Super-Resolution Radar

Reinhard Heckel[1,*], Veniamin I. Morgenshtern[2], and Mahdi Soltanolkotabi[3]

[1]Dept. of Information Technology and Electrical Engineering, ETH Zurich, Switzerland
[2]Dept. of Statistics, Stanford University, CA
[3]Dept. of Electrical Engineering and Computer Science, UC Berkeley, CA

November 2014

Abstract

In this paper we study the identification of a time-varying linear system whose response is a weighted superposition of time and frequency shifted versions of the input signal. This problem arises in a multitude of applications such as wireless communications and radar imaging. Due to practical constraints, the input signal has finite bandwidth B, and the received signal is observed over a finite time interval of length T only. This gives rise to a time and frequency resolution of $1/B$ and $1/T$. We show that this resolution limit can be overcome, i.e., we can recover the exact (continuous) time-frequency shifts and the corresponding attenuation factors, by essentially solving a simple convex optimization problem. This result holds provided that the distance between the time-frequency shifts is at least $2.37/B$ and $2.37/T$, in time and frequency. Furthermore, this result allows the total number of time-frequency shifts to be linear (up to a log-factor) in BT, the dimensionality of the response of the system. More generally, we show that we can estimate the time-frequency components of a signal that is S-sparse in the *continuous* dictionary of time-frequency shifts of a random (window) function, from a number of measurements, that is linear (up to a log-factor) in S.

1 Introduction

The identification of *time-varying* linear systems is a fundamental problem in many engineering applications. Concrete examples include radar and the identification of dispersive communication channels. Radar systems and wireless communication channels are typically modeled as linear systems whose response is a (continuous) weighted superposition of delayed and Doppler shifted versions of the input signal. In general, the response of such a system H to an input signal x is given by

$$(Hx)(t) = \iint s_H(\tau, \nu) x(t-\tau) e^{i2\pi\nu t} d\nu d\tau$$

where $s_H(\tau,\nu)$ denotes the spreading function which characterizes the system. Identification of H amounts to estimate the (unknown) spreading function from an input-output measurement. The input signal $x(t)$ is known and can be controlled by the system engineer. We assume that the spreading function consists of S point scatterers. In radar, those point scatterers correspond to

*Corresponding author

moving targets. Mathematically, this means that the spreading function is of the form

$$s_H(\tau,\nu) = \sum_{n=1}^{S} b_n \delta(\tau - \bar{\tau}_n)\delta(\nu - \bar{\nu}_n).$$

Here, $b_n \in \mathbb{C}$ is a (complex) attenuation factor associated with the time-frequency (delay-Doppler) shift $(\bar{\tau}_n, \bar{\nu}_n) \in \mathbb{R}^2$. With this spreading function, the general input-output relationship above reduces to

$$(Hx)(t) = \sum_{n=1}^{S} b_n x(t - \bar{\tau}_n) e^{i2\pi\bar{\nu}_n t}. \tag{1}$$

In order to identify H we need to estimate the attenuation factors b_n and the corresponding (continuous) time-frequency shifts $(\bar{\tau}_n, \bar{\nu}_n)$ from an input-output measurement.

In practice, such an input-output measurement can only be performed under the constraints that the input signal $x(t)$ has finite bandwidth B and the output signal $(Hx)(t)$ is observed over a finite time interval of length T only. This time and band-limitation determines the "natural" resolution of the system of $1/B$ and $1/T$ in τ- and ν-direction, respectively. Specifically, as we shall explain in more detail in Section 4, the response $(Hx)(t)$ is essentially $L := BT$ dimensional[1], and described by its samples $y_p := (Hx)(p/B)$, which take on the form

$$y_p = \sum_{n=1}^{S} b_n e^{i2\pi p\nu_n} \frac{1}{L} \sum_{\ell,k=-N}^{N} e^{-i2\pi k\tau_n} e^{i2\pi(p-\ell)\frac{k}{L}} a_\ell, \quad p = -N, ..., N. \tag{2}$$

Here, $N := \frac{BT-1}{2}$, $\tau_n := \bar{\tau}_n/T$, $\nu_n := \bar{\nu}_n/B$, and $a_\ell = x(\ell/B)$ are the samples of $x(t)$ at $t = \ell/B$, assumed to be L-periodic. Thus, identification of H under the proviso that $x(t)$ is band-limited and $(Hx)(t)$ is time-limited, amounts to estimating (b_n, τ_n, ν_n) from the samples in (2).

In this paper, we consider the problem of recovering the triplets (b_n, τ_n, ν_n), $(\tau_n, \nu_n) \in [0,1]^2$ from the samples $y_p, p = -N, ..., N$, in (2). We call this the super-resolution radar problem, as recovering the exact time-frequency shifts (τ_n, ν_n) "breaks" the natural resolution limit of $(1/T, 1/B)$.

Alternatively, one can view this estimation problem as the recovery of a signal that is S-sparse in the continuous dictionary of time-frequency shifts of a L-periodic sequence a_ℓ. In order to see this, and to better understand the super-resolution radar problem, we next consider two special cases.

1.1 Time-frequency shifts on a grid

If the time-frequency shifts lie on a $(\frac{1}{B}, \frac{1}{T})$ grid, the super-resolution radar problem reduces to a sparse signal recovery (compressive sensing) problem with a Gabor measurement matrix. To see this, suppose that the time-frequency shifts lie on a $(\frac{1}{B}, \frac{1}{T})$ grid, i.e., $\bar{\tau}_n = \frac{m_n}{B}$, $\bar{\nu}_n = \frac{q_n}{T}$, where $m_n, q_n \in \{0, ..., L-1\}$ are the positions of the time-frequency shifts on the grid. With $\tau_n = \frac{m_n}{BT} = \frac{m_n}{L}$ and $\nu_n = \frac{q_n}{BT} = \frac{q_n}{L}$, (2) reduces to

$$y_p = \sum_{n=1}^{S} b_n e^{i2\pi\frac{q_n p}{L}} a_{p-m_n}, \quad p = -N, ..., N. \tag{3}$$

[1] For simplicity we assume throughout that BT is an odd integer.

Writing (3) in matrix-vector form yields

$$\mathbf{y} = \mathbf{G}\mathbf{b}$$

where $\mathbf{G} \in \mathbb{C}^{L \times L^2}$ is the Gabor matrix with window a_ℓ (cf. (7)), $[\mathbf{y}]_p := y_p$, and the non-zeros and the positions, indexed by (m_n, q_n), of the non-zeros of $\mathbf{b} \in \mathbb{C}^{L^2}$ correspond to the b_n and the (τ_n, ν_n), respectively. Thus, recovery of the (b_n, τ_n, ν_n) amounts to recovering the S-sparse vector $\mathbf{b} \in \mathbb{C}^{L^2}$ from the measurement $\mathbf{y} \in \mathbb{C}^L$. This is a sparse signal recovery problem with a Gabor measurement matrix. A—by now standard—recovery approach is to solve a simple convex ℓ_1-norm-minimization program. From [20, Thm. 5.1] we know that, provided the a_ℓ are i.i.d. sub-Gaussian random variables, and provided that $L \geq O(S \log^2(S) \log^2(L))$, with high probability, all S-sparse vectors $\mathbf{b}$ can be recovered from $\mathbf{y}$ via ℓ_1-minimization.

1.2 Only time or only frequency shifts

We next consider the case of only time or only frequency shifts, and show that in both cases recovery of the (b_n, τ_n) or (b_n, ν_n), is equivalent to the recovery of a weighted superposition of spikes from low-frequency samples. Specifically, if $\tau_n = 0$ for all n, (2) reduces to

$$y_p = a_p \sum_{n=1}^{S} b_n e^{i2\pi p \nu_n}, \quad p = -N, ..., N. \tag{4}$$

The y_p above are samples of a mixture of S complex sinusoids, and estimation of the ν_n corresponds to determining the frequency components of those sinusoids. Expressed differently, the y_p are the lowest L Fourier series coefficients of a signal x that is a weighted (by the b_n) superposition of Dirac measures at locations ν_n. In this sense, the y_p correspond to the low-frequency components of x. Estimation of the ν_n is a line spectral estimation problem, and can be solved using approaches such as Proney's method [12, Ch. 2]. It has been shown recently [8] that recovery of the (b_n, ν_n) can be accomplished by solving a convex total-variation norm minimization program, provided that the minimum separation between any two ν_n is larger than $2/N$. This result is interesting as it shows that the positions of the spices can be identified exactly by solving a (simple) convex program.

An analogous situation arrises in the case of $\tau_n = 0$ for all n: taking the discrete Fourier transform of y_p yields a relation exactly of the form (4).

1.3 Main contribution

In this paper, we consider a random probing signal $x(t)$ by taking its samples a_ℓ to be i.i.d. Gaussian (or sub-Gaussian) random variables. We show that, with probability at least $1 - \delta$, the (b_n, τ_n, ν_n) can be recovered perfectly from the L samples y_p by essentially solving a simple convex program provided that the $(\tau_n, \nu_n) \in [0,1]^2, n = 1, ..., S$, satisfy the minimum distance condition

$$\min(|\tau_n - \tau_m|, |\nu_n - \nu_m|) \geq \frac{2.38}{N}, \text{ for all } n \neq m, \tag{5}$$

where $|\tau_n - \tau_m|$ is the ℓ_∞-distance (i.e., the wrap-around distance on the unit circle), and provided that

$$L \geq Sc \log^3 \left(\frac{c' L^6}{\delta} \right)$$

where c and c' are numerical constants.

Translated to the continuous setup, our result implies that, with probability at least $1-\delta$, we can identify the $(b_n, \bar{\tau}_n, \bar{\nu}_n)$ perfectly provided that

$$|\bar{\tau}_n - \bar{\tau}_m| \geq \frac{4.77}{B} \text{ and } |\bar{\nu}_n - \bar{\nu}_m| \geq \frac{4.77}{T}$$

and

$$BT \geq Sc \log^3\left(\frac{c'(BT)^6}{\delta}\right).$$

This is essentially optimal, as the number S of unknowns can be linear—up to a log-factor—in the dimensionality BT of the observation $(Hx)(t)$.

Finally note that $(\tau_n, \nu_n) \in [0,1]^2$ translates to $(\bar{\tau}_n, \bar{\nu}_n) \in [0,T] \times [0,B]$, i.e., the $(\bar{\tau}_n, \bar{\nu}_n)$ can lie in a rectangle of area $L = BT \gg 1$, i.e., the system H does not need to be underspread[2].

1.4 Notation

We use lowercase boldface letters to denote (column) vectors and uppercase boldface letters to designate matrices. The superscripts ^{T} and ^{H} stand for transposition and Hermitian transposition, respectively. For the vector $\mathbf{x}$, x_q and $[\mathbf{x}]_q$ denote its qth entry, $\|\mathbf{x}\|_2$ its ℓ_2-norm and $\|\mathbf{x}\|_\infty = \max_q |x_q|$ its largest entry. For the matrix $\mathbf{A}$, $\mathbf{A}_{ij}$ and $[\mathbf{A}]_{ij}$ designates the entry in its ith row and jth column, $\|\mathbf{A}\| \coloneqq \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2$ its spectral norm, and $\|\mathbf{A}\|_F \coloneqq (\sum_{i,j} |\mathbf{A}_{ij}|^2)^{1/2}$ its Frobenius norm. The identity matrix is denoted by $\mathbf{I}$. For a complex number b with polar decomposition $b = |b|e^{i2\pi\phi}$, $\text{sign}(b) \coloneqq e^{i2\pi\phi}$. Similarly, for a vector $\mathbf{b}$, $[\text{sign}(\mathbf{b})]_k \coloneqq \text{sign}([\mathbf{b}]_k)$. For the set $\mathcal{T}$, $|\mathcal{T}|$ designates its cardinality and $\overline{\mathcal{T}}$ is its complement. The sinc-function is denoted as $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$. For vectors $\mathbf{r}, \mathbf{r}' \in [0,1]^2$, $|\mathbf{r} - \mathbf{r}'|$ is the ∞-distance. Here, the distance on each coordinate is understood as the wrap-around distance on the unit circle. Throughout, $\mathbf{r}$ denotes a 2-dimensional vector with entries τ and ν, i.e., $\mathbf{r} = [\tau, \nu]^T$. Moreover $c, \tilde{c}, c', c_1, c_2, ...$ are numerical constants that can take on different values at different occurrences. Finally, $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2.

2 Recovery via convex optimization

In this section we present our approach for the recovery of the parameters (b_n, τ_n, ν_n) from the samples y_p in (2). Before we proceed we note that (2) can be rewritten as

$$y_p = \sum_{n=1}^{S} b_n \sum_{\ell, r=-N}^{N} D_N\left(\frac{\ell}{L} - \tau_n\right) D_N\left(\frac{r}{L} - \nu_n\right) a_{p-\ell} e^{i2\pi \frac{rp}{L}}, \quad p = -N, ..., N \tag{6}$$

where

$$D_N(t) \coloneqq \frac{\sin(\pi L t)}{L \sin(\pi t)}$$

is the Dirichlet kernel. Our signal model is a sparse linear combination of time and frequency shifted versions of the sequence a_ℓ. A regularizer that promotes such a sparse linear combination is the atomic norm induced by these signals [10]. We define atoms $\mathbf{a} \in \mathbb{C}^{L^2}$ as

$$[\mathbf{a}(\mathbf{r})]_{(\ell, r)} = D_N\left(\frac{\ell}{L} - \tau\right) D_N\left(\frac{r}{L} - \nu\right), \quad \mathbf{r} = [\tau, \nu]^T, \quad \ell, r = -N, ..., N.$$

[2] A system is called underspread if its spreading function is supported on a rectangle of area much less than one.

Rewriting (6) in matrix-vector form yields

$$\mathbf{y} = \mathbf{G}\mathbf{z}, \quad \mathbf{z} = \sum_{n=1}^{S} |b_n| e^{i2\pi\phi_n} \mathbf{a}(\mathbf{r}_n)$$

where $b_n = |b_n|e^{i2\pi\phi_n}$ is the polar decomposition of b_n and $\mathbf{G} \in \mathbb{C}^{L\times L^2}$ is the Gabor matrix defined by

$$[\mathbf{G}]_{p,(\ell,r)} := a_{p-\ell} e^{i2\pi\frac{rp}{L}}, \quad \ell, r, p = -N, ..., N. \tag{7}$$

The atoms in the set $\mathcal{A} := \{e^{i2\pi\phi}\mathbf{a}(\mathbf{r}), \mathbf{r} \in [0,1]^2, \phi \in [0,1]\}$ are the building blocks of the signal $\mathbf{z}$. The atomic norm $\|\cdot\|_{\mathcal{A}}$ is defined as

$$\|\mathbf{z}\|_{\mathcal{A}} = \inf\{t > 0 \colon \mathbf{z} \in t\,\mathrm{conv}(\mathcal{A})\} = \inf_{b_n\in\mathbb{C}, \mathbf{r}_n\in[0,1]^2} \left\{ \sum_n |b_n| \colon \mathbf{z} = \sum_n b_n \mathbf{a}(\mathbf{r}_n) \right\}$$

where $\mathrm{conv}(\mathcal{A})$ denotes the convex hull of the set $\mathcal{A}$. The atomic norm can enforce sparsity in $\mathcal{A}$ because low-dimensional faces of $\mathrm{conv}(\mathcal{A})$ correspond to signals involving only a few atoms [10, 27]. A natural algorithm for estimating $\mathbf{z}$ is the atomic norm minimization problem [10]

$$\mathrm{AN}(\mathbf{y})\colon \quad \underset{\mathbf{z}}{\text{minimize}}\, \|\mathbf{z}\|_{\mathcal{A}} \ \text{ subject to } \ \mathbf{y} = \mathbf{G}\mathbf{z}. \tag{8}$$

3 Main result

Our main result, stated below, provides conditions, under which atomic norm minimization perfectly recovers $\mathbf{z} = \sum_{n=1}^{S} b_n \mathbf{a}(\mathbf{r}_n)$. Once we obtain $\mathbf{z}$, recovery of the time-frequency shifts is a 2D line spectral estimation problem which can be solved with standard approaches such as Proney's method see e.g. [12, Ch. 2]. We also present a more direct recovery approach in Section 6.2. When the time-frequency shifts $\mathbf{r}_n$ are identified, the coefficients b_n can be obtained by solving the linear system of equations

$$\mathbf{y} = \sum_{n=1}^{S} b_n \mathbf{G}\mathbf{a}(\mathbf{r}_n).$$

At first sight, computation of the atomic norm involves taking the infinimum over infinitely many parameters. However, since the atomic norm can be characterized in terms of linear matrix inequalities, (8) can be formulated as a semidefinite program, which allows to recover $\mathbf{z}$ efficiently. Instead of taking that route, and explicitly stating the corresponding semidefinite program, we show later that the time-frequency shifts $\mathbf{r}_n$ can be identified directly from the dual solution of the atomic norm minimization problem (8). As shown later, the dual of (8) has a semidefinite programming formulation as well.

Theorem 1. *Let $\mathcal{T} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_S\} \subset [0,1]^2$ be any set of points obeying the minimum distance condition*

$$|\mathbf{r}_j - \mathbf{r}_k| \geq \frac{2.38}{N}, \text{ for all } \mathbf{r}_j, \mathbf{r}_k \in \mathcal{T} \text{ with } \mathbf{r}_j \neq \mathbf{r}_k,$$

assume that $N \geq 512$, and let the coefficients $a_\ell, \ell = -N, ..., N$ be chosen i.i.d. from $\mathcal{N}(0, 1/L)$, $L = 2N+1$, and let the $\mathrm{sign}(b_n)$ be i.i.d. uniform on $\{-1, 1\}$. Finally, let $\mathbf{y}$ be the vector with entries y_p as defined in (2), i.e.,

$$\mathbf{y} = \mathbf{G}\mathbf{z}, \quad \mathbf{z} = \sum_{\mathbf{r}_n\in\mathcal{T}} b_n \mathbf{a}(\mathbf{r}_n).$$

Suppose that

$$L \geq Sc \log^3 \left(\frac{c' L^6}{\delta} \right).$$

Then, with probability at least $1 - \delta$, $\mathbf{z}$ *is the unique minimizer of* $\mathrm{AN}(\mathbf{y})$.

The proof of Theorem 1 is based on analyzing the dual problem, specifically we will certify optimality by constructing a dual certificate (see Section 6). The construction of this dual certificate, formalized by Proposition 3 in Section 8, is the main technical contribution of this paper.

The random sign of the coefficients b_n essentially assumes that the time-frequency shifts in (1) have random phase. To keep the proof simple, we assumed that the b_n are real, however the result continues to hold for complex b_n (only the constant 2.38 in (5) changes slightly). The random-phase (i.e., sign) model is in line with standard models in wireless communication and in radar [4], where the b_n are assumed complex Gaussian. Nevertheless, we believe that the random sign assumption is not needed for our statement, and leave a corresponding result to future work.

Finally note that Theorem 1 continues to hold for sub-Gaussian a_ℓ.

4 Detailed problem formulation

In this section we derive the input-output relation (2). As mentioned previously, radar and wireless communication channels are typically modeled as linear systems whose response is a weighted superposition of delayed and Doppler shifted versions of the input signal. In general, the response of the system to the input signal $x(t)$ is given by

$$y(t) = \iint s_H(\tau, \nu) x(t - \tau) e^{i2\pi\nu t} d\nu d\tau \tag{9}$$

where $s_H(\tau, \nu)$ denotes the spreading function associated with the system. The input signal $x(t)$ can be controlled by the system engineer and is known in the channel identification and radar problems. The spreading function depends on the scene and is unknown. We assume that the spreading function consists of S point scatterers. In radar, those point scatterers correspond to moving targets. Mathematically, this means that the spreading function specializes to

$$s_H(\tau, \nu) = \sum_{n=1}^{S} b_n \delta(\tau - \bar{\tau}_n) \delta(\nu - \bar{\nu}_n). \tag{10}$$

Here, b_n, $n = 1, ..., S$, is a (complex) attenuation factor associated with the time frequency-shift $(\bar{\tau}_n, \bar{\nu}_n)$. Owing to path loss and finite velocity of the targets or objects in the scene, we may assume that [26]

$$(\bar{\tau}_n, \bar{\nu}_n) \in [0, \tau_{\max}] \times [0, \nu_{\max}] \tag{11}$$

for some constants $\tau_{\max}, \nu_{\max}$. With (10), (9) reduces to (1), i.e., to

$$y(t) = \sum_{n=1}^{S} b_n x(t - \bar{\tau}_n) e^{i2\pi\bar{\nu}_n t}.$$

The goal is to identify the triplets $(b_n, \bar{\tau}_n, \bar{\nu}_n)$ from the response $y(t)$ to a (known) probing signal $x(t)$, as those parameters characterize the system. For the radar application, this yields the position and relative speed of the objects.

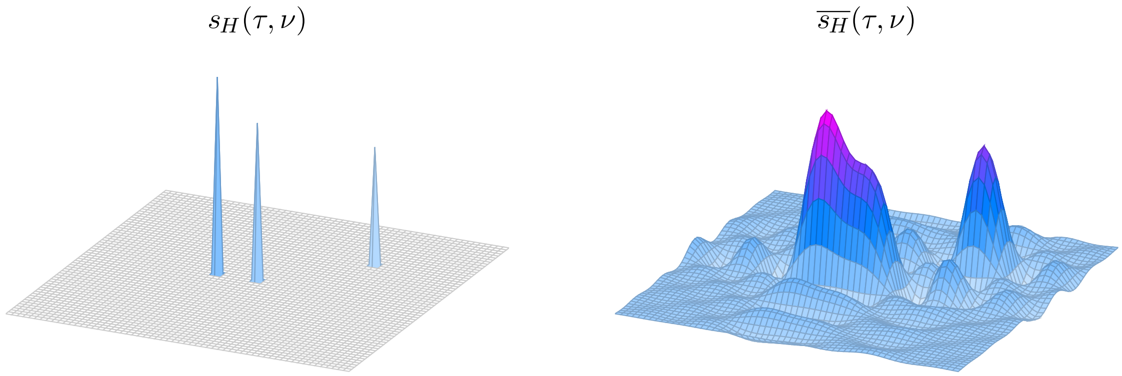

Figure 1: Illustration of the spreading function $s_H(\tau,\nu)$ and the corresponding smeared spreading function $\overline{s_H}(\tau,\nu)$.

In practice, the input signal $x(t)$ has finite bandwidth B and the received signal $y(t)$ can only be observed over a finite time interval of length T. As shown next, this time and band-limitation leads to a discretization of the input-output relation (9) and determines the "natural" resolution of the system of $1/B$ and $1/T$ in τ- and ν-direction, respectively. Specifically, using that $x(t)$ is band-limited to $[0,B)$, and $y(t)$ is time-limited to $[-T/2,T/2)$ (9) becomes [4] (see Appendix A for details)

$$y(t) = \sum_{\ell\in\mathbb{Z}}\sum_{r\in\mathbb{Z}} \overline{s_H}\left(\frac{\ell}{B},\frac{r}{T}\right) x\left(t-\frac{\ell}{B}\right) e^{i2\pi \frac{r}{T}t} \tag{12}$$

where

$$\overline{s_H}(\tau,\nu) := \iint s_H(\tau',\nu')\,\mathrm{sinc}((\tau-\tau')B)\,\mathrm{sinc}((\nu-\nu')T)d\tau' d\nu' \tag{13}$$

is a smeared version of the original spreading function. For points scatterers, i.e., for the spreading function in (10), (13) specializes to

$$\overline{s_H}(\tau,\nu) = \sum_{n=1}^{S} b_n\,\mathrm{sinc}((\tau-\bar{\tau}_n)B)\,\mathrm{sinc}((\nu-\bar{\nu}_n)T). \tag{14}$$

Imagine for a moment that we could measure $\overline{s_H}(\tau,\nu)$ directly. We see that $\overline{s_H}(\tau,\nu)$ is the 2D low-pass-filtered version of the signal $s_H(\tau,\nu)$ in (10), where the filter has resolution $1/B$ in τ direction and resolution $1/T$ in ν direction, see Figure 1 for an illustration. Estimation of the triplets $(b_n,\bar{\tau}_n,\bar{\nu}_n)$, $n=1,\dots,S$, from $\overline{s_H}(\tau,\nu)$ is the classical 2D line spectral estimation problem (see [12] and references therein). In our setup, the situation is further complicated by the fact that we can not measure $\overline{s_H}(\tau,\nu)$ directly. We only get access to $\overline{s_H}(\tau,\nu)$ after the application of the Gabor linear operator in (12).

4.1 Sampling the output

Since $y(t)$ is exactly time-limited to $[-T/2,T/2)$ and approximately band-limited to $[0,B)$ (we assume $\nu_{\max} \ll B$) it is well-described by its samples

$$y\left(\frac{p}{B}\right) = \sum_{\ell,r\in\mathbb{Z}} \overline{s_H}\left(\frac{\ell}{B},\frac{r}{T}\right) x\left(\frac{p-\ell}{B}\right) e^{i2\pi\frac{rp}{BT}}, \quad p=-N,...,N \tag{15}$$

where $N = (BT-1)/2$ (recall that we assume for convenience that BT is an odd integer). Substituting (14) into (15) yields

$$y\left(\frac{p}{B}\right) = \sum_{n=1}^{S} b_n \sum_{\ell,r\in\mathbb{Z}} \operatorname{sinc}(\ell - \tau_n B)\operatorname{sinc}(r-\nu_n T)\, x\left(\frac{p-\ell}{B}\right) e^{i2\pi \frac{rp}{BT}}. \tag{16}$$

We next specify the probing signal $x(t)$ by specifying its samples $x(\ell/B)$. We take the samples of $x(t)$ to be

$$x(\ell/B) = a_\ell$$

where a_ℓ is a L-periodic sequence (recall that $L = 2N+1 = BT$). With $y_p := y\left(\frac{p}{B}\right)$, (16) becomes (see Appendix B for details)

$$y_p = \sum_{n=1}^{S} b_n \sum_{\ell,r=-N}^{N} D_N\left(\frac{\ell}{L} - \tau_n\right) D_N\left(\frac{r}{L} - \nu_n\right) a_{p-\ell} e^{i2\pi\frac{rp}{L}}, \quad p = -N,...,N \tag{17}$$

where $\tau_n = \bar{\tau}_n \frac{B}{L} = \bar{\tau}_n/T$, $\nu_n = \bar{\nu}_n \frac{T}{L} = \bar{\nu}_n/B$ and

$$D_N(t) := \frac{\sin(\pi L t)}{L\sin(\pi t)}.$$

Rewriting the input-output relation (17) yields (2) (see Appendix B for details).

5 Discrete super-resolution radar

An obvious approach to estimate the time-frequency shifts (τ_n, ν_n), from the samples y_p in (2) (recall that once the time frequency shifts are known, estimation of the b_n is trivial) is to suppose the time-frequency shifts lie on a fine grid, and solve the problem on that grid. This leads to a gridding error that becomes small as the grid becomes finer. Our results have immediate consequences for the corresponding (discrete) sparse signal recovery problem, which are discussed in this section.

Suppose we want to recover a sparse discrete signal $s_{m,n} \in \mathbb{C},\ m,n = 0,...,K-1,\ K \geq L = 2N+1$, from samples of the form

$$y_p = \sum_{m,n=0}^{K-1} \left(e^{i2\pi p \frac{m}{K}} \frac{1}{L} \sum_{\ell,k=-N}^{N} e^{-i2\pi k\frac{n}{K}} e^{i2\pi(p-\ell)\frac{k}{L}} a_\ell \right) s_{m,n}, \quad p = -N,...,N. \tag{18}$$

To see the connection to the continuous setup in the previous sections, note that recovery of the S-sparse (discrete) signal $s_{m,n}$ is equivalent to recovery of the (τ_n, ν_n, b_n) from the samples y_p in (2) under the proviso that the (τ_n, ν_n) lie on a $(1/K, 1/K)$ grid[3] (the non-zeros of $s_{m,n}$ correspond to the b_n). Writing the relation (18) in matrix-vector form yields

$$\mathbf{y} = \mathbf{R}\mathbf{s}$$

where $[\mathbf{y}]_p := y_p$, $[\mathbf{s}]_{(m,n)} := s_{m,n}$, and $\mathbf{R} \in \mathbb{C}^{L\times K^2}$ is the matrix with entry in the pth row and (m,n)th column given by the term in the bracket in (18). The matrix $\mathbf{R}$ contains as columns

[3] The discussion in this section could analogously be conduced for a signal on a grid with different spacing in τ and in ν-direction, i.e., $\{(m/K, n/\tilde{K})\}$ with $m = 0,...,K-1$, $n = 0,...,\tilde{K}-1$, and $K \neq \tilde{K}$.

partial time-frequency shifts of the sequence a_ℓ. If $K = L$, $\mathbf{R}$ contains as columns only "whole" time-frequency shifts of the sequence a_ℓ and $\mathbf{R}$ is equal to the Gabor matrix $\mathbf{G}$ defined by (7). In this sense, $K = L$ is the natural grid (cf. Section 1.1) and the ratio K/L can be interpreted as a super resolution factor. The super resolution factor SRF $:= K/L$ determines by how much the $(1/K, 1/K)$ grid is finer than the original $(1/L, 1/L)$ grid.

A standard approach to recover the sparse signal $\mathbf{s}$ from the underdetermined linear system of equations $\mathbf{y} = \mathbf{R}\mathbf{s}$ is to solve the following convex program:

$$\underset{\hat{\mathbf{s}}}{\text{minimize}} \ \|\hat{\mathbf{s}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{R}\hat{\mathbf{s}}. \tag{19}$$

Theorem 2. *Let $\mathcal{T} \subseteq \{0, ..., K-1\}^2$ be the support of the vector $\{[\mathbf{s}]_{(m,n)}\}_{m,n=0}^{K-1}$, obeying*

$$\min_{(m,n),(\tilde{m},\tilde{n})\in\mathcal{T}\colon (m,n)\neq(\tilde{m},\tilde{n})} \frac{1}{K}\max(|m-\tilde{m}|, |n-\tilde{n}|) \geq \frac{2.38}{N}$$

and suppose that the non-zeros of $\mathbf{s}$ have random signs, i.e., $\text{sign}([\mathbf{s}]_{(m,n)}), (m,n) \in \mathcal{T}$ are i.i.d. uniform on $\{-1, 1\}$. Suppose that $N \geq 512$ and let the coefficients $a_\ell, \ell = -N, ..., N$ be chosen i.i.d. from $\mathcal{N}(0, 1/L)$, $L = 2N+1$, and suppose that

$$L \geq Sc\log^3\left(\frac{c' L^6}{\delta}\right).$$

Then, with probability at least $1-\delta$, the solution to (19) is equal to $\mathbf{s}$.

Theorem 2 is proven by constructing a dual certificate (see Appendix C for details). The dual certificate is obtained directly from the dual certificate for the continuous case in Proposition 3 in Section 8.

5.1 Implementation details

The matrix $\mathbf{R}$ has dimension $L \times K^2$, thus as the grid becomes finer (i.e., K becomes larger) the complexity of solving (19) increases. However, the complexity of solving (19) can be reduced as follows. The matrix-vector multiplication $\mathbf{R}\mathbf{s}$ can be implemented efficiently using the fast Fourier transform. This allows to accelerate first-order solvers such as TFOCS [3]. Second, and more importantly, in practice we know that by (11), we have $(\bar{\tau}_n, \bar{\nu}_n) \in [0, \tau_{\max}] \times [0, \nu_{\max}]$, which means that

$$(\tau_n, \nu_n) \in \left[0, \frac{\tau_{\max}}{T}\right] \times \left[0, \frac{\nu_{\max}}{B}\right]. \tag{20}$$

It is therefore sufficient to consider the restriction of $\mathbf{R}$ to the $\frac{K^2}{L\tau_{\max}\nu_{\max}} = \frac{K^2}{BT\tau_{\max}\nu_{\max}}$ columns corresponding to the (τ_n, ν_n) satisfying (20). Since typically $\tau_{\max}\nu_{\max} \ll BT$, this results in a significant reduction of the problem size.

6 Dualization and identification of time-frequency shifts

We next study the dual problem of the atomic norm minimization problem (8). Our proof of Theorem 1 is based on analyzing the dual problem, specifically we will certify optimality by constructing a dual certificate. We will also show that the time-frequency shifts can be obtained from a solution to the dual problem.

The dual norm of $\|\cdot\|_{\mathcal{A}}$ is defined as

$$\|\mathbf{v}\|_{\mathcal{A}^*} = \sup_{\|\mathbf{z}\|_{\mathcal{A}} \leq 1} \mathrm{Re}\, \langle \mathbf{v}, \mathbf{z} \rangle = \sup_{\mathbf{r} \in [0,1]^2, \phi \in [0,1]} \mathrm{Re} \left\langle \mathbf{v}, e^{i2\pi\phi} \mathbf{a}(\mathbf{r}) \right\rangle = \sup_{\mathbf{r} \in [0,1]^2} |\langle \mathbf{v}, \mathbf{a}(\mathbf{r}) \rangle| \,.$$

The dual problem of (8) is [7, Sec. 5.1.16]

$$\underset{\mathbf{q}}{\text{maximize}}\ \mathrm{Re}\, \langle \mathbf{q}, \mathbf{y} \rangle \ \text{ subject to } \left\| \mathbf{G}^H \mathbf{q} \right\|_{\mathcal{A}^*} \leq 1 \tag{21}$$

where $\mathbf{q} = [q_{-N}, ..., q_N]^T$.

6.1 Semidefinte programming formulation of the dual problem

We next show that the dual can be cast as a semidefinite program. The corresponding formulation is similar to the related convex programs in [6, Sec. 3.1], [8, Sec. 4], and [27, Sec. 2.2]. We first show that the constraint in the dual is a 2D trigonometric polynomial that is bounded by one, and can therefore be formulated as a matrix inequality.

The constraint in the dual (21) is

$$\left\| \mathbf{G}^H \mathbf{q} \right\|_{\mathcal{A}^*} = \sup_{\mathbf{r} \in [0,1]^2} \left| \left\langle \mathbf{G}^H \mathbf{q}, \mathbf{a}(\mathbf{r}) \right\rangle \right| \leq 1.$$

The vector $\mathbf{a}(\mathbf{r})$ can be written as

$$\mathbf{a}(\mathbf{r}) = \mathbf{F}^H \mathbf{v}(\mathbf{r})$$

where $\mathbf{F}^H$ is the (inverse) 2D discrete Fourier transform matrix, i.e., $[\mathbf{F}^H]_{(\ell,r),(k,q)} := \frac{1}{L^2} e^{i2\pi \frac{k\ell + qr}{L}}$ and the entires of the vector $\mathbf{v}$ are given by $[\mathbf{v}(\mathbf{r})]_{(k,q)} = e^{-i2\pi(k\tau + q\nu)}$, where $\ell, r, k, q = -N, ..., N$. With these definitions,

$$\left\langle \mathbf{G}^H \mathbf{q}, \mathbf{a}(\mathbf{r}) \right\rangle = \left\langle \mathbf{G}^H \mathbf{q}, \mathbf{F}^H \mathbf{v}(\mathbf{r}) \right\rangle = \left\langle \mathbf{F}\mathbf{G}^H \mathbf{q}, \mathbf{v}(\mathbf{r}) \right\rangle = \sum_{k,q=-N}^{N} [\mathbf{F}\mathbf{G}^H \mathbf{q}]_{(k,q)} e^{i2\pi(k\tau + q\nu)}. \tag{22}$$

Thus, the constraint in the dual (21) says that the 2D trigonometric polynomial in (22) is bounded in magnitude by 1 for $\mathbf{r} \in [0,1]^2$. The following form of the bounded real lemma allows to express this constraint as a matrix inequality.

Proposition 1 ([11, Cor. 4.25, p. 127])**.** *Let P be a bivariate trigonometric polynomial in $\mathbf{r} \in [0,1]^2$*

$$P(\mathbf{r}) = \sum_{k,q=-N}^{N} p_{(k,q)} e^{i2\pi(k\tau + q\nu)}.$$

Then $\sup_{\mathbf{r} \in [0,1]^2} |P(\mathbf{r})| \leq 1$ holds if and only if there exists a Hermitian matrix $\boldsymbol{Q} \in \mathbb{C}^{L^2 \times L^2}$, $L = 2N + 1$, such that

$$\begin{bmatrix} \mathbf{Q} & \mathbf{p} \\ \mathbf{p}^H & 1 \end{bmatrix} \succeq \mathbf{0} \qquad \text{and} \qquad \forall k, q = -N, ..., N, \quad \mathrm{trace}((\boldsymbol{\Theta}_k \otimes \boldsymbol{\Theta}_q)\boldsymbol{Q}) = \begin{cases} 1, & (k,q) = (0,0) \\ 0, & \text{otherwise} \end{cases}$$

where $\boldsymbol{\Theta}_k$ designates the elementary Toeplitz matrix with ones on the k-th diagonal and zeros elsewhere.

By Proposition (1), the constraint of the dual program (21) is satisfied if and only if there exists a matrix $\mathbf{Q}$ such that

$$\begin{bmatrix} \mathbf{Q} & \mathbf{F}\mathbf{G}^H\mathbf{q} \\ \mathbf{q}^H\mathbf{G}\mathbf{F}^H & 1 \end{bmatrix} \succeq \mathbf{0}, \quad \operatorname{trace}((\boldsymbol{\Theta}_k \otimes \boldsymbol{\Theta}_q)\boldsymbol{Q}) = \begin{cases} 1, & (k,q) = (0,0) \\ 0, & \text{otherwise} \end{cases}. \tag{23}$$

Thus, the dual program (21) has the following equivalent semidefinite programming formulation:

$$\underset{\mathbf{q},\mathbf{Q}}{\operatorname{maximize}} \operatorname{Re}\langle \mathbf{q}, \mathbf{y}\rangle \text{ subject to (23)}. \tag{24}$$

6.2 Estimation of the time-frequency shifts from the dual solution

Since the primal problem only has equality constraints, Slater's condition holds which implies strong duality [7, Sec. 5.2.3]. The following proposition is a consequence of strong duality and provides a way to certify optimality of a solution $\mathbf{z}$ to (8). The proof, provided in Appendix E is standard, see e.g., [27, Proof of Prop. 2.4].

Proposition 2. *Let $\mathbf{y} = \mathbf{G}\mathbf{z}$ with $\mathbf{z} = \sum_{\mathbf{r}_n \in \mathcal{T}} b_n \mathbf{a}(\mathbf{r}_n)$. If there exists a dual polynomial*

$$Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle$$

with complex coefficients $\mathbf{q} = [q_{-N}, ..., q_N]^T$ such that

$$Q(\mathbf{r}_n) = \operatorname{sign}(b_n), \ \textit{for all } \mathbf{r}_n \in \mathcal{T}, \ \textit{and } |Q(\mathbf{r})| < 1 \ \textit{for all } \mathbf{r} \in [0,1]^2 \setminus \mathcal{T} \tag{25}$$

then $\mathbf{z}$ is the unique minimizer of $\operatorname{AN}(\mathbf{y})$. Moreover, $\mathbf{q}$ is a dual optimal solution.

The proof of Theorem 1 is based on constructing a dual polynomial satisfying the conditions of Proposition 2, see Section 8.

Proposition 2 suggests that an estimate $\hat{\mathcal{T}}$ of the set of time-frequency shifts $\mathcal{T}$ can be obtained from a dual solution $\mathbf{q}$ by identifying the $\mathbf{r}_n$ with the $\mathbf{r}$ for which the dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle$ achieves magnitude 1. In general, a solution $\hat{\mathbf{q}}$ to (24) is not unique but we can ensure that

$$\mathcal{T} \subseteq \hat{\mathcal{T}} := \{\mathbf{r} \colon |\langle \hat{\mathbf{q}}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle| = 1\}.$$

This is seen as follows. Assume that $\mathcal{T} \setminus \hat{\mathcal{T}} \neq \emptyset$. We then have that

$$\begin{aligned} \operatorname{Re}\langle \hat{\mathbf{q}}, \mathbf{G}\mathbf{z}\rangle &= \operatorname{Re}\left\langle \hat{\mathbf{q}}, \mathbf{G}\sum_{\mathbf{r}_n \in \mathcal{T}} b_n \mathbf{a}(\mathbf{r}_n)\right\rangle \\ &= \sum_{\mathbf{r}_n \in \mathcal{T} \cap \hat{\mathcal{T}}} \operatorname{Re}(b_n^* \langle \hat{\mathbf{q}}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle) + \sum_{\mathbf{r}_n \in \mathcal{T} \setminus \hat{\mathcal{T}}} \operatorname{Re}(b_n^* \langle \hat{\mathbf{q}}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle) \\ &< \sum_{\mathbf{r}_n \in \hat{\mathcal{T}} \cap \mathcal{T}} |b_n| + \sum_{\mathbf{r}_n \in \mathcal{T} \setminus \hat{\mathcal{T}}} |b_n| = \|\mathbf{z}\|_{\mathcal{A}} \end{aligned}$$

where strict inequality follows from $|\langle \hat{\mathbf{q}}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle| < 1$ for $\mathbf{r} \in \mathcal{T} \setminus \hat{\mathcal{T}}$, by definition of the set $\hat{\mathcal{T}}$. This contradicts strong duality, and thus implies that $\mathcal{T} \setminus \hat{\mathcal{T}} = \emptyset$, i.e., we must have $\mathcal{T} \subseteq \hat{\mathcal{T}}$.

In general, we might have $\mathcal{T} \neq \hat{\mathcal{T}}$. However, in "most cases", standard semidefinite programming solvers will yield a solution such that $\mathcal{T} = \hat{\mathcal{T}}$. Specifically, the following can be established (not

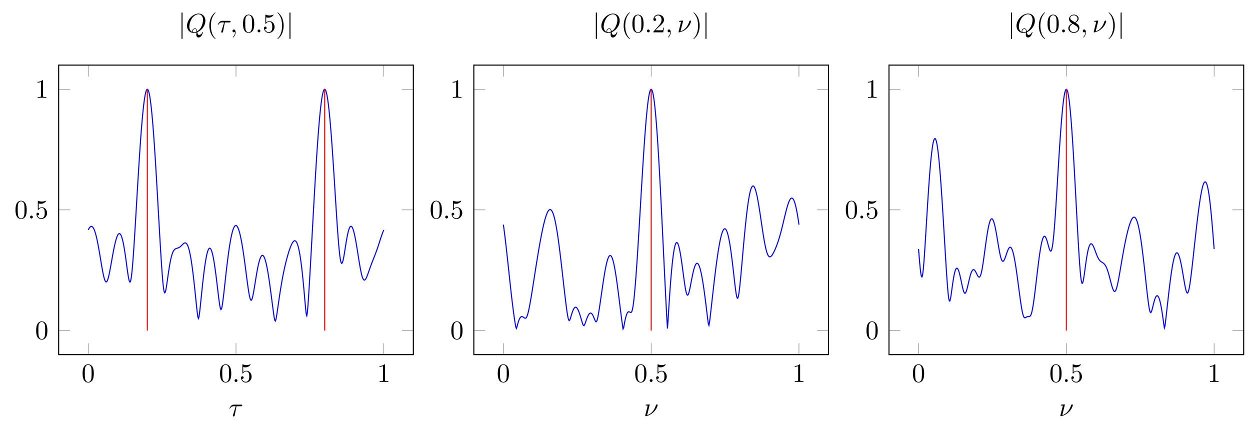

Figure 2: Localization of the time-frequency shifts via the dual polynomial $Q(\tau, \nu)$: $Q(\tau, \nu)$ satisfies $|Q(\tau, \nu)| = 1$ if $(\tau, \nu) \in \{(0.2, 0.5), (0.8, 0.5)\}$ and $|Q(\tau, \nu)| < 1$ otherwise. The red lines show the actual positions of the time-frequency shifts.

shown here; the proof is analogous to that of [27, Prop. 2.5], see also [8, Sec. 4]). Provided there exists a solution $\tilde{\mathbf{q}}$ to the dual (24) such that

$$\langle \tilde{\mathbf{q}}, \mathbf{G}\mathbf{a}(\mathbf{r}_n) \rangle = \operatorname{sign}(b_n), \text{ for all } \mathbf{r}_n \in \mathcal{T}, \text{ and } |\langle \tilde{\mathbf{q}}, \mathbf{G}\mathbf{a}(\mathbf{r}_n) \rangle| < 1 \text{ for all } \mathbf{r} \in [0,1]^2 \setminus \mathcal{T}$$

and we use an interior point method such as SDPT3 to solve (24) we have that $\hat{\mathcal{T}} = \mathcal{T}$.

We next provide a numerical example where the time-frequency shifts can be recovered perfectly from a solution to the semidefinite program (24). We choose $N = 8$, consider the case of two time-frequency shifts, specifically $\mathcal{T} = \{(0.2, 0.5), (0.8, 0.5)\}$, and let the coefficients $a_\ell, \ell = -N, ..., N$ and the $b_n, n = 0, 1$, be i.i.d. uniform on the complex unit sphere. In Figure 2 we plot the dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}) \rangle$ with $\mathbf{q}$ obtained by solving (24) via CVX, a Matlab package for specifying and solving convex programs [13], which calls SDPT3. It can be seen that the time-frequency shifts can be recovered perfectly, i.e., $\hat{\mathcal{T}} = \mathcal{T}$.

6.3 Recovery in the noisy case

In practice, the samples y_p in (2) are corrupted by additive noise. In that case, perfect recovery of the (τ_n, ν_n, b_n) is in general no longer possible, and we can only hope to identify the time-frequency shifts up to an error. In the noisy case, we solve the following convex program:

$$\underset{\mathbf{z}}{\text{minimize}} \, \|\mathbf{z}\|_{\mathcal{A}} \text{ subject to } \|\mathbf{y} - \mathbf{G}\mathbf{z}\|_2 \leq \delta. \tag{26}$$

The semidefinite programing formulation of the dual of (26) takes on the form

$$\underset{\mathbf{q}, \mathbf{Q}}{\text{maximize}} \, \operatorname{Re} \langle \mathbf{q}, \mathbf{y} \rangle - \delta \|\mathbf{q}\|_2 \text{ subject to (23)} \tag{27}$$

and we again estimate the time-frequency shifts $\mathbf{r}_n$ as the $\mathbf{r}$ for which the dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}) \rangle$ achieves magnitude 1. We leave theoretical analysis of this approach to future work, and only provide a numerical example demonstrating that this approach is stable.

We choose $N = 5$, consider the case of one time-frequency shift at $(\tau_1, \nu_1) = (0.5, 0.8)$ (so that the dual polynomial can be easily plotted in 3D) and let the coefficients $a_\ell, \ell = -N, ..., N$ and b_1

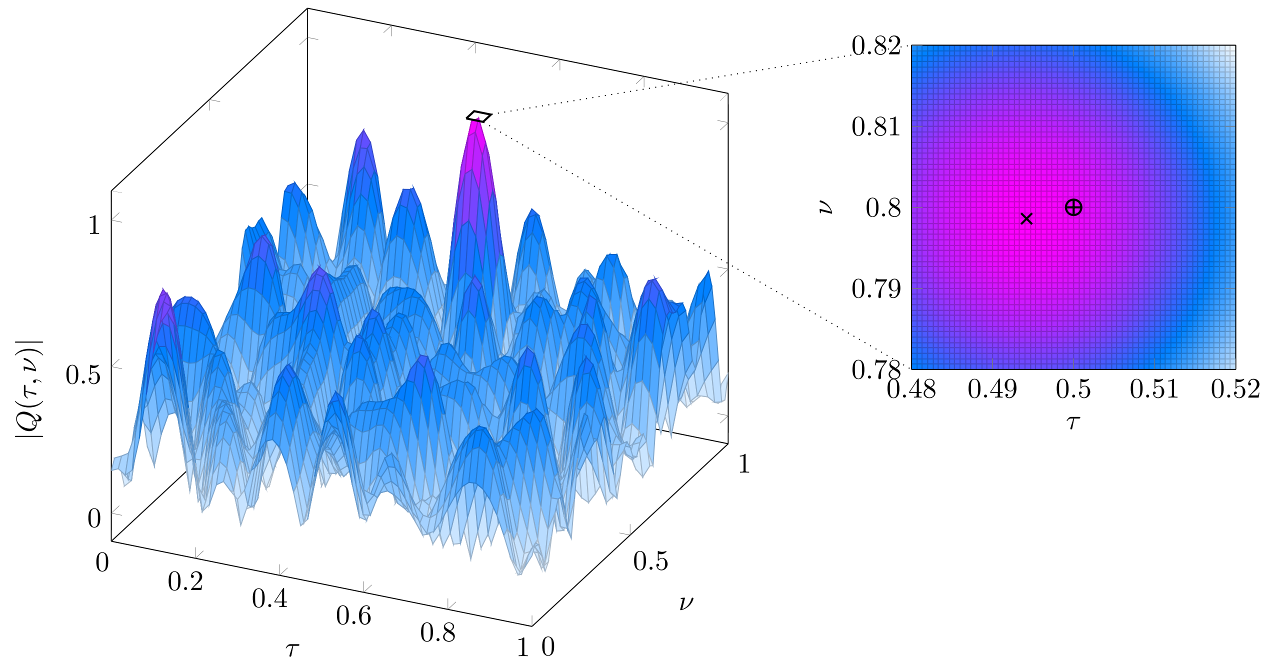

Figure 3: Localization of the time-frequency shifts via the dual polynomial $Q(\tau, \nu)$ from a noisy measurement (10dB noise): $Q(\tau, \nu)$ achieves magnitude 1 for $(\tau, \nu) = (0.4942, 0.7986)$ (marked by $\times$) which is very close to the original time-frequency shift $(0.5, 0.8)$ (marked by $\oplus$).

be i.i.d. uniform on the complex unit sphere. We add 10dB additive complex Gaussian noise to the samples y_p in (2). In Figure 3 we plot the dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{Ga}(\mathbf{r}) \rangle$ with $\mathbf{q}$ obtained by solving (27) (with $\delta = 0.8$) using CVX. The time-frequency shift for which the dual polynomial achieves magnitude 1 is $(0.4942, 0.7986)$ which is very close to the original time-frequency shift $(0.5, 0.8)$.

7 Relationship with previous work

The general problem of extracting the spreading function $s_H(\tau, \nu)$ of a linear time varying system of the form (9) is known as system identification. It has been shown that LTV systems with spreading function compactly supported on a region of area Δ in the time-frequency plane are identifiable if and only if $\Delta \leq 1$ [18, 5, 19, 23]. If the spreading function's support region is unknown, a necessary and sufficient condition for identifiability is $\Delta \leq 1/2$ [15]. In contrast to our work, the input (probing) signal in [18, 5, 19, 23, 15] is not constraint to be band-limited, and the response to the input signal is not constrained to be time-limited. In fact, the probing signal in those works is a (weighted) train of Dirac impulses, which neither decays in time nor in frequency.

Tauböck et al. [28] and Bajwa et al. [2] considered the identification of LTV systems with spreading function compactly supported in a rectangle of area $\Delta \leq 1$. In [28, 2], the time-frequency shifts lie on a (coarse) grid. In our setup, the time frequency shifts must not lie on a grid and may in principle lie in an rectangle of area $L = BT$ that is considerably larger than 1. Herman and Strohmer [16], in the context of compressed sensing radar, and Pfander et al. [22] considered the case where the time-frequency shifts lie on a $(\frac{1}{B}, \frac{1}{T})$ grid, cf. Section 1.1. Baiwa et al. [1] considered the identification of an LTV of the form (1). The approach in [1] requires the time frequency shifts

(τ_n, ν_n) to lie in a rectangle of area much less than 1, i.e., the system needs to be underspread, and requires $(BT)^2 \geq cS$ in general, both assumption are not required here.

In [8], Candès and Fernandez-Granda study the recovery of the frequency components of a mixture of S complex sinusoids from L equally spaced samples (cf. (4)). As mentioned previously, this corresponds to the case of only time or only frequency shifts. Tang et al. [27] study a related problem, namely the recovery of the frequency components from a random subset of the L equally spaced samples. Both [8, 27] study convex algorithms analogous to the algorithm studied here, and the proof techniques of the corresponding performance results inspired the analysis presented in this paper. In [25] Soltanolkotabi improved the results of [8] with simpler proofs by building approximate dual certificates. We believe that one can utilize this result to simplify our proofs and/or remove the random sign assumption. We leave this to future work.

8 Proof of Theorem 1

Theorem 1 is established by constructing a dual polynomial which satisfies the conditions of Proposition 2. This construction is formalized by the following result.

Proposition 3. *Let $\mathcal{T} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_S\} \subset [0,1]^2$ be any set of points obeying the minimum distance condition*

$$|\mathbf{r}_j - \mathbf{r}_k| \geq \frac{2.38}{N}, \text{ for all } \mathbf{r}_j, \mathbf{r}_k \in \mathcal{T} \text{ with } \mathbf{r}_j \neq \mathbf{r}_k,$$

and assume that $N \geq 512$. Let the coefficients $a_\ell, \ell = -N, ..., N$ be chosen i.i.d. from[4] *$\mathcal{N}(0, 1/L)$, $L := 2N+1$, and let the entries of $\mathbf{u} \in \{-1,1\}^S$ be i.i.d. uniform on $\{-1,1\}$. Suppose that*

$$L \geq Sc \log^3\left(\frac{c' L^6}{\delta}\right).$$

Then, with probability at least $1-\delta$, there exists a trigonometric polynomial $Q(\mathbf{r})$, $\mathbf{r} = [\tau, \nu]^T$, of the form

$$Q(\mathbf{r}) = \sum_{p=-N}^{N} \left(e^{-i2\pi p\nu} \sum_{k,\ell=-N}^{N} a_\ell e^{i2\pi(p-\ell)\frac{k}{L}} e^{-i2\pi k\tau} \right) q_p \tag{28}$$

with complex coefficients q_p such that

$$Q(\mathbf{r}_j) = u_j, \text{ for all } \mathbf{r}_j \in \mathcal{T}, \text{ and } |Q(\mathbf{r})| < 1 \text{ for all } \mathbf{r} \in [0,1]^2 \setminus \mathcal{T}.$$

We provide a proof of Proposition 3 by constructing $Q(\mathbf{r})$ explicitly. Our construction of the polynomial $Q(\mathbf{r})$ is inspired by that in [8, 27], and builds on results derived in [8, 27]. We will construct the polynomial $Q(\mathbf{r})$ by interpolating the points $(\mathbf{r}_j, u_j)$, $j = 1, ..., S$, with shifted versions of the kernel $G(\mathbf{r})$ (defined below) and its partial derivatives according to

$$Q(\mathbf{r}) = \sum_{k=1}^{S} \alpha_k G(\mathbf{r} - \mathbf{r}_k) + \beta_{1k} G^{(1,0)}(\mathbf{r} - \mathbf{r}_k) + \beta_{2k} G^{(0,1)}(\mathbf{r} - \mathbf{r}_k) \tag{29}$$

[4]The proposition continues to hold for the a_ℓ zero mean sub-Gaussian with variance $1/L$.

where $G^{(m,n)}(\mathbf{r}) := \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} G(\mathbf{r})$. We will choose the coefficients $\alpha_k, \beta_{1k}, \beta_{2k}$ later in such a way that

$$Q(\mathbf{r}_j) = u_j, \quad Q^{(1,0)}(\mathbf{r}_j) = 0, \text{ and } Q^{(0,1)}(\mathbf{r}_j) = 0, \text{ for all } \mathbf{r}_j \in \mathcal{T} \tag{30}$$

and then show that the resulting polynomial, with high probability, satisfies $|Q(\mathbf{r})| < 1$ for all $\mathbf{r} \notin \mathcal{T}$. Requiring the partial derivatives of $Q(\mathbf{r})$ to be zero on $\mathcal{T}$ implies that the magnitude of $Q(\mathbf{r})$ reaches local maxima on $\mathcal{T}$, which will be important to establish $Q(\mathbf{r}) < 1$ for all $\mathbf{r} \notin \mathcal{T}$.

The polynomial $Q(\mathbf{r})$ is of the form (28), if the kernel $G(\mathbf{r})$ (and its partial derivatives $G^{(1,0)}(\mathbf{r})$ and $G^{(0,1)}(\mathbf{r})$) are of the form (28). To satisfy this, we construct the kernel $G(\mathbf{r})$ as follows. Define the vectors $\mathbf{f}(\mathbf{r}) \in \mathbb{C}^{L^2}$ and $\mathbf{g} \in \mathbb{C}^{L^2}$ as $[\mathbf{f}(\mathbf{r})]_{(k,p)} = e^{i2\pi(k\tau + p\nu)}$ and $[\mathbf{g}]_{(k,p)} = g(p)g(k)$, $k, p = -N, ..., N$, where the $g(k)$ are the coefficients of the squared Fejér kernel, defined in (33) below. Note that, for convenience, we will frequently use a two-dimensional index, e.g., we use (p, k) to index the coefficients of $\mathbf{g}$ and $\mathbf{f}$. For example $\mathbf{g}$ can simply be written as $\mathbf{g} = [g_{(-N,-N)}, g_{(-N,-N+1)}, ..., g_{(-N,N)}, g_{(-N+1,-N)}, ..., g_{(N,N)}]^T$. The matrix $\mathbf{A} \in \mathbb{C}^{L \times L^2}$ is implicitly defined for a vector $\tilde{\mathbf{f}}$ with entries $[\tilde{\mathbf{f}}]_{(p,k)}$ by

$$[\mathbf{A}\tilde{\mathbf{f}}]_p := \sum_{k,\ell=-N}^{N} a_\ell e^{-i2\pi \frac{(p-\ell)k}{L}} [\tilde{\mathbf{f}}]_{(p,k)}, \quad p = -N, ..., N. \tag{31}$$

With this notation, the kernel $G(\mathbf{r})$ is defined by

$$G(\mathbf{r}) := \frac{1}{M^2} \mathbf{f}^H(\mathbf{r}) \mathbf{A}^H \mathbf{A} \mathbf{g} \tag{32}$$

where $M := N/2 + 1$.

First note that, by construction, $G(\mathbf{r})$ (and its partial derivatives $G^{(1,0)}(\mathbf{r})$ and $G^{(0,1)}(\mathbf{r})$) satisfy (28), as desired, since the entries of the vector $\mathbf{f}^H(\mathbf{r})\mathbf{A}^H$ are the terms in the bracket in (28). Due to $\mathbb{E}[\mathbf{A}^H \mathbf{A}] = \mathbf{I}$ (shown later, cf. (36); expectation is with respect to the a_ℓ), we have

$$\bar{G}(\mathbf{r}) := \mathbb{E}[G(\mathbf{r})] = \frac{1}{M^2} \mathbf{f}^H(\mathbf{r}) \mathbf{g} = K(\tau) K(\nu)$$

where $K(t)$ is the squared Fejér kernel, defined as

$$K(t) := \left(\frac{\sin(M\pi t)}{M \sin(\pi t)} \right)^4 = \frac{1}{M} \sum_{k=-N}^{N} g(k) e^{-i2\pi t k}. \tag{33}$$

Note that $K(t)$ is a trigonometric polynomial of degree N with coefficients $g(k)$. $K(t)$ decays rapidly around the origin $t = 0$ ($K(0) = 1$), thus $\bar{G}(\mathbf{r})$ also decays rapidly around the origin $\mathbf{r} = \mathbf{0}$. We will show later that $G(\mathbf{r})$ concentrates tightly around $\bar{G}(\mathbf{r})$, which implies that $\bar{G}(\mathbf{r})$ decays rapidly around the origin as well. In Figure 4 we plot $G(\mathbf{r})$ and $\bar{G}(\mathbf{r})/\bar{G}(\mathbf{0})$ for $N = 60$ and $N = 300$. Note that close to $\mathbf{r} = \mathbf{0}$, the random Kernel $\bar{G}(\mathbf{r})$ and the deterministic Kernel $\bar{G}(\mathbf{r})$ are very close.

We will establish that (30) holds and that $Q(\mathbf{r}) < 1$ for all $\mathbf{r} \notin \mathcal{T}$ by showing that $Q(\mathbf{r})$ is close to the (deterministic) polynomial

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

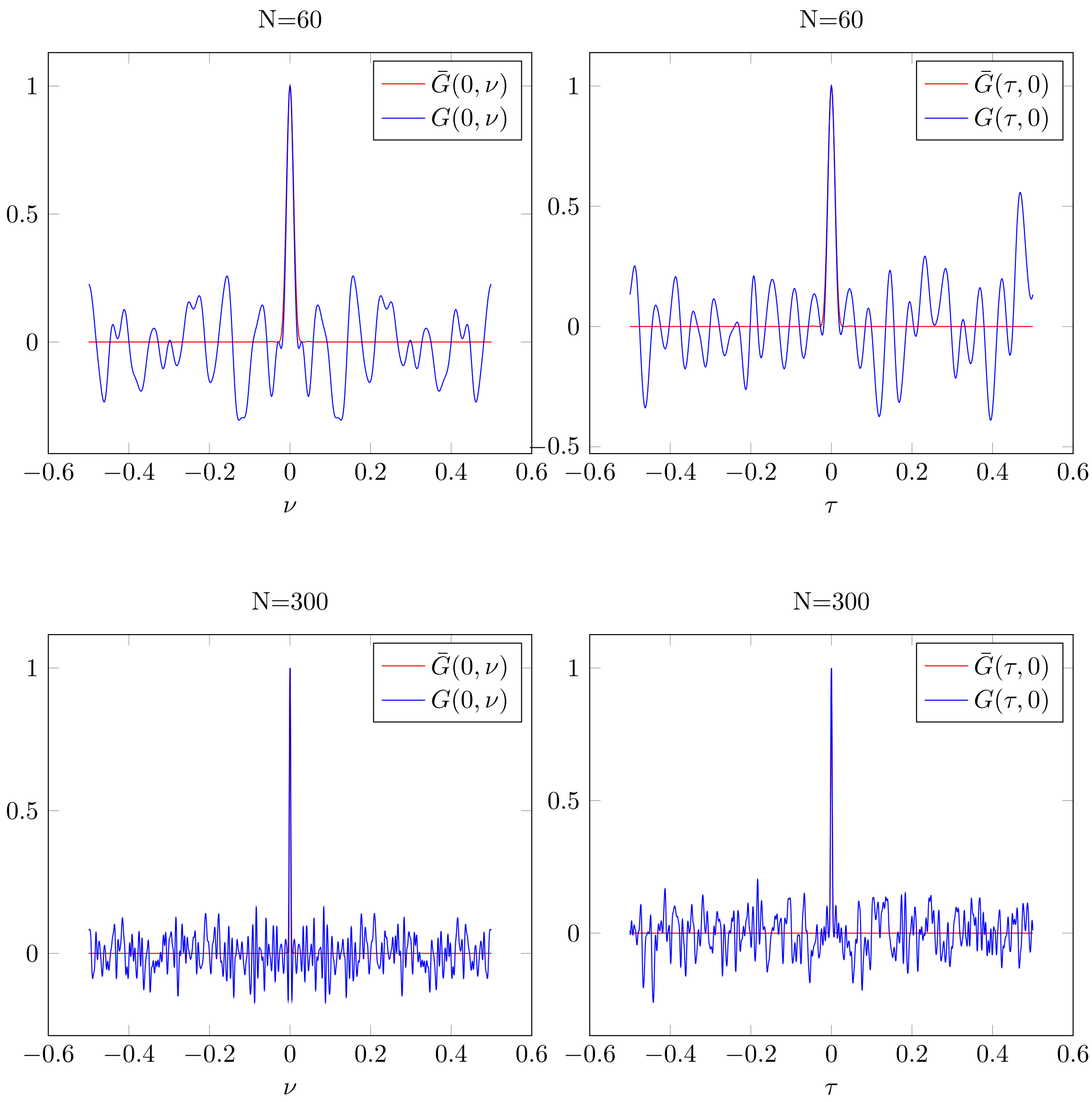

Figure 4: Plots of the random kernel $G(\mathbf{r})$ along with the deterministic kernel $\bar{G}(\mathbf{r})$.

where $\bar{G}^{(m,n)}(\mathbf{r}) := \frac{\partial^m}{\partial \tau^m}\frac{\partial^n}{\partial \nu^n}\bar{G}(\mathbf{r})$, and the coefficients $\bar{\alpha}_k, \bar{\beta}_{1k}$ and $\bar{\beta}_{2k}$ are chosen in a certain way that guarantees that

$$\bar{Q}(\mathbf{r}_j) = u_j, \quad \bar{Q}^{(1,0)}(\mathbf{r}_j) = 0, \text{ and } \bar{Q}^{(0,1)}(\mathbf{r}_j) = 0, \text{ for all } \mathbf{r}_j \in \mathcal{T} \tag{34}$$

and additionally $\bar{Q}(\mathbf{r}) < 1$ for all $\mathbf{r} \in [0,1]^2 \setminus \mathcal{T}$. That such a construction is indeed possible (for any choice of $u_j \in \{-1, 1\}$) is proven in [8] in the context of super-resolution in two dimensions.

The remainder of the proof is organized as follows.

1. We will start by showing that, with high probability, for a fixed $\mathbf{r}$, the kernels $G(\mathbf{r})$ and $\bar{G}(\mathbf{r})$ (and its partial derivatives) are close, i.e., $|G^{(m,n)}(\mathbf{r}) - \bar{G}^{(m,n)}(\mathbf{r})|$ is small.
2. Next, we will show that (with high probability) there exists a choice of coefficients $\alpha_k, \beta_{1k}, \beta_{2k}$ such that (30) is satisfied.
3. Finally we establish that $Q(\mathbf{r}) < 1$ for all $\mathbf{r} \notin \mathcal{T}$ (with the coefficients chosen as in Step 2). This is accomplished using an ϵ-net argument:
   (a) Let $\Omega \subset [0,1]^2$ be a (finite) set of grid points. We will show that $Q(\mathbf{r})$ is close to $\bar{Q}(\mathbf{r})$ for all $\mathbf{r} \in \Omega$ (with high probability).
   (b) Next, we extend this result to hold for all $\mathbf{r} \in [0,1]^2$ using Bernstein's polynomial inequality.
   (c) Combining this result with a result in [8] showing that $\bar{Q}(\mathbf{r}) < 1$ for all $\mathbf{r} \notin \mathcal{T}$, we can conclude that $\bar{Q}(\mathbf{r}) < 1$ for all $\mathbf{r} \notin \mathcal{T}$ (with high probability).

8.1 Step 1: Concentration of $G^{(m,n)}(\mathbf{r})$ around $\bar{G}^{(m,n)}(\mathbf{r})$

In this subsection we establish the following result.

Lemma 1. *For a fixed* $\mathbf{r}$*, for all* $\alpha \geq 0$*, and for all* (m,n) *with* $m + n \leq 4$*,*

$$\mathbb{P}\left[\frac{1}{\kappa^{m+n}}|G^{(m,n)}(\mathbf{r}) - \bar{G}^{(m,n)}(\mathbf{r})| > c_1 12^{\frac{m+n}{2}} \frac{\alpha}{\sqrt{L}}\right] \leq 2\exp\left(-c\min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right)\right) \tag{35}$$

where $\kappa := \sqrt{|K''(0)|}$ *and* c, c_1, c_2 *are numerical constants.*

We first show that $\mathbb{E}[G^{(m,n)}(\mathbf{r})] = \bar{G}^{(m,n)}(\mathbf{r})$. Lemma 1 is proven by expressing $G^{(m,n)}(\mathbf{r})$ as a quadratic form in $\mathbf{a} := [a_{-N}, ..., a_N]^T$, and showing that it does not deviate much from its expectation $\bar{G}^{(m,n)}(\mathbf{r})$ using the Hanson-Wright inequality below.

Theorem 3 (Hanson-Wright inequality [24, Thm. 1.1])**.** *Let* $\mathbf{a} \in \mathbb{R}^L$ *be a random vector with independent zero-mean* K*-sub-Gaussian entries (i.e., the entries obey* $\sup_{p\geq 1} p^{-1}(\mathbb{E}[|a_\ell|^p])^{1/p} \leq K$*), and let* $\mathbf{V}$ *be an* $L \times L$ *matrix. Then, for all* $t \geq 0$*,*

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

where c *is a numerical constant.*

We first show that $\mathbb{E}[\mathbf{A}^H\mathbf{A}] = \mathbf{I}$, which implies

$$\mathbb{E}[G^{(m,n)}(\mathbf{r})] = \bar{G}^{(m,n)}(\mathbf{r}). \tag{36}$$

To this end, note that the partial derivatives of $G(\mathbf{r})$ are

$$G^{(m,n)}(\mathbf{r}) = \frac{1}{M^2} \sum_{k,p=-N}^{N} (-i2\pi k)^m (-i2\pi p)^n e^{-i2\pi(k\tau+p\nu)} [\mathbf{A}^H\mathbf{A}\mathbf{g}]_{(k,p)} \tag{37}$$

and the partial derivatives of $\bar{G}(\mathbf{r})$ are

$$\bar{G}^{(m,n)}(\mathbf{r}) = \frac{1}{M^2} \sum_{k,p=-N}^{N} (-i2\pi k)^m (-i2\pi p)^n e^{-i2\pi(k\tau+p\nu)} g(k)g(p).$$

The matrix $\mathbf{A}^H\mathbf{A} \in \mathbb{C}^{L^2 \times L^2}$ is block diagonal, with the pth block on its diagonal, $\mathbf{B}_p \in \mathbb{C}^{L\times L}$, given by

$$[\mathbf{B}_p]_{k,\tilde{k}} = \sum_{\ell,\tilde{\ell}=-N}^{N} e^{-i2\pi\frac{(p-\ell)k}{L}} e^{i2\pi\frac{(p-\tilde{\ell})\tilde{k}}{L}} a_\ell a_{\tilde{\ell}}^*, \quad k,\tilde{k} = -N,...,N.$$

Using that $\mathbb{E}[a_\ell a_{\tilde{\ell}}^*] = 1/L$ for $\ell = \tilde{\ell}$ and $\mathbb{E}[a_\ell a_{\tilde{\ell}}^*] = 0$ for $\ell \neq \tilde{\ell}$ we obtain that $\mathbb{E}[\mathbf{B}_p] = \mathbf{I}$ and therefore $\mathbb{E}[\mathbf{A}^H\mathbf{A}] = \mathbf{I}$. Plugging this into (37) yields (36).

We next express $G^{(m,n)}(\mathbf{r})$ as a quadratic form in $\mathbf{a}$. To this end, first note that by (31), $\mathbf{A}\mathbf{g} = \mathbf{A}_\mathbf{g}\mathbf{a}$, where $\mathbf{A}_\mathbf{g} \in \mathbb{C}^{L\times L}$ is defined as

$$[\mathbf{A}_\mathbf{g}]_{p,\ell} := \sum_{k=-N}^{N} e^{-i2\pi\frac{(p-\ell)k}{L}} g(p)g(k).$$

Next, define $\mathbf{f}^{(m,n)}(\mathbf{r}) := \frac{\partial^m}{\partial\tau^m}\frac{\partial^n}{\partial\nu^n}\mathbf{f}(\mathbf{r})$, $\mathbf{r} = [\tau,\nu]^T$. We have

$$[\mathbf{A}\mathbf{f}^{(m,n)}]_p = \sum_{k,\ell=-N}^{N} e^{-i2\pi\frac{(p-\ell)k}{L}} (i2\pi k)^m (i2\pi p)^n e^{i2\pi(k\tau+p\nu)} a_\ell$$

thus $(\mathbf{f}^{(m,n)})^H \mathbf{A}^H = \mathbf{a}^H \mathbf{A}_\mathbf{f}^H$ where $\mathbf{A}_\mathbf{f}^H \in \mathbb{C}^{L\times L}$ is defined by

$$[\mathbf{A}_\mathbf{f}^H]_{\tilde{\ell},p} = \sum_{\tilde{k}=-N}^{N} (-i2\pi\tilde{k})^m (-i2\pi p)^n e^{i2\pi\frac{(p-\tilde{\ell})\tilde{k}}{L}} e^{-i2\pi(\tilde{k}\tau+p\nu)}.$$

We therefore obtain the desired representation as a quadratic form

$$G^{(m,n)}(\mathbf{r}) = \frac{1}{M^2}(\mathbf{f}^{(m,n)}(\mathbf{r}))^H \mathbf{A}^H\mathbf{A}\mathbf{g} = \frac{1}{M^2}\mathbf{a}^H\mathbf{A}_\mathbf{f}^H\mathbf{A}_\mathbf{g}\mathbf{a} = \mathbf{a}^H\mathbf{V}^{(m,n)}(\mathbf{r})\mathbf{a} \tag{38}$$

where $\mathbf{V}^{(m,n)}(\mathbf{r}) := \frac{1}{M^2}\mathbf{A}_\mathbf{f}^H\mathbf{A}_\mathbf{g} \in \mathbb{C}^{L\times L}$ has coefficients

$$[\mathbf{V}^{(m,n)}(\mathbf{r})]_{\tilde{\ell},\ell} = \frac{1}{M^2} \sum_{p,k,\tilde{k}=-N}^{N} (-i2\pi\tilde{k})^m (-i2\pi p)^n e^{-i2\pi\frac{(p-\ell)k}{L}} e^{i2\pi\frac{(p-\tilde{\ell})\tilde{k}}{L}} e^{-i2\pi(\tilde{k}\tau+p\nu)} g(p)g(k).$$

In order to evaluate the right hand side (RHS) of the Hanson-Wright inequality, we will need the following upper bound on $\left\|\mathbf{V}^{(m,n)}(\mathbf{r})\right\|_F$.

Lemma 2. *For all* $\mathbf{r}$*, and for all* (m,n) *with* $m+n \leq 4$*,*

$$\left\|\mathbf{V}^{(m,n)}(\mathbf{r})\right\|_F \leq c_1(2\pi N)^{n+m}\sqrt{L}.$$

Proof. We start by upper-bounding $|[\mathbf{V}^{(m,n)}(\mathbf{r})]_{\tilde{\ell},\ell}|$. By definition of $K(t)$ (cf. (33))

$$\begin{aligned}
[\mathbf{V}^{(m,n)}(\mathbf{r})]_{\tilde{\ell},\ell} &= \frac{1}{M}\sum_{p,\tilde{k}=-N}^{N}\left(\frac{1}{M}\sum_{k=-N}^{N} g(k)e^{-i2\pi\frac{(p-\ell)k}{L}}\right)(-i2\pi\tilde{k})^m(-i2\pi p)^n e^{i2\pi\frac{(p-\tilde{\ell})\tilde{k}}{L}} e^{-i2\pi(\tilde{k}\tau+p\nu)}g(p) \\
&= \sum_{p=-N}^{N} K\left(\frac{p-\ell}{L}\right)(-i2\pi p)^n e^{-i2\pi p\nu} g(p)\frac{1}{M}\sum_{\tilde{k}=-N}^{N}(-i2\pi\tilde{k})^m e^{i2\pi\left(\frac{p-\tilde{\ell}}{L}-\tau\right)\tilde{k}}.
\end{aligned}$$

Using that $|g(p)| \leq 1$, for all p, we obtain

$$\begin{aligned}
|[\mathbf{V}^{(m,n)}(\mathbf{r})]_{\tilde{\ell},\ell}| &\leq (2\pi N)^n \sum_{p=-N}^{N}\left|K\left(\frac{p-\ell}{L}\right)\right|\left|\frac{1}{M}\sum_{\tilde{k}=-N}^{N}(-i2\pi\tilde{k})^m e^{i2\pi\left(\frac{p-\tilde{\ell}}{L}-\tau\right)\tilde{k}}\right| \\
&= (2\pi N)^n \sum_{p=-N}^{N}\left|K\left(\frac{p}{L}\right)\right|\left|\frac{1}{M}\sum_{\tilde{k}=-N}^{N}(-i2\pi\tilde{k})^m e^{i2\pi\left(\frac{p+\ell-\tilde{\ell}}{L}-\tau\right)\tilde{k}}\right|
\end{aligned}$$

where we used that the absolute values in the sum above are L-periodic in p (recall that $K(t)$ is 1-periodic). Using that

$$|K(t)| = \left|\frac{\sin(\pi M t)}{M\sin(\pi t)}\right|^4 \leq \min\left(1, \frac{1}{(2Mt)^4}\right)$$

for $t \in [-1/2, 1/2]$ (from $|\sin(\pi t)| \geq 2|t|$) we have

$$\left|K\left(\frac{p}{L}\right)\right| \leq \min\left(1, \frac{1}{(2Mp/L)^4}\right) \leq \min\left(1, \frac{16}{p^4}\right) \leq 16\min\left(1, \frac{1}{p^4}\right)$$

where we used that $\frac{L}{2M} = \frac{2N+1}{N+2} \leq 2$. We thus obtain

$$|[\mathbf{V}^{(m,n)}(\mathbf{r})]_{\tilde{\ell},\ell}| \leq (2\pi N)^{m+n} \underbrace{16(2\pi N)^{-m}\sum_{p=-N}^{N}\min\left(1,\frac{1}{p^4}\right)\left|\frac{1}{M}\sum_{\tilde{k}=-N}^{N}(-i2\pi\tilde{k})^m e^{i2\pi\left(\frac{p+\ell-\tilde{\ell}}{L}-\tau\right)\tilde{k}}\right|}_{U\left(\tau-\frac{\ell-\tilde{\ell}}{L}\right):=}$$

where $U(t)$ is 1-periodic and satisfies $U(t) \leq c\min(1, \frac{1}{L|t|})$ for $|t| \leq 1/2$ as shown in Appendix D. Thus

$$|[\mathbf{V}^{(m,n)}(\mathbf{r})]_{\tilde{\ell},\ell}| \leq (2\pi N)^{m+n} U\left(\tau - \frac{\ell-\tilde{\ell}}{L}\right)$$

which yields

$$\left\|\mathbf{V}^{(m,n)}(\mathbf{r})\right\|_F^2 = \sum_{\ell,\tilde{\ell}=-N}^{N}\left|[\mathbf{V}]_{\tilde{\ell},\ell}\right|^2 \leq (2\pi N)^{2(n+m)}\sum_{\ell,\tilde{\ell}=-N}^{N} U^2\left(\tau - \frac{\ell-\tilde{\ell}}{L}\right). \tag{39}$$

Using that $U(t)$ is 1-periodic, and upper-bounded by $c\min(1, \frac{1}{L|t|})$ we obtain

$$\sum_{\ell=-N}^{N} U^2\left(\tau - \frac{\ell-\tilde{\ell}}{L}\right) \leq \sum_{\ell=-N}^{N}\left(c\min\left(1, \frac{1}{L|\ell/L|}\right)\right)^2 \leq c^2\left(1+2\sum_{\ell\geq 1}\frac{1}{\ell^2}\right) = \underbrace{c^2\left(1+\frac{\pi^2}{3}\right)}_{c_2}. \quad (40)$$

Substituting (40) into (39) yields

$$\left\|\mathbf{V}^{(m,n)}(\mathbf{r})\right\|_F^2 \leq (2\pi N)^{2(n+m)} \sum_{\tilde{\ell}=N}^{N} c_2 \leq (2\pi N)^{2(n+m)} c_2 L \quad (41)$$

where $c_1 = \sqrt{c_2}$. This concludes the proof. □

We are now ready to establish Lemma 1 by application of the Hanson-Wright inequality. With $K''(0) = -\frac{\pi^2}{3}(N^2+4N)$ [8, Eq. 2.23] we have (recall that $\kappa = \sqrt{|K''(0)|}$) that

$$\frac{(2\pi N)^{n+m}}{\kappa^{m+n}} = \frac{(2\pi N)^{n+m}}{(\frac{\pi^2}{3}(N^2+4N))^{(m+n)/2}} \leq 12^{\frac{m+n}{2}}$$

which yields

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{\kappa^{m+n}}|G^{(m,n)}(\mathbf{r}) - \bar{G}^{(m,n)}(\mathbf{r})| > c_1 12^{\frac{m+n}{2}}\frac{\alpha}{\sqrt{L}}\right] &\leq \mathbb{P}\left[|G^{(m,n)}(\mathbf{r}) - \bar{G}^{(m,n)}(\mathbf{r})| > c_1(2\pi N)^{n+m}\frac{\alpha}{\sqrt{L}}\right] \\
&\leq \mathbb{P}\left[|\mathbf{a}^T\mathbf{V}\mathbf{a} - \mathbb{E}[\mathbf{a}^T\mathbf{V}\mathbf{a}]| > \|\mathbf{V}\|_F\frac{\alpha}{L}\right] \quad (42) \\
&\leq 2\exp\left(-c\min\left(\frac{\|\mathbf{V}\|_F^2\alpha^2}{L^2K^4\|\mathbf{V}\|_F^2}, \frac{\|\mathbf{V}\|_F\alpha}{LK^2\|\mathbf{V}\|}\right)\right) \quad (43) \\
&\leq 2\exp\left(-c\min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right)\right) \quad (44)
\end{aligned}$$

where (42) follows from $\|\mathbf{V}\|_F \leq c_1(2\pi N)^{n+m}\sqrt{L}$ (cf. Lemma 2), and from $G^{(m,n)}(\mathbf{r}) = \mathbf{a}^H\mathbf{V}\mathbf{a}$ and $\mathbb{E}[\mathbf{a}^H\mathbf{V}\mathbf{a}] = \mathbb{E}[G^{(m,n)}(\mathbf{r})] = \bar{G}^{(m,n)}(\mathbf{r})$ (cf. (36)). To obtain (43), we used Theorem 3 with $t = \|\mathbf{V}\|_F\frac{\alpha}{L}$, and (44) follows because the sub-Gaussian parameter K of the random variable $[\mathbf{a}]_\ell \sim \mathcal{N}(0, 1/L)$ is given by $K = c_2/\sqrt{L}$ (e.g., [29, Ex. 5.8]) and $\|\mathbf{V}\|_F/\|\mathbf{V}\| \geq 1$.

8.2 Step 2: Choice of the coefficients $\alpha, \beta_{1k}, \beta_{2k}$

We next show that, with high probability, there exists a set of coefficients $\alpha, \beta_{1k}, \beta_{2k}$ such that $Q(\mathbf{r})$ satisfies (30). To this end, we first review the result in [8] with ensures that there exists a set of coefficients $\bar{\alpha}, \bar{\beta}_{1k}, \bar{\beta}_{2k}$ such that (34) is satisfied. Writing (34) in matrix form yields

$$\underbrace{\begin{bmatrix} \bar{\mathbf{D}}_{0,0} & \frac{1}{\kappa}\bar{\mathbf{D}}_{1,0} & \frac{1}{\kappa}\bar{\mathbf{D}}_{0,1} \\ -\frac{1}{\kappa}\bar{\mathbf{D}}_{1,0} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}_{2,0} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}_{1,1} \\ -\frac{1}{\kappa}\bar{\mathbf{D}}_{0,1} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}_{1,1} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}_{0,2} \end{bmatrix}}_{\bar{\mathbf{D}}} \begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \kappa\bar{\boldsymbol{\beta}}_1 \\ \kappa\bar{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad (45)$$

where $[\bar{\mathbf{D}}_{m,n}]_{j,k} \coloneqq \bar{G}^{(m,n)}(\mathbf{r}_j - \mathbf{r}_k)$, $[\bar{\boldsymbol{\alpha}}]_k \coloneqq \bar{\alpha}_k$, $[\bar{\boldsymbol{\beta}}_1]_k \coloneqq \bar{\beta}_{1k}$ and $[\bar{\boldsymbol{\beta}}_2]_k \coloneqq \bar{\beta}_{2k}$. Here we have scaled the entries of $\bar{\mathbf{D}}$ such that its diagonal entries are 1 ($K(0) = 1$, $\kappa^2 = |K''(0)|$, and $K''(0)$ is negative). Since $\bar{\mathbf{D}}_{0,0}, \bar{\mathbf{D}}_{1,1}, \bar{\mathbf{D}}_{2,0}, \bar{\mathbf{D}}_{0,2}$ are symmetric while $\bar{\mathbf{D}}_{1,0}, \bar{\mathbf{D}}_{0,1}$ are antisymmetric, $\bar{\mathbf{D}}$ is symmetric.

The following result, proven in [8, Eq. C6, C7, C8, C9], ensures that $\bar{\mathbf{D}}$ is invertible and thus the coefficients $\bar{\alpha}, \bar{\beta}_{1k}, \bar{\beta}_{2k}$ can be obtained according to

$$\begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \kappa \bar{\boldsymbol{\beta}}_1 \\ \kappa \bar{\boldsymbol{\beta}}_2 \end{bmatrix} = \bar{\mathbf{D}}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix} = \bar{\mathbf{L}} \mathbf{u} \tag{46}$$

where $\bar{\mathbf{L}}$ is the $3S \times S$ submatrix of $\bar{\mathbf{D}}^{-1}$ corresponding to the first S columns of $\bar{\mathbf{D}}^{-1}$.

Proposition 4 ([8, Eq. C6, C7, C8, C9])**.** $\bar{\mathbf{D}}$ *is invertible and*

$$\left\| \mathbf{I} - \bar{\mathbf{D}} \right\| \leq 0.19808 \tag{47}$$

$$\left\| \bar{\mathbf{D}} \right\| \leq 1.19808 \tag{48}$$

$$\left\| \bar{\mathbf{D}}^{-1} \right\| \leq 1.24700. \tag{49}$$

Proof. Since $\bar{\mathbf{D}}$ is real and symmetric, it is normal, and thus its singular values are equal to the absolute values of its eigenvalues. Using that the diagonal entries of $\bar{\mathbf{D}}$ are 1, by Gershgorin's circle theorem [17, Thm. 6.1.1], the eigenvalues of $\bar{\mathbf{D}}$ are in the interval $[1 - \left\| \mathbf{I} - \bar{\mathbf{D}} \right\|_\infty, 1 + \left\| \mathbf{I} - \bar{\mathbf{D}} \right\|_\infty]$, where $\|\mathbf{A}\|_\infty := \max_i \sum_j |[\mathbf{A}]_{i,j}|$. Using that $\left\| \mathbf{I} - \bar{\mathbf{D}} \right\|_\infty \leq 0.19808$ (shown below), it follows that $\bar{\mathbf{D}}$ is invertible and

$$\left\| \bar{\mathbf{D}} \right\| \leq 1 + \left\| \mathbf{I} - \bar{\mathbf{D}} \right\|_\infty \leq 1.19808$$

$$\left\| \bar{\mathbf{D}}^{-1} \right\| \leq \frac{1}{1 - \left\| \mathbf{I} - \bar{\mathbf{D}} \right\|_\infty} \leq 1.2470.$$

The proof is concluded by noting that

$$\begin{aligned} \left\| \mathbf{I} - \bar{\mathbf{D}} \right\|_\infty &= \max \left\{ \|\mathbf{I} - \mathbf{D}_{0,0}\|_\infty + 2 \left\| \frac{1}{\kappa} \bar{\mathbf{D}}_{1,0} \right\|_\infty, \left\| \frac{1}{\kappa} \bar{\mathbf{D}}_{1,0} \right\|_\infty + \left\| \mathbf{I} - \frac{1}{\kappa^2} \mathbf{D}_{2,0} \right\|_\infty + \left\| \frac{1}{\kappa^2} \bar{\mathbf{D}}_{1,1} \right\|_\infty \right\} \\ &\leq 0.19808 \end{aligned}$$

where we used [8, Eq. C6, C7, C8, C9]:

$$\begin{aligned} \|\mathbf{I} - \mathbf{D}_{0,0}\|_\infty &\leq 0.04854 \\ \left\| \frac{1}{\kappa} \bar{\mathbf{D}}_{1,0} \right\|_\infty = \left\| \frac{1}{\kappa} \bar{\mathbf{D}}_{0,1} \right\|_\infty &\leq 0.04258 \\ \left\| \frac{1}{\kappa^2} \bar{\mathbf{D}}_{1,1} \right\|_\infty &\leq 0.04791 \\ \left\| \mathbf{I} - \frac{1}{\kappa^2} \mathbf{D}_{0,2} \right\|_\infty = \left\| \mathbf{I} - \frac{1}{\kappa^2} \mathbf{D}_{2,0} \right\|_\infty &\leq 0.1076. \end{aligned}$$

□

We next show that with high probability there exists a set of coefficients $\alpha, \beta_{1k}, \beta_{2k}$ which satisfies (30). To this end, we write (30) in matrix form:

$$\underbrace{\begin{bmatrix} \mathbf{D}_{0,0} & \frac{1}{\kappa}\mathbf{D}_{1,0} & \frac{1}{\kappa}\mathbf{D}_{0,1} \\ -\frac{1}{\kappa}\mathbf{D}_{1,0} & -\frac{1}{\kappa^2}\mathbf{D}_{2,0} & -\frac{1}{\kappa^2}\mathbf{D}_{1,1} \\ -\frac{1}{\kappa}\mathbf{D}_{0,1} & -\frac{1}{\kappa^2}\mathbf{D}_{1,1} & -\frac{1}{\kappa^2}\mathbf{D}_{0,2} \end{bmatrix}}_{\mathbf{D}} \begin{bmatrix} \boldsymbol{\alpha} \\ \kappa \boldsymbol{\beta}_1 \\ \kappa \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{50}$$

where $[\mathbf{D}_{0,0}]_{j,k} := G^{(m,n)}(\mathbf{r}_j - \mathbf{r}_k)$, $[\boldsymbol{\alpha}]_k := \alpha_k$, $[\boldsymbol{\beta}_1]_k := \beta_{1k}$, and $[\boldsymbol{\beta}_2]_k := \beta_{2k}$ and show that $\mathbf{D}$ can be inverted with high probability. Specifically, we show that the probability of the event

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

is high, and on $\mathcal{E}_\tau$, with $\tau \in [0, 1/4]$, $\mathbf{D}$ is invertible. The fact that $\mathbf{D}$ is invertible on $\mathcal{E}_\tau$, with $\tau \in [0, 1/4]$ follows from the following set of inequalities:

$$\|\mathbf{I} - \mathbf{D}\| \leq \|\mathbf{D} - \bar{\mathbf{D}}\| + \|\bar{\mathbf{D}} - \mathbf{I}\| \leq \tau + 0.1908 \leq 0.4408.$$

Since $\mathbf{D}$ is invertible, the coefficients $\alpha, \beta_{1k}, \beta_{2k}$ can be obtained as

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \kappa\boldsymbol{\beta}_1 \\ \kappa\boldsymbol{\beta}_2 \end{bmatrix} = \mathbf{D}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix} = \mathbf{L}\mathbf{u} \tag{51}$$

where $\mathbf{L}$ is the $3S \times S$ submatrix of $\mathbf{D}^{-1}$ corresponding to the first S columns of $\mathbf{D}^{-1}$. On $\mathcal{E}_\tau$ with $\tau \in [0, 1/4]$, the norm of $\mathbf{L}$ is bounded as well:

$$\|\mathbf{L}\| \leq \|\mathbf{D}^{-1}\| \leq 2\|\bar{\mathbf{D}}^{-1}\| \leq 2.5 \tag{52}$$

where the first inequality follows since $\mathbf{L}$ is a submatrix of $\mathbf{D}^{-1}$, and the second inequality follows from the first part of the lemma below applied to $\mathbf{B} = \mathbf{D}$ and $\mathbf{C} = \bar{\mathbf{D}}$ (by $\|\mathbf{D} - \bar{\mathbf{D}}\| \leq 1/4$ and $\|\bar{\mathbf{D}}^{-1}\| \leq 1.247$, cf. (49), the conditions of the corollary are satisfied). The third inequality again follows (49).

Lemma 3 ([27, Proof of Cor. 4.5])**.** *Suppose that $\mathbf{C}$ is invertible and $\|\mathbf{B} - \mathbf{C}\|\|\mathbf{C}^{-1}\| \leq 1/2$. Then i) $\|\mathbf{B}^{-1}\| \leq 2\|\mathbf{C}^{-1}\|$ and ii) $\|\mathbf{B}^{-1} - \mathbf{C}^{-1}\| \leq 2\|\mathbf{C}^{-1}\|^2\|\mathbf{B} - \mathbf{C}\|$.*

By the second part of the lemma above and again using (49), we obtain that, on $\mathcal{E}_\tau$ with $\tau \in [0, 1/4]$,

$$\|\mathbf{L} - \bar{\mathbf{L}}\| \leq \|\mathbf{D}^{-1} - \bar{\mathbf{D}}^{-1}\| \leq 2\|\bar{\mathbf{D}}^{-1}\|^2\|\mathbf{D} - \bar{\mathbf{D}}\| \leq 2.5\tau. \tag{53}$$

Lemma 4. *For all $\tau > 0$*

$$\mathbb{P}[\mathcal{E}_\tau] \geq 1 - \delta$$

provided that

$$L \geq S\frac{c_4}{\tau^2}\log^2(18S^2/\delta) \tag{54}$$

where c_4 is a numerical constant.

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

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

for m, n with $m+n \leq 2$ and for $j, k = 1, ..., S$. We now have

$$\mathbb{P}\left[\left\|\mathbf{D} - \bar{\mathbf{D}}\right\| \geq \tau\right] \leq \mathbb{P}\left[\sqrt{3s} \max_{j,k,m,n} \frac{1}{\kappa^{m+n}} |[\mathbf{D}_{m,n} - \bar{\mathbf{D}}_{m,n}]_{j,k}| \geq \tau\right] \tag{55}$$

$$\leq \sum_{j,k,m,n} \mathbb{P}\left[\frac{1}{\kappa^{m+n}} |[\mathbf{D}_{m,n} - \bar{\mathbf{D}}_{m,n}]_{j,k}| \geq \frac{\tau}{\sqrt{3s}}\right] \tag{56}$$

$$= \sum_{j,k,m,n} \mathbb{P}\left[\frac{1}{\kappa^{m+n}} |[\mathbf{D}_{m,n} - \bar{\mathbf{D}}_{m,n}]_{j,k}| \geq 12 c_1 \frac{\alpha}{\sqrt{L}}\right] \tag{57}$$

$$\leq \sum_{j,k,m,n} \mathbb{P}\left[\frac{1}{\kappa^{m+n}} |[\mathbf{D}_{m,n} - \bar{\mathbf{D}}_{m,n}]_{j,k}| \geq 12^{\frac{m+n}{2}} c_1 \frac{\alpha}{\sqrt{L}}\right] \tag{58}$$

$$\leq 2(3s)^2 \exp\left(-c \min\left(\frac{\tau^2 L}{c_2^4 c_3^2 3s}, \frac{\tau\sqrt{L}}{c_2^2 c_3 \sqrt{3s}}\right)\right). \tag{59}$$

Here, (55) follows from upper bounding $\left\|\mathbf{D} - \bar{\mathbf{D}}\right\|$ by $\sqrt{3s}$ times the maximum absolute value of $\mathbf{D} - \bar{\mathbf{D}}$. Furthermore, (56) follows from the union bound, (57) follows by setting $\alpha = \frac{\tau\sqrt{L}}{\sqrt{3s}12c_1}$, where c_1 is the constant in Lemma 1, and (58) follows from $12^{\frac{m+n}{2}} \leq 12$, for $m+n \leq 2$. Finally, (59) follows from Lemma 1 (here we set $c_3 := 12c_1$).

The RHS of (59) is smaller than δ, as desired, if

$$\log(18S^2/\delta) \leq c \min\left(\frac{\tau^2 L}{c_2^4 c_3^2 3S}, \frac{\tau\sqrt{L}}{c_2^2 c_3 \sqrt{3S}}\right)$$

which is implied by (54) with $c_4 = 3c_2^4 c_3^2 \max(1/c^2, 1/c)$. □

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

The goal of this section is to prove Lemma 5 below which shows that $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ (and their partial derivatives) are close on a set of (grid) points.

Lemma 5. *Let $\Omega \subset [0,1]^2$ be a finite set of points and pick any $\epsilon \leq 1$ and $\delta > 0$. Suppose that*

$$L \geq \frac{S}{\epsilon^2} \max\left(c_5 \log^2\left(\frac{12S|\Omega|}{\delta}\right) \log\left(\frac{8|\Omega|}{\delta}\right), c \log\left(\frac{4|\Omega|}{\delta}\right) \log\left(\frac{18S^2}{\delta}\right)\right).$$

Then

$$\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \frac{1}{\kappa^{n+m}} \left|Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r})\right| \leq \epsilon\right] \geq 1 - 4\delta.$$

In order to prove Lemma 5, first note that the (m,n)th partial derivative of $Q(\mathbf{r})$ (defined by (29)) is (after normalization with $1/\kappa^{m+n}$)

$$\begin{aligned}\frac{1}{\kappa^{m+n}} Q^{(m,n)}(\mathbf{r}) = \sum_{k=1}^{S} \Big(&\alpha_k \frac{1}{\kappa^{m+n}} G^{(m,n)}(\mathbf{r} - \mathbf{r}_k) \\ &+ \kappa\beta_{1k} \frac{1}{\kappa^{m+n+1}} G^{(m+1,n)}(\mathbf{r} - \mathbf{r}_k) + \kappa\beta_{2k} \frac{1}{\kappa^{m+n+1}} G^{(m,n+1)}(\mathbf{r} - \mathbf{r}_k)\Big) \\ &= \left(\mathbf{v}^{(m,n)}(\mathbf{r})\right)^H \mathbf{Lu}\end{aligned} \tag{60}$$

where we used (51) and defined

$$\left(\mathbf{v}^{(m,n)}\right)^H(\mathbf{r}) := \frac{1}{\kappa^{m+n}}\Big[G^{(m,n)}(\mathbf{r}-\mathbf{r}_1)\cdots G^{(m,n)}(\mathbf{r}-\mathbf{r}_S)\ \frac{1}{\kappa}G^{(m+1,n)}(\mathbf{r}-\mathbf{r}_1)\cdots\frac{1}{\kappa}G^{(m+1,n)}(\mathbf{r}-\mathbf{r}_S)$$
$$\frac{1}{\kappa}G^{(m,n+1)}(\mathbf{r}-\mathbf{r}_1)\cdots\frac{1}{\kappa}G^{(m+1,n+1)}(\mathbf{r}-\mathbf{r}_S)\Big].$$

Since $\mathbb{E}[G^{(m,n)}(\mathbf{r})] = \bar{G}^{(m,n)}(\mathbf{r})$ (cf. (36)), we have

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

where

$$\bar{\mathbf{v}}^H_{(m,n)}(\mathbf{r})\frac{1}{\kappa^{m+n}} = \Big[\bar{G}^{(m,n)}(\mathbf{r}-\mathbf{r}_1)\cdots\bar{G}^{(m,n)}(\mathbf{r}-\mathbf{r}_S)\ \frac{1}{\kappa}\bar{G}^{(m+1,n)}(\mathbf{r}-\mathbf{r}_1)\cdots\frac{1}{\kappa}\bar{G}^{(m+1,n)}(\mathbf{r}-\mathbf{r}_S)$$
$$\frac{1}{\kappa}\bar{G}^{(m,n+1)}(\mathbf{r}-\mathbf{r}_1)\cdots\frac{1}{\kappa}\bar{G}^{(m+1,n+1)}(\mathbf{r}-\mathbf{r}_S)\Big].$$

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

$$\begin{aligned}\frac{1}{\kappa^{m+n}}Q^{(m,n)}(\mathbf{r}) &= \left\langle \mathbf{u}, \mathbf{L}^H\mathbf{v}_{(m,n)}(\mathbf{r})\right\rangle \\ &= \left\langle \mathbf{u}, \bar{\mathbf{L}}^H\bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\rangle + \underbrace{\left\langle \mathbf{u}, \mathbf{L}^H(\mathbf{v}_{(m,n)}(\mathbf{r})-\bar{\mathbf{v}}_{(m,n)}(\mathbf{r}))\right\rangle}_{I_1^{(m,n)}(\mathbf{r})} + \underbrace{\left\langle \mathbf{u}, (\mathbf{L}-\bar{\mathbf{L}})^H\bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\rangle}_{I_2^{(m,n)}(\mathbf{r})} \\ &= \frac{1}{\kappa^{m+n}}\bar{Q}^{(m,n)}(\mathbf{r}) + I_1^{(m,n)}(\mathbf{r}) + I_2^{(m,n)}(\mathbf{r}) \end{aligned} \tag{61}$$

where $\bar{\mathbf{L}}$ was defined below (46). The following two results establish that the perturbations $I_1^{(m,n)}(\mathbf{r})$ and $I_2^{(m,n)}(\mathbf{r})$ are small on a set of (grid) points Ω with high probability.

Lemma 6. *Let $\Omega \subset [0,1]^2$ be a finite set of points, suppose that $m+n \leq 2$. We have, for any $\delta \geq 0$, that*

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

provided that

$$L \geq \frac{c_5}{\epsilon^2}S\log^2\left(\frac{12S|\Omega|}{\delta}\right)\log\left(\frac{8|\Omega|}{\delta}\right).$$

Proof. Set $\Delta\mathbf{v} := \mathbf{v}_{(m,n)}(\mathbf{r}) - \bar{\mathbf{v}}_{(m,n)}(\mathbf{r})$ for notational convenience. By the union bound, we have for all $a, b \geq 0$,

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_1^{(m,n)}(\mathbf{r})| \geq 2.5ab\right] &= \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |\langle \mathbf{u}, \mathbf{L}^H \Delta\mathbf{v}\rangle| \geq 2.5ab\right] \\
&\leq \mathbb{P}\left[\bigcup_{\mathbf{r}\in\Omega} \{|\langle \mathbf{u}, \mathbf{L}^H \Delta\mathbf{v}\rangle| \geq \|\mathbf{L}^H\Delta\mathbf{v}\|_2 b\} \cup \{\|\mathbf{L}^H\Delta\mathbf{v}\|_2 \geq 2.5a\}\right] \\
&\leq \mathbb{P}\left[\bigcup_{\mathbf{r}\in\Omega} \{|\langle \mathbf{u}, \mathbf{L}^H \Delta\mathbf{v}\rangle| \geq \|\mathbf{L}^H\Delta\mathbf{v}\|_2 b\} \cup \{\|\Delta\mathbf{v}\|_2 \geq a\} \cup \{\|\mathbf{L}\| \geq 2.5\}\right] \\
&\leq \mathbb{P}[\|\mathbf{L}\| \geq 2.5] + \sum_{\mathbf{r}\in\Omega} \left(\mathbb{P}[|\langle \mathbf{u}, \mathbf{L}^H \Delta\mathbf{v}\rangle| \geq \|\mathbf{L}^H\Delta\mathbf{v}\|_2 b] + \mathbb{P}[\|\Delta\mathbf{v}\|_2 \geq a]\right) \\
&\leq \mathbb{P}[\overline{\mathcal{E}}_{1/4}] + |\Omega| 4e^{-\frac{b^2}{4}} + \sum_{\mathbf{r}\in\Omega} \mathbb{P}[\|\Delta\mathbf{v}\|_2 \geq a] \qquad (62) \\
&\leq \mathbb{P}[\overline{\mathcal{E}}_{1/4}] + \frac{\delta}{2} + \sum_{\mathbf{r}\in\Omega} \mathbb{P}[\|\Delta\mathbf{v}\|_2 \geq a] \qquad (63)
\end{aligned}$$

where (62) follows from application of Hoeffding's inequality (stated below) and from $\{\|\mathbf{L}\| \geq 2.5\} \subseteq \overline{\mathcal{E}}_{1/4}$ according to (52). For (63), we used $|\Omega| 4e^{-\frac{b^2}{4}} \leq \frac{\delta}{2}$ ensured by choosing $b = 2\sqrt{\log(8|\Omega|/\delta)}$.

Lemma 7 (Hoeffding's inequality)**.** *Suppose the entries of* $\mathbf{u} \in \mathbb{R}^S$ *are i.i.d. with* $\mathbb{P}[u_i = -1] = \mathbb{P}[u_i = 1] = 1/2$*. Then, for all* $t \geq 0$*, and for all* $\mathbf{v} \in \mathbb{C}^S$

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

We next upper-bound $\mathbb{P}[\|\Delta\mathbf{v}\|_2 \geq a]$ in (63). For all $\alpha \geq 0$, using that $12^{\frac{n+m+1}{2}} \leq 12^{\frac{3}{2}}$, we have

$$\begin{aligned}
\mathbb{P}\left[\|\Delta\mathbf{v}\|_2 \geq \frac{\sqrt{3s}}{\sqrt{L}} 12^{\frac{3}{2}} c_1 \alpha\right] &\leq \mathbb{P}\left[\|\Delta\mathbf{v}\|_2 \geq \frac{\sqrt{3s}}{\sqrt{L}} 12^{\frac{n+m+1}{2}} c_1 \alpha\right] = \mathbb{P}\left[\|\Delta\mathbf{v}\|_2^2 \geq \frac{3s}{L} 12^{n+m+1} c_1^2 \alpha^2\right] \\
&\leq \sum_{k=1}^{3s} \mathbb{P}\left[|[\Delta\mathbf{v}]_k|^2 \geq \frac{1}{L} 12^{n+m+1} c_1^2 \alpha^2\right] \qquad (64) \\
&= \sum_{k=1}^{3s} \mathbb{P}\left[|[\Delta\mathbf{v}]_k| \geq \frac{1}{\sqrt{L}} 12^{\frac{n+m+1}{2}} c_1 \alpha\right] \\
&\leq 3s \cdot 2\exp\left(-c\min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right)\right) \qquad (65) \\
&\leq \frac{\delta}{2|\Omega|} \qquad (66)
\end{aligned}$$

where (64) follows from the union bound, (65) follows from Lemma 1, and to obtain (66) we chose $\alpha = \frac{c_2^2}{c} \log\left(\frac{12S|\Omega|}{\delta}\right)$. To see this, note that $\min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right) = \frac{\alpha}{c_2^2}$ as long as $\alpha \geq c_2^2$, which holds since $c \leq 1$. We have established that $\mathbb{P}[\|\Delta\mathbf{v}\|_2 \geq a] \leq \frac{\delta}{2|\Omega|}$ with $a = \frac{\sqrt{3s}}{\sqrt{L}} 12^{\frac{3}{2}} c_1 \frac{c_2^2}{c} \log\left(\frac{12s|\Omega|}{\delta}\right)$.

Substituting (66) into (63) we get

$$\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_1^{(m,n)}(\mathbf{r})| \geq \sqrt{c_5} \frac{\sqrt{s}}{\sqrt{L}} \log\left(\frac{12s|\Omega|}{\delta}\right) \sqrt{\log\left(\frac{8|\Omega|}{\delta}\right)}\right] \leq \delta + \mathbb{P}[\overline{\mathcal{E}}_{1/4}]$$

where $c_5 = (5\sqrt{3}\,12^{\frac{3}{2}} c_1 \frac{c_2^2}{\sqrt{c}})^2$ is a numerical constant. This concludes the proof. □

Lemma 8. *Let $\Omega \subset [0,1]^2$ be a finite set of points. Suppose that $m+n \leq 2$. For all $\epsilon, \delta \geq 0$, and for all $\tau > 0$ with*

$$\tau \leq \frac{\epsilon c_6}{\sqrt{\log\left(\frac{4|\Omega|}{\delta}\right)}} \tag{67}$$

where $c_6 \leq 1/4$ is a numerical constant, we have that

$$\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_2^{(m,n)}(\mathbf{r})| \geq \epsilon \middle| \mathcal{E}_\tau\right] \leq \delta.$$

Proof. By a union bound

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_2^{(m,n)}(\mathbf{r})| \geq \epsilon \middle| \mathcal{E}_\tau\right] &\leq \sum_{\mathbf{r}\in\Omega} \mathbb{P}\left[\left|\left\langle \mathbf{u}, (\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\rangle\right| \geq \epsilon \middle| \mathcal{E}_\tau\right] \\
&\leq \sum_{\mathbf{r}\in\Omega} \mathbb{P}\left[\left|\left\langle \mathbf{u}, (\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\rangle\right| \geq \left\|(\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\|_2 \frac{\epsilon}{c_5\tau}\right] && (68) \\
&\leq |\Omega| 4 e^{-\frac{(\epsilon/(c_5\tau))^2}{4}} && (69) \\
&\leq \delta && (70)
\end{aligned}$$

where (68) follows from (71) below, (69) follows by Hoeffding's inequality (cf. Lemma 7), and to obtain (70) we used the assumption (67) with $c_6 = 1/(2c_5)$.

To complete the proof, note that by (53) we have $\left\|\mathbf{L}-\bar{\mathbf{L}}\right\| \leq 2.5\tau$ on $\mathcal{E}_\tau$. Thus, conditioned on $\mathcal{E}_\tau$,

$$\left\|(\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\|_2 \leq \left\|\mathbf{L}-\bar{\mathbf{L}}\right\| \left\|\bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\|_2 \leq 2.5\tau \left\|\bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\|_1 \leq c_5\tau \tag{71}$$

where we used $\|\cdot\|_2 \leq \|\cdot\|_1$, and the last inequality follows because, for all $\mathbf{r}$,

$$\left\|\bar{\mathbf{v}}_{(m,n)}(\mathbf{r})\right\|_1 = \frac{1}{\kappa^{m+n}} \sum_{k=1}^{S} \left(\left|\bar{G}^{(m,n)}(\mathbf{r}-\mathbf{r}_k)\right| + \left|\frac{1}{\kappa}\bar{G}^{(m+1,n)}(\mathbf{r}-\mathbf{r}_k)\right| + \left|\frac{1}{\kappa}\bar{G}^{(m,n+1)}(\mathbf{r}-\mathbf{r}_k)\right|\right) \leq \frac{c_5}{2.5}$$

where c_5 is a numerical constant, and where we used [8, C.12, Table 6] and $N/\kappa \leq 0.5514$. □

We are now ready to prove the main result of this subsection.

Proof of Lemma 5. From (61), we obtain

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \frac{1}{\kappa^{n+m}} \left|Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r})\right| \geq 2\epsilon\right] &= \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \left|I_1^{(m,n)}(\mathbf{r}) + I_2^{(m,n)}(\mathbf{r})\right| \geq 2\epsilon\right] \\
&\leq \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \left|I_1^{(m,n)}(\mathbf{r})\right| \geq \epsilon\right] + \mathbb{P}\left[\overline{\mathcal{E}}_\tau\right] + \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \left|I_2^{(m,n)}(\mathbf{r})\right| \geq \epsilon | \mathcal{E}_\tau\right] && (72) \\
&\leq 4\delta
\end{aligned}$$

where (72) follows from the union bound and $\mathbb{P}[A] = \mathbb{P}[A \cap \overline{B}] + \mathbb{P}[A \cap B] \leq \mathbb{P}[\overline{B}] + \mathbb{P}[A|B]$ with $B = \mathcal{E}_\tau$ and $A = \left\{\max_{\mathbf{r} \in \Omega} \left| I_2^{(m,n)}(\mathbf{r}) \right| \geq \epsilon \right\}$, and the last inequality follows from Lemmas 4, 6 and 8, respectively.

Specifically, we choose $\tau = \epsilon c_6 \log^{-1/2}\left(\frac{4|\Omega|}{\delta}\right)$. It then follows from Lemma 8 that the third probability in (72) is smaller than δ. With this choice of τ, the condition in Lemma 4 becomes $L \geq S \frac{c_4}{\epsilon^2 c_6^2} \log\left(\frac{4|\Omega|}{\delta}\right) \log\left(\frac{18S^2}{\delta}\right)$, which is satisfied by assumption ($c = \frac{c_4}{c_6^2}$). Moreover, $\tau \leq 1/4$ since $\epsilon \leq 1$ and $c_6 \leq 1/4$. Thus, Lemma 4 yields $\mathbb{P}[\overline{\mathcal{E}_\tau}] \leq \delta$ and $\mathbb{P}[\overline{\mathcal{E}_{1/4}}] \leq \delta$. Finally, observe that the conditions of Lemma 6 are satisfied by assumption, thus the first probability in (72) can be upper bounded by

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

This concludes the proof. □

8.4 Step 3b: $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ are close for all $\mathbf{r}$

We next use an ϵ-net argument together with Lemma 5 to establish that $Q^{(m,n)}(\mathbf{r})$ is close to $\bar{Q}^{(m,n)}(\mathbf{r})$ for all $\mathbf{r} \in [0,1]^2$ with high probability.

Lemma 9. *Let $\epsilon, \delta \geq 0$. If*

$$L \geq S \frac{c}{\epsilon^2} \log^3\left(\frac{c' L^6}{\delta \epsilon^2}\right) \tag{73}$$

then, with probability at least $1 - \delta$,

$$\max_{\mathbf{r} \in [0,1]^2, (m,n):\, m+n \leq 2} \frac{1}{\kappa^{n+m}} \left| Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r}) \right| \leq \epsilon. \tag{74}$$

Proof. We start by choosing a set of points Ω (i.e., the ϵ-net) that is sufficiently dense in the ∞-norm. Specifically, we choose the points in Ω on a rectangular grid such that

$$\max_{\mathbf{r} \in [0,1]^2} \min_{\mathbf{r}_g \in \Omega} |\mathbf{r} - \mathbf{r}_g| \leq \frac{\epsilon}{3\tilde{c} L^{5/2}}. \tag{75}$$

The cardinality of the set Ω is

$$|\Omega| = \left(\frac{3\tilde{c} L^{5/2}}{\epsilon}\right)^2 = c' L^5 / \epsilon^2. \tag{76}$$

We first establish by application of Lemma 5 that $\left| Q^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r}_g) \right|$ is small for all points $\mathbf{r}_g \in \Omega$, and then show that this result continues to hold for all $\mathbf{r} \in [0,1]^2$. We start by noting that the condition of Lemma 5 is satisfied by assumption (73). Using a union bound over all 6 pairs (m,n) with $m + n \leq 2$, it now follows from Lemma 5, that

$$\left\{ \max_{\mathbf{r}_g \in \Omega, m+n \leq 2} \frac{1}{\kappa^{m+n}} \left| Q^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r}_g) \right| \leq \frac{\epsilon}{3} \right\} \tag{77}$$

holds with probability at least $1-6\delta' = 1-\frac{\delta}{2}$ (here, δ' is the original δ in Lemma 5). In order to show that this result continues to hold for all $\mathbf{r} \in [0,1]^2$, we will also need that the event

$$\left\{ \max_{\mathbf{r}\in[0,1]^2, m+n\leq 2} \frac{1}{\kappa^{m+n}} \left| Q^{(m,n)}(\mathbf{r}) \right| \leq \frac{\tilde{c}}{2} L^{3/2} \right\} \tag{78}$$

holds with probability at least $1-\frac{\delta}{2}$. This is shown in Section 8.4.1 below. By the union bound, the events in (77) and (78) hold simultaneously with probability at least $1-\delta$. As we will see in Section 8.4.2, (77) and (78) imply (74) which concludes the proof.

8.4.1 Proof that (78) holds with probability at least $1-\frac{\delta}{2}$

In order to show that (78) holds with probability at least $1-\frac{\delta}{2}$, we first upper-bound $|Q^{(m,n)}(\mathbf{r})|$. By (60),

$$\begin{aligned}
\frac{1}{\kappa^{m+n}} \left| Q^{(m,n)}(\mathbf{r}) \right| &= \left| \left\langle \mathbf{L}\mathbf{u}, \mathbf{v}^{(m,n)}(\mathbf{r}) \right\rangle \right| \\
&\leq \|\mathbf{L}\| \, \|\mathbf{u}\|_2 \left\| \mathbf{v}^{(m,n)}(\mathbf{r}) \right\|_2 \\
&\leq \|\mathbf{L}\| \sqrt{S} \left\| \mathbf{v}^{(m,n)}(\mathbf{r}) \right\|_2 \\
&\leq \|\mathbf{L}\| \sqrt{S}\sqrt{3S} \left\| \mathbf{v}^{(m,n)}(\mathbf{r}) \right\|_\infty \\
&= \|\mathbf{L}\| \sqrt{3}\, S \max_{(m',n')\in\{(m,n),(m+1,n),(m,n+1)\}} \frac{1}{\kappa^{m'+n'}} \left| G^{(m',n')}(\mathbf{r}) \right| \qquad (79)
\end{aligned}$$

where we used $\|\mathbf{u}\|_2 = \sqrt{S}$, since the entries of $\mathbf{u}$ are ± 1. Next, note that, for all $\mathbf{r}$, we have, by (38)

$$\begin{aligned}
\frac{1}{\kappa^{m'+n'}} \left| G^{(m',n')}(\mathbf{r}) \right| &= \frac{1}{\kappa^{m'+n'}} \mathbf{a}^H \mathbf{V}^{(m',n')}(\mathbf{r}) \mathbf{a} \leq \frac{1}{\kappa^{m'+n'}} \|\mathbf{a}\|_2^2 \left\| \mathbf{V}^{(m',n')}(\mathbf{r}) \right\| \\
&\leq c_1 \frac{(2\pi N)^{n'+m'}}{\kappa^{m'+n'}} \sqrt{L} \|\mathbf{a}\|_2^2 \leq c_1 12^{\frac{m'+n'}{2}} \sqrt{L} \|\mathbf{a}\|_2^2 \\
&\leq c_1 12^{\frac{3}{2}} \sqrt{L} \|\mathbf{a}\|_2^2 \qquad (80)
\end{aligned}$$

where we used Lemma 2 to conclude $\left\| \mathbf{V}^{(m',n')} \right\| \leq \left\| \mathbf{V}^{(m',n')} \right\|_F \leq c_1 (2\pi N)^{n'+m'} \sqrt{L}$ and (80) follows from $m'+n' \leq 3$ (recall that $m+n \leq 2$). Substituting (80) into (79) and using that $S \leq L$ (by assumption (73)) yields

$$\frac{1}{\kappa^{m+n}} \left| Q^{(m,n)}(\mathbf{r}) \right| \leq \sqrt{3}\, 12^{\frac{3}{2}} c_1 L^{3/2} \|\mathbf{L}\| \, \|\mathbf{a}\|_2^2.$$

It follows that (with $\frac{\tilde{c}}{2} = 2.5 \cdot 3 \cdot \sqrt{3}\, 12^{\frac{3}{2}} c_1$)

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{r}\in[0,1]^2, m+n\leq 2} \frac{1}{\kappa^{m+n}} \left| Q^{(m,n)}(\mathbf{r}) \right| \geq \frac{\tilde{c}}{2} L^{3/2} \right] &\leq \mathbb{P}\left[\|\mathbf{L}\| \, \|\mathbf{a}\|_2^2 \geq 2.5 \cdot 3 \right] \\
&\leq \mathbb{P}[\|\mathbf{L}\| \geq 2.5] + \mathbb{P}\left[\|\mathbf{a}\|_2^2 \geq 3 \right] \qquad (81) \\
&\leq \frac{\delta}{2} \qquad (82)
\end{aligned}$$

as desired. Here, (81) follows from the union bound and (82) follows from $\mathbb{P}[\|\mathbf{L}\| \geq 2.5] \leq \mathbb{P}\left[\overline{\mathcal{E}}_{1/4}\right] \leq \frac{\delta}{4}$ (by (52) and application of Lemma 4; note that the condition of Lemma 4 is satisfied by (73)) and $\mathbb{P}\left[\|\mathbf{a}\|_2^2 \geq 3\right] \leq \frac{\delta}{4}$, shown below. Using that $4\log(4/\delta) \leq L$ (by (73)), we obtain

$$\begin{aligned}\mathbb{P}\left[\|\mathbf{a}\|_2^2 \geq 3\right] &\leq \mathbb{P}\left[\|\mathbf{a}\|_2^2 \geq 2\left(1 + \frac{2\log(4/\delta)}{L}\right)\right] \\ &\leq \mathbb{P}\left[\|\mathbf{a}\|_2 \geq \left(1 + \frac{\sqrt{2\log(4/\delta)}}{\sqrt{L}}\right)\right] \leq e^{-\frac{2\log(4/\delta)}{2}} = \frac{\delta}{4} \qquad (83)\end{aligned}$$

where we used $\sqrt{2(1+\beta^2)} \geq (1+\beta)$, for all β, and a standard concentration inequality for the norm of a Gaussian random vector, e.g., [21, Eq. 1.6]. This concludes the proof of (78) holding with probability at least $1 - \frac{\delta}{2}$.

8.4.2 Proof that (77) and (78) imply (74)

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

$$\begin{aligned}&\frac{1}{\kappa^{n+m}}\left|Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r})\right| \leq \\ &\quad \frac{1}{\kappa^{n+m}}\left[\left|Q^{(m,n)}(\mathbf{r}) - Q^{(m,n)}(\mathbf{r}_g)\right| + \left|Q^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r}_g)\right| + \left|\bar{Q}^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r})\right|\right]. \qquad (84)\end{aligned}$$

We next upper-bound the terms in (84) separately. With a slight abuse of notation, we write $Q^{(m,n)}(\tau,\nu) = Q^{(m,n)}([\tau,\nu]^T) = Q^{(m,n)}(\mathbf{r})$. The first absolute value in (84) can be upper-bounded according to

$$\begin{aligned}\left|Q^{(m,n)}(\mathbf{r}) - Q^{(m,n)}(\mathbf{r}_g)\right| &= \left|Q^{(m,n)}(\tau,\nu) - Q^{(m,n)}(\tau,\nu_g) + Q^{(m,n)}(\tau,\nu_g) - Q^{(m,n)}(\tau_g,\nu_g)\right| \\ &\leq \left|Q^{(m,n)}(\tau,\nu) - Q^{(m,n)}(\tau,\nu_g)\right| + \left|Q^{(m,n)}(\tau,\nu_g) - Q^{(m,n)}(\tau_g,\nu_g)\right| \\ &\leq |\nu - \nu_g| \sup_z \left|Q^{(m,n+1)}(\tau,z)\right| + |\tau - \tau_g| \sup_z \left|Q^{(m+1,n)}(z,\nu_g)\right| \\ &\leq |\nu - \nu_g| 2\pi N \sup_z \left|Q^{(m,n)}(\tau,z)\right| + |\tau - \tau_g| 2\pi N \sup_z \left|Q^{(m,n)}(z,\nu_g)\right| \qquad (85)\end{aligned}$$

where (85) follows from Bernstein's polynomial inequality, stated below (note that $Q^{(m,n)}(\tau,\nu)$ is a trigonometric polynomial of degree N in both τ and ν).

Proposition 5 (Bernstein's polynomial inequality [14, Cor. 8])**.** *Let p be a trigonometric polynomial of degree N with complex coefficients p_k, i.e., $p(\theta) = \sum_{k=-N}^{N} p_k e^{i2\pi\theta k}$. Then*

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

Substituting (78) into (85) yields that

$$\frac{1}{\kappa^{m+n}}\left|Q^{(m,n)}(\mathbf{r}) - Q^{(m,n)}(\mathbf{r}_g)\right| \leq \frac{\tilde{c}}{2} L^{5/2}(|\tau - \tau_g| + |\nu - \nu_g|) \leq \tilde{c} L^{5/2} |\mathbf{r} - \mathbf{r}_g| \leq \frac{\epsilon}{3} \qquad (86)$$

where the last inequality follows from (75).

We next upper-bound the third absolute value in (84). Using steps analogous to those leading to (86), we obtain

$$\frac{1}{\kappa^{m+n}}\left|\bar{Q}^{(m,n)}(\mathbf{r}_g)-\bar{Q}^{(m,n)}(\mathbf{r})\right|\leq\frac{\epsilon}{3}. \tag{87}$$

Substituting (77), (86), and (87) into (84) yields that

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

with concludes the proof of Lemma 9. □

8.5 Step 3c: Ensuring that $Q(\mathbf{r})<1$ for all $\mathbf{r}\notin\mathcal{T}$

Lemma 10. *Suppose that*

$$L\geq Sc\log^3\left(\frac{c'L^6}{\delta}\right).$$

Then with probability at least $1-\delta$ the following statements hold:

1. *For all $\mathbf{r}$, that satisfy $\min_{\mathbf{r}_j\in\mathcal{T}}|\mathbf{r}-\mathbf{r}_j|\geq 0.2447/N$ we have that $Q(\mathbf{r})<0.9963$.*
2. *For all $\mathbf{r}\notin\mathcal{T}$ that satisfy $0<|\mathbf{r}-\mathbf{r}_j|\leq 0.2447/N$ for some $\mathbf{r}_j\in\mathcal{T}$, we have that $Q(\mathbf{r})<1$.*

Proof. Choose $\epsilon=0.0005$. It follows from Lemma 9 that

$$\frac{1}{\kappa^{n+m}}\left|Q^{(m,n)}(\mathbf{r})-\bar{Q}^{(m,n)}(\mathbf{r})\right|\leq 0.0005,\quad\text{for all } (m,n)\colon m+n\leq 2,\quad\text{and for all } \mathbf{r} \tag{88}$$

with probability at least $1-\delta$. To prove the lemma we will show that Statement 1 and 2 follow from (88) and certain properties of $\bar{Q}^{(m,n)}(\mathbf{r})$ established in [8].

Statement 1 follows directly from combining (88) with the following result via the triangle inequality.

Proposition 6 ([8, Lem. C.4])**.** *For all $\mathbf{r}$, that satisfy $\min_{\mathbf{r}_j\in\mathcal{T}}|\mathbf{r}-\mathbf{r}_j|\geq 0.2447/N$ we have that $Q(\mathbf{r})<0.9958$.*

In order to prove Statement 2, assume without loss of generality (w.l.o.g.) that $\mathbf{0}\in\mathcal{T}$, and consider $\mathbf{r}$ with $|\mathbf{r}|\leq 0.2447/N$. Statement 2 is established by showing that the Hessian matrix of $\tilde{Q}(\mathbf{r}):=|Q(\mathbf{r})|$, i.e.,

$$\mathbf{H}=\begin{bmatrix}\tilde{Q}^{(2,0)}(\mathbf{r}) & \tilde{Q}^{(1,1)}(\mathbf{r})\\ \tilde{Q}^{(1,1)}(\mathbf{r}) & \tilde{Q}^{(0,2)}(\mathbf{r})\end{bmatrix},\quad \tilde{Q}^{(m,n)}(\mathbf{r}):=\frac{\partial^m}{\partial\tau^m}\frac{\partial^n}{\partial\nu^n}\tilde{Q}(\mathbf{r})$$

is negative definite. This is accomplished by showing that

$$\mathrm{trace}(\mathbf{H})=\tilde{Q}^{(2,0)}+\tilde{Q}^{(0,2)}<0 \tag{89}$$

$$\det(\mathbf{H})=\tilde{Q}^{(2,0)}\tilde{Q}^{(0,2)}-(\tilde{Q}^{(1,1)})^2>0 \tag{90}$$

which implies that both eigenvalues of $\mathbf{H}$ are strictly negative. To this end, we will need the following result.

Proposition 7 ([8, Sec. C.2]). *For* $|\mathbf{r}| \leq 0.2447/N$ *and for* $N \geq 512$,

$$1 \geq \bar{Q}(\mathbf{r}) \geq 0.6447 \tag{91}$$

$$\frac{1}{\kappa^2}\bar{Q}^{(2,0)}(\mathbf{r}) \leq -0.3550 \tag{92}$$

$$\frac{1}{\kappa^2}|\bar{Q}^{(1,1)}(\mathbf{r})| \leq 0.3251 \tag{93}$$

$$\frac{1}{\kappa^2}|\bar{Q}^{(1,0)}(\mathbf{r})| \leq 0.3344. \tag{94}$$

Define $Q_R^{(m,n)} = \frac{1}{\kappa^{m+n}}\mathrm{Re}(Q^{(m,n)})$ and $Q_I^{(m,n)} = \frac{1}{\kappa^{m+n}}\mathrm{Im}(Q^{(m,n)})$. We have that

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

therefore

$$\begin{aligned}
\frac{1}{\kappa^2}\tilde{Q}^{(2,0)} &= -\frac{(Q_R Q_R^{(1,0)} + Q_I Q_I^{(1,0)})^2}{|Q|^3} + \frac{|Q^{(1,0)}|^2 + Q_R Q_R^{(2,0)} + Q_I Q_I^{(2,0)}}{|Q|} \\
&= -\frac{Q_R^2 {Q_R^{(1,0)}}^2 + 2Q_R Q_R^{(1,0)} Q_I Q_I^{(1,0)} + Q_I^2 {Q_I^{(1,0)}}^2}{|Q|^3} + \frac{{Q_R^{(1,0)}}^2 + {Q_I^{(1,0)}}^2 + Q_R Q_R^{(2,0)} + Q_I Q_I^{(2,0)}}{|Q|} \\
&= \left(1 - \frac{Q_R^2}{|Q|^2}\right)\frac{{Q_R^{(1,0)}}^2}{|Q|} - \frac{2Q_R Q_R^{(1,0)} Q_I Q_I^{(1,0)} + Q_I^2 {Q_I^{(1,0)}}^2}{|Q|^3} + \frac{{Q_I^{(1,0)}}^2 + Q_I Q_I^{(2,0)}}{|Q|} + \frac{Q_R}{|Q|}Q_R^{(2,0)}
\end{aligned} \tag{95}$$

By Proposition 7, the triangle inequality, and using that $\bar{Q}^{(m,n)}(\mathbf{r})$ is real, the following bounds are in force:

$$\begin{aligned}
Q_R(\mathbf{r}) &\leq \bar{Q}(\mathbf{r}) + \epsilon \leq 1 + \epsilon \\
Q_R(\mathbf{r}) &\geq \bar{Q}(\mathbf{r}) - \epsilon \geq 0.6447 - \epsilon \\
Q_I^{(m,n)} &\leq \epsilon \\
Q_R^{(2,0)}(\mathbf{r}) &\leq \frac{1}{\kappa^2}\bar{Q}^{(2,0)}(\mathbf{r}) + \epsilon \leq -0.3550 + \epsilon \\
|Q_R^{(1,1)}| &\leq \frac{1}{\kappa^2}|\bar{Q}^{(1,1)}(\mathbf{r})| + \epsilon \leq 0.3251 + \epsilon \\
|Q_R^{(1,0)}(\mathbf{r})| &\leq \frac{1}{\kappa^2}|\bar{Q}^{(1,0)}(\mathbf{r})| + \epsilon \leq 0.3344 + \epsilon.
\end{aligned}$$

Using these bounds in (95) with $\epsilon = 0.0005$ yields that $\tilde{Q}^{(2,0)} < -0.3539$, which implies that (89) is satisfied.

It remains to verify (90). First note that

$$\begin{aligned}
&\frac{1}{\kappa^2}\tilde{Q}^{(1,1)} \\
&= \frac{Q_R^{(1,1)}Q_R + Q_R^{(1,0)}Q_R^{(0,1)} + Q_I^{(1,1)}Q_I + Q_I^{(1,0)}Q_I^{(0,1)}}{|Q|} - \frac{(Q_R^{(0,1)}Q_R + Q_I^{(0,1)}Q_I)(Q_R^{(1,0)}Q_R + Q_I^{(1,0)}Q_I)}{|Q|^3} \\
&= Q_R^{(1,1)}\frac{Q_R}{|Q|} + \frac{Q_R^{(1,0)}Q_R^{(0,1)}}{|Q|}\left(1 - \frac{Q_R^2}{|Q|^2}\right) + \frac{Q_I^{(1,1)}Q_I + Q_I^{(1,0)}Q_I^{(0,1)}}{|Q|} \\
&- \frac{Q_R^{(0,1)}Q_R Q_I^{(1,0)}Q_I + Q_I^{(0,1)}Q_I(Q_R^{(1,0)}Q_R + Q_I^{(1,0)}Q_I)}{|Q|^3}. \qquad (96)
\end{aligned}$$

Using the bounds above in (96) yields, with $\epsilon = 0.0005$, that $\frac{1}{\kappa^2}|\tilde{Q}^{(1,1)}| \leq 0.3267$. With $\frac{1}{\kappa^2}\tilde{Q}^{(2,0)} < -0.3539$, it follows that the RHS of (89) can be lower bounded by $\frac{1}{\kappa^2}(0.3539^2 - 0.3267^2) = \frac{1}{\kappa^2}0.01855 > 0$ i.e., (90) holds as well, which concludes the proof of Statement 2. □

Acknowledgments

RH would like to thank Helmut Bölcskei, Céline Aubel, and Emmanuel Candès for helpful discussions. RH would also like to thank Emmanuel Candès for his hospitality during a visit at the Statistics Department at Stanford, and Helmut Bölcskei for his support and for initiating this visit. VM was supported by the Swiss National Science Foundation fellowship for advanced researchers under grant PA00P2_139678.

A Proof of (12)

The proof of the I/O relation (12) appears in [4]. For convenience of the reader, we present the details below. We first write (9) in its equivalent form

$$y(t) = \int L_H(t,f)X(f)e^{i2\pi ft}df \qquad (97)$$

where $X(f) = \int x(t)e^{i2\pi ft}dt$ is the Fourier transform of $x(t)$, and $L_H(t,f)$ is the time-varying transfer function given by

$$L_H(t,f) := \int\int s_H(\tau,\nu)e^{i2\pi(\nu t - \tau f)}d\tau d\nu. \qquad (98)$$

Since $x(t)$ is band-limited to $[0,B)$, we may write

$$X(f) = X(f)H_I(f), \quad H_I(f) := \begin{cases} 1, & 0 \leq f < B \\ 0, & \text{else.} \end{cases}$$

Since $y(t)$ is time-limited to $[-T/2, T/2)$ we may write

$$y(t) = y(t)h_O(t), \quad h_O(t) := \begin{cases} 1, & -T/2 \leq t < T/2 \\ 0, & \text{else.} \end{cases}$$

With the input band-limitation and the output time-limitation, (97) becomes

$$y(t) = \int \overline{L_H}(t,f) X(f) e^{i2\pi f t} df \tag{99}$$

with

$$\overline{L_H}(t,f) \coloneqq L_H(t,f) h_O(t) H_I(f) \tag{100}$$

i.e., the effect of input band-limitation and output time-limitation is accounted for by passing the input signal through a system with time varying transfer function given by $\overline{L_H}$. The spreading function $\overline{s_H}$ of the system (99) and $\overline{L_H}$ are related by the two-dimensional Fourier transform in (98). We see that $\overline{L_H}(t,f)$ "band-limited" with respect to t and f, and hence, by the sampling theorem, can be expressed in terms of its samples as

$$\overline{s_H}(\tau,\nu) = \sum_{m,\ell\in\mathbb{Z}} \overline{s_H}\left(\frac{m}{B}, \frac{l}{T}\right) \operatorname{sinc}\left(\left(\tau - \frac{m}{B}\right) B\right) \operatorname{sinc}\left(\left(\nu - \frac{\ell}{T}\right) T\right). \tag{101}$$

In terms of $\overline{s_H}(\tau,\nu)$ (99) can be written as

$$y(t) = \int\int \overline{s_H}(\tau,\nu) x(t-\tau) e^{i2\pi\nu t} d\nu d\tau \tag{102}$$

and with (101)

$$\begin{aligned} y(t) &= \sum_{m,\ell\in\mathbb{Z}} \overline{s_H}\left(\frac{m}{B}, \frac{\ell}{T}\right) \int \operatorname{sinc}\left(\left(\tau - \frac{m}{B}\right) B\right) x(t-\tau) d\tau \int \operatorname{sinc}\left(\left(\nu - \frac{\ell}{T}\right) T\right) e^{i2\pi\nu t} d\nu \\ &= \sum_{m,\ell\in\mathbb{Z}} \overline{s_H}\left(\frac{m}{B}, \frac{\ell}{T}\right) x\left(t - \frac{m}{B}\right) e^{j2\pi\frac{\ell}{T}t}. \end{aligned}$$

According to (98), (100), and (98), $\overline{s_H}(\tau,\nu)$ and $s_H(\tau,\nu)$ are related as

$$\overline{s_H}(\tau,\nu) = \int\int s_H(\tau',\nu') \operatorname{sinc}((\tau-\tau')B) \operatorname{sinc}((\nu-\nu')T) d\tau' d\nu'. \tag{103}$$

which concludes the proof of (12).

B Proof of (17) and of (2)

We first establish (17). Starting with (16) and changing the order of summation according to $\ell = \tilde{\ell} + Lk$ and $r = \tilde{r} + Lq$ with $\tilde{\ell}, \tilde{r} = N, ..., N$ and $k, q \in \mathbb{Z}$ yields

$$\begin{aligned} y_p &= \sum_{n=1}^{S} b_n \sum_{\tilde{\ell},\tilde{r}=-N}^{N} \sum_{k,q\in\mathbb{Z}} \operatorname{sinc}(\tilde{\ell} + kL - \bar{\tau}_n B) \operatorname{sinc}(\tilde{r} + Lq - \bar{\nu}_n T) a_{p-\tilde{\ell}-kL} e^{i2\pi\frac{(\tilde{r}+Lq)p}{L}} \\ &= \sum_{n=1}^{S} b_n \sum_{\ell,r=-N}^{N} \sum_{k,q\in\mathbb{Z}} \operatorname{sinc}\left(\left(\frac{\ell}{L} - \tau_n + k\right) L\right) \operatorname{sinc}\left(\left(\frac{r}{L} - \nu_n + q\right) L\right) a_{p-\ell} e^{i2\pi\frac{rp}{L}} && (104) \\ &= \sum_{n=1}^{S} b_n \sum_{\ell,r=-N}^{N} D_N\left(\frac{\ell}{L} - \tau_n\right) D_N\left(\frac{r}{L} - \nu_n\right) a_{p-\ell} e^{i2\pi\frac{rp}{L}} && (105) \end{aligned}$$

where (104) follows from a_ℓ being L-periodic and by defining $\tau_n := \bar{\tau}_n \frac{B}{L}$ and $\nu_n := \bar{\nu}_n \frac{T}{L}$. To obtain (105), we used that

$$D_N(t) = \sum_{k \in \mathbb{Z}} \operatorname{sinc}\left(L(t-k)\right) = \frac{\sin(\pi L t)}{L \sin(\pi t)}.$$

Next, we establish (2). Starting from (105) and using that

$$D_N(t) = \frac{1}{L} \sum_{k=-N}^{N} e^{i2\pi t k}$$

yields

$$\begin{aligned}
y_p &= \sum_{n=1}^{S} b_n \sum_{\ell,r=-N}^{N} \left(\frac{1}{L^2} \sum_{k,q=-N}^{N} e^{i2\pi k\left(\frac{\ell}{L}-\tau_n\right)} e^{i2\pi q\left(\frac{r}{L}-\nu_n\right)} \right) a_{p-\ell} e^{i2\pi \frac{rp}{L}} \\
&= \sum_{n=1}^{S} b_n \sum_{\ell=N}^{N} \frac{1}{L} \sum_{k,q=-N}^{N} e^{i2\pi k\left(\frac{\ell}{L}-\tau_n\right)} e^{-i2\pi q \nu_n} a_{p-\ell} \frac{1}{L} \sum_{r=N}^{N} e^{i2\pi \frac{r}{L}(p+q)} \\
&= \sum_{n=1}^{S} b_n e^{i2\pi p \nu_n} \sum_{\ell=N}^{N} \frac{1}{L} \sum_{k=-N}^{N} e^{i2\pi k\left(\frac{\ell}{L}-\tau_n\right)} a_{p-\ell} \\
&= \sum_{n=1}^{S} b_n e^{i2\pi p \nu_n} \frac{1}{L} \sum_{\ell,k=-N}^{N} e^{-i2\pi k \tau_n} e^{i2\pi (p-\ell) \frac{k}{L}} a_\ell.
\end{aligned}$$

C Proof of Theorem 2

The proof follows by establishing exact recovery via construction of a dual certificate. The following proposition is standard, see e.g., [9].

Proposition 8. *Let* $\mathbf{s}$ *be supported on* $\mathcal{T}$, *suppose that* $\mathbf{s}$ *is feasible for* (19) *and assume that* $\mathbf{R}_\mathcal{T}$ *has full column rank. If there exists a vector* $\mathbf{v}$ *in the row space of* $\mathbf{R}$ *with*

$$\mathbf{v}_\mathcal{T} = \operatorname{sign}(\mathbf{s}_\mathcal{T}) \quad \textit{and} \quad \left\|\mathbf{v}_{\overline{\mathcal{T}}}\right\|_\infty < 1 \tag{106}$$

then $\mathbf{s}$ *is the unique minimizer to* (19).

The proof now follows directly from Proposition 3. To see this, set $\mathbf{u} = \operatorname{sign}(\mathbf{s}_\mathcal{T})$ in Proposition 3 and consider the polynomial $Q(\mathbf{r})$ from Proposition 3. Define $\mathbf{v}$ as $[\mathbf{v}]_{(m,n)} = Q([m/K, n/K])$ and note that $\mathbf{v}$ satisfies (106) since $Q([m/K, n/K]) = \operatorname{sign}(\mathbf{s}_{(m,n)})$ for $(m,n) \in \mathcal{T}$ and $Q([m/K, n/K]) < 1$ for $(m,n) \notin \mathcal{T}$.

D Bound on U

We have that

$$U(t) = 16 \sum_{p=-N}^{N} \min\left(1, \frac{1}{p^4}\right) P^{(m)}(p/L - t)$$

with

$$P^{(m)}(t) := \frac{1}{M} \sum_{k=-N}^{N} (-i2\pi k)^m e^{i2\pi tk} = \frac{\partial^m}{\partial t^m} \frac{\sin(L\pi t)}{M \sin(\pi t)}.$$

We start by upper bounding $|P^{(m)}(t)|$. First note that $|P^{(m)}(t)|$ is a 1-periodic and symmetric function, thus in order to upper bound $|P^{(m)}(t)|$, we only need to consider the case $0 \leq t \leq 1/2$.

For $m = 0$, we have that

$$|P^{(0)}(t)| \leq \min\left(4, \frac{1}{M|\sin(\pi t)|}\right).$$

Next, consider the case $m = 1$, and assume that $t \geq 1/L$. We have

$$P^{(1)}(t) = \frac{\cos(L\pi t) L\pi}{M \sin(\pi t)} - \frac{\pi \sin(L\pi t)\cos(\pi t)}{M \sin^2(\pi t)}.$$

Using that $\sin(\pi t) \geq 2t \geq 2/L$ for $1/L \leq t \leq 1/2$ we get

$$|P^{(1)}(t)| \leq \frac{1.5L\pi}{M|\sin(\pi t)|}.$$

Next, consider the case $m = 2$. We have

$$P^{(2)}(t) = \frac{\pi^2(1-L^2)\sin(L\pi t)}{M\sin(\pi t)} - \frac{2L\pi^2\cos(L\pi t)\cos(\pi t)}{M\sin^2(\pi t)} + \frac{2\pi^2\sin(L\pi t)\cos^2(\pi t)}{M\sin^3(\pi t)}.$$

Using again that that $\sin(\pi t) \geq 2t \geq 2/L$ for $1/L \leq t \leq 1/2$ we get

$$|P^{(2)}(t)| \leq \frac{2.5L^2\pi^2}{M|\sin(\pi t)|}.$$

Analogously, we can obtain bounds for $m = 3, 4$. We therefore obtain, for $1/L \leq |t| \leq 1/2$,

$$|P^{(m)}(t)| \leq (L\pi)^m \frac{c_1}{M|\sin(\pi t)|} \leq \underbrace{1.039c_1}_{c_2:=}(2N\pi)^m \frac{1}{M|\sin(\pi t)|}$$

where c_2 is a numerical constant and where we used that $(L/(2M))^m \leq 1.039$ for $N \geq 512$ and $m \leq 4$. Regarding the range $0 \leq |t| \leq 1/L$, simply note that by Bernstein's polynomial inequality (cf. Proposition 5) we have, for all t, from $|P^{(0)}(t)| \leq 4$, that

$$|P^{(m)}(t)| \leq 4(2N\pi)^m.$$

Using that $c_2 \geq 1$, we finally obtain

$$U(t) \leq 16 \sum_{p=-N}^{N} \min\left(1, \frac{1}{p^4}\right) c_2 \begin{cases} 4, & |p/L - t + n| \leq 1/L, n \in \mathbb{Z} \\ \frac{1}{M|\sin(\pi(p/L-t))|}, & \text{else.} \end{cases}$$

The RHS above is 1-periodic (in t) and symmetric around the origin. Thus it suffices to consider $t \in [0, 1/2]$. Assume furthermore that Lt is an even integer, the proof for general t is similar. For

$p \geq 0$, we have that $|p/L-t| \leq 1/2$ and thus $M|\sin(\pi(p/L-t))| \geq M|2(p/L-t)| = 2M/L|p-Lt| \geq 1/2|p-Lt|$. It follows that

$$\begin{aligned}
U(t) &\leq 16c_2 \sum_{p=0}^{N} \min\left(1, \frac{1}{p^4}\right) \min\left(4, \frac{2}{|p-Lt|}\right) \\
&\leq 16c_2 \sum_{p=0}^{Lt/2} \min\left(1, \frac{1}{p^4}\right) \frac{2}{Lt-p} + \sum_{p=Lt/2+1}^{Lt-1} \frac{1}{p^4} \frac{2}{Lt-p} + \sum_{p=Lt}^{N} \frac{1}{p^4} 4 \\
&= \frac{16c_2}{Lt} \left(\sum_{p=0}^{Lt/2} \min\left(1, \frac{1}{p^4}\right) \frac{Lt}{Lt-p} + \sum_{p=Lt/2+1}^{Lt-1} \frac{Lt}{p^4} \frac{2}{Lt-p} + \sum_{p=Lt}^{N} \frac{4Lt}{p^4} \right) \\
&\leq \frac{16c_2}{Lt} \left(\sum_{p=0}^{Lt/2} 2\min\left(1, \frac{1}{p^4}\right) + \sum_{p=Lt/2+1}^{Lt-1} 2\frac{2}{p^3} + \sum_{p=Lt}^{N} \frac{4}{p^3} \right) \\
&\leq \frac{\tilde{c}}{Lt}.
\end{aligned}$$

Analogously we can upper-bound the sum over $p = -N, ..., -1$ which yields $U(t) \leq \frac{c}{L|t|}$, as desired.

E Proof of Proposition 2

The argument is standard, see e.g., [27, Prop. 2.4]. By definition, $\mathbf{q}$ is dual feasible. To see this, note that

$$\left\|\mathbf{G}^H\mathbf{q}\right\|_{\mathcal{A}^*} = \sup_{\mathbf{r}\in[0,1]^2} \left|\left\langle \mathbf{G}^H\mathbf{q}, \mathbf{a}(\mathbf{r}) \right\rangle\right| = \sup_{\mathbf{r}\in[0,1]^2} \left|\left\langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}) \right\rangle\right| = \sup_{\mathbf{r}\in[0,1]^2} |Q(\mathbf{r})| \leq 1 \tag{107}$$

where the last inequality holds by assumption. By (25), we obtain

$$\langle \mathbf{q}, \mathbf{y} \rangle = \left\langle \mathbf{q}, \mathbf{G} \sum_{\mathbf{r}_n\in\mathcal{T}} b_n \mathbf{a}(\mathbf{r}_n) \right\rangle = \sum_{\mathbf{r}_n\in\mathcal{T}} b_n^* \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}_n) \rangle = \sum_{\mathbf{r}_n\in\mathcal{T}} b_n^* \mathrm{sign}(b_n) = \sum_{\mathbf{r}_n\in\mathcal{T}} |b_n| \geq \|\mathbf{z}\|_{\mathcal{A}} \tag{108}$$

where the last inequality holds by definition of the atomic norm. By Hölder's inequality we have that

$$\mathrm{Re}\,\langle \mathbf{q}, \mathbf{y} \rangle = \mathrm{Re}\,\langle \mathbf{q}, \mathbf{G}\mathbf{z} \rangle = \mathrm{Re}\left\langle \mathbf{G}^H\mathbf{q}, \mathbf{z} \right\rangle \leq \left\|\mathbf{G}^H\mathbf{q}\right\|_{\mathcal{A}^*} \|\mathbf{z}\|_{\mathcal{A}} \leq \|\mathbf{z}\|_{\mathcal{A}}$$

where we used (25) for the last inequality. We thus have established that $\mathrm{Re}\,\langle \mathbf{q}, \mathbf{y} \rangle = \|\mathbf{z}\|_{\mathcal{A}}$. Since $(\mathbf{z}, \mathbf{q})$ is primal-dual feasible, it follows from strong duality that $\mathbf{z}$ is a primal optimal solution and $\mathbf{q}$ is a dual optimal solution.

It remains to establish uniqueness. To this end, suppose that $\hat{\mathbf{z}} = \sum_{\mathbf{r}_n\in\hat{\mathcal{T}}} \hat{b}_n \mathbf{a}(\mathbf{r}_n)$ with $\|\hat{\mathbf{z}}\|_{\mathcal{A}} =$

$\sum_{\mathbf{r}_n \in \hat{\mathcal{T}}} |\hat{b}_n|$ and $\hat{\mathcal{T}} \neq \mathcal{T}$ is another optimal solution. We then have

$$\begin{aligned}
\operatorname{Re}\langle \mathbf{q}, \mathbf{G}\hat{\mathbf{z}} \rangle &= \operatorname{Re}\left\langle \mathbf{q}, \mathbf{G} \sum_{\mathbf{r}_n \in \hat{\mathcal{T}}} \hat{b}_n \mathbf{a}(\mathbf{r}_n) \right\rangle \\
&= \sum_{\mathbf{r}_n \in \mathcal{T}} \operatorname{Re}\left(\hat{b}_n^* \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}_n) \rangle \right) + \sum_{\mathbf{r}_n \in \hat{\mathcal{T}} \setminus \mathcal{T}} \operatorname{Re}\left(\hat{b}_n^* \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}_n) \rangle \right) \\
&< \sum_{\mathbf{r}_n \in \mathcal{T}} |\hat{b}_n| + \sum_{\mathbf{r}_n \in \hat{\mathcal{T}} \setminus \mathcal{T}} |\hat{b}_n| \\
&= \|\hat{\mathbf{z}}\|_{\mathcal{A}}
\end{aligned}$$

where we used that $|Q(\mathbf{r})| < 1$ for $\mathbf{r} \notin \mathcal{T}$. This contradicts strong duality and implies that all optimal solutions must be supported on $\mathcal{T}$. Since the set of atoms with $\mathbf{r}_n \in \mathcal{T}$ are linearly independent, it follows that the optimal solution is unique.

References

[1] W. U. Bajwa, K. Gedalyahu, and Y. C. Eldar. Identification of parametric underspread linear systems and super-resolution radar. *IEEE Trans. Signal Process.*, 59(6):2548–2561, June 2011.

[2] W. U. Bajwa, A. M. Sayeed, and R. Nowak. Learning sparse doubly-selective channels. In *Proc. of 46th Allerton Conf. on Commun., Control, and Comput.*, pages 575–582, Monticello, IL, 2008.

[3] S. R. Becker, E. J. Candès, and M. C. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218, September 2011.

[4] P. A. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. Commun. Syst.*, 11(4):360–393, December 1963.

[5] P. A. Bello. Measurement of random time-variant linear channels. *IEEE Trans. Inf. Theory*, 15(4):469–475, July 1969.

[6] B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *arXiv:1204.0562 [cs, math]*, April 2012.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[8] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.*, 67(6):906–956, June 2014.

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

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

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

[12] A. Gershman and N. Sidiropoulos, editors. *Space-Time Processing for MIMO Communications.* Wiley, 2005.

[13] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[14] L. A. Harris. Bernstein's polynomial inequalities and functional analysis. *Irish Math. Soc. Bull.*, 36:19–33, 1996.

[15] R. Heckel and H. Bölcskei. Identification of sparse linear operators. *IEEE Trans. Inf. Theory*, 59(12):7985–8000, 2013.

[16] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.*, 57(6):2275–2284, 2009.

[17] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge; New York, 2 edition edition, December 2012.

[18] T. Kailath. Measurements on time-variant communication channels. *IRE Trans. Inf. Theory*, 8(5):229–236, September 1962.

[19] W. Kozek and G. E. Pfander. Identification of operators with bandlimited symbols. *SIAM J. Math. Anal.*, 37(3):867–888, 2005.

[20] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Communications on Pure and Applied Mathematics*, 67(11):1877–1904, November 2014.

[21] M. Ledoux and M. Talagrand. *Probability in Banach spaces: Isoperimetry and processes.* Springer, Berlin, Heidelberg, 1991.

[22] G. E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.

[23] G. E. Pfander and D. F. Walnut. Measurement of time-variant linear channels. *IEEE Trans. Inf. Theory*, 52(11):4808–4820, December 2006.

[24] M. Rudelson and R. Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18(0), October 2013.

[25] M. Soltanolkotabi. *Algorithms and theory for clustering and nonconvex quadratic programming.* 2014. Stanford Ph.D. Dissertation.

[26] T. Strohmer. Pseudodifferential operators and banach algebras in mobile communications. *Appl. Comput. Harmon. Anal.*, 20(2):237–249, March 2006.

[27] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Inform. Theory*, 59(11):7465–7490, November 2013.

[28] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut. Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing. *IEEE J. Sel. Topics Signal Process.*, 4(2):255–271, 2010.

[29] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing: Theory and applications*, pages 210–268. Cambridge University Press, 2012.