{G} \in \mathbb{C}^{(m+1)\times(m+1)}$ is a *Gram matrix* associated with the trigonometric polynomial $R \in \mathcal{T}_m$ if and only if

$$\forall \omega \in \mathbb{T}, \quad R(\omega) = a_{m,+}(\omega)^* \mathbf{G} a_{m,+}(\omega).$$

Such parametrization is, in general, not unique and we denote by $\mathcal{G}(R)$ the set of matrices satisfying the above relation. $\mathcal{G}(R)$ is called *Gram set* of R.

The next proposition recalled from [27] characterizes the Gram set of a complex trigonometric polynomial taking real values on the unit circle via a simple linear relation.

Proposition 1.2 (Gram's condition)**.** *Let $R \in \mathcal{T}_m$ be a complex trigonometric polynomial taking real values around the unit circle, and let $\mathbf{G} \in \mathbb{C}^{(m+1)\times(m+1)}$. $\mathbf{G} \in \mathcal{G}(R)$ if and only if the relation*

$$\mathcal{H}_m^*(\mathbf{G}) = r_+$$

holds, whereby $r_+ = [r_0, \dots, r_m]^\mathsf{T} \in \mathbb{C}^{m+1}$ is the vector containing the coefficients of R corresponding to its positive exponents.

Finally, the previous considerations allow to show the equivalence between the semi-infinite dual program (1.28) with a finite dimensional convex program involving the maximization of the linear cost function onto a *spectrahedra*: A convex set defined by linear matrix inequalities. This category of convex program, extending linear programs (LP), quadratic programs (QP), and second order cone programing (SOCP) are referred as *semidefinite program* (SDP).

Proposition 1.3 (Semidefinite equivalence)**.** *The dual problem (1.28) is equivalent to the semidefinite program*

$$\begin{aligned} (q_\star, \mathbf{H}_\star) = \arg \max_{q \in \mathbb{C}^{2m+1}} & \Re(\langle q, y \rangle) \\ \textit{subject to} \quad & \begin{bmatrix} \mathbf{H} & q \\ q^* & 1 \end{bmatrix} \succeq 0 \\ & \mathcal{H}_m^*(\mathbf{H}) = e_0. \end{aligned} \tag{1.34}$$

Proof. Since the cost function is unchanged, it is enough to show that the constraint $\|Q\|_\infty \leq 1$ of Program (1.29) is equivalent to the semidefinite constraints of Program (1.34). This can be done by firstly considering the equivalence

$$\|Q\|_\infty \leq 1 \iff \forall \omega \in \mathbb{T}, \quad 1 - |Q(\omega)|^2 \geq 0. \tag{1.35}$$

Since $1 - |Q|^2 \in \mathcal{T}_{2m}$ is positive, the Fejér-Riesz theorem ensures the existence of a trigonometric polynomial $P \in \mathcal{T}_m$ such that

$$\begin{aligned} |P(\omega)|^2 &= \left| e^{-i2\pi m\omega} \tilde{P}(\omega) \right|^2 \\ &= \left| \tilde{P}(\omega) \right|^2 \\ &= 1 - |Q(\omega)|^2, \end{aligned}$$

whereby $\tilde{P}(\omega) = e^{i2\pi m\omega} P(\omega)$ is the half-degree trigonometric polynomial of degree $2m$ obtained by translating the coefficients of P. Letting by $p = [p_{-m}, \cdots, p_m] \in \mathbb{C}^{2m+1}$ the coefficients vector of P, and realizing that $P(-\omega) = \langle a_m(\omega), p \rangle$ for all $\omega \in \mathbb{T}$, one can rewrite the inequalities in (1.35) as

$$\begin{aligned} \|Q\|_\infty \leq 1 &\iff \exists p \in \mathbb{C}^{2m+1}, \forall \omega \in \mathbb{T}, \quad 1 = \langle a_m(\omega), p \rangle^2 + \langle a_m(\omega), q \rangle^2 \\ &\iff \exists p \in \mathbb{C}^{2m+1}, \forall \omega \in \mathbb{T}, \quad 1 = a_m(\omega)^* (pp^* + qq^*)\, a_m(\omega). \end{aligned}$$

Denoting by $\mathbf{H} \in \mathbb{C}^{(2m+1)\times(2m+1)}$ the rank 2 matrix $\mathbf{H} = pp^* + qq^*$, we conclude on the existence of a matrix in $\mathbb{C}^{(2m+1)\times(2m+1)}$ verifying $\mathbf{H} - qq^* \succeq 0$. Furthermore, a simple Schur complement argument yields

$$\begin{bmatrix} \mathbf{H} & q \\ q^* & 1 \end{bmatrix} \succeq 0.$$

Finally, since $\mathbf{H}$ is a Gram matrix of the trigonometric polynomial reaching a constant value equal to 1 everywhere on the torus, Proposition 1.2 ensures that the linear constraint $\mathcal{H}_m^*(\mathbf{H}) = e_0$ holds, concluding on the statement of the proposition. □

1.6.4 Spectral estimation from the dual space

Whenever the relaxation (1.27) is tight, it is possible to relate the support $X_0 \subset \mathbb{T}$ of the output $\mu_0 \in \mathcal{M}(\mathbb{T})$ of Program (1.5) with the output of the dual trigonometric polynomial $Q_\star \in \mathcal{T}_m$ returned by Program (1.28). Indeed, a strong duality argument detailed in [18] guarantees that the modulus of the polynomial $Q_\star$ must reach a value exactly equal to 1 at the location of the spikes in X_0, whenever tightness holds between the Programs (1.5) and (1.27). Hence it is possible to recover X_0 by looking at the roots of the polynomial $R \in \mathcal{T}_{2m}$ given by,

$$\forall \omega \in \mathbb{T}, \quad R(\omega) = 1 - |Q(\omega)|^2.$$

Once the support set $X_\star$ is estimated, the associated complex amplitudes $c_0 \in \mathbb{C}^s$ can be recovered from a simple least square approach.

$$c_\star = \mathbf{V}_m (X_\star)^\dagger y,$$

whereby $\mathbf{V}_m(X_\star)$ is the Vandermonde matrix defined in (1.13). To summarize, the primal optimal μ_{TV} can be reconstructed directly from the solution of the dual problem using Algorithm 1.1. The binary operator $p \odot q$ defined for any two vectors $p, q \in \mathbb{C}^{2m+1}$used in the algorithm is defined by

$$\forall l \in [\![-2m, 2m]\!], \quad h_l = (p \odot q)[l] = \sum_{j \in \mathbb{Z}} \bar{p}_j q_{m-l+j},$$

so that the product $H = P \times Q \in \mathcal{T}_{2m}$ of two polynomials $P, Q \in \mathcal{T}_m$ as for coefficients vector $h \in \mathbb{C}^{4m+1}$ satisfying $h = p \odot q$.

Algorithm 1.1 Frequency recovery from the dual space

$q_\star :=$ Solution of the dual SDP (1.34)
$r := e_0 - q \odot q$
$X_\star :=$ Roots of R of modulus 1
$c_\star := \mathbf{V}_m (X_\star)^\dagger y$
$\mu_{\mathrm{TV}} := \sum_{k=1}^{s} c_k \delta_{x_k}$

1.6.5 Spectral estimation in noise

Up to here, only the case of noise-free spectral estimation has been considered. However, in most of the practical applications, the observations $y \in \mathbb{C}^{2m+1}$ are noisy, and the noise vector $\eta \in \mathbb{C}^{2m+1}$ is often assumed to additive and i.i.d. The noisy counterpart of the observation model (1.3) writes

$$y = \int_{\mathbb{T}} a_m(\omega)\, \mathrm{d}\mu(\omega) + \eta.$$

The Beurling-LASSO estimator [24], also known as atomic soft thresholding (AST) algorithm was introduced in [64] to denoise the spectrum of the observation vector y while promoting a sparse prior structure. The primal AST estimator is defined as follows

$$\mu_{\mathrm{TV}}(\tau) = \arg\min_{\mu \in \mathcal{M}(\mathbb{T})} |\mu|(\mathbb{T}) + \frac{1}{2\tau} \left\| y - \int_{\mathbb{T}} a_m(\omega)\, \mathrm{d}\mu(\omega) \right\|_2^2, \tag{1.36}$$

where $\tau \in \mathbb{R}^+$ acts as a regularization parameter, drawing a trade-off between the sparsity of the solution ($\tau \to 0$) and the fidelity of the solution to the observation ($\tau \to +\infty$). Moreover, a direct analysis ensures that the corresponding dual Beurling-LASSO (or dual AST) program can also be formulated as a semidefinite program of the form

$$\begin{aligned} (q_\star(\tau), \mathbf{H}_\star) = \arg & \max_{q \in \mathbb{C}^{2m+1}} \Re(\langle q, y \rangle) - \frac{\tau}{2} \|q\|_2^2 \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{H} & q \\ q^* & 1 \end{bmatrix} \succeq 0 \\ & \mathcal{H}_m^*(\mathbf{H}) = e_0. \end{aligned} \tag{1.37}$$

The output of the primal and dual programs are still linked by the locations of the roots of the optimal dual polynomial $1 - |Q_\star(\tau)|^2$ in a similar manner than in Section 1.6.4.

Whenever the noise vector $\eta \in \mathbb{C}^{2m+1}$ is assumed to be drawn according to the spherical $2m+1$ dimensional complex Gaussian distribution $\mathcal{N}(0, \sigma^2 I_{2m+1})$, it is shown in [6] that a choice of regularization parameter $\tau = \gamma\sigma\sqrt{m \log m}$, for some $\gamma > 1$, is suitable to guarantee a consistent recovery of the spectral distribution μ_0 in the asymptotic regime where $m \to +\infty$ and under fixed signal-to-noise ratio, while providing accelerated rates of convergence. More recent results, studying the robustness of the TV regularization approach (1.37) in the non-asymptotic regime were given in [46]. In particular, a tradeoff between the resolution $\Delta_{\mathbb{T}}(X)$ and the robustness of the convex estimator to Gaussian noise was drawn. Under a minimal separation constraint on the ground truth support set, it is certified that

- The ℓ_2 distance between the ground truth signal $y_0 = \int_{\mathbb{T}} a_m(\omega)\, \mathrm{d}\mu_0(\omega)$ and the estimate $y_\star$ is a bounded function of the SNR and the dynamic range of the complex amplitudes $c_0 \in \mathbb{C}^s$.

- The distance between any spike of the ground truth support set X_0 and its estimation

$X_\star$ are bounded by the same parameters.

- There is no spurious estimated spike.

1.6.6 Relationship with the atomic norm

1.6.6.1 Atomic norm minimization

There is a vast literature on the reconstruction of a *non-negative* combination of a few number of elements drawn from a continuous dictionary [7], [22], [24], [25], [57]. One assumes to dispose of observations $y \in E$ belonging to some normed vectorial space E of the form

$$y = \sum_{k=1}^{s} c_k \mathbf{a}_k, \quad \mathbf{a}_k \in \mathcal{A},\ c_k \geq 0 \tag{1.38}$$

where $\mathcal{A} \subset E$ is called the *atomic set*, and acts as a dictionary of potentially infinite cardinality for guiding the reconstruction of the atoms $\{\mathbf{a}_k\}_{k=1}^{s}$ building the observation vector y. One of the most popular instance of this settings is the *compressive sensing* problem, whereby $\mathcal{A}$ is assumed to be the set of $2n$ signed canonical vectors $\{e_1, -e_1, e_2, -e_2, \cdots, -e_n\}$ in of $\mathbb{R}^n$ and $y \in \mathbb{C}^r$ will be a $s-$sparse vector of $\mathbb{C}^r$. Low-rank matrix reconstruction is another well studied example that can be modeled by choosing $\mathcal{A} \in \mathbb{C}^{n\times n}$ to be the set of matrices with rank one.

Denote by $\mathrm{conv}(\mathcal{A})$ the convex hull of the atomic set $\mathcal{A}$, if $\mathcal{A} \subset E$ is symmetric and homogeneous, i.e. if the condition

$$\forall \mathbf{a} \in \mathcal{A}, \forall \theta : |\theta| = 1, \quad \theta \mathbf{a} \in \mathcal{A} \tag{1.39}$$

is verified, the gauge function or Minkowski functional $\mathcal{G}_\mathcal{A}$ of $\mathcal{A}$ defined by

$$\begin{aligned} \mathcal{G}_\mathcal{A} : E &\to \mathbb{R}^+ \\ x &\mapsto \inf\{t > 0 : x \in t\mathrm{conv}(\mathcal{A})\} \end{aligned}$$

is a norm on the vectorial space E. This norm is referred as the *atomic norm* of the set $\mathcal{A}$ over E, and is often denoted $\|\cdot\|_\mathcal{A}$. This associated dual norm $\|\cdot\|_\mathcal{A}^*$ is defined by

$$\begin{aligned} \|\cdot\|_\mathcal{A}^* : E &\to \mathbb{R}^+ \\ q &\mapsto \sup_{\|x\|_\mathcal{A} \leq 1} |\langle q, x\rangle| . \end{aligned} \tag{1.40}$$

It is interesting to notice that the atomic ball $\mathcal{B}_\mathcal{A}$ defined by

$$\begin{aligned} \mathcal{B}_\mathcal{A} &= \{y \in E :\ \|y\|_\mathcal{A} = 1\} \\ &= \{y \in E :\ y \in \mathrm{conv}(\mathcal{A})\} \\ &= \mathrm{conv}(\mathcal{A}) \end{aligned} \tag{1.41}$$

is by construction the smallest convex set that contains all the atomic elements, and shares the same geometric properties than the TV ball given in Equation (1.26). The *atomic* *norm minimization* (ANM) program was proposed in [22] as a convex approach to solve a wide range of linear inverse problems of the form

$$\text{find } y \text{ such that } \tilde{y} = y + \eta \text{ and } y \sim (1.38)$$

where $\eta \in E$ if a noise term. The atomic norm minimization consists in computing

$$y_{\mathcal{A}} = \arg\min_{x \in E} \|y\|_{\mathcal{A}} + \frac{1}{2\tau} \|\tilde{y} - y\|_2^2, \tag{1.42}$$

whereby $\tau \in \mathbb{R}^+$ is a parameter to choose adequately. However, one of the major downside of the program (1.42) is that it requires to numerically evaluate to atomic norm of an arbitrary vector $y \in \mathbb{C}^n$, which can be numerically costly and unstable under classic numerical precision depending on the complexity of the atomic set $\mathcal{A}$ to denoise.

1.6.6.2 Equivalence between atomic norm minimization and TV regularization

In the case of line spectral estimation, one seeks to reconstruct a sparse measure $\mu \in \mathcal{M}(\mathbb{T})$ following Model (1.2). The observation vector $y \in \mathbb{C}^{2m+1}$ can be reformulated

$$\begin{aligned} y &= \int_{\mathbb{T}} a_m(\omega)\, \mathrm{d}\mu(\omega) \\ &= \sum_{k=1}^{s} c_k a_m(x_k), \end{aligned} \tag{1.43}$$

for some *complex* amplitude stacked in the vector $c \in \mathbb{C}^s$, and for some support $X = \{x_k\}_{k=1}^s \subset \mathbb{T}$. Since, the atomic norm framework (1.2) requires each atom to be associated with a *positive* coefficient, one can turn expression (1.43) into a well defined atomic representation by decomposing complex amplitudes under the modulus/phase product $c_k = \tilde{c}_k e^{i2\pi\phi_k}$ whereby $\tilde{c}_k \in \mathbb{R}^+$ and $\phi_k \in \mathbb{T}$ yielding

$$\begin{aligned} y &= \sum_{k=1}^{s} \tilde{c}_k e^{i2\pi\phi_k} a_m(x_k) \\ &= \sum_{k=1}^{s} \tilde{c}_k e^{i2\pi\phi_k} a_m(x_k, 0) \\ &= \sum_{k=1}^{s} \tilde{c}_k a_m(x_k, \phi_k), \end{aligned}$$

by introducing the atomic set $\mathcal{A} = \left\{a_m(\omega, \phi) : \quad (\omega, \phi) \in \mathbb{T}^2\right\}$ parametrized by the two elements $(\omega, \phi) \in \mathbb{T}^2$ whereby each atom $a_m(\cdot, \cdot) \in \mathbb{C}^{2m+1}$ reads

$$a_m(\omega, \phi) = \left[e^{i2\pi(-m\omega+\phi)}, e^{i2\pi(-(m-1)\omega+\phi)}, \cdots, e^{i2\pi(m\omega+\phi)}\right].$$

It is straight-forward to verify that that the set $\mathcal{A}$ verifies the homogeneity condition (1.39), therefore its Gauge function induces an atomic norm over $\mathbb{C}^{2m+1}$. In particular, if a measure $\mu \in \mathcal{M}(\mathbb{T})$ follows the spike model (1.2) for $s \leq 2m$, then the atomic norm of the observation vector (1.43) reads

$$\begin{aligned}\|y\|_{\mathcal{A}} &= \left\| \sum_{k=1}^{s} \tilde{c}_k a_m(x_k, \phi_k) \right\|_{\mathcal{A}} \\ &= \sum_{k=1}^{s} \tilde{c}_k = \sum_{k=1}^{s} |c_k| \\ &= |\mu|(\mathbb{T}),\end{aligned}$$

and one concludes on the equivalence between the atomic norm minimization (1.42) and the Beurling-LASSO (1.36).

1.6.6.3 The moment curve and primal semidefinite representability

When ϕ is equal to 0, the set $\{a_m(\omega, 0) : \ \omega \in \mathbb{T}\}$ forms a one-dimensional variety of $\mathbb{C}^{2m+1}$ called *moment curve* of order m. It is well understood that the convex hull of the moment curve is a body that can be parametrized by a set of linear matrix inequalities [56]. We recall the following result from [67, Proposition II.1] that gives a semidefinite representation of the atomic norm for the set $\mathcal{A}$.

Proposition 1.4 (Semidefine representation of the atomic norm)**.** *Let $y \in \mathbb{C}^{2m+1}$, one has*

$$\|y\|_{\mathcal{A}} = \inf_{\substack{u \in \mathbb{C}^{m+1} \\ t > 0}} \left\{ \frac{1}{2m+1} \operatorname{tr}(\mathcal{H}_m(u)) + \frac{1}{2} t : \ \begin{bmatrix} \mathcal{H}_m(u) & y \\ y^* & t \end{bmatrix} \succeq 0 \right\}.$$

The above proposition and the equivalence between ANM and the Beurling-LASSO provides a SDP representation of the primal program (1.36) of the form

$$\begin{aligned} y_{\mathrm{TV}} = \arg &\min_{\substack{y \in \mathbb{C}^{2m+1} \\ u \in \mathbb{C}^{m+1}}} \frac{1}{2m+1} \operatorname{tr}(\mathcal{H}_m(u)) + \frac{1}{2} t \\ \text{subject to} \quad & \begin{bmatrix} \mathcal{H}_m(u) & y \\ y^* & t \end{bmatrix} \succeq 0 \end{aligned} \tag{1.44}$$

whereby y_{TV} and μ_{TV} are linked by the linear integral relation (1.3).

Chapter 2

A tight converse to the spectral resolution limits of TV regularization

2.1 The necessary separation for TV regularization

2.1.1 Previous results

It is shown in [29], [57] that the total variational framework discussed in Subsection 1.6.6 can succeed without the need of assuming any kind of separation between the spikes to recover for a wide range of measurement operators, provided that the measure to reconstruct is *positive-valued*. However, the picture looks different when considering the reconstruction of *complex-valued* (or *signed*) Radon measures, and TV regularization is known to fail to reconstruct complex measures if certain minimal separation criteria are not met. Necessary conditions illustrating this fact were given in [65] for a wide range of inverse problems using the compacity properties of the derivation operator over certain associated dual spaces of functions. In particular, if the dual space $\mathcal{D}$ defining the dual norm (1.40) verifies a Bernstein type inequality of the form

$$\forall q \in \mathcal{D}, \quad \left\|q'\right\|_{L_\infty} \leq C\left(\mathcal{D}\right)\left\|q\right\|_{L_\infty} \tag{2.1}$$

for a Bernstein constant $C\left(\mathcal{D}\right) > 0$, any certificate satisfying the property $\left\|q\right\|_{L_\infty} \leq 1$ will have a derivative bounded in modulus by $C\left(\mathcal{D}\right)$. One may concludes that a minimal distance $\Delta_{\mathcal{D}}$ of at least $\Delta_{\mathcal{D}} \geq \frac{2}{C(\mathcal{D})}$ is necessary to interpolate two spikes with antagonists signs, *e.g.* 1 and -1.

This chapter focuses on tightening the *necessary minimal separation* $\Delta_{\mathbb{T}}\left(X\right)$ for the success of the TV regularized Program (1.27) in the line spectral estimation framework defined in Section 1.3. In our settings, the dual space is the space of trigonometric polynomials of bounded degree $\mathcal{D} = \mathcal{T}_m$, which is a Bernstein space for the constant $\mathcal{C}\left(\mathcal{T}_m\right) = 2\pi m$. One can apply the generic result [65] and conclude that (1.27) can fail whenever $\Delta_{\mathbb{T}}\left(X\right) < \frac{1}{\pi m}$. However, this result is clearly suboptimal, since there is no trigonometric polynomial other than the null polynomial that can effectively saturate the bound (2.1) everywhere on the torus.

The necessary separation was sharpened to $\Delta_{\mathbb{T}}(X) < \frac{1}{2m}$ in [28], which constituted the best result before the introduction of Theorem 2.1. The proof technique relies on an argument from Turán on the decay rate of trigonometric polynomials around their supremal values [72], whose statement is recalled in the following.

Lemma 2.1 (Turán '46)**.** *Let $Q \in \mathcal{T}_m$ be a trigonometric polynomial of degree m whose modulus achieves its maximal value at the point $x_0 \in \mathbb{T}$. We have that*

$$\forall \varepsilon \in \left[-\frac{1}{4m}, \frac{1}{4m}\right], \quad |Q(x_0 + \varepsilon) - Q(x_0)| \leq |Q(x_0)| \left|\sin(2\pi m \varepsilon)\right|, \tag{2.2}$$

moreover equality can be achieved in (2.2) if and only if Q is a monomial of degree m, i.e., if and only if Q writes

$$\forall \omega \in \mathbb{T}, \quad Q(\omega) = A\cos(2\pi m\omega) + B\sin(2\pi m\omega)$$

for some $(A, B) \in \mathbb{C}^2$.

Lemma 2.1 guarantees that no trigonometric polynomial can interpolate two nodes with opposite signs when their minimal distance is smaller than $\frac{1}{2m}$, leading to the desired conclusion.

2.1.2 Main statement

Theorem 2.1 proposes an improvement of the previous results by showing the existence of measures having a minimal separation asymptotically close to $\frac{1}{m}$ for which the convex approach fails. This tight result validates one side of Conjecture 1.1 on the achievable spectral resolution limit through TV regularization. Moreover it constitutes a significant step towards a complete understanding of the predicted phase transition.

Theorem 2.1 (Necessary separation for TV regularization)**.** *For every real $\delta > 2$, there exists $M_\delta \in \mathbb{N}$, such that for every $m \geq M_\delta$, there exists a set $X_m = \left\{x_k^{(m)}\right\}_{k=1}^{s_m} \subset \mathbb{T}$ verifying $\Delta_{\mathbb{T}}(X_m) \geq \frac{1}{m} - \frac{\delta}{m^2}$ and a measure $\mu_m = \sum_{k=1}^{s_m} c_k^{(m)} \delta_{x_k^{(m)}}$ for some $c^{(m)} \in \mathbb{C}^{s_m}$ such that the solution of Program (1.27) is not equal to μ_m.*

The demonstration of this result is provided in Section 2.3, and is based on a construction of a specific sequence of measures $\{\mu_m\}_{m\in\mathbb{N}}$ for which we show the *non-existence* of an associated dual certificate. To reach this result, we introduce in Section 2.2 the notion of *bounded diagonalizing families* of trigonometric polynomials and highlight their relationships with the existence of dual certificates. Theorem 2.2 states that such families cannot exist if the support set is not separated enough.

2.1.3 Impact of the second order term

Figure 2.1 presents *sufficient* values of the parameter M_δ defined in Theorem 2.1 for different choices of the second order term δ. Those results are a by-product of the analysis

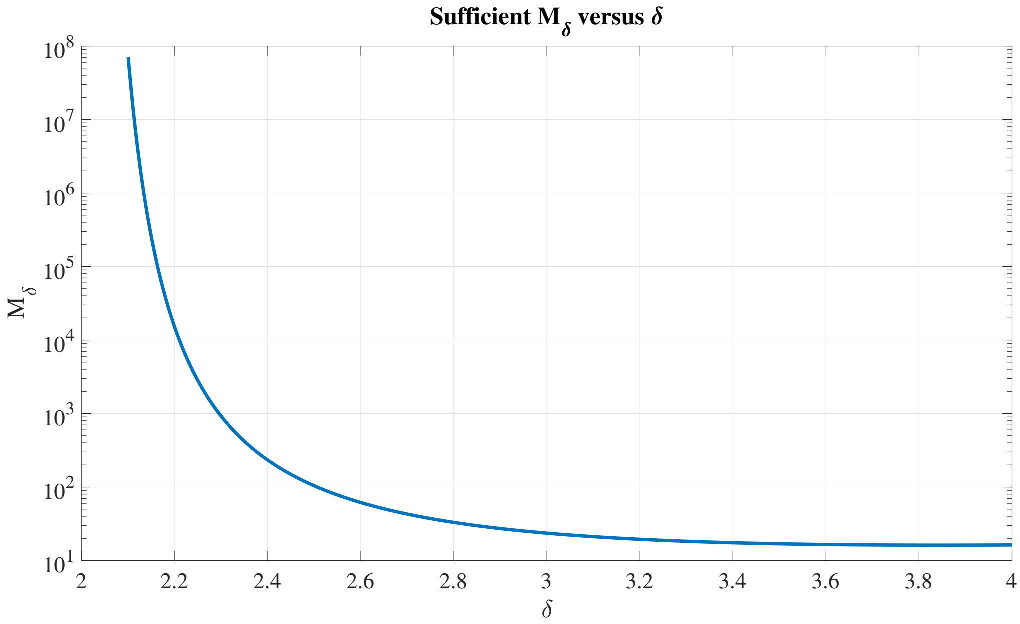

Figure 2.1: Upper bound on the minimal number of observations M_δ requested by Theorem 2.1 against the second order term δ . The curve admits a vertical asymptote of equation $\log(M_\delta) = \Theta\left((\delta - 2)^{-1}\right)$ at $\delta \to 2$.

(2.12) in the proof of Theorem 2.2, and are presented for illustration purposes. However, the present curve has a priori no reason to act as a sharp bound on the minimal achievable value of M_δ.

2.2 Diagonalizing families of trigonometric polynomials

In this section, we introduce an intermediate notion of diagonalizing families of trigonometric polynomials. A graphical representation of diagonalizing polynomials is provided in Figure 2.2. The structure and factorization properties of those polynomials are highlighted in Lemma 2.2. It will be shown in Section 2.3 that such objects play an important role on the existence of dual certificates for the line spectral estimation problem, and will be used to construct a new family of dual certificates in Chapter 3. We start by introducing the following definitions.

Definition 2.1 (Diagonalizing family)**.** Let $X = \{x_k\}_{k=1}^s$ be a finite subset of $\mathbb{T}$. A *first order diagonalizing family* of X over $\mathcal{T}_m$ is a set of s elements $\mathcal{P}_X = \{P_l\}_{l=1}^s$ of $\mathcal{T}_m$ satisfying

$$\forall (l,k) \in [\![s]\!]^2, \quad \begin{cases} P_l(x_k) = \delta_{l=k}, \\ P_l'(x_k) = 0. \end{cases} \tag{2.3}$$

Definition 2.2 (Bounded diagonalizing family)**.** A first order diagonalizing family $\mathcal{P}_X = \{P_l\}_{l=1}^s$ of X is said to be *bounded* if and only if $\|P_l\|_{L_\infty} = 1$ for all $l \in [\![s]\!]$.

Lemma 2.2 (Factorization lemma)**.** *Let $\mathcal{P} = \{P_l\}_{l=1}^s$ be a first order diagonalizing family of trigonometric polynomials of a set X of cardinality s over $\mathcal{T}_m$. If $s \leq m$, then any*

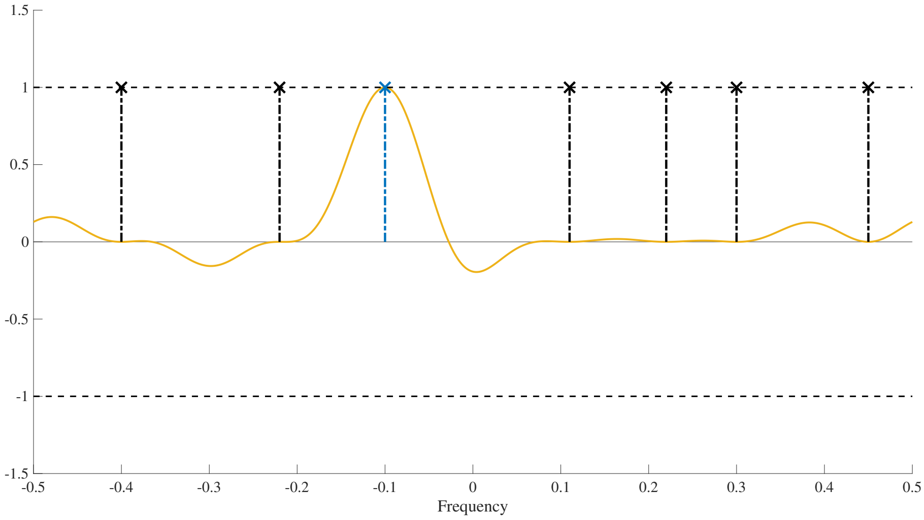

Figure 2.2: A sparse measure (in black), and one element of a diagonalizing trigonometric polynomial for this measure (yellow curve). The trigonometric polynomial has a modulus bounded by 1.

element $P_l \in \mathcal{P}$ can be factored under the form

$$\forall \omega \in \mathbb{T}, \quad P_l(\omega) = Z_{X,l}(\omega) R_l(\omega), \tag{2.4}$$

whereby $Z_{X,l} \in \mathcal{T}_{s-1}$ is the minimal annihilating polynomial on $X \setminus \{x_l\}$ defined by

$$\forall \omega \in \mathbb{T}, \quad Z_{X,l}(\omega) = \prod_{\substack{1 \le k \le s \\ k \ne l}} \frac{\sin^2(\pi(\omega - x_k))}{\sin^2(\pi(x_l - x_k))}, \tag{2.5}$$

and whereby the second factor $R_l \in \mathcal{T}_{m-s+1}$ must satisfy the interpolation conditions

$$\begin{cases} R_l(x_l) = 1 \\ R_l'(x_l) = -2\pi \sum_{\substack{1 \le k \le s \\ k \ne l}} \cot(\pi(x_l - x_k)). \end{cases} \tag{2.6}$$

Proof. Since, the l^{th} element $P_l \in \mathcal{T}_m$ of the diagonalizing family $\mathcal{P} = \{P_l\}_{l=1}^{s}$ has roots with multiplicity two at each of the locations x_k for $k \ne l$, P_l can be factorized in the complex plane under the form

$$\begin{aligned} \forall \omega \in \mathbb{T}, \quad P_l(\omega) &= \prod_{\substack{1 \le k \le s \\ k \ne l}} \left(e^{i2\pi\omega} - e^{i2\pi x_k}\right)^2 \times \bar{R}_l(\omega) \\ &= \prod_{\substack{1 \le k \le s \\ k \ne l}} 4\sin^2(\pi(\omega - x_k)) \times \bar{R}_l(\omega) \end{aligned}$$

whereby $\bar{R}_l$ is a trigonometric polynomial. The left hand side of the factorization is a product of exactly $s-1$ trigonometric polynomial of degree 1, hence $\bar{R}_l$ has a trigonometric degree that cannot exceed $m - s + 1$. One obtains the desired factorization by rescaling the factor $\bar{R}_l$ into R_l as follows

$$\forall \omega \in \mathbb{T}, \quad R_l(\omega) = 4^{s-1} \prod_{\substack{1 \leq k \leq s \\ k \neq l}} \sin^2(\pi(x_l - x_k)) \bar{R}_l(\omega).$$

It remains to prove that R_l satisfies Equation (2.6). Since P_l is a first order diagonalizing polynomial of the support set X, P_l verifies by assumptions Equation (2.3), yielding

$$\begin{aligned} 1 &= P_l(x_l) \\ &= Z_{X,l}(x_l) R_l(x_l) = R_l(x_l). \end{aligned}$$

Secondly, a direct calculation of the derivative of $Z_{X,l}$ at the point x_l leads to

$$\begin{aligned} Z'_{X,l}(x_l) &= \frac{1}{\prod_{\substack{1 \leq k \leq s \\ k \neq l}} \sin^2(\pi(x_l - x_k))} \times \\ &\sum_{\substack{1 \leq k \leq s \\ k \neq l}} \left\{ 2\pi \cos(\pi(x_l - x_k)) \sin(\pi(x_l - x_k)) \prod_{\substack{1 \leq p \leq s \\ p \neq l \\ p \neq k}} \sin^2(\pi(x_l - x_p)) \right\} \\ &= 2\pi \sum_{\substack{1 \leq k \leq s \\ k \neq l}} \frac{\cos(\pi(x_l - x_k))}{\sin(\pi(x_l - x_k))} \\ &= 2\pi \sum_{\substack{1 \leq k \leq s \\ k \neq l}} \cot(\pi(x_l - x_k)). \end{aligned}$$

Furthermore, by assumptions (2.3), the derivative of P_l must cancel at x_l, leading to

$$\begin{aligned} 0 = P'_{X,l}(x_l) &= Z_{X,l}(x_l) R'_l(x_l) + Z'_{X,l}(x_l) R_l(x_l) \\ &= R'_l(x_l) + 2\pi \sum_{\substack{1 \leq k \leq s \\ k \neq l}} \cot(\pi(x_l - x_k)), \end{aligned}$$

which concludes on the statement of the lemma. □

2.3 Proof of Theorem 2.1

2.3.1 Relationship with diagonalizing families

We first start by demonstrating the following lemma, which draws an important connection between the existence of a dual certificate for a measure μ and the existence of a bounded diagonalizing family on its support.

Lemma 2.3. *Let $X = \{x_k\}_{k=1}^{s}$ be a discrete subset of $\mathbb{T}$ with cardinality $s \leq m$. Suppose* *that for every* $u \in \mathbb{U}^s$*, there exists* $Q_u \in \mathcal{T}_m$ *such that*

$$\begin{cases} Q_u(x_k) = u_k, & \forall k \in [\![s]\!] \\ |Q_u(x)| < 1, & \forall x \notin X, \end{cases} \tag{2.7}$$

then X *admits at least one first order bounded diagonalizing family over* $\mathcal{T}_m$*.*

Proof. First for all, if $s = 2$, the diagonalizing family $\mathcal{P}_X = \{P_1, P_2\}$ can be trivially built by considering an appropriate linear combinations of the polynomials $Q_{[1,1]}$ and $Q_{[1,-1]}$. The rest of the proof details the more complicated case where $s \neq 2$.

Denote by $\mathcal{U} = \left\{u^{(k)}\right\}_{k=1}^{s} \subset \mathbb{U}^s$ the set of vectors defined by

$$\forall k \in [\![s]\!], \quad \begin{cases} u_k^{(k)} = 1 \\ u_l^{(k)} = -1, & \forall l \neq k. \end{cases}$$

By assumption, there exists a family $\mathcal{Q} = \{Q_{u^{(k)}}\}_{k=1}^{s} \subset \mathcal{T}_m$ of trigonometric polynomials satisfying the properties of (2.7) for each $u^{(k)} \in \mathbb{C}^s$. Moreover, since the vectors $\left\{u^{(k)}\right\}_{k=1}^{s}$ are real, and noticing that

$$\forall u \in \mathbb{C}^s, \forall \omega \in \mathbb{T}, \quad |\Re(Q_u(\omega))| \leq |Q_u(\omega)|,$$

the family of *real* trigonometric polynomials $\bar{\mathcal{Q}} = \left\{\bar{Q}_{u^{(k)}}\right\}_{k=1}^{s} = \{\Re(Q_{u^{(k)}})\}_{k=1}^{s}$ also verifies the properties of (2.7) for each $u^{(k)} \in \mathbb{C}^s$. In the rest of the proof, we aim to build a bounded diagonalizing family $\mathcal{P}_X$ of X lying in the span of the family $\bar{\mathcal{Q}}$. Namely, we construct

$$\forall l \in [\![s]\!], \quad P_l = \sum_{k=1}^{s} a_k^{(l)} \bar{Q}_{u^{(k)}},$$

whereby $\left\{a^{(l)}\right\}_{l=1}^{s} \subset \mathbb{C}^s$ are coefficients to be determined.

Let $\mathbf{U}_s$ be the matrix $\mathbf{U}_s = \left[u^{(1)}, \cdots, u^{(s)}\right] \in \mathbb{C}^{s\times s}$. One can equivalently write

$$\mathbf{U}_s = 2\mathbf{I}_s - \mathbf{J}_s,$$

whereby $\mathbf{I}_s, \mathbf{J}_s \in \mathbb{C}^{s\times s}$ denote respectively the identity and the all one matrix of size s. Furthermore, the matrix $\mathbf{U}_s$ is invertible for every $s \neq 2$ and one has

$$\forall s \in \mathbb{N} \backslash \{2\}, \quad \mathbf{U}_s^{-1} = \frac{1}{2}\mathbf{I}_s - \frac{1}{2(s-2)}\mathbf{J}_s. \tag{2.8}$$

Hence, the matrix $\mathbf{U}_s$ is full rank, and the s polynomials $\bar{\mathcal{Q}} = \left\{\bar{Q}_{u^{(k)}}\right\}_{k=1}^{s}$ are linearly independent and span a subspace of $\mathcal{T}_m$ of dimension s. Each vector $a^{(l)}$ is the unique solution of the linear system

$$\forall l \in [\![s]\!], \quad \delta_{k=l} = P_l(x_k) = \sum_{k=1}^{s} a_k^{(l)} \bar{Q}_{u^{(k)}}(x_l) = a_l^{(l)} - \sum_{k \neq l} a_k^{(l)},$$

which reformulates for every $l \in [\![s]\!]$ under the matrix form $\mathbf{U}_s a^{(l)} = e_l$, whereby e_l is the l^{th} vector of the canonical basis of $\mathbb{C}^s$. Consequently one has $a^{(l)} = \mathbf{U}_s^{-1} e_l$ for every $l \in [\![s]\!]$, and each polynomial P_l reads

$$\forall l \in [\![s]\!], \quad P_l = \frac{1}{2}\left(1 - \frac{1}{(s-2)}\right) \bar{Q}_{u^{(l)}} - \frac{1}{2(s-2)} \sum_{k \neq l} \bar{Q}_{u^{(k)}}. \tag{2.9}$$

The constructed family $\mathcal{P}_X = \{P_1, \cdots, P_s\} \subset \mathcal{T}_m$ verifies by construction the first condition of Equation (2.3).

Next, since $\left|\bar{Q}_u(x_k)\right| = |u_k| = 1$ and $\left|\bar{Q}_u(\omega)\right| < 1$ for every element ω lying in a small open ball centered on x_k, one may conclude that $Q_u'(x_k) = 0$ for all $k \in [\![s]\!]$. Hence, by linearity, P_l also satisfies $P_l'(x_k) = 0$ for all $k \in [\![s]\!]$. The second condition of (2.3) is verified and $\mathcal{P}_X$ is a first order diagonalizing family for X over $\mathcal{T}_m$.

Furthermore, one proves that the family$\mathcal{P}_X$ is bounded by applying the triangular inequality to Equation (2.9)

$$\begin{aligned}\forall \omega \in \mathbb{T}, \quad |P_l(\omega)| &\leq \frac{1}{2}\left(1 - \frac{1}{(s-2)}\right)\left|\bar{Q}_{u^{(l)}}(\omega)\right| + \frac{1}{2(s-2)} \sum_{k \neq l} \left|\bar{Q}_{u^{(k)}}(\omega)\right| \\ &\leq \frac{1}{2}\left(1 - \frac{1}{(s-2)}\right) + \frac{s-1}{2(s-2)} \\ &= 1,\end{aligned}$$

which ensures that $\|P_l\|_{L_\infty} \leq 1$. Finally, since $|P_l(x_l)| = 1$, one has $\|P_l\|_{L_\infty} = 1$, and the boundedness property of $\mathcal{P}_X$ follows. □

2.3.2 Existence of bounded diagonalizing families

It is worth noticing that, by a classic linear algebra argument, any set $X \subset \mathbb{T}$ with cardinality $s \leq m$ admits infinitely many diagonalizing families. However, the existence of a bounded one is not necessary guaranteed. Theorem 2.2 states that there exist sequences of sets with asymptotic minimal distance $\frac{1}{m}$ that do not admit a bounded diagonalizing family over $\mathcal{T}_m$. Its demonstration is delayed to Section 2.4 for readability.

Theorem 2.2 (Non-existence of a bounded diagonalizing family)**.** *For every real $\delta > 2$, there exists $M_\delta \in \mathbb{N}$, such that for every $m \geq M_\delta$, there exists a set $X_m = \left\{x_k^{(m)}\right\}_{k=1}^{s_m} \subset \mathbb{T}$ such that $\Delta_{\mathbb{T}}(X_m) \geq \frac{1}{m} - \frac{\delta}{m^2}$ and there is no bounded diagonalizing family of X over $\mathcal{T}_m$.*

2.3.3 Conclusion on Theorem 2.1

We now have all the elements to complete the proof of Theorem 2.1. Let $\delta > 2$ and $m \in \mathbb{N}$ with $m \geq M_\delta$ so that one can pick a subset $X_m = \{x_k\}_{k=1}^{s_m} \subset \mathbb{T}$ as in Theorem 2.2. Using the contraposition of Lemma 2.3 on X_m, there must exist one sign pattern $u \in \mathbb{U}^{s_m}$ such that there is no trigonometric polynomial verifying the condition of Theorem 1.1. Consider a measure μ_m of the form $\mu_m = \sum_{k=1}^{s_m} \tau_k u_k \delta_{x_k}$, whereby $\{\tau_k\}_{k=1}^{s_m}$ is a set of *strictly positive* reals. One has $\operatorname{sign}(\tau_k u_k) = u_k$, and we conclude using the negation of Proposition 1.1 that the measure μ_m is not solution of Program (1.27). □

2.4 Proof of Theorem 2.2

Let $m \in \mathbb{N}$, and let $X = \{x_k\}_{k=1}^{s}$ be a subset of $\mathbb{T}$ with cardinality s. We aim to construct a well-separated subset $X^{(m)} = \left\{x_k^{(m)}\right\}_{k=1}^{s} \subseteq \mathbb{T}$ for which every diagonalizing family $\mathcal{P}_X$ on the support set X is *not* bounded in the sense of Definition 2.2. More precisely, the goal is to prove that every family $\mathcal{P}_X$ should have at least one element $P_l \in \mathcal{P}_X$ so that $\left\|P_l^{(m)}\right\|_\infty > 1$.

For convenience, we restrict our analysis to odd trigonometric degrees $m = 2K+1$, and claim that the result is also extendable for even values of m. Let the parameter $\alpha_m \in (0,1)$ be such that $\frac{\alpha_m}{m+1} \triangleq \frac{1}{m} - \frac{\delta}{m^2}$ for some $\delta > 1$ and consider a subset $X^{(m,\delta)} = \left\{x_k^{(m,\delta)}\right\}_{k=-K}^{K}$ of m equispaced elements of the form

$$\forall k \in [\![-K, K]\!], \quad x_k^{(m,\delta)} \triangleq \frac{k\alpha_m}{m+1}.$$

It is clear that for every m the minimal distance of $X^{(m,\delta)}$ reads

$$\Delta_{\mathbb{T}}\left(X^{(m,\delta)}\right) = \frac{\alpha_m}{m+1} = \frac{1}{m} - \frac{\delta}{m^2}.$$

We show that the element $P_0 \in \mathcal{T}_m$ associated with the element $x_0^{(m)} = 0$ can be unstable for a large enough value of the parameter δ. By Lemma 2.2, P_0 can be factorized under the form $P_0 = Z_{m,\delta} \times R_0$, where $Z_{m,\delta} \triangleq Z_{X^{(m,\delta)},0} \in \mathcal{T}_{m-1}$ is the minimal polynomial (2.5) that vanishes on $X^{(m,\delta)} \setminus \{0\}$ and whereby $R_0 \in \mathcal{T}_1$ verifies the conditions (2.6). By symmetry of $X^{(m,\delta)}$ around 0, $R_0'(0) = 0$ and every trigonometric polynomial of degree 1 satisfying (2.6) can be written under the form

$$\forall \omega \in \mathbb{T}, \quad R_\gamma(\omega) \triangleq (1-\gamma) + \gamma\cos(2\pi\omega), \tag{2.10}$$

for some $\gamma \in \mathbb{C}$. Hence P_0 must have a factorization of the form

$$P_0 = P_{m,\delta,\gamma} \triangleq Z_{m,\delta} \times R_\gamma$$

for some $\gamma \in \mathbb{C}$.

It remains to show that if δ is large enough (α_m small enough), every polynomial of the form $P_{m,\delta,\gamma}$ verifies $\|P_{m,\delta,\gamma}\|_{L_\infty} > 1$. Formally, we aim to lower bound the quantity

$$L(m,\delta) \triangleq \inf_{\gamma\in\mathbb{C}} \|P_{m,\delta,\gamma}\|_{L_\infty} = \inf_{\gamma\in\mathbb{C}} \sup_{\omega\in\mathbb{T}} |P_{m,\delta,\gamma}(\omega)|$$

away from 1 for small enough α_m. Intuitively, we expect $Z_{m,\delta}$ to reach large values far away from its roots, at $\omega \simeq \frac{1}{2}$, and expect that the restrictive structure (2.10) on R_γ will not leave enough freedom to drag the product $Z_{m,\delta}(\omega) R_\gamma(\omega)$ below 1.

For ease of calculation, we introduce the translated polynomials $\tilde{Z}_{m,\delta}(\omega) = Z_{m,\delta}\left(\frac{1}{2}-\omega\right)$ and $\tilde{R}_\gamma(\omega) = R_\gamma\left(\frac{1}{2}-\omega\right)$ for all $\omega \in \mathbb{T}$, and let $\Omega_m = \left[-\frac{\alpha_m}{m+1}, \frac{\alpha_m}{m+1}\right] \subset \mathbb{T}$. The two following key lemmas, demonstrated in Section 2.5, provide lower bounds on $\tilde{Z}_{m,\delta}(\omega)$ and $\tilde{R}_\gamma(\omega)$ over the set Ω_m.

Lemma 2.4 (Bound on $\tilde{Z}_{m,\delta}$)**.** *Let $\Omega_m = \left[-\frac{\alpha_m}{m+1}, \frac{\alpha_m}{m+1}\right] \subset \mathbb{T}$ whereby $\frac{\alpha_m}{m+1} = \frac{1}{m} - \frac{\delta}{m^2}$ for some $\delta > 1$. There exists a constant $C(\delta) > 0$ such that*

$$\forall m \in \mathbb{N}, \forall \omega \in \Omega_m,\ \tilde{Z}_{m,\delta}(\omega) \geq C(\delta)(m+1)^{2(\delta-1)}.$$

Lemma 2.5 (Bound on $\tilde{R}_\gamma$)**.** *Let $R_\gamma \in \mathcal{T}_1$ be has in (2.10), then for all $\omega_{\max} \in \left[0, \frac{1}{2}\right]$ one has*

$$\inf_{\gamma\in\mathbb{C}} \sup_{\omega\in[-\omega_{\max},\omega_{\max}]} \left|\tilde{R}_\gamma(\omega)\right| = \frac{\sin^2(\pi\omega_{\max})}{1+\cos^2(\pi\omega_{\max})}. \tag{2.11}$$

One may lower bound the quantity $L(m,\delta)$ by controlling the infimum of each of the factor of $P_{m,\delta,\gamma}$. Applying Lemma 2.4 and Lemma 2.5 leads to

$$\begin{aligned}
L(m,\delta) &= \inf_{\gamma\in\mathbb{C}} \sup_{\omega\in\mathbb{T}} |Z_{m,\delta}(\omega) R_\gamma(\omega)| \\
&= \inf_{\gamma\in\mathbb{C}} \sup_{\omega\in\mathbb{T}} \left|\tilde{Z}_{m,\delta}(\omega) \tilde{R}_\gamma(\omega)\right| \\
&\geq \inf_{\gamma\in\mathbb{C}} \sup_{\omega\in\Omega_m} \left|\tilde{Z}_{m,\delta}(\omega) \tilde{R}_\gamma(\omega)\right| \\
&\geq \inf_{\omega\in\Omega_m} \tilde{Z}_{m,\delta}(\omega) \times \inf_{\gamma\in\mathbb{C}} \sup_{\omega\in\Omega_m} \left|\tilde{R}_\gamma(\omega)\right| \\
&= C(\delta)(m+1)^{2(\delta-1)} \times \frac{\sin^2\left(\pi\frac{\alpha_m}{m+1}\right)}{1+\cos^2\left(\pi\frac{\alpha_m}{m+1}\right)} \\
&\geq \frac{C(\delta)\pi^2\alpha_m^2}{2}(m+1)^{2(\delta-2)} \\
&= \Theta\left(m^{2(\delta-2)}\right),
\end{aligned} \tag{2.12}$$

where we used the fact that $\frac{\sin^2(\pi\omega_{\max})}{1+\cos^2(\pi\omega_{\max})} \geq \frac{\pi^2\omega_{\max}^2}{2}$ for $\omega_{\max} \leq 0.4$. Hence, if $\delta > 2$, $L(m,\delta)$ diverges when m grows large. Consequently, there must exist a value $M_\delta > 0$ such that, for all $m \geq M_\delta$, one has $\|P_{m,\delta,\gamma}\|_\infty > 1$ for every $\gamma \in \mathbb{C}$. Concluding on the proof of the theorem. □

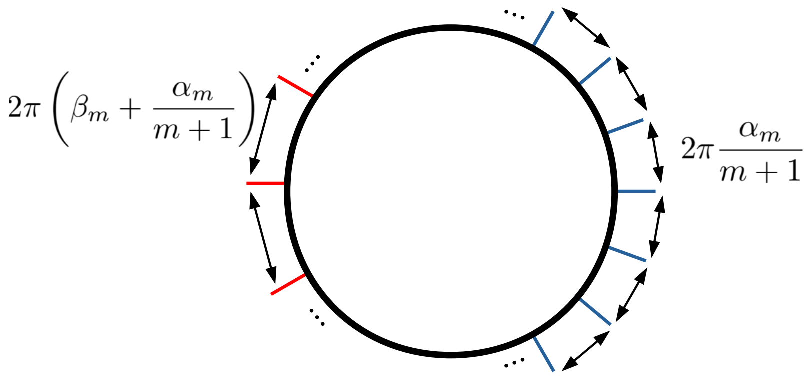

Figure 2.3: A representaion of the parameter α_m and of the associated offset factor β_m.

2.5 Proofs of the auxiliary lemmas

2.5.1 Proof of Lemma 2.4

First of all, the roots $\left\{\tilde{x}_k^{(m,\delta)}\right\}_{|k|=1}^K$ of $\tilde{Z}_{m,\delta}$ are given by the relation

$$\tilde{x}_k^{(m,\delta)} = \frac{1}{2} - x_{K-k+1}^{(m,\delta)},$$

and a direct calculation yields

$$\forall k \in [\![K]\!], \qquad \begin{cases} \tilde{x}_k = \beta_m + \frac{k\alpha_m}{m+1} \\ \tilde{x}_{-k} = -\beta_m - \frac{k\alpha_m}{m+1} \end{cases}$$

whereby $\beta_m \triangleq \frac{1}{2}(1-\alpha_m) > 0$ is an offset factor. Figure 2.3 provides a graphical interpretation of the quantities α_m and β_m for ease of understanding. Using Expression (2.5), one may rearrange $\tilde{Z}_{m,\delta}$ as follows

$$\forall \omega \in \mathbb{T}, \quad \tilde{Z}_{m,\delta}(\omega) = \prod_{k=1}^{K} \frac{\sin^2\left(\pi\left(\beta_m + \frac{k\alpha_m}{m+1} - \omega\right)\right)\sin^2\left(\pi\left(\beta_m + \frac{k\alpha_m}{m+1} + \omega\right)\right)}{\sin^4\left(\pi\frac{k\alpha_m}{m+1}\right)}.$$

The polynomial $\tilde{Z}_{m,\delta}$ has no root over the set Ω_m, hence its logarithm $\tilde{z}_{m,\delta}$ is well defined over Ω_m, and it yields

$$\forall \omega \in \Omega_m,\ \tilde{z}_{m,\delta}(\omega) = -4\ln\sin\left(\pi\frac{k\alpha_m}{m+1}\right) + 2\sum_{k=1}^{K}\ln\sin\left(\pi\left(\beta_m + \frac{k\alpha_m}{m+1} - \omega\right)\right) + \ln\sin\left(\pi\left(\beta_m + \frac{k\alpha_m}{m+1} + \omega\right)\right). \quad (2.13)$$

We derive a lower bound on $\tilde{z}_{m,\delta}$ over Ω_m by using the two elementary lemmas.

Lemma 2.6. *For any $t, h \in \mathbb{R}^+$ such that $t + h \leq \frac{\pi}{2}$, one has*

$$h\cot(t) - \frac{h^2}{2}\csc^2(t) \leq \ln\sin(t+h) - \ln\sin(t) \leq h\cot(a)$$

Proof of Lemma 2.6. First for all, one has $(\ln\sin)' = \cot$ and $(\ln\sin)'' = -\csc^2$. Hence, the function $t \mapsto \ln\sin(t)$ is concave on its definition domain, which is enough to prove the upper inequality. Next, one can apply the fundamental theorem of calculus to derive the lower inequality, which yields that for any $t, h \in \mathbb{R}^+$,

$$\ln\sin(t+h) - \ln\sin(t) = \int_t^{t+h} \cot(u)\,\mathrm{d}u.$$

Moreover, if $t + h \le \frac{\pi}{2}$ then the function $u \to \cot(u)$ is convex over $[t, t+h]$, leading to

$$\begin{aligned}\ln\sin(t+h) - \ln\sin(t) &\ge \int_t^{t+h} \cot(t) - u\csc^2(t)\,\mathrm{d}u \\ &= h\cot(t) - \frac{h^2}{2}\csc^2(t),\end{aligned}$$

concluding on the desired result. □

Lemma 2.7. *For all odd integer $m \in \mathbb{N}$ such that $m = 2K+1$ and all $\alpha \in (0,1)$, the following inequalities hold*

$$\begin{aligned}\sum_{k=1}^{K} \cot\left(\frac{\pi k\alpha}{m+1}\right) &\ge \frac{m+1}{\pi\alpha}\left[\ln(m+1) + \gamma + \ln\left(\frac{1}{\pi}\right)\right] \\ \sum_{k=1}^{K} \csc^2\left(\frac{\pi k\alpha}{m+1}\right) &\le \frac{(m+1)^2}{6\alpha^2},\end{aligned}$$

whereby γ denotes the Euler-Mascheroni constant.

Proof of Lemma 2.7. We start by demonstrating the first inequality. One has

$$\begin{aligned}\sum_{k=1}^{K} \cot\left(\frac{\pi k\alpha}{m+1}\right) &= \sum_{k=1}^{K} \frac{m+1}{\pi k\alpha} + \sum_{k=1}^{K} \cot\left(\frac{\pi k\alpha}{m+1}\right) - \frac{m+1}{\pi k\alpha} \\ &\ge \frac{m+1}{\pi\alpha} H_K + \frac{m+1}{\pi\alpha}\int_0^{\frac{\pi}{2}} \left(\cot(u) - \frac{1}{u}\right)\mathrm{d}u, \end{aligned} \tag{2.14}$$

whereby $H_K = \sum_{k=1}^{K} \frac{1}{k}$ is the K^{th} element of the harmonic series and by comparing the second sum with a Riemann approximation of the integral of the function $u \to \cot(u) - \frac{1}{u}$. Noticing that

$$\int_0^{\frac{\pi}{2}} \left(\cot(u) - \frac{1}{u}\right)\mathrm{d}u = \ln\left(\frac{2}{\pi}\right),$$

and using the fact that $H_K \ge \ln(K) + \gamma + \frac{1}{K}$ for all $K \in \mathbb{N}$, the inequality (2.14) reduces to

$$\begin{aligned}\sum_{k=1}^{K} \cot\left(\frac{\pi k\alpha}{m+1}\right) &\ge \frac{m+1}{\pi\alpha}\left[\ln(K) + \gamma + \frac{1}{K} + \ln\left(\frac{2}{\pi}\right)\right] \\ &= \frac{m+1}{\pi\alpha}\left[\ln\left(\frac{m+1}{2} - 1\right) + \gamma + \frac{2}{m-1} + \ln\left(\frac{2}{\pi}\right)\right] \\ &= \frac{m+1}{\pi\alpha}\left[\ln\left(\frac{m+1}{2}\left(1 - \frac{2}{m+1}\right)\right) + \gamma + \frac{2}{m-1} + \ln\left(\frac{2}{\pi}\right)\right]\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{m+1}{\pi\alpha}\left[\ln\left(\frac{m+1}{2}\right)+\gamma+\frac{2}{m-1}-\frac{2}{m+1}+\ln\left(\frac{2}{\pi}\right)\right] \\
&= \frac{m+1}{\pi\alpha}\left[\ln(m+1)+\gamma+\ln\left(\frac{1}{\pi}\right)\right]
\end{aligned}
$$

which concludes on the first inequality.

One can proceed on a similar way to prove the second inequality of Lemma 2.7.

$$
\begin{aligned}
\sum_{k=1}^{K}\csc^2\left(\frac{\pi k\alpha}{m+1}\right) &= \sum_{k=1}^{K}\left(\frac{m+1}{\pi k\alpha}\right)^2+\sum_{k=1}^{K}\csc^2\left(\frac{\pi k\alpha}{m+1}\right)-\left(\frac{m+1}{\pi k\alpha}\right)^2 \\
&\leq \frac{(m+1)^2}{\pi^2\alpha^2}\left(\frac{\pi^2}{6}-\frac{1}{K}+\frac{1}{2K^2}\right)+\frac{m+1}{\pi\alpha}\int_0^{\frac{\pi}{2}}\left(\csc^2(u)-\frac{1}{u^2}\right)\mathrm{d}u \\
&\leq \frac{(m+1)^2}{\pi^2\alpha^2}\left(\frac{\pi^2}{6}-\frac{2}{m+1}\right)+\frac{2(m+1)}{\pi^2\alpha} \\
&\leq \frac{(m+1)^2}{\pi^2\alpha^2}\left(\frac{\pi^2}{6}-\frac{2(1-\alpha)}{m+1}\right) \\
&\leq \frac{(m+1)^2}{6\alpha^2}
\end{aligned}
$$

whereby we used the fact that $\sum_{k=1}^{K}\frac{1}{k^2}\leq\frac{\pi^2}{6}-\frac{1}{K}+\frac{1}{2K^2}$ and where by viewed the second sum as a lower approximation of the integral

$$
\int_0^{\frac{\pi}{2}}\left(\csc^2(u)-\frac{1}{u^2}\right)\mathrm{d}u=\frac{\pi}{2}.
$$

□

We can now combine the previous results to finish the proof of Lemma 2.4. Since $\pi\left(\frac{K\alpha_m}{m+1}+\beta_m+|\omega|\right)<\frac{\pi}{2}$ for all $\omega\in\Omega_m$, one can apply two times Lemma 2.6 to each term of the sum (2.13), yielding

$$
\begin{aligned}
\tilde{z}_{m,\delta}(\omega) \geq & 4\pi\beta_m\sum_{k=1}^{K}\cot\left(\frac{\pi k\alpha_m}{m+1}\right)-2\pi^2\left(\beta_m^2+\omega^2\right)\sum_{k=1}^{K}\csc^2\left(\frac{\pi k\alpha_m}{m+1}\right) \\
\geq & \frac{4\beta_m(m+1)}{\alpha_m}\left[\ln(m+1)+\gamma+\ln\left(\frac{1}{\pi}\right)\right] \\
& -2\pi^2\left(\beta_m^2+\omega^2\right)\frac{(m+1)^2}{6\alpha^2} \\
\geq & \frac{4\beta_m(m+1)}{\alpha_m}\left[\ln(m+1)+\gamma+\ln\left(\frac{1}{\pi}\right)\right] \\
& -\pi^2\frac{\beta_m^2(m+1)^2}{3\alpha^2}\left(1+\frac{\alpha^2}{\beta_m^2(m+1)^2}\right) \\
\geq & 2(\delta-1)\left[\ln(m+1)+\gamma+\ln\left(\frac{1}{\pi}\right)\right]-\frac{\pi^2(\delta-1)^2}{3}\left(1+\frac{4}{(\delta-1)^2}\right) \\
= & 2(\delta-1)\ln(m+1)-\frac{\pi^2(\delta-1)^2}{3}+2(\delta-1)\left(\gamma+\ln\left(\frac{1}{\pi}\right)\right)-\frac{4\pi^2}{3}
\end{aligned} \tag{2.15}
$$

where we made use of Lemma 2.7, of the fact that $|\omega|\leq\frac{\alpha_m}{m+1}$, and noticing that $\frac{\delta-1}{2}\leq$

$\frac{\beta_m(m+1)}{\alpha_m} \leq \delta - 1$ for all $m \in \mathbb{N}$. Taking back the exponential in (2.15) leads to the desired result whereby

$$C(\delta) = e^{-\frac{\pi^2(\delta-1)^2}{3} + 2(\delta-1)\left(\gamma + \ln\left(\frac{1}{\pi}\right)\right) - \frac{4\pi^2}{3}}$$

is the desired constant. □

2.5.2 Proof of Lemma 2.5

In the following, let $\omega_{\max} \in \left[0, \frac{1}{2}\right]$ and $\Omega = [-\omega_{\max}, \omega_{\max}] \subset \mathbb{T}$, and define $c = \cos^2(\pi\omega_{\max}) \in [0,1]$ for convenience. We aim to find the value of $\gamma \in \mathbb{C}$ for which the supremum of $\left|\tilde{R}_\gamma(\omega)\right|$ is minimal over Ω. Noticing that $\left|\tilde{R}_\gamma(\omega)\right|^2 = (1 - 2|\gamma| c)^2$, the infimum in (2.11) is achieved for some positive real γ, hence

$$\kappa_\Omega \triangleq \inf_{\gamma \in \mathbb{C}} \sup_{\omega \in \Omega} \left|\tilde{R}_t(\omega)\right| = \inf_{\gamma \in \mathbb{R}^+} \sup_{\omega \in \Omega} \left|\tilde{R}_t(\omega)\right|.$$

Moreover, for a fixed value of γ, the symmetry of the function $\left|\tilde{R}_\gamma(\omega)\right|$ and its monotonic behaviors over $[0, \omega_{\max}]$ imply that the supremum is reached either on 0 or on $\omega_{\max}$, leading to

$$\begin{aligned} \sup_{\omega \in \Omega} \left|\tilde{R}_t(\omega)\right| &= \max\left\{\left|\tilde{R}_t(0)\right|, \left|\tilde{R}_t(\omega_{\max})\right|\right\} \\ &= \max\left\{|1 - 2\gamma|, |1 - 2\gamma c|\right\}. \end{aligned} \tag{2.16}$$

Define the auxiliary function g over $\mathbb{R}^+$ as $g(\gamma) = (1-2\gamma)^2 - (1-2\gamma c)^2$. $g(\gamma)$ is positive whenever the maximum (2.16) is reached at 0 and negative whenever it is reached at $\omega_{\max}$. The auxiliary function is parabolic in γ and we have

$$g(\gamma) = \left(1 - c^2\right)\gamma^2 - (1-c)\gamma,$$

which takes positive values for $\gamma \geq \frac{1}{1+c}$ and negative values for $\gamma \leq \frac{1}{1+c}$. Hence it yields

$$\sup_{\omega \in \Omega} \left|\tilde{R}_\gamma(\omega)\right| = \begin{cases} |1 - 2\gamma| & \text{if } \gamma \geq \frac{1}{1+c} \\ |1 - 2\gamma c| & \text{otherwise.} \end{cases} \tag{2.17}$$

The supremum in (2.17) is a piecewise monotonic function in γ. Thus, by similar argument

$$\begin{aligned} \kappa_\Omega &= \min\left\{\left|1 - \frac{2}{1+c}\right|, \left|1 - \frac{2c}{1+c}\right|\right\} \\ &= \frac{1-c}{1+c} \\ &= \frac{\sin^2(\pi\omega_{\max})}{1 + \cos^2(\pi\omega_{\max})}, \end{aligned}$$

concluding on the proof of the lemma. □

Chapter 3

Sufficient separation conditions and the diagonalizing certificate

3.1 The extremal interpolation problem

The fundamental dual certifiability Theorem 1.1 links the success or failure of the TV regularization approach to the sole existence of a *dual certificate*: A trigonometric polynomial satisfying the conditions (1.30) of Theorem 1.1. Although the statement of this theorem is simple and by many means elegant, it is absolutely not a trivial task to understand the mechanics ruling the existence of such polynomial for a given measure $\mu \in \mathcal{M}(\mathbb{T})$.

If it is easy to understand that a dual certificate $Q \in \mathcal{T}_m$ should *at least* belong to the linear subspace of dimension $2(m-s)+1$ defined by the equations

$$\begin{aligned} \forall k \in [\![s]\!], \quad & Q(x_k) = w_k \\ \forall k \in [\![s]\!], \quad & Q'(x_k) = 0, \end{aligned} \tag{3.1}$$

whereby $w \in \mathbb{U}^s$ denotes the sign pattern $w = [\operatorname{sign}(c_1), \cdots, \operatorname{sign}(c_s)]^\intercal$ of the measure to recover. The *extremal constraint*

$$\forall \omega \notin X, \quad |Q(\omega)| < 1 \tag{3.2}$$

is non-linear and hard to verify, since there is no close form expression linking the supremal value of Q with its coefficients vector $q \in \mathbb{C}^{2m+1}$. As explained in Subsection 1.6.2.3, it is still an open problem to determine precisely what are the exact conditions on the measure $\mu \in \mathcal{M}(\mathbb{T})$ under which one can ensure the existence of a dual certificate. Conjecture 1.1 claims that a minimal distance of $\Delta_{\mathbb{T}}(X) > \frac{\alpha}{m}$ on the support set X of the measure μ should be asymptotically enough to guarantee the existence of such an element for every $\alpha > 1$.

In this chapter, we propose a novel *constructive approach* for such certificate Q. The construction of Q is made on top of a certain diagonalizing family of trigonometric polynomials which have been defined in Section 2.2. We demonstrate in Theorem 3.1 the existence of a dual polynomial under near optimal regimes of the kind $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha_{\text{diag}}}{m-1}$

for some $\alpha_{\text{diag}} > 1$. Before going into the details of the proposed construction, we give a review of the one existing in the literature, and provide insights and a justification of our approach.

3.2 Previous results

3.2.1 Construction by translation

In their original analysis of the separation condition [18], the authors proposed a polynomial construction of a certificate working up to twice the TV resolution conjecture $\Delta_{\mathbb{T}}(X) \geq \frac{2}{m}$ and under proviso that m is large enough. This construction, summarized in the following, assumes in a first step to fix a kernel $K_m \in \mathcal{T}_m$ verifying the conditions

$$\forall \omega \in \mathbb{T}, \quad K_m(\omega) = K_m(-\omega)$$
$$\sup_{\omega \in \mathbb{T}} |K_m(\omega)| = K_m(0) = 1,$$

and that decays "fast enough" as ω goes away from 0. The construction of the certificate Q is done in the span of the translation of the kernel K_m and of its derivative K'_m at the location of the spikes $x_k \in X$ to interpolate. Hence, the candidate trigonometric polynomial can be written

$$\forall \omega \in \mathbb{T}, \quad Q(\omega) = \sum_{k=1}^{s} \alpha_k K_m(\omega - x_k) + \beta_k K'_m(\omega - x_k), \tag{3.3}$$

for some coefficients $\{\alpha_k\}_{k=1}^s$ and $\{\beta_k\}_{k=1}^s$. Since Q must at least be a dual pre-certificate in order to be a dual certificate for the desired problem, the coefficients $\{\alpha_k\}_{k=1}^s$ and $\{\beta_k\}_{k=1}^s$ in the construction (3.3) have to be solution of the linear system of $2s$ equations (3.1), and summarized by the matrix expression

$$\begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 \\ \mathbf{G}_1 & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \omega \\ 0 \end{bmatrix}, \tag{3.4}$$

whereby the symmetric matrices $\mathbf{G}_l \in \mathbb{C}^{s \times s}$ for $l \in \{0, 1, 2\}$ have for generic term

$$\forall (i,j) \in [\![s]\!]^2, \quad \mathbf{G}_l(i,j) = K_m^{(l)}(x_j - x_i).$$

The rest of the proof aims to demonstrate that if the kernel K_m is "sharp enough" then the linear system (3.4) is invertible and nearly diagonal so that $\alpha \simeq w$ and $\beta \simeq 0$. The last step of the proof consists in using the *approximate* local expression of the pre-certificate Q and ensuring that its modulus remains below 1. This is mostly done by proving two different inequalities

- *Near region bound:* There exists $C > 0$ such that for all $x_k \in X$, one has on an open ball $\mathcal{B}_{\mathbb{T}}(x_k, \varepsilon_{\text{near}})$ of large enough radius $\varepsilon_{\text{near}}$

$$\forall \varepsilon \leq \varepsilon_{\max}, \quad |Q(x_k + \varepsilon)| \leq 1 - C\varepsilon^2$$

- *Far region bound:* For every $\omega \notin \bigcup_{k=1}^{s} \mathcal{B}(x_k, \varepsilon_{\max})$ belonging to the complementary set one has $|Q(\omega)| \leq C'$ for some $C' \in [0,1)$

which is enough to conclude on the desired result.

The performances guarantees obtained by this construction are heavily relying on the initial choice of the kernel K_m used as a building block for the construction of the certificate Q. The initial proof presented in [18] made use of the Jackson kernel $J_m \in \mathcal{T}_m$ as a building basis, and is defined for every odd trigonometric order m by

$$\forall \omega \in \mathbb{T}, \quad J_m(\omega) = \frac{\sin^4\left(\frac{\pi}{2}(m+1)\,\omega\right)}{(m+1)^4 \sin^4\left(\frac{\pi}{2}\omega\right)}.$$

This construction leads to the sufficient separation guarantee $\Delta_{\mathbb{T}}(X) \geq \frac{2}{m}$ under proviso that $m \geq 128$. This bound was later enhanced in [33] to $\Delta_{\mathbb{T}}(X) \geq \frac{1.26}{m}$ under the stronger proviso that $m \geq 1000$ by replacing the Jackson certificate in the previous construction by a kernel K_m built from the product of three Dirichlet kernels defined in Equation (1.8) of different orders so that

$$\forall \omega \in \mathbb{T}, \quad K_m(\omega) = D_{m_1}(\omega)\, D_{m_2}(\omega)\, D_{m_3}(\omega)$$

for some numerically tuned integers verifying the condition $m_1 + m_2 + m_3 = m$ so that K_m has a trigonometric degree equal to m.

There are several drawbacks that are inherent to the detailed *construction per translation* of the certificate Q. First of all, the choice of the kernels is done on an *arbitrary manner* in both of the discussed works [18], [33], and it is difficult to understand the relation that links the kernel K_m with the final achievable resolution limit $\frac{\alpha(K_m)}{m}$ through this construction. Secondly, both of the previous constructions cannot empirically explain the conjectured phase transition, in the sense that experimental constructions of those certificates fail to verify the supremal condition (3.2) for most of the measures μ with non-critically separated support having minimal distance $\Delta_{\mathbb{T}}(X) \simeq \frac{1.2}{m}$. Finally, this type of construction require to define the certificate Q as a solution of a linear system, which requires to approximate the inverse of the matrix defined in Equation (3.4). Consequently the coefficients vector $q \in \mathbb{C}^{2m+1}$ is not known under a closed form formula.

3.2.2 The Dirichlet certificate and robustness to small noise levels

Understanding the stability of the TV regularization approach in the context of line spectral estimation requires understanding the performance limits of the Beurling-Lasso estimator (1.36) defined in Subsection 1.6.5. Such analysis was presented in [28] for much broader class of linear inverse problem by considering collapsing noise level and by letting the regularization parameter τ of the Beurling-Lasso tend to 0.

Translated in our context, the authors have shown that the solution $q_\star(\tau) \in \mathbb{C}^{2m+1}$ of the dual Beurling-Lasso Problem (1.37) was converging towards a limit $q_{\text{Dir}} \in \mathbb{C}^{2m+1}$ when $\tau \to 0$. Moreover, this limit is known to be the coefficients vector of the trigonometric polynomial $Q_{\text{Dir}} \in \mathcal{T}_m$ achieving minimal L_2 norm over the space of dual pre-certificate, therefore solution of

$$\begin{aligned} q_{\text{Dir}} = \arg \min_{q \in \mathbb{C}^{2m+1}} \|q\|_2 \\ \text{subject to} \quad Q(x_k) = w_k, \quad \forall k \in [\![1, s]\!] \\ Q'(x_k) = 0, \quad \forall k \in [\![1, s]\!]. \end{aligned} \tag{3.5}$$

Consequently, it is "enough" to verify that Q_{Dir} is a valid dual certificate, i.e., to verify that it satisfies the last extremal constraint (3.2) to conclude on the stability of TV regularization to arbitrary small noise levels. Since, the optimal coefficients q_{Dir} are solution of the simple quadratic program (3.5) with a linear constraint of the form $\mathbf{W}q = \begin{bmatrix} w \\ 0 \end{bmatrix}$ for some $\mathbf{W} \in \mathbb{C}^{(2m+1)\times 2s}$, a classic orthogonal projection argument ensures that $q_{\text{Dir}} \in \text{span}(\mathbf{W}^*)$. Consequently, a direct calculation yields the existence of complex coefficients $\{\alpha_k\}_{k=1}^s$ and $\{\beta_k\}_{k=1}^s$ such that

$$\forall \omega \in \mathbb{T}, \quad Q_{\text{Dir}}(\omega) = \sum_{k=1}^{s} \alpha_k D_m(\omega - x_k) + \beta_k D'_m(\omega - x_k), \tag{3.6}$$

where $D_m \in \mathcal{T}_m$ is the Dirichlet kernel defined in Equation (1.8).

It is worth noticing that the Dirichlet certificate Q_{Dir} is itself a construction by translation of the form (3.3). However, the generic recipe presented in Subsection 3.2.1 to derive an associated minimal separation bound is not suited for the kernel D_m due to the high non-locality. In particular, a major drawback for this kernel is that $D_m\left(\frac{\cdot}{m}\right) \sim \text{sinc}(\cdot)$ is not absolutely integrable on its definition domain.

3.3 The diagonalizing certificate

3.3.1 Motivations

A novel framework to construct a dual certificate verifying the conditions of Theorem 1.1 is proposed in this section. The diagonalizing certificates are build upon a well-suited basis of diagonalizing polynomials, whose definitions were discussed in Section 2.2. Diagonalizing basis have the huge advantage of avoiding the tedious inversion step of the linear system (3.4), which is one of the major flaws of the construction by translation presented in Section 3.2.1.

Denote by $X = \{x_1, \cdots, x_s\} \subset \mathbb{T}$ the support, and by $c \in \mathbb{C}^s$ the complex coefficients of the measure $\mu \in \mathcal{M}(\mathbb{T})$ to reconstruct. Moreover denote by $w = [\text{sign}(c_1), \cdots, \text{sign}(c_s)]^\mathsf{T} \in \mathbb{C}^s$ the sign pattern vector of μ. Denote by $\mathcal{P}_X = \{P_l\}_{l=1}^s$ a *diagonalizing family* for the support set X. It is immediate to notice that the trigonometric polynomial $Q_w \in \mathcal{T}_m$ defined by

$$Q_w = \sum_{l=1}^{s} w_l P_l, \tag{3.7}$$

is solution of the system of equations (3.1), and therefore is a *dual pre-certificate* for the

TV regularization problem. No inversion step of the form (3.4) is required to determine the values of the coefficients of Q_w. Once the basis $\mathcal{P}_X$ has been chosen, the expression of Q is given under the *closed form* (3.7) with respect to the sign pattern vector w.

It remains to show that $Q_w \in \mathcal{T}_m$ has a modulus that remains bounded by 1 on the torus to conclude on the tightness of the TV regularization. This chapter focuses on a construction on a *positive diagonalizing family* $\mathcal{P}_X$ for the support set X, and defined as follows.

Definition 3.1 (Positive diagonalizing family)**.** A first order diagonalizing family $\mathcal{P}_X = \{P_l\}_{l=1}^s$ of X is said to be *positive* if and only if $P_l(\omega) \geq 0$ for all $\omega \in \mathbb{T}$.

As stated in the next proposition, the positivity assumption on $\mathcal{P}_X$ allows to reduce the verification of the extremal constraint (3.2) for any sign patterns $w \in \mathbb{C}^s$ to bounding the sum of the elements of the family $\mathcal{P}_X$.

Proposition 3.1 (Universal certifiability on positive diagonalizing families)**.** *Let $X \subset \mathbb{T}$ be a discrete subset of cardinality s. If there exists a positive diagonalizing family of trigonometric polynomials $\mathcal{P}_X = \{P_l\}_{l=1}^s$ of degree m for the support set X verifying*

$$\forall \omega \in \mathbb{T} \backslash X, \quad \sum_{l=1}^{s} P_l(\omega) < 1$$

then the polynomial $Q_w \in \mathcal{T}_m$ defined in Equation (3.7) is a dual certificate for Theorem 1.1 for all $\omega \in \mathbb{U}^s$, and the convex program (1.27) succeed to reconstruct any sparse $\mu_0 \in \mathcal{M}(X)$.

Proof. For a given sign pattern $w \in \mathbb{U}^s$, denote by $Q_w \in \mathcal{T}_m$ defined in (3.7). Since Q_w is a dual pre-certificate for all $w \in \mathbb{U}^s$, it remains to show that $|Q(\omega)| < 1$ over $\mathbb{T} \backslash X$, which is an immediate consequence of the triangular inequality.

$$\begin{aligned} \forall \omega \in \mathbb{T} \backslash X, \quad |Q(\omega)| &= \left| \sum_{l=1}^{s} w_l P_l(\omega) \right| \\ &\leq \sum_{l=1}^{s} |w_l P_l(\omega)| \\ &= \sum_{l=1}^{s} |P_l(\omega)| \\ &= \sum_{l=1}^{s} P_l(\omega) \\ &< 1. \end{aligned}$$

concluding on the desired result. □

3.3.2 Main result

We start by introducing the function h defined by

$$h : \left(\frac{7}{4}, +\infty\right) \to \mathbb{R}^+$$
$$\alpha \mapsto \lim_{p\to+\infty} \left\{ \left(\frac{\pi\alpha}{2p}\right)^{4\alpha-6} \sum_{j=1}^{p} \csc^{4\alpha-6}\left(\pi\frac{j\alpha}{2p}\right) \right\},$$

which is well defined whenever $\alpha > \frac{7}{4}$ by a comparison argument with the Riemann zeta function. And let $\psi^{(1)}$ be the first order polygamma function defined by

$$\psi^{(1)} : \mathbb{R}^+ \to \mathbb{R}^+$$
$$u \mapsto \sum_{j=0}^{\infty} \frac{1}{(u+j)^2}.$$

The function $\alpha \mapsto \psi^{(1)}(1+\alpha)\alpha^2$ is a continuous increasing and unbounded function of the variable α, and the function h is continuous decreasing with

$$\lim_{\alpha\to\frac{7}{4}} h(\alpha) = +\infty$$
$$\lim_{\alpha\to\infty} h(\alpha) = 1.$$

By a classic application of the mean value theorem, the equation

$$\psi^{(1)}(1+\alpha)\alpha^2 = 2h(\alpha), \tag{3.8}$$

has a unique solution α_{diag} on $\left(\frac{7}{4}, +\infty\right)$. A numerical computation leads to the estimate $2.5682 < \alpha_{\text{diag}} < 2.5683$.

The next theorem states that it is possible to build a dual certificate for Theorem 1.1 of the form (3.7) up to the near optimal separation condition $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha_{\text{diag}}}{m-1}$. Moreover, the theorem offers useful properties on the decay rate of the certificate around the interpolated spikes and on its supremal value on the far region. The demonstration of this result is provided in Section 3.5 after introducing the useful notion of saturated diagonalizing families of trigonometric polynomials in Section 3.4.

Theorem 3.1 (Diagonalizing certificate)**.** *Let $\alpha_{\text{diag}} < 2.5683$ be the unique positive solution on $\left(\frac{7}{4}, +\infty\right)$ of Equation (3.8). If a support set $X = \{x_1, \cdots, x_s\}$ of cardinality s verifies the separation condition $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha}{m-1}$ with $\alpha > \alpha_{\text{diag}}$, then the trigonometric polynomial $Q_w \in \mathcal{T}_m$ defined in (3.7) is a dual diagonalizing certificate for Theorem 1.1, and verifies the decay conditions*

$$\begin{cases} \left|Q_w\left(x_k + \frac{\alpha\varepsilon}{m-1}\right)\right| \leq 1 - \left(\psi^{(1)}(1+\alpha)\alpha^2 - 2h(\alpha)\right)\varepsilon^2 & \forall k \in [\![s]\!], \forall \varepsilon \in [-1,1] \\ |Q_w(\omega)| < \frac{2h(\alpha)}{\alpha^2} & \forall \omega \in \Gamma_{\text{far}} \end{cases} \tag{3.9}$$

for any sign pattern $w \in \mathbb{U}^s$, whereby $\Gamma_{\text{far}} = \mathbb{T} \backslash \left\{\bigcup_{k=1}^{s} \mathcal{B}\left(x_k, \frac{\alpha}{m-1}\right)\right\}$.

3.3.3 Discussions

Theorem 3.1 can only certify that the diagonalizing pre-certificate verifies the dual certificate conditions of Theorem 1.1 under a separation limit of the kind $\Delta_{\mathbb{T}}(X) \geq \frac{2.5683}{m-1}$, which is a weaker result than the previous existing bounds in the literature. However, experimental results shown in Figure 3.1 suggest that the proposed construction should remains valid up to the resolution limit $\Delta_{\mathbb{T}}(X) = \frac{1}{m} + o\left(\frac{1}{m}\right)$, which is not the case of the Jackson construction described in Subsection 3.2.1. We believe that the limit $\alpha_{\text{diag}} \approx 2.5683$ is an unnecessary artifact of the technique used to demonstrate Theorem 3.1, and that this analysis can be sharpened.

Moreover, the inequalities (3.9) provided in Theorem 3.1 provide bounds on decay properties verified by the diagonalizing certificates. We believe that the provided bounds can be helpful in a future analysis of the performance of the Beurling-Lasso estimator for reconstruction in noisy environments in the same spirit than the one presented in [46].

3.4 Saturated diagonalizing families

If the support set $X = \{x_1, \cdots, x_s\} \subset \mathbb{T}$ is of cardinality $s \leq m$, then the space of all the trigonometric polynomials P_l verifying the conditions (2.3) of the l^{th} element of a diagonalizing family is an affine subspace of $\mathcal{T}_m$ of dimension $2(m-s)$. Among all those possible elements, one seeks to find an adequate one for serving as a basis element on which the dual pre-certificate $Q_w \in \mathcal{T}_m$ defined by Equation (3.7) is going to be built. As discussed in Subsection 3.2.1 a "good" interpolation basis for the extremal interpolation problem should decay rapidly around its supremal value, while being bounded by a small value far away from the central spike.

To this end, the heuristic construction of *saturated* diagonalizing basis is introduced in the following. The saturated polynomials are constructed by using that $2(m-s)$ extra degrees of freedom in such a way that the resulting roots of the polynomials are *as evenly distributed as possible* around on the unit circle, which will guarantee a strong central decay around x_l, as well as a good control of the polynomial values away from x_l.

For convenience and ease of notations, the rest of the analysis is restricted to odd values of the trigonometric degree $m = 2p + 1$. Denote by $\mathcal{S}$ the subdivision $\mathcal{S} = \{S_{-p}, \cdots, S_{-1}, S_0, S_1, \cdots, S_p\}$ of the torus $\mathbb{T}$ into m intervals of the form

$$\begin{cases} S_0 = \left(-\frac{\alpha}{2p}, \frac{\alpha}{2p}\right) & \\ S_j = \left[\frac{\alpha}{2p} + (j-1)\frac{\beta}{2p}, \frac{\alpha}{2p} + j\frac{\beta}{2p}\right) & \forall j \in [\![1, p]\!] \\ S_{-j} = -S_j & \forall j \in [\![1, p]\!] \end{cases}$$

whereby $\beta = \beta(\alpha)$ is fully determined by the choice of α as follows

$$\beta = 1 - \frac{\alpha}{p} < 1, \tag{3.10}$$

so that $\frac{\alpha}{m-1} + p\frac{\beta}{m-1} = \frac{1}{2}$, and the resulting subdivision $\mathcal{S}$ forms indeed a partition of the

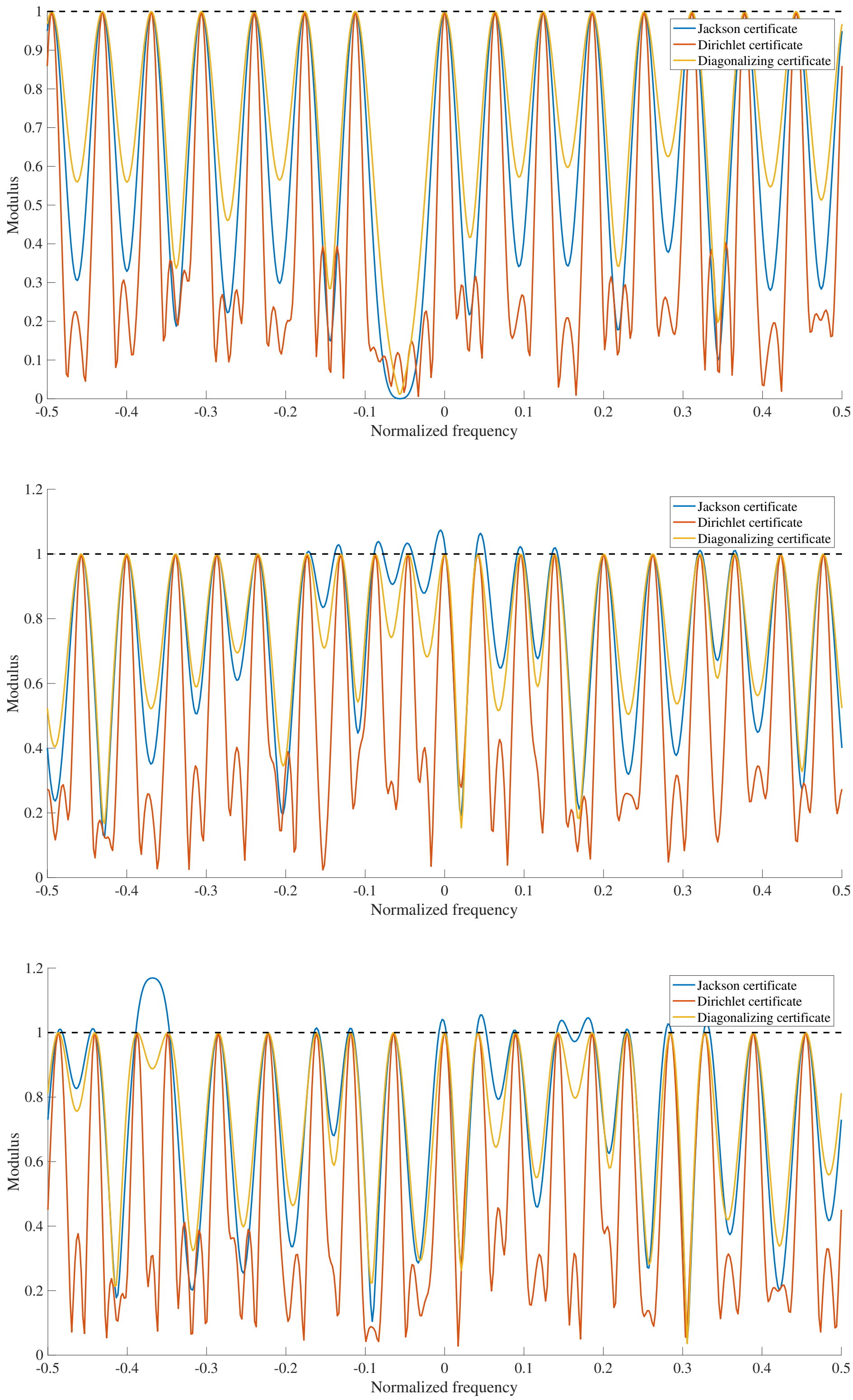

Figure 3.1: The Jackson pre-certificate (in blue), the Dirichlet pre-certificate (in red), and the diagonalizing pre-certificate (in yellow) for $m = 30$. The support set is such that $\Delta_{\mathbb{T}}(X) = \frac{\alpha}{m-1}$ for three different values of α.
Top: $\alpha = 1.8$ The three pre-certificate meets the dual certificate conditions.
Middle: $\alpha = 1.2$ The Jackson pre-certificate fails to meet the dual certificate conditions, while the Dirichlet and the diagonalizing construction validates certificates Conditions (1.30).
Bottom: $\alpha = 1.05$ The Jackson pre-certificate has a maximal modulus even larger than in the previous settings, while the Dirichlet and diagonalizing construction are still meeting the requirements.

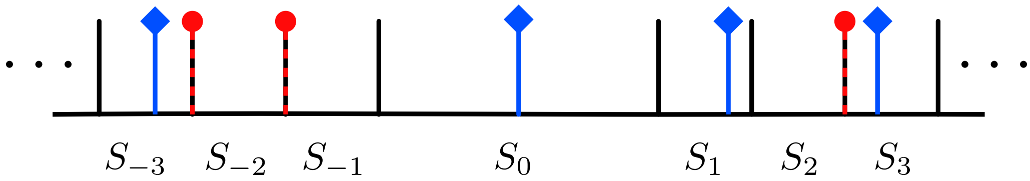

Figure 3.2: A representation of the $\frac{\alpha}{m-1}$-saturated extension of a set X_0. The elements of X_0 corresponds to the blue diamonds. The additional elements of the saturated extension $\tilde{X}_0$ are represented by red circles.
Since X_0 has no element in the bin S_2, an additional one is added to the saturated extension on the right boundary of S_2. Similarly, since S_{-3} and S_{-2} contain no element of X_0, additional elements are added on their left boundary.

torus $\mathbb{T} \simeq \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Definition 3.2 (Saturated extension)**.** Let $X_0 \subseteq \mathbb{T}$ be a set verifying $0 \in X_0$ and with minimal separation $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha}{m-1}$. The $\frac{\alpha}{m-1}$-saturated extension $\tilde{X}_0$ of the set X_0 is the unique subset $\tilde{X}_0 \subseteq \mathbb{T}$ verifying

- $\tilde{X}_0 \cap S_0 = \{0\}$
- For every $j \in [\![1, p]\!]$ if $X_0 \cap S_j = \emptyset$ then $\frac{\alpha + j\beta}{2p} \in \tilde{X}_0$
- For every $j \in [\![1, p]\!]$ if $X_0 \cap S_{-j} = \emptyset$ then $-\frac{\alpha + j\beta}{2p} \in \tilde{X}_0$.

It is easy to see that by construction there is a unique element of $\tilde{X}_0$ in each of the subset S_j for $j \in [\![-p, p]\!]$. The element of $\tilde{X}_0 \cap S_j$ will be denoted $\tilde{x}_j$ so that $\tilde{X}_0 = \{\tilde{x}_s\}_{s=-p}^{p}$. Figure 3.2 draws a representation of a set X_0 and its saturated extension $\tilde{X}_0$. The next proposition provides bounds on the decay rate of at least one trigonometric polynomial associated with the 0^{th} root $\tilde{x}_0 = 0$ of a positive diagonalizing family defined over the $\frac{\alpha}{m-1}$-saturated extension $\tilde{X}_0$ of X_0.

Proposition 3.2 (Properties of the saturated polynomial)**.** *Let $X_0 \subseteq \mathbb{T}$ be a set verifying $0 \in X_0$ and of minimal separation $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha}{m-1}$. Denote by $\tilde{X}_0$ the $\frac{\alpha}{m-1}$-saturated extension of X_0. There exists a positive diagonalizing polynomial $\tilde{P}_0 \in \mathcal{T}_m$ for the set $\tilde{X}_0$ verifying for all $\varepsilon \in [-1, 1]$*

$$\begin{cases} \tilde{P}_0(0) = 1 \\ \tilde{P}_0\left(\frac{\alpha\varepsilon}{m-1}\right) \leq 1 - \psi^{(1)}(1+\alpha)\,\alpha^2\varepsilon^2 \\ \tilde{P}_0\left(\tilde{x}_j + \frac{\varepsilon\alpha}{m-1}\right) \leq \left(\frac{\pi\alpha}{m-1}\csc(\pi\tilde{x}_j)\right)^{4\alpha-6}\varepsilon^2. \end{cases} \tag{3.11}$$

3.5 Proof of Theorem 3.1

Suppose that the support set $X = \{x_1, \cdots, x_s\}$ of cardinality s verifies the hypothesis $\Delta_{\mathbb{T}}(X) \geq \frac{\alpha}{m-1}$. Moreover assume without loss of generality that the points $\{x_1, \cdots, x_s\}$ are in cyclic increasing order. First of all, one must fix the positive diagonalizing basis $\mathcal{P}_X = \{P_l\}_{l=1}^{s} \subset \mathcal{T}_m$ that will be used to build the attempted certificate Q_w as in (3.7). For every $l \in [\![1, s]\!]$, denote by $Y_l = X - x_l$ the translated support centered around x_l.

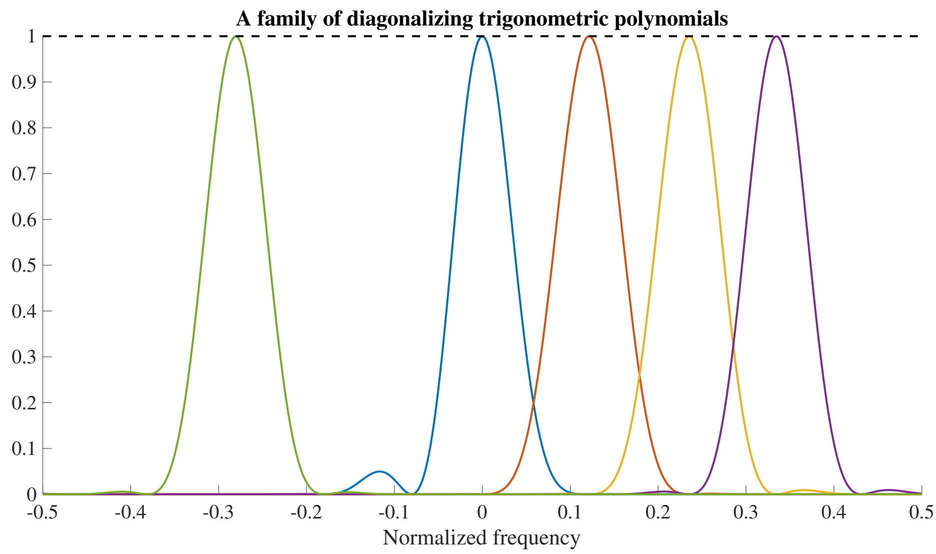

Figure 3.3: Different trigonometric polynomials used as elements of the diagonalizing family $\mathcal{P}_X$.

Since $0 \in Y_l$ by construction, Y_l admits an $\frac{\alpha}{m-1}$-saturated extension denoted $\tilde{Y}_l$. Y_l verifies the hypothesis of Proposition 3.2 and we denote by $\tilde{P}_l \in \mathcal{T}_m$ a polynomial verifying the properties (3.11). Finally, we denote by

$$P_l(\cdot) = \tilde{P}_l(\cdot - x_l)$$

its translation. P_l vanishes on the support $X \backslash \{x_l\}$ up to the first order, and will be used as the l^{th} element of the positive diagonalizing family of trigonometric polynomial $\mathcal{P}_X = \{P_l\}_{l=1}^{s}$. A graphic representation of the basis $\mathcal{P}_X$ is given in Figure 3.3. We start the demonstration by proving the first property of (3.9). For any $k \in [\![s]\!]$, the constructed certificate $Q_w \in \mathcal{T}_m$ defined by Equation (3.7) verifies

$$\begin{aligned} \left| Q_w\left(x_k + \frac{\alpha\varepsilon}{m-1}\right)\right| &= \left|\sum_{l=1}^{s} w_l P_l\left(x_k + \frac{\alpha\varepsilon}{m-1}\right)\right| \\ &\leq \sum_{l=1}^{s} P_l\left(x_k + \frac{\alpha\varepsilon}{m-1}\right) \\ &= \sum_{l=1}^{s} \tilde{P}_l\left(x_k - x_l + \frac{\alpha\varepsilon}{m-1}\right). \\ &= \tilde{P}_l\left(\frac{\alpha\varepsilon}{m-1}\right) + \sum_{\substack{l=1\\ l\neq k}}^{s} \tilde{P}_l\left(x_k - x_l + \frac{\alpha\varepsilon}{m-1}\right). \end{aligned} \tag{3.12}$$

By construction, the elements $x_k - x_l \in Y_l$ for all $k \neq l$. Moreover, using the separation constraint imposes that

$$|x_l - x_k|_{\mathbb{T}} \geq \frac{\alpha\,|l-k|}{m-1},$$

and a direct application of Proposition 3.2 on the elements of the sum (3.12) yields

$$\begin{aligned}\left|Q_w\left(x_k+\frac{\alpha\varepsilon}{m-1}\right)\right| &\leq 1-\psi^{(1)}\left(1+\alpha\right)\alpha^2\varepsilon^2+\sum_{\substack{l=1\\l\neq k}}^{s}\left(\frac{\pi\alpha}{m-1}\csc\left(\pi\left(x_k-x_l\right)\right)\right)^{4\alpha-6}\varepsilon^2\\ &\leq 1-\psi^{(1)}\left(1+\alpha\right)\alpha^2\varepsilon^2+2\sum_{j=1}^{p}\left(\frac{\pi\alpha}{m-1}\csc\left(\pi\frac{j\alpha}{m-1}\right)\right)^{4\alpha-6}\varepsilon^2\\ &\leq 1-\left(\psi^{(1)}\left(1+\alpha\right)\alpha^2-2h\left(\alpha\right)\right)\varepsilon^2.\end{aligned}$$

By assumption $\alpha > \alpha_{\text{diag}}$, and therefore one has

$$\psi^{(1)}\left(1+\alpha\right)\alpha^2-2h\left(\alpha\right)>0.$$

Hence if $\varepsilon \neq 0$, one has

$$\begin{aligned}\left|Q_w\left(x_k+\frac{\alpha\varepsilon}{m-1}\right)\right| &\leq 1-\left(\psi^{(1)}\left(1+\alpha\right)\alpha^2-2h\left(\alpha\right)\right)\varepsilon^2\\ &<1.\end{aligned}$$

Consequently the constructed certificate will have a modulus strictly smaller than 1 on a desired region around x_k, proving the first inequality.

It remains to be shown that $|Q_w(\omega)|$ is small in the far region. If $\omega \in \Gamma_{\text{far}}$, then for every $l \in [\![s]\!]$ the translated point $\omega - x_l$ belongs to some S_j for $j \neq 0$. Therefore ω can be written $\tilde{x}_j + \frac{\varepsilon\alpha}{m-1}$ for some $\varepsilon \in \left[-\frac{\beta}{\alpha}, \frac{\beta}{\alpha}\right]$. Applying Proposition 3.2 yields

$$\begin{aligned}P_l\left(\omega\right) &= P_l\left(\tilde{x}_j+\frac{\varepsilon\alpha}{m-1}\right)\\ &= \tilde{P}_l\left(\tilde{x}_j-x_l+\frac{\varepsilon\alpha}{m-1}\right)\\ &\leq \left(\frac{\beta}{\alpha}\right)^2\left(\frac{\pi\alpha}{m-1}\csc\left(\pi\left(\tilde{x}_j-x_l\right)\right)\right)^{4\alpha-6},\\ &\leq \frac{1}{\alpha^2}\left(\frac{\pi\alpha}{m-1}\csc\left(\pi\left(\tilde{x}_j-x_l\right)\right)\right)^{4\alpha-6}\end{aligned}$$

and one concludes in a similar way that for all $\alpha > \alpha_{\text{diag}}$

$$\begin{aligned}\left|Q_w\left(\omega\right)\right| &\leq \frac{2}{\alpha^2}\left(\frac{\pi\alpha}{m-1}\right)^{(4\alpha-6)}\sum_{j=1}^{p}\csc^{(4\alpha-6)}\left(\pi\frac{j\alpha}{m-1}\right)\\ &\leq \frac{2h\left(\alpha\right)}{\alpha^2}\\ &<1,\end{aligned}$$

concluding the demonstration of the theorem. □

3.6 Proof of Proposition 3.2

3.6.1 Preliminaries

First for all, since $\tilde{P}_0$ diagonalizes the set $\tilde{X}_0$ and $\tilde{P}_0(0) = 1$, Lemma 2.2 ensures the existence of a factorization under the form

$$\tilde{P}_0 = Z_0 R_0 \tag{3.13}$$

such that Z_0 is the first order annihilating polynomial on the saturated partition $\tilde{X}_0 \backslash \{0\}$ given by Equation (2.5). Moreover, Z_0 is by construction a trigonometric polynomial of degree exactly $2p$, and R_0 has to be of maximal degree 1. Denote by $\gamma_{\tilde{X}}$ the quantity

$$\gamma_{\tilde{X}} = \sum_{j=1}^{2p} \cot\left(\pi \tilde{x}_j\right), \tag{3.14}$$

and choose R_0 as follows

$$\begin{aligned} \forall \omega \in \mathbb{T}, \quad R_0(\omega) &= \left(\cos(\pi\omega) - \gamma_{\tilde{X}} \sin(\pi\omega)\right)^2 \\ &= \frac{\sin^2\left(\pi\omega - \operatorname{arccot}\left(\gamma_{\tilde{X}}\right)\right)}{\sin^2\left(\pi \operatorname{arccot}\left(\gamma_{\tilde{X}}\right)\right)} \\ &= \left(1+\gamma_{\tilde{X}}^2\right) \sin^2\left(\pi\omega - \operatorname{arccot}\left(\gamma_{\tilde{X}}\right)\right). \\ &= \left(1+\gamma_{\tilde{X}}^2\right) \sin^2\left(\pi\left(\omega - r_{\tilde{X}}\right)\right), \end{aligned} \tag{3.15}$$

whereby $\pi r_{\tilde{X}} = \operatorname{arccot}\left(\gamma_{\tilde{X}}\right) \in \mathbb{T}$ is the unique root of R_0. It is straight forward to verify that the chosen R_0 verifies the constraint given by Equation (2.6), that it is a positive trigonometric polynomial of degree 1.

We introduce in the following the auxiliary function g_ω defined for every $\omega \in \mathbb{T}$ as follows

$$\begin{aligned} g_\omega : \mathbb{T} \backslash \{0, \omega\} &\to \mathbb{R} \\ x &\mapsto \ln\sin\left(\pi\left|\omega - x\right|\right) - \ln\sin\left(\pi x\right). \end{aligned}$$

It is trivial to verify that the auxiliary function g_ω meets the following variational properties on its definition domain:

- $\lim_{x\to 0} g_\omega(x) = \lim_{x\to 1} g_\omega(x) = +\infty$
- $\lim_{x\to \omega} g_\omega(x) = -\infty$
- g_ω is decreasing on $(0, \omega)$
- g_ω is increasing on $(\omega, 1)$

Since P_0, Z_0 and R_0 are strictly positive nearly everywhere on their definition domains, one can define their logarithms p_0, z_0 and r_0 so that

$$\tilde{p}_0(\omega) = \ln \tilde{P}_0(\omega)$$
$$z_0(\omega) = \ln Z_0(\omega)$$
$$r_0(\omega) = \ln R_0(\omega)$$

for nearly everywhere on the torus. Finally, we assume without loss of generality on ordering of the elements of $\tilde{X}$ such that

$$0 = \tilde{x}_0 < \frac{\alpha}{m-1} \leq \tilde{x}_1 < \cdots < \tilde{x}_{2p} \leq 1 - \frac{\alpha}{m-1}.$$

Subsection 3.6.2 shows the second inequality of (3.11), while Subsection 3.6.3 introduces a demonstration of the third one.

3.6.2 Proof of the central decay

Since $|\omega| < \frac{\alpha}{m-1} \leq x_1$ by assumptions, one can use Lemma 2.6 to bound the quantity $g_\omega(\tilde{x}_k)$ for every $j \in [\![2p]\!]$ as follows

$$\begin{aligned} g_\omega(\tilde{x}_j) &= \ln\sin(\pi(\tilde{x}_j - \omega)) - \ln\sin(\pi\tilde{x}_j) \\ &\leq \pi\omega\cot(\pi\tilde{x}_j) - \frac{\pi^2\omega^2}{2}\csc^2(\pi(\omega - \tilde{x}_j)). \end{aligned}$$

The logarithm $\tilde{p}_0(\omega)$ can be upper bounded by

$$\begin{aligned} \frac{1}{2}\tilde{p}_0(\omega) &= \frac{1}{2}(z_0(\omega) + r_0(\omega)) \\ &= \sum_{j=1}^{2p} g_\omega(\tilde{x}_j) + g_\omega(r_{\tilde{X}}) \\ &\leq \pi\omega\left(\sum_{j=1}^{2p}\cot(\pi\tilde{x}_j) - \cot\operatorname{arccot}(\gamma_{\tilde{X}})\right) \\ &\qquad - \frac{\pi^2\omega^2}{2}\left(\sum_{j=1}^{2p}\csc^2(\pi\tilde{x}_j) + \csc^2(\operatorname{arccot}(\gamma_{\tilde{X}}))\right), \end{aligned} \tag{3.16}$$

whereby $\gamma_{\tilde{X}}$ is defined in Equation (3.14). Moreover, by definition of $\gamma_{\tilde{X}}$,

$$\sum_{j=1}^{2p}\cot(\pi\tilde{x}_j) - \cot\operatorname{arccot}(\gamma_{\tilde{X}}) = 0,$$

and the expression (3.16) reduces to

$$\begin{aligned}
\frac{1}{2}\tilde{p}_0(\omega) &= -\frac{\pi^2\omega^2}{2}\left(\sum_{j=1}^{2p}\csc^2(\pi\tilde{x}_j) + \csc^2(\operatorname{arccot}(\gamma_{\tilde{X}}))\right)\\
&\leq -\frac{\pi^2\omega^2}{2}\left(\sum_{j=1}^{2p}\csc^2(\pi\tilde{x}_j) + 1\right)\\
&\leq -\frac{\pi^2\omega^2}{2}\left(\sum_{j=1}^{2p}\inf_{u_j\in S_j}\csc^2(\pi u_j) + 1\right)\\
&= -\pi^2\omega^2\left(\sum_{j=1}^{p}\csc^2\left(\pi\frac{\alpha+j\beta}{2p}\right) + 1\right).
\end{aligned}$$

The above inequality can be further reduced using the fact that $\csc^2(u) \leq \frac{1}{u^2}$ for all $u \in [0,\pi]$, leading to

$$\begin{aligned}
\frac{1}{2}\tilde{p}_0(\omega) &\leq -\pi^2\omega^2\left(\frac{2p}{\pi\beta}\right)^2\left(\sum_{j=1}^{p}\frac{1}{\left(\frac{\alpha}{\beta}+j\right)^2} + \frac{\pi^2\beta^2}{2p}\right)\\
&\leq -(2p\omega)^2\left(\sum_{j=1}^{p}\frac{1}{(\alpha+j)^2} + \frac{\pi^2\beta^2}{2p}\right)\\
&\leq -(2p\omega)^2\left(\psi^{(1)}(1+\alpha) - \frac{1}{p} + \frac{\pi^2\beta^2}{2p}\right)\\
&\leq -(2p\omega)^2\,\psi^{(1)}(1+\alpha),
\end{aligned}$$

whereby we used the fact that

$$\forall p \in \mathbb{N}, \quad \sum_{j=1}^{p}\frac{1}{(\alpha+j)^2} \geq \psi^{(1)}(1+\alpha) - \frac{1}{p}.$$

One concludes taking back the exponential that

$$\begin{aligned}
\forall\,|\omega| \leq \frac{\alpha}{m-1}, \quad P_0(\omega) &\leq \exp\left(-2(m-1)^2\,\psi^{(1)}(1+\alpha)\,\omega^2\right)\\
&\leq 1-(m-1)^2\,\psi^{(1)}(1+\alpha)\,\omega^2. \qquad (3.17)
\end{aligned}$$

Substituting ω by $\frac{\alpha\varepsilon}{m-1}$ in inequality (3.17) leads to the desired result. □

3.6.3 Proof of the flatness properties

3.6.3.1 Preliminaries

This subsection aims to prove the flatness properties of P_0 around its roots, which is described by the third inequality of (3.11). The essential part of the demonstration aims to bound each elements of the factorization (3.13), which is done in the following lemmas, whose demonstration have been delayed in Subsections 3.6.3.2 and 3.6.3.3 for readability.

Lemma 3.1. *For all $j \in [\![1, 2p]\!]$ and all $\varepsilon \in [-1, 1]$ one has,*

$$Z_0\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1}\right) \leq \left(\frac{\pi\alpha}{m-1}\csc\left(\pi\tilde{x}_j\right)\right)^{4(\alpha-1)} \times \left(\frac{m-1}{\pi\alpha}\right)^2 \sin^2\left(\pi\frac{\alpha\varepsilon}{m-1}\right).$$

Lemma 3.2. *Let R_0 be the residual trigonometric polynomial of the factorization (3.13). For all $j \in [\![1, 2p]\!]$ and all $\varepsilon \in [-1, 1]$ one has,*

$$R_0\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1}\right) \leq \csc^2\left(\pi\frac{\alpha}{m-1}\right)\sin^2\left(\pi\tilde{x}_j\right).$$

The conclusion on the desired result is immediate by making use of the previous lemmas

$$\begin{aligned}
\tilde{P}_0\left(\tilde{x}_j + \frac{\varepsilon\alpha}{m-1}\right) &= Z_0\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1}\right) R_0\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1}\right) \\
&\leq \left(\frac{\pi\alpha}{m-1}\csc\left(\pi\tilde{x}_j\right)\right)^{4\alpha-6} \times \left(\frac{m-1}{\pi\alpha}\right)^2 \sin^2\left(\pi\frac{\alpha\varepsilon}{m-1}\right) \\
&\leq \left(\frac{\pi\alpha}{m-1}\csc\left(\pi\tilde{x}_j\right)\right)^{4\alpha-6} \varepsilon^2.
\end{aligned}$$

3.6.3.2 Proof of Lemma 3.1

We start by bounding the annihilating polynomial on the open interval $I_j = \left(\tilde{x}_j, \tilde{x}_j + \frac{\alpha}{m+1}\right)$ for $j \in [\![1, 2p]\!]$. Next, since $\tilde{x}_l \in S_l$ for every $l \in [\![1, 2p]\!]$, one can further bound the quantity z_0 as follows

$$\begin{aligned}
\frac{1}{2}z_0\left(\omega\right) &= \sum_{l=1}^{2p} g_\omega\left(\tilde{x}_l\right) \\
&= \sum_{l=1}^{j-1} g_\omega\left(\tilde{x}_l\right) + \sum_{l=j+1}^{2p} g_\omega\left(\tilde{x}_l\right) + g_\omega\left(\tilde{x}_j\right) \\
&\leq \sum_{l=1}^{j-1} \sup_{\tilde{x}_l \in S_l} g_\omega\left(\tilde{x}_l\right) + \sum_{l=j+1}^{2p} \sup_{\tilde{x}_l \in S_l} g_\omega\left(\tilde{x}_l\right) + g_\omega\left(\tilde{x}_j\right). \\
&= \sum_{l=1}^{j-1} g_\omega\left(\frac{(\alpha-\beta)+l\beta}{2p}\right) + \sum_{l=1}^{2p-j} g_\omega\left(1 - \frac{(\alpha-\beta)+l\beta}{2p}\right) + g_\omega\left(\omega-\varepsilon\right).
\end{aligned} \tag{3.18}$$

Define the intermediate variable t_j for every $j \in [\![1, 2p]\!]$ to be such that

$$\tilde{x}_j = \frac{\alpha}{2p} + \left(j - 1 + t_j\right)\frac{\beta}{2p}. \tag{3.19}$$

Since by construction of the saturated partition $\tilde{X}_0$ the root $\tilde{x}_j$ belongs to S_j, one has the guarantee that $t_j \in [0, 1]$. Therefore, one can rearrange each of the first sum in the last inequality of (3.18) as

$$\sum_{l=1}^{j-1} g_\omega\left(\frac{(\alpha-\beta)+l\beta}{2p}\right) = \sum_{l=1}^{j-1} \ln\sin\left(\pi\left(\omega - \frac{(\alpha-\beta)+l\beta}{2p}\right)\right) - \ln\sin\left(\pi\frac{(\alpha-\beta)+l\beta}{2p}\right)$$

$$
\begin{aligned}
&= \sum_{l=1}^{j-1} \ln \sin \left(\pi \frac{t_j + \alpha + (j-1)\,\beta - (\alpha - \beta) - l\beta}{2p} \right) \\
&\qquad - \ln \sin \left(\pi \frac{(\alpha - \beta) + l\beta}{2p} \right) \\
&= \sum_{l=1}^{j-1} \ln \sin \left(\pi \frac{t_j + l\beta}{2p} \right) - \ln \sin \left(\pi \frac{(\alpha - \beta) + l\beta}{2p} \right). \qquad (3.20)
\end{aligned}
$$

One could proceed on a similar way to obtain the corresponding equality on the second sum

$$
\sum_{l=1}^{2p-j} g_\omega \left(1 - \frac{(\alpha - \beta) + l\beta}{2p} \right) = \sum_{l=1}^{2p-j} \ln \sin \left(\pi \frac{\beta - t_j + l\beta}{2p} \right) - \ln \sin \left(\pi \frac{(\alpha - \beta) + l\beta}{2p} \right). \tag{3.21}
$$

Making use of Lemma 2.6 and by comparing the integral of $u \to \cot(u)$ with a Riemann sum in the middle, one can further bound the expression (3.20) as

$$
\begin{aligned}
\sum_{l=1}^{j-1} g_\omega \left(\frac{(\alpha - \beta) + l\beta}{2p} \right) &\leq \frac{\pi \left(t_j + \beta - \alpha \right)}{2p} \sum_{l=1}^{j-1} \cot \left(\pi \frac{(\alpha - \beta) + l\beta}{2p} \right) \\
&\leq \frac{t_j + \beta - \alpha}{\beta} \int_{\frac{\pi\alpha}{4p}}^{\pi x_j} \cot(u) \, \mathrm{d}u \\
&= \frac{t_j + \beta - \alpha}{\beta} \left(- \ln \sin \left(\frac{\pi\alpha}{4p} \right) + \ln \sin \left(\pi \tilde{x}_j \right) \right), \qquad (3.22)
\end{aligned}
$$

Following the same steps on Expression (3.22) leads to

$$
\begin{aligned}
\sum_{l=1}^{2p-j} g_\omega \left(1 - \frac{(\alpha - \beta) + l\beta}{2p} \right) &\leq \frac{\pi \left(2\beta - \alpha - t_j \right)}{2p} \sum_{l=1}^{2p-j} \cot \left(\pi \frac{(\alpha - \beta) + l\beta}{2p} \right) \\
&\leq \frac{2\beta - \alpha - t_j}{\beta} \int_{\frac{\pi\alpha}{4p}}^{\pi - \pi x_j} \cot(u) \, \mathrm{d}u \\
&= \frac{2\beta - \alpha - t_j}{\beta} \left(- \ln \sin \left(\frac{\pi\alpha}{4p} \right) + \ln \sin \left(\pi \tilde{x}_j \right) \right). \qquad (3.23)
\end{aligned}
$$

One can inject the bound provided by the inequalities (3.22) and (3.23) into (3.18), which yields

$$
\begin{aligned}
\frac{1}{2} z_0(\omega) &\leq \frac{3\beta - 2\alpha}{\beta} \left(- \ln \sin \left(\frac{\pi\alpha}{4p} \right) + \ln \sin \left(\pi \tilde{x}_j \right) \right) + \ln \sin \left(\pi \frac{\varepsilon}{2p} \right) - \ln \sin \left(\tilde{x}_j \right) \\
&\leq \frac{3\beta - 2\alpha}{\beta} \left(- \ln \left(\frac{\pi\alpha}{2p} \right) + \ln \sin \left(\pi \tilde{x}_j \right) \right) + \ln \sin \left(\pi \frac{\varepsilon}{2p} \right) - \ln \sin \left(\tilde{x}_j \right) \\
&\leq (3 - 2\alpha) \left(- \ln \left(\frac{\pi\alpha}{2p} \right) + \ln \sin \left(\pi \tilde{x}_j \right) \right) + \ln \sin \left(\pi \frac{\varepsilon}{2p} \right) - \ln \sin \left(\tilde{x}_j \right). \\
&= 2\left(1 - \alpha\right) \left(- \ln \left(\frac{\pi\alpha}{2p} \right) + \ln \sin \left(\pi \tilde{x}_j \right) \right) + \ln \sin \left(\pi \frac{\varepsilon}{2p} \right) - \ln \left(\frac{\pi\alpha}{2p} \right)
\end{aligned}
$$

Taking back to exponential yields the desired result.

$$Z_0(\omega) \leq \left(\frac{\pi\alpha}{m-1}\csc(\pi\tilde{x}_j)\right)^{4(\alpha-1)} \times \left(\frac{m-1}{\pi\alpha}\right)^2 \sin^2\left(\pi\frac{\varepsilon}{m-1}\right).$$

□

3.6.3.3 Proof of Lemma 3.2

Since the elements $\tilde{x}_l$ are constrained to belong to the subset S_l, and noticing that $\cot(\cdot)$ is a positive function on the intervals $\{S_1, \cdots, S_p\}$ and a negative function on $\{S_{p+1}, \cdots, S_{2p}\}$ one may provide the following bound on the quantity $\gamma_{\tilde{X}}$

$$\begin{aligned}
|\gamma_{\tilde{X}}| &\leq \sum_{l=1}^{p} \sup_{\tilde{x}_l\in S_l} \cot(\pi\tilde{x}_l) - \sum_{l=p+1}^{2p} \inf_{\tilde{x}_l\in S_l} \cot(\pi\tilde{x}_l) \\
&= \sum_{l=1}^{p} \cot\left(\pi\frac{\alpha-\beta+l\beta}{2p}\right) - \cot\left(\pi\frac{\alpha+l\beta}{2p}\right) \\
&= \cot\left(\pi\frac{\alpha}{2p}\right) - \cot\left(\frac{\pi}{2}\right) \\
&= \cot\left(\pi\frac{\alpha}{2p}\right).
\end{aligned}$$

Consequently, the root $r_{\tilde{X}}$ of the polynomial R verifies

$$\begin{aligned}
|r_{\tilde{X}}| &\geq \frac{1}{\pi}\operatorname{arccot}(\gamma_{\tilde{X}}) \\
&\geq \frac{\alpha}{2p},
\end{aligned}$$

and it comes that, for $\varepsilon \in [-1, 1]$

$$\begin{aligned}
R_0\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1}\right) &= \left(1+\gamma_{\tilde{X}}^2\right)\sin^2\left(\pi\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1} - r_{\tilde{X}}\right)\right) \\
&\leq \left(1+\gamma_{\tilde{X}}^2\right)\sin^2\left(\pi\left(\tilde{x}_j + \frac{\alpha\varepsilon}{m-1} - \frac{\alpha}{m-1}\right)\right) \\
&\leq \left(1+\gamma_{\tilde{X}}^2\right)\sin^2(\pi\tilde{x}_j) \\
&\leq \left(1+\cot^2\left(\pi\frac{\alpha}{m-1}\right)\right)\sin^2(\pi\tilde{x}_j) \\
&= \csc^2\left(\pi\frac{\alpha}{m-1}\right)\sin^2(\pi\tilde{x}_j).
\end{aligned}$$

□

Chapter 4

Dimensionality reduction for TV regularization

4.1 The partial line spectral estimation problem

4.1.1 Definitions and problematics

We consider the problem of estimating a sparse measure $\mu \in \mathcal{M}(\mathbb{T})$ following the spikes model (1.2) from r partial observations constructed linearly from the $2m+1 \geq r$ outputs of a uniform sampler. The partial observation vector $z \in \mathbb{C}^r$ is linked to the uniform measurements $y \in \mathbb{C}^{2m+1}$ by the linear relation $z = \mathbf{M}y$ whereby $\mathbf{M} \in \mathbb{C}^{r \times (2m+1)}$ is the *sub-sampling matrix* of the system, which is assumed to be known.

Similarly to the full measurement case described in Chapter 1, the *partial line spectral estimation problem* consists in finding the sparsest spectral measure $\mu \in \mathcal{M}(\mathbb{T})$ that matches the observations z. Due to the sub-sampling effect, the consistency constraint linking the observations z and the measure μ to recover is slightly different than the one presented before and writes

$$z = \int_{\mathbb{T}} \mathbf{M} a_m(\omega) \, \mathrm{d}\mu(\omega). \tag{4.1}$$

By analogy with the full measurement Program (4.2), the *optimal partial line spectral estimator* can be written as the output of an abstract minimization problem for the L_0 pseudo-norm defined in (1.6) over the set of Radon measures of the form

$$\begin{aligned} \mu_{\mathbf{M},0} = \arg \min_{\mu \in \mathcal{M}(\mathbb{T})} & \|\mu\|_0 \\ \text{subject to} \quad & z = \int_{\mathbb{T}} \mathbf{M} a_m(\omega) \, \mathrm{d}\mu(\omega). \end{aligned} \tag{4.2}$$

Program (4.2) is expected to be hard to solve for a generic matrix $\mathbf{M}$ due to the combinatorial search inherent to the minimization of the pseudo-norm $\|\cdot\|_0$. Therefore, one can naturally introduce the total variation counterpart to this problem in the same manner than what was presented in Section 1.6, yielding

$$\begin{aligned} \mu_{\mathbf{M},\mathrm{TV}} = \arg\min_{\mu \in \mathcal{M}(\mathbb{T})} &\left|\mu\right|(\mathbb{T}) \\ \text{subject to} \quad z = &\int_{\mathbb{T}} \mathbf{M} a_m(\omega)\, \mathrm{d}\mu(\omega). \end{aligned} \tag{4.3}$$

This chapter discusses two fundamental issues arising from the formulation of the convex formulation (4.3):

- *Sampling complexity:* How many measurements r do one need, and what properties must be verified by the sub-sampling $\mathbf{M}$ in order to guarantee that the output of the partial line spectral estimation problem $\mu_{\mathbf{M},\mathrm{TV}}$ is the same as in the full measurement case μ_{TV}?

- *Computational complexity*: Is there a convex algorithm returning the optimum μ_{TV} in a time depending only in the number of observations r and not on the initial dimension $2m+1$?

4.1.2 Lagrangian duality

It has been shown in [67] that the primal problem (4.3) admits for a Lagrange dual problem a certain semidefinite program when the sub-sampling matrix is a selection matrix $\mathbf{C}_{\mathcal{I}}$. This result easily extends in our context for any sub-sampling matrix $\mathbf{M}$ as stated by the following lemma, whose immediate proof has been omitted.

Lemma 4.1 (Partial dual characterization)**.** *The dual feasible set $\mathcal{D}_{\mathbf{M}}$ of Problem (4.3) is characterized by*

$$\mathcal{D}_{\mathbf{M}} = \left\{ v \in \mathbb{C}^r, \begin{cases} q = \mathbf{M}^* v \\ \left\|Q\right\|_\infty \leq 1 \end{cases} \right\},$$

whereby $Q \in \mathcal{T}_m$ is the complex polynomial having for coefficients vector $q \in \mathbb{C}^{2m+1}$. Moreover, the Lagrangian dual of Problem (4.3) is equivalent to the semidefinite program

$$\begin{aligned} (v_\star, \mathbf{H}_\star) = \arg&\max_{v \in \mathbb{C}^r} \Re(\langle v, z\rangle) \\ \textit{subject to} \quad &\begin{bmatrix} \mathbf{H} & q \\ q^* & 1 \end{bmatrix} \succeq 0 \\ &\mathcal{H}_m^*(\mathbf{H}) = e_0 \\ &q = \mathbf{M}^* v. \end{aligned} \tag{4.4}$$

The next proposition provides a counterpart to the dual certifiability Theorem 1.1 in the case of partial measurement, and can be seen as an extension to the previous result proposed in [67] for generic sub-sampling matrices.

Proposition 4.1 (Partial dual certifiability)**.** *If there exists a trigonometric polynomial*

$Q_\star \in \mathcal{T}_m$ with coefficients vector $q_\star \in \mathbb{C}^{2m+1}$ satisfying the conditions

$$\begin{cases} q_\star \in \operatorname{range}(\mathbf{M}^*) \\ Q_\star(x_k) = \operatorname{sign}(c_k), & \forall k \in [\![1, s]\!] \\ |Q_\star(\omega)| < 1, & \forall \omega \notin X \end{cases} \tag{4.5}$$

then the solutions of the Programs (1.5) and (4.3) are unique and one has $\mu_0 = \mu_{\mathbf{M},TV}$.

Proof. Any polynomial $Q_\star$ satisfying the conditions (1.30) maximizes the dual of Problem (1.28) over the feasible set $\mathcal{D}_{\mathbf{I}_n}$, and qualifies as a dual certificate of the same problem. Thus, the solution of Program (1.27) is unique and satisfies $\mu_0 = \mu_{\mathrm{TV}}$ [18]. By strong duality, the primal problem (1.27) and its dual reach the same optimal objective value, denoted $\kappa_\star$.

By the first condition of (4.5), $q_\star = \mathbf{M}^* v_\star$ for some $v_\star \in \mathbb{C}^r$. Since $v \in \mathcal{D}_{\mathbf{M}} \Leftrightarrow \mathbf{M}^* v \in \mathcal{D}_{I_n}$ for all $v \in \mathbb{C}^r$, $v_\star$ is dual optimal for the partial problem (4.3) and reaches the dual objective $\kappa_\star$. By strong duality, $\kappa_\star$ also minimize the primal objective of (4.3). Finally, every feasible point of (4.3) is feasible for (1.27). We conclude by the uniqueness of μ_{TV} on the equality $\mu_0 = \mu_{\mathrm{TV}} = \mu_{\mathbf{M},\mathrm{TV}}$. □

Any polynomial $Q_\star$ satisfying the conditions (4.5) will be called *dual certificate* for the partial line spectral estimation problem. Finding meaningful sufficient conditions for the existence of such dual certificate is a difficult problem in the general case. One might expect their existence under two main conditions: The support set X of the measure μ to reconstruct has to obey a minimal separability condition similar to the one stated in Conjecture 1.1 (for a potentially different constant); The sub-sampling matrix $\mathbf{M}$ has to preserve the geometric structures of the problem for an input y of the form (1.3).

Sufficient conditions for the existence of a dual certificate are recalled for some specific categories of matrices in Subsection 4.2.1. Chapter 5 will provide a certifiability result for a different class of operators. Generic certifiability results holding for arbitrary sub-sampling matrix $\mathbf{M}$ are still lacking and remain as of today an open area of research.

4.2 The sampling and computational complexities of TV regularization

4.2.1 Compressed sensing off-the-grid

The well explored theory of compressed sensing has shown that it is possible to reconstruct sparse *vectors* from relatively few linear measurements, provided that the observation matrix satisfies the restricted isometry property or some incoherence property [16], [20] or whenever the matrix is drawn at random [20]. Moreover, a consequent part of the literature focuses on establishing the minimal amount for measurements that is required for a random operator to reconstruct a sparse signal with high probability or asymptotically almost surely. This critical number of observations is often referred as the *sampling complexity* of the problem.

Recent advances were made to extend the theory of compressed sensing to inverse problems defined on the continuum [70], [71]. In particular, it is shown in [67] that the sampling complexity of *random selection sampling* for the partial line spectral estimation problem can be made logarithmic in the number of initial measurements $2m+1$ when the support set X to reconstruct is of fixed size. The formal statement of this result is recalled in the following.

Theorem 4.1 (Compressed sensing off the grid)**.** *Suppose that the observation set* $\mathcal{I} \subseteq [\![-m, m]\!]$ *of size* r *is uniformly selected at random. Assume that the complex phase* $\operatorname{sign}(c_k)$ *are drawn i.i.d. uniformly at random over the complex unit circle and let* $\delta > 0$*. If* $\Delta_{\mathbb{T}}(X) \geq \frac{2}{m}$*, then there exists a constant* C *such that*

$$r \geq C \max\left\{\log^2 \frac{m}{\delta}, s \log \frac{s}{\delta} \log \frac{m}{\delta}\right\}$$

suffices to guarantee that the output of Program (4.3) with associated subsampling matrix $\mathbf{M} = \mathbf{C}_{\mathcal{I}}$ *is equal to* μ_0 *with probability at least* $1-\delta$*.*

The assumptions of the previous results were broaden to the wider framework of subsampling operator with low coherence $\xi(\mathbf{M})$ discussed in Subsection 1.4.2 in [36].

4.2.2 Geometry of inverse problems and computational cost

The *computational complexity* of a problem refers to the asymptotic amount of time and memory that are requested by an algorithm in order to output the solution of a problem. If there is no difference between the sampling and the computational complexity in the finite dimensional inverse problem framework when seeking for a reconstruction using convex ℓ_1-minimization techniques, the picture is rather different when dealing with reconstruction over continuous spaces. Indeed, reducing the number of observations from an order of magnitude to another does not necessarily implies the existence of an algorithm that will output the desired solution in a time, or with memory usage, of the order of the reduced observations. In order to illustrate this fact, an analogy with the finite dimensional ℓ_1 minimization is drawn in the following.

Consider a finite dimensional inverse problem where one seeks to reconstruct a sparse vector x in dimension d through an observation matrix $\mathbf{M} \in \mathbb{C}^{r \times d}$. One can approach this classic inverse problem by minimizing the ℓ_1-norm among all possible x of the Euclidean space $\mathbb{C}^d$ [15], leading to a *linear program* of the form

$$\begin{aligned} x_\star = \arg\min_{x \in \mathbb{C}^d} \|x\|_1 \\ \text{subject to} \qquad z = \mathbf{M}x. \end{aligned} \tag{4.6}$$

Observing the above program through the lens of the Lagrangian duality theory yields the following linear program

$$
\begin{aligned}
q_\star = \arg\min_{q \in \mathbb{C}^d} & \; \Re \langle q, x \rangle \\
\text{subject to} \quad & \|q\|_\infty \le 1 \\
& \mathbf{M}^* q = 0.
\end{aligned} \tag{4.7}
$$

In geometric terms, the dual linear program (4.7) seeks to maximize a linear cost function over the intersection of a polyhedron in dimension d (the unitary cube in $\mathbb{C}^d$) and an hyperplane of dimension r driven by the span of the matrix $\mathbf{M}^*$. Since polyhedra are *closed under linear projections*, this intersection is itself a polyhedron of dimension r, and one is guaranteed that the output of Program (4.7) is equivalent to the following low dimensional program

$$
\begin{aligned}
u_\star = & \arg\min_{u \in \mathbb{C}^r} \Re \langle u, z \rangle \\
\text{s.t.} \quad & \|\mathbf{M}u\|_\infty \le 1
\end{aligned} \tag{4.8}
$$

which can be solved in a time driven by the actual number of samples r. Hence, in the presented finite-dimensional framework, the computational complexity is always equal to the sampling complexity.

However, the geometric picture is completely different when considering sparse reconstruction over infinite-dimensional spaces. The feasible set of the dual program (4.4) is given by the intersection of a spectrahedron – a set defined by linear matrix inequalities – with an hyperplane defined by the two linear constraints

$$
\begin{cases}
\mathcal{H}_m^* (\mathbf{H}) = e_0 \\
q = \mathbf{M}^* c.
\end{cases}
$$

However, spectrahedra are, in general, not closed under taking linear projection [8], [75], implying that one cannot directly guarantee that the dual program (4.4) can be reformulated as an SDP involving linear matrix inequalities of essential dimension driven by r. Hence, it is not possible to conclude that reducing the number of samples will lead to a reduction of the computational cost requested by the TV regularization to output an estimate of the ground truth measure.

4.3 Low-dimensional semidefinite representations

4.3.1 Main result

A first approach to solve the partial line spectral estimation is to compute the output $v_\star \in \mathbb{C}^r$ of the semidefinite program (4.4). One can then reconstruct the dual polynomial $Q_\star \in \mathcal{T}_m$, where $q_\star = \mathbf{M}^* v_\star$, and run Algorithm 1.1 on $q_\star$ in order to estimate the parameters of the measure μ. However, this method is not satisfactory on a computational point of view, since the SDP (4.4) involves linear matrix inequalities of dimension $2(m+1)$, while the *essential dimension* of partial recovery problem (4.3) is equal $r \leq 2(m+1)$.

We aim to study the existence of low-dimensional semidefinite programs that can output a similar dual polynomial $Q_\star \in \mathcal{T}_m$. It was discussed in Subsection 1.6.3 that the existence of such semidefinite representations was closely related to the Gram parametrization of trigonometric polynomials, who links the positive of polynomials with some sets of matrices. As stated in Lemma 4.1, in the partial line spectral estimation context, the trigonometric polynomials that need to be represented belong to a low dimensional structure characterized by the sub-sampling matrix $\mathbf{M} \in \mathbb{C}^{n \times m}$. Hence, finding *compact* Gram representations, involving matrices of small dimensions is of crucial interest for building an equivalence between the initial dual Problem (4.4) of dimension $2(m+1)$ with a lower dimensional semidefinite program.

We introduce in the following the definition of two new linear operators, that are involved in the novel dimensional reduction approach proposed in Theorem 4.2. For any matrix $\mathbf{M} \in \mathbb{C}^{r \times (2m+1)}$ of maximal rank $r \leq 2(m+1)$, we define the adjoint of the partial Toeplitz operator $\mathcal{R}^*_{\mathbf{M}}(\cdot)$ associated with the matrix $\mathbf{M} \in \mathbb{C}^{r \times 2(m+1)}$ by

$$\begin{aligned} \mathcal{R}^*_{\mathbf{M}} : \mathbb{C}^{r \times r} &\to \mathbb{C}^{m+1} \\ \mathbf{S} &\mapsto \mathcal{R}^*_{\mathbf{M}}(\mathbf{H}) = \mathcal{H}^*_m(\mathbf{M}^*\mathbf{S}\mathbf{M}), \end{aligned} \tag{4.9}$$

whereby $\mathcal{H}^*_m(\cdot)$ has been introduced in Section 1.2. Moreover, the matrix projection operator $\mathcal{P}_{\mathbf{M}}$, which projects orthogonally the rows and column of a matrix $\mathbf{H} \in \mathbb{C}^{(2m+1)\times(2m+1)}$ in the span of $\mathbf{M}^*$, is defined as follows

$$\begin{aligned} \mathcal{P}_{\mathbf{M}} : \mathbb{C}^{(2m+1)\times(2m+1)} &\to \mathbb{C}^{(2m+1)\times(2m+1)} \\ \mathbf{H} &\mapsto \mathcal{P}_{\mathbf{M}}(\mathbf{H}) = \mathbf{M}^*(\mathbf{M}\mathbf{M}^*)^{-1}\mathbf{M}\mathbf{H}\mathbf{M}^*(\mathbf{M}\mathbf{M}^*)^{-1}\mathbf{M}. \end{aligned} \tag{4.10}$$

The next theorem states the conditions under which the dual semidefinite program (4.4) is equivalent to a low-dimensional semidefinite program involving a matrix inequality of size $r+1$.

Theorem 4.2. *Let* $\mathbf{M} \in \mathbb{C}^{r \times (2m+1)}$ *be a matrix of maximal rank with* $r \leq 2(m+1)$. *Suppose that there exists a pair of solution* $(v_\star, \mathbf{H}_\star)$ *of Program (4.3) so that the matrix* $\mathbf{H}_\star$ *verifies the relation*

$$\mathcal{P}_{\mathbf{M}}(\mathbf{H}_\star) = \mathbf{H}_\star, \tag{4.11}$$

then Program (4.4) is equivalent to the low-dimensional semidefinite program

$$\begin{aligned} (v_\star, \mathbf{S}_\star) = \arg\max_{v \in \mathbb{C}^r} &\Re(\langle v, z \rangle) \\ \textit{subject to} \quad &\begin{bmatrix} \mathbf{S} & v \\ v^* & 1 \end{bmatrix} \succeq 0 \\ &\mathcal{R}^*_{\mathbf{M}}(\mathbf{S}) = e_0, \end{aligned} \tag{4.12}$$

in the sense that they reach the same optimal value $\kappa_\star$ *for at least one pair of solutions of* *the form* $(v_\star, \mathbf{H}_\star)$ *and* $(v_\star, \mathbf{S}_\star)$*, whereby*

$$\mathbf{S}_\star = (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M}\mathbf{H}_\star\mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1}.$$

Proof. Denote by $\kappa_\star$ and $\kappa'_\star$ the respective optimal values of Programs (4.4) and (4.12). We start the demonstration by verifying that if $(v, \mathbf{S})$ is in the feasible set of Program (4.12) then there exists a matrix $\mathbf{H}$ so that the pair $(v, \mathbf{H})$ is in the feasible set of Program (4.4). If $(v, \mathbf{S})$ is feasible for Program (4.12) one has, using a Schur complement argument

$$\begin{aligned} \mathbf{S} - vv^* \succeq 0 &\implies \mathbf{M}^* (\mathbf{S} - vv^*) \mathbf{M} \succeq 0 \\ &\implies \mathbf{M}^*\mathbf{S}\mathbf{M} - qq^* \succeq 0 \\ &\implies \begin{bmatrix} \mathbf{M}^*\mathbf{S}\mathbf{M} & q \\ q^* & 1 \end{bmatrix} \succeq 0. \end{aligned}$$

Moreover, by assumption, one has $\mathcal{R}^*_{\mathbf{M}}(\mathbf{S}) = e_0$ and it comes that

$$\begin{aligned} e_0 &= \mathcal{R}^*_{\mathbf{M}}(\mathbf{S}) \\ &= \mathcal{T}^*_m (\mathbf{M}^*\mathbf{S}\mathbf{M}). \end{aligned}$$

Hence the pair $(v, \mathbf{M}^*\mathbf{S}\mathbf{M})$ belongs to the feasible set of Program (4.4). Moreover, since the cost functions of the two programs are equal, one can conclude that $\kappa'_\star \leq \kappa_\star$.

Next, we show that if there exists a pair $(v_\star, \mathbf{H}_\star)$ solution of Program (4.4) verifying the linear constraint (4.11), then the pair $(v_\star, \mathbf{M}\mathbf{H}_\star\mathbf{M}_\star)$ is feasible for Problem (4.12). We start by verifying the inequality constraint. Since $\mathbf{H}_\star - q_\star q_\star^* \succeq 0$ is positive by assumption where $q_\star = \mathbf{M}^* v$, it yields

$$\begin{aligned} \mathbf{H}_\star - q_\star q_\star^* \succeq 0 &\implies \mathbf{H}_\star - \mathbf{M}^* v_\star v_\star^* \mathbf{M} \succeq 0 \\ &\implies \mathcal{P}_{\mathbf{M}}(\mathbf{H}_\star) - \mathbf{M}^* v_\star v_\star^* \mathbf{M} \succeq 0 \\ &\implies \mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M}\mathbf{H}\mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M} - \mathbf{M}^* v_\star v_\star^* \mathbf{M} \succeq 0 \\ &\implies \mathbf{M}^* (\mathbf{S}_\star - v_\star v_\star^*) \mathbf{M} \succeq 0 \\ &\implies \mathbf{S}_\star - v_\star v_\star^* \succeq 0 \\ &\implies \begin{bmatrix} \mathbf{S}_\star & v_\star \\ v_\star^* & 1 \end{bmatrix} \succeq 0. \end{aligned}$$

whereby we used the fact that $\mathbf{M}^*$ has maximal rank in the fifth implication. It remains to show that $\mathbf{S}_\star$ satisfies the linear constraint $\mathcal{R}^*_{\mathbf{M}}(\mathbf{S}_\star) = e_0$. One has

$$\begin{aligned} e_0 &= \mathcal{T}^*_m (\mathbf{H}_\star) \\ &= \mathcal{T}^*_m (\mathcal{P}_{\mathbf{M}}(\mathbf{H}_\star)) \\ &= \mathcal{T}^*_m \left(\mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M}\mathbf{H}_\star\mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M}\right) \\ &= \mathcal{T}^*_m (\mathbf{M}^*\mathbf{S}_\star\mathbf{M}) \\ &= \mathcal{R}^*_{\mathbf{M}}(\mathbf{S}_\star), \end{aligned}$$

which certifies that the pair $(v_\star, \mathbf{M}\mathbf{H}_\star\mathbf{M}^*)$ is feasible for Program (4.12). Finally, the objective of Program (4.12) reaches a value $\kappa_\star$ at point $(v_\star, \mathbf{M}\mathbf{H}_\star\mathbf{M}^*)$ hence $\kappa'_\star \geq \kappa_\star$. It comes that $\kappa'_\star = \kappa_\star$, and therefore that $(v_\star, \mathbf{M}\mathbf{H}_\star\mathbf{M}^*)$ is optimal for Program (4.12), concluding the proof of the theorem. □

The linear constraint $\mathcal{R}^*_{\mathbf{M}}(\mathbf{S}) = e_0$ of Program (4.12) has an explicit dimension that is still equal to $2m+1$. However, one can restrict its definition to the span of $\mathcal{R}^*_{\mathbf{M}}(\mathbf{S})$ which is of dimension at most equal to $\min\left\{2m+1, r^2\right\}$. An explicit characterization of this constraint is provided in Section 4.3.3 whenever $\mathbf{M} = \mathbf{C}_{\mathcal{I}}$ is a selection matrix.

4.3.2 The sparse Fejér-Riesz condition

One need to ensure the existence of dual optimal pair $(v_\star, \mathbf{H}_\star)$ of Program (4.4) with $\mathbf{H}_\star$ lying on the span of the projection $\mathcal{P}_{\mathbf{M}}$ in order to conclude on the tightness of the low dimension Program (4.12). It is shown in this subsection that this problem shares some common roots with the Gram parametrization theory and presented in Subsection 1.6.3.

A novel criterion called sparse Fejér-Riesz condition is introduced in Definition 4.1. Moreover Proposition 4.2 gives conditions on a dual certificate to ensure that the hypothesis of Theorem 4.2 are met, and guarantying the tightness and the uniqueness of the solution of the low-dimensional Program (4.12).

Definition 4.1 (Sparse Fejér-Riesz condition)**.** Let $\mathbf{M} \in \mathbb{C}^{r\times(2m+1)}$ be a matrix of maximal rank with $r \leq 2(m+1)$. A positive trigonometric polynomial $R \in \mathcal{T}_{2m}$ is said to satisfy the *sparse Fejér-Riesz condition* for the matrix $\mathbf{M}$ if there exists $\{P_j\}_{j=1}^{l} \subset \mathcal{T}_m$ for some $l \in \mathbb{N}$ with respective coefficients vectors $\{p_j\}_{j=1}^{l} \subset \mathbb{C}^{2m+1}$ verifying

$$\forall j \in [\![1, l]\!], \quad p_j \in \operatorname{range}(\mathbf{M}^*)$$

such that R can be decomposed as the sum of squares

$$\forall \omega \in \mathbb{T}, \quad R(\omega) = \sum_{j=1}^{l} |P_j(\omega)|^2 . \tag{4.13}$$

Proposition 4.2 (Low-dimensional dual certifiability)**.** *Let* $\mathbf{M} \in \mathbb{C}^{r\times(2m+1)}$ *be a matrix of maximal rank with* $r \leq 2(m+1)$. *If there exists a trigonometric polynomial* $Q_\star \in \mathcal{T}_m$ *with coefficients vector* $q_\star \in \mathbb{C}^{2m+1}$ *satisfying the conditions (4.5) and such that the trigonometric polynomial* $R_\star \in \mathcal{T}_{2m}$ *defined by*

$$\forall \omega \in \mathbb{T}, \quad R_\star(\omega) = 1 - |Q_\star(\omega)|^2 ,$$

verifies the sparse Fejér-Riesz condition for the matrix $\mathbf{M}$ *then:*

- *The solutions of the Programs (1.5) and (4.3) are unique and one has* $\mu_0 = \mu_{\mathbf{M},TV}$.
- *The vector* $q_\star$ *can be written has* $q_\star = \mathbf{M}^* v_\star$ *whereby* $v_\star \in \mathbb{C}^r$ *is a solution of the low-dimensional semidefinite program (4.12).*

Proof. The proof of the tightness of the primal problem is an immediate consequence of Proposition 4.1. It remains to show the existence of a pair $(q_\star, \mathbf{H}_\star)$ that is solution of the low-dimensional semidefinite program (4.12) with $\mathbf{H}_\star \in \mathbb{C}^{(2m+1)\times(2m+1)}$ verifying the relation (4.11). The conclusion follows via a direct application of Theorem 4.2. Since $R_\star$ verifies the sparse Fejér-Riesz condition, one can write

$$\begin{aligned}\forall \omega \in \mathbb{T}, \quad 1 - |Q_\star(\omega)|^2 &= R_\star(\omega) \\ &= \sum_{j=1}^{l} |P_j(\omega)|^2,\end{aligned}$$

for some $\{P_j\}_{j=1}^{l} \subset \mathcal{T}_m$ with respective coefficients vectors $p_j = \mathbf{M}^* u_j$ whereby $u_j \in \mathbb{C}^r$ for all $j \in [\![l]\!]$. Noticing that $|P(-\omega)|^2 = a_m(\omega)\, pp^* a_m(\omega)$ one has that

$$\begin{aligned}\forall \omega \in \mathbb{T}, \quad 1 &= |Q_\star(-\omega)|^2 + \sum_{j=1}^{l} |P_j(-\omega)|^2 \\ &= a_m(\omega)^* \left(q_\star q_\star^* + \sum_{j=1}^{l} p_j p_j^* \right) a_m(\omega) \\ &= a_m(\omega)^* \mathbf{M}^* \left(v_\star v_\star^* + \sum_{j=1}^{l} u_j u_j^* \right) \mathbf{M} a_m(\omega) \\ &= a_m(\omega)^* \mathbf{M}^* \mathbf{S}_\star \mathbf{M} a_m(\omega),\end{aligned}$$

whereby $\mathbf{S}_\star = \left(v_\star v_\star^* + \sum_{j=1}^{l} u_j u_j^* \right)$. By Proposition 1.2 the matrix

$$\mathbf{H}_\star = \mathbf{M}^* \mathbf{S}_\star \mathbf{M} = q_\star q_\star^* + \sum_{j=1}^{l} p_j p_j^* \tag{4.14}$$

verifies the condition $\mathcal{T}_m(\mathbf{H}) = e_0$ and by construction one has

$$\mathbf{H}_\star - q_\star q_\star^* = \sum_{j=1}^{l} p_j p_j^* \succeq 0,$$

concluding on the fact that the construction pair $(v_\star, \mathbf{H}_\star)$ is in optimal solution of Program (4.4).

Finally, it remains to verify that $\mathbf{H}_\star$ verifies the condition $\mathcal{P}_\mathbf{M}(\mathbf{H}_\star) = \mathbf{H}_\star$ in order to conclude. This is immediate since

$$\begin{aligned}\mathcal{P}_\mathbf{M}(\mathbf{H}_\star) &= \mathcal{P}_\mathbf{M}(\mathbf{M}^* \mathbf{S}_\star \mathbf{M}) \\ &= \mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M} (\mathbf{M}^* \mathbf{S}_\star \mathbf{M}) \mathbf{M}^* (\mathbf{M}\mathbf{M}^*)^{-1} \mathbf{M} \\ &= \mathbf{M}^* \mathbf{S}_\star \mathbf{M} \\ &= \mathbf{H}_\star.\end{aligned}$$

□

It remains, as of today, very challenging to understand the conditions for the existence of a trigonometric polynomial verifying the conditions of Proposition 4.2. However experimental results show that their existence should be verified with high probability when considering a random selection subsampling scheme discussed in Subsection 4.2.1 if the number of retained measurement r is large enough, and provided a large enough separation condition. Figure 4.1 compares the performances of the low-dimensional SDP (4.12) with the output of the full size Programs (1.34) and (4.4).

4.3.3 Characterization for the case of selection sub-sampling matrices

Selection matrices constitutes a interesting type of sub-sampling matrices, that arises in many practical applications. They occur naturally in signal processing when dealing with sampling models with missing entries. In this section, we highlight fundamental properties of the partial line spectrum estimation problem from selection based sub-sampling. For convenience, we shorten the notation $\mathcal{R}_{\mathcal{I}} = \mathcal{R}_{\mathbf{C}_{\mathcal{I}}}$ and $\mathcal{R}^*_{\mathcal{I}} = \mathcal{R}^*_{\mathbf{C}_{\mathcal{I}}}$ of a selection matrix $\mathbf{C}_{\mathcal{I}}$. We start by recasting the linear constraint $\mathcal{R}^*_{\mathcal{I}}(\mathbf{S}) = r$ into a more friendly set of equations.

Proposition 4.3 (Properties of the linear constraint for selection matrices)**.** *Let $\mathcal{I} \subset [\![-m, m]\!]$ be a subset of cardinality r and consider any selection matrix $\mathbf{C}_{\mathcal{I}} \in \mathbb{C}^{r\times(2m+1)}$ for this subset. Define by $\mathcal{J}$ the set of its pairwise differences $\mathcal{J} = \mathcal{I} - \mathcal{I} \subseteq [\![-2m, 2m]\!]$, and by $\mathcal{J}_+ = \{j \in \mathcal{J},\ j \geq 0\}$ its positive elements. There exists a skew-symmetric partition of the set $[\![1, r]\!]^2$ into $p = |\mathcal{J}_+|$ subsets $\{\mathbf{J}_k,\ k \in \mathcal{J}_+\}$ given by the support of the matrices $\{\mathbf{C}_{\mathcal{I}}\mathbf{\Theta}_k\mathbf{C}^*_{\mathcal{I}}\}_{k\in\mathcal{J}_+}$ satisfying*

$$\begin{cases} \mathbf{J}_k \cap \mathbf{J}_l = \emptyset, & \forall\,(k,l) \in \mathcal{J}_+^2,\ k \neq l, \\ (i,j) \in \bigcup_{k\in\mathcal{J}_+} \mathbf{J}_k \Leftrightarrow (j,i) \notin \bigcup_{k\in\mathcal{J}_+} \mathbf{J}_k, & \forall\,(i,j) \in [\![-m,m]\!]^2,\ i \neq j, \\ (i,i) \in \bigcup_{k\in\mathcal{J}_+} \mathbf{J}_k, & \forall i \in [\![1,m]\!], \end{cases}$$

such that for every Hermitian matrices $\mathbf{S} \in \mathbb{C}^{r\times r}$

$$\mathcal{R}^*_{\mathcal{I}}(\mathbf{S}) = \sum_{k\in\mathcal{J}_+} \left(\sum_{(l,r)\in J_k} \mathbf{S}_{l,r} \right) e_k, \tag{4.15}$$

whereby $e_k \in \mathbb{C}^{(2m+1)}$ is the k^{th} vector of the canonical basis indexed in $[\![-m, m]\!]$.

Proof. Using the adjoint decomposition of the operator $\mathcal{R}^*_{\mathcal{I}}$ on the canonical basis one has, for every Hermitian matrices $\mathbf{S} \in \mathbb{C}^{r\times r}$

$$\begin{aligned} \mathcal{R}^*_{\mathcal{I}}(\mathbf{S}) &= \sum_{k=-m}^{m} \langle \mathcal{R}_{\mathcal{I}}(e_k), \mathbf{S} \rangle e_k \\ &= \sum_{k=-m}^{m} \langle \mathbf{C}_{\mathcal{I}}\mathbf{\Theta}_k\mathbf{C}^*_{\mathcal{I}}, \mathbf{S} \rangle e_k. \end{aligned} \tag{4.16}$$

Let by $\mathbf{M}_k \in \mathbb{C}^{r\times r}$ the matrix given by $\mathbf{M}_k = \mathbf{C}_{\mathcal{I}}\mathbf{\Theta}_k\mathbf{C}^*_{\mathcal{I}}$ for all $k \in [\![-m, m]\!]$. It remains to show that the support of the matrices $\{\mathbf{M}_k\}_{k\in[\![-m,m]\!]}$ are forming the desired partition.

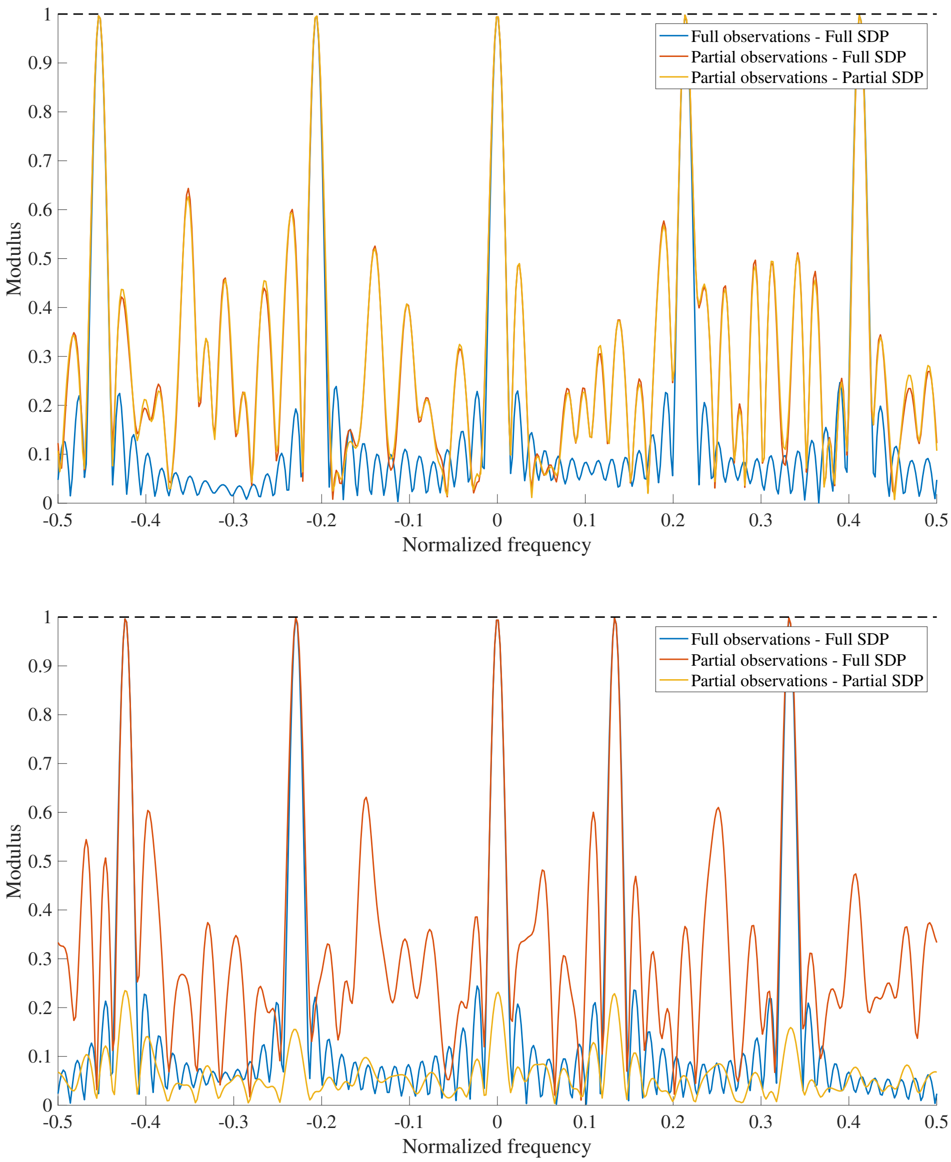

Figure 4.1: Comparison of the outputs of the dual program (1.34) under full observations (in blue), the full size dual program (4.4) under partial observations (in red), and the low-dimension dual program (4.12) (in yellow). The trigonometric order is fixed to $m = 32$ and two different subsampling ratios are considered.
Top: When keeping 70% of the measurements, the three program return a dual polynomial locating the frequencies.
Bottom: When keeping 50% of the measurements, the sparse certificate solution of the full size program can still locate the frequencies of the measure to estimate, while the sparse Fejér-Riesz certificate, solution of the low dimension program (4.12) fails to reach a modulus equal to one at the correct frequencies.

The general term of matrix $\mathbf{M}_k$, obtained by direct calculation, reads

$$\forall (i,j) \in [\![1,m]\!]^2, \quad \mathbf{M}_k(i,j) = \begin{cases} 1 & \text{if } \mathcal{I}[j] - \mathcal{I}[i] = k \\ 0 & \text{otherwise,} \end{cases} \tag{4.17}$$

for all $k \in [\![-m,m]\!]$, whereby $\mathcal{I}[j]$ represents the j^{th} element of the index set $\mathcal{I}$ for the ordering induced by the matrix $\mathbf{C}_{\mathcal{I}}$. The general term (4.17) ensures that

$$\begin{cases} \mathbf{M}_0(i,i) = 1, & \forall i \in [\![1,r]\!] \\ \sum_{k=-m}^{m} \mathbf{M}_k(i,j) = 1 \Leftrightarrow \sum_{k=-m}^{m} \mathbf{M}_k(j,i) = 0, & \forall (i,j) \in [\![1,m]\!]^2, \ i \neq j, \\ k \notin \mathcal{J}_+ \Leftrightarrow \mathbf{M}_k = \mathbf{0}_r, & \forall k \in [\![-m,m]\!], \end{cases}$$

where $\mathbf{0}_r$ is the null square matrix in dimension r. Since the matrices $\{\mathbf{M}_k\}_{k\in[\![-m,m]\!]}$ have boolean entries, the two first assertions yields the set of supports $\{\mathbf{J}_k\}_{k\in[\![-m,m]\!]}$ of $\{\mathbf{M}_k\}_{k\in[\![-m,m]\!]}$ forms a skew-symmetric partition of $[\![1,r]\!]^2$. The third one states that only $p = |\mathcal{J}_+|$ elements of this partition are non-trivial. After removing those null matrices, the set $\{\mathbf{J}_k\}_{k\in\mathcal{J}_+}$ remains a partition of $[\![1,r]\!]^2$. We conclude using Equation (4.16) that, for every Hermitian matrices $\mathbf{S} \in \mathbb{C}^{r\times r}$,

$$\begin{aligned} \mathcal{R}_{\mathcal{I}}^{*}(\mathbf{S}) &= \sum_{k\in\mathcal{J}_+} \langle \mathbf{M}_k, \mathbf{S} \rangle e_k \\ &= \sum_{k\in\mathcal{J}_+} \left(\sum_{(l,r)\in\mathbf{J}_k} \mathbf{S}_{l,r} \right) e_k. \end{aligned}$$

□

This proposition highlights several major properties of the equation $\mathcal{R}_{\mathcal{I}}^{*}(\mathbf{S}) = u$:

- The linear equation is solvable if and only if u is supported on the set $\mathcal{J}_+$, and since $\mathbf{M}_0 = I_r$ the component u_0 has to be real.
- If so, the equation is equivalent to solve $p = |\mathcal{J}_+|$ linear forms. Those p forms are independent one from the other in the sense that they are acting on disjoint extractions of the matrix $\mathbf{S}$.
- The order of each of those forms is smaller that r, i.e., each form involves at most r terms of $\mathbf{S}$.
- The total number of unknowns appearing in this system is exactly $\frac{r(r+1)}{2}$.

In Section 4.4, a highly scalable algorithm to solve the semidefinite program (4.12) for selection matrices, taking advantage of the hereby presented properties, will be presented.

4.4 Acceleration via the alternating direction method of multipliers

4.4.1 Interior point methods and ADMM

Computing the solution of semidefinite program (4.18) using out of the box SDP solvers such as SuDeMi [63] or SDPT3 [69] requires at most $\mathcal{O}\left(\left(d_{\text{lmi}}^2 + d_{\text{lin}}\right)^{3.5}\right)$ operations where d_{lmi} is the dimension of the linear matrix inequality, and d_{lin} the dimension of the linear constraints. For the dual-AST program (1.37) one has, $d_{\text{lmi}} = r + 1$ and $d_{\text{lin}} \leq \frac{r(r+1)}{2}$, and approaching the optimal dual solution will cost $\mathcal{O}\left(r^7\right)$ operations using those interior point methods. It appears to be unrealistic to recover a sparse line spectrum that way when the number of observations exceeds a few hundreds.

In the same spirit than in [64], we derive the steps and update equations to approach the optimal solution via the alternating direction method of multipliers (ADMM). Unlike the original work, we choose to perform ADMM on the dual space instead of the primal one, and adjust the update steps in order to take advantage of the low dimensionality of (4.12). The overall idea of this algorithm is to cut the augmented Lagrangian of the problem into a sum of separable sub-functions. Each iteration consists in performing independent local minimization on each of those quantities. The interested reader can find a detailed survey of this method in [11].

We restrict our analysis to the case of partially observed systems where the subsampling matrix is a selection matrix $\mathbf{C}_{\mathcal{I}} \in \{0,1\}^{r \times (2m+1)}$ for some subset $\mathcal{I} \subseteq [\![-m, m]\!]$ of cardinality r. We will see that the properties of such matrices detailed in Section 4.3.3 will help breaking down the iterative steps of dual ADMM on an elegant manner. Before any further analysis, the low dimensional dual-AST has to be restated into a more friendly form to derive the ADMM update equations. In our approach, we propose the following augmented formulation

$$\begin{aligned} v_\star = \arg\min_{c \in \mathbb{C}^r} & -\Re\left(\langle v, z\rangle\right) + \frac{\tau}{2}\left\|v\right\|_2^2 \\ \text{subject to} \quad & \mathbf{Z} \succeq 0 \\ & \mathbf{Z} = \begin{bmatrix} \mathbf{S} & v \\ v^* & 1 \end{bmatrix} \\ & \sum_{(i,j) \in J_k} \mathbf{S}_{i,j} = \delta_k, \quad k \in \mathcal{J}_+. \end{aligned} \tag{4.18}$$

It is immediate, using Proposition 4.3, to verify that Problems (4.12) and (4.18) are actually equivalent.

4.4.2 Lagrangian separability

We denote by L the restricted Lagrangian of the Problem (4.18), obtained by ignoring the semidefinite constraint $\mathbf{Z} \succeq 0$. In order to ensure plain differentiability with respect to the variables $\mathbf{S}$ and $\mathbf{Z}$, ADMM seeks to minimize an augmented version L_+ of L, with respect to the semidefinite inequality constraint that was put apart. This augmented Lagrangian

L_+ is introduced as follows

$$L_+(\mathbf{P},\mathbf{S},v,\mathbf{\Lambda},\gamma) = L(\mathbf{P},\mathbf{S},v,\mathbf{\Lambda},\gamma) + \frac{\rho}{2}\left\|\mathbf{P} - \begin{bmatrix}\mathbf{S} & v\\ v^* & 1\end{bmatrix}\right\|_F^2 + \frac{\rho}{2}\sum_{k\in\mathcal{J}_+}\left(\sum_{(i,j)\in J_k}\mathbf{S}_{i,j} - \delta_k\right)^2,$$

whereby the variable $\mathbf{\Lambda}$ is an Hermitian matrix of dimension $r+1$ and $\gamma \in \mathbb{C}^{|\mathcal{J}_+|}$ denote respectively the Lagrange multipliers associated with the first and the second equality constraints of Problem (4.18). The regularizing parameter $\rho > 0$ is set to ensure a well conditioned differentiability and to fasten the convergence speed of the alternating minimization towards the global optimum of the cost function L_+. For clarity and convenience, the following decompositions of the parameters $\mathbf{P}$ and $\mathbf{\Lambda}$ are introduced

$$\mathbf{P} = \begin{bmatrix}\mathbf{P}_0 & p\\ p^* & \zeta\end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix}\mathbf{\Lambda}_0 & \lambda\\ \lambda^* & \eta\end{bmatrix}.$$

Moreover, for any square matrix $\mathbf{A} \in \mathbb{C}^{r\times r}$, we let by $\mathbf{A}_{J_k} \in \mathbb{C}^{|J_k|}$ the vector constituted of the terms $\{\mathbf{A}_{i,j},\ (i,j) \in J_k\}$. The order in which the elements of J_k are extracted and placed in this vector has no importance, as long as, once chosen, it remains the same for every matrix $\mathbf{A}$. This allows to decompose the augmented Lagrangian into

$$L_+(\mathbf{P},\mathbf{S},v,\mathbf{\Lambda},\gamma) = L_v(p,v,\lambda) + L_\gamma(\zeta,\eta) + \sum_{k\in\mathcal{J}_+} L_k\left(\mathbf{P}_{0,J_k},\mathbf{S}_{J_k},\mathbf{\Lambda}_{0,J_k}\right),$$

whereby each of the sub-functions reads

$$\begin{aligned}
L_v(p,v,\lambda) &= -\Re\left(z^\mathsf{T} v\right) + \frac{\tau}{2}\|v\|_2^2 + 2\langle\lambda, p-v\rangle + \rho\|p-v\|_2^2\\
L_\gamma(\zeta,\eta) &= \langle\eta,\zeta-1\rangle + \frac{\rho}{2}(\zeta-1)^2\\
\forall k\in\mathcal{J}_+,\quad L_k\left(\mathbf{P}_{0,J_k},\mathbf{S}_{J_k},\mathbf{\Lambda}_{0,J_k}\right) &= \langle\mathbf{\Lambda}_{0,J_k},\mathbf{P}_{0,J_k}-\mathbf{S}_{J_k}\rangle + \gamma_k\left(\sum_{(i,j)\in J_k}\mathbf{S}_{i,j}-\delta_k\right)\\
&\quad + \frac{\rho}{2}\|\mathbf{P}_{0,J_k}-\mathbf{S}_{J_k}\|_2^2 + \frac{\rho}{2}\left(\sum_{(i,j)\in J_k}\mathbf{S}_{i,j}-\delta_k\right)^2.
\end{aligned}$$

4.4.3 Update rules

The ADMM will consist in successively performing the following decoupled update steps:

$$\begin{aligned}
v^{t+1} &\leftarrow \arg\min_v L_v\left(z^t, v, \lambda^t\right)\\
\forall k\in\mathcal{J}_+,\quad \mathbf{S}_{J_k}^{t+1} &\leftarrow \arg\min_{S_{J_k}} L_k\left(\mathbf{P}_{0,J_k}^t,\mathbf{S}_{J_k},\mathbf{\Lambda}_{0,J_k}^t\right)\\
\mathbf{S}_{j,i}^{t+1} &\leftarrow \overline{\mathbf{S}_{i,j}^{t+1}},\quad \forall (i,j)\in\bigcup_{k\in\mathcal{J}_+} J_k
\end{aligned}$$

$$\mathbf{P}^{t+1} \leftarrow \arg\min_{\mathbf{P}\succeq 0} L_+\left(\mathbf{P}, \mathbf{S}^{t+1}, v^{t+1}, \mathbf{\Lambda}^t, \mu^t\right)$$

$$\mathbf{\Lambda}^{t+1} \leftarrow \mathbf{\Lambda}^t + \rho\left(\mathbf{P}^{t+1} - \begin{bmatrix} \mathbf{S}^{t+1} & v^{t+1} \\ v^{t+1^*} & 1 \end{bmatrix}\right)$$

$$\forall k \in \mathcal{J}_+, \quad \gamma_k^{t+1} \leftarrow \gamma_k^t + \rho\left(\sum_{(i,j)\in J_k} \mathbf{S}_{i,j}^{t+1} - \delta_k\right).$$

Since the linear constraint $\mathcal{R}_{\mathcal{I}}^*(\mathbf{S}) = e_0$ has an effect limited to the subspace $\{\mathcal{R}_{\mathcal{I}}(e_k)\}_{k\in\mathcal{J}_+}$, the third update step is necessary to maintain the Hermitian structure of the matrix $\mathbf{S}^{t+1}$ at every iteration. The update steps for the variables v^{t+1} and $\left\{\mathbf{S}_{J_k}^{t+1}\right\}_{k\in\mathcal{J}_+}$ are performed at each iteration by canceling the gradient of their partial augmented Lagrangian and admit, in the presented settings, closed form expressions given by

$$v^{t+1} = \frac{1}{2\rho+\tau}\left(\bar{z} + 2\rho p^t + 2\lambda^t\right)$$

$$\forall k \in \mathcal{J}_+, \quad \mathbf{S}_{J_k}^{t+1} = \left(\mathbf{P}_0^t + \frac{1}{\rho}\mathbf{\Lambda}_0^t\right)_{J_k} - \left(\sum_{(i,j)\in J_k}\left(\mathbf{P}_0^t + \frac{\mathbf{\Lambda}_0^t}{\rho}\right)_{i,j} - \left(\delta_k - \frac{\gamma_k^t}{\rho}\right)\right) j_{|J_k|}$$

whereby j_n is the all-one vector of $\mathbb{C}^n$. The update $\mathbf{P}^{t+1}$ reads at the t^{th} iteration

$$\mathbf{P}^{t+1} \in \arg\min_{Z\succeq 0} \left\|\mathbf{P} - \mathbf{Y}^t\right\|_F^2$$

$$\mathbf{Y}^t = \begin{bmatrix} \mathbf{S}^{t+1} & v^{t+1} \\ v^{t+1^*} & 1 \end{bmatrix} - \frac{\mathbf{\Lambda}^t}{\rho},$$

which can be interpreted as an orthogonal projection of $\mathbf{Y}^t$ onto the cone of positive Hermitian matrices in dimension $r+1$ for the Frobenius inner product. This projection can be computed by looking for the eigenpairs of $\mathbf{Y}^t$, and setting all negative eigenvalues to 0. More precisely, denoting $\mathbf{Y}^t = \mathbf{V}^t\mathbf{D}^t\,\mathbf{V}^{t\,*}$ an eigen-decomposition of $\mathbf{Y}^t$, one get $\mathbf{Z}^{t+1} = \mathbf{V}^t\mathbf{D}_+^t\,\mathbf{V}^{t\,*}$ where $\mathbf{D}_+^t$ is a diagonal matrix whose j^{th} diagonal entry $d_+^t[j]$ satisfies $d_+^t[j] = \max\left\{d^t[j], 0\right\}$.

4.4.4 Computational complexity

On the computational point of view, at each step of ADMM, the update v^{t+1} is a vector addition and performed in a linear time $\mathcal{O}(r)$. On every extractions $\mathbf{S}_{J_k}^{t+1}$ of $\mathbf{S}^{t+1}$, the update equation is assimilated to a vector averaging requiring $\mathcal{O}(|J_k|)$ operations when firstly calculating the common second term of the addition. Since $\bigcup_{k\in\mathcal{J}_+} J_k = \frac{r(r+1)}{2}$, we conclude that the global update of the matrix $\mathbf{S}^{t+1}$ is done in $\mathcal{O}(r^2)$. The update of $\mathbf{Z}^{t+1}$ requires the computation of its spectrum, which can be done in $\mathcal{O}(r^3)$ via power method. Finally updating the Lagrange multipliers $\mathbf{\Lambda}^{t+1}$ and γ^{t+1} consist in simple matrix and vector additions, thus of order $\mathcal{O}(r^2)$.

To summarize, the projection is the most costly operation of the loop. Each step of ADMM method runs in $\mathcal{O}(r^3)$ operations, which is a significant improvement compared to the infeasible path approached used by SDP solvers requiring around $\mathcal{O}\left(r^7\right)$ operations.

Chapter 5

Spectral estimation for multirate sampling systems

5.1 Introduction to multirate sampling

5.1.1 Observation model

In this section, we consider the joint reconstruction problem of a continuous time signal $x(\cdot)$ whose continuous Fourier spectrum is sparse on *the real line* from multiple synchronized samplers. The generalized continuous Fourier transform $\tilde{\mu} \in \mathcal{M}(\mathbb{R})$ of the time signal $x(\cdot)$ is a Radon measure over the real line. The two quantities are linked by the linear relation

$$\forall t \in \mathbb{R}, \quad x(t) = \int_{-\infty}^{\infty} e^{i2\pi\xi t} \mathrm{d}\tilde{\mu}(\xi). \tag{5.1}$$

A multirate sampling system (MRSS) acting on a continuous time signal $x(\cdot)$ is defined by a set $\mathbb{A}$ of p distinct grids (or samplers) $\mathcal{A}_j$, $j \in [\![1, p]\!]$. Each grid is assimilated to a triplet $\mathcal{A}_j = (f_j, \gamma_j, n_j)$, where $f_j \in \mathbb{R}^+$ is its sampling frequency, $\gamma_j \in \mathbb{R}$ is its processing delay, expressed in sample unit for normalization purposes, and $n_j \in \mathbb{N}$ the number of measurements acquired by the grid. We assume those intrinsic characteristics to be known. As a result, the output $y_j \in \mathbb{C}^{n_j}$ of the grid $\mathcal{A}_j$ sampling a complex time signal x reads

$$\forall j \in [\![1, p]\!], \quad y_j = \mathcal{L}_j(\tilde{\mu}), \tag{5.2}$$

whereby $\mathcal{L}_j$ is the linear integral operator defining the effect of the generalized Fourier spectrum on the measurement of the j^{th} grid, and is given by

$$\begin{aligned} \mathcal{L}_j : \mathcal{M}(\mathbb{R}) &\to \mathbb{C}^{n_j} \\ \tilde{\mu} &\mapsto \int_{-\infty}^{\infty} \tilde{a}_j(\xi)\, \mathrm{d}\tilde{\mu}(\xi), \end{aligned} \tag{5.3}$$

whereby the *atomic vectors* $\tilde{a}_j(\cdot) \in \mathbb{C}^{n_j}$ associated with the j^{th} array are given by

$$\forall r \in [\![0, n_j - 1]\!], \forall \xi \in \mathbb{R}, \quad \tilde{a}_j(\xi)[l] = e^{i2\pi \frac{\xi}{f_j}(r - \gamma_j)}. \tag{5.4}$$

For ease of understanding, and calculations, the operators $\mathcal{L}_j$ can be expended as a composition of two elementary operations as follows

$$\forall j \in [\![1,p]\!]\,, \quad \mathcal{L}_j = \mathcal{F}_{n,f_j} \circ \mathcal{S}_{\frac{\gamma_j}{f_j}}\,, \tag{5.5}$$

where $\mathcal{F}_{n,f}$ is the adjoint of the discrete time Fourier transform on $n \in \mathbb{N}$ data points and at the sampling frequency $f \in \mathbb{R}^+$, defined by

$$\begin{aligned} \mathcal{F}_{n,f} : \mathcal{M}(\mathbb{R}) &\to \mathbb{C}^n \\ \mu &\mapsto y = \int_{-\infty}^{\infty} \tilde{a}_{n,f}(\xi)\, \mathrm{d}\tilde{\mu}(\xi) \end{aligned} \tag{5.6}$$

for atoms for the form $\tilde{a}_{n,f}(\xi) = \left[1, e^{i2\pi\frac{\xi}{f}}, e^{i2\pi\frac{2\xi}{f}}, \cdots, e^{i2\pi\frac{(n-1)\xi}{f}}\right] \in \mathbb{C}^n$ for all $\xi \in \mathbb{R}$, and whereby the operator $\mathcal{S}_\tau$, $\tau \in \mathbb{R}$ denotes the temporal shift (or spectral modulation) operator

$$\begin{aligned} \mathcal{S}_\tau : \mathcal{M}(\mathbb{R}) &\to \mathcal{M}(\mathbb{R}) \\ \mu &\mapsto e^{-i2\pi\tau \mathrm{Id}}\mu \end{aligned} \tag{5.7}$$

Multirate sampling systems can be seen as a more generalist version of the coprime sampling systems studied in [52], [54] in the sense that samplers are allowed to work at different frequencies *and* different delays.

On the practical side, applications of the MRSS framework are numerous in signal processing. It occurs when sampling in parallel the output of a common channel in order to get benefits from cleverly designed sampling frequencies and delays; such designs occur, for example, in modern digitalization with variable bit-rates and analysis of video and audio streams, or whenever one seeks to expend the spectral range of an acquisition system while limiting the total amount of samples acquired. The MRSS framework is also naturally fitted to describe sampling processes in distributed sensor networks: each node, with limited processing capabilities, samples at its own rate, a delayed version of a complex signal. Collected data are then sent and merged at a higher level processing unit, performing a global estimation of the spectral distribution on a joint manner. Figure 5.1 provides a schematic representation of this application scenario.

5.1.2 Joint spectral estimation

The line spectral estimation problem will consist in the MRSS context in finding the sparsest spectral density $\tilde{\mu}_0 \in \mathcal{M}(\mathbb{R})$ that *jointly matches* the p observation vectors y_j under the consistency constraint (5.2) for all $j \in [\![1,p]\!]$. Using the same paradigm as the one used to build the initial Program (1.5), one can formulate the optimal estimator as a combinatorial minimization program involving the L_0 pseudo-norm over the set of Radon

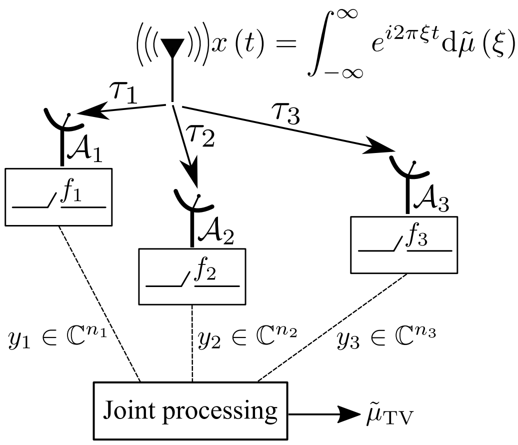

Figure 5.1: Joint spectral estimation in a sensor network.

measure defined over $\mathbb{R}$, leading to

$$\begin{aligned} \tilde{\mu}_0 = \arg \min_{\mu \in \mathcal{M}(\mathbb{R})} \|\mu\|_0 \\ \text{subject to} \quad y_j = \mathcal{L}_j(\mu), \quad \forall j \in [\![1, p]\!], \end{aligned} \tag{5.8}$$

and one can consider its convex relaxation for the total variation norm

$$\begin{aligned} \tilde{\mu}_{\mathrm{TV}} = \arg \min_{\mu \in \mathcal{M}(\mathbb{R})} |\mu|(\mathbb{R}) \\ \text{subject to} \quad y_j = \mathcal{L}_j(\mu), \quad \forall j \in [\![1, p]\!], \end{aligned} \tag{5.9}$$

whereby the total-variation over the real line $|\mu|(\mathbb{R})$ is defined analogously to $|\mu|(\mathbb{T})$ by integrating over $\mathbb{R}$ instead of $\mathbb{T}$ in the definition (1.25).

It is important to notice that two different grids $\mathcal{A}_j$ and $\mathcal{A}_{j'}$ may sample a value of the signal $x(\cdot)$ at the same time instant on the respective sampling indexes r and r', enforcing a relation of the kind $y_j[r] = y_{j'}[r']$. In the following we denote by $\tilde{n} = \sum_{j=1}^{p} n_j$ the total number of samples acquired by the system $\mathbb{A}$, and by $n \leq \tilde{n}$ the *net number of observations* obtained after removing such sampling overlaps, so that n is the number of *independent observation constraints* of the sampling system. The joint measurement vector is denoted $\tilde{y} = \left[y_1^{\mathsf{T}}, \ldots, y_p^{\mathsf{T}}\right]^{\mathsf{T}} \in \mathbb{C}^{\tilde{n}}$. Similarly, its net counterpart $y \in \mathbb{C}^n$ is obtained by discarding the redundancies of $\tilde{y}$, so that $y = \mathbf{C}_{\mathbb{A}} \tilde{y}$ for some selection matrix $\mathbf{C}_{\mathbb{A}} \in \{0,1\}^{n \times \tilde{n}}$. The joint linear measurement constraint of Problem (5.8) can then be reformulated

$$y = \mathcal{L}(\mu), \tag{5.10}$$

whereby the operator $\mathcal{L} \in (\mathcal{M}(\mathbb{R}) \mapsto \mathbb{C}^n)$ admits the operators $\{\mathcal{L}_j\}_{j \in [\![1,p]\!]}$ as restrictions on the p subspaces induced by the construction of the net observation vector y.

5.2 Common grid expansion and SDP formulation

5.2.1 The common grid hypothesis

It is been shown in Chapter 4 that the dual problem can take the form of a low-dimensional SDP (4.12) whenever the observation operator $\mathcal{L}$ can be written under the form $\mathcal{L}(\mu) = \int_{\mathbb{T}} \mathbf{M} a_m(\omega)\,\mathrm{d}\mu$ for some low-dimensional measurement matrix $\mathbf{M} \in \mathbb{C}^{r\times(2m+1)}$, and provided the existence of a solution satisfying the sparse Fejér-Riesz condition introduced in Definition 4.1. This remarkable property is due to the polynomial nature of the adjoint measurement operator $\mathcal{L}^*$ allowing a semidefinite representability of the dual feasible set. However, in the MRSS context, the dual observation operator, given by

$$\forall v \in \mathbb{C}^n, \quad \mathcal{L}^*(v) = \sum_{j=1}^{m} \mathcal{L}_j^*(v_j)$$

does not take such polynomial form in the general case. A direct calculation reveals that $\mathcal{L}^*(v)$ is instead an exponential polynomial[1] for all $v \in \mathbb{C}^n$. Up to our knowledge, there is no welcoming algebraic characterization for optimization purposes of the dual feasible set $\mathcal{D}_{\mathbb{A}} = \{v \in \mathbb{C}^n, \|\mathcal{L}^*(v)\|_\infty \leq 1\}$. Therefore, the theory of sum-of-squares and semidefinite representations developed in Chapter 1 and Chapter 4 cannot be directly transcribed in the MRSS framework.

To bridge this concern, we restrict our analysis to the case where the observation operator admits a factorization of the form $\mathcal{L}(\mu) = \int_{\mathbb{R}} \mathbf{M}\tilde{a}_{n,f}(\xi)\,\mathrm{d}\mu$ for some $n \in \mathbb{N}$, $f \in \mathbb{R}^+$ and for a full row rank matrix $\mathbf{M}$. The following aims to provide an algebraic criterion in terms of the parameters $\{(f_j, \gamma_j, n_j)\}$ of the sampling system $\mathbb{A}$ for this hypothesis to hold. We will see that this extra hypothesis consists in supposing that the samples acquired by $\mathbb{A}$ can by virtually aligned at a higher rate on others grids $\mathcal{A}_+$. Such grids will be called common supporting grid for $\mathbb{A}$, and are defined as follows.

Definition 5.1 (Common supporting grid)**.** A grid $\mathcal{A}_+ = (f_+, \gamma_+, n_+)$ is said to be a *common supporting grid* for a set of sampling grids $\mathbb{A} = \{\mathcal{A}_j\}_{j\in[\![1,p]\!]}$ if and only if the set of samples acquired by the MRSS induced by $\mathbb{A}$ is a subset of the one acquired by $\mathcal{A}_+$. In formal terms, the definition is equivalent to,

$$\left\{\frac{1}{f_j}(r_j - \gamma_j),\, j \in [\![1,p]\!],\, r_j \in [\![0, n_j - 1]\!]\right\} \subseteq \left\{\frac{1}{f_+}(r_+ - \gamma_+),\, r_+ \in [\![0, n_+ - 1]\!]\right\}. \quad (5.11)$$

The set of common supporting grids of $\mathbb{A}$ is denoted by $\mathcal{C}(\mathbb{A})$. Moreover, a common supporting grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ for $\mathbb{A}$ is said to be *minimal* if and only it satisfies the minimality condition,

$$\forall \mathcal{A}_+ \in \mathcal{C}(\mathbb{A}), \quad n_\Phi \leq n_+.$$

Finally, the *equivalent observation set* of the minimal common grid $\mathcal{A}_\Phi$, denoted by $\mathcal{I}$, is the subset of $[\![0, n_\Phi - 1]\!]$ of cardinality m, formed by the l's for which the time instants $\frac{1}{f_\Phi}(l_+ - \gamma_\Phi)$ for $l_\Phi \in [\![0, n_\Phi - 1]\!]$ are acquired by $\mathbb{A}$.

[1] A function f of the complex variable z of the form $f(z) = \sum_{k=1}^{m} c_k z^{\gamma_k}$ for some $\{\gamma_k\}_{[\![1,m]\!]} \subset \mathbb{R}$.

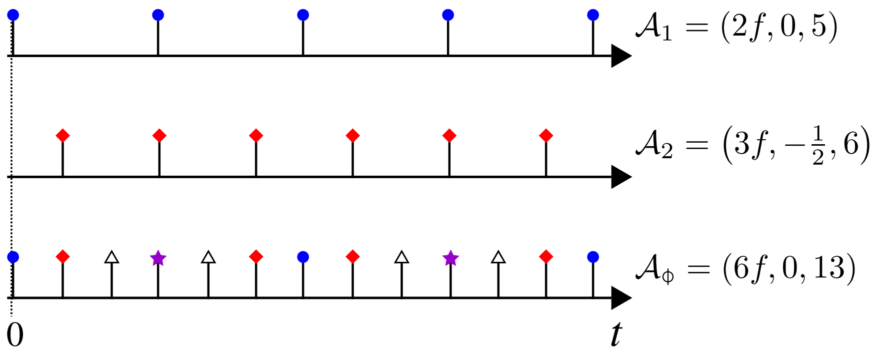

Figure 5.2: A representation of a multirate sampling system $\mathbb{A}$ composed of two arrays $(\mathcal{A}_1, \mathcal{A}_2)$, and its associated minimal common grid $\mathcal{A}_\Phi$. Purple stars in the common grid correspond to time instant acquired multiple times by the system $\mathbb{A}$, and blank triangles to omitted samples. In this example, the dimension of the minimal common grid is $n_\Phi = 13$, The total number of observation of $\mathbb{A}$, $\tilde{n} = 5 + 6 = 11$, and the net number of observations is $n = 9$. Finally the equivalent observation set of the common grid is $\mathcal{I} = \{0, 1, 3, 5, 6, 7, 9, 11, 12\}$.

It is clear that if $\mathcal{C}(\mathbb{A})$ is not empty then the minimal common supporting grid for $\mathbb{A}$ exists and is unique. For ease of understanding, Figure 5.2 illustrates the notion of common supporting grid by showing a MRSS formed by two arrays and their minimal common grid. Proposition 5.1 states necessary and sufficient conditions in terms of the parameters of $\mathbb{A}$ such that the set $\mathcal{C}(\mathbb{A})$ is not empty. The proof of this proposition is technical and delayed to Section 5.7 for readability.

Proposition 5.1 (Existence of common supporting grids)**.** *Given a set of p grids $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j \in [\![1,p]\!]}$, the set $\mathcal{C}(\mathbb{A})$ is not empty if and only if there exist $f_+ \in \mathbb{R}^+$, $\gamma_+ \in \mathbb{R}$, a set of p positive integers $\{l_j\} \in \mathbb{N}^p$, and a set of p integers $\{a_j\} \in \mathbb{Z}^p$ satisfying $f_+ = l_j f_j$ and $\gamma_+ = l_j \gamma_j - a_j$ for all $j \in [\![1,p]\!]$. Moreover a common grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ is minimal, if and only if*

$$\begin{cases} \gcd\left(\{a_j\}_{j \in [\![1,p]\!]} \cup \{l_j\}_{j \in [\![1,p]\!]}\right) = 1 \\ \gamma_\Phi = \max_{j \in [\![1,p]\!]} \{l_j \gamma_j\} \\ n_\Phi = \max_{j \in [\![1,p]\!]} \{l_j (n_j - 1) - a_j\}. \end{cases}$$

Although the conditions of Proposition 5.1 appear to be strong since one get $\mathcal{C}(\mathbb{A}) = \emptyset$ almost surely in the Lebesgue sense when the sampling frequencies and delays are drawn at random, assuming the existence of a common supporting grid for $\mathbb{A}$ is not meaningless in our context. By density, one can approximately align the system $\mathbb{A}$ on an arbitrary fine grid $\mathcal{A}_\varepsilon$, for any given maximal jitter $\varepsilon > 0$, and perform the proposed super-resolution on this common grid. The resulting error from this approximation can be interpreted as a "basis mismatch". The detailed analysis of this approach will not be covered in this work, however, similar approximations can be found in the literature for the analogue atomic norm minimization view of the reconstruction problem [6]. We claim that those results extend in our settings and that the approximation error vanishes in the noiseless settings when going to the limit $\varepsilon \to 0$.

5.2.2 Relationship with partial line spectral estimation

Up to here, the formulation proposed to recover sparse spectra from multirate measurements seems very different from the partial line spectral estimation problem exposed in Chapter 4. The major difference between the formulation of the total variation estimators (5.9) and (4.3) resides in the *definition domain* of the measure to reconstruct. More specifically, the Nyquist-Shannon sampling theorem [51], [59] ensures that the line spectral estimation problem from the output of a single sampler can always be reduced to a reconstruction problem over the torus $\mathbb{T}$, since there is an *aliasing ambiguity* on the reconstructed spectrum modulo the sampling frequency $f \in \mathbb{R}^+$. Hence, one can only estimate the true frequencies within the quotient group $\mathbb{R}/f\mathbb{Z} \simeq \mathbb{T}$.

The next proposition builds a link between the partial line spectral estimation problem (4.3) and TV regularization for MRSS (5.9) whenever the sampling system $\mathbb{A}$ admits a common grid in the sense of Definition 5.1. The proof of this proposition is detailed in Section 5.5.

Proposition 5.2 (Equivalence with the partial line spectral estimation problem). *Let $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j\in[\![1,p]\!]}$ be a set of p arrays, and suppose that the net number of observations $n \in \mathbb{N}$ of the array $\mathbb{A}$ is an odd integer. If the set $\mathcal{C}(\mathbb{A})$ is not empty then there exists a subset $\mathcal{I} \subseteq [\![0, n_\Phi]\!]$ and a selection matrix $\mathbf{C}_\mathcal{I} \in \{0,1\}^{n\times n_\Phi}$ such that the output $\tilde{\mu}_{\mathrm{TV}}$ of Problem (5.8) and the output $\mu_{\mathbf{C}_\mathcal{I},\mathrm{TV}}$ of Program (4.3) with partial measurement matrix $\mathbf{C}_\mathcal{I} \in \{0,1\}^{n\times n_\Phi}$ are equal up to an aliasing factor modulo f_Φ.*

5.3 Dual certifiability and sub-Nyquist guarantees

In this section, sufficient conditions are presented to ensure that the conditions of Proposition 4.1 are fulfilled. Those conditions guarantee the tightness of the total variation relaxation and the optimality and uniqueness of the recovery $\mu_0 = \mu_{\mathbf{C}_\mathcal{I},\mathrm{TV}}$. In addition to this result, it provides mild conditions to ensure a sub-Nyquist recovery of the spectral spikes at a rate f_Φ from measurements taken at various lower rates $\{f_j\}_{j\in[\![1,p]\!]}$. The proof of this result, presented in Section 5.6, relies on previous polynomial construction methods presented in [6], [18], [67].

Theorem 5.1 (Tightness of TV regularization for MRSS). *Let $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j\in[\![1,p]\!]}$ be a set of sampling arrays. Suppose that $\mathcal{C}(\mathbb{A})$ is not empty, and denote by $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ the minimal common supporting grid of $\mathbb{A}$. Assume that the system induced by $\mathbb{A}$ satisfies at least one of the two following separability conditions,*

- Strong condition:

$$\forall j \in [\![1,p]\!], \qquad \begin{cases} \Delta_\mathbb{T}\left(\frac{\tilde{X}}{f_j}\right) \geq \frac{2.52}{n_j - 1} \\ n_j > 2000, \end{cases}$$

- Weak condition:

$$\exists j \in [\![1, p]\!], \quad \begin{cases} \Delta_{\mathbb{T}}\left(\frac{\tilde{X}}{f_j}\right) \geq \frac{2.52}{n_j - 1} \\ n_j > 2000 \\ n \geq (l_j + 1)\, s, \end{cases}$$

then there exists a polynomial $Q_\star$ verifying the conditions (4.5) of Proposition 4.1 for a subsampling matrix $\mathbf{C}_{\mathcal{I}}$. Consequently $\mu_0 = \mu_{\mathbf{C}_{\mathcal{I}},\mathrm{TV}}$. Moreover, $\mu_{\mathbf{C}_{\mathcal{I}},\mathrm{TV}}$ is equal to the ground truth measure $\tilde{\mu}_0$ up to an aliasing factor modulo f_Φ.

Remark 5.1. Under the weaker proviso $n_j > 256$, the above results still hold in both cases when $\tilde{X}$ satisfies the more restrictive separability criterion $\Delta_{\mathbb{T}}\left(\frac{\tilde{X}}{f_j}\right) \geq \frac{4}{n_j - 1}$.

The strong condition for Theorem 5.1 is restrictive and do not particularly highlight any benefits from jointly estimating the spectral support compared to merging the p spectral estimates obtained by simple individual estimation at each sampler. However, the weak condition guarantees that frequencies of the time signal x can be recovered with an ambiguity modulo f_Φ when jointly resolving the MRSS, while individual estimations would guarantee to recover them with an ambiguity modulo $\max_{j \in [\![p]\!]} \{f_j\} \leq f_\Phi$. The weak condition requires a standard spectral separation assumption from a single array $\mathcal{A}_j$, and sufficient net measurements n of the time signal. The extra measurements $n - n_j$ corresponding to the other grids are *not uniformly aligned* with the sampler $\mathcal{A}_j$. Therefore the sampling system induced by $\mathbb{A}$ achieves sub-Nyquist spectral recovery of the spectral spikes, and pushes away the classic spectral range f_j from a factor $\frac{f_\Phi}{f_j} = l_j$. Nevertheless, the provided construction of the dual certificate results in a polynomial having a modulus close to unity on the aliasing frequencies induced by the zero forcing upscaling from f_j to f_Φ. Consequently, one can expect to obtain degraded performances in noisy environments when the sub-sampling factor l_j becomes large.

5.4 Benefits of multirate measurements

5.4.1 Frequency range and resolution tradeoff

Multirate sampling has been used in many applications arising from signal processing and telecommunications in order to reduce either the number of required measurements or the processing complexity [52]. There are three major benefits of making use of MRSS acquisition in the line spectral estimation problem. One might just think MRSS has an obvious way of increasing the number of samples acquired by system compared to a single grid measurement $\mathcal{A}_j \in \mathbb{A}$. This naturally leads to an enhanced *noise robustness*. More importantly, MRSS acquisition brings benefits in terms of *spectral range* extension, and *spectral resolution* improvement. The spectral range extension (or sub-Nyquist) capabilities have been described in Theorem 5.1. Also this topic will not be covered in this thesis, the spectral resolution — the minimal distance on the torus between two spectral spikes to guarantee their recovery —, is also expected to be enhanced in MRSS acquisition due to the observation of delayed versions of the time signal x, which virtually enlarges the global observation window.

For the sake of clarity, Figure 5.3 proposes a comprehensive illustration of the tradeoff between range extension and resolution improvement for a delay-only MRSS constituted of two samplers $\mathcal{A}_1$ and $\mathcal{A}_2$. In Figure 5.3 (a), the delay between the two samplers is such that the joint uniform grid $\mathcal{A}_\Phi$ has no missing observations with a double sampling frequency. One trivially expects to recover the spikes locations of the measure $\tilde{\mu}$ with aliasing ambiguity modulo $2f$. In Figure 5.3 (b), the delay of $\mathcal{A}_2$ is set such that the resulting minimal common grid has a doubled observation window. $\mathcal{A}_\Phi$ fits again in the uniform observation framework analyzed in [18], and the achievable resolution is expected to be twice smaller. Finally a hybrid case is presented in Figure 5.3 (c), where one expect to get some spectral range and resolution improvements from a joint recovery approach.

5.4.2 Complexity improvements

If the set $\mathcal{C}(\mathbb{A})$ is not empty, the joint spectral estimation can be done by solving the semidefinite program (4.4) involving a linear matrix inequality of dimension of $n_\Phi + 1$. The dimensional n_Φ of the minimal common grid is fully determined by the observation pattern induced by $\mathbb{A}$, and reads

$$n_\Phi = \max_{j \in [\![1,p]\!]} \left\{ l_j (n_j - 1) - a_j \right\},$$

whereby the parameters $\{(a_j, l_j)\}_{j \in [\![1,p]\!]}$ are defined in Proposition 5.1. This is particularly disappointing since n_Φ grows at a speed driven by the product of the n_j's, whereas the *essential dimension* m of the problem is given by the number of net observations acquired by the grid $n \leq \tilde{n} = \sum_{j=1}^{p} n_j$. We study the asymptotic ratio $\frac{n}{n_\Phi}$ when the number grids p grows large in two different idealized instances of MRSS to illustrate that the reduced SDP formulation (4.12) brings *orders of magnitude* changes to the computational complexity of the line spectral estimation problem.

Suppose a delay-only MRSS, where $\mathbb{A}$ is constituted of p grids given by $\mathcal{A}_1 = (f, 0, n_0)$ and $\mathcal{A}_j = \left(f, -\frac{1}{b_j}, n_0\right)$ for all $j \in [\![2,p]\!]$. Moreover suppose that $\{b_j\}_{j \in [\![2,p]\!]}$ are jointly coprime. It is easy to verify that $\mathcal{C}(\mathbb{A})$ is not empty in those settings, and that the minimal common grid $\mathcal{A}_\Phi$ is given by $\mathcal{A}_\Phi = \left(\left(\prod_{j=2}^{p} b_j\right) f, 0, \left(\prod_{j=2}^{p} b_j\right) n_0\right)$. One has $n_\Phi = \Omega\left(b^p n_0\right)$ for some constant $b \in \mathbb{R}^+$, while $n = pn_0$. The ratio $\frac{m}{n_\Phi} = o\left(\frac{p}{b^p}\right)$ tends to 0 exponentially fast with the number of samplers m of the system.

On the other hand, suppose a synchronous coprime sampling system between the time instants 0 and T, where $\mathcal{A}_j = (k_j f, 0, k_j f T)$ for all $j \in [\![1,p]\!]$ with $\gcd\{k_j, j \in [\![1,p]\!]\} = 1$. Once again $\mathcal{C}(\mathbb{A})$ is not empty, and the minimal grid is characterized by the parameters $\mathcal{A}_\Phi = \left(\left(\prod_{j=1}^{p} k_j\right) f, 0, \left(\prod_{j=1}^{p} k_j\right) fT\right)$. Consequently the ratio $\frac{n}{n_\Phi} = \frac{\sum_{j=1}^{p} k_j}{\prod_{j=1}^{p} k_j}$ decreases in $o\left(k^{-p}\right)$ for a judicious choice of the integers $\{k_j\}_{j \in [\![1,p]\!]}$.

5.5 Proof of Proposition 5.2

We recall from equation (5.5) that for all $\tilde{\mu} \in \mathcal{M}(\mathbb{R})$, one has,

$$\forall j \in [\![1,m]\!], \forall r_j \in [\![0, n_j - 1]\!], \quad \mathcal{L}_j[k] = \int_{\mathbb{R}} e^{i2\pi \frac{\xi}{f_j}(r_j - \gamma_j)} \mathrm{d}\tilde{\mu}(\xi).$$

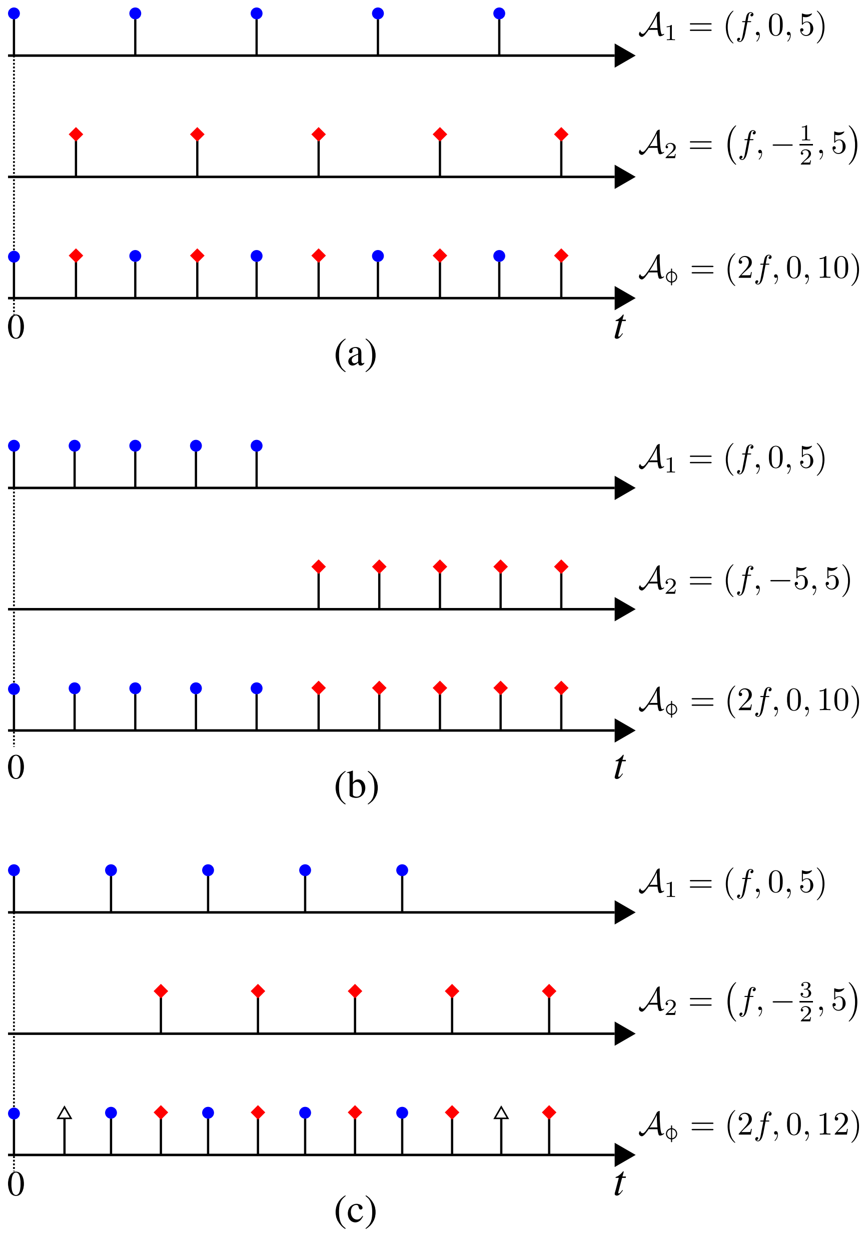

Figure 5.3: A representation of three delay-only MRSS in different settings. In (a), the delay between the two samplers is exactly of half-unit, resulting in a doubled frequency range in the joint analysis. In (b), this delay is such that the overall process equivalently acquires samples on a doubled time frame, resulting in a doubled spectral resolution. Sub-figure (c) represents an hybrid case where both resolution improvement and spectral range extension are expected.

Suppose that $\mathcal{C}(\mathbb{A})$ is not empty, so that the minimal common supporting grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ for $\mathbb{A}$ exists. It comes by equation (5.11) that

$$\begin{aligned}\forall j \in [\![1,m]\!], \forall r_j \in [\![0,n_j-1]\!], \exists r_\Phi \in [\![0,n_\Phi-1]\!], \quad \mathcal{L}_j(\tilde{\mu})[r_j] &= \int_{\mathbb{R}} e^{i2\pi\frac{\xi}{f_\Phi}(r_\Phi-\gamma_\Phi)}\mathrm{d}\tilde{\mu}(\xi) \\ &= \int_{\mathbb{R}} e^{i2\pi\frac{\xi}{f_\Phi}r_\Phi}\mathrm{d}\left(e^{-i2\pi\frac{\xi\gamma_\Phi}{f_\Phi}}\tilde{\mu}(\xi)\right) \\ &= \mathcal{F}_{n_\Phi}\circ\mathcal{S}_{\frac{\gamma_\Phi}{f_\Phi}}(\tilde{\mu})[k].\end{aligned}$$

Let $\mathcal{I} \subseteq [\![0,n_\Phi-1]\!]$ be the equivalent observation set of $\mathcal{A}_\Phi$ introduced in Definition 5.1 and consider a selection matrix $\mathbf{C}_\mathcal{I} \in \mathbb{C}^{n\times n_\Phi}$ for this set. The above equality ensures the measurement operator admit a factorization of the form

$$\mathcal{L} = \mathbf{C}_\mathcal{I}\left(\mathcal{F}_{n_\Phi,f_\Phi}\circ\mathcal{S}_{\frac{\gamma_\Phi}{f_\Phi}}\right). \tag{5.12}$$

Next, let $\tilde{\mu}^\sharp \in \mathcal{M}(\mathbb{R})$ be the offset free surrogate of the measure $\tilde{\mu}$ defined by

$$\tilde{\mu}^\sharp = \mathcal{S}_{\frac{\gamma_\Phi-\frac{n_\Phi-1}{2}}{f_\Phi}}\tilde{\mu},$$

which is a simple modulation of the complex amplitudes of the frequencies in the support set $\tilde{X}$ of the measure μ. Furthermore, define by $\mu_\Phi, \mu_\Phi^\sharp \in \mathcal{M}(\mathbb{T})$ the respective $f_\Phi-$aliased version of the real measures $\tilde{\mu}, \tilde{\mu}^\sharp \in \mathcal{M}(\mathbb{R})$ given by

$$\begin{aligned}\forall\omega\in\mathbb{T}, \quad \mu_\Phi(\omega) &= \sum_{j\in\mathbb{Z}}\tilde{\mu}(f_\Phi(\omega+k)), \\ \forall\omega\in\mathbb{T}, \quad \mu_\Phi^\sharp(\omega) &= \sum_{j\in\mathbb{Z}}\tilde{\mu}^\sharp(f_\Phi(\omega+k)).\end{aligned}$$

It is enough to show that the consistency constraint $\mathcal{L}(\tilde{\mu}_\Phi)$ can be written under the form (4.1) with respect to the aliased version of the measure $\mu_\Phi^\sharp$ to demonstrate this result. Equation (5.12) leads to

$$\begin{aligned} z &= \mathcal{L}(\tilde{\mu}) \\ &= \mathbf{C}_\mathcal{I}\left(\mathcal{F}_{n_\Phi,f_\Phi}\circ\mathcal{S}_{\frac{\gamma_\Phi}{f_\Phi}}(\tilde{\mu})\right) \\ &= \mathbf{C}_\mathcal{I}\left(\mathcal{F}_{n_\Phi,f_\Phi}\circ\mathcal{S}_{-\frac{n_\Phi-1}{2f_\Phi}}\left(\tilde{\mu}^\sharp\right)\right) \\ &= \mathbf{C}_\mathcal{I}\int_{-\infty}^{\infty}\tilde{a}_\Phi^\sharp(\xi)\,\mathrm{d}\tilde{\mu}^\sharp(\xi)\end{aligned}$$

whereby the atoms associated to the common grid $\tilde{a}_\Phi^\sharp(\cdot) \in \mathbb{C}^{n_\Phi}$ reads,

$$\begin{aligned}\forall r_\Phi \in [\![0,n_\Phi-1]\!], \forall\xi\in\mathbb{R}, \quad \tilde{a}_\Phi^\sharp(\xi)[r] =& e^{i2\pi\frac{\xi}{f_\Phi}\left(r_\Phi-\frac{n_\Phi-1}{2}\right)} \\ =& a_{\frac{n_\Phi-1}{2}}\left(\frac{\xi}{f_\Phi}\right).\end{aligned}$$

The primal consistency constraint of Program (5.9) writes

$$\begin{aligned} z &= \mathbf{C}_{\mathcal{I}} \int_{\mathbb{T}} a_{\frac{n_\Phi - 1}{2}}(\omega)\, \mathrm{d}\mu_\Phi^\sharp(\omega) \\ &= \int_{\mathbb{T}} \mathbf{C}_{\mathcal{I}} a_{\frac{n_\Phi - 1}{2}}(\omega)\, \mathrm{d}\mu_\Phi^\sharp(\omega) \end{aligned}$$

which can be identified to the consistency constraint of the partial line spectral estimation problem (4.1), concluding the proof of the proposition. □

5.6 Proof of Theorem 5.1

5.6.1 Preliminaries

In both strong and weak condition cases, the proof relies on previous works presented in [18], [67], and is achieved by constructing a polynomial $Q_\star$ satisfying the conditions (4.5). For simplicity, it will be assumed that $n_j = 2m_j + 1$ for all $j \in [\![1, p]\!]$ and $n_\Phi = 2m_\Phi + 1$ are odd integers for simplicity purposes. We claim that this does not affect the generality of the proof and free the following demonstration of the burden of considering half-integer degree trigonometric polynomials. Since, it has been previously shown in Section 5.5 that shifting the signal in the time domain leave the dual feasible set invariant, and we will assume without loss of generality that $\gamma_\Phi = m_\Phi$ so that

$$\begin{aligned} \forall \mu \in \mathcal{M}(\mathbb{T}), \quad \mathcal{L}(\mu) &= \mathbf{C}_{\mathcal{I}} \mathcal{F}_{n_\Phi}(\mu) \\ &= \mathbf{C}_{\mathcal{I}} \int_{\mathbb{T}} a_{m_\Phi}(\omega)\, \mathrm{d}\mu(\omega)\,. \end{aligned}$$

Before starting the proof, we introduce the notations

$$\begin{aligned} X_\Phi &= \frac{\tilde{X}}{f_\Phi} = \left\{ \frac{\xi}{f_\Phi},\ \xi \in \tilde{X} \right\} \\ \forall j \in [\![1, p]\!], \quad X_j &= \frac{\tilde{X}}{f_j} = \left\{ \frac{\xi}{f_j},\ \xi \in \tilde{X} \right\} \\ \forall j \in [\![1, p]\!], \quad \tilde{X}_j &= \left\{ \frac{\xi}{f_\Phi} + \frac{r_j}{l_j},\ \xi \in \tilde{X}, r_j \in [\![0, l_j - 1]\!] \right\}. \end{aligned}$$

In the above, X_Φ and X_j are the sets of the reduced frequencies of the spectral support $\tilde{X}$ of the measure $\tilde{\mu}$ for the respective sampling frequencies f_Φ and f_j, while $\tilde{X}_j$ is the aliased set of X_j resulting from a zero-forcing upsampling from the rate f_j to the rate f_Φ.

We recall from [67] Proposition II.4, using the improved separability conditions taken from [33] Proposition 4.1, that if $\Delta_{\mathbb{T}}(X_j) \geq \frac{2.52}{n_j - 1}$, then one can build a polynomial $P_{j,\star} \in \mathcal{T}_{m_j}$ satisfying the interpolating conditions

$$\begin{cases} P_{j,\star}\left(\frac{\tilde{x}_k}{f_j}\right) = \operatorname{sign}\left(e^{i2\pi \frac{a_j}{l_j} \frac{\tilde{x}_k}{f_j}} c_k\right), & \forall \frac{\tilde{x}_k}{f_j} \in X_j \\ |P_{j,\star}(\omega)| < 1, & \forall \omega \in \mathbb{T} \backslash X_j \\ \frac{\mathrm{d}^2 |P_{j,\star}|}{\mathrm{d}\omega^2}\left(\frac{\tilde{x}_k}{f_j}\right) \leq -\eta, & \forall \frac{\xi_r}{f_j} \in X_j, \end{cases} \tag{5.13}$$

provided that $n_j > 2 \times 10^3$, for some $\eta > 0$ ($\eta = 7.865\,10^{-2}$ in the original proof presented in [33]), and whereby $\{(a_j, l_j)\}_{j\in[\![1,p]\!]}$ are the pairs of parameters defined in the statement of Proposition 5.1 characterizing the expansion of the array $\mathcal{A}_j$ into the minimal common grid $\mathcal{A}_\Phi$. If the polynomial $P_{j,\star}$ exists, we further introduce the polynomial $Q_{j,\star} \in \mathcal{T}_{m_\Phi}$ defined by

$$\forall \omega \in \mathbb{C}, \quad Q_{j,\star}(\omega) = e^{-i2\pi a_j} P_{j,\star}(l_j\omega). \tag{5.14}$$

By construction, $Q_{j,\star}$ is a sparse polynomial with monomial support on the equivalent observation set $\mathcal{I}$ of the minimal common grid introduced in Definition 5.1. Its coefficients vector $q_{j,\star}$ satisfies the relation $q_{j,\star} = \mathbf{C}_{\mathcal{I}}^* v_{j,\star}$ for some $v_{j,\star} \in \mathbb{C}^n$. It is easy to notice that $Q_{j,\star}$ is $\frac{1}{l_j}$-periodic due to the upscaling effect $\omega \leftarrow l_j\omega$ in equation (5.14). Consequently the polynomial $Q_{j,\star}$ reaches a modulus equal to 1 on every point of $\tilde{X}_j$, with value satisfying

$$\begin{aligned}
\forall \omega \in \tilde{X}_j, \quad Q_j(\omega) &= Q_{j,\star}\left(\frac{\tilde{x}_k}{f_\Phi} + \frac{r_j}{l_j}\right) \\
&= e^{-i2\pi a_j\left(\frac{\tilde{x}_k}{f_\Phi} + \frac{r_j}{l_j}\right)} P_{j,\star}\left(\left(\frac{l_j\tilde{x}_k}{f_\Phi} + r_j\right)\right) \\
&= e^{-i2\pi a_j\left(\frac{\tilde{x}_k}{f_\Phi} + \frac{r_j}{l_j}\right)} \operatorname{sign}\left(e^{i2\pi\frac{a_j}{l_j}\frac{\tilde{x}_k}{f_j}} c_k\right) \\
&= e^{-i2\pi a_j\frac{r_j}{l_j}} \operatorname{sign}(c_k),
\end{aligned}$$

whereby $\frac{\tilde{x}_k}{f_\Phi} \in X_\Phi$ and $r_j \in [\![0, l_j - 1]\!]$. It comes that the constructed polynomial verifies the interpolation conditions

$$\begin{cases}
Q_{j,\star}(\omega) = \operatorname{sign}(c_k), & \forall \omega \in X_\Phi \\
Q_{j,\star}(\omega) = e^{-i2\pi a_j\frac{r_j}{l_j}} \operatorname{sign}(c_k), & \forall \frac{\tilde{x}_k}{f_j} \in \tilde{X}_j \\
|Q_{j,\star}(\omega)| < 1, & \forall \omega \in \mathbb{T}\backslash\tilde{X}_j \\
\frac{\mathrm{d}^2|Q_{j,\star}|}{\mathrm{d}\nu^2}(\omega) \leq -l_j\eta, & \forall \omega \in \tilde{X}_j,
\end{cases} \tag{5.15}$$

where the second equality stand for some $\frac{\tilde{x}_k}{f_\Phi} \in X_\Phi$ and $r_j \in [\![0, l_j - 1]\!]$ such that $\frac{\tilde{x}_k}{f_\Phi} + \frac{r_j}{l_j} \in \tilde{X}_j$.

Under both strong and weak assumptions, we aim to build a sparse polynomial $Q_\star \in \mathcal{T}_{m_\Phi}$ verifying the conditions (4.5). If the existence of such polynomial is verified Proposition 4.1 applies and the desired conclusion follows.

5.6.2 Construction under the strong condition

Suppose that $\Delta_{\mathbb{T}}(X_j) \geq \frac{2.52}{n_j-1}$ and $n_j > 2 \times 10^3$, for all $j \in [\![1, p]\!]$. As explained above, one can find p polynomials $Q_{j,\star} \in \mathcal{T}_m$ satisfying the interpolation properties (5.15), and define their mean by $Q_\star \in \mathcal{T}_{m_\Phi}$ such that

$$\forall \omega \in \mathbb{T}, \quad Q_\star(\omega) = \frac{1}{p}\sum_{j=1}^{p} Q_{j,\star}(\omega).$$

It is clear, by stability through linear combinations, that $Q_\star$ is still sparse and supported over the subset $\mathcal{I}$, ensuring the existence of an element $v_\star \in \mathbb{C}^n$ such that $q_\star = \mathbf{C}_{\mathcal{I}}^* v_\star$. Moreover, it is immediate to verify that $Q_\star$ satisfies

$$|Q_\star(\omega)| = 1 \Leftrightarrow \left(\omega \in \bigcap_{j=1}^{p} \tilde{X}_j \text{ and } \forall j \in [\![1,p]\!],\ Q_{j,\star}(\omega) = u(\omega)\right) \tag{5.16}$$

for some value $u(\omega) \in \mathbb{C}$ of modulus 1, $|u(\omega)| = 1$. Let us denote by $\Gamma \subset \mathbb{T}$ the set of frequencies satisfying where $Q_\star$ has a modulus equal to 1. From Equations (4.5) and (5.15), $Q_\star$ is a dual certificate if and only if $\Gamma = X_\Phi$. One has $X_\Phi \subseteq \Gamma$, thus it remains to prove $\Gamma \subseteq X_\Phi$ to finish the certificate construction under the strong condition. Using the definition of $\tilde{X}_j$ and the interpolation properties (5.15), we have that $\omega \in \Gamma$ is equivalent to

$$\omega \in \bigcap_{j=1}^{p} \tilde{X}_j \Longleftrightarrow \forall (j,j') \in [\![1,p]\!]^2, \exists (k,k') \in [\![1,s]\!]^2, \exists r_j \in [\![0,l_j-1]\!], \exists r_{j'} \in [\![0,l_{j'}-1]\!],$$
$$e^{-i2\pi a_j \frac{r_j}{l_j}} \operatorname{sign}(c_k) = e^{-i2\pi a_{j'} \frac{r_{j'}}{l_{j'}}} \operatorname{sign}(c_{k'}),$$

leading to

$$\omega \in \bigcap_{j=1}^{m} \tilde{X}_j \Longleftrightarrow \forall (j,j') \in [\![1,p]\!]^2, \exists (k,k') \in [\![1,s]\!]^2, \exists r_j \in [\![0,l_j-1]\!], \exists r_j \in [\![0,l_{j'}-1]\!],$$
$$\exists b \in \mathbb{Z}, a_j \frac{r_j}{l_j} + \frac{\arg(c_k)}{2\pi} = a_{j'} \frac{r_{j'}}{l_{j'}} + \frac{\arg(c_{k'})}{2\pi} + b. \tag{5.17}$$

The equality in the RHS of Equatiton (5.17) may occur for all pairs $(j,j') \in [\![1,p]\!]^2$ if and only if $k = k'$, and the above reduces to

$$\omega \in \bigcap_{j=1}^{m} \tilde{X}_j \Longleftrightarrow \forall (j,j') \in [\![1,p]\!]^2, \exists r \in [\![0,l_j-1]\!], \exists r' \in [\![0,l_{j'}-1]\!],$$
$$\exists b \in \mathbb{Z}, \quad \frac{a_j r_j}{l_j} = \frac{a_{j'} r_{j'}}{l_{j'}} + b,$$

which holds if and only if

$$\forall (j,j') \in [\![1,p]\!]^2, \exists r_j \in [\![0,l_j-1]\!],\ l_j \mid a_j l_{j'} r_j.$$

Recalling from the minimality condition of the common grid $\mathcal{A}_\Phi$ detailed in Proposition 5.1 that $\gcd\left(\{a_j\}_{j\in[\![1,p]\!]} \cup \{l_j\}_{j\in[\![1,p]\!]}\right) = 1$, one derives by application of the Gauss theorem

$$\exists j \in [\![1,p]\!], \quad l_j \mid r_j.$$

Since $r_j \in [\![0,l_j-1]\!]$, one has $r_j = 0$. We deduce that there must exists $k \in [\![1,s]\!]$ such that $\omega = \frac{\tilde{x}_k}{f_\Phi} + \frac{0}{l_j}$, thus $\omega \in X_\Phi$. Consequently, $\Gamma \subseteq X_\Phi$, and finally one has $\Gamma = X_\Phi$, which concludes the proof for the strong condition. □

5.6.3 Construction under the weak condition

Suppose that $\Delta_{\mathbb{T}}(X_j) \geq \frac{2.52}{n_j-1}$ and $n_j > 2\times 10^3$ for some $j \in [\![1,p]\!]$, and define the polynomial $Q_{j,\star} \in \mathcal{T}_{m_\Phi}$ as in equation (5.14). Moreover, we define by $\mathcal{H}_j(\mathbb{A}, X_\Phi)$ the affine subspace of elements $v \in \mathbb{C}^n$ such that $q = \mathbf{C}_{\mathcal{I}}^* v$ induces a sparse polynomial $Q \in \mathcal{T}_{m_\Phi}$ supported by monomials taken over the subset $\mathcal{I}$ and satisfying the interpolation conditions

$$\begin{cases} Q(\omega) = \operatorname{sign}(c_k), & \forall \omega \in X_\Phi \\ Q'(\omega) = 0, & \forall \omega \in X_\Phi \\ Q(\omega) = 0, & \forall \omega \in \tilde{X}_j \backslash X_\Phi. \end{cases}$$

The subspace $\mathcal{H}_j(\mathbb{A}, w)$ can be parametrized by the linear equality

$$\mathcal{H}_j(\mathbb{A}, q) = \left\{ v \in \mathbb{C}^n, \mathbf{V}_j(\mathbb{A}, X_\Phi)\, \mathbf{C}_{\mathcal{I}}^* v = w \right\},$$

whereby $w = [\operatorname{sign}(c_1), \dots, \operatorname{sign}(c_s)]^\mathsf{T} \in \mathbb{C}^s$, and for some matrix $\mathbf{V}_j(\mathbb{A}, X_\Phi) \in \mathbb{C}^{(l_j+1)s\times n_\Phi}$ defining the interpolation conditions. Interpolation theory guarantees that $\mathbf{V}_j(\mathbb{A}, X_\Phi)$ is full rank, and therefore the subspace $\mathcal{H}_j(\mathbb{A}, X_\Phi)$ is non-trivial with dimension $n - (l_j+1)s$, provided that $n \geq (l_j+1)s$. We fix an element $t \in \mathcal{H}_j(\mathbb{A}, X_\Phi)$, and denote by $R \in \mathcal{T}_{m_\Phi}$ the polynomial having for coefficients vector $r = \mathbf{C}_{\mathcal{I}}^* t$. In the rest of this proof, we seek to build a dual certificate $Q_\star \in \mathcal{T}_{m_\Phi}$ under the form of a convex combination between R and $Q_{j,\star}$ so that

$$Q_\star = \beta R + (1-\beta) Q_{j,\star}, \quad \beta \in [0,1].$$

First of all, by construction, R and $Q_{j,\star}$ both interpolate the frequencies of X_Φ with values $w_k = \operatorname{sign}(c_k)$, and one has

$$\forall \frac{\tilde{x}_k}{f_\Phi} \in X_\Phi, \quad Q_\star\left(\frac{\tilde{x}_k}{f_\Phi}\right) = w_k. \tag{5.18}$$

Consequently, it remains to derive sufficient conditions on β for the optimality condition $|Q_\star(\omega)| < 1$ to hold everywhere else on $\mathbb{T}\backslash X_\Phi$ to conclude that $Q_\star$ is a dual certificate. To do so, we partition the set $\mathbb{T}$ into three non-intersecting sets $\mathbb{T} = \Gamma_{\text{near}} \cup \Gamma_{\text{alias}} \cup \Gamma_{\text{far}}$, where Γ_{near} is a union of s open balls of small radii $0 < \varepsilon_{\text{near}}$ centered around the frequencies in X_Φ, and where Γ_{alias} is an open set containing the elements of $\tilde{X}_j\backslash X_\Phi$. The set Γ_{far} is defined by the complementary of the two previous in $\mathbb{T}$. The conditions on β for $Q_\star$ to be bounded away from 1 in modulus are derived independently on each of those sets.

We start the analysis on Γ_{near}. For any complex polynomial Q, we respectively denote by $Q_\Re(\nu) = \Re(Q(\omega))$ and $Q_\Im(\nu) = \Im(Q(\omega))$ for all $\omega \in \mathbb{T}$, its real and imaginary part around the unit circle. Moreover, we recall that

$$\frac{\mathrm{d}^2|Q|}{\mathrm{d}\omega^2}(\omega) = -\frac{\left(Q_\Re(\omega)\, Q_\Re'(\omega) + Q_\Im(\omega)\, Q_\Im'(\omega)\right)^2}{|Q(\omega)|^3} + \frac{|Q'(\omega)|^2 + Q_\Re(\omega)\, Q_\Re''(\omega) + Q_\Im(\omega)\, Q_\Im''(\omega)}{|Q(\omega)|}, \tag{5.19}$$

for all $\omega \in \mathbb{T}$. By construction, the derivative of R and $Q_{j,\star}$ cancels on X_Φ and by linearity

$$\forall \omega \in X_\Phi, \quad Q'_\star(\omega) = 0. \tag{5.20}$$

Injecting equations (5.18) and (5.20) into (5.19) leads to

$$\forall \omega \in X_\Phi, \quad \frac{\mathrm{d}^2 |Q_\star|}{\mathrm{d}\omega^2}(\omega) = \cos(w_r) Q''_{*\Re}(\omega) + \sin(w_r) Q''_{*\Im}(\omega).$$

Thus, the operator $\frac{\mathrm{d}^2|\cdot|}{\mathrm{d}\omega^2}$ acts linearly on the polynomial $Q_\star$ at the points in X_Φ, and one has

$$\begin{aligned}\forall \omega \in X_\Phi, \quad \frac{\mathrm{d}^2 |Q_\star|}{\mathrm{d}\omega^2}(\omega) &= \beta \frac{\mathrm{d}^2 |R|}{\mathrm{d}\omega^2}(\omega) + (1-\beta) \frac{\mathrm{d}^2 |Q_{j,\star}|}{\mathrm{d}\omega^2}(\omega) \\ &\leq \beta \frac{\mathrm{d}^2 |R|}{\mathrm{d}\omega^2}(\omega) - (1-\beta) l_j \eta,\end{aligned}$$

using the interpolation properties of equation (5.15). The inequalities

$$\forall \omega \in X_\Phi, \quad \frac{\mathrm{d}^2 |Q_\star|}{\mathrm{d}\omega^2}(\omega) < 0$$

can be jointly satisfied, for a choice of β

$$\beta < \frac{l_j \eta}{\mathcal{U}''_\Phi(R) + l_j \eta}, \tag{5.21}$$

where

$$\mathcal{U}''_\Phi(R) = \max_{\omega \in X_\Phi} \frac{\mathrm{d}^2 |R|}{\mathrm{d}\omega^2}(\omega).$$

Under the condition (5.21) , $|Q_\star| - 1$ has s non-nodal roots on X_Φ, and by continuity of $Q_\star$ there must exist a radius $0 < \varepsilon_{\text{near}}$ such that the inequality

$$\forall \omega \in \Gamma_{\text{near}} \backslash X_\Phi, \quad |Q_\star(\omega)| < 1$$

holds whereby $\Gamma_{\text{near}} = \bigcup_{r=1}^{s} \mathcal{B}\left(\frac{\xi_r}{f_\Phi}, \varepsilon_{\text{near}}\right)$, whereby $\mathcal{B}(\omega, \varepsilon)$ is defined on Page xvii.

We continue the proof by bounding $|Q_\star|$ away from 1 on the set Γ_{alias}. Fix any $0 < \delta < 1$ and let $\Gamma_{\text{alias}} = \{\omega, \ |R(\omega)| < \delta\}$. By continuity of R, Γ_{alias} is an open set verifying $\left(\tilde{X}_j \backslash X_\Phi\right) \subset \Gamma_{\text{alias}}$. Moreover one can impose $\Gamma_{\text{alias}} \cap \Gamma_{\text{near}} = \emptyset$ for a small enough δ. The value of $|Q_\star|$ over Γ_{alias} can be bounded by

$$\begin{aligned}\forall \omega \in \Gamma_{\text{alias}}, \quad |Q_\star(\omega)| &\leq \beta |R(\omega)| + (1-\beta) |Q_{j,\star}(\omega)| \\ &< \beta\delta + (1-\beta).\end{aligned}$$

Consequently, $|Q|$ is smaller than 1 on Γ_{alias} as long as $\beta > 0$.

It remains to prove that $|Q|$ can also be bounded by 1 in the rest of the torus $\Gamma_{\text{far}} = \overline{\mathbb{T} \backslash (\Gamma_{\text{true}} \cup \Gamma_{\text{alias}})}$. Let by $\mathcal{U}_{\text{far}}(R)$ and $\mathcal{U}_{\text{far}}(Q_{j,\star})$ the respective suprema of R and $Q_{j,\star}$ over Γ_{far}. Γ_{far} is a closed set, and thus compact. It comes that the suprema of R and Q are reached in some points inside Γ_{far}. Moreover introducing the suprema of $Q_{j,\star}$ over this set

$$\mathcal{U}_{\text{far}}\left(Q_{j,\star}\right) = \sup_{\omega \in \Gamma_{\text{far}}} \left\{\left|Q_{j,\star}\left(\omega\right)\right|\right\} < 1,$$

since $\tilde{X}_j \not\subseteq \Gamma_{\text{far}}$, leads to

$$\begin{aligned}\forall \omega \in \Gamma_{\text{far}}, \quad \left|Q_{\star}\left(\omega\right)\right| &\leq \beta \left|R\left(\omega\right)\right| + (1-\beta)\left|Q_{j,\star}\left(\omega\right)\right| \\ &< \beta \mathcal{U}_{\text{far}}\left(R\right) + (1-\beta)\,\mathcal{U}_{\text{far}}\left(Q_{j,\star}\right)\end{aligned}$$

for all $\omega \in \Gamma_{\text{far}}$, and thus $\left|Q_{\star}\left(\omega\right)\right| < 1$ can be achieved everywhere on Γ_{far} provided a choice of β verifying

$$\beta < \frac{1 - \mathcal{U}_{\text{far}}\left(Q_{j,\star}\right)}{\mathcal{U}_{\text{far}}\left(R\right) - \mathcal{U}_{\text{far}}\left(Q_{j,\star}\right)}.$$

We conclude that for any coefficient β satisfying

$$0 < \beta < \min\left\{\frac{l_j \eta}{\mathcal{U}_{\Phi}''\left(R\right) + l_j \eta}, \frac{1 - \mathcal{U}_{\text{far}}\left(Q_{j,\star}\right)}{\mathcal{U}_{\text{far}}\left(R\right) - \mathcal{U}_{\text{far}}\left(Q_{j,\star}\right)}\right\},$$

the polynomial $Q_{\star}$ meet the conditions (4.5) and thus qualifies as a dual certificate. □

5.7 Proof of Proposition 5.1

5.7.1 Existence of a common grid

Suppose that $\mathcal{A}_+$ is a common supporting grid for the set of arrays $\mathbb{A}$. The relation (5.11) ensures that

$$\forall j \in [\![1, p]\!],\ \forall r \in [\![0, n_j - 1]\!],\ \exists q_j\left[r\right] \in [\![0, n_+ - 1]\!] \quad \text{s.t.} \quad \frac{1}{f_j}\left(r - \gamma_j\right) = \frac{1}{f_+}\left(q_j\left[r\right] - \gamma_+\right), \tag{5.22}$$

whereby each integer $q_j\left[r\right]$ represents the position of the r^{th} samples of the j^{th} grid in the common grid. By subtracting two instances of (5.22) applied to the grid j and for the samples of order r and $r+1$ one gets

$$\forall j \in [\![1, p]\!], \forall r \in [\![0, n_j - 1]\!], \quad \frac{f_+}{f_j} = q_j\left[r+1\right] - q_j\left[r\right] \triangleq l_j,$$

where $\left\{l_j\right\}_{j \in [\![1,p]\!]}$ are positive integers since q_j is an increasing sequence for all $j \in [\![1, p]\!]$. It comes that $\left\{q_j\right\}_{j \in [\![1,p]\!]}$ are p arithmetic progressions with respective increment l_j

$$\forall j \in [\![1, p]\!], \forall r \in [\![0, n_j - 1]\!], \quad q_j\left[k\right] = q_j\left[0\right] + l_j r.$$

Reporting those results in equation (5.22) leads to

$$\forall j \in [\![1, p]\!], \quad \gamma_+ = q_j\left[0\right] + l_j \gamma_j.$$

Letting $a_j = -q_j\left[0\right]$ for all $j \in [\![1, p]\!]$ proofs the necessity part.

On the other hand, suppose now the existence of positive integers $\{l_j\} \in \mathbb{N}^p$ and integers $\{a_j\} \in \mathbb{Z}^p$ such that the relations

$$\begin{cases} f_+ = l_j f_j, & \forall j \in [\![1, p]\!] \\ \gamma_+ = l_j \gamma_j - a_j, & \forall j \in [\![1, p]\!], \end{cases} \tag{5.23}$$

hold for some $f_+ \in \mathbb{R}^+$ and $\gamma_+ \in \mathbb{R}$. It comes that

$$\begin{aligned} \forall j \in [\![1, p]\!], \forall r \in [\![0, n_j - 1]\!], \quad \frac{1}{f_j}(r - \gamma_j) &= \frac{1}{f_j}(r - l_j a_j - l_j \gamma_+) \\ &= \frac{1}{f_+}(l_j r - a_j - \gamma_+). \end{aligned} \tag{5.24}$$

Defining the quantities

$$\begin{cases} q_j[r] = l_j r - a_j, & \forall j \in [\![1, p]\!] \\ n_+ \geq \max_{j \in [\![1,p]\!]} \{q_j [n_j - 1]\}, \end{cases} \tag{5.25}$$

ensures that the grid $\mathcal{A}_+ = (f_+, \gamma_+, n_+)$ supports the system defined by $\mathbb{A}$. This achieves the sufficiency part, and thus the characterization of the existence of a common grid.

5.7.2 Conditions for minimality

Suppose that $\mathbb{A}$ admits a common grid, it is clear that exactly one element of $\mathcal{C}(\mathbb{A})$ reaches the minimal order n_Φ. Denote by $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ this element. Moreover, denote by $\{l_j\} \in \mathbb{N}^p$ and $\{a_j\} \in \mathbb{Z}^p$ the elements characterizing the grid expansion of $\mathbb{A}$ onto $\mathcal{A}_\Phi$ defined in (5.23), and let $\delta = \gcd\left(\{a_j\}_{j \in [\![1,p]\!]} \cup \{l_j\}_{j \in [\![1,p]\!]}\right)$. By equation (5.24), one has

$$\forall j \in [\![1, p]\!], \forall r \in [\![0, n_j - 1]\!], \quad \frac{1}{f_j}(r - \gamma_j) = \frac{\delta}{f_\Phi}\left(\frac{l_j}{\delta} r - \frac{a_j}{\delta} - \frac{\gamma_\Phi}{\delta}\right),$$

Thus the grid $\mathcal{A}_\Phi = \left(\frac{f_\Phi}{\delta}, \frac{\gamma_\Phi}{\delta}, \lceil \frac{n_\Phi}{\delta} \rceil\right)$ supports $\mathbb{A}$ and belongs to $\mathcal{C}(\mathbb{A})$. My minimality of $\mathcal{A}_\Phi$ one has $\lceil \frac{n_\Phi}{\delta} \rceil \geq n_\Phi$ and we conclude that $\delta = 1$. Moreover, the minimality implies that the first and the last samples of the grid $\mathcal{A}_\Phi$ must be acquired by an element of $\mathbb{A}$, otherwise the shorter grids $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi - 1, n_\Phi - 1)$, or $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi + 1, n_\Phi - 1)$ would also support $\mathbb{A}$. Using the properties (5.25)

$$\begin{cases} \forall j \in [\![1, p]\!], & \gamma_\Phi = l_j \gamma_j - a_j \\ \exists j \in [\![1, p]\!], & a_j = 0 \\ \forall j \in [\![1, p]\!], & a_j \leq 0, \end{cases}$$

which implies $\gamma_\Phi = \max_{j \in [\![1,p]\!]} \{l_j \gamma_j\}$, ensuring that the conditions describing the minimal grid stated in Proposition 5.1 are necessary.

For the sufficiency, consider the grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ of $\mathcal{C}(\mathbb{A})$ where $\gamma_\Phi = \max_{j \in [\![1,p]\!]} \{l_j \gamma_j\}$

and with expansion parameters $\{l_j\} \in \mathbb{N}^p$ and $\{a_j\} \in \mathbb{Z}^p$ satisfying

$$\gcd\left(\{a_j\} \cup \{l_j\},\ j \in [\![1,p]\!]\right) = 1.$$

Let $\mathcal{A}' = (f', \gamma', n') \in \mathcal{C}(\mathbb{A})$ be any other grid and let by δ' its corresponding greatest common divisor. δ' devises every integer linear combination of $\{a_j\} \cup \{l_j\}$, and in particular every elements of the set $\{l_j r - a_j :\ j \in [\![1,p]\!], r \in [\![0, n_j - 1]\!]\}$. Therefore (f', γ') is identifiable to $(\delta' f_\Phi, \delta' \gamma_\Phi - b)$ for some $b \in \mathbb{Z}$. Moreover since γ_Φ is maximum, the grid $\mathcal{A}_\Phi$ samples an element of $\mathbb{A}$ at index 0, and thus $\mathcal{A}' \in \mathcal{C}(\mathbb{A})$ if only and only if $b \geq 0$. Finally it comes from equation (5.25) that n' must satisfy

$$\begin{aligned}
n' &\geq \max_{j \in [\![1,p]\!]} \left\{ q'_j \left[n_j - 1\right] \right\} \\
&\geq \max_{j \in [\![1,p]\!]} \left\{ \delta' l_j \left(n_j - 1\right) - \delta' l_j a_j + b \right\} \\
&\geq \max_{j \in [\![1,p]\!]} \left\{ l_j \left(n_j - 1\right) - a_j \right\} \\
&\geq n_\Phi,
\end{aligned}$$

demonstrating the sufficiency part, and concluding the proof of Proposition 5.1. □

Chapter 6

Conclusion

6.1 Summary

The line spectral estimation problem is an inverse problem consisting in recovering a sparse measure from the observation of its first trigonometric moments. This thesis aims to extend recent advances related to the resolvability of the line spectral estimation problem via the means of convex relaxations, and a principal focus is placed on the so-called total variation (TV) approach.

Chapter 1 settles the context of this thesis. The line spectrum estimation problem is defined in Section 1.3. The notion of *statistical spectral resolution limit*: the distance under which two spikes cannot be distinguished, is discussed in Section 1.4 and formally stated in Equation (1.11). The major existing algorithms to solve the line spectral estimation problem are presented, and a comprehensive background on the TV regularization approach is given in Section 1.6. A dual certifiability result, stated in Theorem 1.1, and linking the success of TV regularization to output the sparsest possible measure with the existence of a *dual certificate* is recalled from [18]. Moreover, the spectral resolution conjecture for TV regularization is formulated in Subsection 1.6.2.3 and states that the method should be successful to reconstruct any given sparse measure with support having a minimal distance exactly *twice larger* than the theorized statistical resolution limit when the number of measurement grows large. The semidefinite formulation of the TV approach is recalled in Proposition 1.3 and finds its roots unto the Fejér-Riesz Theorem 1.2.

Chapter 2 discusses the necessary minimal separation that is requested by the total variation approach in order to output the desired result. The previous existing bounds are discussed in Section 2.1. A novel result, showing the existence of sparse measures with minimal distance asymptotically close to twice the statistical resolution limit for which TV regularization is guaranteed to fail is introduced in Theorem 2.1 and *closes the necessary side of the spectral resolution conjecture 1.1.* The rest of the chapter aims to provide a demonstration of this result. In particular, the proposed proof technique relies on the introduction of an intermediate notion of *diagonalizing families* of trigonometric polynomials, which are intimately linked with the existence of a dual certificate to the convex problem, as stated in Theorem 2.2.

Chapter 3 focuses on the sufficiency condition of the spectral resolution conjecture. Generic properties of the *extremal interpolation problem* that a dual certificate has to verify

Tang '14 [65]	$\Delta_{\mathbb{T}}(X) > \frac{1}{\pi m}$
Duval, Peyré '14 [28]	$\Delta_{\mathbb{T}}(X) > \frac{1}{2m}$
Ferreira Da Costa, Dai '18	$\Delta_{\mathbb{T}}(X) > \frac{1}{m} - \frac{\delta}{m^2}$ and $m \geq M_\delta$

Table 6.1: Review of the existing bounds for the *necessary* separation condition

Candès, Fernandes-G. '12 [18]	$\Delta_{\mathbb{T}}(X) > \frac{2}{m}$ and $m \geq 128$
Fernandes-Granda '14 [33]	$\Delta_{\mathbb{T}}(X) > \frac{1.26}{m}$ and $m \geq 1000$
Ferreira Da Costa, Dai '18	$\Delta_{\mathbb{T}}(X) > \frac{2.5683}{m-1}$

Table 6.2: Review of the existing bounds for the *sufficient* separation condition

to guarantee the success of the convex approach are given in Section 3.1. A tour of the existing constructions of such certificates is given in Section 3.2. A novel construction of a *diagonalizing certificate* is proposed in Section 3.3 in order to overcome the flaws of the previous constructions. The novel Theorem 3.1 guarantees that the diagonalizing construction satisfies the hypothesis of the dual certifiability Theorem 1.1 *under a mild separation constraint* on the spikes of the measure to reconstruct. Also Theorem 3.1 cannot guarantee the success of the proposed construction up to the limit stated in Conjecture 1.1, empirical results highlights, unlike the previous existing constructions, that the diagonalizing certificate can guarantee an exact reconstruction up to the phase transition. The rest of the chapter proposes a proof of Theorem 3.1, which also has the advantage to be much lighter and direct than previous results in the literature.

Chapter 4 introduces the *partial line spectral estimation problem* in Section 4.1, which consists in projecting the moments into a low-dimensional subspace. It is shown in Section 4.2 that the particular semidefinite geometry of the problem does not allow to immediately conclude that a reduction of the number of observations implies a reduction of the *computation complexity* of the problem. The novel Theorem 4.2 states the conditions under which the partial line spectral estimation problem can be reformulated into a low-dimensional semidefinite program whose dimension only depends on the actual number of observations. Proposition 4.2 relates the tightness of this program to the existence of a dual certificate satisfying the *sparse Fejér-Riesz condition* given in Definition 4.1. Section 4.4 proposes to further improve the computational efficiency of the problem by deriving the steps of the scalable *alternative direction method of multipliers* in the case where the subsampling operator is a selection matrix.

Finally, Chapter 5 discusses an extension of the spectral estimation problem to the case of *multirate sampling systems* (MRSS), composed of multiple synchronized samplers processing the time domain at different frequencies and with different delays. It is shown in Section 5.2 that, if the samplers admit a common supporting grid introduced in Definition 5.1, spectral estimation can be *jointly performed* by solving a semidefinite problem similar to the one analyzed in Chapter 4. Theorem 5.1 guarantees that sparse spectra verifying certain separation conditions can be reconstructed at a sub-Nyquist sampling rate via the use of MRSS. The benefits in terms of spectral resolution, noise robustness and computational costs of the partial semidefinite approach are discussed in Section 5.4.

6.2 Open problems in line spectral estimation

Closing the spectral resolution conjecture. If the diagonalizing certificate construction proposed in Chapter 3 appears to satisfy the dual certificate conditions of Theorem 1.1 up to the conjectured resolution limit $\Delta_{\mathbb{T}}(X) > \frac{1}{m}$, the demonstration provided for Theorem 3.1 does not allow to conclude on the achievability of the limit. Instead, guarantees are only provided for $\Delta_{\mathbb{T}}(X) > \frac{\alpha_{\mathrm{res}}}{m}$ where $\alpha_{\mathrm{res}} \leq 2.5683$. Fixing this current gap, that is believed to be artificial, is a problem of major interests, since it will fully close the necessary side of the spectral resolution conjecture for the total variation approach.

The diagonalizing certificate and the resolution/precision tradeoff. A recent analysis of the Beurling-LASSO estimator in noisy environments (1.36) presented in [46] relates the precision of the reconstruction with the decay rate of the dual certificate at the spikes locations and its flatness in the regions that are far away from any elements in the support set X. As the analysis is conducted for the Jackson construction presented in Subsection 3.2.1 and requires a tedious inversion step, one can wonder how the use of the diagonalizing certificate could help to derive enhanced performance guarantees. In particular, in the non-collapsing regime where $\Delta_{\mathbb{T}}(X) = \omega\left(\frac{\alpha}{m}\right)$, the diagonalizing certificate is excepted to be sharper than the Jackson one, yielding potentially better robustness guarantees.

Compressed sensing and the sparse Fejér-Riesz condition. The sampling complexity of the line spectral estimation problem discussed in Subsection 4.2.1 is proven in [67] to be as low as logarithmic orders in the initial number of measurement without degrading the resolution order when considering random subsampling. However, it is important to understand up to which subsampling order the computational complexity can be reduced. This requires to extend the initial proof in [67] and show whether there exists random dual certificates verifying the sparse Fejér-Riesz condition introduced in Definition 4.1 with high probability if the number of samples is large enough.

6.3 Beyond line spectral estimation

Statistical resolution limits from generalized measurements The proof of the statistical resolution limit in the context of line spectral estimation was proven using the help of *extremal functions* preconditioning the bandlimited spectra [2]. A natural research direction would be to understand whether there is any possible extension of this theory to a broader class of point spread functions to highlight the existence of resolution limits in diverse sparsity models. In particular, a notion of generalized Vandermonde matrices can be formulated for shift invariant point spread functions, splines models, and for T-systems (*e.g.* Laplace measurements, Gaussian moments, Stieljies transform), which are used as sparsity prior in various signal processing scenarios.

Doubly gridless bilinear inverse problems The continuous blind gain phase calibration (BGPC) and the blind deconvolution (BD) problems are two instances of bilinear inverse problems occurring in image and array processing characterized by bilinear observation maps of the form

$$y = \mathcal{B}_{\mathrm{BGPC}}\left(\mathbf{X}, u\right) = \mathrm{diag}\left(u\right)\mathbf{F}\mathbf{X}, \tag{BGPC}$$

$$y = \mathcal{B}_{\mathrm{BD}}\left(\mathbf{X}, u\right) = \mathbf{F}^{*}\mathrm{diag}\left(u\right)\mathbf{F}\mathbf{X} \tag{BD}$$

whereby $\mathbf{F}$ is often assumed to be a discrete Fourier transform matrix, while $\mathbf{X}$ represents the system inputs. One commonly assumes a subspace or low-rank prior on the inputs to ensure the uniqueness of the solution up to some transforms. Recovery and robustness guarantees by ℓ_1-minimization approaches have been provided for those discrete problems [76]. One could study whether such results are extendable to the gridless case where $\mathbf{X}$ becomes a finite set of functions lying in some Hilbert space (*e.g.* bandlimited functions), $\mathbf{F}$ becomes a discrete-time Fourier operator, and tackle the problem though the lense of TV regularization.

Bibliography

[1] P. Akhilesh and D. S. Ramana, "A remark on the Beurling-Selberg function", *Acta Mathematica Hungarica*, vol. 139, no. 4, pp. 354–362, 2013.

[2] C. Aubel and H. Bölcskei, "Vandermonde matrices with nodes in the unit disk and the large sieve", *Applied and Computational Harmonic Analysis*, 2017.

[3] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances", *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[4] T. Bendory, S. Dekel, and A. Feuer, "Exact recovery of non-uniform splines from the projection onto spaces of algebraic polynomials", *Journal of Approximation Theory*, vol. 182, pp. 7–17, 2014.

[5] ——, "Robust recovery of stream of pulses using convex optimization", *Journal of Mathematical Analysis and Applications*, vol. 442, no. 2, pp. 511–536, 2016.

[6] 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, Dec. 2013.

[7] ——, "Atomic norm denoising with applications to line spectral estimation", *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[8] G. Blekherman, P. A. Parrilo, and R. R. Thomas, *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.

[9] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations", *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.

[10] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems", *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

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

[13] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.

[14] J.-F. Cai, X. Qu, W. Xu, and G.-B. Ye, "Robust recovery of complex exponential signals from random gaussian projections via low rank hankel matrix reconstruction", *Applied and Computational Harmonic Analysis*, vol. 41, no. 2, pp. 470–490, 2016.

[15] E. J. Candes and T. Tao, "Decoding by linear programming", *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[16] E. J. Candès, "The restricted isometry property and its implications for compressed sensing", *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[17] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data", *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[18] ——, "Towards a mathematical theory of super-resolution", *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[19] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements", *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[20] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[21] C. Carathéodory, "Über den variabilitätsbereich der fourier'schen konstanten von positiven harmonischen funktionen", *Rendiconti Del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 193–217, 1911.

[22] 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, 2012.

[23] S. Chrétien and H. Tyagi, "Multi-kernel unmixing and super-resolution using the modified matrix pencil method", *arXiv:1807.02862*, 2018.

[24] Y. De Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation", *Journal of Mathematical Analysis and Applications*, vol. 395, no. 1, pp. 336–354, 2012.

[25] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures", *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, 2017.

[26] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix", *IEEE Transactions on signal processing*, vol. 55, no. 5, pp. 1741–1757, 2007.

[27] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*. Springer, 2007, vol. 103.

[28] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution", *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[29] A. Eftekhari, J. Tanner, A. Thompson, B. Toader, and H. Tyagi, "Non-negative super-resolution is stable", 2018.

[30] A. Eftekhari and A. Thompson, "A bridge between past and present: Exchange and conditional gradient methods are equivalent", *arXiv:1804.10243*, 2018.

[31] Y. C. Eldar, *Sampling theory: Beyond bandlimited systems.* Cambridge University Press, 2015.

[32] C. Fernandez-Granda, "Super-resolution of point sources via convex programming", *Information and Inference: A Journal of the IMA*, vol. 5, no. 3, pp. 251–303, 2016.

[33] ——, "Super-resolution of point sources via convex programming", *Information and Inference: A Journal of the IMA*, vol. 5, no. 3, pp. 251–303, 2016.

[34] J. Fourier, *Théorie analytique de la chaleur.* Chez Firmin Didot, père et fils, 1822.

[35] R. Heckel, "Super-resolution mimo radar", in *Information Theory (ISIT), 2016 IEEE International Symposium on*, IEEE, 2016, pp. 1416–1420.

[36] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization", *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4001–4023, Sep. 2018.

[37] W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik", in *Original Scientific Papers Wissenschaftliche Originalarbeiten*, Springer, 1985, pp. 478–504.

[38] H. Hindi, "A tutorial on convex optimization II: Duality and interior point methods", in *2006 American Control Conference*, Jun. 2006.

[39] B. Huang, M. Bates, and X. Zhuang, "Super-resolution fluorescence microscopy", *Annual review of biochemistry*, vol. 78, pp. 993–1016, 2009.

[40] A. E. Ingham, "Some trigonometrical inequalities with applications to the theory of series", *Mathematische Zeitschrift*, vol. 41, no. 1, pp. 367–379, 1936.

[41] S. M. Kay, *Modern spectral estimation.* Pearson Education India, 1988.

[42] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions", *Acta Mathematica*, vol. 117, no. 1, pp. 37–52, 1967.

[43] J. B. Lasserre, "Global optimization with polynomials and the problem of moments", *SIAM Journal on optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[44] J.-B. Lasserre, *Moments, positive polynomials and their applications.* World Scientific, 2010, vol. 1.

[45] M. A. Lerma, "Some applications of extremal functions in Fourier analysis", *Analysis*, pp. 1–21, 1996.

[46] Q. Li and G. Tang, "Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision", *arXiv preprint arXiv:1612.01459*, 2016.

[47] W. Liao and A. Fannjiang, “MUSIC for single-snapshot spectral estimation: Stability and super-resolution”, *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, 2016.

[48] J. Lindberg, “Mathematical concepts of optical superresolution”, *Journal of Optics*, vol. 14, no. 8, 2012.

[49] A. Moitra, “Super-resolution, extremal functions and the condition number of Vandermonde matrices”, arXiv: `1408.1681`.

[50] M. Negreanu and E. Zuazua, “Discrete ingham inequalities and applications”, *Comptes Rendus Mathematique*, vol. 338, no. 4, pp. 281–286, 2004.

[51] H. Nyquist, “Certain topics in telegraph transmission theory”, *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[52] P. Pal and P. P. Vaidyanathan, “Coprime sampling and the MUSIC algorithm”, in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, IEEE, 2011, pp. 289–294.

[53] G. R. de Prony, “Essai expérimental et analytique: Sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l’eau et de la vapeur de l’alkool, à différentes températures”, *J. de l’École Polytechnique*, vol. 1, pp. 24–76, 1795.

[54] S. Qin, Y. D. Zhang, M. G. Amin, and A. M. Zoubir, “Generalized coprime sampling of Toeplitz matrices for spectrum estimation”, *IEEE Transactions on Signal Processing*, vol. 65, no. 1, pp. 81–94, Jan. 2017.

[55] L. Rayleigh, “Investigations in optics, with special reference to the spectroscope”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, no. 49, pp. 261–274, 1879.

[56] R. Sanyal, F. Sottile, and B. Sturmfels, “Orbitopes”, *Mathematika*, vol. 57, no. 2, pp. 275–314, 2011.

[57] G. Schiebinger, E. Robeva, and B. Recht, “Superresolution without separation”, *Information and Inference: A Journal of the IMA*, vol. 7, no. 1, pp. 1–30, 2017.

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

[59] C. E. Shannon, “Communication in the presence of noise”, *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[60] A. Shapiro, “Semi-infinite programming, duality, discretization and optimality conditions”, *Optimization*, vol. 58, no. 2, pp. 133–161, Feb. 2009.

[61] 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.

[62] D. Slepian and H. O. Pollak, “Prolate spheroidal wave functions, Fourier analysis and uncertainty - I”, *Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[63] 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, Aug. 1999.

[64] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation", *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

[65] G. Tang, "Resolution limits for atomic decompositions via Markov-Bernstein type inequalities", *2015 International Conference on Sampling Theory and Applications, SampTA 2015*, pp. 548–552, May 2015.

[66] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries-just discretize", in *Signals, Systems and Computers, 2013 Asilomar Conference on*, 2013, pp. 1043–1047.

[67] 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, Nov. 2013.

[68] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The probability of a subspace swap in the SVD", *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, Mar. 1995.

[69] 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, Aug. 2001.

[70] Y. Traonmilin and R. Gribonval, "Stable recovery of low-dimensional cones in Hilbert spaces: One rip to rule them all", *Applied and Computational Harmonic Analysis*, 2016.

[71] Y. Traonmilin, G. Puy, R. Gribonval, and M. E. Davies, "Compressed sensing in Hilbert spaces", in *Compressed Sensing and its Applications*, Springer, 2017, pp. 359–384.

[72] P. Turán, "On rational polynomials", *Acta Univ. Szeged, Sect. Sci. Math*, pp. 106–113, 1946.

[73] M. Unser, "Sampling—50 years after Shannon", *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[74] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation", *IEEE transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[75] C. Vinzant, "What is... a spectrahedron", *Notices Amer. Math. Soc*, vol. 61, no. 5, pp. 492–494, 2014.

[76] L. Wang and Y. Chi, "Blind deconvolution from multiple sparse inputs", *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1384–1388, 2016.

[77] X. Wei and P. L. Dragotti, "Guaranteed performance in the FRI setting", *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1661–1665, Oct. 2015.

[78] A. Zygmund, *Trigonometric series*. Cambridge university press, 2002, vol. 1.