

Generalized Line Spectral Estimation via Convex Optimization

Reinhard Heckel* and Mahdi Soltanolkotabi†

September 26, 2016

Abstract

Line spectral estimation is the problem of recovering the frequencies and amplitudes of a mixture of a few sinusoids from equispaced samples. However, in a variety of signal processing problems arising in imaging, radar, and localization we do not have access directly to such equispaced samples. Rather we only observe a severely undersampled version of these observations through linear measurements. This paper is about such *generalized line spectral estimation* problems. We reformulate these problems as sparse signal recovery problems over a *continuously indexed* dictionary which can be solved via a convex program. We prove that the frequencies and amplitudes of the components of the mixture can be recovered perfectly from a near-minimal number of observations via this convex program. This result holds provided the frequencies are sufficiently separated, and the linear measurements obey natural conditions that are satisfied in a variety of applications.

1 Introduction

Many signals of interest in practical scenarios can be decomposed into short sums of simple objects. Signals arising in imaging applications, radar, and direction of arrival estimation are a few examples. Often, these signals can be represented as a sparse linear combination of elements in a *continuously indexed* dictionary, with the index corresponding for example to the location of an object. Such sparse representations typically allow these signals to be recovered from very few observations.

Traditional approaches to recover signals that are sparse in a continuously dictionary are spectral estimation techniques like Prony’s method, MUSIC, and ESPRIT [SM05]. However, these techniques are only applicable to a few specific continuously indexed dictionaries.

A different line of research in the area of compressive sensing has focused on recovery and denoising of signals that are sparse in a discretely indexed dictionary via greedy methods [Tro04] and convex programs [CRT06]. In principle, these approaches can also be applied to the recovery of signals that are sparse in continuously indexed dictionaries by discretizing the continuous parameter space. However, this discretization step induces a gridding error. While in practice—provided the grid is chosen sufficiently fine—the gridding error is negligible, fine discretization leads to dictionaries with extremely correlated, i.e., coherent, columns. Most of the theory of compressive sensing relies on the dictionary to be incoherent and therefore does not apply to fine grids.

In recent advances convex programming techniques have been shown to succeed for signals that have a sparse representation in continuously indexed dictionaries. Specifically, in [CFG14;

*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA

†Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA

[\[CFG13; FG16\]](#) it is shown that the frequencies of a weighted superposition of complex sinusoids can be recovered perfectly from evenly spaced samples by solving a convex total-variation norm minimization program. This result holds provided that the frequencies of the sinusoids in the mixture are sufficiently separated. A similar statement continues to hold when only a random subset of the uniformly spaced samples are available [\[Tan+13\]](#). Related convex programs have been studied for denoising [\[BTR13\]](#), signal recovery from short-time Fourier measurements [\[ASB15\]](#), deducing time-frequency components of a signal with applications in super-resolution radar [\[HMS16\]](#), and line spectral estimation from multiple measurements [\[LC16\]](#).

The goal of this paper is to characterize an even larger class of continuously indexed dictionaries, for which such convex programs provably succeed. This class includes several practically relevant dictionaries, such as those arising in super-resolution radar and multiple-input multiple-output (MIMO) radar. While our analytical results are for the noiseless case, we provide simulations showing that our approach is robust to noise. We want to highlight that a major advantage of the convex optimization based approach we use in this paper is its wide and straightforward applicability. Indeed, for a number of applications such as super-resolution MIMO radar, we are not aware of any other algorithm that provably succeeds under mild conditions.

1.1 The generalized line spectral estimation problem

In this section we formally introduce the problem considered in this paper. To this aim, consider a discrete vector $\mathbf{z} \in \mathbb{C}^{2N+1}$ consisting of equispaced sampled of a mixture of complex sinusoids

$$\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\nu_k), \quad \text{with } \mathbf{f}(\nu) := [e^{-i2\pi N\nu}, e^{-i2\pi(N-1)\nu}, \dots, e^{i2\pi N\nu}]^T. \quad (1)$$

Here, $\{b_k\}$ are complex-valued coefficients, $\{\nu_k\}$ are frequency parameters that can take any value in $[0, 1]$, and S is the (unknown) number of components in the mixture. In this paper, the goal is to deduce these coefficients and frequency parameters from linear observations of this mixture. More specifically, we consider the problem of recovering $\{b_k\}$ and $\{\nu_k\}$ from measurements of the form

$$\mathbf{y} = \mathbf{A}\mathbf{z} \in \mathbb{C}^M, \quad (2)$$

where $\mathbf{A} \in \mathbb{C}^{M \times L}$ is a sensing matrix and $L := 2N + 1$. The focus of this paper is on the regime where $M < L$; thus \mathbf{A} is not invertible. If \mathbf{A} were invertible, recovery of the unknowns from \mathbf{y} reduces to recovery from \mathbf{z} . The latter is known as a *line spectral estimation* problem. We thus refer to the more general problem of recovering the parameters $\{(b_k, \nu_k)\}$ from $\mathbf{y} = \mathbf{A}\mathbf{z}$ as the *generalized line spectral estimation* problem.

Line spectral estimation is a classical signal processing problem that can be efficiently solved with standard techniques such as Prony's method, MUSIC, and ESPRIT [\[SM05\]](#). However, these standard spectral estimation techniques are in general not applicable to the generalized line spectral estimation problem, in particular not when $M < L$. Recently, there has been a surge of interest in using ideas based on convex optimization for solving the classical line spectral estimation problem and variants thereof. Specifically, [\[CFG14\]](#) show that a convex optimization scheme provably recovers the mixture coefficients and frequencies from \mathbf{z} provided the frequencies are sufficiently separated. As mentioned before, related convex programs have also been studied for diverse problems arising in applications from denoising to radar [\[Tan+13; BTR13; ASB15; HMS16; Cai+16\]](#).

All these problems are special instances of the generalized line spectral estimation problem or the higher dimensional versions thereof, each corresponding to a specific matrix \mathbf{A} in (2). We wish to emphasize that in particular many relevant problems in radar and localization correspond to the high dimensional version (see Equation (7) in Section 1.4) of the one-dimensional relation (2).

Note that the generalized line spectral estimation problem is a sparse signal recovery problem: due to $\mathbf{y} = \sum_{k=1}^S b_k \mathbf{Af}(\nu_k)$, recovery of the parameters $\{(b_k, \nu_k)\}$ corresponds to the recovery of a signal that is sparse in the continuously indexed dictionary $\{\mathbf{Af}(\nu), \nu \in [0, 1]\}$. As it turns out, many practically relevant *continuously* indexed dictionaries that occur in applications such as localization and radar are of the form $\{\mathbf{Af}(\nu), \nu \in [0, 1]\}$.

The main goal of this paper is to show that for a large class of measurement matrices \mathbf{A} , and therefore for a large class of continuously indexed dictionaries, a convex program provably recovers the coefficients $\{b_k\}$ and frequencies $\{\nu_k\}$ from the measurements \mathbf{y} . Our results encompass those in [Tan+13; ASB15; HMS16], but allow for many new measurement models which are of interest in diverse applications such as MIMO radar.

1.2 What mixtures?

Throughout we assume that the frequencies of the components of the mixture are sufficiently separated. Specifically, we require that

$$|\nu_k - \nu_{k'}| \geq \frac{2}{N}, \quad \text{for all } k, k' \in \{1, \dots, S\}, k \neq k'. \quad (3)$$

Here, and throughout, $|\nu_k - \nu_{k'}|$ is the wrap-around distance on the unit circle. For example, $|3/4 - 1/2| = 1/4$ but $|5/6 - 1/6| = 1/3 \neq 2/3$. We note that some form of separation between the frequencies is necessary for *stable* recovery. This follows from identification of the mixture components being extremely ill conditioned when the frequencies are clustered closely together, even when $\mathbf{A} = \mathbf{I}$. Specifically, if there are S' components with frequencies ν_k in an interval of length smaller than $\frac{2S'}{L}$, and S' is large, then in the presence of even a tiny amount of noise, stable recovery from \mathbf{z} is not possible (e.g. see [Don92, Thm. 1.1] and [CFG14, Sec. 1.7]). Condition (3) allows us to have $0.5 S'$ time-frequency shifts in an interval of length $\frac{2S'}{L}$, which is optimal up to the constant 0.5.

1.3 What measurements?

From (1) we know that the signal \mathbf{z} can be written as a sparse linear combination of at most S atoms $\mathbf{f}(\nu)$ in the set $\mathcal{A} := \{\mathbf{f}(\nu), \nu \in [0, 1]\}$. Clearly, one can not hope to solve the generalized line spectral estimation for all measurement matrices even when there are many measurements. For instance, if all the measurement vectors (rows of \mathbf{A}) are orthogonal to the elements of \mathcal{A} then these measurements contain no information about the signal and one can not possibly hope to do spectral estimation.

We next describe a large class of measurement matrices \mathbf{A} that allow to recover the parameters $\{(b_k, \nu_k)\}$ from \mathbf{y} . For $r = 0, 1, \dots, M - 1$, let $\mathbf{a}_r \in \mathbb{C}^L$ denote the rows of \mathbf{A} (a.k.a. the measurement vectors). We assume that the measurement vectors are drawn i.i.d. according to a distribution F obeying two natural conditions, specifically the isotropy and incoherence properties discussed below. We note that we do not actually require the rows of the measurement matrix \mathbf{A} to be i.i.d. Indeed, our results continue to hold if the empirical distribution of the rows of \mathbf{A} obeys the isotropy property

and incoherence properties stated below. It is therefore sufficient if F corresponds to a distribution that picks a vector $\mathbf{a} \in \mathbb{C}^L$ uniformly at random from the rows of \mathbf{A} .

Isotropy property: The distribution F obeys the isotropy property if the measurement vectors $\mathbf{a} \sim F$ obey

$$\mathbb{E}_{\mathbf{a} \sim F} [\mathbf{a}\mathbf{a}^H] = \frac{1}{R} \mathbf{I}. \quad (4)$$

This condition ensures diversity in the measurements. Indeed, if all the measurement vectors are skewed towards certain directions, i.e., $\mathbf{a}_r \approx \mathbf{a}$ for some vector \mathbf{a} then an increase in the number of measurements does not reveal additional information about the signal. In this paper we assume that the isotropy condition holds. We note however that our results hold more generally with the identity matrix replaced with any invertible covariance matrix.

Incoherence property: The distribution F obeys the incoherence property with parameter μ if for all $\mathbf{a} \sim F$ we have

$$\sup_{\mathbf{f} \in \mathbb{C}^L : \|\mathbf{f}\|_\infty \leq 1} |\langle \mathbf{f}, \mathbf{a} \rangle|^2 = \|\mathbf{a}\|_{\ell_1}^2 \leq \frac{L}{M} \mu. \quad (5)$$

Small values of μ imply that the sensing vector \mathbf{a} is incoherent with all atoms in the set

$$\mathcal{A} = \{\mathbf{f}(\nu), \nu \in [0, 1]\} \subset \{\mathbf{f} \in \mathbb{C}^L : \|\mathbf{f}\|_\infty \leq 1\}.$$

We note that the incoherence condition can be relaxed in the sense that (5) does not need to hold deterministically and it is sufficient that it holds with high probability. In Section 1.3.3, we discuss sensing matrices that are incoherent in a probabilistic sense. We note that $\mu \geq 1$ as by the isotropy condition we have

$$\mathbb{E} \left[\sup_{\mathbf{f} \in \mathbb{C}^L : \|\mathbf{f}\|_\infty \leq 1} |\langle \mathbf{f}, \mathbf{a} \rangle|^2 \right] = \mathbb{E} \left[\|\mathbf{a}\|_1^2 \right] \geq \mathbb{E} \left[\|\mathbf{a}\|_2^2 \right] = \text{trace} (\mathbb{E} [\mathbf{a}\mathbf{a}^H]) = \frac{\text{trace}(\mathbf{I})}{M} = \frac{L}{M},$$

which in turn implies $\mu \geq 1$.

Our first result (Theorem 1 stated in Section 3) guarantees that with high probability, the vector \mathbf{z} (and in fact the coefficient/frequency pairs $\{(b_k, \nu_k)\}$ of (1)) can be recovered from M linear measurements of the form $\mathbf{y} = \mathbf{Az} \in \mathbb{C}^M$ by solving a convex program. This result holds provided that the minimum separation condition (3) holds, and provided that the empirical distribution F of the rows of the measurement matrix \mathbf{A} obeys the isotropy and incoherence property with number of measurements satisfying

$$M \geq c\mu S \log^2(L),$$

with c a fixed numerical constant.

Our result states that the more incoherent the sensing vectors are to the atoms in \mathcal{A} , the fewer samples we need for signal recovery. This results is optimal up to logarithmic factors, since the minimum number of measurements required for any method to succeed is at least on the order of S . Our result parallels well known results pertaining to ℓ_1 -minimization for signal recovery for

the case where \mathcal{A} is a basis for \mathbb{C}^L , see [CRT06; CP11], and [FR13, Ch. 12]. Remarkable about our result is that the intuition developed in [CRT06; CP11] carries over, even though \mathcal{A} is not a basis and contains infinitely many points. The only additional condition that we need compared to [CRT06; CP11], [FR13, Ch. 12] is the minimum separation condition. We next discuss three particular classes of measurement modalities which obey the isotropy and incoherence properties.

1.3.1 Subsampled orthogonal matrices

Let \mathbf{U} be an orthogonal basis obeying $\mathbf{U}^H \mathbf{U} = \frac{L}{M} \mathbf{I}$ and let F be the distribution that drawn the sensing vectors independently and uniformly at random from the rows of \mathbf{U} . In this case the distribution F is isotropic and incoherent with incoherence parameter $\mu = \max_k \frac{M}{L} \|\mathbf{u}_k\|_1^2$, where \mathbf{u}_k is the k -th row of \mathbf{U} . As a more specific example, assume $\mathbf{U} = \sqrt{\frac{L}{M}} \mathbf{I}$, where \mathbf{I} is the identity matrix. In this case, $\mu = 1$. This corresponds to the “compressive sensing off the grid” [Tan+13] measurement model, which is the continuous version of the classic compressive sensing problem studied in the seminal work of Candès, Romberg and Tao [CRT06].

1.3.2 Random sampling of sparse trigonometric polynomials

Let us consider another example, not covered by previous results in the literature. Assume we have a continuous signal $z(t) = \sum_{k=1}^S b_k e^{i2\pi\nu_k t}$, and suppose we take M independent measurement of z by sampling z at locations t_r chosen uniformly at random in the interval $[0, L - 1]$. This is equivalent to sampling from a distribution F where the r -th sensing vector is equal to

$$\mathbf{a}_r = \frac{1}{\sqrt{M}} \mathbf{F}^H \mathbf{f}(t_r/L),$$

with t_r drawn independently and uniformly from $[0, L - 1]$. Here, $\mathbf{F} \in \mathbb{C}^{L \times L}$ is the Discrete Fourier Transform (DFT) matrix with entries $[\mathbf{F}]_{p,\ell} = \frac{1}{\sqrt{L}} e^{-i2\pi p\ell/L}$, and the entries of $\mathbf{f}(t_r/L)$ are given by $[\mathbf{f}(t_r/L)]_p = e^{i2\pi p \frac{t_r}{L}}$ with $p = -N, \dots, N$. Note that

$$\mathbb{E} [\mathbf{a}_r \mathbf{a}_r^H] = \frac{1}{M} \mathbf{F}^H \mathbb{E} [\mathbf{f}(t_r/L) \mathbf{f}(t_r/L)^H] \mathbf{F} = \frac{1}{M} \mathbf{F}^H \mathbf{F} = \frac{1}{M} \mathbf{I},$$

so that the distribution F is isotropic. Moreover, the distribution is incoherent with coherence parameter $c \log^2(L)$, where c is a numerical constant. To see that latter, first note that

$$\|\mathbf{a}_r\|_1 = \sqrt{\frac{L}{M}} \sum_{\ell=-N}^N \left| \frac{1}{L} \sum_{p=-N}^N e^{-i2\pi p \frac{\ell-t_r}{L}} \right|.$$

The term in the absolute value is the Dirichlet kernel evaluated at $\frac{\ell-t_r}{L}$. Since the L_1 -norm of the Dirichlet kernel is upper bounded by $\tilde{c} \log(L)$, the incoherence parameter can be bounded by $\mu = c \log^2(L)$. It then follows from our result (i.e. Theorem 1) that, with high probability, the frequencies ν_k can be perfectly recovered from the samples $y_r = z(t_r)$ for $r = 0, \dots, M - 1$, provided that $M \geq S c' \log^4(L)$, and the minimum separation condition holds. This generalizes the main result in [Rau07], which assumes the frequencies ν_k lie on a grid with spacing $1/L$.

1.3.3 Random projection matrices

We next note that our results also hold for a large class of random projection matrices, including matrices with i.i.d. sub-Gaussian¹ entries [FR13, Lem. 9.8]. For instance, matrices with i.i.d. Gaussian entries and i.i.d. entries that are uniformly distributed on $\{-1/\sqrt{M}, 1/\sqrt{M}\}$. While those matrices are not deterministically incoherent in the sense of definition (14), they are in a probabilistic sense incoherent with respect to *any* vector and in particular with respect to the vectors $\{\mathbf{f} \in \mathbb{C}^L : \|\mathbf{f}\|_\infty \leq 1\}$.

We can also guarantee exact signal recovery when using such matrices (see Theorem 2 in Section 3 for a formal statement). Specifically, our result guarantee that with high probability, the pairs $\{(b_k, \nu_k)\}$ and in turn the signal \mathbf{z} in (1) can be perfectly recovered by a convex program, provided that the minimum separation condition (3) holds and the number of measurements satisfies

$$M \geq cS \log(L),$$

where c is a fixed numerical constant.

1.4 Extensions to higher dimensions

The results in Section 1.3 generalize to higher dimensions. This is particularly important, as many applications are two or three dimensional generalized line spectral estimation problems. Specifically, consider the signal

$$\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\mathbf{r}_k), \quad (6)$$

where the complex valued coefficients $\{b_k\}$ and the vectors $\{\mathbf{r}_k\}$ in $[0, 1]^d$ containing continuous frequencies are the parameters we wish to estimate. Here, the entries of $\mathbf{f}(\mathbf{r})$ are given by $[\mathbf{f}(\mathbf{r})]_{\mathbf{p}} = e^{i2\pi \langle \mathbf{p}, \mathbf{r} \rangle}$ where \mathbf{p} an integer vector with ℓ -th entry, $\ell = 1, \dots, d$, given by $[\mathbf{p}]_\ell = -N, \dots, N$. We observe measurements of the form

$$\mathbf{y} = \mathbf{A}\mathbf{z} \in \mathbb{C}^M, \quad (7)$$

where $\mathbf{A} \in \mathbb{C}^{M \times L^d}$ is the measurement matrix with $L = 2N + 1$, as before. In the d -dimensional case, we require the minimum separation condition

$$\max_{\ell=1, \dots, d} |[\mathbf{r}_k]_\ell - [\mathbf{r}_{k'}]_\ell| \geq \frac{c(d)}{N}, \quad (8)$$

with $c(d)$ a constant only depending on the dimension d , to hold. We note that the minimum separation condition (8) only requires the \mathbf{r}_k to be separated in one coordinate, but not in all coordinates simultaneously. Motivated by applications such as MIMO radar in this paper we only focus on $d \leq 3$. In these cases one can assume $c(d) = 5$. However, our results easily extend to higher dimensions ($d > 3$). A variety of signal processing problems arising in imaging, radar, and localization can be phrased as a generalized line spectral estimation problem in higher dimensions. We next discuss two particular radar examples.

¹A random variable x is sub-Gaussian [FR13, Sec. 7.4] if its tail probability satisfies $\mathbb{P}[|x| > t] \leq c_1 e^{-c_2 t^2}$ for constants $c_1, c_2 > 0$.

1.4.1 Super-resolution radar

A radar aims to identify the relative velocities and distances of moving targets, often modeled as point scatterers. To this end, the radar transmits a suitably chosen probing signal, and in response receives a weighted superposition of delayed and Doppler shifted versions of the probing signal. Estimating the delays and Doppler shifts corresponding to the different targets from the response signal at the receiver yields the relative distances and velocities of the targets. As shown in [HMS16], the response to a band-limited probing signal observed within a certain time interval is described by the input-output relation

$$\mathbf{y} = \sum_{k=1}^S b_k \mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}, \quad (9)$$

where $\mathbf{x} \in \mathbb{C}^L$ is the probing signal and \mathcal{F}_ν and \mathcal{T}_τ are time and frequency shift operators². As shown in [HMS16, Sec. 6]

$$\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x} = \mathbf{G}_x \mathbf{F}^H \mathbf{f}(\mathbf{r}), \quad \mathbf{r} = [\tau, \nu]^T, \quad (10)$$

where $\mathbf{G}_x \in \mathbb{C}^{L \times L^2}$ is the Gabor matrix defined by

$$[\mathbf{G}_x]_{p,(k,\ell)} := x_{p-\ell} e^{i2\pi \frac{kp}{L}}, \quad k, \ell, p = -N, \dots, N, \quad (11)$$

and \mathbf{F}^H is the inverse 2D discrete Fourier transform matrix with the entry in the (k, ℓ) -th row and (r, q) -th column given by $[\mathbf{F}^H]_{(k,\ell),(r,q)} := \frac{1}{L^2} e^{i2\pi \frac{qk+r\ell}{L}}$. Note that the entries x_ℓ of the vector \mathbf{x} in (11) are assumed to be L -periodic, therefore $x_\ell = x_{(\ell \bmod L)}$. Thus, (9) is of the form $\mathbf{y} = \mathbf{A}\mathbf{z}$ with $\mathbf{A} = \mathbf{G}_x \mathbf{F}^H$. It is shown in [HMS16] that for a Gaussian probing signal \mathbf{x} , the triplets (b_k, τ_k, ν_k) can be recovered perfectly with high probability, provided the $[\tau_k, \nu_k]$ obey the minimum separation condition (8) and the number of measurements satisfies $L \geq cS \log^3(L)$. While the matrix $\mathbf{A} = \mathbf{G}_x \mathbf{F}^H$ does not obey the exact conditions in Section 1.3, the result in [HMS16] can be derived from the mathematical framework developed in this paper (in particular from Lemma 1 in Section 6).

We finally remark that previous works have considered the case where the $[\tau_k, \nu_k]$ lie on a coarse grid [HS09; HB13] (specifically a grid with spacing $(1/L, 1/L)$). Theorem 1 in [KMR14] guarantees that in this case, ℓ_1 -minimization provably succeeds.

1.4.2 Super-resolution MIMO radar

A standard or Single-Input Single-Output (SISO) radar can determine the relative distances and velocities of the targets by estimating the induced delays and Doppler shifts. A SISO radar is however incapable of determining the actual positions of the objects with a single transmit/receive antenna. MIMO radar systems [BF03; LS07] use multiple antennas to transmit multiple probing signals simultaneously and record the reflections from the targets with multiple receive antennas. As a result a MIMO radar can resolve the relative angles along with the relative distances and

²The time and frequency shift operators are formally defined as $[\mathcal{F}_\nu \mathbf{x}_p]_p = x_p e^{i2\pi p\nu}$ and $\mathcal{T}_\tau = \mathbf{F}^H \mathcal{F}_\tau \mathbf{F}$ where \mathbf{F} is a one dimensional Discrete Fourier Transform (DFT) matrix, with the entry in the k -th row and r -th column given by $[\mathbf{F}]_{k,r} := \frac{1}{\sqrt{L}} e^{-i2\pi \frac{kr}{L}}$, $k, r = 0, \dots, L-1$.

velocities of targets. Let \mathbf{x}_j be the samples of the probing signal transmitted by the j -th transmit antenna, with $j = 0, \dots, N_T - 1$, where N_T is the number of transmit antennas. As explained in more detail in [Hec16], the response to those probing signals, recorded at the r -th antenna, $r = 0, \dots, N_R - 1$, where N_R is the number of receive antennas, is given by

$$\mathbf{y}_r = \sum_{k=1}^S b'_k e^{i2\pi r N_T \beta_k} \sum_{j=0}^{N_T-1} e^{i2\pi j \beta_k} \mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_j, \quad (12)$$

where the parameters $\tau_k, \nu_k, \beta_k \in [0, 1]$ correspond to delays, Doppler shifts, and angle. Again, by setting $\mathbf{y} := [\mathbf{y}_0^T, \dots, \mathbf{y}_{M-1}^T]^T$ and defining $\mathbf{A} \in \mathbb{C}^{N_R L \times N_R N_T L^2}$ appropriately, the input-output relation (12) obeys (6) with $d = 3$, as detailed in Section 6.3 (here, we assume for notational simplicity that $N_R N_T = L$). In this paper, we prove that with high probability the parameters (β_k, τ_k, ν_k) can be recovered perfectly, provided the minimum separation condition (8) holds, from a near-minimal number of measurements.

1.5 Outline

In Section 2 we present an algorithm for the generalized line spectral estimation problem, and Section 3 contains the corresponding formal recovery guarantees. In Section 4 we prove results demonstrating that if the frequencies lie on an arbitrarily fine grid, ℓ_1 -minimization techniques provably succeed for generalized line spectral estimation. We illustrate the effectiveness of our methods with various numerical experiments in Section 5. Finally, we prove our results in Sections 6 and 7.

1.6 Notation

We provide a brief summary of some of the notation used throughout the paper. The superscripts T and H stand for transposition and Hermitian transposition, respectively. For the vector \mathbf{x} , $[\mathbf{x}]_q$ denotes its q -th entry, $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_1$ its ℓ_2 - and ℓ_1 -norm, and $\|\mathbf{x}\|_\infty = \max_q |x_q|$ its largest entry. For the matrix \mathbf{A} , $[\mathbf{A}]_{ij}$ designates the entry in its i -th row and j -th column, $\|\mathbf{A}\| := \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2$ its spectral norm, and $\|\mathbf{A}\|_F := (\sum_{i,j} |[\mathbf{A}]_{ij}|^2)^{1/2}$ its Frobenius norm. The identity matrix is denoted by \mathbf{I} . For convenience, we will frequently use a two- or three-dimensional index for vectors and matrices, e.g., we write $[\mathbf{g}]_{(k,\ell)}, k, \ell = -N, \dots, N$ for $\mathbf{g} = [g_{(-N,-N)}, g_{(-N,-N+1)}, \dots, g_{(-N,N)}, g_{(-N+1,-N)}, \dots, g_{(N,N)}]^T$. For the set \mathcal{S} , $|\mathcal{S}|$ designates its cardinality and \mathcal{S}^c is its complement. Finally, $c, \tilde{c}, c', c_1, c_2, \dots$ are numerical constants that can take on different values at different occurrences.

2 Generalized line spectral estimation via atomic norm minimization

In this section, we present the recovery algorithm considered in this paper. We consider the d -dimensional case, and state formal results for $d \leq 3$. Recall that recovery of the coefficients $\{b_k\}$ and frequencies $\{\mathbf{r}_k\}$ from \mathbf{z} in (6) is a d -dimensional line spectral estimation problem that can be solved with standard spectral estimation techniques such as Prony's method [SM05], [GS05, Ch. 2]. Therefore, it suffices to recover $\mathbf{z} \in \mathbb{C}^{L^d}$ from the linear measurements $\mathbf{y} = \mathbf{A}\mathbf{z} \in \mathbb{C}^M$. To this aim,

we use the fact that \mathbf{z} is a sparse linear combination of atoms from the set $\mathcal{A} := \{\mathbf{f}(\mathbf{r}), \mathbf{r} \in [0, 1]^d\}$. A regularizer that promotes such a sparse linear combination is the atomic norm induced by these signals [Cha+12], defined as

$$\|\mathbf{z}\|_{\mathcal{A}} := \inf_{b_k \in \mathbb{C}, \mathbf{r}_k \in [0, 1]^d} \left\{ \sum_k |b_k| : \mathbf{z} = \sum_k b_k \mathbf{f}(\mathbf{r}_k) \right\}.$$

We estimate \mathbf{z} by solving the basis pursuit type atomic norm minimization problem

$$\text{AN}(\mathbf{y}): \underset{\tilde{\mathbf{z}}}{\text{minimize}} \|\tilde{\mathbf{z}}\|_{\mathcal{A}} \text{ subject to } \mathbf{y} = \mathbf{A}\tilde{\mathbf{z}}. \quad (13)$$

To summarize, we estimate $\{(b_k, \mathbf{r}_k)\}$ from \mathbf{y} by

1. solving $\text{AN}(\mathbf{y})$ in order to obtain \mathbf{z} ,
2. estimating $\{\mathbf{r}_k\}$ from \mathbf{z} by solving the corresponding line spectral estimation problem, and
3. solving the linear system of equations $\mathbf{y} = \sum_{k=1}^S b_k \mathbf{A}\mathbf{f}(\mathbf{r}_k)$ for the b_k .

We remark that the $\{\mathbf{r}_k\}$ may be obtained more directly from a solution to the dual of (13); see [BTR13, Sec. 3.1], [CFG14, Sec. 4], [Tan+13, Sec. 2.2], and [HMS16, Sec. 6] for details on this approach applied to related problems. This approach does, however, require identification of the zeros of a d -dimensional trigonometric polynomial.

Since computation of the atomic norm involves taking the infimum over infinitely many parameters, finding a solution to $\text{AN}(\mathbf{y})$ may appear to be daunting. For the 1D case (i.e., $d = 1$), the atomic norm can be characterized in terms of linear matrix inequalities (LMIs) [Tan+13, Prop. 2.1]. This characterization is based on the Vandermonde decomposition lemma for Toeplitz matrices, and allows to formulate the atomic norm minimization program as a semidefinite program that can be solved in polynomial time. While this lemma generalizes to higher dimensions [YXS16, Thm. 1], there is no polynomial bound on the dimension of the corresponding LMIs. Nevertheless, using [YXS16, Thm. 1] one can obtain a convex relaxation of $\text{AN}(\mathbf{y})$, which is a Semidefinite Program (SDP) and can be solved in polynomial time. Similarly, a solution of the dual of $\text{AN}(\mathbf{y})$ can be found with a SDP relaxation.

Since the computational complexity of those SDPs is often prohibitive in practice, in particular for $d > 1$, we do not detail these SDPs. Instead, we show in Section 4 that the frequencies $\{\mathbf{r}_k\}$ of the mixtures can be recovered on an arbitrarily fine grid via ℓ_1 -minimization. While this approach introduces a gridding error, the grid may be chosen sufficiently fine for the gridding error to be negligible compared to the error induced by additive noise (in practice, there is typically additive noise).

3 Main results

In this section we state our main results which show that under mild conditions, the solution to $\text{AN}(\mathbf{y})$ in (13) yields the unknown signal \mathbf{z} . As discussed in the previous section, from \mathbf{z} we can then extract the coefficients and frequency parameters $\{(b_k, \mathbf{r}_k)\}$. In the introduction, we informally stated our results for the 1D case. As mentioned in the introduction, those results continue to hold

for higher dimensions. In an effort to minimize redundancy and for the sake of concreteness, we state and prove our results for isotropic and incoherent matrices for the 3D case. We choose the 3D case since the MIMO radar problem is a three dimensional problem, and therefore the corresponding proof requires deriving technical results for the 3D case. We state results for three different classes of random matrices.

3.1 Isotropic and incoherent matrices

We start by stating our main result for a matrix \mathbf{A} obeying isotropy and incoherence conditions.

Theorem 1. *Let $\mathbf{A} \in \mathbb{C}^{M \times L^3}$ with $L \geq 1024$, be a random matrix with rows \mathbf{a}_r chosen independently from a distribution obeying the isotropy and incoherence properties*

$$\mathbb{E} [\mathbf{a}_r \mathbf{a}_r^H] = \frac{1}{M} \mathbf{I} \quad \text{and} \quad \sup_{\mathbf{f} \in \mathbb{C}^{L^3}: \|\mathbf{f}\|_\infty \leq 1} |\langle \mathbf{f}, \mathbf{a}_r \rangle|^2 \leq \frac{L^3}{M} \mu, \quad (14)$$

for some fixed $\mu \geq 1$. Furthermore, assume we have M measurements $\mathbf{y} = \mathbf{Az}$ from an unknown signal of the form $\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\mathbf{r}_k)$. Also assume that the signs $\text{sign}(b_k)$ of the coefficients $b_k \in \mathbb{C}$ are chosen independently from symmetric distributions on the complex unit circle, and the vectors $\mathbf{r}_k = (\beta_k, \tau_k, \nu_k) \in [0, 1]^3$ obey the minimum separation condition

$$\max(|\beta_k - \beta_{k'}|, |\tau_k - \tau_{k'}|, |\nu_k - \nu_{k'}|) \geq \frac{5}{N} \text{ for all } k, k' \text{ with } k \neq k'. \quad (15)$$

Then as long as

$$M \geq c\mu S \log^2(L/\delta), \quad (16)$$

with c a fixed numerical constant, \mathbf{z} is the unique minimizer of $\text{AN}(\mathbf{y})$ in (13) with probability at least $1 - \delta$.

Our main results states that when the measurement matrix obeys the isotropy and incoherence conditions and the frequencies of the mixture are sufficiently separated then we can recover the signal \mathbf{z} via atomic norm minimization, and in turn recover the parameters $\{(b_k, \mathbf{r}_k)\}$. We note that any algorithm (regardless of tractability) requires at least $M \gtrsim S \log(L/S)$ measurements for *stable* recovery, and $M \geq S$ measurements for recovery in general, so that our results are optimal up to the additional logarithmic factor.

3.2 Random projection matrices

Our next class of matrices are random projection matrices that preserve Euclidean norms. These matrices are not covered directly by Theorem 1 as they are not deterministically incoherent in the sense of definition (14). However, these matrices are still incoherent in a probabilistic sense with respect to *any* vector and in particular with respect to the vectors $\{\mathbf{f} \in \mathbb{C}^L: \|\mathbf{f}\|_\infty \leq 1\}$.

For simplicity we state our result for a Gaussian random matrix.

Theorem 2. *Let $\mathbf{A} \in \mathbb{C}^{M \times L}$ be a random projection matrix with i.i.d. $\mathcal{N}(0, 1/M)$ entries. Furthermore, assume we have M measurements $\mathbf{y} = \mathbf{Az}$ from an unknown signal of the form $\mathbf{z} =$*

$\sum_{k=1}^S b_k \mathbf{f}(\nu_k), L \geq 512$. Also assume that the frequencies $\nu_k \in [0, 1]$ obey the minimum separation condition (3). Then as long as

$$M \geq cS \log(L), \quad (17)$$

with c a fixed numerical constant, \mathbf{z} is the unique minimizer of $\text{AN}(\mathbf{y})$ in (13) with probability at least $e^{-(M-2)/8}$.

Note that this statement is superior to our previous statements in that it does not require the sign of the coefficients $\text{sign}(b_k)$ to be random. The proof in Section 7 reveals that the result follows from Gordon's Lemma [Gor88; Gor85] carefully applied to this problem. Indeed, this result holds more broadly for all matrices that preserve Euclidean norms of vectors over a general set (i.e., matrices obeying Gordon's Lemma), for instance sub-Gaussian random matrices and certain more structured random projection matrices defined below. The corresponding proof is analogous to that for a Gaussian matrix presented in Section 7 by using results in [Dir15] and [ORS15a, Thm. 3.3] ensuring that the corresponding random matrices behave like Gaussian random matrices in terms of dimensionality reduction.

Isotropic sub-Gaussian matrices: The first is the class of isotropic sub-Gaussian matrices [ORS15b]. A matrix $\mathbf{A} \in \mathbb{C}^{M \times L}$ is Δ -sub-Gaussian if its rows are independent of each other and for all $r = 0, \dots, M - 1$, the r -th row \mathbf{a}_r^H satisfies

$$\mathbb{E}[\mathbf{a}_r] = \mathbf{0}, \quad \mathbb{E}[\mathbf{a}_r \mathbf{a}_r^H] = \mathbf{I},^3 \quad \text{and for any vector } \mathbf{v}, \mathbb{P}[|\langle \mathbf{v}, \mathbf{a}_r \rangle| \geq t \|\mathbf{v}\|_2] \leq e^{-\frac{t^2}{\Delta^2}}.$$

Theorem 2 continues to hold with condition (17) replaced by $M \geq cS \log(L)/\Delta^2$.

Subsampled Orthogonal with Random Sign (SORS): The second class of matrices to which Theorem 2 can be extended to is the class of SORS matrices [ORS15a]. Let $\mathbf{U} \in \mathbb{R}^{L \times L}$ denote an orthonormal matrix obeying

$$\mathbf{U}^H \mathbf{U} = \mathbf{I} \quad \text{and} \quad \max_{i,j} |\mathbf{F}_{ij}| \leq \frac{\Delta}{\sqrt{n}}. \quad (18)$$

Define the random subsampled matrix $\mathbf{H} \in \mathbb{R}^{M \times L}$ with i.i.d. rows chosen uniformly at random from the rows of \mathbf{U} . A SORS matrix is defined as $\mathbf{A} = \mathbf{H}\mathbf{D}$, where $\mathbf{D} \in \mathbb{R}^{L \times L}$ is a random diagonal matrix with the diagonal entries i.i.d. ± 1 with equal probability. Theorem 2 extends to SORS matrices with condition (17) replaced by $M \geq c \frac{S \log^4(L)}{\Delta^2}$. Note that such random matrices include the fast random projection matrices in [AL13; KW11]. Specifically, with \mathbf{H} the Hadamard matrix, the random projection $\mathbf{A} = \mathbf{H}\mathbf{D}$ can be applied in time $O(L \log L)$ as opposed to time $O(ML)$ for the realizations of general sub-Gaussian random matrices.

³The assumption that the population covariance matrix is identity is not necessary. In fact our result generalizes to any covariance matrix. The only change is that the required number of measurements will now scale with the condition number of the covariance matrix.

3.3 Superresolution MIMO radar matrices

We next state our main result for random matrices appearing in MIMO radar problems. A simple calculation shows (see Section 6.3) that the MIMO radar input-output relation (12) can be viewed as taking $N_R L$ linear measurements

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{N_R-1} \end{bmatrix} = \mathbf{A}\mathbf{z}, \quad (19)$$

from a vector \mathbf{z} of the form (6) with $d = 3$. Here, $\mathbf{A} \in \mathbb{C}^{N_R L \times N_R N_T L^2}$ is a function of the transmitted signals \mathbf{x}_j and is given by

$$\mathbf{A} := \begin{bmatrix} \mathbf{B} & & \\ & \ddots & \\ & & \mathbf{B} \end{bmatrix}, \quad \mathbf{B} := [\mathbf{G}_{\mathbf{x}_0} \mathbf{F}^H \dots \mathbf{G}_{\mathbf{x}_{N_T-1}} \mathbf{F}^H] \in \mathbb{C}^{L \times L^2 N_T}, \quad (20)$$

where $\mathbf{G}_{\mathbf{x}} \mathbf{F}^H$ is the matrix in (10) that appeared in the super-resolution radar problem before.

Theorem 3. *Assume that the probing signals $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N_T-1} \in \mathbb{R}^L$ are i.i.d. $\mathcal{N}(\mathbf{0}, \frac{1}{N_T L} \mathbf{I})$ random vectors with $L = 2N + 1 \geq 1024$ and $N_T N_R \geq 1024$. Let $\mathbf{y} = \mathbf{A}\mathbf{z}$ with \mathbf{A} defined in (20) and $\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\mathbf{r}_k)$. Also, assume that the signs $\text{sign}(b_k)$ of the coefficients $b_k \in \mathbb{C}$ are chosen independently from symmetric distributions on the complex unit circle, and the vectors $\mathbf{r}_k = (\beta_k, \tau_k, \nu_k) \in [0, 1]^3$ obey the minimum separation condition*

$$|\beta_k - \beta_{k'}| \geq \frac{10}{N_T N_R - 1} \quad \text{or} \quad |\tau_k - \tau_{k'}| \geq \frac{5}{N} \quad \text{or} \quad |\nu_k - \nu_{k'}| \geq \frac{5}{N}, \quad \text{for all } k, k' : k \neq k'. \quad (21)$$

Then as long as

$$\min(L, N_T N_R) \geq S c \log^3(L/\delta), \quad (22)$$

with c a fixed numerical constant, \mathbf{z} is the unique minimizer of $\text{AN}(\mathbf{y})$ in (13) with probability at least $1 - \delta$.

Even though this result does not allow S to be near linear in the total number of measurements $N_R L$, it is optimal up to log-factors. Indeed, $S \leq \min(L, N_T N_R)$ is in general a necessary condition to uniquely recover the b_k even when the frequencies of the mixture components (β_k, τ_k, ν_k) are known [Hec16].

We note that Theorems 1 and 3 assume that the coefficients $\{b_k\}$ have random sign. In the radar model, where the b_k describe the attenuation factors, this is a common and well justified assumption [Bel63]. While this assumption is reasonable for several applications, we believe that it is not necessary for our results to hold. Indeed, Theorem 2 does not require this assumption. We hope to remove this assumption in future publications.

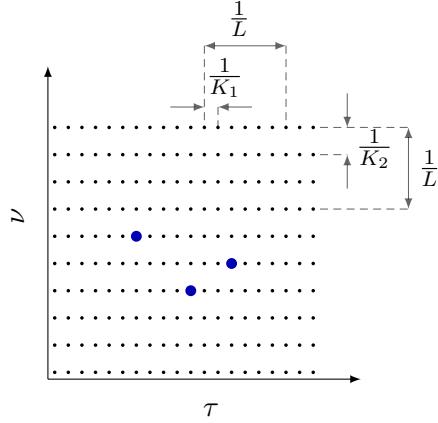


Figure 1: Illustration of the (fine) grid in two dimensions. In this example, the non-zeros, marked by blue dots, violate the minimum separation condition (24).

4 Recovery on a fine grid

A natural approach for estimating the parameters $\{\mathbf{r}_k\}$ from $\mathbf{y} = \mathbf{A}\mathbf{z}$, with \mathbf{z} defined in (6), is to assume the parameters $\{\mathbf{r}_k\}$ lie on a *fine* grid, and solve the generalized line spectral estimation problem on that grid. When the frequencies do not lie on a grid this of course leads to a “gridding error”. However, this error is negligible when the grid is very fine. When the frequencies do lie on a grid the generalized line spectral estimation problem can be formulated as a discrete sparse signal recovery problem, albeit with a very coherent dictionary. In this section we discuss this approach in detail focusing on the 3D-case.

Suppose the vectors $\mathbf{r}_k = (\beta_k, \tau_k, \nu_k)$ lie on a grid with spacing $(1/K_1, 1/K_2, 1/K_3)$, where K_1, K_2, K_3 are integers obeying $K_1, K_2, K_3 \geq L = 2N + 1$, see Figure 1 for an illustration. With this assumption, recovery of the parameters $\{(b_k, \mathbf{r}_k)\}$ from \mathbf{y} reduces to the recovery of the sparse (discrete) signal $\mathbf{s} \in \mathbb{C}^{K_1 K_2 K_3}$ from the measurement

$$\mathbf{y} = \mathbf{A}\mathbf{F}_{\text{grid}}\mathbf{s},$$

where $\mathbf{F}_{\text{grid}} \in \mathbb{C}^{L^3 \times K_1 K_2 K_3}$ is the matrix with (n_1, n_2, n_3) -th column given by $\mathbf{f}(n_1/K_1, n_2/K_2, n_3/K_3)$. The non-zeros of \mathbf{s} and its indices correspond to the $\{b_k\}$ and the $\{\mathbf{r}_k\}$ on the grid, respectively. A standard approach for recovering the sparse signal \mathbf{s} from the underdetermined linear system of equations $\mathbf{y} = \mathbf{A}\mathbf{F}_{\text{grid}}\mathbf{s}$ is to solve the following convex program:

$$\text{L1}(\mathbf{y}): \underset{\tilde{\mathbf{s}}}{\text{minimize}} \|\tilde{\mathbf{s}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{F}_{\text{grid}}\tilde{\mathbf{s}}. \quad (23)$$

The following theorems discuss our main results for recovery on the fine grid. For brevity we only state the discrete analogue to Theorems 1 and 3. A similar result holds for Theorem 2. Our first result is the discrete analogue of Theorem 1.

Theorem 4. *Let \mathbf{s} be an S -sparse vector with non-zeros indexed by the support set $\mathcal{S} \subseteq [K_1] \times$*

$[K_2] \times [K_3]$, $[K] := \{0, \dots, K-1\}$, obeying the minimum separation condition

$$\max_{\ell \in \{1,2,3\}} \frac{|n_\ell - n'_\ell|}{K_\ell} \geq \frac{5}{N}, \quad \text{for all triplets } [n_1, n_2, n_3], [n'_1, n'_2, n'_3] \in \mathcal{S}. \quad (24)$$

Furthermore, assume that the signs of the non-zeros of \mathbf{s} are chosen independently from symmetric distributions on the complex unit circle. Also assume we have M linear measurements $\mathbf{A} \in \mathbb{C}^{M \times L^3}$ with $L \geq 1024$ of the form $\mathbf{y} = \mathbf{AF}_{\text{grids}}$ with the matrix \mathbf{A} obeying the conditions of Theorem 1, in particular the isotropy and incoherence conditions. Then as long as

$$M \geq c\mu S \log^2(L/\delta),$$

with c a fixed numerical constant, \mathbf{s} is the unique minimizer of $\text{L1}(\mathbf{y})$ with probability at least $1 - \delta$.

Our second result is the discrete analogue of Theorem 3.

Theorem 5. Let \mathbf{s} be a S -sparse vector with non-zeros indexed by the support set $\mathcal{S} \subseteq [K_1] \times [K_2] \times [K_3]$, obeying the minimum separation condition, for all triplets $[n_1, n_2, n_3], [n'_1, n'_2, n'_3] \in \mathcal{S}$,

$$\frac{|n_1 - n'_1|}{K_1} \geq \frac{10}{N_T N_R - 1} \quad \text{or} \quad \frac{|n_2 - n'_2|}{K_1} \geq \frac{5}{N} \quad \text{or} \quad \frac{|n_3 - n'_3|}{K_1} \geq \frac{5}{N}.$$

Furthermore, assume that the signs of the non-zeros of \mathbf{s} are chosen independently from symmetric distributions on the complex unit circle. Suppose that \mathbf{A} satisfies the conditions of Theorem 3 and assume (22) holds for some fixed $\delta > 0$. Then, with probability at least $1 - \delta$, \mathbf{s} is the unique minimizer of $\text{L1}(\mathbf{y})$.

Note that Theorems 4 and 5 do not impose any restrictions on K_1, K_2 , and K_3 , in particular they can be arbitrarily large. The proof of those results, presented in Appendix B, is closely linked to that of Theorems 1 and 3. Specifically, the existence of a certain dual certificate guarantees that \mathbf{s} is the unique minimizer of $\text{L1}(\mathbf{y})$ in (23). These dual certificates are obtained directly from the dual polynomials for the continuous case in the proofs of Theorems 1 and 3. See Appendix B for further details.

5 Numerical results

In this section, we perform some numerical experiments to evaluate the resolution obtained by our approach, to demonstrate that it is robust to noise, and also to compare it to a competing algorithm. We focus on the MIMO radar problem, as results for the other matrices covered by our theory are similar. We set $N_T = 3, N_R = 3, L = 41$, and draw $S = 5$ target locations (β_k, τ_k, ν_k) and corresponding attenuation factors b_k uniformly at random from $[0, 1] \times [0, 2/\sqrt{L}]^2$ and from the complex unit disc, respectively. Moreover, we choose $K_1 = \text{SRF} \cdot N_T N_R, K_2 = \text{SRF} \cdot L$, and $K_3 = \text{SRF} \cdot L$, where $\text{SRF} \geq 1$ can be interpreted as a super-resolution factor determining by how much the $(1/K_1, 1/K_2, 1/K_3)$ grid is finer than the “coarse” grid $(1/(N_T N_R), 1/L, 1/L)$. To account for additive noise, we solve the following modification of $\text{L1}(\mathbf{y})$ in (23)

$$\text{L1-ERR: minimize}_{\tilde{\mathbf{s}}} \|\tilde{\mathbf{s}}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{R}\tilde{\mathbf{s}}\|_2^2 \leq \eta, \quad (25)$$

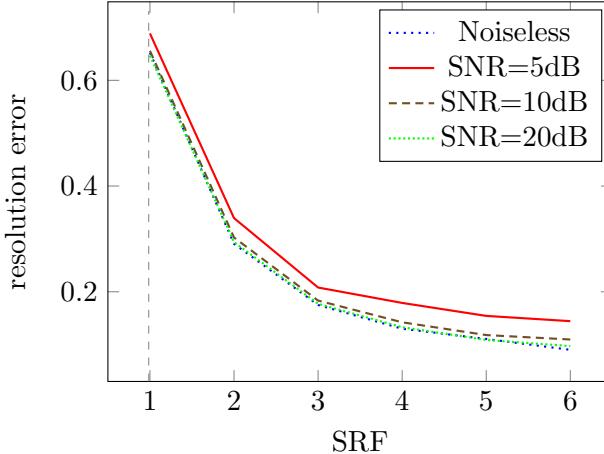


Figure 2: Resolution error for the recovery of $S = 5$ targets from the samples \mathbf{y} with and without additive Gaussian noise \mathbf{n} of a certain signal-to-noise ratio $\text{SNR} = \|\mathbf{y}\|_2^2/\|\mathbf{n}\|_2^2$, for varying super-resolution factors (SRFs). The resolution error is defined as the average over $(N_T^2 N_R^2 (\hat{\beta}_k - \beta_k)^2 + L^2 (\hat{\tau}_k - \tau_k)^2 + L^2 (\hat{\nu}_k - \nu_k)^2)^{1/2}$, $k = 1, \dots, S$, where $(\hat{\beta}_k, \hat{\tau}_k, \hat{\nu}_k)$ are the locations obtained by solving L1-ERR.

with η chosen on the order of the noise variance. There are two error sources incurred by this approach: the gridding error obtained by assuming the points lie on a grid with spacing $(1/K_1, 1/K_2, 1/K_3)$, and the additive noise error. The former decreases when SRF increases, while the latter only depends on the noise variance. The results, depicted in Figure 2, show that the target resolution of our approach is significantly better than the compressed sensing based approach proposed in [DR16; SW15] corresponding to recovery on the coarse grid, i.e., $\text{SRF} = 1$. Moreover, the results show that our approach is robust to noise.

We now compare our approach to the Iterative Adaptive Approach (IAA) [Yar+10], proposed for MIMO radar in [Rob+10]. IAA is based on weighted least squares and has been proposed in the array processing literature, where typically multiple measurements, i.e., multiple observations \mathbf{y} are available. IAA can work well with even one measurement vector \mathbf{y} and can therefore be directly applied to the problems considered in this paper. However, to the best of our knowledge no rigorous performance guarantees are available for IAA. We compare the IAA algorithm [Yar+10, Table II, “The IAA-APES Algorithm”] to L1-ERR in (25). We use the same problem parameters as above, but with $\text{SRF} = 3$ and $(\beta_k, \tau_k, \nu_k) = (k/(N_R N_t), k/L, k/L)$, $k = 0, \dots, S - 1$, so that the location parameters lie on the coarse grid and are therefore separated. As before, we draw the corresponding attenuation factors b_k i.i.d. uniformly at random from the complex unit disc. Our results, depicted in Figure 3, show that L1-ERR performs better in this experiment than IAA specially for small signal-to-noise ratios.

6 Proofs of Theorems 1 and 3

Theorems 1 and 3 are proven by constructing an appropriate dual certificate. The existence of this certificate guarantees that the solution to $\text{AN}(\mathbf{y})$ in (13) is \mathbf{z} , as formalized by Proposition 1 below. Proposition 1 is a consequence of strong duality, and well known for the discrete setting

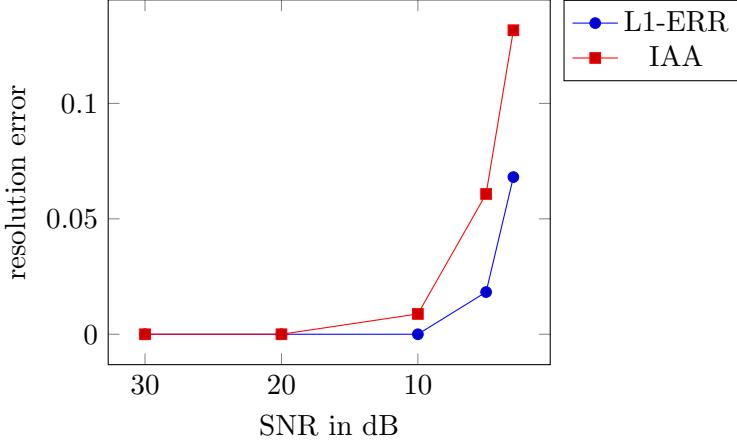


Figure 3: Resolution error achieved by L1-ERR and by IAA applied to $\mathbf{y} + \mathbf{n}$, where $\mathbf{n} \in \mathbb{C}^R$ is additive Gaussian noise, such that the signal-to-noise ratio is $\text{SNR} := \|\mathbf{y}\|_2^2 / \|\mathbf{n}\|_2^2$. As before, the resolution error is defined as $(N_T^2 N_R^2 (\hat{\beta}_k - \beta_k)^2 + L^2 (\hat{\tau}_k - \tau_k)^2 + L^2 (\hat{\nu}_k - \nu_k)^2)^{1/2}$, where $(\hat{\beta}_k, \hat{\tau}_k, \hat{\nu}_k)$ are the locations obtained by using L1-ERR or IAA. The resolution error is averaged over 100 trials.

from the compressed sensing literature [CRT06]. The proof of Proposition 1 is standard, see for example [Tan+13, Proof of Prop. 2.4].

Proposition 1. *Let $\mathbf{y} = \mathbf{Az}$ with $\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\mathbf{r}_k)$. If there exists a dual certificate*

$$Q(\mathbf{r}) = \langle \mathbf{A}^H \mathbf{q}, \mathbf{f}(\mathbf{r}) \rangle \quad (26)$$

with complex coefficients $\mathbf{q} \in \mathbb{C}^M$ such that

$$Q(\mathbf{r}_k) = \text{sign}(b_k), \text{ for all } k = 1, \dots, S, \quad \text{and } |Q(\mathbf{r})| < 1 \text{ for all } \mathbf{r} \in [0, 1]^3 \setminus \{\mathbf{r}_1, \dots, \mathbf{r}_S\}, \quad (27)$$

then \mathbf{z} is the unique minimizer of $\text{AN}(\mathbf{y})$.

We now need to show that such a dual certificate exists under appropriate conditions. This is the subject of the next section.

6.1 Sufficient conditions for the existence of dual certificates

In this section we state the main technical result of this paper—Lemma 1 below—which provides sufficient conditions on \mathbf{A} that guarantee the existence of a dual certificate as described in Proposition 1. The proofs of Theorems 1 and 3 are based on verifying these conditions.

To derive this lemma, we construct the dual certificate Q of Proposition 1 explicitly. Our construction of Q is similar to that in [HMS16], and is inspired by the construction of related polynomials in [CFG14; Tan+13]. We briefly outline the construction of Q , in order to introduce the notation required for stating our key lemma. To this aim, first note that the constraint (26) requires Q to be a 3D-trigonometric polynomial with coefficients $\mathbf{A}^H \mathbf{q}$. To build Q we therefore need to construct a 3D-trigonometric polynomial that satisfies the conditions (27), and whose

coefficients are constraint to be of the form $\mathbf{A}^H \mathbf{q}$, for some $\mathbf{q} \in \mathbb{C}^M$. Since \mathbf{A} is a random matrix, Q is a *random* trigonometric polynomial.

It is instructive to first consider the construction of a *deterministic* 3D trigonometric polynomial of the form $\bar{Q}(\mathbf{r}) = \langle \bar{\mathbf{q}}, \mathbf{f}(\mathbf{r}) \rangle$ with deterministic coefficients $\bar{\mathbf{q}}$ that satisfies the conditions (27), but whose coefficients $\bar{\mathbf{q}} \in \mathbb{C}^{L^3}$ are *not* required to be of the form $\mathbf{A}^H \mathbf{q}$. Such a construction has been established (provided a minimum separation condition on the \mathbf{r}_k holds) by [CFG14, Prop. 2.1, Prop. C.1] for the 1D and 2D case. The 3D result follows in a similar manner, we therefore omit the details. To construct $\bar{Q}(\mathbf{r})$, Candès and Fernandez-Granda [CFG14] interpolate the points $u_k := \text{sign}(b_k)$ with a fast-decaying kernel \bar{G} (defined below) and slightly adopt this interpolation near the \mathbf{r}_k with the partial derivatives $\bar{G}^{(n_1, n_2, n_3)}(\mathbf{r}) := \frac{\partial^{n_1}}{\partial \beta^{n_1}} \frac{\partial^{n_2}}{\partial \tau^{n_2}} \frac{\partial^{n_3}}{\partial \nu^{n_3}} \bar{G}(\mathbf{r})$ (recall that $\mathbf{r} = [\beta, \tau, \nu]^T$) to ensure that local maxima are achieved at the \mathbf{r}_k . Specifically, \bar{Q} takes the form

$$\bar{Q}(\mathbf{r}) = \sum_{k=1}^S \bar{\alpha}_k \bar{G}(\mathbf{r} - \mathbf{r}_k) + \bar{\alpha}_{1k} \bar{G}^{(1,0,0)}(\mathbf{r} - \mathbf{r}_k) + \bar{\alpha}_{2k} \bar{G}^{(0,1,0)}(\mathbf{r} - \mathbf{r}_k) + \bar{\alpha}_{3k} \bar{G}^{(0,0,1)}(\mathbf{r} - \mathbf{r}_k). \quad (28)$$

Here, $\bar{G}^{(n_1, n_2, n_3)}(\mathbf{r}) := \frac{\partial^{n_1}}{\partial \beta^{n_1}} \frac{\partial^{n_2}}{\partial \tau^{n_2}} \frac{\partial^{n_3}}{\partial \nu^{n_3}} \bar{G}(\mathbf{r})$ and $\bar{G}(\mathbf{r}) := F(\tau)F(\nu)F(\beta)$, where F is the squared Fejér kernel defined as

$$F(t) := \left(\frac{\sin(R\pi t)}{R \sin(\pi t)} \right)^4, \quad R := \frac{N}{2} + 1.$$

The Fejér kernel is a trigonometric polynomial of degree $N/2$ which implies that F is a trigonometric polynomial of degree N . F can be written in the form

$$F(t) = \frac{1}{R} \sum_{k=-N}^N g_k e^{i2\pi t k}, \quad (29)$$

with coefficients g_k obeying $|g_k| \leq 1$. Observe that shifted versions of F (i.e., $F(t - t_0)$) and the derivatives of F are also 1D trigonometric polynomials of degree N . Therefore the kernel \bar{G} , its partial derivatives, and shifted versions are also 3D trigonometric polynomials of the form $\langle \bar{\mathbf{q}}, \mathbf{f}(\mathbf{r}) \rangle$, as desired.

\bar{Q} is not a valid dual certificate for the generalized line spectral estimation problem as it is not of the form (26). However, inspired by the construction of \bar{Q} we build Q by interpolating the points $u_k = \text{sign}(b_k)$ at \mathbf{r}_k with the functions

$$G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_k) := \langle \mathbf{A} \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \mathbf{A} \mathbf{f}(\mathbf{r}) \rangle. \quad (30)$$

Here, $\mathbf{g}_{\mathbf{m}}(\mathbf{r})$, $\mathbf{m} := [m_1, m_2, m_3]^T$ is the vector with (v, k, p) -th coefficient given by

$$\frac{1}{R^3} g_v g_k g_p (i2\pi v)^{m_1} (i2\pi k)^{m_2} (i2\pi p)^{m_3} e^{-i2\pi(\beta v + \tau k + \nu p)}, \quad (31)$$

where the g_k are the coefficients of the squared Fejér kernel F in (29). Assuming $\mathbb{E}[\mathbf{A}^H \mathbf{A}] = \mathbf{I}$ we have

$$\mathbb{E}[G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_k)] = \frac{\partial^{m_1}}{\partial \beta^{m_1}} \frac{\partial^{m_2}}{\partial \tau^{m_2}} \frac{\partial^{m_3}}{\partial \nu^{m_3}} F(\beta - \beta_k) F(\nu - \nu_k) F(\tau - \tau_k) = \bar{G}^{(\mathbf{m})}(\mathbf{r} - \mathbf{r}_k). \quad (32)$$

We construct Q by interpolating the u_k at \mathbf{r}_k with the functions $G_{(0,0,0)}(\mathbf{r}, \mathbf{r}_k)$, $k = 1, \dots, S$, and slightly adopt this interpolation near the \mathbf{r}_k with linear combinations of the functions $G_{(1,0,0)}(\mathbf{r}, \mathbf{r}_k)$, $G_{(0,1,0)}(\mathbf{r}, \mathbf{r}_k)$, and $G_{(0,0,1)}(\mathbf{r}, \mathbf{r}_k)$, in order to ensure that local maxima of Q are achieved exactly at the \mathbf{r}_k :

$$Q(\mathbf{r}) = \sum_{k=1}^S \alpha_k G_{(0,0,0)}(\mathbf{r}, \mathbf{r}_k) + \alpha_{1k} G_{(1,0,0)}(\mathbf{r}, \mathbf{r}_k) + \alpha_{2k} G_{(0,1,0)}(\mathbf{r}, \mathbf{r}_k) + \alpha_{3k} G_{(0,0,1)}(\mathbf{r}, \mathbf{r}_k). \quad (33)$$

Since $Q(\mathbf{r})$ is a linear combination of the functions $G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_k) = \langle \mathbf{A}^H \mathbf{A} \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \mathbf{f}(\mathbf{r}) \rangle$, it is of the form $\langle \mathbf{A}^H \mathbf{q}, \mathbf{f}(\mathbf{r}) \rangle$ as required by (26).

We are now ready to state our key lemma, whose proof is deferred to Appendix A, providing sufficient conditions on \mathbf{A} ensuring that Q is indeed an appropriate dual certificate.

Lemma 1. *Let $\mathcal{S} = \{\mathbf{r}_1, \mathbf{r}_1, \dots, \mathbf{r}_S\} \subset [0, 1]^3$ be an arbitrary set of points obeying the minimum separation condition (15). Let $\mathbf{u} \in \mathbb{C}^S$ be a random vector, whose entries are chosen independently from symmetric distributions on the complex unit circle. Pick $\delta > 0$, and let \mathbf{A} be a random matrix that obeys*

$$\mathbb{P} \left[\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} |\langle (\mathbf{A}^H \mathbf{A} - \mathbf{I}) \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \mathbf{f}^{\mathbf{n}}(\mathbf{r}) \rangle| \geq \frac{c_1}{\sqrt{S \log(L/\delta)}} \right] \leq c_3 \frac{\delta}{L^{15}}, \quad (34a)$$

for all $\mathbf{r}_k, \mathbf{r} \in [0, 1]^3$, all $\mathbf{m}, \mathbf{n} \in \{0, 1, 2\}^3$ with $\|\mathbf{m}\|_1, \|\mathbf{n}\|_1 \leq 2$.

Here, $[\mathbf{f}^{\mathbf{n}}(\mathbf{r})]_{(v,k,p)} := (-i2\pi v)^{n_1} (-i2\pi k)^{n_2} (-i2\pi p)^{n_3} e^{-i2\pi(\beta v + \tau k + \nu p)}$ and $\kappa := \sqrt{|F^{(2)}(0)|}$. Moreover, suppose that for $\mathbf{r}_k \in [0, 1]^3$, there is a constant \hat{c} such that

$$\mathbb{P} \left[\max_{\mathbf{r} \in [0, 1]^3} |\langle \mathbf{A} \mathbf{g}_{\mathbf{0}}(\mathbf{r}_k), \mathbf{A} \mathbf{f}(\mathbf{r}) \rangle| \geq \frac{\hat{c}}{S} L^3 \right] \leq \frac{\delta}{8}. \quad (34b)$$

Then, with probability at least $1 - \delta$, there exists a trigonometric polynomial $Q(\mathbf{r}) = \langle \mathbf{A}^H \mathbf{q}, \mathbf{f}(\mathbf{r}) \rangle$, $\mathbf{r} = [\tau, \nu, \beta]^T$, with complex coefficients $\mathbf{q} \in \mathbb{C}^M$ obeying

$$Q(\mathbf{r}_k) = u_k, \text{ for all } \mathbf{r}_k \in \mathcal{S}, \text{ and } |Q(\mathbf{r})| < 1 \text{ for all } \mathbf{r} \in [0, 1]^3 \setminus \mathcal{S}. \quad (35)$$

Theorems 1 and 3 are proven in the coming sections by showing that the conditions (34a) and (34b) are satisfied for the corresponding random matrices \mathbf{A} .

6.2 Proof of Theorem 1: Isotropic and incoherent matrices

We prove Theorem 1 by showing that the isotropy and incoherence conditions imply conditions (34a) and (34b) of Lemma 1, which in turn implies the existence of an appropriate dual certificate satisfying the conditions of Proposition 1.

We start with verifying (34a). Note that

$$\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} \langle (\mathbf{A}^H \mathbf{A} - \mathbf{I}) \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \mathbf{f}^{\mathbf{n}}(\mathbf{r}) \rangle = \sum_{r=0}^{M-1} \underbrace{\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} (\mathbf{f}^{\mathbf{n}}(\mathbf{r}))^H \left(\mathbf{a}_r \mathbf{a}_r^H - \frac{1}{M} \mathbf{I} \right) \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k)}_{v_r}, \quad (36)$$

where \mathbf{a}_r^H are the rows of \mathbf{A} . From the isotropy property ($\mathbb{E} [\mathbf{a}_r \mathbf{a}_r^H] = \frac{1}{M} \mathbf{I}$), we have $\mathbb{E} [v_r] = 0$. Thus, (36) is a sum of independent zero-mean random variables, and we will use Bernstein's inequality to provide an upper bound. Define $\tilde{\mathbf{f}} := \frac{1}{\kappa \|\mathbf{n}\|_1 \sqrt{L^3}} \mathbf{f}^n(\mathbf{r})$ and $\tilde{\mathbf{g}} := \sqrt{L^3} \frac{1}{\kappa \|\mathbf{m}\|_1} \mathbf{g}_m(\mathbf{r}_k)$, for notational convenience. In order to apply Bernstein's inequality we need an upper bound on $|v_r|$. To derive this bound, first note that a simple application of Hölder's inequality yields

$$|v_r| \leq \left| \tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}} \right| \leq \left\| \tilde{\mathbf{f}} \right\|_\infty \left\| \tilde{\mathbf{g}} \right\|_\infty \left\| \mathbf{a}_r \right\|_1^2. \quad (37)$$

To bound $|v_r|$ we proceed by bounding each of the terms in (37). To this aim, first note that by using the fact that $\kappa = \sqrt{F''(0)} = \sqrt{\frac{\pi^2}{3}(N^2 + 4N)}$ ([CFG14, Eq. 2.23]) we can conclude that

$$\left(\frac{2\pi N}{\kappa} \right)^{\|\mathbf{m}+\mathbf{n}\|_1} \leq c', \quad (38)$$

holds for \mathbf{m}, \mathbf{n} obeying $\|\mathbf{m}\|_1, \|\mathbf{n}\|_1 \leq 2$. Using the latter inequality we arrive at

$$\left\| \tilde{\mathbf{f}} \right\|_\infty^2 = \left(\frac{\|\mathbf{f}^n(\mathbf{r})\|_\infty}{\kappa \|\mathbf{n}\|_1} \right)^2 \frac{1}{L^3} = \left(\frac{(2\pi N)^{\|\mathbf{n}\|_1}}{\kappa \|\mathbf{n}\|_1} \right)^2 \frac{1}{L^3} \leq \frac{\tilde{c}}{L^3}. \quad (39)$$

Similarly, the definition of $\mathbf{g}_m(\mathbf{r})$ in (31) together with the coefficients g_k obeying $|g_k| \leq 1$, implies that $\|\mathbf{g}_m(\mathbf{r}_k)\|_\infty \leq (2\pi N)^{\|\mathbf{m}\|_1} / R^3$. Thus, using the latter inequality together with inequality (38) and the fact that $\frac{L}{R} = \frac{2N+1}{N/2+1} \leq 4$ yields

$$\left\| \tilde{\mathbf{g}} \right\|_\infty^2 = L^3 \left(\frac{\|\mathbf{g}_m(\mathbf{r}_k)\|_\infty}{\kappa \|\mathbf{m}\|_1} \right)^2 \leq \frac{4^6}{L^3} \left(\frac{(2\pi N)^{\|\mathbf{m}\|_1}}{\kappa \|\mathbf{m}\|_1} \right)^2 \leq \frac{\tilde{c}}{L^3}. \quad (40)$$

Now utilizing (39) and (40) together with the incoherence condition (14) in (37) yields

$$|v_r| \leq \frac{c}{L^3} \left\| \mathbf{a}_r \right\|_1^2 \leq c \frac{\mu}{M} =: B. \quad (41)$$

In order to apply Bernstein's inequality, we also need an upper bound on $\sum_{r=0}^{M-1} \mathbb{E} [|v_r|^2]$. To obtain such a bound, note that

$$\begin{aligned} \sum_{r=0}^{M-1} \mathbb{E} [|v_r|^2] &\leq \sum_{r=0}^{M-1} \mathbb{E} [(\tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}})^2] \\ &\leq \sum_{r=0}^{M-1} \left\| \tilde{\mathbf{f}} \right\|_\infty \left\| \tilde{\mathbf{g}} \right\|_\infty \mathbb{E} [\left\| \mathbf{a}_r \right\|_1^2 |\tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}}|] \end{aligned} \quad (42)$$

$$\leq M \sqrt{\frac{\tilde{c}}{L^3}} \sqrt{\frac{\bar{c}}{L^3}} \mu \frac{L^3}{M} \mathbb{E} [|\tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}}|] = c' \mu \mathbb{E} [|\tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}}|]. \quad (43)$$

Here, (42) follows from Hölder's inequality, and for (43) we used the upper bounds (39) and (40)

together with the incoherence assumption (14). Finally note that

$$\begin{aligned} \mathbb{E} \left[\left| \tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}} \right| \right] &\leq \mathbb{E} \left[\max \left(\tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{f}}, \tilde{\mathbf{g}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}} \right) \right] \\ &\leq \mathbb{E} \left[\tilde{\mathbf{f}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{f}} + \tilde{\mathbf{g}}^H \mathbf{a}_r \mathbf{a}_r^H \tilde{\mathbf{g}} \right] \\ &\leq \tilde{\mathbf{f}}^H \mathbb{E} [\mathbf{a}_r \mathbf{a}_r^H] \tilde{\mathbf{f}} + \tilde{\mathbf{g}}^H \mathbb{E} [\mathbf{a}_r \mathbf{a}_r^H] \tilde{\mathbf{g}} \\ &= \frac{1}{M} \left(\|\tilde{\mathbf{f}}\|_2^2 + \|\tilde{\mathbf{g}}\|_2^2 \right) \end{aligned} \quad (44)$$

$$\leq \frac{c}{M}. \quad (45)$$

Here, (44) follows from the isotropy assumption, and for (45) we used the upper bounds (39) and (40) on the maximal value of $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ together with $\|\cdot\|_2 \leq \sqrt{L^3} \|\cdot\|_\infty$. Plugging (45) in (43) yields

$$\sum_{r=0}^{M-1} \mathbb{E} [|v_r|^2] \leq c \frac{\mu}{M} =: \sigma^2. \quad (46)$$

We now have all the elements to apply Bernstein's inequality. Before applying Bernstein's inequality first note that by assumption (16), we have

$$\frac{c_1}{\sqrt{S \log(L/\delta)}} \geq t := \sqrt{c' \frac{\mu}{M} \log(L/\delta)}. \quad (47)$$

Thus an application of Bernstein's inequality yields

$$\begin{aligned} \mathbb{P} \left[\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} |\langle (\mathbf{A}^H \mathbf{A} - \mathbf{I}) \mathbf{g}_m(\mathbf{r}_k), \mathbf{f}^n(\mathbf{r}) \rangle| \geq \frac{c_1}{\sqrt{S \log(L/\delta)}} \right] &\leq \mathbb{P} \left[\left| \sum_{r=0}^{M-1} v_r \right| \geq t \right] \\ &\leq 2 \exp \left(- \frac{\frac{1}{2} t^2}{\sigma^2 + \frac{1}{3} B t} \right) \\ &\leq 2 e^{-15 \log(L/\delta)} \\ &\leq 2 \frac{\delta}{L^{15}}. \end{aligned}$$

This concludes the proof of condition (34a).

We next show that condition (34b) holds. Note that

$$\begin{aligned} |\langle \mathbf{A} \mathbf{g}_0(\mathbf{r}_k), \mathbf{A} \mathbf{f}(\mathbf{r}) \rangle| &\leq \sum_{r=0}^{M-1} |\mathbf{f}^H(\mathbf{r}) \mathbf{a}_r \mathbf{a}_r^H \mathbf{g}_0(\mathbf{r}_k)| \\ &\leq \sum_{r=0}^{M-1} \|\tilde{\mathbf{f}}\|_\infty \|\mathbf{a}_r\|_1^2 \|\tilde{\mathbf{g}}\|_\infty \leq c \mu \leq \frac{M}{S \log^2(L/\delta)} \leq \frac{L^3}{S}. \end{aligned}$$

Here, we used incoherence ($\|\mathbf{a}_r\|_1^2 \leq \frac{L^3}{M} \mu$), the upper bounds (39) and (40), and finally assumption (16) and $M \leq L^3$. This concludes the proof of (34b).

6.3 Proof of Theorem 3: MIMO radar

We first show that the MIMO-radar input-output relation (12) obeys $\mathbf{y} = \mathbf{A}\mathbf{z}$, with $\mathbf{A} \in \mathbb{C}^{N_R L \times N_R N_T L^2}$ defined in (20). Recall that we assume that $L = N_R N_T$. First note that equation (12) can be written as

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_{M-1} \end{bmatrix} = \sum_{k=1}^S b'_k \begin{bmatrix} e^{i2\pi 0\beta_k} \sum_{j=0}^{N_T-1} e^{i2\pi j\beta} \mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_j \\ \vdots \\ e^{i2\pi N_T(N_R-1)\beta_k} \sum_{j=0}^{N_T-1} e^{i2\pi j\beta_k} \mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_j \end{bmatrix} \\ &= \sum_{k=1}^S b'_k e^{i2\pi \beta_k N} \underbrace{\begin{bmatrix} [\mathcal{F}_{\nu} \mathcal{T}_{\tau_k} \mathbf{x}_0 \dots \mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_{N_T-1}] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & [\mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_0 \dots \mathcal{F}_{\nu_k} \mathcal{T}_{\tau_k} \mathbf{x}_{N_T-1}] \end{bmatrix}}_{b_k} \mathbf{f}(\beta), \end{aligned} \quad (48)$$

where $[\mathbf{f}(\beta)]_j = e^{i2\pi \beta j}$, $j = -N, \dots, N$. With

$$\mathcal{F}_{\nu} \mathcal{T}_{\tau} \mathbf{x} = \mathbf{G}_{\mathbf{x}} \mathbf{F}^H \mathbf{f}([\tau, \nu]),$$

from (10), it is seen that equality (48) becomes

$$\mathbf{y} = \sum_{k=1}^S b_k \mathbf{A} \mathbf{f}(\mathbf{r}), \quad \mathbf{A} := \begin{bmatrix} \mathbf{B} & & \\ & \ddots & \\ & & \mathbf{B} \end{bmatrix}. \quad (49)$$

This concludes the proof of the MIMO-radar input-output relation (12) obeying relation $\mathbf{y} = \mathbf{A}\mathbf{z}$ with \mathbf{z} defended in (6).

The remainder of the proof of Theorem 3 is again similar to the proof of Theorem 1 in Section 6.2. Specifically, we show that the conditions (34a) and (34b) of Lemma 1 are satisfied, which in turn implies the existence of an appropriate dual certificate. For notational convenience, we prove that those conditions are satisfied for the normalized measurement model

$$\bar{\mathbf{y}} = \bar{\mathbf{A}} \mathbf{z}, \text{ where } \bar{\mathbf{y}} := \sqrt{L^2 N_T} \mathbf{y} \text{ and } \bar{\mathbf{A}} := \sqrt{L^2 N_T} \mathbf{A}.$$

In a nutshell, we first show that $\langle \bar{\mathbf{A}} \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \bar{\mathbf{A}} \mathbf{f}^n(\mathbf{r}) \rangle$ in conditions (34a) and (34b) can be expressed as the quadratic form (see Section 6.3.1)

$$\langle \bar{\mathbf{A}} \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \bar{\mathbf{A}} \mathbf{f}^n(\mathbf{r}) \rangle = \mathbf{x}^H \mathbf{V}_{\mathbf{m}}^{(n)}(\mathbf{r}, \mathbf{r}_k) \mathbf{x}, \quad (50)$$

where $\mathbf{V}_{\mathbf{m}}^{(n)}(\mathbf{r}, \mathbf{r}_k) \in \mathbb{C}^{LN_T \times LN_T}$ is defined in equation (64) in Section 6.3.1 below, and $\mathbf{x} := [\mathbf{x}_0^T, \dots, \mathbf{x}_{N_T-1}^T]^T$ is a Gaussian random vector (recall that the probing signals are Gaussian). Conditions (34a) and (34b) then follow by utilizing the Hanson-Wright inequality stated in the lemma below, which ensures that this quadratic form does not deviate too much from its expectation

$$\mathbb{E} [\langle \bar{\mathbf{A}} \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \bar{\mathbf{A}} \mathbf{f}^n(\mathbf{r}) \rangle] = \langle \mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \mathbf{f}^n(\mathbf{r}) \rangle. \quad (51)$$

Equality in (51) holds by $\mathbb{E} [\bar{\mathbf{A}}^H \bar{\mathbf{A}}] = \mathbf{I}$, which in turn follows from $\mathbb{E} [\mathbf{G}_{\mathbf{x}_k}^H \mathbf{G}_{\mathbf{x}_k}] = \frac{1}{N_T} \mathbf{I}$ [HMS16, Sec. 8.1] by noting that

$$\mathbb{E} [(\mathbf{G}_{\mathbf{x}_k} \mathbf{F}^H)^H \mathbf{G}_{\mathbf{x}_k} \mathbf{F}^H] = \mathbf{F} \mathbb{E} [\mathbf{G}_{\mathbf{x}_k}^H \mathbf{G}_{\mathbf{x}_k}] \mathbf{F}^H = \frac{1}{L^2 N_T} \mathbf{I}$$

and

$$\mathbb{E} \left[(\mathbf{G}_{\mathbf{x}_k} \mathbf{F}^H)^H \mathbf{G}_{\mathbf{x}_{k'}} \mathbf{F}^H \right] = \mathbf{0}, \quad \text{for } k \neq k'.$$

Lemma 2 (Hanson-Wright inequality [RV13, Thm. 1.1]). *Let $\mathbf{x} \in \mathbb{R}^{LN_T}$ be a random vector with independent zero-mean K -sub-Gaussian entries (i.e., the entries obey $\sup_{p \geq 1} p^{-1} (\mathbb{E} [|x_\ell|^p])^{1/p} \leq K$), and let \mathbf{V} be an $LN_T \times LN_T$ matrix. Then, for all $t \geq 0$,*

$$\mathbb{P} [| \mathbf{x}^T \mathbf{V} \mathbf{x} - \mathbb{E} [\mathbf{x}^T \mathbf{V} \mathbf{x}] | > t] \leq 2 \exp \left(-\tilde{c} \min \left(\frac{t^2}{K^4 \|\mathbf{V}\|_F^2}, \frac{t}{K^2 \|\mathbf{V}\|} \right) \right),$$

where \tilde{c} is a numerical constant.

We next detail how conditions (34a) and (34b) follow from the Hanson-Wright inequality. In order to evaluate the RHS of the Hanson-Wright inequality, we need the following upper bound on the norm of $\mathbf{V}_m^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_k)$, proven in Section 6.3.2 below.

Lemma 3. *For all $\mathbf{r}, \mathbf{r}_k \in [0, 1]^3$, and for all positive integer vectors \mathbf{n}, \mathbf{m} with $\|\mathbf{n} + \mathbf{m}\|_1 \leq 4$, we have*

$$\left\| \mathbf{V}_m^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_k) \right\|_F \leq c_4 (2\pi N)^{\|\mathbf{m} + \mathbf{n}\|_1} N_T \sqrt{L}. \quad (52)$$

We are now ready to establish (34a) by applying the Hanson-Wright inequality. Setting $\mathbf{V} := \mathbf{V}_m^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_k)$ for ease of presentation, it follows that

$$\begin{aligned} & \mathbb{P} \left[\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} |\langle (\bar{\mathbf{A}}^H \bar{\mathbf{A}} - \mathbf{I}) \mathbf{g}_m(\mathbf{r}_k), \mathbf{f}^n(\mathbf{r}) \rangle| > \frac{c_1}{\sqrt{S \log(L/\delta)}} \right] \\ & \leq \mathbb{P} \left[\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} |\langle (\bar{\mathbf{A}}^H \bar{\mathbf{A}} - \mathbf{I}) \mathbf{g}_m(\mathbf{r}_k), \mathbf{f}^n(\mathbf{r}) \rangle| > c_4 c' \frac{\alpha}{\sqrt{L}} \right] \end{aligned} \quad (53)$$

$$\leq \mathbb{P} \left[|\langle (\bar{\mathbf{A}}^H \bar{\mathbf{A}} - \mathbf{I}) \mathbf{g}_m(\mathbf{r}_k), \mathbf{f}^n(\mathbf{r}) \rangle| > c_4 (2\pi N)^{\|\mathbf{m} + \mathbf{n}\|_1} \frac{\alpha}{\sqrt{L}} \right] \quad (54)$$

$$\leq \mathbb{P} \left[|\mathbf{x}^T \mathbf{V} \mathbf{x} - \mathbb{E} [\mathbf{x}^T \mathbf{V} \mathbf{x}]| > \|\mathbf{V}\|_F \frac{\alpha}{LN_T} \right] \quad (55)$$

$$\leq 2 \exp \left(-\tilde{c} \min \left(\frac{\|\mathbf{V}\|_F^2 \alpha^2}{(LN_T)^2 K^4 \|\mathbf{V}\|_F^2}, \frac{\|\mathbf{V}\|_F \alpha}{LN_T K^2 \|\mathbf{V}\|} \right) \right) \quad (56)$$

$$\leq 2 \exp \left(-\tilde{c} \min \left(\frac{\alpha^2}{c_5^4}, \frac{\alpha}{c_5^2} \right) \right). \quad (57)$$

$$\leq c_3 \frac{\delta}{L^{15}} \quad (58)$$

Here, (53) follows by assumption (22):

$$\frac{c_1}{\sqrt{S \log(L/\delta)}} \geq \frac{c_1}{\sqrt{\log(L/\delta)}} \frac{\sqrt{c \log^3(L/\delta)}}{\sqrt{L}} \geq c_4 c' \frac{\alpha}{\sqrt{L}}, \quad (59)$$

where we set $\alpha = \frac{c_5^2}{\hat{c}} \log(L^{15}/(2c_3\delta))$. Equation (54) follows from $(2\pi N)^{\|\mathbf{m}+\mathbf{n}\|_1} \leq c'\kappa^{\|\mathbf{m}+\mathbf{n}\|_1}$, by (38). Moreover, (55) follows from (50), (51), and (52). To obtain (56), we used Lemma 2 with $t = \|\mathbf{V}\|_F \frac{\alpha}{LN_T}$. Equation (57) holds because the sub-Gaussian parameter K of the random variable $[\mathbf{x}]_\ell \sim \mathcal{N}(0, 1/(LN_T))$ is given by $K = c_5/\sqrt{LN_T}$ (see e.g., [Ver12, Ex. 5.8]) and $\|\mathbf{V}\|_F/\|\mathbf{V}\| \geq 1$. Finally, (58) follows by using the fact that for $\alpha \geq c_5^2$, we have $\min\left(\frac{\alpha^2}{c_5^4}, \frac{\alpha}{c_5^2}\right) = \frac{\alpha}{c_5^2}$.

We next verify (34b). We have

$$\begin{aligned} \mathbb{P} \left[\max_{\mathbf{r} \in [0,1]^3} |\langle \bar{\mathbf{A}}\mathbf{g}_0(\mathbf{r}_k), \bar{\mathbf{A}}\mathbf{f}(\mathbf{r}) \rangle| \geq \frac{\hat{c}}{S} L^3 \right] &= \mathbb{P} \left[\max_{\mathbf{r} \in [0,1]^3} \left| \mathbf{x}^H \mathbf{V}_0^{(0)}(\mathbf{r}, \mathbf{r}_k) \mathbf{x} \right| \geq \frac{\hat{c}}{S} L^3 \right] \\ &\leq \mathbb{P} \left[\left\| \mathbf{V}_0^{(0)}(\mathbf{r}, \mathbf{r}_k) \right\| \|\mathbf{x}\|_2^2 \geq \frac{\hat{c}}{S} L^3 \right] \\ &\leq \mathbb{P} \left[c_4 N_T \sqrt{L} \|\mathbf{x}\|_2^2 \geq \frac{\hat{c}}{S} L^3 \right] \end{aligned} \quad (60)$$

$$\leq \mathbb{P} \left[c_4 N_T \sqrt{L} \|\mathbf{x}\|_2^2 \geq \frac{\hat{c} c \log^3(L/\delta)}{L} L^3 \right] \quad (61)$$

$$\begin{aligned} &\leq \mathbb{P} \left[\|\mathbf{x}\|_2^2 \geq c_5 \log^3(L/\delta) \sqrt{L} \right] \\ &\leq \frac{\delta}{8}. \end{aligned} \quad (62)$$

Here (60) follows from (52) and $\|\cdot\| \leq \|\cdot\|_F$, and (61) follows by the assumption (22). Finally, (62) follows from a standard concentration bound of a χ^2 -random variable.

6.3.1 Proof of Equation (50)

In this section we prove that $\langle \bar{\mathbf{A}}\mathbf{g}_m(\mathbf{r}_k), \bar{\mathbf{A}}\mathbf{f}^n(\mathbf{r}) \rangle$ can be expressed as the quadratic form in \mathbf{x} . To this end, we note that the vector $\mathbf{g}_m(\mathbf{r}_k)$ defined in (31) can be written as

$$\mathbf{g}_m(\mathbf{r}_k) = \mathbf{g}_{m_1}(\beta_k) \otimes \mathbf{g}_{m_3}(\nu_k) \otimes \mathbf{g}_{m_2}(\tau_k) \in \mathbb{C}^{L^3}, \quad \mathbf{r}_k = [\tau_k, \nu_k, \beta_k]^T,$$

where \otimes denotes the Kronecker product, and

$$[\mathbf{g}_m(\tau)]_r := \frac{1}{R} g_r e^{-i2\pi\tau r} (i2\pi r)^m, \quad r = -N, \dots, N.$$

As shown the paper [HMS16, p. 25])

$$\begin{aligned} &[L\mathbf{G}_{\mathbf{x}_j} \mathbf{F}^H \mathbf{g}_{m_3}(\nu) \otimes \mathbf{g}_{m_2}(\tau)]_p \\ &= \frac{1}{L} \sum_{k,\ell=-N}^N \left(\sum_{r,q=-N}^N g_r g_q e^{-i2\pi(\tau r + \nu q)} (i2\pi r)^{m_2} (i2\pi q)^{m_3} e^{i2\pi \frac{qk+r\ell}{L}} \right) [\mathbf{x}_j]_{p-\ell} e^{i2\pi \frac{kp}{L}} \\ &= \sum_{\ell=-N}^N [\mathbf{x}_j]_\ell \underbrace{\sum_{r=-N}^N e^{i2\pi \frac{r(p-\ell)}{L}} g_r g_p e^{-i2\pi(\tau r - \nu p)} (i2\pi r)^{m_2} (-i2\pi p)^{m_3}}_{h_{p,\ell} :=}, \end{aligned} \quad (63)$$

where we used that $\frac{1}{L} \sum_{k=-N}^N e^{i2\pi \frac{k(p+q)}{L}}$ is equal to 1 if $p = -q$ and equal to 0 otherwise, together with the fact that $[\mathbf{x}_j]_\ell$ is L -periodic in ℓ . We next write (63) in matrix-vector form according to

$$L\mathbf{G}_{\mathbf{x}_j}\mathbf{F}^H(\mathbf{g}_{m_3}(\nu) \otimes \mathbf{g}_{m_2}(\tau)) = \mathbf{H}_{\mathbf{g}}\mathbf{x}_j,$$

where the (p, ℓ) -th entry of $\mathbf{H}_{\mathbf{g}} \in \mathbb{C}^{L \times L}$ is given by $h_{p,\ell}$ (note that $\mathbf{H}_{\mathbf{g}}$ is a function of $\mathbf{g}_{m_3}(\nu) \otimes \mathbf{g}_{m_2}(\tau)$). Recall that $\mathbf{x} = [\mathbf{x}_0^T, \dots, \mathbf{x}_{N_T-1}^T]^T \in \mathbb{C}^{LN_T}$. With this notation, we have

$$\begin{aligned} & L\mathbf{A}\mathbf{g}_{\mathbf{m}}(\mathbf{r}_j) \\ &= L \begin{bmatrix} [\mathbf{G}_{\mathbf{x}_0}\mathbf{F}^H, \dots, \mathbf{G}_{\mathbf{x}_{N_T-1}}\mathbf{F}^H] \\ \ddots \\ [\mathbf{G}_{\mathbf{x}_0}\mathbf{F}^H, \dots, \mathbf{G}_{\mathbf{x}_{N_T-1}}\mathbf{F}^H] \end{bmatrix} \mathbf{g}_{m_1}(\beta_j) \otimes \mathbf{g}_{m_3}(\nu_j) \otimes \mathbf{g}_{m_2}(\tau_j) \\ &= L \begin{bmatrix} \ddots \\ [\dots \mathbf{G}_{\mathbf{x}_k}\mathbf{F}^H(\mathbf{g}_{m_3}(\nu_j) \otimes \mathbf{g}_{m_2}(\tau_j)) \dots] \\ \ddots \end{bmatrix} \mathbf{g}_{m_1}(\beta_j) \\ &= \begin{bmatrix} \ddots \\ [\mathbf{H}_{\mathbf{g}}\mathbf{x}_0, \dots, \mathbf{H}_{\mathbf{g}}\mathbf{x}_{N_T-1}] \\ \ddots \end{bmatrix} \mathbf{g}_{m_1}(\beta_j) \\ &= \mathbf{B}_{\mathbf{g}}\mathbf{x}, \end{aligned}$$

where $\mathbf{B}_{\mathbf{g}} \in \mathbb{C}^{N_R L \times N_T L}$ is a block matrix where the (ℓ, k) -th block, $\ell = -\frac{N_R-1}{2}, \dots, \frac{N_R-1}{2}$, $k = -\frac{N_T-1}{2}, \dots, \frac{N_T-1}{2}$, is given by $e^{i2\pi\beta(k+\ell N_T)}\mathbf{H}_{\mathbf{g}}$.

Analogously, we define the matrix $\mathbf{H}_{\mathbf{f}} \in \mathbb{C}^{L \times L}$ such that

$$L\mathbf{G}_{\mathbf{x}_j}\mathbf{F}^{(m_3)}(\nu) \otimes \mathbf{f}^{(m_2)}(\tau) = \mathbf{H}_{\mathbf{f}}\mathbf{x}_j,$$

and a block matrix $\mathbf{B}_{\mathbf{f}} \in \mathbb{C}^{N_R L \times N_T L}$ with (ℓ, k) -th block given by $e^{i2\pi\beta(k+\ell N_T)}\mathbf{H}_{\mathbf{f}}$, such that

$$L\mathbf{A}\mathbf{f}^{(\mathbf{n})}(\mathbf{r}) = \mathbf{B}_{\mathbf{f}}\mathbf{x}.$$

With this notation,

$$\langle \bar{\mathbf{A}}\mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), \bar{\mathbf{A}}\mathbf{f}^{(\mathbf{n})}(\mathbf{r}) \rangle = N_T \langle L\mathbf{A}\mathbf{g}_{\mathbf{m}}(\mathbf{r}_k), L\mathbf{A}\mathbf{f}^{(\mathbf{n})}(\mathbf{r}) \rangle = \mathbf{x}^H \mathbf{V}_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j) \mathbf{x},$$

with

$$\mathbf{V}_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j) := N_T \mathbf{B}_{\mathbf{f}}^H \mathbf{B}_{\mathbf{g}}. \quad (64)$$

This concludes the proof of equation (50).

6.3.2 Proof of Lemma 3

We are now ready to prove Lemma 3. First note that from the previous section,

$$\mathbf{V}_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j) = \frac{N_T}{R} \mathbf{D} \otimes \frac{1}{R^2} \mathbf{H}_{\mathbf{f}}^H \mathbf{H}_{\mathbf{g}},$$

with $\mathbf{D} \in \mathbb{C}^{N_T \times N_T}$ defined as

$$\mathbf{D} := \frac{N_T}{R} \sum_{\ell=-(N_R-1)/2}^{(N_R-1)/2} \begin{bmatrix} & \vdots \\ (i2\pi(k+\ell N_T))^{n_1} e^{i2\pi\beta(k+\ell N_T)} & \\ & \vdots \end{bmatrix} [\dots (i2\pi(k'+\ell N_T))^{m_1} g_{k'+\ell N_T} \dots],$$

where $k, k' = -(N_T-1)/2, \dots, (N_T-1)/2$ are the indices of the vectors above. With $k+\ell N_T \leq N$, and using that $g_k \leq 1$ we have

$$\|\mathbf{D}\|_F \leq \frac{N_T}{R} N_R \sqrt{N_T} (2\pi N)^{n_1} \sqrt{N_T} (2\pi N)^{m_1} \leq 4(2\pi N)^{m_1+n_1} N_T,$$

where we used that $N_R N_T = L = 2N + 1 \leq 4R$. It follows that

$$\left\| \mathbf{V}_m^{(n)}(\mathbf{r}, \mathbf{r}_j) \right\|_F = \left\| \frac{N_T}{R} \mathbf{D} \right\|_F \left\| \frac{1}{R^2} \mathbf{H}_f^H \mathbf{H}_g \right\|_F \leq c_1 (2\pi N)^{\|\mathbf{m}+\mathbf{n}\|_1} N_T \sqrt{L},$$

which concludes the proof. For the inequality above, we used [HMS16, Lem. 3] which states that $\left\| \frac{1}{R^2} \mathbf{H}_f^H \mathbf{H}_g \right\|_F \leq c_2 (2\pi N)^{m_2+m_3+n_2+n_3} \sqrt{L}$.

7 Proof of Theorem 2

To prove Theorem 2 we need a few definitions.

Given a convex function f , the *set of descent* of f at a point \mathbf{x} is defined as

$$\mathcal{D}_f(\mathbf{x}) := \left\{ \mathbf{h} : f(\mathbf{x} + \mathbf{h}) \leq f(\mathbf{x}) \right\}.$$

The *tangent cone* $\mathcal{T}_f(\mathbf{x})$ is the conic hull of the descent set. That is, the smallest closed cone $\mathcal{T}_f(\mathbf{x})$ obeying $\mathcal{D}_f(\mathbf{x}) \subset \mathcal{T}_f(\mathbf{x})$. The Gaussian width of a set $\mathcal{C} \subset \mathbb{R}^n$ is defined as:

$$\omega(\mathcal{C}) := \mathbb{E} \left[\sup_{\mathbf{z} \in \mathcal{C}} \langle \mathbf{w}, \mathbf{z} \rangle \right],$$

where the expectation is taken over $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For the remainder of this section, we set $f(\cdot) = \|\cdot\|_{\mathcal{A}}$.

We now state a theorem from [Cha+12] characterizing the sample complexity of structured signal recovery problems from linear measurements. This theorem is based on Gordon's escape through the mesh lemma [Gor88] and a standard Gaussian concentration inequality for Lipschitz functions.

Theorem 6 ([Cha+12, Cor. 3.3]). *Let $\mathbf{A} \in \mathbb{R}^{M \times L}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1/M)$ entries. Assume we have M linear measurements of the form $\mathbf{y} = \mathbf{A}\mathbf{z}$ from an unknown signal \mathbf{z} . Let $\hat{\mathbf{z}}$ be a solution to the atomic norm minimization problem $\text{AN}(\mathbf{y})$, i.e.,*

$$\hat{\mathbf{z}} = \arg \min_{\bar{\mathbf{z}}} \|\mathbf{z}\|_{\mathcal{A}} \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\bar{\mathbf{z}}.$$

Fix $\eta > 0$. Then, as long as

$$M \geq (\omega(\mathcal{T}_{\mathcal{A}}(\mathbf{z}) \cap \mathbb{S}^{L-1}) + \eta)^2 + 1, \tag{65}$$

with probability at least $1 - e^{-\frac{\eta^2}{2}}$ the optimization problem $\text{AN}(\mathbf{z})$ recovers the unknown signal, i.e., $\hat{\mathbf{z}} = \mathbf{z}$. Here, \mathbb{S}^{L-1} is the unit sphere and $\mathcal{T}_{\mathcal{A}}(\mathbf{z})$ is the tangent cone associated with the atomic norm, i.e., $\mathcal{T}_{\mathcal{A}}(\mathbf{z}) = \mathcal{T}_f(\mathbf{z})$ with $f(\cdot) = \|\cdot\|_{\mathcal{A}}$.

As long as

$$M \geq 4\omega^2(\mathcal{T}_{\mathcal{A}}(\mathbf{z}) \cap \mathbb{S}^{L-1}), \quad (66)$$

setting $\eta = \sqrt{M}/2$ and applying the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$\begin{aligned} M &= \frac{M}{2} + \frac{M-2}{2} + 1 \geq 2\omega^2(\mathcal{T}_{\mathcal{A}}(\mathbf{z}) \cap \mathbb{S}^{L-1}) + \frac{M-2}{2} + 1 \\ &\geq \left(\omega(\mathcal{T}_{\mathcal{A}}(\mathbf{z}) \cap \mathbb{S}^{L-1}) + \frac{\sqrt{M-2}}{2} \right)^2 + 1 = (\omega(\mathcal{T}_{\mathcal{A}}(\mathbf{z}) \cap \mathbb{S}^{L-1}) + \eta)^2 + 1. \end{aligned}$$

Thus, by Theorem 6 condition (66) ensures exact recovery via atomic norm minimization with probability at least $1 - e^{-(M-2)/8}$.

We next upper bound the Gaussian width using a connection to denoising with ℓ_1 -regularized minimization from Oymak and Hassibi [OH13]. To this end, we define the proximity operator

$$\tilde{\mathbf{z}}_f(\mathbf{y}, \lambda) = \arg \min_{\bar{\mathbf{z}}} \frac{1}{2} \|\mathbf{y} - \bar{\mathbf{z}}\|_2^2 + \lambda f(\bar{\mathbf{z}}).$$

Oymak and Hassibi [OH13, Thm. 1.1, Prop. 5.2] characterize the mean-square distance (the mean square distance of a set \mathcal{C} is defined as $D(\mathcal{C}) := \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0, \mathbf{I})} [\min_{\mathbf{s} \in \mathcal{C}} \|\mathbf{w} - \mathbf{s}\|_2^2]$) of the scaled sub-differential $\partial f(\mathbf{z})$ of f at \mathbf{x} as

$$D(\lambda' \partial f(\mathbf{z})) = \max_{\sigma > 0} \frac{\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0, \mathbf{I})} [\|\tilde{\mathbf{z}}_f(\mathbf{z} + \sigma \mathbf{w}, \sigma \lambda') - \mathbf{z}\|_2^2]}{\sigma^2}.$$

Due to

$$D(\lambda' \partial f(\mathbf{z})) \geq D(\text{cone}(\partial f(\mathbf{z}))) = D(\mathcal{T}_f^\circ(\mathbf{z})) \geq \omega^2(\mathcal{T}_f(\mathbf{z}) \cap \mathbb{S}^{L-1}), \quad (67)$$

this yields an upper bound on the Gaussian width. Here, the first inequality and equality, respectively, can be found in [OH13, Eq. (3.5)] and [OH13, p. 18] (note that $\text{cone}(\partial f(\mathbf{z}))$ is a closed, convex cone, and \mathcal{C}° denotes the dual cone of \mathcal{C}).

Thus, the condition

$$M \geq 4 \min_{\lambda'} \max_{\sigma > 0} \frac{\mathbb{E}_{\mathbf{x} \sim \mathcal{N}(0, \mathbf{I})} [\|\tilde{\mathbf{z}}_f(\mathbf{z} + \sigma \mathbf{w}, \sigma \lambda') - \mathbf{z}\|_2^2]}{\sigma^2} \quad (68)$$

implies (66), and therefore, (68) is sufficient to guarantee that the estimate $\hat{\mathbf{z}}$ of $\text{AN}(\mathbf{y})$ is exact, i.e., $\hat{\mathbf{z}} = \mathbf{z}$ with probability at least $1 - e^{-(M-2)/8}$.

We now state a lemma that controls the right hand-side of (68) when the frequencies of the mixture components are sufficiently separated. The proof of this lemma is essentially a consequence of results established in [TBR15] and is deferred to Section C.

Lemma 4. Let \mathbf{z} be of the form (1) with the frequencies of the mixtures obeying the minimum separation condition (3). Consider the proximity operator with $f(\cdot) = \|\cdot\|_{\mathcal{A}}$, defined as

$$\tilde{\mathbf{z}}(\mathbf{y}, \lambda) = \arg \min_{\bar{\mathbf{z}}} \frac{1}{2} \|\mathbf{y} - \bar{\mathbf{z}}\|_2^2 + \lambda \|\bar{\mathbf{z}}\|_{\mathcal{A}}. \quad (69)$$

Let $\lambda = c\sigma\sqrt{L \log L}$, where c is a fixed numerical constant. Then the proximity operator obeys

$$\frac{\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0, \mathbf{I})} \left[\|\tilde{\mathbf{z}}(\mathbf{z} + \sigma\mathbf{w}, \lambda) - \mathbf{z}\|_2^2 \right]}{\sigma^2} \leq cS \log L. \quad (70)$$

Combining Lemma 4 above with (68) completes the proof. All that remains is to prove Lemma 4 which is the subject of Appendix C.

8 Acknowledgements

The work of RH was supported by the Swiss National Science Foundation under grant P2EZP2_159065. Theorems 3 and 5 on the MIMO radar problem and the numerical results in Section 5 were presented at the 2016 IEEE International Symposium on Information Theory (ISIT) [Hec16].

References

- [AL13] N. Ailon and E. Liberty. “An almost optimal unrestricted fast Johnson-Lindenstrauss transform”. In: ACM Trans. Algorithms 9.3 (2013), 21:1–21:12.
- [ASB15] C. Aubel, D. Stotz, and H. Bölcskei. “A theory of super-resolution from short-time Fourier transform measurements”. In: arXiv:1509.01047 (2015).
- [Bel63] P. A. Bello. “Characterization of randomly time-variant linear channels”. In: IEEE Trans. Commun. Syst. 11.4 (1963), pp. 360–393.
- [BTR13] B. N. Bhaskar, G. Tang, and B. Recht. “Atomic norm denoising with applications to line spectral estimation”. In: IEEE Trans. Signal Process. 61.23 (2013), pp. 5987–5999.
- [BF03] D. W. Bliss and K. W. Forsythe. “Multiple-input multiple-output (MIMO) radar and imaging”. In: Asilomar Conf. on Signals, Syst. and Comput. 2003, pp. 54–59.
- [Cai+16] J.-F. Cai, X. Qu, W. Xu, and G.-B. Ye. “Robust recovery of complex exponential signals from random Gaussian projections via low rank Hankel matrix reconstruction”. In: Appl. Comput. Harm. Anal. 41.2 (2016), pp. 470–490.
- [CFG14] E. J. Candès and C. Fernandez-Granda. “Towards a mathematical theory of super-resolution”. In: Comm. Pure Appl. Math. 67.6 (2014), pp. 906–956.
- [CFG13] E. J. Candès and C. Fernandez-Granda. “Super-resolution from noisy data”. In: J. Fourier Anal. Appl. 19.6 (2013), pp. 1229–1254.
- [CRT06] E. J. Candès, J. Romberg, and T. Tao. “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. In: IEEE Trans. Inf. Theory 52.2 (2006), pp. 489–509.

- [CP11] E. J. Candès and Y. Plan. “A probabilistic and RIPless theory of compressed sensing”. In: IEEE Trans. Inf. Theory 57.11 (2011).
- [Cha+12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. “The convex geometry of linear inverse problems”. In: Found. Comput. Math. 12.6 (2012), pp. 805–849.
- [Dir15] S. Dirksen. “Tail bounds via generic chaining”. In: Electron. J. Probab. 20 (2015).
- [Don92] D. L. Donoho. “Superresolution via sparsity constraints”. In: SIAM J. on Math. Anal. 23.5 (1992), pp. 1309–1331.
- [DR16] D. Dorsch and H. Rauhut. “Refined analysis of sparse MIMO radar”. In: J. Fourier Anal. Appl. (2016).
- [FG16] C. Fernandez-Granda. “Super-resolution of point sources via convex programming”. In: Inf. Inference (2016).
- [FR13] S. Foucart and Rauhut, H. *A Mathematical Introduction to Compressive Sensing*. New York, NY: Springer, 2013.
- [GS05] A. Gershman and N. Sidiropoulos, eds. *Space-Time Processing for MIMO Communications*. English. Chichester; Hoboken, NJ: Wiley, 2005.
- [Gor88] Y. Gordon. “On Milman’s inequality and random subspaces which escape through a mesh in \mathbb{R}^n ”. In: Geom. Aspects of Func. Anal. Lect. Notes Math. 1317. Springer, 1988, pp. 84–106.
- [Gor85] Y. Gordon. “Some inequalities for Gaussian processes and applications”. In: Israel J. Math 50.4 (1985), pp. 265–289.
- [Har96] L. A. Harris. “Bernstein’s polynomial inequalities and functional analysis”. In: Irish Math. Soc. Bull. 36 (1996), pp. 19–33.
- [Hec16] R. Heckel. “Super-resolution MIMO radar”. In: IEEE Int. Symp. Inf. Theory. 2016, pp. 1416–1420.
- [HB13] R. Heckel and H. Bölskei. “Identification of sparse linear operators”. In: IEEE Trans. Inf. Theory 59.12 (2013), pp. 7985–8000.
- [HMS16] R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi. “Super-resolution Radar”. In: Inf. Inference 5.1 (2016), pp. 22–75.
- [HS09] M. Herman and T. Strohmer. “High-resolution Radar via compressed sensing”. In: IEEE Trans. Signal Process. 57.6 (2009), pp. 2275–2284.
- [HJ12] R. A. Horn and C. R. Johnson. *Matrix Analysis*. English. 2nd. New York, NY: Cambridge University Press, 2012.
- [KMR14] F. Krahmer, S. Mendelson, and H. Rauhut. “Suprema of chaos processes and the restricted isometry property”. In: Commun. Pur. Appl. Math. 67.11 (2014), pp. 1877–1904.
- [KW11] F. Krahmer and R. Ward. “New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property”. In: SIAM J. on Math. Anal. 43.3 (2011), pp. 1269–1281.
- [LS07] J. Li and P. Stoica. “MIMO Radar with colocated antennas”. In: IEEE Signal Process. Mag. 24.5 (2007), pp. 106–114.

- [LC16] Y. Li and Y. Chi. “Off-the-grid line spectrum denoising and estimation with multiple measurement vectors”. In: IEEE Trans. Signal Process. 64.5 (2016), pp. 1257–1269.
- [OH13] S. Oymak and B. Hassibi. “Sharp MSE bounds for proximal denoising”. In: Found. Comut. Math. 16.4 (2013), pp. 965–1029.
- [ORS15a] S. Oymak, B. Recht, and M. Soltanolkotabi. “Isometric sketching of any set via the Restricted Isometry Property”. In: arXiv:1506.03521 (2015).
- [ORS15b] S. Oymak, B. Recht, and M. Soltanolkotabi. “Sharp time–data tradeoffs for linear inverse problems”. In: arXiv:1507.04793 (2015).
- [Rau07] H. Rauhut. “Random sampling of sparse trigonometric polynomials”. In: Appl. Comput. Harmon. Anal. 22.1 (2007), pp. 16–42.
- [Rob+10] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. Sadjadi. “Iterative adaptive approaches to MIMO radar imaging”. In: IEEE J. Sel. Topics Signal Process 4.1 (2010), pp. 5–20.
- [RV13] M. Rudelson and R. Vershynin. “Hanson-Wright inequality and sub-Gaussian concentration”. In: Electron. Commun. Probab. 18.0 (2013).
- [SM05] P. Stoica and R. L. Moses. *Spectral Analysis of Signals*. P.H., 2005.
- [SW15] T. Strohmer and H. Wang. “Adventures in compressive sensing based MIMO Radar”. In: *Excursions in Harm. Anal.* Appl. Num. Harm. Anal. 2015, pp. 285–326.
- [TBR15] G. Tang, B. N. Bhaskar, and B. Recht. “Near minimax line spectral estimation”. In: IEEE Trans. Inf. Theory 61.1 (2015), pp. 499–512.
- [Tan+13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. “Compressed sensing off the grid”. In: IEEE Trans. Inform. Theory 59.11 (2013), pp. 7465–7490.
- [Tro04] J. A. Tropp. “Greed is good: Algorithmic results for sparse approximation”. In: IEEE Trans. Inform. Theory 50.10 (2004), pp. 2231–2242.
- [Ver12] R. Vershynin. “Introduction to the non-asymptotic analysis of random matrices”. In: *Compressed sensing: Theory and applications*. New York, NY: Cambridge University Press, 2012, pp. 210–268.
- [YXS16] Z. Yang, L. Xie, and P. Stoica. “Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution”. In: IEEE Trans. Inf. Theory 62.6 (2016), pp. 3685–3701.
- [Yar+10] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer. “Source Localization and Sensing: A Nonparametric Iterative Adaptive Approach Based on Weighted Least Squares”. In: IEEE Transactions on Aerospace and Electronic Systems 46.1 (2010), pp. 425–443.

A Proof of Lemma 1

In this section we prove our main technical result. As explained in Section 6.1, we construct Q explicitly according to (33). We proceed as follows:

Step 1: We show that, with probability at least $1 - \frac{\delta}{2}$ there exists a choice of coefficients $\alpha_k, \alpha_{1k}, and \alpha_{2k}$ and α_{3k} such that

$$Q(\mathbf{r}_k) = u_k \quad \text{and} \quad \nabla Q(\mathbf{r}_k) = \mathbf{0} \quad \text{for all } \mathbf{r}_k \in \mathcal{S}. \quad (71)$$

This is necessary to ensure that Q reaches local maxima at the \mathbf{r}_k , which is required for the interpolation and boundedness condition in (35) to hold simultaneously.

Step 2: We conclude the proof by showing that with the coefficients from Step 1, with probability at least $1 - \frac{\delta}{2}$, the polynomial Q with the coefficients from step 1 obeys $|Q(\mathbf{r})| < 1$ uniformly for all $\mathbf{r} \in [0, 1]^3 \setminus \mathcal{S}$. This is accomplished using an ϵ -net argument:

Step 2a: Let $\Omega \subset [0, 1]^3$ be a (finite) set of grid points that is sufficiently dense in the ℓ_∞ -norm. Specifically, we choose the points in Ω on a rectangular grid such that

$$\max_{\mathbf{r} \in [0, 1]^3} \min_{\mathbf{r}_g \in \Omega} |\mathbf{r} - \mathbf{r}_g| \leq \frac{\epsilon}{3\tilde{c}L^4}, \quad (72)$$

where $\epsilon := 0.0005$. The cardinality of the set Ω is

$$|\Omega| = \left(\frac{3\tilde{c}L^4}{\epsilon} \right)^3 = c'L^{12}/\epsilon^3 = \tilde{c}L^{12}. \quad (73)$$

For every $\mathbf{r} \in \Omega$, we show that $Q(\mathbf{r})$ is “close” to $\bar{Q}(\mathbf{r})$ with high probability.

Step 2b: We use that $Q(\mathbf{r})$ is close to $\bar{Q}(\mathbf{r})$ for all $\mathbf{r} \in \Omega$ combined with Bernstein’s polynomial inequality to conclude that $Q(\mathbf{r})$ is close to $\bar{Q}(\mathbf{r})$ uniformly for all $\mathbf{r} \in [0, 1]^3$.

Step 2c: Finally, we combine this result with the properties of $Q(\mathbf{r})$ established in Section A.1 (in particular $|\bar{Q}(\mathbf{r})| < 1$ for all $\mathbf{r} \in [0, 1]^3 \setminus \mathcal{S}$) to conclude that $|Q(\mathbf{r})| < 1$ holds with high probability uniformly for all $\mathbf{r} \in [0, 1]^3 \setminus \mathcal{S}$.

For this argument to work, we will use that, by assumption (34a), $G_{\mathbf{m}}(\mathbf{r}, \mathbf{r}_k) = \langle \mathbf{Ag}_m(\mathbf{r}_k), \mathbf{Af}(\mathbf{r}) \rangle$ concentrates around the deterministic function $\bar{G}^{(n)}(\mathbf{r} - \mathbf{r}_k) = \langle \mathbf{g}_m(\mathbf{r}_k), \mathbf{f}(\mathbf{r}) \rangle$.

A.1 Construction of a certificate with unconstraint, deterministic coefficients

The first building block of our construction of Q is a *deterministic* 3D trigonometric polynomial $\bar{Q}(\mathbf{r}) = \langle \bar{\mathbf{q}}, \mathbf{f}(\mathbf{r}) \rangle$ that satisfies the interpolation and boundedness conditions (35), but whose coefficients $\bar{\mathbf{q}}$ are *not* constraint to be of the form $\mathbf{A}^H \mathbf{q}$. As mentioned before, corresponding constructions for 1D and 2D trigonometric polynomials have been derived by Candès and Fernandez-Granda [CFG14, Prop. 2.1, Prop. C.1]. As remarked in [CFG14], those results generalize to higher dimensions, albeit with a change in the numerical constant in the minimum separation condition (i.e., the constant 5 in (15)). The corresponding generalization is stated below.

Proposition 2. *Let $\mathcal{S} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_S\} \subset [0, 1]^3$ be an arbitrary set of points obeying the minimum separation condition (15). Then, for all $u_k \in \mathbb{C}, k = 1, \dots, S$, there exists a trigonometric polynomial $\bar{Q}(\mathbf{r}) = \langle \bar{\mathbf{q}}, \mathbf{f}(\mathbf{r}) \rangle$ obeying the interpolation and boundedness conditions (35).*

We provide a few details of the proof of Proposition 2 in order to collect properties of the polynomial \bar{Q} that are needed later for constructing Q ; however we omit proofs of the technical details as they are similar to the proof of [CFG14, Prop. C.1].

As mentioned in Section 6.1, the polynomial \bar{Q} is constructed by interpolating the points u_k as indicated in (28). To ensure that $\bar{Q}(\mathbf{r})$ reaches local maxima, which is necessary for the interpolation and boundedness condition in (35) to hold simultaneously, we choose the coefficients $\bar{\alpha}_k, \bar{\alpha}_{1k}, \bar{\alpha}_{2k}, \bar{\alpha}_{3k}$ in (28) such that

$$\bar{Q}(\mathbf{r}_k) = u_k \quad \text{and} \quad \nabla \bar{Q}(\mathbf{r}_k) = \mathbf{0} \quad \text{for all } \mathbf{r}_k \in \mathcal{S}. \quad (74)$$

We first establish that there exist coefficients such that (74) holds.

Lemma 5. *Under the hypothesis of Proposition 2, there exist coefficient vectors $\bar{\alpha} = [\bar{\alpha}_1, \dots, \bar{\alpha}_S]^T$ and $\bar{\alpha}_\ell = [\bar{\alpha}_{\ell 1}, \dots, \bar{\alpha}_{\ell S}]^T$ such that the polynomial $\bar{Q}(\mathbf{r})$ in (28) obeys (74). Moreover, those coefficients satisfy*

$$\begin{aligned} \|\bar{\alpha}\|_\infty &\leq 1.0021, \quad \|\bar{\alpha}_\ell\|_\infty \leq 0.0132, \quad \ell = 1, 2, 3, \\ \text{and, if } u_k &= 1, \operatorname{Re}(\alpha_k) \geq 0.9868, |\operatorname{Im}(\alpha_k)| \leq 0.0132. \end{aligned} \quad (75)$$

Proof outline of Lemma 5. Writing (74) in matrix form yields

$$\underbrace{\begin{bmatrix} \bar{\mathbf{D}}^{(0,0,0)} & \frac{1}{\kappa} \bar{\mathbf{D}}^{(1,0,0)} & \frac{1}{\kappa} \bar{\mathbf{D}}^{(0,1,0)} & \frac{1}{\kappa} \bar{\mathbf{D}}^{(0,0,1)} \\ -\frac{1}{\kappa} \bar{\mathbf{D}}^{(1,0,0)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(2,0,0)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(1,1,0)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(1,0,1)} \\ -\frac{1}{\kappa} \bar{\mathbf{D}}^{(0,1,0)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(1,1,0)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(0,2,0)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(0,1,1)} \\ -\frac{1}{\kappa} \bar{\mathbf{D}}^{(0,0,1)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(1,0,1)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(0,1,1)} & -\frac{1}{\kappa^2} \bar{\mathbf{D}}^{(0,0,2)} \end{bmatrix}}_{\bar{\mathbf{D}}} \begin{bmatrix} \bar{\alpha} \\ \kappa \bar{\alpha}_1 \\ \kappa \bar{\alpha}_2 \\ \kappa \bar{\alpha}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix},$$

where $[\bar{\mathbf{D}}^{(n_1, n_2, n_3)}]_{j,k} := \bar{G}^{(n_1, n_2, n_3)}(\mathbf{r}_j - \mathbf{r}_k)$ and $\kappa := \sqrt{|F^{(2)}(0)|}$. Here, we have scaled the entries of $\bar{\mathbf{D}}$ such that its diagonal entries are 1 ($F(0) = 1$, $\kappa^2 = |F^{(2)}(0)|$, and $F^{(2)}(0)$ is negative). The existence of coefficients such that \bar{Q} obeys (74) follows from $\bar{\mathbf{D}}$ being invertible, which is established next.

Proposition 3. $\bar{\mathbf{D}}$ is invertible and obeys

$$\|\mathbf{I} - \bar{\mathbf{D}}\| \leq 0.03254 \quad (76)$$

$$\|\bar{\mathbf{D}}^{-1}\| \leq 1.03363. \quad (77)$$

The proof of Proposition 3, not detailed here, shows that $\bar{\mathbf{D}}$ is close to the identity matrix \mathbf{I} by using that $G(\mathbf{r} - \mathbf{r}_k)$ and its derivatives decay fast around \mathbf{r}_k , and that the points $\mathbf{r}_k \in \mathcal{S}$ are sufficiently separated, by the minimum separation condition (15).

Since $\bar{\mathbf{D}}$ is invertible, the coefficients in Lemma 5 are given by:

$$\begin{bmatrix} \bar{\alpha} \\ \kappa \bar{\alpha}_1 \\ \kappa \bar{\alpha}_2 \\ \kappa \bar{\alpha}_3 \end{bmatrix} = \bar{\mathbf{D}}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \bar{\mathbf{L}} \mathbf{u}, \quad (78)$$

where $\bar{\mathbf{L}}$ is the $4S \times S$ submatrix of $\bar{\mathbf{D}}^{-1}$ corresponding to the first S columns of $\bar{\mathbf{D}}^{-1}$. It follows from $\bar{\mathbf{D}}$ being close to identity that the ℓ_∞ -norm of $\bar{\mathbf{D}}$ is small, which establishes (75). \square

The proof of Proposition 2 is concluded by showing that $\bar{Q}(\mathbf{r})$ satisfies the boundedness conditions in (35) as well. This is formalized by the following lemma.

Lemma 6. *Let $Q(\mathbf{r})$ be the polynomial in (28) with coefficients given by (78). Under the hypothesis of Proposition 2,*

- i. $|Q(\mathbf{r})| < 1$ for all $\mathbf{r} \in [0, 1]^3 \setminus \mathcal{S}$ that satisfy $\min_{\mathbf{r}_k \in \mathcal{S}} |\mathbf{r} - \mathbf{r}_k| \leq 0.18/N$,
- ii. for all \mathbf{r} that satisfy $\min_{\mathbf{r}_k \in \mathcal{S}} |\mathbf{r} - \mathbf{r}_k| \geq 0.18/N$ it follows that $|Q(\mathbf{r})| < 0.99$.

A.2 Step 1: Choice of the coefficients of \bar{Q}

We next show that, with high probability, we can select the coefficients of Q such that it satisfies (71). To this end, we write (71) in matrix form:

$$\underbrace{\begin{bmatrix} \mathbf{D}^{(0,0,0)} & \frac{1}{\kappa} \mathbf{D}^{(0,0,0)} & \frac{1}{\kappa} \mathbf{D}^{(0,0,0)} & \frac{1}{\kappa} \mathbf{D}^{(0,0,0)} \\ \frac{1}{\kappa} \mathbf{D}^{(1,0,0)} & -\frac{1}{\kappa^2} \mathbf{D}^{(1,0,0)} & -\frac{1}{\kappa^2} \mathbf{D}^{(1,0,0)} & -\frac{1}{\kappa^2} \mathbf{D}^{(1,0,0)} \\ -\frac{1}{\kappa} \mathbf{D}^{(0,1,0)} & -\frac{1}{\kappa^2} \mathbf{D}^{(0,1,0)} & -\frac{1}{\kappa^2} \mathbf{D}^{(0,1,0)} & -\frac{1}{\kappa^2} \mathbf{D}^{(0,1,0)} \\ -\frac{1}{\kappa} \mathbf{D}^{(0,0,1)} & -\frac{1}{\kappa^2} \mathbf{D}^{(0,0,1)} & -\frac{1}{\kappa^2} \mathbf{D}^{(0,0,1)} & -\frac{1}{\kappa^2} \mathbf{D}^{(0,0,1)} \end{bmatrix}}_{\mathbf{D} :=} \begin{bmatrix} \boldsymbol{\alpha} \\ \kappa \boldsymbol{\alpha}_1 \\ \kappa \boldsymbol{\alpha}_2 \\ \kappa \boldsymbol{\alpha}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad (79)$$

where $[\mathbf{D}_{(m_1, m_2, m_3)}^{(n_1, n_2, n_3)}]_{j,k} := G_{(m_1, m_2, m_3)}^{(n_1, n_2, n_3)}(\mathbf{r}_j, \mathbf{r}_k)$, $[\boldsymbol{\alpha}]_k := \alpha_k$, $[\boldsymbol{\alpha}_i]_k := \beta_{ik}$, $i = 1, 2, 3$. To show that the system of equations (79) has a solution, and in turn (71) holds, we will show that, with high probability, \mathbf{D} is invertible. To this end, we show that

$$\mathcal{E}_\xi := \{\|\mathbf{D} - \bar{\mathbf{D}}\| \leq \xi\}$$

occurs with high probability, and that \mathbf{D} is invertible on \mathcal{E}_ξ for all $\xi \in (0, 1/4]$. The fact that \mathbf{D} is invertible on \mathcal{E}_ξ for all $\xi \in (0, 1/4]$ follows with (76) by noting that:

$$\|\mathbf{I} - \mathbf{D}\| \leq \|\mathbf{D} - \bar{\mathbf{D}}\| + \|\bar{\mathbf{D}} - \mathbf{I}\| \leq \xi + 0.03254 \leq 0.28254.$$

Since \mathbf{D} is invertible, the coefficients of Q can be selected as

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \kappa \boldsymbol{\alpha}_1 \\ \kappa \boldsymbol{\alpha}_2 \\ \kappa \boldsymbol{\alpha}_3 \end{bmatrix} = \mathbf{D}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \mathbf{L} \mathbf{u}, \quad (80)$$

where \mathbf{L} is the $4S \times S$ submatrix of \mathbf{D}^{-1} consisting of the first S columns of \mathbf{D}^{-1} . In the remainder of this proof, we assume that the coefficients are given by (80). We will need the following bounds on the norm of \mathbf{L} and its deviation from $\bar{\mathbf{L}}$ later.

Lemma 7. *On the event \mathcal{E}_ξ with $\xi \in (0, 1/4]$ the following identities hold*

$$\|\mathbf{L}\| \leq 2.07, \quad (81)$$

$$\|\mathbf{L} - \bar{\mathbf{L}}\| \leq 2.14\xi. \quad (82)$$

Lemma 7 follows with (76) from matrix inequalities, see [HMS16, Lem. 4] for details. In the remainder of this proof, we set

$$\xi = \epsilon c_6 \log^{-1/2} \left(\frac{4|\Omega|}{\delta'} \right), \quad \delta' := \delta/176, \quad (83)$$

with Ω as defined in Appendix A, Step 2.

It remains to establish that the event \mathcal{E}_ξ occurs with high probability.

Lemma 8. *Under the assumptions of Lemma 1,*

$$\mathbb{P}[\mathcal{E}_\xi] \geq 1 - \delta'.$$

Proof. We will upper-bound $\|\mathbf{D} - \bar{\mathbf{D}}\|$ by upper-bounding the largest entry of $\mathbf{D} - \bar{\mathbf{D}}$. To this end, first note that the entries of $\mathbf{D} - \bar{\mathbf{D}}$ are given by

$$\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} [\mathbf{D}_{\mathbf{m}}^{(\mathbf{n})} - \bar{\mathbf{D}}^{(\mathbf{m} + \mathbf{n})}]_{j,k} = \frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} (G_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}_j, \mathbf{r}_k) - \bar{G}^{(\mathbf{m} + \mathbf{n})}(\mathbf{r}_j - \mathbf{r}_k)),$$

for $\|\mathbf{m}\|_\infty, \|\mathbf{n}\|_\infty \leq 1$ and $\|\mathbf{m} + \mathbf{n}\|_1 \leq 2$ and for $j, k = 1, \dots, S$. We now have

$$\mathbb{P}[\|\mathbf{D} - \bar{\mathbf{D}}\| \geq \xi] \leq \mathbb{P} \left[\sqrt{4S} \max_{j,k,\mathbf{m},\mathbf{n}} \frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} |[\mathbf{D}_{\mathbf{m}}^{(\mathbf{n})} - \bar{\mathbf{D}}^{(\mathbf{m} + \mathbf{n})}]_{j,k}| \geq \xi \right] \quad (84)$$

$$\leq \sum_{j,k,\mathbf{m},\mathbf{n}} \mathbb{P} \left[\frac{1}{\kappa \|\mathbf{m} + \mathbf{n}\|_1} |[\mathbf{D}_{\mathbf{m}}^{(\mathbf{n})} - \bar{\mathbf{D}}^{(\mathbf{m} + \mathbf{n})}]_{j,k}| \geq \frac{c_6 \epsilon}{\sqrt{4S \log(4|\Omega|/\delta)}} \right] \quad (85)$$

$$\leq 16(4S)^2 c_3 \frac{\delta}{L^{15}} \leq \delta'. \quad (86)$$

Here, (84) follows from the fact that \mathbf{D} and $\bar{\mathbf{D}}$ are $4S \times 4S$ matrices, (85) follows from the union bound, and (86) follows from assumption (34a). Finally, for the last inequality in (86), we used $S \leq L^3$ and that c_3 is sufficiently small. \square

A.3 Step 2a: $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ are close on a grid

We next show that $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ and their partial derivative are close on a set of grid points. This results is used in Step 2b in an ϵ -net argument to prove that $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ are close for all \mathbf{r} .

Lemma 9. *Suppose that $\|\mathbf{n}\|_1 \leq 2$. Under the assumptions of Lemma 1,*

$$\mathbb{P} \left[\max_{\mathbf{r} \in \Omega} \frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}) \right| \leq \epsilon \right] \geq 1 - \frac{\delta}{44}.$$

In order to prove Lemma 9, we first note that the \mathbf{n} -th partial derivative of $Q(\mathbf{r})$ (defined by (33)) after normalization with $1/\kappa^{\|\mathbf{n}\|_1}$ is given by

$$\frac{1}{\kappa^{\|\mathbf{n}\|_1}} Q^{(\mathbf{n})}(\mathbf{r}) = (\mathbf{v}^{(\mathbf{n})}(\mathbf{r}))^H \mathbf{L} \mathbf{u}. \quad (87)$$

Here, we used that the coefficients are given by (80) and we defined

$$(\mathbf{v}^{(\mathbf{n})}(\mathbf{r}))^H := \frac{1}{\kappa^{\|\mathbf{n}\|_1}} \left[G_{(0,0,0)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_1), \dots, G_{(0,0,0)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_S), \frac{1}{\kappa} G_{(1,0,0)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_1), \dots, \frac{1}{\kappa} G_{(1,0,0)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_S), \right. \\ \left. \frac{1}{\kappa} G_{(0,1,0)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_1), \dots, \frac{1}{\kappa} G_{(0,1,0)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_S), \frac{1}{\kappa} G_{(0,0,1)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_1), \dots, \frac{1}{\kappa} G_{(0,0,1)}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_S) \right].$$

Since $\mathbb{E} \left[G_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j) \right] = \bar{G}^{(\mathbf{n}+\mathbf{m})}(\mathbf{r} - \mathbf{r}_j)$ (by (51)), we have

$$\mathbb{E} \left[\mathbf{v}^{(\mathbf{n})}(\mathbf{r}) \right] = \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r}),$$

where

$$(\bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r}))^H := \frac{1}{\kappa^{\|\mathbf{n}\|_1}} \left[\bar{G}^{(\mathbf{n})}(\mathbf{r} - \mathbf{r}_1), \dots, \bar{G}^{(\mathbf{n})}(\mathbf{r} - \mathbf{r}_S), \frac{1}{\kappa} \bar{G}^{(n_1+1, n_2, n_3)}(\mathbf{r} - \mathbf{r}_1), \dots, \frac{1}{\kappa} \bar{G}^{(n_1+1, n_2, n_3)}(\mathbf{r} - \mathbf{r}_S), \right. \\ \left. \frac{1}{\kappa} \bar{G}^{(n_1, n_2+1, n_3)}(\mathbf{r} - \mathbf{r}_1), \dots, \frac{1}{\kappa} \bar{G}^{(n_1, n_2+1, n_3)}(\mathbf{r} - \mathbf{r}_S), \frac{1}{\kappa} \bar{G}^{(n_1, n_2, n_3+1)}(\mathbf{r} - \mathbf{r}_1), \dots, \frac{1}{\kappa} \bar{G}^{(n_1, n_2, n_3+1)}(\mathbf{r} - \mathbf{r}_S) \right].$$

Next, we decompose the derivatives of $Q(\mathbf{r})$ according to

$$\begin{aligned} \frac{1}{\kappa^{\|\mathbf{n}\|_1}} Q^{(\mathbf{n})}(\mathbf{r}) &= \left\langle \mathbf{u}, \mathbf{L}^H \mathbf{v}^{(\mathbf{n})}(\mathbf{r}) \right\rangle \\ &= \left\langle \mathbf{u}, \bar{\mathbf{L}}^H \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r}) \right\rangle + \underbrace{\left\langle \mathbf{u}, \mathbf{L}^H (\mathbf{v}^{(\mathbf{n})}(\mathbf{r}) - \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})) \right\rangle}_{I_1^{(\mathbf{n})}(\mathbf{r})} + \underbrace{\left\langle \mathbf{u}, (\mathbf{L} - \bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r}) \right\rangle}_{I_2^{(\mathbf{n})}(\mathbf{r})} \\ &= \frac{1}{\kappa^{\|\mathbf{n}\|_1}} \bar{Q}^{(\mathbf{n})}(\mathbf{r}) + I_1^{(\mathbf{n})}(\mathbf{r}) + I_2^{(\mathbf{n})}(\mathbf{r}), \end{aligned}$$

where $\bar{\mathbf{L}}$ was defined in (78). It follows that

$$\begin{aligned} \mathbb{P} \left[\max_{\mathbf{r} \in \Omega} \frac{1}{\kappa^{\|\mathbf{n}\|_1}} |Q^{(\mathbf{n})}(\mathbf{r}) - \bar{Q}^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon \right] &= \mathbb{P} \left[\max_{\mathbf{r} \in \Omega} |I_1^{(\mathbf{n})}(\mathbf{r}) + I_2^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon \right] \\ &\leq \mathbb{P} \left[\max_{\mathbf{r} \in \Omega} |I_1^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2 \right] + \mathbb{P} [\mathcal{E}_\xi^c] + \mathbb{P} \left[\max_{\mathbf{r} \in \Omega} |I_2^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2 | \mathcal{E}_\xi \right] \quad (88) \end{aligned}$$

$$\leq 4\delta' = \frac{\delta}{44}, \quad (89)$$

which concludes the proof of Lemma 9. Here, (88) follows from the union bound and the fact that $\mathbb{P}[A] = \mathbb{P}[A \cap B^c] + \mathbb{P}[A \cap B] \leq \mathbb{P}[B^c] + \mathbb{P}[A|B]$ with $B = \mathcal{E}_\xi$ and $A = \left\{ \max_{\mathbf{r} \in \Omega} |I_2^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2 \right\}$. Equation (89) follows from

$$\mathbb{P} \left[\max_{\mathbf{r} \in \Omega} |I_1^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2 \right] \leq \delta' + \mathbb{P} [\mathcal{E}_{1/4}^c], \quad (90)$$

$$\mathbb{P} \left[\max_{\mathbf{r} \in \Omega} |I_2^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2 | \mathcal{E}_\xi \right] \leq \delta', \quad (91)$$

and Lemma 8 (note that $\xi \leq 1/4$ since $\epsilon/2 \leq 1$ and $c_6 \leq 1/4$, thus, Lemma 8 yields $\mathbb{P}[\mathcal{E}_\xi^c] \leq \delta'$ and $\mathbb{P}[\mathcal{E}_{1/4}^c] \leq \delta'$). Inequalities (90) and (91), proven below ensure that the perturbations $I_1^{(\mathbf{n})}(\mathbf{r})$ and $I_2^{(\mathbf{n})}(\mathbf{r})$ are small on a set of (grid) points Ω , with high probability.

Proof of (90)

Set $\Delta \mathbf{v}^{(\mathbf{n})} := \mathbf{v}^{(\mathbf{n})}(\mathbf{r}) - \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})$ and $\omega := \frac{\epsilon}{20\sqrt{\log(8|\Omega|/\delta')}}$ for notational convenience. We have

$$\begin{aligned} \mathbb{P}\left[\max_{\mathbf{r} \in \Omega} |I_1^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2\right] &= \mathbb{P}\left[\max_{\mathbf{r} \in \Omega} |\langle \mathbf{u}, \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \rangle| \geq 5\omega \cdot 2\sqrt{\log(8|\Omega|/\delta')}\right] \\ &\leq \mathbb{P}\left[\bigcup_{\mathbf{r} \in \Omega} \left\{ |\langle \mathbf{u}, \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \rangle| \geq \left\| \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 2\sqrt{\log(8|\Omega|/\delta')} \right\} \cup \left\{ \left\| \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 \geq 5\omega \right\} \right] \\ &\leq \mathbb{P}\left[\bigcup_{\mathbf{r} \in \Omega} \left\{ |\langle \mathbf{u}, \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \rangle| \geq \left\| \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 2\sqrt{\log(8|\Omega|/\delta')} \right\} \cup \left\{ \left\| \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 \geq 2\omega \right\} \cup \{\|\mathbf{L}\| \geq 2.5\} \right] \\ &\leq \mathbb{P}[\|\mathbf{L}\| \geq 2.5] + \sum_{\mathbf{r} \in \Omega} \left(\mathbb{P}\left[|\langle \mathbf{u}, \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \rangle| \geq \left\| \mathbf{L}^H \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 2\sqrt{\log(8|\Omega|/\delta')} \right] + \mathbb{P}\left[\left\| \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 \geq 2\omega \right] \right) \end{aligned} \quad (92)$$

$$\begin{aligned} &\leq \mathbb{P}\left[\mathcal{E}_{1/4}^c\right] + |\Omega|4e^{-\log(8|\Omega|/\delta')} + \sum_{\mathbf{r} \in \Omega} \mathbb{P}\left[\left\| \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2 \geq 2\omega\right] \\ &\leq \mathbb{P}\left[\mathcal{E}_{1/4}^c\right] + \frac{\delta'}{2} + \frac{\delta'}{2}, \end{aligned} \quad (93)$$

where (92) follows from application of Hoeffding's inequality (stated below, cf. Lemma 10) and from $\{\|\mathbf{L}\| \geq 2.5\} \subseteq \mathcal{E}_{1/4}^c$ according to (81). For (93), we used

$$\begin{aligned} \mathbb{P}\left[\left\| \Delta \mathbf{v}^{(\mathbf{n})} \right\|_2^2 \geq 4\omega^2\right] &\leq \sum_k \mathbb{P}\left[(\Delta \mathbf{v}^{(\mathbf{n})})_k^2 \geq \omega^2/S\right] = \sum_k \mathbb{P}\left[(\Delta \mathbf{v}^{(\mathbf{n})})_k \geq \frac{\epsilon}{20\sqrt{S \log(8|\Omega|/\delta')}}\right] \\ &\leq \sum_k \mathbb{P}\left[(\Delta \mathbf{v}^{(\mathbf{n})})_k \geq \frac{c_1}{\sqrt{S \log(L/\delta')}}\right] \end{aligned} \quad (94)$$

$$\leq 4S \frac{c_3 \delta'}{L^{15}} \leq \frac{\delta'}{2\tilde{c}L^{12}} = \frac{\delta'}{2|\Omega|}, \quad (95)$$

where (94) follows from $\frac{\epsilon}{20\sqrt{\log(8|\Omega|/\delta')}} = \frac{\epsilon}{20\sqrt{\log(8\tilde{c}L^{12}/\delta')}} \geq \frac{\epsilon}{20\sqrt{12\log((8\tilde{c})^{1/12}L/\delta')}} \geq \frac{c_1}{\sqrt{\log(L/\delta')}}$, provided that $1 \geq (8\tilde{c})^{1/12}$, and $\frac{\epsilon}{20\sqrt{12}} \geq c_1$, where c_1 is the constant in (34a). Moreover, (95) follows from the assumption (34a), with $c_3 \leq \frac{1}{8\tilde{c}}$.

Lemma 10 (Hoeffding's inequality). *Suppose the entries of $\mathbf{u} \in \mathbb{R}^S$ are sampled independently from symmetric distributions on the complex unit disc. Then, for all $t \geq 0$, and for all $\mathbf{v} \in \mathbb{C}^S$*

$$\mathbb{P}[|\langle \mathbf{u}, \mathbf{v} \rangle| \geq \|\mathbf{v}\|_2 t] \leq 4e^{-\frac{t^2}{4}}.$$

Proof of (91)

By the union bound

$$\begin{aligned} \mathbb{P}\left[\max_{\mathbf{r} \in \Omega} |I_2^{(\mathbf{n})}(\mathbf{r})| \geq \epsilon/2 \mid \mathcal{E}_\xi\right] &\leq \sum_{\mathbf{r} \in \Omega} \mathbb{P}\left[\left|\left\langle \mathbf{u}, (\mathbf{L} - \bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r}) \right\rangle\right| \geq \epsilon/2 \mid \mathcal{E}_\xi\right] \\ &\leq \sum_{\mathbf{r} \in \Omega} \mathbb{P}\left[\left|\left\langle \mathbf{u}, (\mathbf{L} - \bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r}) \right\rangle\right| \geq \left\|(\mathbf{L} - \bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})\right\|_2 \frac{\epsilon c_6}{\xi}\right] \end{aligned} \quad (96)$$

$$\leq |\Omega| 4e^{-\frac{(\epsilon c_6/\xi)^2}{4}} \quad (97)$$

$$\leq \delta', \quad (98)$$

where (96) follows from (99) below, (97) follows by Hoeffding's inequality (Lemma 10), and (98) holds by definition of ξ in (83).

To complete the proof, note that by (82) we have $\|\mathbf{L} - \bar{\mathbf{L}}\| \leq 2.14\xi$ on \mathcal{E}_ξ . Thus, conditioned on \mathcal{E}_ξ ,

$$\left\|(\mathbf{L} - \bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})\right\|_2 \leq \|\mathbf{L} - \bar{\mathbf{L}}\| \left\|\bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})\right\|_2 \leq 2.14\xi \left\|\bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})\right\|_1 \leq c_5 \xi \leq \frac{\xi}{2c_6}, \quad (99)$$

where we used $\|\cdot\|_2 \leq \|\cdot\|_1$, and the third inequality follows from the fact that, for all \mathbf{r} ,

$$\begin{aligned} \left\|\bar{\mathbf{v}}^{(\mathbf{n})}(\mathbf{r})\right\|_1 &= \frac{1}{\kappa \|\mathbf{n}\|_1} \sum_{k=1}^S \left(\left| \bar{G}^{(\mathbf{n})}(\mathbf{r} - \mathbf{r}_k) \right| + \left| \frac{1}{\kappa} \bar{G}^{(n_1+1, n_2, n_3)}(\mathbf{r} - \mathbf{r}_k) \right| + \left| \frac{1}{\kappa} \bar{G}^{(n_1, n_2+1, n_3)}(\mathbf{r} - \mathbf{r}_k) \right| \right. \\ &\quad \left. + \left| \frac{1}{\kappa} \bar{G}^{(n_1, n_2, n_3+1)}(\mathbf{r} - \mathbf{r}_k) \right| \right) \leq \frac{c_5}{2.14}. \end{aligned}$$

Here, c_5 is a numerical constant; the inequality can be shown analogously to corresponding bounds in [CFG14, Appendix C.2] pertaining to the 2D case. To prove those bounds we use the minimum separation condition and that the functions $\bar{G}^{(\mathbf{n})}(\mathbf{r} - \mathbf{r}_k)$ decay quickly around \mathbf{r}_k . Finally, the last inequality in (99) follows from choosing c_6 sufficiently small.

A.4 Step 2b: $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ are close for all \mathbf{r}

We next use an ϵ -net argument together with Lemma 9 to establish that $Q^{(\mathbf{n})}(\mathbf{r})$ is close to $\bar{Q}^{(\mathbf{n})}(\mathbf{r})$ with high probability uniformly for all $\mathbf{r} \in [0, 1]^3$.

Lemma 11. *With probability at least $1 - \frac{\delta}{2}$,*

$$\max_{\mathbf{r} \in [0, 1]^3, \mathbf{n}: \|\mathbf{n}\|_1 \leq 2} \frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}) \right| \leq \epsilon. \quad (100)$$

Proof. First, we use Lemma 9 to show that $|Q^{(\mathbf{n})}(\mathbf{r}_g) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}_g)|$ is small for all $\mathbf{r}_g \in \Omega$. Using the union bound over all 11 vectors \mathbf{n} with non-negative integer entries obeying $\|\mathbf{n}\|_1 \leq 2$, it follows from Lemma 9, that

$$\left\{ \max_{\mathbf{r}_g \in \Omega, \mathbf{n}: \|\mathbf{n}\|_1 \leq 2} \frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}_g) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}_g) \right| \leq \frac{\epsilon}{3} \right\} \quad (101)$$

holds with probability at least $1 - \frac{\delta}{4}$.

We will show in Section A.4.2 below that (101) together with

$$\left\{ \max_{\mathbf{r} \in [0,1]^3, \mathbf{n}: \|\mathbf{n}\|_1 \leq 2} \frac{1}{\kappa \|\mathbf{n}\|_1} |Q^{(\mathbf{n})}(\mathbf{r})| \leq \frac{\tilde{c}}{3} L^3 \right\} \quad (102)$$

implies (100). In Section A.4.1 below we show that (102) holds with probability at least $1 - \frac{\delta}{4}$. By the union bound, the events in (101) and (102) hold simultaneously with probability at least $1 - \frac{\delta}{2}$, which concludes the proof.

A.4.1 Proof of the fact that (102) holds with probability at least $1 - \frac{\delta}{4}$:

We start by upper-bounding $|Q^{(\mathbf{n})}(\mathbf{r})|$. By (87),

$$\begin{aligned} \frac{1}{\kappa \|\mathbf{n}\|_1} |Q^{(\mathbf{n})}(\mathbf{r})| &= \left| \langle \mathbf{L}\mathbf{u}, \mathbf{v}^{(\mathbf{n})}(\mathbf{r}) \rangle \right| \\ &\leq \|\mathbf{L}\| \|\mathbf{u}\|_2 \left\| \mathbf{v}^{(\mathbf{n})}(\mathbf{r}) \right\|_2 \\ &\leq \|\mathbf{L}\| \sqrt{S} \sqrt{4S} \left\| \mathbf{v}^{(\mathbf{n})}(\mathbf{r}) \right\|_\infty \\ &= \|\mathbf{L}\| \sqrt{4S} \max_{j, \mathbf{m} \in \{(0,0,0), (1,0,0), (0,1,0), (0,0,1)\}} \frac{1}{\kappa \|\mathbf{n} + \mathbf{m}\|_1} |G_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j)| \\ &\leq \|\mathbf{L}\| \sqrt{4S} \frac{\tilde{c}}{S} L^3 \end{aligned} \quad (103)$$

$$\leq 2.5 \sqrt{4} \tilde{c} L^3 \quad (104)$$

where we used $\|\mathbf{u}\|_2 = \sqrt{S}$, since the entries of \mathbf{u} are on the unit disc; (103) holds with probability at least $1 - \delta/8$ according to (105) below, and (104) holds with probability at least $1 - \delta/8$ due to $\mathbb{P}[\|\mathbf{L}\| \geq 2.5] \leq \mathbb{P}[\mathcal{E}_{1/4}^c] \leq \frac{\delta}{8}$ (by (81) and application of Lemma 8). This concludes the proof of (102) holding with probability at least $1 - \frac{\delta}{4}$.

It is left to prove that

$$\left\{ \max_{\mathbf{r}, \mathbf{r}_j \in [0,1]^3, \mathbf{n}, \mathbf{m}: \|\mathbf{n} + \mathbf{m}\|_1 \leq 3} \frac{1}{\kappa \|\mathbf{n} + \mathbf{m}\|_1} |G_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j)| \leq \frac{\tilde{c}}{S} L^3 \right\} \quad (105)$$

holds with probability at least $1 - \delta/8$. To this end, note that $G_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j)$ is a trigonometric polynomial in \mathbf{r}, \mathbf{r}_j . Thus, by Bernstein's inequality, we have

$$\max_{\mathbf{r}, \mathbf{r}_j} \frac{1}{\kappa \|\mathbf{n} + \mathbf{m}\|_1} |G_{\mathbf{m}}^{(\mathbf{n})}(\mathbf{r}, \mathbf{r}_j)| \leq \frac{(2\pi N)^{\|\mathbf{n} + \mathbf{m}\|_1}}{\kappa \|\mathbf{n} + \mathbf{m}\|_1} \max_{\mathbf{r}, \mathbf{r}_j} |G_{\mathbf{0}}^{(\mathbf{0})}(\mathbf{r}, \mathbf{r}_j)| \leq c' \max_{\mathbf{r}, \mathbf{r}_j} |G_{\mathbf{0}}^{(\mathbf{0})}(\mathbf{r}, \mathbf{r}_j)| \leq \frac{\tilde{c}}{S} L^3$$

where the last inequality holds with probability at least $1 - \delta/8$, by assumption (34b).

A.4.2 Proof of the fact that (101) and (102) imply (100):

Consider a point $\mathbf{r} \in [0, 1]^3$ and let \mathbf{r}_g be the point in Ω closest to \mathbf{r} in ℓ_∞ -distance. By the triangle inequality,

$$\frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}) \right| \leq \frac{1}{\kappa \|\mathbf{n}\|_1} \left[\left| Q^{(\mathbf{n})}(\mathbf{r}) - Q^{(\mathbf{n})}(\mathbf{r}_g) \right| + \left| Q^{(\mathbf{n})}(\mathbf{r}_g) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}_g) \right| + \left| \bar{Q}^{(\mathbf{n})}(\mathbf{r}_g) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}) \right| \right]. \quad (106)$$

We next upper-bound the terms in (106) separately. With a slight abuse of notation, we write $Q^{(\mathbf{n})}(\beta, \tau, \nu) = Q^{(\mathbf{n})}([\beta, \tau, \nu]^T) = Q^{(\mathbf{n})}(\mathbf{r})$. The first absolute value in (106) can be upper-bounded according to

$$\begin{aligned} \left| Q^{(\mathbf{n})}(\mathbf{r}) - Q^{(\mathbf{n})}(\mathbf{r}_g) \right| &\leq \left| Q^{(\mathbf{n})}(\beta, \tau, \nu) - Q^{(\mathbf{n})}(\beta, \tau, \nu_g) \right| + \left| Q^{(\mathbf{n})}(\beta, \tau, \nu_g) - Q^{(\mathbf{n})}(\beta, \tau_g, \nu_g) \right| \\ &\quad + \left| Q^{(\mathbf{n})}(\beta, \tau_g, \nu_g) - Q^{(\mathbf{n})}(\beta_g, \tau_g, \nu_g) \right| \\ &\leq |\beta - \beta_g| \sup_z \left| Q^{(n_1+1, n_2, n_3)}(z, \tau, \nu) \right| + |\tau - \tau_g| \sup_z \left| Q^{(n_1, n_2+1, n_3)}(\beta, z, \nu) \right| \\ &\quad + |\nu - \nu_g| \sup_z \left| Q^{(n_1, n_2, n_3+1)}(\beta, z, \nu) \right| \\ &\leq |\beta - \beta_g| 2\pi N \sup_z \left| Q^{(\mathbf{n})}(z, \tau, \nu) \right| + |\tau - \tau_g| 2\pi N \sup_z \left| Q^{(\mathbf{n})}(\beta, z, \nu) \right| \\ &\quad + |\nu - \nu_g| 2\pi N \sup_z \left| Q^{(\mathbf{n})}(\beta, \tau, z) \right|, \end{aligned} \quad (107)$$

where (107) follows from Bernstein's polynomial inequality, stated below (recall that $Q^{(\mathbf{n})}(\beta, \tau, \nu)$ is a trigonometric polynomial of degree N in β, τ and ν).

Proposition 4 (Bernstein's polynomial inequality [Har96, Cor. 8]). *Let $p(\theta)$ be a trigonometric polynomial of degree N with complex coefficients p_k , i.e., $p(\theta) = \sum_{k=-N}^N p_k e^{i2\pi\theta k}$ where p_k are complex coefficients. Then*

$$\sup_{\theta} \left| \frac{d}{d\theta} p(\theta) \right| \leq 2\pi N \sup_{\theta} |p(\theta)|.$$

On the event (102) we have, by (107), that

$$\frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}) - Q^{(\mathbf{n})}(\mathbf{r}_g) \right| \leq \frac{\tilde{c}}{3} L^4 (|\beta + \beta_g| + |\tau - \tau_g| + |\nu - \nu_g|) \leq \tilde{c} L^4 |\mathbf{r} - \mathbf{r}_g| \leq \frac{\epsilon}{3}, \quad (108)$$

where the last inequality follows from (72).

We next upper-bound the third absolute value in (106). Using steps analogous to those leading to (108), we obtain

$$\frac{1}{\kappa^{m+n}} \left| \bar{Q}^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r}) \right| \leq \frac{\epsilon}{3}. \quad (109)$$

Substituting (101), (108), and (109) into (106) yields that

$$\frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}) \right| \leq \epsilon, \text{ for all } \mathbf{n}: \|\mathbf{n}\|_1 \leq 2 \text{ and for all } \mathbf{r} \in [0, 1]^2.$$

This concludes the proof. \square

A.5 Step 2c: $|Q(\mathbf{r})| < 1$ for all $\mathbf{r} \in [0, 1]^3 \setminus \mathcal{S}$

Lemma 12. *With probability at least $1 - \frac{\delta}{2}$ the following statements hold simultaneously:*

- i. *For all \mathbf{r} , that satisfy $\min_{\mathbf{r}_j \in \mathcal{S}} |\mathbf{r} - \mathbf{r}_j| \geq 0.18/N$ we have that $|Q(\mathbf{r})| < 0.9995$.*
- ii. *For all $\mathbf{r} \notin \mathcal{S}$ that satisfy $0 < |\mathbf{r} - \mathbf{r}_j| \leq 0.18/N$ for some $\mathbf{r}_j \in \mathcal{S}$, we have that $|Q(\mathbf{r})| < 1$.*

Proof. By Lemma 11, (recall that $\epsilon = 0.0005$), with probability $\geq 1 - \frac{\delta}{2}$,

$$\frac{1}{\kappa \|\mathbf{n}\|_1} \left| Q^{(\mathbf{n})}(\mathbf{r}) - \bar{Q}^{(\mathbf{n})}(\mathbf{r}) \right| \leq 0.0005, \quad (110)$$

for all nonnegative integer vectors \mathbf{n} : $\|\mathbf{n}\|_1 \leq 2$, and for all $\mathbf{r} \in [0, 1]^3$. Statement i follows directly by combining (110) with Lemma 6 via the triangle inequality.

We next show that Statement ii follows from (110) and certain properties of \tilde{Q} . We assume without loss of generality that $\mathbf{0} \in \mathcal{S}$, and that $Q(\mathbf{0}) = 1$, and consider a vector \mathbf{r} with $|\mathbf{r}| \leq 0.18/N$. Statement ii now follows by showing that the Hessian matrix $\tilde{\mathbf{H}}$ of $\tilde{Q}(\mathbf{r}) := |Q(\mathbf{r})|$ is negative definite. By Sylvester's criterion [HJ12, Thm. 7.2.5], $\tilde{\mathbf{H}}$ is negative definite if the leading principal minors denoted by $\tilde{\mathbf{H}}_1$, $\tilde{\mathbf{H}}_2$, and $\tilde{\mathbf{H}}_3 = \tilde{\mathbf{H}}$, have alternating sign, which is implied by

$$\det(\tilde{\mathbf{H}}_1) = \tilde{Q}^{(2,0,0)}(\mathbf{r}) < 0 \quad (111)$$

$$\det(\tilde{\mathbf{H}}_2) = \tilde{Q}^{(2,0,0)}(\mathbf{r})\tilde{Q}^{(0,2,0)}(\mathbf{r}) - \tilde{Q}^{(1,1,0)}(\mathbf{r})\tilde{Q}^{(1,1,0)}(\mathbf{r}) > 0 \quad (112)$$

$$\begin{aligned} \det(\tilde{\mathbf{H}}) &= \tilde{Q}^{(2,0,0)}(\mathbf{r})\tilde{Q}^{(0,2,0)}(\mathbf{r})\tilde{Q}^{(0,0,2)}(\mathbf{r}) + 2\tilde{Q}^{(1,1,0)}(\mathbf{r})\tilde{Q}^{(0,1,1)}(\mathbf{r})\tilde{Q}^{(1,0,1)}(\mathbf{r}) \\ &\quad - \tilde{Q}^{(1,0,1)}(\mathbf{r})\tilde{Q}^{(0,2,0)}(\mathbf{r})\tilde{Q}^{(1,0,1)}(\mathbf{r}) - \tilde{Q}^{(1,1,0)}(\mathbf{r})\tilde{Q}^{(1,1,0)}(\mathbf{r})\tilde{Q}^{(0,0,2)}(\mathbf{r}) \\ &\quad - \tilde{Q}^{(2,0,0)}(\mathbf{r})\tilde{Q}^{(0,1,1)}(\mathbf{r})\tilde{Q}^{(0,1,1)}(\mathbf{r}) < 0. \end{aligned} \quad (113)$$

The inequalities (111), (112), and (113), follow directly from

$$\tilde{Q}^{(2,0,0)}, \tilde{Q}^{(0,2,0)}, \tilde{Q}^{(0,0,2)} \leq -0.842\kappa^2, \quad |\tilde{Q}^{(1,1,0)}|, |\tilde{Q}^{(1,0,1)}|, |\tilde{Q}^{(0,1,1)}| \leq 0.2113\kappa^2. \quad (114)$$

To prove (114), we define $Q_R^{(\mathbf{n})} := \frac{1}{\kappa \|\mathbf{n}\|_1} \text{Re}(Q^{(\mathbf{n})})$, $Q_I^{(\mathbf{n})} := \frac{1}{\kappa \|\mathbf{n}\|_1} \text{Im}(Q^{(\mathbf{n})})$ and note that

$$\frac{1}{\kappa} \tilde{Q}^{(1,0,0)} = \frac{Q_R^{(1,0,0)} Q_R + Q_I^{(1,0,0)} Q_I}{|Q|}.$$

After simplification, we obtain

$$\begin{aligned} \frac{1}{\kappa^2} \tilde{Q}^{(2,0,0)} &= \left(1 - \frac{Q_R^2}{|Q|^2}\right) \frac{Q_R^{(1,0,0)}{}^2}{|Q|} - \frac{2Q_R Q_R^{(1,0,0)} Q_I Q_I^{(1,0,0)} + Q_I^2 Q_I^{(1,0,0)}{}^2}{|Q|^3} \\ &\quad + \frac{Q_I^{(1,0,0)}{}^2 + Q_I Q_I^{(2,0,0)}}{|Q|} + \frac{Q_R}{|Q|} Q_R^{(2,0,0)}, \end{aligned} \quad (115)$$

and

$$\begin{aligned} \frac{1}{\kappa^2} \tilde{Q}^{(1,1,0)} &= Q_R^{(1,1,0)} \frac{Q_R}{|Q|} + \frac{Q_R^{(1,0,0)} Q_R^{(0,1,0)}}{|Q|} \left(1 - \frac{Q_R^2}{|Q|^2}\right) + \frac{Q_I^{(1,1,0)} Q_I + Q_I^{(1,0,0)} Q_I^{(0,1,0)}}{|Q|} \\ &\quad - \frac{Q_R^{(0,1,0)} Q_R Q_I^{(1,0,0)} Q_I + Q_I^{(0,1,0)} Q_I (Q_R^{(1,0,0)} Q_R + Q_I^{(1,0,0)} Q_I)}{|Q|^3}. \end{aligned} \quad (116)$$

The upper bounds in (114) now follow by applying the following inequalities with $\epsilon = 0.0005$ to (115) and (116):

$$\begin{aligned} Q_R(\mathbf{r}) &\leq |\bar{Q}(\mathbf{r})| + \epsilon \leq 1 + \epsilon, \quad Q_R(\mathbf{r}) \geq |\bar{Q}(\mathbf{r})| - \epsilon \geq 0.9298 - \epsilon, \quad |Q_I(\mathbf{r})| \leq \bar{Q}_I(\mathbf{r}) + \epsilon \leq 0.0702 \\ |Q_R^{(1,0,0)}(\mathbf{r})| &\leq |\bar{Q}_R^{(1,0,0)}(\mathbf{r})| + \epsilon \leq 0.3756 + \epsilon, \quad |Q_I^{(1,0,0)}(\mathbf{r})| \leq |\bar{Q}_I^{(1,0,0)}(\mathbf{r})| + \epsilon \leq 0.0541 + \epsilon. \\ Q_R^{(2,0,0)}(\mathbf{r}) &\leq \bar{Q}_R^{(2,0,0)}(\mathbf{r}) + \epsilon \leq -0.8572 + \epsilon, \quad |Q_I^{(2,0,0)}(\mathbf{r})| \leq |\bar{Q}_I^{(2,0,0)}(\mathbf{r})| + \epsilon \leq -0.1005 + \epsilon \\ |Q_R^{(1,1,0)}| &\leq |\bar{Q}_R^{(1,1,0)}(\mathbf{r})| + \epsilon \leq 0.1967 + \epsilon, \quad |Q_I^{(1,1,0)}| \leq |\bar{Q}_I^{(1,1,0)}(\mathbf{r})| + \epsilon \leq 0.0921 + \epsilon. \end{aligned}$$

Here, $\bar{Q}_R^{(\mathbf{n})} = \frac{1}{\kappa\|\mathbf{n}\|_1} \operatorname{Re}(\bar{Q}^{(\mathbf{n})})$ and $\bar{Q}_I^{(\mathbf{n})} = \frac{1}{\kappa\|\mathbf{n}\|_1} \operatorname{Im}(\bar{Q}^{(\mathbf{n})})$. The inequalities above follow from combining (110) with Proposition 5 below using the triangle inequality (recall that $\bar{Q}^{(\mathbf{n})}(\mathbf{r})$ is real).

Proposition 5. *Suppose that $\mathbf{0} \in \mathcal{S}$, and let $Q(\mathbf{0}) = 1$. Then, for $|\mathbf{r}| \leq 0.18/N$,*

$$\begin{aligned} |\bar{Q}(\mathbf{r})| &\leq 1, \quad \bar{Q}_R(\mathbf{r}) \geq 0.92982 \\ \bar{Q}_R^{(2,0,0)}(\mathbf{r}) &\leq -0.8572, \quad |\bar{Q}_I^{(2,0,0)}(\mathbf{r})| \leq 0.1005 \\ |\bar{Q}_R^{(1,1,0)}| &\leq 0.1967, \quad |\bar{Q}_I^{(1,1,0)}| \leq 0.0921 \\ |\bar{Q}_R^{(1,0,0)}(\mathbf{r})| &\leq 0.3756, \quad |\bar{Q}_I^{(1,0,0)}(\mathbf{r})| \leq 0.0541. \end{aligned}$$

The proof of Proposition 5 is analogous to corresponding results established for the 1D and 2D case in [CFG14, Proofs of Lemmas 2.3 and 2.4, Appendix C.2], and therefore omitted. \square

B Proof of discrete results (Theorems 4 and 5)

The following proposition—standard in the theory of compressed sensing (see e.g., [CRT06])—shows that the existence of a certain dual polynomial guarantees that $L1(\mathbf{y})$ in (23) succeeds in reconstructing \mathbf{s} .

Proposition 6. *Let $\mathbf{y} = \mathbf{A}\mathbf{F}_{\text{grid}}\mathbf{s}$ and let \mathcal{S} denote the support of \mathbf{s} . Assume that the columns of $\mathbf{A}\mathbf{F}_{\text{grid}}$ indexed by \mathcal{S} are linearly independent. If there exists a vector \mathbf{v} in the row space of $\mathbf{A}\mathbf{F}_{\text{grid}}$ with*

$$\mathbf{v}_{\mathcal{S}} = \operatorname{sign}(\mathbf{s}_{\mathcal{S}}) \quad \text{and} \quad \|\mathbf{v}_{\mathcal{S}^c}\|_{\infty} < 1 \tag{117}$$

then \mathbf{s} is the unique minimizer of $L1(\mathbf{y})$ in (23). Here, $\mathbf{v}_{\mathcal{S}}$ is the sub-vector of \mathbf{v} corresponding to the entries indexed by \mathcal{S} .

The proof now follows directly from Lemma 1, and from the conditions of Lemma 1 being satisfied by the assumptions of Theorems 1 and 3, as shown in Sections 6.2 and 6.3. To see this, set $\mathbf{u} = \operatorname{sign}(\mathbf{s}_{\mathcal{S}})$ in Lemma 1 and consider the polynomial $Q(\mathbf{r})$ from Lemma 1. Define \mathbf{v} as $[\mathbf{v}]_{(n_1, n_2, n_3)} = Q([n_1/K_1, n_2/K_2, n_3/K_3])$ and note that it satisfies (117) since $Q([n_1/K_1, n_2/K_2, n_3/K_3]) = \operatorname{sign}(\mathbf{s}_{(n_1, n_2, n_3)})$ for $[n_1, n_2, n_3] \in \mathcal{S}$ and $|Q([n_1/K_1, n_2/K_2, n_3/K_3])| < 1$ for $[n_1, n_2, n_3] \notin \mathcal{S}$.

C Proof of Lemma 4

Lemma 4 is essentially a consequence of results established in [TBR15], in particular the following proposition.

Proposition 7 ([TBR15, Thm. 1]). *Let $\mathbf{z} = \sum_{k=1}^S b_k \mathbf{f}(\nu_k)$, and suppose the ν_1, \dots, ν_S obey the minimum separation condition (3). Suppose that the dual norm of $\|\cdot\|_{\mathcal{A}}$, defined as $\|\cdot\|_{\mathcal{A}}^* := \sup_{\nu \in [0,1]} |\langle \cdot, \mathbf{f}(\nu) \rangle|$, obeys*

$$\|\mathbf{w}\|_{\mathcal{A}}^* \leq \frac{\lambda}{\eta}. \quad (118)$$

Then, for $\eta > 1$ sufficiently large, the proximity operator in (69) with regularization parameter $\lambda = \eta\sigma\sqrt{4L\log(L)}$ applied to $\mathbf{y} = \mathbf{z} + \mathbf{w}$ with $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ obeys

$$\|\tilde{\mathbf{z}}(\mathbf{z} + \mathbf{w}, \lambda) - \mathbf{z}\|_2^2 \leq c\sigma^2 S \log L.$$

Since (118) holds with high probability [BTR13, App. C] for our choice of λ , Theorem 7 provides an error bound that holds with high-probability. However, we need to characterize the expectation of the estimation error, thus slightly more work is required to establish Lemma 4.

For notational convenience, we set $\tilde{\mathbf{z}} = \tilde{\mathbf{z}}(\mathbf{z} + \mathbf{w}, \lambda)$ and let $\mathbf{e} := \mathbf{z} - \tilde{\mathbf{z}}$ be the error vector. Note that

$$\mathbb{E} [\|\mathbf{e}\|_2^2] = \mathbb{E} [\|\mathbf{e}\|_2^2 \mathbb{1}\{\|\mathbf{w}\|_{\mathcal{A}}^* < \lambda/\eta\}] + \mathbb{E} [\|\mathbf{e}\|_2^2 \mathbb{1}\{\|\mathbf{w}\|_{\mathcal{A}}^* \geq \lambda/\eta\}]. \quad (119)$$

The lemma follows by upper bounding the terms on the RHS. The first term is smaller than $c\sigma^2 S \log L$ by Proposition 7. We next upper bound the second term. By the Cauchy-Schwarz inequality, we have

$$\mathbb{E} [\|\mathbf{e}\|_2^2 \mathbb{1}\{\|\mathbf{w}\|_{\mathcal{A}}^* \geq \lambda/\eta\}] \leq \sqrt{\mathbb{E} [\|\mathbf{e}\|_2^4]} \sqrt{\mathbb{P} [\|\mathbf{w}\|_{\mathcal{A}}^* \geq \lambda/\eta]} \leq c\sigma^2 \log(L), \quad (120)$$

where the last inequality follows from Lemma 13, which provides a crude upper bound on $\mathbb{E} [\|\mathbf{e}\|_2^2]$, and Lemma 14, which states an upper bound on the probability for the atomic norm of \mathbf{w} being large.

Lemma 13.

$$\sqrt{\mathbb{E} [\|\mathbf{e}\|_2^4]} \leq c\sigma^2 L \log(L).$$

Lemma 14.

$$\mathbb{P} [\|\mathbf{w}\|_{\mathcal{A}}^* \geq \lambda/\eta] \leq \frac{1}{L^2}.$$

It remains to prove Lemmas 13 and 14.

Proof of Lemma 13. We show below that

$$\|\mathbf{e}\|_2^2 \leq 8\|\mathbf{w}\|_2^2 + c\lambda^2 \frac{S}{L} = 8\|\mathbf{w}\|_2^2 + c\eta^2 \sigma^2 \log(L)S. \quad (121)$$

It follows that

$$\mathbb{E} \left[\|\mathbf{e}\|_2^4 \right] \leq 64\mathbb{E} \left[\|\mathbf{w}\|_2^4 \right] + 16\mathbb{E} \left[\|\mathbf{w}\|_2^2 \right] \eta^2 \sigma^2 4L \log(L) + c^2 \eta^4 \sigma^4 L^2 \log^2(L) \leq c' \eta^4 \sigma^4 L^2 \log^2(L).$$

Here, we used that $\|\mathbf{w}\|_2^2$ is χ^2 distributed and the second moment of a χ^2 random variable with L degrees of freedom is $L(L+2)$. This proves the lemma.

It remains to prove (121). First note that $\mathbf{z} = \int_0^1 \mathbf{f}(\nu) \mu(d\nu)$, where μ is the measure defined as $\mu(\nu) = \sum_{k=1}^S b_k \delta(\nu - \nu_k)$, where δ is the Dirac measure. Next, let $\tilde{\mu}$ be the representing measure of the estimate $\tilde{\mathbf{z}}$ with minimum total variation norm (for a definition of the total variation norm, denoted by $\|\cdot\|_{TV}$ see [CFG14, App. A]), and note that $\|\tilde{\mu}\|_{TV} = \|\tilde{\mathbf{z}}\|_{\mathcal{A}}$. By optimality of $\tilde{\mathbf{z}}$, we have

$$\frac{1}{2} \|\mathbf{y} - \tilde{\mathbf{z}}\|_2^2 + \lambda \|\tilde{\mu}\|_{TV} \leq \frac{1}{2} \|\mathbf{y} - \mathbf{z}\|_2^2 + \lambda \|\mu\|_{TV}.$$

Denote by $P_{\mathcal{S}}(\tilde{\mu})$ the projection of the measure $\tilde{\mu}$ on the support set of μ , i.e., $\mathcal{S} = \{\nu_1, \dots, \nu_S\}$, likewise $P_{\mathcal{S}^c}(\tilde{\mu})$ is the projection onto the complement of \mathcal{S} . With this notation and $\mathbf{y} = \mathbf{z} + \mathbf{w}$, the former inequality is equivalent to

$$\frac{1}{2} \|\mathbf{e}\|_2^2 + \lambda (\|P_{\mathcal{S}}(\tilde{\mu})\|_{TV} + \|P_{\mathcal{S}^c}(\tilde{\mu})\|_{TV}) \leq \langle \mathbf{e}, \mathbf{w} \rangle + \lambda \|\mu\|_{TV}. \quad (122)$$

We next lower bound $\|P_{\mathcal{S}}(\tilde{\mu})\|_{TV}$. To this end, note that Proposition 2.1 in [CFG14] guarantees that there exists a polynomial $Q(\nu) = \langle \mathbf{q}, \mathbf{f}(\nu) \rangle$ that interpolates the sign pattern of μ and obeys $|Q(\nu)| \leq 1$. Let $\xi := \tilde{\mu} - \mu$ be the difference measure. Since Q interpolates the sign of μ ,

$$\|\mu\|_{TV} + \int_0^1 Q(\nu) P_{\mathcal{S}}(\xi)(d\nu) \leq \|P_{\mathcal{S}}(\tilde{\mu})\|_{TV}.$$

Using this inequality in (122) and rearranging terms yields

$$\begin{aligned} \frac{1}{2} \|\mathbf{e}\|_2^2 + \lambda \|P_{\mathcal{S}^c}(\tilde{\mu})\|_{TV} &\leq \langle \mathbf{e}, \mathbf{w} \rangle - \lambda \int_0^1 Q(\nu) P_{\mathcal{S}}(\xi)(d\nu) \\ &= \langle \mathbf{e}, \mathbf{w} \rangle - \lambda \int_0^1 Q(\nu) \xi(d\nu) + \lambda \int_0^1 Q(\nu) P_{\mathcal{S}^c}(\xi)(d\nu) \\ &= \langle \mathbf{e}, \mathbf{w} \rangle - \lambda \langle \mathbf{e}, \mathbf{q} \rangle + \lambda \int_0^1 Q(\nu) P_{\mathcal{S}^c}(\xi)(d\nu) \\ &\leq \langle \mathbf{e}, \mathbf{w} \rangle - \lambda \langle \mathbf{e}, \mathbf{q} \rangle + \lambda \|P_{\mathcal{S}^c}(\xi)\|_{TV}. \end{aligned}$$

where the penultimate inequality follows from $Q(\nu) = \langle \mathbf{q}, \mathbf{f}(\nu) \rangle$ and Parseval's identity, and the last inequality holds by $|Q(\nu)| \leq 1$. We have shown that

$$\frac{1}{2} \|\mathbf{e}\|_2^2 \leq \langle \mathbf{e}, \mathbf{w} \rangle - \lambda \langle \mathbf{e}, \mathbf{q} \rangle. \quad (123)$$

Next, using the inequality $ab \leq \frac{1}{2}(a^2 + b^2)$ with $a = \|\mathbf{e}\|_2/(2\sqrt{\lambda})$, $b = 2\sqrt{\lambda} \|\mathbf{q}\|_2$ we obtain

$$-\lambda \langle \mathbf{e}, \mathbf{q} \rangle \leq \lambda \|\mathbf{e}\|_2 \|\mathbf{q}\|_2 \leq \frac{1}{8} \|\mathbf{e}\|_2^2 + 2\lambda^2 \|\mathbf{q}\|_2^2.$$

Using this in (123) yields

$$\begin{aligned}\frac{3}{8}\|\mathbf{e}\|_2^2 &\leq \langle \mathbf{e}, \mathbf{w} \rangle + 2\lambda^2\|\mathbf{q}\|_2^2 \\ &\leq \|\mathbf{e}\|_2\|\mathbf{w}\|_2 + 2\lambda^2\|\mathbf{q}\|_2^2.\end{aligned}$$

Application of $ab \leq \frac{1}{2}(a^2 + b^2)$ once again yields $\|\mathbf{e}\|_2\|\mathbf{w}\|_2 \leq \frac{1}{8}\|\mathbf{e}\|_2^2 + 2\|\mathbf{w}\|_2^2$ which in turn yields

$$\frac{1}{4}\|\mathbf{e}\|_2^2 \leq 2\|\mathbf{w}\|_2^2 + 2\lambda^2\|\mathbf{q}\|_2^2.$$

By Parseval's theorem, and Hölder's inequality we have

$$\|\mathbf{q}\|_2^2 = \langle Q(\nu), Q(\nu) \rangle \leq \|Q\|_1\|Q\|_\infty \leq \frac{cS}{L},$$

where the last inequality follows from $\|Q\|_1 \leq \frac{cS}{L}$ [TBR15, Lem. 4]. This concludes the proof of inequality (121). \square

Proof of Lemma 14. As shown in [BTR13, C.4, with $N = n$], a consequence of Bernstein's polynomial inequality is that

$$\|\mathbf{w}\|_{\mathcal{A}}^* \leq (1 + 4\pi) \max_{m=0, \dots, L-1} \underbrace{\left| \sum_{\ell=0}^{L-1} w_\ell e^{i2\pi\ell(m/L)} \right|}_{Z_m}.$$

The random variable Z_m is stochastically upper bounded by a Gaussian random variable with variance $\sigma^2 L$. It follows that, for any $\beta > 1/\sqrt{2\pi}$,

$$\mathbb{P} [|Z_m| \geq \sigma\sqrt{L}\beta] \leq 2e^{-\beta^2}.$$

The proof is concluded by noting that, by the union bound

$$\mathbb{P} [\|\mathbf{w}\|_{\mathcal{A}}^* \geq \sigma\sqrt{L}\sqrt{4\log(L)}] \leq 2Le^{-4\log(L)} \leq \frac{1}{L^2}.$$

\square