Demixing Sines and Spikes: Robust Spectral Super-resolution in the Presence of Outliers

Carlos Fernandez-Granda*, Gongguo Tang†, Xiaodong Wang‡ and Le Zheng‡

September 2016

Abstract

We consider the problem of super-resolving the line spectrum of a multisinusoidal signal from a finite number of samples, some of which may be completely corrupted. Measurements of this form can be modeled as an additive mixture of a sinusoidal and a sparse component. We propose to demix the two components and super-resolve the spectrum of the multisinusoidal signal by solving a convex program. Our main theoretical result is that– up to logarithmic factors– this approach is guaranteed to be successful with high probability for a number of spectral lines that is linear in the number of measurements, even if a constant fraction of the data are outliers. The result holds under the assumption that the phases of the sinusoidal and sparse components are random and the line spectrum satisfies a minimum-separation condition. We show that the method can be implemented via semidefinite programming and explain how to adapt it in the presence of dense perturbations, as well as exploring its connection to atomic-norm denoising. In addition, we propose a fast greedy demixing method which provides good empirical results when coupled with a local nonconvex-optimization step.

Keywords. Atomic norm, continuous dictionary, convex optimization, greedy methods, line spectra estimation, outliers, semidefinite programming, sparse recovery, super-resolution.

1 Introduction

The goal of *spectral super-resolution* is to estimate the spectrum of a multisinusoidal signal from a finite number of samples. This is a problem of crucial importance in signal-processing applications, such as target identification from radar measurements [3, 21], digital filter design [59], underwater acoustics [2], seismic imaging [6], nuclear-magnetic-resonance spectroscopy [72] and power electronics [43]. In this paper, we study spectral super-resolution in the presence of perturbations that completely corrupt a subset of the data. The corrupted samples can be interpreted as *outliers* that do not follow the same multisinusoidal model as the rest of the measurements and complicate significantly the task of super-resolving the spectrum of the signal of interest. Depending on the application, outliers may appear due to sensor failures, interference from other signals or impulsive noise. For instance, radar measurements can be corrupted by lightning discharges, spurious radio emissions or telephone switching transients [36, 45].

*Courant Institute of Mathematical Sciences and Center for Data Science, NYU, New York, NY
†Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO
‡Electrical Engineering Department, Columbia University, New York, NY

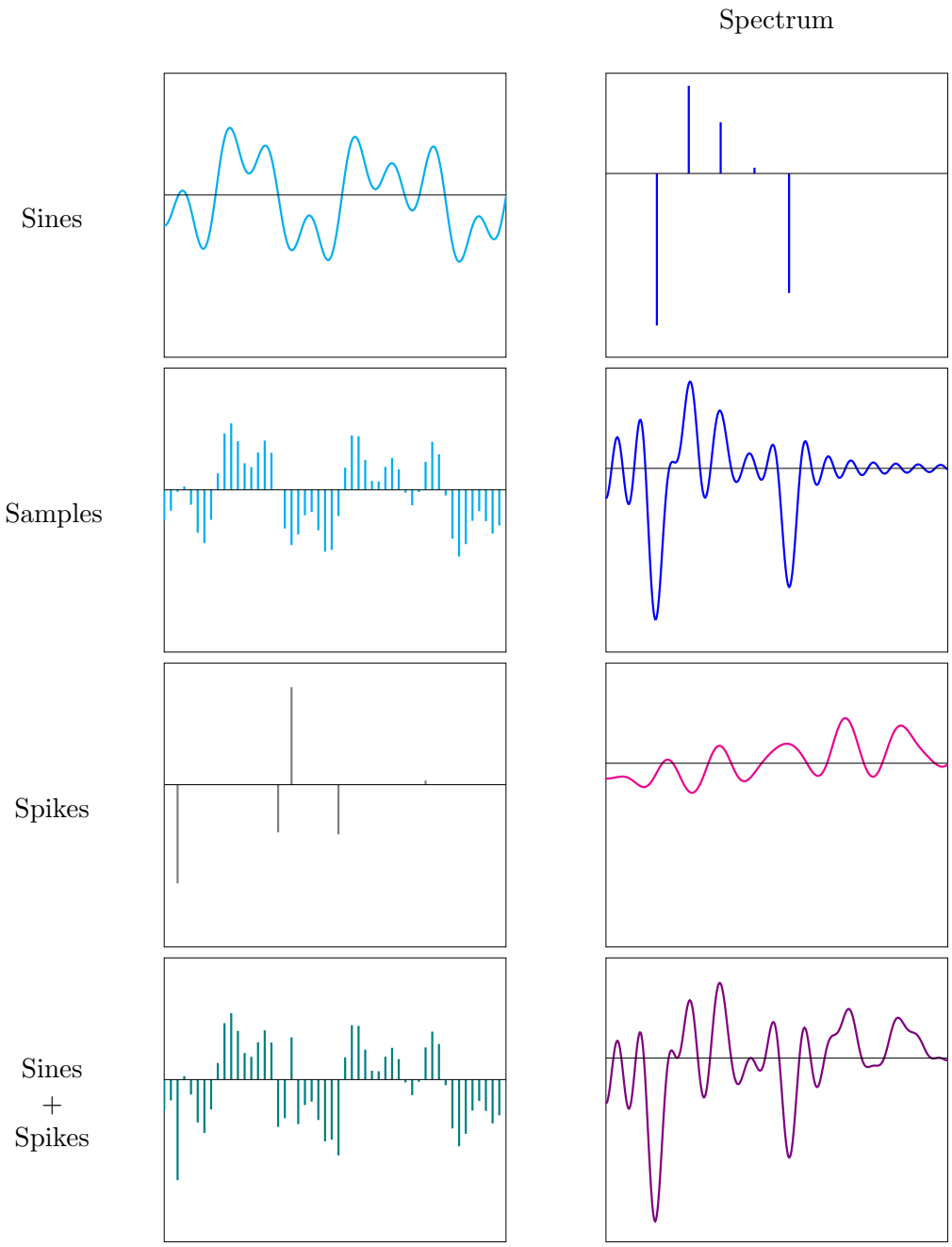

Figure 1: The top row shows a multisinusoidal signal (left) and its sparse spectrum (right). The minimum separation of the spectrum is $2.8/(n-1)$ (see Section 2.2). On the second row, truncating the signal to a finite interval after measuring $n := 101$ samples at the Nyquist rate (left) results in aliasing in the frequency domain (right). The third row shows some impulsive noise (left) and its corresponding spectrum (right). The last row shows the superposition of the multisinusoidal signal and the sparse noise, which yields a mixture of *sines* and *spikes* depicted in the time (left) and frequency domains (right). For ease of visualization, the amplitudes of the spectrum of the sines and of the spikes are real (we only show half of the spectrum and half of the spikes because their amplitudes and positions are symmetric).

Figure 1 illustrates the problem of performing spectral super-resolution in the presence of outliers. The top row shows a superposition of sinusoids and its corresponding sparse spectrum. In the second row, the multisinusoidal signal is sampled at the Nyquist rate over a finite interval, which induces spectral aliasing and makes it challenging to resolve the individual spectral lines. The sparse signal in the third row represents an additive perturbation that corrupts some of the samples. Finally, the bottom row shows the available measurements: a mixture of *sines* (samples from the multisinusoidal signal) and *spikes* (the sparse perturbation). Our objective is to *demix* these two components and super-resolve the spectrum of the sines.

Broadly speaking, there are three main approaches to spectral super-resolution: linear nonparametric methods [62], techniques based on Prony's method [27, 62] and optimization-based methods [4, 38, 65]. The first three rows of Figure 2 show the results of applying a representative of each approach to a spectral super-resolution problem when there are no outliers in the data (left column) and when there are (right column).

In the absence of corruptions, the periodogram– a linear nonparametric technique that uses windowing to reduce spectral aliasing [41]– locates most of the relevant frequencies, albeit at a coarse resolution. In contrast, both the Prony-based approach– represented by the Multiple Signal Classification (MUSIC) algorithm [5, 57]– and the optimization-based method– based on total-variation norm minimization [4, 13, 65]– recover the true spectrum of the signal perfectly. All of these techniques are designed to allow for small Gaussian-like perturbations to the data and hence their performance degrades gracefully when such noise is present (not shown in the figure). However, as we can see in the right column of Figure 2, when outliers are present in the data their performance is severely affected: none of the methods detect the fourth spectral line of the signal and they all hallucinate two large spurious spectral lines to the right of the true spectrum.

The subject of this paper is an optimization-based method that leverages sparsity-inducing norms to perform spectral super-resolution and simultaneously detect outliers in the data. The bottom row of Figure 2 shows that this approach is capable of super-resolving the spectrum of the multisinusoidal signal in Figure 1 exactly from the corrupted measurements, in contrast to techniques that do not account for the presence of outliers in the data. Below is a brief roadmap of the paper.

- Section 2 describes our methods and main results. In Section 2.1 we introduce a mathematical model of the spectral super-resolution problem. Section 2.2 justifies the need of a minimum-separation condition on the spectrum of the signal for spectral super-resolution to be well posed. In Section 2.3 we present our optimization-based method and provide a theoretical characterization of its performance. Section 2.4 discusses the robustness of the technique to the choice of regularization parameter. Section 2.5 explains how to adapt the method when the data are perturbed by dense noise. Section 2.6 establishes a connection between our method and atomic-norm denoising. Finally, in Section 2.7 we review the related literature.

- Our main theoretical contribution– Theorem 2.2– establishes that solving the convex program introduced in Section 2.3 allows to super-resolve up to k spectral lines exactly in the presence of s outliers (i.e. when s measurements are completely corrupted) with high probability from a number of data that is linear both in k and s up to logarithmic factors. Section 3 is dedicated to the proof of this result, which is non-asymptotic and holds under several assumptions that are described in Section 2.3.

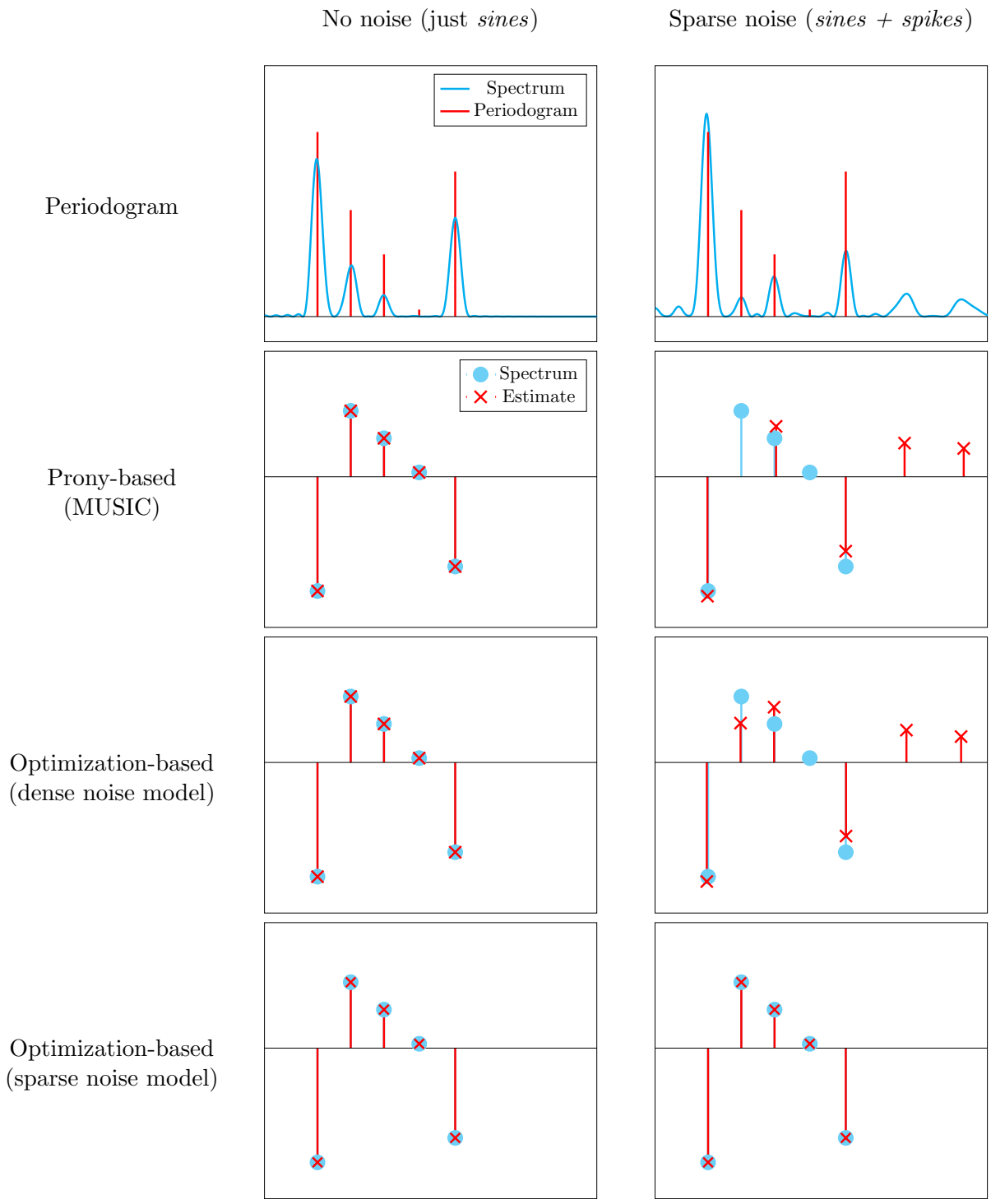

Figure 2: Estimate of the sparse spectrum of the multisinusoidal signal from Figure 1 when outliers are absent from the data (left column) and when they are present (right column). The estimates are shown in red; the true location of the spectra is shown in blue. Methods that do not account for outliers fail to recover all the spectral lines when impulsive noise corrupts the data, whereas an optimization-based estimator incorporating a sparse-noise model still achieves exact recovery.

- Section 4 focuses on demixing algorithms. In Sections 4.1 and 4.2 we explain how to implement the methods discussed in Sections 2.3 and 2.5 respectively by recasting the dual of the corresponding optimization problems as a tractable semidefinite program. In Section 4.3 we propose a greedy demixing technique that achieves good empirical results when combined with a local nonconvex-optimization step. Section 4.4 describes the implementation of atomic-norm denoising in the presence of outliers using semidefinite programming. Matlab code of all the algorithms discussed in this section is available online[1].
- Section 5 reports numerical experiments illustrating the performance of the proposed approach. In Section 5.1 we investigate under what conditions our optimization-based method achieves exact demixing empirically. In Section 5.2 we compare atomic-norm denoising to an alternative approach based on matrix completion.
- We conclude the paper outlining several future research directions in Section 6.

2 Robust spectral super-resolution via convex programming

2.1 Mathematical model

We model the multisinusoidal signal of interest as a superposition of k complex exponentials

$$g(t) := \sum_{j=1}^{k} \boldsymbol{x}_j \exp(i2\pi f_j t), \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{C}^k$ is the vector of complex amplitudes and $\boldsymbol{x}_j$ is its jth entry. The spectrum of g consists of spectral lines, modeled by Dirac measures that are supported on a subset $T := \{f_1, \ldots, f_k\}$ of the unit interval $[0, 1]$

$$\mu = \sum_{f_j \in T} \boldsymbol{x}_j \, \delta(f - f_j), \tag{2.2}$$

where $\delta(f - f_j)$ denotes a Dirac measure located at f_j. Sparse spectra of this form are often called *line spectra* in the literature. Note that a simple change of variable allows to apply this model to signals with spectra restricted to any interval $[f_{\min}, f_{\max}]$.

By the Nyquist-Shannon sampling theorem we can recover g, and consequently μ, from an infinite sequence of regularly spaced samples $\{g(l),\ l \in \mathbb{Z}\}$ by sinc interpolation. The aim of spectral super-resolution is to estimate the support of the line spectrum T and the amplitude vector $\boldsymbol{x}$ from a *finite* set of n contiguous samples instead. Note that $\{g(l),\ l \in \mathbb{Z}\}$ are the Fourier-series coefficients of μ, so mathematically we seek to recover an atomic measure from a subset of its Fourier coefficients. As described in the introduction, we are interested in tackling this problem when a subset of the data is completely corrupted. These corruptions are modeled as additive impulsive noise, represented by a sparse vector $\boldsymbol{z} \in \mathbb{C}^n$ with s nonzero entries. The data are consequently of the form

$$\boldsymbol{y}_l = g(l) + \boldsymbol{z}_l, \quad 1 \leq l \leq n. \tag{2.3}$$

[1] http://www.cims.nyu.edu/~cfgranda/scripts/spectral_superres_outliers.zip

To represent the measurement model more compactly, we define an operator $\mathcal{F}_n$ that maps a measure to its first n Fourier series coefficients,

$$\boldsymbol{y} = \mathcal{F}_n \, \mu + \boldsymbol{z}. \tag{2.4}$$

Intuitively, $\mathcal{F}_n$ maps the spectrum μ to n regularly spaced samples of the signal g in the time domain.

2.2 Minimum-separation condition

Even in the absence of any noise, the problem of recovering a signal from n samples is vastly underdetermined: we can fill in the missing samples $g\left(0\right), g\left(-1\right), \ldots$ and $g\left(n+1\right), g\left(n+2\right), \ldots$ any way we like and then apply sinc interpolation to obtain an estimate that is consistent with the data. For the inverse problem to make sense we need to leverage additional assumptions about the structure of the signal. In spectral super-resolution the usual assumption is that the spectrum of the signal is sparse. This is reminiscent of compressed sensing [17], where signals are recovered robustly from randomized measurements by exploiting a sparsity prior.

A crucial insight underlying compressed-sensing theory is that the randomized operator obeys the *restricted-isometry property* (RIP), which ensures that the measurements preserve the energy of any sparse signal with high probability [18]. Unfortunately, this is not the case for our measurement operator of interest. The reason is that signals consisting of clustered spectral lines may lie almost in the null space of the sampling operator, even if the number of spectral lines is small. Additional conditions beyond sparsity are necessary to ensure that the problem is well posed. To this end, we define the *minimum separation* of the support of a signal, as introduced in [12].

Definition 2.1 (Minimum separation)**.** *For a set of points $T \subset [0, 1]$, the minimum separation (or minimum distance) is defined as the closest distance between any two elements from T,*

$$\Delta(T) = \inf_{(f_1, f_2) \in T \,:\, f_1 \neq f_2} |f_2 - f_1|. \tag{2.5}$$

To be clear, this is the wrap-around distance so that the distance between $f_1 = 0$ and $f_2 = 3/4$ is equal to $1/4$.

If the minimum distance is too small with respect to the number of measurements then it may be impossible to resolve a signal even under very small levels of noise. A fundamental limit in this sense is $\Delta^* := \frac{2}{n-1}$, which is the width of the main lobe of the periodized sinc kernel that is convolved with the spectrum when we truncate the number of samples to n. This limit arises because for minimum separations just below $\Delta^*/2$ there exist signals that are *almost* suppressed by the sampling operator $\mathcal{F}_n$. If such a signal d corresponds to the difference between two different signals s_1 and s_2 so that $s_1 - s_2 = d$, it will be very challenging to distinguish s_1 and s_2 from the available data[2]. This phenomenon can be characterized theoretically in an asymptotic setting using Slepian's prolate-spheroidal sequences [58] (see also Section 3.2 in [12]). More recently, Theorem 1.3 of [49] provides a non-asymptotic analysis and other works have obtained lower bounds on the minimum separation necessary for convex-programming approaches to succeed [33, 64].

[2]For a concrete example of two signals with a minimum separation of $0.9\Delta^*$ that are almost indistinguishable from data consisting of $n = 2\,10^3$ samples see Figure 2 of [38]

2.3 Robust spectral super-resolution via convex programming

Spectral super-resolution in the presence of outliers boils down to estimating μ and $\boldsymbol{z}$ in the mixture model (2.4). Without additional constraints, this is not very ambitious: data consistency is trivially achieved, for instance, by setting the sines to zero and declaring every sample to be a spike. Our goal is to fit the two components *in the simplest way possible*, i.e. so that the spectrum of the multisinusoidal signal– the *sines*– is restricted to a small number of frequencies and the impulsive noise– the *spikes*– only affects a small subset of the data.

Many modern signal-processing methods rely on the design of cost functions that (1) encode prior knowledge about signal structure and (2) can be minimized efficiently. In particular, penalizing the ℓ_1-norm is an efficient and robust method for obtaining sparse estimates in denoising [24], regression [69] and inverse problems such as compressed sensing [19, 28]. In order to fit a mixture model where both the spikes and the spectrum of the sines are sparse, we propose minimizing a cost function that penalizes the ℓ_1 norm of both components (or rather a continuous counterpart of the ℓ_1 norm in the case of the spectrum, as we explain below). We would like to note that this approach was introduced by some of the authors of the present paper in [38, 68], but without any theoretical analysis, and applied to multiple-target tracking from radar measurements in [75]. Similar ideas have been previously leveraged to separate low-rank and sparse matrices [14, 23], perform compressed sensing from corrupted data [44] and demix signals that are sparse in different bases [48].

Recall that the spectrum of the sinusoidal component in our mixture model is modeled as a measure that is supported on a continuous interval. Its ℓ_1-norm is therefore not well defined. In order to promote sparsity in the estimate, we resort instead to a continuous version of the ℓ_1 norm: the total-variation (TV) norm[3]. If we consider the space of measures supported on the unit interval, this norm is dual to the infinity norm, so that

$$||\nu||_{\mathrm{TV}} := \sup_{||h||_\infty \leq 1, h \in C(\mathbb{T})} \mathrm{Re}\left[\int_{\mathbb{T}} \overline{h\left(f\right)} \nu\left(\mathrm{d}f\right)\right]. \tag{2.6}$$

for any measure ν (for a different definition see Section A in the appendix of [12]). In the case of a superposition of Dirac deltas as in (2.2), the total-variation norm is equal to the ℓ_1 norm of the coefficients, i.e. $||\mu||_{\mathrm{TV}} = ||\boldsymbol{x}||_1$. Spectral super-resolution via TV-norm minimization, introduced in [12, 26] (see also [11]), has been shown to achieve exact recovery under a minimum separation of $\frac{2.52}{n-1}$ in [38] and to be robust to missing data in [66].

Our proposed method minimizes the sum of the ℓ_1 norm of the spikes and the TV norm of the spectrum of the sines subject to a data-consistency constraint:

$$\min_{\tilde{\mu}, \tilde{\boldsymbol{z}}} ||\tilde{\mu}||_{\mathrm{TV}} + \lambda ||\tilde{\boldsymbol{z}}||_1 \quad \text{subject to} \quad \mathcal{F}_n \, \tilde{\mu} + \tilde{\boldsymbol{z}} = \boldsymbol{y}. \tag{2.7}$$

$\lambda > 0$ is a regularization parameter that governs the weight of each penalty term. This optimization program is convex. Section 4.1 explains how to solve it by reformulating its dual as a semidefinite program. Our main theoretical result is that solving (2.7) achieves perfect demixing with high probability under certain assumptions.

[3] *Total variation* often also refers to the ℓ_1 norm of the discontinuities of a piecewise-constant function, which is a popular regularizer in image processing and other applications [55].

Theorem 2.2 (Proof in Section 3)**.** *Suppose that we observe n samples of the form*

$$\boldsymbol{y} = \mathcal{F}_n \, \mu + \boldsymbol{z}, \tag{2.8}$$

where each entry in $\boldsymbol{z}$ is nonzero with probability $\frac{s}{n}$ (independently of each other) and the support $T := \{f_1, \ldots, f_k\}$ of

$$\mu := \sum_{j=1}^{k} \boldsymbol{x}_j \delta \left(f - f_j \right), \tag{2.9}$$

has a minimum separation lower bounded by

$$\Delta_{\min} := \frac{2.52}{n-1}. \tag{2.10}$$

If the phases of the entries in $\boldsymbol{x} \in \mathbb{C}^k$ and the nonzero entries in $\boldsymbol{z} \in \mathbb{C}^n$ are iid random variables uniformly distributed in $[0, 2\pi]$, then the solution to Problem (2.7) with $\lambda = 1/\sqrt{n}$ is exactly equal to μ and $\boldsymbol{z}$ with probability $1 - \epsilon$ for any $\epsilon > 0$ as long as

$$k \leq C_k \left(\log \frac{n}{\epsilon} \right)^{-2} n, \tag{2.11}$$

$$s \leq C_s \left(\log \frac{n}{\epsilon} \right)^{-2} n, \tag{2.12}$$

for fixed numerical constants C_k and C_s and $n \geq 2 \times 10^3$.

The theorem guarantees that our method is able to super-resolve a number of spectral lines that is proportional to the number of measurements, even if the data contain a *constant fraction of outliers*, up to logarithmic factors. The proof is presented in Section 3; it is based on the construction of a random trigonometric polynomial that certifies exact demixing. Our result is non-asymptotic and holds with high probability under several assumptions, which we now discuss in more detail.

- The support of the sparse corruptions follows a Bernoulli model where each entry is nonzero with probability s/n independently from each other. This model is essentially equivalent to choosing the support of the outliers uniformly at random from all possible subsets of cardinality s, as shown in Section 7.1 of [14] (see also [17, Section 2.3] and [20, Section 8.1]).
- The phases of the amplitudes of the spectral lines are assumed to be iid uniform random variables (note however that the amplitudes can take any value). Modeling the phase of the spectral components of a multisinusoidal signal in this way is a common assumption in signal processing, see for example [62, Chapter 4.1].
- The phases of the amplitudes of the additive corruptions are also assumed to be iid uniform random variables (the amplitudes can again take any value). If we constrain the corruptions to be real, the derandomization argument in [14, Section 2.2] allows to obtain guarantees for arbitrary sign patterns.
- We have already discussed the minimum-separation condition on the spectrum of the multisinusoidal component in Section 2.2.

Our assumptions model a non-adversarial situation where the outliers are not designed to cancel out the samples from the multisinusoidal signal. In the absence of any such assumption it is possible to concoct instances for which the demixing problem is ill posed, even if the number of spectral lines and outliers is small. We illustrate this with a simple example, based on the *picket-fence* sequence used as an extremal function for signal-decomposition uncertainty principles in [29, 30]. Consider k' spectral lines with unit amplitudes with an equispaced support

$$\mu' := \frac{1}{k'} \sum_{j=0}^{k'-1} \delta\left(f - j/k'\right). \tag{2.13}$$

The samples of the corresponding multisinusoidal signal g' are zero except at multiples of k'

$$g'(l) = \begin{cases} 1 & \text{if } l/k' \in \mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

If we choose the corruptions $\boldsymbol{z'}$ to cancel out these nonzero samples

$$\boldsymbol{z'}_l = \begin{cases} -1 & \text{if } l/k' \in \mathbb{Z}, \\ 0 & \text{otherwise,} \end{cases} \tag{2.15}$$

then the corresponding measurements are all equal to zero! For these data the demixing problem is obviously impossible to solve by any method. Set $k' := \sqrt{n}$ so that the number of measurements n equals $(k')^2$. Then the number of outliers is just $n/k' = \sqrt{n}$ and the minimum separation between the spikes is $1/\sqrt{n}$, which amply satisfies the minimum-separation condition 2.10. This shows that additional assumptions beyond the minimum-separation condition are necessary for the inverse problem to make sense. A related phenomenon arises in compressed sensing, where random measurement schemes avoid similar adversarial situations (see [17, Section 1.3] and [70]). An interesting subject for future research is whether it is possible to establish the guarantees for exact demixing provided by Theorem 2.2 without random assumptions on the phase of the different components, or if these assumptions are necessary for the demixing problem to be well posed.

2.4 Regularization parameter

A question of practical importance is whether the performance of our demixing method is robust to the choice of the regularization parameter λ in Problem (2.7). Theorem 2.2 indicates that this is the case in the following sense. If we set λ to a fixed value that is proportional to $1/\sqrt{n}$[4], then exact demixing occurs for a number of spectral lines k and a number of outliers s that range from zero to a certain maximum value proportional to n (up to logarithmic factors).

In this section we provide additional theoretical evidence for the robustness of our method to the choice of λ. If exact recovery occurs for a certain pair $\{\mu, \boldsymbol{z}\}$ and a certain λ then it will also succeed for *any trimmed version* $\{\mu', \boldsymbol{z'}\}$ (obtained by removing some elements of the support of μ or $\boldsymbol{z}$, or both) for *the same value of* λ.

[4]To be precise, Theorem 2.2 assumes $\lambda := 1/\sqrt{n}$, but one can check that the whole proof goes through if we set λ to $c/\sqrt{n}$ for any positive constant c. The only effect is a change in the constants C_s and C_k in (2.11) and (2.12).

Lemma 2.3 (Proof in Section A)**.** *Let $\boldsymbol{z}$ be a vector with support Ω and let μ be an arbitrary measure such that*

$$\boldsymbol{y} = \mathcal{F}_n \, \mu + \boldsymbol{z}. \tag{2.16}$$

Assume that the pair $\{\mu, \boldsymbol{z}\}$ is the unique solution to Problem (2.7) and consider the data

$$\boldsymbol{y'} = \mathcal{F}_n \, \mu' + \boldsymbol{z'}. \tag{2.17}$$

μ' is a trimmed version of μ: it is equal to μ on a subset of its support $T' \subseteq T$ and is zero everywhere else. Similarly, $\boldsymbol{z'}$ equals $\boldsymbol{z}$ on a subset of entries $\Omega' \subseteq \Omega$ and is zero otherwise. For any choice of T' and Ω', $\{\mu, \boldsymbol{z'}\}$ is the unique solution to Problem (2.7) if we set the data vector to equal $\boldsymbol{y'}$ for the same value of λ.

This result and its proof are inspired by Theorem 2.2 in [14]. As illustrated by Figures 12 and 13, our numerical experiments corroborate the lemma: we consistently observe that if exact demixing occurs for most signals with a certain number of spectral lines and outliers, then it also occurs for most signals with less spectral lines and less corruptions (as long as the minimum separation is the same) for a fixed value of λ.

2.5 Stability to dense perturbations

One of the advantages of our optimization-based framework is that we can account for additional assumptions on the problem structure by modifying either the cost function or the constraints of the optimization problem used to perform demixing. In most applications of spectral super-resolution, the data will deviate from the multisinusoidal model (2.1) because of measurement noise and other perturbations, even in the absence of outliers. We model such deviations as a dense additive perturbation $\boldsymbol{w}$, such that $||\boldsymbol{w}||_2 \leq \sigma$ for a certain noise level σ,

$$\boldsymbol{y} = \mathcal{F}_n \, \mu + \boldsymbol{z} + \boldsymbol{w}. \tag{2.18}$$

Problem (2.7) can be adapted to this measurement model by relaxing the equality constraint that enforces data consistency to an inequality which takes into account the noise level

$$\min_{\tilde{\mu}, \tilde{\boldsymbol{z}}} ||\tilde{\mu}||_{\mathrm{TV}} + \lambda \, ||\tilde{\boldsymbol{z}}||_1 \quad \text{subject to} \quad ||\boldsymbol{y} - \mathcal{F}_n \, \tilde{\mu} + \tilde{\boldsymbol{z}}||_2 \leq \sigma. \tag{2.19}$$

Just like Problem (2.7), this optimization problem can be solved by recasting its dual as a tractable semidefinite program, as we explain in detail in Section 4.2.

2.6 Atomic-norm denoising

Our demixing method is closely related to atomic-norm denoising of multisinusoidal samples. Consider the n-dimensional vector $\boldsymbol{g} := \mathcal{F}_n \, \mu$ containing *clean* samples from a signal g defined by (2.1). The assumption that the spectrum μ of g consists of k spectral lines is equivalent to $\boldsymbol{g}$ having

a sparse representation in an infinite dictionary of n-dimensional sinusoidal *atoms* $\boldsymbol{a}(f,\phi) \in \mathbb{C}^n$ parameterized by frequency $f \in [0,1)$ and phase $\phi \in [0,2\pi)$,

$$\boldsymbol{a}(f,\phi)_l := \frac{1}{\sqrt{n}} e^{i\phi} e^{i2\pi l f}, \quad 1 \leq l \leq n. \tag{2.20}$$

Indeed, $\boldsymbol{g}$ can be expressed as a linear combination of k atoms

$$\boldsymbol{g} = \sqrt{n} \sum_{j=1}^{k} |\boldsymbol{x}_j| \, \boldsymbol{a}(f_j,\phi_j), \quad \boldsymbol{x}_j := |\boldsymbol{x}_j| \, e^{i2\pi\phi_j}. \tag{2.21}$$

This representation can be leveraged in an optimization framework using the atomic norm, an idea introduced in [22] and first applied to spectral super-resolution in [4]. The atomic norm induced by a set of atoms $\mathcal{A}$ is equal to the gauge of $\mathcal{A}$ defined by

$$||\boldsymbol{u}||_{\mathcal{A}} := \inf \{t > 0 : \boldsymbol{u} \in t \operatorname{conv}(\mathcal{A})\}, \tag{2.22}$$

which is a norm as long as $\mathcal{A}$ is centrally symmetric around the origin (as is the case for (2.20)). Geometrically, the unit ball of the atomic norm is the convex hull of the atoms in $\mathcal{A}$, just like the ℓ_1-norm ball is the convex hull of unit-norm one-sparse vectors. As a result, signals consisting of a small number of atoms tend to have a smaller atomic norm (just like sparse vectors tend to have a smaller ℓ_1-norm).

Consider the problem of denoising the samples of g from corrupted data of the form (2.4),

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{z}. \tag{2.23}$$

To be clear, the aim is now to separate $\boldsymbol{g}$ from the corruption vector $\boldsymbol{z}$ instead of directly estimating the spectrum of $\boldsymbol{g}$. In order to demix the two signals we penalize the atomic norm of the multisinusoidal component and the ℓ_1 norm of the sparse component,

$$\min_{\tilde{\boldsymbol{g}},\tilde{\boldsymbol{z}}} \frac{1}{\sqrt{n}} ||\tilde{\boldsymbol{g}}||_{\mathcal{A}} + \lambda ||\tilde{\boldsymbol{z}}||_1 \quad \text{subject to} \quad \tilde{\boldsymbol{g}} + \tilde{\boldsymbol{z}} = \boldsymbol{y}, \tag{2.24}$$

where $\lambda > 0$ is a regularization parameter.

Problems 2.19 and 2.24 are closely related. Their convex cost functions are designed to exploit sparsity assumptions on the spectrum of g and on the corruption vector $\boldsymbol{z}$ in ways that are essentially equivalent. More formally, both problems have the same dual, as implied by the following lemma and Lemma 4.1.

Lemma 2.4 (Proof in Section B.1)**.** *The dual of Problem* (2.24) *is*

$$\max_{\boldsymbol{\eta} \in \mathbb{C}^n} \langle \boldsymbol{y}, \boldsymbol{\eta} \rangle \quad \textit{subject to} \quad ||\mathcal{F}_n^* \boldsymbol{\eta}||_\infty \leq 1, \tag{2.25}$$

$$||\boldsymbol{\eta}||_\infty \leq \lambda, \tag{2.26}$$

where the inner product is defined as $\langle \boldsymbol{y}, \boldsymbol{\eta} \rangle := \operatorname{Re}(\boldsymbol{y}^* \boldsymbol{\eta})$.

The fact that the two optimization problems share the same dual has an important consequence established in Section B.2: the same dual certificate can be used to prove that they achieve exact demixing. As a result, the proof of Theorem 2.2 immediately implies that solving Problem (2.24) is successful in separating $\boldsymbol{g}$ and $\boldsymbol{z}$ under the conditions described in Section 2.3.

Corollary 2.5 (Proof in Section B.2)**.** *Under the assumptions of Theorem 2.2,* $\boldsymbol{g} := \mathcal{F}_n \mu$ *and* $\boldsymbol{z}$ *are the unique solutions to Problem* (2.24)*.*

Problem (2.24) can be adapted to denoise data that is perturbed by both outliers and dense noise, which follows the measurement model (2.18). Inspired by previous work on line-spectra denoising via atomic-norm minimization [4, 65], we remove the equality constraint and add a regularization term to ensure consistency with the data,

$$\min_{\tilde{\boldsymbol{g}},\tilde{\boldsymbol{z}}} \frac{1}{\sqrt{n}} \left|\left|\tilde{\boldsymbol{g}}\right|\right|_{\mathcal{A}} + \lambda \left|\left|\tilde{\boldsymbol{z}}\right|\right|_1 + \frac{\gamma}{2} \left|\left|\boldsymbol{y} - \tilde{\boldsymbol{g}} - \tilde{\boldsymbol{z}}\right|\right|_2^2, \tag{2.27}$$

where $\gamma > 0$ is a regularization parameter with a role analogous to σ in Problem (2.19).

In Section 4.4 we discuss how to implement atomic-norm denoising by reformulating Problems 2.24 and 2.27 as semidefinite programs.

2.7 Related work

Most previous works analyzing the problem of demixing sines and spikes make the assumption that the frequencies of the sinusoidal component lie on a grid with step size $1/n$, where n is the number of samples. In that case, demixing reduces to a discrete sparse decomposition problem in a dictionary formed by the concatenation of an identity and a discrete-Fourier-transform matrix [30]. Bounds on the coherence of this dictionary can be used to derive guarantees for basis pursuit [29] and also techniques based on Prony's method [31]. Coherence-based bounds do not reflect the fact that most sparse subsets of the dictionary are well conditioned [70], which can be exploited to obtain stronger guarantees for ℓ_1-norm based methods under random assumptions [44, 63]. In this paper we depart from this previous literature by considering a sinusoidal component whose spectrum may lie on arbitrary points of the unit interval.

Our work draws from recent developments on the super-resolution of point sources and line spectra via convex optimization. In [12] (see also [26]), the authors establish that TV minimization achieves exact recovery of measures satisfying a minimum separation of $\frac{4}{n-1}$, a result that is sharpened to $\frac{2.52}{n-1}$ in [38]. In [66] the method is adapted to a compressed-sensing setting, where a large fraction of the measurements may be missing. The proof of Theorem 2.2 builds upon the techniques developed in [12, 38, 66]. We would like to point out that stability guarantees for TV-norm-based approaches established in subsequent works [1, 13, 33, 37, 65] hold only for small perturbations and do not apply when the data may be perturbed by sparse noise of arbitrary amplitude, as is the case in this paper.

In [25], a spectral super-resolution approach based on robust low-rank matrix recovery is shown to be robust to outliers under some incoherence assumptions, which are empirically related to our minimum-separation condition (see Section A in [25]). Ignoring logarithmic factors, the guarantees in [25] allow for exact denoising of up to $\mathcal{O}\left(\sqrt{n}\right)$ spectral lines in the presence of $\mathcal{O}\left(n\right)$ outliers, where n is the number of measurements. Corollary 2.5, which follows from our main result Theorem 2.2, establishes that our approach succeeds in denoising up to $\mathcal{O}\left(n\right)$ spectral lines also in the presence of $\mathcal{O}\left(n\right)$ outliers (again ignoring logarithmic factors). In Section 5.2 we compare both techniques empirically. Finally, we would like to mention another method exploiting optimization and low-rank matrix structure [74] and an alternative approach to gridless spectral super-resolution [60], which

has been recently adapted to account for missing data and impulsive noise [73]. In both cases, no theoretical results guaranteeing exact recovery in the presence of outliers are provided.

3 Proof of Theorem 2.2

3.1 Dual polynomial

We prove Theorem 2.2 by constructing a trigonometric polynomial whose existence certifies that solving Problem (2.7) achieves exact demixing. We refer to this object as a *dual polynomial*, because its vector of coefficients is a solution to the dual of Problem (2.7). This vector is known as a *dual certificate* in the compressed-sensing literature [17].

Proposition 3.1 (Proof in Section C)**.** *Let $T \subset [0,1]$ be the nonzero support of μ and $\Omega \subset \{1,2,\ldots,n\}$ the nonzero support of $\boldsymbol{z}$. If there exists a trigonometric polynomial of the form*

$$Q(f) = \mathcal{F}_n^* \boldsymbol{q} \tag{3.1}$$

$$= \sum_{j=1}^{n} \boldsymbol{q}_j \, e^{-i2\pi jf}, \tag{3.2}$$

which satisfies

$$Q(f_j) = \frac{\boldsymbol{x}_j}{|\boldsymbol{x}_j|}, \qquad \forall f_j \in T, \tag{3.3}$$

$$|Q(f)| < 1, \qquad \forall f \in T^c, \tag{3.4}$$

$$\boldsymbol{q}_l = \lambda \frac{\boldsymbol{z}_l}{|\boldsymbol{z}_l|}, \qquad \forall l \in \Omega, \tag{3.5}$$

$$|\boldsymbol{q}_l| < \lambda, \qquad \forall l \in \Omega^c, \tag{3.6}$$

then $(\mu, \boldsymbol{z})$ is the unique solution to Problem 2.7 as long as $k + s \leq n$.

The dual polynomial can be interpreted as a subgradient of the TV norm at the measure μ, in the sense that

$$||\mu + \nu||_{\mathrm{TV}} \geq ||\mu||_{\mathrm{TV}} + \langle Q, \nu \rangle, \quad \langle Q, \nu \rangle := \mathrm{Re}\left[\int_{[0,1]} \overline{Q(f)} \, \mathrm{d}\, \nu(f) \right], \tag{3.7}$$

for any measure ν supported in the unit interval. In addition, weighting the coefficients of Q by $1/\lambda$ yields a subgradient of the ℓ_1 norm at the vector $\boldsymbol{z}$. This means that for any other feasible pair $(\mu', \boldsymbol{z}')$ such that $\boldsymbol{y} = \mathcal{F}_n \mu' + \boldsymbol{z}'$

$$\left|\left|\mu'\right|\right|_{\mathrm{TV}} + \lambda \left|\left|\boldsymbol{z}'\right|\right|_1 \geq ||\mu||_{\mathrm{TV}} + \left\langle Q, \mu' - \mu \right\rangle + \lambda \, ||\boldsymbol{z}||_1 + \lambda \left\langle \frac{1}{\lambda} \boldsymbol{q}, \boldsymbol{z}' - \boldsymbol{z} \right\rangle \tag{3.8}$$

$$\geq ||\mu||_{\mathrm{TV}} + \left\langle \mathcal{F}_n^* \, \boldsymbol{q}, \mu' - \mu \right\rangle + \lambda \, ||\boldsymbol{z}||_1 + \left\langle \boldsymbol{q}, \boldsymbol{z}' - \boldsymbol{z} \right\rangle \tag{3.9}$$

$$= ||\mu||_{\mathrm{TV}} + \lambda \, ||\boldsymbol{z}||_1 + \left\langle \boldsymbol{q}, \mathcal{F}_n \left(\mu' - \mu \right) + \boldsymbol{z}' - \boldsymbol{z} \right\rangle \tag{3.10}$$

$$= ||\mu||_{\mathrm{TV}} + \lambda \, ||\boldsymbol{z}||_1 \quad \text{since } \mathcal{F}_n \, \mu' + \boldsymbol{z}' = \mathcal{F}_n \, \mu + \boldsymbol{z}. \tag{3.11}$$

The existence of Q thus implies that $(\mu, \boldsymbol{z})$ is a solution to Problem 2.7. In fact, as stated Proposition 3.1, it implies that $(\mu, \boldsymbol{z})$ is the unique solution. The rest of this section is devoted to showing that a dual polynomial exists with high probability, as formalized by the following proposition.

Proposition 3.2 (Existence of dual polynomial)**.** *Under the assumptions of Theorem 2.2 there exists a dual polynomial associated to μ and $\boldsymbol{z}$ with probability at least $1 - \epsilon$.*

In order to simplify notation in the sequel, we define the vectors $\boldsymbol{h} \in \mathbb{C}^k$ and $\boldsymbol{r} \in \mathbb{C}^s$ and an integer m such that

$$\boldsymbol{h}_j := \frac{\boldsymbol{x}_j}{|\boldsymbol{x}_j|} \quad 1 \leq j \leq k, \tag{3.12}$$

$$\boldsymbol{r}_l := \frac{\boldsymbol{z}_l}{|\boldsymbol{z}_l|} \quad l \in \Omega, \tag{3.13}$$

$$m := \begin{cases} \frac{n-1}{2} & \text{if } n \text{ is odd,} \\ \frac{n}{2} - 1 & \text{if } n \text{ is even.} \end{cases} \tag{3.14}$$

Applying a simple change of variable, we express Q as

$$Q(f) = \sum_{l=-m}^{m} \boldsymbol{q}_l \, e^{-i2\pi l f}. \tag{3.15}$$

In a nutshell, our goal is (1) to construct a polynomial of this form so that Q interpolates $\boldsymbol{h}$ on T and $\boldsymbol{q}$ interpolates $\boldsymbol{r}$ on Ω, and (2) to verify that the magnitude of Q is strictly bounded by one on T^c and the magnitude of $\boldsymbol{q}$ is strictly bounded by λ on Ω^c.

3.2 Construction via interpolation

We now take a brief detour to introduce a basic technique for the construction of dual polynomials. Consider the spectral super-resolution problem when the data are of the form $\bar{\boldsymbol{y}} := \mathcal{F}_n \, \mu$, i.e. when there are no outliers. A simple corollary to Proposition 3.1 is that the existence of a dual polynomial of the form

$$\bar{Q}(f) = \sum_{l=-m}^{m} \bar{\boldsymbol{q}}_l \, e^{-i2\pi l f} \tag{3.16}$$

such that

$$\bar{Q}(f_j) = \boldsymbol{h}_j, \qquad \forall f_j \in T, \tag{3.17}$$

$$\left|\bar{Q}(f)\right| < 1, \qquad \forall f \in T^c, \tag{3.18}$$

implies that TV-norm minimization achieves exact recovery in the absence of noise. In this section we describe how to construct such a polynomial using interpolation. This technique was introduced in [12] to obtain guarantees for super-resolution under a minimum-separation condition.

The basic idea is to use a kernel $\bar{K}$ and its derivative $\bar{K}^{(1)}$ to interpolate $\boldsymbol{h}$ while forcing the derivative of the polynomial to equal zero on T. Setting the derivative to zero induces a local

extremum which ensures that the magnitude of the polynomial stays bounded below one in the vicinity of T (see Figure 11 in [38] for an illustration). More formally,

$$\bar{Q}\left(f\right) := \sum_{j=1}^{k} \bar{\boldsymbol{\alpha}}_j \, \bar{K}\left(f - f_j\right) + \kappa \sum_{j=1}^{k} \bar{\boldsymbol{\beta}}_j \, \bar{K}^{(1)}\left(f - f_j\right), \tag{3.19}$$

where

$$\kappa := \frac{1}{\sqrt{\left|\bar{K}^{(2)}\left(0\right)\right|}} \tag{3.20}$$

is the value of the second derivative of the kernel at the origin. This quantity will appear often in the proof to simplify notation. $\bar{\boldsymbol{\alpha}} \in \mathbb{C}^k$ and $\bar{\boldsymbol{\beta}} \in \mathbb{C}^k$ are coefficient vectors set so that

$$\bar{Q}\left(f_j\right) = \boldsymbol{h}_j, \quad f_j \in T, \tag{3.21}$$

$$\bar{Q}_R^{(1)}\left(f_j\right) + i\, \bar{Q}_I^{(1)}\left(f_j\right) = 0, \quad f_j \in T, \tag{3.22}$$

where $\bar{Q}_R^{(1)}$ denotes the real part of $\bar{Q}^{(1)}$ and $\bar{Q}_I^{(1)}$ the imaginary part. In matrix form, $\bar{\boldsymbol{\alpha}}$ and $\bar{\boldsymbol{\beta}}$ are the solution to the system

$$\begin{bmatrix} \bar{D}_0 & \bar{D}_1 \\ -\bar{D}_1 & \bar{D}_2 \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \bar{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{bmatrix} \tag{3.23}$$

where

$$\left(\bar{D}_0\right)_{jl} = \bar{K}\left(f_j - f_l\right), \quad \left(\bar{D}_1\right)_{jl} = \kappa\, \bar{K}^{(1)}\left(f_j - f_l\right), \quad \left(\bar{D}_2\right)_{jl} = -\kappa^2\, \bar{K}^{(2)}\left(f_j - f_l\right). \tag{3.24}$$

In [12] $\bar{Q}$ is shown to be a valid dual polynomial for a minimum separation equal to $\frac{4}{n-1}$ when the interpolation kernel is a squared Fejér kernel. The required minimum separation is sharpened to $\frac{2.52}{n-1}$ in [38] by using a different kernel, which will be our choice for $\bar{K}$ in this paper. Consider the Dirichlet kernel of order $\tilde{m} > 0$

$$\mathcal{D}_{\tilde{m}}\left(f\right) := \frac{1}{2\,\tilde{m}+1} \sum_{l=-\tilde{m}}^{\tilde{m}} e^{i2\pi l f} = \begin{cases} 1 & \text{if } f = 0 \\ \frac{\sin((2\,\tilde{m}+1)\pi f)}{(2\,\tilde{m}+1)\sin(\pi f)} & \text{otherwise .} \end{cases} \tag{3.25}$$

Following [38], we define $\bar{K}$ as the product of three different Dirichlet kernels with different orders

$$\bar{K}\left(f\right) := \mathcal{D}_{0.247\,m}\left(f\right) \mathcal{D}_{0.339\,m}\left(f\right) \mathcal{D}_{0.414\,m}\left(f\right) \tag{3.26}$$

$$= \sum_{l=-m}^{m} \boldsymbol{c}_l \, e^{i2\pi l f} \tag{3.27}$$

where $\boldsymbol{c} \in \mathbb{C}^n$ is the convolution of the Fourier coefficients of the three Dirichlet kernels. The choice of the width of the three kernels might seem rather arbitrary; it is chosen to optimize the bound on the minimum separation by achieving a good tradeoff between the *spikiness* of $\bar{K}$ in the vicinity

of the origin and its asymptotic decay [38]. For simplicity we assume that $0.247\,m$, $0.339\,m$ and $0.414\,m$ are all integers.[5] Figure 3 shows $\bar{K}$ and its first derivative.

We end the section with two lemmas bounding κ and the magnitude of the coefficients of $\boldsymbol{q}$, which will be useful at different points of the proof.

Lemma 3.3. *If $m \geq 10^3$, the constant κ, defined by (3.20), satisfies*

$$\frac{0.467}{m} \leq \kappa \leq \frac{0.468}{m}. \tag{3.28}$$

Proof. The bound follows from the fact that $\mathcal{D}^{(2)}_{\tilde{m}}(0) := -4\pi^2\tilde{m}(1+\tilde{m})/3$ and equation (C.19) in [38] (see also Lemma 4.8 in [38]). □

Lemma 3.4 (Proof in Section D)**.** *The coefficients of $\bar{K}$ satisfy*

$$||\boldsymbol{c}||_\infty \leq \frac{1.3}{m}. \tag{3.29}$$

3.3 Interpolation with a random kernel

The trigonometric polynomial $\bar{Q}$ defined in the previous section is not a valid certificate when outliers are present in the data; it does not satisfy (3.5) and (3.6). In order to adapt the construction so that it meets these conditions we draw upon techniques developed in [66], which studies spectral super-resolution in a compressed-sensing scenario where a subset $\mathcal{S}$ of the samples is missing. To prove that TV-norm minimization succeeds in such a scenario, the authors of [66] construct a bounded polynomial with coefficients restricted to the complement of $\mathcal{S}$, which interpolates the sign pattern of the line spectra on their support. This is achieved by using an interpolation kernel with coefficients supported on $\mathcal{S}^c$.

We denote our dual-polynomial candidate by Q. Let us begin by decomposing Q into two components

$$Q(f) := Q_{\text{aux}}(f) + R(f), \tag{3.30}$$

such that the coefficients of the first component are restricted to Ω^c ,

$$Q_{\text{aux}}(f) := \sum_{l \in \Omega^c} \boldsymbol{q}_l\, e^{-i2\pi lf}, \tag{3.31}$$

and the coefficients of the second component are restricted to Ω and *fixed to equal* $\lambda \boldsymbol{r}$ (recall that $\lambda = 1/\sqrt{n}$),

$$R(f) := \frac{1}{\sqrt{n}} \sum_{l \in \Omega} \boldsymbol{r}_l\, e^{-i2\pi lf}. \tag{3.32}$$

This immediately guarantees that Q satisfies (3.5). Now our task is to construct Q_{aux} so that Q also meets the rest of conditions in Proposition 3.1.

[5] To avoid this assumption one can adapt the width of the three kernels so that the length of their convolution equals $2m$ and then recompute the bounds that we borrow from [38].

Following the interpolation technique described in Section 3.2, we constrain Q to interpolate $\boldsymbol{h}$ and have zero derivative in T,

$$Q\left(f_j\right) = \boldsymbol{h}_j, \quad f_j \in T, \tag{3.33}$$

$$Q_R^{(1)}\left(f_j\right) + i\, Q_I^{(1)}\left(f_j\right) = 0, \qquad f_j \in T. \tag{3.34}$$

Given that R is fixed, this is equivalent to

$$Q_{\text{aux}}\left(f_j\right) = \boldsymbol{h}_j - R\left(f_j\right), \qquad f_j \in T, \tag{3.35}$$

$$\left(Q_{\text{aux}}\right)_R^{(1)}\left(f_j\right) + i\, \left(Q_{\text{aux}}\right)_I^{(1)}\left(f_j\right) = -R_R^{(1)}\left(f_j\right) - i\, R_I^{(1)}\left(f_j\right), \quad f_j \in T, \tag{3.36}$$

where the subscript R indicates the real part of a number and the subscript I the imaginary part. This interpolation problem is very similar to the one that arises in compressed sensing off the grid [66]: we must interpolate a certain vector with a polynomial whose coefficients are restricted to a certain subset, in our case Ω^c. Following [66] we employ an interpolation kernel K obtained by selecting the coefficients of $\bar{K}$ in Ω^c,

$$K\left(f\right) := \sum_{l \in \Omega^c} \boldsymbol{c}_l\, e^{i2\pi l f} \tag{3.37}$$

$$= \sum_{l=-m}^{m} \delta_{\Omega^c}\left(l\right) \boldsymbol{c}_l\, e^{i2\pi l f}, \tag{3.38}$$

where δ_{Ω^c} is an indicator random variable that is equal to one if $l \in \Omega^c$ and to zero otherwise. Under the assumptions of Theorem 2.2 these are independent Bernoulli random variables with parameter $\frac{n-s}{n}$, so that the mean of K is equal to a scaled version of $\bar{K}$,

$$\mathbb{E}\left(K\left(f\right)\right) := \frac{n-s}{n} \sum_{l=-m}^{m} \boldsymbol{c}_l\, e^{i2\pi l f} \tag{3.39}$$

$$= \frac{n-s}{n} \bar{K}\left(f\right). \tag{3.40}$$

K and its derivatives concentrate around $\bar{K}$ and its derivatives (scaled by $\frac{n-s}{n}$) near the origin, but they don't display the same asymptotic decay. This is illustrated in Figure 3.

Using K and its first derivative $K^{(1)}$ to construct Q_{aux} ensures that its nonzero coefficients are restricted to Ω^c. In more detail, Q_{aux} is a linear combination of shifted and scaled copies of K and $K^{(1)}$,

$$Q_{\text{aux}}\left(f\right) := \sum_{j=1}^{k} \boldsymbol{\alpha}_j\, K\left(f - f_j\right) + \kappa\, \boldsymbol{\beta}_j\, K^{(1)}\left(f - f_j\right), \tag{3.41}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^k$ and $\boldsymbol{\beta} \in \mathbb{C}^k$ are chosen to satisfy (3.35) and (3.36). The corresponding system of equations (3.35) and (3.36) can be recast in matrix form:

$$\begin{bmatrix} D_0 & D_1 \\ -D_1 & D_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h} \\ 0 \end{bmatrix} - \frac{1}{\sqrt{n}} B_\Omega\, \boldsymbol{r}, \tag{3.42}$$

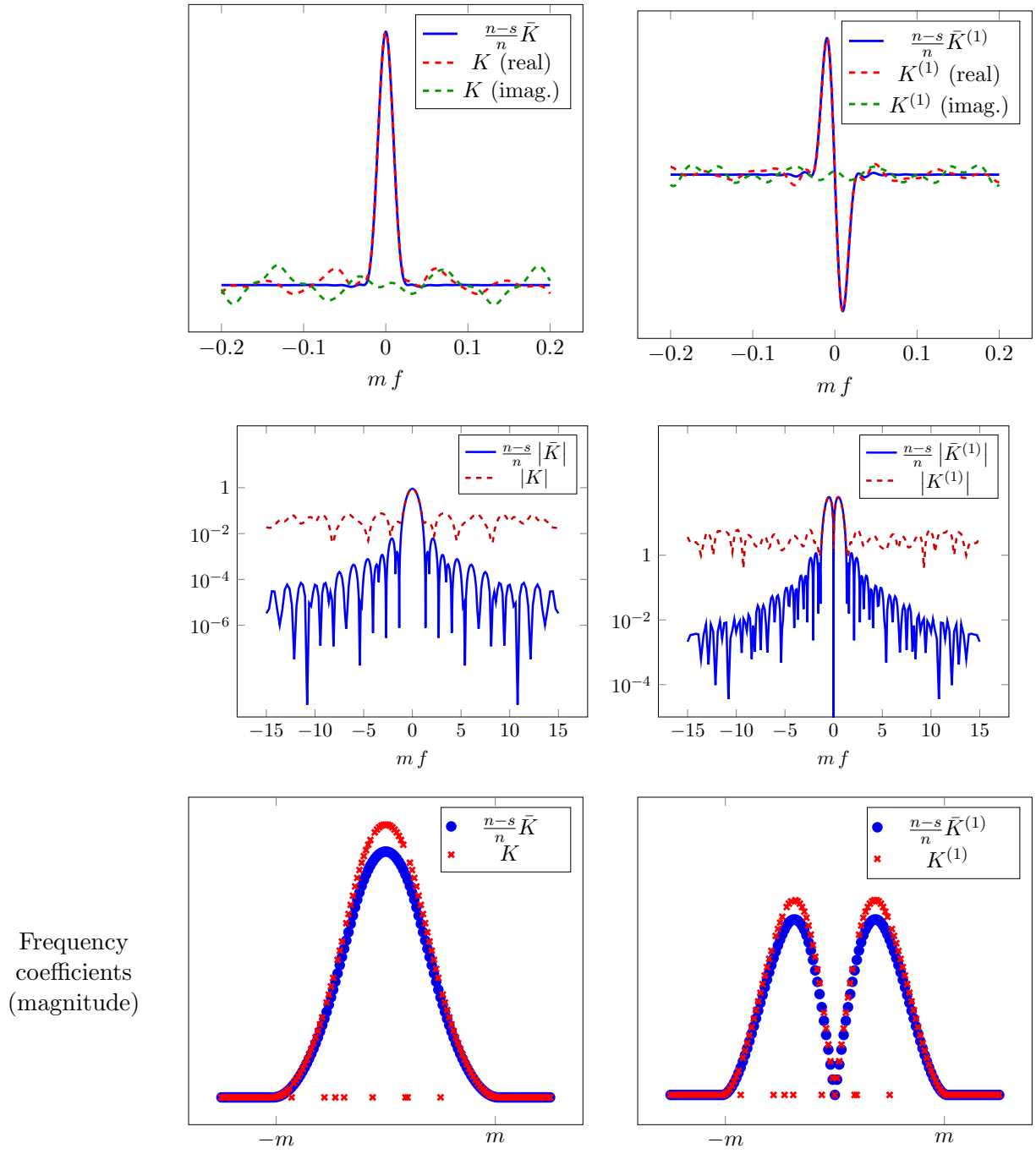

Figure 3: The top row shows the interpolating kernel K and $K^{(1)}$ compared to a scaled version of $\bar{K}$ and $\bar{K}^{(1)}$. In the second row we see the asymptotic decay of the magnitudes of both kernels and their derivatives. The left image in the bottom row illustrates the construction of K: the Fourier coefficients $\boldsymbol{c}$ of $\bar{K}$ that lie in Ω are set to zero. On the right we can see the Fourier coefficients of $K^{(1)}$ and a scaled version of $\bar{K}^{(1)}$.

where

$$(D_0)_{jl} = K(f_j - f_l), \quad (D_1)_{jl} = \kappa K^{(1)}(f_j - f_l), \quad (D_2)_{jl} = -\kappa^2 K^{(2)}(f_j - f_l). \tag{3.43}$$

Note that we have expressed the values of R and $R^{(1)}$ in T in terms of $\boldsymbol{r}$,

$$\frac{1}{\sqrt{n}} B_\Omega \boldsymbol{r} = \begin{bmatrix} R(f_1) & R(f_2) & \cdots & R(f_k) & -\kappa R^{(1)}(f_1) & -\kappa R^{(1)}(f_2) & \cdots & -\kappa R^{(1)}(f_k) \end{bmatrix}^T, \tag{3.44}$$

where

$$\boldsymbol{b}(l) := \begin{bmatrix} e^{-i2\pi l f_1} & e^{-i2\pi l f_2} & \cdots & e^{-i2\pi l f_k} & i2\pi l\kappa\, e^{-i2\pi l f_1} & \cdots & i2\pi l\kappa\, e^{-i2\pi l f_k} \end{bmatrix}^T, \tag{3.45}$$

$$B_\Omega := \begin{bmatrix} \boldsymbol{b}(i_1) & \boldsymbol{b}(i_2) & \cdots & \boldsymbol{b}(i_s) \end{bmatrix}, \quad \Omega = \{i_1, i_2, \ldots i_s\}. \tag{3.46}$$

Solving this system of equations yields $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and fixes the dual-polynomial candidate,

$$Q(f) := \sum_{j=1}^{k} \boldsymbol{\alpha}_j K(f - f_j) + \kappa \sum_{j=1}^{k} \boldsymbol{\beta}_j K^{(1)}(f - f_j) + R(f) \tag{3.47}$$

$$= \boldsymbol{v_0}(f)^T D^{-1} \left(\begin{bmatrix} \boldsymbol{h} \\ 0 \end{bmatrix} - \frac{1}{\sqrt{n}} B_\Omega \boldsymbol{r} \right) + R(f), \tag{3.48}$$

where we define

$$\boldsymbol{v_\ell}(f) := \kappa^\ell \begin{bmatrix} K^{(\ell)}(f - f_1) & \cdots & K^{(\ell)}(f - f_k) & \kappa K^{(\ell+1)}(f - f_1) & \cdots & \kappa K^{(\ell+1)}(f - f_k) \end{bmatrix}^T$$

for $\ell = 0, 1, 2, \ldots$ In the next section we establish that a polynomial of this form is guaranteed to be a valid certificate with high probability. Figure 4 illustrates our construction for a specific example (note that for ease of visualization $\boldsymbol{h}$ is real instead of complex).

Before ending this section, we record three useful lemmas concerning $\boldsymbol{b}$, B_Ω and $\boldsymbol{v_\ell}$. The first bounds the ℓ_2 norm of $\boldsymbol{b}$.

Lemma 3.5. *If $m \geq 10^3$, for $-m \leq l \leq m$*

$$||\boldsymbol{b}(l)||_2^2 \leq 10\, k. \tag{3.49}$$

Proof.

$$||\boldsymbol{b}(l)||_2^2 \leq k \left(1 + \max_{-m \leq l \leq m} (2\pi l \kappa)^2 \right) \leq 9.65\, k \quad \text{by Lemma 3.3.} \tag{3.50}$$

□

The second yields a bound on the operator norm of B_Ω that holds with high probability.

Lemma 3.6 (Proof in Section E)**.** *Under the assumptions of Theorem 2.2, the event*

$$\mathcal{E}_B := \left\{ \|B_\Omega\| > C_B \left(\log \frac{n}{\epsilon} \right)^{-\frac{1}{2}} \sqrt{n} \right\}, \tag{3.51}$$

where C_B is a numerical constant defined by (H.41), occurs with probability at most $\epsilon/5$.

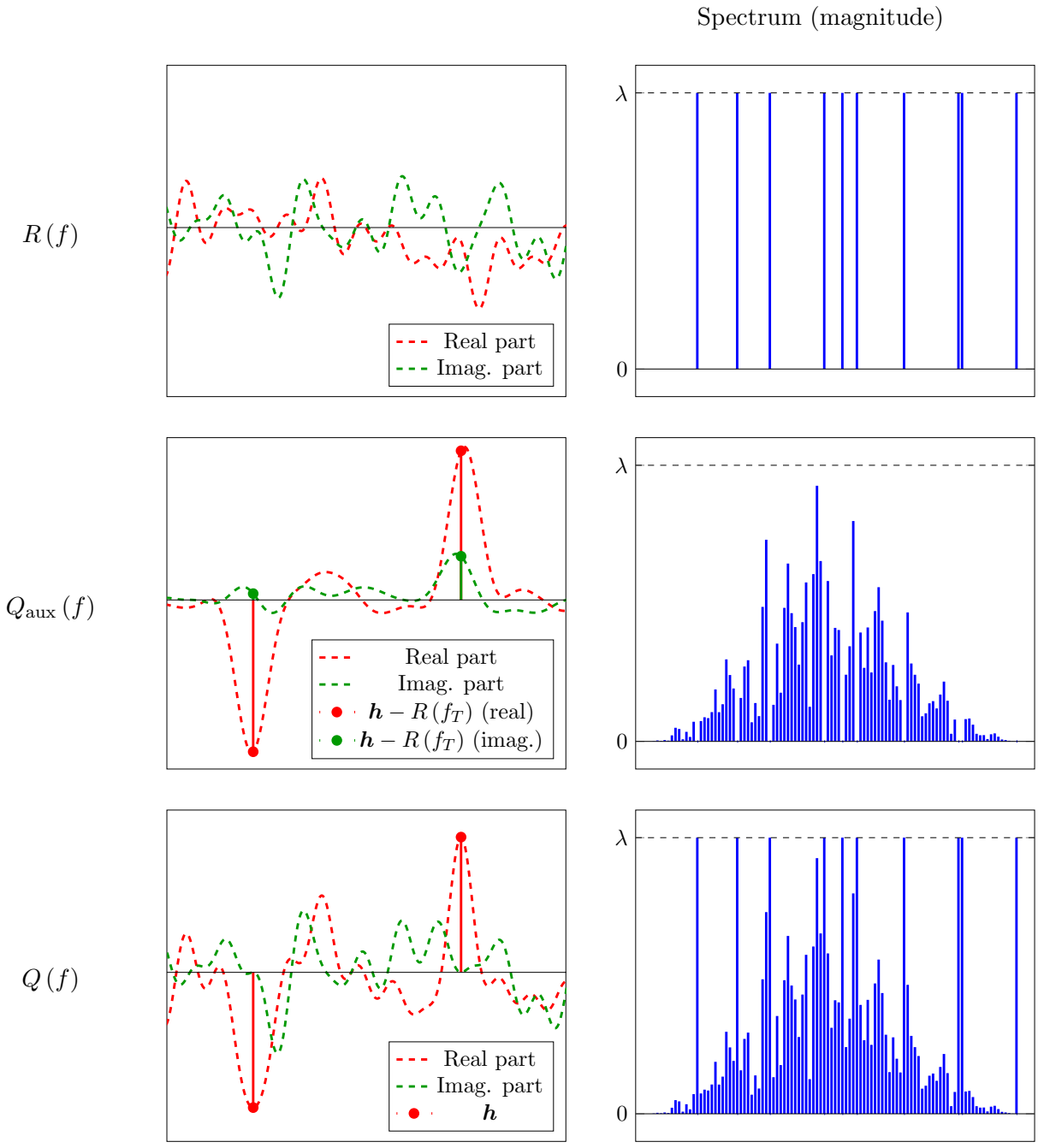

Figure 4: Illustration of our construction of a dual-polynomial candidate Q. The first row shows R, the component that results from fixing the coefficients of Q in Ω to equal $\boldsymbol{r}$. The second row shows Q_{aux}, the component built to ensure that Q interpolates $\boldsymbol{h}$ by correcting for the presence of R. On the right image of the second row, we see that the coefficients of Q_{aux} are indeed restricted to Ω^c. Finally, the last row shows that Q satisfies all of the conditions in Proposition 3.1.

The third allows to control the behavior of $\boldsymbol{v}_{\boldsymbol{\ell}}$, establishing that it does not deviate much from

$$\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f) := \kappa^{\ell}\left[\bar{K}^{(\ell)}(f-f_1) \quad \cdots \quad \bar{K}^{(\ell)}(f-f_k) \quad \kappa\,\bar{K}^{(\ell+1)}(f-f_1) \quad \cdots \quad \kappa\,\bar{K}^{(\ell+1)}(f-f_k)\right]^T$$

on a fine grid with high probability.

Lemma 3.7 (Proof in Section F)**.** *Let $\mathcal{G} \subseteq [0,1]$ be an equispaced grid with cardinality $400\,n^2$. Under the assumptions of Theorem 2.2, the event*

$$\mathcal{E}_v := \left\{ \left\| \boldsymbol{v}_{\boldsymbol{\ell}}(f) - \frac{n-s}{n}\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f) \right\|_2 > C_{\boldsymbol{v}} \left(\log \frac{n}{\epsilon}\right)^{-\frac{1}{2}}, \quad \text{for all } f \in \mathcal{G} \text{ and } \ell \in \{0,1,2,3\} \right\}, \tag{3.52}$$

where $C_{\boldsymbol{v}}$ is a numerical constant defined by (H.45), has probability bounded by $\epsilon/5$.

3.4 Proof of Proposition 3.2

This section summarizes the remaining steps to establish that our proposed construction yields a valid certificate. A detailed description of each step is included in the appendix. First, we show that the system of equations (3.42) has a unique solution with high probability, so that Q is well defined. To alleviate notation, let

$$D := \begin{bmatrix} D_0 & D_1 \\ -D_1 & D_2 \end{bmatrix}, \qquad \bar{D} := \begin{bmatrix} \bar{D}_0 & \bar{D}_1 \\ -\bar{D}_1 & \bar{D}_2 \end{bmatrix}. \tag{3.53}$$

The following result implies that D concentrates around a scaled version of $\bar{D}$. As a result, it is invertible and we can bound the operator norm of its inverse leveraging results from [38].

Lemma 3.8 (Proof in Section G)**.** *Under the assumptions of Theorem 2.2, the event*

$$\mathcal{E}_D := \left\{ \left\| D - \frac{n-s}{n}\bar{D} \right\| \geq \frac{n-s}{4n} \min\left\{ 1, \frac{C_D}{4}\left(\log\frac{n}{\epsilon}\right)^{-\frac{1}{2}} \right\} \right\} \tag{3.54}$$

occurs with probability at most $\epsilon/5$.

In addition, within the event $\mathcal{E}_D^c$, D is invertible and

$$\left\| D^{-1} \right\| \leq 8, \tag{3.55}$$

$$\left\| D^{-1} - \frac{n}{n-s}\bar{D}^{-1} \right\| \leq C_D \left(\log\frac{n}{\epsilon}\right)^{-\frac{1}{2}}, \tag{3.56}$$

where C_D is a numerical constant defined by (H.49).

An immediate consequence of the lemma is that there exists a solution to the system (3.42) and therefore (3.3) holds as long as $\mathcal{E}_D^c$ occurs.

Corollary 3.9. *In $\mathcal{E}_D^c$ Q is well defined and $Q(f_j) = \boldsymbol{h}_j$ for all $f_j \in T$.*

All that remains is to establish that Q meets conditions (3.4) and (3.6); recall that (3.5) is satisfied by construction.

To prove (3.4), we apply a technique from [66]. We first show that Q and its derivatives concentrate around $\bar{Q}$ and its derivatives respectively on a fine grid. Then we leverage Bernstein's inequality to demonstrate that both polynomials and their respective derivatives are close on the whole unit interval. Finally, we borrow some bounds on $\bar{Q}$ and its second derivative from [38] to complete the proof. The details can be found in Section H of the appendix.

Proposition 3.10 (Proof in Section H)**.** *Conditioned on* $\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c$

$$|Q(f)| < 1 \quad \text{for all } f \in T^c \tag{3.57}$$

with probability at least $1 - \epsilon/5$ *under the assumptions of Theorem 2.2.*

Finally, the following proposition establishes that the remaining condition (3.6) holds in $\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c$ with high probability. The proof uses Hoeffding's inequality combined with Lemmas 3.8 and 3.6 to control the magnitude of the coefficients of $\boldsymbol{q}$.

Proposition 3.11 (Proof in Section I)**.** *Conditioned on* $\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c$

$$|\boldsymbol{q}_l| < \frac{1}{\sqrt{n}}, \quad \text{for all } l \in \Omega^c, \tag{3.58}$$

with probability at least $1 - \epsilon/5$ *under the assumptions of Theorem 2.2.*

Now, to complete the proof, let us define $\mathcal{E}_Q$ to be the event that (3.4) holds and $\mathcal{E}_q$ the event that (3.6) holds. Applying De Morgan's laws, the union bound and the fact that for any pair of events $\mathcal{E}_A$ and $\mathcal{E}_B$

$$\mathbb{P}(\mathcal{E}_A) \leq \mathbb{P}(\mathcal{E}_A | \mathcal{E}_B^c) + \mathbb{P}(\mathcal{E}_B). \tag{3.59}$$

we have

$$\begin{aligned}
\mathbb{P}\left((\mathcal{E}_Q \cap \mathcal{E}_q)^c\right) &= \mathbb{P}\left(\mathcal{E}_Q^c \cup \mathcal{E}_q^c\right) && (3.60)\\
&\leq \mathbb{P}\left(\mathcal{E}_Q^c \cup \mathcal{E}_q^c | \mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c\right) + \mathbb{P}\left(\mathcal{E}_B \cup \mathcal{E}_D \cup \mathcal{E}_v\right) && (3.61)\\
&\leq \mathbb{P}\left(\mathcal{E}_Q^c | \mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c\right) + \mathbb{P}\left(\mathcal{E}_q^c | \mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c\right) + \mathbb{P}\left(\mathcal{E}_B\right) + \mathbb{P}\left(\mathcal{E}_D\right) + \mathbb{P}\left(\mathcal{E}_v\right) && (3.62)\\
&\leq \epsilon && (3.63)
\end{aligned}$$

by Lemmas 3.6, 3.7 and 3.8 and Propositions 3.10 and 3.11. We conclude that our construction yields a valid certificate with probability at least $1 - \epsilon$.

4 Algorithms

In this section we discuss how to implement the techniques described in Section 2. In addition, we introduce a greedy demixing method which yields good empirical results. Matlab code implementing all the algorithms presented below is available online[6]. The code allows to reproduce the figures in this section, which illustrate the performance of the different approaches through a running example.

[6] http://www.cims.nyu.edu/~cfgranda/scripts/spectral_superres_outliers.zip

4.1 Demixing via semidefinite programming

The main obstacle to solving Problem (2.7) is that the primal variable $\tilde{\mu}$ is infinite-dimensional. One could tackle this issue by discretizing the possible support of $\tilde{\mu}$ and replacing its TV norm by the ℓ_1 norm of the corresponding vector [67]. Here, we present an alternative approach, originally proposed in [38], that solves the infinite-dimensional optimization problem directly without resorting to discretization. The approach, inspired by a method for TV-norm minimization [12] (see also [4]), relies on the fact that the dual of Problem (2.7) can be recast as a finite-dimensional semidefinite program (SDP).

To simplify notation we introduce the operator $\mathcal{T}$. For any vector $\boldsymbol{u}$ whose first entry $\boldsymbol{u}_1$ is positive and real, $\mathcal{T}(\boldsymbol{u})$ is a Hermitian Toeplitz matrix whose first row is equal to $\boldsymbol{u}^T$. 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 of the different off-diagonal elements of a matrix

$$\mathcal{T}^*(M)_j = \sum_{i=1}^{n-j+1} M_{i,i+j-1}. \tag{4.1}$$

Lemma 4.1. *The dual of Problem (2.7) is*

$$\max_{\boldsymbol{\eta} \in \mathbb{C}^n} \langle \boldsymbol{y}, \boldsymbol{\eta} \rangle \quad \textit{subject to} \quad ||\mathcal{F}_n^* \boldsymbol{\eta}||_\infty \leq 1, \tag{4.2}$$

$$||\boldsymbol{\eta}||_\infty \leq \lambda, \tag{4.3}$$

where the inner product is defined as $\langle \boldsymbol{y}, \boldsymbol{\eta} \rangle := \mathrm{Re}(\boldsymbol{y}^* \boldsymbol{\eta})$. *This problem is equivalent to the semidefinite program*

$$\begin{aligned} \max_{\boldsymbol{\eta} \in \mathbb{C}^n,\, \Lambda \in \mathbb{C}^{n \times n}} \langle \boldsymbol{y}, \boldsymbol{\eta} \rangle \quad \textit{subject to} \quad & \begin{bmatrix} \Lambda & \boldsymbol{\eta} \\ \boldsymbol{\eta}^* & 1 \end{bmatrix} \succeq 0, \\ & \mathcal{T}^*(\Lambda) = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}, \\ & ||\boldsymbol{\eta}||_\infty \leq \lambda, \end{aligned} \tag{4.4}$$

where $\mathbf{0} \in \mathbb{C}^{n-1}$ *is a vector of zeros.*

Lemma 4.1, which follows from Lemma 4.3 below, shows that it is tractable to compute the n-dimensional solution to the dual of Problem (2.7). However, our goal is to obtain the primal solution, which represents the estimate of the line spectrum and the sparse corruptions. The following lemma, which is a consequence of Lemma 4.4, establishes that we can decode the support of the primal solution from the dual solution.

Lemma 4.2. *Let*

$$\hat{\mu} = \sum_{f_j \in \widehat{T}} \hat{\boldsymbol{x}}_j \, \delta(f - f_j), \tag{4.5}$$

and $\hat{\boldsymbol{z}}$ *be a solution to (2.7), such that* $\widehat{T}$ *and* $\widehat{\Omega}$ *are the nonzero supports of the line spectrum* $\hat{\mu}$ *and the spikes* $\hat{\boldsymbol{z}}$ *respectively. If* $\hat{\boldsymbol{\eta}} \in \mathbb{C}^n$ *is a corresponding dual solution, then for any* f_j *in* $\widehat{T}$

$$(\mathcal{F}_n^* \hat{\boldsymbol{\eta}})(f_j) = \frac{\hat{\boldsymbol{x}}_j}{|\hat{\boldsymbol{x}}_j|} \tag{4.6}$$

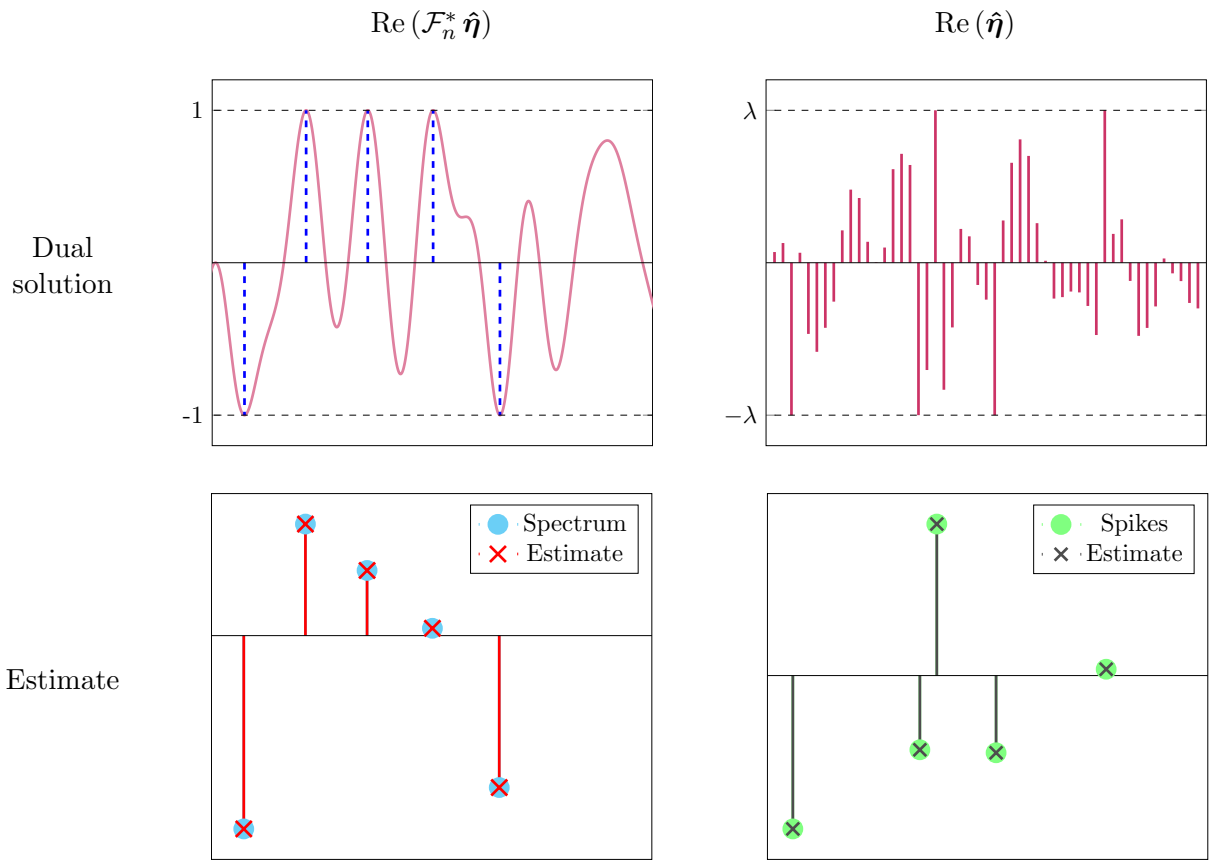

Figure 5: Demixing of the signal in Figure 1 by semidefinite programming. Top left: the polynomial $\mathcal{F}_n^* \hat{\boldsymbol{\eta}}$ (light red), where $\hat{\boldsymbol{\eta}}$ is a solution of Problem (4.4), interpolates the sign of the line spectrum of the sines (dashed red) on their support. Top right: $\lambda^{-1}\hat{\boldsymbol{\eta}}$ interpolates the sign pattern of the spikes on their support. Bottom: locating the support of μ and $\boldsymbol{z}$ allows to demix very accurately (the circular markers represent the original spectrum of the sines and the original spikes and the crosses the corresponding estimates). The parameter λ is set to $1/\sqrt{n}$.

and for any l *in* $\widehat{\Omega}$

$$\hat{\boldsymbol{\eta}}_l = \lambda \frac{\hat{\boldsymbol{z}}_l}{|\hat{\boldsymbol{z}}_l|}. \tag{4.7}$$

In words, the weighted dual solution $\lambda^{-1}\hat{\boldsymbol{\eta}}$ and the corresponding polynomial $\mathcal{F}_n^* \hat{\boldsymbol{\eta}}$ interpolate the sign patterns of the primal-solution components $\hat{\boldsymbol{z}}$ and $\hat{\mu}$ on their respective supports, as illustrated in the top row of Figure 5. This suggests estimating the support of the line spectrum and the outliers in the following way.

1. Solve (4.4) to obtain a dual solution $\hat{\boldsymbol{\eta}}$ and compute $\mathcal{F}_n^* \hat{\boldsymbol{\eta}}$.
2. Set the estimated support of the spikes $\widehat{\Omega}$ to the set of points where $|\hat{\boldsymbol{\eta}}|$ equals λ.
3. Set the estimated support of the line spectrum $\widehat{T}$ to the set of points where $|\mathcal{F}_n^* \hat{\boldsymbol{\eta}}|$ equals one.

4. Estimate the amplitudes of $\hat{\mu}$ and $\hat{\boldsymbol{\eta}}$ on $\hat{T}$ and $\hat{\Omega}$ respectively by solving a system of linear equations $\boldsymbol{y} = \mathcal{F}_n \hat{\mu} + \hat{\boldsymbol{\eta}}$.

Figure 5 shows the results obtained by this method on the data described in Figure 1: both components are recovered very accurately. However, we caution the reader that while the primal solution $(\hat{\mu}, \hat{\boldsymbol{z}})$ is generally unique, the dual solutions are non-unique and some of the dual solutions might produce spurious frequencies and spikes in steps 2 and 3. In fact, the dual solutions form a convex set and only those in the interior of this convex set give exact supports $\hat{\Omega}$ and $\hat{T}$, while those on the boundary generate spurious estimates. When the semidefinite program (4.4) is solved using interior point algorithms as the case in CVX, a dual solution in the interior is returned, generating correct supports as shown in Figure 5. Refer to [66] for a rigorous treatment of this topic for the related missing-data case. Such technical complication will not seriously affect our estimates of the supports since the amplitudes inferred in step 4 will be zero for the extra frequencies and spikes, providing a means to eliminate them.

4.2 Demixing in the presence of dense perturbations

As described in Section 2.5 our demixing method can be adapted to the presence of dense noise in the data by relaxing the equality constraint in Problem 2.7 to an inequality constraint. The only effect on the dual of the optimization problem, which can still be reformulated as an SDP, is an extra term in the cost function.

Lemma 4.3 (Proof in Section J.1)**.** *The dual of Problem (2.19) is*

$$\max_{\boldsymbol{\eta} \in \mathbb{C}^n} \ \langle \boldsymbol{y}, \boldsymbol{\eta} \rangle - \sigma \left|\left| \boldsymbol{\eta} \right|\right|_2 \tag{4.8}$$

$$\text{subject to} \quad \left|\left| \mathcal{F}_n^* \, \boldsymbol{\eta} \right|\right|_\infty \leq 1, \tag{4.9}$$

$$\left|\left| \boldsymbol{\eta} \right|\right|_\infty \leq \lambda. \tag{4.10}$$

This problem is equivalent to the semidefinite program

$$\max_{\boldsymbol{\eta} \in \mathbb{C}^n, \, \Lambda \in \mathbb{C}^{n \times n}} \ \langle \boldsymbol{y}, \boldsymbol{\eta} \rangle - \sigma \left|\left| \boldsymbol{\eta} \right|\right|_2 \quad \text{subject to} \quad \begin{bmatrix} \Lambda & \boldsymbol{\eta} \\ \boldsymbol{\eta}^* & 1 \end{bmatrix} \succeq 0, \tag{4.11}$$

$$\mathcal{T}^* \left(\Lambda \right) = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}, \tag{4.12}$$

$$\left|\left| \boldsymbol{\eta} \right|\right|_\infty \leq \lambda, \tag{4.13}$$

where $\mathbf{0} \in \mathbb{C}^{n-1}$ *is a vector of zeros.*

As in the case without dense noise, the support of the primal solution of Problem (2.19) can be decoded from the dual solution. This is justified by the following lemma, which establishes that the weighted dual solution $\lambda^{-1} \hat{\boldsymbol{\eta}}$ and the corresponding polynomial $\mathcal{F}_n^* \, \hat{\boldsymbol{\eta}}$ interpolate the sign patterns of the primal-solution components $\hat{\boldsymbol{z}}$ and $\hat{\mu}$ on their respective supports.

Lemma 4.4 (Proof in Section J.2)**.** *Let*

$$\hat{\mu} = \sum_{f_j \in \widehat{T}} \hat{\boldsymbol{x}}_j \, \delta \left(f - f_j \right), \tag{4.14}$$

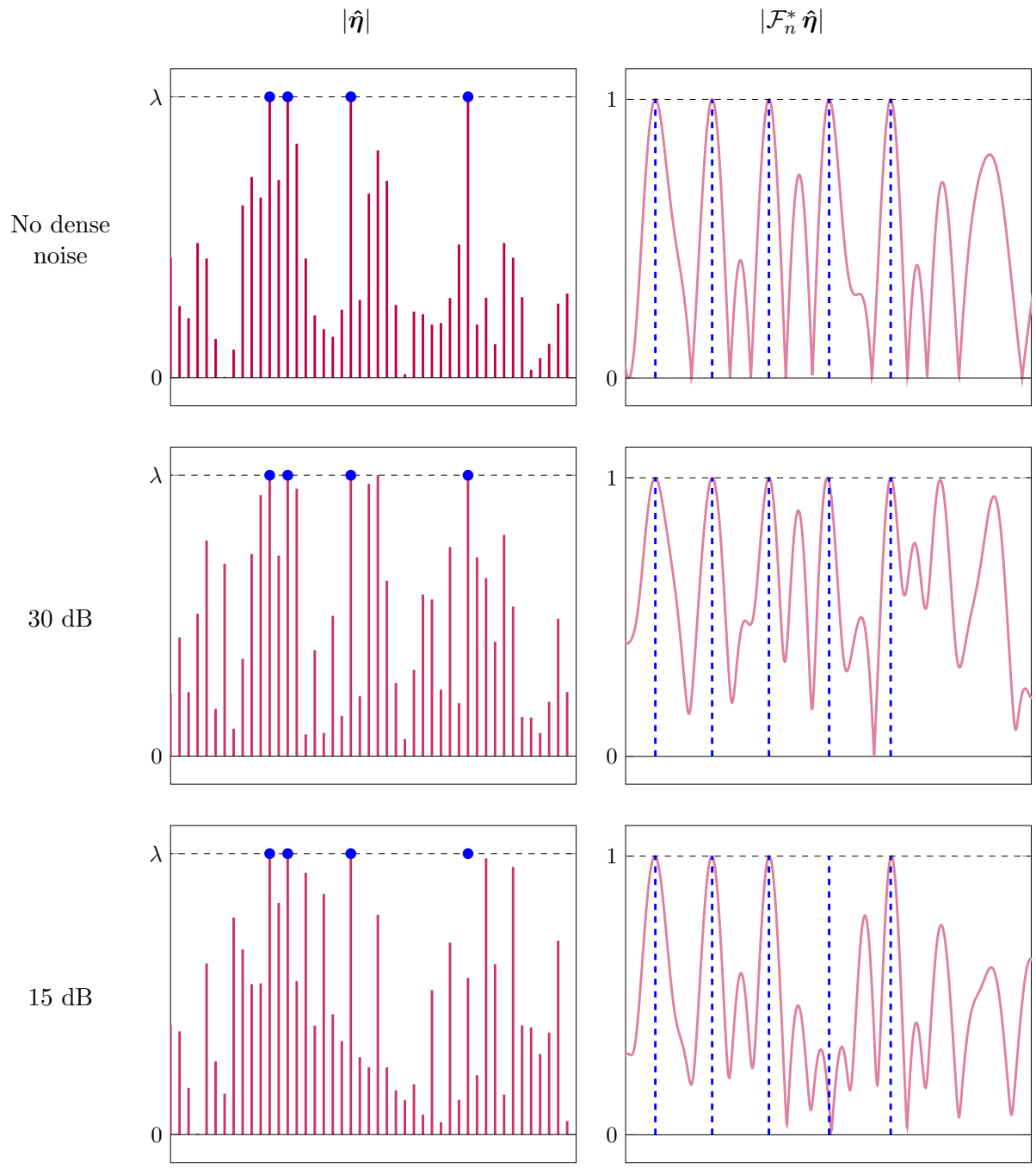

Figure 6: The left column shows the magnitude of the solution to Problem (B.5) (top row) and to Problem 4.8 for different noise levels (second and third rows). $|\hat{\boldsymbol{\eta}}|$ is represented by red lines. Additionally, the support of the sparse perturbation $\boldsymbol{z}$ is marked in blue. The right column shows the trigonometric polynomial corresponding to the dual solutions in red, as well as the support of the spectrum of the multisinusoidal components in blue. The data are the same as in Figure 1 (except for the added noise, which is iid Gaussian). The parameters λ and σ are set to $1/\sqrt{n}$ and 1.5 $||\boldsymbol{w}||_2$ respectively. Note that in practice, the value of the noise level would have to be estimated, for example by cross validation.

and $\hat{\boldsymbol{z}}$ *be a solution to* (2.19)*, such that* $\widehat{T}$ *and* $\widehat{\Omega}$ *are the nonzero supports of the line spectrum* $\hat{\mu}$ *and the spikes* $\hat{\boldsymbol{z}}$ *respectively. If* $\hat{\boldsymbol{\eta}} \in \mathbb{C}^n$ *is a corresponding dual solution, then for any* f_j *in* $\widehat{T}$

$$\left(\mathcal{F}_n^* \, \hat{\boldsymbol{\eta}}\right)(f_j) = \frac{\hat{\boldsymbol{x}}_j}{|\hat{\boldsymbol{x}}_j|} \tag{4.15}$$

and for any l *in* $\widehat{\Omega}$

$$\hat{\boldsymbol{\eta}}_l = \lambda \frac{\hat{\boldsymbol{z}}_l}{|\hat{\boldsymbol{z}}_l|}. \tag{4.16}$$

Figure 6 shows the magnitude of the dual solutions for different values of additive noise. Motivated by the lemma, we propose to estimate the support of the outliers using $\hat{\boldsymbol{\eta}}$ and the support of the spectral lines using $|\mathcal{F}_n^* \, \hat{\boldsymbol{\eta}}|$. Our method to perform spectral super-resolution in the presence of outliers and dense noise consequently consists of the following steps:

1. Solve (4.11) to obtain a dual solution $\hat{\boldsymbol{\eta}}$ and compute $\mathcal{F}_n^* \, \hat{\boldsymbol{\eta}}$.
2. Set the estimated support of the spikes $\widehat{\Omega}$ to the set of points where $|\hat{\boldsymbol{\eta}}|$ equals λ.
3. Set the estimated support of the spectrum $\widehat{T}$ to the set of points where $|\mathcal{F}_n^* \, \hat{\boldsymbol{\eta}}|$ equals one.
4. Estimate the amplitudes of $\hat{\mu}$ by solving a least-squares problem using only the data that do not lie in the estimated support of the spikes $\widehat{\Omega}$.

Figure 7 shows the result of applying our method to data that includes additive iid Gaussian noise with a signal-to-noise ratio (SNR) of 30 and 15 dB. Despite the presence of the dense noise, our method is able to detect all spectral lines at 30 dB and all but one at 15 dB. Additionally, it is capable of detecting most of the spikes correctly: at 30 dB it detects a spurious spike and at 15 dB it misses one. Note that the spike that is not detected when the SNR is 15 dB has a magnitude small enough for it to be considered part of the dense noise.

4.3 Greedy demixing enhanced by local nonconvex optimization

In this section we propose an alternative method for spectral super-resolution in the presence of outliers, which is significantly faster than the SDP-based approach described in the previous sections. In the spirit of matching-pursuit methods [47, 51], the algorithm selects the spectral lines of the signal and the locations of the outliers in a greedy fashion. This is equivalent to choosing atoms from a dictionary of the form

$$\mathcal{D} := \left\{ \boldsymbol{a}\left(f, 0\right), \; f \in [0, 1] \right\} \cup \left\{ \boldsymbol{e}\left(l\right), \; 1 \leq l \leq n \right\}. \tag{4.17}$$

The dictionary includes the multisinusoidal atoms $\boldsymbol{a}\left(f, 0\right)$ defined in (2.20) and n *spiky* atoms $\boldsymbol{e}\left(l\right) \in \mathbb{R}^n$, which are equal to the one-sparse standard-basis vectors. By (2.23), if the data $\boldsymbol{y}$ are of the form (2.3) then they have a $(k+s)$-sparse representation in terms of the atoms in $\mathcal{D}$. Greedy demixing aims to find this sparse representation iteratively.

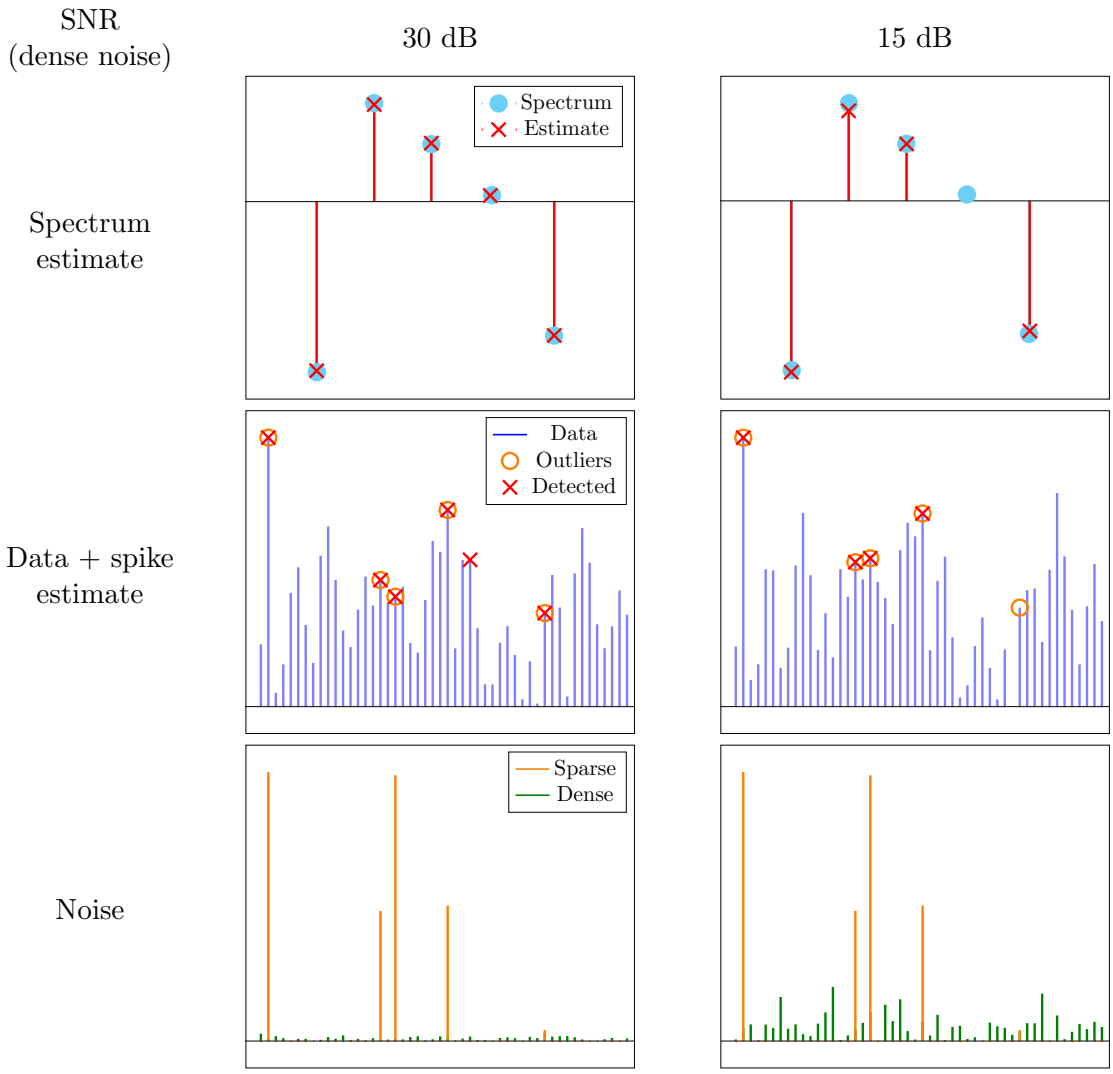

Figure 7: The top row shows the results of applying SDP-based spectral super-resolution in the presence of both dense noise and outliers (bottom row) for two different dense-noise levels (left and right columns). The second row shows the magnitude of the data, the location of the outliers and the outlier estimate produced by the method. In the bottom row we can see the magnitude of the sparse and dense noise (note that when the SNR is 15 dB the smallest sparse-noise components is below the dense-noise level). The signal is the same as in Figure 1 and the data are the same as in Figure 6. The parameter σ is set to 1.5 $||\boldsymbol{w}||_2$ and λ is set to $1/\sqrt{n}$.

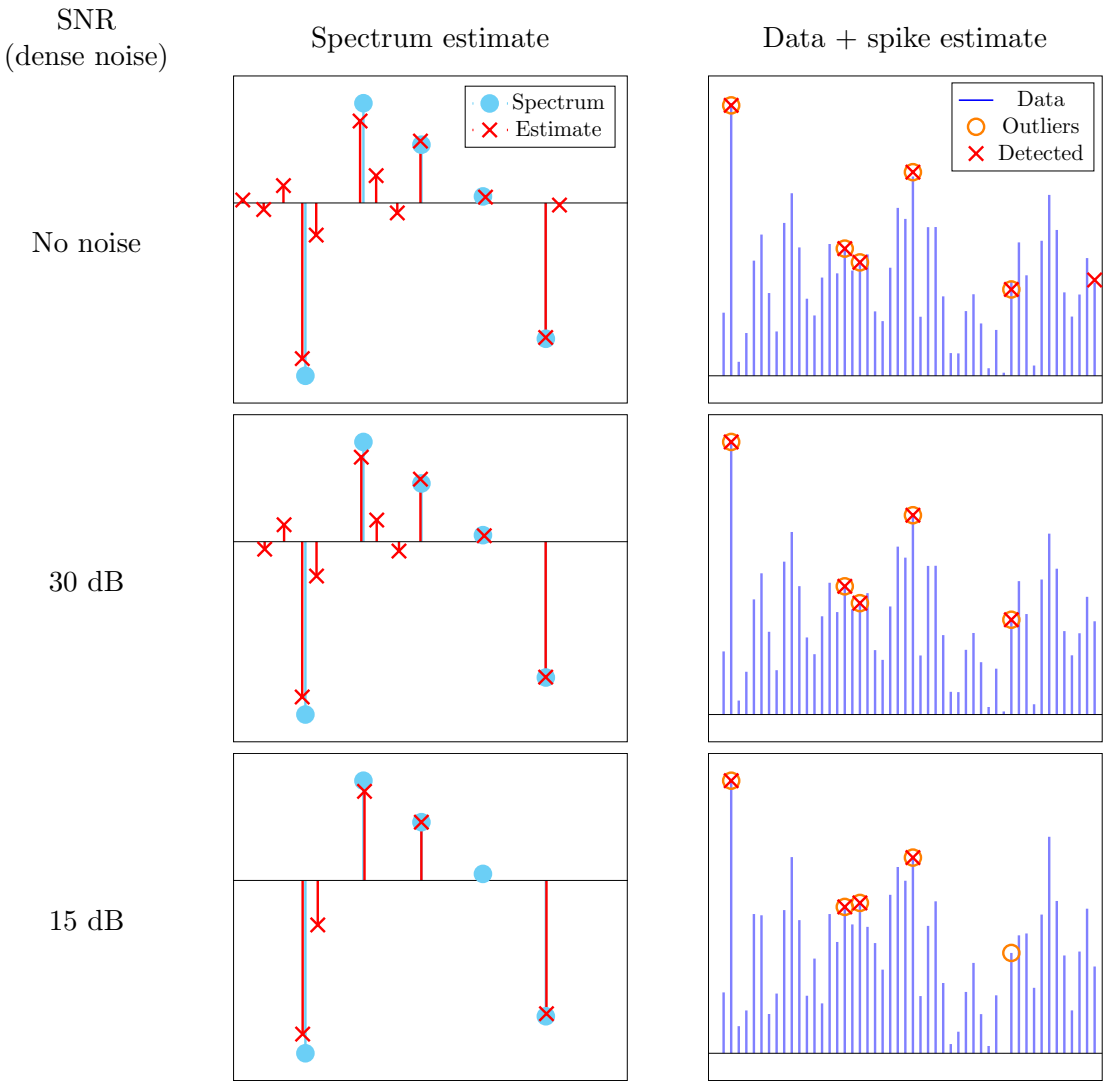

Figure 8: Greedy demixing without a local optimization step. The signal is the same as in Figure 1 and the noisy data are the same as in Figures 6 and 7. The thresholding parameter τ is set depending on the noise level: at 30 dB and in the absence of dense noise it is set small enough not to eliminate the spectral line with the smallest coefficient in the pruning step, whereas at 15 dB it is set so as not to discard the spectral line with the second smallest coefficient.

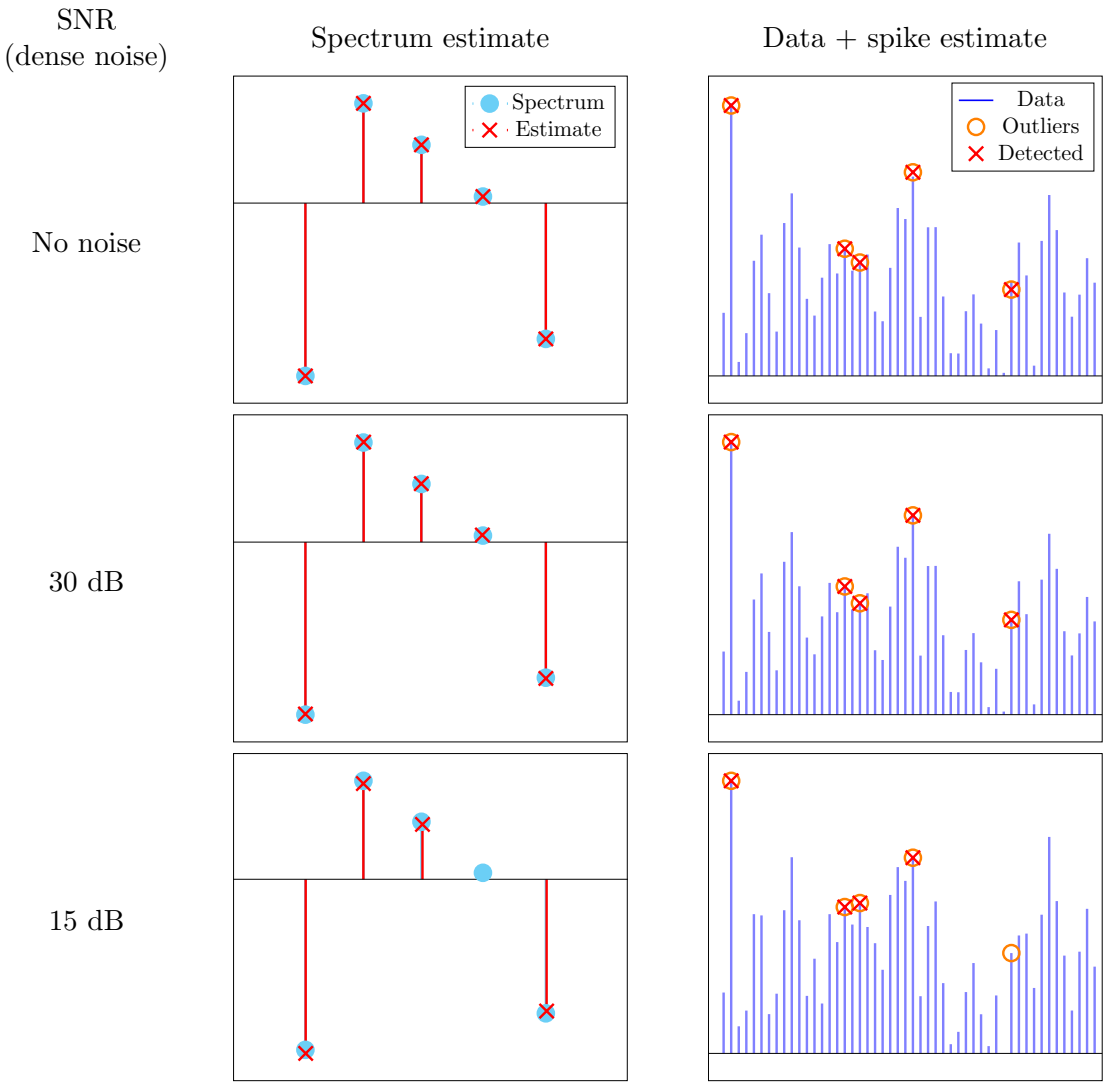

Figure 9: Greedy demixing with a local optimization step. The signal is the same as in Figure 1 and the noisy data are the same as in Figures 6, 7 and 8. The thresholding parameter τ is set as described in the caption of Figure 8.

Inspired by recent work on atomic-norm minimization based on the conditional-gradient method [7, 52, 53], our greedy-demixing procedure includes selection, pruning and local-optimization steps (see also [34, 35, 61] for spectral super-resolution algorithms that leverage a local-optimization step similar to ours).

1. **Initialization:** The residual $\boldsymbol{r} \in \mathbb{C}^n$ is initialized to equal the data vector $\boldsymbol{y}$. The sets of estimated spectral lines $\widehat{T}$ and spikes $\widehat{\Omega}$ are initialized to equal the empty set.

2. **Selection:** At each iteration we compute the atom in $\mathcal{D}$ that has the highest correlation with the current residual $\boldsymbol{r}$ and update either $\widehat{T}$ or $\widehat{\Omega}$. For the *spiky* atoms the correlation is just equal to $||\boldsymbol{r}||_\infty$. For the sinusoidal atoms, we compute the highest correlation by first determining the location f_{grid} of the maximum of the function $corr\,(f) := |\langle \boldsymbol{a}\,(f, 0)\,, \boldsymbol{r}\rangle|$ on a fine grid, which can be done efficiently by computing an oversampled fast Fourier transform, and then finding a local minimum of the function $corr\,(f)$ using a local search method initialized at f_{grid}.

3. **Pruning:** After adding a new atom to $\widehat{T}$ or $\widehat{\Omega}$, we compute the coefficients corresponding to the selected atoms using a least-squares fit. We then remove any atoms whose corresponding coefficients are smaller than a threshold $\tau > 0$.

4. **Local optimization:** We fix the number of selected sinusoidal atoms $\hat{k} := |\widehat{T}|$ and optimize their locations to update $\widehat{T}$ by finding a local minimum of the least-squares cost function

$$ls\left(f_1, \ldots, f_{\hat{k}}\right) := \min_{\hat{\boldsymbol{x}} \in \mathbb{C}^{\hat{k}}, \hat{\boldsymbol{z}} \in \mathbb{C}^{|\widehat{\Omega}|}} \left|\left| \boldsymbol{y} - \sqrt{n} \sum_{j=1}^{\hat{k}} \hat{\boldsymbol{x}}_j\, \boldsymbol{a}\,(f_j, 0) - \sum_{l \in \widehat{\Omega}} \hat{\boldsymbol{z}}_l\, \boldsymbol{e}\,(l) \right|\right|_2, \tag{4.18}$$

   using a local search method[7] initialized at the current estimate $\widehat{T}$. Alternatively, one can use other methods such as gradient descent to find a local minimum of the nonconvex function.

5. The residual is updated by computing the coefficients corresponding to the currently selected atoms using least-squares and subtracting the resulting approximation from $\boldsymbol{y}$.

This algorithm can be applied without any modification to data that are perturbed by dense noise. In Figures 8 and 9 we illustrate the performance of the method on the same data used in Figures 5 and 7. Figure 8 shows what happens if we omit the local-optimization step: the algorithm does not yield exact demixing even in the absence of dense noise. In contrast, in Figure 9 we see that greedy demixing combined with local optimization recovers the two mixed components exactly when no additional noise perturbs the data. In addition, the procedure is robust to the presence of dense noise, as shown in the last two rows of Figure 9.

Intuitively, the greedy method is not able to achieve exact recovery, because it optimizes the position of each spectral line one by one, eventually not being able to make further progress. The local-optimization step refines the fit by optimizing over the positions of the spectral lines simultaneously. This succeeds when the initialization is close enough to a *good* local minimum of the cost function. Our experiments seem to indicate that the greedy scheme provides such an initialization.

[7] We use the Matlab function *fminsearch* based on the simplex search method [42].

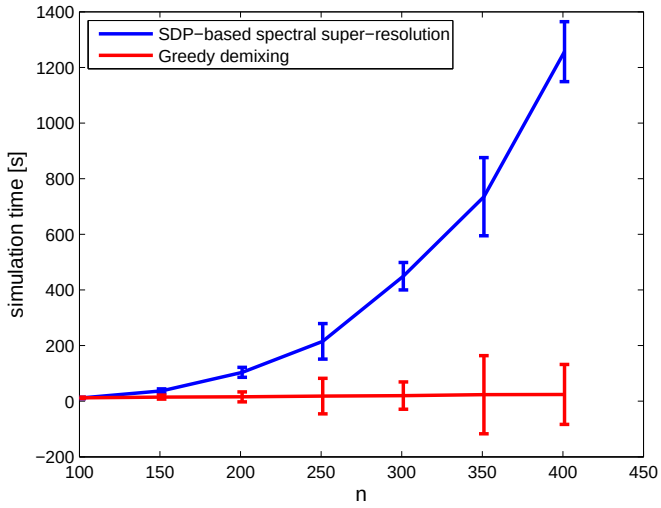

Figure 10: Comparison of average running times for the SDP-based demixing approach described in Section 4.1 and greedy demixing with a local optimization step over 10 tries (the error bars show 95% confidence intervals). The number of spectral lines and of outliers equal 10. The amplitudes of both components are iid Gaussian. The minimum separation of the spectral lines is $2.8/(n+1)$. Both algorithms achieve exact recovery in all instances. The experiments were carried out on a laptop with an Intel Core i5-5300 CPU 2.3GHz and 12G RAM.

As illustrated in Figure 10, the greedy scheme is significantly faster than the SDP-based approach described earlier. These preliminary empirical results show the potential of coupling greedy approaches with local nonconvex optimization. Establishing guarantees for such demixing procedures is an interesting research direction.

4.4 Atomic-norm denoising

In this section, we discuss how to implement the atomic-norm based denoising procedure described in Section 2.6. Our method relies on the fact that the atomic norm has a semidefinite characterization when the dictionary contains sinusoidal atoms of the form (2.20) . This is established in the following proposition, which we borrow from [4, 66].

Proposition 4.5 (Proposition 2.1 [66], [4]). *For $\boldsymbol{g} \in \mathbb{C}^n$*

$$||\boldsymbol{g}||_{\mathcal{A}} = \inf_{t\in\mathbb{R},\, \boldsymbol{u}\in\mathbb{C}^n} \left\{ \frac{n\,\boldsymbol{u}_1 + t}{2} \;:\; \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \boldsymbol{g} \\ \boldsymbol{g}^* & t \end{bmatrix} \succeq 0 \right\}, \tag{4.19}$$

where the operator $\mathcal{T}$ is defined in Section 4.1.

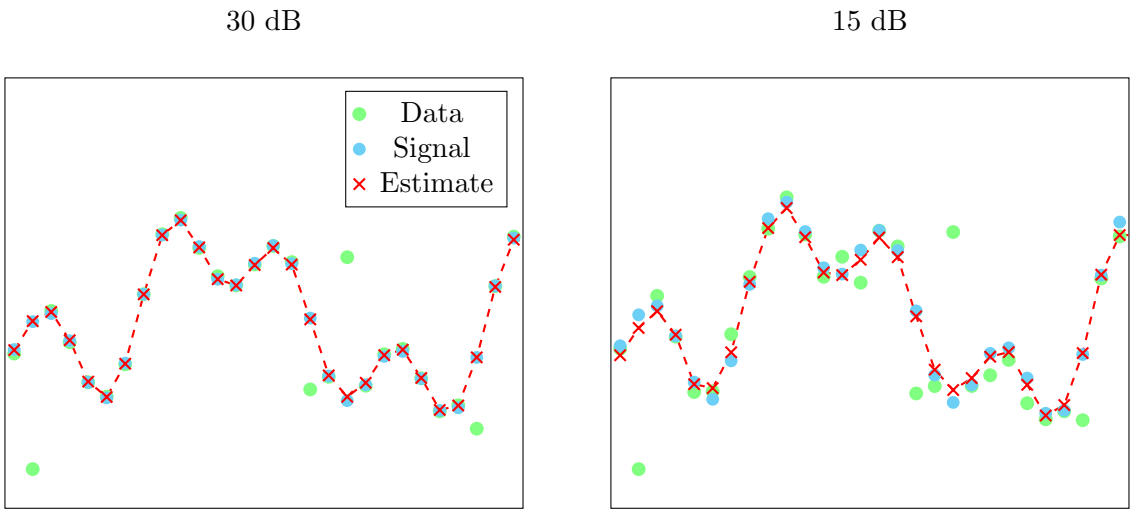

Figure 11: Denoising via atomic-norm minimization in the presence of both outliers and dense noise. The signal is the same as in Figure 1 and the data is the same as in Figures 6 and 7. The parameter λ is set to $1/\sqrt{n}$, whereas γ is set to $1/\left|\left|w\right|\right|_2$ (in practice, we would have to estimate the noise level or set the parameter via cross validation).

This result allows us to rewrite (2.24) as the semidefinite program

$$\min_{\substack{t\in\mathbb{R},\,\boldsymbol{u}\in\mathbb{C}^n,\\ \tilde{\boldsymbol{g}}\in\mathbb{C}^n,\,\tilde{\boldsymbol{z}}\in\mathbb{C}^n}} \frac{n\,\boldsymbol{u}_1+t}{2\sqrt{n}}+\lambda\left|\left|\tilde{\boldsymbol{z}}\right|\right|_1 \quad \text{subject to} \quad \begin{bmatrix}\mathcal{T}\left(\boldsymbol{u}\right) & \tilde{\boldsymbol{g}}\\ \tilde{\boldsymbol{g}}^* & t\end{bmatrix}\succeq 0, \tag{4.20}$$

$$\tilde{\boldsymbol{g}}+\tilde{\boldsymbol{z}}=\boldsymbol{y}, \tag{4.21}$$

which is precisely the dual program of (4.4).

Similarly, Problem (2.27) can be reformulated as the semidefinite program,

$$\min_{\substack{t\in\mathbb{R},\,\boldsymbol{u}\in\mathbb{C}^n,\\ \tilde{\boldsymbol{g}}\in\mathbb{C}^n,\,\tilde{\boldsymbol{z}}\in\mathbb{C}^n}} \frac{n\,\boldsymbol{u}_1+t}{2\sqrt{n}}+\lambda\left|\left|\tilde{\boldsymbol{z}}\right|\right|_1+\frac{\gamma}{2}\left|\left|\boldsymbol{y}-\tilde{\boldsymbol{g}}-\tilde{\boldsymbol{z}}\right|\right|_2^2 \quad \text{subject to} \quad \begin{bmatrix}\mathcal{T}\left(\boldsymbol{u}\right) & \tilde{\boldsymbol{g}}\\ \tilde{\boldsymbol{g}}^* & t\end{bmatrix}\succeq 0 \tag{4.22}$$

This problem can be solved efficiently using the alternating direction method of multipliers [8] (see also [4] for a similar implementation of SDP-based atomic-norm denoising for the case without outliers), as described in detail in Section J.3 of the appendix. Figure 11 shows the results of applying this method to denoise the data used in Figures 7, 8 and 9. In the absence of dense noise, the approach yields perfect denoising (not shown in the figure). When dense noise perturbs the data, the method is still able to perform effective denoising, correcting for the presence of the outliers.

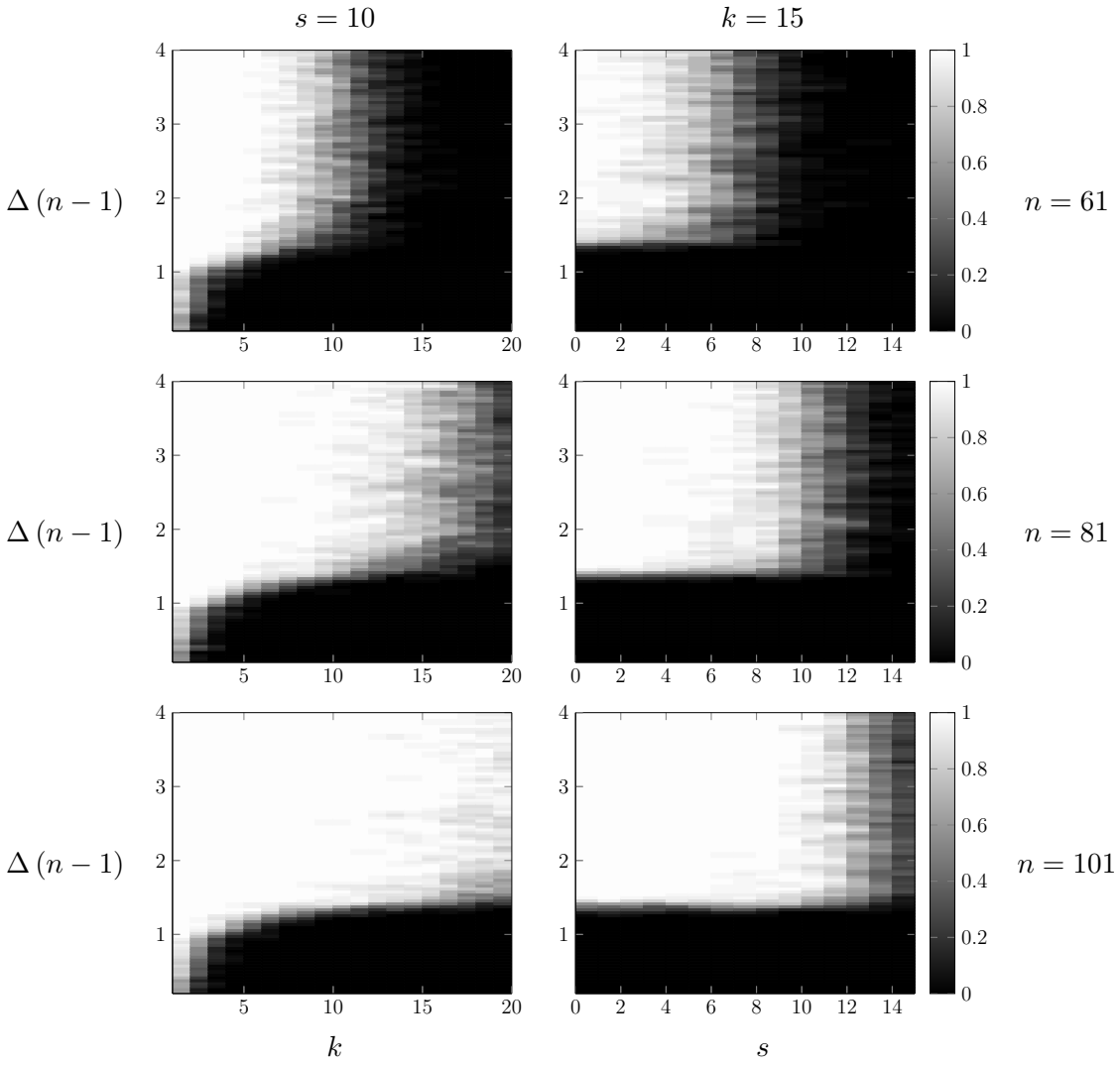

Figure 12: Graphs showing the fraction of times Problem (2.7) achieves exact demixing over 10 trials with random signs and supports for different numbers of spectral lines k (left column) and outliers s (right column), as well as different values of the minimum separation of the spectral lines. Each row shows results for a different number of measurements. The value of the regularization parameter λ is 0.1 for the left column and 0.15 for the second column. The simulations are carried out using CVX [39].

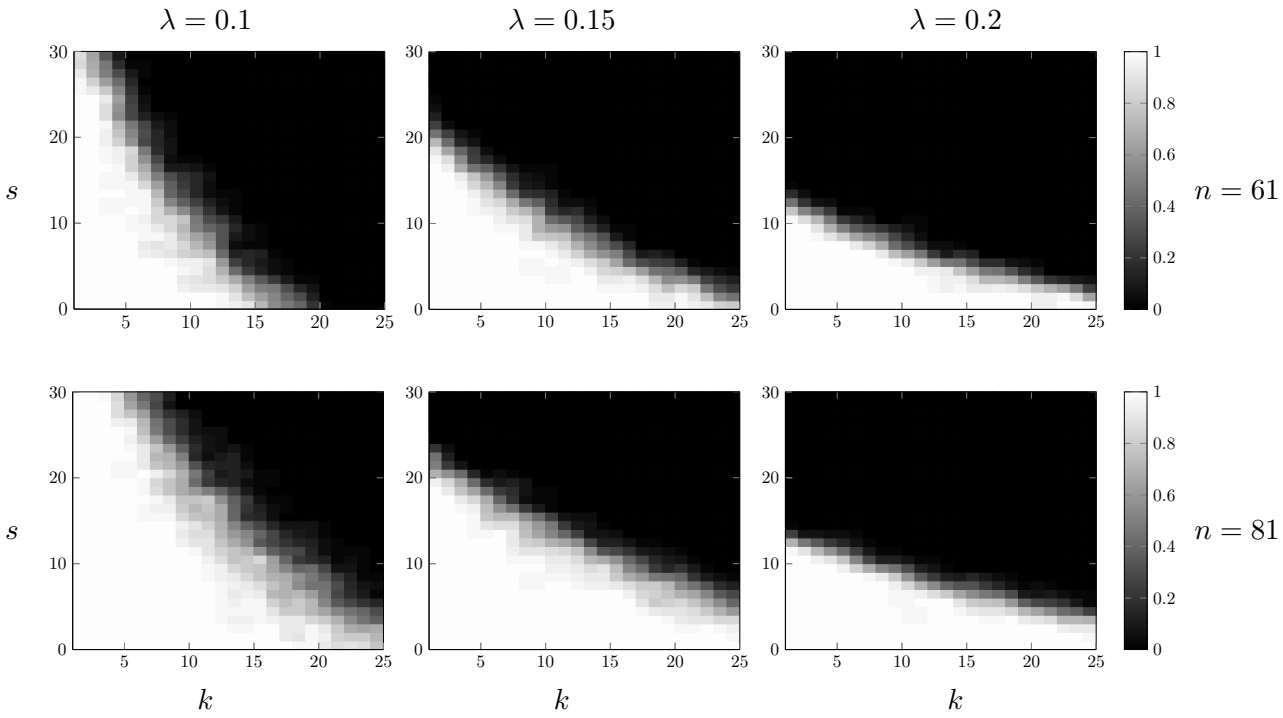

Figure 13: Graphs showing the fraction of times Problem (2.7) achieves exact demixing over 10 trials with random signs and supports for different numbers of spectral lines k and outliers s. The minimum separation of the spectral lines is $2/(n-1)$. Each column shows results for a different value of the regularization parameter λ. Each row shows results for a different number of measurements n. The simulations are carried out using CVX [39].

5 Numerical Experiments

5.1 Demixing via semidefinite programming

In this section we investigate the performance of the method described in Section 2. To do this, we apply the SDP-based approach described in Section 4.1 to data of the form (2.3) varying the different parameters of interest. Fixing either the number of spectral lines k or the number of outliers s allows us to visualize the performance of the method for a range of values of the line spectrum's minimum separation Δ (defined by (2.5)). The results are shown in Figure 12. We observe that in every instance there is a rapid phase transition between the values at which the method always achieves exact demixing and the values at which it fails. The minimum separation at which this phase transition takes place is between $1/(n-1)$ and $2/(n-1)$, which is smaller than the minimum-separation required by Theorem 2.2. We conjecture that if we allow for arbitrary sign patterns, the phase transition would occur near $2/(n-1)$. In fact, if we constrain the amplitudes of the spectral lines to be real instead of complex, the phase transition occurs at a higher minimum separation, as shown in [38, Figure 7].

In order to investigate the effect of the regularization parameter on the performance of the algorithm, we fix Δ and perform demixing for different values of k and s. The results are shown in Figure 13. As suggested by Lemma 2.3, for fixed s the method succeeds for all values of k below a certain limit, and vice versa when we vary s. Since λ weights the effect of the terms that promote sparsity of the two different components in our mixture model, it is no surprise that varying it affects the tradeoff between the number of spectral lines and of spikes that we can demix. For smaller λ the sparsity-inducing term affecting the multisinusoidal component is stronger, so the method succeeds for mixtures with smaller k and larger s. Analogously, for larger λ the sparsity-inducing term affecting the outlier component is stronger, so the method succeeds for mixtures with larger k and smaller s.

5.2 Comparison with matrix-completion based denoising

In this section, we compare the SDP-based atomic-norm denoising method described in Section 4.4 to the matrix-completion based denoising method from [25]. Both algorithms are implemented using CVX [39] and applied to data following model (2.23). In general we observe that both methods either *succeed*, achieving extremely small errors (the relative MSE[8] is smaller than 10^{-8}), or *fail*, producing very large errors. We compare the performance by recording whether the methods succeed or fail in denoising randomly generated signals for different number of spectral lines k and outliers s. To provide a more complete picture, we repeat the simulations for different values of the regularization parameters (λ for atomic-norm denoising and θ for matrix-completion denoising) that govern the sparsity-inducing terms of the corresponding optimization problems. The values of λ and θ are chosen separately to yield the best possible performance.

Figure 14 shows the results. We observe that atomic-norm denoising consistently outperforms matrix-completion denoising across regimes in which the methods achieve different tradeoffs between the values of k and s. In addition, atomic-norm denoising is faster: the average running

[8]The relative MSE is defined as the ratio between the ℓ_2-norm of the difference between the *clean* samples $\boldsymbol{g}$ and the estimate divided by $||\boldsymbol{g}||_2$.

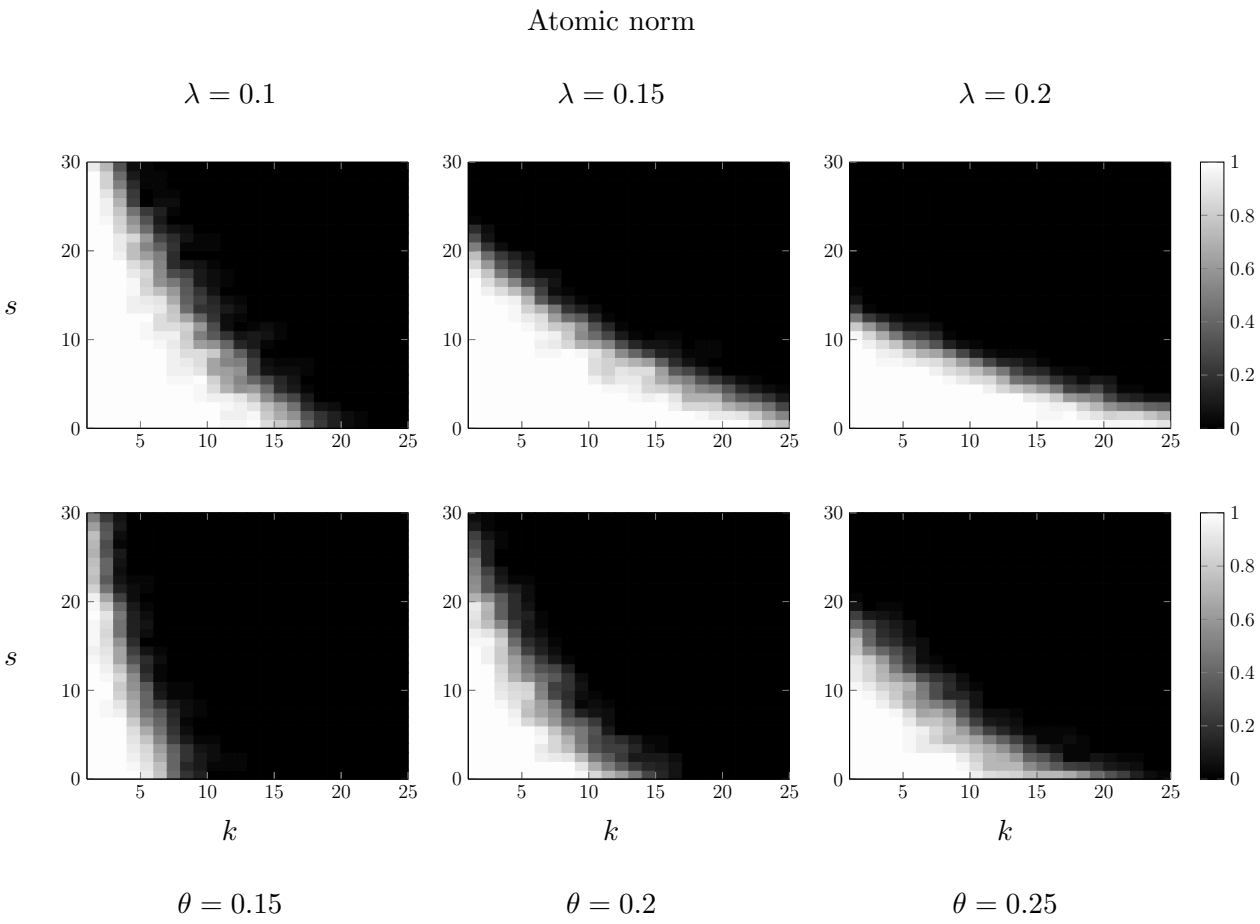

Figure 14: Graphs showing the fraction of times Problem (4.20) (top row) and the matrix-completion approach from [25] (bottom row) achieve exact denoising for different values of their respective regularization parameters over 10 trials with random signs and supports. The minimum separation of the spectral lines is $2/(n-1)$ and the number of data is $n = 61$. The simulations are carried out using CVX [39].

time for each trial is 3.25 seconds with a standard deviation of 0.30 s, whereas the average running time for the matrix-completion approach is of 11.1 s with a standard deviation of 1.32 s. The experiments were carried out on an Intel Xeon desktop computer with a 3.5 GHz CPU and 24 GB of RAM.

6 Conclusion and future research directions

In this work we propose an optimization-based method for spectral super-resolution in the presence of outliers and characterize its performance theoretically. In addition, we describe how to implement the approach using semidefinite programming, discuss its connection to atomic-norm denoising and present a greedy demixing algorithm with a promising empirical performance. Our results suggest the following directions for future research.

- Proving a result similar to Theorem 2.2 without the assumption that the phases of the different components are random. This would require showing that the dual-polynomial construction in Section 3.3 is valid, without leveraging the concentration bounds that we use for our proof. It is unclear whether this is possible because the interpolation kernel K does not display a good asymptotic decay, as shown in Figure 3. Note that if the amplitudes of the sparse noise $\boldsymbol{z}$ are constrained to be real, then a derandomization argument similar to the one in [14, Theorem 2.1] allows to establish the same guarantees as Theorem 2.2 for a sparse perturbation that has an arbitrary deterministic sign pattern.
- Deriving guarantees for spectral super-resolution via the approach described in Section 2.5 in the presence of dense and sparse noise. To achieve this, one could combine our dual polynomial construction with the techniques developed in [13, 37, 65]. In addition, it would be interesting to investigate the application of the method when the level of dense noise is unknown, as in [10].
- Developing fast algorithms to solve the semidefinite programs in Sections 4.1 and 4.2. We have found that ADMM is effective for denoising, but the dual variable converges too slowly for it to be effective in super-resolving the line spectrum.
- Investigating whether greedy demixing techniques, like the one in Section 4.3, can achieve the same performance as our convex-programming approach both empirically and theoretically.
- Considering other structured noise models, beyond sparse perturbations, which could be learnt from data by leveraging techniques such as dictionary learning [46, 50]. For instance, this could allow to deal with recurring interferences in radar applications.

Acknowledgements

C.F. is generously supported by NSF award DMS-1616340. G.T. is generously supported by NSF award CCF-1464205.

References

[1] J.-M. Azais, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

[2] L. G. Beatty, J. D. George, and A. Z. Robinson. Use of the complex exponential expansion as a signal representation for underwater acoustic calibration. *The Journal of the Acoustical Society of America*, 63(6):1782–1794, 1978.

[3] A. J. Berni. Target identification by natural resonance estimation. *IEEE Transactions on Aerospace and Electronic systems*, (2):147–154, 1975.

[4] B. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *Signal Processing, IEEE Transactions on*, 61(23):5987–5999, Dec 2013.

[5] 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.

[6] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman. Imaging and time reversal in random media. *Inverse Problems*, 18(5):1247, 2002.

[7] N. Boyd, G. Schiebinger, and B. Recht. The alternating descent conditional gradient method for sparse inverse problems. Preprint.

[8] 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.

[9] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ Pr, Mar. 2004.

[10] C. Boyer, Y. De Castro, and J. Salmon. Adapting to unknown noise level in sparse deconvolution. Preprint.

[11] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

[12] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, Mar.

[13] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[14] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011.

[15] E. J. Candes and Y. Plan. A probabilistic and ripless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254, 2011.

[16] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, Apr. 2007.

[17] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Thy.*, 52(2):489–509, Feb. 2006.

[18] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Thy.*, 51(12):4203–4215, 2005.

[19] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[20] E. J. Candès and T. Tao. The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Trans. Inf. Thy.*, 56(5):2053–2080, 2010.

[21] R. Carriere and R. L. Moses. High resolution radar target modeling using a modified Prony estimator. *IEEE Transactions on Antennas and Propagation*, 40(1):13–18, 1992.

[22] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

[23] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[24] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[25] Y. Chen and Y. Chi. Robust spectral compressed sensing via structured matrix completion. *Information Theory, IEEE Transactions on*, 60(10):6576–6601, Oct 2014.

[26] Y. De Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354.

[27] B. G. R. de Prony. Essai éxperimental 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'école Polytechnique*, 1(22):24–76, 1795.

[28] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[29] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.

[30] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989.

[31] P. L. Dragotti and Y. M. Lu. On sparse representation in Fourier and local bases. *IEEE Transactions on Information Theory*, 60(12):7888–7899, 2014.

[32] B. Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer Verlag, Feb. 2007.

[33] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, pages 1–41, 2015.

[34] A. Eftekhari and M. B. Wakin. Greed is super: A fast algorithm for super-resolution. Preprint.

[35] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM Journal on Imaging Sciences*, 5(1):179–202, 2012.

[36] Y. Faxin, S. Yiying, and L. Yongtan. An effective method of anti-impulsive-disturbance for ship-target detection in hf radar. In *Radar, 2001 CIE International Conference on, Proceedings*, pages 372–375. IEEE, 2001.

[37] C. Fernandez-Granda. Support detection in super-resolution. In *Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 145–148, 2013.

[38] C. Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference*, 2016.

[39] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `../../cvx`, Apr. 2011.

[40] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Thy.*, 57(3):1548–1566, Mar. 2009.

[41] F. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51 – 83, 1978.

[42] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.

[43] Z. Leonowicz, T. Lobos, and J. Rezmer. Advanced spectrum estimation methods for signal analysis in power electronics. *IEEE Transactions on Industrial Electronics*, 50(3):514–519, 2003.

[44] X. Li. Compressed sensing and matrix completion with constant proportion of corruptions. *Constructive Approximation*, 37(1):73–99, 2013.

[45] X. Lu, J. Wang, A. Ponsford, and R. Kirlin. Impulsive noise excision and performance analysis. In *2010 IEEE Radar Conference*, pages 1295–1300. IEEE, 2010.

[46] J. Mairal, F. Bach, and J. Ponce. Sparse modeling for image and vision processing. *arXiv preprint arXiv:1411.3230*, 2014.

[47] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[48] M. B. McCoy and J. A. Tropp. Sharp recovery bounds for convex demixing, with applications. *Foundations of Computational Mathematics*, 14(3):503–567, 2014.

[49] A. Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, 2015.

[50] B. A. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[51] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44. IEEE, 1993.

[52] N. Rao, P. Shah, and S. Wright. Forward?backward greedy algorithms for signal demixing. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 437–441. IEEE, 2014.

[53] N. Rao, P. Shah, and S. Wright. Forward–backward greedy algorithms for atomic norm regularization. *IEEE Transactions on Signal Processing*, 63(21):5798–5811, 2015.

[54] R. Rockafellar. *Conjugate Duality and Optimization*. Regional conference series in applied mathematics. Society for Industrial and Applied Mathematics, 1974.

[55] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[56] A. Schaeffer. Inequalities of A. Markoff and S. Bernstein for polynomials and related functions. *Bull. Amer. Math. Soc.*, 47, Nov. 1941.

[57] R. Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276–280, 1986.

[58] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case. *Bell System Technical Journal*, 57:1371–1430, 1978.

[59] J. O. Smith. *Introduction to digital filters: with audio applications*, volume 2. Julius Smith, 2008.

[60] P. Stoica, P. Babu, and J. Li. New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data. *IEEE Transactions on Signal Processing*, 59(1):35–47, 2011.

[61] 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.

[62] P. Stoica and R. L. Moses. *Spectral analysis of signals*. Prentice Hall, Upper Saddle River, New Jersey, 1 edition, 2005.

[63] D. Su. Compressed sensing with partially corrupted Fourier measurements. Preprint.

[64] G. Tang. Resolution limits for atomic decompositions via Markov-Bernstein type inequalities. In *Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 548–552, 2015.

[65] G. Tang, B. Bhaskar, and B. Recht. Near minimax line spectral estimation. *Information Theory, IEEE Transactions on*, 61(1):499–512, Jan 2015.

[66] G. Tang, B. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *Information Theory, IEEE Transactions on*, 59(11):7465–7490, Nov 2013.

[67] G. Tang, B. N. Bhaskar, and B. Recht. Sparse recovery over continuous dictionaries-just discretize. In *2013 Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047, Nov 2013.

[68] G. Tang, P. Shah, B. N. Bhaskar, and B. Recht. Robust line spectral estimation. In *Signals, Systems and Computers, 2014 48th Asilomar Conference on*, pages 301–305. IEEE, 2014.

[69] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[70] J. A. Tropp. On the linear independence of spikes and sines. *Journal of Fourier Analysis and Applications*, 14(5-6):838–858, 2008.

[71] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, Aug. 2011.

[72] V. Viti, C. Petrucci, and P. Barone. Prony methods in NMR spectroscopy. *International Journal of Imaging Systems and Technology*, 8(6):565–571, 1997.

[73] Z. Yang and L. Xie. On gridless sparse methods for line spectral estimation from complete and incomplete data. *IEEE Transactions on Signal Processing*, 63(12):3139–3153.

[74] W.-J. Zeng, H. So, and L. Huang. ℓ_p-music: Robust direction-of-arrival estimator for impulsive noise environments. *IEEE Transactions on Signal Processing*, 61:4296–4308, 2013.

[75] L. Zheng and X. Wang. Improved NN-JPDAF for joint multiple target tracking and feature extraction. Preprint.

A Proof of Lemma 2.3

For any vector $\boldsymbol{u}$ and any atomic measure ν, we denote by $\boldsymbol{u}_{\mathcal{S}}$ and $\nu_{\mathcal{S}}$ the restriction of $\boldsymbol{u}$ and ν to the subset of their support indexed by a set $\mathcal{S}$. Let $\{\hat{\mu}, \hat{\boldsymbol{z}}\}$ be any solution to Problem (2.7) applied to $\boldsymbol{y'}$. The pair $\left\{\hat{\mu} + \mu_{T/T'}, \hat{\boldsymbol{z}} + \boldsymbol{z}_{\Omega/\Omega'}\right\}$ is feasible for Problem (2.7) applied to $\boldsymbol{y}$ since

$$\begin{aligned}
\mathcal{F}_n \hat{\mu} + \mathcal{F}_n \mu_{T/T'} + \hat{\boldsymbol{z}} + \boldsymbol{z}_{\Omega/\Omega'} &= \boldsymbol{y'} + \mathcal{F}_n \mu_{T/T'} + \boldsymbol{z}_{\Omega/\Omega'} && \text{(A.1)}\\
&= \mathcal{F}_n \mu' + \mathcal{F}_n \mu_{T/T'} + \boldsymbol{z'} + \boldsymbol{z}_{\Omega/\Omega'} && \text{(A.2)}\\
&= \mathcal{F}_n \mu + \boldsymbol{z} && \text{(A.3)}\\
&= \boldsymbol{y}. && \text{(A.4)}
\end{aligned}$$

By the triangle inequality and the assumption that $\{\mu, \boldsymbol{z}\}$ is the unique solution to Problem (2.7) applied to $\boldsymbol{y'}$, this implies

$$||\mu||_{\mathrm{TV}} + \lambda ||\boldsymbol{z}||_1 < \left|\left|\hat{\mu} + \mu_{T/T'}\right|\right|_{\mathrm{TV}} + \lambda \left|\left|\hat{\boldsymbol{z}} + \boldsymbol{z}_{\Omega/\Omega'}\right|\right|_1 \tag{A.5}$$

$$\leq ||\hat{\mu}||_{\mathrm{TV}} + \left|\left|\hat{\mu}_{T/T'}\right|\right|_{\mathrm{TV}} + \lambda ||\hat{\boldsymbol{z}}||_1 + \lambda \left|\left|\boldsymbol{z}_{\Omega/\Omega'}\right|\right|_1 \tag{A.6}$$

unless $\hat{\mu} + \mu_{T/T'} = \mu$ and $\hat{\boldsymbol{z}} + \boldsymbol{z}_{\Omega/\Omega'} = \boldsymbol{z}$, so that

$$\left|\left|\mu'\right|\right|_{\mathrm{TV}} + \lambda \left|\left|\boldsymbol{z'}\right|\right|_1 = ||\mu||_{\mathrm{TV}} - \left|\left|\mu_{T/T'}\right|\right|_{\mathrm{TV}} + \lambda ||\boldsymbol{z}||_1 - \lambda \left|\left|\boldsymbol{z}_{\Omega/\Omega'}\right|\right|_1 \tag{A.7}$$

$$< ||\hat{\mu}||_{\mathrm{TV}} + \lambda ||\hat{\boldsymbol{z}}||_1 , \tag{A.8}$$

unless $\hat{\mu} = \mu$ and $\hat{\boldsymbol{z}} = \boldsymbol{z'}$. We conclude that $\{\mu', \boldsymbol{z'}\}$ must be the unique solution to Problem (2.7) applied to $\boldsymbol{y'}$.

B Atomic-norm denoising

B.1 Proof of Lemma 2.4

We define a scaled dual norm $\|\cdot\|_{\mathcal{A}'} := \|\cdot\|_{\mathcal{A}}/\sqrt{n}$. The dual norm of $\|\cdot\|_{\mathcal{A}'}$ is

$$\|\boldsymbol{\eta}\|_{\mathcal{A}'}^* = \sup_{||\tilde{\boldsymbol{g}}||_{\mathcal{A}} \leq \sqrt{n}} \langle \boldsymbol{\eta}, \tilde{\boldsymbol{g}} \rangle \tag{B.1}$$

$$= \sup_{\phi \in [0,2\pi), f \in [0,1]} \left\langle \boldsymbol{\eta}, \sqrt{n} e^{i\phi} \boldsymbol{a}\left(f, 0\right) \right\rangle \tag{B.2}$$

$$= \sup_{f \in [0,1]} \left| \left\langle \boldsymbol{\eta}, \sqrt{n} \boldsymbol{a}\left(f, 0\right) \right\rangle \right| \tag{B.3}$$

$$= ||\mathcal{F}_n^* \boldsymbol{\eta}||_\infty . \tag{B.4}$$

The result now follows from the fact that the dual of 2.24 is

$$\max_{\boldsymbol{\eta} \in \mathbb{C}^n} \langle \boldsymbol{y}, \boldsymbol{\eta} \rangle \quad \text{subject to} \quad \|\boldsymbol{\eta}\|_{\mathcal{A}'}^* \leq 1, \tag{B.5}$$

$$||\boldsymbol{\eta}||_\infty \leq \lambda, \tag{B.6}$$

by a standard argument [22, Section 2.1].

B.2 Proof of Corollary 2.5

The corollary is a direct consequence of the following lemma, which establishes that the dual polynomial whose existence we establish in Proposition 3.2 also guarantees that solving Problem (2.24) achieves exact demixing.

Lemma B.1. *If there exists a trigonometric polynomial Q satisfying the conditions listed in Proposition 3.1, then $\boldsymbol{g}$ and $\boldsymbol{z}$ are the unique solutions to Problem (2.24).*

Proof. In the case of the atoms defined by (2.20), the atomic norm is given by

$$||\boldsymbol{u}||_{\mathcal{A}} = \inf_{\substack{\{\tilde{\boldsymbol{x}}_j \geq 0\},\ \{\phi_j \in [0,2\pi)\} \\ \{f_j \in [0,1]\}}} \left\{ \sum_j \tilde{\boldsymbol{x}}_j : \boldsymbol{u} = \sum_j \tilde{\boldsymbol{x}}_j \boldsymbol{a}\left(f_j, \phi_j\right) \right\}, \tag{B.7}$$

so that

$$\begin{aligned} ||\boldsymbol{g}||_{\mathcal{A}} &\leq ||\boldsymbol{x}||_1 \quad \text{due to (2.21)} && \text{(B.8)} \\ &= ||\mu||_{\mathrm{TV}}\,. && \text{(B.9)} \end{aligned}$$

By construction,

$$\begin{aligned} \langle \boldsymbol{q}, \boldsymbol{y} \rangle &= \langle \boldsymbol{q}, \boldsymbol{g} + \boldsymbol{z} \rangle && \text{(B.10)} \\ &= \langle \mathcal{F}_n^* \boldsymbol{q}, \mu \rangle + \langle \boldsymbol{q}, \boldsymbol{z} \rangle && \text{(B.11)} \\ &= \int_{[0,1]} \overline{Q\left(f\right)}\ \mathrm{d}\,\mu\left(f\right) + \lambda \sum_{l=1}^{s} |\boldsymbol{z}_l| && \text{(B.12)} \\ &= ||\mu||_{\mathrm{TV}} + \lambda\, ||\boldsymbol{z}||_1\,. && \text{(B.13)} \end{aligned}$$

Consider an arbitrary feasible pair $\{\boldsymbol{g'}, \boldsymbol{z'}\}$ different from $\{\boldsymbol{g}, \boldsymbol{z}\}$, such that $\boldsymbol{z'}$ has nonzero support Ω' and

$$\boldsymbol{g'} = \sqrt{n} \sum_{f_j \in T'} \boldsymbol{x'}_j \boldsymbol{a}\left(f_j, 0\right), \quad \left|\left|\boldsymbol{g'}\right|\right|_{\mathcal{A}} := \sum_{f_j \in T'} \left|\boldsymbol{x'}_j\right| \tag{B.14}$$

for a sequence of complex coefficients $\boldsymbol{x'}$ and a set of frequency locations $T' \subseteq [0,1]$.

Note that as long as $k + s \leq n$ (recall that $k := |T|$ and $s := |\Omega|$) then either $T \neq T'$ or $\Omega \neq \Omega'$. The reason is that under that condition any set formed by k atoms of the form $\boldsymbol{a}\left(f_j, 0\right)$ and s vectors with cardinality one is linearly independent (this is equivalent to the matrix $\begin{bmatrix} F_T & I_\Omega \end{bmatrix}$ in Section C.1 being full rank), so that if both $T = T'$ and $\Omega = \Omega'$ then $\boldsymbol{g} + \boldsymbol{z} = \boldsymbol{g'} + \boldsymbol{z}$ would imply that $\boldsymbol{g} = \boldsymbol{g'}$ and $\boldsymbol{z} = \boldsymbol{z}$ (and we are assuming this is not the case).

By conditions (3.3) and (3.4)

$$\begin{aligned} \sqrt{n}\, \langle \boldsymbol{q}, \boldsymbol{a}\left(f_j, 0\right) \rangle &= Q\left(f_j\right) && \text{(B.15)} \\ &= \frac{\boldsymbol{x}_j}{|\boldsymbol{x}_j|}, \qquad \forall f_j \in T, && \text{(B.16)} \\ \sqrt{n}\, \langle \boldsymbol{q}, \boldsymbol{a}\left(f_j, 0\right) \rangle &= |Q\left(f\right)| && \text{(B.17)} \\ &< 1, \qquad \forall f \in T^c. && \text{(B.18)} \end{aligned}$$

We have

$$\begin{aligned} ||\boldsymbol{g}||_{\mathcal{A}} + \lambda\, ||\boldsymbol{z}||_1 &\leq \langle \boldsymbol{q}, \boldsymbol{y} \rangle \quad \text{by (B.9) and (B.13)} && \text{(B.19)} \\ &= \left\langle \boldsymbol{q}, \boldsymbol{g'} \right\rangle + \left\langle \boldsymbol{q}, \boldsymbol{z'} \right\rangle && \text{(B.20)} \\ &= \sqrt{n} \sum_{f_j \in T'} \boldsymbol{x'}_j \left\langle \boldsymbol{q}, \boldsymbol{a}\left(f, 0\right) \right\rangle + \left\langle \boldsymbol{q}_{\Omega'}, \boldsymbol{z'} \right\rangle && \text{(B.21)} \\ &< \sqrt{n} \sum_{f_j \in T'} \left|\boldsymbol{x'}_j\right| + \lambda \sum_{l \in \Omega'} \left|\boldsymbol{z'}_j\right| && \text{(B.22)} \\ &= \left|\left|\boldsymbol{g'}\right|\right|_{\mathcal{A}} + \lambda\, \left|\left|\boldsymbol{z'}\right|\right|_1 && \text{(B.23)} \end{aligned}$$

where (B.22) follows from conditions (3.5) and (3.6), (B.16), (B.18) and the fact that either $T \neq T'$ or $\Omega \neq \Omega'$. We conclude that $\{\boldsymbol{g}, \boldsymbol{z}\}$ must be the unique solution to Problem (2.24). $\square$

C Proof of Proposition 3.1

For any vector $\boldsymbol{u}$ and any atomic measure ν, we denote by $\boldsymbol{u}_{\mathcal{S}}$ and $\nu_{\mathcal{S}}$ the restriction of $\boldsymbol{u}$ and ν to the subset of their support indexed by a set $\mathcal{S}$ ($\boldsymbol{u}_{\mathcal{S}}$ has the same dimension as $\boldsymbol{u}$ and $\nu_{\mathcal{S}}$ is still a measure in the unit interval). Let us consider an arbitrary feasible pair μ' and $\boldsymbol{z'}$, such that $\mu' \neq \mu$ or $\boldsymbol{z'} \neq \boldsymbol{z}$. Due to the constraints of the optimization problem, μ' and $\boldsymbol{z'}$ satisfy

$$\boldsymbol{y} = \mathcal{F}_n \, \mu + \boldsymbol{z} = \mathcal{F}_n \, \mu' + \boldsymbol{z'}. \tag{C.1}$$

The following lemma establishes that μ'_{T^c} and $\boldsymbol{z}'_{\Omega^c}$ cannot both equal zero.

Lemma C.1 (Proof in Section C.1)**.** *If $\{\mu', \boldsymbol{z'}\}$ is feasible and μ'_{T^c} and $\boldsymbol{z}'_{\Omega^c}$ both equal zero, then $\mu = \mu'$ and $\boldsymbol{z} = \boldsymbol{z'}$.*

This lemma and the existence of Q imply that the cost function evaluated at $\{\mu', \boldsymbol{z'}\}$ is larger than at $\{\mu, \boldsymbol{z}\}$:

$$\begin{aligned}
||\mu'||_{\mathrm{TV}} + \lambda \, ||\boldsymbol{z'}||_1 &= ||\mu'_T||_{\mathrm{TV}} + ||\mu'_{T^c}||_{\mathrm{TV}} + \lambda \, ||\boldsymbol{z}'_{\Omega}||_1 + \lambda \, ||\boldsymbol{z}'_{\Omega^c}||_1 && \text{(C.2)} \\
&> ||\mu'_T||_{\mathrm{TV}} + \langle Q, \mu'_{T^c} \rangle + \lambda \, ||\boldsymbol{z}'_{\Omega}||_1 + \langle \boldsymbol{q}, \boldsymbol{z}'_{\Omega^c} \rangle \qquad \text{by Lemma C.1, (3.4) and (3.6)} \\
&\geq \langle Q, \mu' \rangle + \langle \boldsymbol{q}, \boldsymbol{z'} \rangle \qquad \text{by (3.3) and (3.5)} && \text{(C.3)} \\
&= \langle \mathcal{F}_n^* \boldsymbol{q}, \mu' \rangle + \langle \boldsymbol{q}, \boldsymbol{z'} \rangle && \text{(C.4)} \\
&= \langle \boldsymbol{q}, \mathcal{F}_n \, \mu' + \boldsymbol{z'} \rangle && \text{(C.5)} \\
&= \langle \boldsymbol{q}, \mathcal{F}_n \, \mu + \boldsymbol{z} \rangle \qquad \text{by (C.1)} && \text{(C.6)} \\
&= \langle \mathcal{F}_n^* \boldsymbol{q}, \mu \rangle + \langle \boldsymbol{q}, \boldsymbol{z} \rangle && \text{(C.7)} \\
&= \langle Q, \mu \rangle + \langle \boldsymbol{q}, \boldsymbol{z} \rangle && \text{(C.8)} \\
&= ||\mu||_{\mathrm{TV}} + \lambda \, ||\boldsymbol{z}||_1 \qquad \text{by (3.3) and (3.5).} && \text{(C.9)}
\end{aligned}$$

We conclude that $\{\mu, \boldsymbol{z}\}$ must be the unique solution.

C.1 Proof of Lemma C.1

If μ'_{T^c} and $\boldsymbol{z}'_{\Omega^c}$ both equal zero, then

$$\begin{aligned}
\mathcal{F}_n \, \mu + \boldsymbol{z} - \mathcal{F}_n \, \mu'_T - \boldsymbol{z}'_{\Omega} &= \mathcal{F}_n \, \mu' + \boldsymbol{z'} - \mathcal{F}_n \, \mu'_T - \boldsymbol{z}'_{\Omega} \qquad \text{by (C.1)} && \text{(C.10)} \\
&= \mathcal{F}_n \, \mu'_{T^c} + \boldsymbol{z}'_{\Omega^c} && \text{(C.11)} \\
&= 0. && \text{(C.12)}
\end{aligned}$$

We index the entries of $\Omega := \{i_1, i_2, \ldots, i_s\}$ and define the matrix $\begin{bmatrix} F_T & I_\Omega \end{bmatrix} \in \mathbb{C}^{n \times (k+s)}$, where

$$(F_T)_{lj} = e^{i 2 \pi l f_j} \quad \text{for } 1 \leq l \leq n, \; 1 \leq j \leq k, \tag{C.13}$$

$$(I_\Omega)_{lj} = \begin{cases} 1 & \text{if } l = i_j \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } 1 \leq l \leq n, \; 1 \leq j \leq s. \tag{C.14}$$

If $k+s \leq n$ then $\begin{bmatrix} F_T & I_\Omega \end{bmatrix}$ is full rank (this follows from the fact that F_T is a submatrix of a Vandermonde matrix). Equation (C.12) implies

$$\begin{bmatrix} F_T & I_\Omega \end{bmatrix} \begin{bmatrix} \boldsymbol{x} - \boldsymbol{x'} \\ \mathcal{P}_\Omega \boldsymbol{z} - \mathcal{P}_\Omega \boldsymbol{z'} \end{bmatrix} = 0, \tag{C.15}$$

where $\mathcal{P}_\Omega \boldsymbol{u'} \in \mathbb{C}^s$ is the subvector of $\boldsymbol{u'}$ containing the entries indexed by Ω and $\boldsymbol{x'} \in \mathbb{C}^T$ is the vector containing the amplitudes of μ' (recall that by assumption $\mu'_{T^c} = 0$). We conclude that $\mu = \mu'$ and $\boldsymbol{z} = \boldsymbol{z'}$.

D Proof of Lemma 3.4

The vector of coefficients $\boldsymbol{c}$ equals the convolution of three rectangles of widths $2 \cdot 0.247\, m + 1$, $2 \cdot 0.339\, m + 1$ and $2 \cdot 0.414\, m + 1$ and amplitudes $(2 \cdot 0.247\, m + 1)^{-1}$, $(2 \cdot 0.339\, m + 1)^{-1}$ and $(2 \cdot 0.414\, m + 1)^{-1}$. Some simple computations show that the amplitude of the convolution of three rectangles with unit amplitudes and widths $a_1 < a_2 < a_3$ is bounded by $a_1 a_2$. An immediate consequence is that the amplitude of $\boldsymbol{c}$ is bounded by

$$\begin{aligned} \|\boldsymbol{c}\|_\infty &\leq \frac{(2 \cdot 0.247\, m + 1)\,(2 \cdot 0.339\, m + 1)}{(2 \cdot 0.247\, m + 1)\,(2 \cdot 0.339\, m + 1)\,(2 \cdot 0.414\, m + 1)} && \text{(D.1)} \\ &\leq \frac{1}{(2 \cdot 0.414\, m + 1)} && \text{(D.2)} \\ &\leq \frac{1.3}{m}. && \text{(D.3)} \end{aligned}$$

E Proof of Lemma 3.6

To bound the operator norm of B_Ω, we control the behavior of

$$\begin{aligned} H &:= B_\Omega B_\Omega^* && \text{(E.1)} \\ &= \sum_{l \in \Omega} \boldsymbol{b}\,(l)\, \boldsymbol{b}\,(l)^*\,, && \text{(E.2)} \end{aligned}$$

which concentrates around a scaled version of

$$\bar{H} := \sum_{l=-m}^{m} \boldsymbol{b}\,(l)\, \boldsymbol{b}\,(l)^*\,. \tag{E.3}$$

The following lemma bounds the operator norm of $\bar{H}$.

Lemma E.1 (Proof in Section E.1)**.** *Under the assumptions of Theorem 2.2*

$$\left\|\bar{H}\right\| \leq 260\, \pi^2 n\, \log k. \tag{E.4}$$

By (2.12) $s \leq C_s\, n \left(\log k \log \frac{n}{\epsilon}\right)^{-1}$ which together with the lemma implies

$$\left\| \frac{s}{n} \bar{H} \right\| \leq \frac{C_B^2\, n}{2} \left(\log \frac{n}{\epsilon}\right)^{-1} \tag{E.5}$$

if we set C_s small enough. The following lemma uses the matrix Bernstein inequality to control the deviation of H from a scaled version of $\bar{H}$.

Lemma E.2 (Proof in Section E.2)**.** *Under the assumptions of Theorem 2.2*

$$\left\| H - \frac{s}{n} \bar{H} \right\| \leq \frac{C_B^2\, n}{2} \left(\log \frac{n}{\epsilon}\right)^{-1} \tag{E.6}$$

with probability at least $1 - \epsilon/5$.

We conclude that

$$\|B_\Omega\| \leq \sqrt{\|H\|} \tag{E.7}$$

$$\leq \sqrt{\frac{s}{n} \|\bar{H}\| + \left\| H - \frac{s}{n} \bar{H} \right\|} \tag{E.8}$$

$$\leq C_B \sqrt{n} \left(\log \frac{n}{\epsilon}\right)^{-\frac{1}{2}} \tag{E.9}$$

with probability at least $1 - \epsilon/5$ by the triangle inequality.

E.1 Proof of Lemma E.1

We express the matrix $\bar{H}$ in terms of the Dirichlet kernel $\mathcal{D}_m$ of order m defined in (3.25) and its derivatives,

$$\bar{H} = n \begin{bmatrix} \bar{H}_0 & \bar{H}_1 \\ -\bar{H}_1 & \bar{H}_2 \end{bmatrix} \tag{E.10}$$

where

$$\left(\bar{H}_0\right)_{jl} = \mathcal{D}_m\left(f_j - f_l\right), \quad \left(\bar{H}_1\right)_{jl} = \kappa\, \mathcal{D}_m^{(1)}\left(f_j - f_l\right), \quad \left(\bar{H}_2\right)_{jl} = -\kappa^2\, \mathcal{D}_m^{(2)}\left(f_j - f_l\right). \tag{E.11}$$

In order to bound the operator norm of $\bar{H}$ we first establish some bounds on $\mathcal{D}_m^{(\ell)}$ for $\ell = 0, 1, 2$. Due to how the kernel is normalized in (3.25), the magnitude of $\mathcal{D}_m$ is bounded by one. This yields a uniform bound on the magnitude of its derivatives by Bernstein's polynomial inequality.

Theorem E.3 (Bernstein's polynomial inequality [56])**.** *For any complex-valued polynomial P of degree N*

$$\sup_{|z| \leq 1} \left| P^{(1)}(z) \right| \leq N \sup_{|z| \leq 1} |P(z)|. \tag{E.12}$$

Applying the theorem, we have

$$\left| \mathcal{D}_m^{(\ell)}(f) \right| \leq (2\pi m)^\ell. \tag{E.13}$$

The following lemma allows us to control the tail of the Dirichlet kernel and its derivatives.

Lemma E.4 ([38, Section C.4]). *If $m \geq 10^3$, for $f \geq 80/m$*

$$\left|\mathcal{D}_m^{(\ell)}(f)\right| \leq \frac{1.1\, 2^{\ell-2}\pi^\ell m^{\ell-1}}{f}. \tag{E.14}$$

We now combine these two bounds to control the sum of the magnitudes of $\mathcal{D}_m^{(\ell)}$ when evaluated at T for $\ell = 0, 1, 2$. By the minimum-separation condition (2.10), if we fix $f_i \in T$ then there are at most 126 other frequencies in T that are at a distance of $80/m$ or less from f_i. We bound those terms using (E.13) and deal with the rest by applying Lemma E.4,

$$\sup_{f_i} \sum_{j=1}^{k} \kappa^\ell \left|\mathcal{D}_m^{(\ell)}(f_i - f_j)\right| \leq 126\pi^\ell\kappa^\ell \sup_f \left|\mathcal{D}_m^{(\ell)}(f)\right| + 2\kappa^\ell \sum_{j=1}^{k} \sup_{|f|\geq j\Delta_{\min}} \left|\mathcal{D}_m^{(\ell)}(f)\right| \tag{E.15}$$

$$\leq 126\pi^\ell + \frac{1}{m^{(l)}} \sum_{j=1}^{k} \frac{1.1\,\pi^\ell m^{\ell-1}}{4j\Delta_{\min}} \quad \text{by Lemma 3.3 and (E.13)} \tag{E.16}$$

$$\leq 130\pi^\ell \log k \quad \text{since } \Delta_{\min} := \frac{1.26}{m} \text{ and } \sum_{j=1}^{k} \frac{1}{j} \leq 1 + \log k \leq 2\log k \tag{E.17}$$

as long as k is larger than 2 (the argument can be easily modified if this is not the case).

By Gershgorin's circle theorem, the eigenvalues of $\bar{H}$, and consequently its operator norm, are bounded by

$$n \max_i \left\{ \sum_{j=1}^{k} \left|\mathcal{D}_m(f_i - f_j)\right| + \sum_{j=1}^{k} \kappa \left|\mathcal{D}_m^{(1)}(f_i - f_j)\right|, \right. \tag{E.18}$$

$$\left. \sum_{j=1}^{k} \kappa \left|\mathcal{D}_m^{(1)}(f_i - f_j)\right| + \sum_{j=1}^{k} \kappa^2 \left|\mathcal{D}_m^{(2)}(f_i - f_j)\right| \right\} \leq 260\,\pi^2 n \log k. \tag{E.19}$$

E.2 Proof of Lemma E.2

Under the assumptions of Theorem 2.2

$$H = \sum_{l=-m}^{m} \delta_\Omega(l)\, \boldsymbol{b}(l)\, \boldsymbol{b}(l)^*, \tag{E.20}$$

where $\delta_\Omega(-m)$, $\delta_\Omega(-m+1)$, $\ldots$, $\delta_\Omega(m)$ are iid Bernouilli random variables with parameter $\frac{s}{n}$. We control this sum of independent random matrices using the matrix Bernstein inequality.

Theorem E.5 (Matrix Bernstein inequality [71, Theorem 1.4]). *Let $\{X_l\}$ be a finite sequence of independent zero-mean self-adjoint random matrices of dimension d such that $\|X_l\| \leq B$ almost surely for a certain constant B. For all $t \geq 0$ and a positive constant σ^2*

$$\mathbb{P}\left\{\left\|\sum_{l=-m}^{m} X_l\right\| \geq t\right\} \leq d \exp\left(\frac{-t^2/2}{\sigma^2 + Bt/3}\right) \quad \text{as long as} \quad \left\|\sum_{l=-m}^{m} \mathbb{E}\left(X_l^2\right)\right\| \leq \sigma^2. \tag{E.21}$$

We apply the matrix Bernstein inequality to the finite sequence of independent adjoint zero-mean random matrices of the form

$$X_l := \left(\delta_\Omega(l) - \frac{s}{n}\right) \boldsymbol{b}(l)\, \boldsymbol{b}(l)^*, \quad -m \leq l \leq m. \tag{E.22}$$

These random matrices satisfy

$$H - \frac{s}{n}\bar{H} = \sum_{l=-m}^{m} X_l. \tag{E.23}$$

By Lemma 3.5

$$\|X_l\| \leq \sup_{-m\leq l\leq m} ||\boldsymbol{b}(l)||_2^2 \tag{E.24}$$

$$\leq B := 10\,k. \tag{E.25}$$

In addition,

$$\sigma^2 := \left\|\textstyle\sum_{l=-m}^{m} \mathbb{E}\left(X_l^2\right)\right\| \tag{E.26}$$

$$= \left\|\textstyle\sum_{l=-m}^{m} \mathbb{E}\left(\left(\bar{\delta}(l) - \frac{s}{n}\right)^2\right) ||\boldsymbol{b}(l)||_2^2\, \boldsymbol{b}(l)\, \boldsymbol{b}(l)^*\right\| \tag{E.27}$$

$$\leq 10\,k\,\frac{s}{n}\left\|\bar{H}\right\| \tag{E.28}$$

$$\leq 10\,C_B^2\,n\,k\left(\log\frac{n}{\epsilon}\right)^{-1} \tag{E.29}$$

by Lemma 3.5, (E.5) and the fact that the variance of a Bernouilli random variable of parameter p equals $p(1-p)$. Setting $t := \frac{C_B^2\,n}{2}\left(\log\frac{n}{\epsilon}\right)^{-1}$ in Theorem E.5, so that $\sigma^2 = 20\,k\,t$, yields

$$\mathbb{P}\left\{\left\|H - \frac{s}{n}\bar{H}\right\| \geq t\right\} \leq 2k\,\exp\left(\frac{-t^2/2}{\sigma^2 + Bt/3}\right) \tag{E.30}$$

$$= 2k\,\exp\left(\frac{-3t}{140\,k}\right). \tag{E.31}$$

The probability is smaller or equal to $\epsilon/5$ as long as

$$k \leq \frac{3\,C_B^2\,n}{280}\left(\log\frac{10\,k}{\epsilon}\log\frac{n}{\epsilon}\right)^{-1} \tag{E.32}$$

which holds by (2.11) if we set C_k small enough.

F Proof of Lemma 3.7

The proof uses the following concentration bound that controls the deviation of a sum of independent vectors.

Theorem F.1 (Vector Bernstein inequality [15, Theorem 2.6], [40, Theorem 12]). *Let $\mathcal{U} \subset \mathbb{R}^d$ be a finite sequence of independent zero-mean random vectors with $||\boldsymbol{u}||_2 \leq B$ almost surely and $\sum_{\boldsymbol{u}\in\mathcal{U}} \mathbb{E}\,||\boldsymbol{u}||_2^2 \leq \sigma^2$ for all $\boldsymbol{u} \in \mathcal{U}$, where B and σ^2 are positive constants. For all $t \geq 0$*

$$\mathbb{P}\left(\left|\left|\sum_{\boldsymbol{u}\in\mathcal{U}} \boldsymbol{u}\right|\right|_2 \geq t\right) \leq \exp\left(-\frac{t^2}{8\sigma^2} + \frac{1}{4}\right) \quad \textit{for} \;\; 0 \leq t \leq \frac{\sigma^2}{B}. \tag{F.1}$$

By the definitions of $\bar{K}$, K and $\boldsymbol{b}$ in (3.27), (3.38) and (3.45),

$$\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f) = \sum_{l=-m}^{m} (i2\pi\kappa l)^{\ell}\, \boldsymbol{c}_l\, e^{i2\pi l f} \boldsymbol{b}(l), \tag{F.2}$$

$$\boldsymbol{v}_{\boldsymbol{\ell}}(f) = \sum_{l=-m}^{m} \delta_{\Omega^c}(l)\, (i2\pi\kappa l)^{\ell}\, \boldsymbol{c}_l\, e^{i2\pi l f} \boldsymbol{b}(l), \tag{F.3}$$

where by assumption $\delta_{\Omega^c}(-m)$, …, $\delta_{\Omega^c}(m)$ are iid Bernoulli random variables with parameter $p := \frac{n-s}{n}$. This implies that the finite collection of zero-mean random vectors of the form

$$\boldsymbol{u}(\ell, l) := (\delta_{\Omega^c}(l) - p)\, (i2\pi\kappa l)^{\ell}\, \boldsymbol{c}_l\, e^{i2\pi l f} \boldsymbol{b}(l), \tag{F.4}$$

satisfy

$$\boldsymbol{v}_{\boldsymbol{\ell}}(f) - p\, \bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f) = \sum_{l=-m}^{m} \boldsymbol{u}(l). \tag{F.5}$$

We have

$$||\boldsymbol{u}(\ell, l)||_2 \leq \pi^3\, ||\boldsymbol{c}||_{\infty} \sup_{-m\leq l\leq m} ||\boldsymbol{b}(l)||_2 \quad \text{by Lemma (3.3) and } \ell \leq 3 \tag{F.6}$$

$$\leq B := \frac{128\sqrt{k}}{m} \quad \text{by Lemmas 3.4 and 3.5,} \tag{F.7}$$

as well as

$$\sum_{l=-m}^{m} \mathbb{E}\,||\boldsymbol{u}(\ell, l)||_2^2 = \sum_{l=-m}^{m} \mathbb{E}\left((\delta_{\Omega^c}(l) - p)^2\right) (2\pi\kappa l)^{2\ell}\, |\boldsymbol{c}_l|^2\, ||\boldsymbol{b}(l)||_2^2 \tag{F.8}$$

$$\leq \pi^6 n\, \mathbb{E}\left((\delta_{\Omega^c}(1) - p)^2\right) ||\boldsymbol{c}||_{\infty}^2 \sup_{-m\leq l\leq m} ||\boldsymbol{b}(l)||_2^2 \quad \text{by Lemma (3.3)} \tag{F.9}$$

$$\leq \sigma^2 := \frac{3.25\, 10^4\, k}{m}, \tag{F.10}$$

where the last inequality follow from Lemmas 3.4 and 3.5 and $\mathbb{E}\left((p - \delta_{\Omega^c}(l))^2\right) = p(1-p)$. By the vector Bernstein inequality for $0 \leq t \leq \sigma^2/B$ and the union bound we have

$$\mathbb{P}\left(\sup_{f\in\mathcal{G}} \|v_{\ell}(f) - p\bar{v}_{\ell}(f)\|_2 \geq t, \quad \ell \in \{0, 1, 2, 3\}\right) \leq 4\,|\mathcal{G}| \exp\left(-\frac{t^2}{8\sigma^2} + \frac{1}{4}\right). \tag{F.11}$$

To make the right-hand side smaller than $\epsilon/5$, we fix t to equal

$$t := \sigma \sqrt{8\left(\frac{1}{4} + \log \frac{20\,|\mathcal{G}|}{\epsilon}\right)}. \tag{F.12}$$

This choice of t is valid because

$$\begin{aligned}
\frac{t}{\sigma} &= \sqrt{8\left(\frac{1}{4} + \log \frac{20\,|\mathcal{G}|}{\epsilon}\right)} && \text{(F.13)}\\
&\leq \sqrt{74 + 16 \log n + 8 \log \frac{1}{\epsilon}} && \text{(F.14)}\\
&\leq 0.315\sqrt{n} + \sqrt{8 \log \frac{1}{\epsilon}} && \text{(F.15)}\\
&\leq 0.32\sqrt{n}. && \text{(F.16)}
\end{aligned}$$

Inequality (F.15) follows from the fact that $\sqrt{74 + 16 \log n} \leq 0.315\sqrt{n}$ for $n \geq 2\,10^3$. Inequality (F.16) holds by (2.11) and (2.12) as long as we set C_k and C_s small enough and either $k \geq 1$ or $s \geq 1$. This establishes that t/σ is smaller than $0.32\sqrt{n} \leq \sigma/B$.

We conclude that the desired bound holds as long as

$$C_{\boldsymbol{v}} \left(\log \frac{n}{\epsilon}\right)^{-\frac{1}{2}} \geq t \geq \sqrt{\frac{2\,10^3\,k}{n}\left(\frac{1}{4} + \log \frac{8\,10^3\,n^2}{\epsilon}\right)}, \tag{F.17}$$

which is the case by (2.11) if we set C_k small enough.

G Proof of Lemma 3.8

The proof is based on the proof of Lemma 4.4 in [66]. The following lemma establishes that $\bar{D}$ is invertible and close to the identity.

Lemma G.1 (Proof in Section G.1)**.** *Under the assumptions of Theorem 2.2*

$$\begin{aligned}
\left\|I - \bar{D}\right\| &\leq 0.468, && \text{(G.1)}\\
\left\|\bar{D}\right\| &\leq 1.468, && \text{(G.2)}\\
\left\|\bar{D}^{-1}\right\| &\leq 1.88. && \text{(G.3)}
\end{aligned}$$

By the definition of $\bar{K}$ and K in (3.27) and (3.38) respectively we can write D and $\bar{D}$ as sums of self-adjoint matrices,

$$\begin{aligned}
\bar{D} &= \sum_{l=-m}^{m} \boldsymbol{c}_l\, \boldsymbol{b}\,(l)\, \boldsymbol{b}\,(l)^*\,, && \text{(G.4)}\\
D &= \sum_{l=-m}^{m} \delta_{\Omega^c}\,(l)\; \boldsymbol{c}_l\, \boldsymbol{b}\,(l)\, \boldsymbol{b}\,(l)^*\,, && \text{(G.5)}
\end{aligned}$$

where by assumption $\delta_{\Omega^c}(-m), \ldots, \delta_{\Omega^c}(m)$ are iid Bernoulli random variables with parameter $p := \frac{n-s}{n}$. In the following lemma we leverage the matrix Bernstein inequality to establish that D concentrates around $p\,\bar{D}$.

Lemma G.2 (Proof in Section G.2)**.** *Under the assumptions of Theorem 2.2*

$$\left\|D - p\bar{D}\right\| \geq \frac{p}{4} \min\left\{1, \frac{C_D}{4}\left(\log\frac{n}{\epsilon}\right)^{-\frac{1}{2}}\right\}. \tag{G.6}$$

with probability at most $\epsilon/5$.

Applying the triangle inequality together with Lemma G.1 allows to lower bound the smallest singular value of D under the assumption that (G.6) holds

$$\begin{aligned}\frac{\sigma_{\min}(D)}{p} &\geq \sigma_{\min}(I) - \left\|I - \bar{D}\right\| - \frac{1}{p}\left\|D - p\bar{D}\right\| &\text{(G.7)}\\ &\geq 0.282. &\text{(G.8)}\end{aligned}$$

This proves that D is invertible. To complete the proof we borrow two inequalities from [66].

Lemma G.3 ([66, Appendix E])**.** *For any matrices* A *and* B *such that* B *is invertible and*

$$\|A - B\|\left\|B^{-1}\right\| \leq \frac{1}{2} \tag{G.9}$$

we have

$$\begin{aligned}\left\|A^{-1}\right\| &\leq 2\left\|B^{-1}\right\|, &\text{(G.10)}\\ \left\|A^{-1} - B^{-1}\right\| &\leq 2\left\|B^{-1}\right\|^2\|A - B\|. &\text{(G.11)}\end{aligned}$$

We set $A := D$ and $B := p\bar{D}$. By Lemmas G.1 and Lemma G.2,

$$\left\|D - p\bar{D}\right\|\left\|\left(p\bar{D}\right)^{-1}\right\| \leq \frac{1}{2} \tag{G.12}$$

with probability at least $1 - \epsilon/5$. Lemmas G.1, G.2 and G.3 then imply

$$\begin{aligned}\left\|D^{-1}\right\| &\leq 2\left\|\left(p\bar{D}\right)^{-1}\right\| &\text{(G.13)}\\ &\leq \frac{4}{p}, &\text{(G.14)}\\ \left\|D^{-1} - \left(p\bar{D}\right)^{-1}\right\| &\leq 2\left\|\left(p\bar{D}\right)^{-1}\right\|^2\left\|D - p\bar{D}\right\| &\text{(G.15)}\\ &\leq \frac{C_D}{2p}\left(\log\frac{n}{\epsilon}\right)^{-\frac{1}{2}}, &\text{(G.16)}\end{aligned}$$

with the same probability. Finally, if $s \leq n/2$, which is the case by (2.12), we have $1/p \leq 2$ and the proof is complete.

G.1 Proof of Lemma G.1

The following bounds on the submatrices of $\bar{D}$ are obtained by combining Lemma 3.3 with some results borrowed from [38].

Lemma G.4 ([38, Section 4.2]). *Under the assumptions of Theorem 2.2*

$$\left|\left|I-\bar{D}_0\right|\right|_\infty \leq 1.855\,10^{-2}, \tag{G.17}$$
$$\left|\left|\bar{D}_1\right|\right|_\infty \leq 5.148\,10^{-2}, \tag{G.18}$$
$$\left|\left|I-\bar{D}_2\right|\right|_\infty \leq 0.416. \tag{G.19}$$

Following a similar argument as in Appendix C of [66] yields the desired result:

$$\left\|I-\bar{D}\right\| \leq \left|\left|I-\bar{D}\right|\right|_\infty \tag{G.20}$$
$$\leq \max\left\{\left|\left|I-\bar{D}_0\right|\right|_\infty + \left|\left|\bar{D}_1\right|\right|_\infty, \left|\left|I-\bar{D}_2\right|\right|_\infty + \left|\left|\bar{D}_1\right|\right|_\infty\right\} \tag{G.21}$$
$$\leq 0.468, \tag{G.22}$$
$$\left\|\bar{D}\right\| \leq 1+\left\|I-\bar{D}\right\| \leq 1.468, \tag{G.23}$$
$$\left\|\bar{D}^{-1}\right\| \leq \frac{1}{1-\left|\left|I-\bar{D}\right|\right|_\infty} \leq 1.88. \tag{G.24}$$

G.2 Proof of Lemma G.2

We define

$$X_l := \left(p-\delta_{\Omega^c}(l)\right)\boldsymbol{c}_l\,\boldsymbol{b}(l)\,\boldsymbol{b}(l)^T, \tag{G.25}$$

which has zero mean since

$$\mathbb{E}(X_l) = \left(p-\mathbb{E}\left(\delta_{\Omega^c}(l)\right)\right)\boldsymbol{c}_l\,\boldsymbol{b}(l)\,\boldsymbol{b}(l)^T \tag{G.26}$$
$$= 0. \tag{G.27}$$

By the proofs of Lemmas 3.4 and 3.5, for any $-m \leq l \leq m$,

$$\|X_l\| \leq \max_{-m\leq l\leq m}\left\|\boldsymbol{c}_l\,\boldsymbol{b}(l)\,\boldsymbol{b}(l)^T\right\| \tag{G.28}$$
$$\leq ||\boldsymbol{c}||_\infty \max_{-m\leq l\leq m}||\boldsymbol{b}(l)||_2^2 \tag{G.29}$$
$$\leq B := \frac{12.6\,k}{m}. \tag{G.30}$$

Also, $\mathbb{E}\left(\left(p-\delta_{\Omega^c}(l)\right)^2\right) = p(1-p)$, which implies

$$\mathbb{E}\left(X_l^2\right) = p(1-p)\,\boldsymbol{c}_l^2\,||\boldsymbol{b}(l)||_2^2\,\boldsymbol{b}(l)\,\boldsymbol{b}(l)^T. \tag{G.31}$$

Since $\boldsymbol{c}_l \geq 0$ for all l ($\boldsymbol{c}$ is the convolution of three positive rectangular pulses),

$$\sum_{l=-m}^{m} \boldsymbol{c}_l^2 \left|\left|\boldsymbol{b}\left(l\right)\right|\right|_2^2 \boldsymbol{b}\left(l\right) \boldsymbol{b}\left(l\right)^T \preceq \left|\left|\boldsymbol{c}\right|\right|_\infty \max_{-m \leq l \leq m} \left|\left|\boldsymbol{b}\left(l\right)\right|\right|_2^2 \sum_{l=-m}^{m} \boldsymbol{c}_l \boldsymbol{b}\left(l\right) \boldsymbol{b}\left(l\right)^T \tag{G.32}$$

$$\preceq \frac{12.6\, k}{m} \bar{D} \quad \text{by Lemma 3.4 and 3.5,} \tag{G.33}$$

so that

$$\sum_{l=-m}^{m} \mathbb{E}\left(X_l^2\right) \leq p \left\| \sum_{l=-m}^{m} \boldsymbol{c}_l^2 \left|\left|\boldsymbol{b}\left(l\right)\right|\right|_2^2 \boldsymbol{b}\left(l\right) \boldsymbol{b}\left(l\right)^T \right\| \tag{G.34}$$

$$\leq \frac{12.6\, p\, k \left\|\bar{D}\right\|}{m} \tag{G.35}$$

$$\leq \sigma^2 := \frac{18.5\, p\, k}{m} \quad \text{by Lemma G.1.} \tag{G.36}$$

Setting $t = \frac{p}{4} C_{\min} \left(\log \frac{n}{\epsilon}\right)^{-\frac{1}{2}}$ where $C_{\min} := \min\left\{1, C_D/4\right\}$, the matrix Bernstein inequality from Theorem E.5 implies that

$$\mathbb{P}\left\{\left\|D^{-1} - p\bar{D}^{-1}\right\| > t\right\} \leq 2\, k\, \exp\left(-\frac{C_{\min}^2 p\, m}{32\, k}\left(18.5 \log\frac{n}{\epsilon} + 1.05\, C_{\min}\sqrt{\log\frac{n}{\epsilon}}\right)^{-1}\right)$$

$$\leq 2\, k\, \exp\left(-\frac{C_D'\left(n-s\right)}{k \log\frac{n}{\epsilon}}\right) \tag{G.37}$$

for a small enough constant C_D'. This probability is smaller than $\epsilon/5$ as long as

$$k \leq \frac{C_D'\, n}{2}\left(\log\frac{10k}{\epsilon}\log\frac{n}{\epsilon}\right)^{-1}, \tag{G.38}$$

$$s \leq \frac{n}{2}, \tag{G.39}$$

which holds by (2.11) and (2.12) if we set C_k and C_s small enough.

H Proof of Proposition 3.10

We begin by expressing $Q^{(\ell)}$ and $\bar{Q}^{(\ell)}$ in terms of $\boldsymbol{h}$ and $\boldsymbol{r}$,

$$\kappa^\ell\, \bar{Q}^{(\ell)}\left(f\right) := \kappa^\ell \sum_{j=1}^{k} \bar{\boldsymbol{\alpha}}_j\, \bar{K}^{(\ell)}\left(f - f_j\right) + \kappa^{\ell+1} \sum_{j=1}^{k} \bar{\boldsymbol{\beta}}_j\, \bar{K}^{(\ell+1)}\left(f - f_j\right) \tag{H.1}$$

$$= \bar{\boldsymbol{v}}_{\boldsymbol{\ell}}\left(f\right)^T \bar{D}^{-1} \begin{bmatrix} \boldsymbol{h} \\ 0 \end{bmatrix}, \tag{H.2}$$

$$\kappa^\ell\, Q^{(\ell)}\left(f\right) := \kappa^\ell \sum_{j=1}^{k} \boldsymbol{\alpha}_j\, K^{(\ell)}\left(f - f_j\right) + \kappa^{\ell+1} \sum_{j=1}^{k} \boldsymbol{\beta}_j\, K^{(\ell+1)}\left(f - f_j\right) + \kappa^\ell\, R^{(\ell)}\left(f\right) \tag{H.3}$$

$$= \boldsymbol{v}_{\boldsymbol{\ell}}\left(f\right)^T D^{-1}\left(\begin{bmatrix} \boldsymbol{h} \\ 0 \end{bmatrix} - \frac{1}{\sqrt{n}} B_\Omega\, \boldsymbol{r}\right) + \kappa^\ell\, R^{(\ell)}\left(f\right). \tag{H.4}$$

The difference between $Q^{(\ell)}$ and $\bar{Q}^{(\ell)}$ can be decomposed into several terms,

$$\kappa^\ell \, Q^{(\ell)}(f) = \kappa^\ell \, \bar{Q}^{(\ell)}(f) + \kappa^\ell \, R^{(\ell)}(f) + I_1^{(\ell)}(f) + I_2^{(\ell)}(f) + I_3^{(\ell)}(f), \tag{H.5}$$

$$I_1^{(\ell)}(f) := -\frac{1}{\sqrt{n}} \boldsymbol{v}_{\boldsymbol{\ell}}(f)^T D^{-1} B_\Omega \, \boldsymbol{r}, \tag{H.6}$$

$$I_2^{(\ell)}(f) := \left(\boldsymbol{v}_{\boldsymbol{\ell}}(f) - \frac{n-s}{n} \bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f) \right)^T D^{-1} \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{bmatrix}, \tag{H.7}$$

$$I_3^{(\ell)}(f) := \frac{n-s}{n} \bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)^T \left(D^{-1} - \frac{n}{n-s} \bar{D}^{-1} \right) \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{bmatrix}. \tag{H.8}$$

The following lemma provides bounds on these terms that hold with high probability in every point of a grid $\mathcal{G}$ that discretizes the unit interval.

Lemma H.1 (Proof in Section H.1)**.** *Conditioned on $\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c$, the events*

$$\mathcal{E}_R := \left\{ \sup_{f \in \mathcal{G}} \left| \kappa^\ell R^{(\ell)}(f) \right| \geq \frac{10^{-2}}{8}, \ell = 0, 1, 2, 3 \right\} \tag{H.9}$$

and

$$\mathcal{E}_i := \left\{ \sup_{f \in \mathcal{G}} \left| I_i^{(\ell)}(f) \right| \geq \frac{10^{-2}}{8}, \ell = 0, 1, 2, 3 \right\} \quad i = 1, 2, 3 \tag{H.10}$$

where $\mathcal{G} \subseteq [0, 1]$ is an equispaced grid with cardinality $|\mathcal{G}| = 400n^2$ occur each with probability at most $\epsilon/20$ under the assumptions of Theorem 2.2.

By the triangle inequality, Lemma H.1 implies

$$\sup_{f \in \mathcal{G}} \left| \kappa^\ell \, Q^{(\ell)}(f) - \kappa^\ell \, \bar{Q}^{(\ell)}(f) \right| \leq \frac{10^{-2}}{2} \tag{H.11}$$

with probability at least $1 - \epsilon/5$ conditioned on $\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c$.

We have controlled the deviation between $Q^{(\ell)}$ and $\bar{Q}^{(\ell)}$ on a fine grid. The following result extends the bound to the whole unit interval.

Lemma H.2 (Proof in Section H.3)**.** *Under the assumptions of Theorem 2.2*

$$\left| \kappa^\ell Q^{(\ell)}(f) - \kappa^\ell \bar{Q}^{(\ell)}(f) \right| \leq 10^{-2} \quad \text{for } \ell \in \{0, 1, 2\}. \tag{H.12}$$

This bound suffices to establish the desired result for values of f that lie away from T. Let us define

$$\mathcal{S}_{\text{near}} := \{ f \mid |f - f_j| \leq 0.09 \text{ for some } f_j \in T \}, \tag{H.13}$$

$$\mathcal{S}_{\text{far}} := [0, 1] / \mathcal{S}_{\text{near}}. \tag{H.14}$$

Section 4 of [38] provides a bound on $\bar{Q}$ which holds over all of $\mathcal{S}_{\text{far}}$ under the minimum-separation condition (2.10) (see Figure 12 in [38] as well as the code that supplements [38]).

Proposition H.3 (Bound on $\bar{Q}$ [38, Section 4])**.** *Under the assumptions of Theorem 2.2*

$$\left|\bar{Q}\left(f\right)\right| < 0.99 \quad f \in \mathcal{S}_{\text{far}}. \tag{H.15}$$

Combining Lemma H.2 and Proposition H.3

$$\begin{aligned} \left|Q\left(f\right)\right| &\leq \left|\bar{Q}\left(f\right)\right| + 10^{-2} && \text{(H.16)} \\ &< 1 \quad \text{for all } f \in \mathcal{S}_{\text{far}}. && \text{(H.17)} \end{aligned}$$

To bound Q in $\mathcal{S}_{\text{near}}$ we recall that by Corollary 3.9 in $\mathcal{E}_D^c$ $\left|Q\left(f_j\right)\right|^2 = 1$ and

$$\begin{aligned} \frac{\mathrm{d}\left|Q\left(f_j\right)\right|^2}{\mathrm{d}f} &= 2\,Q_R^{(1)}\left(f_j\right)Q_R\left(f_j\right) + 2\,Q_I^{(1)}\left(f_j\right)Q_I\left(f_j\right) && \text{(H.18)} \\ &= 0 && \text{(H.19)} \end{aligned}$$

for every f_j in T. Let $\tilde{f}$ be the element in T that is closest to an arbitrary f belonging to $\mathcal{S}_{\text{near}}$. The second-order bound

$$\left|Q\left(f\right)\right|^2 \leq 1 + \left(f - \tilde{f}\right)^2 \sup_{f \in \mathcal{S}_{\text{near}}} \frac{\mathrm{d}^2\left|Q\left(f\right)\right|^2}{\mathrm{d}f^2} \tag{H.20}$$

implies that we only need to show that $|Q|^2$ is concave in $\mathcal{S}_{\text{near}}$ to complete the proof. First, we bound the derivatives of $\bar{Q}$ and Q using Bernstein's polynomial inequality.

Lemma H.4. *Under the assumptions of Theorem 2.2, for any $\ell = 0, 1, 2, \ldots$*

$$\begin{aligned} \sup_{f \in [0,1]} \left|\kappa^\ell \bar{Q}^{(\ell)}\left(f\right)\right| &\leq 1, && \text{(H.21)} \\ \sup_{f \in [0,1]} \left|\kappa^\ell Q^{(\ell)}\left(f\right)\right| &\leq 1.01. && \text{(H.22)} \end{aligned}$$

Proof. $\bar{Q}$ is a trigonometric polynomial of degree m and its magnitude is bounded by one (see Proposition 2.3 in [38]). Combining Theorem E.3 and Lemma 3.3 yields (H.21). The triangle inequality, Lemma H.2 and (H.21) imply (H.22). □

Section 4 of [38] also provides a bound on the second derivative of $\left|\bar{Q}\right|^2$ which holds over all of $\mathcal{S}_{\text{near}}$ under the minimum-separation condition (2.10) (again, see Figure 12 in [38] as well as the code that supplements [38]).

Proposition H.5 (Bound on the second derivative of $\left|\bar{Q}\right|$ [38, Section 4])**.** *Under the assumptions of Theorem 2.2*

$$\frac{\mathrm{d}^2\left|\bar{Q}\left(f\right)\right|^2}{\mathrm{d}f^2} \leq -0.8\,m^2 \quad f \in \mathcal{S}_{\text{near}}. \tag{H.23}$$

Combining Proposition H.5, Lemma H.4 and the triangle inequality, as well as the lower bound on κ from Lemma 3.3, allows us to conclude that the second derivative of $\left|\bar{Q}\right|^2$ is negative in $\mathcal{S}_{\text{near}}$. Indeed, for any $f \in \mathcal{S}_{\text{near}}$

$$\frac{\kappa^2}{2}\frac{\mathrm{d}^2\left|Q\left(f\right)\right|^2}{\mathrm{d}f^2} = \kappa^2 Q_R^{(2)}\left(f\right)Q_R\left(f\right) + \kappa^2 Q_I^{(2)}\left(f\right)Q_I\left(f\right) + \left|\kappa\, Q^{(1)}\left(f\right)\right|^2 \tag{H.24}$$

$$\begin{aligned}
&\leq \frac{\kappa^2}{2}\frac{\mathrm{d}^2\left|\bar{Q}\left(f\right)\right|^2}{\mathrm{d}f^2} + 2\left|\kappa^2 Q^{(2)}\left(f\right) - \kappa^2\bar{Q}^{(2)}\left(f\right)\right|\sup_{f'}\left|Q\left(f'\right)\right| \\
&\quad + 2\left|Q\left(f\right) - \bar{Q}\left(f\right)\right|\sup_{f'}\left|\kappa^2\bar{Q}^{(2)}\left(f'\right)\right| \\
&\quad + 2\left|\kappa\, Q^{(1)}\left(f\right) - \kappa\,\bar{Q}^{(1)}\left(f\right)\right|\left(\sup_{f'}\left|\kappa\, Q^{(1)}\left(f'\right)\right| + \sup_{f'}\left|\kappa\,\bar{Q}^{(1)}\left(f'\right)\right|\right) \qquad \text{(H.25)} \\
&\leq -0.087 + 2\cdot 10^{-2}(4 + 2\cdot 10^{-2}) \qquad \text{(H.26)} \\
&< 0. \qquad \text{(H.27)}
\end{aligned}$$

H.1 Proof of Lemma H.1

Following an argument used in [66] (see also [16]), we use Hoeffding's inequality to bound the different terms.

Theorem H.6 (Hoeffding's inequality)**.** *Let the components of $\tilde{\boldsymbol{u}}$ be sampled i.i.d. from a symmetric distribution on the complex unit circle. For any $t > 0$ and any vector $\boldsymbol{u}$*

$$\mathbb{P}\left(\left|\langle\tilde{\boldsymbol{u}}, \boldsymbol{u}\rangle\right| \geq \tilde{\epsilon}\right) \leq 4\exp\left(-\frac{\tilde{\epsilon}^2}{4\left|\left|\boldsymbol{u}\right|\right|_2^2}\right). \tag{H.28}$$

Corollary H.7. *Let the components of $\tilde{\boldsymbol{u}}$ be sampled i.i.d. from a symmetric distribution on the complex unit circle. For any finite collection of vectors $\mathcal{U}$ with cardinality $4\left|G\right| = 1600n^2$ the event*

$$\mathcal{E} := \left\{\left|\langle\tilde{\boldsymbol{u}}, \boldsymbol{u}\rangle\right| > \frac{10^{-2}}{8} \quad \textit{for all} \quad \boldsymbol{u} \in \mathcal{U}\right\} \tag{H.29}$$

has probability at most $\epsilon/20$ as long as

$$\left|\left|\boldsymbol{u}\right|\right|_2^2 \leq C_{\mathcal{U}}^2\left(\log\frac{n}{\epsilon}\right)^{-1} \quad \textit{for all} \quad \boldsymbol{u} \in \mathcal{U}, \tag{H.30}$$

where $C_{\mathcal{U}} := 1/5000$.

Proof. The result follows directly from the proposition and the union bound. □

Bound on $\mathbb{P}\left(\mathcal{E}_R|\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c\right)$

We consider the family of vectors

$$\boldsymbol{u}\left(\ell, f\right) := \frac{\kappa^\ell}{\sqrt{n}}\left[\left(i2\pi l_1\right)^\ell e^{i2\pi l_1 f} \quad \left(i2\pi l_2\right)^\ell e^{i2\pi l_2 f} \quad \cdots \quad \left(i2\pi l_s\right)^\ell e^{i2\pi l_s f}\right]^T \tag{H.31}$$

where $\ell \in \{0,1,2,3\}$ and f belongs to $\mathcal{G}$, so that $|\mathcal{U}| = 4\,|\mathcal{G}|$. We have

$$||\boldsymbol{u}\,(\ell,f)||_2^2 \leq \frac{\kappa^{2\ell}\,(2\pi m)^{2\ell}\,s}{n} \tag{H.32}$$

$$\leq \frac{\pi^6 s}{n} \quad \text{by Lemma 3.3} \tag{H.33}$$

$$\leq C_{\mathcal{U}}^2 \left(\log \frac{n}{\epsilon}\right)^{-1} \quad \text{by (2.12) if we set } C_s \text{ small enough.} \tag{H.34}$$

The desired result follows by Corollary H.7 because

$$\kappa^{\ell} R^{(\ell)}\,(f) = \langle \boldsymbol{r}, \boldsymbol{u}\,(\ell,f)\rangle\,. \tag{H.35}$$

Bound on $\mathbb{P}\,(\mathcal{E}_1|\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c)$

We have

$$I_1^{(\ell)}\,(f) = \langle \boldsymbol{u}\,(\ell,f)\,, \boldsymbol{r}\rangle\,, \qquad \boldsymbol{u}\,(\ell,f) := -\frac{1}{\sqrt{n}}\,B_\Omega^*\,D^{-1}\boldsymbol{v_\ell}\,(f)\,, \tag{H.36}$$

where $\ell \in \{0,1,2,3\}$ and f belongs to $\mathcal{G}$, so that $|\mathcal{U}| = 4\,|\mathcal{G}|$.

To bound $||\boldsymbol{u}\,(\ell,f)||_2$ we leverage a bound on the ℓ_2 norm of $\boldsymbol{v_\ell}$ which follows from Lemma 3.7 and the following bound on the ℓ_2 norm of $\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}$.

Lemma H.8 (Proof in Section H.2)**.** *Under the assumptions of Theorem 2.2, there is a fixed numerical constant* $C_{\bar{\boldsymbol{v}}}$ *such that for any* f

$$||\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}\,(f)||_2 \leq C_{\bar{\boldsymbol{v}}}. \tag{H.37}$$

Corollary H.9. *In* $\mathcal{E}_v^c$ *for any* $f \in \mathcal{G}$

$$||\boldsymbol{v_\ell}\,(f)||_2 \leq C_{\bar{\boldsymbol{v}}} + C_{\boldsymbol{v}}. \tag{H.38}$$

Proof. The result follows from the lemma, the triangle inequality and Lemma 3.7. □

Combining Lemma 3.8 and Corollary H.9 yields

$$||\boldsymbol{u}\,(\ell,f)||_2 \leq \frac{1}{\sqrt{n}}\,\|B_\Omega\|\,\left\|D^{-1}\right\|\,||\boldsymbol{v_\ell}\,(f)||_2 \tag{H.39}$$

$$\leq \frac{8\,(C_{\bar{\boldsymbol{v}}} + C_{\boldsymbol{v}})\,\|B_\Omega\|}{\sqrt{n}} \tag{H.40}$$

in $\mathcal{E}_D^c \cap \mathcal{E}_v^c$. Corollary H.7 implies the desired result if

$$\|B_\Omega\| \leq C_B \left(\log \frac{n}{\epsilon}\right)^{-\frac{1}{2}} \sqrt{n}, \quad C_B := \frac{C_{\mathcal{U}}}{8\,(C_{\bar{\boldsymbol{v}}} + C_{\boldsymbol{v}})}, \tag{H.41}$$

which is the case in $\mathcal{E}_B^c$ by Lemma 3.6.

Bound on $\mathbb{P}\left(\mathcal{E}_2|\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c\right)$

We have

$$I_2^{(\ell)}(f) = \langle \boldsymbol{u}(\ell, f), \boldsymbol{h} \rangle, \qquad \boldsymbol{u}(\ell, f) := PD^{-1}\left(\boldsymbol{v}_{\boldsymbol{\ell}}(f) - \frac{n-s}{n}\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)\right) \tag{H.42}$$

where $P \in \mathbb{R}^{k \times 2k}$ is the projection matrix that selects the first k entries in a vector, $\ell \in \{0, 1, 2, 3\}$ and f belongs to $\mathcal{G}$, so that $|\mathcal{U}| = 4|\mathcal{G}|$.

Since $\|P\| = 1$, by Lemma 3.8 in $\mathcal{E}_D^c$

$$||\boldsymbol{u}(\ell, f)||_2 \leq \|P\| \left\|D^{-1}\right\| \left|\left|\boldsymbol{v}_{\boldsymbol{\ell}}(f) - \frac{n-s}{n}\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)\right|\right|_2 \tag{H.43}$$

$$\leq 8 \left|\left|\boldsymbol{v}_{\boldsymbol{\ell}}(f) - \frac{n-s}{n}\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)\right|\right|_2. \tag{H.44}$$

The desired result holds if

$$\left|\left|\boldsymbol{v}_{\boldsymbol{\ell}}(f) - \frac{n-s}{n}\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)\right|\right|_2 \leq C_{\boldsymbol{v}}\left(\log\frac{n}{\epsilon}\right)^{-\frac{1}{2}}, \quad C_{\boldsymbol{v}} := \frac{C_{\mathcal{U}}}{8}, \tag{H.45}$$

which is the case in $\mathcal{E}_v^c$ by Lemma 3.7.

Bound on $\mathbb{P}\left(\mathcal{E}_3|\mathcal{E}_B^c \cap \mathcal{E}_D^c \cap \mathcal{E}_v^c\right)$

We have

$$I_3^{(\ell)}(f) = \langle \boldsymbol{u}(\ell, f), \boldsymbol{h} \rangle, \qquad \boldsymbol{u}(\ell, f) := \frac{n-s}{n} P\left(D^{-1} - \frac{n}{n-s}\bar{D}^{-1}\right)\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f) \tag{H.46}$$

where $\ell \in \{0, 1, 2, 3\}$ and f belongs to $\mathcal{G}$, so that $|\mathcal{U}| = 4|\mathcal{G}|$.

Since $\|P\| = 1$, by Lemma 3.7

$$||\boldsymbol{u}(\ell, f)||_2 \leq \|P\| \left\|D^{-1} - \frac{n}{n-s}\bar{D}^{-1}\right\| ||\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)||_2 \tag{H.47}$$

$$\leq C_{\bar{v}} \left\|D^{-1} - \frac{n}{n-s}\bar{D}^{-1}\right\|. \tag{H.48}$$

The desired result holds if

$$\left\|D^{-1} - \frac{n}{n-s}\bar{D}^{-1}\right\| \leq C_D\left(\log\frac{n}{\epsilon}\right)^{-\frac{1}{2}}, \quad C_D := \frac{C_{\mathcal{U}}}{C_{\bar{v}}}, \tag{H.49}$$

for a fixed numerical constant C_D, which is the case in $\mathcal{E}_D^c$ by Lemma 3.8.

H.2 Proof of Lemma H.8

We use the ℓ_1 norm to bound the ℓ_2 norm of $\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)$:

$$||\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)||_2 \leq ||\bar{\boldsymbol{v}}_{\boldsymbol{\ell}}(f)||_1 \tag{H.50}$$

$$= \sum_{j=1}^{k} \kappa^{\ell} \left| \bar{K}^{(\ell)}(f - f_j) \right| + \sum_{j=1}^{k} \kappa^{\ell+1} \left| \bar{K}^{(\ell+1)}(f - f_j) \right|. \tag{H.51}$$

To bound the sum on the right we leverage some results from [38].

Lemma H.10.

$$\kappa^{\ell} \left| \bar{K}^{(\ell)}(f) \right| \leq \begin{cases} C_1 & \forall f \in [-\frac{1}{2}, \frac{1}{2}], \\ C_2\, m^{-3}\, |f|^{-3} & \text{if} \quad \frac{80}{m} \leq |f| \leq \frac{1}{2}, \end{cases} \tag{H.52}$$

for suitably chosen numerical constant C_1 and C_2.

Proof. The constant bound on the kernel follows from Corollary 4.5, Lemma 4.6 and Lemma C.2 in [38] (see also Figures 14 and 15 in the same paper). The bound for large f follows from Lemma C.2 in [38]. □

By the minimum-separation condition (2.10), there are at most 127 elements of T that are at a distance of $80/m$ or less from f. We use the first bound in (H.52) to control the contribution of those elements and the second bound to deal with the remaining terms,

$$\sum_{j=1}^{k} \kappa^{\ell} \left| \bar{K}^{(\ell)}(f - f_j) \right| \leq \sum_{j : |f - f_j| < \frac{80}{m}} C_1 + \sum_{j : \frac{80}{m} \leq |f - f_j| \leq \frac{1}{2}} \frac{C_2}{m^3 \, |f - f_j|^3} \tag{H.53}$$

$$\leq 127\, C_1 + 2\, C_2 \sum_{j=1}^{\infty} \frac{1}{m^3 (j \Delta_{\min})^3} \tag{H.54}$$

$$\leq 127\, C_1 + 2\, C_2 \sum_{j=1}^{\infty} \frac{1}{j^3} \tag{H.55}$$

$$= 127\, C_1 + 2\, C_2\, \zeta(3), \tag{H.56}$$

where $\zeta(3)$ is Apéry's constant, which is bounded by 1.21. This completes the proof.

H.3 Proof of Lemma H.2

The proof follows a similar argument to the proof of Proposition 4.12 in [66]. We begin by bounding the deviations of $Q^{(\ell)}$ and $\bar{Q}^{(\ell)}$ on neighboring points.

Lemma H.11 (Proof in Section H.3.1)**.** *Under the assumptions of Theorem 2.2, for any f_1, f_2 in the unit interval*

$$\left| \kappa^{\ell} Q^{(\ell)}(f_2) - \kappa^{\ell} Q^{(\ell)}(f_1) \right| \leq n^2 \, |f_2 - f_1| \,, \tag{H.57}$$

$$\left| \kappa^{\ell} \bar{Q}^{(\ell)}(f_2) - \kappa^{\ell} \bar{Q}^{(\ell)}(f_1) \right| \leq n^2 \, |f_2 - f_1| \,. \tag{H.58}$$

For any f in the unit interval there exists a grid point $f_{\mathcal{G}}$ such that the distance between the two points is smaller than the step size $\left(400\, n^2\right)^{-1}$. This allows to establish the desired result by combining (H.11) with Lemma H.11 and the triangle inequality,

$$\begin{aligned}
\left|\kappa^{\ell} Q^{(\ell)}(f) - \kappa^{\ell} \bar{Q}^{(\ell)}(f)\right| &\leq \left|\kappa^{\ell} Q^{(\ell)}(f) - \kappa^{\ell} Q^{(\ell)}(f_{\mathcal{G}})\right| + \left|\kappa^{\ell} Q^{(\ell)}(f_{\mathcal{G}}) - \kappa^{\ell} \bar{Q}^{(\ell)}(f_{\mathcal{G}})\right| && \text{(H.59)}\\
&\quad + \left|\kappa^{\ell} \bar{Q}^{(\ell)}(f_{\mathcal{G}}) - \kappa^{\ell} \bar{Q}^{(\ell)}(f)\right| && \text{(H.60)}\\
&\leq 2\, n^2 \left|f - f_{\mathcal{G}}\right| + 5\, 10^{-3} && \text{(H.61)}\\
&\leq 10^{-2}. && \text{(H.62)}
\end{aligned}$$

H.3.1 Proof of Lemma H.11

We first derive a coarse uniform bound on $Q^{(\ell)}$ for $\ell \in \{0, 1, 2, 3\}$. For this we need bounds on the ℓ_2 norm of $\boldsymbol{v_\ell}(f)$ and the magnitude of $R^{(\ell)}(f)$ that hold over the whole unit interval, not only on a discrete grid. By the definitions of K and $\boldsymbol{b}(j)$ in (3.38) and (3.45), for any f

$$\begin{aligned}
||\boldsymbol{v_\ell}(f)||_2 &= \left|\left|\sum_{l \in \Omega^c} (i2\pi\kappa l)^{\ell}\, \boldsymbol{c}_l e^{i2\pi l f} \boldsymbol{b}(l)\right|\right|_2 && \text{(H.63)}\\
&\leq \pi^{\ell} n\, ||c||_{\infty} \sup_{-m \leq l \leq m} ||\boldsymbol{b}(l)||_2 \quad \text{by Lemma 3.3} && \text{(H.64)}\\
&\leq \frac{1.3\, \pi^3 n \sqrt{10k}}{m} \quad \text{by Lemmas 3.4 and 3.5} && \text{(H.65)}\\
&\leq 256\sqrt{k}. && \text{(H.66)}
\end{aligned}$$

Similarly, for any f

$$\begin{aligned}
\left|\kappa^{\ell} R^{(\ell)}(f)\right| &= \left|\lambda\, \kappa^{\ell} \sum_{l \in \Omega} (-i2\pi l)^{\ell}\, \boldsymbol{r}_l\, e^{-i2\pi l f}\right| && \text{(H.67)}\\
&\leq \frac{\kappa^{\ell}\, (2\pi)^{\ell}}{\sqrt{n}} \sum_{l \in \Omega} l^{\ell} && \text{(H.68)}\\
&\leq \frac{\kappa^{\ell}\, (2\pi)^{\ell}\, s\, m^{\ell}}{\sqrt{n}} && \text{(H.69)}\\
&\leq \frac{4\pi^3 s}{\sqrt{n}} \quad \text{by Lemma 3.3.} && \text{(H.70)}
\end{aligned}$$

We also derive a coarse bound on the operator norm B_Ω

$$\begin{aligned}
\|B_\Omega\| &\leq \sqrt{\left\|\bar{H}\right\|} && \text{(H.71)}\\
&\leq \sqrt{260\, \pi^2 n\, \log k} \quad \text{by Lemma E.1} && \text{(H.72)}
\end{aligned}$$

which holds because B_Ω is a submatrix of a matrix $\bar{B}$ such that $\bar{H} = \bar{B}\bar{B}^*$. These bounds together with (H.4), the Cauchy-Schwarz inequality and the triangle inequality imply that in $\mathcal{E}_D^c$

$$\left|\kappa^\ell Q^{(\ell)}(f)\right| \leq ||\boldsymbol{v_\ell}(f)||_2 \left\|D^{-1}\right\| \left(||\boldsymbol{h}||_2 + \frac{1}{\sqrt{n}} \|B_\Omega\| \, ||\boldsymbol{r}||_2\right) + \left|\kappa^\ell R^{(\ell)}(f)\right| \tag{H.73}$$

$$\leq 5\,10^5 \left(k + \sqrt{ks\log k}\right) \tag{H.74}$$

$$\leq \frac{n}{7} \quad \text{by (2.11) and (2.12) if we set } C_k \text{ and } C_s \text{ small enough .} \tag{H.75}$$

Finally, if we interpret $Q^{(\ell)}(z)$ as a function of $z \in \mathbb{C}$, a generalization of the mean-value theorem yields

$$\left|\kappa^\ell Q^{(\ell)}(f_2) - \kappa^\ell Q^{(\ell)}(f_1)\right| \leq \kappa^\ell \left|e^{i2\pi f_2} - e^{i2\pi f_1}\right| \sup_{z'} \left|\frac{\mathrm{d}Q^{(\ell)}(z')}{\mathrm{d}z}\right| \tag{H.76}$$

$$\leq \frac{2\pi\,|f_2 - f_1|}{\kappa} \sup_f \left|\kappa^{\ell+1} Q^{(\ell+1)}(f)\right| \tag{H.77}$$

$$\leq n^2\,|f_2 - f_1| \quad \text{by (H.75) for } \ell \in \{0,1,2\}. \tag{H.78}$$

The bound on the deviation of $\bar{Q}^\ell$ is obtained using exactly the same argument together with the bound (H.21). In the case of $\bar{Q}$ the bound is extremely coarse, but it suffices for our purpose.

I Proof of Proposition 3.11

Let l be an arbitrary element of Ω^c. We express the corresponding coefficient $\boldsymbol{q}_l$ in terms of the sign patterns $\boldsymbol{h}$ and $\boldsymbol{r}$,

$$\boldsymbol{q}_l = \boldsymbol{c}_l \left(\sum_{j=1}^k \boldsymbol{\alpha}_j\, e^{i2\pi l f_j} + i2\pi l\kappa \sum_{j=1}^k \boldsymbol{\beta}_j\, e^{i2\pi l f_j}\right) \tag{I.1}$$

$$= \boldsymbol{c}_l\, \boldsymbol{b}(l)^* \begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix} \tag{I.2}$$

$$= \boldsymbol{c}_l\, \boldsymbol{b}(l)^*\, D^{-1} \left(\begin{bmatrix}\boldsymbol{h}\\ 0\end{bmatrix} - \frac{1}{\sqrt{n}} B_\Omega\, \boldsymbol{r}\right) \tag{I.3}$$

$$= \boldsymbol{c}_l \left(\left\langle PD^{-1}\boldsymbol{b}(l), \boldsymbol{h}\right\rangle + \frac{1}{\sqrt{n}} \left\langle B_\Omega^* D^{-1}\boldsymbol{b}(l), \boldsymbol{r}\right\rangle\right), \tag{I.4}$$

where $P \in \mathbb{R}^{k\times 2k}$ is the projection matrix that selects the first k entries in a vector.

The bounds

$$\left|\left|PD^{-1}\boldsymbol{b}(l)\right|\right|_2^2 \leq \|P\|^2 \left\|D^{-1}\right\|^2 ||\boldsymbol{b}(l)||_2^2 \tag{I.5}$$

$$\leq 640k \quad \text{in } \mathcal{E}_D^c \text{ by Lemmas 3.5 and 3.8} \tag{I.6}$$

$$\leq \frac{0.18^2 n}{\log\frac{40}{\epsilon}} \text{by (2.11) if we set } C_k \text{ small enough,} \tag{I.7}$$

and

$$\left|\left|B_\Omega^* D^{-1} \boldsymbol{b}\left(l\right)\right|\right|_2^2 \leq \left\|B_\Omega\right\|^2 \left\|D^{-1}\right\|^2 \left|\left|\boldsymbol{b}\left(l\right)\right|\right|_2^2 \tag{I.8}$$

$$\leq 640\, C_B^2\, kn \quad \text{in } \mathcal{E}_B^c \cap \mathcal{E}_D^c \text{ by Lemmas 3.6 and 3.8} \tag{I.9}$$

$$\leq \frac{0.18^2 n^2}{\log \frac{40}{\epsilon}} \quad \text{by (2.11) if we set } C_k \text{ small enough,} \tag{I.10}$$

imply by Hoeffding's inequality (Theorem H.6) that the probability of each of the events

$$\left|\left\langle P D^{-1} \boldsymbol{b}\left(l\right), \boldsymbol{h}\right\rangle\right| > 0.18\sqrt{n}, \tag{I.11}$$

$$\left|\left\langle B_\Omega^* D^{-1} \boldsymbol{b}\left(l\right), \boldsymbol{r}\right\rangle\right| > 0.18n \tag{I.12}$$

is bounded by $\epsilon/10$. By Lemma 3.4 and the union bound, this implies

$$\left|\boldsymbol{q}_l\right| \leq \left|\left|\boldsymbol{c}\right|\right|_\infty \left(\left|\left\langle D^{-1} \boldsymbol{b}\left(l\right), \begin{bmatrix} \boldsymbol{h} \\ 0 \end{bmatrix} \right\rangle\right| + \frac{\left|\left\langle B_\Omega^* D^{-1} \boldsymbol{b}\left(l\right), \boldsymbol{r}\right\rangle\right|}{\sqrt{n}}\right) \tag{I.13}$$

$$\leq \frac{2.6}{n}\left(0.18\sqrt{n} + 0.18\sqrt{n}\right) \tag{I.14}$$

$$< \frac{1}{\sqrt{n}} \tag{I.15}$$

with probability at least $1 - \epsilon/5$.

J Algorithms

J.1 Proof of Lemma 4.3

The problem is equivalent to

$$\min_{\tilde{\mu}, \tilde{\boldsymbol{z}}, \boldsymbol{u}} \left|\left|\tilde{\mu}\right|\right|_{\mathrm{TV}} + \lambda \left|\left|\tilde{\boldsymbol{z}}\right|\right|_1 \qquad \text{subject to } \left|\left|\boldsymbol{y} - \boldsymbol{u}\right|\right|_2^2 \leq \sigma^2 \tag{J.1}$$

$$\mathcal{F}_n\, \tilde{\mu} + \tilde{\boldsymbol{z}} = \boldsymbol{u}, \tag{J.2}$$

where we have introduced an auxiliary primal variable $\boldsymbol{u} \in \mathbb{C}^n$. Let us define the dual variables $\boldsymbol{\eta} \in \mathbb{C}^n$ and $\nu \geq 0$. The Lagrangian is equal to

$$\mathcal{L}\left(\tilde{\mu}, \tilde{\boldsymbol{z}}, \boldsymbol{\eta}\right) = \left|\left|\tilde{\mu}\right|\right|_{\mathrm{TV}} + \lambda \left|\left|\tilde{\boldsymbol{z}}\right|\right|_1 + \left\langle \boldsymbol{u} - \mathcal{F}_n\, \tilde{\mu} - \tilde{\boldsymbol{z}}, \boldsymbol{\eta}\right\rangle + \nu\left(\left|\left|\boldsymbol{y} - \boldsymbol{u}\right|\right|_2^2 - \sigma^2\right) \tag{J.3}$$

$$= \left|\left|\tilde{\mu}\right|\right|_{\mathrm{TV}} - \left\langle \tilde{\mu}, \mathcal{F}_n^* \boldsymbol{\eta}\right\rangle + \lambda \left|\left|\tilde{\boldsymbol{z}}\right|\right|_1 - \left\langle \tilde{\boldsymbol{z}}, \boldsymbol{\eta}\right\rangle + \left\langle \boldsymbol{u}, \boldsymbol{\eta}\right\rangle + \nu\left(\left|\left|\boldsymbol{y} - \boldsymbol{u}\right|\right|_2^2 - \sigma^2\right) \tag{J.4}$$

where $\eta \in \mathbb{C}^n$ is the dual variable.

To compute the Lagrange dual function we minimize the value of the Lagrangian over the primal variables [9]. The minimum of

$$\left|\left|\tilde{\mu}\right|\right|_{\mathrm{TV}} - \left\langle \tilde{\mu}, \mathcal{F}_n^* \boldsymbol{\eta}\right\rangle \tag{J.5}$$

over $\tilde{\mu}$ is $-\infty$ unless (4.9) holds. Moreover, if (4.9) holds then the minimum is at $\tilde{\mu} = 0$ by Hölder's inequality. Similarly, minimizing

$$\lambda \left|\left|\tilde{\boldsymbol{z}}\right|\right|_1 - \langle \tilde{\boldsymbol{z}}, \boldsymbol{\eta} \rangle \tag{J.6}$$

over $\boldsymbol{z}$ yields $-\infty$ unless (4.10) holds, whereas if (4.10) holds the minimum is attained at $\tilde{\boldsymbol{z}} = 0$. All that remains is to minimize

$$\langle \boldsymbol{u}, \boldsymbol{\eta} \rangle + \nu \left(\left|\left| \boldsymbol{y} - \boldsymbol{u} \right|\right|_2^2 - \sigma^2 \right) \tag{J.7}$$

with respect to $\boldsymbol{u}$ (note that (4.9) and (4.10) do not involve $\boldsymbol{u}$). The function is convex with respect to $\boldsymbol{u}$ so we set the gradient to zero to deduce that the minimum is at $\boldsymbol{u} = \boldsymbol{y} - \frac{1}{2\nu}\eta$. Plugging in this value yields the Lagrange dual function

$$\langle \boldsymbol{y}, \boldsymbol{\eta} \rangle - \frac{1}{4\nu} \left|\left| \eta \right|\right|_2^2 - \nu\sigma^2. \tag{J.8}$$

The dual problem consists of maximizing the Lagrange dual function subject to $\nu \geq 0$, (4.9) and (4.10). For any fixed value of $\tilde{\eta}$, maximizing over ν is easy, the expression is convex in the half plane $\nu \geq 0$ and the derivative is zero at $\left|\left|\eta\right|\right|_2 / 2\sigma$. Plugging this into (J.8) yields the dual problem (4.8).

The reformulation of (4.8) as a semidefinite program is an immediate consequence of the following proposition.

Proposition J.1 (Semidefinite characterization [32, Theorem 4.24], [38, Proposition 2.4])**.** *Let* $\boldsymbol{\eta} \in \mathbb{C}^n$*,*

$$\left| \left(\mathcal{F}_n^* \, \boldsymbol{\eta} \right) (f) \right| \leq 1 \quad \textit{for all } f \in [0, 1]$$

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

$$\begin{bmatrix} \Lambda & \boldsymbol{\eta} \\ \boldsymbol{\eta}^* & \mathrm{I} \end{bmatrix} \succeq 0, \qquad \mathcal{T}^* \left(\Lambda \right) = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}, \tag{J.9}$$

where $\mathbf{0} \in \mathbb{C}^{n-1}$ *is a vector of zeros.*

J.2 Proof of Lemma 4.4

The interior of the feasible set of Problem (4.8) contains the origin and is therefore non empty, so strong duality holds by a generalized Slater condition [54] and we have

$$\sum_{f_j \in \widehat{T}} |\hat{\boldsymbol{x}}_j| + \lambda \sum_{l \in \widehat{\Omega}} |\hat{\boldsymbol{z}}_l| = \left|\left| \hat{\mu} \right|\right|_{\mathrm{TV}} + \lambda \left|\left| \hat{\boldsymbol{z}} \right|\right|_1 = \langle \hat{\boldsymbol{\eta}}, \boldsymbol{y} \rangle - \sigma \left|\left| \boldsymbol{\eta} \right|\right|_2 \tag{J.10}$$

$$\leq \langle \hat{\boldsymbol{\eta}}, \boldsymbol{y} \rangle - \langle \hat{\boldsymbol{\eta}}, \boldsymbol{y} - \mathcal{F}_n \, \hat{\mu} - \hat{\boldsymbol{z}} \rangle \tag{J.11}$$

$$= \langle \hat{\boldsymbol{\eta}}, \mathcal{F}_n \, \hat{\mu} + \hat{\boldsymbol{z}} \rangle \tag{J.12}$$

$$= \mathrm{Re} \left[\sum_{f_j \in \widehat{T}} |\hat{\boldsymbol{x}}_j| \, \overline{\left(\mathcal{F}_n^* \, \hat{\boldsymbol{\eta}} \right) (f_j)} \frac{\hat{\boldsymbol{x}}_j}{|\hat{\boldsymbol{x}}_j|} + \sum_{l \in \widehat{\Omega}} |\hat{\boldsymbol{z}}_l| \, \overline{\hat{\boldsymbol{\eta}}_l} \frac{\hat{\boldsymbol{z}}_l}{|\hat{\boldsymbol{z}}_l|} \right]. \tag{J.13}$$

The inequality (J.11) follows from the Cauchy-Schwarz inequality because $\{\hat{\mu}, \hat{\boldsymbol{z}}\}$ is primal feasible and hence $||\boldsymbol{y} - \mathcal{F}_n \hat{\mu} - \hat{\boldsymbol{z}}||_2 \leq \sigma$. Due to the constraints (4.9) and (4.10) and Hölder's inequality, the inequality that we have established is only possible if (4.15) and (4.16) hold. The proof is complete.

J.3 Atomic-noise denoising via the alternating direction method of multipliers

We rewrite Problem (4.22) as

$$\min_{\substack{t\in\mathbb{R},\, \boldsymbol{u}\in\mathbb{C}^n,\\ \tilde{\boldsymbol{g}}\in\mathbb{C}^n,\, \tilde{\boldsymbol{z}}\in\mathbb{C}^n\\ \Psi\in\mathbb{C}^{n+1\times n+1}}} \frac{\xi}{2}\left(n\,\boldsymbol{u}_1 + t\right) + \lambda' ||\tilde{\boldsymbol{z}}||_1 + \frac{1}{2}||\boldsymbol{y} - \tilde{\boldsymbol{g}} - \tilde{\boldsymbol{z}}||_2^2 \quad \text{subject to} \quad \Psi = \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \tilde{\boldsymbol{g}} \\ \tilde{\boldsymbol{g}}^* & t \end{bmatrix}, \tag{J.14}$$

$$\Psi \succeq 0, \tag{J.15}$$

where $\xi := \frac{1}{\gamma\sqrt{n}}$ and $\lambda' := \frac{\lambda}{\gamma}$. The augmented Lagrangian for this problem is of the form

$$\mathcal{L}_\rho\left(t, \boldsymbol{u}, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{z}}, \Upsilon, \Psi\right) := \frac{\xi}{2}\left(n\,\boldsymbol{u}_1 + t\right) + \lambda' ||\tilde{\boldsymbol{z}}||_1 + \frac{1}{2}||\boldsymbol{y} - \tilde{\boldsymbol{g}} - \tilde{\boldsymbol{z}}||_2^2 + \left\langle \Upsilon, \Psi - \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \tilde{\boldsymbol{g}} \\ \tilde{\boldsymbol{g}}^* & t \end{bmatrix} \right\rangle \tag{J.16}$$

$$+ \frac{\rho}{2}\left|\left|\Psi - \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \tilde{\boldsymbol{g}} \\ \tilde{\boldsymbol{g}}^* & t \end{bmatrix}\right|\right|_F^2, \tag{J.17}$$

where $\rho > 0$ is a parameter. The alternating direction method of multipliers (ADMM) minimizes the augmented Lagrangian by iteratively applying the updates:

$$t^{(l+1)} := \arg\min_t \mathcal{L}_\rho\left(t, \boldsymbol{u}^{(l)}, \tilde{\boldsymbol{g}}^{(l)}, \tilde{\boldsymbol{z}}^{(l)}, \Upsilon^{(l)}, \Psi^{(l)}\right), \tag{J.18}$$

$$\boldsymbol{u}^{(l+1)} := \arg\min_{\boldsymbol{u}} \mathcal{L}_\rho\left(t^{(l)}, \boldsymbol{u}, \tilde{\boldsymbol{g}}^{(l)}, \tilde{\boldsymbol{z}}^{(l)}, \Upsilon^{(l)}, \Psi^{(l)}\right), \tag{J.19}$$

$$\tilde{\boldsymbol{g}}^{(l+1)} := \arg\min_{\tilde{\boldsymbol{g}}} \mathcal{L}_\rho\left(t^{(l)}, \boldsymbol{u}^{(l)}, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{z}}^{(l)}, \Upsilon^{(l)}, \Psi^{(l)}\right), \tag{J.20}$$

$$\tilde{\boldsymbol{z}}^{(l+1)} := \arg\min_{\tilde{\boldsymbol{z}}} \mathcal{L}_\rho\left(t^{(l)}, \boldsymbol{u}^{(l)}, \tilde{\boldsymbol{g}}^{(l)}, \tilde{\boldsymbol{z}}, \Upsilon^{(l)}, \Psi^{(l)}\right), \tag{J.21}$$

$$\Psi^{(l+1)} := \arg\min_{\Psi} \mathcal{L}_\rho\left(t^{(l)}, \boldsymbol{u}^{(l)}, \tilde{\boldsymbol{g}}^{(l)}, \tilde{\boldsymbol{z}}^{(l)}, \Upsilon^{(l)}, \Psi\right), \tag{J.22}$$

$$\Upsilon^{(l+1)} := \Upsilon^{(l)} + \rho\left(\Psi^{(l+1)} - \begin{bmatrix} \mathcal{T}\left(\boldsymbol{u}^{(l+1)}\right) & \tilde{\boldsymbol{g}}^{(l+1)} \\ \left(\tilde{\boldsymbol{g}}^{(l+1)}\right)^* & t^{(l+1)} \end{bmatrix}\right), \tag{J.23}$$

where l indicates the iteration number. We refer the interested reader to the tutorial [8] and references therein for a justification of these steps and more information on ADMM.

For the method to be practical, we need an efficient implementation of all the updates. The augmented Lagrangian is convex and differentiable with respect to t, $\boldsymbol{u}$ and $\tilde{\boldsymbol{g}}$, so for these variables

we just need to compute their gradient and set it to zero. This yields the closed-form updates:

$$t^{(l+1)} = \Psi_{n+1}^{(l)} + \frac{1}{\rho}\left(\Upsilon_{n+1}^{(l)} - \frac{\xi}{2}\right), \tag{J.24}$$

$$\boldsymbol{u}^{(l+1)} = M\,\mathcal{T}^*\left(\Psi_0^{(l)} + \frac{\Upsilon_0^{(l)}}{\rho}\right) - \frac{\xi}{2\rho}\boldsymbol{e}\,(1)\,, \tag{J.25}$$

$$\tilde{\boldsymbol{g}}^{(l+1)} = \frac{1}{2\rho+1}\left(\boldsymbol{y} - \tilde{\boldsymbol{z}}^{(l)} + 2\rho\,\boldsymbol{\psi}^{(l)} + 2\boldsymbol{\upsilon}^{(l)}\right), \tag{J.26}$$

where $\boldsymbol{e}\,(1) := [1, 0, 0, ..., 0]^T$, $\mathcal{T}^*$ outputs a vector whose j-th element is the trace of the $(j-1)$-th subdiagonal of the input matrix, M is a diagonal matrix such that

$$M_{j,j} = \frac{1}{n-j+1}, \quad j = 1, ...n, \tag{J.27}$$

and

$$\Psi^{(l)} := \begin{bmatrix} \Psi_0^{(l)} & \boldsymbol{\psi}^{(l)} \\ (\boldsymbol{\psi}^{(l)})^* & \Psi_{n+1}^{(l)} \end{bmatrix}, \qquad \Upsilon^{(l)} := \begin{bmatrix} \Upsilon_0^{(l)} & \boldsymbol{\upsilon}^{(l)} \\ (\boldsymbol{\upsilon}^{(l)})^* & \Upsilon_{n+1}^{(l)} \end{bmatrix}. \tag{J.28}$$

$\Psi_0^{(l)}$ and $\Upsilon_0^{(l)}$ are $n \times n$ matrices, $\boldsymbol{\psi}^{(l)}$ and $\boldsymbol{\upsilon}^{(l)}$ are n-dimensional vectors and $\Psi_{n+1}^{(l)}$ and $\Upsilon_{n+1}^{(l)}$ are scalars.

Updating $\tilde{\boldsymbol{z}}$ requires solving the problem

$$\min_{\tilde{\boldsymbol{z}}} \lambda' \|\tilde{\boldsymbol{z}}\|_1 + \frac{1}{2}\|\boldsymbol{y} - \tilde{\boldsymbol{g}}^{(l)} - \tilde{\boldsymbol{z}}\|_2^2, \tag{J.29}$$

which is easily achieved by the applying a proximal operator

$$\tilde{\boldsymbol{z}}^{(l+1)} := \text{prox}_{\lambda'}(\boldsymbol{y} - \tilde{\boldsymbol{g}}^{(l)}), \tag{J.30}$$

where for $1 \leq j \leq n$

$$\text{prox}_{\lambda'}\left(\tilde{\boldsymbol{z}}\right)_j := \begin{cases} \text{sign}\left(\tilde{\boldsymbol{z}}_{\boldsymbol{j}}\right)\left(|\tilde{\boldsymbol{z}}_j| - \lambda'\right) & \text{if } |\tilde{\boldsymbol{z}}_j| > \lambda' \\ 0 & \text{otherwise.} \end{cases} \tag{J.31}$$

Finally, the update of $\Psi^{(l)}$ amounts to a projection onto the positive semidefinite cone

$$\Psi^{(l+1)} = \arg\min_{\Psi \succeq 0} \left\| \Psi - \begin{bmatrix} \mathcal{T}\left(\boldsymbol{u}^{(l)}\right) & \tilde{\boldsymbol{g}}^{(l)} \\ \left(\tilde{\boldsymbol{g}}^{(l)}\right)^* & t^{(l)} \end{bmatrix} + \frac{1}{\rho}\Upsilon^{(l)} \right\|_F^2, \tag{J.32}$$

which can be accomplished by computing the eigenvalue decomposition of the matrix and setting all negative eigenvalues to zero.