A CONVEX-PROGRAMMING FRAMEWORK FOR SUPER-RESOLUTION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Carlos Fernandez-Granda

August 2014

A mis padres y mi abuela Adela

Abstract

This thesis proposes a general framework to extract information from low-resolution data, a crucial challenge in applications ranging from microscopy, astronomy and medical imaging to geophysics, signal processing and spectroscopy. Assume that we only have information about the spectrum of a superposition of point sources in the low-frequency band $[-f_{\text{lo}}, f_{\text{lo}}]$. We show that as long as the sources are separated by $2/f_{\text{lo}}$, solving a simple convex program achieves exact recovery, in the sense that the original signal is the unique solution to the optimization problem. This is established by leveraging a novel proof technique which constructs a dual certificate by interpolation. In addition, we describe how to implement the recovery algorithm by solving a finite-dimensional semidefinite program, which locates the support of the signal with infinite precision. The guarantees extend to higher dimensions and other models. They imply, for instance, that we can recover a piecewise-smooth function from bandlimited data by super-resolving the discontinuity points. We also provide an analysis of the stability of our method in the presence of noise. On the one hand, we prove that it is possible to extrapolate the spectrum of the original signal up to a frequency $f_{\text{hi}} > f_{\text{lo}}$ to obtain an approximation error between the higher-resolution reconstruction and the truth that is proportional to the noise level times the square of the *super-resolution factor* (SRF) $f_{\text{hi}}/f_{\text{lo}}$. On the other hand, we derive support-detection guarantees that quantify the precision with which we can determine the location of each individual point source in the signal.

Acknowledgements

First and foremost I am deeply grateful to my advisor Emmanuel Candès. Throughout my PhD he has pushed me to work on hard important problems and at the same time given me the freedom to pursue my own interests. I cannot thank him enough for his constant support and for making me question *why we do what we do* at every step.

I am very thankful to Abbas El Gamal for his support and for teaching me about teaching; to Yi Ma for hosting me at Microsoft Research and for his contagious enthusiasm about computer vision; and to Yves Meyer, Gil Strang and Thomas Strohmer for their kindness. I would also like to thank Stephen Boyd, Andrea Montanari and George Papanicolau for being part of my thesis committee. In addition, I am indebted to Stephen and to Michael Grant for CVX, which has been very useful in my research.

I am happy to acknowledge many interesting interactions with my fellow group members over the years: Mahdi, Yaniv, David Gross, Ewout, Carlos, Xiaodong, Veniamin, Mark, Lester, Rina, Deanna, Jérôme, Kahye, Alexandra, Vlad, Weijie, Reinhard, Kenji and Raf.

I gratefully acknowledge financial support from Fundación Caja Madrid and from Fundación La Caixa, which does a great job of supporting Spanish students abroad.

On a more personal note, I would like to thank Mahdi for our conversations about castles and lawnmowers; Ricardo and Juan for some memorable nights; little Dani, Uri, Félix, Idoia and Míkel for making Stanford feel more like home; Jacobo for those dinners; and Álvaro, Gardeta, Isaías, Jalvo and Nico for becoming the dysfunctional siblings I never had.

I cannot end without thanking Nadine for making me smile and my parents, to whom I owe everything.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As Lord Rayleigh pointed out in his seminal 1891 paper On Pin-hole Photography [82], *it has long been known that the resolving power of lenses, however perfect, is limited*. Indeed, diffraction imposes an inflexible limit on the resolution of any optical system [73]. However we are often interested in information that is only apparent beyond this limit. For instance, in microscopy [74], astronomy [81] or medical imaging [58] it may be challenging to discern cellular structures, celestial bodies or incipient tumors from the available data. This is illustrated by the image on the left of Figure 1.1[1], which shows measurements of the interior of a cell obtained by fluorescence microscopy. The limited resolution of the microscope produces aliasing artifacts that completely obscure the fine-scale details of the image. The aim of super-resolution is to uncover such fine-scale structure from coarse-scale measurements. This is a fundamental problem in optics [73], electronic imaging [75], where photon shot noise constrains the minimum possible pixel size, and many other areas including spectroscopy [60], radar [76], non-optical medical imaging [67] and seismology [68].

The main objective of this thesis is to develop and analyze effective methods for signal estimation from low-resolution measurements; a problem that we call *super-resolution*. We warn the reader that this word has different meanings in different disciplines. In optics it is usually used to describe data-acquisition techniques designed to overcome the diffraction limit [73]. In image-processing and computer-vision

[1] I would like to thank Veniamin Morgenshtern for these two images.

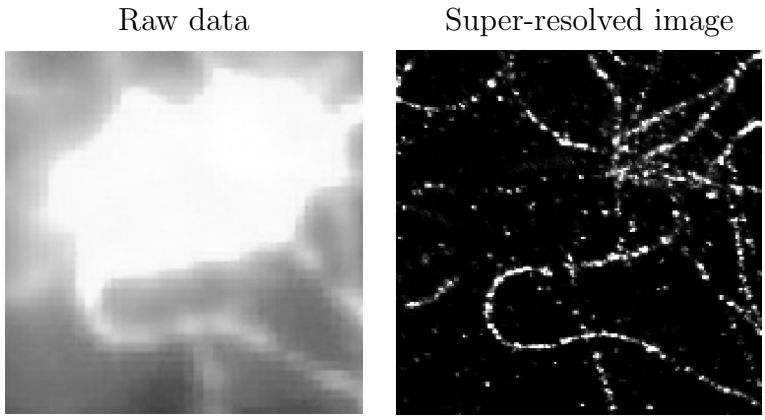


Figure 1.1: An example of super-resolution applied to cell imaging using fluorescence microscopy data. The measurements on the left are contaminated by noise and heavy aliasing. Super-resolving the probes reveals the fine-scale details of the cell, as we can see on the right.

applications the term tends to refer to the problem of obtaining a high-resolution image either from several low-resolution images [77] or by upsampling a single image while preserving its edges and hallucinating textures in a reasonable way [52]. Here, super-resolution denotes the inverse problem of estimating a signal from low-pass measurements. In contrast to the previous definitions, we assume that the data-acquisition mechanism is *fixed* and we aim to *recover* rather than hallucinate the lost high-resolution information.

1.1 Super-resolution of point sources

In order to super-resolve a signal it is necessary to leverage some prior knowledge about its structure. Otherwise the problem is hopelessly ill posed; the missing spectrum can be filled in arbitrarily to produce estimates that correspond to the data. In this work we consider signals that are well modeled as superpositions of point sources, although our analysis extends to other classes of signals such as piecewise-differentiable functions. Point sources are an important model in many applications.

They are used to represent celestial bodies in astronomy [55], neuron spikes in neuroscience [83] or line spectra in signal processing and spectroscopy [70, 95]. In addition, locating pointwise fluorescent probes is a crucial step in some optical super-resolution procedures capable of handling more complicated objects. Techniques such as photoactivated localization microscopy (PALM) [13, 62] or stochastic optical reconstruction microscopy (STORM) [88] are based on localizing probes that switch randomly between a fluorescent and a non-fluorescent state. To obtain an image of a certain object, multiple frames are gathered. Each frame consists of a superposition of blurred light sources that correspond to the active probes. Deblurring these sources and combining the results allows to super-resolve the object of interest. The image to the right of Figure 1.1 was generated in this way.

Mathematically, we will be interested on signals modeled as sums of Dirac measures or *spikes* supported on a set T

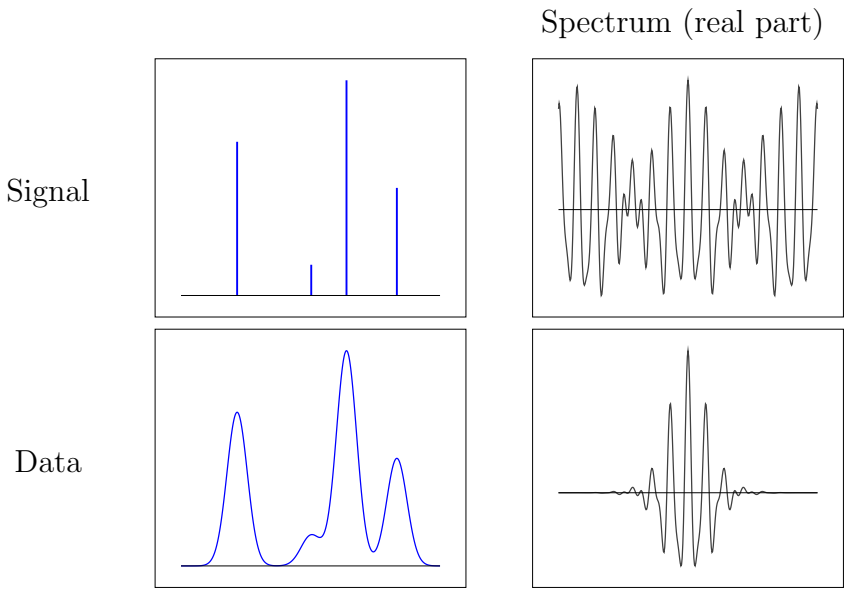
$$x := \sum_{t_j \in T} c_j \delta_{t_j}, \tag{1.1.1}$$

where δ_τ is a Dirac measure at τ and the amplitudes a_j may be complex valued. In many of the applications mentioned previously a reasonable model for the measurement process is the convolution of the signal with a low-pass point spread function (PSF) ϕ. This convolution smooths out the fine-scale details producing a low-resolution version of the original signal,

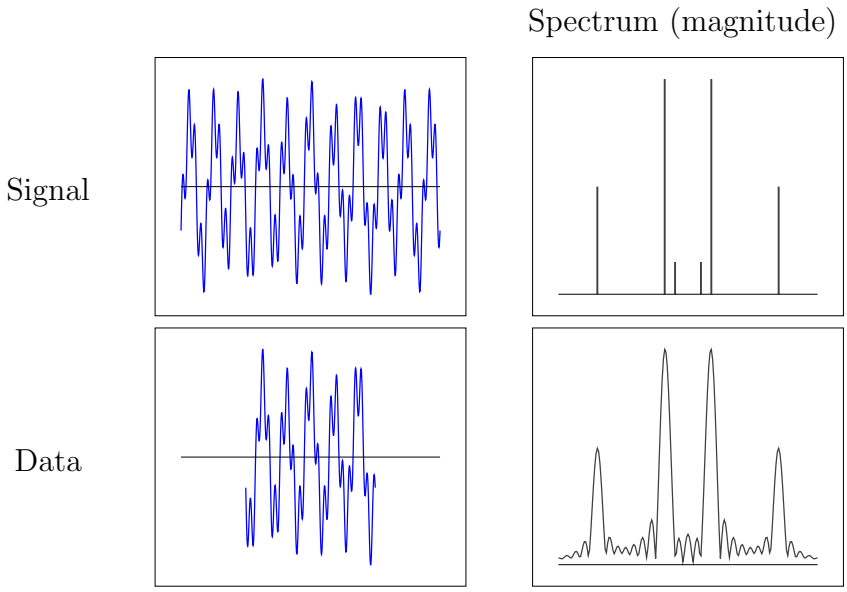
$$x_{\mathrm{LR}}(t) := \phi * x(t) = \sum_{t_j \in T} c_j \phi(t - t_j), \tag{1.1.2}$$

as illustrated at the top of Figure 1.2. The aim of *spatial super-resolution* is to estimate x from x_{LR}, a problem which is often also referred to as *deconvolution*. Recovering the lost fine-scale information amounts to extrapolating the spectrum of x. This becomes apparent when we consider the sensing mechanism in the frequency domain. If the cut-off frequency of the PSF is equal to f_{lo}, the spectrum of x_{LR}

$$\hat{x}_{\mathrm{LR}} = \widehat{\phi}\, \hat{x} = \widehat{\phi}\, \Pi_{[-f_{\mathrm{lo}}, f_{\mathrm{lo}}]^d}\, \hat{x},$$



Spatial super-resolution



Spectral super-resolution

Figure 1.2: Schematic illustration of spatial and spectral super-resolution.

where d is the ambient dimension, $\hat{f}$ denotes the Fourier transform of a function or measure f and $\Pi_{[-f_{\rm lo},f_{\rm lo}]^d}$ is an indicator function that is zero out of the set $[-f_{\rm lo}, f_{\rm lo}]^d$. Consequently, the high frequency information in the signal is suppressed in the data. Super-resolution aims to extrapolate this information from the low-pass measurements.

In this thesis, we assume that the PSF of the sensing mechanism is known. This is usually a realistic assumption in point-source super-resolution as long as the measurement process is indeed space-invariant. In such cases the PSF can be estimated by locating an isolated blurred source in the data. However, for general images or PSFs that are not low pass, which arise for instance due to motion blurring, it is necessary to jointly estimate the PSF and the signal; a problem known as *blind deconvolution* [22].

An important instance of the model given by (1.1.1) and (1.1.2) is when the signal is one dimensional and the Fourier transform of the PSF ϕ is constant over $[-f_{\rm lo}, f_{\rm lo}]$, i.e. ϕ is a periodized sinc or Dirichlet kernel. In this case, since the support of $x_{\rm LR}$ is restricted to the unit interval, it follows from the sampling theorem that its spectrum is completely determined by the discrete samples

$$y(k) = \hat{x}_{\rm LR}(k) = \int_0^1 e^{-i2\pi kt} x(\mathrm{d}t) = \sum_j c_j e^{-i2\pi k t_j}, \quad k \in \mathbb{Z},\, |k| \leq f_{\rm lo}, \tag{1.1.3}$$

where we assume for simplicity that $f_{\rm lo}$ is an integer. In a more compact form,

$$y = \mathcal{F}_n\, x$$

where $y \in \mathbb{C}^n$ and $\mathcal{F}_n$ is the linear operator that maps a measure or function to its lowest $n := 2f_{\rm lo} + 1$ Fourier coefficients.

The interest of analyzing data of the form (1.1.3) is twofold. First, many of the conclusions obtained from such an analysis translate almost directly to cases where the sensing mechanism is based on low-pass PSFs other than the periodized sinc. Second, the model is of prime importance in signal-processing applications. Indeed, if we interpret x as the atomic spectrum of a multi-sinusoidal signal then the data y have a very natural interpretation: they correspond to a finite number of samples of

the signal measured at the Nyquist rate. Truncating the samples in the time domain is equivalent to convolving the spectrum of the signal with a periodized sinc, which induces aliasing in the frequency domain as depicted in the lower half of Figure 1.2. *Spectral super-resolution* is the problem of estimating the line spectra x from such measurements.

In the signal-processing literature, spectral super-resolution has particular significance in direction-of-arrival (DOA) problems where the aim is to determine the direction from which a propagating wave arrives at an array of sensors. In fact, some of the most popular techniques for spectral super-resolution, such as ESPRIT [85], have been motivated by this problem. We refer the reader to [1, 108] for a detailed account of DOA problems and their connection to spectral super-resolution.

Before presenting an outline of the thesis and an overview of the main results, we provide some context by describing the state of the art for tackling the problems of spatial and spectral super-resolution.

1.2 Previous art

Existing super-resolution methods can be divided into two main categories: nonparametric approaches, widely used in both spatial and spectral super-resolution, and parametric approaches based on Prony's method, which are popular mainly in spectral super-resolution. The distinction is based on whether the algorithm receives the number of sources present in the original signal as input. If this is the case, the method is considered parametric.

1.2.1 Nonparametric estimation

Nonparametric techniques can be classified in turn depending on the PSF of the sensing process. We begin by describing techniques applied in spatial super-resolution and then move on to those adapted to spectral super-resolution, where the PSF is a periodized sinc.

Spatial super-resolution

In this section, we assume that the low-resolution measurements of our signal of interest x are obtained as in (1.1.2) and in addition are corrupted by a certain perturbation z,

$$x_{\text{data}}(t) := x_{\text{LR}}(t) + z(t) = \phi * x(t) + z(t). \tag{1.2.1}$$

Imagine that we know that the signal consists of a single blurred source. In this case, it seems reasonable to estimate the location of the source by fitting the PSF to the data. The fit can be carried out, for instance, by minimizing a least-squares cost function. This is a common choice for two reasons: first, the minimizer can be computed very efficiently, as we explain below, and second, the result corresponds to the maximum-likelihood estimate under Gaussian noise. For a single source, performing a least-squares fit amounts to finding the shift that maximizes the inner product between the data and the PSF ϕ shifted by $\tilde{t}$, which is denoted by $\phi_{\tilde{t}}$. More precisely, the estimated location

$$\begin{aligned} t_{\text{est}} &= \arg\min_{\tilde{t}} \min_{\alpha \in \mathbb{R}} ||x_{\text{data}} - \alpha\, \phi_{\tilde{t}}||_2 \\ &= \arg\max_{\tilde{t}} |\langle x_{\text{data}}, \phi_{\tilde{t}} \rangle| . \end{aligned} \tag{1.2.2}$$

Here $\langle \cdot, \cdot \rangle$ is the inner product in $\mathcal{L}_2$ (or $\mathbb{R}^N$ for some integer N if the data are discretized). The procedure is equivalent to performing detection using matched filters [54], a popular technique in signal processing and communications.

When more than one point source is present, estimating the signal support may be significantly more difficult, depending on to what extent the data corresponding to the different point sources can be teased apart. If the sources are far enough with respect to the decay of the PSF, the aliasing caused by neighboring blurred sources may be negligible. In this case, one can just fit the PSF locally to each of them. This approach is often followed in fluorescence microscopy [13, 62, 88], where a Gaussian parametric model can be used to approximate the PSF [110].

If there is interference between the blurred sources, then the problem becomes

much more challenging. An option is to apply (1.2.2) iteratively to estimate the sources one at a time in a greedy fashion. In [5] this scheme is applied to seismography and in [63] to astronomy, where it is known as CLEAN. Assuming a statistical model on the distribution of the spikes and on the noise allows to develop more sophisticated schemes. For example, a common choice is to model the noise as Gaussian and the spikes as following a Bernouilli prior. An approximation to the maximum a posteriori (MAP) estimate can then be obtained by expectation maximization [31] or by successively optimizing the estimate with respect to the location of one of the spikes while fixing the rest, a technique known as Single Most Likely Replacement (SMLR) [69] which is equivalent to maximizing the MAP criterion by coordinate ascent. Although these methods may work well in some cases, for many applications super-resolution of sources that are clustered is very challenging. As a result, in fluorescence microscopy the data is usually measured at a rate that ensures a certain separation between the active fluorophores [13, 62, 88]. Developing methods capable of operating at higher densities is an active research area (see for example [112] for an approach that is very related to the methods developed in this thesis).

Spectral super-resolution

In this section we focus on the case where ϕ is a periodized sinc, so that the measurements are described by (1.1.3) and we have access to

$$y_{\text{data}} = \mathcal{F}_n \, x + z, \tag{1.2.3}$$

where $y_{\text{data}} \in \mathbb{C}^n$ and $z \in \mathbb{C}^n$ is an unknown perturbation. By far the most popular nonparametric method for analyzing this kind of data is the periodogram [91] (see also Chapter 2 in [98] for a thorough analysis). The periodogram is computed by projecting the data onto signal space, in order to obtain an approximation to the spectrum of the original signal,

$$P = \mathcal{F}_n^* \, y_{\text{data}}.$$

In a high signal-to-noise regime,

$$P(t) \approx \left(\mathcal{F}_n^* \mathcal{F}_n \, x\right)(t) = \sum_{t_j \in T} c_j D_{f_{\text{lo}}}(t - t_j),$$

where $D_{f_{\text{lo}}}$ is the periodized sinc or Dirichlet kernel

$$D_{f_{\text{lo}}}(t) := \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} e^{i2\pi kt} = \begin{cases} 1 & \text{if } t = 0 \\ \frac{\sin((2f_{\text{lo}}+1)\pi t)}{(2f_{\text{lo}}+1)\sin(\pi t)} & \text{otherwise .} \end{cases}$$

Just to clarify, in most applications t would index the frequency domain, *not* the time domain, but we keep this notation for the sake of consistency. Also, f_{lo} no longer has a physical meaning (beyond determining the number of time-domain samples) in contrast to spatial super-resolution where it represents the cut-off frequency of the sensing mechanism.

In the absence of noise, the periodogram is consequently the result of convolving the line spectra with a periodized sinc, which is equivalent to the projection of the original line spectra onto a space of low-pass functions. Since $D_{f_{\text{lo}}}$ is unimodal, a straightforward way to detect the line spectra is to locate local maxima of P that are far apart. The problem with this approach is that the side lobes corresponding to large blurred spikes may mask the presence of smaller spikes. As a result, the periodogram is not very useful if the spikes are not far enough from each other or if their amplitudes differ substantially, *even if no noise is present in the data*. The image at the center of Figure 1.3 illustrates this: detecting some of the lower-amplitude spikes from the periodogram is impossible.

In order to alleviate this problem one can apply a window function $\hat{w} \in \mathbb{C}^n$ to the data before computing the periodogram,

$$y_{\hat{w}} = y_{\text{data}} \cdot \hat{w},$$

where $\cdot$ denotes pointwise multiplication. If the noise is small, the windowed periodogram

$$P_{\hat{w}}(f) = \mathcal{F}_n^* \, y_{\hat{w}} \approx \sum_{t_j \in T} c_j w\left(t - t_j\right),$$

where w denotes the inverse Fourier transform of the window function. Ideally, w should be as *spiky* as possible to make it easier to locate the support of the line spectra from the windowed periodogram. However, this is challenging due to the constraint that $\hat{w}$ has finite support and hence w is a low-pass function. In the image on the top right of Figure 1.3 we apply a Gaussian window to the data. To be more precise, we set $\hat{w}$ to be a truncated Gaussian, so that w is equal to the convolution between a periodized sinc and a periodized Gaussian. The resulting periodogram, shown at the center of Figure 1.3, has much less *spectral leakage* from the largest line spectra, due to the fact that the Gaussian window has lower side lobes than the periodized sinc. However, the latter is spikier at the origin, which allows to better distinguish neighboring line spectra with similar amplitudes. This is also apparent in the image, especially for the line spectra on the right. In general, designing an adequate window implies finding a good tradeoff between the width of the main lobe and the height of the side lobes. We refer the reader to [59] for a detailed account of design considerations and types of window function.

To conclude, the periodogram is a useful method to obtain a general idea of the spectral structure of the signal and can provide insight as to the number of sources and their approximate location, especially if the data is processed with a suitable window function. However, the method does not provide a precise localization of the line spectra of the original signal even in the absence of noise. This is the aim of the techniques described in the following section.

1.2.2 Parametric methods for spectral super-resolution

In this section we review parametric methods for spectral analysis based on Prony's method. These algorithms take the number of line spectra of the original signal

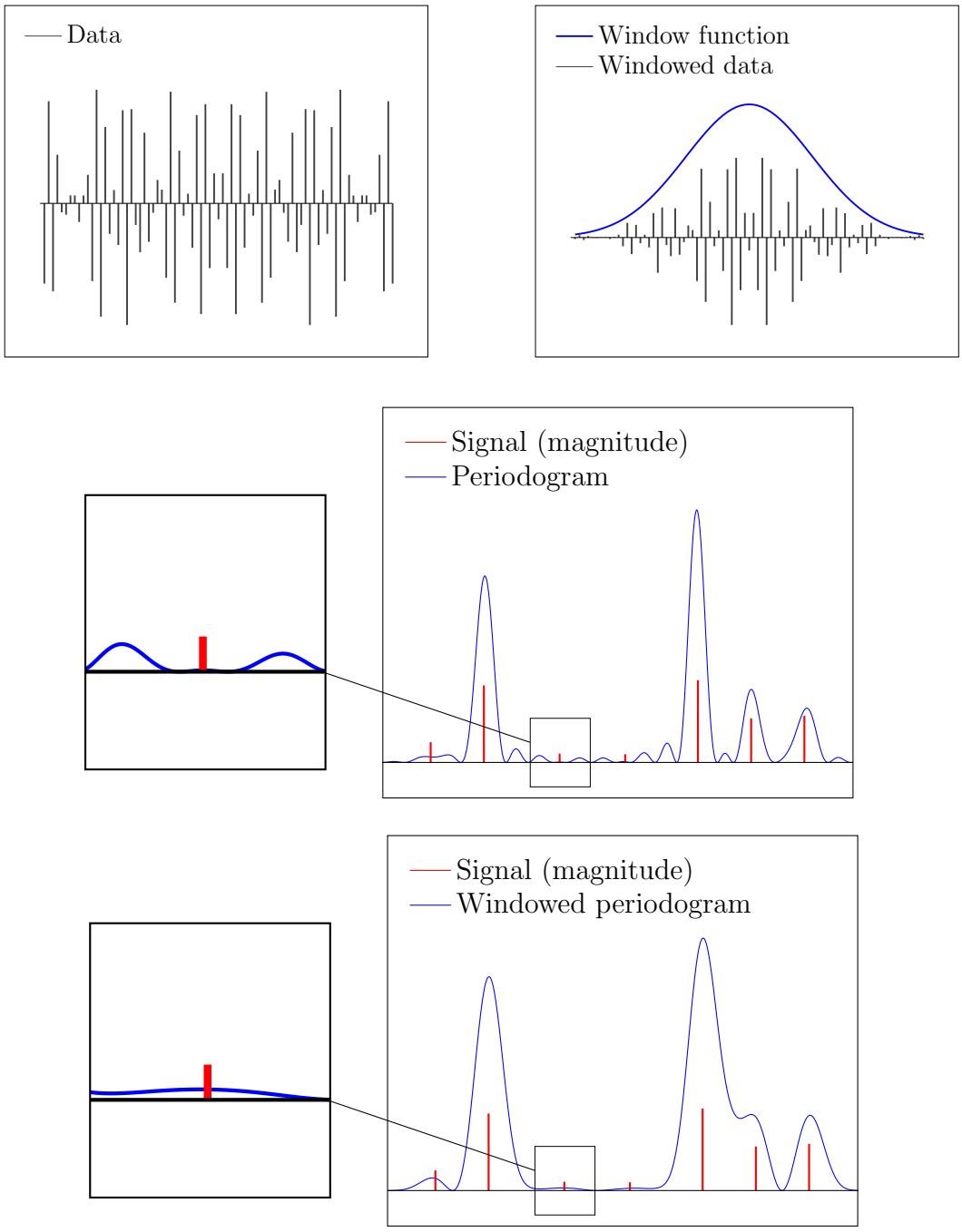


Figure 1.3: Example of data in spectral super-resolution generated according to (1.1.3) before (top left) and after applying a window function (top right). No noise is added to the data. Below we can see the periodogram (center) and windowed periodogram (bottom) computed from the data. The scaling of the periodograms is set so that both the large and small peaks can be seen on the plot.

as input and return an estimate of their location. We would also like to mention that algebraic techniques related to these parametric methods have been applied to the location of singularities in the reconstruction of piecewise polynomial functions from a finite number of Fourier coefficients (see [3, 6, 47] and references therein). The theoretical analysis of these methods proves their accuracy up to a certain limit related to the number of measurements.

Prony's method

The following simple observation is the key idea behind Prony's method [80]: the position of s line spectra can be encoded as the zeros of a trigonometric polynomial of order s. In fact, in the absence of noise, it turns out that we can always find such a polynomial and achieve perfect spectral super-resolution by locating its roots. Unfortunately, this good news comes with a caveat: the method can be extremely unstable in the presence of noise, as we will show in a moment.

A first question that arises is whether a polynomial of order s with roots on the support of the signal actually exists for any possible support. The answer is yes: assume that x is as in (1.1.1) and $|T| = s$, then

$$P_{\text{Prony}}(t) := \prod_{j=1}^{s} \left(1 - e^{i2\pi(t-t_j)}\right) = 1 + \sum_{l=1}^{s} v_l e^{i2\pi lt}, \quad v_0 := 1, \tag{1.2.4}$$

is a polynomial of order s with roots located exactly on T.

Now, how can we find such a polynomial? We must relate the coefficients of the polynomial to the available data. In spectral super-resolution the data correspond to regular samples of the multi-sinusoidal signal, or equivalently to the Fourier coefficients of the line spectra. We assume that we have access to n such measurements $\hat{x}_1, \ldots, \hat{x}_n$. This model is exactly equivalent to (1.1.3). Observe that for any integer k,

$$\int_0^1 e^{i2\pi kt} P_{\text{Prony}}(t) x(\mathrm{d}t) = \sum_{j=0}^{s} c_j e^{i2\pi k t_j} \left(1 - e^{i2\pi(t_j - t_j)}\right) = 0.$$

By Parseval's Theorem this implies

$$\sum_{l=-\infty}^{\infty} v_l \hat{x}_{k-l} = \sum_{l=0}^{s} v_l \hat{x}_{k-l} = 0, \tag{1.2.5}$$

which provides us with an equation linking the coefficients of the Prony polynomial and the Fourier coefficients of the signal for each k. The coefficients of the polynomial can be interpreted as a filter that suppresses the signal x. This motivates the use of the term *annihilating filter* in works on estimation of signals with a *finite rate of innovation* [17, 109]. These are signals with a parametric representation that has a finite number of degrees of freedom. The superpositions of Dirac measures that we consider in spectral super-resolution belong to this class. For them, this approach reduces to Prony's method (see Section 3 in [109]).

If we select k to be between $s+1$ and $n+s$ then (1.2.5) only involves the available data. This provides a system of n equations, so that as long as $n \geq 2s$, we have enough equations to solve for the s unknown coefficients of the Prony polynomial (recall that we fix $v_0 = 1$). Prony's method consists of solving this system and then decoding the location of the line spectra by rooting the corresponding polynomial. For simplicity we assume that $n = 2s$:

Input: Number of line spectra s, regular time samples $\hat{x}_1, \ldots, \hat{x}_n$.

1. Solve the system of equations

$$\begin{bmatrix} \hat{x}_s & \ldots & \hat{x}_2 & \hat{x}_1 \\ \hat{x}_{s+1} & \ldots & \hat{x}_3 & \hat{x}_2 \\ \ldots & \ldots & \ldots & \ldots \\ \hat{x}_{n-1} & \ldots & \hat{x}_{n-s+1} & \hat{x}_{n-s} \end{bmatrix} \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \ldots \\ \tilde{v}_s \end{bmatrix} = \begin{bmatrix} \hat{x}_{s+1} \\ \hat{x}_{m+1} \\ \ldots \\ \hat{x}_n \end{bmatrix} \tag{1.2.6}$$

   and set $\tilde{v}_0 = 1$.

2. Root the polynomial corresponding to $\tilde{v}_0, \ldots, \tilde{v}_s$ to obtain its s roots $z_1, \ldots, z_s$ (for instance by building the companion matrix of the polynomial [79]).

3. For every root on the unit circle $z_j = e^{i2\pi(t_{\text{est}})_j}$ output $(t_{\text{est}})_j$.

It is straightforward to prove that this procedure achieves exact localization of the line spectra in the absence of noise.

Lemma 1.2.1. *In the absence of noise, the output of Prony's method is equal to the support of the line spectra of the original signal.*

Proof. The coefficients of the polynomial (1.2.4) are a feasible solution for the system of equations (1.2.6). In fact, they are the unique solution. To show this we compute the factorization

$$\begin{bmatrix} \hat{x}_s & \dots & \hat{x}_2 & \hat{x}_1 \\ \hat{x}_{s+1} & \dots & \hat{x}_3 & \hat{x}_2 \\ \dots & \dots & \dots & \dots \\ \hat{x}_{n-1} & \dots & \hat{x}_{n-s+1} & \hat{x}_{n-s} \end{bmatrix} =$$

$$\begin{bmatrix} e^{-i2\pi t_1} & e^{-i2\pi t_2} & \dots & e^{-i2\pi t_s} \\ e^{-i2\pi 2t_1} & e^{-i2\pi 2t_2} & \dots & e^{-i2\pi 2t_s} \\ \dots & \dots & \dots & \dots \\ e^{-i2\pi s t_1} & e^{-i2\pi s t_2} & \dots & e^{-i2\pi s t_s} \end{bmatrix} \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c_s \end{bmatrix} \begin{bmatrix} e^{-i2\pi(s-1)t_1} & \dots & e^{-i2\pi t_1} & 1 \\ e^{-i2\pi(s-1)t_2} & \dots & e^{-i2\pi 2t_2} & 1 \\ \dots & \dots & \dots & \dots \\ e^{-i2\pi(s-1)t_s} & \dots & e^{-i2\pi s t_s} & 1 \end{bmatrix}.$$

The diagonal matrix is full rank as long as all the coefficients c_j are non zero, whereas the two remaining matrices are submatrices of Vandermonde matrices, and consequently also full rank. As a result, the matrix in (1.2.6) is full rank, so the system of equations has a unique solution equal to (1.2.4). This completes the proof, as rooting (1.2.4) obviously yields the support of the line spectra. □

Unfortunately, Prony's method as presented above cannot be applied to real data even if the signal-to-noise ratio is exceptionally high. The image on the left of Figure 1.4 shows how the Prony polynomial allows to super-resolve the location of the line spectra to very high accuracy from noiseless data. However, on the right we can see the result of applying the method to data that have a very small amount of noise (the ratio between the ℓ_2 norm of the noise and the noiseless data is around 10^{-8}!). The roots of the Prony polynomial are perturbed away from the points of the unit circle that correspond to the line spectra, so that it is no longer possible to accurately

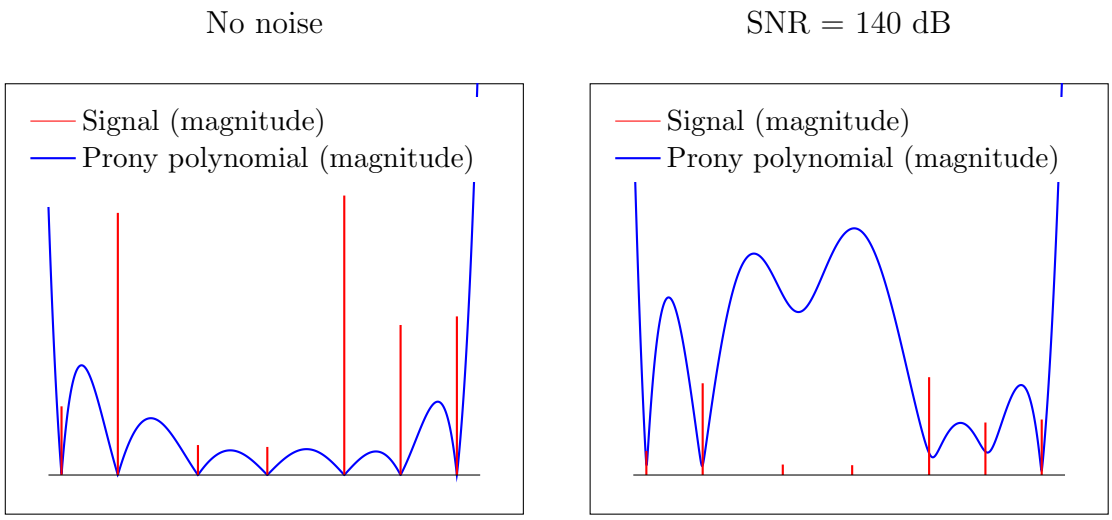


Figure 1.4: Prony polynomial applied to the data in Figure 1.3 (left). On the right we can see the effect of adding a very small quantity of noise to the data. The roots of the polynomial no longer coincide with the line spectra of the original signal. Note that the vertical axis is scaled differently in the two figures.

locate the support of the signal. It is consequently necessary to adapt the method to deal with noisy data if there is to be any hope of applying it in any realistic scenario. The following sections discuss such extensions.

Subspace methods

In this section we describe one of the most popular approaches for robustifying Prony's method. We will assume that the signal and the noise follow probabilistic models that satisfy two assumptions:

- **Assumption 1:** The different sinusoidal components in the original signal are of the form

$$x = \sum_{t_j \in T} c_j \delta_{t_j} = \sum_{t_j \in T} |c_j| \, e^{i\phi_j} \delta_{t_j},$$

  where the phases ϕ_j are independent and uniformly distributed in the interval $[0, 2\pi]$, whereas the amplitudes are arbitrary and deterministic. Note that this

implies that the expectation of x is equal to zero and that the covariance matrix $\mathbb{E}\left[cc^*\right] = D_c$, where

$$D_c := \begin{bmatrix} |c_1|^2 & 0 & \dots & 0 \\ 0 & |c_2|^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & |c_s|^2 \end{bmatrix}.$$

- **Assumption 2:** The measurements are corrupted by white Gaussian noise with zero mean and standard deviation σ, which is independent from the signal. At time k we measure

$$y_k = \int_0^1 e^{-i2\pi kt} x(\mathrm{d}t) + z_k, \tag{1.2.7}$$

where $z \sim N(0, \sigma^2 I)$ is a zero-mean Gaussian random vector with covariance matrix $\sigma^2 I$.

Let us compute the covariance matrix of the data under these assumptions. To ease notation, for two integers $n_1 < n_2$ we define

$$a_{n_1:n_2}(t) := \begin{bmatrix} e^{-i2\pi n_1 t} \\ e^{-i2\pi (n_1+1)t} \\ \dots \\ e^{-i2\pi n_2 t} \end{bmatrix}, \qquad A_{n_1:n_2} := \begin{bmatrix} a_{n_1:n_2}(t_1) & a_{n_1:n_2}(t_2) & \cdots & a_{n_1:n_2}(t_s) \end{bmatrix}. \tag{1.2.8}$$

If the data consists of a vector $y \in \mathbb{C}^m$ of m samples of the form (1.2.7), we have

$$y = A_{1:m} c + z,$$

where $c \in \mathbb{C}^s$ is the vector of signal coefficients and $z \in \mathbb{C}^m$ is a noise vector. The samples are indexed from 1 to m. This choice is made to simplify notation; we could have taken any m contiguous integers (recall that we can interpret these measurements

as time samples of a multisinusoidal signal, so choosing a different set of contiguous integers merely introduces a time shift). Due to our two assumptions,

$$\begin{aligned}\mathbb{E}\left[yy^*\right] &= \mathbb{E}\left[A_{1:m}cc^*A_{1:m}^* + A_{1:m}cz^* + zc^*A_{1:m}^* + zz^*\right] \\ &= A_{1:m}\,\mathbb{E}\left[cc^*\right]A_{1:m}^* + A_{1:m}\,\mathbb{E}\left[c\right]\mathbb{E}\left[z^*\right] + \mathbb{E}\left[z\right]\mathbb{E}\left[c^*\right]A_{1:m}^* + \mathbb{E}\left[zz^*\right] \\ &= A_{1:m}D_cA_{1:m}^* + \sigma^2 I.\end{aligned}$$

The following lemma characterizes the eigendecomposition of the covariance matrix.

Lemma 1.2.2. *The eigendecomposition of* $\mathbb{E}\left[yy^*\right]$ *is equal to*

$$\mathbb{E}\left[yy^*\right] = \begin{bmatrix}U_S & U_N\end{bmatrix}\begin{bmatrix}\Lambda + \sigma^2 I_s & 0 \\ 0 & \sigma^2 I_{n-s}\end{bmatrix}\begin{bmatrix}U_S^* \\ U_N^*\end{bmatrix},$$

where $U_S \in \mathbb{C}^{m\times s}$ *is an orthonormal matrix that spans the column space of* $A_{1:m}$*,* $U_N \in \mathbb{C}^{m\times(m-s)}$ *is an orthonormal matrix spanning its orthogonal complement,* $\sigma^2 I_k \in \mathbb{C}^{k\times k}$ *is a diagonal matrix with diagonal entries equal to* σ^2 *and*

$$\Lambda = \begin{bmatrix}\lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_s\end{bmatrix}, \qquad \lambda_j > 0 \quad \textit{for } 1 \le j \le s.$$

Proof. We begin by writing the full eigendecomposition of AD_cA^*, which is a symmetric matrix of rank s. By the spectral theorem, this matrix has a singular value decomposition of the form

$$A_{1:m}D_cA_{1:m}^* = \begin{bmatrix}U_S & U_N\end{bmatrix}\begin{bmatrix}\Lambda & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}U_S^* \\ U_N^*\end{bmatrix},$$

where U_S, U_N and Λ are as defined above. To complete the proof, note that

$$\begin{aligned}\mathbb{E}\left[yy^*\right] &= A_{1:m} D_c A_{1:m}^* + \sigma^2 I \\ &= \begin{bmatrix} U_S & U_N \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_S^* \\ U_N^* \end{bmatrix} + \begin{bmatrix} U_S & U_N \end{bmatrix} \begin{bmatrix} \sigma^2 I_s & 0 \\ 0 & \sigma^2 I_{n-s} \end{bmatrix} \begin{bmatrix} U_S^* \\ U_N^* \end{bmatrix}.\end{aligned}$$

□

Let us assume that we have available $\Sigma \in \mathbb{C}^{m \times m}$, an approximation to the real covariance matrix $\mathbb{E}\left[yy^*\right]$. The theorem suggests performing an eigendecomposition of this matrix and then exploiting the fact that the eigenspace associated to the largest eigenvalues should approximately correspond to the *signal subspace*, whereas the remaining eigenvectors should belong to the *noise subspace*. As a final step, we would somehow decode the position of the line spectra from these subspaces; hence the term *subspace methods*.

A common approach used to approximate the covariance matrix is to compute the empirical covariance matrix of the data. Assume that we have access to n measurements of the form (1.2.7), where the realizations of the noise component in the measurements are independent. For a fixed integer m, we define the empirical covariance matrix

$$\Sigma\left(m\right) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} y_{[j:j+m-1]} y_{[j:j+m-1]}^*, \tag{1.2.9}$$

where $y_{[j:j+m-1]} := \begin{bmatrix} y_j & y_{j+1} & \cdots & y_{j+m-1} \end{bmatrix}^T$. Asymptotically, when the number of measurements tends to infinity, this estimate converges to the true covariance matrix of the data (see Section 4.9.1 in [98]). The left column of Figure 1.5 shows the eigenvalues of this approximation for different values of the signal-to-noise ratio and the parameter m.

We are now ready to describe the arguably most popular parametric spectral-analysis method: multiple-signal classification (MUSIC) [4, 16, 90].

Input: Number of line spectra s, value of m, noisy data $y_1, \ldots, y_n$.

1. Build the empirical covariance matrix $\Sigma(m)$ as in (1.2.9).

2. Compute the eigendecomposition of $\Sigma(m)$ and select the matrix of eigenvectors $\tilde{U}_N$ corresponding to the $m-s$ smallest eigenvalues.

3. Output an estimate of s estimated positions for the line spectra by:

   (a) **Spectral MUSIC** [16,90]: Locating the s highest peaks of the *pseudospectrum*

   $$P_{\text{SMUSIC}}(t) = \frac{1}{a_{1:m}(t)^* \tilde{U}_N \tilde{U}_N^* a_{1:m}(t)}.$$

   (b) **Root MUSIC** [4]: Finding the s pairs of reciprocal roots of the polynomial

   $$P_{\text{RMUSIC}}(t) = a_{1:m}(t)^* \tilde{U}_N \tilde{U}_N^* a_{1:m}(t)$$

   which are closer to the unit circle.

If we set $m = s+1$ this is equivalent to the Pisarenko method [78]. In the absence of noise, MUSIC allows us to locate the line spectra exactly, as shown in the following lemma. This is not surprising, as Root MUSIC is essentially equivalent to Prony's method if no noise is added to the data.

Lemma 1.2.3. *For noiseless data, the subspace corresponding to the matrix $\tilde{U}_N$ used in the MUSIC algorithm, corresponds exactly to the noise subspace, i.e. it is the orthogonal complement to the subspace spanned by $a_{1:m}(t_1)$, ..., $a_{1:m}(t_s)$, where t_1, ..., t_s are the true locations of the line spectra.*

Proof. In the absence of noise, we can write $\Sigma = XX^*$, where

$$X = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \dots & \hat{x}_{n-m+1} \\ \hat{x}_2 & \hat{x}_3 & \dots & \hat{x}_{n-m+2} \\ \dots & \dots & \dots & \dots \\ \hat{x}_m & \hat{x}_{m+1} & \dots & \hat{x}_n \end{bmatrix} = A_{1:m} \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c_s \end{bmatrix} A_{0:n-m}^T$$

□

The matrix X in the proof of the lemma is simultaneously Toeplitz and low rank. This motivates a denoising procedure that can be used to preprocess the data prior to applying MUSIC in the presence of noise:

1. Compute a rank-s approximation of X by truncating its singular value decomposition. In general, this low-rank approximation will not be Toeplitz due to the noise in the data.

2. Project the result onto the space of Toeplitz matrices by averaging its diagonals.

This approach was originally proposed by Tufts and Kumaresan [106] and is also known as Cadzow denoising [21].

When noise is present, MUSIC will be effective as long as the empirical covariance matrix– or the denoised approximation if we choose to apply preprocessing– is close to the true covariance matrix. This will be the case if the noise is small. The theoretical analysis that is currently available is based on an asymptotic characterization of the sample covariance matrix under Gaussian noise [34,100]. This allows to obtain explicit guarantees on the recovery error in simple cases involving one or two spikes. More recently, some steps towards analyzing the algorithm in a non-asymptotic regime have been taken in [48].

Figure 1.5 shows the performance of Root MUSIC for a specific example. On the left column, we can see the decay of the eigenvalues of the empirical covariance matrix. At high signal-to-noise ratios (SNR) there is a clear transition between the eigenvalues corresponding to the *signal subspace* (in this case $s = 7$) and the *noise subspace*, but this is no longer necessarily the case when the noise is increased. On the right column, we can see the performance of the algorithm for different values of the SNR and the parameter m. At relatively high SNRs MUSIC is an effective algorithm as long as the assumptions on the signal (random phases), noise (Gaussian) and measurement model (equispaced time samples) are satisfied. In Figure 1.6 we show the result of running the algorithm for the wrong value of the parameter s. If the value is not too different to s and the SNR not too low, the method is still capable of approximately locating the support.

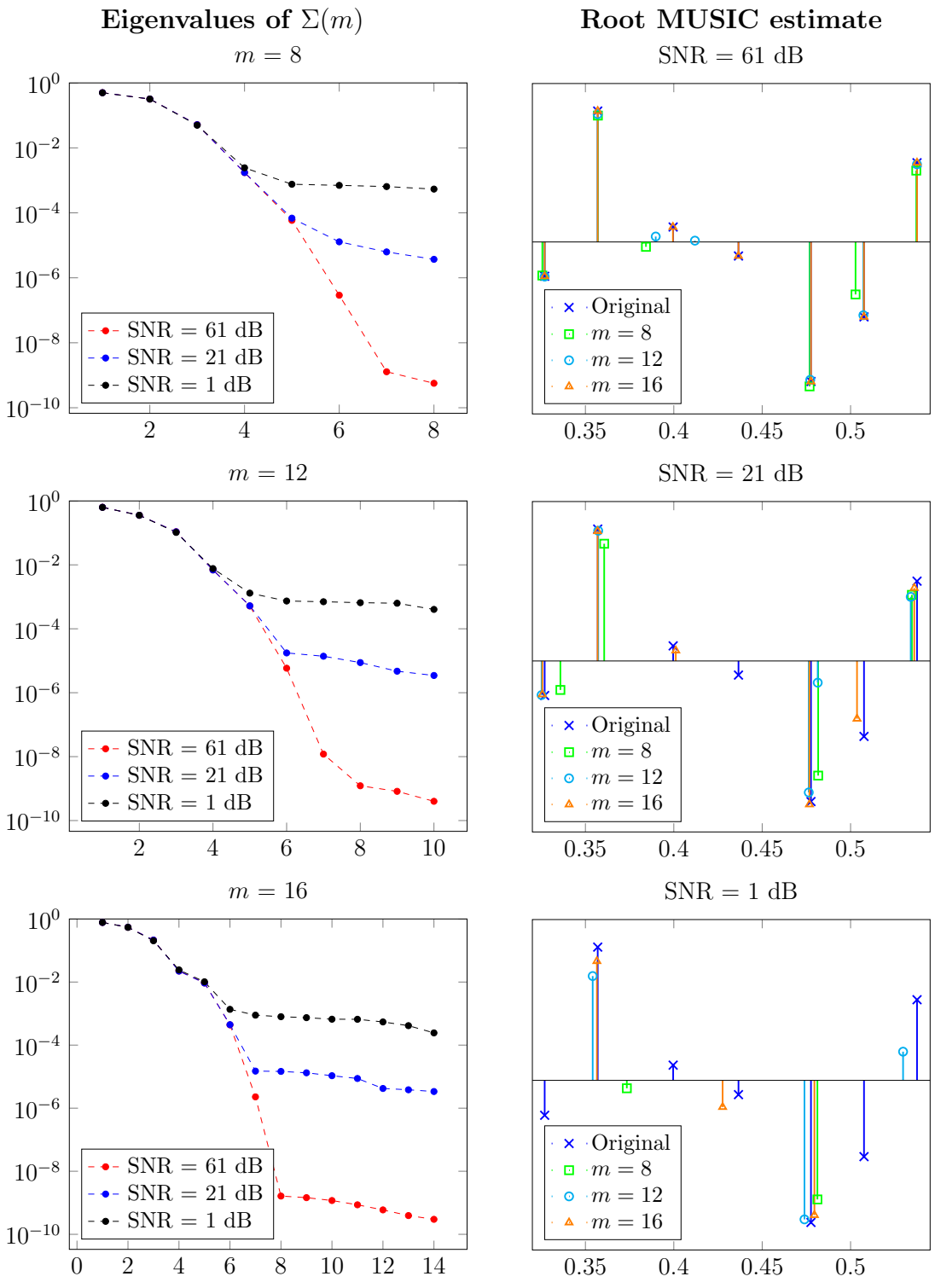


Figure 1.5: Eigenvalues of the autocorrelation matrix used for the Root MUSIC algorithm (left) and corresponding line-spectra estimates (right) for different values of the parameter m and of the SNR respectively. The noise is i.i.d. Gaussian with zero mean.

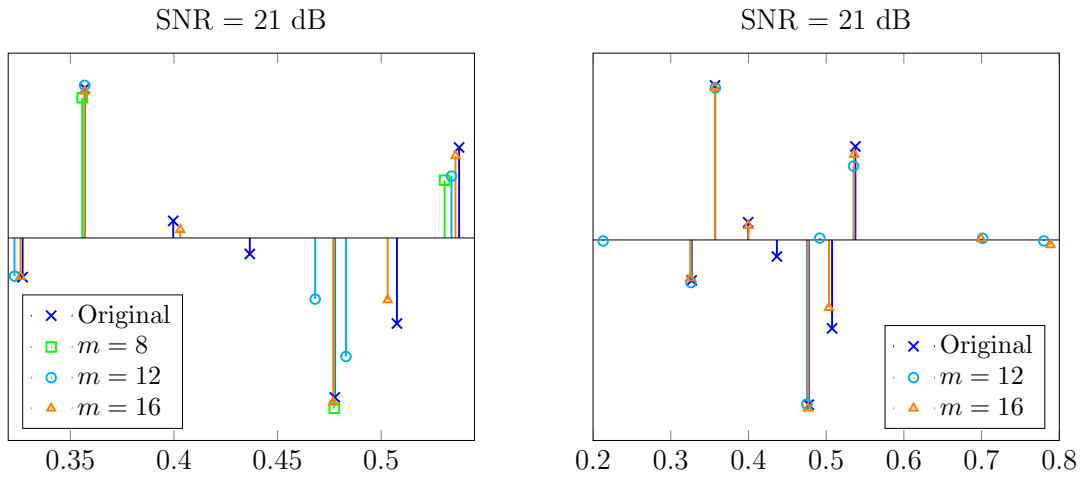


Figure 1.6: Line-spectra estimates obtained by Root MUSIC when the estimated number of sources is equal to $s-1$ (left) and $s+1$ (right) for the same data as in Figure 1.5.

Matrix-pencil methods

In this section we describe an alternative approach to perform spectral super-resolution by robustifying Prony's method. We assume that the data consists of a vector $y \in \mathbb{C}^{n+1}$ of $n+1$ samples of the form (1.2.7), $y = A_{0:n}c + z$, where $c \in \mathbb{C}^s$ is the vector of signal coefficients and $z \in \mathbb{C}^{n+1}$ is a noise vector, and define

$$Y_0 = \begin{bmatrix} y_0 & y_1 & \cdots & y_{n-m} \\ y_1 & y_2 & \cdots & y_{n-m+1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{m-1} & y_m & \cdots & y_{n-1} \end{bmatrix}, \qquad Y_1 = \begin{bmatrix} y_1 & y_2 & \cdots & y_{n-m+1} \\ y_2 & y_3 & \cdots & y_{n-m+2} \\ \cdots & \cdots & \cdots & \cdots \\ y_m & y_{m+1} & \cdots & y_n \end{bmatrix},$$

for a certain integer $s < m < n$. Without further ado, we present the matrix-pencil method for spectral super-resolution [65]:

Input: Number of line spectra s, value of m, noisy data $y_0, \ldots, y_n$.

1. Build the matrices Y_0 and Y_1.
2. Compute the s largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s$ of the matrix $Y_0^\dagger\, Y_1$, where $Y_0^\dagger$ indicates the Moore-Penrose pseudoinverse of the matrix Y_0.

3. Output the phase of λ_1, λ_2, ..., λ_s divided by 2π.

At a first glance, it is not very clear why the method should produce a good estimate of the line spectra, i.e. why $\lambda_j \approx e^{i2\pi t_j}$ for $t_j \in T$, or for that matter why the method has anything to do with matrix pencils. To clarify this, let us once again consider what occurs in the noiseless case. In the absence of noise, Y_0 and Y_1 are equal to

$$X_0 = \begin{bmatrix} \hat{x}_0 & \hat{x}_1 & \dots & \hat{x}_{n-m} \\ \hat{x}_1 & \hat{x}_2 & \dots & \hat{x}_{n-m+1} \\ \dots & \dots & \dots & \dots \\ \hat{x}_{m-1} & \hat{x}_m & \dots & \hat{x}_{n-1} \end{bmatrix}, \qquad X_1 = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \dots & \hat{x}_{n-m+1} \\ \hat{x}_2 & \hat{x}_3 & \dots & \hat{x}_{n-m+2} \\ \dots & \dots & \dots & \dots \\ \hat{x}_m & \hat{x}_{m+1} & \dots & \hat{x}_n \end{bmatrix}$$

respectively. The matrix pencil of two matrices M_1 and M_2 is defined as the matrix-valued function

$$L_{M_1,M_2}(\mu) := M_2 - \mu M_1$$

for any complex-valued μ. For square matrices, the eigenvalues of $M_1^\dagger M_2$ are called *generalized eigenvalues* of the matrix pencil. In our case, the matrices are not rectangular, but we can still define a *rank-reducing value* of the matrix pencil as a value of μ for which $L_{M_1,M_2}(\mu)$ has a lower rank than M_2. The following lemma establishes that obtaining these values allows to super-resolve the line spectra if no noise is added to the data.

Lemma 1.2.4. *$e^{i2\pi t_1}$, $e^{i2\pi t_2}$, ..., $e^{i2\pi t_s}$ are rank-reducing values for the matrix pencil $X_1 - \mu X_0$, which can be obtained by computing the eigenvalues of the matrix $X_0^\dagger X_1$.*

Proof. Recall the definition of $A_{n_1:n_2}$ (1.2.8). We can write

$$X_0 = A_{0:m-1} C A_{0:n-m}^T.$$

Let $\tilde{U}\,\tilde{\Sigma}\,\tilde{V}^*$ be the singular value decomposition of $C\,A_{0:n-m}^T$, where $\tilde{U} \in \mathbb{C}^{s\times s}$, $\tilde{\Sigma} \in$

$\mathbb{C}^{s\times s}$ and $\tilde{V} \in \mathbb{C}^{n-m+1\times s}$. We have

$$\begin{aligned} X_1 &= A_{1:m}\, C\, A_{0:n-m}^T \\ &= A_{0:m-1}\, \Phi\, C\, A_{0:n-m}^T \\ &= X_0\, \tilde{V}\, \tilde{\Sigma}^{-1}\, \tilde{U}^*\, \Phi\, \tilde{U}\, \tilde{\Sigma}\, \tilde{V}^*. \end{aligned}$$

Now, note that the row space of X_0 is spanned by $A_{0:n-m}^T$. This implies that $\tilde{V} = X_0^\dagger\, X_0 \tilde{V}$ so that

$$X_0^\dagger\, X_1 = \tilde{V}\, \tilde{\Sigma}^{-1}\, \tilde{U}^*\, \Phi\, \tilde{U}\, \tilde{\Sigma}\, \tilde{V}^* = P^{-1} \begin{bmatrix} \Phi & 0 \\ 0 & 0 \end{bmatrix} P,$$

where

$$P = \begin{bmatrix} \tilde{U} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{V}^* \\ \tilde{V}_\perp^* \end{bmatrix}$$

and $\tilde{V}_\perp$ is an orthonormal matrix whose column space is the orthogonal complement to the column space of $\tilde{V}$. In words, computing the eigendecomposition of $X_0^\dagger\, X_1$ yields Φ as stated by the lemma. □

An analogous procedure to the matrix-pencil method can be carried out using the *signal subspace* obtained from an eigendecomposition of the empirical covariance matrix (see the previous section on subspace methods). This algorithm is known as ESPRIT (Estimation of Signal Parameters via Rotational Invariance Technique) and was originally introduced in the context of direction-of-arrival problems [85]. For more information on ESPRIT, we refer the interested reader to Chapter 4 of [98], where it is reported that matrix-pencil algorithms exhibit a similar performance to subspace methods in practice.

1.3 Outline and Contributions

At an abstract level, the deconvolution of point sources or spikes from bandlimited data is an instance of a central question in modern data processing: how to recover a low-dimensional object embedded in a high-dimensional space from incomplete linear measurements. Nonparametric techniques based on convex optimization have had great success in tackling problems of this flavor. Notable examples include sparse regression in high-dimensional settings [103], compressed sensing [28, 39] and matrix completion [25]. The main objective of this thesis is to explore the application of these ideas to the problem of super-resolution. In the following outline we highlight some of the main contributions:

- In Chapter 2 we propose a nonparametric algorithm to perform super-resolution of point sources and provide theoretical guarantees for its success in the absence of noise under a condition on the support of the signal. We also describe extensions of our results to a multidimensional setting and to the super-resolution of piecewise-smooth functions.

- Chapter 3 is dedicated to analyzing the stability of the proposed method. First, we show that the problem of super-resolution may be hopelessly ill posed if the point sources are too clustered. Then we provide robustness guarantees that bound the estimation error at resolutions beyond that of the available data under the same condition on the signal support as in the noiseless case. Finally, we present guarantees on the support-detection capability of our method in the presence of noise.

- In Chapter 4 discusses how to implement super-resolution via convex programming, offering two alternatives that are based on the discretization of the domain and on the reformulation of an infinite-dimensional program into a tractable finite-dimensional semidefinite program.

- Chapter 5 proves the main result in Chapter 2. In it we present a novel proof technique which allows to construct a polynomial that certifies exact recovery and is also useful to establish stability guarantees.

- Chapter 6 proves that exact recovery also occurs in a two-dimensional setting, extending the techniques presented in Chapter 5.
- Chapter 7 proves stability guarantees presented in Chapter 3 which bound the approximation error at higher resolutions.
- Chapter 8 proves the support-detection guarantees presented in Chapter 3.
- Chapter 9 contains the conclusion and outlines several future research directions and open problems.

Chapter 2

Super-resolution via convex optimization

In this chapter we propose a nonparametric algorithm to perform super-resolution of point sources and provide theoretical guarantees for its success in the absence of noise. In Section 2.1 we begin by motivating the condition under which these results hold, which precludes the support of the signal from being too clustered. In Section 2.2 we present our main result concerning exact recovery in a noiseless setting. The extension to multiple dimensions is analyzed in Section 2.3, whereas Section 2.4 discusses a discrete setting which prompts the definition of super-resolution factor in Section 2.5. Section 2.6 provides numerical experiments to characterize the performance of the algorithm. Sections 2.7 and 2.8 survey related work in the literature. Finally, Section 2.9 describes the extension of our results to the super-resolution of piecewise-smooth functions.

2.1 Sparsity is not enough

In contrast to compressed sensing [26], where randomized measurements preserve the energy of arbitrary sparse signals with high probability (this is commonly known as

The results presented in this section are joint work with Emmanuel Candès and have been published in [23].

the *restricted isometry property* [27]), sparsity is not a strong enough prior to ensure that the super-resolution problem described in Section 1.1 is not extremely ill posed. Indeed, clustered signals can be nearly completely annihilated by the low-pass sensing mechanism. This can be characterized by means of the seminal work of Slepian [94] on discrete prolate spheroidal sequences, as surveyed in Section 3.1, but we give here a concrete example to drive home this important point .

To keep things simple and coherent with Slepian's work, let us assume that (1) the signal is supported on an infinite unit-step grid, not on a continuous interval, and (2) we have access to a low-pass band $[-W, W]$ of the discrete-time Fourier transform of the signal $\hat{x}(\omega) = \sum_{t\in\mathbb{Z}} x_t e^{-i2\pi\omega t}$, $\omega \in [-1/2, 1/2]$. The corresponding measurement operator is denoted by $\mathcal{P}_W$, and is equivalent to a convolution with a sinc kernel, not a periodized sinc or Dirichlet kernel as in Section 1.1. The quantity $\mathrm{SRF} = 1/2W$ can be interpreted as a *super-resolution factor*, that quantifies to what extent we need to extrapolate the spectrum in order to recover the signal. This definition of the super-resolution factor coincides with the one in Section 2.5 when $N \to \infty$. Set a mild level of super-resolution to fix ideas,

$$\mathrm{SRF} = 4.$$

Now the work of Slepian shows that there is a k-sparse signal supported on $[0, \ldots, k-1]$ obeying

$$\mathcal{P}_W x = \lambda\, x, \quad \lambda \approx 5.22\, \sqrt{k+1}\, e^{-3.23(k+1)}. \tag{2.1.1}$$

For $k = 48$,

$$\lambda \leq 7 \times 10^{-68}. \tag{2.1.2}$$

Even knowing the support ahead of time, how are we going to recover such signals from noisy measurements? For a very mild super-resolution factor of just $\mathrm{SRF} = 1.05$ (we only seek to extend the spectrum by 5%), (2.1.1) becomes

$$\mathcal{P}_W x = \lambda\, x, \quad \lambda \approx 3.87\, \sqrt{k+1}\, e^{-0.15(k+1)}, \tag{2.1.3}$$

which implies that there exists a unit-norm signal with at most 256 consecutive

nonzero entries such that $||\mathcal{P}_W x||_2 \leq 1.2 \times 10^{-15}$. Of course, as the super-resolution factor increases, the ill-posedness gets worse. For large values of SRF, there is x obeying (2.1.1) with

$$\log \lambda \approx -(0.4831 + 2\log(\text{SRF}))k. \tag{2.1.4}$$

It is important to emphasize that this is not a worst-case analysis. In fact, with $k = 48$ and SRF $= 4$, Slepian shows that there is a large-dimensional subspace of signals supported on $\mathbb{C}^k$ spanned by orthonormal eigenvectors with eigenvalues of magnitudes nearly as small as (2.1.2).

The conclusion is that if signals are allowed to be arbitrarily clustered, then super-resolution can be completely ill posed. We need additional conditions on the signals of interest beyond sparsity for the problem to make sense. Here we use a *minimum-separation condition*, illustrated in Figure 2.1, for this purpose.

Definition 2.1.1 (Minimum separation)**.** *Let $\mathbb{T}$ be the circle obtained by identifying the endpoints on $[0, 1]$. For a family of points $T \subset \mathbb{T}$, the minimum separation is defined as the closest distance between any two elements from T,*

$$\Delta(T) = \inf_{(t,t') \in T \,:\, t \neq t'} |t - t'|. \tag{2.1.5}$$

where $|t - t'|$ is the ℓ_∞ distance (maximum deviation in any coordinate). To be clear, this is the wrap-around distance so that the distance between $t = 0$ and $t' = 3/4$ is equal to $1/4$.

In the rest of this thesis we will show that constraining the support of the signals of interest to have a certain minimum separation allows to prove that a nonparametric method based on convex optimization is capable of producing accurate estimates even in the presence of noise.

2.2 Total-variation norm minimization

To perform super-resolution of point sources we propose to minimize a continuous counterpart of the ℓ_1 norm, the total-variation norm, subject to data constraints.

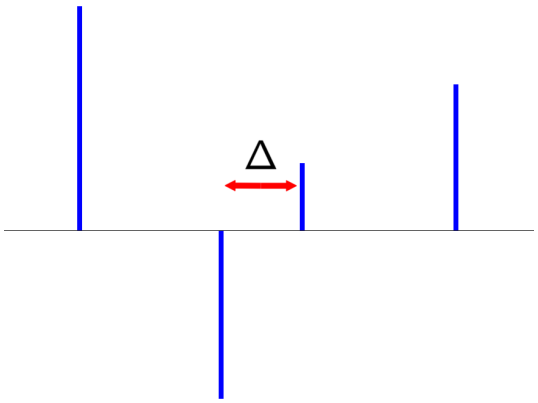


Figure 2.1: Illustration of the minimum-separation condition (2.1.5).

With $\mathbb{T} = [0, 1]$, the total variation of a complex measure ν on a set $B \in \mathcal{B}(\mathbb{T})$ is defined by

$$|\nu|(B) = \sup \sum_{j=1}^{\infty} |\nu(B_j)|,$$

where the supremum is taken over all partitions of B into a finite number of disjoint measurable subsets. The total variation $|\nu|$ is a positive measure on $\mathcal{B}(\mathbb{T})$ and can be used to define the total-variation norm on the space of complex measures on $\mathcal{B}(\mathbb{T})$,

$$||\nu||_{\mathrm{TV}} = |\nu|(\mathbb{T}).$$

For further details, we refer the reader to [87]. As mentioned above, for the purpose of this work, one can think of this norm as a continuous analog to the ℓ_1 norm for discrete signals. In fact, with x as in (1.1.1), $||x||_{\mathrm{TV}}$ is equal to the ℓ_1 norm of the amplitudes $||c||_1 = \sum_j |c_j|$. We would like to emphasize that it should not be confused with the total variation of a piecewise constant function, which is a popular regularizer in image processing [86].

To recover the signal of interest x from low-pass data we propose solving the convex program

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad \mathcal{F}_n \tilde{x} = y, \tag{2.2.1}$$

where the minimization is carried out over the set of all finite complex measures $\tilde{x}$ supported on $[0, 1]$. Chapter 4 discusses ways of computing solutions to this optimization problem.

One of our main theoretical results is that minimizing the total-variation norm of the signal estimate achieves exact recovery in the absence of noise, as long as the original signal satisfies a certain minimum separation. Its proof is provided in Chapter 5.

Theorem 2.2.1. *Let $T = \{t_j\}$ be the support of x. If the minimum distance obeys*

$$\Delta(T) \geq 2 \,/ f_{\mathrm{lo}} := 2 \, \lambda_{\mathrm{lo}}, \tag{2.2.2}$$

then x is the unique solution to (2.2.1)*. This holds with the proviso that $f_{lo} \geq 128$. If x is known to be real-valued, then the minimum gap can be lowered to $1.87 \, \lambda_{lo}$.*

At first sight, this result might appear quite unexpected. The total-variation norm makes no real assumption about the structure of the signal. Yet, not knowing that there are any spikes, let alone how many there are, total-variation minimization locates the position of those spikes with *infinite precision*! It is interesting to note that the theorem does not depend on the amplitudes of the signal and applies to situations where we have both very large and very small spikes, or– put in signal-processing terms– a large dynamic range in the input signal.

The information we have about x is equivalent to observing the projection of x onto its low-frequency components, i.e. the constraint in (2.2.1) is the same as $\mathcal{P}_n \tilde{x} = \mathcal{P}_n x$, where $\mathcal{P}_n = \mathcal{F}_n^* \mathcal{F}_n$. As is well known, this projection is the convolution with the Dirichlet kernel, which has slowly decaying side lobes. The width of the main lobe of this convolution kernel is equal to λ_{lo} as shown in Figure 2.2. The theorem states that exact recovery will occur if the spikes are $2\lambda_{\mathrm{lo}}$ apart. Numerical simulations provided in Section 2.6 indicate that the actual minimum separation at which total-variation norm minimization is guaranteed to achieve exact recovery is λ_{lo}. In the literature $\lambda_{\mathrm{lo}}/2$ is known as the Rayleigh resolution limit.

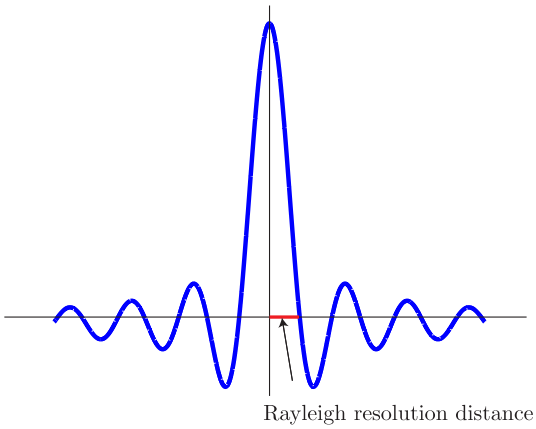


Figure 2.2: λ_{lo} corresponds to the width of the main lobe of the convolution kernel.

2.3 Super-resolution in higher dimensions

Our results extend to higher dimensions and reveal the same dependence between the minimum separation and the measurement resolution as in one dimension. For concreteness, we discuss the 2-dimensional setting and emphasize that the situation in d dimensions is similar. Here, we have a measure

$$x = \sum_j c_j \delta_{t_j},$$

as before but in which the $t_j \in [0,1]^2$. We are given information about x in the form of low-frequency samples of the form

$$y(k) = \int_{[0,1]^2} e^{-i2\pi\langle k,t\rangle} x(\mathrm{d}t) = \sum_j c_j e^{-i2\pi\langle k,t_j\rangle}, \quad k = (k_1,k_2) \in \mathbb{Z}^2,\ |k_1|, |k_2| \le f_{\text{lo}}. \tag{2.3.1}$$

This again introduces a physical resolution of about $\lambda_{\text{lo}} = 1/f_{\text{lo}}$. In this context, we may think of our problem as imaging point sources in the 2D plane—such as idealized stars in the sky—with an optical device with resolution about λ_{lo}—such as a diffraction-limited telescope. Our next result states that it is possible to locate the point sources without any error whatsoever if they are separated by a distance of $2.38\,\lambda_{\text{lo}}$ simply by minimizing the total variation.

Theorem 2.3.1. *Let $T = \{t_j\} \subset [0,1]^2$ identified with $\mathbb{T}^2$ be the support of x obeying the separation condition*[1]

$$\Delta(T) \geq 2.38 \,/ f_{\mathrm{lo}} = 2.38 \, \lambda_{\mathrm{lo}}. \tag{2.3.2}$$

Then if x is real valued, it is the unique minimum total-variation solution among all real objects obeying the data constraints (2.3.1)*. Hence, the recovery is exact. For complex measures, the same statement holds but with a slightly different constant.*

Whereas we have tried to optimize the constant in one dimension, we have not really attempted to do so here in order to keep the proof reasonably short and simple. Hence, this theorem is subject to improvement.

Theorem 2.3.1 is proved for real-valued measures in Chapter 6. However, the proof techniques can be applied to higher dimensions and complex measures almost directly. In details, suppose we observe the discrete Fourier coefficients of a d-dimensional object at $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ corresponding to low frequencies $0 \leq |k_1|, \ldots, |k_d| \leq f_{\mathrm{lo}}$. Then the minimum total-variation solution is exact provided that the minimum distance obeys $\Delta(T) \geq C_d \, \lambda_{\mathrm{lo}}$, where C_d is some positive numerical constant depending only on the dimension. Finally, as the proof makes clear, extensions to other settings, in which one observes Fourier coefficients if and only if the ℓ_2 norm of k is less or equal to a frequency cut-off, are straightforward.

2.4 Discrete super-resolution

Our continuous theory immediately implies analogous results for finite signals. Suppose we wish to recover a discrete signal $x \in \mathbb{C}^N$ from low-frequency data. Just as before, we could imagine collecting low-frequency samples of the form

$$y_k = \sum_{t=0}^{N-1} x_t e^{-i2\pi kt/N}, \quad |k| \leq f_{\mathrm{lo}}; \tag{2.4.1}$$

[1] Recall that distance is measured in ℓ_∞.

the connection with the previous sections is obvious since x might be interpreted as samples of a discrete signal on a grid $\{t/N\}$ with $t = 0, 1, \ldots, N-1$. In fact, the continuous-time setting is the limit of infinite resolution in which N tends to infinity while the number of samples remains constant (i.e. f_{lo} is fixed). Instead, we can choose to study the regime in which the ratio between the actual resolution of the signal $1/N$ and the resolution of the data defined as $1/f_{\text{lo}}$ is constant. This gives the corollary below.

Corollary 2.4.1. *Let $T \subset \{0, 1, \ldots, N-1\}$ be the support of $\{x_t\}_{t=0}^{N-1}$ obeying*

$$\min_{t,t' \in T : t \neq t'} \frac{1}{N} |t - t'| \geq 2\,\lambda_{\text{lo}} = 2\,/f_{\text{lo}}. \tag{2.4.2}$$

Then the solution to

$$\min ||\tilde{x}||_1 \quad \textit{subject to} \quad F_n \tilde{x} = y \tag{2.4.3}$$

in which F_n is the partial Fourier matrix in (2.4.1) *is exact.*

2.5 The super-resolution factor

In the discrete framework, we wish to resolve a signal on a fine grid with spacing $1/N$. However, we observe the lowest $n = 2f_{\text{lo}} + 1$ Fourier coefficients so that in principle, one can only hope to recover the signal on a coarser grid with spacing $1/n$ as shown in Figure 2.3. Hence, the factor N/n, or equivalently, the ratio between the spacings in the coarse and fine grids, can be interpreted as a super-resolution factor (SRF). Below, we set

$$\text{SRF} = N/n \approx N/2f_{\text{lo}}; \tag{2.5.1}$$

when the SRF is equal to 5 as in the figure, we are looking for a signal at a resolution 5 times higher than what is stricto senso permissible. One can then recast Corollary 2.4.1 as follows: if the nonzero components of $\{x_t\}_{t=0}^{N-1}$ are separated by at least $4 \times \text{SRF}$, perfect super-resolution via ℓ_1-norm minimization occurs.

The reason for introducing the SRF is that with inexact data, we obviously cannot hope for infinite resolution. Indeed, noise will ultimately limit the resolution one can

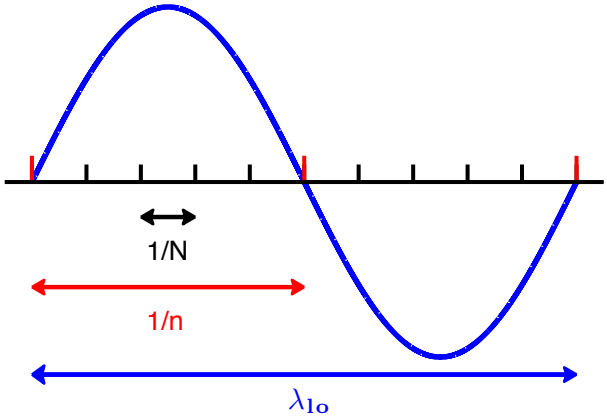


Figure 2.3: Fine grid with spacing $1/N$. We only observe frequencies between $-f_{\text{lo}}$ and f_{lo}, $f_{\text{lo}}/N \approx \frac{1}{2}\text{SRF}^{-1}$, so that the highest frequency sine wave available has wavelength $1/f_{\text{lo}} = \lambda_{\text{lo}}$. These data only allow a Nyquist sampling rate of $\lambda_{\text{lo}}/2 \approx 1/n$. In this sense, we can interpret the super-resolution factor N/n as the ratio between these two resolutions.

ever hope to achieve and, therefore, the question of interest is to study the accuracy one might expect from a practical super-resolution procedure as a function of both the noise level and the SRF. In Chapter 3 we will elaborate on this and in Section 3.2.3 we provide a definition of the super-resolution factor that is applicable for the case when the point sources in the signals are located on a continuous interval rather than on a discretized grid.

2.6 Numerical evaluation of the minimum distance

To evaluate the minimum distance needed to guarantee exact recovery by ℓ_1 minimization of any signal in $\mathbb{C}^N$, for a fixed N, we propose the following heuristic scheme:

- For a super-resolution factor $\text{SRF} = N/n$, we work with a partial DFT matrix F_n with frequencies up to $f_{\text{lo}} = \lfloor n/2 \rfloor$. Fix a candidate minimum distance Δ.
- Using a greedy algorithm, construct an adversarial support with elements separated by at least Δ by sequentially adding elements to the support. Each new element is chosen to minimize the condition number formed by the columns corresponding to the selected elements.

- Take the signal x to be the singular vector corresponding to the smallest singular value of F_n restricted to T.
- Solve the ℓ_1-minimization problem (2.4.3) and declare that exact recovery occurs if the normalized error is below a threshold (in our case 10^{-4}).

This construction of an adversarial signal was found to be better adapted to the structure of our measurement matrix than other methods proposed in the literature such as [43]. We used this scheme and a simple binary search to determine a lower bound for the minimum distance that guarantees exact recovery for $N = 4096$, super-resolution factors of 8, 16, 32 and 64 and support sizes equal to 2, 5, 10, 20 and 50. The simulations were carried out in Matlab, using `CVX` [57] to solve the optimization problem. Figure 2.4 shows the results, which suggest that on the discrete grid we need at least a minimum distance equal to twice the super-resolution factor in order to guarantee reconstruction of the signal (red curve). Translated to the continuous setting, in which the signal would be supported on a grid with spacing $1/N$, this implies that $\Delta \geq \lambda_{\text{lo}}$ is a necessary condition for exact recovery.

2.7 Comparison with related work

The use of ℓ_1 minimization for the recovery of sparse spike trains from noisy bandlimited measurements has a long history and was proposed in the 1980s by researchers in seismic prospecting [33, 72, 89]. For finite signals and under the rather restrictive assumption that the signal is real valued and nonnegative, [53] and [41] prove that k spikes can be recovered from $2k + 1$ Fourier coefficients by this method, a result very related to Carathéodory's work from the beginning of the 20th century [29, 30]. The work [36] extends this result to the continuous setting by using total-variation minimization, dubbed as *Beurling minimal extrapolation* in honor of Beurling's work on extrapolating a function from a portion of its spectrum [14]. In contrast to these previous works, our results require a minimum distance between spikes but allow for arbitrary complex amplitudes, which is crucial in certain applications (especially those concerning spectral super-resolution).

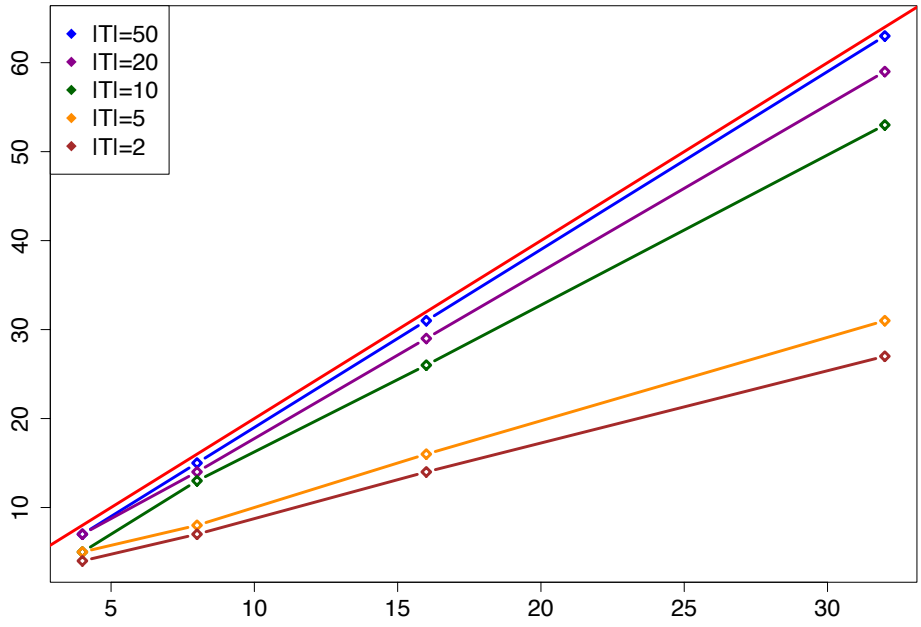


Figure 2.4: Minimum distances (vertical axis) at which exact recovery by ℓ_1 minimization occurs for the adversarial complex sign patterns against the corresponding super-resolution factors. At the red curve, the minimum distance would be exactly equal to twice the super-resolution factor. Signal length is $N = 4096$.

The only theoretical guarantee we are aware of concerning the recovery of spike trains with general amplitudes is very recent and due to Kahane [66]. Kahane offers variations on compressive sensing results in [26] and studies the reconstruction of a function with lacunary Fourier series coefficients from its values in a small contiguous interval, a setting that is equivalent to that of Corollary 2.4.1 when the size N of the fine grid tends to infinity. With our notation, whereas we require a minimum distance equal to $4 \times \mathrm{SRF}$, this work shows that a minimum distance of $10 \times \mathrm{SRF}\sqrt{\log \mathrm{SRF}}$ is sufficient for exact recovery. Although the log factor might seem unimportant at first glance, it in fact precludes extending Kahane's result to the continuous setting of Theorem 2.2.1. Indeed, by letting the resolution factor tend to infinity so as to approach the continuous setting, the spacing between consecutive spikes would need

to tend to infinity as well.

2.8 Connections to sparse recovery literature

Theorem 2.2.1 and Corollary 2.4.1 can be interpreted in the framework of sparse signal recovery. For instance, by swapping time and frequency, Corollary 2.4.1 asserts that one can recover a sparse superposition of tones with arbitrary frequencies from n time samples of the form

$$y_t = \sum_{j=0}^{N-1} x_j e^{-i2\pi t \omega_j}, \quad t = 0, 1, \ldots, n-1$$

where the frequencies are of the form $\omega_j = j/N$. Since the spacing between consecutive frequencies is not $1/n$ but $1/N$, we have an oversampled discrete Fourier transform, where the oversampling ratio is equal to the super-resolution factor. In this context, our results imply that a sufficient condition for perfectly super-resolving these tones is a minimum separation of $4/n$. Moreover, Theorem 2.2.1 extends this to continuum dictionaries where tones ω_j can take on arbitrary real values.

In the literature, there are several conditions that guarantee perfect signal recovery by ℓ_1-norm minimization. The results obtained from their application to our problem are, however, very weak.

- The matrix with normalized columns $f_j = \{e^{-i2\pi t \omega_j}/\sqrt{n}\}_{t=0}^{n-1}$ does not obey the restricted isometry property [27] since a submatrix composed of a very small number of contiguous columns is already very close to singular, see [94] and Section 3.1 for related claims. For example, with $N = 512$ and a modest SRF equal to 4, the smallest singular value of submatrices formed by eight consecutive columns is $3.32\ 10^{-5}$.

- Applying the discrete uncertainty principle proved in [40], we obtain that recovery by ℓ_1 minimization succeeds as long as

$$2\,|T|\,(N-n) < N.$$

If $n < N/2$, i.e. SRF > 2, this says that $|T|$ must be zero. In other words, to recover one spike, we would need at least half of the Fourier samples.

- Other guarantees based on the coherence of the dictionary yield similar results. A popular condition [105] requires that

$$|T| < \frac{1}{2}(M^{-1} + 1), \tag{2.8.1}$$

where M is the coherence of the system defined as $\max_{i \neq j} |\langle f_i, f_j \rangle|$. When $N = 1024$ and SRF $= 4$, $M \approx 0.9003$ so that this becomes $|T| \leq 1.055$, and we can only hope to recover one spike.

There are slightly improved versions of (2.8.1). In [42], Dossal studies the deconvolution of spikes by ℓ_1 minimization. This work introduces the weak exact recovery condition (WERC) defined as

$$\mathrm{WERC}\,(T) = \frac{\beta\,(T)}{1 - \alpha\,(T)},$$

where

$$\alpha(T) = \sup_{i \in T} \sum_{j \in T/\{i\}} |\langle f_i, f_j \rangle|\,, \quad \beta(T) = \sup_{i \in T^c} \sum_{j \in T} |\langle f_i, f_j \rangle|\,.$$

The condition $\mathrm{WERC}\,(T) < 1$ guarantees exact recovery. Considering three spikes and using Taylor expansions to bound the sine function, the minimum distance needed to ensure that $\mathrm{WERC}\,(T) < 1$ may be lower bounded by $24\,\mathrm{SRF}^3/\pi^3 - 2\,\mathrm{SRF}$. This is achieved by considering three spikes at $\omega \in \{0, \pm\Delta\}$, where $\Delta = (k + 1/2)/n$ for some integer k; we omit the details. If $N = 20,000$ and the number of measurements is $1,000$, this allows for the recovery of at most 3 spikes, whereas Corollary 2.4.1 implies that it is possible to reconstruct at least $n/4 = 250$. Furthermore, the cubic dependence on the super-resolution factor means that if we fix the number of measurements and let $N \to \infty$, which is equivalent to the continuous setting of Theorem 2.2.1, the separation needed becomes infinite and we cannot guarantee the recovery of even two spikes.

Finally, we would also like to mention some very recent work on sparse recovery

in highly coherent frames by modified greedy compressed sensing algorithms [44, 49]. Interestingly, these approaches explicitly enforce conditions on the recovered signals that are similar in spirit to our minimum distance condition. As opposed to ℓ_1-norm minimization, such greedy techniques may be severely affected by large dynamic ranges (see [49]). Understanding under what conditions their performance may be comparable to that of convex programming methods is an interesting research direction.

2.9 Extensions

Our results and techniques can be extended to super-resolve many other types of signals. We just outline such a possible extension. Suppose $x : [0,1] \to \mathbb{C}$ is a periodic piecewise smooth function with period 1, defined by

$$x(t) = \sum_{t_j \in T} \mathbf{1}_{(t_{j-1},t_j)} p_j(t);$$

on each time interval (t_{j-1}, t_j), x is polynomial of degree ℓ. For $\ell = 0$, we have a piecewise constant function, for $\ell = 1$, a piecewise linear function and so on. Also suppose x is globally $\ell - 1$ times continuously differentiable (as for splines). We observe

$$y_k = \int_{[0,1]} x(t)\, e^{-i2\pi kt} \mathrm{d}t \ , \ |k| \le f_{\mathrm{lo}}.$$

The $(\ell+1)$th derivative of x (in the sense of distributions) denoted by $x^{(\ell+1)}$ is an atomic measure supported on T and equal to

$$x^{(\ell+1)} = \sum_j c_j \delta_{t_j}, \quad c_j = p_{j+1}^{(\ell)}(t_j) - p_j^{(\ell)}(t_j).$$

Hence, we can imagine recovering $x^{(\ell+1)}$ by solving

$$\min \|\tilde{x}^{(\ell+1)}\|_{\mathrm{TV}} \quad \text{subject to} \quad F_n \tilde{x} = y. \tag{2.9.1}$$

Standard Fourier analysis gives that the kth Fourier coefficient of this measure is given by

$$y_k^{(\ell+1)} = (i2\pi k)^{\ell+1} y_k, \quad k \neq 0. \tag{2.9.2}$$

Hence, we observe the Fourier coefficients of $x^{(\ell+1)}$ except that corresponding to $k = 0$, which must vanish since the periodicity implies $\int_0^1 x^{(\ell+1)}(\mathrm{d}t) = 0 = \int_0^1 x^{(j)}(t)\mathrm{d}t$, $1 \leq j \leq \ell$. Hence, it follows from Theorem 2.2.1 that (2.9.1) recovers $x^{(\ell+1)}$ exactly as long as the discontinuity points are at least $2\lambda_{\mathrm{lo}}$ apart. Because x is $\ell - 1$ times continuously differentiable and periodic, $x^{(\ell+1)}$ determines x up to a shift in function value, equal to its mean. However, we can read the mean value of x off $y_0 = \int_0^1 x(t)\mathrm{d}t$ and, therefore, (2.9.1) achieves perfect recovery.

Corollary 2.9.1. *If $T = \{t_j\}$ obeys* (2.2.2)*, x is determined exactly from y by solving* (2.9.1)*.*

Extensions to non-periodic functions and other types of discontinuities and smoothness assumptions are also possible using similar arguments.

Chapter 3

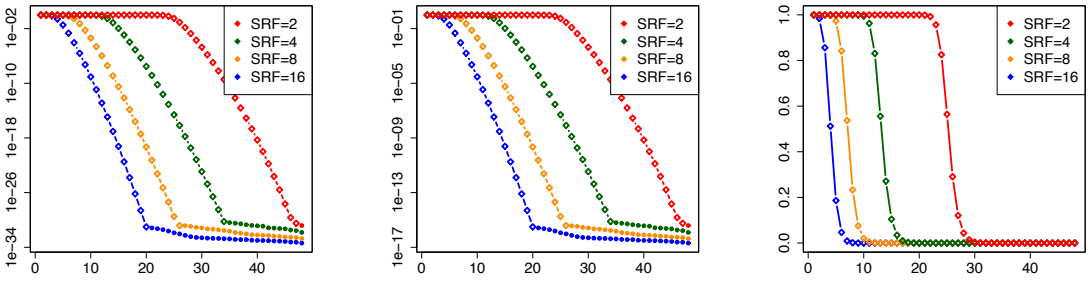
Stability guarantees

In Chapter 2 we studied the problem of recovering superpositions of point sources in a noiseless setting, where one has perfect low-frequency information. In this chapter we consider a setting where the data are contaminated with noise, a situation which is unavoidable in practical applications. In a nutshell, Chapter 2 proves that with noiseless data, one can recover a superposition of point sources exactly, namely, with arbitrary high accuracy, by solving a simple convex program. This phenomenon holds as long as the spacing between the sources is on the order of the resolution limit. With noisy data it is of course no longer possible to achieve infinite precision, but we can still evaluate the robustness of the method.

We begin by justifying the need for conditions on the signal support beyond sparsity to ensure that the problem is well posed when noise is added to the data in Section 3.1. Under such a condition, stability guarantees are then derived in two different ways: by quantifying the estimation accuracy at a higher resolution in Section 3.2 and by studying the support-detection accuracy in Section 3.3. At the end of the chapter, relevant references are discussed in Section 3.4.

3.1 Sparsity and stability

Consider the vector space $\mathbb{C}^{48}$ of sparse signals of length $N = 4096$ supported on a certain interval of length 48. Figure 3.1 shows the eigenvalues of the low-pass



(a) (b) (c)

Figure 3.1: (a) Eigenvalues of P_n acting on signals supported on a contiguous interval of length 48 for super-resolution factors of 2, 4, 8 and 16 and a signal length of 4096. (b) Singular values of $\frac{1}{\sqrt{N}}F_n$ on a logarithmic scale. (c) Same as (b) but on a linear scale. Due to limited numerical precision (machine precision), the smallest singular values, marked with circular dots on the graphs, are of course not computed accurately.

filter $P_n = \frac{1}{N}F_nF_n^*$ acting on $\mathbb{C}^{48}$ for different values of the super-resolution factor. Recall the definition of super-resolution factor (SRF) in Section 2.5. For SRF $= 4$, there exists a subspace of dimension 24 such that any unit-normed signal ($\|x\|_2 = 1$) belonging to it obeys

$$\|P_n x\|_2 \leq 2.52\ 10^{-15} \quad \Leftrightarrow \quad \frac{1}{\sqrt{N}}\|F_n x\|_2 \leq 5.02\ 10^{-8}.$$

For SRF $= 16$ this is true of a subspace of dimension 36, two thirds of the total dimension. Such signals can be completely canceled out by perturbations of norm $5.02\ 10^{-8}$, so that even at signal-to-noise ratios (SNR) of more than 145 dB, recovery is impossible *by any method* whatsoever.

Interestingly, the sharp transition shown in Figure 3.1 between the first singular values almost equal to one and the others, which rapidly decay to zero, can be characterized asymptotically by using the work of Slepian on prolate spheroidal sequences [94]. Introduce the operator $\mathcal{T}_k$, which sets the value of an infinite sequence to zero on the complement of an interval T of length k. With the notation of Section 2.1, the eigenvectors of the operator $\mathcal{P}_W\mathcal{T}_k$ are the discrete prolate spheroidal

sequences $\{s_j\}_{j=1}^k$ introduced in [94],

$$\mathcal{P}_W \mathcal{T}_k s_j = \lambda_j s_j, \quad 1 > \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0. \tag{3.1.1}$$

Set $v_j = \mathcal{T}_k s_j / \sqrt{\lambda_j}$, then by (3.1.1), it is not hard to see that

$$\mathcal{T}_k \mathcal{P}_W v_j = \lambda_j v_j, \qquad ||v_j||_2 = 1.$$

In fact, the v_j's are also orthogonal to each other [94], and so they form an orthonormal basis of $\mathbb{C}^k$ (which can represent any sparse vector supported on T). For values of j near k, the value of λ_j is about $A_j e^{-\gamma(k+1)}$ for

$$A_j = \frac{\sqrt{\pi} 2^{\frac{14(k-j)+9}{4}} \alpha^{\frac{2(k-j)+1}{4}} (k+1)^{k-j+0.5}}{(k-j)! \, (2-\alpha)^{k-j+0.5}},$$

where

$$\alpha = 1 + \cos 2\pi W, \quad \gamma = \log\left(1 + \frac{2\sqrt{\alpha}}{\sqrt{2} - \sqrt{\alpha}}\right).$$

Therefore, for a fixed value of SRF $= 1/2W$, and $k \geq 20$, the small eigenvalues are equal to zero for all practical purposes. In particular, for SRF $= 4$ and SRF $= 1.05$ we obtain (2.1.1) and (2.1.3) in Section 2.1 respectively. Additionally, a Taylor series expansion of γ for large values of SRF yields (2.1.4).

Since $\|\mathcal{P}_W v_j\|_{L_2} = \sqrt{\lambda_j}$, the bound on λ_j for j near k directly implies that some sparse signals are essentially zeroed out, even for small super-resolution factors. However, Figure 3.1 suggests an even stronger statement: as the super-resolution factor increases not only *some*, but *most* signals supported on T seem to be almost completely suppressed by the low pass filtering. Slepian provides an asymptotic characterization for this phenomenon. Indeed, just about the first $2kW$ eigenvalues of $\mathcal{P}_W \mathcal{T}_k$ cluster near one, whereas the rest decay abruptly towards zero. To be concrete, for any $\epsilon > 0$ and $j \geq 2kW(1+\epsilon)$, there exist positive constants C_0, and γ_0 (depending on ϵ and W) such that

$$\lambda_j \leq C_0 e^{-\gamma_0 k}.$$

This holds for all $k \geq k_0$, where k_0 is some fixed integer. This implies that for any interval T of length k, there exists a subspace of signals supported on T with dimension asymptotically equal to $(1 - 1/\mathrm{SRF})\, k$, which is obliterated by the measurement process. This has two interesting consequences. First, even if the super-resolution factor is just barely above one, asymptotically there will always exist an irretrievable vector supported on T. Second, if the super-resolution factor is two or more, most of the information encoded in clustered sparse signals is lost. Consider for instance a random sparse vector x supported on T with i.i.d. entries. Its projection onto a fixed subspace of dimension about $(1 - 1/\mathrm{SRF})\, k$ (corresponding to the negligible eigenvalues) contains most of the energy of the signal with high probability. However, this component is practically destroyed by low-pass filtering. Hence, super-resolving almost any tightly clustered sparse signal in the presence of noise is hopeless. This justifies the need for a minimum separation between nonzero components.

3.2 Accuracy at a higher resolution

Suppose that we aim to super-resolve a signal and that the noise level and sensing resolution are fixed. Then one expects that it will become increasingly hard to recover the fine details of the signal as the scale of these features become finer. There should be a tradeoff between the resolution at which we aim to recover the signal and the precision with which we are able to do so, which will depend on the noise level. The goal of this section is to make this vague statement mathematically precise. We begin with an analysis of the discrete setting and then provide a more general result valid for signals supported on a continuous domain.

3.2.1 A discrete result

In this section we study the robustness to noise of super-resolution via convex optimization in the discrete setting of Section 2.4. For simplicity, we study a deterministic scenario in which the projection of the noise onto the signal space has bounded ℓ_1

The results presented in this section are joint work with Emmanuel Candès and have been published in [23] and [24].

norm but is otherwise arbitrary and can be adversarial. The observations are consequently of the form

$$y = F_n x + z, \quad \frac{1}{N} \, ||F_n^* z||_1 \leq \delta \tag{3.2.1}$$

for some $\delta \geq 0$, where F_n is the partial Fourier matrix in (2.4.1). Letting P_n be the orthogonal projection of a signal onto the first n Fourier modes, $P_n = \frac{1}{N} F_n^* F_n$, we can view (3.2.1) as an input noise model since with $z = F_n w$, we have

$$y = F_n(x + w), \quad ||z||_1 \leq \delta.$$

Another way to write this model with arbitrary input noise $w \in \mathbb{C}^N$ is

$$y = F_n(x + w), \quad ||P_n w||_1 \leq \delta$$

since the high-frequency part of w is filtered out by the measurement process. Finally, with $\tilde{y} = N^{-1} F_n^* y$, (3.2.1) is equivalent to

$$\tilde{y} = P_n x + P_n w, \quad ||P_n w||_1 \leq \delta. \tag{3.2.2}$$

In words, we observe a low-pass version of the signal corrupted with an additive low-pass error whose ℓ_1 norm is at most δ. In the case where $n = N$, $P_n = I$, and our model becomes

$$\tilde{y} = x + w, \quad \|w\|_1 \leq \delta.$$

In this case, one cannot hope for a reconstruction with an error in the ℓ_1 norm less than the noise level δ. We now wish to understand how quickly the recovery error deteriorates as the super-resolution factor increases.

We propose studying the relaxed version of the noiseless problem (2.4.3)

$$\min_{\tilde{x}} ||\tilde{x}||_1 \quad \text{subject to} \quad ||P_n \tilde{x} - \tilde{y}||_1 \leq \delta. \tag{3.2.3}$$

We show that this recovers x with a precision inversely proportional to δ and to the square of the super-resolution factor.

Theorem 3.2.1. *Assume that x obeys the separation condition* (2.4.2)*. Then with the noise model* (3.2.2)*, the solution x_{est} to* (3.2.3) *obeys*

$$||x_{\text{est}} - x||_1 \leq C_0 \, \text{SRF}^2 \, \delta, \tag{3.2.4}$$

for some positive constant C_0.

This theorem, which shows the simple dependence upon the super-resolution factor and the noise level, is proved in the next section. Clearly, plugging in $\delta = 0$ in (3.2.4) gives Corollary 2.4.1. In the rest of Section 3.2, we show that the same dependence between the error at a higher resolution and the super-resolution factor holds for signals supported on continuous domains, as long as the definition of the SRF is adapted to this setting.

3.2.2 Proof of Theorem 3.2.1

The proof is a fairly simple consequence of the following lemma.

Lemma 3.2.2. *Under the assumptions of Theorem 3.2.1, any vector h such that $F_n h = 0$ obeys*

$$||P_T h||_1 \leq \rho \, ||P_{T^c} h||_1 , \tag{3.2.5}$$

for some numerical constant ρ obeying $0 < \rho < 1$. This constant is of the form $1 - \rho = \alpha/\text{SRF}^2$ for some positive $\alpha > 0$. If SRF ≥ 3.03, we can take $\alpha = 0.0883$.

Proof. Let $P_T h_t = |P_T h_t| \, e^{i\phi_t}$ be the polar decomposition of $P_T h$, and consider the low-frequency polynomial $q(t)$ in Proposition 5.2.1 interpolating $v_t = e^{-i\phi_t}$. We shall abuse notations and set $q = \{q_t\}_{t=0}^{N-1}$ where $q_t = q(t/N)$. For $t \notin T$, $|q(t/N)| = |q_t| \leq \rho < 1$. By construction $q = P_n q$, and thus $\langle q, h \rangle = \langle q, P_n h \rangle = 0$. Also,

$$\langle P_T q, P_T h \rangle = \|P_T h\|_1.$$

The conclusion follows from

$$\begin{aligned} 0 &= \langle q, h \rangle \\ &= \langle P_T q, P_T h \rangle + \langle P_{T^c} q, P_{T^c} h \rangle \\ &\geq \|P_T h\|_1 - \|P_{T^c} q\|_\infty \|P_{T^c} h\|_1 \\ &\geq \|P_T h\|_1 - \rho \|P_{T^c} h\|_1. \end{aligned}$$

For the numerical constant, we use Lemma 5.2.5 which says that if $0 \in T$ and $1/N \leq 0.1649 \lambda_{\text{lo}}$, which is about the same as $1/\text{SRF} \approx 2f_c/N \leq 2 \times 0.1649$ or $\text{SRF} > 3.03$, we have

$$|q(1/N)| \leq 1 - 0.3533 \, (f_c/N)^2 \approx 1 - 0.0883/\text{SRF}^2 = \rho.$$

This applies directly to any other t such that $\min_{\tau \in T} |t - \tau| = 1$. Also, for all t at distance at least 2 from T, Lemma 5.2.5 implies that $|q(t/N)| \leq \rho$. This completes the proof. □

Set $h = x_{\text{est}} - x$ and decompose the error into its low-pass and high-pass components

$$h_L = P_n h, \quad h_H = h - h_L.$$

The high-frequency part is in the null space of P_n and (3.2.5) gives

$$||P_T h_H||_1 \leq \rho \, ||P_{T^c} h_H||_1 \, . \tag{3.2.6}$$

For the low-frequency component we have

$$\|h_L\|_1 = \|P_n(x_{\text{est}} - x)\|_1 \leq \|P_n x_{\text{est}} - s\|_1 + \|s - P_n x\|_1 \leq 2\delta. \tag{3.2.7}$$

To bound $\|P_{T^c} h_H\|_1$, we exploit the fact that x_{est} has minimum ℓ_1 norm. We have

$$\begin{aligned} ||x||_1 \geq ||x+h||_1 &\geq ||x+h_H||_1 - \|h_L\|_1 \\ &\geq \|x\|_1 - ||P_T h_H||_1 + ||P_{T^c} h_H||_1 - \|h_L\|_1 \\ &\geq \|x\|_1 + (1-\rho)\,||P_{T^c} h_H||_1 - \|h_L\|_1, \end{aligned}$$

where the last inequality follows from (3.2.6). Hence,

$$||P_{T^c} h_H||_1 \leq \frac{1}{1-\rho}\|h_L\|_1 \quad \Rightarrow \quad \|h_H\|_1 \leq \frac{1+\rho}{1-\rho}\,\|h_L\|_1.$$

To conclude,

$$\|h\|_1 \leq \|h_L\|_1 + \|h_H\|_1 \leq \frac{2}{1-\rho}\|h_L\|_1 \leq \frac{4\delta}{1-\rho},$$

where the last inequality follows from (3.2.7).

Since from Lemma 3.2.2, we have $1-\rho = \alpha/\text{SRF}^2$ for some numerical constant α, the upper bound is of the form $4\alpha^{-1}\,\text{SRF}^2\,\delta$. For $\Delta(T) \geq 2.5\lambda_{\text{lo}}$, we have $\alpha^{-1} \approx 11.235$.

3.2.3 The super-resolution factor for continuous domains

Let us consider a continuous setting where we have observations about an object x of the form

$$y(t) = (Q_{\text{lo}} x)(t) + z(t), \tag{3.2.8}$$

where t is a continuous parameter (time, space, and so on) belonging to the d-dimensional cube $[0,1]^d$. Above, z is a noise term which can either be stochastic or deterministic, and Q_{lo} is a bandlimiting operator with a frequency cut-off equal to $f_{\text{lo}} = 1/\lambda_{\text{lo}}$. Here, λ_{lo} is a positive parameter representing the finest scale at which x is observed. To make this more precise, we take Q_{lo} to be a low-pass filter of width λ_{lo} as illustrated at the top of Figure 3.2; that is,

$$(Q_{\text{lo}} x)(t) = (K_{\text{lo}} * x)(t)$$

such that in the frequency domain the convolution equation becomes

$$(\widehat{Q_{\mathrm{lo}}x})(f) = \widehat{K}_{\mathrm{lo}}(f)\hat{x}(f), \quad f \in \mathbb{Z}^d,$$

and the setting is the same as the one described in Section 1.1 with $K_{\mathrm{lo}} = \phi$. Recall that we denote the usual Fourier transform of a measure or function g, provided that it exists, by $\hat{g}(f) = \int e^{-i2\pi\langle f,t\rangle} g(\mathrm{d}t)$ and note that we assume that the spectrum of the low-pass kernel $\widehat{K}_{\mathrm{lo}}(f)$ vanishes outside of the cell $[-f_{\mathrm{lo}}, f_{\mathrm{lo}}]^d$.

Our goal is to resolve the signal x at a finer scale $\lambda_{\mathrm{hi}} \ll \lambda_{\mathrm{lo}}$. In other words, we would like to obtain a *high-resolution* estimate x_{est} such that $Q_{\mathrm{hi}}\, x_{\mathrm{est}} \approx Q_{\mathrm{hi}}\, x$, where Q_{hi} is a bandlimiting operator with cut-off frequency $f_{\mathrm{hi}} := 1/\lambda_{\mathrm{hi}} > f_{\mathrm{lo}}$. This is illustrated at the bottom of Figure 3.2, which shows the convolution between K_{hi} and x. A different way to pose the problem is as follows: we have noisy data about the spectrum of an object of interest in the low-pass band $[-f_{\mathrm{lo}}, f_{\mathrm{lo}}]$, and would like to estimate the spectrum in the possibly much wider band $[-f_{\mathrm{hi}}, f_{\mathrm{hi}}]$. We redefine the super-resolution factor (SRF) as

$$\mathrm{SRF} := \frac{f_{\mathrm{hi}}}{f_{\mathrm{lo}}} = \frac{\lambda_{\mathrm{lo}}}{\lambda_{\mathrm{hi}}}; \tag{3.2.9}$$

in words, we wish to double the resolution if the SRF is equal to two, to quadruple it if the SRF equals four, and so on. Given the notorious ill-posedness of spectral extrapolation, a natural question is how small the error at scale λ_{hi} between the estimated and the true signal $K_{\mathrm{hi}} * (x_{\mathrm{est}} - x)$ can be? In particular, how does it scale with both the noise level and the SRF? The rest of this section addresses this important question.

3.2.4 Stability guarantees at a higher resolution

We consider the signal model (1.1.1). Although we focus on the one-dimensional case, our methods extend in a straightforward manner to the multidimensional case, as we shall make precise later on. In addition, we assume the measurement model (3.2.8) in which $t \in [0, 1]$, which from now on we identify with the unit circle $\mathbb{T}$, and $z(t)$ is

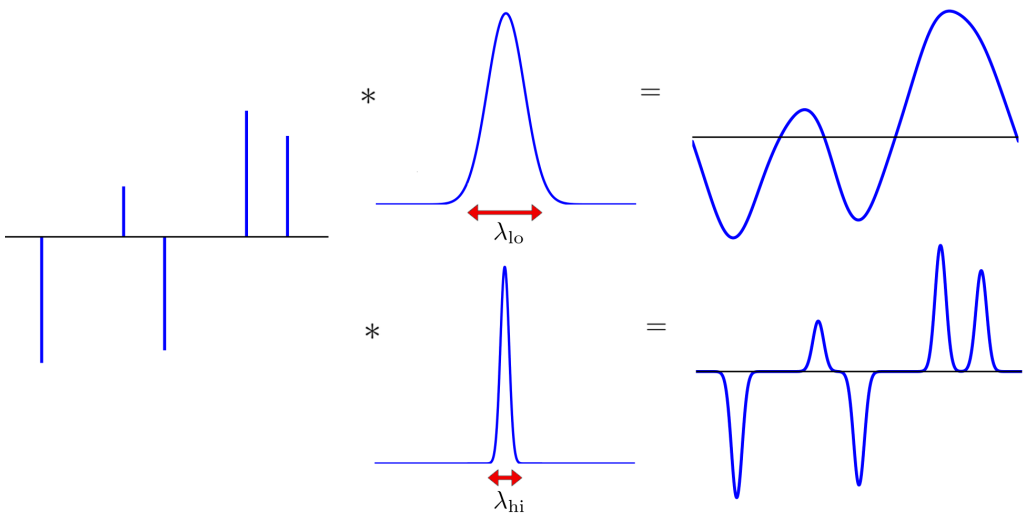


Figure 3.2: Illustration of the super-resolution factor (SRF). A signal (left) is measured at a low resolution by a convolution with a kernel (top middle) of width λ_{lo} (top right). Super-resolution aims at approximating the outcome of a convolution with a much narrower kernel (bottom middle) of width λ_{hi}. Hence, the goal is to recover the bottom right curve.

a bandlimited error term obeying

$$\|z\|_{L_1} = \int_{\mathbb{T}} |z(t)| \, \mathrm{d}t \leq \delta. \tag{3.2.10}$$

The measurement error z is otherwise arbitrary and can be adversarial. For concreteness, we set K_{lo} to be the periodic Dirichlet kernel

$$K_{\text{lo}}(t) = \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} e^{i2\pi kt} = \frac{\sin(\pi(2f_{\text{lo}}+1)t)}{\sin(\pi t)}. \tag{3.2.11}$$

By definition, for each $f \in \mathbb{Z}$, this kernel obeys $\widehat{K}_{\text{lo}}(f) = 1$ if $|f| \leq f_{\text{lo}}$ whereas $\widehat{K}_{\text{lo}}(f) = 0$ if $|f| > f_{\text{lo}}$. We emphasize, however, that our results hold for other low-pass filters. Indeed, our model (3.2.8) can be equivalently written in the frequency domain as $\hat{y}(f) = \hat{x}(f) + \hat{z}(f)$, $|f| \leq f_{\text{lo}}$. Hence, if the measurements are of the form $y = G_{\text{lo}} * x + z$ for some other low-pass kernel G_{lo}, we can filter them linearly to obtain $\hat{y}_G(f) := \hat{y}(f)/\widehat{G}_{\text{lo}}(f) = \hat{x}(f) + \hat{z}(f)/\widehat{G}_{\text{lo}}(f)$. Our results can then be applied

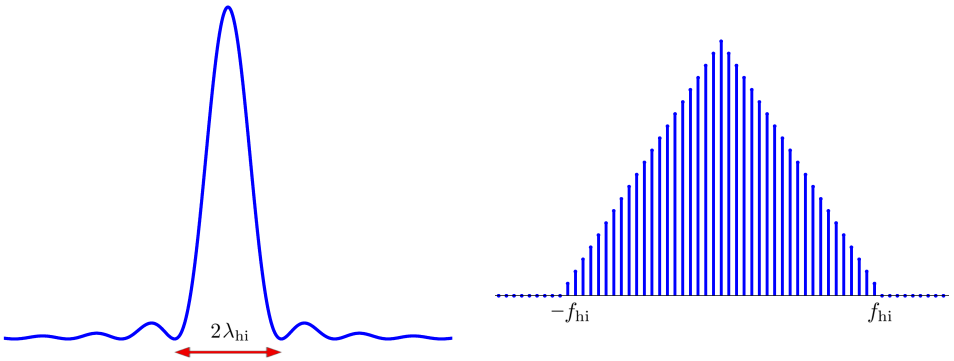


Figure 3.3: The Fejér kernel (3.2.13) (a) with half width about λ_{hi}, and its Fourier series coefficients (b). The kernel is bandlimited since the Fourier coefficients vanish beyond the cut-off frequency f_{hi}.

to this formulation if the weighted perturbation $\hat{z}(f)/\widehat{G}_{\text{lo}}(f)$ is bounded.

To perform recovery, we propose solving

$$\min_{\tilde{x}} ||\tilde{x}||_{\text{TV}} \quad \text{subject to} \quad ||Q_{\text{lo}}\tilde{x} - y||_{\mathcal{L}_1} \leq \delta. \tag{3.2.12}$$

This is a relaxed version of Problem (2.2.1) analyzed in Chapter 2. It is important to observe that the recovery algorithm is completely agnostic to the target resolution λ_{hi}, so our results hold simultaneously for any value of $\lambda_{\text{hi}} > \lambda_{\text{lo}}$.

Our objective is to approximate the signal up until a certain resolution determined by the width of the smoothing kernel $\lambda_{\text{hi}} > \lambda_{\text{lo}}$ used to compute the error. To fix ideas, we set

$$K_{\text{hi}}(t) = \frac{1}{f_{\text{hi}}+1} \sum_{k=-f_{\text{hi}}}^{f_{\text{hi}}} (f_{\text{hi}}+1-|k|)\, e^{i2\pi kt} = \frac{1}{f_{\text{hi}}+1} \left(\frac{\sin(\pi(f_{\text{hi}}+1)t)}{\sin(\pi t)} \right)^2 \tag{3.2.13}$$

to be the Fejér kernel with cut-off frequency $f_{\text{hi}} = 1/\lambda_{\text{hi}}$. Figure 3.3 shows this kernel together with its spectrum.

Our model (3.3.1) asserts that we can achieve a low-resolution error obeying

$$||K_{\text{lo}} * (x_{\text{est}} - x)||_{\mathcal{L}_1} \leq \delta,$$

but that we cannot do better as well. The main question is: how does this degrade when we substitute the low-resolution with the high-resolution kernel?

Theorem 3.2.3. *Assume that the support T of x obeys the separation condition*

$$\Delta(T) \geq 2\lambda_{\mathrm{lo}}. \tag{3.2.14}$$

Then under the noise model (3.3.1)*, any solution x_{est} to problem* (3.2.12)[1] *obeys*

$$||K_{\mathrm{hi}} * (x_{\mathrm{est}} - x)||_{\mathcal{L}_1} \leq C_0 \, \mathrm{SRF}^2 \, \delta,$$

where C_0 is a positive numerical constant.

Thus, minimizing the total-variation norm subject to data constraints yields a stable approximation of any superposition of Dirac measures obeying the minimum-separation condition. When $z = 0$, setting $\delta = 0$ and letting $\mathrm{SRF} \to \infty$, this recovers the conclusion of Theorem 2.2.1 which shows that $x_{\mathrm{est}} = x$, i.e. we achieve infinite precision. What is interesting here is the quadratic dependence of the estimation error in the super-resolution factor. Chapter 7 contains the proof of the result.

3.2.5 Application to stochastic noise

We have chosen to analyze problem (3.2.12) and a perturbation with bounded L_1 norm for simplicity, but our techniques can be adapted to other recovery schemes and noise models. For instance, suppose we observe noisy samples of the spectrum

$$\eta(k) = \int_{\mathbb{T}} e^{-i2\pi kt} \, x(dt) + \epsilon_k, \quad k = -f_{\mathrm{lo}}, -f_{\mathrm{lo}} + 1, \ldots, f_{\mathrm{lo}}, \tag{3.2.15}$$

where ϵ_k is an iid sequence of complex-valued $\mathcal{N}(0, \sigma^2)$ variables (this means that the real and imaginary parts are independent $\mathcal{N}(0, \sigma^2)$ variables). This is equivalent to a line-spectra estimation problem with additive Gaussian white noise, as we explain

[1] To be precise, the theorem holds for any feasible point $\tilde{x}$ obeying $||\tilde{x}||_{\mathrm{TV}} \leq ||x||_{\mathrm{TV}}$; this set is not empty since it contains x.

below. In order to super-resolve the signal under this model, we propose the following convex program

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad ||Q_{\mathrm{lo}}\tilde{x} - y||_{\mathcal{L}_2} \leq \delta, \tag{3.2.16}$$

which can be implemented using off-the-shelf software as discussed in Section 4.3. A corollary to our main theorem establishes that with high probability solving this problem allows to super-resolve the signal despite the added perturbation with an error that scales with the square of the super-resolution factor and is proportional to the noise level.

Corollary 3.2.4. *Fix $\gamma > 0$. Under the stochastic noise model* (3.2.15)*, the solution to problem* (3.2.16) *with $\delta = (1+\gamma)\,\sigma\sqrt{4f_{\mathrm{lo}}+2}$ obeys*

$$||K_{\mathrm{hi}} * (x_{\mathrm{est}} - x)||_{\mathcal{L}_1} \leq C_0\ (1+\gamma)\ \sqrt{4f_{\mathrm{lo}}+2}\,\mathrm{SRF}^2\,\sigma \tag{3.2.17}$$

with probability at least $1 - e^{-2f_{\mathrm{lo}}\gamma^2}$.

Proof. The proof of Theorem 3.2.3 relies on two identities

$$||x_{\mathrm{est}}||_{\mathrm{TV}} \leq ||x||_{\mathrm{TV}}\,, \tag{3.2.18}$$

$$||Q_{\mathrm{lo}}\,(x_{\mathrm{est}} - x)||_{\mathcal{L}_1} \leq 2\delta, \tag{3.2.19}$$

which suffice to establish

$$||K_{\mathrm{hi}} * (x_{\mathrm{est}} - x)||_{\mathcal{L}_1} \leq C_0\,\mathrm{SRF}^2\,\delta.$$

To prove the corollary, we show that (3.2.18) and (3.2.19) hold. Due to the fact that $||\epsilon||_2^2$ follows a χ^2-distribution with $4f_{\mathrm{lo}}+2$ degrees of freedom, we have

$$\mathbb{P}\left(||\epsilon||_2 > (1+\gamma)\,\sigma\sqrt{4f_{\mathrm{lo}}+2} = \delta\right) < e^{-2f_{\mathrm{lo}}\gamma^2},$$

for any positive γ by a concentration inequality (see [71, Section 4]). By Parseval, this implies that with high probability $||Q_{\mathrm{lo}}x - y||_{\mathcal{L}_2} = ||\epsilon||_2 \leq \delta$. As a result, x_{est} is

feasible, which implies (3.2.18) and furthermore

$$\begin{aligned} ||Q_{\text{lo}}\left(x_{\text{est}} - x\right)||_{\mathcal{L}_1} &\leq ||Q_{\text{lo}}\left(x_{\text{est}} - x\right)||_{\mathcal{L}_2} \\ &\leq ||Q_{\text{lo}} x - y||_{\mathcal{L}_2} + ||y - Q_{\text{lo}} x_{\text{est}}||_{\mathcal{L}_2} \\ &\leq 2\delta, \end{aligned}$$

since by the Cauchy-Schwarz inequality $||f||_{\mathcal{L}_1} \leq ||f||_{\mathcal{L}_2}$ for any function f with bounded L_2 norm supported on the unit interval. Thus, (3.2.19) also holds and the proof is complete. □

3.2.6 Extensions

Other high-resolution kernels. We work with the high-resolution Fejér kernel but our results hold for any symmetric kernel that obeys the properties (3.2.20) and (3.2.21) below, since our proof only uses these simple estimates. The first reads

$$\int_{\mathbb{T}} |K_{\text{hi}}\left(t\right)| \, \mathrm{d}t \leq C_0, \qquad \int_{\mathbb{T}} |K'_{\text{hi}}\left(t\right)| \, \mathrm{d}t \leq C_1 \, \lambda_{\text{hi}}^{-1}, \qquad \sup |K''_{\text{hi}}\left(t\right)| \leq C_2 \, \lambda_{\text{hi}}^{-3}, \tag{3.2.20}$$

where C_0, C_1 and C_2 are positive constants independent of λ_{hi}. The second is that there exists a nonnegative and nonincreasing function $f : [0, 1/2] \to \mathbb{R}$ such that

$$|K''_{\text{hi}}\left(t + \lambda_{\text{hi}}\right)| \leq f(t), \quad 0 \leq t \leq 1/2,$$

and

$$\int_0^{1/2} f(t) \mathrm{d}t \leq C_3 \, \lambda_{\text{hi}}^{-2}. \tag{3.2.21}$$

This is to make sure that (7.1.6) holds. (For the Fejér kernel, we can take f to be quadratic in $[0, 1/2 - \lambda_{\text{hi}}]$ and constant in $[1/2 - \lambda_{\text{hi}}, 1/2]$.)

Higher dimensions. Our techniques can be applied to establish robustness guarantees for the recovery of point sources in higher dimensions. The only parts of the proof of Theorem 3.2.3 that do not generalize directly are Lemmas 7.2.1, 7.2.2 and 7.2.4. However, the methods used to prove these lemmas can be extended without much

difficulty to multiple dimensions as described in Section 7.3.

Spectral line estimation. Swapping time and frequency, Theorem 3.2.3 can be immediately applied to the estimation of spectral lines in which we observe

$$y(t) = \sum_j \alpha_j e^{i2\pi\omega_j t} + z(t), \quad t = 0, 1, \ldots, n-1,$$

where α is a vector of complex-valued amplitudes and z is a noise term. Here, our work implies that a nonparametric method based on convex optimization is capable of approximating the spectrum of a multitone signal with arbitrary frequencies, as long as these frequencies are sufficiently far apart, and furthermore that the reconstruction is stable. In this setting, the smoothed error quantifies the quality of the approximation windowed at a certain spectral resolution.

3.3 Support-detection accuracy

In this section we consider the problem of super-resolving the support of a superposition of point sources x, denoted by T. Our aim is to quantify how accurately we can estimate T from the lower end of the spectrum of x when the data are perturbed by noise. We will assume measurements of the form $y = \mathcal{F}_n x + z$, as in (1.1.3). We model the perturbation $z \in \mathbb{C}^n$ as having bounded ℓ_2 norm,

$$||z||_2 \leq \delta. \tag{3.3.1}$$

The noise is otherwise arbitrary and can be adversarial.

To recover x we again propose relaxing Problem 2.2.1 to account for our prior knowledge about the noise:

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad ||\mathcal{F}_n \tilde{x} - y||_2 \leq \delta. \tag{3.3.2}$$

This problem is equivalent to 3.2.16 and can be solved tractably using semidefinite

The results presented in this section have been published in [51].

programming as detailed in Section 4.3.

Chapter 2 established that TV-norm minimization achieves exact recovery in a noiseless setting under the minimum-separation condition. Section 3.2 characterizes the reconstruction error for noisy measurements as the target resolution increases. In this section, we study support detection. If the original signal contains a spike of a certain amplitude we ask: *How accurately can we recover the position of the spike? How does the accuracy depend on the noise level, the amplitude of the spike and the amplitude of the signal at other locations?*

The following theorem characterizes the error between the estimated support T_{est} and the original support T in terms of the estimated spikes that are *near* or *far* from the locations of the original spikes.

Theorem 3.3.1. *Consider the noise model* (3.3.1) *and assume the support T satisfies the minimum-separation condition* (3.2.14)*. The solution to problem* (3.2.12)[2]

$$x_{\text{est}} = \sum_{t_k^{est} \in T_{\text{est}}} c_k^{\text{est}} \delta_{t_k^{est}}$$

with support T_{est} obeys the properties

$$\left| c_j - \sum_{\left\{ t_l^{\text{est}} \in T_{est} : \left| t_l^{\text{est}} - t_j \right| \leq c\lambda_{\text{lo}} \right\}} c_l^{\text{est}} \right| \leq C_1 \delta \quad \forall t_j \in T, \tag{3.3.3}$$

$$\sum_{\left\{ t_l^{\text{est}} \in T_{est},\, t_j \in T :\, \left| t_l^{\text{est}} - t_j \right| \leq c\lambda_{\text{lo}} \right\}} \left| c_l^{\text{est}} \right| \left(t_l^{\text{est}} - t_j \right)^2 \leq C_2 \lambda_{\text{lo}}^2 \delta, \tag{3.3.4}$$

$$\sum_{\left\{ t_l^{\text{est}} \in T_{est} :\, \left| t_l^{\text{est}} - t_j \right| > c\lambda_{\text{lo}}\, \forall t_j \in T \right\}} \left| c_l^{\text{est}} \right| \leq C_3 \delta, \tag{3.3.5}$$

where C_1, C_2 and C_3 are positive numerical constants and $c = 0.1649$.

In plain words, the energy of the estimate is guaranteed to cluster tightly around

[2]This solution can be shown to be an atomic measure with discrete support under very general conditions.

locations where the amplitude of the original signal is above the noise level. Property (3.3.3) implies that each individual spike t_j in the original support T is well approximated by spikes that belong to a small neighborhood around t_j of width about $0.16\lambda_{\text{lo}}$. Property (3.3.4) strengthens this statement, by showing that the farther away the estimated spikes are from t_j the smaller their amplitude has to be. Finally, Property (3.3.5) establishes that any spurious spikes that do not belong to a neighborhood of the original support must have small amplitude. These bounds are essentially optimal for the case of adversarial noise, which can be highly concentrated. The main novelty of the result is that it provides local stability guarantees that only depend on the value of the original signal at a given location. This allows to quantify the accuracy of support detection for each individual spike in the signal.

Corollary 3.3.2. *Under the conditions of Theorem 3.3.1, for any element t_i in the support of x such that $c_i > C_1\delta$ there exists an element t_i^{est} in the support of the estimate x_{est} satisfying*

$$\left|t_i - t_i^{\text{est}}\right| \leq \sqrt{\frac{C_2\delta}{|c_i| - C_1\delta}}\lambda_{\text{lo}}.$$

Despite the aliasing effect of the low-pass filter applied to the signal, the bound on the support-detection error *only depends on the amplitude of the corresponding spike* (and the noise level). This does not follow from previous analysis. In particular, the bound on the weighted $\mathcal{L}_1$ norm of the error derived in Section 3.2 does not allow to derive such local guarantees. As explained in Chapter 8, obtaining detection guarantees that only depend on the amplitude of the spike of interest is made possible by the existence of a certain low-frequency polynomial, constructed in Lemma 8.1.2.

We illustrate the significance of this result with the example shown in Figure 3.4. The signal in the figure has a source with very large amplitude which induces very heavy aliasing. Our theory predicts that nonetheless applying our convex-programming approach will yield an accurate estimate as long as the smallest spikes are sufficiently large with respect to the noise level. In the example, we add white Gaussian noise to induce a signal-to-noise ratio of 20 dB, which would actually be 15 dB without

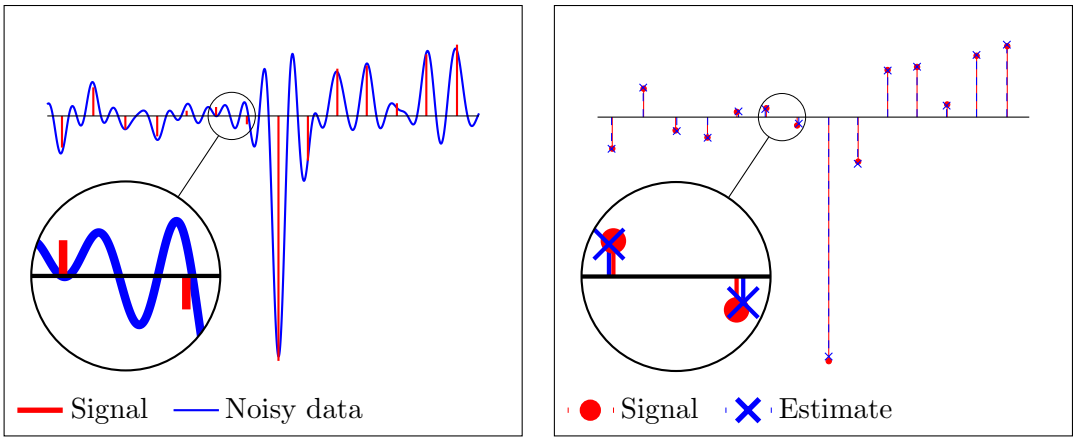


Figure 3.4: Super-resolution of a signal with a large dynamic range. We add white Gaussian noise to induce a signal-to-noise ratio of 20 dB, which would actually be 15 dB without the presence of the large spike. To compute the estimate we solve Problem (3.3.2) as explained in Section 4.3.

the presence of the large spike. Our numerical experiment shows that convex programming is indeed capable of super-resolving the spikes despite the aliasing and the noise. This is interesting because such situations can be highly problematic for more traditional nonparametric super-resolution methods such as those described in Section 1.2.1, as shown in Figure 1.3, and may also be challenging for the parametric methods described in Section 1.2.2, since the aliasing might make it difficult to estimate the number of point sources beforehand.

3.4 Related work

Earlier work on the super-resolution problem in the presence of noise studied under which conditions recovery is not ill-posed. Donoho [38] studies the modulus of continuity of the recovery of a signed measure on a discrete lattice from its spectrum on the interval $[-f_{\text{lo}}, f_{\text{lo}}]$, a setting which is also equivalent to that of Corollary 2.4.1 when the size N of the fine grid tends to infinity. More precisely, if the support of the measure is constrained to contain at most ℓ elements in any interval of length $2/(\ell\, f_{\text{lo}})$, then the modulus of continuity is of order $O\left(\text{SRF}^{2\ell+1}\right)$ as SRF grows to

infinity (note that for $\ell = 1$ the constraint reduces to a minimum distance condition between spikes, which is comparable to the separation condition (2.2.2)). This means that if the ℓ_2 norm of the difference between the measurements generated by two signals satisfying the support constraints is known to be at most δ, then the ℓ_2 norm of the difference between the signals may be of order $O\left(\mathrm{SRF}^{2\ell+1}\,\delta\right)$. This result suggests that, in principle, the super-resolution of spread-out signals is not hopelessly ill-conditioned. Having said this, it does not propose any practical recovery algorithm (a brute-force search for sparse measures obeying the low-frequency constraints would be computationally intractable). More recently, [7] studies the local stability of the super-resolution problem in a continuous domain, but also does not provide any tractable algorithms to perform recovery.

Other works have explored the trade-off between resolution and signal-to-noise ratio for the detection of two closely-spaced line spectra [93] or light sources [61, 92]. A recent reference [50], which focuses mainly on the related problem of imaging point scatterers, analyzes the performance of a parametric method in the case of signals sampled randomly from a discrete grid under the assumption that the sample covariance matrix is close enough to the true one. In general, parametric techniques require prior knowledge of the model order and rely heavily on the assumption that the noise is white or at least has known spectrum (so that it can be whitened) as described in Section 1.2.2. An alternative approach that overcomes the latter drawback is to perform nonlinear least-squares estimation of the model parameters [99]. Unfortunately, the resulting optimization problem has an extremely multimodal cost function, which makes it very sensitive to initialization [97].

Previous theoretical work on the stability of methods based on ℓ_1-norm regularization is limited to a discrete and finite-dimensional setting, where the support of the signal of interest is restricted to a finer uniform grid [23]. However, even if we discretize the dictionary, previous stability results for sparse recovery in redundant dictionaries do not apply due to the high coherence between the atoms of the overcomplete Fourier dictionary. In addition, it is worth mentioning that working with a discrete dictionary can easily degrade the quality of the estimate [32] (see [96] for a related discussion concerning grid selection for spectral analysis), which highlights

the importance of analyzing the problem in the continuous domain. This observation has spurred the appearance of modified compressed-sensing techniques specifically tailored to the task of spectral estimation [44, 49, 64]. Proving stability guarantees for these methods under conditions on the support or the dynamic range of the signal is an interesting research direction.

Finally, we would like to mention some very recent work on super-resolution via convex programming. In [2], the authors bound the support-detection error of a convex program similar to Problem (3.2.16) in the presence of stochastic noise, but the bound depends on the amplitude of the solution rather than on the amplitude of the original spike. In [15], the authors consider the problem of denoising samples taken from signals with sparse spectra using convex programming. In more detail, for a signal composed of s sinusoids supported on an arbitrary set T of the unit interval, assume that we have access to n measurements of the form

$$y_k = w_k + z_k, \quad 1 \leq k \leq n,$$

where w_k denotes the noiseless sample

$$w_k = \sum_{j=1}^{s} c_j e^{-i2\pi k t_j}, \quad 1 \leq k \leq n,$$

and z is a vector of i.i.d. Gaussian noise with standard deviation equal to σ. Using an estimator that combines a least-squares term with total-variation norm regularization, [15] shows that it is possible to obtain an estimate w_{est} satisfying

$$\frac{||w - w_{\text{est}}||_2^2}{n} \lesssim \sigma \sqrt{\frac{\log n}{n}} \sum_{j=1}^{s} |c_j| \,.$$

In later work [101], this result is improved to

$$\frac{||w - w_{\text{est}}||_2^2}{n} \lesssim \sigma^2 \frac{k \log n}{n}$$

under the minimum-separation condition (3.2.14) by building upon the techniques

presented in Chapters 5 and 7. It is important to emphasize that these results concern the *denoising* of the available samples, which in the case of spatial super-resolution would correspond to the low-pass part of the spectrum, in contrast to the methods analyzed in this chapter, which aim to achieve the *extrapolation* of the spectrum of the signal up to a higher resolution. However, [15] does include an extension of the results of Section 3.3 to sparse superpositions of point sources measured under Gaussian perturbations.

Chapter 4

Total-variation minimization

In this chapter we discuss how to solve Problem (2.2.1), which may seem challenging at first because its primal variable is infinite-dimensional. We describe two approaches: Section 4.1 explains how to approximate the solution to the primal problem by discretizing the support of the estimate, whereas in Section 4.2 we show that it is actually possible to solve the continuous program directly by reformulating its dual as a semidefinite program. The latter approach is extended to a noisy setting in Section 4.3. An alternative method to perform total-variation minimization, which we do not discuss, would be to approximate the solution by estimating the support of the signal in an iterative fashion as proposed in [20].

4.1 Discretization of the primal problem

For superpositions of Dirac measures lying on a finite grid, the total-variation norm reduces to the ℓ_1 norm, for which many efficient optimization algorithms have been developed in the last decade. We refer the interested reader to [8–10,18,35,56,107] and references therein. In more detail, if we discretize the unit interval into an N-point

The results presented in this chapter are joint work with Emmanuel Candès and have been published in [23] and [24].

grid, the operator $\mathcal{F}_n$ becomes a matrix $F_n \in \mathbb{C}^{n \times N}$ and Problem (2.2.1) becomes

$$\min_{\tilde{x} \in \mathbb{C}^N} ||\tilde{x}||_1 \quad \text{subject to} \quad F_n \, \tilde{x} = y.$$

This approach is particularly appropriate for applications such as imaging where the original data may correspond to discrete samples of the low-resolution version of the signal x_{LR}, which we denote by $x_{\mathrm{LR},N}$ (for simplicity we assume that the grids for x and x_{LR} are the same). In such applications, the point spread function ϕ in (1.1.2) is usually not well approximated by a sinc function. For instance, in fluorescence microscopy a Gaussian parametric model is often employed for this purpose [110]. We can incorporate a discretized estimate ϕ_N of the point spread function of the sensing process by recasting the problem as

$$\min_{\tilde{x} \in \mathbb{C}^N} ||\tilde{x}||_1 \quad \text{subject to} \quad \phi_N * \tilde{x} = x_{\mathrm{LR},N}. \tag{4.1.1}$$

A possible disadvantage of discretizing the primal problem is the higher computational cost that is incurred when a very high accuracy is desired in the support estimate. The computational complexity will depend on N, instead of on the number of measurements n or the number of spikes, which may be undesirable for certain applications. In such cases, the approach described in the following section would be more appropriate, as long as the sensing mechanism is well modeled as a convolution with a sinc or a Dirichlet kernel, which is the case in spectral super-resolution as described in Section 1.2.1.

4.2 Reformulation of the dual problem

In this section we take an alternative route to the discretization of the primal problem, showing that (2.2.1) can be cast as a semidefinite program with just $(n+1)^2/2$ variables, and that highly accurate solutions can be found rather easily. This formulation is similar to that in [15] which concerns a related infinite-dimensional convex program.

The convex program dual to

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad \mathcal{F}_n \, \tilde{x} = y$$

is

$$\max_{\nu} \ \mathrm{Re}\langle y, \nu \rangle \quad \text{subject to} \quad ||\mathcal{F}_n^* \, \nu||_\infty \leq 1; \tag{4.2.1}$$

the constraint says that the trigonometric polynomial $(\mathcal{F}_n^* \, \nu)(t) = \sum_{|k| \leq f_{\mathrm{lo}}} \nu_k e^{i2\pi kt}$ has a modulus uniformly bounded by 1 over the interval $[0, 1]$. The interior of the feasible set contains the origin and is consequently non empty, so that strong duality holds by a generalized Slater condition [84]. The cost function involves a finite vector of dimension n, but the problem is still infinite dimensional due to the constraint.

Theorem 4.24 in [45] allows to express this constraint as the intersection between the cone of positive semidefinite matrices $\{X : X \succeq 0\}$ and an affine hyperplane. The result follows from the Fejér-Riesz Theorem. We provide a proof for the sake of completeness. To ease notation we define the matrix-valued operator $\mathcal{T} : \mathbb{C}^n \to \mathbb{C}^{n \times n}$. For any vector u such that u_1 is positive and real, $\mathcal{T}(u)$ is a Hermitian Toeplitz matrix whose first row is equal to u. The adjoint of $\mathcal{T}$ with respect to the usual matrix inner product $\langle M_1, M_2 \rangle = \mathrm{Tr}\,(M_1^* M_2)$, extracts the sums of the diagonal and off-diagonal elements of a matrix

$$\mathcal{T}^*(M)_j = \sum_{i=1}^{n-j+1} M_{i,i+j-1}.$$

Proposition 4.2.1. *For any $\nu \in \mathbb{C}^n$ and any positive constant d,*

$$|(\mathcal{F}_n^* \, \nu)(t)| \leq d \quad \textit{for all } t \in \mathbb{T}$$

if and only if there exists a Hermitian matrix $Q \in \mathbb{C}^{n \times n}$ obeying

$$\begin{bmatrix} Q & \nu \\ \nu^* & 1 \end{bmatrix} \succeq 0, \qquad \mathcal{T}^*(Q) = d^2 e_1, \tag{4.2.2}$$

where e_1 is the first vector of the canonical basis of $\mathbb{R}^n$.

Proof. Let us define

$$z_0(t) := \begin{bmatrix} e^{i2\pi f_c t} & e^{i2\pi(f_c-1)t} & \dots & e^{-i2\pi f_c t} \end{bmatrix}^T,$$
$$z_1(t) := \begin{bmatrix} 1 & e^{i2\pi t} & \dots & e^{i2\pi n t} \end{bmatrix}^T.$$

Note that $(\mathcal{F}_n^* c)(t) = z_0(t)^* c$ and $\mathcal{T}(z_1(t)) = z_0(t)^* z_0(t)$.

If (4.2.2) holds then $Q - \nu\nu^* \succeq 0$ and $\mathcal{T}^*(Q) = d^2 e_1$. This implies

$$\begin{aligned} |(\mathcal{F}_n^* \nu)(t)|^2 &= z_0(t)^* \nu\nu^* z_0(t) \\ &\leq z_0(t)^* Q z_0(t) \\ &= \operatorname{Tr}(z_0(t)^* Q z_0(t)) \\ &= \langle Q, z_0(t) z_0(t)^* \rangle \\ &= \langle Q, \mathcal{T}(z_1(t)) \rangle \\ &= \langle \mathcal{T}^*(Q), z_1(t) \rangle \\ &= d^2 (z_1(t))_1 = d^2. \end{aligned}$$

For the converse, $|(\mathcal{F}_n^* \nu)(t)| \leq d$ implies that the trigonometric polynomial $d^2 - z_0(t)^* \nu\nu^* z_0(t)$ is non-negative. By the Fejér-Riesz Theorem there exists a polynomial $P(t) = \tilde{\nu}^* z_0(t)$ such that

$$d^2 - z_0(t)^* \nu\nu^* z_0(t) = |P(t)|^2 = z_0(t)^* \tilde{\nu}\tilde{\nu}^* z_0(t). \tag{4.2.3}$$

Now let us set $Q = \nu\nu^* + \tilde{\nu}\tilde{\nu}^*$. $Q - \nu\nu^*$ is obviously positive semidefinite and consequently so is

$$\begin{bmatrix} Q & \nu \\ \nu^* & 1 \end{bmatrix} \succeq 0.$$

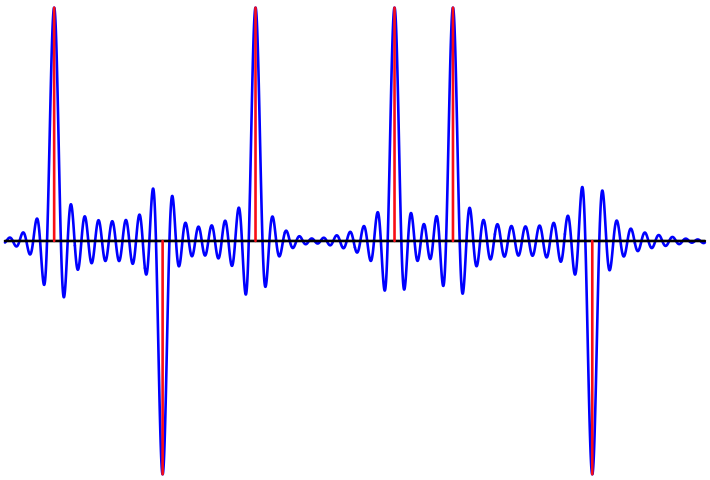


Figure 4.1: The sign of a real atomic measure x is plotted in red. The trigonometric polynomial $\mathcal{F}_n^* \nu$ where ν is a solution to the dual problem (4.2.4) is plotted in blue. Note that $\mathcal{F}_n^* \nu$ interpolates the sign of x. Here, $f_{\text{lo}} = 50$ so that we have $n = 101$ low-frequency coefficients.

Furthermore, by (4.2.3)

$$\begin{aligned} d^2 &= z_0(t)^* Q z_0(t) \\ &= \langle Q, z_0(t) z_0(t)^* \rangle \\ &= \langle Q, \mathcal{T}(z_1(t)) \rangle \\ &= \langle \mathcal{T}^*(Q), z_1(t) \rangle \end{aligned}$$

for all t, which is only possible if $\mathcal{T}^*(Q) = d^2 e_1$. This completes the proof. □

Setting $d = 1$ in Proposition 4.2.1, Problem (4.2.1) is consequently equivalent to

$$\max_{\nu \in \mathbb{C}^n, Q \in \mathbb{C}^{n \times n}} \operatorname{Re}\langle y, \nu \rangle \qquad \text{subject to} \quad \begin{bmatrix} Q & \nu \\ \nu^* & 1 \end{bmatrix} \succeq 0, \quad \mathcal{T}^*(Q) = e_1. \tag{4.2.4}$$

To be complete, the decision variables are an Hermitian matrix $Q \in \mathbb{C}^{n \times n}$ and a vector of coefficients $\nu \in \mathbb{C}^n$. The finite-dimensional semidefinite program can be solved by off-the-shelf convex programming software.

f_{lo}	25	50	75	100
Average error	$6.66\,10^{-9}$	$1.70\,10^{-9}$	$5.58\,10^{-10}$	$2.96\,10^{-10}$
Maximum error	$1.83\,10^{-7}$	$8.14\,10^{-8}$	$2.55\,10^{-8}$	$2.31\,10^{-8}$

Table 4.1: Support-estimation accuracy achieved by solving (4.2.4) via CVX [57]. For each value of the cut-off frequency f_{lo}, 100 signals were generated with random complex amplitudes situated at approximately $f_{\text{lo}}/4$ random locations in the unit interval separated by at least $2/f_{\text{lo}}$. The table shows the errors in estimating the support locations.

The careful reader will observe that we have just shown how to compute the optimal value of (2.2.1), but we are really interested in the solution of the primal instead of the dual problem, as the theoretical results in Chapters 2 and 3 suggest that the former will be a good estimate of the signal of interest. Fortunately, due to strong duality the finite-dimensional solution has a very useful interpretation: it corresponds to the coefficients of a polynomial that interpolates the sign of the primal solution.

In more detail, to find a primal solution, we abuse notation by letting ν be the solution to (4.2.4) and consider the trigonometric polynomial

$$p_{2n-2}(e^{i2\pi t}) = 1 - |(\mathcal{F}_n^*\nu)(t)|^2 = 1 - \sum_{k=-2f_{\text{lo}}}^{2f_{\text{lo}}} u_k e^{i2\pi kt}, \qquad u_k = \sum_j \nu_j \bar{\nu}_{j-k}. \tag{4.2.5}$$

Note that $z^{2f_{\text{lo}}}p_{2n-2}(z)$, where $z \in \mathbb{C}$, is a polynomial of degree $4f_{\text{lo}} = 2(n-1)$ with the same roots as $p_{2n-2}(z)$—besides the trivial root $z = 0$. Hence, $p_{2n-2}(e^{i2\pi t})$ has at most $2n-2$ roots. By construction $p_{2n-2}(e^{i2\pi t})$ is a real-valued and nonnegative trigonometric polynomial; in particular, it cannot have single roots on the unit circle since the existence of single roots would imply that $p_{2n-2}(e^{i2\pi t})$ takes on negative values. Therefore, $p_{2n-2}(e^{i2\pi t})$ is either equal to zero everywhere or has at most $n-1$

roots on the unit circle. By strong duality, any solution x_{est} to (2.2.1) obeys

$$\begin{aligned}
\text{Re}\langle y, \nu \rangle &= \text{Re}\langle \mathcal{F}_n \, x_{\text{est}}, \nu \rangle \\
&= \text{Re}\langle x_{\text{est}}, \mathcal{F}_n^* \, \nu \rangle \\
&= \text{Re}\left[\int_0^1 \overline{\left(\mathcal{F}_n^* \nu\right)(t)} \, x_{\text{est}}(\mathrm{d}t) \right] \\
&= ||x_{\text{est}}||_{\text{TV}} \, ,
\end{aligned}$$

which implies that the trigonometric polynomial $\mathcal{F}_n^* \nu$ is exactly equal to the sign of x_{est} when x_{est} is not vanishing. This is illustrated in Figure 4.1. Thus, to recover the support of the solution to the primal problem, we must simply locate the roots of p_{2n-2} on the unit circle, for instance by computing the eigenvalues of its companion matrix [79]. As shown in Table 4.1, this scheme allows to recover the support with very high precision. Having obtained the estimate for the support T_{est}, the amplitudes of the signal can be reconstructed by solving the system of equations $\sum_{t_j \in T_{\text{est}}} c_j e^{-i2\pi k t_j} = y_k$, $|k| \leq f_c$, using the method of least squares. There is a unique solution as we have at most $n-1$ columns which are linearly independent since one can add columns to form a Vandermonde system.[1] Figure 4.2 illustrates the accuracy of this procedure; a Matlab script reproducing this example is available at `http://www-stat.stanford.edu/~candes/superres_sdp.m`.

In summary, in the usual case when p_{2n-2} has less than n roots on the unit circle, we have explained how to retrieve the minimum total-variation norm solution. It remains to address the situation in which p_{2n-2} vanishes everywhere. In principle, this could happen even if a primal solution to (2.2.1) is an atomic measure supported on a set T obeying $|T| < n$. For example, let x be a positive measure satisfying the conditions of Theorem 2.2.1, which implies that it is the unique solution to (2.2.1). Consider a vector $\nu \in \mathbb{C}^n$ such that $(\mathcal{F}_n^* \nu)(t) = 1$; i.e. the trigonometric polynomial

[1] The set of roots contains the support of a primal optimal solution; if it is a strict superset, then some amplitudes will vanish.

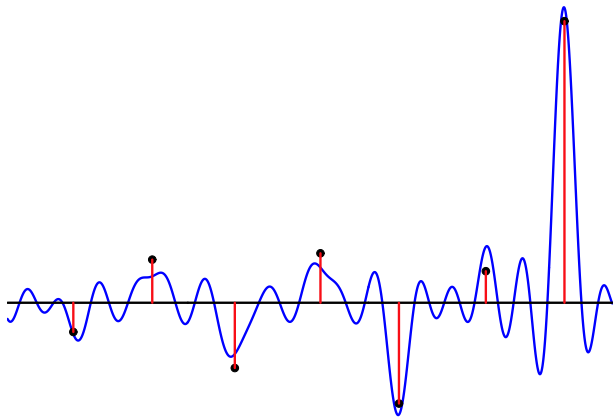


Figure 4.2: There are 21 spikes situated at arbitrary locations separated by at least $2\lambda_c$ and we observe 101 low-frequency coefficients ($f_{\text{lo}} = 50$). In the plot, seven of the original spikes (black dots) are shown along with the corresponding low resolution data (blue line) and the estimated signal (red line).

is constant. Then

$$\text{Re}\langle y, \nu \rangle = \text{Re}\langle \mathcal{F}_n \, x, \nu \rangle = \text{Re}\langle x, \mathcal{F}_n^* \, \nu \rangle = ||x||_{\text{TV}} \, ,$$

which shows that ν is a solution to the dual (4.2.4) that does not carry any information about the support of x. Fortunately, this situation is highly unusual in practice. In fact, it does not occur as long as

$$\text{there exists a solution } \tilde{\nu}, \tilde{Q} \text{ to (4.2.4) obeying } |(\mathcal{F}_n^* \tilde{\nu})(t)| < 1 \text{ for some } t \in [0,1], \tag{4.2.6}$$

and we use interior point methods as in SDPT3 [104] to solve (4.2.4). (Our simulations use CVX which in turn calls SDPT3.) This phenomenon is explained below. At the moment, we would like to remark that Condition (4.2.6) is sufficient for the primal problem (2.2.1) to have a unique solution, and holds except in very special cases. To illustrate this, suppose y is a random vector, *not* a measurement vector corresponding to a sparse signal. In this case, we typically observe dual solutions as shown in Figure 4.3 (non-vanishing polynomials with at most $n-1$ roots). To be sure, we have solved 400 instances of (4.2.4) with different values of f_{lo} and random data y. In

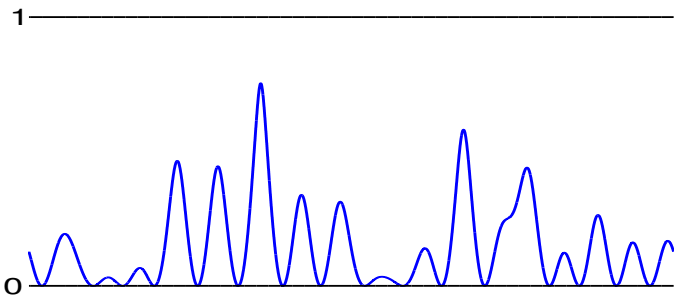


Figure 4.3: The trigonometric polynomial $p_{2n-2}(e^{i2\pi t})$ computed from random data $y \in \mathbb{C}^{21}$ ($n = 21$ and $f_{\text{lo}} = 10$) with i.i.d. complex Gaussian entries. The polynomial has 16 roots.

every single case, Condition (4.2.6) held so that we could construct a primal feasible solution x with a duality gap below 10^{-8}, see Figure 4.4. In all instances, the support of x was constructed by determining roots of $p_{2n-2}(z)$ at a distance at most 10^{-4} from the unit circle.

Interior point methods approach solutions from the interior of the feasible set by solving a sequence of optimization problems in which an extra term, a scaled *barrier function*, is added to the cost function [19]. To be more precise, in our case (4.2.4) would become

$$\max_{\nu,Q} \ \operatorname{Re}\left[y^*\nu\right] + t \log\det\left(\begin{bmatrix} Q & \nu \\ \nu^* & 1 \end{bmatrix}\right) \quad \text{subject to} \quad (4.2.2), \tag{4.2.7}$$

where t is a positive parameter that is gradually reduced towards zero in order to approach a solution to (4.2.4). Let λ_k, $1 \leq k \leq n$, denote the eigenvalues of $Q - \nu\nu^*$. By Schur's formula (Theorem 1.1 in [111]) we have

$$\log\det\left(\begin{bmatrix} Q & \nu \\ \nu^* & 1 \end{bmatrix}\right) = \log\det\left(Q - \nu\nu^*\right) = \sum_{k=1}^{n} \log\lambda_k.$$

Suppose Condition (4.2.6) holds. Lemma 4.2.2 below states that there exists a solution to Problem (4.2.4) $\tilde{\nu}$, $\tilde{Q}$ with the property that at least one eigenvalue of $\tilde{Q} - \tilde{\nu}\tilde{\nu}^*$ is bounded away from zero. This immediately implies that the magnitude of the corresponding polynomial $\mathcal{F}_n^*\tilde{\nu}$ is not equal to one all over the unit interval and therefore

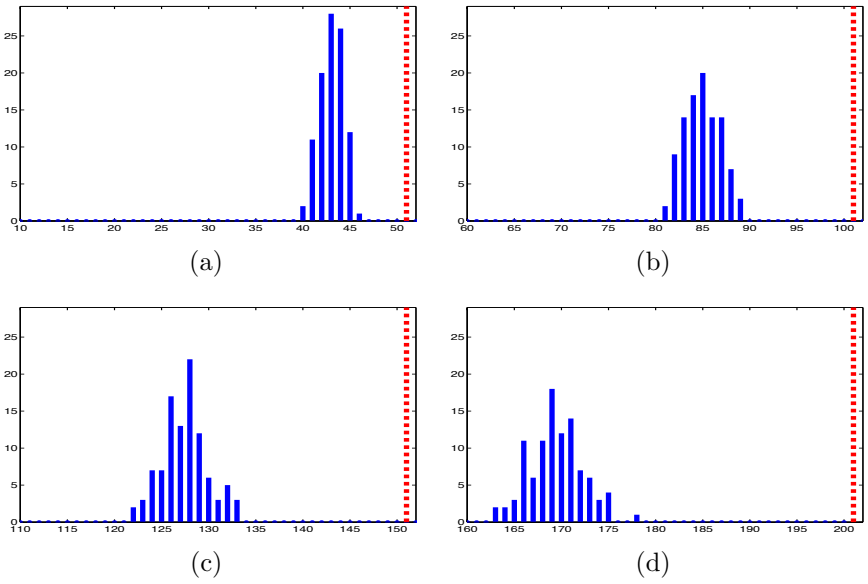


Figure 4.4: Primal feasible points x with a duality gap below 10^{-8} are constructed from random data y sampled with i.i.d. complex Gaussian entries. A dual gap below 10^{-8} implies that $\|x\|_{\mathrm{TV}} - \|x_{\mathrm{est}}\|_{\mathrm{TV}} \leq 10^{-8}$, where x_{est} is any primal optimal solution. (For reference, the optimal value $\|x_{\mathrm{est}}\|_{\mathrm{TV}}$ is on the order of 10 in all cases.) Each figure plots the frequency of occurrence of support cardinalities out of 100 realizations. For example, in (a) we obtained a support size equal to 44 in 25 instances out of 100. The value of n is the same in each plot and is marked by a dashed red line; (a) $n = 51$, (b) $n = 101$, (c) $n = 151$, (d) $n = 201$.

the solution will be informative if we use an interior-point method and will allow us to estimate the support. More precisely, in the limit $t \to 0$, (4.2.7) will construct a non-vanishing polynomial p_{2n-2} with at most $n-1$ roots on the unit circle rather than the trivial solution $p_{2n-2} = 0$ since in the latter case, all the eigenvalues of $\tilde{Q} - \tilde{\nu}\tilde{\nu}^*$ vanish. Hence, an interior-point method can be said to solve the primal problem (2.2.1) provided Condition (4.2.6) holds.

Lemma 4.2.2. *The solution $\tilde{\nu}$, $\tilde{Q}$ from Condition* (4.2.6) *is such that*

$$\tilde{Q} - \tilde{\nu}\tilde{\nu}^* \succ 0.$$

Proof. Let $\tilde{t}$ be the point at which $\left|(\mathcal{F}_n^*\tilde{\nu})\left(\tilde{t}\right)\right| < 1$. Then,

$$\begin{aligned} z_0\left(\tilde{t}\right)^*\left(\tilde{Q}-\tilde{\nu}\tilde{\nu}^*\right)z_0\left(\tilde{t}\right) &= \mathrm{Tr}\left(z_0\left(\tilde{t}\right)^*\tilde{Q}z_0\left(\tilde{t}\right)\right) - z_0\left(\tilde{t}\right)^*\tilde{\nu}\tilde{\nu}^*z_0\left(\tilde{t}\right) \\ &= \left\langle \tilde{Q}, z_0\left(\tilde{t}\right)z_0\left(\tilde{t}\right)^*\right\rangle - \left|(\mathcal{F}_n^*\tilde{\nu})\left(\tilde{t}\right)\right|^2 \\ &= \left\langle \tilde{Q}, \mathcal{T}\left(z_1\left(\tilde{t}\right)\right)\right\rangle - \left|(\mathcal{F}_n^*\tilde{\nu})\left(\tilde{t}\right)\right|^2 \\ &= \left\langle \mathcal{T}^*\left(\tilde{Q}\right), z_1\left(\tilde{t}\right)\right\rangle - \left|(\mathcal{F}_n^*\tilde{\nu})\left(\tilde{t}\right)\right|^2 \\ &= 1 - \left|(\mathcal{F}_n^*\tilde{\nu})\left(\tilde{t}\right)\right|^2 > 0. \end{aligned}$$

□

4.3 Reformulation of the dual problem for super-resolution from noisy data

In this section we discuss how to solve a total-variation norm minimization problem with relaxed constraints that account for perturbations in the data. The dual problem of the convex program

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad ||\mathcal{F}_n\tilde{x} - y||_2 \leq \delta, \tag{4.3.1}$$

takes the form

$$\max_{\nu\in\mathbb{C}^n} \mathrm{Re}\left[y^*\nu\right] - \delta\,||\nu||_2 \quad \text{subject to} \quad ||\mathcal{F}_n^*\,\nu||_\infty \leq 1.$$

This dual problem can be recast as the semidefinite program

$$\max_{\nu,Q} \mathrm{Re}\left[y^*\nu\right] - \delta\,||\nu||_2 \quad \text{subject to} \quad \begin{bmatrix} Q & \nu \\ \nu^* & 1 \end{bmatrix} \succeq 0, \quad \mathcal{T}^*\left(Q\right) = e_1, \tag{4.3.2}$$

where Q is an $n\times n$ Hermitian matrix, leveraging Proposition 4.2.1. As in the noiseless case, our aim is to obtain a primal solution, which corresponds to the estimate of our signal of interest. This means that after solving Problem (4.3.2) we still need a

procedure to compute a solution to the primal. The following lemma provides such a procedure. Similarly to the noiseless case, the argument is based on strong duality.

Lemma 4.3.1. *Let* $(x_{\text{est}}, \nu_{\text{est}})$ *be a primal-dual pair of solutions to* (4.3.1)–(4.3.2). *For any* $t \in \mathbb{T}$ *with* $x_{\text{est}}(t) \neq 0$,

$$(\mathcal{F}_n^* \nu_{\text{est}})(t) = \text{sign}(x_{\text{est}}(t)).$$

Proof. Since x_{est} is feasible, $\|y - \mathcal{F}_n x_{\text{est}}\|_{\ell_2} \leq \delta$. Strong duality holds here by a generalized Slater condition [84]. Hence, we have that

$$\begin{aligned} ||x_{\text{est}}||_{\text{TV}} &= \text{Re}\left[y^* \nu_{\text{est}}\right] - \delta ||\nu_{\text{est}}||_2 \\ &= \langle \mathcal{F}_n x_{\text{est}}, \nu_{\text{est}} \rangle + \langle y - \mathcal{F}_n x_{\text{est}}, \nu_{\text{est}} \rangle - \delta ||\nu_{\text{est}}||_2 \\ &\leq \langle x_{\text{est}}, \mathcal{F}_n^* \nu_{\text{est}} \rangle . \end{aligned}$$

Since the Cauchy-Schwarz inequality gives

$$\delta ||\nu_{\text{est}}||_2 \geq ||y - \mathcal{F}_n x_{\text{est}}||_2 ||\nu_{\text{est}}||_2 \geq \langle y - \mathcal{F}_n x_{\text{est}}, \nu_{\text{est}} \rangle .$$

By Hölder's inequality and the constraint on $\mathcal{F}_n^* \nu_{\text{est}}$, $||x_{\text{est}}||_{\text{TV}} \geq \langle x_{\text{est}}, \mathcal{F}_n^* \nu_{\text{est}} \rangle$ so that equality holds. This is only possible if $\mathcal{F}_n^* \nu_{\text{est}}$ equals the sign of x_{est} at every point where x_{est} is nonzero. □

This result implies that it is usually possible to determine the support of the primal solution by locating those points where the polynomial $q(t) = (\mathcal{F}_n^* \nu_{\text{est}})(t)$ has modulus equal to one. Once the support is estimated accurately, a solution to the primal problem can be found by solving a discrete least-squares problem.

Figure 4.5 shows the results of some numerical experiments that illustrate this approach. On the left column we see the projection of the data onto the signal space compared to the noiseless low-pass projection of the signal for different values of the signal-to-noise ratio induced by adding i.i.d. white Gaussian noise. On the right column, we see the result of solving the semidefinite program (4.3.2), locating the support by determining at what points the modulus of $\mathcal{F}_n^* \nu_{\text{est}}(t)$ is equal to one and

fitting the amplitude of the spikes using least squares. For large signal-to-noise ratios the estimate produced in this way is very accurate. For lower signal-to-noise ratios we are still able to locate the larger spikes. This behavior is in agreement with the theoretical results on support detection provided in Section 3.3.

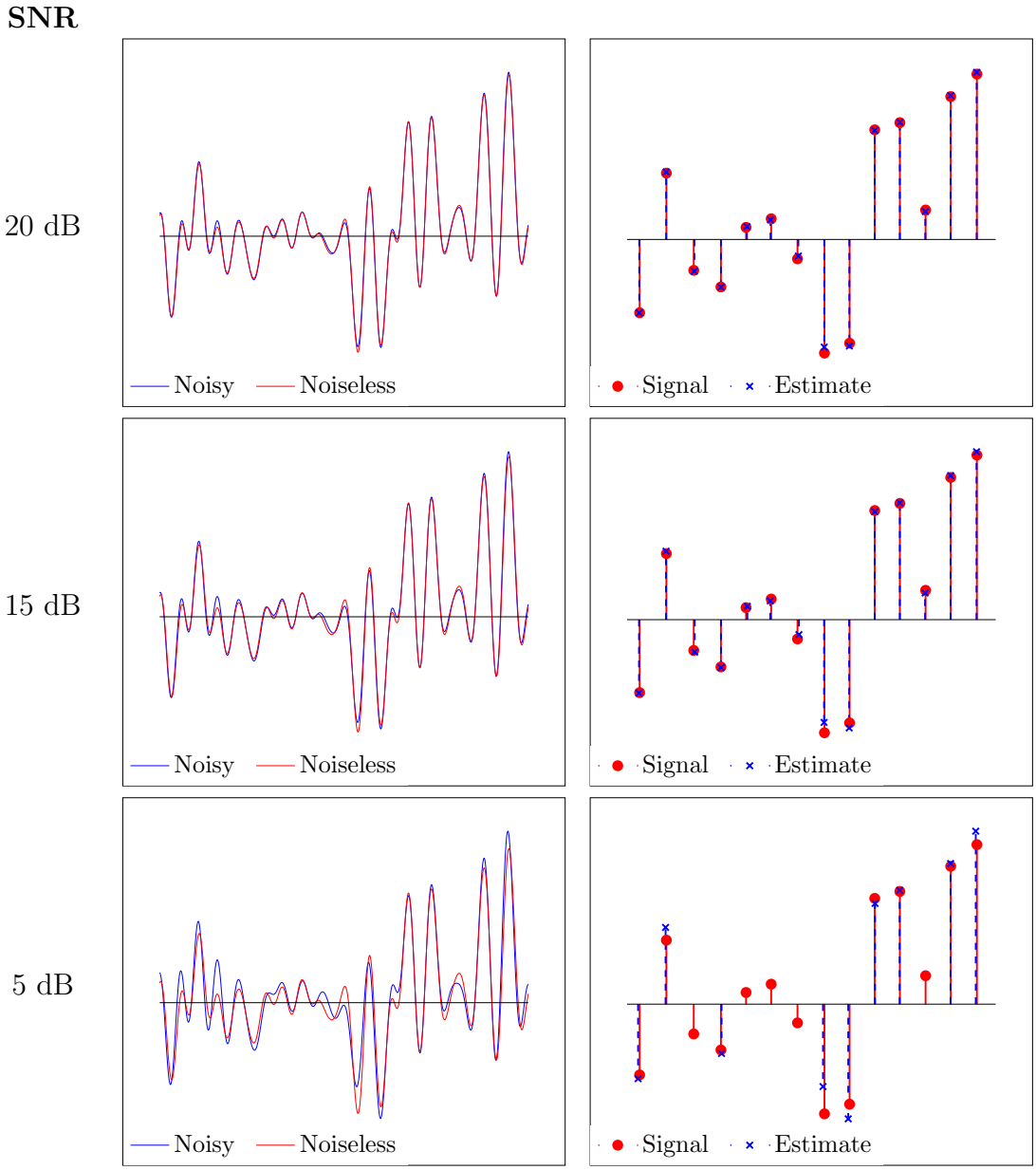


Figure 4.5: Results of super-resolving a signal by applying semidefinite programming for different values of the signal-to-noise ratio induced by adding i.i.d. white Gaussian noise.

Chapter 5

Proof of Exact Recovery

This chapter proves that total-variation norm minimization achieves exact recovery under a minimum-separation condition on the signal support, as stated in Theorem 2.2.1. Section 5.1 explains how the proof relies on the construction of a certain polynomial that certifies exact recovery. Section 5.2 presents a novel proof technique which allows to construct such a polynomial when the support of the signal is not too clustered together. A number of subsequent works build upon this construction to prove the support-detection capability of super-resolution via convex programming [2, 46], guarantees on denoising of line spectra via convex programming [101], recovery of sparse signals from a random subset of their low-pass Fourier coefficients [102], recovery of non-uniform splines from their projection onto spaces of algebraic polynomials [12, 37] and recovery of point sources from their projection onto spherical harmonics [11]. Finally, the lemmas used to establish the result are proved in Section 5.3. Throughout the chapter we write $\Delta = \Delta(T) \geq \Delta_{\min} = 2\,\lambda_{\text{lo}}$. Also, we identify the interval $[0, 1)$ with the circle $\mathbb{T}$.

5.1 Dual certificates

In the discrete setting, the compressed sensing literature has made clear that the existence of a certain *dual certificate* guarantees that the ℓ_1 solution is exact [26]. In the continuous setting, a sufficient condition for the success of the total-variation

solution also concerns the existence of a solution to the dual of the convex program, which corresponds to a low-pass polynomial that interpolates the sign pattern of the signal. This is made precise by the following proposition.

Proposition 5.1.1. *Suppose that for any vector $v \in \mathbb{C}^{|T|}$ with unit-magnitude entries, there exists a low-frequency trigonometric polynomial*

$$q(t) = \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} a_k e^{i2\pi kt} \tag{5.1.1}$$

obeying the following properties

$$\begin{cases} q(t_j) = v_j, & t_j \in T, \\ |q(t)| < 1, & t \in \mathbb{T} \setminus T. \end{cases} \tag{5.1.2}$$

Then x is the unique solution to Problem (2.2.1)*.*

Proof. The proof is a variation on the well-known argument for finite signals, and we note that a proof for continuous-time signals, similar to that below, can be found in [36]. Let x_{est} be a solution to (2.2.1) and set $x_{\text{est}} = x + h$. Consider the Lebesgue decomposition of h relative to $|x|$,

$$h = h_T + h_{T^c},$$

where (1) h_T and h_{T^c} is a unique pair of complex measures on $\mathcal{B}(\mathbb{T})$ such that h_T is absolutely continuous with respect to $|x|$, and (2) h_{T^c} and $|x|$ are mutually singular. It follows that h_T is concentrated on T while h_{T^c} is concentrated on T^c. Invoking a corollary of the Radon-Nykodim Theorem (see Theorem 6.12 in [87]), it is possible to perform a polar decomposition of h_T,

$$h_T = e^{i2\pi\phi(t)} |h_T| ,$$

such that $\phi(t)$ is a real function defined on $\mathbb{T}$. We can now choose v such that $v_j = e^{-i2\pi\phi(t_j)}$ for all $t_j \in T$, so that $\int_{\mathbb{T}} q(t) h_T(\mathrm{d}t) = ||h_T||_{\text{TV}}$. The existence of the

corresponding q suffices to establish a valuable inequality between the total-variation norms of h_T and h_{T^c}. Begin with

$$0 = \int_{\mathbb{T}} q(t)h(\mathrm{d}t) = \int_{\mathbb{T}} q(t)h_T(\mathrm{d}t) + \int_{\mathbb{T}} q(t)h_{T^c}(\mathrm{d}t) = ||h_T||_{\mathrm{TV}} + \int_{\mathbb{T}} q(t)h_{T^c}(\mathrm{d}t)$$

and observe that

$$\left| \int_{\mathbb{T}} q(t)h_{T^c}(\mathrm{d}t) \right| < ||h_{T^c}||_{\mathrm{TV}}$$

provided $h_{T^c} \neq 0$. This gives

$$||h_T||_{\mathrm{TV}} \leq ||h_{T^c}||_{\mathrm{TV}}$$

with a strict inequality if $h \neq 0$. Assuming $h \neq 0$, we have

$$||x||_{\mathrm{TV}} \geq ||x+h||_{\mathrm{TV}} = ||x+h_T||_{\mathrm{TV}} + ||h_{T^c}||_{\mathrm{TV}} \geq ||x||_{\mathrm{TV}} - ||h_T||_{\mathrm{TV}} + ||h_{T^c}||_{\mathrm{TV}} > ||x||_{\mathrm{TV}}\,.$$

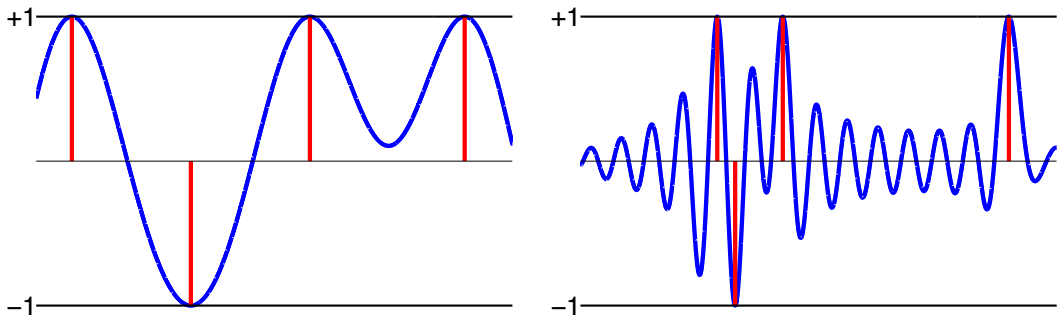
This is a contradiction and thus $h = 0$. In other words, x is the unique minimizer. □

Constructing a bounded low-frequency polynomial interpolating the sign pattern of certain signals becomes increasingly difficult if the minimum distance separating the spikes is too small. This is illustrated in Figure 5.1, where we show that if spikes are very near, it would become in general impossible to find an interpolating low-frequency polynomial obeying (5.1.2).

5.2 Proof of Theorem 2.2.1

Theorem 2.2.1 is a direct consequence of combining Proposition (5.1.1) and the proposition below, which establishes the existence of a valid dual certificate provided the elements in the support are sufficiently spaced.

Proposition 5.2.1. *Let $v \in \mathbb{C}^{|T|}$ be an arbitrary vector obeying $|v_j| = 1$. Then under the hypotheses of Theorem 2.2.1, there exists a low-frequency trigonometric polynomial obeying* (5.1.2).



(a) (b)

Figure 5.1: (a) Low-frequency polynomial interpolating a sign pattern in which the support is well separated, and obeying the off-support condition (5.1.2). In (b), we see that if the spikes become very near, we would need a rapidly (high-frequency) interpolating polynomial in order to achieve (5.1.2). This is the reason why there must be a minimum separation between consecutive spikes.

The remainder of the chapter proves this proposition. Our method consists in interpolating v on T with a low-frequency kernel and correcting the interpolation to ensure that the derivative of the dual polynomial is zero on T. The kernel we employ is

$$K(t) = \left[\frac{\sin\left(\left(\frac{f_{\text{lo}}}{2} + 1\right)\pi t\right)}{\left(\frac{f_{\text{lo}}}{2} + 1\right)\sin\left(\pi t\right)} \right]^4, \quad 0 < t < 1, \tag{5.2.1}$$

and $K(0) = 1$. If f_{lo} is even, $K(t)$ is the square of the Fejér kernel which is a trigonometric polynomial with frequencies obeying $|k| \leq f_{\text{lo}}/2$. As a consequence, K is of the form (5.1.1). The careful reader might remark that the choice of the interpolation kernel seems somewhat arbitrary. In fact, one could also use the Fejér kernel or any other power of the Fejér kernel using almost identical proof techniques. We have found that the second power nicely balances the trade-off between localization in time and in frequency, and thus yields a good constant.

To construct the dual polynomial, we interpolate v with both K and its derivative K',

$$q(t) = \sum_{t_j \in T} \alpha_j K(t - t_j) + \beta_j K'(t - t_j), \tag{5.2.2}$$

where $\alpha, \beta \in \mathbb{C}^{|T|}$ are coefficient sequences. The polynomial q is as in (5.1.1). In

order to make it obey $q(t_k) = v_k$, we impose

$$q(t_k) = \sum_{t_j \in T} \alpha_j K\left(t_k - t_j\right) + \beta_j K'\left(t_k - t_j\right) = v_k, \quad \forall t_k \in T, \tag{5.2.3}$$

whereas to ensure that it verifies $|q(t)| < 1$ for $t \in T^c$, we impose $q'(t_k) = 0$,

$$q'(t_k) = \sum_{t_j \in T} \alpha_j K'\left(t_k - t_j\right) + \beta_j K''\left(t_k - t_j\right) = 0, \quad \forall t_k \in T. \tag{5.2.4}$$

As we will see, this implies that the magnitude of q reaches a local maximum at those points, which in turn can be used to show that (5.1.2) holds.

The proof of Proposition 5.2.1 consists of three lemmas, which are proved in the following section. The first one establishes that if the support is spread out, it is possible to interpolate any sign pattern exactly.

Lemma 5.2.2. *Under the hypotheses of Proposition 5.2.1, there exist coefficient vectors α and β obeying*

$$\begin{aligned} ||\alpha||_\infty &\leq \alpha^\infty := 1 + 8.824\ 10^{-3}, \\ ||\beta||_\infty &\leq \beta^\infty := 3.294\ 10^{-2}\, \lambda_{\text{lo}}, \end{aligned} \tag{5.2.5}$$

such that (5.2.3)–(5.2.4) *hold. Further, if $v_1 = 1$,*

$$\begin{aligned} \operatorname{Re} \alpha_1 &\geq 1 - 8.824\ 10^{-3}, \\ |\operatorname{Im} \alpha_1| &\leq 8.824\ 10^{-3}. \end{aligned} \tag{5.2.6}$$

To complete the proof, Lemmas 5.2.3 and 5.2.4 show that $|q\,(t)\,| < 1$ in the unit interval.

Lemma 5.2.3. *Fix $\tau \in T$. Under the hypotheses of Proposition 5.2.1, $|q(t)| < 1$ for $|t - \tau| \in (0, 0.1649\, \lambda_{\text{lo}}]$.*

Lemma 5.2.4. *Fix $\tau \in T$. Then under the hypotheses of Proposition 5.2.1, $|q(t)| < 1$ for $|t - \tau| \in [0.1649\, \lambda_{\text{lo}}, \Delta/2]$. This can be extended as follows: letting τ_+ be the closest spike to the right, i. e. $\tau_+ = \min\{t \in T : t > \tau\}$. Then $|q(t)| < 1$ for all t obeying $0 < t - \tau \leq (\tau_+ - \tau)/2$, and likewise for the left side.*

In addition, we record a useful lemma to derive stability results.

Lemma 5.2.5. *If $\Delta(T) \geq 2.5\,\lambda_{\mathrm{lo}}$, then for any $\tau \in T$,*

$$|q(t)| \leq 1 - 0.3353 f_{\mathrm{lo}}^2 (t-\tau)^2, \quad \text{for all } t:\ |t-\tau| \leq 0.1649\,\lambda_{\mathrm{lo}}. \tag{5.2.7}$$

Further, for $\min_{\tau\in T} |t-\tau| > 0.1649\,\lambda_{\mathrm{lo}}$, $|q(t)|$ is upper bounded by the right-hand side above evaluated at $0.1649\,\lambda_{\mathrm{lo}}$.

Finally, Section 5.3.5 describes how the proof can be adapted to obtain a slightly smaller bound on the minimum distance for real-valued signals.

5.3 Proofs of lemmas

The proofs of the three main lemmas of the previous section make repeated use of the fact that the interpolation kernel and its derivatives decay rapidly away from the origin. The intermediate result below, proved in Section 5.3.6, quantifies this by establishing upper bounds on the magnitude of the kernel.

Lemma 5.3.1. *For $\ell \in \{0,1,2,3\}$, let $K^{(\ell)}$ be the ℓth derivative of K ($K = K^{(0)}$). For $\frac{1}{2} f_{\mathrm{lo}}^{-1} = \frac{1}{2}\lambda_{\mathrm{lo}} \leq t \leq \frac{1}{2}$, we have*

$$\left|K^{(\ell)}(t)\right| \leq B_\ell(t) = \begin{cases} \tilde{B}_\ell(t) = \frac{\pi^\ell H_\ell(t)}{(f_{\mathrm{lo}}+2)^{4-\ell}\, t^4}, & \frac{1}{2}\lambda_{\mathrm{lo}} \leq t \leq \sqrt{2}/\pi, \\ \frac{\pi^\ell H_\ell^\infty}{(f_{\mathrm{lo}}+2)^{4-\ell}\, t^4}, & \sqrt{2}/\pi \leq t < \frac{1}{2}, \end{cases}$$

where $H_0^\infty = 1$, $H_1^\infty = 4$, $H_2^\infty = 18$, $H_3^\infty = 77$,

$$\begin{aligned} H_0(t) &= a^4(t), \\ H_1(t) &= a^4(t)\left(2 + 2b(t)\right), \\ H_2(t) &= a^4(t)\left(4 + 7b(t) + 6b^2(t)\right), \\ H_3(t) &= a^4(t)\left(8 + 24b(t) + 30b^2(t) + 15b^3(t)\right), \end{aligned}$$

and

$$a(t) = \frac{2}{\pi\left(1 - \frac{\pi^2 t^2}{6}\right)}, \qquad b(t) = \frac{1}{f_{\text{lo}}}\frac{a(t)}{t}.$$

For each ℓ, the bound on the magnitude of $B_\ell(t)$ is nonincreasing in t and $\tilde{B}_\ell(\Delta - t) + \tilde{B}_\ell(\Delta + t)$ is increasing in t for $0 \le t < \Delta/2$ if $0 \le \Delta + t \le \sqrt{2}/\pi$.

This result can be used to control quantities of the form $\sum_{t_i \in T\setminus\{\tau\}} |K(t - t_i)|$ for $\tau \in T$ as shown in the following lemma.

Lemma 5.3.2. *Suppose $0 \in T$. Then for all $t \in [0, \Delta/2]$,*

$$\sum_{t_i \in T\setminus\{0\}} \left|K^{(\ell)}(t - t_i)\right| \le F_\ell(\Delta, t) = F_\ell^+(\Delta, t) + F_\ell^-(\Delta, t) + F_\ell^\infty(\Delta_{\min}),$$

where

$$F_\ell^+(\Delta, t) = \max\left\{\max_{\Delta \le t_+ \le 3\Delta_{\min}} \left|K^{(\ell)}(t - t_+)\right|, B_\ell(3\Delta_{\min} - t)\right\} + \sum_{j=2}^{20} \tilde{B}_\ell(j\Delta_{\min} - t),$$

$$F_\ell^-(\Delta, t) = \max\left\{\max_{\Delta \le t_- \le 3\Delta_{\min}} \left|K^{(\ell)}(t_-)\right|, B_\ell(3\Delta_{\min})\right\} + \sum_{j=2}^{20} \tilde{B}_\ell(j\Delta_{\min} + t),$$

$$F_\ell^\infty(\Delta_{\min}) = \frac{\kappa\,\pi^\ell H_\ell^\infty}{(f_{\text{lo}} + 2)^{4-\ell}\Delta_{\min}^4}, \qquad \kappa = \frac{\pi^4}{45} - 2\sum_{j=1}^{19}\frac{1}{j^4} \le 8.98\,10^{-5}.$$

Moreover, $F_\ell(\Delta, t)$ is nonincreasing in Δ for all t, and $F_\ell(\Delta_{\min}, t)$ is nondecreasing in t.

Proof. We consider the sum over positive $t_i \in T$ first and denote by t_+ the positive element in T closest to 0. We have

$$\sum_{t_i \in T:\, 0 < t_i \le 1/2} \left|K^{(\ell)}(t - t_i)\right| = \left|K^{(\ell)}(t - t_+)\right| + \sum_{t_i \in T\setminus\{t_+\}:\, 0 < t_i \le 1/2} \left|K^{(\ell)}(t - t_i)\right|. \quad (5.3.1)$$

Let us assume $t_+ < 2\Delta_{\min}$ (if $t_+ > 2\Delta_{\min}$ the argument is very similar). Note that the

assumption that $f_{\text{lo}} \geq 128$ implies $21\Delta_{\min} < 0.33 < \sqrt{2}/\pi$. By Lemma 5.3.1 and the minimum-separation condition, this means that the second term on the right-hand side is at most

$$\sum_{j=2}^{20} \tilde{B}_\ell(j\Delta_{\min} - t) + \frac{\pi^\ell}{(f_{\text{lo}}+2)^{4-\ell}} \sum_{j=21}^{\infty} \frac{H_\ell^\infty}{(j\Delta_{\min} \pm t)^4}, \tag{5.3.2}$$

which can be upper bounded using the fact that

$$\begin{aligned}
\sum_{j=21}^{\infty} \frac{H_\ell^\infty}{(j\Delta_{\min} \pm t)^4} &\leq \sum_{j=20}^{\infty} \frac{H_\ell^\infty}{(j\Delta_{\min})^4} \\
&= \frac{H_\ell^\infty}{\Delta_{\min}^4} \left(\sum_{j=1}^{\infty} \frac{1}{j^4} - \sum_{j=1}^{19} \frac{1}{j^4} \right) \\
&= \frac{H_\ell^\infty}{\Delta_{\min}^4} \left(\frac{\pi^4}{90} - \sum_{j=1}^{19} \frac{1}{j^4} \right) \\
&= \frac{\kappa H_\ell^\infty}{2\Delta_{\min}^4};
\end{aligned}$$

the first inequality holds because $t < \Delta_{\min}$ and the last because the Riemann zeta function is equal to $\pi^4/90$ at 4. Also,

$$\left|K^{(\ell)}(t - t_+)\right| \leq \begin{cases} \max_{\Delta \leq t_+ \leq 3\Delta_{\min}} \left|K^{(\ell)}(t - t_+)\right|, & t_+ \leq 3\Delta_{\min}, \\ B_\ell(3\Delta_{\min} - t), & t_+ > 3\Delta_{\min}. \end{cases}$$

Hence, the quantity in (5.3.1) is bounded by $F_\ell^+(\Delta, t) + F_\ell^\infty(\Delta_{\min})/2$. A similar argument shows that the sum over negative $t_i \in T$ is bounded by $F_\ell^-(\Delta, t) + F_\ell^\infty(\Delta_{\min})/2$.

To verify the claim about the monotonicity with respect to Δ, observe that both terms

$$\max \left\{ \max_{\Delta \leq t_+ \leq 3\Delta_{\min}} \left|K^{(\ell)}(t - t_+)\right|, B_\ell(3\Delta_{\min} - t) \right\}$$

and

$$\max\left\{\max_{\Delta\le t_-\le 3\Delta_{\min}}\left|K^{(\ell)}\left(t_-\right)\right|, B_\ell\left(3\Delta_{\min}\right)\right\}$$

are nonincreasing in Δ.

Fix $\Delta=\Delta_{\min}$ now. Since $\tilde{B}_\ell\left(j\Delta-t\right)+\tilde{B}_\ell\left(j\Delta+t\right)$ is increasing in t for $j\le 20$ (recall that $21\Delta_{\min}<\sqrt{2}/\pi$), we only need to check that the first term in the expression for F_ℓ^+ is nondecreasing in t. To see this, rewrite this term (with $\Delta=\Delta_{\min}$) as

$$\max\left\{\max_{\Delta_{\min}-t\le u\le 3\Delta_{\min}-t}\left|K^{(\ell)}\left(u\right)\right|, B_\ell\left(3\Delta_{\min}-t\right)\right\}.$$

Now set $t'>t$. Then by Lemma 5.3.1,

$$B_\ell\left(3\Delta_{\min}-t'\right)\ge\begin{cases}B_\ell\left(3\Delta_{\min}-t\right),\\ \left|K(u)\right|, & u\ge 3\Delta_{\min}-t'.\end{cases}$$

Also, we can verify that

$$\max_{\Delta_{\min}-t'\le u\le 3\Delta_{\min}-t'}\left|K^{(\ell)}\left(u\right)\right|\ge\max_{\Delta_{\min}-t\le u\le 3\Delta_{\min}-t'}\left|K^{(\ell)}\left(u\right)\right|.$$

This concludes the proof. □

In the proof of Lemmas 5.2.3 and 5.2.4, it is necessary to find a numerical upper bound on $F_\ell(\Delta_{\min},t)$ at $t\in\{0,0.1649\,\lambda_{\text{lo}},0.4269\,\lambda_{\text{lo}},0.7559\,\lambda_{\text{lo}}\}$ (for the last two points, we only need bounds for $\ell=0,1$). For a fixed t, it is easy to find the maximum of $\left|K^{(\ell)}\left(t-t_+\right)\right|$ where t_+ ranges over $[\Delta_{\min},3\Delta_{\min}]$ since we have expressions for the smooth functions $K^{(\ell)}$ (see Section 5.3.6). For reference, these functions are plotted in Figure 5.2. The necessary upper bounds are gathered in Table 5.1.

Finally, a last fact we shall use is that $K\left(0\right)=1$ is the global maximum of K and $\left|K''\left(0\right)\right|=\left|-\pi^2 f_{\text{lo}}\left(f_{\text{lo}}+4\right)/3\right|$ the global maximum of $|K''|$.

t/λ_{lo}	$F_0\left(1.98\lambda_{\text{lo}}, t\right)$	$F_1\left(1.98\lambda_{\text{lo}}, t\right)$	$F_2\left(1.98\lambda_{\text{lo}}, t\right)$	$F_3\left(1.98\lambda_{\text{lo}}, t\right)$
0	$6.253\ 10^{-3}$	$7.639\ 10^{-2} f_{\text{lo}}$	$1.053\ f_{\text{lo}}^2$	$8.078\ f_{\text{lo}}^3$
0.1649	$6.279\ 10^{-3}$	$7.659\ 10^{-2} f_{\text{lo}}$	$1.055\ f_{\text{lo}}^2$	$18.56\ f_{\text{lo}}^3$
0.4269	$8.029\ 10^{-3}$	$0.3042 f_{\text{lo}}$		
0.7559	$5.565\ 10^{-2}$	$1.918 f_{\text{lo}}$		

Table 5.1: Numerical upper bounds on $F_\ell(1.98\lambda_{\text{lo}}, t)$.

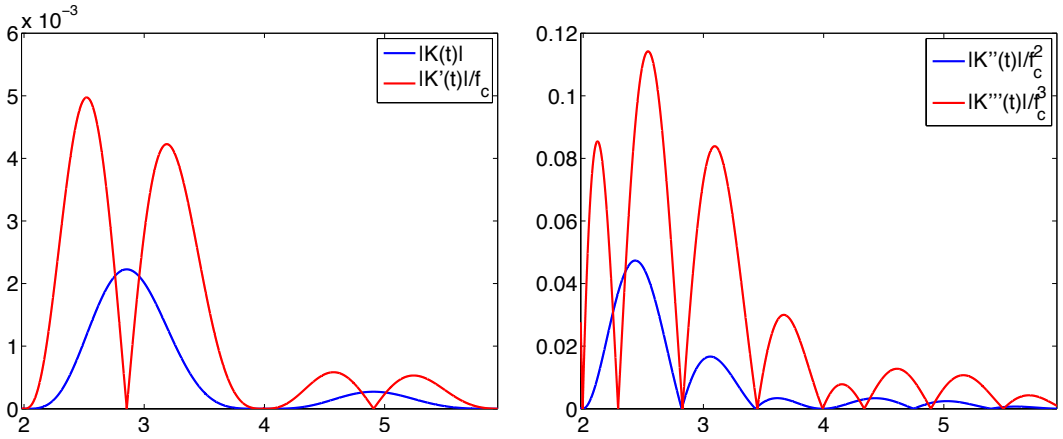


Figure 5.2: $\left|K^{(\ell)}(t)\right|$ for $t \in [\Delta_{\min}, 3\Delta_{\min}]$. The scaling of the x-axis is in units of λ_{lo}.

5.3.1 Proof of Lemma 5.2.2

Set

$$(D_0)_{jk} = K\left(t_j - t_k\right), \quad (D_1)_{jk} = K'\left(t_j - t_k\right), \quad (D_2)_{jk} = K''\left(t_j - t_k\right),$$

where j and k range from 1 to $|T|$. With this, (5.2.3) and (5.2.4) become

$$\begin{bmatrix} D_0 & D_1 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix}.$$

A standard linear algebra result asserts that this system is invertible if and only if D_2 and its Schur complement $D_0 - D_1 D_2^{-1} D_1$ are both invertible. To prove that this

is the case we can use the fact that a symmetric matrix M is invertible if

$$||I - M||_\infty < 1, \tag{5.3.3}$$

where $\|A\|_\infty$ is the usual infinity norm of a matrix defined as

$$\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty = \max_i \sum_j |A_{ij}|.$$

This follows from $M^{-1} = (I - H)^{-1} = \sum_{k \geq 0} H^k$, $H = I - M$, where the series is convergent since $||H||_\infty < 1$. In particular,

$$\left|\left|M^{-1}\right|\right|_\infty \leq \frac{1}{1 - ||I - M||_\infty}. \tag{5.3.4}$$

We also make use of the inequalities below, which follow from Lemma 5.3.2,

$$||I - D_0||_\infty \leq F_0\left(\Delta_{\min}, 0\right) \leq 6.253\ 10^{-3}, \tag{5.3.5}$$

$$||D_1||_\infty \leq F_1\left(\Delta_{\min}, 0\right) \leq 7.639\ 10^{-2}\ f_{\text{lo}}, \tag{5.3.6}$$

$$||K''(0)\, I - D_2||_\infty \leq F_2\left(\Delta_{\min}, 0\right) \leq 1.053\ f_{\text{lo}}^2. \tag{5.3.7}$$

Note that D_2 is symmetric because the second derivative of the interpolation kernel is symmetric. The bound (5.3.7) and the identity $K''(0) = -\pi^2 f_{\text{lo}}\left(f_{\text{lo}} + 4\right)/3$ give

$$\left|\left|I - \frac{D_2}{K''(0)}\right|\right|_\infty < 1,$$

which implies the invertibility of D_2. The bound (5.3.4) then gives

$$\left|\left|D_2^{-1}\right|\right|_\infty \leq \frac{1}{|K''(0)| - ||K''(0)\, I - D_2||_\infty} \leq \frac{0.4275}{f_{\text{lo}}^2}. \tag{5.3.8}$$

Combining this with (5.3.5) and (5.3.6) yields

$$\begin{aligned}\left|\left|I - \left(D_0 - D_1 D_2^{-1} D_1\right)\right|\right|_\infty &\leq ||I - D_0||_\infty + ||D_1||_\infty^2 \left|\left|D_2^{-1}\right|\right|_\infty \\ &\leq 8.747\ 10^{-3} < 1. \end{aligned} \tag{5.3.9}$$

Note that the Schur complement of D_2 is symmetric because D_0 and D_2 are both symmetric whereas $D_1^T = -D_1$ since the derivative of the interpolation kernel is odd. This shows that the Schur complement of D_2 is invertible and, therefore, the coefficient vectors α and β are well defined.

There just remains to bound the interpolation coefficients, which can be expressed as

$$\begin{bmatrix}\alpha \\ \beta\end{bmatrix} = \begin{bmatrix} I \\ -D_2^{-1} D_1 \end{bmatrix} C^{-1} v, \quad C := D_0 - D_1 D_2^{-1} D_1,$$

where C is the Schur complement. The relationships (5.3.4) and (5.3.9) immediately give a bound on the magnitude of the entries of α

$$||\alpha||_\infty = \left|\left|C^{-1} v\right|\right|_\infty \leq \left|\left|C^{-1}\right|\right|_\infty \leq 1 + 8.824\ 10^{-3}.$$

Similarly, (5.3.6), (5.3.8) and (5.3.9) allow to bound the entries of β:

$$\begin{aligned}||\beta||_\infty &\leq \left|\left|D_2^{-1} D_1 C^{-1}\right|\right|_\infty \\ &\leq \left|\left|D_2^{-1}\right|\right|_\infty ||D_1||_\infty \left|\left|C^{-1}\right|\right|_\infty \leq 3.294\ 10^{-2}\, \lambda_{\text{lo}}. \end{aligned}$$

Finally, with $v_1 = 1$, we can use (5.3.9) to show that α_1 is almost equal to 1. Indeed,

$$\alpha_1 = 1 - \gamma_1, \quad \gamma_1 = [(I - C^{-1})v]_1,$$

$|\gamma_1| \leq ||I - C^{-1}||_\infty$, and

$$\left|\left|I - C^{-1}\right|\right|_\infty = \left|\left|C^{-1}(I - C)\right|\right|_\infty \leq \left|\left|C^{-1}\right|\right|_\infty ||I - C||_\infty \leq 8.824\ 10^{-3}.$$

This concludes the proof.

5.3.2 Proof of Lemma 5.2.3

We assume without loss of generality that $\tau = 0$ and $q(0) = 1$. By symmetry, it suffices to show the claim for $t \in (0, 0.1649\,\lambda_{\text{lo}}]$. Since $q'(0) = 0$, local strict concavity would imply that $|q(t)| < 1$ near the origin. We begin by showing that the second derivative of $|q|$ is strictly negative in the interval $(0, 0.1649\,\lambda_{\text{lo}})$. This derivative is equal to

$$\frac{\mathrm{d}^2 |q|}{\mathrm{d}t^2}(t) = -\frac{\left(q_R(t)\,q_R'(t) + q_I(t)\,q_I'(t)\right)^2}{|q(t)|^3} + \frac{|q'(t)|^2 + q_R(t)\,q_R''(t) + q_I(t)\,q_I''(t)}{|q(t)|},$$

where q_R is the real part of q and q_I the imaginary part. As a result, it is sufficient to show that

$$q_R(t)\,q_R''(t) + |q'(t)|^2 + |q_I(t)|\,|q_I''(t)| < 0, \tag{5.3.10}$$

as long as $|q(t)|$ is bounded away from zero. In order to bound the different terms in (5.3.10), we use the series expansions of the interpolation kernel and its derivatives around the origin to obtain the inequalities, which hold for all $t \in [-1/2, 1/2]$,

$$K(t) \geq 1 - \frac{\pi^2}{6} f_{\text{lo}}(f_{\text{lo}} + 4)\,t^2, \tag{5.3.11}$$

$$|K'(t)| \leq \frac{\pi^2}{3} f_{\text{lo}}(f_{\text{lo}} + 4)\,t, \tag{5.3.12}$$

$$K''(t) \leq -\frac{\pi^2}{3} f_{\text{lo}}(f_{\text{lo}} + 4) + \frac{\pi^4}{6}(f_{\text{lo}} + 2)^4\,t^2, \tag{5.3.13}$$

$$|K''(t)| \leq \frac{\pi^2}{3} f_{\text{lo}}(f_{\text{lo}} + 4), \tag{5.3.14}$$

$$|K'''(t)| \leq \frac{\pi^4}{3}(f_{\text{lo}} + 2)^4\,t. \tag{5.3.15}$$

The lower bounds are decreasing in t, while the upper bounds are increasing in t, so we can evaluate them at $0.1649\,\lambda_{\text{lo}}$ to establish that for all $t \in [0, 0.1649\,\lambda_{\text{lo}}]$,

$$\begin{aligned} &K(t) \geq 0.9539, && K''(t) \leq -2.923\,f_{\text{lo}}^2, \\ &|K'(t)| \leq 0.5595\,f_{\text{lo}}, && |K''(t)| \leq 3.393\,f_{\text{lo}}^2, && |K'''(t)| \leq 5.697\,f_{\text{lo}}^3. \end{aligned} \tag{5.3.16}$$

We combine this with Lemmas 5.3.2 and 5.2.2 to control the different terms in (5.3.10) and begin with $q_R(t)$. Here,

$$\begin{aligned} q_R(t) &= \sum_{t_j \in T} \mathrm{Re}(\alpha_j) K(t-t_j) + \mathrm{Re}(\beta_j) K'(t-t_j) \\ &\geq \mathrm{Re}(\alpha_1) K(t) - ||\alpha||_\infty \sum_{t_j \in T \setminus \{0\}} |K(t-t_j)| - ||\beta||_\infty \sum_{t_j \in T} |K'(t-t_j)| \\ &\geq \mathrm{Re}(\alpha_1) K(t) - \alpha^\infty F_0(\Delta, t) - \beta^\infty (|K'(t)| + F_1(\Delta, t)) \\ &\geq \mathrm{Re}(\alpha_1) K(t) - \alpha^\infty F_0(\Delta_{\min}, t) - \beta^\infty (|K'(t)| + F_1(\Delta_{\min}, t)) \\ &\geq 0.9182. \end{aligned} \tag{5.3.17}$$

The third inequality follows from the monotonicity of F_ℓ in Δ, and the last from (5.3.16) together with the monotonicity of $F_1(\Delta_{\min}, t)$ in t, see Lemma 5.3.2, so that we can plug in $t = 0.1649\,\lambda_{\mathrm{lo}}$. Observe that this shows that q is bounded away from zero since $|q(t)| \geq q_R(t) \geq 0.9198$. Very similar computations yield

$$\begin{aligned} |q_I(t)| &= \left| \sum_{t_j \in T} \mathrm{Im}(\alpha_j) K(t-t_j) + \mathrm{Im}(\beta_j) K'(t-t_j) \right| \\ &\leq |\mathrm{Im}(\alpha_1)| + ||\alpha||_\infty \sum_{t_j \in T \setminus \{0\}} |K(t-t_j)| + ||\beta||_\infty \sum_{t_j \in T} |K'(t-t_j)| \\ &\leq |\mathrm{Im}(\alpha_1)| + \alpha^\infty F_0(\Delta_{\min}, t) + \beta^\infty (|K'(t)| + F_1(\Delta_{\min}, t)) \\ &\leq 3.611\ 10^{-2} \end{aligned}$$

and

$$\begin{aligned} q_R''(t) &= \sum_{t_j \in T} \mathrm{Re}(\alpha_j) K''(t-t_j) + \sum_{t_j \in T} \mathrm{Re}(\beta_j) K'''(t-t_j) \\ &\leq \mathrm{Re}(\alpha_1) K''(t) + ||\alpha||_\infty \sum_{t_j \in T \setminus \{0\}} |K''(t-t_j)| + ||\beta||_\infty \sum_{t_j \in T} |K'''(t-t_j)| \\ &\leq \mathrm{Re}(\alpha_1) K''(t) + \alpha^\infty F_2(\Delta_{\min}, t) + \beta^\infty (|K'''(t)| + F_3(\Delta_{\min}, t)) \\ &\leq -1.034\, f_{\mathrm{lo}}^2. \end{aligned} \tag{5.3.18}$$

Similarly,

$$\begin{aligned}
|q_I''(t)| &= \left| \sum_{t_j \in T} \mathrm{Im}(\alpha_j) K''(t-t_j) + \sum_{t_j \in T} \mathrm{Im}(\beta_j) K'''(t-t_j) \right| \\
&\leq \mathrm{Im}(\alpha_1) |K''(t)| + ||\alpha||_\infty \sum_{t_j \in T \setminus \{0\}} |K''(t-t_j)| + ||\beta||_\infty \sum_{t_j \in T} |K'''(t-t_j)| \\
&\leq \mathrm{Im}(\alpha_1) |K''(t)| + \alpha^\infty F_2(\Delta_{\min}, t) + \beta^\infty (|K'''(t)| + F_3(\Delta_{\min}, t)) \\
&\leq 1.893\, f_{\mathrm{lo}}^2
\end{aligned}$$

and

$$\begin{aligned}
|q'(t)| &= \left| \sum_{t_j \in T} \alpha_j K'(t-t_j) + \beta_j K''(t-t_j) \right| \\
&\leq ||\alpha||_\infty \sum_{t_j \in T} |K'(t-t_j)| + ||\beta||_\infty \sum_{t_j \in T} |K''(t-t_j)| \\
&\leq \alpha^\infty |K'(t)| + \alpha^\infty F_1(\Delta_{\min}, t) + \beta^\infty (|K''(t)| + F_2(\Delta_{\min}, t)) \\
&\leq 0.7882\, f_{\mathrm{lo}}.
\end{aligned}$$

These bounds allow us to conclude that $|q|''$ is negative on $[0, 0.1649\lambda_{\mathrm{lo}}]$ since

$$q_R(t)\, q_R''(t) + |q'(t)|^2 + |q_I(t)|\, |q_I''(t)| \leq -9.291\ 10^{-2} f_{\mathrm{lo}}^2 < 0.$$

This completes the proof.

5.3.3 Proof of Lemma 5.2.4

As before, we assume without loss of generality that $\tau = 0$ and $q(0) = 1$. We use Lemma 5.3.2 again to bound the absolute value of the dual polynomial on

$[0.1649\lambda_{\text{lo}}, \Delta/2]$ and write

$$\begin{aligned}
|q(t)| &= \Big|\sum_{t_j \in T} \alpha_j K(t-t_j) + \beta_j K'(t-t_j)\Big| \\
&\leq ||\alpha||_\infty \Big[|K(t)| + \sum_{t_j \in T\setminus\{0\}} |K(t-t_j)|\Big] + ||\beta||_\infty \Big[|K'(t)| + \sum_{t_j \in T\setminus\{0\}} |K'(t-t_j)|\Big] \\
&\leq \alpha^\infty |K(t)| + \alpha^\infty F_0(\Delta_{\min}, t) + \beta^\infty |K'(t)| + \beta^\infty F_1(\Delta_{\min}, t). \qquad (5.3.19)
\end{aligned}$$

Note that we are assuming adversarial sign patterns and as a result we are unable to exploit cancellations in the coefficient vectors α and β. To control $|K(t)|$ and $|K'(t)|$ between $0.1649\,\lambda_{\text{lo}}$ and $0.7559\,\lambda_{\text{lo}}$, we use series expansions around the origin which give

$$\begin{aligned}
K(t) &\leq 1 - \frac{\pi^2 f_{\text{lo}}(f_{\text{lo}}+4)t^2}{6} + \frac{\pi^4 (f_{\text{lo}}+2)^4 t^4}{72} \\
|K'(t)| &\leq \frac{\pi^2 f_{\text{lo}}(f_{\text{lo}}+4)t}{3},
\end{aligned} \qquad (5.3.20)$$

for all $t \in [-1/2, 1/2]$. Put

$$L_1(t) = \alpha^\infty\Big[1 - \frac{\pi^2 f_{\text{lo}}(f_{\text{lo}}+4)t^2}{6} + \frac{\pi^4 (f_{\text{lo}}+2)^4 t^4}{72}\Big] + \beta^\infty \frac{\pi^2 f_{\text{lo}}(f_{\text{lo}}+4)t}{3},$$

with derivative equal to

$$L_1'(t) = -\alpha^\infty\Big[\frac{\pi^2 f_{\text{lo}}(f_{\text{lo}}+4)t}{3} - \frac{\pi^4 (f_{\text{lo}}+2)^4 t^3}{18}\Big] + \beta^\infty \frac{\pi^2 f_{\text{lo}}(f_{\text{lo}}+4)}{3}.$$

This derivative is strictly negative between $0.1649\,\lambda_{\text{lo}}$ and $0.7559\,\lambda_{\text{lo}}$, which implies that $L_1(t)$ is decreasing in this interval. Put

$$L_2(t) = \alpha^\infty F_0(\Delta_{\min}, t) + \beta^\infty F_1(\Delta_{\min}, t).$$

By Lemma 5.3.2, this function is increasing. With (5.3.19), this gives the crude bound

$$|q(t)| \leq L_1(t) + L_2(t) \leq L_1(t_1) + L_2(t_2) \quad \text{for all } t \in [t_1, t_2]. \qquad (5.3.21)$$

$t_1/\lambda_{\mathrm{lo}}$	$t_2/\lambda_{\mathrm{lo}}$	$L_1(t_1)$	$L_2(t_2)$
0.1649	0.4269	0.9818	$1.812\,10^{-2}$
0.4269	0.7559	0.7929	0.2068

Table 5.2: Numerical quantities used in (5.3.21).

Table 5.2 shows that taking $\{t_1, t_2\} = \{0.1649\,\lambda_{\mathrm{lo}}, 0.4269\,\lambda_{\mathrm{lo}}\}$ and then $\{t_1, t_2\} = \{0.4269\,\lambda_{\mathrm{lo}}, 0.7559\,\lambda_{\mathrm{lo}}\}$ proves that $|q(t)| < 1$ on $[0.1649\lambda_{\mathrm{lo}}, 0.7559\lambda_{\mathrm{lo}}]$. For $0.7559\lambda_{\mathrm{lo}} \leq t \leq \Delta/2$, we apply Lemma 5.3.1 and obtain

$$
\begin{aligned}
|q(t)| &\leq \alpha^{\infty}\Big[B_0(t) + B_0(\Delta - t) + \sum_{j=1}^{\infty} B_0(j\Delta_{\min} + \Delta - t) + \sum_{j=1}^{\infty} B_0(j\Delta_{\min} + t)\Big] \\
&\quad + \beta^{\infty}\Big[B_1(t) + B_1(\Delta - t) + \sum_{j=1}^{\infty} B_1(j\Delta_{\min} + \Delta - t) + \sum_{j=1}^{\infty} B_1(j\Delta_{\min} + t)\Big] \\
&\leq \alpha^{\infty}\Big[B_0(0.7559\lambda_{\mathrm{lo}}) + \sum_{j=1}^{\infty} B_0(j\Delta_{\min} - 0.7559\lambda_{\mathrm{lo}}) + \sum_{j=1}^{\infty} B_0(j\Delta_{\min} + 0.7559\lambda_{\mathrm{lo}})\Big] \\
&\quad + \beta^{\infty}\Big[B_1(0.7559\lambda_{\mathrm{lo}}) + \sum_{j=1}^{\infty} B_1(j\Delta_{\min} - 0.7559\lambda_{\mathrm{lo}}) + \sum_{j=1}^{\infty} B_1(j\Delta_{\min} + 0.7559\lambda_{\mathrm{lo}})\Big] \\
&\leq 0.758;
\end{aligned}
$$

here, the second step follows from the monotonicity of B_0 and B_1. Finally, for $\Delta/2 \leq t \leq t_+/2$, this last inequality applies as well. This completes the proof.

5.3.4 Proof of Lemma 5.2.5

Replacing $\Delta = 1.98\,\lambda_{\mathrm{lo}}$ by $\Delta = 2.5\,\lambda_{\mathrm{lo}}$ and going through exactly the same calculations as in Sections 5.3.1 and 5.3.2 yields that for any t obeying $0 \leq |t - \tau| \leq 0.1649\,\lambda_{\mathrm{lo}}$,

$$
\frac{\mathrm{d}^2\,|q|}{\mathrm{d}t^2}(t) \leq -0.6706\,f_{\mathrm{lo}}^2.
$$

t/λ_{lo}	$F_0\,(2.5\lambda_{\text{lo}}, t)$	$F_1\,(2.5\lambda_{\text{lo}}, t)$	$F_2\,(2.5\lambda_{\text{lo}}, t)$	$F_3\,(2.5\lambda_{\text{lo}}, t)$
0	$5.175\ 10^{-3}$	$6.839\ 10^{-2} f_{\text{lo}}$	$0.8946\, f_{\text{lo}}^2$	$7.644\, f_{\text{lo}}^3$
0.1649	$5.182\ 10^{-3}$	$6.849\ 10^{-2} f_{\text{lo}}$	$0.9459\, f_{\text{lo}}^2$	$7.647\, f_{\text{lo}}^3$

Table 5.3: Numerical upper bounds on $F_\ell(2.5\lambda_{\text{lo}}, t)$.

t/λ_{lo}	$F_0\,(1.87\lambda_{\text{lo}}, t)$	$F_1\,(1.87\lambda_{\text{lo}}, t)$	$F_2\,(1.87\lambda_{\text{lo}}, t)$	$F_3\,(1.87\lambda_{\text{lo}}, t)$
0	$6.708\ 10^{-3}$	$7.978\ 10^{-2} f_{\text{lo}}$	$1.078\, f_{\text{lo}}^2$	$16.01\, f_{\text{lo}}^3$
0.17	$6.747\ 10^{-3}$	$0.1053 f_{\text{lo}}$	$1.081\, f_{\text{lo}}^2$	$41.74\, f_{\text{lo}}^3$

Table 5.4: Numerical upper bounds on $F_\ell(1.87\lambda_{\text{lo}}, t)$.

For reference, we have computed in Table 5.3 numerical upper bounds on $F_\ell(2.5\lambda_{\text{lo}}, t)$ at $t = \{0, 0.1649\,\lambda\}$. Since $|q(0)| = 1$ and $q'(0) = 0$, it follows that

$$|q\,(t)| \leq |q\,(0)| - \frac{1}{2}\,0.6706\, f_{\text{lo}}^2 t^2. \tag{5.3.22}$$

At a distance of $0.1649\,\lambda_{\text{lo}}$, the right-hand side is equal to 0.9909. The calculations in Section 5.3.3 with $\Delta = 2.5\,\lambda_{\text{lo}}$ imply that the magnitude of $q(t)$ at locations at least $0.1649\,\lambda_{\text{lo}}$ away from an element of T is bounded by 0.9843. This concludes the proof.

5.3.5 Improvement for real-valued signals

The proof for real-valued signals is almost the same as the one we have discussed for complex-valued signals—only simpler. The only modification to Lemmas 5.2.2 and 5.2.4 is that the minimum distance is reduced to $1.87\,\lambda_{\text{lo}}$, and that the bound in Lemma 5.2.4 is shown to hold starting at $0.17\,\lambda_{\text{lo}}$ instead of $0.1649\,\lambda_{\text{lo}}$. For reference, we provide upper bounds on $F_\ell(1.87\lambda_{\text{lo}}, t)$ at $t \in \{0, 0.17\,\lambda_{\text{lo}}\}$ in Table 5.4. As to Lemma 5.2.3, the only difference is that to bound $|q|$ between the origin and $0.17\,\lambda_{\text{lo}}$, it is sufficient to show that the second derivative of q is negative and make sure that $q > -1$. Computing (5.3.18) for $\Delta = 1.87\lambda_{\text{lo}}$ for $t \in [0, 0.17\,\lambda_{\text{lo}}]$, we obtain $q'' < -0.1181$. Finally, (5.3.17) yields $q > 0.9113$ in $[0, 0.17\,\lambda_{\text{lo}}]$.

5.3.6 Proof of Lemma 5.3.1

The first inequality in the lemma holds due to two lower bounds on the sine function:

$$|\sin(\pi t)| \geq 2|t|, \text{ for all } t \in [-1/2, 1/2] \tag{5.3.23}$$

$$\sin(\pi t) \geq \pi t - \frac{\pi^3 t^3}{6} = \frac{2t}{a(t)}, \text{ for all } t \geq 0. \tag{5.3.24}$$

The proof for these expressions, which we omit, is based on concavity of the sine function and on a Taylor expansion around the origin. Put $f = f_{\text{lo}}/2 + 1$ for short. Some simple calculations give $K'(0) = 0$ and for $t \neq 0$,

$$K'(t) = 4\pi \left(\frac{\sin(f\pi t)}{f\sin(\pi t)}\right)^3 \left(\frac{\cos(f\pi t)}{\sin(\pi t)} - \frac{\sin(f\pi t)\cos(\pi t)}{f\sin^2(\pi t)}\right). \tag{5.3.25}$$

Further calculations show that the value of the second derivative of K at the origin is $-\pi^2 f_{\text{lo}}(f_{\text{lo}} + 4)/3$, and for $t \neq 0$,

$$K''(t) = \frac{4\pi^2 \sin^2(f\pi t)}{f^2 \sin^4(\pi t)} \left[3\left(\cos(f\pi t) - \frac{\sin(f\pi t)\cos(\pi t)}{f\sin(\pi t)}\right)^2 - \sin^2(f\pi t) - \frac{\sin(2f\pi t)}{f\tan(\pi t)} + \frac{\sin^2(f\pi t)}{f^2\tan^2(\pi t)} + \frac{\sin^2(f\pi t)}{f^2\sin^2(\pi t)}\right]. \tag{5.3.26}$$

It is also possible to check that the third derivative of K is zero at the origin, and for $t \neq 0$,

$$K'''(t) = \frac{4\pi^3 \sin(f\pi t)}{f\sin^4(\pi t)} \left(6G_1(t) + 9\sin(f\pi t)\, G_2(t) + \sin^2(f\pi t)\, G_3(t)\right) \tag{5.3.27}$$

with

$$G_1(t) = \left(\cos(f\pi t) - \frac{\sin(f\pi t)\cos(\pi t)}{f\sin(\pi t)}\right)^3,$$

$$
\begin{aligned}
G_2(t) &= \left(\cos\left(f\pi t\right) - \frac{\sin\left(f\pi t\right)\cos\left(\pi t\right)}{f\sin\left(\pi t\right)}\right) \\
&\quad \left(-\sin\left(f\pi t\right) - \frac{2\cos\left(f\pi t\right)}{f\tan\left(\pi t\right)} + \frac{\sin\left(f\pi t\right)}{f^2\tan^2\left(\pi t\right)} + \frac{\sin\left(f\pi t\right)}{f^2\sin^2\left(\pi t\right)}\right), \\
G_3(t) &= \frac{3\cos\left(f\pi t\right)\left(1+\cos^2\left(\pi t\right)\right)}{f^2\sin^2\left(\pi t\right)} - \cos\left(f\pi t\right) + \frac{3\sin\left(f\pi t\right)}{f\tan\left(\pi t\right)} \\
&\quad - \frac{\sin\left(f\pi t\right)\left(1+5\cos\left(\pi t\right)\right)}{f^3\sin^3\left(\pi t\right)}.
\end{aligned}
$$

The remaining inequalities in the lemma are all almost direct consequences of plugging (5.3.23) and (5.3.24) into (5.3.25), (5.3.26) and (5.3.27). The bounds are nonincreasing in t because the derivative of $b(t)$ is negative between zero and $\sqrt{2}/\pi$ and one can check that $H_\ell(\sqrt{2}/\pi) < H_\ell^\infty$ for $f_{\text{lo}} \geq 128$. Additionally, $b^k(t)$ is strictly convex for positive t and $k \in \{1, 2, 3\}$, so the derivative with respect to τ of $b^k(\Delta - \tau) + b^k(\Delta + \tau)$ is positive for $0 \leq \tau < \Delta/2$, which implies that $\tilde{B}_\ell(\Delta - \tau) + \tilde{B}_\ell(\Delta + \tau)$ is increasing in τ.

Chapter 6

Proof of exact recovery in 2D

This chapter provides the proof of Theorem 2.3.1, which establishes exact recovery in two dimensions. Section 6.1 describes the main argument of the proof, using several lemmas that are proved in Section 6.2. Throughout the chapter we write $\Delta = \Delta(T) \geq \Delta_{\min} = 2.38\,\lambda_{\text{lo}}$. Unless specified otherwise, $|r - r'|$ is the ∞ distance.

6.1 Outline of proof

Note that Proposition 5.1.1 also holds in multiple dimensions. It states that if there exists a low-pass polynomial that interpolates the sign pattern of a signal and has a magnitude strictly smaller than one on the off-support, then total-variation norm minimization achieves exact recovery. As a result, Theorem 2.3.1 follows from Proposition 6.1.1 below, which guarantees the existence of such a dual certificate.

Proposition 6.1.1. *Let $T = \{r_1, r_2, \ldots\} \subset \mathbb{T}^2$ be any family of points obeying the minimum distance condition*

$$|r_j - r_k| \geq 2.38\,\lambda_{lo}, \quad r_j \neq r_k \in T.$$

Assume $f_{lo} \geq 512$. Then for any vector $v \in \mathbb{R}^{|T|}$ with $|v_j| = 1$, there exists a trigonometric polynomial q with Fourier series coefficients supported on $\{-f_c, -f_{lo} +$

$1, \ldots, f_c\}^2$ *with the property*

$$\begin{cases} q(r_j) = v_j, & t_j \in T, \\ |q(r)| < 1, & t \in \mathbb{T}^2 \setminus T. \end{cases} \tag{6.1.1}$$

The proof is similar to that of Proposition 5.2.1 in that we shall construct the dual polynomial q by interpolation with a low-pass, yet rapidly decaying two-dimensional kernel. Here, we consider

$$K^{\mathrm{2D}}(r) = K(x)\, K(y),$$

obtained by tensorizing the square of the Fejer kernel (5.2.1). (For reference, if we had data in which $y(k)$ is observed if $\|k\|_2 \leq f_c$, we would probably use a radial kernel.) Just as before, we have fixed K somewhat arbitrarily, and it would probably be possible to optimize this choice to improve the constant factor in the expression for the minimum distance. We interpolate the sign pattern using K^{2D} and its partial derivatives, denoted by $K^{\mathrm{2D}}_{(1,0)}$ and $K^{\mathrm{2D}}_{(0,1)}$ respectively, as follows:

$$q(r) = \sum_{r_j \in T} \left[\alpha_j K^{\mathrm{2D}}(r - r_j) + \beta_{1j} K^{\mathrm{2D}}_{(1,0)}(r - r_j) + \beta_{2j} K^{\mathrm{2D}}_{(0,1)}(r - r_j) \right],$$

and we fit the coefficients so that for all $t_j \in T$,

$$\begin{aligned} q(t_j) &= v_j, \\ \nabla q(t_j) &= 0. \end{aligned} \tag{6.1.2}$$

The first intermediate result shows that the dual polynomial is well defined, and also controls the magnitude of the interpolation coefficients.

Lemma 6.1.2. *Under the hypotheses of Proposition 6.1.1, there are vectors α, β_1 and β_2 obeying* (6.1.2) *and*

$$\begin{aligned} ||\alpha||_\infty &\leq 1 + 5.577\ 10^{-2}, \\ ||\beta||_\infty &\leq 2.930\ 10^{-2}\, \lambda_{lo}, \end{aligned} \tag{6.1.3}$$

where $\beta = (\beta_1, \beta_2)$. *Further, if* $v_1 = 1$,

$$\alpha_1 \geq 1 - 5.577\, 10^{-2}. \tag{6.1.4}$$

Proposition 6.1.1 is now a consequence of the two lemmas below which control the size of q near a point $r_0 \in T$. Without loss of generality, we can take $r_0 = 0$.

Lemma 6.1.3. *Assume* $0 \in T$. *Then under the hypotheses of Proposition 6.1.1,* $|q(r)| < 1$ *for all* $0 < |r| \leq 0.2447\, \lambda_{lo}$.

Lemma 6.1.4. *Assume* $0 \in T$. *Then under the conditions of Proposition 6.1.1,* $|q(r)| < 1$ *for all* r *obeying* $0.2447\, \lambda_{lo} \leq |r| \leq \Delta/2$. *This also holds for all* r *that are closer to* $0 \in T$ *(in the* ∞ *distance) than to any other element in* T.

6.2 Proof of lemmas

6.2.1 Proof of Lemma 6.1.2

To express the interpolation constraints in matrix form, define

$$\begin{aligned}
(D_0)_{jk} &= K^{\mathrm{2D}}(r_j - r_k), & \left(D_{(1,0)}\right)_{jk} &= K^{\mathrm{2D}}_{(1,0)}(r_j - r_k), \\
\left(D_{(0,1)}\right)_{jk} &= K^{\mathrm{2D}}_{(0,1)}(r_j - r_k), & \left(D_{(1,1)}\right)_{jk} &= K^{\mathrm{2D}}_{(1,1)}(r_j - r_k), \\
\left(D_{(2,0)}\right)_{jk} &= K^{\mathrm{2D}}_{(2,0)}(r_j - r_k), & \left(D_{(0,2)}\right)_{jk} &= K^{\mathrm{2D}}_{(0,2)}(r_j - r_k).
\end{aligned}$$

To be clear, $K^{\mathrm{2D}}_{(\ell_1,\ell_2)}$ means that we are taking ℓ_1 and ℓ_2 derivatives with respect to the first and second variables. Note that D_0, $D_{(2,0)}$, $D_{(1,1)}$ and $D_{(0,2)}$ are symmetric, while $D_{(1,0)}$ and $D_{(0,1)}$ are antisymmetric, because K and K'' are even while K' is odd. The interpolation coefficients are solutions to

$$\begin{bmatrix} D_0 & D_{(1,0)} & D_{(0,1)} \\ D_{(1,0)} & D_{(2,0)} & D_{(1,1)} \\ D_{(0,1)} & D_{(1,1)} & D_{(0,2)} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} v \\ 0 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix} D_0 & -\tilde{D}_1^T \\ \tilde{D}_1 & \tilde{D}_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix}, \tag{6.2.1}$$

where we have defined two new matrices $\tilde{D}_1$ and $\tilde{D}_2$. The norm of these matrices can be bounded by leveraging 1D results. For instance, consider

$$||I - D_0||_\infty = \sum_{r_j \in T \setminus \{0\}} \left| K^{\text{2D}}(r_j) \right|.$$

We split this sum into different regions corresponding to whether $|x_j|$ or $|y_j| \leq \Delta/2$ and to $\min(|x_j|, |y_j|) \geq \Delta/2$. First,

$$\sum_{r_j \neq 0 \,:\, |y_j| < \Delta/2} \left| K^{\text{2D}}(r_j) \right| \leq \sum_{r_j \neq 0 \,:\, |y_j| < \Delta/2} B_0(x_j) \leq 2 \sum_{j \geq 1} B_0(j\Delta).$$

This holds because the x_j's must be at least Δ apart, B_0 is nonincreasing and the absolute value of K^{2D} is bounded by one. The region $\{r_j \neq 0, |x_j| < \Delta/2\}$ yields the same bound. Now observe that Lemma 6.2.1 below combined with Lemma 5.3.1 gives

$$\sum_{r_j \neq 0 \,:\, \min(x_j, y_j) \geq \Delta/2} \left| K^{\text{2D}}(r_j) \right| \leq \sum_{r_j \neq 0 \,:\, \min(x_j, y_j) \geq \Delta/2} B_0(x_j) B_0(y_j) \leq \Big[\sum_{j_1 \geq 0} B_0(\Delta/2 + j_1 \Delta) \Big]^2.$$

To bound this expression, we apply the exact same technique as for (5.3.2) in Section 5.3, starting at $j = 0$ and setting $j_0 = 20$. This gives

$$||I - D_0||_\infty \leq 4 \sum_{j \geq 1} B_0(j\Delta) + 4 \Big[\sum_{j \geq 0} B_0(\Delta/2 + j\Delta) \Big]^2 \leq 4.854 \; 10^{-2}. \tag{6.2.2}$$

Lemma 6.2.1. *Suppose $x \in \mathbb{R}^2_+$ and $f(x) = f_1(x_1) f_2(x_2)$ where both f_1 and f_2 are nonincreasing. Consider any collection of points $\{x_j\} \subset \mathbb{R}^2_+$ for which $|x_i - x_j| \geq 1$. Then*

$$\sum_j f(x_j) \leq \sum_{j_1 \geq 0} f_1(j_1) \sum_{j_2 \geq 0} f_2(j_2).$$

Proof. Consider the mapping $x \in \mathbb{R}^2_+ \mapsto (\lfloor x_1 \rfloor, \lfloor x_2 \rfloor)$. This mapping is injective over our family of points. (Indeed, two points cannot be mapped to the same pair of integers (j_1, j_2) as it would otherwise imply that they are both inside the square $[j_1 + 1) \times [j_2 + 1)$, hence violating the separation condition.) Therefore, the monotonicity

assumption gives

$$\sum_j f(x_j) \leq \sum_j f_1(\lfloor x_{j,1} \rfloor) f_2(\lfloor x_{j,2} \rfloor) \leq \sum_{j_1, j_2 \geq 0} f_1(j_1) f_2(j_2),$$

which proves the claim. □

Applying the same reasoning, we obtain

$$\begin{aligned} \left|\left|D_{(1,0)}\right|\right|_\infty \leq 2 \sum_{j \geq 1} B_1(j\Delta) + 2\|K'\|_\infty \sum_{j \geq 1} B_0(j\Delta) \\ + 4\Big[\sum_{j \geq 0} B_0(\Delta/2 + j\Delta)\Big]\Big[\sum_{j \geq 0} B_1(\Delta/2 + j\Delta)\Big]. \end{aligned}$$

In turn, the same upper-bounding technique yields

$$\left|\left|D_{(1,0)}\right|\right|_\infty \leq 7.723\ 10^{-2}\, f_{\text{lo}}, \tag{6.2.3}$$

where we have used the fact that $\|K'\|_\infty \leq 2.08\,(f_{\text{lo}} + 2)$, which follows from combining Lemma 5.3.1 with (5.3.12). Likewise,

$$\left|\left|D_{(1,1)}\right|\right|_\infty \leq 4\|K'\|_\infty \sum_{j \geq 1} B_1(j\Delta) + 4\Big[\sum_{j \geq 0} B_1(\Delta/2 + j\Delta)\Big]^2 \leq 0.1576\, f_{\text{lo}}^2, \tag{6.2.4}$$

and finally,

$$\begin{aligned} \left|\left|\left|K^{\text{2D}}_{(2,0)}\,(0)\right|\, I - D_{(2,0)}\right|\right|_\infty \leq 2 \sum_{j \geq 1} B_2(j\Delta) + 2\|K''\|_\infty \sum_{j \geq 1} B_0(j\Delta) \\ + 4\Big[\sum_{j \geq 0} B_0(\Delta/2 + j\Delta)\Big]\Big[\sum_{j \geq 0} B_2(\Delta/2 + j\Delta)\Big] \leq 0.3539\, f_{\text{lo}}^2, \end{aligned} \tag{6.2.5}$$

since $\|K''\|_\infty = \pi^2 f_{\text{lo}}\,(f_{\text{lo}} + 4)\,/3$, as $|K''|$ reaches its global maximum at the origin.

We use these estimates to show that the system (6.2.1) is invertible and the

coefficient sequences are bounded. To ease notation, set

$$\begin{aligned} S_1 &= D_{(2,0)} - D_{(1,1)} D_{(0,2)}^{-1} D_{(1,1)}, \\ S_2 &= D_{(1,0)} - D_{(1,1)} D_{(0,2)}^{-1} D_{(0,1)}, \\ S_3 &= D_0 + S_2^T S_1^{-1} S_2 - D_{(0,1)} D_{(0,2)}^{-1} D_{(0,1)}. \end{aligned}$$

Note that S_1 is a Schur complement of $D_{(0,2)}$ and that a standard linear algebra identity gives

$$\tilde{D}_2^{-1} = \begin{bmatrix} S_1^{-1} & -S_1^{-1} D_{(1,1)} D_{(0,2)}^{-1} \\ -D_{(0,2)}^{-1} D_{(1,1)} S_1^{-1} & D_{(0,2)}^{-1} + D_{(0,2)}^{-1} D_{(1,1)} S_1^{-1} D_{(1,1)} D_{(0,2)}^{-1} \end{bmatrix}.$$

Using this and taking the Schur complement of $\tilde{D}_2$, which is equal to S_3, the solution to (6.2.1) can be written as

$$\begin{aligned} &\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} I \\ -\tilde{D}_2^{-1} \tilde{D}_1 \end{bmatrix} \left(D_0 + \tilde{D}_1^T \tilde{D}_2^{-1} \tilde{D}_1 \right)^{-1} v \\ \Leftrightarrow \quad &\begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} I \\ -S_1^{-1} S_2 \\ D_{(0,2)}^{-1} \left(D_{(1,1)} S_1^{-1} S_2 - D_{(0,1)} \right) \end{bmatrix} S_3^{-1} v. \end{aligned}$$

Applying (5.3.4) from Section 5.3.1, we obtain

$$\left|\left| D_{(0,2)}^{-1} \right|\right|_\infty \leq \frac{1}{\left| K_{(0,2)}^{\mathrm{2D}}(0) \right| - \left|\left| \left| K_{(0,2)}^{\mathrm{2D}}(0) \right| I - D_{(0,2)} \right|\right|_\infty} \leq \frac{0.3399}{f_{\mathrm{lo}}^2}, \tag{6.2.6}$$

which together with $K_{(2,0)}^{\mathrm{2D}}(0) = -\pi^2 f_{\mathrm{lo}} (f_{\mathrm{lo}} + 4)/3$ and (6.2.4) imply

$$\begin{aligned} \left|\left| \left| K_{(2,0)}^{\mathrm{2D}}(0) \right| I - S_1 \right|\right|_\infty &\leq \left|\left| \left| K_{(2,0)}^{\mathrm{2D}}(0) \right| I - D_{(2,0)} \right|\right|_\infty + \left|\left| D_{(1,1)} \right|\right|_\infty^2 \left|\left| D_{(0,2)}^{-1} \right|\right|_\infty \\ &\leq 0.33624\, f_{\mathrm{lo}}^2. \end{aligned}$$

Another application of (5.3.4) then yields

$$\left|\left|S_1^{-1}\right|\right|_\infty \leq \frac{1}{\left|K_{(2,0)}^{\mathrm{2D}}(0)\right| - \left|\left|\left|K_{(2,0)}^{\mathrm{2D}}(0)\right| I - S_1\right|\right|_\infty} \leq \frac{0.3408}{f_{\mathrm{lo}}^2}. \tag{6.2.7}$$

Next, (6.2.3), (6.2.4) and (6.2.6) allow to bound S_2,

$$||S_2||_\infty \leq \left|\left|D_{(1,0)}\right|\right|_\infty + \left|\left|D_{(1,1)}\right|\right|_\infty \left|\left|D_{(0,2)}^{-1}\right|\right|_\infty \left|\left|D_{(0,1)}\right|\right|_\infty \leq 8.142\ 10^{-2}\, f_{\mathrm{lo}},$$

which combined with (6.2.2), (6.2.3), (6.2.6) and (6.2.7) implies

$$||I - S_3||_\infty \leq ||I - D_0||_\infty + ||S_2||_\infty^2 \left|\left|S_1^{-1}\right|\right|_\infty + \left|\left|D_{(0,1)}\right|\right|_\infty^2 \left|\left|D_{(0,2)}^{-1}\right|\right|_\infty \leq 5.283\ 10^{-2}.$$

The results above allow us to derive bounds on the coefficient vectors by applying (5.3.4) one last time, establishing

$$\begin{aligned}
||\alpha||_\infty &\leq \left|\left|S_3^{-1}\right|\right|_\infty \leq 1 + 5.577\ 10^{-2}, \\
||\beta_1||_\infty &\leq \left|\left|S_1^{-1} S_2 S_3^{-1}\right|\right|_\infty \leq \left|\left|S_1^{-1}\right|\right|_\infty ||S_2||_\infty \left|\left|S_3^{-1}\right|\right|_\infty \leq 2.930\ 10^{-2}\, \lambda_{\mathrm{lo}}, \\
\alpha_1 &= v_1 - \left(\left(I - S_3^{-1}\right) v\right)_1 \geq 1 - \left|\left|S_3^{-1}\right|\right|_\infty ||I - S_3||_\infty \geq 1 - 5.577\ 10^{-2},
\end{aligned}$$

where the last lower bound holds if $v_1 = 1$. The derivation for $||\beta_2||_\infty$ is identical and we omit it.

6.2.2 Proof of Lemma 6.1.3

Since v is real valued, α, β and q are all real valued. For $|r| \leq 0.2447\, \lambda_{\mathrm{lo}}$, we show that the Hessian matrix of q,

$$H = \begin{bmatrix} q_{(2,0)}(r) & q_{(1,1)}(r) \\ q_{(1,1)}(r) & q_{(0,2)}(r) \end{bmatrix}$$

is negative definite. In what follows, it will also be useful to establish bounds on the kernel and its derivatives near the origin. Using (5.3.11)–(5.3.15), we obtain

$$K^{\mathrm{2D}}(x,y) \geq \left(1-\frac{\pi^2 f_{\mathrm{lo}}(f_{\mathrm{lo}}+4)\,x^2}{6}\right)\left(1-\frac{\pi^2 f_{\mathrm{lo}}(f_{\mathrm{lo}}+4)\,y^2}{6}\right)$$
$$K^{\mathrm{2D}}_{(2,0)}(x,y) \leq \left(-\frac{\pi^2 f_{\mathrm{lo}}(f_{\mathrm{lo}}+4)}{3}+\frac{(f_{\mathrm{lo}}+2)^4\,\pi^4 x^2}{6}\right)\left(1-\frac{\pi^2 f_{\mathrm{lo}}(f_{\mathrm{lo}}+4)\,y^2}{6}\right)$$

and

$$\left|K^{\mathrm{2D}}_{(1,0)}(x,y)\right| \leq \frac{\pi^2 f_{\mathrm{lo}}(f_{\mathrm{lo}}+4)\,x}{3}, \qquad \left|K^{\mathrm{2D}}_{(1,1)}(x,y)\right| \leq \frac{\pi^4 f_{\mathrm{lo}}^2(f_{\mathrm{lo}}+4)^2\,xy}{9},$$
$$\left|K^{\mathrm{2D}}_{(2,1)}(x,y)\right| \leq \frac{\pi^4 f_{\mathrm{lo}}^2(f_{\mathrm{lo}}+4)^2\,y}{9}, \qquad \left|K^{\mathrm{2D}}_{(3,0)}(x,y)\right| \leq \frac{\pi^4(f_{\mathrm{lo}}+2)^4\,x}{3}.$$

These bounds are all monotone in x and y so we can evaluate them at $x = 0.2447\,\lambda_{\mathrm{lo}}$ and $y = 0.2447\,\lambda_{\mathrm{lo}}$ to show that for any $|r| \leq 0.2447\,\lambda_{\mathrm{lo}}$,

$$\begin{aligned} &K^{\mathrm{2D}}(r) \geq 0.8113 && \left|K^{\mathrm{2D}}_{(1,0)}(r)\right| \leq 0.8113 && K^{\mathrm{2D}}_{(2,0)}(r) \leq -2.097\,f_{\mathrm{lo}}^2, \\ &\left|K^{\mathrm{2D}}_{(1,1)}(r)\right| \leq 0.6531\,f_{\mathrm{lo}}, && \left|K^{\mathrm{2D}}_{(2,1)}(r)\right| \leq 2.669\,f_{\mathrm{lo}}^2, && \left|K^{\mathrm{2D}}_{(3,0)}(r)\right| \leq 8.070\,f_{\mathrm{lo}}^3. \end{aligned} \tag{6.2.8}$$

The bounds for $K^{\mathrm{2D}}_{(1,0)}$, $K^{\mathrm{2D}}_{(2,0)}$, $K^{\mathrm{2D}}_{(2,1)}$ and $K^{\mathrm{2D}}_{(3,0)}$ of course also hold for $K^{\mathrm{2D}}_{(0,1)}$, $K^{\mathrm{2D}}_{(0,2)}$, $K^{\mathrm{2D}}_{(1,2)}$ and $K^{\mathrm{2D}}_{(0,3)}$. Additionally, it will be necessary to bound sums of the form $\sum_{r_j \in T/\{0\}} \left|K^{\mathrm{2D}}_{(\ell_1,\ell_2)}(r-r_j)\right|$ for r such that $|r| \leq \Delta/2$ and $\ell_1, \ell_2 = 0,1,2,3$. Consider the case $(\ell_1,\ell_2) = (0,0)$. Without loss of generality, let $r = (x,y) \in \mathbb{R}^2_+$. By Lemma 6.2.1, the contribution of those r_j's belonging to the three quadrants $\{|r| > \Delta/2\} \setminus \mathbb{R}^2_+$ obeys

$$\sum_{|r_j|>\Delta/2, r_j \notin \mathbb{R}^2_+} \left|K^{\mathrm{2D}}(r-r_j)\right| \leq 3\Big[\sum_{j\geq 0} B_0(\Delta/2 + j\Delta)\Big]^2.$$

Similarly, the contribution from the bands where either $|r_{j,1}|$ or $|r_{j,2}| \leq \Delta/2$ obeys

$$\sum_{|r_{j,1}|\leq\Delta/2 \text{ or } |r_{j,2}|\leq\Delta/2} \left|K^{\mathrm{2D}}(r-r_j)\right| \leq 2\sum_{j\geq 1} B_0(j\Delta - |r|).$$

It remains to bound the sum over r_j's lying in the positive quadrant $\{|r| > \Delta/2\} \cap \mathbb{R}^2_+$. To do this, let $f_1(t)$ be equal to one if $|t| \leq \Delta$ and to $B_0(\Delta t - |r|)$ otherwise. Taking $f_2 = f_1$, Lemma 6.2.1 gives

$$\sum_{|r_j|>\Delta/2, r_j \in \mathbb{R}^2_+} \left|K^{\mathrm{2D}}(r - r_j)\right| \leq \sum_{j\geq 1} B_0(j\Delta - |r|) + \Big[\sum_{j\geq 1} B_0(j\Delta - |r|)\Big]^2.$$

We can apply exactly the same reasoning to the summation of $K^{\mathrm{2D}}_{(\ell_1,\ell_2)}$ for other values of ℓ_1 and ℓ_2, and obtain that for any r such that $|r| \leq \Delta/2$,

$$\sum_{r_j \in T\setminus\{0\}} \left|K^{\mathrm{2D}}_{(\ell_1,\ell_2)}(r - r_j)\right| \leq Z_{(\ell_1,\ell_2)}(|r|); \tag{6.2.9}$$

here, for $u \geq 0$,

$$\begin{aligned} Z_{(\ell_1,\ell_2)}(u) = {} & 2\sum_{j\geq 1} K^{(\ell_1)}_\infty B_{\ell_2}(j\Delta - u) + 2K^{(\ell_2)}_\infty B_{\ell_1}(j\Delta - u) + K^{(\ell_1)}_\infty \sum_{j\geq 1} B_{\ell_2}(j\Delta) \\ & + K^{(\ell_2)}_\infty \sum_{j\geq 1} B_{\ell_1}(j\Delta) + 3\Big[\sum_{j\geq 0} B_{\ell_1}(\Delta/2 + j\Delta)\Big]\Big[\sum_{j\geq 0} B_{\ell_2}(\Delta/2 + j\Delta - u)\Big] \\ & + \Big[\sum_{j\geq 1} B_{\ell_1}(j\Delta - u)\Big]\Big[\sum_{j\geq 1} B_{\ell_2}(j\Delta)\Big] \end{aligned}$$

in which $K^{(\ell_1)}_\infty$ is a bound on the global maximum of $K^{(\ell_1)}$. The absolute value of the kernel K and its second derivative reach their global maxima at the origin, so $K^{(0)}_\infty = 1$ and $K^{(2)}_\infty = \pi^2 f_{\mathrm{lo}}(f_{\mathrm{lo}} + 4)/3$. Combining the bounds on $|K'|$ and $|K'''|$ in Lemma 5.3.1 with (5.3.12) and (5.3.15), we can show that $K^{(1)}_\infty = 2.08(f_{\mathrm{lo}} + 2)$ and $K^{(3)}_\infty = 25.3(f_{\mathrm{lo}} + 2)^3$ if $f_{\mathrm{lo}} \geq 512$. Since $Z_{(\ell_1,\ell_2)} = Z_{(\ell_2,\ell_1)}$, we shall replace $Z_{(\ell_1,\ell_2)}$ for which $\ell_1 > \ell_2$ by $Z_{(\ell_2,\ell_1)}$.

Since

$$q_{(2,0)}(r) = \sum_{r_j \in T} \alpha_j K^{\mathrm{2D}}_{(2,0)}(r - r_j) + \beta_{1j} K^{\mathrm{2D}}_{(3,0)}(r - r_j) + \beta_{2j} K^{\mathrm{2D}}_{(2,1)}(r - r_j)$$

$Z_{(0,0)}(u)$	$Z_{(0,1)}(u)$	$Z_{(1,1)}(u)$	$Z_{(0,2)}(u)$	$Z_{(1,2)}(u)$	$Z_{(0,3)}(u)$
$6.405\ 10^{-2}$	$0.1047\, f_{\text{lo}}$	$0.1642\, f_{\text{lo}}^2$	$0.4019\, f_{\text{lo}}$	$0.6751\, f_{\text{lo}}^3$	$1.574 f_{\text{lo}}^3$

Table 6.1: Upper bounds on $Z_{(\ell_1,\ell_2)}(u)$ at $0.2447\,\lambda_{\text{lo}}$.

it follows from (6.2.8) and (6.2.9) that

$$\begin{aligned}
q_{(2,0)}(r) &\leq \alpha_1 K^{2\text{D}}_{(2,0)}(r) + ||\alpha||_\infty \sum_{r_j \in T\setminus\{0\}} \left|K^{2\text{D}}_{(2,0)}(r-r_j)\right| \\
&\quad + ||\beta||_\infty \Big[\left|K^{2\text{D}}_{(3,0)}(r)\right| + \sum_{r_j \in T\setminus\{0\}} \left|K^{2\text{D}}_{(3,0)}(r-r_j)\right| \\
&\quad + \left|K^{2\text{D}}_{(2,1)}(r)\right| + \sum_{r_j \in T\setminus\{0\}} \left|K^{2\text{D}}_{(2,1)}(r-r_j)\right|\Big] \\
&\leq \alpha_1 K^{2\text{D}}_{(2,0)}(r) + ||\alpha||_\infty Z_{(0,2)}(|r|) + ||\beta||_\infty \Big(\left|K^{2\text{D}}_{(3,0)}(r)\right| + Z_{(0,3)}(|r|) \\
&\quad + \left|K^{2\text{D}}_{(2,1)}(r)\right| + Z_{(1,2)}(|r|)\Big) \\
&\leq -1.175\, f_{\text{lo}}^2.
\end{aligned}$$

The last inequality uses values of $Z_{(\ell_1,\ell_2)}(u)$ at $u = 0.2447\,\lambda_{\text{lo}}$ reported in Table 6.1. By symmetry, the same bound holds for $q_{(0,2)}$. Finally, similar computations yield

$$\begin{aligned}
\left|q_{(1,1)}(r)\right| &= \sum_{r_j \in T} \alpha_j K^{2\text{D}}_{(1,1)}(r-r_j) + \beta_{1j} K^{2\text{D}}_{(2,1)}(r-r_j) + \beta_{2j} K^{2\text{D}}_{(1,2)}(r-r_j) \\
&\leq ||\alpha||_\infty \left[\left|K^{2\text{D}}_{(1,1)}(r)\right| + Z_{(1,1)}(|r|)\right] \\
&\quad + ||\beta||_\infty \left[\left|K^{2\text{D}}_{(2,1)}(r)\right| + \left|K^{2\text{D}}_{(1,2)}(r)\right| + 2Z_{(1,2)}(|r|)\right] \\
&\leq 1.059\, f_{\text{lo}}^2.
\end{aligned}$$

Since $\text{Tr}(H) = q_{(2,0)} + q_{(0,2)} < 0$ and $\det(H) = |q_{(2,0)}||q_{(0,2)}| - |q_{(1,1)}|^2 > 0$, both eigenvalues of H are strictly negative.

We have shown that q decreases along any segment originating at 0. To complete

the proof, we must establish that $q > -1$ in the square. Similar computations show

$$\begin{aligned} q(r) &= \sum_{r_j \in T} \alpha_j K^{\mathrm{2D}}(r - r_j) + \beta_{1j} K^{\mathrm{2D}}_{(1,0)}(r - r_j) + \beta_{2j} K^{\mathrm{2D}}_{(0,1)}(r - r_j) \\ &\geq \alpha_1 K^{\mathrm{2D}}(r) - ||\alpha||_\infty Z_{(0,0)}(|r|) - ||\beta||_\infty \left(\left| K^{\mathrm{2D}}_{(0,1)}(r) \right| + \left| K^{\mathrm{2D}}_{(1,0)}(r) \right| + 2Z_{(0,1)}(|r|) \right) \\ &\geq 0.6447. \end{aligned}$$

6.2.3 Proof of Lemma 6.1.4

For $0.2447\,\lambda_{\mathrm{lo}} \leq |r| \leq \Delta/2$,

$$\begin{aligned} |q| &\leq \left| \sum_{r_j \in T} \alpha_j K^{\mathrm{2D}}(r - r_j) + \beta_{1j} K^{\mathrm{2D}}_{(1,0)}(r - r_j) + \beta_{2j} K^{\mathrm{2D}}_{(0,1)}(r - r_j) \right| \\ &\leq ||\alpha||_\infty \left[\left| K^{\mathrm{2D}}(r) \right| + Z_{(0,0)}(|r|) \right] + ||\beta||_\infty \left[\left| K^{\mathrm{2D}}_{(1,0)}(r) \right| + \left| K^{\mathrm{2D}}_{(0,1)}(r) \right| + 2Z_{(0,1)}(|r|) \right]. \end{aligned}$$

Using the series expansion around the origin of K and K' (5.3.20), we obtain that for $t_1 \leq |r| \leq t_2$,

$$\begin{aligned} \left| K^{\mathrm{2D}}(r) \right| &\leq \left(1 - \frac{\pi^2 f_{\mathrm{lo}} (f_{\mathrm{lo}} + 4) x^2}{6} + \frac{\pi^4 (f_{\mathrm{lo}} + 2)^4 x^4}{72} \right) \\ &\qquad \left(1 - \frac{\pi^2 f_{\mathrm{lo}} (f_{\mathrm{lo}} + 4) y^2}{6} + \frac{\pi^4 (f_{\mathrm{lo}} + 2)^4 y^4}{72} \right) \\ &\leq \left(1 - \frac{\pi^2 \left(1 + \frac{2}{f_{\mathrm{lo}}}\right)^2 t_1^2}{6} \left(1 - \frac{\pi^2 \left(1 + \frac{2}{f_{\mathrm{lo}}}\right)^2 t_2^2}{12} \right) \right)^2, \\ \left| K^{\mathrm{2D}}_{(1,0)}(r) \right| &\leq \left(\frac{\pi^2 f_{\mathrm{lo}} (f_{\mathrm{lo}} + 4) t_2}{3} \right)^2. \end{aligned}$$

The same bound holds for $K^{\mathrm{2D}}_{(0,1)}$. Now set

$$W(r) = \alpha^\infty \left| K^{\mathrm{2D}}(r) \right| + 2\beta^\infty \left| K^{\mathrm{2D}}_{(1,0)}(r) \right|$$

t_1/λ_{lo}	t_2/λ_{lo}	Upper bound on $W(r)$	$Z_{(0,0)}(t_2)$	$Z_{(0,1)}(t_2)$
0.2447	0.27	0.9203	$6.561\,10^{-2}$	0.1074
0.27	0.36	0.9099	$7.196\,10^{-2}$	0.1184
0.36	0.56	0.8551	$9.239\,10^{-2}$	0.1540
0.56	0.84	0.8118	0.1490	0.2547

Table 6.2: Numerical quantities used to bound $|q|$ between $0.2447\,\lambda_{\text{lo}}$ and $0.84\,\lambda_{\text{lo}}$.

where α^∞ and β^∞ are the upper bounds from Lemma 6.1.2. The quantities reported in Table 6.2 imply that setting $\{t_1, t_2\}$ to $\{0.1649\,\lambda_{\text{lo}}, 0.27\,\lambda_{\text{lo}}\}$, $\{0.27\,\lambda_{\text{lo}}, 0.36\,\lambda_{\text{lo}}\}$, $\{0.36\,\lambda_{\text{lo}}, 0.56\,\lambda_{\text{lo}}\}$ and $\{0.56\,\lambda_{\text{lo}}, 0.84\,\lambda_{\text{lo}}\}$ yields $|q| < 0.9958$, $|q| < 0.9929$, $|q| < 0.9617$ and $|q| < 0.9841$ respectively in the corresponding intervals. Finally, for $|r|$ between $0.84\,\lambda_{\text{lo}}$ and $\Delta/2$, applying Lemma (5.3.1) allows to show that $W(r) \leq 0.5619$, $Z_{(0,0)}(0.84\,\lambda_{\text{lo}}) \leq 0.3646$ and $Z_{(0,1)}(0.84\,\lambda_{\text{lo}}) \leq 0.6502\,f_{\text{lo}}$, so that $|q| \leq 0.9850$. These last bounds also apply to any location beyond $\Delta/2$ closer to 0 than to any other element of T because of the monotonicity of B_0 and B_1. This concludes the proof.

Chapter 7

Proof of stability guarantees

This chapter contains the proof of Theorem 3.2.3. The main argument is laid out in Section 7.1, while Section 7.2 provides proofs for all the supporting lemmas. Finally, Section 7.3 discusses how to extend the result to a multidimensional setting.

7.1 Proof of Theorem 3.2.3

It is useful to first introduce various objects we shall need in the course of the proof. We let $T = \{t_j\}$ be the support of x and define the disjoint subsets

$$\begin{aligned} S^{\lambda}_{\text{near}}(j) &:= \{t \ : \ |t - t_j| \leq 0.16\lambda\}, \\ S^{\lambda}_{\text{far}} &:= \{t \ : \ |t - t_j| > 0.16\lambda, \ \forall t_j \in T\}; \end{aligned}$$

here, $\lambda \in \{\lambda_{\text{lo}}, \lambda_{\text{hi}}\}$, and j ranges from 1 to $|T|$. We write the union of the sets $S^{\lambda}_{\text{near}}(j)$ as

$$S^{\lambda}_{\text{near}} := \cup_{j=1}^{|T|} S^{\lambda}_{\text{near}}(j)$$

and observe that the pair $(S^{\lambda}_{\text{near}}, S^{\lambda}_{\text{far}})$ forms a partition of $\mathbb{T}$. The value of the constant 0.16 is not important and chosen merely to simplify the argument. We denote the restriction of a measure μ with finite total variation on a set S by $P_S\mu$ (note that in

contrast we denote the low-pass projection in the frequency domain by Q_{lo}). This restriction is well defined for the above sets, as one can take the Lebesgue decomposition of μ with respect to a positive σ-finite measure supported on any of them [87]. To keep some expressions in compact form, we set

$$I_{S^{\lambda}_{\mathrm{near}}(j)}(\mu) := \frac{1}{\lambda_{\mathrm{lo}}^2} \int_{S^{\lambda}_{\mathrm{near}}(j)} (t - t_j)^2 \, |\mu| \, (\mathrm{d}t),$$
$$I_{S^{\lambda}_{\mathrm{near}}}(\mu) := \sum_{t_j \in T} I_{S^{\lambda}_{\mathrm{near}}(j)}(\mu)$$

for any measure μ and $\lambda \in \{\lambda_{\mathrm{lo}}, \lambda_{\mathrm{hi}}\}$. Finally, we reserve the symbol C to denote a numerical constant whose value may change at each occurrence.

Set $h = x - x_{\mathrm{est}}$. The error obeys

$$||Q_{\mathrm{lo}} h||_{\mathcal{L}_1} \leq ||Q_{\mathrm{lo}} x - y||_{\mathcal{L}_1} + ||y - Q_{\mathrm{lo}} x_{\mathrm{est}}||_{\mathcal{L}_1} \leq 2\delta,$$

and has bounded total-variation norm since $||h||_{\mathrm{TV}} \leq ||x||_{\mathrm{TV}} + ||x_{\mathrm{est}}||_{\mathrm{TV}} \leq 2\,||x||_{\mathrm{TV}}$. Our aim is to bound the L_1 norm of the smoothed error $e := K_{\mathrm{hi}} * h$,

$$||e||_{\mathcal{L}_1} = \int_{\mathbb{T}} \left| \int_{\mathbb{T}} K_{\mathrm{hi}}(t - \tau) \, h(\mathrm{d}\tau) \right| \mathrm{d}t.$$

We begin with a lemma bounding the total-variation norm of h *away* from T.

Lemma 7.1.1. *Under the conditions of Theorem 3.2.3, there exist positive constants C_a and C_b such that*

$$\left|\left| P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h) \right|\right|_{\mathrm{TV}} + I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}}(h) \leq C_a \, \delta,$$
$$\left|\left| P_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{far}}}(h) \right|\right|_{\mathrm{TV}} \leq C_b \, \mathrm{SRF}^2 \, \delta.$$

This lemma is proved in Section 7.2.1 and relies on the existence of the dual certificate constructed in Chapter 5 to guarantee exact recovery in the noiseless setting.

To develop a bound about $\|e\|_{L_1}$, we begin by applying the triangle inequality to

obtain

$$\begin{aligned}|e(t)| &= \left|\int_{\mathbb{T}} K_{\mathrm{hi}}(t-\tau)\, h(\mathrm{d}\tau)\right| \\ &\leq \left|\int_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}} K_{\mathrm{hi}}(t-\tau)\, h(\mathrm{d}\tau)\right| + \left|\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}} K_{\mathrm{hi}}(t-\tau)\, h(\mathrm{d}\tau)\right|. \end{aligned} \tag{7.1.1}$$

By a corollary of the Radon-Nykodim Theorem (see Theorem 6.12 in [87]), it is possible to perform the polar decomposition $P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)(\mathrm{d}\tau) = e^{i2\pi\theta(\tau)}\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)\right|(\mathrm{d}\tau)$ such that $\theta(\tau)$ is a real function and $\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)\right|$ is a positive measure. Then

$$\begin{aligned}\int_{\mathbb{T}}\left|\int_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}} K_{\mathrm{hi}}(t-\tau)\, h(\mathrm{d}\tau)\right|\mathrm{d}t &\leq \int_{\mathbb{T}}\int_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}} |K_{\mathrm{hi}}(t-\tau)|\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)\right|(\mathrm{d}\tau)\,\mathrm{d}t \\ &= \int_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}\left(\int_{\mathbb{T}}|K_{\mathrm{hi}}(t-\tau)|\,\mathrm{d}t\right)\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)\right|(\mathrm{d}\tau) \\ &\leq C_0\left|\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)\right|\right|_{\mathrm{TV}}, \end{aligned} \tag{7.1.2}$$

where we have applied Fubini's theorem and (3.2.20) (note that the total-variation norm of $\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h)\right|$ is bounded by $2\,||x||_{\mathrm{TV}} < \infty$).

In order to control the second term in the right-hand side of (7.1.1), we use a first-order approximation of the super-resolution kernel provided by the Taylor series expansion of $\psi(\tau) = K_{\mathrm{hi}}(t-\tau)$ around t_j: for any τ such that $|\tau - t_j| \leq 0.16\lambda_{\mathrm{hi}}$, we have

$$|K_{\mathrm{hi}}(t-\tau) - K_{\mathrm{hi}}(t-t_j) - K'_{\mathrm{hi}}(t-t_j)(t_j-\tau)| \leq \sup_{u:|t-t_j-u|\leq 0.16\lambda_{\mathrm{hi}}} \frac{1}{2}|K''_{\mathrm{hi}}(u)|\,(\tau-t_j)^2.$$

Applying this together with the triangle inequality, and setting $t_j = 0$ without loss of

generality, gives

$$\int_{\mathbb{T}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} K_{\mathrm{hi}}(t-\tau)\, h(\mathrm{d}\tau) \right| \mathrm{d}t \leq \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} K_{\mathrm{hi}}(t)\, h(\mathrm{d}\tau) \right| \mathrm{d}t$$
$$+ \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} K'_{\mathrm{hi}}(t)\, \tau h(\mathrm{d}\tau) \right| \mathrm{d}t + \frac{1}{2} \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} \sup_{|t-u| \leq 0.16\lambda_{\mathrm{hi}}} |K''_{\mathrm{hi}}(u)|\, \tau^2 |h|(\mathrm{d}\tau) \right| \mathrm{d}t. \tag{7.1.3}$$

(To be clear, we do not lose generality by setting $t_j = 0$ since the analysis is invariant by translation; in particular by a translation placing t_j at the origin. To keep things as simple as possible, we shall make a frequent use of this argument.)

We then combine Fubini's theorem with (3.2.20) to obtain

$$\begin{aligned} \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} K_{\mathrm{hi}}(t)\, h(\mathrm{d}\tau) \right| \mathrm{d}t &\leq \int_{\mathbb{T}} |K_{\mathrm{hi}}(t)|\, \mathrm{d}t \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} h(\mathrm{d}\tau) \right| \\ &\leq C_0 \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} h(\mathrm{d}\tau) \right| \end{aligned} \tag{7.1.4}$$

and

$$\begin{aligned} \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} K'_{\mathrm{hi}}(t)\, \tau h(\mathrm{d}\tau) \right| \mathrm{d}t &\leq \int_{\mathbb{T}} |K'_{\mathrm{hi}}(t)|\, \mathrm{d}t \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} \tau h(\mathrm{d}\tau) \right| \\ &\leq \frac{C_1}{\lambda_{\mathrm{hi}}} \left| \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} \tau h(\mathrm{d}\tau) \right|. \end{aligned} \tag{7.1.5}$$

Some simple calculations show that (3.2.20) and (3.2.21) imply

$$\int_{\mathbb{T}} \sup_{|t-u| \leq 0.16\lambda_{\mathrm{hi}}} |K''_{\mathrm{hi}}(u)|\, \mathrm{d}t \leq \frac{C_4}{\lambda^2_{\mathrm{hi}}} \tag{7.1.6}$$

for a positive constant C_4. This together with Fubini's theorem yield

$$\int_{\mathbb{T}} \left| \int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} |K_{\text{hi}}''(u)|\, \tau^2 |h|\,(\mathrm{d}\tau) \right| \mathrm{d}t \leq \int_{\mathbb{T}} \sup_{|t-u| \leq 0.16\lambda_{\text{hi}}} |K_{\text{hi}}''(t)|\, \mathrm{d}t \left| \int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} \tau^2 |h|\,(\mathrm{d}\tau) \right|$$
$$\leq C_4\, \mathrm{SRF}^2\, I_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)}(h) \tag{7.1.7}$$

for any u. In order to make use of these bounds, it is necessary to control the local action of the measure h on a constant and a linear function. The following two lemmas are proved in Sections 7.2.2 and 7.2.3.

Lemma 7.1.2. *Take T as in Theorem 3.2.3 and any measure h obeying $||Q_{\text{lo}}h||_{\mathcal{L}_1} \leq 2\delta$. Then*

$$\sum_{t_j \in T} \left| \int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} h\,(\mathrm{d}\tau) \right| \leq 2\delta + \left|\left| P_{S_{\text{far}}^{\lambda_{\text{hi}}}}(h) \right|\right|_{\text{TV}} + C\, I_{S_{\text{near}}^{\lambda_{\text{hi}}}}(h)\,.$$

Lemma 7.1.3. *Take T as in Theorem 3.2.3 and any measure h obeying $||Q_{\text{lo}}h||_{\mathcal{L}_1} \leq 2\delta$. Then*

$$\sum_{t_j \in T} \left| \int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} (\tau - t_j)\, h\,(\mathrm{d}\tau) \right| \leq C\Big(\lambda_{\text{lo}}\, \delta + \lambda_{\text{lo}} \left|\left| P_{S_{\text{far}}^{\lambda_{\text{lo}}}}(h) \right|\right|_{\text{TV}} + \lambda_{\text{lo}}\, I_{S_{\text{near}}^{\lambda_{\text{lo}}}}(h)$$
$$+ \lambda_{\text{hi}}\, \mathrm{SRF}^2\, I_{S_{\text{near}}^{\lambda_{\text{lo}}}}(h) \Big).$$

We may now conclude the proof of our main theorem. Indeed, the inequalities (7.1.2), (7.1.3), (7.1.4), (7.1.5) and (7.1.7) together with $I_{S_{\text{near}}^{\lambda_{\text{hi}}}}(h) \leq I_{S_{\text{near}}^{\lambda_{\text{lo}}}}(h)$ and Lemmas 7.1.2 and 7.1.3 imply

$$||e||_{\mathcal{L}_1} \leq C \left(\mathrm{SRF}\, \delta + \left|\left| P_{S_{\text{far}}^{\lambda_{\text{hi}}}}(h) \right|\right|_{\text{TV}} + \mathrm{SRF} \left|\left| P_{S_{\text{far}}^{\lambda_{\text{lo}}}}(h) \right|\right|_{\text{TV}} + \mathrm{SRF}^2\, I_{S_{\text{near}}^{\lambda_{\text{lo}}}}(h) \right)$$
$$\leq C\, \mathrm{SRF}^2 \delta,$$

where the second inequality follows from Lemma 7.1.1.

7.2 Proof of lemmas

7.2.1 Proof of Lemma 7.1.1

The proof relies on the existence of a certain low-frequency polynomial, characterized in the following lemma which recalls results from Proposition 5.2.1 and Lemma 5.2.5.

Lemma 7.2.1. *Suppose T obeys the separation condition* (3.2.14) *and take any $v \in \mathbb{C}^{|T|}$ with $|v_j| = 1$. Then there exists a low-frequency trigonometric polynomial*

$$q(t) = \sum_{k=-f_{lo}}^{f_{lo}} c_k e^{i2\pi kt}$$

obeying the following properties:

$$q(t_j) = v_j, \quad t_j \in T, \tag{7.2.1}$$

$$|q(t)| \leq 1 - \frac{C_a \left(t - t_j\right)^2}{\lambda_{lo}^2}, \quad t \in S_{\text{near}}^{\lambda_{lo}}(j), \tag{7.2.2}$$

$$|q(t)| < 1 - C_b, \quad t \in S_{\text{far}}^{\lambda_{lo}}, \tag{7.2.3}$$

with $0 < C_b \leq 0.16^2 C_a < 1$.

Invoking a corollary of the Radon-Nykodim Theorem (see Theorem 6.12 in [87]), it is possible to perform a polar decomposition of $\mathcal{P}_T h$,

$$\mathcal{P}_T h = e^{i\phi(t)} \left|\mathcal{P}_T h\right|,$$

such that $\phi(t)$ is a real function defined on $\mathbb{T}$. To prove Lemma 7.1.1, we work with $v_j = e^{-i\phi(t_j)}$. Since q is low frequency,

$$\left|\int_{\mathbb{T}} q(t) \mathrm{d}h(t)\right| = \left|\int_{\mathbb{T}} q(t) Q_{\text{lo}} h(t)\, \mathrm{d}t\right| \leq ||q||_{\infty} \left|\left|Q_{\text{lo}} h\right|\right|_{\mathcal{L}_1} \leq 2\delta. \tag{7.2.4}$$

Next, since q interpolates $e^{-i\phi(t)}$ on T,

$$\begin{aligned}
||\mathcal{P}_T h||_{\mathrm{TV}} &= \int_{\mathbb{T}} q(t)\mathcal{P}_T h\,(\mathrm{d}t) \\
&\leq \left|\int_{\mathbb{T}} q(t)h\,(\mathrm{d}t)\right| + \left|\int_{T^c} q(t)h\,(\mathrm{d}t)\right| \\
&\leq 2\delta + \sum_{j\in T}\left|\int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)\setminus\{t_j\}} q(t)h\,(\mathrm{d}t)\right| + \left|\int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}} q(t)h\,(\mathrm{d}t)\right|. \qquad (7.2.5)
\end{aligned}$$

Applying (7.2.3) in Lemma 7.2.1 and Hölder's inequality, we obtain

$$\begin{aligned}
\left|\int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}} q(t)h\,(\mathrm{d}t)\right| &\leq \left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}q\right|\right|_{\infty}\left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}} \\
&\leq (1-C_b)\left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}}. \qquad (7.2.6)
\end{aligned}$$

Set $t_j = 0$ without loss of generality. The triangle inequality and (7.2.2) in Lemma 7.2.1 yield

$$\begin{aligned}
\left|\int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)\setminus\{0\}} q(t)h\,(\mathrm{d}t)\right| &\leq \int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)\setminus\{0\}} |q(t)|\,|h|\,(\mathrm{d}t) \\
&\leq \int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)\setminus\{0\}} \left(1-\frac{C_a t^2}{\lambda_{\mathrm{lo}}^2}\right)|h|\,(\mathrm{d}t) \\
&\leq \int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)\setminus\{0\}} |h|\,(\mathrm{d}t) - C_a I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)}(h). \qquad (7.2.7)
\end{aligned}$$

Combining (7.2.5), (7.2.6) and (7.2.7) gives

$$||\mathcal{P}_T h||_{\mathrm{TV}} \leq 2\delta + ||\mathcal{P}_{T^c} h||_{\mathrm{TV}} - C_b\left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}} - C_a I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}}(h).$$

Observe that we can substitute λ_{lo} with λ_{hi} in (7.2.5) and (7.2.7) and obtain

$$||\mathcal{P}_T h||_{\mathrm{TV}} \leq 2\delta + ||\mathcal{P}_{T^c} h||_{\mathrm{TV}} - 0.16^2\, C_a\,\mathrm{SRF}^{-2}\left|\left|P_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}} - C_a I_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}}(h).$$

This follows from using (7.2.2) instead of (7.2.3) to bound the magnitude of q on $S^{\lambda_{\mathrm{hi}}}_{\mathrm{far}}$.

These inequalities can be interpreted as a generalization of the strong null-space

property used to obtain stability guarantees for super-resolution on a discrete grid (see Lemma 3.2.2). Combined with the fact that x_{est} has minimal total-variation norm among all feasible points, they yield

$$\begin{aligned}
||x||_{\text{TV}} &\geq ||x+h||_{\text{TV}} \\
&\geq ||x||_{\text{TV}} - ||\mathcal{P}_T h||_{\text{TV}} + ||\mathcal{P}_{T^c} h||_{\text{TV}} \\
&\geq ||x||_{\text{TV}} - 2\delta + C_b \left|\left| P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + C_a I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h).
\end{aligned}$$

As a result, we conclude that

$$C_b \left|\left| P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + C_a I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h) \leq 2\delta,$$

and by the same argument,

$$0.16^2\, C_a\, \text{SRF}^{-2} \left|\left| P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + C_a I_{S^{\lambda_{\text{hi}}}_{\text{near}}}(h) \leq 2\delta.$$

This finishes the proof.

7.2.2 Proof of Lemma 7.1.2

The proof of this lemma relies upon the low-frequency polynomial from Lemma 7.2.1 and the fact that $q(t)$ is almost constant locally near the support of the signal of interest, as illustrated by Figure 7.1. As a result the polynomial is a good approximation to the chosen sign pattern when we only take into account a neighborhood of the support. This is shown by the following lemma, which we prove in Section 7.2.4.

Lemma 7.2.2. *There is a polynomial q satisfying the properties from Lemma 7.2.1 and, additionally,*

$$|q(t) - v_j| \leq \frac{C\,(t-t_j)^2}{\lambda_{\text{lo}}^2}, \qquad \textit{for all } t \in S^{\lambda_{\text{lo}}}_{\text{near}}(j).$$

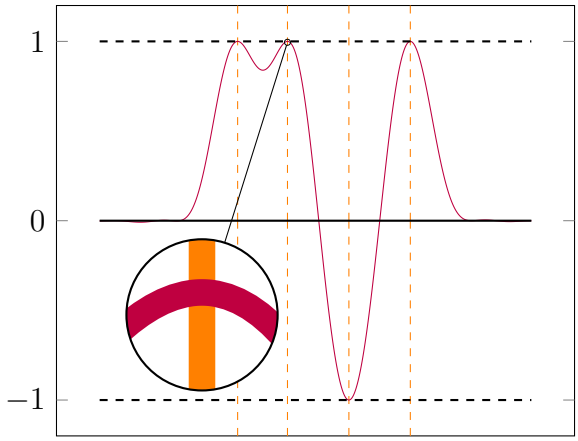


Figure 7.1: The low-frequency polynomial from Lemma 7.2.1 is almost constant in a neighborhood of the support of the signal.

Consider the polar form

$$\int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} h\left(\mathrm{d}\tau\right) = \left|\int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} h\left(\mathrm{d}\tau\right)\right| e^{i\theta_j},$$

where $\theta_j \in [0, 2\pi)$. We set $v_j = e^{i\theta_j}$ in Lemma 7.2.1 and apply the triangular inequality to obtain

$$\begin{aligned}\left|\int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} h\left(\mathrm{d}\tau\right)\right| &= \int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} e^{-i\theta_j} h\left(\mathrm{d}\tau\right) \\ &\leq \int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} q\left(\tau\right) h\left(\mathrm{d}\tau\right) + \left|\int_{S_{\text{near}}^{\lambda_{\text{hi}}}(j)} \left(q\left(\tau\right) - e^{-i\theta_j}\right) h\left(\mathrm{d}\tau\right)\right|, \end{aligned} \tag{7.2.8}$$

for all $t_j \in T$. By another application of the triangle inequality and (7.2.4)

$$\int_{S_{\text{near}}^{\lambda_{\text{hi}}}} q\left(\tau\right) h\left(\mathrm{d}\tau\right) \leq \left|\int_{\mathbb{T}} q\left(\tau\right) h\left(\mathrm{d}\tau\right)\right| + \left|\int_{S_{\text{far}}^{\lambda_{\text{hi}}}} q\left(\tau\right) h\left(\mathrm{d}\tau\right)\right| \leq 2\delta + \left|\left|P_{S_{\text{far}}^{\lambda_{\text{hi}}}}\left(h\right)\right|\right|_{\text{TV}}. \tag{7.2.9}$$

To bound the remaining term in (7.2.8), we apply Lemma 7.2.2 with $t_j = 0$ (this is

no loss of generality),

$$\begin{aligned}\left|\int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} \left(q(t)-e^{-i\theta_j}\right) h\left(\mathrm{d}t\right)\right| &\leq \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} \left|q(t)-e^{-i\theta_j}\right| |h|\left(\mathrm{d}t\right) \\ &\leq \int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} \frac{Ct^2}{\lambda_{\mathrm{lo}}^2} |h|\left(\mathrm{d}t\right) = CI_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)}(h)\,.\end{aligned}$$

It follows from this, (7.2.8) and (7.2.9) that

$$\left|\int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}} h\left(\mathrm{d}\tau\right)\right| \leq 2\delta + \left|\left|P_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}} + CI_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}}(h)\,.$$

The proof is complete.

7.2.3 Proof of Lemma 7.1.3

We record a simple lemma.

Lemma 7.2.3. *For any measure μ and $t_j = 0$,*

$$\left|\int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \tau\mu\left(\mathrm{d}\tau\right)\right| \leq 6.25\,\lambda_{\mathrm{hi}}\,\mathrm{SRF}^2\, I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)}(\mu)\,.$$

Proof. Note that in the interval $[0.16\lambda_{\mathrm{hi}}, 0.16\lambda_{\mathrm{lo}}]$, $t/0.16\,\lambda_{\mathrm{hi}} \geq 1$, whence

$$\begin{aligned}\left|\int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \tau\mu\left(\mathrm{d}\tau\right)\right| &\leq \int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \tau\,|\mu|\left(\mathrm{d}\tau\right) \\ &\leq \int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \frac{\tau^2}{0.16\lambda_{\mathrm{hi}}} |\mu|\left(\mathrm{d}\tau\right) \\ &\leq \frac{\lambda_{\mathrm{lo}}^2}{0.16\,\lambda_{\mathrm{hi}}} I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}(j)}(\mu)\,.\end{aligned}$$

□

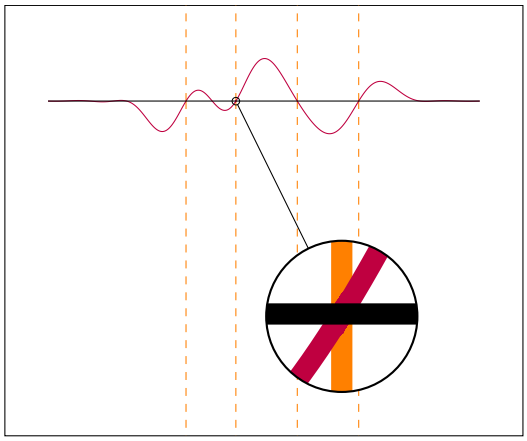


Figure 7.2: The low-frequency polynomial from Lemma 7.2.4 is almost linear in a neighborhood of the support of the signal.

We now turn our attention to the proof of Lemma 7.1.3. By the triangle inequality,

$$\sum_{t_j \in T} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} (\tau - t_j)\, h\,(\mathrm{d}\tau) \right| \leq$$
$$\sum_{t_j \in T} \left| \int_{S^{\lambda_{\text{lo}}}_{\text{near}}(j)} (\tau - t_j)\, h\,(\mathrm{d}\tau) \right| + \sum_{t_j \in T} \left| \int_{0.16\lambda_{\text{hi}} \leq |\tau - t_j| \leq 0.16\lambda_{\text{lo}}} (\tau - t_j)\, h\,(\mathrm{d}\tau) \right|. \quad (7.2.10)$$

The second term is bounded via Lemma 7.2.3. For the first, we use an argument very similar to the proof of Lemma 7.1.2. Here, we exploit the existence of a low-frequency polynomial that is almost linear in the vicinity of the elements of T. The result below, proved in Section 7.2.5, characterizes the polynomial, which is portrayed in Figure 7.2.

Lemma 7.2.4. *Suppose T obeys the separation condition* (3.2.14) *and take any $v \in \mathbb{C}^{|T|}$ with $|v_j| = 1$. Then there exists a low-frequency trigonometric polynomial*

$$q_1(t) = \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} c_k e^{i2\pi k t}$$

obeying

$$|q_1(t) - v_j\,(t - t_j)\,| \leq \frac{C_a\,(t-t_j)^2}{\lambda_{\rm lo}}, \quad t \in S^{\lambda_{\rm lo}}_{\rm near}\,(j)\,, \tag{7.2.11}$$

$$|q_1(t)| \leq C_b \lambda_{\rm lo}, \quad t \in S^{\lambda_{\rm lo}}_{\rm far}, \tag{7.2.12}$$

for positive constants C_a, C_b.

Consider the polar decomposition of

$$\int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} (\tau - t_j)\,h\,(\mathrm{d}\tau) = \left|\int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} (\tau - t_j)\,h\,(\mathrm{d}\tau)\right| e^{i\theta_j},$$

where $\theta_j \in [0, 2\pi)$, $t_j \in T$, and set $v_j = e^{i\theta_j}$ in Lemma 7.2.4. Again, suppose $t_j = 0$. Then

$$\begin{aligned}\left|\int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} \tau h\,(\mathrm{d}\tau)\right| &= \int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} e^{-i\theta_j}\tau h\,(\mathrm{d}\tau) \\ &\leq \left|\int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} \left(q_1\,(\tau) - e^{-i\theta_j}\tau\right) h\,(\mathrm{d}\tau)\right| + \int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} q_1\,(\tau)\,h\,(\mathrm{d}\tau)\,.\end{aligned} \tag{7.2.13}$$

The inequality (7.2.11) and Hölder's inequality allow to bound the first term in the right-hand side of (7.2.13),

$$\begin{aligned}\left|\int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} \left(q_1\,(\tau) - e^{-i\theta_j}\tau\right) h\,(\mathrm{d}\tau)\right| &\leq \int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} \left|q_1\,(\tau) - e^{-i\theta_j}\tau\right| |h|\,(\mathrm{d}\tau) \\ &\leq \frac{C_a}{\lambda_{\rm lo}} \int_{S^{\lambda_{\rm lo}}_{\rm near}(j)} \tau^2\,|h|\,(\mathrm{d}\tau) \\ &\leq C_a\,\lambda_{\rm lo}\,I_{S^{\lambda_{\rm lo}}_{\rm near}(j)}\,(h)\,.\end{aligned} \tag{7.2.14}$$

Another application of the triangular inequality yields

$$\int_{S^{\lambda_{\rm lo}}_{\rm near}} q_1\,(\tau)\,h\,(\mathrm{d}\tau) \leq \left|\int_{\mathbb{T}} q_1\,(\tau)\,h\,(\mathrm{d}\tau)\right| + \int_{S^{\lambda_{\rm lo}}_{\rm far}} q_1\,(\tau)\,h\,(\mathrm{d}\tau)\,. \tag{7.2.15}$$

We employ Hölder's inequality, (7.2.4), (7.2.11) and (7.2.12) to bound each of the

terms in the right-hand side. First,

$$\left|\int_{\mathbb{T}} q_1(\tau)\, h(\mathrm{d}\tau)\right| \leq ||q_1||_\infty \, ||Q_{\mathrm{lo}} h||_{\mathcal{L}_1} \leq C\, \lambda_{\mathrm{lo}}\, \delta. \tag{7.2.16}$$

Second,

$$\int_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}} q_1(\tau)\, h(\mathrm{d}\tau) \leq \left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(q_1)\right|\right|_\infty \left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}} \leq C_b\, \lambda_{\mathrm{lo}} \left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}}. \tag{7.2.17}$$

Combining (7.2.10) with these estimates gives

$$\begin{aligned}\sum_{t_j \in T} \left|\int_{S^{\lambda_{\mathrm{hi}}}_{\mathrm{near}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau)\right| \leq C\Big(&\lambda_{\mathrm{lo}}\, \delta + \lambda_{\mathrm{lo}} \left|\left|P_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{far}}}(h)\right|\right|_{\mathrm{TV}} + \lambda_{\mathrm{lo}}\, I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}}(h) \\ &+ \lambda_{\mathrm{hi}}\, \mathrm{SRF}^2\, I_{S^{\lambda_{\mathrm{lo}}}_{\mathrm{near}}}(h)\Big),\end{aligned}$$

as desired.

7.2.4 Proof of Lemma 7.2.2

We use the construction described in Section 5.2. In more detail,

$$q(t) = \sum_{t_k \in T} \alpha_k K(t - t_k) + \beta_k K^{(1)}(t - t_k),$$

where $\alpha, \beta \in \mathbb{C}^{|T|}$ are coefficient vectors, and K is defined in (5.2.1). Recall that we denote here, the ℓth derivative of K by $K^{(\ell)}$. By construction, the coefficients α and β are selected such that for all $t_j \in T$,

$$\begin{aligned} q(t_j) &= v_j \\ q'(t_j) &= 0. \end{aligned}$$

Without loss of generality we consider $t_j = 0$ and bound $q(t) - v_j$ in the interval $[0, 0.16\lambda_{\mathrm{lo}}]$. To ease notation, we define $w(t) = q(t) - v_j = w_R(t) + i\, w_I(t)$, where w_R is the real part of w and w_I the imaginary part. Leveraging different results from

Section 5.1 (in particular the equations in (5.3.16) and Lemmas 5.2.2 and 5.3.2), we have

$$\begin{aligned}
|w_R''(t)| &= \left| \sum_{t_k \in T} \mathrm{Re}\,(\alpha_k)\, K^{(2)}(t - t_k) + \sum_{t_k \in T} \mathrm{Re}\,(\beta_k)\, K^{(3)}(t - t_k) \right| \\
&\leq ||\alpha||_\infty \sum_{t_k \in T} \left| K^{(2)}(t - t_k) \right| + ||\beta||_\infty \sum_{t_k \in T} \left| K^{(3)}(t - t_k) \right| \\
&\leq C_\alpha \left(\left| K^{(2)}(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| K^{(2)}(t - t_k) \right| \right) \\
&\quad + C_\beta \lambda_{\mathrm{lo}} \left(\left| K^{(3)}(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| K^{(3)}(t - t_k) \right| \right) \\
&\leq C\, f_{\mathrm{lo}}^2.
\end{aligned}$$

The same bound holds for w_I''. Since $w_R(0)$, $w_R'(0)$, $w_I(0)$ and $w_I'(0)$ are all equal to zero, this implies $|w_R(t)| \leq C' f_{\mathrm{lo}}^2 t^2$ and $|w_I(t)| \leq C' f_{\mathrm{lo}}^2 t^2$ in the interval of interest, which allows the conclusion

$$|w(t)| \leq C\, f_{\mathrm{lo}}^2 t^2.$$

7.2.5 Proof of Lemma 7.2.4

The proof is similar to the technique proposed in Section 5.2, where a low-frequency kernel and its derivative are used to interpolate an arbitrary sign pattern on a support satisfying the minimum-distance condition. More precisely, we set

$$q_1(t) = \sum_{t_k \in T} \alpha_k K(t - t_k) + \beta_k K^{(1)}(t - t_k), \tag{7.2.18}$$

where $\alpha, \beta \in \mathbb{C}^{|T|}$ are coefficient vectors, and K is defined by (5.2.1). Note that K, $K^{(1)}$ and, consequently, q_1 are trigonometric polynomials with cut-off frequency f_{lo}.

By Lemma 5.3.2, it holds that for any $t_0 \in T$ and $t \in \mathbb{T}$ obeying $|t - t_0| \leq 0.16\lambda_{\text{lo}}$,

$$\sum_{t_k \in T \setminus \{t_0\}} \left| K^{(\ell)}\left(t - t_k\right) \right| \leq C_\ell f_{\text{lo}}^\ell, \tag{7.2.19}$$

where C_ℓ is a positive constant for $\ell = 0, 1, 2, 3$; in particular, $C_0 \leq 0.007$, $C_1 \leq 0.08$ and $C_2 \leq 1.06$. In addition, there exist other positive constants C_0' and C_1', such that for all $t_0 \in T$ and $t \in \mathbb{T}$ with $|t - t_0| \leq \Delta/2$,

$$\sum_{t_k \in T \setminus \{t_0\}} \left| K^{(\ell)}\left(t - t_k\right) \right| \leq C_\ell' f_{\text{lo}}^\ell \tag{7.2.20}$$

for $\ell = 0, 1$.

In order to satisfy (7.2.11) and (7.2.12), we constrain q_1 as follows: for each $t_j \in T$,

$$\begin{aligned} q_1(t_j) &= 0, \\ q_1'(t_j) &= v_j. \end{aligned}$$

Intuitively, this forces q_1 to approximate the linear function $v_j\left(t - t_j\right)$ around t_j. These constraints can be expressed in matrix form,

$$\begin{bmatrix} D_0 & D_1 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ v \end{bmatrix},$$

where

$$\left(D_0\right)_{jk} = K\left(t_j - t_k\right), \quad \left(D_1\right)_{jk} = K^{(1)}\left(t_j - t_k\right), \quad \left(D_2\right)_{jk} = K^{(2)}\left(t_j - t_k\right),$$

and j and k range from 1 to $|T|$. Lemma 5.2.2 establishes that under the minimum-separation condition this system is invertible, so that α and β are well defined. These coefficient vectors can consequently be expressed as

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} -D_0^{-1} D_1 \\ I \end{bmatrix} S^{-1} v, \quad S := D_2 - D_1 D_0^{-1} D_1,$$

where S is the Schur complement. Inequality (7.2.19) implies

$$||I - D_0||_\infty \leq C_0, \tag{7.2.21}$$

$$||D_1||_\infty \leq C_1 f_{\text{lo}}, \tag{7.2.22}$$

$$||\kappa I - D_2||_\infty \leq C_2 f_{\text{lo}}^2, \tag{7.2.23}$$

where $\kappa = \left|K^{(2)}(0)\right| = \pi^2 f_{\text{lo}}(f_{\text{lo}} + 4)/3$.

Recall that if $||I - M||_\infty < 1$, the series $M^{-1} = (I - (I - M))^{-1} = \sum_{k\geq 0} (I - M)^k$ is convergent and we have

$$\left|\left|M^{-1}\right|\right|_\infty \leq \frac{1}{1 - ||I - M||_\infty}.$$

This, together with (7.2.21), (7.2.22) and (7.2.23) implies

$$\begin{aligned}
\left|\left|D_0^{-1}\right|\right|_\infty &\leq \frac{1}{1 - ||I - D_0||_\infty} \\
&\leq \frac{1}{1 - C_0}, \\
||\kappa I - S||_\infty &\leq ||\kappa I - D_2||_\infty + ||D_1||_\infty \left|\left|D_0^{-1}\right|\right|_\infty ||D_1||_\infty \\
&\leq \left(C_2 + \frac{C_1^2}{1 - C_0}\right) f_{\text{lo}}^2,
\end{aligned}$$

$$\begin{aligned}
\left|\left|S^{-1}\right|\right|_\infty &= \kappa^{-1} \left|\left|\left(\frac{S}{\kappa}\right)^{-1}\right|\right|_\infty \\
&\leq \frac{1}{\kappa - ||\kappa I - S||_\infty} \\
&\leq \left(\kappa - \left(C_2 + \frac{C_1^2}{1 - C_0}\right) f_{\text{lo}}^2\right)^{-1} \\
&\leq C_\kappa \lambda_{\text{lo}}^2,
\end{aligned}$$

for a certain positive constant C_κ. Note that due to the numeric upper bounds on

the constants in (7.2.19) C_κ is indeed a positive constant as long as $f_{\text{lo}} \geq 1$. Finally, we obtain a bound on the magnitude of the entries of α

$$\begin{aligned} ||\alpha||_\infty &= \left|\left|D_0^{-1} D_1 S^{-1} v\right|\right|_\infty \\ &\leq \left|\left|D_0^{-1} D_1 S^{-1}\right|\right|_\infty \\ &\leq \left|\left|D_0^{-1}\right|\right|_\infty ||D_1||_\infty \left|\left|S^{-1}\right|\right|_\infty \\ &\leq C_\alpha \lambda_{\text{lo}}, \end{aligned} \tag{7.2.24}$$

where $C_\alpha = C_\kappa C_1 / (1 - C_0)$, and on the entries of β

$$\begin{aligned} ||\beta||_\infty &= \left|\left|S^{-1} v\right|\right|_\infty \\ &\leq \left|\left|S^{-1}\right|\right|_\infty \\ &\leq C_\beta \lambda_{\text{lo}}^2, \end{aligned} \tag{7.2.25}$$

for a positive constant $C_\beta = C_\kappa$. Combining these inequalities with (7.2.20) and the fact that the absolute values of $K(t)$ and $K^{(1)}(t)$ are bounded by one and $7 f_{\text{lo}}$ respectively (see also Section 6.2.1), we obtain the bound below, valid for any t. We use t_i to denote the element in T nearest to t (note that all other elements are at least $\Delta/2$ away). This establishes (7.2.12).

$$\begin{aligned} |q_1(t)| &= \left| \sum_{t_k \in T} \alpha_k K\left(t - t_k\right) + \sum_{t_k \in T} \beta_k K^{(1)}\left(t - t_k\right) \right| \\ &\leq ||\alpha||_\infty \sum_{t_k \in T} \left|K\left(t - t_k\right)\right| + ||\beta||_\infty \sum_{t_k \in T} \left|K^{(1)}\left(t - t_k\right)\right| \\ &\leq C_\alpha \lambda_{\text{lo}} \left(\left|K\left(t\right)\right| + \sum_{t_k \in T \setminus \{t_i\}} \left|K\left(t - t_k\right)\right| \right) \\ &\quad + C_\beta \lambda_{\text{lo}}^2 \left(\left|K^{(1)}\left(t\right)\right| + \sum_{t_k \in T \setminus \{t_i\}} \left|K^{(1)}\left(t - t_k\right)\right| \right) \\ &\leq C \lambda_{\text{lo}}. \end{aligned} \tag{7.2.26}$$

The proof is completed by the following lemma, which proves (7.2.11).

Lemma 7.2.5. *For any $t_j \in T$ and $t \in \mathbb{T}$ obeying $|t - t_j| \leq 0.16\lambda_{\text{lo}}$, we have*

$$|q_1(t) - v_j\,(t - t_j)\,| \leq \frac{C\,(t - t_j)^2}{\lambda_{\text{lo}}}.$$

Proof. We assume without loss of generality that $t_j = 0$. By symmetry, it suffices to show the claim for $t \in (0, 0.16\,\lambda_{\text{lo}}]$. To ease notation, we define $w(t) = v_j t - q_1(t) = w_R(t) + i\,w_I(t)$, where w_R is the real part of w and w_I the imaginary part. Leveraging (7.2.24), (7.2.25) and (7.2.19) together with the fact that $K^{(2)}(t)$ and $K^{(3)}(t)$ are bounded by $4f_{\text{lo}}^2$ and $6f_{\text{lo}}^3$ respectively if $|t| \leq 0.16\lambda_{\text{lo}}$ (see Section 5.3.2), we obtain

$$\begin{aligned}
|w_R''\,(t)| &= \left| \sum_{t_k \in T} \text{Re}\,(\alpha_k)\,K^{(2)}\,(t - t_k) + \sum_{t_k \in T} \text{Re}\,(\beta_k)\,K^{(3)}\,(t - t_k) \right| \\
&\leq ||\alpha||_\infty \sum_{t_k \in T} \left| K^{(2)}\,(t - t_k) \right| + ||\beta||_\infty \sum_{t_k \in T} \left| K^{(3)}\,(t - t_k) \right| \\
&\leq C_\alpha \lambda_{\text{lo}} \left(\left| K^{(2)}\,(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| K^{(2)}\,(t - t_k) \right| \right) \\
&\quad + C_\beta \lambda_{\text{lo}}^2 \left(\left| K^{(3)}\,(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| K^{(3)}\,(t - t_k) \right| \right) \\
&\leq C\,f_{\text{lo}}.
\end{aligned}$$

The same bound applies to w_I. Since $w_R(0)$, $w_R'(0)$, $w_I(0)$ and $w_I'(0)$ are all equal to zero, this implies $|w_R(t)| \leq C f_{\text{lo}} t^2$—and similarly for $|w_I(t)|$—in the interval of interest. Whence, $|w(t)| \leq C f_{\text{lo}} t^2$. □

7.3 Extension to multiple dimensions

The extension of the proof hinges on establishing versions of Lemmas 7.2.1, 7.2.2 and 7.2.4 for multiple dimensions. These lemmas construct bounded low-frequency polynomials which interpolate a sign pattern on a well-separated set of points S and

have bounded second derivatives in a neighborhood of S. In the multidimensional case, we need the directional derivative of the polynomials to be bounded in any direction, which can be ensured by bounding the eigenvalues of their Hessian matrix evaluated on the support of the signal. To construct such polynomials one can proceed in a way similar to the proof of Lemmas 7.2.1 and 7.2.4, namely, by using a low-frequency kernel constructed by tensorizing several squared Fejér kernels to interpolate the sign pattern, while constraining the first-order derivatives to either vanish or have a fixed value. As in the one-dimensional case, one can set up a system of equations and prove that it is well conditioned using the rapid decay of the interpolation kernel away from the origin. Finally, one can verify that the construction satisfies the required conditions by exploiting the fact that the interpolation kernel and its derivatives are locally quadratic and rapidly decaying. This is spelled out in Section 6.1 to prove a version of Lemma 7.2.1 in two dimensions. In order to clarify further how to adapt our techniques to a multidimensional setting we provide below a sketch of the proof of the analog of Lemma 7.1.1 in two dimensions. In particular, this illustrates how the increase in dimension does not change the exponent of the SRF in our recovery guarantees.

7.3.1 Proof of an extension of Lemma 7.1.1 to two dimensions

We now have $t \in \mathbb{T}^2$. As a result, we redefine

$$\begin{aligned} S^{\lambda}_{\text{near}}(j) &:= \left\{ t \ : \ ||t - t_j||_{\infty} \leq w\,\lambda \right\}, \\ S^{\lambda}_{\text{far}} &:= \left\{ t \ : \ ||t - t_j||_{\infty} > w\,\lambda, \ \forall t_j \in T \right\}, \\ I_{S^{\lambda}_{\text{near}}(j)}(\mu) &:= \frac{1}{\lambda_{\text{lo}}^2} \int_{S^{\lambda}_{\text{near}}(j)} ||t - t_j||_2^2 \, |\mu| \, (\mathrm{d}t), \end{aligned}$$

where w is a constant.

The proof relies on the existence of a low-frequency polynomial

$$q(t) = \sum_{k_1=-f_{\text{lo}}}^{f_{\text{lo}}} \sum_{k_2=-f_{\text{lo}}}^{f_{\text{lo}}} c_{k_1,k_2} e^{i2\pi(k_1 t_1 + k_2 t_2)}$$

satisfying

$$q(t_j) = v_j, \quad t_j \in T, \tag{7.3.1}$$

$$|q(t)| \leq 1 - \frac{C'_a \, ||t - t_j||_2^2}{\lambda_{\text{lo}}^2}, \quad t \in S_{near}^{\lambda_{\text{lo}}}(j), \tag{7.3.2}$$

$$|q(t)| < 1 - C'_b, \quad t \in S_{far}^{\lambda_{\text{lo}}}, \tag{7.3.3}$$

where C'_a and C'_b are constants. Proposition 6.1.1 establishes the existence of such a polynomial. Under a minimum distance condition, which constrains the elements of T to be separated by $2.38\,\lambda_{\text{lo}}$ in infinity norm (as mentioned before this choice of norm is arbitrary and could be changed to the ℓ_2 norm), in Section 6.1 we show that q satisfies (7.3.1) and (7.3.3) and that both eigenvalues of its Hessian matrix evaluated on T are of order f_{lo}^2, which implies (7.3.2).

As in one dimension, we perform a polar decomposition of $\mathcal{P}_T h$,

$$\mathcal{P}_T h = e^{i\phi(t)} \, |\mathcal{P}_T h| \, ,$$

and work with $v_j = e^{-i\phi(t_j)}$. The rest of the proof is almost identical to the 1D case. Since q is low frequency,

$$\left| \int_{\mathbb{T}^2} q(t) \mathrm{d}h(t) \right| \leq 2\delta. \tag{7.3.4}$$

Next, since q interpolates $e^{-i\phi(t)}$ on T,

$$||\mathcal{P}_T h||_{\text{TV}} = \int_{\mathbb{T}^2} q(t) \mathcal{P}_T h \, (\mathrm{d}t) \leq 2\delta + \sum_{j \in T} \left| \int_{S_{\text{near}}^{\lambda_{\text{lo}}}(j) \setminus \{t_j\}} q(t) h \, (\mathrm{d}t) \right| + \left| \int_{S_{\text{far}}^{\lambda_{\text{lo}}}} q(t) h \, (\mathrm{d}t) \right| . \tag{7.3.5}$$

Applying (7.3.3) and Hölder's inequality, we obtain

$$\left| \int_{S_{\text{far}}^{\lambda_{\text{lo}}}} q(t) h \, (\mathrm{d}t) \right| \leq (1 - C'_b) \left|\left| P_{S_{\text{far}}^{\lambda_{\text{lo}}}}(h) \right|\right|_{\text{TV}} . \tag{7.3.6}$$

Setting $t_j = (0,0)$ without loss of generality, the triangle inequality and (7.3.2) yield

$$\left| \int_{S^{\lambda_{\text{lo}}}_{\text{near}}(j)\setminus\{(0,0)\}} q(t)h\,(\mathrm{d}t) \right| \leq \int_{S^{\lambda_{\text{lo}}}_{\text{near}}(j)\setminus\{(0,0)\}} |h|\,(\mathrm{d}t) - C'_a I_{S^{\lambda_{\text{lo}}}_{\text{near}}(j)}(h)\,. \tag{7.3.7}$$

Combining (7.3.5), (7.3.6) and (7.3.7) gives

$$||\mathcal{P}_T h||_{\text{TV}} \leq 2\delta + ||\mathcal{P}_{T^c} h||_{\text{TV}} - C'_b \left|\left| P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} - C'_a I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h)$$

and similarly

$$||\mathcal{P}_T h||_{\text{TV}} \leq 2\delta + ||\mathcal{P}_{T^c} h||_{\text{TV}} - w^2\, C'_a\, \text{SRF}^{-2} \left|\left| P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} - C'_a I_{S^{\lambda_{\text{hi}}}_{\text{near}}}(h)\,.$$

By the same argument as in the 1D case, the fact that x_{est} has minimal total-variation norm is now sufficient to establish

$$C'_b \left|\left| P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + C'_a I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h) \leq 2\delta,$$

and

$$w^2\, C'_a\, \text{SRF}^{-2} \left|\left| P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + C'_a I_{S^{\lambda_{\text{hi}}}_{\text{near}}}(h) \leq 2\delta.$$

Chapter 8

Proof of support-detection guarantees

In this chapter we prove Theorem 3.3.1. Section 8.1 provides the outline of the proof and Sections 8.2 and 8.3 contain the proofs of the two main lemmas used to establish the result.

8.1 Main argument

We begin with an intermediate result proved in Section 8.2.

Lemma 8.1.1. *Under the assumptions of Theorem 3.3.1*

$$\sum_{t_k^{\text{est}} \in T_{\text{est}}} \left| c_k^{\text{est}} \right| \min \left\{ C_a, \frac{C_b \, d\left(t_k^{\text{est}}, T\right)}{\lambda_{\text{lo}}^2} \right\} \leq 2\delta,$$

where C_a and C_b are positive numerical constants and

$$d\left(t, T\right) := \min_{t_i \in T} \left(t - t_i\right)^2.$$

Properties (3.3.4) and (3.3.5) are direct corollaries of Lemma (8.1.1). To establish Property (3.3.3) we need the following key lemma, proved in Section 8.3.

Lemma 8.1.2. *Suppose T obeys condition* (3.2.14) *and $f_{\text{lo}} \geq 10$. Then for any $t_j \in T$ there exists a low-pass polynomial*

$$q_{t_j}(t) = \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} b_k e^{i2\pi kt},$$

$b \in \mathbb{C}^n$, such that $|q_{t_j}(t)| < 1$ for all $t \neq t_j$ and

$$\begin{aligned}
&q_{t_j}(t_j) = 1, \\
&q_{t_j}(t_l) = 0 \quad t_l \in T \setminus \{t_j\}, \\
&|1 - q_{t_j}(t)| \leq \frac{C_1' (t - t_j)^2}{\lambda_{\text{lo}}^2} \quad \textit{for} \quad |t - t_j| \leq c\lambda_{\text{lo}}, && (8.1.1) \\
&|q_{t_j}(t)| \leq \frac{C_1' (t - t_l)^2}{\lambda_{\text{lo}}^2} \quad \textit{for} \quad t_l \in T \setminus \{t_j\}, \ |t - t_l| \leq c\lambda_{\text{lo}}, && (8.1.2) \\
&|q_{t_j}(t)| < C_2' \quad \textit{if} \quad |t - t_l| > c\lambda_{\text{lo}} \ \forall \ t_l \in T, && (8.1.3)
\end{aligned}$$

where $0 < c^2 C_2' \leq C_1' < 1$.

The polynomial q_{t_j} provided by this lemma is designed to satisfy $\int_{\mathbb{T}} q_{t_j}(t) x(\mathrm{d}t) = c_j$ and vanish on the rest of the support of the signal. This allows to decouple the estimation error at t_j from the amplitude of the rest of the spikes. Since x and x_{est} are feasible for (3.2.12), we can apply Parseval's Theorem and the Cauchy-Schwarz inequality to obtain

$$\begin{aligned}
\left| \int_{\mathbb{T}} q_{t_j}(t) x(\mathrm{d}t) - \int_{\mathbb{T}} q_{t_j}(t) x_{\text{est}}(\mathrm{d}t) \right| &= \left| \sum_{k=-f_{lo}}^{f_{lo}} b_k \mathcal{F}_n(x - x_{\text{est}})_k \right| \\
&\leq \left|\left| q_{t_j} \right|\right|_{\mathcal{L}_2} \left|\left| \mathcal{F}_n(x - x_{\text{est}}) \right|\right|_2 \\
&\leq 2\delta, && (8.1.4)
\end{aligned}$$

where we have used that the absolute value and consequently the $\mathcal{L}_2$ norm of q_{t_j} is

bounded by one. In addition, by Lemmas 8.1.2 and 8.1.1 we have

$$\begin{aligned}
&\Big| \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|>c\lambda_{\text{lo}}\right\}} c_k^{\text{est}} q_{t_j}(t_k^{\text{est}}) + \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|\le c\lambda_{\text{lo}}\right\}} c_k^{\text{est}} \left(q_{t_j}(t_k^{\text{est}})-1\right) \Big| \\
&\quad \le \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|>c\lambda_{\text{lo}}\right\}} \left|c_k^{\text{est}}\right| \left|q_{t_j}(t_k^{\text{est}})\right| + \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|\le c\lambda_{\text{lo}}\right\}} \left|c_k^{\text{est}}\right| \left|1-q_{t_j}(t_k^{\text{est}})\right| \\
&\quad \le \sum_{t_k^{\text{est}}\in T_{\text{est}}} \left|c_k^{\text{est}}\right| \min\left\{C_2', \frac{C_1' d\left(t_k^{\text{est}}, T\right)}{\lambda_{\text{lo}}^2}\right\} \\
&\quad \le C\delta, \qquad (8.1.5)
\end{aligned}$$

for a positive numerical constant C. Finally, Lemma 8.1.2, the triangle inequality, (8.1.4) and (8.1.5) yield

$$\begin{aligned}
\Big|c_j - \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|\le c\lambda_{lo}\right\}} c_k^{\text{est}}\Big| &= \Big| \int_{\mathbb{T}} q_{t_j}(t) x(\mathrm{d}t) - \int_{\mathbb{T}} q_{t_j}(t) x_{\text{est}}(\mathrm{d}t) \\
&\quad + \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|>c\lambda_{\text{lo}}\right\}} c_k^{\text{est}} q_{t_j}(t_k^{\text{est}}) \\
&\quad + \sum_{\left\{k:\, \left|t_k^{\text{est}}-t_j\right|\le c\lambda_{\text{lo}}\right\}} c_k^{\text{est}} \left(q_{t_j}(t_k^{\text{est}})-1\right) \Big| \\
&\le C'\delta,
\end{aligned}$$

for a positive numerical constant C'.

8.2 Proof of Lemma 8.1.1

The proof relies on the dual certificate constructed in Chapter 5. Recall that Proposition 5.2.1 implies that if T obeys Condition (3.2.14), for any $v \in \mathbb{C}^{|T|}$ such that

$|v_j| = 1$ for all entries v_j there exists a low-pass polynomial

$$q(t) = \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} d_k e^{i2\pi kt},$$

where $d \in \mathbb{C}^n$ that satisfies

$$\begin{aligned}
&q(t_j) = v_j, \qquad t_j \in T, \\
&|q(t)| < 1 - C_a, \qquad |t - t_j| > c\lambda_{\text{lo}} \quad \forall t_j \in T, \\
&|q(t)| \leq 1 - \frac{C_b \left(t - t_j\right)^2}{\lambda_{\text{lo}}^2}, \qquad |t - t_j| \leq c\lambda_{\text{lo}},\, t_j \in T,
\end{aligned}$$

with $0 < c^2 C_b \leq C_a < 1$.

We set $v_j = \overline{c_j} / |c_j|$. The proposition implies that

$$\begin{aligned}
\int_{\mathbb{T}} q(t) x_{\text{est}} \left(\mathrm{d}t\right) &\leq \sum_k \left|c_k^{\text{est}}\right| \left|q(t_k^{\text{est}})\right| \\
&\leq \sum_k \left(1 - \min\left\{C_a, \frac{C_b\, d\left(t_k^{\text{est}}, T\right)}{\lambda_{\text{lo}}^2}\right\}\right) \left|c_k^{\text{est}}\right|. \qquad (8.2.1)
\end{aligned}$$

The same argument used to prove (8.1.4) yields

$$\left| \int_{\mathbb{T}} q(t) x_{\text{est}}(\mathrm{d}t) - \int_{\mathbb{T}} q(t) x(\mathrm{d}t) \right| \leq 2\delta.$$

Now, taking into account that $\int_{\mathbb{T}} q(t) x \left(\mathrm{d}t\right) = ||x||_{\text{TV}}$ by construction and $||x_{\text{est}}||_{\text{TV}} \leq ||x||_{\text{TV}}$, we have

$$\begin{aligned}
\int_{\mathbb{T}} q(t) x_{\text{est}} \left(\mathrm{d}t\right) &= \int_{\mathbb{T}} q(t) x \left(\mathrm{d}t\right) + \int_{\mathbb{T}} q(t) x_{\text{est}}(\mathrm{d}t) - \int_{\mathbb{T}} q(t) x(\mathrm{d}t) \\
&\geq ||x||_{\text{TV}} - 2\delta \geq ||x_{\text{est}}||_{\text{TV}} - 2\delta \\
&= \sum_k \left|c_k^{\text{est}}\right| - 2\delta.
\end{aligned}$$

Combining this with (8.2.1) completes the proof.

8.3 Proof of Lemma 8.1.2

The proof is similar to the technique proposed in Section 5.2, where a low-frequency kernel and its derivative are used to interpolate an arbitrary sign pattern on a support satisfying the minimum-distance condition (3.2.14). More precisely, we set

$$q_{t_j}(t) = \sum_{t_k \in T} \alpha_k K(t - t_k) + \beta_k K^{(1)}(t - t_k), \tag{8.3.1}$$

where $\alpha, \beta \in \mathbb{C}^{|T|}$ are coefficient vectors, and K is defined in (5.2.1). Recall that we denote the ℓth derivative of K by $K^{(\ell)}$. Note that K, $K^{(1)}$ and, consequently, q_{t_j} are trigonometric polynomials of the required degree.

We impose

$$\begin{aligned} q_{t_j}(t_j) &= 1, \\ q_{t_j}(t_l) &= 0, \qquad t_l \in T / \{t_j\}, \\ q'_{t_j}(t_k) &= 0, \qquad t_k \in T. \end{aligned}$$

We express these constraints in matrix form. Let $e_{t_j} \in \mathbb{R}^{|T|}$ denote the one-sparse vector with one nonzero entry at the position corresponding to t_j. Then,

$$\begin{bmatrix} D_0 & D_1 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} e_{t_j} \\ 0 \end{bmatrix},$$

where

$$\begin{aligned} (D_0)_{lk} &= K(t_l - t_k), \quad (D_1)_{lk} = K^{(1)}(t_l - t_k), \\ (D_2)_{lk} &= K^{(2)}(t_l - t_k), \end{aligned}$$

and l and k range from 1 to $|T|$. Lemma 5.2.2 establishes that under the minimum-separation condition this system is invertible. As a result α and β are well defined and q_{t_j} satisfies $q_{t_j}(t_j) = 1$ and $q_{t_j}(t_l) = 0$ for $t_l \in T / \{t_j\}$. The coefficient vectors can

be expressed as

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = \begin{bmatrix} I \\ -D_2^{-1}D_1\end{bmatrix} S^{-1}e_{t_j}, \quad S := D_0 - D_1 D_2^{-1} D_1,$$

where S is the Schur complement. We borrow some results from Section 5.3.1,

$$\begin{aligned}
||I-S||_\infty &\leq 8.747\,10^{-3},\\
\left|\left|S^{-1}\right|\right|_\infty &\leq 1 + 8.824\,10^{-3},\\
\left|\left|I-S^{-1}\right|\right|_\infty &\leq \left|\left|S^{-1}\right|\right|_\infty ||S-I||_\infty \leq 8.825\,10^{-3},\\
\left|\left|\alpha - e_{t_j}\right|\right|_\infty &\leq \left|\left|I-S^{-1}\right|\right|_\infty \left|\left|e_{t_j}\right|\right|_\infty\\
&\leq 8.825\,10^{-3}, \qquad (8.3.2)\\
||\beta||_\infty &\leq 3.294\,10^{-2}\lambda_{\text{lo}}. \qquad (8.3.3)
\end{aligned}$$

Lemma 5.3.1 allows to obtain

$$K(t) \leq \frac{1}{(f_{\text{lo}}t)^4} \leq 0.333, \quad K'(t) \leq \frac{4\pi}{f_{\text{lo}}^3 t^4} \leq 4.18\, f_{\text{lo}},$$

for $|t| > c\lambda_{\text{lo}}$ as long $f_{\text{lo}} \geq 10$. By the same lemma, if we set the minimum separation $\Delta_{\min}$ to $2/f_c$

$$\begin{aligned}
\sum_{t_k \in T\setminus\{t_a,t_b\}} |K(t-t_k)| &\leq \sum_{l=0}^{\infty} \frac{1}{\left(f_{\text{lo}}\Delta_{\min}(\frac{1}{2}+l)\right)^4} + \sum_{l=0}^{\infty}\frac{1}{(f_{\text{lo}}\Delta_{\min}l)^4}\\
&\leq 1.083,
\end{aligned}$$

$$\begin{aligned}
\sum_{t_k \in T\setminus\{t_a,t_b\}} \left|K^{(1)}(t-t_k)\right| &\leq \sum_{l=0}^{\infty} \frac{4\pi}{f_{\text{lo}}^3\left(\Delta_{\min}(\frac{1}{2}+l)\right)^4} + \sum_{j=0}^{\infty}\frac{4\pi}{f_{\text{lo}}^3(\Delta_{\min}l)^4}\\
&\leq 1.75\, f_{\text{lo}},
\end{aligned}$$

where t_a and t_b are the two spikes nearest to t. Let t_i be the element of $T/\{t_j\}$ that

is nearest to t. Combining these inequalities with (8.3.2) and (8.3.3) proves that

$$\begin{aligned}
|q_{t_j}(t)| &= \Big| \sum_{t_k \in T} \alpha_k K\left(t - t_k\right) + \sum_{t_k \in T} \beta_k K^{(1)}\left(t - t_k\right) \Big| \\
&\leq \left|K\left(t - t_j\right)\right| + \left|\left|\alpha - e_{t_j}\right|\right|_\infty \Big(\left|K\left(t - t_j\right)\right| \\
&\quad + \left|K\left(t - t_i\right)\right| + \sum_{t_k \in T \setminus \{t_i, t_j\}} \left|K\left(t - t_k\right)\right| \Big) \\
&\quad + \left|\left|\beta\right|\right|_\infty \Big(\left|K^{(1)}\left(t - t_j\right)\right| + \left|K^{(1)}\left(t - t_i\right)\right| \\
&\quad + \sum_{t_k \in T \setminus \{t_i, t_j\}} \left|K^{(1)}\left(t - t_k\right)\right| \Big) \\
&\leq 0.69,
\end{aligned}$$

if $|t - t_k| > c\lambda_{\text{lo}}$ for all $t_k \in T$ so that (8.1.3) holds. The proof is completed by two lemmas which prove (8.1.1) and (8.1.2) and $|q_{t_j}(t)| < 1$ for any t. They rely on the following bounds borrowed from equation (5.3.16),

$$\begin{aligned}
K\left(t\right) &\geq 0.9539, & K^{(2)}\left(t\right) &\leq -2.923\, f_{\text{lo}}^2, \\
\left|K^{(1)}\left(t\right)\right| &\leq 0.5595\, f_{\text{lo}}, & \left|K^{(2)}\left(t\right)\right| &\leq 3.393\, f_{\text{lo}}^2, \\
\left|K^{(3)}\left(t\right)\right| &\leq 5.697\, f_{\text{lo}}^3,
\end{aligned} \tag{8.3.4}$$

and on the fact that, due to Lemma 5.3.2, for any $t_0 \in T$ and $t \in \mathbb{T}$ obeying $|t - t_0| \leq c\lambda_{\text{lo}}$,

$$\sum_{t_k \in T \setminus \{t_0\}} \left|K^{(2)}\left(t - t_k\right)\right| \leq 1.06\, f_{\text{lo}}^2 \tag{8.3.5}$$

$$\sum_{t_k \in T \setminus \{t_0\}} \left|K^{(3)}\left(t - t_k\right)\right| \leq 18.6\, f_{\text{lo}}^3. \tag{8.3.6}$$

Lemma 8.3.1. *For any t such that $|t - t_j| \leq c\lambda_{\text{lo}}$,*

$$1 - 4.07\left(t - t_j\right)^2 f_{\text{lo}}^2 \leq q_{t_j}(t) \leq 1 - 2.30\left(t - t_j\right)^2 f_{\text{lo}}^2.$$

Proof. We assume without loss of generality that $t_j = 0$. By symmetry, it suffices to

show the claim for $t \in (0, c\,\lambda_{\text{lo}}]$. By (8.3.2), (8.3.3), (8.3.4), (8.3.5) and (8.3.6),

$$\begin{aligned} q_0''(t) &= \sum_{t_k \in T} \alpha_k K^{(2)}(t - t_k) + \sum_{t_k \in T} \beta_k K^{(3)}(t - t_k) \\ &\leq \left(1 + \left|\left|\alpha - e_{t_j}\right|\right|_\infty\right) K^{(2)}(t) \\ &\quad + \left|\left|\alpha - e_{t_j}\right|\right|_\infty \sum_{t_k \in T \setminus \{0\}} \left|K^{(2)}(t - t_k)\right| \\ &\quad + ||\beta||_\infty \left(\left|K^{(3)}(t)\right| + \sum_{t_k \in T \setminus \{0\}} \left|K^{(3)}(t - t_k)\right| \right) \\ &\leq -2.30\, f_{\text{lo}}^2. \end{aligned}$$

Similar computations yield $|q_0''(t)| \leq 4.07\, f_{\text{lo}}^2$. This together with $q_0(0) = 1$ and $q_0'(0) = 0$ implies the desired result. □

Lemma 8.3.2. *For any $t_l \in T \setminus \{t_j\}$ and t obeying $|t - t_l| \leq c\lambda_{\text{lo}}$, we have*

$$|q_{t_j}(t)| \leq 16.64\,(t - t_l)^2\, f_{\text{lo}}^2.$$

Proof. We assume without loss of generality that $t_l = 0$ and prove the claim for $t \in (0, c\,\lambda_{\text{lo}}]$. By (8.3.2), (8.3.3), (8.3.4), (8.3.5) and (8.3.6)

$$\begin{aligned} \left|q_{t_j}''(t)\right| &= \left| \sum_{t_k \in T} \alpha_k K^{(2)}(t - t_k) + \sum_{t_k \in T} \beta_k K^{(3)}(t - t_k) \right| \\ &\leq \left(1 + \left|\left|\alpha - e_{t_j}\right|\right|_\infty\right) \left|K^{(2)}(t - t_j)\right| \\ &\quad + \left|\left|\alpha - e_{t_j}\right|\right|_\infty \left(\left|K^{(2)}(t)\right| + \sum_{t_k \in T \setminus \{0, t_j\}} \left|K^{(2)}(t - t_k)\right| \right) \\ &\quad + ||\beta||_\infty \left(\left|K^{(3)}(t)\right| + \sum_{t_k \in T \setminus \{0\}} \left|K^{(3)}(t - t_k)\right| \right) \\ &\leq 16.64\, f_{\text{lo}}^2, \end{aligned}$$

since in the interval of interest $\left|K^{(2)}(t - t_j)\right| \leq \frac{18\pi^2}{f_{\text{lo}}^2(\Delta_{\min} - 0.16 f_{\text{lo}})^4} \leq 15.67\, f_{\text{lo}}^2$ due to Lemma 5.3.1. This together with $q_{t_j}(0) = 0$ and $q_{t_j}'(0) = 0$ implies the desired

result. □

Chapter 9

Conclusion

In this thesis, we have developed a framework to perform super-resolution based on convex programming. In particular, we have shown that we can super-resolve *events* such as point sources, discontinuity points, and so on with infinite precision from just a few low-frequency samples. This holds in any dimension provided that the distance between events is proportional to $1/f_{\text{lo}} = \lambda_{\text{lo}}$, where f_{lo} is the highest observed frequency; for instance, in one dimension, a sufficient condition is that the distance between events is at least $2\lambda_{\text{lo}}$. Furthermore, we have proved that when such condition holds, stable recovery is possible in the sense that we can (1) bound the approximation error at resolutions beyond λ_{lo} and (2) obtain guarantees on the accuracy of the recovered support.

Many interesting open questions on super-resolution via convex programming remain. We finish by listing some interesting research directions.

- *Conditions beyond minimum separation:* For signals that consist of small clusters of point sources, the minimum-separation condition is somewhat too strict, as convex-programming can often still be used to obtain an accurate estimate, even if the condition is violated. Because of this, it would be useful to derive recovery guarantees based on alternative conditions that limit the density of point sources, rather than just the minimum separation between them.

- *Minimum separation in multi-dimensional settings:* As discussed in Chapter 2,

in one dimension it seems clear that the minimum separation at which convex programming begins to fail to achieve exact recovery is the inverse of the cut-off frequency of the data. We can apply our techniques to prove that in multiple dimensions it is sufficient for the minimum separation to be proportional to the inverse of the cut-off frequency for exact recovery to take place, but the exact value of the constant is yet to be determined.

- *Super-resolution of curves:* Super-resolving curves or equivalently sharp edges is a fundamental problem in computer vision and signal processing. Unfortunately, it is more challenging than point-source super-resolution. For instance, it is unclear how to even define a condition in the spirit of minimum-separation under which to prove exact recovery.

- *Blind deconvolution:* Joint estimation of the signal of interest and the point-spread function of the sensing process is of great interest in applications where motion blur might be present. In computer vision, a common approach to this problem is to locate edges and then determine their orientation and the shape of the point-spread function simultaneously, so this problem is very much related to edge super-resolution.

Bibliography

[1] R. Adve. Direction of arrival estimation. Online notes available at `http://www.comm.utoronto.ca/~rsadve/Notes/DOA.pdf`.

[2] J.-M. Azais, Y. de Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 2014.

[3] N. S. Banerjee and J. F. Geer. Exponentially accurate approximations to periodic Lipschitz functions based on Fourier series partial sums. *J. Sci. Comput.*, 13:419–460, 1998.

[4] A. Barabell. Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 8, pages 336–339, 1983.

[5] I. Barrodale, C. A. Zala, and N. R. Chapman. Comparison of the ℓ_1 and ℓ_2 norms applied to one-at-a-time spike extraction from seismic traces. *Geophysics*, 49(11):2048–2052, 1984.

[6] D. Batenkov and Y. Yomdin. Algebraic Fourier reconstruction of piecewise smooth functions. *Math. Comput.*, 81(277), 2012.

[7] D. Batenkov and Y. Yomdin. On the accuracy of solving confluent Prony systems. *SIAM Journal on Applied Mathematics*, 73(1):134–154, 2013.

[8] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[9] S. Becker, J. Bobin, and E. J. Candès. NESTA: a fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39, 2011.

[10] S. R. Becker, E. J. Candès, and M. C. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218, 2011.

[11] T. Bendory, S. Dekel, and A. Feuer. Exact recovery of dirac ensembles from the projection onto spaces of spherical harmonics. Preprint.

[12] 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*, 182:7–17, 2014.

[13] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

[14] A. Beurling. Sur les intégrales de fourier absolument convergentes et leur application à une transformation fonctionelle. In *Proc. Scandinavian Math., Helsinki, Finland*, 1938.

[15] B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Transactions on Signal Processing*, 61(23):5987–5999, 2013.

[16] G. Bienvenu. Influence of the spatial coherence of the background noise on high resolution passive methods. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 306 – 309, 1979.

[17] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. *IEEE Signal Processing Magazine*, 25(2):31–40, 2008.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

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

[20] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 2012.

[21] J. A. Cadzow. Signal enhancement- a composite property mapping algorithm. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(1):49–62, 1988.

[22] P. Campisi and K. Egiazarian. *Blind image deconvolution: theory and applications*. CRC press, 2007.

[23] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956.

[24] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[25] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[26] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[27] E. J. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[28] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions in Information Theory*, 52:5406–5425, 2006.

[29] C. Carathéodory. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Mathematische Annalen*, 64(1):95–115, 1907.

[30] C. Carathéodory. Über den Variabilitätsbereich Fourierschen Konstanten von positiven harmonischen Funktionen. *Rend. Circ. Mat. Palermo*, 32:193–217, 1911.

[31] F. Champagnat, Y. Goussard, and J. Idier. Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE Transactions on Signal Processing*, 44(12):2988–2998, 1996.

[32] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Transactions on Signal Processing*, 59(5):2182–2195, 2011.

[33] J. F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(5):826–844, 1973.

[34] H. Clergeot, S. Tressens, and A. Ouamri. Performance of high resolution frequencies estimation methods compared to the Cramér-Rao bounds. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(11):1703–1720, 1989.

[35] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[36] Y. de Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354.

[37] Y. De Castro and G. Mijoule. Non-uniform spline recovery from small degree polynomial approximation. Preprint.

[38] D. L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.

[39] D. L. Donoho. Compressed sensing. *IEEE Transactions in Information Theory*, 52(4):1289 –1306, 2006.

[40] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49:906–931, 1989.

[41] D. L. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. In *Proceedings of the National Academy of Sciences*, pages 9446–9451, 2005.

[42] C. Dossal. *Estimation de fonctions géométriques et déconvolution.* PhD thesis, École Polytechnique, 2005.

[43] C. Dossal, G. Peyré, and J. Fadili. A numerical exploration of compressed sampling recovery. In *Proceedings of SPARS'09*, 2009.

[44] M. F. Duarte and R. G. Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, To appear.

[45] B. Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer, 2007.

[46] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. Preprint.

[47] K. S. Eckhoff. Accurate reconstructions of functions of finite regularity from truncated Fourier series expansions. *Math. Comput.*, 64:671–690, 1995.

[48] A. Fannjiang and W. Liao. MUSIC for single-snapshot spectral estimation: stability and super-resolution.

[49] A. Fannjiang and W. Liao. Coherence-pattern guided compressive sensing with unresolved grids. *SIAM Journal on Imaging Sciences*, 5:179, 2012.

[50] A. C. Fannjiang. The MUSIC algorithm for sparse objects: a compressed sensing analysis. *Inverse Problems*, 27(3):035013, 2011.

[51] C. Fernandez-Granda. Support detection in super-resolution. In *Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 145–148, 2013.

[52] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[53] J. J. Fuchs. Sparsity and uniqueness for some specific underdetermined linear systems. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 729–732, 2005.

[54] L. George. An introduction to matched filters. *IRE Transactions on Information Theory*, 6(3):311–329, 1960.

[55] A. Ghez, G. Duchêne, K. Matthews, S. Hornstein, A. Tanner, J. Larkin, M. Morris, E. Becklin, S. Salim, T. Kremenek, et al. The first measurement of spectral lines in a short-period star bound to the Galaxy's central black hole: a paradox of youth. *The Astrophysical Journal Letters*, 586(2):L127, 2003.

[56] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[57] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx`, 2011.

[58] H. Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52:43–63, 2009.

[59] F. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51 – 83, 1978.

[60] T. D. Harris, R. D. Grober, J. K. Trautman, and E. Betzig. Super-resolution imaging spectroscopy. *Applied Spectroscopy*, 48(1):14A–21A, 1994.

[61] C. Helstrom. The detection and resolution of optical signals. *IEEE Transactions on Information Theory*, 10(4):275 – 287, 1964.

[62] S. T. Hess, T. P. Girirajan, and M. D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258, 2006.

[63] J. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15:417, 1974.

[64] L. Hu, Z. Shi, J. Zhou, and Q. Fu. Compressed sensing of complex sinusoids: An approach based on dictionary refinement. *IEEE Transactions on Signal Processing*, 60(7):3809 –3822, 2012.

[65] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(5):814–824, 1990.

[66] J. P. Kahane. Analyse et synthèse harmoniques. Preprint.

[67] J. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari. Super-resolution in PET imaging. *IEEE Transactions on Medical Imaging*, 25(2):137 –147, 2006.

[68] V. Khaidukov, E. Landa, and T. J. Moser. Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. *Geophysics*, 69(6):1478–1490, 2004.

[69] J. J. Kormylo and J. M. Mendel. Maximum likelihood detection and estimation of Bernoulli-Gaussian processes. *IEEE Transactions on Information Theory*, 28(3):482–488, 1982.

[70] L. H. Lajunen and P. Perämäki. *Spectrochemical analysis by atomic absorption and emission*. Royal Society of Chemistry, 2004.

[71] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 28(5):1302–1053, 1992.

[72] S. Levy and P. K. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46(9):1235–1243, 1981.

[73] J. Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001, 2012.

[74] C. W. McCutchen. Superresolution in microscopy and the Abbe resolution limit. *Journal of the Optical Society of America*, 57(10):1190–1190, 1967.

[75] P. Milanfar, editor. *Super-Resolution Imaging*. Series: Digital Imaging and Computer Vision, 2010.

[76] J. Odendaal, E. Barnard, and C. Pistorius. Two-dimensional superresolution radar imaging using the MUSIC algorithm. *IEEE Transactions on Antennas and Propagation*, 42(10):1386 –1391, 1994.

[77] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21 – 36, 2003.

[78] V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysical Journal of the Royal Astronomical Society*, 33(3):347–366, 1973.

[79] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C (2nd ed.): the art of scientific computing*. 1992.

[80] R. Prony. Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'Ecole Polytechnique*, 1(2):24–76, 1795.

[81] K. G. Puschmann and F. Kneer. On super-resolution in astronomical imaging. *Astronomy and Astrophysics*, 436:373–378, 2005.

[82] L. Rayleigh. On pin-hole photography. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 31(189):87–99, 1891.

[83] F. Rieke. *Spikes: exploring the neural code*. The MIT Press, 1999.

[84] R. Rockafellar. *Conjugate Duality and Optimization*. Regional conference series in applied mathematics. Society for Industrial and Applied Mathematics, 1974.

[85] R. Roy and T. Kailath. ESPRIT- estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984 –995, 1989.

[86] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[87] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, 3rd edition, 1987.

[88] M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–796, 2006.

[89] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986.

[90] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276 – 280, 1986.

[91] A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3(1):13–41, 1898.

[92] M. Shahram and P. Milanfar. Imaging below the diffraction limit: a statistical analysis. *IEEE Transactions on Image Processing*, 13(5):677 –689, 2004.

[93] M. Shahram and P. Milanfar. On the resolvability of sinusoids with nearby frequencies in the presence of noise. *IEEE Transactions on Signal Processing*, 53(7):2579 – 2588, 2005.

[94] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case. *Bell System Technical Journal*, 57:1371–1430, 1978.

[95] L. F. Smith and L. H. Aller. On the classification of emission-line spectra of planetary nuclei. *The Astrophysical Journal*, 157:1245, 1969.

[96] P. Stoica and P. Babu. Sparse estimation of spectral lines: Grid selection problems and their solutions. *IEEE Transactions on Signal Processing*, 60(2):962 –967, 2012.

[97] P. Stoica, R. Moses, B. Friedlander, and T. Soderstrom. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(3):378–392, 1989.

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

[99] P. Stoica and A. Nehorai. Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case. *Circuits, Systems, and Signal Processing*, 8(1):3–15, 1989.

[100] P. Stoica and T. Soderstrom. Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies. *IEEE Transactions on Signal Processing*, 39(8):1836–1847, 1991.

[101] G. Tang, B. N. Bhaskar, and B. Recht. Near minimax line spectral estimation. Preprint.

[102] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.

[103] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[104] 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*, 11(1):545–581, 1999.

[105] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030 –1051, 2006.

[106] D. W. Tufts and R. Kumaresan. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE*, 70(9):975–989, 1982.

[107] E. Van Den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[108] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. John Wiley & Sons, 2004.

[109] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, 2002.

[110] B. Zhang, J. Zerubia, and J. C. Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819–1829, 2007.

[111] F. Zhang. *The Schur Complement and Its Applications*. Springer Science, 2005.

[112] L. Zhu, W. Zhang, D. Elnatan, and B. Huang. Faster STORM using compressed sensing. *Nature methods*, 9(7):721–723, 2012.

Carlos Fernandez-Granda

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Emmanuel J. Candès) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Abbas El Gamal)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Stephen P. Boyd)

Approved for the Stanford University Committee on Graduate Studies