Stable Super-Resolution of Positive Sources: the Discrete Setup

Veniamin I. Morgenshtern[1] and Emmanuel J. Candès[2]

[1]Dept. of Statistics, Stanford University, CA
[2]Dept. of Mathematics and Dept. of Statistics, Stanford University, CA

December 19, 2014

Abstract

In single-molecule microscopy it is necessary to reconstruct a signal that consists of positive point sources from noisy observations of the spectrum of the signal in the low-frequency band $[-f_c, f_c]$. It is shown that the problem can be solved using convex optimization in a stable fashion. The stability of reconstruction depends on Rayleigh-regularity of the support of the signal, i.e., on how many point sources can occur within an interval of length $1.87/f_c$. The stability estimate is complimented by a converse result: the performance of convex algorithm is nearly optimal. The results are generalized to multi-dimension signals.

1 Introduction

The goal of super-resolution is to recover high-frequency information about an object from low-frequency observations only. Such problems arise in many areas of science and engineering, for example, in mass-spectrometry, radar imaging, and wireless communication. The model problem for the analysis in this paper is super-resolution in optical microscopy. As we will see next, due to diffraction of light, the natural resolution of microscopes is inversely proportional to the wavelength of light used for observation. This makes it fundamentally difficult to study sub-wavelength features of the object.

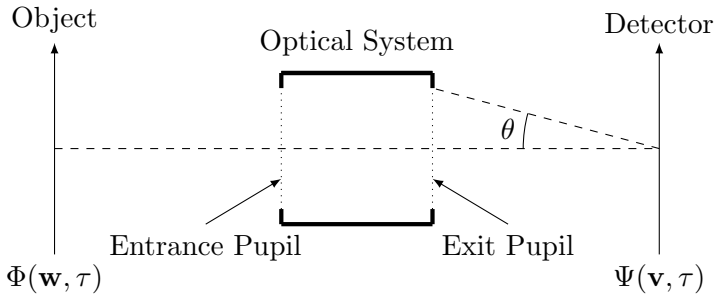

Figure 1: Model of an Optical System

Equations of optics are linear in terms of electric and magnetic fields. Vectorial nature of the fields can be neglected in Fourier optics and the physics can be fully described via the time-varying phasor [15, Sec. 3.2], i.e., one of the 6 components of electric/magnetic fields. Assume

that narrow-band (not necessarily monochromatic) light is used for illumination. Consider an optical microscope depicted in Figure 1. Let $\Phi(\mathbf{w},\tau)$ and $\Psi(\mathbf{v},\tau)$ denote the input/output phasors describing the field emitted by the object being imaged and the field generated at the receiver of the system, respectively. Above, $\mathbf{w},\mathbf{v}\in\mathbb{R}^2$ are indexing spacial coordinates in the object plane and in the detector plane, respectively, and $\tau\in\mathbb{R}$ is indexing time. We assume, for convenience, that the phasors $\Phi(\mathbf{w},\tau),\Psi(\mathbf{v},\tau)$ have been frequency-shifted (as a function of τ) to be centered around the mean frequency of the optical wave [15, p. 132], so that, for example, $E(\mathbf{w},\tau)=\Re[\Phi(\mathbf{w},\tau)e^{2\pi \mathrm{i}\bar{\nu}\tau}]$, where E is one the components of the electric field and $\bar{\nu}$ is the average frequency of emitted light. The diffraction of light in the optical system can be described by the Fraunhofer approximation to the Maxwell equations, leading to the following model [15, Eq (6-6)]

$$\Psi(\mathbf{v},\tau)=\kappa_1\int h(\mathbf{v}-\mathbf{w})\Phi(\mathbf{w},\tau)d\mathbf{w}, \tag{1}$$

where κ_1 is a normalization constant and [15, Sec. 6.2.2]

$$h(\mathbf{v})=\frac{1}{\sqrt{2\bar{f}_c}}\frac{\sin(2\pi\bar{f}_c v_1)}{\pi v_1}\frac{1}{\sqrt{2\bar{f}_c}}\frac{\sin(2\pi\bar{f}_c v_2)}{\pi v_2},\quad \mathbf{v}=[v_1,v_2]^{\mathsf{T}}, \tag{2}$$

is the point-spread function introduced by diffraction. In (2) we assumed that the entrance and the exit pupils in Figure 1 are rectangular. The spacial frequency cut-off of the optical system is given by

$$\bar{f}_c=\frac{\sin(\theta)}{\lambda}, \tag{3}$$

where λ is the wavelength of emitted light (average wavelength in the narrow-band illumination case) and θ is half of the angle spanned by the exit pupil as seen from the center of the image plane (see Figure 1). Note, that due to narrow-band illumination assumption, $h(\cdot)$ only depends on the average wavelength of the optical wave, but does not depend on the specific frequencies in the illuminating spectrum so that the system model is described by the simple convolution equation (1).

In optics, the carrier frequency $\bar{\nu}\sim 500\,\mathrm{THz}$ is much higher than the frequency $f_{\mathrm{HET}}\sim 10\,\mathrm{GHz}$ to which electronic components can respond (for example the frequency of heterodyne, used to down-convert the signal). The consequence of this is the fact that in optics only the time-average of the instantaneous intensity of received light (called received intensity) is directly observable [15, Eq (6-8)]:

$$\tilde{s}(\mathbf{v})\triangleq\langle\Psi(\mathbf{v},\tau)\Psi^*(\mathbf{v},\tau)\rangle \tag{4}$$

where $\Psi^*(\cdot)$ denotes the complex conjugate of $\Psi(\cdot)$ and $\langle\cdot\rangle$ stands for the time averaging:

$$\langle g(\tau)\rangle=f_{\mathrm{HET}}\int_0^{1/f_{\mathrm{HET}}} g(\tau)d\tau. \tag{5}$$

In majority of microscopy applications, the object emits incoherent light. Mathematically, this situation is described by assuming that frequencies of spatially separated emitters vary in statistically independent fashions. This idealized property may be represented by the equation [15, Eq (6-14)]

$$\left\langle\Phi(\mathbf{w},\tau)\Phi^*(\mathbf{w}',\tau)\right\rangle=\kappa_2\delta(\mathbf{w}-\mathbf{w}')x(\mathbf{w}), \tag{6}$$

where κ_2 is a normalization constant. The quantity $x(\mathbf{w})$ is the time-average of the instantaneous intensity of light emitted by the object; it is called emitted intensity. Substituting (1) into (4) and then using (6) we obtain the following input-output relation

$$\tilde{s}(\mathbf{v})=\kappa\int f_{\mathrm{low}}(\mathbf{v}-\mathbf{w})x(\mathbf{w})d\mathbf{w},\quad f_{\mathrm{low}}(\mathbf{v}-\mathbf{w})\triangleq|h(\mathbf{v}-\mathbf{w})|^2 \tag{7}$$

where $\kappa = \kappa_1^2 \kappa_2$ is a normalization constant. Observe that (7) is a linear convolution equation w.r.t. emitted intensity; compare to (1), which is a linear convolution equation w.r.t. components of the emitted field. The low-frequency kernel $f_{\text{low}}(\mathbf{v})$ is the square of the two-dimensional (2D) sinc kernel (2) and has the spacial frequency cut-off at $f_c = 2\bar{f}_c$ (twice that of the kernel $h(\mathbf{v})$). The emitted intensity $x(\mathbf{w})$ is a nonnegative function, a property that is crucially important for all results in this paper. The l1-norm of the signal $x(\mathbf{w})$

$$\|x(\cdot)\|_1 = \int |x(\mathbf{w})| d\mathbf{w} \tag{8}$$

has the meaning of cumulative emitted intensity, or, in other words, the total energy of light emitted per second. As a side remark, note that when the sample is illuminated by coherent light, as in X-ray crystallography, the resulting input-output relations is no longer linear, in stark contrast to (7), and the phase retrieval problem needs to be solved. For the interested reader, this point is explained in Appendix B. More details on the material reviewed so far can be found in an excellent book by Goodman [15].

Our goal is to reconstruct the signal $x(\mathbf{w})$ from the observations $\tilde{s}(\mathbf{v})$ in (7). Without additional structural assumptions on $x(\mathbf{w})$, this is clearly not possible, because the hi-frequency components of $x(\mathbf{w})$ are lost. The details of $x(\mathbf{w})$ that are smaller than the Rayleigh diffraction limit[1], $1.22/f_c$, can not be distinguished [15, Sec 6.5.2].

The model application for our analysis is single-molecule microscopy [3,11,19], a modern imaging technique, in which the signal $x(\mathbf{w})$ consists of several disjoint molecules emitting light. The size of each molecule is much smaller than $1/f_c$ and it is necessary to estimate the locations of the molecules with precision that is significantly higher than the Rayleigh diffraction limit $\sim 1/f_c$. Mathematically, we assume that the signal $x(\mathbf{w})$ consists of several Dirac spikes:

$$x(\mathbf{w}) = \sum_i x_i \delta(\mathbf{w} - \mathbf{w}_i), \tag{9}$$

where the locations of the spikes correspond to the locations of the molecules. In microscopy the observations in (7) are contaminated by noise. The main contribution of this paper is to show that under the structural model (9), it is possible to estimate $x(\mathbf{w})$ stably (w.r.t. to noise) with resolution beyond the diffraction limit via convex optimization. The stability of estimation fundamentally depends on how regularly (in the sense explained in Section 2) the spikes are distributed in the image domain.

1.1 Mathematical model

The super-resolution theory developed in this paper is discrete, which means that the input signal $x(\mathbf{w})$ is assumed to be supported on a grid. In optics, there is no grid, of course; the spikes in (9) can be in arbitrary (continuous) locations. In the companion paper [21], we show how to generalize our key results, Theorem 1 and Theorem 2, to the continuous setup. The analysis of the problem in the continuous setup is much more complex than the one presented here; the final result, however, is essentially the same. The interested reader is referred to [21] for details. To be able to explain the key concepts in the simplest possible way, we now switch to a much-simpler-to-analyze discrete setup.

[1]The specific value of the constant, 1.22, is largely a historical convention; the point here is that the details of the image that are much smaller than $1/f_c$ are blurred.

1.1.1 1D model

We begin by writing a one-dimensional (1D) discrete analog of (7). To do so, observe that (7) contains a convolution with a square of the sinc kernel. The frequency response of the sinc kernel is a box function (ideal low-pass filter). The frequency response of the square of the the sinc kernel is a convolution of the box function with itself, which is a triangle function. Therefore, the discrete and 1D analog of (7) is:

$$\tilde{\mathbf{s}} = \underbrace{\mathbf{F}^{\mathsf{H}}\hat{\mathbf{Q}}\mathbf{F}}_{\mathbf{Q}=[\mathbf{q}_0\cdots\mathbf{q}_{N-1}]}\mathbf{x}, \tag{10}$$

where $\mathbf{F} = \left[e^{-\mathrm{i}2\pi kl/N}/\sqrt{N}\right]_{-N/2\le k\le N/2-1,\ 0\le l\le N-1}$ is the discrete Fourier transform (DFT) matrix and $\hat{\mathbf{Q}} = \mathrm{diag}([\hat{q}_{-N/2+1}\cdots\hat{q}_{N/2}]^{\mathsf{T}})$ with

$$\hat{q}_k = \begin{cases} 1 - \frac{|k|}{f_c+1}, & k = -f_c, \ldots, f_c \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Above, $\mathbf{x} = [x_0\cdots x_{N-1}]^{\mathsf{T}} \in \mathbb{R}_+^N$ is the unknown signal, the nonzero elements of which represent the molecules in the corresponding location (on the grid); the elements of $\tilde{\mathbf{s}}$ represent the intensity of light observed at the corresponding pixel at the detector; and we assumed that N is even for simplicity. The wavelength $\lambda_c \triangleq 1/f_c$ quantifies the width of the convolution kernel represented by $\mathbf{Q}$.

1.1.2 2D model

Similarly, in the 2D case the discrete analog of (7) reads

$$\tilde{\mathbf{s}}_{2\mathrm{D}} = \underbrace{\mathbf{F}_{2\mathrm{D}}^{\dagger}\hat{\mathbf{Q}}_{2\mathrm{D}}\mathbf{F}_{2\mathrm{D}}}_{\mathbf{Q}_{2\mathrm{D}}}\mathbf{x}_{2\mathrm{D}}, \tag{12}$$

where $\mathbf{F}_{2\mathrm{D}} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N \times \mathbb{C}^N$ is the linear operator that implements the 2D Fourier transform and acts according to

$$[\mathbf{F}_{2\mathrm{D}}\mathbf{x}_{2\mathrm{D}}]_{k_1,k_2} = \frac{1}{N}\sum_{l_1=0}^{N-1}\sum_{l_2=0}^{N-1} x_{l_1,l_2}e^{-\mathrm{i}2\pi(k_1l_1+k_2l_2)/N} \tag{13}$$

and $\hat{\mathbf{Q}}_{2\mathrm{D}} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N \times \mathbb{C}^N$ is the diagonal operator in the 2D Fourier domain that acts according to

$$[\hat{\mathbf{Q}}_{2\mathrm{D}}\mathbf{y}_{2\mathrm{D}}]_{k_1,k_2} = \hat{q}_{k_1}\hat{q}_{k_2}[\mathbf{y}_{2\mathrm{D}}]_{k_1,k_2}. \tag{14}$$

1.1.3 Intensity normalization

According to the discussion above, we interpret $\|\mathbf{x}\|_1$ and $\|\mathbf{x}_{2\mathrm{D}}\|_1$ as the total intensity of light emitted by the object. Similarly, we interpret $\|\tilde{\mathbf{s}}\|_1$ and $\|\tilde{\mathbf{s}}_{2\mathrm{D}}\|_1$ as the total intensity of light observed at the receiver. By convolution theorem, equation (11) guarantees $\|\mathbf{q}_l\|_1 = 1$ for all l. Therefore,

$$\|\tilde{\mathbf{s}}\|_1 = \|\sum_{l=0}^{N-1}\mathbf{q}_l x_l\|_1 = \sum_{l=0}^{N-1}\|\mathbf{q}_l\|_1 x_l = \sum_{l=0}^{N-1} x_l = \|\mathbf{x}\|_1 \tag{15}$$

and similarly $\|\tilde{\mathbf{s}}_{2\mathrm{D}}\|_1 = \|\mathbf{x}_{2\mathrm{D}}\|_1$. Hence, our normalization is such that the intensity of light (emitted energy per second) is conserved in the system.

1.1.4 Noise

All imaging systems are fundamentally noisy. In modern microscopy applications, the intensities of emitted/received light are very low. In low-intensity regime, the dominant source of noise is due to quantum-mechanical effects. Consider the 1D case. In quantum-mechanical picture, $\tilde{s}_l$ represents the expected number of photons to be recorded (in one second) at the l-th pixel of the detector. The actual number of photons detected (in one second) is a random variable that is Poisson-distributed according to $s_l \sim \mathrm{Pois}\,(\tilde{s}_l)$ or, equivalently,

$$\mathbf{s} \sim \mathrm{Pois}\,(\tilde{\mathbf{s}})\,. \tag{16}$$

For our purposes it is convenient to introduce signal-dependent additive noise $\mathbf{z} = \mathbf{s} - \mathbf{Q}\mathbf{x}$ so that the input-output (IO) relation (16) becomes

$$\mathbf{s} = \mathbf{Q}\mathbf{x} + \mathbf{z}. \tag{17}$$

In exactly the same way, in the 2D case, the noisy IO relation reads

$$\mathbf{s}_{2\mathrm{D}} = \mathbf{Q}_{2\mathrm{D}}\mathbf{x}_{2\mathrm{D}} + \mathbf{z}_{2\mathrm{D}}. \tag{18}$$

Clearly, $\mathbf{z}$ and $\mathbf{z}_{2\mathrm{D}}$ are signal-dependent. This does not complicate matters because the stability estimates derived in this paper are valid for arbitrary $\mathbf{z}$ and $\mathbf{z}_{2\mathrm{D}}$, as long as $\|\mathbf{z}\|_1 \leq \delta$ and $\|\mathbf{z}_{2\mathrm{D}}\|_1 \leq \delta$. To account for randomness in $\mathbf{z}$ and $\mathbf{z}_{2\mathrm{D}}$, one should choose δ large enough so that $\|\mathbf{z}\|_1 \leq \delta$ and $\|\mathbf{z}_{2\mathrm{D}}\|_1 \leq \delta$ with high probability. This is easy to do using concentration inequalities.

1.2 Mathematical model with flat spectrum

We have seen that it is important to study IO relations (17) and (18) because they model the physics of imaging in optics. From the mathematical point of view, we will see that our main results are practically independent of the exact shape of the spectrum in (11). Concretely, only a constant in stability estimates changes with the shape of the spectrum. To analyze the models (17) and (18), it is convenient to begin by analyzing a model in which the spectrum of the convolution kernel is flat and has a sharp cut-off at f_c. In the 1D case this model is

$$\mathbf{s} = \mathbf{P}\mathbf{x} + \mathbf{z}, \tag{19}$$

where $\mathbf{P} \triangleq \mathbf{F}^{\mathsf{H}}\hat{\mathbf{P}}\mathbf{F}$ and $\hat{\mathbf{P}} \triangleq \mathrm{diag}([\hat{p}_{-N/2+1} \cdots \hat{p}_{N/2}]^{\mathsf{T}})$ with

$$\hat{p}_k \triangleq \begin{cases} 1, & k = -f_c, \ldots, f_c \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

In the 2D case the corresponding model is

$$\mathbf{s}_{2\mathrm{D}} \triangleq \mathbf{P}_{2\mathrm{D}}\mathbf{x}_{2\mathrm{D}} + \mathbf{z}_{2\mathrm{D}}, \tag{21}$$

where $\mathbf{P}_{2\mathrm{D}} \triangleq \mathbf{F}_{2\mathrm{D}}^{\dagger}\hat{\mathbf{P}}_{2\mathrm{D}}\mathbf{F}_{2\mathrm{D}}$ and $\hat{\mathbf{P}}_{2\mathrm{D}} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N \times \mathbb{C}^N$ is the diagonal operator in the 2D Fourier domain that acts according to

$$[\hat{\mathbf{P}}_{2\mathrm{D}}\mathbf{y}_{2\mathrm{D}}]_{k_1,k_2} \triangleq \hat{p}_{k_1}\hat{p}_{k_2}[\mathbf{y}_{2\mathrm{D}}]_{k_1,k_2}. \tag{22}$$

1.3 Notation

Sets are denoted by calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots$ Boldface letters $\mathbf{A}, \mathbf{B}, \ldots$ and $\mathbf{a}, \mathbf{b}, \ldots$ denote matrices (or linear operators) and vectors, respectively. The element in the i-th row and j-th column of a matrix $\mathbf{A}$ is a_{ij} or $[\mathbf{A}]_{i,j}$, and the i-th element of the vector $\mathbf{a}$ is a_i or $[\mathbf{a}]_i$. For a vector $\mathbf{a}$, $\mathrm{diag}(\mathbf{a})$ stands for the diagonal matrix that has the entries of $\mathbf{a}$ on its main diagonal. The superscripts $^{\mathsf{T}}$ and $^{\mathsf{H}}$ stand for transposition and Hermitian transposition, respectively. For a linear operator $\mathbf{A}$, its adjoint is denoted as $\mathbf{A}^{\dagger}$. For a finite subset of the set of natural numbers, $\mathcal{I} \subset \mathbb{N}$, we write $\big|\mathcal{I}\big|$ for the cardinality of $\mathcal{I}$. For two functions $f(\cdot)$ and $g(\cdot)$, the notation $f(\cdot) = \mathcal{O}(g(\cdot))$ means that $\lim_{t\to\infty} \big|f(t)/g(t)\big|$ is bounded. For $x \in \mathbb{R}$, $\lceil x \rceil \triangleq \min\{m \in \mathbb{Z} \mid m \geq x\}$. We use $[l : k]$ to designate the set of natural numbers $\{l, l+1, \ldots, k\}$. The expectation operator is designated by $\mathbb{E}\big[\cdot\big]$. For vector $\mathbf{a} \in \mathbb{C}^K$, $\|\mathbf{a}\|_1 = \sum_{j=0}^{K-1} \big|a_j\big|$ and $\|\mathbf{a}\|_2 = \sqrt{\sum_{j=0}^{K-1} \big|a_j\big|^2}$ denote l1- and l2- norms, respectively; $\|\mathbf{a}\|$ means either $\|\mathbf{a}\|_1$ or $\|\mathbf{a}\|_2$. The number of nonzero elements of vector $\mathbf{a}$ is denoted $\|\mathbf{a}\|_0$. For a matrix $\mathbf{A} \in \mathbb{C}^{K\times K}$, the operator norm is defined as $\|\mathbf{A}\|_{1,op} = \max_{i:i\in[0\,:\,K-1]} \sum_{j=0}^{K-1} \big|a_{ij}\big|$.

1.4 Organization of the paper

In Section 2 we define the fundamental notion of Rayleigh-regularity and present our main results. In Section 3 we briefly review the literature and explain the novelty of this paper. Our main results, Theorem 1 and Theorem 2, are proven in Section 4; the converse, Theorem 3, is proven in Appendix A. We conclude in Section 5.

2 Main results

Our main goal is to estimate $\mathbf{x}$, $\mathbf{x}_{2\mathrm{D}}$ from $\mathbf{s}$, $\mathbf{s}_{2\mathrm{D}}$ in a stable way. This corresponds to resolving the continuous signal in (9) from low-frequency observations beyond the diffraction limit. The nonzero elements of $\mathbf{x}$ (or $\mathbf{x}_{2\mathrm{D}}$) will be suggestively called "spikes" (see Figure 2 for illustration). Consider the 1D model for concreteness. From (11), (20) we see that we have access to $n = 2f_c + 1$ low-frequency observations. The total number of degrees-of-freedom in $\mathbf{x}$ is N. The ratio $\mathrm{SRF} \triangleq N/n$ is called the super-resolution factor (SRF), it quantifies the degree of under-sampling.

As we will review below, in the noiseless case, the sparsity condition $\|\mathbf{x}\|_0 < n/2$ is sufficient for recovery of $\mathbf{x}$. If there is noise, it turns out that sparsity is not enough. We will see that our ability to estimate $\mathbf{x}$, $\mathbf{x}_{2\mathrm{D}}$ from $\mathbf{s}$, $\mathbf{s}_{2\mathrm{D}}$ in a stable way fundamentally depends on how regular the positions of the spikes are, i.e., how many spikes may be clustered close together. We next define this notion of regularity formally.

2.1 Rayleigh-regularity in 1D model

Let $\mathbb{T}$ be a circle of length one, i.e., the interval $[0, 1]$ with points 0 and 1 identified. It will be convenient to think of the signal $\mathbf{x} \in \mathbb{C}^N$ as a signal supported on the grid $\{0, 1/N, \ldots, 1-1/N\} \subset \mathbb{T}$ (see Figure 2). The following definition, inspired by [12, Def. 1], is needed to formally state the regularity conditions we will use.

Definition 1. Fix N, n. We say that index-set $\mathcal{T} \subset \{0, 1/N, \ldots, 1 - 1/N\} \subset \mathbb{T}$ is Rayleigh-regular with parameters (d, r) and write $\mathcal{T} \in \mathcal{R}^{\mathrm{idx}}(d, r; N, n)$ if every interval $(a, b) \subset \mathbb{T}$ of length $\mu[(a, b)] = d/(n-1) = d\lambda_c/2$ contains no more than r elements of $\mathcal{T}$:

$$|\mathcal{T} \cap (a, b)| \leq r. \tag{23}$$

To be clear, (a, b) is a wrap-around interval on the circle $\mathbb{T}$, or an arc, and $\mu[(a, b)]$ is the length of this arc. When no ambiguity arises, we will shortly write $\mathcal{R}^{\text{idx}}(d, r)$ instead of $\mathcal{R}^{\text{idx}}(d, r; N, n)$.

It is also convenient to define a set of Rayleigh-regular signals (and nonnegative Rayleigh-regular signals) with parameters (d, r):

$$\mathcal{R}(d, r) = \{\mathbf{x} \in \mathbb{C}^N : \text{supp}(\mathbf{x}) \in \mathcal{R}^{\text{idx}}(d, r)\} \tag{24}$$

$$\mathcal{R}_+(d, r) = \{\mathbf{x} \in \mathbb{R}_+^N : \text{supp}(\mathbf{x}) \in \mathcal{R}^{\text{idx}}(d, r)\}, \tag{25}$$

where $\text{supp}(\mathbf{x}) = \{l/N : x_l \neq 0\}$.

In the special case $r = 1$, we say that the spikes of the signal $\mathbf{x} \in \mathcal{R}^{\text{idx}}(d, 1)$ satisfy the minimum separation constraint: every two spikes are separated by at least $d\lambda_c/2$.

Figure 2 illustrates these concepts for different values of parameters. In the figure, point 0 should be identified with point 1. Intuitively, $\mathbf{x} \in \mathcal{R}(d, r)$ simply means that the signal $\mathbf{x}$, depicted on the grid $\{0, 1/N, \ldots, 1-1/N\} \subset \mathbb{T}$ contains no more than r spikes in any d consecutive Nyquist intervals; Nyquist interval is the interval of length $\lambda_c/2$. We mention in passing that $\mathcal{R}(d, r_1) \subset \mathcal{R}(d, r_2)$ for $r_1 \leq r_2$ and $\mathcal{R}(d_1, r) \subset \mathcal{R}(d_2, r)$ for $d_1 \geq d_2$.

Consider examples. First take $\mathbf{x} \in \mathcal{R}(1, 1)$. This signal may contain one spike per one Nyquist interval. Each spike has two unknown parameters associated with it: location and magnitude. There are $2n$ unknown variables in total, which is more than the number of observations n (cf. (11), (20)). Hence, recovery of $\mathbf{x} \in \mathcal{R}(1, 1)$ is in general not possible even in the noiseless case. If we knew the locations of the spikes but not the magnitudes, we could recover the signal $\mathbf{x} \in \mathcal{R}(1, 1)$ by solving a system of linear equations.

Next take $\mathbf{x} \in \mathcal{R}(2, 1)$. This signal may only contain one spike per two Nyquist intervals. Hence, the total number of unknown parameters, n, is equal to the number of observations and recovery of $\mathbf{x} \in \mathcal{R}(2, 1)$ is barely possible in the noiseless case. For example, as we will discuss in Section 3, $\mathbf{x}$ can be recovered by Prony method. In general, $\mathbf{x} \in \mathcal{R}(2r, r)$ is the absolute limit for recovery of complex-valued signals in the noiseless case, i.e., $\mathbf{x} \in \mathcal{R}(2r - \epsilon, r), \epsilon > 0$, is in general not recoverable.

For positive signals, $\mathbf{x} \in \mathbb{R}_+^N$, strictly speaking, the general dimension-counting considerations above do not hold because the positivity of $\mathbf{x}$ supplies extra information. Still, it is possible to construct adversarial signals $\mathbf{x} \in \mathcal{R}_+(2r - \epsilon, r)$ that will not be recoverable. Looking ahead, in this paper we will show that $\mathbf{x} \in \mathcal{R}_+(3.74r, r)$ can be recovered stably, even if there is noise.

2.2 Rayleigh-regularity in 2D model

In the 2D case, it is convenient to think of the signal $\mathbf{x}_{2\text{D}} \in \mathbb{C}^N \times \mathbb{C}^N$ as a signal supported on a grid $\{0, 1/N, \ldots, 1-1/N\} \times \{0, 1/N, \ldots, 1-1/N\} \subset \mathbb{T} \times \mathbb{T}$, where $\mathbb{T} \times \mathbb{T}$ is the torus, i.e., the square $[0, 1] \times [0, 1]$ with the corresponding sides identified. We generalize the notion of Rayleigh-regularity as follows.

Definition 2. Fix N, n. We say that index-set $\mathcal{T} \subset \{0, 1/N, \ldots, 1 - 1/N\} \times \{0, 1/N, \ldots, 1 - 1/N\} \subset \mathbb{T} \times \mathbb{T}$ is Rayleigh-regular with parameters (d, r) and write $\mathcal{T} \in \mathcal{R}_{2\text{D}}^{\text{idx}}(d, r; N, n)$ if it can be represented as a union of r sets $\mathcal{T} = \mathcal{T}_1 \cup \ldots \cup \mathcal{T}_r$ that are nonintersecting and satisfy the 2D minimum separation constraint. Precisely:

1. for all $1 \leq i < j \leq r$, $\mathcal{T}_i \cap \mathcal{T}_j = \emptyset$;
2. for all $1 \leq i \leq r$, $\mathcal{T}_i$ satisfies: for all square subsets $\mathcal{D} \subset \mathbb{T}^2$ of dimensions $d\lambda_c/2 \times d\lambda_c/2$,

$$|\mathcal{T}_i \cap \mathcal{D}| \leq 1. \tag{26}$$

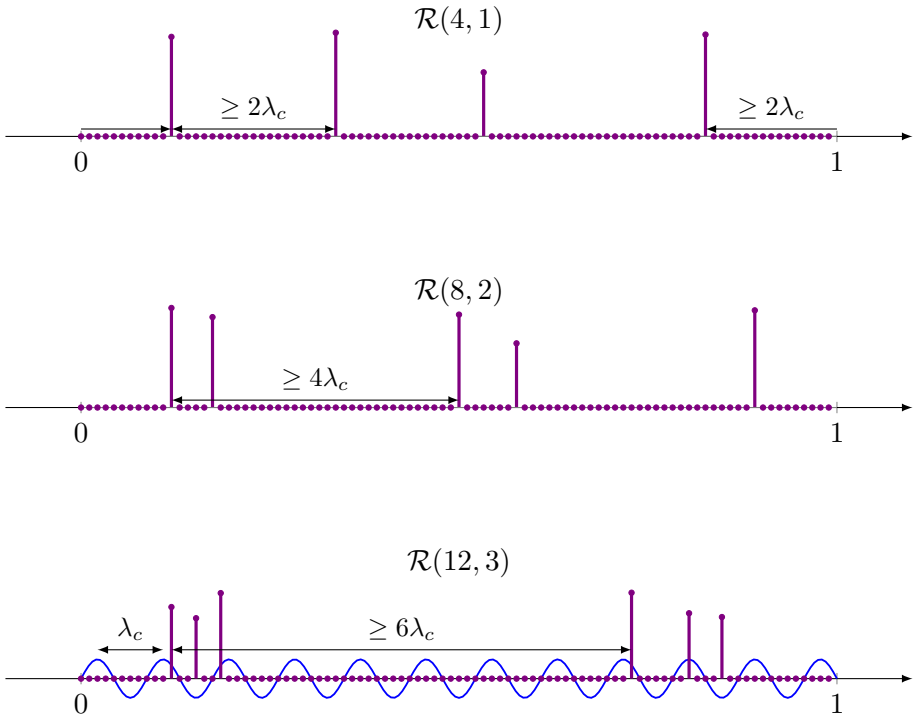

Figure 2: Examples of discrete N dimensional signals from Rayleigh classes $\mathcal{R}(4,1)$, $\mathcal{R}(8,2)$, $\mathcal{R}(12,3)$ depicted on the grid $\{0, 1/N, \ldots, 1-1/N\} \subset \mathbb{T}$. The function $\sin(2\pi f_c t)$ is depicted on the third panel for reference. Here, $N = 92$ and $n = 23$, so that SRF $= 4$ and $\lambda_c = 1/11$. Points 0 and 1 should be identified.

For a signal $\mathbf{x}_{2\mathrm{D}} \in \mathbb{C}^N \times \mathbb{C}^N$ and $\mathbf{x}_{2\mathrm{D}} \in \mathbb{R}_+^N \times \mathbb{R}_+^N$ we write $\mathbf{x}_{2\mathrm{D}} \in \mathcal{R}_{2\mathrm{D}}(d,r)$ and $\mathbf{x}_{2\mathrm{D}} \in \mathcal{R}_{+,2\mathrm{D}}(d,r)$, respectively, and say that $\mathbf{x}_{2\mathrm{D}}$ is Rayleigh-regular if $\mathrm{supp}(\mathbf{x}_{2\mathrm{D}}) \in \mathcal{R}_{2\mathrm{D}}^{\mathrm{idx}}(d,r;N,n)$.

Remark 1. Note the difference between the definitions of Rayleigh-regularity in the 1D case, Definition 1, and in the 2D case, Definition 2. For the proofs in both 1D and 2D cases we will need a property that the support set $\mathcal{T}$ can be represented as a union of r sets that are nonintersecting and satisfy the minimum separation constraint, as required in Definition 2. In the 1D case, this property is a simple consequence of the requirement in Definition 1: no more than r spikes in d consecutive Nyquist intervals. In the 2D case, this is not true: the condition of Definition 2 does not necessarily follow from a 2D version of the condition in Definition 1, as can be seen by constructing a simple counter-example (we have a counter-example for $r = 4$).

2.3 Recovery algorithms

2.3.1 P and $\mathbf{P}_{2\mathrm{D}}$ models

Assume that in (19) and (21) the noise is bounded according to $\|\mathbf{z}\|_1 \leq \delta$ and $\|\mathbf{z}_{2\mathrm{D}}\|_1 \leq \delta$, respectively. We propose to estimate $\mathbf{x}$ and $\mathbf{x}_{2\mathrm{D}}$ by solving the following convex optimization problem

$$\text{find} \quad \hat{\mathbf{x}} \quad \text{s.t.} \quad \|\mathbf{s} - \mathbf{P}\hat{\mathbf{x}}\|_1 \leq \delta, \quad \hat{\mathbf{x}} \geq \mathbf{0}, \tag{CVX$_{\mathbf{P},1\mathrm{D}}$}$$

in the 1D case and

$$\text{find} \quad \hat{\mathbf{x}}_{2\mathrm{D}} \quad \text{s.t.} \quad \|\mathbf{s} - \mathbf{P}_{2\mathrm{D}}\hat{\mathbf{x}}_{2\mathrm{D}}\|_1 \leq \delta, \quad \hat{\mathbf{x}}_{2\mathrm{D}} \geq \mathbf{0}, \tag{CVX$_{\mathbf{P},2\mathrm{D}}$}$$

in the 2D case.

2.3.2 Q and $\mathbf{Q}_{2\mathrm{D}}$ models

Assume that in (17) the noise is bounded according to $\|\mathbf{z}\|_1 \leq \delta$. To estimate $\mathbf{x}$ stably we propose a two-step procedure. First, we transform the problem by flattening the spectrum of the operator $\mathbf{Q}$. Second, we apply convex optimization to the result. To control the l1-norm of the transformed noise, we need to be careful and avoid discontinuities in the spectrum of the transformation. Specifically, choose $1/2 \leq \alpha < 1$ (a parameter we can optimize) so that αf_c is an integer. Define the filter

$$\hat{r}_k \triangleq \begin{cases} \frac{f_c+1}{f_c+1-|k|}, \ k = -\alpha f_c, \ldots, \alpha f_c \\ (f_c+1)\left(a|k| + b\right), \ |k| = \alpha f_c + 1, \ldots, f_c \\ 0, \ \text{otherwise.} \end{cases} \tag{27}$$

with

$$a \triangleq -\frac{1}{f_c(1-\alpha)+1}\frac{1}{f_c(1-\alpha)} \quad \text{and} \quad b \triangleq \frac{1}{f_c(1-\alpha)+1}\frac{1}{1-\alpha}. \tag{28}$$

Set $\hat{\mathbf{R}} \triangleq \mathrm{diag}(\hat{\mathbf{r}})$ with $\hat{\mathbf{r}} \triangleq [\hat{r}_{-N/2+1} \cdots \hat{r}_{N/2}]^{\mathsf{T}}$ and let $\mathbf{R} \triangleq \mathbf{F}^{\mathsf{H}}\hat{\mathbf{R}}\mathbf{F}$. Multiplying both sides of (17) by $\mathbf{R} = \mathbf{F}^{\mathsf{H}}\hat{\mathbf{R}}\mathbf{F}$ we get

$$\mathbf{R}\mathbf{s} = \mathbf{F}^{\mathsf{H}}\hat{\mathbf{R}}\mathbf{F}\mathbf{F}^{\mathsf{H}}\hat{\mathbf{Q}}\mathbf{F}\mathbf{x} + \mathbf{R}\mathbf{z} \tag{29}$$

$$= \mathbf{F}^{\mathsf{H}}\hat{\mathbf{R}}\hat{\mathbf{Q}}\mathbf{F}\mathbf{x} + \mathbf{R}\mathbf{z} \tag{30}$$

$$= \underbrace{\mathbf{F}^{\mathsf{H}}\hat{\mathbf{T}}\mathbf{F}}_{\mathbf{T}}\mathbf{x} + \mathbf{R}\mathbf{z}, \tag{31}$$

where $\hat{\mathbf{T}} \triangleq \mathrm{diag}([\hat{t}_{-N/2+1} \cdots \hat{t}_{N/2}]^{\mathsf{T}})$ with

$$\hat{t}_k \triangleq \begin{cases} 1, \ k = -\alpha f_c, \ldots, \alpha f_c \\ \left(f_c + 1 - |k|\right)\left(a|k| + b\right), \ |k| = \alpha f_c + 1, \ldots, f_c \\ 0, \ \text{otherwise.} \end{cases} \tag{32}$$

Note that the spectrum of $\mathbf{T}$ is flat in the region $\alpha f_c \leq k \leq \alpha f_c$; this will be important in the proof of our stability results. We will see in Section 4.2 that the transformed noise in (31) satisfies $\|\mathbf{R}\mathbf{z}\|_1 \leq C(\alpha)\delta$, where

$$C(\alpha) \triangleq 2\alpha + \frac{2}{1-\alpha} + \frac{1.11}{2(1-\alpha)^2}. \tag{33}$$

Note that $C(\alpha)$ is finite as long as $\alpha < 1$. For $\alpha = 1/2$, $C(\alpha) = 7.22$. For $\alpha = 0.75$, $C(\alpha) = 18.38$. The estimator $\hat{\mathbf{x}}$ is a solution to the following convex optimization problem

$$\text{find} \quad \hat{\mathbf{x}} \quad \text{s.t.} \quad \|\mathbf{R}\mathbf{s} - \mathbf{T}\hat{\mathbf{x}}\|_1 \leq C(\alpha)\delta, \quad \hat{\mathbf{x}} \geq \mathbf{0}. \tag{CVX$_{\mathbf{Q},1\mathrm{D}}$}$$

Similarly, assuming that the noise in (18) is bounded according to $\|\mathbf{z}_{2\mathrm{D}}\|_1 \leq \delta$, we estimate $\mathbf{x}_{2\mathrm{D}}$ by solving

$$\text{find} \quad \hat{\mathbf{x}}_{2\mathrm{D}} \quad \text{s.t.} \quad \|\hat{\mathbf{R}}_{2\mathrm{D}}\mathbf{s}_{2\mathrm{D}} - \hat{\mathbf{T}}_{2\mathrm{D}}\hat{\mathbf{x}}_{2\mathrm{D}}\|_1 \leq C_{2\mathrm{D}}(\alpha)\delta, \quad \hat{\mathbf{x}}_{2\mathrm{D}} \geq \mathbf{0}, \tag{CVX$_{\mathbf{Q},2\mathrm{D}}$}$$

where $\mathbf{R}_{2D}, \mathbf{T}_{2D} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N \times \mathbb{C}^N$ are diagonal operators in 2D Fourier domain that act according to $\mathbf{R}_{2D} = \mathbf{F}_{2D}^{\mathsf{H}} \hat{\mathbf{R}}_{2D} \mathbf{F}_{2D}, \mathbf{T}_{2D} = \mathbf{F}_{2D}^{\mathsf{H}} \hat{\mathbf{T}}_{2D} \mathbf{F}_{2D}$ with

$$[\hat{\mathbf{R}}_{2D} \mathbf{y}_{2D}]_{k_1,k_2} = \hat{r}_{k_1} \hat{r}_{k_2} [\mathbf{y}_{2D}]_{k_1,k_2} \tag{34}$$

$$[\hat{\mathbf{T}}_{2D} \mathbf{y}_{2D}]_{k_1,k_2} = \hat{t}_{k_1} \hat{t}_{k_2} [\mathbf{y}_{2D}]_{k_1,k_2} \tag{35}$$

and $C_{2D}(\alpha)$ depends on α only and is finite for $\alpha < 1$.

2.4 Main results

We are now ready to state our main results.

Theorem 1.
Part P: *Let* $\mathbf{x} \in \mathcal{R}_+(3.74r, r)$ *with* $r \geq 1$ *and assume that* $\mathbf{s}$ *is given by (19) with* $\|\mathbf{z}\|_1 \leq \delta$. *Assume*[2] $\mathrm{SRF} \geq 3.03/r$ *and* $f_c \geq 128r$. *Assume, for simplicity, that* r *divides* S *and* r *divides* f_c. *Then, every solution* $\hat{\mathbf{x}}$ *of (*$\mathrm{CVX}_{\mathbf{P},1D}$*) satisfies*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq C_1(r) \left(\frac{N}{n-1}\right)^{2r} \delta \approx C_1(r) \mathrm{SRF}^{2r} \delta, \tag{36}$$

where $C_1(r)$, *defined in (91), depends on* r *only.*

Part Q: *Let* $\mathbf{x} \in \mathcal{R}_+(3.74r/\alpha, r)$ *with* $r \geq 1$ *and* $1/2 \leq \alpha < 1$. *Assume that* $\mathbf{s}$ *is given by (17) with* $\|\mathbf{z}\|_1 \leq \delta$. *Assume* $\mathrm{SRF} \geq 3.03/r$ *and* $f_c \geq 256r$. *Then, every solution* $\hat{\mathbf{x}}$ *of (*$\mathrm{CVX}_{\mathbf{Q},1D}$*) satisfies*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq C_1(r, \alpha) \left(\frac{N}{n-1}\right)^{2r} \delta \approx C_1(r, \alpha) \mathrm{SRF}^{2r} \delta, \tag{37}$$

where $C_1(r, \alpha)$, *defined in (104), depends on* r *and* α *only and is finite for all* $\alpha < 1$.

Theorem 2.
Part P: *Let* $\mathbf{x}_{2D} \in \mathcal{R}_{+,2D}(4.76r, r)$ *with* $r \geq 1$ *and assume that* $\mathbf{s}_{2D}$ *is given by (21) with* $\|\mathbf{z}_{2D}\|_1 \leq \delta$. *Assume* $f_c \geq 512r$. *Then, every solution* $\hat{\mathbf{x}}_{2D}$ *of (*$\mathrm{CVX}_{\mathbf{P},2D}$*) satisfies*

$$\|\hat{\mathbf{x}}_{2D} - \mathbf{x}_{2D}\|_1 \leq C_2(r) \left(\frac{N}{n-1}\right)^{2r} \delta \approx C_2(r) \mathrm{SRF}^{2r} \delta, \tag{38}$$

where $C_2(r)$, *defined in (208), depends on* r *only.*

Part Q: *Let* $\mathbf{x}_{2D} \in \mathcal{R}_{+,2D}(4.76r/\alpha, r)$ *with* $r \geq 1$ *and assume that* $\mathbf{s}_{2D}$ *is given by (12) with* $\|\mathbf{z}_{2D}\|_1 \leq \delta$. *Assume,* $f_c \geq 1024r$. *Then, every solution* $\hat{\mathbf{x}}_{2D}$ *of (*$\mathrm{CVX}_{\mathbf{Q},2D}$*) satisfies*

$$\|\hat{\mathbf{x}}_{2D} - \mathbf{x}_{2D}\|_1 \leq C_2(r, \alpha) \left(\frac{N}{n-1}\right)^{2r} \delta \approx C_2(r, \alpha) \mathrm{SRF}^{2r} \delta, \tag{39}$$

where $C_2(r, \alpha)$ *depends on* r *and* α *only and is finite for all* $\alpha < 1$.

The proof of Theorem 1 and Theorem 2 is the main technical novelty of this paper and is presented in Section 4.

[2]The mild condition $\mathrm{SRF} \geq 3.03/r$ is only needed to obtain the specific value for $C_1(r)$.

2.5 Interpretation of the results

2.5.1 Implications for single-molecule microscopy

Consider Theorem 2, Part Q, for concreteness. Remember that each element $[\mathbf{z}_{2D}]_{k_1,k_2}$ of the signal $\mathbf{z}_{2D}$ contains the difference in intensity of light between the noiseless (ideal) output $[\mathbf{Q}_{2D}\mathbf{x}_{2D}]_{k_1,k_2}$ of the optical system and the real noisy signal $[\mathbf{s}_{2D}]_{k_1,k_2}$ recorded at the detector located at (k_1, k_2). Hence, $\|\mathbf{z}_{2D}\|_1$ is the cumulative difference in light intensity between the noiseless (ideal) and the real observation. Similarly, $[\mathbf{x}_{2D}]_{k_1,k_2}$ is the intensity of light emitted from the location in space with coordinates (k_1, k_2) and $[\hat{\mathbf{x}}_{2D}]_{k_1,k_2}$ is the estimated light intensity for this location. Hence $\|\mathbf{x}_{2D} - \hat{\mathbf{x}}_{2D}\|_1$ is the cumulative error in light intensity generated by the recovery algorithm. Theorem 2, Part Q, tells us that the cumulative error in light intensity in signal estimates is bounded by the amplified version of the cumulative error in light intensity in the data. The noise amplification factor (NAF) behaves as SRF^{2r}, where r is the parameter describing the regularity of the signal's support. If the noise level δ is sufficiently small and the signal is sufficiently regular (r is small), i.e., not too many molecules are clustered close together, and SRF is modest, then the algorithm ($\mathrm{CVX}_{\mathbf{Q},2D}$) is guaranteed to achieve excellent super-resolution results. As we will explain in Section 2.7, no algorithm can perform substantially better. We stress the importance of stability estimates given in terms of l1 norm (as apposed to l2 norm). As we see, in incoherent optics, l1 norm has a direct physical interpretation (energy), and l2 norm does not.

2.5.2 Noise amplification factor

The discussion above can be formalized as follows.

Definition 3. Let $\|\cdot\|$ be a norm, $\mathcal{C}$ be a signal class, and $\mathbf{B}$ be a linear operator. Suppose that for the model $\mathbf{s} = \mathbf{B}\mathbf{x} + \mathbf{z}$ an algorithm A produces an estimator $\hat{\mathbf{x}}(\mathbf{s})$ that satisfies a uniform stability guarantee

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \leq \mathrm{NAF}[A, \mathcal{C}, \mathbf{B}]\delta \tag{40}$$

for all $\mathbf{x} \in \mathcal{C}$ and all $\mathbf{z}$ such that $\|\mathbf{z}\| \leq \delta$. Then we say that NAF of A is (at most) $\mathrm{NAF}[A, \mathcal{C}, \mathbf{B}]$.

With this notation, taking $\|\cdot\| = \|\cdot\|_1$ the result in Theorem 2, Part P, can written concisely as follows

$$\mathrm{NAF}[\mathrm{CVX}_{\mathbf{P},1D}, \mathcal{C}, \mathbf{P}] = C_1(r)\mathrm{SRF}^{2r}, \quad \mathcal{C} = \mathcal{R}_+(3.74r, r) \tag{41}$$

and similarly for all other cases.

2.6 Remarks

Remark 2. By comparing Part P to Part Q in Theorem 1, we see that the results are not sensitive to the choice of the kernel: the only difference between stability estimates in (36) and (37) is in the constant ($C_1(r)$ vs. $C_1(r, \alpha)$). As will be clear from Section 4, the results remain valid for just about any other low-frequency kernel.

Remark 3. Consider the noiseless setup $\delta = 0$. Theorems 1 and 2 guarantee that the signals, $\mathbf{x}$ and $\mathbf{x}_{2D}$, can be recovered exactly via convex optimization. In the 1D case, as we will review in Section 3, this result is well known, [13, 14]. In the 2D case, this is new: as we will explain in Section 4, this result cannot be obtained by a straightforward generalization of the technique in [13, 14].

Remark 4. For concreteness, Theorem 2 is stated for 2D signals, but the result can be straightforwardly generalized to multi-dimensional signals.

Remark 5. Results in Theorem 1 and 2 have a natural probabilistic interpretation. The signal $\mathbf{x} \in \mathbb{R}_+^N$ is interpreted as probability measure: it is positive and can be normalized to sum to one. According to (15), $\tilde{\mathbf{s}}$ is also a probability measure. The stability estimate, a bound on $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$, is a bound on the total variation distance between probability distributions.

2.7 Tightness of the results

In this section we argue that our results in Theorem 1, Part Q, and Theorem 2, Part Q, are nearly tight. To be specific, we focus on the 1D case and analyze the question: How fundamental are the stability estimates in Theorem 1, Part Q?

To simplify the discussion set $\alpha = 1 - \epsilon \approx 1$, the largest value of α for which the constant $C(\alpha)$ in (33) and hence $C_1(r, \alpha)$ in (37) is finite. The question consists of two parts:

(i) Can the assumption $\mathcal{C} = \mathcal{R}_+(3.74r, r)$ be substituted with $\mathcal{C} = \mathcal{R}_+(d, r)$ with $d < 3.74r$ without changing the bound

$$\mathrm{NAF}[\mathrm{CVX}_{\mathbf{Q},1\mathrm{D}}, \mathcal{C}, \mathbf{Q}] \leq c\mathrm{SRF}^{2r}? \tag{42}$$

(ii) Can the exponent $2r$ in (42) be made smaller?

2.7.1 Tightness of the length of the interval

To answer question (i), note that we have already argued in Section 2.1 that even in the noiseless case it is not possible to recover many of the signals $\mathbf{x} \in \mathcal{R}_+(d, r)$ with $d < 2r$. Hence, $d = 3.74r$ is within factor 1.87 of the optimum. As we will review in Section 3, a simple consequence of the results in [13] and [14] guarantees us that when $\mathbf{x} \in \mathcal{R}_+(2r, r)$ it is possible to recover $\mathbf{x}$ exactly by convex programming in the noiseless case. We pay the factor 1.87 to be able to achieve stability.

2.7.2 Tightness of the exponent

To answer question (ii) above, we need the concept of modulus of continuity (MC).

Definition 4. Let $\|\cdot\|$ be a norm and $\mathbf{B}$ be a linear operator. Consider the model $\mathbf{s} = \mathbf{B}\mathbf{x} + \mathbf{z}$. The MC of a signal class $\mathcal{C}$ (w.r.t. to this norm) is defined as

$$\mathrm{MC}[\mathcal{C}, \mathbf{B}] \triangleq \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}} \frac{\|\mathbf{x}_1 - \mathbf{x}_2\|}{\|\mathbf{B}(\mathbf{x}_1 - \mathbf{x}_2)\|}. \tag{43}$$

The MC is related to the NAF via the following simple facts.

1. If the NAF of an algorithm A is at most $\mathrm{NAF}[\mathrm{A}, \mathcal{C}, \mathbf{B}]$, then

$$\mathrm{NAF}[\mathrm{A}, \mathcal{C}, \mathbf{B}] \geq \mathrm{MC}[\mathcal{C}, \mathbf{B}]. \tag{44}$$

2. Consider exhaustive search algorithm (in general intractable) for super-resolution of signals in $\mathcal{C}$:

$$\text{find} \quad \hat{\mathbf{x}} \in \mathcal{C} \quad \text{s.t.} \quad \|\mathbf{s} - \mathbf{B}\hat{\mathbf{x}}\| \leq \delta. \tag{ES}$$

The NAF of this algorithm satisfies

$$\mathrm{NAF}[\mathrm{ES}, \mathcal{C}, \mathbf{B}] \leq 2\mathrm{MC}[\mathcal{C}, \mathbf{B}]. \tag{45}$$

We now provide a lower bound on the MC that shows, according to (44), that if the noise is arbitrary, no algorithm can have a NAF that is smaller than $c\mathrm{SRF}^{2r-1}$. Therefore, the exponent $2r$ in (37) and in (39) is nearly optimal.

Theorem 3. *Fix r and let $\mathcal{C} = \mathcal{R}_+(d, r)$ for some d. Then there exist signals $\mathbf{x} = [x_0 \cdots x_{N-1}]^\mathsf{T}, \tilde{\mathbf{x}} = [\tilde{x}_0 \cdots \tilde{x}_{N-1}]^\mathsf{T} \in \mathcal{C}$ s.t.*

$$\|\mathbf{x} - \tilde{\mathbf{x}}\|_1 = 1 \tag{46}$$

and when $N, n \to \infty$, $N/n \to \mathrm{SRF}$

$$\|\mathbf{Q}(\mathbf{x} - \tilde{\mathbf{x}})\|_1 \to \chi(r, \mathrm{SRF})\mathrm{SRF}^{2r-1}. \tag{47}$$

For $\mathrm{SRF} \to \infty$,

$$\chi(r, \mathrm{SRF}) \to c_L(r) \tag{48}$$

where $c_L(r)$ depends on r only and is given explicitly in (141). Consequently, letting $\|\cdot\|$ be $\|\cdot\|_1$ in Definition 4,

$$\mathrm{MC}[\mathcal{C}, \mathbf{Q}] \geq c_L(r)\mathrm{SRF}^{2r-1} \tag{49}$$

when N, n and $\mathrm{SRF} = N/n$ are large.

The proof of Theorem 3 is given in Appendix A. It relies on an explicit construction of nonnegative signals $\mathbf{x}$ and $\tilde{\mathbf{x}}$ such that supports of $\mathbf{x}$ and $\tilde{\mathbf{x}}$ are disjoint and, at the same time, the spikes in $\mathbf{x} - \tilde{\mathbf{x}}$ cancel our as much as possible after the convolution with the kernel represented by $\mathbf{Q}$.

Theorem 3 is interesting. It tells us that the MC increases exponentially with r. For example, for a practically interesting case $\mathrm{SRF} = 8$, it is not difficult to estimate from (49) that super-resolution could only be possible if $r \leq 5$. For $r > 5$, MC is greater than 10^5, which makes the requirements on how small δ should be unrealistically strict for any practical system. This is an optimistic estimate. In reality, $r = 1, 2, 3$ is the regime where super-resolution will be possible.

Compare Theorem 3 to Theorem 1, Part Q. We see that the exponent of SRF in the right-hand side (RHS) of (37) is within one of the best possible. It is important to point out that the convex optimization algorithm in ($\mathrm{CVX}_{\mathbf{Q},1\mathrm{D}}$) knows nothing at all about the regularity of the signal class $\mathcal{C}$. Yet, surprisingly, it has a nearly optimal stability guarantee for all signals in $\mathcal{C}$.

It is not known, whether there is an algorithm that has a stability guarantee with SRF exponent matching that in the RHS (49), even if one allows for intractable exhaustive search methods that have access to information about the regularity of $\mathcal{C}$. In the very special case, when the signal contains exactly one spike, it is not difficult to see that a simple matched-filter will have NAF $\sim$ SRF, matching the exponent in the RHS of (49). This can be used in the setting where the spikes are guaranteed to be so far apart, that the overlap between their images in the output space can be neglected; this only happens when the distance between neighboring spikes far exceeds $\lambda_c/2$ and all the spikes have roughly the same magnitude. In general, in the interesting case where the images of neighboring spikes can overlap in the output space, it is not clear how one could close the small gap between (49) and (37). In fact, it is possible that the exponent in (49) can be made larger. As we shall see, to construct adversarial signals $\mathbf{x}, \tilde{\mathbf{x}}$ in the proof of Theorem 3, we only use signals that contain exactly r spikes each. However, the signals in $\mathcal{R}_+(d, r)$ can have more than r spikes, of course, which could allow one to construct pairs $\mathbf{x}, \tilde{\mathbf{x}}$ that give larger bound than that in the RHS of (49).

3 Literature review and innovations

3.1 Prior art

Algebraic methods. Since the work of Prony [25] in 1975 there has been a massive amount of research on the problem of super-resolution. Prony developed a way to solve the problem in the noiseless 1D case (19) ($\mathbf{z} = \mathbf{0}$) by an algebraic method. The data $\mathbf{s}$ is used to form a trigonometric polynomial, zeros of which coincide with the locations of spikes in $\mathbf{x}$. The polynomial is then factored and the locations of spikes are found. Once this is done, the amplitudes can be estimated by solving a system of linear equations. In the noiseless case, under signal sparsity constraint $\|\mathbf{x}\|_0 < n/2$, Prony method recovers $\mathbf{x}$ perfectly. Except for sparsity, no further Rayleigh-regularity assumptions on the signal's support are needed. Also, the amplitudes in $\mathbf{x}$ may be complex. With noise, the performance of Prony method degrades sharply. The difficulty comes from the fact that zeros of the trigonometric polynomial constructed by an algebraic method can shift significantly even with small changes in the data.

Many noise-aware versions of Prony method have been developed and are used routinely in engineering applications, for example in radar (see [30, Chapter 6]). The most popular methods are MUSIC and its numerous variations [1, 4, 27, 24, 34, 5], matrix-pensil [17], and ESPRIT [23, 26]. For more details on algebraic methods we refer the interested reader to the excellent book [30, Chapter 4].

The stability of noise-aware algebraic methods is not theoretically well-understood. Asymptotic results (at high SNR) on stability of MUSIC in the presence of Gaussian noise are derived in [10,33]. More recently, some steps towards analyzing the algorithm in a non-asymptotic regime have been taken in [20]. Nevertheless, to the best of our knowledge, no strong theoretical stability guarantees like those in Theorem 1 are available for algebraic methods. Hence, the search for super-resolution methods that are stable in practice and have sharp theoretical stability guarantees is an important open problem.

Algebraic methods have been generalized to the multi-dimensional case. Surprisingly, the generalizations are not straightforward and many methods ([9], [8], [30, Sec. 4.9.7]) have very restrictive sparsity constraints: $\|\mathbf{x}_{2\mathrm{D}}\|_0 < n$. Recall, that in the 2D case the total number of observations is n^2. In [18] it has been shown in the noiseless case that $\|\mathbf{x}_{2\mathrm{D}}\|_0$ can be as large as $n^2/4$, as one would expect from dimension-counting considerations. Clearly, our discussion about instability of algebraic methods applies in the multi-dimensional case.

Fundamental limits. In the ground-breaking work [12], Donoho studied the fundamental limits of stability of 1D super-resolution in the presence of noise. The main findings of [12] can be summarized as follows. Let $\|\cdot\| = \|\cdot\|_2$ in the definition of NAF.

- Let $\mathcal{C} = \mathcal{R}(4r, r)$, then the NAF of the exhaustive search algorithm (ES) [with $\mathbf{B} = \mathbf{P}$] is bounded according to
$$\mathrm{NAF}[\mathrm{ES}, \mathcal{C}, \mathbf{P}] \leq c(r)\mathrm{SRF}^{2r+1}, \tag{50}$$
where $c(r)$ is a positive constant that might depend on r but not on N or n.

- Let $\mathcal{C} = \mathcal{R}(d, r)$, then
$$\mathrm{MC}[\mathcal{C}, \mathbf{P}] \geq c(r)\mathrm{SRF}^{2r-1}, \tag{51}$$
where $c(r)$ is a positive constant that might depend on r but not on N or n.

To the best of our knowledge, the fundamental analysis in [12] has not been generalized to the multi-dimensional case. Unfortunately, no tractable algorithm is proposed in [12] that would have NAF bounded above by the RHS of (50). The algorithm (ES) is not feasible in general, because $\mathcal{C}$ is not convex. The key question posed in [12] is whether a feasible algorithm that achieves stability in (50) exists.

Other works [32, 31, 2] study stability of super-resolution problem in the presence of noise, but also do not provide a tractable algorithm to perform recovery. Work in [28,29,16] analyzes detection and separation of two closely-spaced spikes, but does not generalize to the case when there are more than two spikes in the signal.

Super-resolution under minimum separation constraint. A progress towards resolving the question posed [12] in the general case, $\mathbf{x} \in \mathbb{C}^N$, (recall, in this paper we consider the case $\mathbb{R}^N_+$ only) has recently been made in [7]. Let $\|\cdot\| = \|\cdot\|_1$ in the definition of the NAF and consider $\mathcal{C} = \mathcal{R}(4, 1)$. It was shown that the NAF of l1-minimization algorithm

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 \quad \text{s.t.} \quad \|\mathbf{s} - \mathbf{P}\hat{\mathbf{x}}\|_1 \leq \delta \tag{l1}$$

is at most

$$\mathrm{NAF}[\mathrm{l1}, \mathcal{C}, \mathbf{P}] \leq c\mathrm{SRF}^2, \tag{52}$$

where c is a positive numerical constant[3]. The condition $\mathbf{x} \in \mathcal{C} = \mathcal{R}(4, 1)$ is restrictive because it means that the signal $\mathbf{x}$ cannot contain spikes that are closer than $2\lambda_c$ apart (when depicted on $\mathbb{T}$ as in Figure 2), the minimum separation constraint. [For the real-valued signals $\mathbf{x} \in \mathbb{R}^N$, bound (52) holds as long as $\mathbf{x} \in \mathcal{C} = \mathcal{R}(3.74, 1)$.] This is a serious limitation for many applications, for example, for single-molecule microscopy. Strictly speaking, it is usually understood that the goal of super-resolution is to distinguish spikes that are (significantly) closer than the Rayleigh diffraction limit, i.e., (significantly) closer than $\sim \lambda_c$ apart. Unfortunately, if there are spikes closer than λ_c apart, l1-minimization algorithm does not, in general, return the correct solution even in the noiseless case when $\mathbf{x} \in \mathbb{C}^N$. Results in [7] also cover the multi-dimensional case under the 2D minimum separation constraint.

Super-resolution of noiseless nonnegative signals. The case of 1D nonnegative signal, $\mathbf{x} \in \mathbb{R}^N_+$, was analyzed in [13]. See also [14] for a shorter exposition of the same idea. Adapting to our setting, the result in [13] can be summarized as follows. Let $\|\cdot\| = \|\cdot\|_1$ in the definition of NAF. Let $\mathcal{C}$ be the class of all signals with $\|\mathbf{x}\|_0 < n/2$. Then, $\mathrm{NAF}[\mathrm{CVX}_{\mathbf{P},1\mathrm{D}}, \mathcal{C}, \mathbf{P}]$, the NAF of the convex optimization algorithm ($\mathrm{CVX}_{\mathbf{P},1\mathrm{D}}$) is a finite positive constant. The exact dependence of $\mathrm{NAF}^{\mathrm{cvx}}$ on N and n is not specified in [13]. As we will see, further examination of the proof technique used in [13] leads to the bound of the form

$$\mathrm{NAF}^{\mathrm{cvx}}[\mathcal{C}, \mathbf{P}] \leq (cN)^{2\|\mathbf{x}\|_0}, \tag{53}$$

where c is a numerical constant. This result does not depend on the Rayleigh-regularity of $\mathbf{x}$. However, by comparing (53) to (50) and (52) we see that the bound in (53) is weak. To see this, consider the interesting case $N, n \to \infty$ with $N/n = \mathrm{SRF}$ kept constant. In this case the bounds in (50) and (52) remain finite, whereas the RHS of (53) converges to $+\infty$. In summary, we expect the NAF to depend on SRF only, but not on N or n separately.

[3]Numerical constant means a finite number, independent of any parameters; constant c has different value on each appearance in the paper.

3.2 Innovations

Compared to the previous literature, the novelty of the results presented in Section 2 can be summarized as follows.

- As compared to algebraic methods, our Theorems 1 and 2 show that fast convex optimization algorithms can recover the signal in a provably stable way. As we discussed earlier, strong worst-case stability guarantees are not available for algebraic methods. The price we pay is that our results crucially rely on non-negativity of the signal; algebraic methods do not need this assumption.

- As compared to work in [12], our recovery algorithms are convex and, hence, feasible. The exhaustive search method of [12] can not be used in practice. The difference between stability exponents in (36) and in (50) is because the stability of estimates in [12] are in terms of l2-norm and our estimates are in terms of l1-norm. We point out that stability bounds for the exhaustive search algorithm in [12] do not rely on the nonnegativity of the signal.

- As compared to work in [7], our results do not rely on the restrictive minimum-separation assumption. At the same time, results in [7] are for the general case $\mathbf{x} \in \mathbb{C}^N$, results in this paper crucially rely on nonnegativity of $\mathbf{x}$. Mathematics developed in [7] serves as a building block for our construction.

- As compared to work in [13], our stability estimates are much stronger, because they depend on SRF, but not on N. Further, if one tries to use the proof technique used in [13] to generalize the noiseless results in [13], [14] to the 2D case, one would need to assume that $\|\mathbf{x}_{2\mathrm{D}}\|_0 < n/2$. This is too restrictive. As we see from Theorem 2, if the signal's support is Rayleigh-regular, $\|\mathbf{x}_{2\mathrm{D}}\|_0$ can be of the order of n^2 in the 2D case.

4 Proofs of main results

4.1 Proof of Theorem 1, Part P

The proof of the theorem is based on the following lemma.

Lemma 1. *Assume that the assumptions of Theorem 1 are satisfied. Set*

$$\mathbf{h} = [h_0 \cdots h_{N-1}]^{\mathsf{T}} = \hat{\mathbf{x}} - \mathbf{x}. \tag{54}$$

and

$$\mathcal{T} = \{l/N : h_l < 0\}. \tag{55}$$

Assume that there exists $\mathbf{q} = [q_0 \cdots q_{N-1}]^{\mathsf{T}} \in \mathbb{R}^N$ *such that* $\mathbf{Pq} = \mathbf{q}$ *and*

$$q_l < -\rho, \quad l/N \in \mathcal{T} \tag{56}$$

$$q_l > \rho, \quad l/N \in \{0, 1/N, \ldots, 1 - 1/N\} \setminus \mathcal{T}. \tag{57}$$

Let $R = \max_{l \in [0\,:\,N-1]} \big|q_l\big|$. *Then,*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \le 2\delta \frac{R}{\rho}. \tag{58}$$

Proof. On the one hand,

$$\begin{aligned}|\langle \mathbf{q}, \mathbf{h}\rangle| &= |\langle \mathbf{P}\mathbf{q}, \mathbf{h}\rangle| = |\langle \mathbf{q}, \mathbf{P}\mathbf{h}\rangle| \\ &\leq \|\mathbf{q}\|_\infty \|\mathbf{P}\mathbf{h}\|_1 \leq \|\mathbf{q}\|_\infty \|\mathbf{P}\mathbf{x} - \mathbf{s} + \mathbf{s} - \mathbf{P}\hat{\mathbf{x}}\|_1 \\ &\leq \|\mathbf{q}\|_\infty \|\mathbf{P}\mathbf{x} - \mathbf{s}\|_1 + \|\mathbf{q}\|_\infty \|\mathbf{s} - \mathbf{P}\hat{\mathbf{x}}\|_1 \leq 2\delta R. \end{aligned} \tag{59}$$

On the other hand,

$$|\langle \mathbf{q}, \mathbf{h}\rangle| = \left| \sum_{l=0}^{N-1} q_l h_l \right| = \sum_{l=0}^{N-1} q_l h_l = \sum_{l=0}^{N-1} |q_l||h_l| \geq \rho \|\mathbf{h}\|_1, \tag{60}$$

where we have used that $\operatorname{sign}(q_l) = \operatorname{sign}(h_l)$ for all $l = 0, \ldots, N-1$. Bound (58) follows by combining (59) and (60). □

4.1.1 Intuition of the proof of Theorem 1

First observe that since $\mathbf{x}, \hat{\mathbf{x}} \geq \mathbf{0}$ it follows that $\mathcal{T}$ in (55) satisfies $\left|\mathcal{T}\right| \leq \|\mathbf{x}\|_0 < n/2$. The idea now is to construct a real trigonometric polynomial of largest frequency f_c

$$q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-\mathrm{i}2\pi k t} \in \mathbb{R} \quad \text{for all} \quad t \tag{61}$$

that satisfies

$$q(t_0) = 0, \quad \text{for all} \quad t_0 \in \mathcal{T} \tag{62}$$

$$q(t) > 0, \quad \text{for all} \quad t \notin \mathcal{T} \tag{63}$$

$$q(t) \leq 1, \quad \text{for all} \quad t \tag{64}$$

and set

$$q_l = q(l/N) - \rho, \quad l = 0, \ldots, N-1, \tag{65}$$

where

$$\rho = \frac{1}{2} \arg\min_{l/N \notin \mathcal{T}} \{q(l/N)\}. \tag{66}$$

Set $\mathbf{q} = [q_0 \cdots q_{N-1}]^\mathsf{T}$ and observe that $\mathbf{q}$ satisfies the conditions of Lemma 1 with $R = 1 - \rho$ and ρ defined in (66), which leads to stability estimate (58).

In order to get a tight bound, $q(t)$ should be chosen such that the RHS of (66) is as large as possible. A classical approach to constructing $q(t)$ that satisfies (62)–(64) is to set

$$q(t) = \prod_{t_0 \in \mathcal{T}} \frac{1}{2} \left[\cos(2\pi(t + 1/2 - t_0)) + 1\right]. \tag{67}$$

This approach works whenever $\left|\mathcal{T}\right| < n/2$ with no further assumptions on the regularity of $\mathcal{T}$. It was used in [13, 14]. The problem with the approach is that $q(t)$ in (67) grows extremely slowly around its zeros, making the RHS of (66) very small, which translates into highly suboptimal stability estimates. To demonstrate this, assume that $\mathcal{T} = \{0\}$, i.e., $\left|\mathcal{T}\right| = 1$. Then (see Figure 3)

$$q(t) = \frac{1}{2} \left[\cos(2\pi(t + 1/2)) + 1\right] \tag{68}$$

so that

$$q(1/N) \leq \frac{\pi^2}{N^2} \tag{69}$$

and $\rho \leq \frac{\pi^2}{2N^2}$. In this case $R = 1$. Plugging this into (58) we get an estimate no better than

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \delta \frac{2}{\pi^2} N^2. \tag{70}$$

This is weak. In the case when $\mathbf{x}$ has one spike, the separation condition $\mathcal{T} \in \mathcal{R}^{\text{idx}}(3.74, 1)$ of [7] is trivially satisfied. The results of [7] guarantee (if l1 minimization reconstruction algorithm is used)

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq c\delta N^2/n^2 = c\delta \text{SRF}^2, \tag{71}$$

where c is a numerical constant. In the interesting case $N, n \to \infty$ with $\text{SRF} = N/n$ fixed, (71) is much stronger than (70). The reason why [7] provides stability guarantees stronger than (70) is that the trigonometric polynomial $q(t)$ constructed in [7] grows around its zeros much faster than $q(t)$ in (68). We review the behavior of $q(t)$ constructed in [7] in Lemma 2 below and illustrate the difference between this trigonometric polynomial and the one in (68) in Figure 3. Based on the results of [7], we then present a novel construction for $q(t)$. Crucially, the novel construction does not rely on the minimal separation condition $\mathcal{T} \in \mathcal{R}^{\text{idx}}(3.74, 1)$ needed in [7] and works for all signals with Rayleigh-regular support of the type $\mathcal{T} \in \mathcal{R}^{\text{idx}}(3.74r, r)$. At the same time, the new polynomial $q(t)$ grows fast around its zeros, which allows us to derive strong stability guarantees.

4.1.2 Main building block: $q(t)$ under separation

The following lemma is an immediate consequence of [7, Lm. 2.5], adapted to the case of real-valued signals as explained in [7, Sec. 2.5].

Lemma 2. *Fix N, n. Assume $\mathcal{T} \in \mathcal{R}^{\text{idx}}(3.74, 1; N, n)$. Set, as before, $f_c = (n-1)/2, \lambda_c = 1/f_c$. Assume $f_c \geq 128$. Then, there exist a real trigonometric polynomial $q(t; N, n) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-\text{i}2\pi kt}$ such that*

$$q(t) \leq 1, \quad \textit{for all} \quad t \tag{72}$$

$$q(t_0) = 0, \quad \textit{for all} \quad t_0 \in \mathcal{T} \tag{73}$$

$$q(t) \geq \phi(t), \quad \textit{for all} \quad t, \tag{74}$$

where (see Figure 3)

$$\phi(t) = \begin{cases} c_1 f_c^2 (t_0 - t)^2, \ \textit{for all } t \textit{ s.t. there exists } t_0 \in \mathcal{T} \textit{ s.t. } \left|t - t_0\right| \leq c_2 \lambda_c \\ c_1 f_c^2 (c_2 \lambda_c)^2 = c_1 c_2^2, \ \textit{for all } t \in \left\{\tilde{t} : \left|\tilde{t} - t_0\right| \geq c_2 \lambda_c \ \forall \ t_0 \in \mathcal{T}\right\}, \end{cases} \tag{75}$$

and $c_1 = 0.029$, $c_2 = 0.17$.

As discussed in the previous section, the significance of the lemma is that when the elements in $\mathcal{T}$ are well-separated, it guarantees the existence of a trigonometric polynomial $q(t; N, n)$ that satisfies (62) and (63) and grows much faster around its zeros than the trigonometric polynomial in (67) [see Figure 3]. In fact, even more is true: the growth of $q(t)$ around its zeros guaranteed by the first line in (75) is optimal up to a constant. Suppose we tried to construct $q(t)$, a real nonnegative trigonometric polynomial with highest frequency f_c of the form (61) s.t. $q(t) \leq 1$ for all t, that

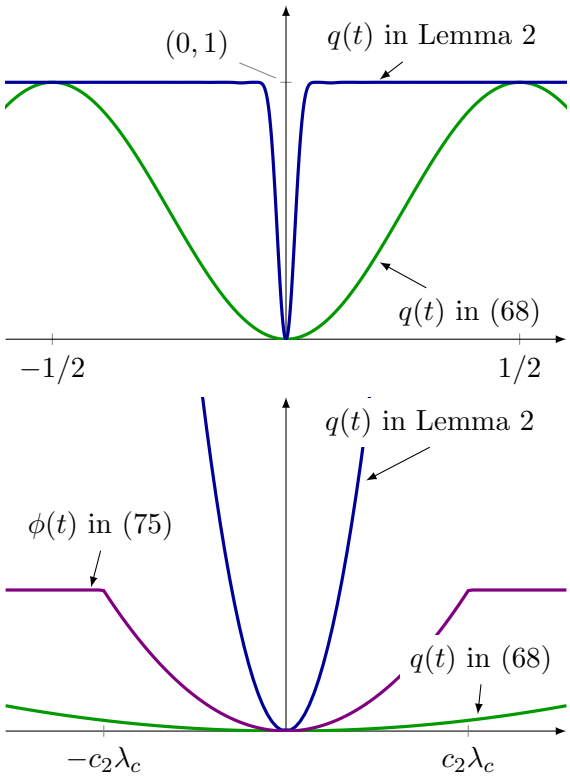

Figure 3: Comparison of trigonometric polynomials (two different zoom levels): trigonometric polynomial constructed in [7] (see Lemma 2) bounces off zero much faster than the one used in [13, 14]. The analytic lower bound $\phi(t)$ defined in (75) is also displayed.

grows around its zero as fast as possible. How fast could it possibly grow? Since the polynomial $q(t)$ is a superposition of harmonic functions, it could not grow faster that the pure harmonic function with the highest available frequency (the harmonic function should be normalized to be positive and never exceed one). Hence, we cannot hope for faster growth than that of

$$\frac{1}{2}\left[\cos(2\pi f_c t + 1/2)) + 1\right] \approx \pi^2 f_c^2 t^2 \quad \text{for small } t. \tag{76}$$

Comparing (76) to (75), we see that Lemma 2 provides a construction that is optimal up to a numerical constant.

We now show how to extend the construction in Lemma 2 to the case when the elements of $\mathcal{T}$ are not necessarily well-separated, but $\mathcal{T}$ is Rayleigh-regular. Together with Lemma 1, this will prove Theorem 1. The proof below is illustrated on Figure 4 and we encourage the reader to consult the figure while following the proof.

4.1.3 Construction of $q(t)$ without separation

Let $\mathbf{x} \in \mathcal{R}_+(3.74r, r)$. Assume $\mathbf{h}$ is given by (54) and $\mathcal{T} = \{t_k\}_{k=1}^S$ is defined in (55) and is ordered: $t_k < t_l$ for $k < l$. Since h_l can only be negative for $l/N \in \mathrm{supp}(\mathbf{x})$, it follows $\mathcal{T} \in \mathcal{R}^{\mathrm{idx}}(3.74r, r)$.

We now represent $\mathcal{T}$ as a union of non-intersecting subsets $\mathcal{T} = \cup_{i=1}^r \mathcal{T}_i$, where $\mathcal{T}_i = \{t_{rk+i}\}_{k=0}^{S/r-1}$, $i = 1, \ldots, r$. Since $\mathcal{T} \in \mathcal{R}^{\mathrm{idx}}(3.74r, r)$, we have that $\mathcal{T}_i \in \mathcal{R}^{\mathrm{idx}}(3.74r, 1)$, $i = 1, \ldots, r$. Observe, by

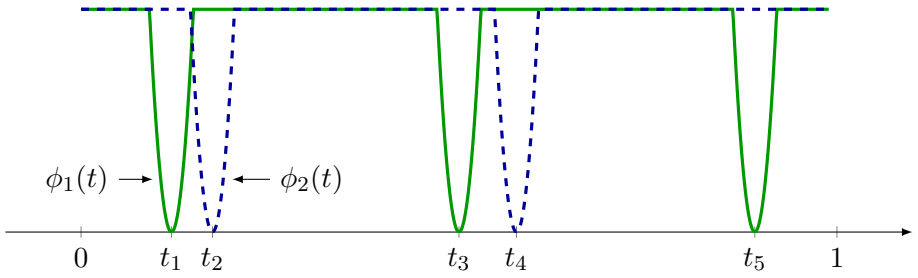

Figure 4: Illustration of the proof of Theorem 1 for $r = 2$; $\mathcal{T}_1 = \{t_1, t_2, t_3\}$; $\mathcal{T}_2 = \{t_2, t_4\}$. The trigonometric polynomials $q_1(t)$, $q_2(t)$ satisfy $q_1(t) = 0$ for all $t \in \mathcal{T}_1$ and $q_2(t) = 0$ for all $t \in \mathcal{T}_2$; they are not displayed. The lower bounds $\phi_1(t)$ and $\phi_2(t)$ defined in (81) are depicted.

rescaling,

$$\mathcal{R}^{\text{idx}}(3.74r, 1; N, n) = \mathcal{R}^{\text{idx}}(3.74, 1; N, \tilde{n}) \tag{77}$$

where $\tilde{n} = (n-1)/r+1$. Set $\tilde{f}_c = (\tilde{n}-1)/2 = (n-1)/(2r) = f_c/r$ and $\tilde{\lambda}_c = 1/\tilde{f}_c = r/f_c$. By Lemma 2, there exist trigonometric polynomials $q_i(t; N, \tilde{n}) = \sum_{k=-\tilde{f}_c}^{\tilde{f}_c} \hat{q}_{ik} e^{-\mathrm{i}2\pi kt}$, $i = 0, \ldots, r-1$, such that

$$q_i(t) \le 1, \quad \text{for all} \quad t \tag{78}$$
$$q_i(t_0) = 0, \quad \text{for all} \quad t_0 \in \mathcal{T}_i \tag{79}$$
$$q_i(t) \ge \phi_i(t), \quad \text{for all} \quad t, \tag{80}$$

where (see Figure 4)

$$\phi_i(t) = \begin{cases} c_1 \tilde{f}_c^{\,2} (t_0 - t)^2, \text{ for all } t \text{ s.t. there exists } t_0 \in \mathcal{T}_i \text{ s.t. } \left|t - t_0\right| \le c_2 \tilde{\lambda}_c \\ c_1 \tilde{f}_c^{\,2} (c_2 \tilde{\lambda}_c)^2 = c_1 c_2^2, \text{ otherwise.} \end{cases} \tag{81}$$

The trigonometric polynomial we seek is obtained by taking the product of q_i:

$$q(t) = \prod_{i=1}^{r} q_i(t; N, \tilde{n}). \tag{82}$$

By construction, $q(t)$ is band-limited, i.e., $q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-\mathrm{i}2\pi kt}$, and it satisfies

$$q(t) \le 1, \quad \text{for all} \quad t \tag{83}$$
$$q(t_0) = 0, \quad \text{for all} \quad t_0 \in \mathcal{T} \tag{84}$$
$$q(t) \ge \prod_{i=1}^{r} \phi_i(t) \tag{85}$$

and $\prod_{i=1}^{r} \phi_i(t)$ can be further lower bounded as follows

$$\prod_{i=1}^{r} \phi_i(t) \ge \begin{cases} c_1^r \tilde{f}_c^{\,2r} (t_0 - t)^{2r}, \text{ for all } t \text{ s.t. there exists } t_0 \in \mathcal{T} \text{ s.t. } \left|t - t_0\right| \le c_2 \tilde{\lambda}_c \\ (c_1 c_2^2)^r, \text{ otherwise.} \end{cases} \tag{86}$$

The assumption SRF $\geq 3/r$ implies SRF $= N/n > 1/(2rc_2) \approx 2.94/r$, which mean $\frac{1}{N} < c_2 r \frac{2}{n}$ so that $\frac{1}{N} < c_2 \tilde{\lambda}_c$. Therefore, from (85) and (86), it follows that

$$\rho = \frac{1}{2} \arg\min_{l/N \notin \mathcal{T}} \{q(l/N)\} \tag{87}$$

$$\geq c_1^r \frac{1}{2} \tilde{f}_c^{\,2r} \frac{1}{N^{2r}} \tag{88}$$

$$= c_1^r \frac{1}{2r^{2r}} f_c^{2r} \frac{1}{N^{2r}} \tag{89}$$

$$= c_1^r \frac{1}{2(2r)^{2r}} \left(\frac{n-1}{N} \right)^{2r}. \tag{90}$$

Plugging this into (58) we obtain the final result

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \underbrace{4 c_3^r r^{2r}}_{C_1(r)} \left(\frac{N}{n-1} \right)^{2r} \delta, \tag{91}$$

where $c_3 = 4/c_1 = 67.79$. This completes the proof. □

4.1.4 Possible Improvement

In the statement of Theorem 1 we assumed $\mathbf{x} \in \mathcal{R}_+(3.74r, r)$. The constant 3.74 comes from the fact that our construction is built on top of Lemma 2 borrowed from [7, Sec. 2.5]. If the constant 3.74 in Lemma 2 is reduced, all our results automatically improve also. Carlos Fernandez-Granda privately shared with us that it is possible to substitute 3.74 by 2.88 in Lemma 2. Any future improvement in Lemma 2 is accommodated by our technique with no modifications.

4.1.5 Possible Extension

The result in Theorem 1, Part P, is for positive signals $\mathbf{x} \in \mathcal{R}_+(3.74r, r)$. As we have already discussed, if the signal is complex and satisfies the minimum separation constraint, $\mathbf{x} \in \mathcal{R}_+(3.74, 1)$, according to [7], it can be recovered by l1-minimization (l1). The proof technique developed in this paper can be used to generalize both results as follows. Consider a signal that consists of clusters of spikes as depicted in Figure 5. The signal does not satisfy the minimum separation constraint. The spikes within each cluster have the same sign, but across clusters the signs can be different. Our proof technique can be used to show that if the clusters are sufficiently separated, then the signal can be recovered stably by l1-minimization (l1). Details of this derivation are omitted.

Figure 5: Signal with clustered spikes that have the same sign within clusters but different signs across clusters

4.2 Proof of Theorem 1, Part Q

To Prove Theorem 1, Part Q, we need to modify Lemma 1 as follows.

Lemma 3. *Assume that the assumptions of Theorem 1, Part Q, are satisfied. Set*

$$\mathbf{h} = [h_0 \cdots h_{N-1}]^{\mathsf{T}} = \hat{\mathbf{x}} - \mathbf{x}. \tag{92}$$

and

$$\mathcal{T} = \{l \in [0 : N-1] : h_l < 0\}. \tag{93}$$

Assume that there exists $\mathbf{q} = [q_0 \cdots q_{N-1}]^{\mathsf{T}} \in \mathbb{R}^N$ *such that* $\mathbf{Pq} = \mathbf{q}$ *and*

$$q_l < -\rho, \quad l \in \mathcal{T} \tag{94}$$

$$q_l > \rho, \quad l \in [0 : N-1] \setminus \mathcal{T}. \tag{95}$$

Let $R = \max_{l \in [0\,:\,N-1]} \left|q_l\right|$. *Then,*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq 2C(\alpha)\delta\frac{R}{\rho}. \tag{96}$$

Proof. First we need to upper-bound $\|\mathbf{Rz}\|_1$. Note that

$$\|\mathbf{Rz}\|_1 \leq \|\mathbf{R}\|_{1,op}\|\mathbf{z}\|_1. \tag{97}$$

By definition, $\mathbf{R} = [\mathbf{r}_0 \cdots \mathbf{r}_{N-1}]$ is a circulant matrix, and, hence, $\|\mathbf{R}\|_{1,op} = \|\mathbf{r}_0\|_1$. Further, by properties of circulant matrices,

$$\hat{\mathbf{r}} = \sqrt{N}\mathbf{F}\mathbf{r}_0 \tag{98}$$

or, equivalently,

$$\mathbf{r}_0 = \frac{1}{\sqrt{N}}\mathbf{F}^{\mathsf{H}}\hat{\mathbf{r}}. \tag{99}$$

so that

$$\|\mathbf{R}\|_{1,op} = \frac{1}{\sqrt{N}}\|\mathbf{F}^{\mathsf{H}}\hat{\mathbf{r}}\|_1. \tag{100}$$

Since, $\hat{\mathbf{r}}$ contains samples of a continuous nonnegative function, we expect that $\frac{1}{\sqrt{N}}\|\mathbf{F}^{\mathsf{H}}\hat{\mathbf{r}}\|_1$ is bounded above by a constant that depends on α only. Indeed, using standard methods from Fourier analysis we show in Appendix C that for all N and all SRF

$$\frac{1}{\sqrt{N}}\|\mathbf{F}^{\mathsf{H}}\hat{\mathbf{r}}\|_1 \leq C(\alpha), \tag{101}$$

where $C(\alpha)$ is defined in (33). Next, we proceed as in the proof of Lemma 1. First note that $\mathbf{Pq} = \mathbf{q}$ implies that $\mathbf{Tq} = \mathbf{q}$. On the one hand,

$$\begin{aligned} |\langle \mathbf{q}, \mathbf{h}\rangle| &= |\langle \mathbf{Tq}, \mathbf{h}\rangle| = |\langle \mathbf{q}, \mathbf{Th}\rangle| \\ &\leq \|\mathbf{q}\|_\infty \|\mathbf{Th}\|_1 \leq \|\mathbf{q}\|_\infty \|\mathbf{Tx} - \mathbf{Rs} + \mathbf{Rs} - \mathbf{T}\hat{\mathbf{x}}\|_1 \\ &\leq \|\mathbf{q}\|_\infty \underbrace{\|\mathbf{Tx} - \mathbf{Rs}\|_1}_{\|\mathbf{Rz}\|_1 \leq C(\alpha)\delta} + \|\mathbf{q}\|_\infty \underbrace{\|\mathbf{Rs} - \mathbf{T}\hat{\mathbf{x}}\|_1}_{\leq C(\alpha)\delta} \leq 2C(\alpha)\delta R. \end{aligned} \tag{102}$$

On the other hand,

$$|\langle \mathbf{q}, \mathbf{h}\rangle| = \sum_{l=0}^{N-1} |q_l||h_l| \geq \rho\|\mathbf{h}\|_1. \tag{103}$$

Bound (96) follows by combining (102) and (103). □

Finally, following the same steps as in Section 4.1 we obtain (37) with

$$C_1(r, \alpha) = C_1(r)C(\alpha)\left(\frac{1}{\alpha}\right)^{2r}. \tag{104}$$

4.3 Remarks on the proof of Theorem 2

The proof of Theorem 2 closely mimics that of Theorem 1 and is given in Appendix D. To the best of our knowledge, Theorem 2 is the first result, in the noisy and in the noiseless setting, that shows that super-resolution problem can be solved in the 2D case via convex optimization when the signal is nonnegative, even if the minimum separation condition is not satisfied. It is instructive to discuss this point in details as the discussion reveals interesting insights about the super-resolution problem in multi-dimensional case.

Suppose one would like to obtain a noiseless result, similar to the one in [13], [14] and show that super-resolution problem can be solved via convex optimization in the noiseless setting when the signal is nonnegative. Following our discussion in Section 4.1.1 one could take a 2D version of the trigonometric polynomial in (68),

$$\frac{1}{4}\left[\cos(2\pi(t_1+1/2))+1\right]\left[\cos(2\pi(t_2+1/2))+1\right] \tag{105}$$

and then form a product of such terms to build a low-frequency trigonometric polynomial $q_{2\mathrm{D}}(t_1,t_2)$ as done in (67). What is the largest number of terms the product in (67) could contain in the 2D case? To be able to guarantee that $q_{2\mathrm{D}}(t_1,t_2)$ has frequency no larger than f_c in both variables, one can have no more that $n/2$ terms of the form (105). Hence, the largest sparsity allowed to be able to guarantee that the convex method works would need to satisfy $\|\mathbf{x}_{2\mathrm{D}}\|_0 < n/2$. This is discouraging: recall that in the 2D case we have n^2 observations, so we would like to be able to recover on the order of n^2 spikes. It is easy to see that $\|\mathbf{x}_{2\mathrm{D}}\|_0 < n/2$ is a tight bound in the worst case: think about the situation where the signal's support is located along a line parallel to coordinate axis in the domain $\mathbb{T}\times\mathbb{T}$. In this case, even though we have n^2 observations, the problem is essentially one-dimensional with n observations and no more that $n/2$ can possibly be resolved. But what happens in the typical situation where the spikes are Rayleigh-regularly spread over the domain $\mathbb{T}\times\mathbb{T}$? In this case, Theorem 2 shows that $\|\mathbf{x}_{2\mathrm{D}}\|_0$ can be as large as $\frac{(n-1)^2}{4.76^2r^2}r \approx \frac{n^2}{4.76^2r}$, i.e. $\|\mathbf{x}_{2\mathrm{D}}\|_0$ grows linearly with the number of observations, n^2, when r is small (i.e., signal's support is sufficiently regular). In the worst case, when signal's support is not regular, $r\sim n$, we are back to the bound $\|\mathbf{x}_{2\mathrm{D}}\|_0 \le \frac{n^2}{4.76^2r} \sim n$. The key reason why we are able to obtain a significantly better result under Rayleigh-regularity than what one would get by taking the approach of [13], [14] is that we only have r terms in the product in (199) as compared to order n terms in (67). In turn, each term in (199) interpolates $\sim \frac{n^2}{r^2}$ points of the signal's support; in (67), each term interpolates only one such point.

5 Conclusion

In this paper we have seen that when the signal is positive and Rayleigh-regular, the super-resolution problem can be solved stably via convex optimization. A simple convex optimization reconstruction algorithm achieves near optimal worst-case stability. In this paper we made the simplifying assumption that the signal is supported on a discrete grid. We refer the interested reader to the companion paper [21] that covers the continuous setup, where the signal is not restricted to the grid.

A widely open research problem is this. Suppose a 1D signal is complex-valued and the noise is arbitrary. It is shown in [12] that if the signal belongs to $\mathcal{R}(4r,r)$, then stable super-resolution is possible via exhaustive search in this class. If the signal belongs to $\mathcal{R}(4,1)$, it is shown in [7] that stable super-resolution can be achieved via convex l1-minimization algorithm. Is there a

computationally feasible algorithm that achieves stable super-resolution for signals in $\mathcal{R}(4r, r)$ with $r > 1$? If no such algorithm is found, is it possible to show that this problem is fundamentally computationally difficult in some sense?

Acknowledgments

VM was supported by the Swiss National Science Foundation fellowship for advanced researchers under grant PA00P2_139678.

A Proof of Theorem 3

The proof uses the idea of [13, Th. 4.1] with the following important difference. In [13, Th. 4.1], authors provide a lower bound for the MC defined as $\mathrm{MC}[\mathcal{C}] \triangleq \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{C}} \frac{\|\mathbf{x}_1-\mathbf{x}_2\|_1}{\|\mathbf{L}(\mathbf{x}_1-\mathbf{x}_2)\|_2}$. To be able to assess the accuracy of (42) we need to find a lower bound for MC defined as $\mathrm{MC}[\mathcal{C}, \mathbf{Q}] \triangleq \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{C}} \frac{\|\mathbf{x}_1-\mathbf{x}_2\|_1}{\|\mathbf{Q}(\mathbf{x}_1-\mathbf{x}_2)\|_1}$.

A.1 Intuition of the proof of Theorem 3

The best way to understand the idea behind Theorem 3 is to consider the following back-of-the-envelope calculation. Let

$$\mathbf{h} = [h_0 \cdots h_{N-1}]^{\mathsf{T}} = \mathbf{x} - \tilde{\mathbf{x}}, \tag{106}$$

$$\mathbf{s} = [s_0 \cdots s_{N-1}]^{\mathsf{T}} = \mathbf{Q}\mathbf{h}. \tag{107}$$

Then we see that

$$s_m = \frac{1}{(f_c+1)N} \sum_{k=-f_c}^{f_c} e^{\mathrm{i}2\pi mk/N} (f_c + 1 - |k|) \sum_{l=0}^{N-1} e^{-\mathrm{i}2\pi lk/N} h_l \tag{108}$$

$$= \frac{1}{(f_c+1)N} \sum_{l=0}^{N-1} \sum_{k=-f_c}^{f_c} (f_c + 1 - |k|) e^{\mathrm{i}2\pi(m-l)k/N} h_l \tag{109}$$

$$= \frac{1}{(f_c+1)N} \sum_{l=0}^{N-1} \left(\sum_{k=-f_c/2}^{f_c/2} e^{\mathrm{i}2\pi(m-l)k/N} \right)^2 h_l \tag{110}$$

$$= \sum_{l=0}^{N-1} g_{\mathrm{low}}\left(\frac{m-l}{N}\right) h_l, \tag{111}$$

where

$$g_{\mathrm{low}}(t) = \frac{1}{(f_c+1)N} \left(\sum_{k=-f_c/2}^{f_c/2} e^{\mathrm{i}2\pi tk} \right)^2 = \frac{1}{(1+f_c)N} \left(\frac{\sin((1+f_c)\pi t)}{\sin(\pi t)} \right)^2 \tag{112}$$

is the Fejér kernel. The idea is to construct $\mathbf{h}$, $\|\mathbf{h}\|_1 = 1$, with at most $2r$ nonzero elements, in such a way that for each m the terms in the sum

$$s_m = \sum_{l=0}^{N-1} g_{\mathrm{low}}\left(\frac{m-l}{N}\right) h_l \tag{113}$$

cancel each other out as much as possible, so that $\left|s_m\right|$ is small for all m. One way to construct $\mathbf{h}$ with these properties in a systematic way is to set

$$x_k = \begin{cases} \frac{1}{2^{2r-1}}\binom{2r-1}{k}, \ k = 2l, \ l = 0, \ldots, r-1 \\ 0, \ \text{otherwise} \end{cases}$$
$$\tilde{x}_k = \begin{cases} \frac{1}{2^{2r-1}}\binom{2r-1}{k}, \ k = 2l+1, \ l = 0, \ldots, r-1 \\ 0, \ \text{otherwise} \end{cases}, \tag{114}$$

so that $\|\mathbf{x}\|_0 = \|\tilde{\mathbf{x}}\|_0 = r$. Set, for convenience, $\omega = 2r-1$ and note that s_m becomes

$$s_m = \sum_{l=0}^{\omega} \underbrace{(-1)^l \frac{1}{2^\omega}\binom{\omega}{l}}_{h_l} g_{\text{low}}\left(\frac{m-\omega}{N} + \frac{\omega - l}{N}\right) \tag{115}$$
$$= \frac{1}{2^\omega}\Delta^\omega_{1/N}[g_{\text{low}}]\left(\frac{m-\omega}{N}\right), \tag{116}$$

where

$$\Delta^\omega_\delta[g_{\text{low}}](t) = \sum_{l=0}^{\omega}(-1)^l\binom{\omega}{l} g_{\text{low}}(t + (\omega - l)\delta)\,. \tag{117}$$

is the finite-difference operator of order ω, applied to the kernel $g_{\text{low}}(\cdot)$. Furthermore,

$$\|\mathbf{h}\|_1 = \|\mathbf{x} - \tilde{\mathbf{x}}\|_1 = \frac{1}{2^\omega}\sum_{l=0}^{\omega}\binom{\omega}{l} = 1, \tag{118}$$

as required. We next estimate $\|\mathbf{s}\|_1$ by a back-of-the-envelope calculation that will be justified in Section A.2. For large N and large SRF,

$$\Delta^\omega_{1/N}[g_{\text{low}}](t) \approx \frac{1}{N^\omega}\frac{d^\omega g_{\text{low}}}{dt^\omega}(t). \tag{119}$$

Therefore, for large N and large SRF,

$$\begin{aligned} \|\mathbf{s}\|_1 &= \sum_{m=0}^{N-1}|s_m| \\ &\approx \frac{1}{2^\omega}\frac{1}{N^{\omega-1}}\frac{1}{N}\sum_{m=0}^{N-1}\left|\frac{d^\omega g_{\text{low}}}{dt^\omega}\left(\frac{m}{N} - \frac{\omega}{N}\right)\right| \\ &\approx \frac{1}{2^\omega}\frac{1}{N^{\omega-1}}\int_0^1\left|\frac{d^\omega g_{\text{low}}}{dt^\omega}\left(t - \frac{\omega}{N}\right)\right|dt \\ &= \frac{1}{2^\omega}\frac{1}{N^{\omega-1}}\int_{-1/2}^{1/2}\left|\frac{d^\omega g_{\text{low}}}{dt^\omega}(t)\right|dt. \end{aligned} \tag{120}$$

For large n, and $t \in (-1/2, 1/2)$,

$$g_{\text{low}}(t) = \frac{1}{(f_c+1)N}\left(\frac{\sin((f_c+1)\pi t)}{\sin(\pi t)}\right)^2 = \frac{(f_c+1)}{N}\left(\frac{\sin((f_c+1)\pi t)}{(f_c+1)\pi t}\right)^2 = \frac{(f_c+1)}{N}g((f_c+1)t) \tag{121}$$

where

$$g(t) = \left(\frac{\sin(\pi t)}{\pi t}\right)^2, \tag{122}$$

which implies

$$\frac{d^\omega g_{\text{low}}}{dt^\omega}(t) \approx \frac{(f_c+1)}{N}\frac{d^\omega g((f_c+1)t)}{dt^\omega}(t) \tag{123}$$

$$= \frac{(f_c+1)^{\omega+1}}{N}\frac{d^\omega g((f_c+1)t)}{d((f_c+1)t)^\omega}(t) \tag{124}$$

$$= \frac{(f_c+1)^{\omega+1}}{N}\frac{d^\omega g}{dt^\omega}((f_c+1)t). \tag{125}$$

Therefore,

$$\begin{aligned}\|\mathbf{s}\|_1 &\approx \frac{1}{2^\omega}\frac{(f_c+1)^{\omega+1}}{N^\omega}\int_{-1/2}^{1/2}\left|\frac{d^\omega g}{dt^\omega}((f_c+1)t)\right|dt \\ &\le \frac{1}{2^\omega}\frac{(f_c+1)^{\omega}}{N^\omega}\int_{-\infty}^{\infty}\left|\frac{d^\omega g}{dt^\omega}((f_c+1)t)\right|d((f_c+1)t) \\ &\approx \frac{1}{\text{SRF}^{2r-1}}\frac{1}{c_L(r)}\end{aligned} \tag{126}$$

where

$$c_L(r) \triangleq \left(\frac{1}{4^{2r-1}}\int_{-\infty}^{\infty}\left|\frac{d^{2r-1}g}{dt^{2r-1}}(t)\right|dt\right)^{-1}. \tag{127}$$

The bound (49) now follows from (118) and (126) by the definition of the MC.

A.2 Proof of Theorem 3

Using the intuition developed in the previous section we now formally prove Theorem 3.

Proof. Begin by writing s_m as an inverse DFT of its spectrum:

$$s_m = \frac{1}{N}\sum_{k=-f_c}^{f_c} e^{\mathrm{i}2\pi mk/N} q\left(\frac{k}{f_c+1}\right) p_\omega\left(\frac{k}{N}\right) \tag{128}$$

where

$$q(f) = \begin{cases} 1-\left|f\right|, & f\in[-1,1] \\ 0, & \text{otherwise.}\end{cases} \tag{129}$$

and

$$p_\omega(f) = \sum_{l=0}^{\omega} e^{-\mathrm{i}2\pi lf} h_l. \tag{130}$$

Now let

$$t_m = \int_{-(f_c+1)/N}^{(f_c+1)/N} e^{\mathrm{i}2\pi mf} q\left(f\frac{N}{f_c+1}\right) p_\omega(f)df. \tag{131}$$

It is not difficult to see that for all N

$$\sum_{m=-N/2+1}^{N/2} |t_m - s_m| \le \sum_{|m|\ge N/2} |t_m|, \tag{132}$$

and therefore, since the series $\sum_{m=-\infty}^{\infty} \left|t_m\right|$ converges, we conclude

$$\|\mathbf{s}\|_1 = \sum_{m=0}^{N-1} |s_m| = \sum_{m=-N/2+1}^{N/2} |s_m| \to \sum_{m=-\infty}^{\infty} |t_m| \tag{133}$$

when $N, n \to \infty$ with $N/n = \text{SRF}$ fixed. Using the fact that $q(f) = 0$ for $\left|f\right| > 1/2$ and changing variables in the integral in (131) we can write

$$t_m = \int_{-(f_c+1)/N}^{(f_c+1)/N} e^{\mathrm{i}2\pi m f} q\left(\frac{fN}{f_c+1}\right) p_\omega(f) df \tag{134}$$

$$= \frac{f_c+1}{N} \int_{-1}^{1} e^{\mathrm{i}2\pi (m(f_c+1)/N) f} q(f)\, p_\omega\left(\frac{f(f_c+1)}{N}\right) df. \tag{135}$$

We conclude, when $N, n \to \infty$ with $N/n = \text{SRF}$ fixed

$$\|\mathbf{s}\|_1 \to \left(\frac{1}{\text{SRF}}\right)^{\omega} \frac{1}{\chi(r, \text{SRF})} \tag{136}$$

where

$$\frac{1}{\chi(r,\eta)} \triangleq \frac{1}{2^\omega} \frac{1}{2\eta} \sum_{m=-\infty}^{\infty} \left| \int_{-1}^{1} e^{\mathrm{i}2\pi (m/(2\eta)) f} q(f)\, (2\eta)^{2r-1} p_{2r-1}(f/(2\eta)) df \right|. \tag{137}$$

Since the finite difference operator converges to the derivative operator as $\delta \to 0$:

$$\frac{1}{\delta^\omega} \Delta_\delta^\omega [g_{\text{low}}](\cdot) \to \frac{d^\omega(\cdot)}{dt^\omega}, \quad \delta \to 0, \tag{138}$$

it follows that for every fixed f

$$\eta^\omega p_\omega(f/\eta) \to \frac{1}{2^\omega} (\mathrm{i}2\pi f)^\omega, \quad \eta \to \infty. \tag{139}$$

Therefore, when $\text{SRF} \to \infty$,

$$\chi(r, \text{SRF}) \to c_L(r) \tag{140}$$

where

$$c_L(r) \triangleq 2^{2r-1} \left(\pi^{2r-1} \int_{-\infty}^{\infty} \left| \int_{-1}^{1} e^{\mathrm{i}2\pi t f} q(f)\, f^{2r-1} df \right| dt \right)^{-1}. \tag{141}$$

Direct numerical computation reveals

$$c_L(r) \geq \begin{cases} 1.66, & r = 1 \\ 1.44, & r = 2 \\ 0.92, & r = 3 \\ 0.48, & r = 4 \\ 0.24, & r = 5. \end{cases} \tag{142}$$

□

B Coherent Optics

When the illumination is perfectly coherent, the time-varying phasor amplitudes across the object plane differ only by complex constants so that we can write

$$\Phi(\mathbf{w},\tau) = \Phi(\mathbf{w})\frac{\Phi(0,\tau)}{\sqrt{\left\langle \left|\Phi(0,\tau)\right|^2 \right\rangle}}. \tag{143}$$

Plugging this into (4) we obtain

$$\tilde{s}_{\mathrm{coh}}(\mathbf{v}) = \left| \kappa_1 \int h(\mathbf{v}-\mathbf{w})\Phi(\mathbf{w})d\mathbf{w} \right|^2. \tag{144}$$

We see that the coherent imaging system is described by a linear model (1) in terms of the phasor quantities that cannot be directly observed. The directly observable received intensity, $\tilde{s}_{\mathrm{coh}}(\mathbf{v})$, is a nonlinear (quadratic) function (144) of the signal $\Phi(\mathbf{w})$.

C Proof of (101)

We will use the following lemma.

Lemma 4. *Assume $\hat{u}_l$ is a discrete periodic signal with period N. Let*

$$\hat{v}_l = \hat{u}_l - \hat{u}_{l-1} \quad \textit{for all} \quad l \tag{145}$$

$$\hat{w}_l = \hat{v}_l - \hat{v}_{l-1} \quad \textit{for all} \quad l \tag{146}$$

be the first and second differences of $\hat{u}_l$, respectively. Let u_k, v_k, w_k denote N-periodic sequences of inverse DFT coefficients of $\hat{u}_l, \hat{v}_l, \hat{w}_l$, respectively. For example,

$$u_k = \frac{1}{\sqrt{N}} \sum_{l=-N/2+1}^{N/2} \hat{u}_l e^{2\pi \mathrm{i} l k/N} \quad \textit{for all} \quad k. \tag{147}$$

Assume that

$$\sum_{k=-N/2+1}^{N/2} |\hat{w}_k| \le A \tag{148}$$

Then,

$$|u_k| \le \frac{1}{\sqrt{N}} \frac{A}{2-2\cos(k2\pi/N)} \quad \textit{for all} \quad k \neq 0 \mod N. \tag{149}$$

Proof. By the theorem about the DFT of first differences [22, p.223],

$$w_k = \left(1 - e^{2\pi \mathrm{i} k/N}\right)^2 u_k \quad \text{for all} \quad k \tag{150}$$

and, consequently,

$$u_k = \frac{1}{\left(1 - e^{2\pi \mathrm{i} k/N}\right)^2} w_k \quad \text{for all} \quad k \neq 0, \pm N, \pm 2N, \ldots. \tag{151}$$

Next, observe that

$$|w_k| \leq \frac{1}{\sqrt{N}} \sum_{l=-N/2+1}^{N/2} |\hat{w}_l| \leq \frac{A}{\sqrt{N}} \tag{152}$$

The proof follows by substituting (152) into (151) and using that

$$\left| \left(1 - e^{2\pi \mathrm{i} k/N}\right)^2 \right| = \left(1 - e^{-2\pi \mathrm{i} k/N}\right)\left(1 - e^{2\pi \mathrm{i} k/N}\right) \tag{153}$$

$$= 2 - 2\cos(2\pi k/N)\,. \tag{154}$$

□

Next, set $\hat{u}_l = \hat{r}_l$, continued periodically with period N, and define $\hat{v}_l, \hat{w}_l$ as in (145) and (146). Observe the following facts:

$$\hat{v}_k = 0, \quad k \in [-N/2+1 : -\beta f_c] \tag{155}$$

$$\hat{v}_k = (f_c+1)|a|, \quad k \in [-f_c+1 : -\alpha f_c] \tag{156}$$

$$\hat{v}_{-\alpha f_c+1} = -\frac{f_c+1}{((1-\alpha)f_c+1)((1-\alpha)f_c+2)} = -\hat{v}_{\alpha f_c} \tag{157}$$

$$\hat{v}_k = -(f_c+1)|a|, \quad k \in [\alpha f_c+1 : f_c] \tag{158}$$

$$\hat{v}_k = 0, \quad k \in [f_c+1 : N/2] \tag{159}$$

and note that $\hat{v}_k$ is monotonically increasing on the intervals

$$[-N/2+1 : -\alpha f_c] \tag{160}$$

$$[-\alpha f_c+1 : \alpha f_c] \tag{161}$$

$$[\alpha f_c+1 : N/2]. \tag{162}$$

From this it immediately follows that

$$\sum_{k=-N/2+1}^{-\alpha f_c} |\hat{w}_k| = \sum_{k=-f_c+1}^{-\alpha f_c} |\hat{w}_k| = \hat{v}_{-\alpha f_c} - \hat{v}_{-f_c} = (1+f_c)|a| \tag{163}$$

$$|\hat{w}_{-\alpha f_c+1}| = \frac{f_c+1}{((1-\alpha)f_c+1)((1-\alpha)f_c+2)} + (f_c+1)|a| = |\hat{w}_{\alpha f_c+1}| \tag{164}$$

$$\sum_{k=-\alpha f_c+2}^{\alpha f_c} |\hat{w}_k| = \hat{v}_{\alpha f_c} - \hat{v}_{-\alpha f_c+1} = 2\frac{f_c+1}{((1-\alpha)f_c+1)((1-\alpha)f_c+2)} \tag{165}$$

$$\sum_{k=\alpha f_c+2}^{N/2} |\hat{w}_k| = \hat{v}_{f_c+1} - \hat{v}_{\alpha f_c+1} = (f_c+1)|a| \tag{166}$$

so that

$$\sum_{k=-N/2+1}^{N/2} |\hat{w}_k| = 4\frac{f_c+1}{((1-\alpha)f_c+1)((1-\alpha)f_c+2)} + 4(f_c+1)|a| \leq A \tag{167}$$

where

$$A \triangleq \frac{1}{(f_c+1)} \underbrace{\frac{4}{(1-\alpha)^2}}_{D} + \frac{2}{f_c} \underbrace{\frac{2}{(1-\alpha)^2}}_{E}\,. \tag{168}$$

Direct calculation reveals

$$|u_k| \leq \frac{1}{\sqrt{N}} \sum_{l=-N/2+1}^{N/2} |\hat{u}_l| \tag{169}$$

$$= \frac{1}{\sqrt{N}} \left[1 + 2 \left(\sum_{l=1}^{\alpha f_c} |\hat{u}_l| + \sum_{l=\alpha f_c+1}^{f_c} |\hat{u}_l| \right) \right] \tag{170}$$

$$\leq \frac{1}{\sqrt{N}} \left[1 + 2 \left(\frac{\alpha f_c}{2} \left(1 + \frac{f_c+1}{(1-\alpha)f_c+1} \right) + \frac{f_c+1}{2((1-\alpha)f_c+2)} - \frac{f_c+1}{((1-\alpha)f_c+1)((1-\alpha)f_c+2)} \right) \right] \tag{171}$$

$$\leq \frac{f_c}{2\sqrt{N}} \underbrace{\left(2\alpha + \frac{2}{1-\alpha} \right)}_{B} \tag{172}$$

for all k. Finally, formula (101) is obtained as follows

$$\frac{1}{\sqrt{N}} \|\mathbf{F}^{\mathsf{H}} \hat{\mathbf{r}}\|_1 = \frac{1}{\sqrt{N}} \sum_{k=-N/2+1}^{N/2} |u_k| \tag{173}$$

$$\leq \frac{2}{\sqrt{N}} \sum_{k=0}^{N/(f_c+1)} |u_k| + \frac{2}{\sqrt{N}} \sum_{k=N/(f_c+1)+1}^{N/2} |u_k| \tag{174}$$

$$\leq \frac{f_c}{N} \frac{N}{f_c+1} B + \frac{2A}{N} \sum_{k=N/(f_c+1)+1}^{N/2} \frac{1}{2 - 2\cos(k2\pi/N)} \tag{175}$$

$$\leq B + \frac{D}{2\pi^2} + \frac{E}{2\pi^2} \frac{2(f_c+1)}{f_c} \tag{176}$$

$$\leq B + \frac{D}{2\pi^2} + 3\frac{E}{2\pi^2}, \tag{177}$$

where we used that

$$\sum_{k=N/(f_c+1)+1}^{N/2} \frac{1}{2 - 2\cos(k2\pi/N)} \leq \int_{N/(f_c+1)}^{N/2} \frac{1}{2 - 2\cos(f2\pi/N)} df \tag{178}$$

$$= \frac{N \cot(\pi/(f_c+1))}{4\pi} \tag{179}$$

$$= \frac{N(f_c+1)}{4\pi} \frac{\cot(\pi/(f_c+1))}{(f_c+1)} \tag{180}$$

and

$$\frac{\cot(\pi/(f_c+1))}{(f_c+1)} \leq \frac{1}{\pi} \quad \text{for all} \quad (f_c+1) \geq 0. \tag{181}$$

D Proof of Theorem 2

We begin by proving Part P. A straightforward modification of Lemma 1 shows that to prove the theorem, we need to construct a 2D dual vector $\mathbf{q}_{2\mathrm{D}} \in \mathbb{C}^{N^2}$ that contains the elements $\{q_{l_1,l_2}\}_{l_1,l_2=1}^{N}$

such that it satisfies

$$\mathbf{P}_{2\mathrm{D}}\mathbf{q}_{2\mathrm{D}} = \mathbf{q}_{2\mathrm{D}} \tag{182}$$

and

$$q_{l_1,l_2} < -\rho, \ \left(\frac{l_1}{N}, \frac{l_2}{N}\right) \in \mathcal{T}$$
$$q_{l_1,l_2} > \rho, \ \left(\frac{l_1}{N}, \frac{l_2}{N}\right) \in \left\{0, \frac{1}{N}, \ldots, 1 - \frac{1}{N}\right\} \times \left\{0, \frac{1}{N}, \ldots, 1 - \frac{1}{N}\right\} \setminus \mathcal{T},$$

where $\mathcal{T}$ is the support of $\mathbf{h}_{2\mathrm{D}} = \hat{\mathbf{x}}_{2\mathrm{D}} - \mathbf{x}_{2\mathrm{D}}$.

Following the approach taken in the 1D case, we now construct a 2D real trigonometric polynomial of the largest frequency f_c in both variables:

$$q(t_1, t_2) = \sum_{k_1=-f_c}^{f_c} \sum_{k_2=-f_c}^{f_c} \hat{q}_{k_1,k_2} e^{-\mathrm{i}2\pi(k_1 t_1 + k_2 t_2)} \in \mathbb{R}^2 \quad \text{for all} \quad (t_1, t_2) \tag{183}$$

that satisfies

$$q(t_1, t_2) = 0, \quad \text{for all} \quad (t_1, t_2) \in \mathcal{T} \tag{184}$$
$$q(t_1, t_2) > 0, \quad \text{for all} \quad (t_1, t_2) \notin \mathcal{T} \tag{185}$$
$$q(t_1, t_2) \leq 1, \quad \text{for all} \quad (t_1, t_2) \tag{186}$$

and set

$$q_{l_1,l_2} = q(l_1/N, l_2/N) - \rho, \quad l_1, l_2 = 0, \ldots, N-1, \tag{187}$$

where

$$\rho = \frac{1}{2} \underset{(l_1/N, l_2/N) \notin \mathcal{T}}{\arg\min} \{q(l_1/N, l_2/N)\}. \tag{188}$$

Just as in the 1D case, to construct $q(t_1, t_2)$ we start with a 2D version of Lemma 2, proven in [7, Prop. C.1], [6, Sec. D.1].

Lemma 5. *Fix N, n. Assume $\mathcal{T} \in \mathcal{R}_{2\mathrm{D}}^{\mathrm{idx}}(4.76, 1; N, n)$. Set, as before, $f_c = (n-1)/2, \lambda_c = 1/f_c$. Assume $f_c \geq 512$. Then, there exist a real trigonometric polynomial*

$$q(\mathbf{t}; N, n) = \sum_{k_1=-f_c}^{f_c} \sum_{k_2=-f_c}^{f_c} \hat{q}_{k_1,k_2} e^{-\mathrm{i}2\pi(k_1 t_1 + k_2 t_2)}, \quad \mathbf{t} = [t_1, t_2]^{\mathsf{T}}, \tag{189}$$

such that

$$q(\mathbf{t}) \leq 1, \quad \textit{for all} \quad \mathbf{t} \tag{190}$$
$$q(\mathbf{t}) = 0, \quad \textit{for all} \quad \mathbf{t} \in \mathcal{T} \tag{191}$$
$$q(\mathbf{t}) \geq \phi(\mathbf{t}), \quad \textit{for all} \quad \mathbf{t}, \tag{192}$$

where

$$\phi(\mathbf{t}) = \begin{cases} c_1 f_c^2 \|\mathbf{t}_0 - \mathbf{t}\|_2^2, \ \textit{for all } \mathbf{t} \ \textit{s.t. there exists } \mathbf{t}_0 \in \mathcal{T} \ \textit{s.t. } \|\mathbf{t} - \mathbf{t}_0\|_\infty \leq c_2 \lambda_c \\ c_3 \ \textit{for all } \mathbf{t} \in \left\{\tilde{\mathbf{t}} : \|\tilde{\mathbf{t}} - \mathbf{t}_0\|_\infty \geq c_2 \lambda_c \ \textit{for all } \mathbf{t}_0 \in \mathcal{T}\right\}, \end{cases} \tag{193}$$

where c_1, c_2, and c_3 are numerical constants.

Since $\mathbf{x}_{2\mathrm{D}} \in \mathcal{C} = \mathcal{R}_{+,2\mathrm{D}}(4.76r, r)$, we can represent $\mathcal{T}$ as a union of non-intersecting subsets $\mathcal{T} = \cup_{i=1}^{r} \mathcal{T}_i$, such that $\mathcal{T}_i \in \mathcal{R}_{2\mathrm{D}}^{\mathrm{idx}}(4.76r, 1)$, $i = 1, \ldots, r$. Observe, by rescaling,

$$\mathcal{R}_{2\mathrm{D}}^{\mathrm{idx}}(4.76r, 1; N, n) = \mathcal{R}_{2\mathrm{D}}^{\mathrm{idx}}(4.76, 1; N, \tilde{n}) \tag{194}$$

where $\tilde{n} = (n-1)/r+1$. Set $\tilde{f}_c = (\tilde{n}-1)/2 = (n-1)/(2r) = f_c/r$ and $\tilde{\lambda_c} = 1/\tilde{f}_c = r/f_c$. By Lemma 5, there exist trigonometric polynomials $q_i(\mathbf{t}; N, \tilde{n}) = \sum_{k_1=-\tilde{f}_c}^{\tilde{f}_c} \sum_{k_2=-\tilde{f}_c}^{\tilde{f}_c} \hat{q}_{i,k_1,k_2} e^{-\mathrm{i}2\pi(k_1 t_1 + k_2 t_2)}$, $i = 0, \ldots, r-1$, such that

$$q_i(\mathbf{t}) \leq 1, \quad \text{for all} \quad \mathbf{t} \tag{195}$$

$$q_i(\mathbf{t}) = 0, \quad \text{for all} \quad \mathbf{t} \in \mathcal{T}_i \tag{196}$$

$$q_i(\mathbf{t}) \geq \phi_i(\mathbf{t}), \quad \text{for all} \quad \mathbf{t}, \tag{197}$$

where

$$\phi_i(\mathbf{t}) = \begin{cases} c_1 \tilde{f}_c^{\,2} \|\mathbf{t}_0 - \mathbf{t}\|_2^2, \text{ for all } \mathbf{t} \text{ s.t. there exists } \mathbf{t}_0 \in \mathcal{T} \text{ s.t. } \|\mathbf{t} - \mathbf{t}_0\|_\infty \leq c_2 \tilde{\lambda_c} \\ c_3 \text{ for all } \mathbf{t} \in \left\{ \tilde{\mathbf{t}} : \|\tilde{\mathbf{t}} - \mathbf{t}_0\|_\infty \geq c_2 \lambda_c \text{ for all } \mathbf{t}_0 \in \mathcal{T} \right\}. \end{cases} \tag{198}$$

The trigonometric polynomial we seek is obtained by taking the product of q_i:

$$q(\mathbf{t}) = \prod_{i=1}^{r} q_i(\mathbf{t}; N, \tilde{n}). \tag{199}$$

By construction, $q(\mathbf{t})$ is band-limited, i.e., $q(\mathbf{t}) = \sum_{k_1=-f_c}^{f_c} \sum_{k_2=-f_c}^{f_c} \hat{q}_{k_1,k_2} e^{-\mathrm{i}2\pi(k_1 t_1 + k_2 t_2)}$, and it satisfies

$$q(\mathbf{t}) \leq 1, \quad \text{for all} \quad \mathbf{t} \tag{200}$$

$$q(\mathbf{t}_0) = 0, \quad \text{for all} \quad \mathbf{t}_0 \in \mathcal{T} \tag{201}$$

$$q(\mathbf{t}) \geq \prod_{i=1}^{r} \phi_i(\mathbf{t}) \tag{202}$$

and $\prod_{i=1}^{r} \phi_i(\mathbf{t})$ can be further lower bounded as follows

$$\prod_{i=1}^{r} \phi_i(\mathbf{t}) \geq \begin{cases} c_1^r \tilde{f}_c^{\,2r} \|\mathbf{t}_0 - \mathbf{t}\|_2^{2r}, \text{ for all } \mathbf{t} \text{ s.t. there exists } \mathbf{t}_0 \in \mathcal{T} \text{ s.t. } \|\mathbf{t} - \mathbf{t}_0\|_\infty \leq c_2 \tilde{\lambda_c} \\ c_3^r, \text{ otherwise.} \end{cases} \tag{203}$$

From (202) and (203), it follows that

$$\rho = \frac{1}{2} \arg\min_{l/N \notin \mathcal{T}} \{q(l/N)\} \tag{204}$$

$$\geq c_4^r \frac{1}{2} \tilde{f}_c^{\,2r} \frac{1}{N^{2r}} \tag{205}$$

$$= c_4^r \frac{1}{2r^{2r}} f_c^{2r} \frac{1}{N^{2r}} \tag{206}$$

$$= c_4^r \frac{1}{2(2r)^{2r}} \left(\frac{n-1}{N} \right)^{2r}, \tag{207}$$

where $c_4 = \min(c_1, c_3)$. Using the modification of Lemma 1 we obtain the final result

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \underbrace{4c_5^r r^{2r}}_{C_2(r)} \left(\frac{N}{n-1}\right)^{2r} \delta, \tag{208}$$

where $c_5 = 4/c_4$.

The proof of Part Q follows the steps outlined in Section 4.2 with slight modifications. Omitted.

References

[1] A. Barabell. Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, volume 8, pages 336–339, 1983.

[2] Dmitry Batenkov and Yosef Yomdin. On the accuracy of solving confluent prony systems. *SIAM Journal on Applied Mathematics*, 73(1):134–154, January 2013.

[3] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, September 2006.

[4] Georges Bienvenu. Influence of the spatial coherence of the background noise on high resolution passive methods. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, volume 4, pages 306–309, 1979.

[5] James A. Cadzow. Signal enhancement—A composite property mapping algorithm. *IEEE Trans. Acoust., Speech, Signal Process.*, 36(1):49–62, January 1988.

[6] Emmanuel J. Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *Fourier analysis and its applications*, 19(6):1229–1254, December 2013.

[7] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, June 2014.

[8] Michael Clark, Lars Eldén, and Petre Stoica. A computationally efficient implementation of 2-D IQML. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, volume 2, pages 1730–1734, November 1997.

[9] Michael P. Clark and Louis L. Scharf. Two-dimensional modal analysis based on maximum likelihood. *IEEE Trans. Signal Process.*, 42(6):1443–1452, June 1994.

[10] H Clergeot, Sara Tressens, and A Ouamri. Performance of high resolution frequencies estimation methods compared to the cramer-rao bounds. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(11):1703–1720, November 1989.

[11] Robert M Dickson, Andrew B Cubitt, Roger Y Tsien, and WE Moerner. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640):355–358, July 1997.

[12] David L. Donoho. Superresolution via sparsity constraints. *SIAM J. Math. Anal.*, 23(5):1309–1331, September 1992.

[13] David L. Donoho, Iain M. Johnstone, Jeffrey C. Hoch, and Alan S. Stern. Maximum entropy and the nearly black object. *Journal of the Royal Statistical Society: Series B*, 54(1):41–81, June 1992.

[14] Jean-Jacque Fuchs. Sparsity and uniqueness for some specific under-determined linear systems. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, volume 5, pages v/729–v/732, 2005.

[15] Josheph W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, 1988.

[16] Carl W. Helstrom. The detection and resolution of optical signals. *IEEE Trans. Inf. Theory*, 10(4):275–287, October 1964.

[17] Yingbo Hua and Tapan K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust., Speech, Signal Process.*, 38(5):814–824, May 1990.

[18] Tao Jiang, Nicholas D. Sidiropoulos, and Jos M. F. ten Berge. Almost-sure identifiability of multidimensional harmonic retrieval. *IEEE Trans. Signal Process.*, 49(9):1849–1859, September 2001.

[19] Thomas A Klar, Stefan Jakobs, Marcus Dyba, Alexander Egner, and Stefan W Hell. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences*, 97(15):8206–8210, July 2000.

[20] Wenjing Liao and Albert Fannjiang. MUSIC for single-snapshot spectral estimation: stability and super-resolution. 2014.

[21] Veniamin I. Morgenshtern and Emmanuel J. Candès. Stable super-resolution of positive sources: the continuous setup. 2014. To be submitted, preprint is available upon request.

[22] Allan W. Oppenheim, Allan S. Willsky, and S. Hamid. *Signals and Systems*. Prentice Hall, 2nd edition, 1996.

[23] Arogyaswami Paulraj, Richard Roy, and Thomas Kailath. A subspace rotation approach to signal parameter estimation. *Proc. IEEE*, 74(7):1044–1046, July 1986.

[24] Vladilen F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysical Journal International*, 33(3):347–366, 1973.

[25] R. Prony. Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'Ecole Polytechnique*, 1(2):24–76, 1795.

[26] Richard Roy and Thomas Kailath. ESPRIT – estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(7):984–995, July 1989.

[27] Ralph O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propagat.*, AP-34(3):276–280, March 1986.

[28] Morteza Shahram and Peyman Milanfar. Imaging below the diffraction limit: a statistical analysis. *IEEE Transactions on Image Processing*, 13(5):677–689, May 2004.

[29] Morteza Shahram and Peyman Milanfar. On the resolvability of sinusoids with nearby frequencies in the presence of noise. *IEEE Trans. Signal Process.*, 53(7):2579–2588, July 2005.

[30] Petre Stoica and Randolph Moses. *Spectral Analysis of Signals*. Prentice Hall, 2005.

[31] Petre Stoica, Randolph L. Moses, Benjamin Friedlander, and Torsten Söderström. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(3):378–392, 1989.

[32] Petre Stoica and Arye Nehorai. Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case. *Circuits, Systems and Signal Processing*, 8(1):3–15, 1989.

[33] Petre Stoica and Torsten Söderström. Statistical analysis of music and subspace rotation estimates of sinusoidal frequencies. *IEEE Trans. Signal Process.*, 39(8):1836–1847, August 1991.

[34] Donald W. Tufts and Ramdas Kumaresan. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE*, 70(9):975–989, September 1982.