Super-resolution of point sources via convex programming

Carlos Fernandez-Granda

Courant Institute of Mathematical Sciences and Center for Data Science, NYU, New York, NY, USA

†Corresponding author. Emails: cfgranda@cims.nyu.edu, cfgranda@gmail.com

[Received on 4 August 2015; revised on 11 February 2016; accepted on 19 February 2016]

We consider the problem of recovering a signal consisting of a superposition of point sources from low-resolution data with a cutoff frequency \( f_c \). If the distance between the sources is under \( 1/f_c \), this problem is not well posed in the sense that the low-pass data corresponding to two different signals may be practically the same. We show that minimizing a continuous version of the \( \ell_1 \)-norm achieves exact recovery as long as the sources are separated by at least \( 1.26/f_c \). The proof is based on the construction of a dual certificate for the optimization problem, which can be used to establish that the procedure is stable to noise. Finally, we illustrate the flexibility of our optimization-based framework by describing extensions to the demixing of sines and spikes and to the estimation of point sources that share a common support.

Keywords: super-resolution; line-spectra estimation; convex optimization; dual certificates; sparse recovery; overcomplete dictionaries; group sparsity; multiple measurements.

1. Introduction

Extracting fine-scale information from low-resolution data is a major challenge in many areas of the applied sciences. In microscopy, astronomy and any other application employing an optical device, spatial resolution is fundamentally limited by diffraction [50]. Figure 1 illustrates a popular model for the data-acquisition process in such cases: the object of interest is convolved with a point-spread function (PSF) that blurs the fine-scale details, acting essentially as a low-pass filter. The problem of super-resolution is that of reconstructing the original image from the blurred measurements. An analogous challenge often arises in signal-processing: estimating the spectrum of a signal from a finite number of samples. Truncating the signal in the time domain limits the spectral resolution, as shown in the lower half of Fig. 1. Spectral super-resolution, or equivalently line-spectra estimation, is the problem of recovering the spectrum of the original signal from the truncated data.

By super-resolution we mean the inverse problem of estimating a signal from low-resolution measurements, but the term may have other meanings in different contexts. In optics, it often refers to the problem of overcoming the diffraction limit by modifying the data-acquisition mechanism [50]. In image-processing, it denotes the problem of upsampling an image onto a finer grid, while preserving its edge structure and hallucinating high-frequency textures in a reasonable way [52]. To be clear, in this work we focus on recovering the lost fine-scale features without altering the low-pass sensing process.

In order to super-resolve a signal, it is necessary to leverage some prior knowledge about its structure. Otherwise the problem is hopelessly ill posed, since the missing spectrum can be filled in arbitrarily to produce estimates that correspond to the data. Here, we consider signals that may be represented as superpositions of point sources, such as celestial bodies in astronomy [41], neuron spikes in neuroscience [56] or line spectra in signal processing and spectroscopy [45, 66]. In addition, locating point-wise fluorescent probes is a crucial step in some optical super-resolution procedures capable of handling...
Fig. 1. Schematic illustration of spatial and spectral super-resolution.
more complicated objects, such as photoactivated localization microscopy \cite{9,44} or stochastic optical reconstruction microscopy \cite{60}.

At an abstract level, the deconvolution of point sources or spikes from bandlimited data are an instance of a central question in modern data processing: how to recover a low-dimensional object embedded in a high-dimensional space from incomplete linear measurements. Non-parametric techniques based on convex optimization have had great success in tackling problems of this flavor. Notable examples include sparse regression in high-dimensional settings \cite{74}, compressed sensing \cite{24,32} and matrix completion \cite{21}. The interest of developing optimization-based methods for super-resolution lies in their robustness to noise and in their flexibility to account for different structural assumptions on the signal, noise and measurement model. In recent work, convex programming has been shown to recover a superposition of point sources exactly from bandlimited data, as long as the sources are separated by a minimum distance of \(2/f_c\), where \(f_c\) is the cutoff frequency of the sensing process \cite{18}. Subsequent publications \cite{3,19,39,54,71} have established that the method is robust to noise in non-asymptotic regimes.

The goal of the present paper is to further develop this line of research through two main contributions:

- We establish that it is possible to super-resolve signals with minimum separations above \(1.26/f_c\) via convex programming. Section 2 provides the context for this result by describing the basic super-resolution problem and our optimization-based approach. The proof, which is based on the construction of a novel dual certificate that also allows us to extend previous stability results, is presented in Section 4.

- In Section 3, we illustrate the flexibility of our approach by adapting it to two related problems: demixing of sines and spikes and super-resolution of multiple signals that share a common support. In each case, we propose a optimization program tailored to the problem, discuss how to solve it, analyze its optimality conditions and provide some numerical simulations.

2. Super-resolution of point sources

2.1 Basic model

We model a superposition of point sources as a sum of weighted Dirac measures supported on a subset \(T\) of the unit interval

\[
x := \sum_{t_j \in T} a_j \delta_{t_j},
\]

where \(\delta_{\tau}\) is a Dirac measure at \(\tau\) and the amplitudes \(a_j\) may be complex valued. We study the problem of estimating such a signal from low-resolution measurements, which correspond to the convolution between the signal and a low-pass PSF \(\phi\),

\[
x_{LR} (t) := \phi \ast x (t) = \sum_{t_j \in T} a_j \phi (t - t_j),
\]

as illustrated at the top of Fig. 1. If the cutoff frequency of the PSF is equal to \(f_c\), in the frequency domain the measurements are of the form

\[
\mathcal{F} x_{LR} = \mathcal{F} \phi \mathcal{F} x = \hat{\phi} \Pi_{[-f_c,f_c]} \mathcal{F} x,
\]
where \( \mathcal{F}f \) denotes the Fourier transform of a function or measure \( f \) and \( \Pi_{[-f_c,f_c]} \) is an indicator function that is zero out of the interval \([−f_c,f_c]\). For ease of exposition, we assume that the Fourier transform of the PSF is constant over \([−f_c,f_c]\), i.e., the PSF is a periodized sinc or Dirichlet kernel, but our results hold for any PSF with a known low-pass spectrum. Since the support of \( x_{LR} \) is restricted to the unit interval, it follows from the sampling theorem that its spectrum is completely determined by the discrete samples

\[
y(k) = \mathcal{F}x_{LR}(k) = \int_0^1 e^{-i2\pi kt} x(dt) = \sum_j a_j e^{-i2\pi kj}, \quad k \in \mathbb{Z}, \ |k| \leq f_c,
\]

where we assume for simplicity that \( f_c \) is an integer. In a more compact form, the sensing process can be represented as

\[
y = \mathcal{F}_n x,
\]

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

If the signal \( x \) is used to model a superposition of line spectra, equation (2.4) has a very natural interpretation: the data \( y \) correspond to a finite number of samples of the signal in the time domain. As sketched in the lower half of Fig. 1, truncating the signal in the time domain is equivalent to convolving its spectrum with a sinc function. Our model can consequently be applied directly to spectral super-resolution, where the aim is to estimate sparse line spectra from time-domain samples.

### 2.2 Minimum separation

In contrast to compressed sensing, where randomized measurements preserve the energy of arbitrary sparse signals with high probability (this is commonly known as the restricted isometry property [23]), sparsity is not a strong enough prior to ensure that the super-resolution problem is well posed. Indeed, low-pass filtering may suppress sparse signals almost entirely if their support is too clustered together. As a result, in order to derive meaningful guarantees for super-resolution, it is necessary to impose conditions on the signal support. To this end, we define the minimum separation of the support of a signal, as introduced in [18].

**Definition 2.1 (Minimum separation)** Let \( \mathbb{T} \) be the circle obtained by identifying the endpoints on \([0, 1]\). For a family of points \( T \subset \mathbb{T} \), the minimum separation (or minimum distance) is defined as the closest distance between any two elements from \( T \),

\[
\Delta(T) = \inf_{(t,t') \in T \times T} |t - t'|,
\]

where \( |t - t'| \) is the \( \ell_{\infty} \) distance (maximum deviation in any coordinate). To be clear, this is the wrap-around distance, so that the distance between \( t = 0 \) and \( t' = \frac{3}{4} \) is equal to \( \frac{1}{4} \).

If the minimum distance is too small with respect to the cutoff frequency of the data, it may become impossible to estimate the signal even under very small perturbations to the data. A fundamental limit in this sense is \( \lambda_c := 1/f_c \), the inverse of the cutoff frequency, which also corresponds to the width of the main lobe of the PSF \( \phi \). The reason is that for minimum separations just below \( \lambda_c/2 \), there exist signals that lie almost in the null space of the low-pass operator defined by (2.4). 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

\[\text{\footnote{\( \lambda_c/2 \) is the notorious Rayleigh resolution limit [31], below which it is challenging to even distinguish two neighboring sources.}}\]
distinguish $s_1$ and $s_2$ from low-resolution data. Figure 2 illustrates this: the measurements corresponding to two signals with disjoint supports and a minimum distance of $0.9\lambda_c$ for $f_c = 10^3$ are indeed almost indistinguishable. The phenomenon can be characterized theoretically in an asymptotic setting using Slepian’s prolate-spheroidal sequences [65] (see also Section 3.2 in [18]). More recently, Theorem 1.3 of [53] provides a non-asymptotic analysis. Finally, other works have obtained lower bounds on the minimum separation necessary for convex-programming approaches to succeed [36,70].

2.3 Optimization-based super-resolution

Our approach to super-resolution is based on optimization: we estimate the signal by minimizing a sparsity-inducing norm. Since we are interested in point sources that may be supported at arbitrary
locations within a continuous interval, we resort to a continuous counterpart of the $\ell_1$-norm known as the total-variation norm.\(^2\) If we consider the space of measures supported on the unit interval, this norm is dual to the infinity-norm (see Section A in the appendix of [18] for a different definition), so that for a measure $x$ we have

$$\|x\|_{TV} = \sup_{\|f\|_\infty \leq 1} \Re \left[ \int_T f(t) x(dt) \right].$$

For a superposition of Dirac deltas $\sum_j a_j \delta_{t_j}$, the total-variation norm is equal to the $\ell_1$-norm of the coefficients, i.e., $\|x\|_{TV} = \sum_j |a_j|$. Our super-resolution method consists in minimizing the total-variation norm of the estimate subject to data constraints, as proposed in [18],

$$\min_{\tilde{x}} \|\tilde{x}\|_{TV} \text{ subject to } \mathcal{F}_n \tilde{x} = y, \quad (2.6)$$

where the minimization is carried out over the set of all finite complex measures $\tilde{x}$ supported on $[0, 1]$. Section 2.7 discusses how to solve this optimization problem.

In [18], it was established that if the minimum separation of the support of a signal $\Delta(T)$ is greater or equal to $2\lambda_c$, TV-norm minimization achieves exact recovery. However, numerical experiments indicate that the actual limit at which super-resolution via TV-norm regularization may fail is $\lambda_c$. This is shown in Fig. 3; see also Section 5 of [18]. As explained in Section 2.2, $\lambda_c$ is a natural limit in the sense that below that minimum separation the problem may become ill posed. Our main result is that the guarantee for exact recovery can be extended to a minimum separation of just $1.26\lambda_c$.

**Theorem 2.2** Let $T = \{t_j\}$ be the support of $x$. If the minimum separation obeys

$$\Delta(T) \geq 1.26/f_c = 1.26\lambda_c, \quad (2.7)$$

then $x$ is the unique solution to (2.6). This holds as long as $f_c \geq 10^3$.

---

\(^2\) The term *total variation* may also refer to the $\ell_1$-norm of the discontinuities of a piecewise-constant function, which is a popular regularizer in image processing and other applications [59].

---

Fig. 3. Graphs showing the fraction of times Problem (2.6) achieves exact recovery over 10 trials with random signs and supports. A phase transition occurs near $\lambda_c$. The simulations are carried out using the high-precision semi-definite-programming solver CVX [42].
Since the signal is assumed to be supported on the unit interval, the result implies that it is possible to recover a number of point sources that is directly proportional to the cutoff frequency.

2.4 Dual certificate

Theorem 2.2 is a direct consequence of Proposition 2.3, which establishes the existence of a certain subgradient of the TV-norm that is orthogonal to the null space of the measurement operator. Such an object is often referred to as a dual certificate in the compressed-sensing literature [22] because it certifies that exact recovery occurs, and its coefficients are a solution to the dual of Problem (2.6).

**Proposition 2.3** Under the conditions of Theorem 2.2 for any sign pattern \( v \in \mathbb{C}^{|T|} \), such that \( |v_j| = 1 \) for all \( j \), there exists a low-pass trigonometric polynomial

\[
q(t) = \sum_{k=-f_c}^{f_c} c_k e^{i2\pi kt} \tag{2.8}
\]

obeying

\[
q(t_j) = v_j, \quad t_j \in T, \tag{2.9}
\]

\[
|q(t)| < 1, \quad t_j \notin T. \tag{2.10}
\]

In addition, there exist numerical constants \( C_0 \in (0, 1) \) and \( C_1 \) and \( C_2 \) such that

\[
1 - C_1 f_c^2 (t - t_j)^2 \leq |q(t)| \leq 1 - C_2 f_c^2 (t - t_j)^2, \quad |t - t_j| \leq C_0 \lambda_c, \quad t_j \in T. \tag{2.11}
\]

By (2.9) and (2.10), the polynomial is a subgradient of the total-variation norm at the original signal \( x \). By (2.8) it is also low pass, which means that it is orthogonal to the null space of the measurement operator, as any signal in the null space is high pass. This immediately implies that for any signal \( h \) in the null space

\[
\|x + h\|_{TV} \geq \|x\|_{TV} + \langle q, h \rangle = \|x\|_{TV}.
\]

The bound on the off-support (2.10) actually implies that this inequality is strict, so that \( x \) is the unique solution to Problem 2.6. A complete proof of the fact that Theorem 2.2 follows from Proposition 2.3 is provided in Section A of the appendix of [18]. The quadratic bound (2.11) is key in establishing robustness guarantees, see [19, 39]. Section 4 is devoted to proving Proposition 2.3.

In the case of super-resolution, the dual certificate is a trigonometric polynomial with cutoff frequency \( f_c \) that interpolates the sign of the original signal on its support, and has magnitude strictly bounded by one on the off-support (see Proposition 2.3). In [18], such a dual certificate is constructed using interpolation with a low-pass kernel and its derivative. The proof of Proposition 2.3 generalizes this approach, allowing us to optimize the choice of the interpolation kernel by using sharp non-asymptotic bounds on the Dirichlet kernel; we defer the details to Section 4.

Finally, we would like to emphasize that the structure of the dual polynomial reveals what signals will be more challenging for the optimization-based procedure. If the support of the signal is cluttered together and the sign of its coefficients varies rapidly, it may not be possible to achieve the interpolation with a bounded low-pass polynomial. In Sections 3.1.3 and 3.2.3, we show that similar insights arise...
when we derive dual certificates for other optimization programs designed to tackle extensions of the basic super-resolution problem.

2.5 Robustness to noise

In any problem involving real data, it is necessary to account for perturbations and model imperfections. In the case of super-resolution, we can adapt Problem (2.6) by using an inequality constraint to quantify the uncertainty,

$$\min_{\bar{x}} \|\bar{x}\|_{TV} \quad \text{subject to} \quad \|F_n\bar{x} - y\|_2^2 \leq \delta,$$  \hspace{1cm} (2.12)

where $\delta$ is an estimate of the noise level. Alternatively, we could also consider a Lagrangian formulation of the form

$$\min_{\bar{x}} \|\bar{x}\|_{TV} + \gamma \|F_n\bar{x} - y\|_2^2,$$  \hspace{1cm} (2.13)

where the regularization parameter $\gamma > 0$ governs the tradeoff between data fidelity and the sparsity of the estimate.

Recent works [3,19,39,71] derive non-asymptotic guarantees on the estimation error achieved when solving these problems to perform super-resolution from noisy data. The proofs of these stability guarantees rely in part on the dual certificate constructed in [18] and on generalizations of this construction. As a result, the guarantees only hold under the proviso that the minimum separation is greater or equal to $2\lambda_c$.

The techniques developed to prove Theorem 2.2, which are presented in Section 4, allow us to extend these results to minimum separations of just $1.26\lambda_c$. In more detail, Lemma 2.7 in [19] constructs a low-pass polynomial $q$ such that $q(t_j) = 0$ and $|q'(t_j)| = 1$ for all $t_j$ belonging to the support of the original signal (i.e., the polynomial is locally linear). The polynomial is built through interpolation, using the same low-pass kernel as in [18]. The construction can consequently be adapted by using the interpolation kernel described in Section 4, along with the bounds provided in Section 4.1. The same holds for the results in [39]. In this case, a low-pass polynomial that is equal to one on a certain element of the support and to zero on the rest is used to obtain support-detection guarantees. The polynomial is constructed in Lemma 2.2 of [39], again using the same interpolation kernel as in [18]. The kernel and bounds described in the present work can be leveraged to build such a polynomial for supports with minimum separations above $1.26\lambda_c$. Finally, our techniques can also be used to sharpen the analysis in [54], which studies the super-resolution of point sources with positive amplitudes.

2.6 Extensions

Apart from the extensions discussed in Section 3, there are several interesting extensions to the basic super-resolution problem. In compressed sensing off the grid the aim is to super-resolve a sparse signal from a random subset of its low-pass Fourier coefficients. Exact recovery via convex programming has been shown to occur with high probability for a number of measurements that is proportional to the sparsity level (up to logarithmic factors) as long as the support of the signal has a minimum separation of $2\lambda_c$ [72]. This bound can be lowered to $1.26\lambda_c$ using the dual-certificate construction in Section 4. Our construction also allows us to sharpen results pertaining to the recovery of non-uniform splines from their projection onto spaces of algebraic polynomials [8,29]. Finally, it is straightforward to apply our results to the recovery of piecewise-constant or piecewise-smooth functions from low-pass data. The details can be found in Section 1.10 of [18].
2.7 Implementation

Solving Problem (2.6) may seem challenging at first because its primal variable is infinite-dimensional. This may be overcome by discretizing the unit interval into a grid and solving an $\ell_1$-norm minimization problem. However, it is also possible to solve the problem exactly without resorting to discretization. A strategy to achieve this is to recast the dual problem as a tractable semi-definite program, and then decode the support of the primal variable from the dual solution. The dual problem is

$$\max_{c \in \mathbb{C}^n} \langle y, c \rangle \quad \text{subject to} \quad \|F_n^* c\|_\infty \leq 1, \quad (2.14)$$

where the inner product is defined as $\langle y, c \rangle := \text{Re} \, (y^* c)$. The dual variable $c$ is finite-dimensional, but the constraint is infinite-dimensional; the magnitude of the trigonometric polynomial $(F_n^* c)(t) = \sum_{|k| \leq f} c_k e^{2\pi i k t}$ must be bounded by 1. Fortunately, this condition has a semi-definite representation provided by the following proposition, which is a consequence of the Fejér–Riesz Theorem (see Section A.1 of the Appendix for the proof and Theorem 4.24 in [35] for a more general result).

**Proposition 2.4** Let $C \in \mathbb{C}^{n \times m}$ and let $C_k$ denote the $k$th column of $C$,

$$\sum_{k=1}^{m} \left| (F_n^* C_k)(t) \right|^2 \leq 1 \quad \text{for all } t \in T$$

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

$$\begin{bmatrix} \Lambda & C \\ C^* & I \end{bmatrix} \succeq 0, \quad T^* (\Lambda) = e_1, \quad (2.15)$$

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

For any vector $u$ such that $u_1$ is positive and real, $T^*(u)$ is a Hermitian Toeplitz matrix whose first row is equal to $u$. The adjoint of $T$ with respect to the usual matrix inner product $\langle M_1, M_2 \rangle = \text{Tr} \, (M_1^* M_2)$, extracts the sums of the diagonal and off-diagonal elements of a matrix

$$T^* (M) = \sum_{i=1}^{n-1} M_{i,j} \sum_{j=1}^{n-i} 1, \quad (2.16)$$

Setting $m = 1$ in Proposition 2.4, Problem (2.14) is equivalent to

$$\max_{c \in \mathbb{C}^n, \Lambda \in \mathbb{C}^{n \times n}} \langle y, c \rangle \quad \text{subject to} \quad \begin{bmatrix} \Lambda & c \\ c^* & 1 \end{bmatrix} \succeq 0, \quad T^* (\Lambda) = e_1. \quad (2.16)$$

In order to extract an estimate of the support of the primal variable, we construct a support-locating polynomial from the solution to (2.16), $\hat{c}$,

$$P_{\hat{c}} (t) = (F_n^* \hat{c})(t). \quad (2.17)$$
By strong duality, which holds because the interior of the feasible set of Problem (2.14) contains the origin and is consequently non-empty \([57]\), any solution \(\hat{x}\) to (2.6) obeys
\[
\text{Re} \left[ \int_0^1 \hat{P}_c(t)\hat{\chi}(dt) \right] = \langle \mathcal{F}_n^\ast \hat{c}, x \rangle = \langle \hat{c}, \mathcal{F}_n x \rangle = \langle y, \hat{c} \rangle = \|\hat{\chi}\|_{TV},
\]
which implies that \(P_c\) is equal to the sign of the primal solution \(\hat{x}\) at any point where the latter is non-zero. This suggests super-resolving the signal by applying the following scheme:

1. Solve the finite-dimensional semi-definite program (2.16).
2. Construct the support-locating polynomial and determine a set of points where its magnitude is equal to one, to produce an estimate of the signal support.
3. Estimate the amplitude of the signal by solving the corresponding system of equations.

Figure 4 shows an example. This approach was proposed in [18] and extended to a noisy setting in [19]. We refer the reader to Section 4 in [18] and Section 3 in [19] for numerical simulations and a more detailed discussion (see also [12], where a related semi-definite program is applied to the denoising of line spectra). Sections 3.1.2 and 3.2.2 describe similar algorithms for demixing sines and spikes, and super-resolving point sources with a common support. Finally, we would like to mention that other strategies to overcome discretization have been proposed [15,16].

2.8 Related work

A popular approach to the estimation of point sources from bandlimited data is to fit an estimate of the PSF of the sensing device to each source individually; see [44] for an example in fluorescence microscopy. This is essentially equivalent to matched-filtering techniques used in communications and radar, as well as linear non-parametric spectral-analysis algorithms such as the periodogram [68]. These simple methods perform well, as long as the dynamic range of the signal is not very high and the sources are well separated. However, if this is not the case, their performance can be degraded due to aliasing, even in the absence of noise.

Prony’s method is a classical technique that exploits the algebraic structure of the data to tackle the problem of spectral super-resolution [55]. The method is able to super-resolve the support of the sparse
spectrum of a signal from measurements following (2.3), by building a polynomial whose zeros are located exactly on the support. This motivates the use of the term annihilating filter in the finite rate of innovation (FRI) framework [14,34,77]. The FRI framework allows us to super-resolve signals with a parametric representation that has a finite number of degrees of freedom. The superpositions of Dirac measures that we consider in spectral super-resolution belong to this class. For this type of signal, FRI reduces exactly to Prony’s method (see Section 3 in [77] or Section 1 in [14]). We would like to mention that algebraic techniques have also been applied to the location of singularities in the reconstruction of piecewise polynomial functions from a finite number of Fourier coefficients (see [4,6,37] and references therein). The theoretical analysis of these methods proves their accuracy up to a certain limit related to the number of measurements.

In the absence of noise, Prony’s method is capable of super-resolving $s$ spikes from $2s$ measurements without any minimum-separation condition. However, this is no longer the case when the data are noisy, as small perturbations may produce substantial changes in the position of the zeros of the annihilating polynomial. A popular approach to overcome this issue is to estimate the location of the sources from the eigendecomposition of the empirical covariance matrix of the data. Such techniques, which include MUSIC [13,64] and ESPRIT [58], have widespread popularity in the signal-processing community. We refer the reader to [68] for a detailed description. A variant that is often used by FRI methods fits a low-rank Toeplitz model to a matrix constructed from the data, see Section 2 of [14]. This technique is due to Tufts and Kumaresan [76], and is also known as Cadzow denoising [17].

Most theoretical work analyzing the robustness of covariance-based methods for spectral super-resolution is based on the asymptotic characterization of the sample covariance matrix under Gaussian noise [27,69]. However, some recent stability analysis has shown that these methods are robust for signals with a minimum separation above $\lambda_c$ [53], see also [30,49]. In addition, [49] shows that numerically the noise level tolerated by MUSIC for small signal separations obeys a power law. In practice, the performance of classical spectral methods, such as MUSIC, Cadzow’s method or ESPRIT seems to be comparable to the optimization-based approach which is the focus of the present paper, see [12] for some numerical comparisons.

The idea of applying $\ell_1$-norm minimization and related optimization programs to the super-resolution problem dates back to Beurling [10,11] and researchers in geophysics from the 1970s [26,46,61]. Theoretical guarantees for positive signals in the absence of noise were obtained in [28,33,40] (see also [63] for an interesting recent result). As explained previously in this section, [18] proved that exact recovery via convex optimization for measures with arbitrary complex amplitudes satisfying a minimum separation of $2\lambda_c$ and provided stability guarantees for a discretized version of the problem. Subsequent works then established stability guarantees in a continuous setting [3,19,36,39,71] and analyzed related problems using similar techniques [2,7,43]. Finally, we would like to mention an application of super-resolution via convex programming to fluorescence microscopy [80], another optimization-based approach to super-resolve point sources on a continuous domain [38], and an alternative proof of exact recovery via TV-norm minimization [67] that allows us to sharpen the result in [18] to $1.56\lambda_c$.

3. A general framework for super-resolution

In practical applications, it is often desirable to adapt the signal, noise or measurement model to incorporate prior information. Optimization-based methods are well suited for this purpose: tailoring the cost function and the constraints of the optimization program allows us to account for application-specific structure. We illustrate this by deriving optimization-based algorithms to demix sines and spikes, and to super-resolve multiple signals with a common support. In both cases, we provide an implementation
based on semi-definite programming, derive the corresponding dual certificate and report some experimental results.

3.1 Super-resolution from corrupted data: demixing of sines and spikes

As mentioned previously, the super-resolution problem discussed in Section 2.1 can be used to model line-spectra estimation. In that case, the measurements correspond to equispaced samples from a multi-sinusoidal signal measured at the Nyquist rate over a limited period of time. We now consider line-spectra estimation, when some of the data may be completely corrupted. This could be due to the presence of sparse noise, to the intermittent failure of a sensor or to some other impulsive phenomenon. We model the corruptions as a vector \( s \in \mathbb{C}^n \) with support \( S \). In the resulting measurement model, the data \( y \) are of the form

\[
y = \mathcal{F}_n \tilde{x} + s.
\]

(3.1)

As shown in Fig. 5, impulsive corruptions produce additional aliasing that further masks the line-spectra locations when we visualize the data in the frequency domain. We aim to demix the two components: the sines, whose spectrum is denoted by \( x \), and the spikes \( s \).

3.1.1 Estimation via convex programming

Convex programming has proved very effective in separating incoherent low-dimensional components \([20, 25, 47, 51]\). Inspired by these works, we design a simple convex program that allows us to achieve effective demixing without any previous knowledge of the number of spikes or sines. To separate the two components, we penalize sparsity-inducing norms in both domains: the \( \ell_1 \)-norm of the spikes and the TV-norm the spectrum of the sines, which is modeled as an atomic measure. These two norms are combined additively. More explicitly, the estimate is computed by solving

\[
\min_{\tilde{x}, \tilde{s} \in \mathbb{C}^n} \|\tilde{x}\|_{\text{TV}} + \eta \|\tilde{s}\|_1 \quad \text{subject to } \mathcal{F}_n \tilde{x} + \tilde{s} = y,
\]

(3.2)

where \( \eta > 0 \) is a real-valued parameter that determines the tradeoff between the two terms in the cost function. This formulation was previously proposed in \([73]\).

3.1.2 Implementation

As in the case of our basic model (see Section 2.7), we have two options to solve Problem (3.2). The first is to discretize the support of the primal variable that represents the line-spectra estimate and then solve an \( \ell_1 \)-norm minimization problem. The second is to consider the dual of Problem (3.2), provided by the following lemma (see Section A.2 of the Appendix for the proof).

**Lemma 3.1** The dual of Problem (3.2) is

\[
\max_{c \in \mathbb{C}^n} \langle y, c \rangle \quad \text{subject to } \|\mathcal{F}_n^* c\|_\infty \leq 1 \quad \text{and} \quad \|c\|_\infty \leq \eta.
\]

(3.3)

Problem 3.3 can be recast as the semi-definite program

\[
\max_{c \in \mathbb{C}^n, A \in \mathbb{C}^{nxn}} \langle A, c \rangle \quad \text{subject to } \begin{bmatrix} A & c^* \\ c & 1 \end{bmatrix} \succeq 0, \quad T^* (A) = e_1 \quad \text{and} \quad \|c\|_\infty \leq \eta,
\]

(3.4)
Fig. 5. Mixture of sines and spikes: samples from a multi-sinusoidal signal (top) are combined with impulsive events (middle) to yield the data (bottom).

by invoking Proposition 2.4. After solving (3.4), we still need to decode the support of the primal solution from the dual solution. This can be done by leveraging the following lemma, proved in Section A.3 of Appendix.

**Lemma 3.2** Let

\[ \hat{x} = \sum_{t_j \in \hat{T}} \hat{a}_j \delta_{t_j} = \sum_{t_j \in \hat{T}} |\hat{a}_j| e^{i\phi_j} \delta_{t_j} \quad \text{and} \quad \hat{s}_l = |\hat{s}_l| e^{i\psi_l}, \ l \in \hat{S} \subseteq \{1, \ldots, n\}, \]

be a solution to (3.2), such that \( \hat{T} \) and \( \hat{S} \) are the non-zero supports of the line spectra \( \hat{x} \) and the spikes \( \hat{s} \), respectively. If \( \hat{c} \in \mathbb{C}^n \) is a corresponding dual solution, then for any \( t_j \in \hat{T} \)

\[ (\mathcal{F}_n^* \hat{c}) (t_j) = e^{i\phi_j} \]
Fig. 6. Demixing of the signal in Fig. 5 by semi-definite programming. Top left: the polynomial $\mathcal{F}_c^* \hat{c}$ (green) corresponding to the solution of (3.4) $\hat{c}$ interpolates the sign of the line spectra of the sines (orange) on their support. Top right: $\eta^{-1} \hat{c}$ interpolates the sign pattern of the spikes on their support. Bottom: locating the support of $x$ and $s$ allows us to demix very accurately.

and for any $l$, $1 \leq l \leq n$ in $\hat{S}$

$$\hat{c}_l = \eta e^{i \psi_l}.$$ 

In words, the weighted dual solution $\eta^{-1} \hat{c}$ and the corresponding polynomial $\mathcal{F}_n^* \hat{c}$ interpolate the sign patterns of the primal-solution components $\hat{x}$ and $\hat{s}$ on their respective supports. This suggests the following demixing scheme:

1. Solve the finite-dimensional semi-definite program (3.4) to obtain a dual solution $\hat{c}$.
2. The support of $\hat{s}$ is given by the points at which $\hat{c}$ equals $\eta$.
3. Construct the support-locating polynomial $\mathcal{F}_c^* \hat{c}$. The support of $\hat{x}$ is given by the points at which its magnitude equals one.
4. Estimate the amplitudes of $\hat{x}$ and $\hat{s}$ by solving the corresponding system of equations.

The resulting $\hat{x}$ and $\hat{s}$ are our estimates of the line spectra and of the spikes, respectively. Figure 6 shows the results obtained by this method on the data described in Fig. 5: both components are recovered very accurately.³

³ A Matlab script reproducing this example is available at http://www.cims.nyu.edu/~cfgranda/scripts/sines_and_spikes.m.
3.1.3 Theoretical analysis: dual certificate

The success or failure of the demixing procedure is determined by the structure of the signal that we aim to recover. As in the case of our basic model, this is reflected by the dual certificate corresponding to Problem (3.2).

Lemma 3.3 Let $y = \mathcal{F}^* x + s$, where

$$x = \sum_{t_j \in T} a_j \delta_{t_j} = \sum_{t_j \in T} |a_j| e^{i\phi_j} \delta_{t_j} \quad \text{and} \quad s_l = |w_l| e^{i\psi_l}$$

for $s \in \mathbb{C}^n$ with support $S \subseteq \{1, \ldots, n\}$. If $|T| + |S| \leq n$ and there exists a low-pass polynomial $q : T \to \mathbb{C}$

$$q(t) = \mathcal{F}^* c = \sum_{l=-f_c}^{f_c} c_l e^{i2\pi l t}, \quad (3.5)$$

where $c \in \mathbb{C}^n$, such that

$$q(t_j) = e^{i\phi_j} \quad t_j \in T, \quad |q(t)| < 1 \quad t \notin T, \quad (3.6)$$

$$c_l = \eta e^{i\psi_l} \quad l \in S, \quad |c_j| < \eta \quad j \notin S, \quad (3.7)$$

then $x$ and $s$ are the unique solutions to Problem (3.2).

The lemma, proved in Section A.4 of Appendix, shows that the dual certificate is a low-pass polynomial which interpolates the sign pattern of the line spectra, just like the dual certificate corresponding to Problem (2.6). As a result, the approach might fail for line spectra with very clustered supports, which makes sense as such problems are challenging even in the absence of corruptions. The dual certificate satisfies two additional conditions that depend on the spikes/corruptions: its coefficients interpolate the sign pattern of the spikes on their support and their magnitude is bounded by $\eta$. If there are too many spikes, it may become impossible to build a bounded low-pass polynomial under these constraints that also interpolates the sign of the line spectra. There is consequently a tradeoff between the number of sines and spikes that may be recovered simultaneously. In the following section, we provide some experimental results that illustrate this tradeoff. Characterizing it theoretically is an interesting research direction.

3.1.4 Numerical results

Figure 7 shows the result of applying our demixing procedure to signals with different minimum separations of the line spectra $x$, as well as different cardinalities of the supports of $x$ and $s$. In each case, we observe a rapid transition between a regime in which exact recovery almost always occurs and a regime in which the method fails. For a fixed minimum separation, the procedure succeeds for a number of spikes which decreases as we increase the number of sine components. In contrast, if we fix the cardinality of $x$, the phase transition occurs at a value of the minimum separation that does not seem to depend on the number of spikes, as long as there are not too many. Similarly, for a fixed number of spikes the minimum separation necessary to achieve exact recovery seems to be independent of the cardinality of $x$ below a certain limit.

3.2 Point sources with a common support

In this section, we consider the problem of super-resolving $m$ signals consisting of superpositions of point sources, which share a common support. This is a model of interest in fluorescence microscopy,
Fig. 7. Graphs showing the fraction of times Problem (3.2) achieves exact recovery over 10 trials with random signs and supports for different values of the minimum separation of the line spectra $x$ (left), as well as different cardinalities of the support of the line spectra $x$ denoted by $|T|$ (center) and of the spikes $s$ (right). The simulations are carried out using CVX [42].

where the data are often well modeled as low-resolution frames, where fluorescence probes flicker stochastically [9,44,60], and in sparse-channel estimation for wireless communications [5].

The signals are modeled as superpositions of Dirac measures with support $T$:

$$X_k(t) = \sum_{j \in T} A_{jk} \delta_{t_j}, \quad 1 \leq k \leq m,$$

where $A \in \mathbb{C}^{|T| \times m}$. The low-resolution measurements consist of the $n = 2f_c + 1$ first Fourier coefficients of each signal. We represent these low-pass data with cutoff frequency $f_c$ as a matrix $Y \in \mathbb{C}^{n \times m}$,

$$Y_{lk} = \int_0^1 e^{-i2\pi lt} X_k(\mathrm{d}t) = \sum_{j=1}^s A_{jk} e^{-i2\pi t_j} = (\mathcal{F}_{f_c} X_k)_l, \quad |l| \leq f_c, \quad 1 \leq k \leq m.$$  

(3.9)

To be clear, the $k$th column of $Y$ contains the low-pass Fourier coefficients of the signal $X_k$.

3.2.1 Estimation via convex programming  

Note that we could just apply the approach based on TV-norm minimization described in Section 2 to super-resolve each signal separately. However, our goal is to exploit the diversity of the amplitudes $A_{jk}$ to obtain a better estimate of the common support $T$ by processing all the data simultaneously. This can be achieved by leveraging a generalization of the TV-norm, which we call group total-variation norm (gTV). Formally, we define the gTV-norm as

$$\|X\|_{gTV} := \sup_{F: T \to \mathbb{C}^n, \|F(t)\|_2 \leq 1} \sum_{k=1}^m \text{Re} \left\{ \int_T \overline{F_k(t)} X_k(\mathrm{d}t) \right\}.$$  

We would like to stress that the gTV-norm is a well-known quantity. In fact, it is commonly used as a definition for the total variation of vector-valued signals [1]. Our intent is not to propose a new name for it, but rather to use an abbreviation that distinguishes it from the TV-norm while highlighting the connection to group sparsity. Intuitively, the gTV-norm is the continuous counterpart of the mixed...
The $\ell_1/\ell_2$-norm, which is used to promote group sparsity [75,79]. For an arbitrary matrix $M$

$$\|M\|_{\ell_1/\ell_2} := \sum_j \|M_j\|_2,$$

where $M_j$ denotes its $j$th row. If the measures $X_k$ are superpositions of point sources then

$$\|X\|_{gTV} = \|A\|_{\ell_1/\ell_2}.$$

Let us provide some intuition as to why the $\ell_1/\ell_2$-norm promotes joint sparsity. Consider a matrix $M$ which has both zero and non-zero elements. If we choose an element $M_{jk}$ from an empty row $M_j$ and set its value to $\epsilon > 0$, then $\|M\|_{\ell_1/\ell_2}$ increases by $\epsilon$. However, if the row is not empty and $\|M_j\|_2 > 0$, then $\|M\|_{\ell_1/\ell_2}$ increases by just $\epsilon^2/2\|M_j\|_2$ as long as $\epsilon \ll \|M_j\|_2$ (this follows from a Taylor expansion around $\epsilon = 0$ of $(\|M_j\|_2^2 + \epsilon^2)^{1/2}$). It is consequently much cheaper to introduce new elements in non-empty rows than in empty ones, which promotes a common sparsity pattern in the columns.

To super-resolve the signals jointly, we propose minimizing the continuous version of the $\ell_1/\ell_2$-norm subject to data constraints

$$\min_{\tilde{X}} \|\tilde{X}\|_{gTV} \text{ subject to } [\mathcal{F}_n\tilde{X}_1 \mathcal{F}_n\tilde{X}_2 \cdots \mathcal{F}_n\tilde{X}_m] = Y. \quad (3.10)$$

This cost function favors solutions that have a sparse common support, allowing us to jointly super-resolve signals that are very challenging to super-resolve individually, as illustrated in Fig. 8. We would like to note that this approach is equivalent to the one proposed by two recent preprints [48,78] from the perspective of atomic-norm minimization.

### 3.2.2 Implementation

If we discretize the domain, Problem (3.10) reduces to an $\ell_1/\ell_2$-norm minimization problem. However, as in the case of Problems (2.6) and (3.2), it is possible to obtain a solution over the continuous domain directly.

**Lemma 3.4** The dual of Problem (3.10) is

$$\max_{C \in \mathbb{C}^{n \times m}} \langle Y, C \rangle \text{ subject to } \sup_t \sum_{k=1}^m \left| \left( \mathcal{F}_n^* C_k \right)(t) \right|^2 \leq 1, \quad (3.11)$$

where $C_k$ denotes column $k$ of $C$.

The proof of the lemma can be found in Section A.5 of Appendix. By Proposition 2.4, we can recast (3.11) as the semi-definite program

$$\max_{C \in \mathbb{C}^{n \times m}, A \in \mathbb{C}^{n \times n}} \langle Y, C \rangle \text{ subject to } \begin{bmatrix} A & C \\ C^* & I_n \end{bmatrix} \succeq 0, \quad T^*(A) = e_1, \quad (3.12)$$

where $I_n$ is the identity matrix of dimensions $n \times n$.

To recover the support of the primal solution from a dual solution $\hat{C}$, we compute a generalization of the polynomial (2.17) defined in Section 2.7

$$P_{\hat{C}}(t) = \sum_{k=1}^m \left| \left( \mathcal{F}_n^* \hat{C}_k \right)(t) \right|^2. \quad (3.13)$$
The following lemma establishes that we can use this polynomial to locate the support of the primal solution. We defer the proof to Section A.6 in Appendix.

**Lemma 3.5** Let $\hat{X}$ be a solution to Problem (3.10) such that $\hat{X}_k(t) = \sum_{t_j \in \hat{T}} \hat{A}_{jk} \delta_{t_j}$ for $1 \leq k \leq m$, where $\hat{A} \in \mathbb{C}^{|\hat{T}| \times m}$ and $\hat{T}$ denotes the support of the solution. Then, for any corresponding dual solution $|C|$, the support-locating polynomial satisfies

$$P_{\hat{C}}(t_j) = 1 \quad \text{for all } t_j \in \hat{T}.$$ 

The careful reader will observe that the lemma does not imply that the support-locating polynomial cannot be equal to one all over the support, and hence completely uninformative. However, in the proof of the lemma we show that the polynomial $F_n^* \hat{C}_k$, which is restricted to the hypercylinder of radius one by constraint (3.11), is equal to $A_{jk} / \|A_j\|_2$ on $\hat{T}$ (recall that $A_j$ is the $j$th row of $A$). This means that,
except in pathological cases where the polynomial stays on the surface of the hypercylinder over all of its domain, the support-locating polynomial is informative. Indeed, this is what we observe in numerical simulations. See Section 4 of [18] for a related discussion.

Finally, the procedure to estimate the signals is as follows:

1. Solve the finite-dimensional semi-definite program (3.12).
2. Construct the support-locating polynomial (3.13). Set \( \hat{T} \) to the set of points where its magnitude is equal to one.
3. Estimate the amplitudes of the individual signals by solving the corresponding systems of equations.

Figure 8 shows the polynomial \( P_\hat{C} \) for an example, where \( m = 3, s = 15, f_c = 40 \) and the signal amplitudes are real valued and generated at random. The minimum separation of the common support is \( 0.7\lambda_c \). The polynomial \( P_\hat{C} \) obtained from (3.12) allows us to locate the support of the signal with great precision. In contrast, the individual support-locating polynomials obtained from solving the individual TV-norm minimization problems do not allow us to locate the support: they miss some spikes and also produce false positives.\(^4\)

3.2.3 Theoretical analysis: dual certificate

As in the case of Problems (2.6) and (3.2), the dual certificate corresponding to Problem (3.10) provides some useful intuition as to the performance we can expect from the optimization-based method depending on the structure of the signals that generate the data.

**Lemma 3.6** For \( X \) defined by (3.8) and \( Y \) defined by (3.9), if there exists an \( m \)-dimensional low-pass polynomial \( Q : \mathbb{T} \to \mathbb{C}^m \),

\[
Q_k(t) = (\mathcal{F}_n^* C_k)(t) = \sum_{l=-f_c}^{f_c} C_{lk} e^{i2\pi lt}, \quad 1 \leq k \leq m, \quad C = [C_1 \cdots C_m] \in \mathbb{C}^{n \times m},
\]

such that

\[
Q_k(t_j) = \frac{A_{jk}}{\sqrt{\sum_{k'=1}^{m} |A_{jk'}|^2}}, \quad t_j \in T, \quad \max_{t \in \hat{T}} \sum_{k=1}^{m} |Q_k(t)|^2 < 1,
\]

then \( X \) is the unique solution to Problem (3.10).

The dual certificate provided by this lemma (see Section A.7 in Appendix for the proof) is an \( m \)-dimensional low-pass polynomial \( Q \) that interpolates points from the \( m \)-dimensional unit hypersphere \( S^m \) on the support of the signal, while remaining strictly inside the hypercylinder \( \mathbb{T} \times S^m \) elsewhere. If the amplitudes of the \( m \) signals are in a general position, then it is plausible that interpolating these

\(^4\) A Matlab script reproducing this example is available at [http://www.cims.nyu.edu/~cfgranda/scripts/superres_common_support.m](http://www.cims.nyu.edu/~cfgranda/scripts/superres_common_support.m).
Fig. 9. Graphs showing the fraction of times Problem (3.10) achieves exact recovery over 10 trials with random signs and supports for different values of \( m \). As \( m \) grows, the minimum separation at which exact recovery occurs tends to \( \lambda_c/2 \). The simulations are carried out using CVX [42].

points may become easier as the dimension \( m \) of the hypercylinder increases. This would imply that using a larger number of signals may allow us to recover common supports with smaller minimum separations. This is confirmed by the numerical simulations in Fig. 9, which show that exact recovery occurs for decreasing values of the minimum distance as we increase \( m \).

3.2.4 Numerical results  Figure 9 shows the fraction of successes achieved by gTV minimization for different values of the minimum distance, of the number of signals \( m \) and of the number of point sources in each signal [7]. As in the case of TV minimization for a single signal (see Fig. 3), a phase transition occurs at a certain value of the minimum distance. This critical value of the minimum distance decreases as \( m \) increases, saturating at \( \lambda_c/2 \). Below this value, there exist signals that may lie almost in the null space of the measurement operator (note that this is a stronger statement than the difference of two signals being almost in the null space, see Section 2.2). Characterizing the minimum separation at which gTV-norm minimization succeeds for amplitudes in a general position is an intriguing research problem.
4. Proof of Proposition 2.3

Our aim is to construct a low pass polynomial that interpolates an arbitrary sign pattern \( v \in \mathbb{C}^{|T|} \) on a set of points \( T \) separated by a minimum distance of \( 1.26/f_c \). As a first idea, we could consider interpolating the sign pattern with a Dirichlet kernel,

\[
K(f, t) := \frac{1}{2f + 1} \sum_{k=-f}^{f} e^{i2\pi kt},
\]

which can also be written as

\[
K(f, t) = \begin{cases} 
1 & \text{if } t = 0 \\
\frac{1}{(2f + 1) \sin (\pi t)} & \text{otherwise}.
\end{cases} 
\]

The result of the interpolation would satisfy (2.8) and (2.9) by construction, but controlling its magnitude on the off-support would be problematic because of the slow decay of the tail of \( K \). An alternative is to use a faster-decaying kernel instead, as in [18], which uses the fourth power of the Dirichlet kernel, also known as Jackson kernel. Taking the fourth power of a kernel is equivalent to repeatedly convolving it with itself in the frequency domain, which smooths out the discontinuities of its spectrum. The resulting kernel is consequently not as spiky at the origin, but has a faster decaying tail. Generalizing this idea, we can take the product of \( p \) kernels with different bandwidths \( \gamma_1 f_c, \gamma_2 f_c, \ldots, \gamma_p f_c \) to achieve a better tradeoff between the behavior of the resulting kernel at the origin and the decay of its tail. This yields a kernel of the form

\[
K_{\gamma}(t) := \prod_{j=1}^{p} K(\gamma_j f_c, t) = \sum_{k=-f_c}^{f_c} c_k e^{i2\pi kt},
\]

where \( c \in \mathbb{C}^n \) is the convolution of the Fourier coefficients of \( K(\gamma_1 f_c, t), K(\gamma_2 f_c, t), \ldots, K(\gamma_p f_c, t) \). As long as \( \sum_{j=1}^{p} \gamma_j = 1 \), the cutoff frequency of the kernel is indeed \( f_c \). We fix \( p = 3 \) and \( \gamma_1 = 0.247 \), \( \gamma_2 = 0.339 \) and \( \gamma_3 = 0.414 \). Figure 10 shows \( K_{\gamma} \) and its spectrum.
Fig. 11. Dual-polynomial if we only use $K_γ$ to interpolate the sign pattern (left), as opposed to also incorporating its derivative into the construction, and forcing the derivative of the polynomial to be zero at the interpolation points (right).

Interpolating the sign pattern with an adequately selected kernel is not sufficient to produce a valid construction. The reason is that the magnitude of the polynomial will tend to exceed one near the elements of the support $T$. This can be avoided, however, by forcing the derivative of the kernel to be zero at those points, an idea introduced in [18]. As a result, the magnitude of the polynomial has a local extremum at the interpolation point and obeys condition (2.11) for a sufficiently large minimum separation, as illustrated in Fig. 11. In order to enforce the extra constraint on the derivative of $q$, we need more degrees of freedom in the construction. For this purpose, we incorporate the derivative of the kernel, so that the construction is of the form

$$q(t) = \sum_{t_j \in T} \alpha_j K_γ(t - t_j) + \beta_j K_γ^{(1)}(t - t_j), \quad (4.3)$$

where $\alpha, \beta \in \mathbb{C}^{\lvert T \rvert}$ are coefficient sequences satisfying

$$q(t_k) = \sum_{t_j \in T} \alpha_j K_γ(t_k - t_j) + \beta_j K_γ^{(1)}(t_k - t_j) = v_k, \quad t_k \in T, \quad (4.4)$$

$$q_R^{(1)}(t_k) + \imath q_I^{(1)}(t_k) = \sum_{t_j \in T} \alpha_j K_γ^{(1)}(t_k - t_j) + \beta_j K_γ^{(2)}(t_k - t_j) = 0, \quad t_k \in T, \quad (4.5)$$

and $q_R$ and $q_I$ are the real and imaginary parts of $q$.

This construction trivially satisfies (2.8) because both $K_γ$ and $K_γ^{(1)}$ are low pass. The following lemma, proved in Section 4.2, establishes that (2.9) holds and derives some bounds on the interpolation coefficients.

**Lemma 4.1** Under the hypotheses of Theorem 2.2, there exist coefficient vectors $\alpha$ and $\beta$ obeying (4.4) and (4.5), such that

$$\|\alpha\|_{\infty} \leq \alpha_U := 1 + 2.370 \times 10^{-2},$$

$$\|\beta\|_{\infty} \leq \beta_U := 4.247 \times 10^{-2}\lambda_c.$$
In addition, if $v_1 = 1$,

$$\text{Re} \alpha_1 \geq \alpha_1^L := 1 - 2.370 \times 10^{-2}, \quad |\text{Im} \alpha_1| \leq \alpha_1^U := 2.370 \times 10^{-2}. \quad (4.7)$$

The proof is completed by two lemmas, proved in Sections 4.3 and 4.4, which show that our construction satisfies (2.10) and (2.11).

**Lemma 4.2** Fix $t_0 \in T$. Under the hypotheses of Theorem 2.2, $|q(t)| < 1$ for $|t - t_0| \in (0, 0.288316 \lambda_c]$ and there exist numerical constants $C_0 \in (0, 1)$, $C_1$ and $C_2$ such that (2.11) holds.

**Lemma 4.3** Fix $t_0 \in T$. Let $t_+$ be the nearest element of $T$ such that $t_+ > t_0$ and $t_-$ be the nearest element of $T$ such that $t_- < t_0$. Under the hypotheses of Theorem 2.2, $|q(t)| < 1$ for $t_0 - (t_0 - t_-)/2 \leq t < t_0$ and $t_0 < t \leq t_0 + (t_+ - t_0)/2$.

The proofs of Lemmas 4.1–4.3 make heavy use of bounds on $K_\gamma$ that are presented in Section 4.1.

The proofs of the different lemmas required to establish our main result involve some rather intricate calculations. In order to ease the task of checking the proof, we provide a Matlab script that performs these calculations at http://www.cims.nyu.edu/~cfgranda/scripts/superres_proof.zip.

### 4.1 Bounds on the interpolation kernel

Our results rely on sharp non-asymptotic bounds on the Dirichlet kernel $K$ (4.1) and its derivatives, which can in turn be used to establish bounds on the interpolation kernel $K_\gamma$ (4.2) and its derivatives. To avoid excessive clutter, we defer the precise expressions and their proofs to the Appendix. The following lemma, proved in Section C.1, provides upper and lower bounds near the origin.

**Lemma 4.4** For $f_c \geq 10^3$ and $\ell \in \{0, 1, 2, 3\}$, for any $t$ in the unit interval if $\tau$ is such that $|f_c \tau - \tau| \leq \varepsilon$, then

$$B_{\gamma, \ell}^L(\tau) - (2\pi)^{\ell+1} f_c^{\ell+1} \varepsilon \leq K_\gamma^{(\ell)}(t) \leq B_{\gamma, \ell}^U(\tau) + (2\pi)^{\ell+1} f_c^{\ell+1} \varepsilon,$$

where $B_{\gamma, \ell}^L$ and $B_{\gamma, \ell}^U$ are defined in Section B.1 of Appendix.

A direct corollary yields a sharp bound on the magnitude of the kernel and its derivatives.

**Corollary 4.5** For $f_c \geq 10^3$, $\ell \in \{0, 1, 2, 3\}$ and any $t$ in the unit interval, if $\tau$ is such that $|f_c \tau - \tau| \leq \varepsilon$, then

$$|K_\gamma^{(\ell)}(t)| \leq B_{\gamma, \ell}^\infty(\tau, \varepsilon) := \max \{|B_{\gamma, \ell}^L(\tau)|, |B_{\gamma, \ell}^U(\tau)|\} + (2\pi)^{\ell+1} f_c^{\ell+1} \varepsilon.$$

Away from the origin, we establish bounds on the absolute value of $K_\gamma$ and its derivatives that are weaker, but decreasing. The following lemma is proved in Section C.2 of Appendix.
LEMMA 4.6 For $f_c \geq 10^3$, $\ell \in \{0, 1, 2, 3\}$ and $|t| < 450/f_c$

$$|K^{(\ell)}_\gamma (t)| \leq b_{\gamma, \ell} (f_c t),$$

where $b_{\gamma, \ell}$ are decreasing functions defined in Section B.2 of Appendix.

Corollary 4.5 and Lemma 4.6 are combined in the next lemma, proved in Section C.3, to bound the sum of absolute values of shifted copies of the kernel (and of its derivatives) if they are separated by the minimum distance. The lemma is stated for an interval around the origin, but it holds for any position in the unit interval.

LEMMA 4.7 Assume $0 \in T$ and let $\tau_{\min} := \Delta_{\min} / \lambda_c = 1.26$ denote the minimum separation of the set $T$, scaled so that $\Delta (T) \geq \tau_{\min} / f_c$. If $f_c \geq 10^3$ and $\gamma = [0.247, 0.339, 0.414]^T$, for any $0 \leq t \leq \tau_{\min} / 2$, $\epsilon > 0$ and $\tau$ such that $\tau - \epsilon \leq f_c t \leq \tau$

$$\sum_{t_j \in T \setminus \{0\}} |K^{(\ell)}_\gamma (t - t_j)| \leq H_\ell (\tau) + H_\ell (-\tau),$$

where

$$H_\ell (\tau) := \sum_{j=1}^{20} \max \left\{ \max_{u \in G_{j, \epsilon}} B^{\infty}_{\gamma, \ell} (u, \epsilon), b_{\gamma, \ell} (j + 4 \tau_{\min}) \right\} + \tilde{C}_\ell,$$

where $G_{j, \epsilon}$ is an $\epsilon$-grid of equispaced points covering the interval $[j \tau_{\min} - \tau, (j + 4) \tau_{\min}]$ and

$$\tilde{C}_\ell := \sum_{j=21}^{400} b_{\gamma, \ell} (j - 1/2 \tau_{\min}) + C_\ell,$$

$C_0 = 8.85210^{-7}$, $C_1 = 5.56210^{-6}$, $C_2 = 3.49510^{-5}$, $C_3 = 2.19610^{-4}$.

4.2 Proof of Lemma 4.1

Set

$$(D_0)_{jk} = K_{\gamma} (t_j - t_k), \quad (D_1)_{jk} = K^{(1)}_{\gamma} (t_j - t_k), \quad (D_2)_{jk} = K^{(2)}_{\gamma} (t_j - t_k),$$

where $j$ and $k$ range from 1 to $|T|$. With this, (4.4) and (4.5) become

$$\begin{bmatrix} D_0 & D_1 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix}.$$
where $\|A\|_\infty$ is the usual infinity-norm of a matrix defined as $\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty = \max_i \sum_j |a_{ij}|$. This follows from $M^{-1} = (I - H)^{-1} = \sum_{k \geq 0} H^k$, $H = I - M$, where the series is convergent since $\|H\|_\infty < 1$. In particular,

$$\|M^{-1}\|_\infty \leq \frac{1}{1 - \|I - M\|_\infty}.$$  

(4.9)

By Lemma 4.7,

$$\|I - D_0\|_\infty \leq \max_{t_0 \in T} \sum_{t_j \in T \setminus \{t_0\}} |K_{y'} (t_0 - t_j)| \leq 2H_0 (0) \leq 1.85510^{-2},$$  

(4.10)

$$\|D_1\|_\infty \leq \max_{t_0 \in T} \sum_{t_j \in T \setminus \{t_0\}} |K_{y'}^{(1)} (t_0 - t_j)| \leq 2H_1 (0) \leq 0.1110 f_c.$$  

(4.11)

$$\|K_{y'}^{(2)} (0) I - D_2\|_\infty \leq \max_{t_0 \in T} \sum_{t_j \in T \setminus \{t_0\}} |K_{y'}^{(2)} (t_0 - t_j)| \leq 2H_2 (0) \leq 1.895 f_c.$$  

(4.12)

Note that $D_2$ is symmetric because the second derivative of the interpolation kernel $K_{y'}^{(2)}$ is symmetric. The bound (4.12) implies

$$\left\| I - \frac{D_2}{K_{y'}^{(2)} (0)} \right\| < 1,$$

which implies the invertibility of $D_2$. To control the infinity-norm of the inverse of $D_2$ we need bounds on the value of $K_{y'}^{(2)}$ at the origin.

LEMMA 4.8 If $f_c \geq f_{\text{min}}$

$$K_{y'}^{(2)} (0) \leq B_{y',2}^{0,U} < 0, \quad \text{where} \quad B_{y',2}^{0,U} = \sum_{j=1}^{p} B_{2}^{0,U} (\gamma_j) \quad \text{and} \quad B_{2}^{0,U} (\gamma_j) = -\frac{4\pi^2 \gamma_j^2 f_c^2}{3}.$$  

Proof. The result is a consequence of $K (f, 0) = 1$, $K^{(1)} (f, 0) = 0$, $K^{(2)} (f, 0) = -4\pi^2 f (1 + f) / 3$ and (C.19). 

By the lemma and (4.9), we have

$$\left\| D_2^{-1} \right\|_\infty \leq \frac{1}{\left\| K_{y'}^{(2)} (0) I - D_2 \right\|_\infty} \leq \frac{1}{\left\| B_{y',2}^{0,U} \right\| - \left\| K_{y'}^{(2)} (0) I - D_2 \right\|_\infty} \leq 0.3738 \lambda_{c}^2.$$  

(4.13)
Combining this with (4.10) and (4.11) yields
\[ \| I - (D_0 - D_1D_2^{-1}D_1) \|_\infty \leq \| I - D_0 \|_\infty + \| D_1 \|_\infty^2 \| D_2^{-1} \|_\infty \leq 2.31510^{-2} < 1. \] (4.14)

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

There just remains to bound the interpolation coefficients, which can be expressed as
\[
\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} I \\ -D_2^{-1}D_1 \end{bmatrix} C^{-1} v, \quad C := D_0 - D_1D_2^{-1}D_1,
\]
where \( C \) is the Schur complement. The relationships (4.9) and (4.14) immediately give a bound on the magnitude of the entries of \( \alpha \)
\[ \| \alpha \|_\infty = \| C^{-1} v \|_\infty \leq \| C^{-1} \|_\infty \leq 1 + 2.37010^{-2}. \]

Similarly, (4.11), (4.13) and (4.14) allow us to bound the entries of \( \beta \):
\[
\| \beta \|_\infty \leq \| D_2^{-1}D_1C^{-1} \|_\infty \\
\leq \| D_2^{-1} \|_\infty \| D_1 \|_\infty \| C^{-1} \|_\infty \leq 4.24710^{-2} \lambda_c.
\]

Finally, with \( v_1 = 1 \), we can use (4.14) to show that \( \alpha_1 \) is almost equal to 1. Indeed,
\[ |1 - \alpha_1| = |1 - (C^{-1} v)_1| \\
= |(I - C^{-1} v)_1| \\
\leq \| I - C^{-1} \|_\infty \\
= \| C^{-1} (I - C) \|_\infty \\
\leq \| C^{-1} \|_\infty \| I - C \|_\infty \leq 2.37010^{-2}. \]

This concludes the proof.

4.3 Proof of Lemma 4.2

Without loss of generality, we map the unit circle \( \mathbb{T} \) to the interval \([-\frac{1}{2}, \frac{1}{2}]\), and we set \( t_0 = 0 \) and \( v_0 = 1 \).

In order to bound the magnitude of the dual polynomial locally, we control the second derivative of its square around the origin. In particular, we bound it on an interval of the form \( f(t) \in [j\epsilon, (j + 1)\epsilon] \) for an integer \( j \geq 0 \) and \( \epsilon > 0 \). Let us decompose \( q = q_r + iq_i \) into its real and imaginary parts. The derivative is equal to
\[
\frac{d^2}{dt^2} \frac{q_r}{|q|^2} (t) = 2 \left(q_r t q_i^2 t + q_i t q_r^2 t + q_r^4 t^2 + q_i^4 t^2\right) \\
= 2 \left(q_r t q_i^2 t + q_i t q_r^2 t + |q| t^2\right).
\]
where \( \tilde{q}_1 (t) := q_R^{(1)} + iq_I^{(1)} \). To alleviate notation we define

\[
G_{\ell,j,e} (a_1, a_2) := a_1 B_{j,e}^\infty (je, e) + a_2 (H_{\ell} - j + 1 \epsilon + H_{dj} + 1 \epsilon),
\]

where \( B_{j,e}^\infty \) is defined in Corollary 4.5. By Lemmas 4.1, 4.4 and 4.7 and Corollary 4.5, we have

\[
q_R (t) = \sum_{t_j \in T} \text{Re} \left( \alpha_j \right) K_{j} (t - t_j) + \text{Re} \left( \beta_j \right) K_{j}^{(1)} (t - t_j)
\]

\[
\geq \text{Re} (\alpha_0) K_{j} (t) - \| \alpha \|_\infty \sum_{t_j \in T \setminus \{0\}} \left| K_{j} (t - t_j) \right| - \| \beta \|_\infty \sum_{t_j \in T} \left| K_{j}^{(1)} (t - t_j) \right|
\]

\[
\geq a_R^L \left( B_{j,e}^L, 2 \pi \epsilon \right) - \alpha^U (H_{0} - j + 1 \epsilon + H_{dj} + 1 \epsilon) - G_{1,j,e} (\beta^U, \beta^U)
\]

\[
:= J_{0,R}^L (j),
\]

for \( j \epsilon \leq f \leq (j + 1) \epsilon \) since \( B_{j,e}^L \) is positive in the interval of interest. Similarly,

\[
|q_R (t)| \leq G_{0,j,e} (\alpha^U, \alpha^U) + G_{1,j,e} (\beta^U, \beta^U) := J_{0,R}^U (j),
\]

\[
|q_I (t)| \leq G_{0,j,e} (\alpha_I^U, \alpha^U) + G_{1,j,e} (\beta^U, \beta^U) := J_{0,R}^I (j),
\]

\[
|\tilde{q}_1 (t)| \leq G_{1,j,e} (\alpha^U, \alpha^U) + G_{2,j,e} (\beta^U, \beta^U) := J_{1,R} (j),
\]

\[
\left| q_R^{(2)} (t) \right| \leq G_{2,j,e} (\alpha^U, \alpha^U) + G_{3,j,e} (\beta^U, \beta^U) := J_{2,R} (j),
\]

\[
\left| q_I^{(2)} (t) \right| \leq G_{2,j,e} (\alpha_I^U, \alpha^U) + G_{3,j,e} (\beta^U, \beta^U) := J_{2,I} (j),
\]

and if \( B_{j,e}^U (j \epsilon) + (2 \pi)^3 \epsilon < 0 \)

\[
q_R^{(2)} (t) \leq a_R^L \left( B_{j,e}^L, 2 \pi \epsilon + 2 \pi^3 \epsilon \right) + \alpha^U (H_{2} - j + 1 \epsilon + H_{2j} + 1 \epsilon) + G_{3,j,e} (\beta^U, \beta^U) := J_{2,R}^{\text{near}} (j).
\]

Combining these expressions, we obtain a bound on the second derivative of the square of the polynomial \( q \),

\[
\frac{d^2 t}{d |q|^2} (t) \leq F_{2,j},
\]

of the form

\[
F_{2,j} := \begin{cases} 
2 \left( J_{2,R}^{\text{near}} J_{0,R}^L j + J_{2,R} J_{0,R} j + J_{0,R}^2 \right), & \text{if } B_{j,e}^U (j \epsilon) < 0 \\
2 \left( J_{2,R} J_{0,R}^L j + J_{2,R} J_{0,R} j + J_{0,R}^2 \right), & \text{otherwise},
\end{cases}
\]

(4.16)

which holds for any \( t \) such that \( f_\epsilon \leq t \leq (j + 1) \epsilon \). This bound allows us to control the curvature of the squared magnitude of \( |q| \) on a fine grid, which can in turn be leveraged to obtain a bound on \( |q| \). This is made precise by the following lemma, proved in Section C.5 of Appendix.
LEMMA 4.9  If $F$ is a twice continuously differentiable real-valued function such that

$$F(0) = 1, \quad F^{(1)}(0) = 0, \quad \sup_{j \leq i \leq (j+1)\epsilon} F^{(2)}(\tilde{i}) \leq F_{2j},$$

for a certain positive real number $\epsilon$ and all non-negative integers $j \geq 0$, then if $k \epsilon \leq t \leq (k + 1) \epsilon$ for an integer $k$, we have

$$F(t) \leq F_k^\infty := 1 + \epsilon^2 \left( \sum_{j=0}^{k-1} \sum_{l=1}^{j-1} F_{2j} + \mathbb{1}_{F_{2j}>0}F_{2j} + \mathbb{1}_{F_{2j}>0}F_2 \right),$$

where

$$F_2 = \sum_{j=0}^{k-1} F_{2j} + \mathbb{1}_{F_{2j}>0}F_{2k}.$$

We apply the lemma by setting $F = |q|^2$. By Lemma 4.1, $F(0) = |q(0)|^2 = 1$ and $F^{(1)}(0) = 2q_R(0)q_R^{(1)}(0) + 2q_I(0)q_I^{(1)}(0) = 0$. To obtain a bound on $|q|^2$, we evaluate the bound computationally on a grid with step size $\epsilon = 10^{-6}$ covering the interval $[0, 0.3]$. The corresponding values of $F_k^\infty$ are plotted on the left of Fig. 12. Since $F_k^\infty$ lies below $1 - 6.39410^{-7}$ between $k = 1$ and $k = 0.288316/\epsilon$, we have

$$|q(t)| \leq 1 - 3.19710^{-7}, \quad |t - t_k| \leq 0.288316\lambda, \quad t_k \in T,$$

since we fixed $t_0 = 0$ without loss of generality.

For $t$ belonging to the interval $(0, \epsilon)$ we apply the following inequality, which follows from the fundamental theorem of calculus, and the facts that $F'(0) = 0$ and $F_{2.0} \leq -3.23f_c^2$ (the proof of
Lemma 4.9 is based on a similar argument)

\[
\sup_{t \in (0, \epsilon)} |q(t)|^2 \leq \sup_{t \in (0, \epsilon)} F(0) + t^2 \sup_{u \in [0, \epsilon]} F''(u)
\]

\[
= \sup_{t \in (0, \epsilon)} 1 + F_{2,0} t^2
\]

\[
\leq \sup_{t \in (0, \epsilon)} 1 - 3.23 f^2 c t^2 < 1.
\]

Finally, to establish the quadratic bound (2.11) on the magnitude of \( q \), we prove that in the interval the second derivative of \( |q| \) lies strictly below zero. In fact, we can make use of the bound (4.16). Since

\[
\frac{d^2 t}{d|q|^2}(t) = -\frac{(q_R(q_R^2 + q_Iq_I^2) + q_Iq_I^2)}{|q(t)|^3} + \frac{q_R(t) q_R^{(2)}(t) + q_I(t) q_I^{(2)}(t) |\bar{q}_1(t)|^2}{|q(t)|} - \frac{1}{2} \frac{d^2 t}{d|q|^2}(t)
\]

this second derivative is negative if \( d^2 |q|^2/dt^2 \) is negative. To bound this latter quantity, we evaluate the bound \( F_{2,j} \) on a grid with step size \( \epsilon = 10^{-6} \) covering the interval \([0, 0.2] \). The result is plotted on the right of Fig. 12. Between 0 and 0.110497 the bound is strictly negative, which immediately implies that this is also the case for the second derivative of \( |q| \). This establishes the upper bound in (2.11), because \( q(t_0) = 1 \) and the derivative of \( |q| \) is zero by Lemma 4.1. To prove the lower bound, we just need to bound the second derivative of the real and imaginary parts of \( q - v_j \). Since the argument is almost identical to the one used to bound the second derivative of \( |q|^2 \), we omit the details.

4.4 Proof of Lemma 4.3

By Lemmas 4.1, 4.4 and 4.7 and Corollary 4.5

\[
|q(t)| \leq G_{0,j,\epsilon}(\alpha^U, \alpha^U) + G_{1,j,\epsilon}(\beta^U, \beta^U),
\]

for any \( t \leq \Delta_{\min}/2 \) in \([j, (j+1) \epsilon] \), where \( G_{j,\epsilon} \) is defined in (4.15). We evaluate the bound numerically on a grid with step size \( \epsilon = 10^{-6} \) to establish that in the interval \([0.288316, \tau_{\min}/2] \)

\[
|q(t)| \leq 1 - 4.64110^{-3}.
\]

The numerical values are plotted in Fig. 13.

Together with Lemma 4.2, this completes the proof, except for the following case: The first spike to the right of \( t_0 \), which we denote by \( t_1 \), might be at a larger distance than \( \Delta_{\min} \), so it is still necessary to
bound \( q \) in the interval \([\Delta_{\text{min}}/2, t_0 + (t_1 - t_0)/2]\). Let \( t \) be a point in this interval, then

\[
|q(t)| = \left| \sum_{t_j \in T} a_j K_y(t - t_j) + \beta_j K_y^{(1)}(t - t_j) \right|
\]

\[
\leq \alpha U \left( K_y(t - t_0) + \sum_{t_j \in T \cap [t_0]} |K_y(t - t_j)| \right) + \beta U \left( K_y^{(1)}(t_0) + \sum_{t_j \in T \cap [t_0]} |K_y^{(1)}(t - t_j)| \right)
\]

\[
\leq \alpha U \left( \max_{u \in G} B_{\gamma,0}(u, \epsilon), b_{\gamma,0}(2\tau_{\text{min}}) \right) + H_0 \left( \tau_{\text{min}}/2 \right) + H_0(0)
\]

\[
+ \beta U \left( \max_{u \in G} B_{\gamma,1}(u, \epsilon), b_{\gamma,1}(2\tau_{\text{min}}) \right) + H_1 \left( \tau_{\text{min}}/2 \right) + H_1(0),
\]

where \( G \) is a grid with step size \( \epsilon \) covering the interval \([\tau_{\text{min}}/2, 2\tau_{\text{min}}]\). The bound follows from Lemmas 4.1, 4.4 and 4.7 and Corollary 4.5. In particular, from the definition of \( H_\ell \) in Lemma 4.7 it is not difficult to show that if \( \tau_{\text{min}}/2 \leq t \leq t_1 - \tau_{\text{min}}/2 \), then

\[
\sum_{t_j \in T \cap (0, \frac{1}{2})} |K_y^{(\ell)}(t - t_j)| \leq H_\ell(\tau_{\text{min}}/2) \quad \text{and} \quad \sum_{t_j \in T \cap [-\frac{1}{2}, 0)} |K_y^{(\ell)}(t - t_j)| \leq H_\ell(0).
\]

Finally, we evaluate the bound numerically for a step size of \( \epsilon = 10^{-6} \) to prove that

\[
|q(t)| \leq 0.896
\]

in \([\Delta_{\text{min}}/2, t_0 + (t_1 - t_0)/2]\).
5. Conclusion and future research

In this work, we provide a reasonably tight characterization of the performance of super-resolution via convex programming when the signal consists of point sources in one dimension that are separated by a minimum distance. Interesting research directions include developing a sharper bound than the one provided by [18] for two-dimensional signals, studying other conditions that allow for a small degree of clustering in the support of the signal as in [54], and extending the analysis to other signal models such as piecewise-constant signals in two dimensions. We have also derived optimization-based algorithms to tackle the demixing of sines and spikes, and the super-resolution of point sources with a common support. In both cases, it would be interesting to derive theoretical guarantees by building the appropriate dual certificates, as outlined in Sections 3.1.3 and 3.2.3. Finally, applying optimization-based approaches in real applications will require developing efficient algorithms to solve the corresponding optimization programs on large-scale datasets (see [15] for recent work in this direction that is relevant to super-resolution).

Acknowledgements

The author is grateful to Emmanuel Candès for useful discussions and to the anonymous reviewers for helpful suggestions.

References


Appendix A. Proof of auxiliary results in Sections 2 and 3

A.1 Proof of Proposition 2.4

Let us define

\[ z_0(t) := [e^{i2\pi f_1 t} \quad e^{i2\pi (f_1 - 1) t} \ldots e^{-i2\pi f_1 t}]^\top, \]
\[ z_1(t) := [1 \quad e^{i2\pi t} \ldots e^{i2\pi nt}]^\top. \]

Note that \( (F_n^* C_k)(t) = z_0(t)^* C_k \), so that \( \sum_k |F_n^* C_k|^2 = z_0(t)^* C C^* z_0(t) \). If (2.15) holds, then \( A - C C^* \succeq 0 \) and \( T^*(A) = e_1 \). This implies

\[
\sum_k \left| (F_n^* C_k)(t) \right|^2 = z_0(t)^* C C^* z_0(t) \leq z_0(t)^* A z_0(t) = \text{Tr} \left( z_0(t)^* A z_0(t) \right) = \langle A, z_0(t)^* z_0(t) \rangle = \langle A, T(z_1(t)) \rangle.
\]
\[
\langle T^* (A), z_1 (t) \rangle = (z_1(t))_1 = 1.
\]

For the converse, \( \sum_k |F^*_n C_k|^2 \leq 1 \) implies that the trigonometric polynomial \( 1 - z_0(t)^* C C^* z_0(t) \) is non-negative. By the Fejér–Riesz Theorem there exists a polynomial \( P(t) = \tilde{c}^* z_0(t) \) such that
\[
1 - z_0(t)^* C C^* z_0(t) = |P(t)|^2 = z_0(t)^* \tilde{c} \tilde{c}^* z_0(t).
\] (A.1)

Now let us set \( \Lambda = C C^* + \tilde{c} \tilde{c}^* \). \( \Lambda \) and \( \Lambda - C C^* \) are positive semi-definite by construction, which implies
\[
\begin{bmatrix}
\Lambda \\
C^* \\
I
\end{bmatrix} \succeq 0.
\]

Furthermore, by (A.1)
\[
1 = z_0(t)^* A z_0(t) = \langle A, z_0(t) z_0(t)^* \rangle = \langle A, T (z_1(t)) \rangle = \langle T^* (A), z_1 (t) \rangle
\]
for all \( t \), which is only possible if \( T^* (A) = e_1 \). This completes the proof.

A.2 Proof of Lemma 3.1

The Lagrangian is equal to
\[
\mathcal{L}(\hat{x}, \hat{s}, c) = \|\hat{x}\|_{TV} + \eta \|\hat{s}\|_1 + \langle c, y - F_n \hat{x} - \hat{s} \rangle
\]
\[
= \langle c, y \rangle + \|\hat{x}\|_{TV} - \langle F^*_n c, \hat{x} \rangle + \eta \|\hat{s}\|_1 - \langle c, \hat{s} \rangle,
\]

where \( c \in \mathbb{C}^n \) is the dual variable. The lower bound on the cost function provided by minimizing the Lagrangian over \( \hat{x} \) and \( \hat{s} \) is trivial \((-\infty)\) unless \( |\langle F^*_n c \rangle (t)\| \leq 1 \) for all \( t \) and \( \|c\|_\infty \leq \eta \). Under these constraints, the minimum is attained for \( \hat{x} = 0 \) and \( \hat{s} = 0 \) for any value of \( c \). Maximizing the remaining expression over \( c \) yields the dual problem.

A.3 Proof of Lemma 3.2

The interior of the feasible set of Problem (3.4) contains the origin and is consequently non-empty, so strong duality holds by a generalized Slater condition [57], and we have
\[
\sum_{t_j \in \hat{T}} |\hat{a}_j| + \eta \sum_{l \in \hat{S}} |\hat{s}_l| = \|\hat{x}\|_{TV} + \eta \|\hat{s}\|_1 = \langle \hat{c}, y \rangle
\]
\[
= \langle \hat{c}, F_n \hat{x} + \hat{s} \rangle
\]
\[
= \text{Re} \left[ \sum_{t_j \in \hat{T}} |\hat{a}_j| (F^*_n \hat{c}) (t_j) e^{i\phi_j} + \sum_{l \in \hat{S}} |\hat{s}_l| \tilde{c}_l e^{i\psi_l} \right].
\]
By Hölder’s inequality and the constraints on \( \hat{c} \), this implies the result (otherwise equality would not be possible).

### A.4 Proof of Lemma 3.3

For a vector \( v \), an atomic measure \( z \) and a set \( A \), we denote by \( v_A \) and \( z_A \) the restriction of \( v \) and \( z \) to the subset of their support indexed by \( A \), respectively. Let us consider an arbitrary solution \( x' \) and \( s' \), such that \( x' \neq x \) or \( s' \neq s \). Due to the constraints of the optimization problem, \( x' \) and \( s' \) satisfy

\[
y = \mathcal{F}_n x + s = \mathcal{F}_n x' + s'. \tag{A.2}
\]

We define the matrix \([F_T \quad I_S]\), where

\[
(F_T)_{lj} = e^{i2\pi lj} \quad \text{for } 1 \leq l \leq n, 1 \leq j \leq |T|,
\]

\[
(I_S)_{lj} = \begin{cases} 1 & \text{if } l = s(j) \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 1 \leq l \leq n, 1 \leq j \leq |S|.
\]

If \(|T| + |S| \leq n\), then \([F_T \quad I_S]\) is full rank (this follows from the fact that \( F_T \) is a submatrix of a Vandermonde matrix). As a result \( x_T' \) and \( s_S' \) are not both equal to zero. Indeed,

\[
\mathcal{F}_n x_T' + s_S' = \mathcal{F}_n x' + s' - \mathcal{F}_n x_T' + s_S' \quad \text{by (A.2)}
\]

\[
= [F_T \quad I_S] \begin{bmatrix} a_T - a_T' \\ s_S - s_S' \end{bmatrix} \neq 0, \tag{A.3}
\]

as this would imply \( x = x' \) and \( s = s' \) (\( a \) is the vector containing the coefficients of \( x \)). Finally, we conclude

\[
\|x'\|_{TV} + \eta \|s'\|_1 = \|x_T'\|_{TV} + \|x_T'\|_{TV} + \eta \|s_S'\|_1 + \eta \|s_S'\|_1
\]

\[
> \|x_T'\|_{TV} + \langle q, x_T' \rangle + \eta \|s_S'\|_1 + \langle c, s_S' \rangle \quad \text{by (3.6) and (3.7)}
\]

\[
\geq \langle q, x' \rangle + \langle c, s' \rangle \quad \text{by (3.6) and (3.7)}
\]

\[
= \langle \mathcal{F}_n^* c, x' \rangle + \langle c, s' \rangle
\]

\[
= \langle c, \mathcal{F}_n x' + s' \rangle \quad \text{by (A.2)}
\]

\[
= \langle c, \mathcal{F}_n x + s \rangle \quad \text{by (A.2)}
\]

\[
= \langle \mathcal{F}_n^* c, x \rangle + \langle c, s \rangle
\]

\[
= \langle q, x \rangle + \langle c, s \rangle
\]

\[
\geq \|x\|_{TV} + \eta \|s\|_1 \quad \text{by (3.6), (3.7) and Hölder’s inequality.}
\]
A.5 Proof of Lemma 3.4

The Lagrangian is equal to

\[ \mathcal{L}(\tilde{X}, C) = \|\tilde{X}\|_{gTV} + \langle C, Y - [\mathcal{F}_n\tilde{X}_1 \quad \mathcal{F}_n\tilde{X}_2 \quad \cdots \quad \mathcal{F}_n\tilde{X}_m] \rangle \]

\[ = \langle C, Y \rangle + \|\tilde{X}\|_{gTV} - \sum_{k=1}^{m} \langle C_k, \mathcal{F}_n\tilde{X}_k \rangle \]

\[ = \langle C, Y \rangle + \|\tilde{X}\|_{gTV} - \sum_{k=1}^{m} \langle \mathcal{F}_n^* C_k, \tilde{X}_k \rangle \]

\[ = \langle C, Y \rangle + \sup_{\|\mathcal{F}(t)\|_2 \leq 1} \sum_{k=1}^{m} \text{Re} \left\{ \int_{T} \mathcal{F}_k^* (t) X_k (dt) \right\} - \sum_{k=1}^{m} \text{Re} \left\{ \int_{T} \mathcal{F}_n^* C_k^* (t) X_k (dt) \right\}, \]  

where \( C \in \mathbb{C}^{n \times m} \) is the dual variable. The lower bound on the cost function provided by minimizing the Lagrangian over \( \tilde{X} \) is trivial \((-\infty)\) unless

\[ \sum_{k=1}^{m} |(\mathcal{F}_n^* C_k)(t)|^2 \leq 1 \]

for all \( t \). Under these constraints, the minimum is attained for \( \tilde{X} = 0 \). Maximizing the remaining expression over \( C \) yields the dual problem.

A.6 Proof of Lemma 3.5

Let us define the matrix \( P \in \mathbb{C}^{[T] \times m} \),

\[ P_{jk} := \left( \mathcal{F}_n^* \hat{C}_k \right) (t_j). \]

The result is an immediate consequence of the following equality

\[ P_j = \frac{A_j}{\|A_j\|}, \quad (A.4) \]

where \( M_j \) denotes the \( j \)th row of a matrix \( M \).

By strong duality, which holds by a generalized Slater condition [57], we have that

\[ \|\hat{X}\|_{gTV} = \langle Y, \hat{C} \rangle. \]

On the one hand,

\[ \|\hat{X}\|_{gTV} = \sum_{j=1}^{[T]} \|A_j\|_2. \]
On the other,

\[
\langle Y, \hat{C} \rangle = \sum_{k=1}^{m} \langle Y_k, \hat{C}_k \rangle \\
= \sum_{k=1}^{m} \langle F_n \hat{X}_k, \hat{C}_k \rangle \\
= \sum_{k=1}^{m} \langle \hat{X}_k, F_n^* \hat{C}_k \rangle \\
= \sum_{j=1}^{|T|} \sum_{k=1}^{m} \text{Re} \left\{ \hat{A}_{jk} F_n^* \hat{C}_k (\hat{t}_j) \right\} \\
= \sum_{j=1}^{|T|} \langle A_j, P_j \rangle.
\]

This combined with the following lemma implies (A.4).

**Lemma A.1** Let \( M, N \in \mathbb{C}^{n_1 \times n_2} \), such that the rows of \( N \) have non-zero norm, if

\[
\sum_{j=1}^{n_1} \langle M_{j,:}, N_{j,:} \rangle = \sum_{j=1}^{n_1} \| N_j \|_2, \tag{A.5}
\]

\[
\max_{1 \leq j \leq n_1} \| M_j \|_2 \leq 1 \tag{A.6}
\]

then

\[
M_{jk} = \frac{N_{jk}}{\| N_j \|_2}.
\]

**Proof.** By (A.6) and the Cauchy–Schwarz inequality, \( \langle M_{j,:}, N_{j,:} \rangle \leq \| N_j \|_2 \) for all \( j \). If for any of them \( \langle M_{j,:}, N_{j,:} \rangle < \| N_j \|_2 \) then (A.5) cannot hold, so

\[
\langle M_{j,:}, N_{j,:} \rangle = \| N_j \|_2 \quad \text{for all } j. \tag{A.7}
\]

This directly implies the desired result because for any \( j \)

\[
\left\| M_{j,:} - \frac{N_{j,:}}{\| N_j \|_2} \right\|_2^2 = \left\langle M_{j,:} - \frac{N_{j,:}}{\| N_j \|_2}, M_{j,:} - \frac{N_{j,:}}{\| N_j \|_2} \right\rangle \\
= \| M_{j,:} \|_2^2 - 2 \frac{\langle M_{j,:}, N_{j,:} \rangle}{\| N_j \|_2} + 1 \\
= \| M_{j,:} \|_2^2 - 1 = 0.
\]

\( \square \)
A.7 Proof of Lemma 3.6

Any $m$-dimensional measure that is feasible for Problem (3.10) can be written as $X + H$, where $H$ is an $m$-dimensional measure such that $\mathcal{F}_n H_k = 0$ for $1 \leq k \leq m$. This implies that

$$\langle Q, H \rangle = \sum_{k=1}^{m} \langle Q_k, H_k \rangle = \sum_{k=1}^{m} \langle \mathcal{F}^*_n C_k, H_k \rangle = \sum_{k=1}^{m} \langle C_k, \mathcal{F}_n H_k \rangle = 0. \tag{A.8}$$

For any $m$-dimensional measure $\tilde{X}$ composed by $m$ measures $\tilde{X}_k$ with bounded TV-norm supported on the unit interval $T$, we denote its restriction to a subset $S$ of $T$ by $\tilde{X}_S$ and its restriction to the complement of $S$ by $\tilde{X}_S^c$. Since $\|Q(t)\|_2 \leq 1$ for any $t$, we have $\langle Q, X + H_T \rangle = \|X + H_T\|_{gTV}$. Similarly, since $\|Q(t)\|_2 < 1$ for any $t$ in $T^c$, $\langle Q, H_T^c \rangle < \|H_T^c\|_{gTV}$. This combined with the fact that the $gTV$-norm is separable and that by construction $\langle Q, X \rangle = \|X\|_{gTV}$ allows us to conclude that

$$\|X + H\|_{gTV} = \|X + H_T\|_{gTV} + \|H_T^c\|_{gTV}$$

$$> \langle Q, X + H_T \rangle + \langle Q, H_T^c \rangle$$

$$= \langle Q, X \rangle + \langle Q, H \rangle$$

$$= \|X\|_{gTV},$$

as long as $H_T^c$ is non-zero. Since the measurements are injective on $T$ (the restriction of the measurement operator is a submatrix of a Vandermonde matrix and thus full rank) $H_T^c = 0$ implies $H = 0$ and the proof is complete.

Appendix B. Bounds on the interpolation kernel

B.1 Upper and lower bounds near the origin in Lemma 4.4

We use $f_{min}$ to denote a lower bound on the cutoff frequency $f_c$. The letter U denotes an upper bound and the letter L denotes a lower bound. The bounds on $K_\gamma$ are

$$B_{\gamma,0}^U (\tau) := \max_{A \in \{L,U\}^p} \prod_{j=1}^{p} B_{0}^{A_j} (\gamma_j, \tau),$$

$$B_{\gamma,0}^L (\tau) := \min_{A \in \{L,U\}^p} \prod_{j=1}^{p} B_{0}^{A_j} (\gamma_j, \tau), \tag{B.1}$$

where $\{L, U\}^p$ denotes the set of strings of length $p$, where each character is equal to U or L. $B_{\ell}^U$ and $B_{\ell}^L$ for $\ell = 0, 1, 2, 3$ are bounds on the Dirichlet kernel $K$ and its derivatives which are defined below. The bounds on $K_{\gamma}^{(1)}, K_{\gamma}^{(2)}$ and $K_{\gamma}^{(3)}$, which follow, also depend on these functions.

$$B_{\gamma,1}^U (\tau) := \sum_{i=1}^{p} \max_{A \in \{L,U\}^p} B_{1}^{A_i} (\gamma_i, \tau) \prod_{j=1, j \neq i}^{p} B_{0}^{A_j} (\gamma_j, \tau),$$

$$B_{\gamma,1}^L (\tau) := \sum_{i=1}^{p} \min_{A \in \{L,U\}^p} B_{1}^{A_i} (\gamma_i, \tau) \prod_{j=1, j \neq i}^{p} B_{0}^{A_j} (\gamma_j, \tau), \tag{B.2}$$
\[ B^U_{\tau,2}(\tau) := \sum_{i=1}^{p} \max_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) \prod_{j=1,j\neq i}^{p} B^A_j(\gamma_j, \tau) \]
\[ + \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \max_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) B^A_j(\gamma_j, \tau) \prod_{k=1,k\neq(i,j)}^{p} B^A_k(\gamma_k, \tau), \]
\[ B^L_{\tau,2}(\tau) := \sum_{i=1}^{p} \min_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) \prod_{j=1,j\neq i}^{p} B^A_j(\gamma_j, \tau) \]
\[ + \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \min_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) B^A_j(\gamma_j, \tau) \prod_{k=1,k\neq(i,j)}^{p} B^A_k(\gamma_k, \tau), \]
\[ B^U_{\tau,3}(\tau) := \sum_{i=1}^{p} \max_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) \prod_{j=1,j\neq i}^{p} B^A_j(\gamma_j, \tau) \]
\[ + 3 \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \max_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) B^A_j(\gamma_j, \tau) \prod_{k=1,k\neq(i,j)}^{p} B^A_k(\gamma_k, \tau) \]
\[ + \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \sum_{k=1,k\neq(i,j)}^{p} \max_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) B^A_j(\gamma_j, \tau) B^A_k(\gamma_k, \tau) \]
\[ \times \prod_{l=1,l\neq(i,j,k)}^{p} B^A_l(\gamma_l, \tau), \]
\[ B^L_{\tau,3}(\tau) := \sum_{i=1}^{p} \min_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) \prod_{j=1,j\neq i}^{p} B^A_j(\gamma_j, \tau) \]
\[ + 3 \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \min_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) B^A_j(\gamma_j, \tau) \prod_{k=1,k\neq(i,j)}^{p} B^A_k(\gamma_k, \tau) \]
\[ + \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \sum_{k=1,k\neq(i,j)}^{p} \min_{A \in [L,U]^p} B^A_i(\gamma_i, \tau) B^A_j(\gamma_j, \tau) B^A_k(\gamma_k, \tau) \]
\[ \times \prod_{l=1,l\neq(i,j,k)}^{p} B^A_l(\gamma_l, \tau). \]

These bounds are plotted for \( \gamma = [0.247, 0.339, 0.414]^T \) in Fig. B1.

We separate the bounds on \( K, B^U_0 \) and \( B^L_0 \), into two intervals.

\[ B^U_0(\gamma_0, \tau) := \begin{cases} B^U_{\text{near},0}(\gamma_0, \tau) & \text{if } 0 \leq |\tau| \leq \tau_0(\gamma_0) \\ B^U_{\text{far},0}(\gamma_0, \tau) & \text{if } |\tau| > \tau_0(\gamma_0), \end{cases} \]

and do exactly the same for \( B^L_0 \). The intervals depend on the value of \( \gamma_0 \). In particular: \( \tau_0(0.247) = 0.9112, \tau_0(0.339) = 0.6615, \tau_0(0.414) = 0.5401. \)
The bounds near the origin are of the form

\[ B_{\text{near},0}^{U}(\gamma_{0}, \tau) \]

\[ := \frac{1 - 2\pi^{2}\gamma_{0}^{2}\tau^{2}/3 + 2\pi^{4}\gamma_{0}^{4}\tau^{4}/15 + (1 + 1/2\gamma_{0}f_{\text{min}})^{4} - 4\pi^{6}\gamma_{0}^{6}\tau^{6}/315 + 2\pi^{8}\gamma_{0}^{8}\tau^{8}/2835 (1 + 1/2\gamma_{0}f_{\text{min}})^{8}}{1 - 2\pi^{2}\gamma_{0}^{2}\tau^{2}/3 + 2\pi^{4}\gamma_{0}^{4}\tau^{4}/15 - 4\pi^{6}\gamma_{0}^{6}\tau^{6}/315} \left( 1 - \pi^{2}\tau^{2}/6f_{\text{min}}^{2} \right) \times \frac{\sin (2\pi\gamma_{0}\tau)}{2\pi\gamma_{0}\tau} \]  

(B.6)

\[ B_{\text{near},0}^{L}(\gamma_{0}, \tau) \]

\[ := \frac{1 - 2\pi^{2}\gamma_{0}^{2}\tau^{2}/3 + 2\pi^{4}\gamma_{0}^{4}\tau^{4}/15 + (1 + 1/2\gamma_{0}f_{\text{min}})^{4} - 4\pi^{6}\gamma_{0}^{6}\tau^{6}/315 + 2\pi^{8}\gamma_{0}^{8}\tau^{8}/2835 (1 + 1/2\gamma_{0}f_{\text{min}})^{6} \sin (2\pi\gamma_{0}\tau)}{2\pi\gamma_{0}\tau} \].

(B.7)
Farther from the origin,

\[
B_{\text{far},0}^U (\gamma_0, \tau) := \frac{\sin (2\pi \gamma_0 \tau)}{2\pi \gamma_0 \tau} \left( 1 - \mathbb{I}_{\sin(2\pi \gamma_0 \tau) < 0} \left( \pi^2 \tau^2 / 2f_{\text{min}}^2 \right) \right) + \frac{\cos (2\pi \gamma_0 \tau)}{1 + 2\gamma_0 f_{\text{min}}} ,
\]

\[
B_{\text{far},0}^L (\gamma_0, \tau) := \frac{\sin (2\pi \gamma_0 \tau)}{2\pi \gamma_0 \tau} \left( 1 - \mathbb{I}_{\sin(2\pi \gamma_0 \tau) \geq 0} \left( \pi^2 \tau^2 / 2f_{\text{min}}^2 \right) \right) + \frac{\cos (2\pi \gamma_0 \tau)}{1 + 2\gamma_0 f_{\text{min}}} ,
\]

where \( \mathbb{I}_S \) is an indicator function that is equal to one when statement \( S \) holds and zero otherwise.

The bounds on \( K^{(1)}, K^{(2)} \) and \( K^{(3)} \) are

\[
B_1^U (\gamma_0, \tau) := \frac{f_c (\cos (2\pi \gamma_0 \tau) - B_0^L (\gamma_0, \tau))}{\tau} \left( 1 - \mathbb{I}_{\cos(2\pi \gamma_0 \tau) \leq B_0^L (\gamma_0, \tau)} \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) \]

\[\quad - \frac{\pi f_c}{\tau} \mathbb{I}_{\sin(2\pi \gamma_0 \tau) < 0} \left( \pi^2 \tau^2 / 2f_{\text{min}}^2 \right) ,
\]

\[
B_1^L (\gamma_0, \tau) := \frac{f_c (\cos (2\pi \gamma_0 \tau) - B_0^U (\gamma_0, \tau))}{\tau} \left( 1 - \mathbb{I}_{\cos(2\pi \gamma_0 \tau) \geq B_0^U (\gamma_0, \tau)} \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) \]

\[\quad - \frac{\pi f_c}{\tau} \mathbb{I}_{\sin(2\pi \gamma_0 \tau) \geq 0} \left( \pi^2 \tau^2 / 2f_{\text{min}}^2 \right) ,
\]

\[
B_2^U (\gamma_0, \tau) := \frac{2f_c^2 (B_0^L (\gamma_0, \tau) - \cos (2\pi \gamma_0 \tau))}{\tau^2} \left( 1 - \mathbb{I}_{\cos(2\pi \gamma_0 \tau) \geq B_0^L (\gamma_0, \tau)} \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) \]

\[\quad - 4\pi^2 \gamma_0^2 f_c^2 B_0^L (\gamma_0, \tau) + \mathbb{I}_{h_2^L (\gamma_0, \tau) \geq 0} f_c^2 h_2^U (\gamma_0, \tau) ,
\]

\[
B_2^L (\gamma_0, \tau) := \frac{2f_c^2 (B_0^U (\gamma_0, \tau) - \cos (2\pi \gamma_0 \tau))}{\tau^2} \left( 1 - \mathbb{I}_{\cos(2\pi \gamma_0 \tau) \leq B_0^U (\gamma_0, \tau)} \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) \]

\[\quad - 4\pi^2 \gamma_0^2 f_c^2 B_0^U (\gamma_0, \tau) + \mathbb{I}_{h_2^U (\gamma_0, \tau) < 0} f_c^2 h_2^L (\gamma_0, \tau) ,
\]

where

\[
h_2^U (\gamma_0, \tau) = \frac{4\pi^2 \gamma_0}{f_{\text{min}}} \left( \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) \mathbb{I}_{\sin(2\pi \gamma_0 \tau) < 0} \left( \frac{\sin (2\pi \gamma_0 \tau)}{2\pi \gamma_0 \tau} - B_0^L (\gamma_0, \tau) \right) ,
\]

\[
h_2^L (\gamma_0, \tau) = \frac{4\pi^2 \gamma_0}{f_{\text{min}}} \left( \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) \mathbb{I}_{\sin(2\pi \gamma_0 \tau) \geq 0} \left( \frac{\sin (2\pi \gamma_0 \tau)}{2\pi \gamma_0 \tau} - B_0^U (\gamma_0, \tau) \right) ,
\]

\[
B_3^U (\gamma_0, \tau) := -\frac{4\pi^2 \gamma_0^2 f_c^2 B_1^L (\gamma_0, \tau)}{\tau^2} \left( 1 + \mathbb{I}_{h_1^L (\gamma_0, \tau) < 0} \left( \frac{1}{\gamma_0 f_{\text{min}}} - \frac{1}{2\gamma_0^2 f_{\text{min}}^2} \right) \right)
\]

\[\quad + \frac{2f_c h_1^U (\gamma_0, \tau)}{\tau} \left( 1 - \mathbb{I}_{h_1^L (\gamma_0, \tau) < 0} \frac{\pi^2 \tau^2}{2f_{\text{min}}^2} \right) ,
\]
\[
B^I_3 (\gamma_0, \tau) := -4\pi^2 \gamma_0^2 f_c^2 B^U_1 (\gamma_0, \tau) \left( 1 + \mathbb{1}_{B^U_1 (\gamma_0, \tau) > 0} \left( \frac{1}{\gamma_0 f_{\min}} - \frac{1}{2\gamma_0^2 f_{\min}^2} \right) \right) \\
+ \frac{2f_c h^I_3 (\gamma_0, \tau)}{\tau} \left( 1 - \mathbb{1}_{h^I_3 (\gamma_0, \tau) > 0} \frac{\pi^2 \tau^2}{2f_{\min}^2} \right),
\]

(B.15)

where the definition of \(h^U_3 (\gamma_0, \tau)\) and \(h^I_3 (\gamma_0, \tau)\) is separated into two intervals:

\[
h^U_3 (\gamma_0, \tau) := \begin{cases} 
    h^U_{\text{near,3,a}} (\gamma_0, \tau) \mathbb{1}_{h^U_{\text{near,3,a}}(\gamma_0, \tau) > 0} - h^U_{\text{near,3,b}} (\gamma_0, \tau) & \text{if } 0 \leq |\tau| \leq \tau_3 (\gamma_0), \\
    h^U_{\text{far,3}} (\gamma_0, \tau) & \text{if } |\tau| > \tau_3 (\gamma_0),
\end{cases}
\]

\[
h^L_3 (\gamma_0, \tau) := \begin{cases} 
    h^L_{\text{near,3,a}} (\gamma_0, \tau) \mathbb{1}_{h^L_{\text{near,3,a}}(\gamma_0, \tau) < 0} - h^L_{\text{near,3,b}} (\gamma_0, \tau) & \text{if } 0 \leq |\tau| \leq \tau_3 (\gamma_0), \\
    h^L_{\text{far,3}} (\gamma_0, \tau) & \text{if } |\tau| > \tau_3 (\gamma_0).
\end{cases}
\]

In particular, \(\tau_3 (0.247) = 0.4614, \tau_3 (0.339) = 0.2346\) and \(\tau_3 (0.414) = 0.3051\). Finally, we have

\[
h^U_{\text{near,3,a}} (\gamma_0, \tau) := \frac{4\pi^2 \gamma_0 f_c^2}{f_{\min}} \left( B^U (\gamma_0, \tau) - \frac{\sin (2\pi \gamma_0 \tau)}{2\pi \gamma_0 \tau} \left( 1 - \mathbb{1}_{\sin(2\pi \gamma_0 \tau) > 0} \frac{\pi^2 \tau^2}{2f_{\min}^2} \right) \right),
\]

\[
h^L_{\text{near,3,a}} (\gamma_0, \tau) := \frac{4\pi^2 \gamma_0 f_c^2}{f_{\min}} \left( B^L (\gamma_0, \tau) - \frac{\sin (2\pi \gamma_0 \tau)}{2\pi \gamma_0 \tau} \left( 1 - \mathbb{1}_{\sin(2\pi \gamma_0 \tau) < 0} \frac{\pi^2 \tau^2}{2f_{\min}^2} \right) \right),
\]

\[
h^U_{\text{near,3,b}} (\gamma_0, \tau) := \frac{f_c^2 (31.3\pi^4 \gamma_0^4 \tau^2 - 16\pi^6 \gamma_0^6 \tau^4 (1 + 2/\gamma_0 f_{\min}) (1 - \pi^2 \tau^2/2f_{\min}^2)^2}{15 (2 + 1/\gamma_0 f_{\min})}.
\]

\[
h^L_{\text{near,3,b}} (\gamma_0, \tau) := \frac{2\pi^4 \gamma_0^2 f_c^2 \tau^2 (8.17 + 20/\gamma_0 f_{\min})}{15 (1 - \pi^2 \tau^2/6f_{\min}^2)^2},
\]

\[
h^U_{\text{far,3}} (\gamma_0, \tau) := \frac{f_c B^U_1 (\gamma_0, \tau)}{\tau} \left( 1 - \mathbb{1}_{B^U_1 (\gamma_0, \tau) > 0} \frac{\pi^2 \tau^2}{2f_{\min}^2} \right) - B^U_2 (\gamma_0, \tau),
\]

\[
h^L_{\text{far,3}} (\gamma_0, \tau) := \frac{f_c B^L_1 (\gamma_0, \tau)}{\tau} \left( 1 - \mathbb{1}_{B^L_1 (\gamma_0, \tau) > 0} \frac{\pi^2 \tau^2}{2f_{\min}^2} \right) - B^L_2 (\gamma_0, \tau).
\]

(B.16)

B.2 Bounds on the tail of the kernel in Lemma 4.6

The bounds are plotted for \(\gamma = [0.247, 0.339, 0.414]^{\top}\) in Fig. B2.

\[
b_{\gamma,0} (\tau) := \prod_{j=1}^{p} b_0 (\gamma_j, \tau),
\]

\[
b_{\gamma,1} (\tau) := \sum_{i=1}^{p} b_1 (\gamma_i, \tau) \prod_{j=1, j \neq i}^{p} b_0 (\gamma_j, \tau),
\]
Fig. B2. Non-increasing upper bounds on the absolute value of $K_\gamma$ and its derivatives. Best viewed on a computer screen.

$$b_{\gamma,2}(\tau) := \sum_{i=1}^{p} b_2(\gamma_i, \tau) \prod_{j=1,j\neq i}^{p} b_0(\gamma_j, \tau) + \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} b_1(\gamma_i, \tau) b_1(\gamma_j, \tau) \prod_{k=1,k\neq[i,j]}^{p} b_0(\gamma_k, \tau),$$

$$b_{\gamma,3}(\tau) := \sum_{i=1}^{p} b_3(\gamma_i, \tau) \prod_{j=1,j\neq i}^{p} b_0(\gamma_j, \tau) + 3 \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} b_2(\gamma_i, \tau) b_1(\gamma_j, \tau) b_1(\gamma_k, \tau) \prod_{k=1,k\neq[i,j]}^{p} b_0(\gamma_k, \tau)$$

$$+ \sum_{i=1}^{p} \sum_{j=1,j\neq i}^{p} \sum_{k=1,k\neq[i,j]}^{p} b_1(\gamma_i, \tau) b_1(\gamma_j, \tau) b_1(\gamma_k, \tau) \prod_{l=1,l\neq[i,j]}^{p} b_0(\gamma_l, \tau),$$

(B.17)
where
\[
b_0(\gamma_0, \tau) := \frac{1}{2\pi \gamma_0 \tau \left(1 - \pi^2 \tau^2 / 6f_{\text{min}}^2\right)},
\]
\[
b_1(\gamma_0, \tau) := \frac{f_c (1 + b_0(\gamma_0, \tau))}{\tau \left(1 - \pi^2 \tau^2 / 6f_{\text{min}}^2\right)},
\]
\[
b_2(\gamma_0, \tau) := 4\pi^2 \gamma_0 f_c^2 b_0(\gamma_0, \tau) \left(1 + \frac{1}{\gamma_0 f_{\text{min}}}\right) + \frac{2f_c b_1(\gamma_0, \tau)}{\tau},
\]
\[
b_3(\gamma_0, \tau) := 4\pi^2 \gamma_0 f_c^2 b_1(\gamma_0, \tau) \left(1 + \frac{1}{\gamma_0 f_{\text{min}}}\right) + \frac{2f_c}{\tau} \left(b_2(\gamma_0, \tau) + \frac{f_c b_1(\gamma_0, \tau)}{\tau}\right).
\]

### Appendix C. Proof of the bounds on the interpolation kernel

#### C.1 Proof of Lemma 4.4

We assume that \(t\) is between 0 and 0.5. The bounds for \(t\) between \(-0.5\) and 0 follow by symmetry (\(K^2/\) and \(K^{(2)}\) are even, \(K^{(1)}\) and \(K^{(3)}\) are odd). The following local bounds, which are derived from Taylor series expansions around the origin, will be used often during the proof. If \(u \geq 0\)

\[
u - \frac{u^3}{6} \leq \sin(u) \leq u, \tag{C.1}
\]
\[
u - \frac{u^3}{6} + \frac{u^5}{120} - \frac{u^7}{5040} \leq \sin(u) \leq u - \frac{u^3}{6} + \frac{u^5}{120} - \frac{u^7}{5040} + \frac{u^9}{362880}, \tag{C.2}
\]
\[
u - \frac{u^2}{2} \leq \cos(u) \leq 1, \tag{C.3}
\]
\[
u - \frac{u^2}{2} + \frac{u^4}{24} - \frac{u^6}{720} \leq \cos(u) \leq 1 - \frac{u^2}{2} + \frac{u^4}{24}, \tag{C.4}
\]

and if \(0 \leq u \leq \sqrt{2}\)

\[
u \leq \tan(u) \leq \frac{u}{1 - \frac{u^2}{2}}, \tag{C.5}
\]
\[
u^2 \leq \sin(u) \tan(u) \leq \nu^2 + \frac{u^4}{2 - u^2}, \tag{C.6}
\]
\[
u^3 \leq \sin(u) \tan^2(u) \leq \nu^3 + \frac{u^5 (4 - u^2)}{(2 - u^2)^2}. \tag{C.7}
\]

Let us fix a value \(0 < \gamma_0 < 1\) and set \(f_0 = \gamma_0 f_c\) to alleviate notation. Henceforth, we assume that \(f_c \geq f_{\text{min}}\). By (C.1) and (C.5)

\[
\frac{1}{\pi t} \leq \frac{1}{\sin(\pi t)} \leq \pi t \left(1 - \pi^2 f_0^2 t^2 / 6f_{\text{min}}^2\right), \tag{C.8}
\]
\[
\frac{1 - \pi^2 f_0^2 t^2 / 2f_{\text{min}}^2}{\pi t} \leq \frac{1}{\tan(\pi t)} \leq \frac{1}{\pi t}. \tag{C.9}
\]
We begin by deriving bounds on $K(\gamma_0 f_c, t)$ near the origin. By (C.2)

$$\frac{\sin((2f_0 + 1)\pi t)}{2f_0 + 1} \frac{2f_0}{\sin(2\pi f_0 t)} \leq \frac{1 - \pi^2 (2f_0 + 1)^2 t^2 / 6 + \pi^4 (2f_0 + 1)^4 t^4 / 120 - \pi^6 (2f_0 + 1)^6 t^6 / 5040 + \pi^8 (2f_0 + 1)^8 t^8 / 362880}{1 - 2\pi^2 f_0^2 t^2 / 3 + 2\pi^4 f_0^4 t^4 / 15 - 4\pi^6 f_0^6 t^6 / 315},$$

$$\frac{\sin((2f_0 + 1)\pi t)}{2f_0 + 1} \frac{2f_0}{\sin(2\pi f_0 t)} \geq \frac{1 - \pi^2 (2f_0 + 1)^2 t^2 / 6 + \pi^4 (2f_0 + 1)^4 t^4 / 120 - \pi^6 (2f_0 + 1)^6 t^6 / 5040}{1 - 2\pi^2 f_0^2 t^2 / 3 + 2\pi^4 f_0^4 t^4 / 15 - 4\pi^6 f_0^6 t^6 / 315 + 2\pi^8 f_0^8 t^8 / 2835}.$$

Combining this with (C.8) establishes

$$B^L_{\text{near},0}(\gamma_0 f_c, t) \leq K(\gamma_0 f_c, t) \leq B^U_{\text{near},0}(\gamma_0 f_c, t). \quad (C.10)$$

To prove

$$B^L_{\text{far},0}(\gamma_0 f_c, t) \leq K(\gamma_0 f_c, t) \leq B^U_{\text{far},0}(\gamma_0 f_c, t), \quad (C.11)$$

we express the kernel in the following form

$$K(f_0, t) = \frac{\sin(2\pi f_0 t) \cos(\pi t)}{(2f_0 + 1) \sin(\pi t)} + \frac{\cos(2\pi f_0 t)}{(2f_0 + 1)},$$

and apply (C.3) and (C.9).

Some basic trigonometric identities yield

$$K^{(1)}(f_0, t) = \frac{\pi}{\sin(\pi t)} \left( \cos((2f_0 + 1)\pi t) - \frac{\sin((2f_0 + 1)\pi t) \cos(\pi t)}{(2f_0 + 1) \sin(\pi t)} \right)$$

$$= \frac{\pi}{\sin(\pi t)} \left( \cos((2f_0 + 1)\pi t) - K(f_0, t) \cos(\pi t) \right) \quad (C.12)$$

$$= \frac{\pi}{\sin(\pi t)} \left( \cos(2\pi f_0 t) \cos(\pi t) - \sin(2\pi f_0 t) \sin(\pi t) - K(f_0, t) \cos(\pi t) \right) \quad (C.13)$$

$$= \pi \left( \frac{\cos(2\pi f_0 t) - K(f_0, t)}{\tan(\pi t)} - \sin(2\pi f_0 t) \right).$$

This implies

$$K^{(1)}(f_0, t) \leq \pi \left( \frac{\cos(2\pi f_0 t) - B^U_{\text{far},0}(\gamma_0 f_0 t)}{\tan(\pi t)} - \sin(2\pi f_0 t) \right),$$

$$K^{(1)}(f_0, t) \geq \pi \left( \frac{\cos(2\pi f_0 t) - B^U_{\text{far},0}(\gamma_0 f_0 t)}{\tan(\pi t)} - \sin(2\pi f_0 t) \right).$$

Applying (C.9) then establishes

$$B^U_1(\gamma_0 f_c, t) \leq K^{(1)}(\gamma_0 f_c, t) \leq B^U_1(\gamma_0 f_c, t). \quad (C.14)$$
To bound the second derivative of the kernel we again leverage some trigonometric identities,

\[
K^{(2)}(f_0, t) = -\pi^2 K(f_0, t) \left( (2f_0 + 1)^2 - 1 \right) - \frac{2\pi K^{(1)}(f_0, t)}{\tan(\pi t)}
\]

\[
= -4\pi^2 K(f_0, t) \left( f_0^2 + f_0 \right) - \frac{2\pi^2}{\tan(\pi t)} \left( \cos(2\pi f_0 t) - K(f_0, t) \right) - \sin(2\pi f_0 t)
\]

\[
= -4\pi^2 f_0^2 K(f_0, t) - \frac{2\pi^2 (\cos(2\pi f_0 t) - K(f_0, t))}{\tan^2(\pi t)} + 4\pi^2 f_0 \left( \frac{\sin(2\pi f_0 t)}{2f_0 \tan(\pi t)} - K(f_0, t) \right).
\]  
(C.15)

This implies

\[
K^{(2)}(f_0, t) \leq -4\pi^2 f_0^2 B_0^L(\gamma_0, f_0 t) + \frac{2\pi^2 (B_0^U(\gamma_0, f_0 t) - \cos(2\pi f_0 t))}{\tan^2(\pi t)} + 4\pi^2 f_0 \left( \frac{\sin(2\pi f_0 t)}{2f_0 \tan(\pi t)} - B_0^L(\gamma_0, f_0 t) \right).
\]

\[
K^{(2)}(f_0, t) \geq -4\pi^2 f_0^2 B_0^U(\gamma_0, f_0 t) + \frac{2\pi^2 (B_0^U(\gamma_0, f_0 t) - \cos(2\pi f_0 t))}{\tan^2(\pi t)} + 4\pi^2 f_0 \left( \frac{\sin(2\pi f_0 t)}{2f_0 \tan(\pi t)} - B_0^U(\gamma_0, f_0 t) \right).
\]

Applying (C.9) yields

\[
B_2^L(\gamma_0, f_0 t) \leq K^{(2)}(\gamma_0, f_0 t) \leq B_2^U(\gamma_0, f_0 t).
\]  
(C.16)

Using the fact that the derivative of $1/\tan(u)$ is $-1 - 1/\tan^2(u)$, we rewrite the third derivative of the kernel as

\[
K^{(3)}(f_0, t) = -\pi^2 K^{(1)}(f_0, t) \left( (2f_0 + 1)^2 - 3 \right) - \frac{2\pi}{\tan(\pi t)} \left( K^{(2)}(f_0, t) - \pi K^{(1)}(f_0, t) \right) 
\]

\[
= -2\pi^2 K^{(1)}(f_0, t) \left( 2f_0^2 + f_0 - 1 \right) - \frac{2\pi^2}{\tan(\pi t)} \left( -4\pi K(f_0, t) \left( f_0^2 + f_0 \right) - 3K^{(1)}(f_0, t) \right) 
\]

\[
= -2\pi^2 K^{(1)}(f_0, t) \left( 2f_0^2 + f_0 - 1 \right) - \frac{2\pi^2}{\tan(\pi t)} \left( 4f_0 \left( \frac{\sin(2\pi f_0 t)}{2f_0 \tan(\pi t)} - K(f_0, t) \right) + z(t) \right).
\]  
(C.17)

To bound $z(t)$ we perform some algebraic manipulations and then apply (C.2), (C.4), (C.6) and (C.7),

\[
z(t) = -4f_0^2 K(f_0, t) + \frac{\sin(2\pi f_0 t)}{\tan(\pi t)} - \frac{3(\cos(2\pi f_0 t) - K(f_0, t))}{\tan^2(\pi t)} 
\]

\[
= \left( -4f_0^2 + 2f_0 + 1 \right) \sin(2\pi f_0 t) - \frac{4f_0^2 \cos(2\pi f_0 t)}{(2f_0 + 1) \tan(\pi t)} - \frac{3((2f_0 + 1) \cos(2\pi f_0 t) \sin(\pi t) - \sin((2f_0 + 1) \pi t))}{(2f_0 + 1) \sin(\pi t) \tan^2(\pi t)}.
\]
\[
\frac{\sin^2(2\pi f_0 t)}{2f_0 + 1} \sin^2(\pi t) \tan(\pi t) + 3 \cos(\pi t)
\]

\[
+ \frac{\cos^2(2\pi f_0 t)(-4f_0^2 \sin(\pi t) \tan^2(\pi t) - 6f_0 \sin(\pi t))}{(2f_0 + 1) \sin(\pi t) \tan^2(\pi t)}
\]

\[
= \frac{(2\pi f_0 t - 4\pi^3 f_0^3 t^3/3 + 4\pi^5 f_0^5 t^5/15 - \epsilon_1) \((-4f_0^2 + 2f_0 + 1)(\pi^2 t^2 + \epsilon_2) + 3(1 - \pi^2 t^2/2 + \epsilon_3))}{(2f_0 + 1) \sin(\pi t) \tan^2(\pi t)}
\]

\[
+ \frac{(1 - 2\pi^2 f_0^2 t^2 + 2\pi^4 f_0^4 t^4/3 - \epsilon_4)(-4f_0^2(\pi^3 t^3 + \epsilon_5) - 6f_0(\pi t - \pi^3 t^5/6 + \epsilon_6))}{(2f_0 + 1) \sin(\pi t) \tan^2(\pi t)}
\]

\[
= \frac{4\pi^5 f_0^3 t^5 (8 + 20/f_0 - 14f_0^2 + 8f_0^2 + z_0(f_0, t, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6)) - 8\pi^7 f_0^7 t^7}{15(2f_0 + 1) \sin(\pi t) \tan^2(\pi t)},
\]

where

\[
|\epsilon_1| \leq \frac{8\pi^7 f_0^7 t^7}{315}, \quad |\epsilon_2| \leq \pi^4 t^4, \quad |\epsilon_3| \leq \frac{\pi^4 t^4}{24},
\]

\[
|\epsilon_4| \leq \frac{4\pi^6 f_0^6 t^6}{45}, \quad |\epsilon_5| \leq 4\pi^5 t^5, \quad |\epsilon_6| \leq \frac{\pi^5 t^5}{120}.
\]

If \(f_0 t \leq \frac{1}{2}\) and \(f_{\min} \geq 10^3\) some straightforward, yet tedious, computations yield

\[
|z_0(f_0, t, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6)| \leq 0.17,
\]

which together with (C.8) and (C.9) implies\(^5\)

\[
\frac{31.3\pi^2 f_0^4 t^2 - 16\pi^4 f_0^6 t^4(1 + 2/\gamma_0 f_{\min})}{15(2 + 1/\gamma_0 f_{\min})} \leq z(t) \leq \frac{2\pi^2 f_0^4 t^2 (8.17 + 20/\gamma_0 f_{\min})}{15(1 - \pi^2 f_0^2 t^2/2 + 6\gamma_0^2 f_{\min}^2)}.
\]

This establishes

\[
B_{\gamma_j}^{(3)}(\gamma_0, f_0, t) \leq K^{(3)}(\gamma_0 f_0, t) \leq B_{\gamma_j}^{(3)}(\gamma_0, f_0, t)
\]

for \(0 \leq f_0 t \leq \tau_3(\gamma_0)\). For \(f_0 t \geq \tau_3(\gamma_0)\) the bounds are obtained by combining (C.17) with (C.9).

The bounds (C.10), (C.11), (C.14), (C.16) and (C.18) combined with the identities:

\[
K^{(1)}_y(t) := \sum_{i=1}^{p} K^{(1)}(\gamma_0 f_c, t) \prod_{j \neq i} K(\gamma_j f_c, t),
\]

\[
K^{(2)}_y(t) := \sum_{i=1}^{p} K^{(2)}(\gamma_0 f_c, t) \prod_{j \neq i} K(\gamma_j f_c, t) + \sum_{i=1}^{p} K^{(1)}(\gamma_0 f_c, t) \sum_{j \neq i} K^{(1)}(\gamma_j f_c, t) \prod_{k \neq \{i,j\}} K(\gamma_k f_c, t),
\]

\(^5\) In this step, we assume that \(\gamma_0 \geq 1/32c_0\), which holds for the values of \(\gamma_0\) we are interested in. Otherwise, we would have to add a term \(-5/8f_0^2\) in the first factor of the numerator of the lower bound. For the upper bound, we assume \(\gamma_0 \geq 1/2c_0^2\). Similarly, this holds for the values of \(\gamma_0\) that we consider, but we could add an extra term to the numerator of the right-hand side if this were not the case.
\[ K_{\gamma}^{(3)}(t) := \sum_{i=1}^{p} K_{(3)}(\gamma f_c, t) \prod_{j \neq i} K_{(1)}(\gamma f_c, t) \sum_{j=1}^{p} K_{(2)}(\gamma f_c, t) \prod_{k \neq (i,j)} K_{(1)}(\gamma f_c, t) \]

\[ + \sum_{i=1}^{p} K_{(1)}(\gamma f_c, t) \sum_{j \neq i} K_{(1)}(\gamma f_c, t) \prod_{k \neq (i,j)} K_{(1)}(\gamma f_c, t) \]  

(C.19)

yield

\[ B_{\gamma,\ell}^L(f_c t) \leq K_{\gamma}^{(\ell)}(t) \leq B_{\gamma,\ell}^U(f_c t). \]  

(C.20)

Finally, we control the deviation of the kernel and its derivatives from their values on a grid. Bernstein’s inequality allows us to bound the \(\ell\)th derivative of a trigonometric polynomial.

**Theorem C.1 (Theorem 4 [62])**  Let \( P \) be a trigonometric polynomial of order \( N \)

\[ P(\theta) = \sum_{j=1}^{N} a_j \cos(j\theta) + b_j \sin(j\theta), \]  

(C.21)

such that \(|P(\theta)| \leq 1\) for all real \(\theta\). Then,

\[ |P^{(\ell)}(\theta)| \leq N^\ell \quad \text{for all real } \theta. \]

Applying the theorem to \( K_{\gamma}(t) \), which can be written as \( P(2\pi t) \) for \( P \) of the form (C.21) for \( N = f_c \) and has magnitude bounded by 1, we obtain

\[ |K_{\gamma}^{(\ell)}(t)| \leq (2\pi)^\ell f_c^\ell. \]

As a result, for any \( t \) such that \(|t - t^*| \leq \epsilon / f_c \)

\[ |K_{\gamma}^{(\ell)}(t) - K_{\gamma}^{(\ell)}(t^*)| \leq (2\pi)^{\ell+1} f_c^\ell \epsilon \]

by the mean-value theorem. Combining this with (C.20) completes the proof.

### C.2 Proof of Lemma 4.6

By (C.8)

\[ |K(f_0, t)| = \left| \frac{\sin((2f_0 + 1)\pi t)}{(2f_0 + 1)\sin(\pi t)} \right| \leq \frac{1}{2\pi f_0 t \left( 1 - \pi^2 f_0^2 t^2 / 6y_0^2 f_{\min}^2 \right)}, \]  

(C.22)
As long as \( t < \sqrt{2f_{\min}/\pi f_c} \) this bound is decreasing because the derivative of the denominator is positive. By (C.12) and (C.8), we have

\[
|K^{(1)}(f_0, t)| \leq \frac{\pi}{2} \frac{|\cos((2f_0 + 1)\pi t)| + |Kf_0, t| \sin(\pi t)|}{|\sin(\pi t)|} \\
\leq \frac{1 + b_0(\gamma_0, f_c t)}{t (1 - \pi^2 f_0^2 t^2/6\gamma_0^2 f_{\min}^2)}.
\]

(C.23)

The bound is again decreasing if \( t < \sqrt{2f_{\min}/\pi f_c} \) because the derivative of the denominator is positive and \( b_0 \) is decreasing. By (C.15) and (C.9),

\[
|K^{(2)}(f_0, t)| \leq 4\pi^2 f_0^2 b_0(\gamma_0, f_c t) \left(1 + \frac{1}{\gamma_0 f_{\min}}\right) + \frac{2b_1(\gamma_0, f_c t)}{t}.
\]

(C.24)

If \( t < \sqrt{2f_{\min}/\pi f_c} \), the bound is decreasing because so are \( b_0 \) and \( b_1 \). The same holds for the next bound, which results from combining (C.17) and (C.9),

\[
|K^{(3)}(f_0, t)| \leq 4\pi^2 f_0^2 b_1(\gamma_0, f_c t) \left(1 + \frac{1}{\gamma_0 f_{\min}}\right) + \frac{2b_1(\gamma_0, f_c t)}{t}.
\]

(C.25)

These inequalities together with (C.19) establish the bounds in the lemma, which are decreasing in \( \tau \) if \( \tau \leq 450 < \sqrt{2f_{\min}/\pi} \) (which implies \( t < \sqrt{2f_{\min}/\pi f_c} \) if \( f_{\min} \geq 10^3 \)) because they are products of positive decreasing functions.

C.3 Proof of Lemma 4.7

We bound the first term of the decomposition

\[
\sum_{t_j \in T \setminus \{0\}} |K^{(1)}(f_0, t_j)| = \sum_{t_j \in T \cap ((0, \frac{1}{2}))} |K^{(1)}(f_0, t_j)| + \sum_{t_j \in T \cap [-1/2, 0)} |K^{(1)}(f_0, t_j)|;
\]

the argument for the second term is almost identical.

Without loss of generality, we assume that the elements of \( T \cap \left(0, \frac{1}{2}\right) \) are ordered so that \( 0 < t_1 < t_2 < \cdots < \frac{1}{2} \). By Corollary 4.5,

\[
|K^{(1)}(f_0, t_j)| \leq B^\infty_{y,c}(u, \epsilon)
\]

for any \( u \) such that \( f_c(t_j - t) - \epsilon \leq u \leq f_c(t_j - t) + \epsilon \). Maximizing this quantity on an equispaced grid \( G_{j,t} \) with separation \( \epsilon \) that covers the interval \( (\tau_{\min} - \tau, (j + 4) \tau_{\min}] \) for \( \tau - \epsilon \leq f_c t \leq \tau \) yields a bound that is valid for all values of \( t_j \) such that \( t_j - t \leq (j + 4) \Delta_{\min} \), since \( t_j - t \geq (\Delta_{\min} - t) \) due to the
minimum-separation condition. If \( t_j - t > (j + 4) \Delta_{\text{min}} \), then by Lemma 4.6

\[
|K_{\gamma}^{(t)}(t - t_j)| \leq \delta b_{\gamma, \ell}(j + 4 \tau_{\text{min}}).
\]

Taking the maximum of the two bounds allows us to control the contribution of the kernel (or the \( \ell \)th derivative of the kernel) centered at \( t_j \). We take advantage of this for \( 1 \leq j \leq 20 \) and then apply the coarser bound

\[
|K_{\gamma}^{(t)}(t - t_j)| \leq b_{\gamma, \ell}(j - 1/2 \tau_{\text{min}})
\]

for \( 21 \leq j \leq 400 \). This bound follows from Lemma 4.6 and the fact that \( t_j - t \geq (j - 1/2) \Delta_{\text{min}} \) by the minimum-separation condition. The remaining terms can be controlled using the following lemma, which we prove in Section C.4.

**Lemma C.2** For \( f_c \geq 10^3 \), \( \ell \in \{0, 1, 2, 3\} \), \( \gamma = [0.247, 0.339, 0.414]^T \) and \( 80/f_c \leq |t| \leq 1/2 \)

\[
|K_{\gamma}^{(t)}(t)| \leq b_{\gamma, \ell}^{\infty}(t) := \left( \frac{2\pi f_c}{\ell} \right)^{\ell} \left( \frac{1.1}{4f_c t} \right)^p.
\]

By the lemma,

\[
\sum_{t \in T \cap (400 \Delta_{\text{min}}, 1/2)} |K_{\gamma}^{(t)}(t - t_j)| \leq \sum_{k=400}^{\infty} b_{\gamma, \ell}^{\infty}(k \tau_{\text{min}})
\]

\[
= \sum_{k=1}^{\infty} b_{\gamma, \ell}^{\infty}(k \tau_{\text{min}}) - \sum_{k=1}^{400} b_{\gamma, \ell}^{\infty}(k \tau_{\text{min}})
\]

\[
= \left( \frac{2\pi f_c}{\ell} \right)^{\ell} \left( \frac{1.1}{4\tau_{\text{min}}} \right)^p \left( \sum_{k=1}^{\infty} \frac{1}{k^p} - \sum_{k=1}^{400} \frac{1}{k^p} \right) := C_{\ell}.
\]

The infinite series \( \sum_{k=1}^{\infty} k^{-p} \) is equal to the Riemann zeta function \( \zeta(p) \), so we can easily compute \( C_{\ell} \) for the desired value of \( \gamma \) and conclude that

\[
\sum_{t \in T \cap (0, 1/2)} |K_{\gamma}^{(t)}(t - t_j)| \leq H(t).
\]

**C.4 Proof of Lemma C.2**

To obtain bounds that are valid for large \( t \), we use a simple lower bound on the sine function that holds between 0 and \( \pi/2 \)

\[
2t \leq \sin(\pi t).
\]
Combining it with (C.8) and the bounds (C.22–C.25) on the magnitude of the kernel and its derivatives from Section C.2 yields

\[
|K(f_0, t)| \leq \frac{1}{(2f_0 + 1) 2t} \leq \frac{1}{4f_0 t},
\]

\[
|K^{(1)}(f_0, t)| \leq \frac{\pi}{2t} \left( 1 + \frac{1}{4f_0 t} \right),
\]

\[
|K^{(2)}(f_0, t)| \leq \frac{\pi^2 f_0^2}{t} \left( 1 + \frac{1}{\gamma_0 f_{\text{min}}} \right) + \frac{\pi}{t^2} \left( 1 + \frac{1}{4f_0 t} \right) \leq \frac{\pi^2 f_0^2}{t} \left( 1 + \frac{1 + 1/4f_0 t}{\gamma_0 f_{\text{min}}} + \frac{1 + 1/4f_0 t}{\pi f_0 t} \right),
\]

\[
|K^{(3)}(f_0, t)| \leq \frac{2\pi^3 f_0^2}{t} \left( \left( 1 + \frac{1 + 1/4f_0 t}{\gamma_0 f_{\text{min}}} \right) + \frac{1}{\pi f_0 t} \left( 1 + \frac{1 + 1/4f_0 t}{\gamma_0 f_{\text{min}}} + \frac{1 + 1/4f_0 t}{\pi f_0 t} \right) \right)
\]

\[
+ \frac{1}{2\pi^2 f_0^2 t^2} \left( 1 + \frac{1}{4f_0 t} \right).
\]

If \( \gamma_0 \geq 0.25 \), then \( 1/f_0 t \leq 0.05 \) for \( t \geq 80/f_c \), which allows us to coarsely summarize these inequalities as

\[
|K^{(\ell)}(f_0, t)| \leq \frac{1.12 \ell^{-2} \pi^{\ell} f_0^{\ell-1}}{t}
\]

for \( j \in 0, 1, 2, 3 \) and \( f_{\text{min}} \geq 10^3 \). Together with (C.19) this implies

\[
|K_\gamma(t)| \leq \left( \frac{1.1}{4f_c t} \right)^p \prod_{i=1}^p \frac{1}{\gamma_i},
\]

\[
|K_\gamma^{(1)}(t)| \leq 2\pi f_c \left( \frac{1.1}{4f_c t} \right)^p \sum_{i=1}^p \prod_{j=1,j \neq i}^p \frac{1}{\gamma_j},
\]

\[
|K_\gamma^{(2)}(t)| \leq (2\pi f_c)^2 \left( \frac{1.1}{4f_c t} \right)^p \sum_{i=1}^p \left( \frac{\gamma_i}{\prod_{j=1,j \neq i}^p \gamma_j} + \sum_{j=1,j \neq i}^p \prod_{k=1,k \neq j}^p \frac{1}{\gamma_k} \right),
\]

\[
|K_\gamma^{(3)}(t)| \leq (2\pi f_c)^3 \left( \frac{1.1}{4f_c t} \right)^p \sum_{i=1}^p \left( \frac{\gamma_i^2}{\prod_{j=1,j \neq i}^p \gamma_j} + \sum_{j=1,j \neq i}^p \prod_{k=1,k \neq j}^p \frac{3\gamma_i}{\gamma_k} + \sum_{k=1,k \neq i}^p \prod_{m=1,m \neq i,k}^p \frac{1}{\gamma_m} \right).
\]

Since by definition \( \sum_{j=1}^p \gamma_j = 1 \), the following identities complete the proof,

\[
\sum_{i=1}^p \prod_{j=1,j \neq i}^p \frac{1}{\gamma_j} = \sum_{i=1}^p \frac{\gamma_i}{\prod_{j=1}^p \gamma_j} = \frac{1}{\prod_{j=1}^p \gamma_j}.
\]
\[
\sum_{i=1}^{p} \left( \frac{\gamma_i}{\prod_{j=1,j \neq i}^{p} \gamma_j} + \sum_{j=1,j \neq i}^{p} \prod_{k=1,k \neq i,j}^{p} \frac{1}{\gamma_k} \right) = \frac{1}{\prod_{j=1}^{p} \gamma_j} \sum_{i=1}^{p} \left( \gamma_i^2 + \gamma_i \sum_{j=1,j \neq i}^{p} \gamma_j \right)
\]

\[
= \frac{1}{\prod_{j=1}^{p} \gamma_j} \sum_{i=1}^{p} \gamma_i^2 + \gamma_i \left( 1 - \gamma_i \right) = \frac{1}{\prod_{j=1}^{p} \gamma_j},
\]

\[
\sum_{i=1}^{p} \left( \frac{\gamma_i^2}{\prod_{j=1,j \neq i}^{p} \gamma_j} + \sum_{j=1,j \neq i}^{p} \left( \frac{3\gamma_i}{\prod_{k=1,k \neq i,j}^{p} \gamma_k} + \sum_{k=1,k \neq i,j}^{p} \prod_{m=1,m \neq i,k}^{p} \frac{1}{\gamma_m} \right) \right)
\]

\[
= \sum_{i=1}^{p} \left( \frac{\gamma_i^2}{\prod_{j=1,j \neq i}^{p} \gamma_j} + \sum_{j=1,j \neq i}^{p} \left( 2\gamma_i + \gamma_i \right) \sum_{k=1,k \neq i,j}^{p} \gamma_k \right)
\]

\[
= \frac{1}{\prod_{j=1}^{p} \gamma_j} \sum_{i=1}^{p} \left( \gamma_i^3 + (\gamma_i^2 + \gamma_i) \sum_{j=1,j \neq i}^{p} \gamma_j \right) = \frac{1}{\prod_{j=1}^{p} \gamma_j}.
\]

C.5 Proof of Lemma 4.9

By the fundamental theorem of calculus, if \( t' \) is such that \( j\epsilon \leq t' \leq (j+1)\epsilon \) for some integer \( j \)

\[
F'(t') = F'(0) + \int_{0}^{t'} F''(u) \, du
\]

\[
= \sum_{i=0}^{j-1} \int_{i\epsilon}^{(i+1)\epsilon} F''(u) \, du + \int_{j\epsilon}^{t'} F''(u) \, du.
\]

This implies

\[
F'(t') \leq \epsilon \left( \sum_{i=0}^{j-1} \mathbb{I}_{F_{2,i} > 0} F_{2,i} \right).
\]  \hspace{1cm} (C.27)

Now, recall that \( k\epsilon \leq t \leq (k+1)\epsilon \). By the fundamental theorem of calculus, the assumptions on \( F \) and (C.27),

\[
F(t) = F(0) + \int_{0}^{t} F'(u) \, du
\]

\[
= 1 + \sum_{j=0}^{k-1} \int_{j\epsilon}^{(j+1)\epsilon} F'(u) \, du + \int_{k\epsilon}^{t} F'(u) \, du
\]

\[
= 1 + \epsilon^2 \left( \sum_{j=0}^{k-1} \left( \sum_{i=1}^{j-1} \mathbb{I}_{F_{2,i} > 0} F_{2,i} \right) + \mathbb{I}_{F_{2} > 0} \right).
\]