Information and Inference: A Journal of the IMA (2015) Page 1 of 53
doi:10.1093/imaiai/xxx000

Super-Resolution Radar

REINHARD HECKEL*
Dept. of Electrical Engineering and Computer Science, University of California, Berkeley, CA∗
*Corresponding author: reinhard.heckel@gmail.com

VENIAMIN I. MORGENSHTERN
Dept. of Statistics, Stanford University, CA

AND

MAHDI SOLTANOLKOTABI
Ming Hsieh Dept. of Electrical Engineering, University of Southern California, Los Angeles, CA

[Received on 26 December 2015]

In this paper we study the identification of a time-varying linear system from its response to a known input signal. More specifically, we consider systems whose response to the input signal is given by a weighted superposition of delayed and Doppler shifted versions of the input. 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 delay and Doppler resolution of $1/B$ and $1/T$. We show that this resolution limit can be overcome, i.e., we can exactly recover the continuous delay-Doppler pairs and the corresponding attenuation factors, by solving a convex optimization problem. This result holds provided that the distance between the delay-Doppler pairs is at least $2.37/B$ in time or $2.37/T$ in frequency. Furthermore, this result allows the total number of delay-Doppler pairs to be linear up to a log-factor in BT, the dimensionality of the response of the system, and thereby the limit for identifiability. Stated differently, 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.

Keywords: Super-resolution; radar; convex programming; compressed sensing; sparsity; line spectral estimation; linear time-varying system

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. In this paper, we study the problem of identifying a system H whose response $y = Hx$ to the probing

*R. Heckel was with the Dept. of Information Technology and Electrical Engineering, ETH Zurich, Switzerland, and is now with the Dept. of Electrical Engineering and Computer Science, University of California, Berkeley, CA

signal x can be described by finitely many delays and Doppler shifts:

$$y(t) = \sum_{j=1}^{S} b_j x(t - \bar{\tau}_j) e^{i2\pi\bar{\nu}_j t}. \tag{1.1}$$

Here, b_j is the attenuation factor corresponding to the delay-Doppler pair $(\bar{\tau}_j, \bar{\nu}_j)$. In radar imaging, for example, this input-output relation corresponds to a scene consisting of S moving targets modeled by point scatters, where the input x is the probing signal transmitted by the radar, and the output y is the superposition of the reflections of the probing signal by the point scatters. The relative distances and velocities of the targets can be obtained from the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$.

In order to identify the system H (e.g., to locate the targets in radar) we need to estimate the continuous delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ and the corresponding attenuation factors b_j from a single input-output measurement, i.e., from the response y to a known and suitably selected probing signal x. There are, however, important constraints on the type of input-output measurements that can be performed in practice: The probing signal x must be band-limited and approximately time-limited. Also, the response y can be observed only over a finite time interval. For concreteness, we assume that we observe the response y over an interval of length T and that x has bandwidth B and is approximately supported on a time interval of length proportional to T. This time- and band-limitation determines the "natural" resolution of the system, i.e., the accuracy up to which the delay-Doppler pairs can be identified is proportional to $1/B$ and $1/T$ in τ- and ν-directions, respectively. This resolution is achieved by a standard pulse-Doppler radar that performs digital matched filtering in order to detect the delay-Doppler pairs.

From (1.1), it is evident that band- and approximate time-limitation of x implies that y is band- and approximately time-limited as well—provided that the delay-Doppler pairs are compactly supported. In radar, due to path loss and finite velocity of the targets or objects in the scene this is indeed the case [44]. Throughout, we will therefore assume that $(\bar{\tau}_j, \bar{\nu}_j) \in [-T/2, T/2] \times [-B/2, B/2]$. This is not a restrictive assumption since it only says that the delays and Doppler shifts must be smaller than the effective support of the probing signal in time and frequency, respectively. Note that the region in the (τ, ν)-plane where the delay-Doppler pairs may be located can have area as large as $BT \gg 1$. For certain applications, it is reasonable to assume that the system is *underspread*, i.e., that the delay Doppler pairs lie in a region of area $\ll 1$ [47, 2, 1]. We do not need the underspread assumption in this paper.

Since y is band-limited and approximately time-limited, by the $2WT$-Theorem [41, 17], it is essentially characterized by on the order of BT coefficients. We therefore sample y in the interval $[-T/2, T/2]$ at rate $1/B$, so as to collect $L \coloneqq BT$ samples[1]. Furthermore, we choose x partially periodic by taking its samples $x_\ell = x(\ell/B)$ to be L-periodic for $3L$ many samples, and zero otherwise, so that x is essentially supported on an interval of length $3T$. For the readers familiar with wireless communication, we point out that the partial periodization of x serves a similar purpose as the cyclic prefix used in OFDM systems. As detailed in Section 4, the corresponding samples $y_p \coloneqq y(p/B)$ in the interval $p/B \in [-T/2, T/2]$ are given by

$$y_p = \sum_{j=1}^{S} b_j [\mathscr{F}_{\nu_j} \mathscr{T}_{\tau_j} \mathbf{x}]_p, \quad p = -N, ..., N, \quad N \coloneqq \frac{L-1}{2}, \tag{1.2}$$

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

where

$$[\mathscr{T}_\tau \mathbf{x}]_p := \frac{1}{L}\sum_{k=-N}^{N}\left[\left(\sum_{\ell=-N}^{N} x_\ell e^{-i2\pi\frac{\ell k}{L}}\right)e^{-i2\pi k\tau}\right]e^{i2\pi\frac{pk}{L}} \quad \text{and} \quad [\mathscr{F}_\nu \mathbf{x}]_p := x_p e^{i2\pi p\nu}. \tag{1.3}$$

Here, we defined[2] the time-shifts $\tau_j := \bar{\tau}_j/T$ and frequency-shifts $\nu_j := \bar{\nu}_j/B$. Since $(\bar{\tau}_j, \bar{\nu}_j) \in [-T/2, T/2] \times [-B/2, B/2]$ we have $(\tau_j, \nu_j) \in [-1/2, 1/2]^2$. Since $\mathscr{T}_\tau \mathbf{x}$ and $\mathscr{F}_\nu \mathbf{x}$ are 1-periodic in τ and ν, we can assume in the remainder of the paper that $(\tau_j, \nu_j) \in [0,1]^2$. The operators $\mathscr{T}_\tau$ and $\mathscr{F}_\nu$ can be interpreted as fractional time and frequency shift operators in $\mathbb{C}^L$. If the (τ_j, ν_j) lie on a $(1/L, 1/L)$ grid, the operators $\mathscr{F}_\nu$ and $\mathscr{T}_\tau$ reduce to the "natural" time frequency shift operators in $\mathbb{C}^L$, i.e., $[\mathscr{T}_\tau \mathbf{x}]_p = x_{p-\tau L}$ and $[\mathscr{F}_\nu \mathbf{x}]_p = x_p e^{i2\pi p\frac{\nu L}{L}}$. The definition of a time shift in (1.3) as taking the Fourier transform, modulating the frequency, and taking the inverse Fourier transform is a very natural definition of a *continuous* time-shift $\tau_j \in [0,1]$ of a *discrete* signal $\mathbf{x} = [x_0, ..., x_{L-1}]^T$. Finally note that to obtain (1.2) (see Section 4) from (1.1), we approximate a periodic sinc function with a finite sum of sinc functions (this is where partial periodization of x becomes relevant). Thus (1.2) does not hold exactly if we take the probing signal to be essentially time-limited. However, in Section 4 we show that the incurred relative error decays as $1/\sqrt{L}$ and is therefore negligible for large L. The numerical results in Section 5.2 indeed confirm that this error is negligible. If we took x to be T-periodic on $\mathbb{R}$, (1.2) becomes exact, but at the cost of x not being time-limited.

The problem of identifying the system H with input-output relation (1.1) under the constraints that the probing signal x is band-limited and the response to the probing signal $y = Hx$ is observed on a finite time interval now reduces to the estimation of the triplets (b_j, τ_j, ν_j) from the samples in (1.2). Motivated by this connection to the continuous system model, in this paper, we consider the problem of recovering the attenuation factors b_j and the corresponding time-frequency shifts $(\tau_j, \nu_j) \in [0,1]^2, j = 1, ..., S$, from the samples $y_p, p = -N, ..., N$, in (1.2). We call this the super-resolution radar problem, as recovering the exact time-frequency shifts (τ_j, ν_j) "breaks" the natural resolution "limit" of $(1/B, 1/T)$ achieved by standard pulse-Doppler radar.

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

1.1 *Time-frequency shifts on a grid*

If the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ lie on a $(\frac{1}{B}, \frac{1}{T})$ grid or equivalently if the time-frequency shifts (τ_j, ν_j) lie on a $(\frac{1}{L}, \frac{1}{L})$ grid, the super-resolution radar problem reduces to a sparse signal recovery problem with a Gabor measurement matrix. To see this, note that in this case $\tau_j L$ and $\nu_j L$ are integers in $\{0, ..., L-1\}$, and (1.2) reduces to

$$y_p = \sum_{j=1}^{S} b_j x_{p-\tau_j L} e^{i2\pi\frac{(\nu_j L)p}{L}}, \quad p = -N, ..., N. \tag{1.4}$$

Equation (1.4) can be written in matrix-vector form

$$\mathbf{y} = \mathbf{G}\mathbf{s}.$$

[2]To avoid ambiguity, we refer to $(\bar{\tau}_j, \bar{\nu}_j)$ as delay-Doppler pair and to (τ_j, ν_j) as time-frequency shift.

Here, $[\mathbf{y}]_p := y_p$, $\mathbf{G} \in \mathbb{C}^{L\times L^2}$ is the Gabor matrix with window $\mathbf{x}$, where the entry in the pth row and $(\tau_j L, \nu_j L)$-th column is $x_{p-\tau_j L} e^{i2\pi \frac{(\nu_j L)p}{L}}$, and $\mathbf{s} \in \mathbb{C}^{L^2}$ is a sparse vector where the j-th non-zero entry is given by b_j and is indexed by $(\tau_j L, \nu_j L)$. Thus, the recovery of the triplets (b_j, τ_j, ν_j) amounts to recovering the S-sparse vector $\mathbf{s} \in \mathbb{C}^{L^2}$ from the measurement vector $\mathbf{y} \in \mathbb{C}^L$. This is a sparse signal recovery problem with a Gabor measurement matrix. A—by now standard—recovery approach is to solve a convex ℓ_1-norm-minimization program. From [30, Thm. 5.1] we know that, provided the x_ℓ are i.i.d. sub-Gaussian random variables, and provided that $S \leqslant cL/(\log L)^4$ for a sufficiently small numerical constant c, with high probability, all S-sparse vectors $\mathbf{s}$ can be recovered from $\mathbf{y}$ via ℓ_1-minimization. Note that the result [30, Thm. 5.1] only applies to the Gabor matrix $\mathbf{G}$ and therefore does not apply to the super-resolution problem where the "columns" $\mathscr{F}_\nu \mathscr{T}_\tau \mathbf{x}$ are highly correlated.

1.2 *Only time or only frequency shifts*

Next, we consider the case of only time or only frequency shifts, and show that in both cases recovery of the (b_j, τ_j) and the (b_j, ν_j), is equivalent to the recovery of a weighted superposition of spikes from low-frequency samples. Specifically, if $\tau_j = 0$ for all j, (1.2) reduces to

$$y_p = x_p \sum_{j=1}^{S} b_j e^{i2\pi p \nu_j}, \quad p = -N, ..., N. \tag{1.5}$$

The y_p above are samples of a mixture of S complex sinusoids, and estimation of the (b_j, ν_j) corresponds to determining the magnitudes and the frequency components of these sinusoids. Estimation of the (b_j, ν_j) is known as a line spectral estimation problem, and can be solved using approaches such as Prony's method [21, Ch. 2]. Recently, an alternative approach for solving this problem has been proposed, specifically in [9] it is shown that exact recovery of the (b_j, ν_j) is possible by solving a convex total-variation norm minimization program. This results holds provided that the minimum separation between any two ν_j is larger than $2/N$. An analogous situation arises when there are only time shifts ($\nu_j = 0$ for all j) as taking the discrete Fourier transform of y_p yields a relation exactly of the form (1.5).

1.3 *Main contribution*

In this paper, we consider a random probing signal by taking the x_ℓ in (1.2) to be i.i.d. Gaussian (or sub-Gaussian) random variables. We show that with high probability, the triplets (b_j, τ_j, ν_j) can be recovered perfectly from the L samples y_p by (essentially) solving a convex program. This holds provided that two conditions are satisfied:

- *Minimum separation condition:* We assume the time-frequency shifts $(\tau_j, \nu_j) \in [0,1]^2$, $j = 1, ..., S$, satisfy the minimum separation condition

$$\max(|\tau_j - \tau_{j'}|, |\nu_j - \nu_{j'}|) \geqslant \frac{2.38}{N}, \text{ for all } j \neq j', \tag{1.6}$$

  where $|\tau_j - \tau_{j'}|$ is the wrap-around distance on the unit circle. For example, $|3/4 - 1/2| = 1/4$ but $|5/6 - 1/6| = 1/3 \neq 2/3$. Note that the time-frequency shifts must not be separated in both time *and* frequency, e.g., (1.6) can hold even when $\tau_j = \tau_{j'}$ for some $j \neq j'$.

- *Sparsity:* We also assume that the number of time-frequency shifts S obeys

$$S \leqslant c \frac{L}{(\log L)^3},$$

where c is a numerical constant.

This result is essentially optimal in terms of the allowed sparsity level, as the number S of unknowns can be linear—up to a log-factor—in the number of observations L. Even when we are given the time-frequency shifts (τ_j, ν_j), we can only hope to recover the corresponding attenuation factors $b_j, j = 1, ..., S$, by solving (1.2), provided that $S \leqslant L$.

We note that some form of separation between the time-frequency shifts is necessary for stable recovery. To be specific, we consider the simper problem of line spectral estimation (cf. Section 1.2) that is obtained from our setup by setting $\tau_j = 0$ for all j. Clearly, any condition necessary for the line spectral estimation problem is also necessary for the super-resolution radar problem. Consider an interval of the ν-axis of length $\frac{2S'}{L}$. If there are more than S' frequencies ν_j in this interval, then the problem of recovering the (b_j, ν_j) becomes extremely ill-posed when S' is large [15, Thm. 1.1], [35], [9, Sec. 1.7], [6]. Hence, in the presence of even a tiny amount of noise, stable recovery is not possible. The condition in (1.6) allows us to have $0.42\,S'$ time-frequency shifts in an interval of length $\frac{2S'}{L}$, an optimal number of frequencies up to the constant 0.42. We emphasize that while some sort of separation between the time-frequency shifts is necessary, the exact form of separation required in (1.6) may not be necessary for stable recovery and less restrictive conditions may suffice. Indeed, in the simpler problem of line spectral estimation (i.e., $\tau_j = 0$ for all j), Donoho [15] showed that stable super-resolution is possible via an exhaustive search algorithm even when condition (1.6) is violated locally as long as every interval of the ν-axis of length $\frac{2S'}{L}$ contains less than $S'/2$ frequencies ν_j and S' is small (in practice, think of $S' \lesssim 10$). The exhaustive search algorithm is infeasible in practice and an important open question in the theory of line spectral estimation is to develop a feasible algorithm that achieves the stability promised in [15]. In the special case when the b_j are real and positive, stable recovery can be achieved by convex optimization [35, 11, 20], see also [40] for recent results on more general positive linear combinations of waveforms.

Translated to the continuous setup, our result implies that with high probability we can identify the triplets $(b_j, \bar{\tau}_j, \bar{\nu}_j)$ perfectly provided that

$$|\bar{\tau}_j - \bar{\tau}_{j'}| \geqslant \frac{4.77}{B} \text{ or } |\bar{\nu}_j - \bar{\nu}_{j'}| \geqslant \frac{4.77}{T}, \quad \text{for all } j \neq j' \tag{1.7}$$

and $S \leqslant c\frac{BT}{(\log(BT))^3}$. Since we can exactly identify the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$, our result offers a significant improvement in resolution over conventional radar techniques. Specifically, with a standard pulse-Doppler radar that samples the received signal and performs digital matched filtering in order to detect the targets, the delay-Dopper shifts $(\bar{\tau}_j, \bar{\nu}_j)$ can only be estimated up to an uncertainty of about $(1/T, 1/B)$.

We hasten to add that in the radar literature, the term super-resolution is often used for the ability to resolve very close targets, specifically targets even closer than the Rayleigh resolution limit [38] that is proportional to $1/B$ and $1/T$ for delay and Doppler resolution, respectively. Our work permits identification of a *each* target with precision much higher than $1/B$ and $1/T$ as long as other targets are not too close, specifically other targets should be separated by a constant multiple of the Rayleigh resolution limit (cf. (1.7)).

Finally recall that $(\tau_j, \nu_j) \in [0,1]^2$ translates to $(\bar{\tau}_j, \bar{\nu}_j) \in [-T/2, T/2] \times [-B/2, B/2]$, i.e., the $(\bar{\tau}_j, \bar{\nu}_j)$ can lie in a rectangle of area $L = BT \gg 1$, i.e., the system H does not need to be underspread[3]. The ability to handle systems that are *overspread* is important in radar applications. Here, we might need

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

to resolve targets with large relative distances and relative velocities, resulting in delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ that lie in a region of area larger than 1 in the time-frequency plane.

We finally note that standard non-parametric estimation methods such as the MUSIC algorithm can in general *not* be directly applied to the super-resolution radar problem. The reason is that MUSIC requires *multiple* (sets of) measurements (snapshots) [43, Sec. 6.3], whereas we assume only a *single* (set of) measurement $\{y_p, p = -N,...,N\}$ to be available. However, by choosing the probing signal $\mathbf{x}$ in (1.2) to be periodic, the single measurement $\{y_p, p = -N,...,N\}$ can be transformed to multiple, smaller (sets of) measurements and MUSIC can then be applied. This approach, discussed in detail in Appendix H, has a few limitations. In particular, it requires the ν_j to be distinct, the time-shifts τ to lie in a significantly smaller range than $[0,1]$, $S < \sqrt{L}$, and appears to be significantly more sensitive to noise compared to our convex programming based approach. We note that when multiple sets of measurements are available, e.g., by observing distinct paths of a signal by an array of antennas, subspace methods such as MUSIC have been studied for delay-Doppler estimation [27].

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 $[\mathbf{x}]_q$ denotes its q-th entry, $\|\mathbf{x}\|_2$ its ℓ_2-norm and $\|\mathbf{x}\|_\infty = \max_q |x_q|$ its largest entry. For the matrix $\mathbf{A}$, $[\mathbf{A}]_{ij}$ designates the entry in its i-th row and j-th column, $\|\mathbf{A}\| := \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2$ its spectral norm, $\|\mathbf{A}\|_F := (\sum_{i,j} |[\mathbf{A}]_{ij}|^2)^{1/2}$ its Frobenius norm, and $\mathbf{A} \succeq 0$ signifies that $\mathbf{A}$ is positive semidefinite. The identity matrix is denoted by $\mathbf{I}$. For convenience, we will frequently use a two-dimensional index for vectors and matrices, e.g., we write $[\mathbf{g}]_{(k,\ell)}, k,\ell = -N,...,N$ for $\mathbf{g} = [g_{(-N,-N)}, g_{(-N,-N+1)}, \cdots, g_{(-N,N)}, g_{(-N+1,-N)}, \cdots, g_{(N,N)}]^T$. For a complex number b with polar decomposition $b = |b|e^{i2\pi\phi}$, $\operatorname{sign}(b) := e^{i2\pi\phi}$. Similarly, for a vector $\mathbf{b}$, $[\operatorname{sign}(\mathbf{b})]_k := \operatorname{sign}([\mathbf{b}]_k)$. For the set $\mathscr{S}$, $|\mathscr{S}|$ designates its cardinality and $\overline{\mathscr{S}}$ is its complement. The sinc-function is denoted as $\operatorname{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$. For vectors $\mathbf{r}, \mathbf{r}' \in [0,1]^2$, $|\mathbf{r} - \mathbf{r}'| = \max(|r_1 - r_1'|, |r_2 - r_2'|)$. Here, $|x - y|$ is the wrap-around distance on the unit circle between two scalars x and y. For example, $|3/4 - 1/2| = 1/4$ but $|5/6 - 1/6| = 1/3 \neq 2/3$. 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, $\mathscr{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2.

2. Recovery via convex optimization

In this section we present our approach to the recovery of the parameters (b_j, τ_j, ν_j) from the samples y_p in (1.2). Before we proceed we note that (1.2) can be rewritten as (see Appendix A for a detailed derivation)

$$y_p = \sum_{j=1}^{S} b_j \sum_{k,\ell=-N}^{N} D_N\left(\frac{\ell}{L} - \tau_j\right) D_N\left(\frac{k}{L} - \nu_j\right) x_{p-\ell} e^{i2\pi \frac{pk}{L}}, \quad p = -N,...,N, \tag{2.1}$$

where

$$D_N(t) := \frac{1}{L} \sum_{k=-N}^{N} e^{i2\pi t k} \tag{2.2}$$

is the Dirichlet kernel. We define atoms $\mathbf{a} \in \mathbb{C}^{L^2}$ as

$$[\mathbf{a}(\mathbf{r})]_{(k,\ell)} = D_N\left(\frac{\ell}{L} - \tau\right) D_N\left(\frac{k}{L} - \nu\right), \quad \mathbf{r} = [\tau, \nu]^T, \quad k, \ell = -N, ..., N. \tag{2.3}$$

Rewriting (2.1) in matrix-vector form yields

$$\mathbf{y} = \mathbf{G}\mathbf{z}, \quad \mathbf{z} = \sum_{j=1}^{S} |b_j| e^{i2\pi\phi_j} \mathbf{a}(\mathbf{r}_j), \quad \mathbf{r}_j = [\tau_j, \nu_j]^T,$$

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

$$[\mathbf{G}]_{p,(k,\ell)} \coloneqq x_{p-\ell} e^{i2\pi \frac{kp}{L}}, \quad k, \ell, p = -N, ..., N. \tag{2.4}$$

The signal $\mathbf{z}$ is a sparse linear combination of time and frequency shifted versions of the atom $\mathbf{a}(\mathbf{r})$. A regularizer that promotes such a sparse linear combination is the atomic norm induced by these signals [12]. The atoms in the set $\mathscr{A} \coloneqq \{e^{i2\pi\phi}\mathbf{a}(\mathbf{r}), \mathbf{r} \in [0,1]^2, \phi \in [0,1]\}$ are the building blocks of the signal $\mathbf{z}$. The atomic norm $\|\cdot\|_{\mathscr{A}}$ is defined as

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

where $\mathrm{conv}(\mathscr{A})$ denotes the convex hull of the set $\mathscr{A}$. The atomic norm can enforce sparsity in $\mathscr{A}$ because low-dimensional faces of $\mathrm{conv}(\mathscr{A})$ correspond to signals involving only a few atoms [12, 46]. A natural algorithm for estimating $\mathbf{z}$ is the atomic norm minimization problem [12]

$$\mathrm{AN}(\mathbf{y}) \colon \quad \underset{\tilde{\mathbf{z}}}{\text{minimize}} \|\tilde{\mathbf{z}}\|_{\mathscr{A}} \ \text{ subject to } \ \mathbf{y} = \mathbf{G}\tilde{\mathbf{z}}. \tag{2.5}$$

Once we obtain $\mathbf{z}$, the recovery of the time-frequency shifts is a 2D line spectral estimation problem that can be solved with standard approaches such as Prony's method, see e.g. [21, Ch. 2]. In Section 6.2, we will present a more direct approach for recovering the continuous time-frequency shifts $\mathbf{r}_j$, and in Section 5 we show that the time-frequency shifts can be recovered on an arbitrarily fine grid via ℓ_1-minimization. When the time-frequency shifts $\mathbf{r}_j = [\tau_j, \nu_j]^T$ are identified, the coefficients b_j can be obtained by solving the linear system of equations

$$\mathbf{y} = \sum_{j=1}^{S} b_j \mathscr{F}_{\nu_j} \mathscr{T}_{\tau_j} \mathbf{x}.$$

Computation of the atomic norm involves taking the infimum over infinitely many parameters and may appear to be daunting. For the case of only time or only frequency shifts (cf. Section 1.2), the atomic norm can be characterized in terms of linear matrix inequalities [46, Prop. 2.1]; this allows us to formulate the atomic norm minimization program as a semidefinite program that can be solved efficiently. The characterization [46, Prop. 2.1] relies on a classical Vandermonde decomposition lemma for Toeplitz matrices by Carathéodory and Fejér. While this lemma generalizes to higher dimensions [50, Thm. 1], this generalization fundamentally comes with a rank constraint on the corresponding Toeplitz matrix. This appears to prohibit a characterization of the atomic norm paralleling that of [46, Prop. 2.1] which explains why no semidefinite programming formulation of the atomic norm minimization problem (2.5)

is known, to the best of our knowledge. Nevertheless, based on [50, Thm. 1], one can obtain a semidefinite programming *relaxation* of AN($\mathbf{y}$).

Instead of taking that route, and explicitly stating the corresponding semidefinite program, we show in Section 6.2 that the time-frequency shifts $\mathbf{r}_j$ can be identified directly from the dual solution of the atomic norm minimization problem AN($\mathbf{y}$) in (2.5), and propose a semidefinite programming *relaxation* that allows us to find a solution of the dual efficiently.

3. Main result

Our main result, stated below, provides conditions guaranteeing that the solution to AN($\mathbf{y}$) in (2.5) is $\mathbf{z}$ (perfect recovery). As explained in Section 2, from $\mathbf{z}$ we can obtain the triplets (b_j, τ_j, ν_j) easily.

THEOREM 3.1 Assume that the samples of the probing signal $x_\ell, \ell = -N, ..., N$, are i.i.d. $\mathcal{N}(0, 1/L)$ random variables, $L = 2N+1$. Let $\mathbf{y} \in \mathbb{C}^L$, with $L \geqslant 1024$, contain the samples of the output signal obeying the input-output relation (1.2), i.e.,

$$\mathbf{y} = \sum_{j=1}^{S} b_j \mathscr{F}_{\nu_j} \mathscr{T}_{\tau_j} \mathbf{x} = \mathbf{G}\mathbf{z}, \quad \mathbf{z} = \sum_{\mathbf{r}_j \in \mathscr{S}} b_j \mathbf{a}(\mathbf{r}_j)$$

where $\mathbf{G}$ is the Gabor matrix of time-frequency shifts of the input sequence x_ℓ defined in (2.4). Assume that the sign(b_j) are i.i.d. uniform on $\{-1, 1\}$ and that the set of time-frequency shifts $\mathscr{S} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_S\} \subset [0,1]^2$ obeys the minimum separation condition

$$\max(|\tau_j - \tau_{j'}|, |\nu_j - \nu_{j'}|) \geqslant \frac{2.38}{N} \text{ for all } [\tau_j, \nu_j], [\tau_{j'}, \nu_{j'}] \in \mathscr{S} \text{ with } j \neq j'. \tag{3.1}$$

Furthermore, choose $\delta > 0$ and assume

$$S \leqslant c \frac{L}{(\log(L^6/\delta))^3},$$

where c is a numerical constant. Then, with probability at least $1 - \delta$, $\mathbf{z}$ is the unique minimizer of AN($\mathbf{y}$) in (2.5).

Recall that the complex-valued coefficients b_j in (1.1) in the radar model describe the attenuation factors. Therefore, it is natural to assume that the phases of different b_j are independent from each other and are uniformly distributed on the unit circle of the complex plane. Indeed, in standard models in wireless communication and radar [4], the b_j are assumed to be complex Gaussian. To keep the proof simple, in Theorem 3.1 we assume that the b_j are real-valued. The assumption that the coefficients b_j have random sign is the real-valued analogue of the random phase assumption discussed above. Theorem 3.1 continues to hold for complex-valued b_j (only the constant 2.38 in (3.1) changes slightly). While this random sign assumption is natural for many applications, we believe that it is not necessary for our result to hold. Finally, we would like to point out that Theorem 3.1 continues to hold for sub-Gaussian sequences x_ℓ with trivial modifications to our proof.

The proof of Theorem 3.1 is based on analyzing the dual of AN($\mathbf{y}$). Specifically, we will construct an appropriate dual certificate; the existence of this certificate guarantees that the solution to AN($\mathbf{y}$) in (2.5) is $\mathbf{z}$. This is a standard approach, e.g., in the compressed sensing literature, the existence of a related dual certificate guarantees that the solution to ℓ_1-minimization is exact [10]. The dual of AN($\mathbf{y}$)

in (2.5) is [8, Sec. 5.1.16]

$$\underset{\mathbf{q}}{\text{maximize}}\ \mathrm{Re}\,\langle \mathbf{q},\mathbf{y}\rangle \ \text{ subject to } \left\|\mathbf{G}^H\mathbf{q}\right\|_{\mathscr{A}^*} \leqslant 1, \tag{3.2}$$

where $\mathbf{q} = [q_{-N},...,q_N]^T \in \mathbb{C}^L$ and

$$\|\mathbf{v}\|_{\mathscr{A}^*} = \sup_{\|\mathbf{z}\|_{\mathscr{A}}\leqslant 1} \mathrm{Re}\,\langle \mathbf{v},\mathbf{z}\rangle = \sup_{\mathbf{r}\in[0,1]^2} |\langle \mathbf{v},\mathbf{a}(\mathbf{r})\rangle|$$

is the dual norm. Note that the constraint of the dual (3.2) can be rewritten as:

$$\left\|\mathbf{G}^H\mathbf{q}\right\|_{\mathscr{A}^*} = \sup_{\mathbf{r}\in[0,1]^2} |\langle \mathbf{q},\mathbf{G}\mathbf{a}(\mathbf{r})\rangle| = \sup_{[\tau,\nu]\in[0,1]^2} |\langle \mathbf{q},\mathscr{F}_\nu\mathscr{T}_\tau\mathbf{x}\rangle| \leqslant 1, \tag{3.3}$$

where we used $\mathbf{G}\mathbf{a}(\mathbf{r}) = \mathscr{F}_\nu\mathscr{T}_\tau\mathbf{x}$. By definition of the time and frequency shifts in (1.3) it is seen that $\langle \mathbf{q},\mathscr{F}_\nu\mathscr{T}_\tau\mathbf{x}\rangle$ is a 2D trigonometric polynomial (in τ,ν, see (8.2) for its specific form). The constraint in the dual is therefore equivalent to the requirement that the absolute value of a specific 2D trigonometric polynomial is bounded by one. A sufficient condition for the success of atomic norm minimization is given by the existence of a certain dual certificate of the form $\langle \mathbf{q},\mathscr{F}_\nu\mathscr{T}_\tau\mathbf{x}\rangle$. This is formalized by Proposition 3.2 below and is a consequence of strong duality. Strong duality is implied by Slater's conditions being satisfied [8, Sec. 5.2.3] (the primal problem only has equality constraints).

PROPOSITION 3.2 Let $\mathbf{y} = \sum_{j=1}^S b_j\mathscr{F}_{\nu_j}\mathscr{T}_{\tau_j}\mathbf{x} = \mathbf{G}\mathbf{z}$ with $\mathbf{z} = \sum_{\mathbf{r}_j\in\mathscr{S}} b_j\mathbf{a}(\mathbf{r}_j)$. If there exists a dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q},\mathscr{F}_\nu\mathscr{T}_\tau\mathbf{x}\rangle$with complex coefficients $\mathbf{q} = [q_{-N},...,q_N]^T$ such that

$$Q(\mathbf{r}_j) = \mathrm{sign}(b_j), \text{ for all } \mathbf{r}_j\in\mathscr{S}, \text{ and } |Q(\mathbf{r})| < 1 \text{ for all } \mathbf{r}\in[0,1]^2\setminus\mathscr{S} \tag{3.4}$$

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

The proof of Proposition 3.2 is standard, see e.g., [46, Proof of Prop. 2.4], and is provided in Appendix G for convenience of the reader. The proof of Theorem 3.1 consists of constructing a dual polynomial satisfying the conditions of Proposition 3.2, see Section 8.

4. Relationship between the continuous time and discrete time models

In this section we discuss in more detail how the discrete time model (1.2) follows from the continuous time model (1.1) through band- and time-limitations. This section is geared towards readers with interest in radar and wireless communication applications. Readers not interested in the detailed justification of (1.2) may wish to skip this section on a first reading. We start by explaining the effect of band- and time-limitation in continuous time. We show that (1.2) holds *exactly* when the probing signal is T-periodic, and holds *approximately* when the probing signal is essentially time-limited on an interval of length $3T$, as discussed in the introduction. Finally, we explicitly quantify the corresponding approximation error.

As mentioned previously, radar systems and wireless communication channels are typically modeled as linear systems whose response is a weighted superposition of delayed and Doppler-shifted versions of the probing signal. In general, the response $y = Hx$ of the system H to the probing signal x is given by

$$y(t) = \iint s_H(\tau,\nu)x(t-\tau)e^{i2\pi\nu t}d\nu d\tau, \tag{4.1}$$

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

$$s_H(\tau,\nu) = \sum_{j=1}^{S} b_j \delta(\tau-\bar{\tau}_j)\delta(\nu-\bar{\nu}_j). \tag{4.2}$$

Here, $b_j,\ j=1,...,S$, are (complex-valued) attenuation factors associated with the delay-Doppler pair $(\bar{\tau}_j,\bar{\nu}_j)$. With (4.2), (4.1) reduces to the input-output relation (1.1) stated in the introduction, i.e., to

$$y(t) = \sum_{j=1}^{S} b_j x(t-\bar{\tau}_j) e^{i2\pi\bar{\nu}_j t}.$$

4.1 *The effect of band- and time-limitations*

In practice, the probing signal x has finite bandwidth B and the received signal y can only be observed over a finite time interval of length T. We refer to this as time-limitation, even though y is non-zero outside of the time interval of length T. As shown next, this band- and time-limitation lead to a discretization of the input-output relation (4.1) and determines the "natural" resolution of the system of $1/B$ and $1/T$ in τ- and ν-directions, respectively. Specifically, using the fact that x is band-limited to $[-B/2,B/2]$, (4.1) can be rewritten in the form

$$y(t) = \sum_{k\in\mathbb{Z}}\sum_{\ell\in\mathbb{Z}} \overline{s_H}\left(\frac{\ell}{B},\frac{k}{T}\right) x\left(t-\frac{\ell}{B}\right) e^{i2\pi\frac{k}{T}t}, \tag{4.3}$$

with $t\in[-T/2,T/2]$. Here,

$$\overline{s_H}(\tau,\nu) := \iint s_H(\tau',\nu')\,\mathrm{sinc}((\tau-\tau')B)\,\mathrm{sinc}((\nu-\nu')T)\,d\tau' d\nu' \tag{4.4}$$

is a smeared version of the original spreading function. The relation (4.3) appears in [4]; for the sake of completeness, we detail the derivations leading to (4.3) in Appendix B. For point scatterers, i.e., for the spreading function in (4.2), (4.4) specializes to

$$\overline{s_H}(\tau,\nu) = \sum_{j=1}^{S} b_j\,\mathrm{sinc}((\tau-\bar{\tau}_j)B)\,\mathrm{sinc}((\nu-\bar{\nu}_j)T). \tag{4.5}$$

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 (4.2), 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 $(b_j,\bar{\tau}_j,\bar{\nu}_j),\ j=1,\ldots,S$, from $\overline{s_H}(\tau,\nu)$ is the classical 2D line spectral estimation problem. 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 (4.3).

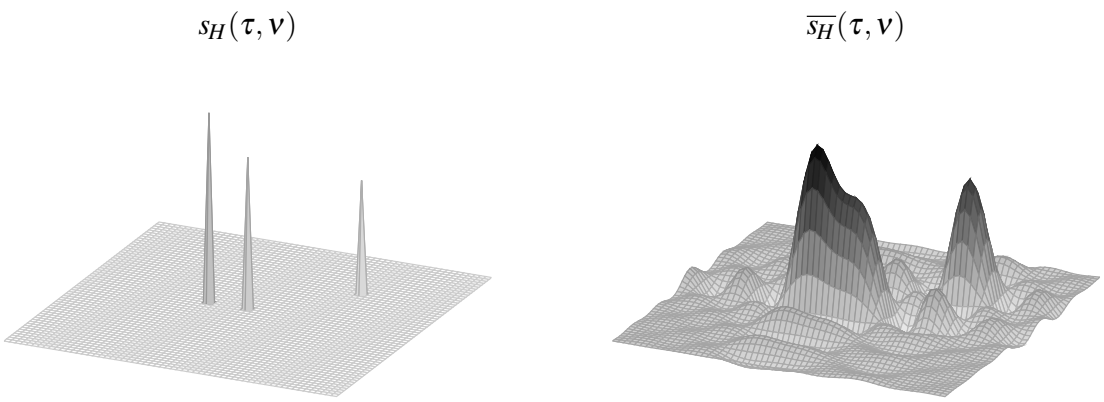

FIG. 1. Illustration of the spreading function $s_H(\tau,\nu)$ and the corresponding smeared spreading function $\overline{s_H}(\tau,\nu)$.

4.2 *Choice of probing signal*

Next, we consider a probing signal that is band-limited and essentially time-limited to an interval of length $3T$. To be concrete, we consider the signal

$$\tilde{x}(t) = \sum_{\ell=-L-N}^{L+N} x_\ell \operatorname{sinc}(tB-\ell) \tag{4.6}$$

where the coefficients x_ℓ are L-periodic, with the $x_\ell, \ell = -N,...,N$, i.i.d. $\mathcal{N}(0,1/L)$. A realization of the random signal $\tilde{x}(t)$ along with its Fourier transform is depicted in Figure 2. Since the sinc-kernel above is band-limited to $[-\frac{B}{2},\frac{B}{2}]$, $\tilde{x}$ is band-limited to $[-\frac{B}{2},\frac{B}{2}]$ as well. As the sinc-kernel decays relatively fast, $\tilde{x}$ is essentially supported on the interval $[-\frac{3T}{2},\frac{3T}{2}]$. We hasten to add that there is nothing fundamental about using the sinc-kernel to construct $\tilde{x}$ here; we choose it out of mathematical convenience, and could as well use a kernel that decays faster in time. For the readers familiar with wireless communication, we point out that the partial periodization of $\tilde{x}$ serves a similar purpose as the cyclic prefix used in OFDM systems.

4.3 *Sampling the output*

Recall that delay-Doppler pairs satisfy, by assumption, $(\bar{\tau}_j,\bar{\nu}_j) \in [-\frac{T}{2},\frac{T}{2}] \times [\frac{B}{2},\frac{B}{2}]$. Thus, $H\tilde{x}$ is band-limited to $[-B,B]$ and approximately time-limited to $[-2T,2T]$. According to the $2WT$-Theorem [41, 17], $H\tilde{x}$ has on the order of BT degrees of freedom so that one can think of $H\tilde{x}$ as having effective dimensionality on the order of BT. Therefore, by sampling $H\tilde{x}$ at rate $1/B$ in the interval $[-T/2,T/2]$, we collect the number of samples that matches the dimensionality of $H\tilde{x}$ up to a constant. The samples $\tilde{y}_p \coloneqq (H\tilde{x})(p/B)$ are given by

$$\tilde{y}_p = \sum_{k,\ell\in\mathbb{Z}} \overline{s_H}\left(\frac{\ell}{B},\frac{k}{T}\right) \tilde{x}\left(\frac{p-\ell}{B}\right) e^{i2\pi\frac{kp}{BT}}, \quad p=-N,...,N, \tag{4.7}$$

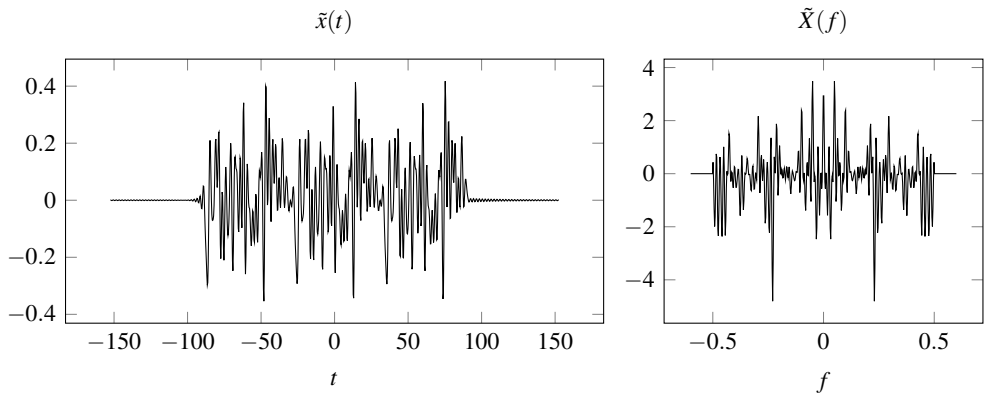

FIG. 2. The probing signal $\tilde{x}(t)$ and the real part of its Fourier transform $\tilde{X}(f)$ for $B = 1, T = 61, L = 61$: $\tilde{x}(t)$ is essentially time-limited on an interval of length $3T$ and band-limited to $[-B/2, B/2]$.

where $N = (L-1)/2$ and $L = BT$. Substituting (4.5) into (4.7) and defining $\tilde{x}_\ell := \tilde{x}(\ell/B)$ yields

$$\tilde{y}_p = \sum_{j=1}^{S} b_j \sum_{k,\ell\in\mathbb{Z}} \operatorname{sinc}(\ell - \bar{\tau}_j B)\operatorname{sinc}(k - \bar{\nu}_j T)\,\tilde{x}_{p-\ell} e^{i2\pi\frac{kp}{BT}}. \tag{4.8}$$

Next, we rewrite (4.8) in terms of the following equivalent expression of the Dirichlet kernel defined in (2.2)

$$D_N(t) = \sum_{k\in\mathbb{Z}} \operatorname{sinc}\left(L\,(t-k)\right), \tag{4.9}$$

and a truncated version

$$\tilde{D}_N(t) := \sum_{k=-1}^{1} \operatorname{sinc}\left(L\,(t-k)\right), \tag{4.10}$$

as

$$\tilde{y}_p = \sum_{k,\ell=-N}^{N} \sum_{j=1}^{S} b_j \tilde{D}_N\left(\frac{p-\ell}{L} - \tau_j\right) D_N\left(\frac{k}{L} - \nu_j\right) x_\ell e^{i2\pi\frac{kp}{L}}, \quad p = -N, ..., N, \tag{4.11}$$

where $\tau_j = \bar{\tau}_j/T$, $\nu_j = \bar{\nu}_j/B$, as before (see Appendix C for details). For $t \in [-1.5, 1.5]$, $\tilde{D}_N(t)$ is well approximated by $D_N(t)$, therefore (4.11) is well approximated by

$$y_p = \sum_{k,\ell=-N}^{N} \sum_{j=1}^{S} b_j D_N\left(\frac{\ell}{L} - \tau_j\right) D_N\left(\frac{k}{L} - \nu_j\right) x_{p-\ell} e^{i2\pi\frac{kp}{L}}, \quad p = -N, ..., N. \tag{4.12}$$

Here, y_p is equivalent to (1.2), as already mentioned in Section 2 and shown in Appendix A. Note that (4.12) can be viewed as the periodic equivalent of (4.3) (with $\overline{s_H}(\tau, \nu)$ given by (4.5)). We show in Appendix C that the discretization (4.12) can be obtained from (4.3) without any approximation if the probing signal $x(t)$ is chosen to be T-periodic with its samples selected as $x(\ell/B) = x_\ell$, for all $\ell \in \mathbb{Z}$.

Recall that the samples of the quasi-periodic probing signal satisfy $\tilde{x}(\ell/B) = x_\ell$ for $\ell \in [-N-L, L+N]$ and $\tilde{x}(\ell/B) = 0$ otherwise. Clearly, the periodic probing signal is not time-limited and, hence, can not be used in practice. The error we make by approximating $\tilde{y}_p$ with y_p is very small, as shown by the following proposition, proven in Appendix D. We also show numerically in Section 5.2 that in practice, the error we make by approximating $\tilde{y}_p$ with y_p is negligible.

PROPOSITION 4.1 Let the x_ℓ be i.i.d. $\mathcal{N}(0, 1/L)$ and let the sign of b_j be i.i.d. uniform on $\{-1, 1\}$. For all $\alpha > 0$, the difference between $\tilde{y}_p$ in (4.11) and y_p in (4.12) satisfies

$$\mathbb{P}\Big[|y_p - \tilde{y}_p| \geqslant c\frac{\alpha}{L}\|\mathbf{b}\|_2\Big] \leqslant (4 + 2L^2)e^{-\frac{\alpha}{2}},$$

where $\mathbf{b} = [b_1, ..., b_S]^T$ and c is a numerical constant.

Proposition 4.1 ensures that with high probability, the $\tilde{y}_p$ are very close to the y_p. Note that, under the conditions of Theorem 3.1, $\|\mathbf{y}\|_2 \approx \|\mathbf{b}\|_2, \mathbf{y} = [y_{-N}, ..., y_N]^T$, with high probability (not shown here). Since it follows from Proposition 4.1 that

$$\|\mathbf{y} - \tilde{\mathbf{y}}\|_2 \leqslant \frac{c_1}{\sqrt{L}}\|\mathbf{b}\|_2$$

holds with high probability, we can also conclude that $\frac{\|\mathbf{y}-\tilde{\mathbf{y}}\|_2}{\|\mathbf{y}\|_2} \leqslant \frac{c_2}{\sqrt{L}}, \tilde{\mathbf{y}} = [\tilde{y}_{-N}, ..., \tilde{y}_N]^T$ holds with high probability. That is, the relative error we make in approximating the $\tilde{y}_p$ with the y_p tends to zero in ℓ_2 norm as L grows.

5. Super-resolution radar on a grid

One approach to estimate the triplets (b_j, τ_j, ν_j) from the samples y_p in (1.2) is to suppose the time-frequency shifts lie on a *fine* grid, and solve the problem on that grid. When the time-frequency shifts do not exactly lie on a grid this leads to an error that becomes small as the grid becomes finer. In Section 5.2 below we study numerically the error incurred by this approach; for an analysis of the corresponding error in related problems see [45]. Our results have immediate consequences for the corresponding (discrete) sparse signal recovery problem, as we discuss next.

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

$$y_p = \sum_{m,n=0}^{K-1} s_{m,n}[\mathscr{F}_{m/K}\mathscr{T}_{n/K}\mathbf{x}]_p, \quad p = -N, ..., N. \tag{5.1}$$

$$= \sum_{m,n=0}^{K-1} \left(e^{i2\pi p\frac{m}{K}} \frac{1}{L} \sum_{\ell,k=-N}^{N} e^{-i2\pi k\frac{n}{K}} e^{i2\pi(p-\ell)\frac{k}{L}} x_\ell \right) s_{m,n}, \quad p = -N, ..., N. \tag{5.2}$$

To see the connection to the gridless setup in the previous sections, note that the recovery of the S-sparse (discrete) signal $s_{m,n}$ is equivalent to the recovery of the triplets (b_j, τ_j, ν_j) from the samples y_p in (1.2) if we assume that the (τ_j, ν_j) lie on a $(1/K, 1/K)$ grid (the non-zeros of $s_{m,n}$ correspond to the b_j). Writing the relation (5.2) 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 $\mathbf{R} \in \mathbb{C}^{L\times K^2}, K \geqslant L$, is the matrix with (m,n)-th column given by $\mathscr{F}_{m/K}\mathscr{T}_{n/K}\mathbf{x}$. The matrix $\mathbf{R}$ contains as columns "fractional" time-frequency shifts of the sequence x_ℓ. If $K = L$, $\mathbf{R}$ contains as columns only "whole" time-frequency shifts of the sequence x_ℓ and $\mathbf{R}$ is equal to the Gabor matrix $\mathbf{G}$ defined by (2.4). In this sense, $K = L$ is the natural grid (as noted before, see Section 1.1) and the ratio $\mathrm{SRF} := K/L$ can be interpreted as a super-resolution factor. The super-resolution factor determines by how much the $(1/K, 1/K)$ grid is finer than the natural $(1/L, 1/L)$ grid.

A standard approach to the recovery of 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:

$$\mathrm{L1}(\mathbf{y}): \quad \underset{\tilde{\mathbf{s}}}{\text{minimize}} \ \|\tilde{\mathbf{s}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{R}\tilde{\mathbf{s}}. \tag{5.3}$$

The following theorem is our main result for recovery on the fine grid.

THEOREM 5.1 Assume that the samples of the probing signal $x_\ell = -N, ..., N$, in (5.2) are i.i.d. $\mathscr{N}(0, 1/L)$ random variables, $L = 2N + 1$. Let $\mathbf{y} \in \mathbb{C}^L$, with $L \geqslant 1024$ be the samples of the output signal obeying the input-output relationship (5.2), i.e., $\mathbf{y} = \mathbf{R}\mathbf{s}$. Let $\mathscr{S} \subseteq \{0, ..., K-1\}^2$ be the support of the vector $[\mathbf{s}]_{(m,n)}$, $m, n = 0, ..., K-1$, and suppose that it satisfies the minimum separation condition

$$\min_{(m,n),(m',n')\in\mathscr{S}:\ (m,n)\neq(m',n')} \frac{1}{K}\max(|m-m'|, |n-n'|) \geqslant \frac{2.38}{N}.$$

Moreover, suppose that the non-zeros of $\mathbf{s}$ have random sign, i.e., $\mathrm{sign}([\mathbf{s}]_{(m,n)}), (m,n) \in \mathscr{S}$ are i.i.d. uniform on $\{-1, 1\}$. Choose $\delta > 0$ and assume

$$S \leqslant c\frac{L}{(\log(L^6/\delta))^3},$$

where c is a numerical constant. Then, with probability at least $1-\delta$, $\mathbf{s}$ is the unique minimizer of $\mathrm{L1}(\mathbf{y})$ in (5.3).

Note that Theorem 5.1 does not impose any restriction on K, in particular K and therefore the super-resolution factor $\mathrm{SRF} = K/L$ can be arbitrarily large. The proof of Theorem 5.1, presented in Appendix E, is closely linked to that of Theorem 3.1. As reviewed in Appendix E, the existence of a certain dual certificate guarantees that $\mathbf{s}$ is the unique minimizer of $\mathrm{L1}(\mathbf{y})$ in (5.3). The dual certificate is obtained directly from the dual certificate for the continuous case (i.e., from Proposition 8.1 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 (5.3) increases. The complexity of solving (5.3) can be managed as follows. First, the complexity of first-order convex optimization algorithms (such as TFOCS [3]) for solving (5.3) is dominated by multiplications with the matrices $\mathbf{R}$ and $\mathbf{R}^H$. Due to the structure of $\mathbf{R}$, those multiplications can be done very efficiently by utilizing the fast Fourier transform. Second, in practice we have $(\bar{\tau}_j, \bar{\nu}_j) \in [0, \tau_{\max}] \times [0, \nu_{\max}]$, which means that

$$(\tau_j, \nu_j) \in \left[0, \frac{\tau_{\max}}{T}\right] \times \left[0, \frac{\nu_{\max}}{B}\right]. \tag{5.4}$$

It is therefore sufficient to consider the restriction of $\mathbf{R}$ to the $\frac{\tau_{\max}\nu_{\max}K^2}{BT} = \tau_{\max}\nu_{\max}L \cdot \mathrm{SRF}^2$ many columns corresponding to the (τ_j, ν_j) satisfying (5.4). Since typically $\tau_{\max}\nu_{\max} \ll BT = L$, this results in a significant reduction of the problem size.

5.2 *Numerical results*

We next evaluate numerically the resolution obtained by our approach. We consider identification of the time-frequency shifts from the response to the (essentially) time-limited signal in (4.6), which corresponds to identification from the samples $\tilde{y}_p, p = -N, ..., N$ given by (4.11), without and with additive Gaussian noise. We also consider identification from the response to a signal with L-periodic samples $x(\ell/B) = x_\ell$, which corresponds to identification from the samples $y_p, p = -N, ..., N$ in (4.12). To account for additive noise, we solve the following modification of L1($\mathbf{y}$) in (5.3)

$$\text{L1-ERR}\colon \underset{\tilde{\mathbf{s}}}{\text{minimize}} \ \|\tilde{\mathbf{s}}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{R}\tilde{\mathbf{s}}\|_2^2 \leqslant \delta, \tag{5.5}$$

with δ chosen on the order of the noise variance. We choose $L = 201$, and for each problem instance, we draw $S = 10$ time-frequency shifts (τ_j, ν_j) uniformly at random from $[0, 2/\sqrt{201}]^2$, which amounts to drawing the corresponding delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ from $[0,2] \times [0,2]$. We use SPGL1 [48] to solve L1-ERR. The attenuation factors b_j corresponding to the time-frequency shifts (τ_j, ν_j) are drawn uniformly at random from the complex unit disc, independently across j. In Figure 3 we plot the average resolution error versus the super-resolution factor SRF $= K/L$. The resolution error is defined as the average over $j = 1, ..., S$ of $L\sqrt{(\hat{\tau}_j - \tau_j)^2 + (\hat{\nu}_j - \nu_j)^2}$, where the $(\hat{\tau}_j, \hat{\nu}_j)$ are the time-frequency shifts obtained by solving (5.5). There are three error sources incurred by this approach. The first is the gridding error obtained by assuming the points lie on a fine grid with grid constant $(1/K, 1/K)$. The second is the model error from approximating the $\tilde{y}_p$ in (4.11), obtained by sending an essentially time-limited probing signal (cf. (4.6)), with the y_p in (4.12), obtained by sending a truly periodic input signal $x(t)$. The third is the additive noise error. Note that the resolution attained at SRF $= 1$ corresponds to the resolution attained by matched filtering and by the compressive sensing radar architecture [25] discussed in Section 1.1. We see that for all SRF > 1, the resolution is significantly improved using our super-resolution radar approach. We furthermore observe that for low values of SRF, the gridding error dominates, while for large values of SRF, the additive noise error dominates. By looking at the noiseless case, it is seen that the gridding error decays as $1/\text{SRF}$, e.g., at SRF $= 20$, the error is about $0.4/20$. This demonstrates empirically that in practice solving the super-resolution radar problem on a fine grid is essentially as good as solving it on the continuum–provided the super-resolution factor is chosen sufficiently large. Finally, we observe that the model error is negligible, even for large signal-to-noise ratios (the curves in Figure 3 corresponding to the noiseless and the periodic case are indistinguishable).

6. Identification of time-frequency shifts

In this section we show that the time-frequency shifts can be obtained from a solution to the dual problem, and present a semidefinite programming relaxation to obtain a solution of the dual efficiently.

6.1 *Semidefinite programming relaxation of the dual*

Our relaxation is similar in spirit to related convex programs in [7, Sec. 3.1], [9, Sec. 4], and [46, Sec. 2.2]. We show that the constraint in the dual is equivalent to the requirement that the absolute value of a specific 2D trigonometric polynomial is bounded by one, and, therefore, this constraint can be formulated as a linear matrix inequality. The dimensions of the corresponding matrices are, however, unspecified. Choosing a certain relaxation degree for those matrices, and substituting the constraint in the dual with the corresponding matrix inequality leads to a semidefinite programming relaxation of the dual.

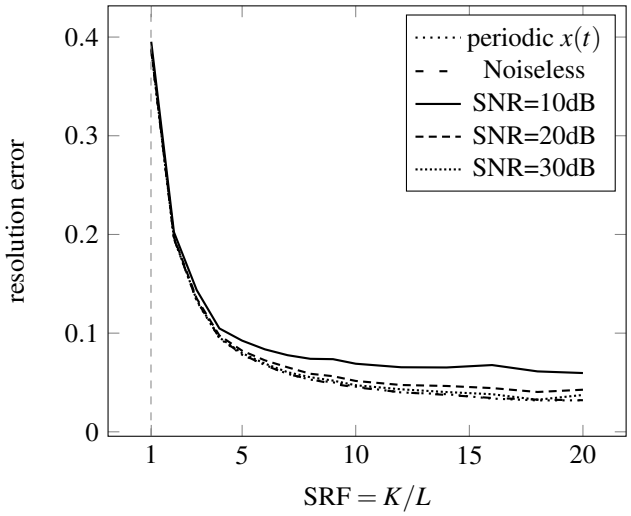

FIG. 3. Resolution error $L\sqrt{(\hat{\tau}_j-\tau_j)^2+(\hat{\nu}_j-\nu_j)^2}$ for the recovery of $S=10$ time-frequency shifts from the samples $y_p, p=-N,...,N$ in (4.12) (periodic input signal $x(t)$), and identification from the samples $\tilde{y}_p, p=-N,...,N$ in (4.11) (essentially time-limited input signal $\tilde{x}(t)$) with and without additive Gaussian noise n_p of a certain signal-to-noise ratio SNR $:=$ $\|[\tilde{y}_{-N},...,\tilde{y}_N]\|_2^2/\|[n_{-N},...,n_N]\|_2^2$, by solving L1-ERR.

Recall that the constraint of the dual program (3.2) is

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

By definition of the Dirichlet kernel in (2.2), the vector $\mathbf{a}(\mathbf{r})$ defined in (2.3) can be written as

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

where $\mathbf{F}^H$ is the (inverse) 2D discrete Fourier transform matrix with the entry in the (k,ℓ)-th row and (r,q)-th column given by $[\mathbf{F}^H]_{(k,\ell),(r,q)} := \frac{1}{L^2}e^{i2\pi\frac{qk+r\ell}{L}}$ and the entries of the vector $\mathbf{f}$ are given by $[\mathbf{f}(\mathbf{r})]_{(r,q)} := e^{-i2\pi(r\tau+q\nu)}$, $k,\ell,q,r=-N,...,N$, $\mathbf{r}=[\tau,\nu]^T$. With these definitions,

$$\left\langle \mathbf{G}^H\mathbf{q}, \mathbf{a}(\mathbf{r})\right\rangle = \left\langle \mathbf{G}^H\mathbf{q}, \mathbf{F}^H\mathbf{f}(\mathbf{r})\right\rangle = \left\langle \mathbf{F}\mathbf{G}^H\mathbf{q}, \mathbf{f}(\mathbf{r})\right\rangle = \sum_{r,q=-N}^{N} [\mathbf{F}\mathbf{G}^H\mathbf{q}]_{(r,q)} e^{i2\pi(r\tau+q\nu)}. \tag{6.1}$$

Thus, the constraint in the dual (3.2) says that the 2D trigonometric polynomial in (6.1) is bounded in magnitude by 1 for all $\mathbf{r}\in[0,1]^2$. The following form of the bounded real lemma allows us to approximate this constraint by a linear matrix inequality.

PROPOSITION 6.1 ([16, Cor. 4.25, p. 127]) Let P be a bivariate trigonometric polynomial in $\mathbf{r}=[\tau,\nu]^T$

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

If

$$\sup_{\mathbf{r}\in[0,1]^2} |P(\mathbf{r})| < 1$$

then there exists a matrix $\mathbf{Q} \succeq 0$ such that

$$\begin{bmatrix} \mathbf{Q} & \mathbf{p} \\ \mathbf{p}^H & 1 \end{bmatrix} \succeq \mathbf{0} \qquad \text{and} \qquad \forall k,\ell = -N,...,N, \quad \operatorname{trace}((\Theta_k \otimes \Theta_\ell)\mathbf{Q}) = \begin{cases} 1, & (k,\ell) = (0,0) \\ 0, & \text{otherwise} \end{cases} \tag{6.2}$$

where Θ_k designates the elementary Toeplitz matrix with ones on the k-th diagonal and zeros elsewhere. The vector $\mathbf{p}$ contains the coefficients $p_{(k,\ell)}$ of P as entries, and is padded with zeros to match the dimension of $\mathbf{Q}$.

Reciprocally, if there exists a matrix $\mathbf{Q} \succeq 0$ satisfying (6.2), then

$$\sup_{\mathbf{r}\in[0,1]^2} |P(\mathbf{r})| \leqslant 1.$$

Contrary to the corresponding matrix inequality for a 1D trigonometric polynomial of degree L [7, 9, 46], where the size of the matrix $\mathbf{Q}$ is fixed to $L \times L$, here, the size of the matrix $\mathbf{Q}$ is not determined and may, in principle, be significantly larger than the minimum size $L^2 \times L^2$. This stems from the fact that the sum-of-squares representation of a positive trigonometric polynomial of degree (L,L) possibly involves factors of degree larger than (L,L) (see, e.g., [16, Sec. 3.1]). Therefore, Proposition 6.1 only gives a sufficient condition and can not be used to *exactly* characterize the constraint of the dual program (3.2). Fixing the degree of the matrix $\mathbf{Q}$ to the minimum size of $L^2 \times L^2$ yields a *relaxation* of the constraint of the dual in (3.2) that leads to the following semidefinite programming *relaxation* of the dual program (3.2):

$$\underset{\mathbf{q},\mathbf{Q}\in\mathbb{C}^{L^2\times L^2},\mathbf{Q}\succeq 0}{\text{maximize}} \quad \operatorname{Re}\langle \mathbf{q}, \mathbf{y}\rangle \ \text{ subject to (6.4).} \tag{6.3}$$

$$\begin{bmatrix} \mathbf{Q} & \mathbf{F}\mathbf{G}^H\mathbf{q} \\ \mathbf{q}^H\mathbf{G}\mathbf{F}^H & 1 \end{bmatrix} \succeq \mathbf{0}, \quad \operatorname{trace}((\Theta_k \otimes \Theta_\ell)\mathbf{Q}) = \begin{cases} 1, & (k,\ell) = (0,0) \\ 0, & \text{otherwise} \end{cases}. \tag{6.4}$$

Note that we could also use a higher relaxation degree than (L,L), which would, in general, lead to a better approximation of the original problem. However, relaxations of minimal degree often yield optimal solutions in practice, as, e.g., observed in the related problem of 2D FIR filter design [16]. In Section 6.2, we report an example that shows that the relaxation of minimal degree also yields optimal solutions for our problem in practice.

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

Proposition 3.2 suggests that an estimate $\hat{\mathscr{S}}$ of the set of time-frequency shifts $\mathscr{S}$ can be obtained from a dual solution $\mathbf{q}$ by identifying the $\mathbf{r}_j$ for which the dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q}, \mathscr{T}_\nu \mathscr{T}_\tau \mathbf{x}\rangle = \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle$ has magnitude 1. In general, the solution $\mathbf{q}$ to (3.2) is not unique but we can ensure that

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

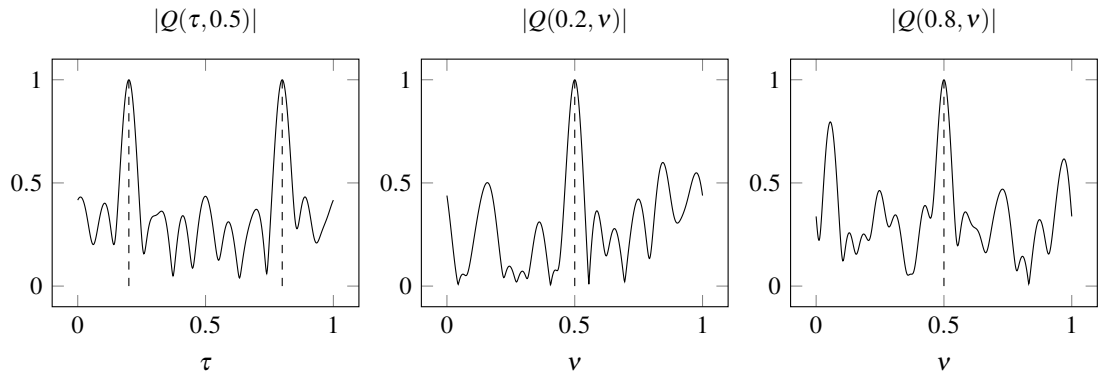

FIG. 4. Localization of the time-frequency shifts via the estimated dual polynomial $Q(\tau,\nu)$ obtained by solving (6.3) with noiseless measurements. The red lines show the actual positions of the time-frequency shifts located at $(0.2,0.5)$ and $(0.8,0.5)$. Note that the estimated dual polynomial satisfies $|Q(\tau,\nu)|=1$ if $(\tau,\nu)\in\{(0.2,0.5),(0.8,0.5)\}$ and $|Q(\tau,\nu)|<1$ otherwise, thereby providing accurate identification of the time-frequency shifts.

To see this assume that $\mathscr{S}\setminus\hat{\mathscr{S}}\neq\emptyset$. Then, we have

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

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

In general, we might have $\mathscr{S}\neq\hat{\mathscr{S}}$. However, (provided the solutions of (3.2) and its relaxation (6.3) coincide), in "most cases", standard semidefinite programming solvers (e.g., SDPT3) will yield a solution such that $\mathscr{S}=\hat{\mathscr{S}}$, see [46, Prop. 2.5] and [9, Sec. 4] for formal results on related problems.

We next provide a numerical example where the time-frequency shifts can be recovered perfectly from a solution to the semidefinite program (6.3). We choose $N=8$, consider the case of two time-frequency shifts, specifically $\mathscr{S}=\{(0.2,0.5),(0.8,0.5)\}$, and let the coefficients $x_\ell,\ell=-N,...,N$ and the $b_j,j=1,2$, be i.i.d. uniform on the complex unit sphere. In Figure 4 we plot the dual polynomial $Q(\mathbf{r})=\langle\mathbf{q},\mathbf{G}\mathbf{a}(\mathbf{r})\rangle$ with $\mathbf{q}$ obtained by solving (6.3) via CVX [22] (CVX calls the SDPT3 solver). It can be seen that the time-frequency shifts can be recovered perfectly, i.e., $\hat{\mathscr{S}}=\mathscr{S}$.

6.3 *Recovery in the noisy case*

In practice, the samples y_p in (1.2) are corrupted by additive noise. In that case, perfect recovery of the (b_j,τ_j,ν_j) is in general no longer possible, and we can only hope to identify the time-frequency shifts

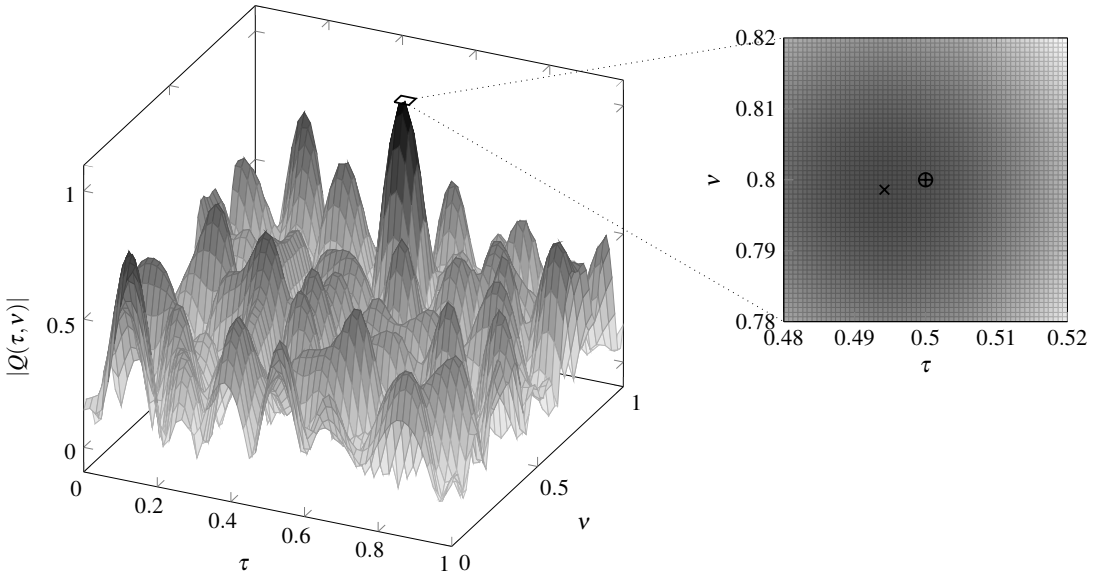

FIG. 5. Localization of the time-frequency shifts via the estimated dual polynomial $Q(\tau,\nu)$ obtained by solving (6.6) using noisy measurements (10dB noise). The estimated dual polynomial satisfies $|Q(\tau,\nu)| = 1$ for $(\tau,\nu) = (0.4942, 0.7986)$ (marked by $\times$); this is very close to the original time-frequency shift $(0.5, 0.8)$ (marked by $\oplus$).

up to an error. In the noisy case, we solve the following convex program:

$$\underset{\tilde{\mathbf{z}}}{\text{minimize}} \|\tilde{\mathbf{z}}\|_{\mathscr{A}} \;\text{ subject to }\; \|\mathbf{y} - \mathbf{G}\tilde{\mathbf{z}}\|_2 \leqslant \delta. \tag{6.5}$$

The semidefinite programing relaxation of the dual of (6.5) takes the form

$$\underset{\mathbf{q},\mathbf{Q}}{\text{maximize}} \operatorname{Re}\langle \mathbf{q}, \mathbf{y}\rangle - \delta\|\mathbf{q}\|_2 \text{ subject to (6.4)} \tag{6.6}$$

and we again estimate the time-frequency shifts $\mathbf{r}_j$ 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 visualized in 3D) and let the coefficients $x_\ell, \ell = -N, ..., N$ and b_1 be random variables that are i.i.d. uniform on the complex unit sphere but normalized to have variance $1/L$. We add i.i.d. complex Gaussian noise n_p to the samples samples y_p in (1.2), such that the signal-to-noise ratio (SNR), defined as $\text{SNR} = \|[y_{-N}, ..., y_N]\|_2^2 / \|[n_{-N}, ..., n_N]\|_2^2$, is 10dB (we express the SNR in decibels computed as $10\log_{10}(\text{SNR})$). In Figure 5 we plot the dual polynomial $Q(\mathbf{r}) = \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r})\rangle$ with $\mathbf{q}$ obtained by solving (6.6) (with $\delta = 0.8$) using CVX. The time-frequency shift for which the dual polynomial achieves magnitude 1 is $(0.4942, 0.7986)$; it is very close to the original time-frequency shift $(0.5, 0.8)$.

7. Relation to previous work

The general problem of extracting the spreading function $s_H(\tau,\nu)$ of a linear time varying system of the form (4.1) is known as system identification. It has been shown that LTV systems with spreading

function compactly supported on a known region of area Δ in the time-frequency plane are identifiable if and only if $\Delta \leqslant 1$ as shown by Kailath, Bello, Kozek, and Pfander in [28, 5, 29, 37]. If the spreading function's support region is unknown, a necessary and sufficient condition for identifiability is $\Delta \leqslant 1/2$ [24]. In contrast to our work, the input (probing) signal in [28, 5, 29, 37, 24] is not band-limited, and the response to the probing signal is not time-limited. In fact, the probing signal in those works is a (weighted) train of Dirac impulses and therefore decays neither in time nor in frequency. If it is necessary to use a band-limited probing signal and observe the response to the probing signal for a finite amount time (as is the case in practice), it follows from the theory developed in [28, 5, 29, 37, 24] that $s_H(\tau,\nu)$ can be identified with precision proportional to $1/B$ and $1/T$ in τ and ν directions, respectively. This is good enough if the function $s_H(\tau,\nu)$ is smooth in the sense that it does not vary (too) much at the scale $1/B$ and $1/T$ in τ and ν directions, respectively. The assumptions we make about $s_H(\tau,\nu)$ in this paper are very far from smoothness: Our $s_H(\tau,\nu)$ consists of Dirac delta functions, and hence is peaky. In this sense, the results in this paper are complementary to those in [28, 5, 29, 37, 24].

Tauböck et al. [47] and Bajwa et al. [2] considered the identification of LTV systems with spreading function compactly supported in a rectangle of area $\Delta \leqslant 1$. In [47, 2], the time-frequency shifts lie on a *coarse* grid (i.e., the grid in Section 1.1). In our setup, the time-frequency shifts need not lie on a grid and may in principle lie in a rectangle of area $L = BT$ that is considerably larger than 1. In a related direction, Herman and Strohmer [25] (in the context of compressed sensing radar) and Pfander et al. [36] considered the case where the time-frequency shifts lie on a coarse grid (see Section 1.1 for details).

Bajwa et al. [1] also considered the identification of an LTV of the form (1.1). The approach in [1] requires that the delay-Doppler pairs $(\bar{\tau}_j, \bar{\nu}_j)$ lie in a rectangle of area much smaller than 1 (this is required for a certain approximation in [1, Eq. 5] to be valid), and also requires in the worst case $(BT)^2 \geqslant cS$. Both assumption are not required here.

In [9], Candès and Fernandez-Granda study the recovery of the frequency components of a mixture of S complex sinusoids from L equally spaced samples as in (1.5) (see Section 1.2). As mentioned in Section 1.2, this corresponds to the case of only time or only frequency shifts. Recently, Fernandez-Granda [19] improved the main result of [9] by providing an tighter constant in the minimum separation condition. Tang et al. [46] study a related problem, namely the recovery of the frequency components from a random subset of the L equally spaced samples. Both [9, 46] study convex algorithms analogous to the algorithm studied here, and the proof techniques developed in these papers inspired the analysis presented here. In [42] the author improved the results of [9] with simpler proofs by building approximate dual polynomials. 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. Finally, we would like to mention a few recent papers that use algorithms not based on convex optimization for the super-resolution problem [14, 13, 18, 33, 34]. We should note that some of these approaches can handle smaller separation compared to convex optimization based approaches but the stability of these approaches to noise is not well understood, they do not as straightforwardly generalize to higher dimensions, and often need the model order (e.g., the number of frequencies) as input parameter. To the best of our knowledge, the approaches in [14, 13, 18, 33, 34] have not been generalized to the super-resolution radar problem studied in this paper.

8. Construction of the dual polynomial

In this section we prove Theorem 3.1 by constructing a dual polynomial that satisfies the conditions of Proposition 3.2. Existence of the dual polynomial is guaranteed by the following proposition, which is the main technical result of this paper.

PROPOSITION 8.1 Assume that the samples of the probing signal $x_\ell, \ell = -N, ..., N$, are i.i.d. $\mathcal{N}(0, 1/L)$ random variables and $L = 2N+1 \geqslant 1024$. Let $\mathscr{S} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_S\} \subset [0,1]^2$ be an arbitrary set of points obeying the minimum separation condition

$$\max(|\tau_j - \tau_{j'}|, |\nu_j - \nu_{j'}|) \geqslant \frac{2.38}{N} \text{ for all } [\tau_j, \nu_j], [\tau_{j'}, \nu_{j'}] \in \mathscr{S} \text{ with } j \neq j'. \tag{8.1}$$

Let $\mathbf{u} \in \{-1,1\}^S$ be a random vector, whose entries are i.i.d. and uniform on $\{-1,1\}$. Choose $\delta > 0$ and assume

$$S \leqslant c \frac{L}{\log^3\left(\frac{L^6}{\delta}\right)}$$

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

$$Q(\mathbf{r}) = \langle \mathbf{q}, \mathscr{F}_\nu \mathscr{T}_\tau \mathbf{x} \rangle = \sum_{p=-N}^{N} [\mathscr{F}_\nu \mathscr{T}_\tau \mathbf{x}]_p^* q_p = \sum_{k,\ell=-N}^{N} \underbrace{\left(\frac{1}{L} x_\ell^* \sum_{p=-N}^{N} e^{i2\pi(p-\ell)\frac{k}{L}} q_p \right)}_{q_{k,\ell} :=} e^{-i2\pi(k\tau + p\nu)} \tag{8.2}$$

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

$$Q(\mathbf{r}_j) = u_j, \text{ for all } \mathbf{r}_j \in \mathscr{S}, \text{ and } |Q(\mathbf{r})| < 1 \text{ for all } \mathbf{r} \in [0,1]^2 \setminus \mathscr{S}. \tag{8.3}$$

We provide a proof of Proposition 8.1 by constructing $Q(\mathbf{r})$ explicitly. Our construction of the polynomial $Q(\mathbf{r})$ is inspired by that in [9, 46]. To built $Q(\mathbf{r})$ we need to construct a 2D-trigonometric polynomial that satisfies (8.3), and whose coefficients $q_{k,\ell}$ are constraint to be of the form (8.2). It is instructive to first consider the construction of a 2D trigonometric polynomial

$$\bar{Q}(\mathbf{r}) = \sum_{k,\ell=-N}^{N} \bar{q}_{k,\ell} e^{-i2\pi(k\tau + \ell\nu)} \tag{8.4}$$

satisfying (8.3) without any constraint on the coefficients $\bar{q}_{k,\ell}$. To this end, we next review the construction in [9] establishing that there exists a 2D trigonometric polynomial $\bar{Q}$ satisfying simultaneously the interpolation condition $\bar{Q}(\mathbf{r}_j) = u_j$, for all $\mathbf{r}_j \in \mathscr{S}$, and the boundedness condition $\left|\bar{Q}(\mathbf{r})\right| < 1$, for all $\mathbf{r} \notin \mathscr{S}$, provided the minimum separation condition (8.1) holds. In order to construct $\bar{Q}$, Candès and Fernandez-Granda [9] interpolate the points u_j with a fast-decaying kernel $\bar{G}$ and its partial derivatives according to

$$\bar{Q}(\mathbf{r}) = \sum_{k=1}^{S} \bar{\alpha}_k \bar{G}(\mathbf{r} - \mathbf{r}_k) + \bar{\beta}_{1k} \bar{G}^{(1,0)}(\mathbf{r} - \mathbf{r}_k) + \bar{\beta}_{2k} \bar{G}^{(0,1)}(\mathbf{r} - \mathbf{r}_k). \tag{8.5}$$

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

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

For N even, the Fejér kernel is a trigonometric polynomial of degree $N/2$. It follows that $F(t)$ is a trigonometric polynomial of degree N (with certain coefficients g_k), i.e.,

$$F(t) = \frac{1}{M} \sum_{k=-N}^{N} g_k e^{i2\pi tk}. \tag{8.6}$$

Since shifted versions of $F(t)$ and the derivatives of $F(t)$ are also 1D trigonometric polynomials of degree N, it follows that the kernel $\bar{G}$, its partial derivatives, and shifted versions thereof, are 2D trigonometric polynomials of the form (8.4). Since $F(t)$ decays rapidly around the origin $t = 0$ ($F(0) = 1$), $\bar{G}(\mathbf{r})$ decays rapidly around the origin $\mathbf{r} = \mathbf{0}$ as well. To ensure that $\bar{Q}(\mathbf{r})$ reaches local maxima, which is necessary for the interpolation and boundedness conditions to hold simultaneously, the coefficients $\bar{\alpha}_k, \bar{\beta}_{1k}$ and $\bar{\beta}_{2k}$ are chosen in a specific way guaranteeing that

$$\bar{Q}(\mathbf{r}_k) = u_k, \quad \bar{Q}^{(1,0)}(\mathbf{r}_k) = 0, \text{ and } \bar{Q}^{(0,1)}(\mathbf{r}_k) = 0, \text{ for all } \mathbf{r}_k \in \mathscr{S}, \tag{8.7}$$

where $\bar{Q}^{(m,n)}(\mathbf{r}) \coloneqq \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} \bar{Q}(\mathbf{r})$. The idea of this construction is to interpolate the u_j with the functions $\bar{G}(\mathbf{r} - \mathbf{r}_j)$ (the α_k are in fact close to the u_j) and to slightly adopt this interpolation near the $\mathbf{r}_j$ with the functions $\bar{G}^{(1,0)}(\mathbf{r} - \mathbf{r}_j)$, $\bar{G}^{(0,1)}(\mathbf{r} - \mathbf{r}_j)$ to ensure that local maxima are achieved at the $\mathbf{r}_j$ (the β_{1k}, β_{2k} are in fact very small). The key properties of the interpolating functions used in this construction are that $G(\mathbf{r} - \mathbf{r}_j)$ decays fast around $\mathbf{r}_j$, as this enables a good constant in the minimum separation condition (8.1), and that the "correction" functions $\bar{G}^{(1,0)}(\mathbf{r} - \mathbf{r}_j)$, $\bar{G}^{(0,1)}(\mathbf{r} - \mathbf{r}_j)$ are small at $\mathbf{r}_j$, but sufficiently large in a small region relatively close to $\mathbf{r}_j$, and decay fast far from $\mathbf{r}_j$ as well.

A first difficulty with generalizing this idea to the case where the coefficients $q_{k,\ell}$ of the 2D trigonometric polynomial have the special form (8.2) is this: Since $\mathbf{x}$ is a random vector our interpolation and correction functions are naturally non-deterministic and thus showing the equivalent to (8.7), namely (8.9), requires a probabilistic analysis. Our analysis therefore relies on concentration of measure results. A second difficulty is that interpolating the points u_j with shifted versions of a *single* function will not work, as shifted versions of a function of the special form (8.2), playing the role of $\bar{G}$, are in general not of the form (8.2) (the time and frequency shift operators $\mathscr{T}_\tau$ and $\mathscr{F}_\nu$ do not commute). As a result we have to work with different interpolating functions for different $\mathbf{r}_j$'s. These function reach their maxima at or close to the $\mathbf{r}_j$'s. A third difficulty is that we cannot simply use the derivatives of our interpolation functions as "correction" functions, because the derivatives of a polynomial of the form (8.2) are in general not of the form (8.2).

We will construct the polynomial $Q(\mathbf{r})$ by interpolating the points $(\mathbf{r}_k, u_k)$ with functions $G_{(m,n)}(\mathbf{r}, \mathbf{r}_k)$, $m, n = 0, 1$ that have the form (8.2):

$$Q(\mathbf{r}) = \sum_{k=1}^{S} \alpha_k G_{(0,0)}(\mathbf{r}, \mathbf{r}_k) + \beta_{1k} G_{(1,0)}(\mathbf{r}, \mathbf{r}_k) + \beta_{2k} G_{(0,1)}(\mathbf{r}, \mathbf{r}_k). \tag{8.8}$$

Choosing the functions $G_{(m,n)}(\mathbf{r}, \mathbf{r}_k)$ in (8.8) to be of the form (8.2) ensures that $Q(\mathbf{r})$ itself is of the form (8.2). We will show that, with high probability, there exists a choice of coefficients $\alpha_k, \beta_{1k}, \beta_{2k}$ such that

$$Q(\mathbf{r}_k) = u_k, \quad Q^{(1,0)}(\mathbf{r}_k) = 0, \text{ and } Q^{(0,1)}(\mathbf{r}_k) = 0, \text{ for all } \mathbf{r}_k \in \mathscr{S}, \tag{8.9}$$

where $Q^{(m,n)}(\mathbf{r}) \coloneqq \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} Q(\mathbf{r})$. This ensures that $Q(\mathbf{r})$ reaches local maxima at the $\mathbf{r}_k$. We will then show that with this choice of coefficients, the resulting polynomial satisfies $|Q(\mathbf{r})| < 1$ for all $\mathbf{r} \notin \mathscr{S}$.

Now that we have stated our general strategy, we turn our attention to the construction of the interpolating and correction functions $G_{(m,n)}(\mathbf{r},\mathbf{r}_k)$. We will chose $G_{(0,0)}(\mathbf{r},\mathbf{r}_k)$ such that its peak is close to $\mathbf{r}_k$ and such that it decays fast around $\mathbf{r}_k$. For $G_{(0,0)}(\mathbf{r},\mathbf{r}_k)$ to have the form (8.2) it must be random (due to $\mathbf{x}$); we will show that the properties just mentioned are satisfied with high probability. We start by recalling (cf. Section 6)

$$\mathscr{F}_\nu \mathscr{T}_\tau \mathbf{x} = \mathbf{G}\mathbf{F}^H \mathbf{f}(\mathbf{r}),$$

where $\mathbf{G}$ is the Gabor matrix (cf. 2.4). Here, $\mathbf{F}^H$ is the inverse 2D discrete Fourier transform matrix with the entry in the (k,ℓ)-th row and (r,q)-th column given by $[\mathbf{F}^H]_{(k,\ell),(r,q)} \coloneqq \frac{1}{L^2} e^{i2\pi \frac{qk+r\ell}{L}}$ and $[\mathbf{f}(\mathbf{r})]_{(r,q)} = e^{-i2\pi(r\tau+q\nu)}$ with $k,\ell,q,r = -N,...,N$. Next, define the vector $\mathbf{g}_{(m,n)}(\mathbf{r}_j) \in \mathbb{C}^{L^2}$ as

$$[\mathbf{g}_{(m,n)}(\mathbf{r}_j)]_{(r,q)} = g_r g_q e^{-i2\pi(\tau_j r + \nu_j q)} (i2\pi r)^m (i2\pi q)^n, \quad r,q = -N,...,N, \quad \mathbf{r}_j = [\tau_j,\nu_j]^T.$$

Here, the g_r are the coefficients of the squared Fejér kernel in (8.6). With this notation, we define

$$\begin{aligned} G_{(m,n)}(\mathbf{r},\mathbf{r}_j) &\coloneqq \frac{L^2}{M^2} \left\langle \mathbf{G}\mathbf{F}^H \mathbf{g}_{(m,n)}(\mathbf{r}_j), \mathscr{F}_\nu \mathscr{T}_\tau \mathbf{x} \right\rangle \\ &= \frac{L^2}{M^2} \mathbf{f}^H(\mathbf{r}) \mathbf{F}\mathbf{G}^H \mathbf{G}\mathbf{F}^H \mathbf{g}_{(m,n)}(\mathbf{r}_j). \end{aligned} \tag{8.10}$$

By identifying $\mathbf{q}$ in (8.2) with $\mathbf{G}\mathbf{F}^H \mathbf{g}_{(m,n)}(\mathbf{r}_j)$, we immediately see that $G_{(m,n)}(\mathbf{r},\mathbf{r}_j)$ and in turn $Q(\mathbf{r})$ has the form (8.2), as desired. The particular choice of $G_{(m,n)}(\mathbf{r},\mathbf{r}_j)$ is motivated by the fact—made precise later—that $G_{(m,n)}(\mathbf{r},\mathbf{r}_j)$ concentrates around the deterministic function $G^{(m,n)}(\mathbf{r}-\mathbf{r}_j)$ of (8.5). Thus, $G_{(0,0)}(\mathbf{r},\mathbf{r}_j)$ decays rapidly around $\mathbf{r}_j$ with high probability. To demonstrate this empirically, in Figure 6 we plot $\bar{G}(\mathbf{r})$ and $G_{(0,0)}(\mathbf{r},\mathbf{0})/G_{(0,0)}(\mathbf{0},\mathbf{0})$ for $N = 60$ and $N = 300$. Note that close to $\mathbf{r} = \mathbf{0}$, the random function $G_{(0,0)}(\mathbf{r},\mathbf{0})$ and the deterministic kernel $\bar{G}(\mathbf{r})$ are very close. A simple calculation shows that the expected value of $G_{(m,n)}(\mathbf{r},\mathbf{r}_j)$ with respect to $\mathbf{x}$ is equal to $G^{(m,n)}(\mathbf{r}-\mathbf{r}_j)$. Specifically, as shown later on in Section 8.1, $\mathbb{E}[\mathbf{G}^H\mathbf{G}] = \mathbf{I}$. This immediately implies that

$$\begin{aligned} \mathbb{E}[G_{(m,n)}(\mathbf{r},\mathbf{r}_j)] &= \frac{1}{M^2} \mathbf{f}^H(\mathbf{r}) \mathbf{g}_{(m,n)}(\mathbf{r}_j) \\ &= \frac{1}{M^2} \sum_{r,q=-N}^{N} e^{i2\pi(r\tau+q\nu)} g_r g_q e^{-i2\pi(\tau_j r+\nu_j q)} (i2\pi r)^m (i2\pi q)^n \\ &= \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} \frac{1}{M^2} \sum_{r,q=-N}^{N} e^{i2\pi(r(\tau-\tau_j)+q(\nu-\nu_j))} g_r g_q \\ &= \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} F(\tau-\tau_j) F(\nu-\nu_j) \\ &= \bar{G}^{(m,n)}(\mathbf{r}-\mathbf{r}_j). \end{aligned} \tag{8.11}$$

The remainder of the proof is organized as follows.

Step 1: We will show that for every $\mathbf{r} \in [0,1]^2$ the function $G_{(0,0)}(\mathbf{r},\mathbf{r}_j)$ is close to $\bar{G}(\mathbf{r}-\mathbf{r}_j)$ with high probability, i.e., $|G_{(0,0)}(\mathbf{r},\mathbf{r}_j) - \bar{G}(\mathbf{r}-\mathbf{r}_j)|$ is small.

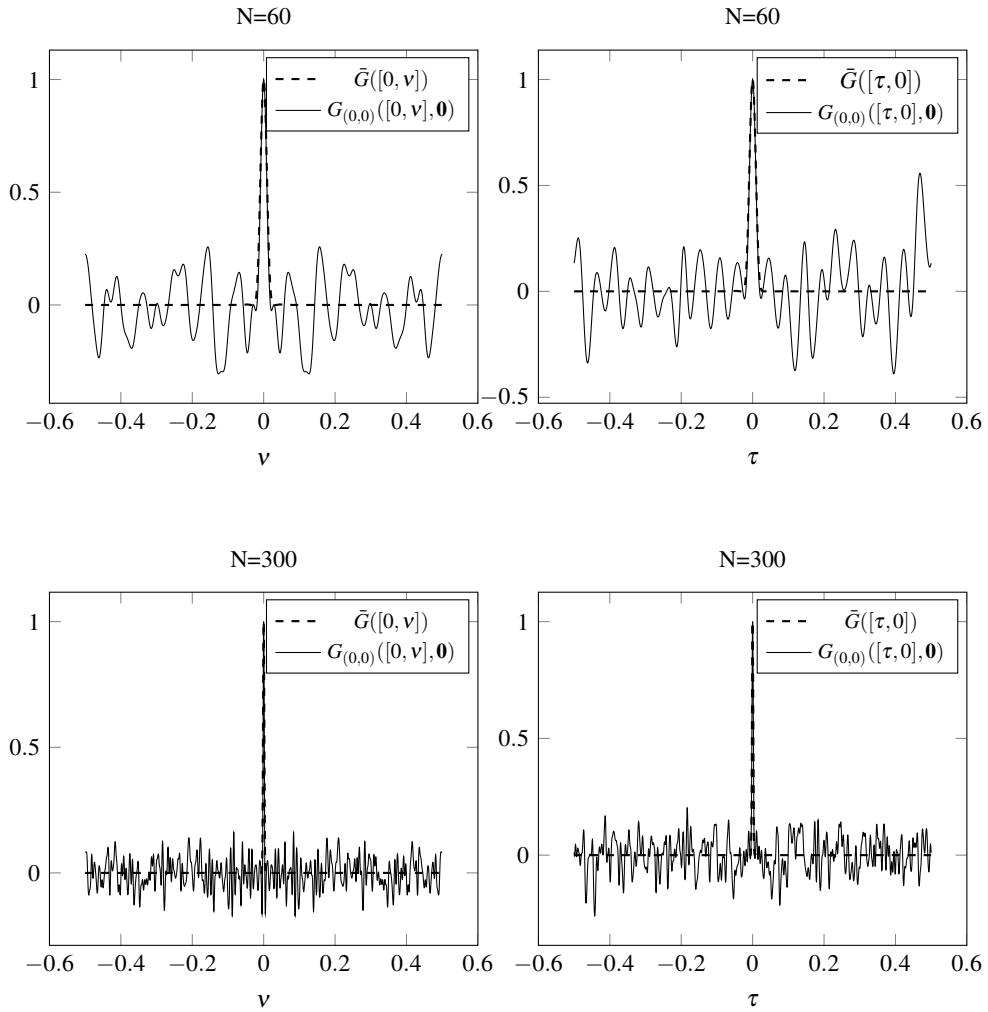

FIG. 6. Plots of the random kernel $G_{(0,0)}(\mathbf{r},\mathbf{0})/G_{(0,0)}(\mathbf{0},\mathbf{0})$ along with the deterministic kernel $\bar{G}(\mathbf{r})$.

Step 2: We will then show that for a randomly chosen $\mathbf{x}$, with high probability, there exists a specific choice of coefficients $\alpha_k, \beta_{1k}, \beta_{2k}$ guaranteeing that (8.9) is satisfied.

Step 3: We conclude the proof by showing that with the coefficients chosen as in Step 2, $|Q(\mathbf{r})| < 1$ with high probability uniformly for all $\mathbf{r} \notin \mathscr{S}$. This is accomplished using an ε-net argument.

- Step 3a: Let $\Omega \subset [0,1]^2$ be a (finite) set of grid points. For every $\mathbf{r} \in \Omega$, we show that $Q(\mathbf{r})$ is "close" to $\bar{Q}(\mathbf{r})$ with high probability.
- Step 3b: We use Bernstein's polynomial inequality to conclude that this result holds with high probability uniformly for all $\mathbf{r} \in [0,1]^2$.
- Step 3c: Finally, we combine this result with a result in [9] that shows that $\left|\bar{Q}(\mathbf{r})\right| < 1$ for all $\mathbf{r} \notin \mathscr{S}$ to conclude that $|Q(\mathbf{r})| < 1$ holds with high probability uniformly for all $\mathbf{r} \notin \mathscr{S}$.

8.1 *Step 1: Concentration of $G_{(0,0)}(\mathbf{r}, \mathbf{r}_j)$ around $\bar{G}(\mathbf{r} - \mathbf{r}_j)$*

In this subsection we establish the following result.

LEMMA 8.1 Let $G^{(m,n)}_{(m',n')}(\mathbf{r}, \mathbf{r}_j) = \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} G_{(m',n')}(\mathbf{r}, \mathbf{r}_j)$ and fix $\mathbf{r}, \mathbf{r}_j \in [0,1]^2$. For all $\alpha \geqslant 0$, and for all nonnegative integers m, m', n, n' with $m + m' + n + n' \leqslant 4$,

$$\mathbb{P}\left[\frac{1}{\kappa^{m+m'+n+n'}} |G^{(m,n)}_{(m',n')}(\mathbf{r}, \mathbf{r}_j) - \bar{G}^{(m+m',n+n')}(\mathbf{r} - \mathbf{r}_j)| > c_1 12^{\frac{m+m'+n+n'}{2}} \frac{\alpha}{\sqrt{L}} \right] \leqslant 2 \exp\left(-c \min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2} \right) \right), \tag{8.12}$$

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

To this aim, first note that, by the definition of $G_{(m',n')}(\mathbf{r}, \mathbf{r}_j)$ in (8.10) we have

$$G^{(m,n)}_{(m',n')}(\mathbf{r}, \mathbf{r}_j) = \frac{\partial^m}{\partial \tau^m} \frac{\partial^n}{\partial \nu^n} G_{(m',n')}(\mathbf{r}, \mathbf{r}_j) = \frac{L^2}{M^2} \left(\mathbf{f}^{(m,n)}(\mathbf{r})\right)^H \mathbf{F} \mathbf{G}^H \mathbf{G} \mathbf{F}^H \mathbf{g}_{(m',n')}(\mathbf{r}_j),$$

where for $r, q = -N, \ldots, N$ we define $[\mathbf{f}^{(m,n)}(\mathbf{r})]_{(r,q)} := (-i2\pi r)^m (-i2\pi r)^n e^{-i2\pi(r\tau + q\nu)}$. Lemma 8.1 is now proven in two steps. First, we show that $\mathbb{E}[\mathbf{G}^H \mathbf{G}] = \mathbf{I}$. From this, following calculations similar to (8.11), we obtain

$$\mathbb{E}[G^{(m,n)}_{(m',n')}(\mathbf{r}, \mathbf{r}_j)] = \bar{G}^{(m+m',n+n')}(\mathbf{r} - \mathbf{r}_j). \tag{8.13}$$

Second, we express $G^{(m,n)}_{(m',n')}(\mathbf{r}, \mathbf{r}_j)$ as a quadratic form in $\mathbf{x} := [x_{-N}, \ldots, x_N]^T$, and utilize the Hanson-Wright inequality stated in the lemma below to show that $G^{(m,n)}_{(m',n')}(\mathbf{r}, \mathbf{r}_j)$ does not deviate too much from its expected value $\bar{G}^{(m+m',n+n')}(\mathbf{r} - \mathbf{r}_j)$.

LEMMA 8.2 (Hanson-Wright inequality [39, Thm. 1.1]) Let $\mathbf{x} \in \mathbb{R}^L$ be a random vector with independent zero-mean K-sub-Gaussian entries (i.e., the entries obey $\sup_{p \geqslant 1} p^{-1} (\mathbb{E}[|x_\ell|^p])^{1/p} \leqslant K$), and let $\mathbf{V}$ be an $L \times L$ matrix. Then, for all $t \geqslant 0$,

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

where c is a numerical constant.

We first establish $\mathbb{E}[\mathbf{G}^H\mathbf{G}] = \mathbf{I}$. By definition of the Gabor matrix in (2.4), the entry in the (k,ℓ)-th row and (k',ℓ')-th column of $\mathbf{G}^H\mathbf{G}$ is given by

$$[\mathbf{G}^H\mathbf{G}]_{(k,\ell),(k',\ell')} = \sum_{p=-N}^{N} x^*_{p-\ell} x_{p-\ell'} e^{-i2\pi\frac{kp}{L}} e^{i2\pi\frac{k'p}{L}}.$$

Noting that $\mathbb{E}[x_\ell] = 0$, we conclude that $\mathbb{E}[[\mathbf{G}^H\mathbf{G}]_{(k,\ell),(k',\ell')}] = 0$ for $\ell \neq \ell'$. For $\ell = \ell'$, using the fact that $\mathbb{E}[x^*_{p-\ell}x_{p-\ell}] = 1/L$, we arrive at

$$\mathbb{E}[[\mathbf{G}^H\mathbf{G}]_{(k,\ell),(k',\ell')}] = \frac{1}{L}\sum_{p=-N}^{N} e^{i2\pi\frac{(k'-k)p}{L}}.$$

The latter is equal to 1 for $k = k'$ and 0 otherwise. This concludes the proof of $\mathbb{E}[\mathbf{G}^H\mathbf{G}] = \mathbf{I}$.

We now turn our attention to the concentration part of the argument, where we express $G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)$ as a quadratic form in $\mathbf{x}$ and apply Lemma 8.2. To this end, first note that

$$\begin{aligned}[L\mathbf{G}\mathbf{F}^H\mathbf{g}_{(m',n')}(\mathbf{r}_j)]_p &= \frac{1}{L}\sum_{k,\ell=-N}^{N}\left(\sum_{r,q=-N}^{N} g_r g_q e^{-i2\pi(\tau_j r+\nu_j q)}(i2\pi r)^{m'}(i2\pi q)^{n'} e^{i2\pi\frac{qk+r\ell}{L}}\right) x_{p-\ell}e^{i2\pi\frac{kp}{L}} \\ &= \sum_{\ell=-N}^{N} x_\ell \sum_{r=-N}^{N} e^{i2\pi\frac{r(p-\ell)}{L}} g_r g_p e^{-i2\pi(\tau_j r-\nu_j p)}(i2\pi r)^{m'}(-i2\pi p)^{n'}, \end{aligned} \tag{8.14}$$

where we used that $\frac{1}{L}\sum_{k=-N}^{N} e^{i2\pi\frac{k(p+q)}{L}}$ is equal to 1 if $p = -q$ and equal to 0 otherwise, together with the fact that x_ℓ is L-periodic. We next write (8.14) in matrix-vector form. For the ease of presentation, we define the matrix $\mathbf{A}(\mathbf{g}_{(m',n')}(\mathbf{r}_j)) \in \mathbb{C}^{L\times L}$ (note that $\mathbf{A}$ is a function of $\mathbf{g}_{(m',n')}(\mathbf{r}_j)$) by

$$[\mathbf{A}(\mathbf{g}_{(m',n')}(\mathbf{r}_j))]_{p,\ell} \coloneqq \sum_{k=-N}^{N} e^{i2\pi\frac{k(p-\ell)}{L}} g_k g_p e^{-i2\pi(\tau_j k-\nu_j p)}(i2\pi k)^{m'}(-i2\pi p)^{n'}.$$

Utilizing this shorthand, writing (8.14) in matrix-vector form yields

$$L\mathbf{G}\mathbf{F}^H\mathbf{g}_{(m',n')}(\mathbf{r}_j) = \mathbf{A}(\mathbf{g}_{(m',n')}(\mathbf{r}_j))\mathbf{x},$$

where $\mathbf{x} = [x_{-N},\ldots,x_N]^T$.

Analogously as in (8.14), we have

$$[L\mathbf{G}\mathbf{F}^H\mathbf{f}^{(m,n)}(\mathbf{r})]_p = \sum_{\ell=-N}^{N} x_\ell \sum_{k=-N}^{N} e^{i2\pi\frac{k(p-\ell)}{L}} e^{-i2\pi(k\tau-p\nu)}(-i2\pi k)^m(i2\pi p)^n. \tag{8.15}$$

Defining the matrix $\mathbf{A}^H(\mathbf{f}^{(m,n)}(\mathbf{r})) \in \mathbb{C}^{L\times L}$ by

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

allows us to express (8.15) in matrix-vector form according to

$$L(\mathbf{f}^{(m,n)}(\mathbf{r}))^H \mathbf{F}\mathbf{G}^H = \mathbf{x}^H \mathbf{A}^H(\mathbf{f}^{(m,n)}(\mathbf{r})).$$

This allows us to represent $G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)$ in the desired quadratic form

$$G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j) = \frac{L^2}{M^2}(\mathbf{f}^{(m,n)}(\mathbf{r}))^H \mathbf{F}\mathbf{G}^H\mathbf{G}\mathbf{F}^H\mathbf{g}_{(m',n')}(\mathbf{r}_j) = \mathbf{x}^H \underbrace{\frac{1}{M^2}\mathbf{A}^H(\mathbf{f}^{(m,n)}(\mathbf{r}))\mathbf{A}(\mathbf{g}_{(m',n')}(\mathbf{r}_j))}_{\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j):=}\mathbf{x}, \quad (8.16)$$

where

$$[\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)]_{\tilde{\ell},\ell} = \frac{1}{M^2}\sum_{p,k,\tilde{k}=-N}^{N} e^{i2\pi\frac{(p-\ell)k}{L}} e^{-i2\pi\frac{(p-\tilde{\ell})\tilde{k}}{L}} e^{i2\pi(\tilde{k}\tau - p(v-v_j)-k\tau_j)} g_p g_k (i2\pi\tilde{k})^m (i2\pi k)^{m'} (-i2\pi p)^{n+n'}.$$

In order to evaluate the RHS of the Hanson-Wright inequality, we will need the following upper bound on $\left\|\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right\|_F$. We defer the proof to Section 8.1.1.

LEMMA 8.3 For all $\mathbf{r}$ and $\mathbf{r}_j$, and for all non-negative m,m',n,n' with $m+m'+n+n' \leqslant 4$,

$$\left\|\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right\|_F \leqslant c_1(2\pi N)^{m+m'+n+n'}\sqrt{L}. \quad (8.17)$$

We are now ready to establish Lemma 8.1 by applying the Hanson-Wright inequality. To this end note that using $\kappa = \sqrt{F''(0)} = \sqrt{\frac{\pi^2}{3}(N^2+4N)}$ and utilizing [9, Eq. 2.23] we have

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

Setting $\mathbf{V} := \mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)$ for ease of presentation, we have

$$\begin{aligned}
&\mathbb{P}\left[\frac{1}{\kappa^{m+m'+n+n'}}|G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j) - \bar{G}^{(m+m',n+n')}(\mathbf{r}-\mathbf{r}_j)| > c_1 12^{\frac{m+m'+n+n'}{2}}\frac{\alpha}{\sqrt{L}}\right] \\
&\quad\leqslant \mathbb{P}\left[|G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j) - \bar{G}^{(m+m',n+n')}(\mathbf{r}-\mathbf{r}_j)| > c_1(2\pi N)^{m+m'+n+n'}\frac{\alpha}{\sqrt{L}}\right] \\
&\quad\leqslant \mathbb{P}\left[|\mathbf{x}^T\mathbf{V}\mathbf{x} - \mathbb{E}[\mathbf{x}^T\mathbf{V}\mathbf{x}]| > \|\mathbf{V}\|_F\frac{\alpha}{L}\right] \quad (8.18) \\
&\quad\leqslant 2\exp\left(-c\min\left(\frac{\|\mathbf{V}\|_F^2\alpha^2}{L^2K^4\|\mathbf{V}\|_F^2}, \frac{\|\mathbf{V}\|_F\alpha}{LK^2\|\mathbf{V}\|}\right)\right) \quad (8.19) \\
&\quad\leqslant 2\exp\left(-c\min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right)\right). \quad (8.20)
\end{aligned}$$

Here, (8.18) follows from (8.16) and (8.17), together with the fact that $\mathbb{E}[\mathbf{x}^H\mathbf{V}\mathbf{x}] = \mathbb{E}[G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)] = \bar{G}^{(m+m',n+n')}(\mathbf{r}-\mathbf{r}_j)$ (cf. (8.13)). To obtain (8.19), we used Lemma 8.2 with $t = \|\mathbf{V}\|_F \frac{\alpha}{L}$. Finally, (8.20) holds because the sub-Gaussian parameter K of the random variable $[\mathbf{x}]_\ell \sim \mathcal{N}(0,1/L)$ is given by $K = c_2/\sqrt{L}$ (e.g., [49, Ex. 5.8]) and $\|\mathbf{V}\|_F/\|\mathbf{V}\| \geqslant 1$.

8.1.1 *Proof of Lemma 8.3:* We start by upper-bounding $|[\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)]_{\tilde{\ell},\ell}|$. By definition of $F(t)$ (cf. (8.6))

$$\begin{aligned}
&[\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)]_{\tilde{\ell},\ell} \\
&= \sum_{p=-N}^{N} \left(\frac{1}{M} \sum_{k=-N}^{N} g_k (i2\pi k)^{m'} e^{i2\pi\left(\frac{p-\ell}{L}-\tau_j\right)k} \right) \left(\frac{1}{M} \sum_{\tilde{k}=-N}^{N} (i2\pi\tilde{k})^{m} e^{-i2\pi\left(\frac{p-\tilde{\ell}}{L}-\tau\right)\tilde{k}} \right) \cdot \\
&\qquad\qquad \cdot g_p (-i2\pi p)^{n+n'} e^{-i2\pi(\nu-\nu_j)p} \\
&= \sum_{p=-N}^{N} F^{(m')}\left(\frac{p-\ell}{L}-\tau_j\right) \left(\frac{1}{M} \sum_{\tilde{k}=-N}^{N} (i2\pi\tilde{k})^{m} e^{-i2\pi\left(\frac{p-\tilde{\ell}}{L}-\tau\right)\tilde{k}} \right) g_p (-i2\pi p)^{n+n'} e^{-i2\pi(\nu-\nu_j)p},
\end{aligned}$$

where $F^{(m)}(t) \coloneqq \frac{\partial^m}{\partial\tau^m} F(t)$. Since $|g_p| \leqslant 1$ holds for all p, we obtain

$$\begin{aligned}
|[\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)]_{\tilde{\ell},\ell}| &\leqslant (2\pi N)^{n+n'} \sum_{p=-N}^{N} \left| F^{(m')}\left(\frac{p-\ell}{L}-\tau_j\right)\right| \left| \frac{1}{M} \sum_{\tilde{k}=-N}^{N} (-i2\pi\tilde{k})^{m} e^{i2\pi\left(\frac{p-\tilde{\ell}}{L}-\tau\right)\tilde{k}} \right| \\
&= (2\pi N)^{n+n'} \sum_{p=-N}^{N} \left| F^{(m')}\left(\frac{p}{L}+s/L-\tau_j\right)\right| \left| \frac{1}{M} \sum_{\tilde{k}=-N}^{N} (-i2\pi\tilde{k})^{m} e^{i2\pi\left(\frac{p+\ell+s-\tilde{\ell}}{L}-\tau\right)\tilde{k}} \right|,
\end{aligned} \tag{8.21}$$

where we choose s as the integer minimizing $|s/L-\tau_j|$ and used the fact that the absolute values in the sum above are L-periodic in p (recall that $F(t)$ is 1-periodic).

We proceed by upper-bounding $|F^{(m)}(t)|$. To this aim, we use Bernstein's polynomial inequality, (cf. Proposition 8.3 below), to conclude that

$$\sup_t \left|F^{(m)}(t)\right| \leqslant (2\pi N)^m \sup_t |F(t)| = (2\pi N)^m. \tag{8.22}$$

Also note that, from [9, Lem. 2.6] we know that for $|t| \in [1/(2N), 1/2]$, there exists a numerical constant $\tilde{c}$ such that

$$|F^{(m)}(t)| \leqslant \tilde{c}(2\pi N)^m \frac{1}{(2Mt)^4}. \tag{8.23}$$

Combining (8.22) and (8.23) we arrive at

$$|F^{(m)}(t)| \leqslant H^{(m)}(t) \coloneqq \tilde{c}(2\pi N)^m \min\left(1, \frac{1}{(2Mt)^4}\right). \tag{8.24}$$

Utilizing the latter inequality we have

$$\begin{aligned}\left|F^{(m)}\left(\frac{p}{L}+s/L-\tau_j\right)\right| &\leqslant H^{(m)}\left(\frac{p}{L}+s/L-\tau_j\right)\\ &\leqslant c' H^{(m)}\left(\frac{p}{L}\right) &(8.25)\\ &\leqslant c'\bar{c}(2\pi N)^m \min\left(1,\frac{1}{(2Mp/L)^4}\right) &(8.26)\\ &\leqslant c'\bar{c}(2\pi N)^m \min\left(1,\frac{16}{p^4}\right) \leqslant 16c'\bar{c}(2\pi N)^m \min\left(1,\frac{1}{p^4}\right). &(8.27)\end{aligned}$$

Here, (8.25) holds because when s is the integer minimizing $|s/L-\tau_j|$, we have that $|s/L-\tau_j| \leqslant 1/(2L)$. Therefore, for all p with $|p|>0$, $2M(p/L-s/L-\tau_j)$ is within a constant factor of $2Mp/L$, which proves that (8.25) holds for a numerical constant c'. To obtain (8.26) we used (8.24). Finally, (8.27) follows from $\frac{L}{2M}=\frac{2N+1}{N+2}<2$.

Plugging (8.27) into (8.21) we obtain

$$|[\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)]_{\tilde{\ell},\ell}| \leqslant (2\pi N)^{m+m'+n+n'} \underbrace{\hat{c}(2\pi N)^{-m}\sum_{p=-N}^{N}\min\left(1,\frac{1}{p^4}\right)\left|\frac{1}{M}\sum_{\tilde{k}=-N}^{N}(-i2\pi\tilde{k})^m e^{i2\pi\left(\frac{p+s+\ell-\tilde{\ell}}{L}-\tau\right)\tilde{k}}\right|}_{U\left(\tau-\frac{s+\ell-\tilde{\ell}}{L}\right):=}, \tag{8.28}$$

where $\hat{c}=16c'\bar{c}$. We show in Appendix F that $U(t)$ is 1-periodic and satisfies $U(t)\leqslant c\min(1,\frac{1}{L|t|})$ for $|t|\leqslant 1/2$. Using this bound together with (8.28) we conclude that

$$\begin{aligned}\left\|\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right\|_F^2 &= \sum_{\ell,\tilde{\ell}=-N}^{N}\left|[\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)]_{\tilde{\ell},\ell}\right|^2\\ &\leqslant (2\pi N)^{2(m+m'+n+n')}\sum_{\ell,\tilde{\ell}=-N}^{N}U^2\left(\tau-\frac{s+\ell-\tilde{\ell}}{L}\right)\\ &\leqslant (2\pi N)^{2(m+m'+n+n')}\sum_{\tilde{\ell}=-N}^{N}\sum_{\ell=-N}^{N}\left(c\min\left(1,\frac{1}{L|\ell/L|}\right)\right)^2\\ &\leqslant c^2(2\pi N)^{2(m+m'+n+n')}\sum_{\tilde{\ell}=-N}^{N}\left(1+2\sum_{\ell\geqslant 1}\frac{1}{\ell^2}\right)\\ &= c^2 L(2\pi N)^{2(m+m'+n+n')}\left(1+\frac{\pi^2}{3}\right).\end{aligned}$$

The proof is now complete by setting $c_1=c\sqrt{1+\pi^2/3}\sqrt{L}$.

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

We next show that, with high probability, it is possible to select the coefficients $\alpha_k,\beta_{1k},\beta_{2k}$ such that $Q(\mathbf{r})$ satisfies (8.9). To this end, we first review the result in [9] that ensures that there exists a set of

coefficients $\bar{\alpha}_k, \bar{\beta}_{1k}, \bar{\beta}_{2k}$ such that (8.7) is satisfied. Specifically, writing (8.7) in matrix form yields

$$\underbrace{\begin{bmatrix} \bar{\mathbf{D}}^{(0,0)} & \frac{1}{\kappa}\bar{\mathbf{D}}^{(1,0)} & \frac{1}{\kappa}\bar{\mathbf{D}}^{(0,1)} \\ -\frac{1}{\kappa}\bar{\mathbf{D}}^{(1,0)} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}^{(2,0)} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}^{(1,1)} \\ -\frac{1}{\kappa}\bar{\mathbf{D}}^{(0,1)} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}^{(1,1)} & -\frac{1}{\kappa^2}\bar{\mathbf{D}}^{(0,2)} \end{bmatrix}}_{\bar{\mathbf{D}}} \begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \kappa\bar{\boldsymbol{\beta}}_1 \\ \kappa\bar{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{8.29}$$

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 ($F(0) = 1$, $\kappa^2 = |F''(0)|$, and $F''(0)$ 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 and $\bar{\mathbf{D}}^{(1,0)}, \bar{\mathbf{D}}^{(0,1)}$ are antisymmetric, $\bar{\mathbf{D}}$ is symmetric.

The following result, which directly follows from [9, Eq. C6, C7, C8, C9], ensures that $\bar{\mathbf{D}}$ is invertible and thus the coefficients $\bar{\alpha}_k, \bar{\beta}_{1k}, \bar{\beta}_{2k}$ can be obtained according to

$$\begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \kappa\bar{\boldsymbol{\beta}}_1 \\ \kappa\bar{\boldsymbol{\beta}}_2 \end{bmatrix} = \bar{\mathbf{D}}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix} = \bar{\mathbf{L}}\mathbf{u}, \tag{8.30}$$

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 8.2 $\bar{\mathbf{D}}$ is invertible and

$$\left\|\mathbf{I} - \bar{\mathbf{D}}\right\| \leqslant 0.19808 \tag{8.31}$$

$$\left\|\bar{\mathbf{D}}\right\| \leqslant 1.19808 \tag{8.32}$$

$$\left\|\bar{\mathbf{D}}^{-1}\right\| \leqslant 1.24700. \tag{8.33}$$

Proof. The proof of this proposition is an immediate consequence of [9, Eq. C6, C7, C8, C9]. 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 [26, 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 $\|\mathbf{A}\|_\infty \coloneqq \max_i \sum_j |[\mathbf{A}]_{i,j}|$. Using that $\left\|\mathbf{I} - \bar{\mathbf{D}}\right\|_\infty \leqslant 0.19808$ (shown below), it follows that $\bar{\mathbf{D}}$ is invertible and

$$\begin{aligned} \left\|\bar{\mathbf{D}}\right\| &\leqslant 1 + \left\|\mathbf{I} - \bar{\mathbf{D}}\right\|_\infty \leqslant 1.19808 \\ \left\|\bar{\mathbf{D}}^{-1}\right\| &\leqslant \frac{1}{1 - \left\|\mathbf{I} - \bar{\mathbf{D}}\right\|_\infty} \leqslant 1.2470. \end{aligned}$$

The proof is concluded by noting that

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

where we used [9, Eq. C6, C7, C8, C9]:

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

□

We next select the coefficients $\alpha_k, \beta_{1k}, \beta_{2k}$ such that $Q(\mathbf{r})$ satisfies the interpolation conditions (8.9). To this end, we write (8.9) in matrix form:

$$\underbrace{\begin{bmatrix}\mathbf{D}^{(0,0)}_{(0,0)} & \frac{1}{\kappa}\mathbf{D}^{(0,0)}_{(1,0)} & \frac{1}{\kappa}\mathbf{D}^{(0,0)}_{(0,1)}\\ -\frac{1}{\kappa}\mathbf{D}^{(1,0)}_{(0,0)} & -\frac{1}{\kappa^2}\mathbf{D}^{(1,0)}_{(1,0)} & -\frac{1}{\kappa^2}\mathbf{D}^{(1,0)}_{(0,1)}\\ -\frac{1}{\kappa}\mathbf{D}^{(0,1)}_{(0,0)} & -\frac{1}{\kappa^2}\mathbf{D}^{(0,1)}_{(1,0)} & -\frac{1}{\kappa^2}\mathbf{D}^{(0,1)}_{(0,1)}\end{bmatrix}}_{\mathbf{D}}\begin{bmatrix}\boldsymbol{\alpha}\\ \kappa\boldsymbol{\beta}_1\\ \kappa\boldsymbol{\beta}_2\end{bmatrix} = \begin{bmatrix}\mathbf{u}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \tag{8.34}$$

where $[\mathbf{D}^{(m,n)}_{(m',n')}]_{j,k} \coloneqq G^{(m,n)}_{(m',n')}(\mathbf{r}_j, \mathbf{r}_k)$, $[\boldsymbol{\alpha}]_k \coloneqq \alpha_k$, $[\boldsymbol{\beta}_1]_k \coloneqq \beta_{1k}$, and $[\boldsymbol{\beta}_2]_k \coloneqq \beta_{2k}$. To show that the system of equations (8.34) has a solution, and in turn (8.9) can be satisfied, we will show that, with high probability, $\mathbf{D}$ is invertible. To this end, we show that the probability of the event

$$\mathscr{E}_\xi = \{\left\|\mathbf{D}-\bar{\mathbf{D}}\right\| \leqslant \xi\}$$

is high, and $\mathbf{D}$ is invertible on $\mathscr{E}_\xi$ for all $\xi \in (0, 1/4]$. The fact that $\mathbf{D}$ is invertible on $\mathscr{E}_\xi$ for all $\xi \in (0, 1/4]$ follows from the following set of inequalities:

$$\|\mathbf{I}-\mathbf{D}\| \leqslant \left\|\mathbf{D}-\bar{\mathbf{D}}\right\| + \left\|\bar{\mathbf{D}}-\mathbf{I}\right\| \leqslant \xi + 0.1908 \leqslant 0.4408.$$

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

$$\begin{bmatrix}\boldsymbol{\alpha}\\ \kappa\boldsymbol{\beta}_1\\ \kappa\boldsymbol{\beta}_2\end{bmatrix} = \mathbf{D}^{-1}\begin{bmatrix}\mathbf{u}\\ \mathbf{0}\end{bmatrix} = \mathbf{L}\mathbf{u}, \tag{8.35}$$

where $\mathbf{L}$ is the $3S \times S$ submatrix of $\mathbf{D}^{-1}$ corresponding to the first S columns of $\mathbf{D}^{-1}$. We record two useful inequalities about $\mathbf{L}$ and its deviation from $\bar{\mathbf{L}}$ on the event $\mathscr{E}_\xi$ in the lemma below.

LEMMA 8.4 On the event $\mathscr{E}_\xi$ with $\xi \in (0, 1/4]$ the following identities hold

$$\|\mathbf{L}\| \leqslant 2.5, \tag{8.36}$$

$$\left\|\mathbf{L}-\bar{\mathbf{L}}\right\| \leqslant 2.5\xi. \tag{8.37}$$

Proof. We will make use of the following lemma.

LEMMA 8.5 ([46, Proof of Cor. 4.5]) Suppose that $\mathbf{C}$ is invertible and $\|\mathbf{B}-\mathbf{C}\|\left\|\mathbf{C}^{-1}\right\| \leqslant 1/2$. Then i) $\left\|\mathbf{B}^{-1}\right\| \leqslant 2\left\|\mathbf{C}^{-1}\right\|$ and ii) $\left\|\mathbf{B}^{-1}-\mathbf{C}^{-1}\right\| \leqslant 2\left\|\mathbf{C}^{-1}\right\|^2\|\mathbf{B}-\mathbf{C}\|$.

First note that since $\left\|\mathbf{D}-\bar{\mathbf{D}}\right\| \leqslant 1/4$ and $\left\|\bar{\mathbf{D}}^{-1}\right\| \leqslant 1.247$ (cf. (8.33)) the conditions of Lemma 8.5 with $\mathbf{B}=\mathbf{D}$ and $\mathbf{C}=\bar{\mathbf{D}}$ are satisfied.

Equation (8.36) now readily follows:

$$\|\mathbf{L}\| \leqslant \left\|\mathbf{D}^{-1}\right\| \leqslant 2\left\|\bar{\mathbf{D}}^{-1}\right\| \leqslant 2.5. \tag{8.38}$$

Specifically, for the first inequality, we used that $\mathbf{L}$ is a submatrix of $\mathbf{D}^{-1}$, the second inequality follows by application of part *i*) of Lemma 8.5, and the last inequality follows from (8.33).

To prove (8.37), we use part *ii*) of the lemma above together with (8.33), to conclude that

$$\left\|\mathbf{L}-\bar{\mathbf{L}}\right\| \leqslant \left\|\mathbf{D}^{-1}-\bar{\mathbf{D}}^{-1}\right\| \leqslant 2\left\|\bar{\mathbf{D}}^{-1}\right\|^2\left\|\mathbf{D}-\bar{\mathbf{D}}\right\| \leqslant 2.5\xi. \tag{8.39}$$

holds on $\mathscr{E}_\xi$ with $\xi \in (0,1/4]$. □

It only remains to prove that the event $\mathscr{E}_\xi$ does indeed have high probability.

LEMMA 8.6 For all $\xi > 0$

$$\mathbb{P}\left[\mathscr{E}_\xi\right] \geqslant 1-\delta$$

provided that

$$L \geqslant S\frac{c_4}{\xi^2}\log^2(18S^2/\delta), \tag{8.40}$$

where c_4 is a numerical constant.

Proof. We will upper-bound $\left\|\mathbf{D}-\bar{\mathbf{D}}\right\|$ 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

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

for $m,m',n,n' \leqslant 1$ and $m+m'+n+n' \leqslant 2$ and for $j,k=1,\ldots,S$. We now have

$$\begin{aligned}
\mathbb{P}\left[\left\|\mathbf{D}-\bar{\mathbf{D}}\right\| \geqslant \xi\right] &\leqslant \mathbb{P}\left[\sqrt{3S}\max_{j,k,m,m',n,n'}\frac{1}{\kappa^{m+m'+n+n'}}|[\mathbf{D}^{(m,n)}_{(m',n')}-\bar{\mathbf{D}}^{(m+m',n+n')}]_{j,k}| \geqslant \xi\right] && (8.41)\\
&\leqslant \sum_{j,k,m,m',n,n'}\mathbb{P}\left[\frac{1}{\kappa^{m+m'+n+n'}}|[\mathbf{D}^{(m,n)}_{(m',n')}-\bar{\mathbf{D}}^{(m+m',n+n')}]_{j,k}| \geqslant \frac{\xi}{\sqrt{3S}}\right] && (8.42)\\
&= \sum_{j,k,m,m',n,n'}\mathbb{P}\left[\frac{1}{\kappa^{m+m'+n+n'}}|[\mathbf{D}^{(m,n)}_{(m',n')}-\bar{\mathbf{D}}^{(m+m',n+n')}]_{j,k}| \geqslant 12c_1\frac{\alpha}{\sqrt{L}}\right] && (8.43)\\
&\leqslant \sum_{j,k,m,m',n,n'}\mathbb{P}\left[\frac{1}{\kappa^{m+m'+n+n'}}|[\mathbf{D}^{(m,n)}_{(m',n')}-\bar{\mathbf{D}}^{(m+m',n+n')}]_{j,k}| \geqslant 12^{\frac{m+m'+n+n'}{2}}c_1\frac{\alpha}{\sqrt{L}}\right] && (8.44)\\
&\leqslant 2(3S)^2\exp\left(-c\min\left(\frac{\xi^2 L}{c_2^4c_3^2 3S},\frac{\xi\sqrt{L}}{c_2^2c_3\sqrt{3S}}\right)\right). && (8.45)
\end{aligned}$$

Here, (8.41) follows from the fact that $\mathbf{D}$ and $\bar{\mathbf{D}}$ are $3S \times 3S$ matrices; (8.42) follows from the union bound; (8.43) follows by setting $\alpha := \frac{\xi\sqrt{L}}{\sqrt{3S}12c_1}$, where c_1 is the constant in Lemma 8.1; (8.44) follows from the fact that $12^{\frac{m+m'+n+n'}{2}} \leqslant 12$ holds for $m+m'+n+n' \leqslant 2$; and (8.45) follows from Lemma 8.1 (here we set $c_3 := 12c_1$).

To show that the RHS of (8.45) is smaller than δ, as desired, it suffices to show

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

which is a consequence of (8.40) 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 8.7 below that shows that $Q(\mathbf{r})$ and $\bar{Q}(\mathbf{r})$ and their partial derivatives are close on a set of (grid) points.

LEMMA 8.7 Let $\Omega \subset [0,1]^2$ be a finite set of points. Fix $0 < \varepsilon \leqslant 1$ and $\delta > 0$. Suppose that

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

Then,

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

In order to prove Lemma 8.7, first note that the (m,n)-th partial derivative of $Q(\mathbf{r})$ (defined by (8.8)) after normalization with $1/\kappa^{m+n}$ is given by

$$\begin{aligned}\frac{1}{\kappa^{m+n}} Q^{(m,n)}(\mathbf{r}) &= \frac{1}{\kappa^{m+n}} \sum_{k=1}^{S} \left(\alpha_k G_{(0,0)}^{(m,n)}(\mathbf{r},\mathbf{r}_k) + \kappa\beta_{1k} \frac{1}{\kappa} G_{(1,0)}^{(m,n)}(\mathbf{r},\mathbf{r}_k) + \kappa\beta_{2k} \frac{1}{\kappa} G_{(0,1)}^{(m,n)}(\mathbf{r},\mathbf{r}_k)\right) \\ &= \left(\mathbf{v}^{(m,n)}(\mathbf{r})\right)^H \mathbf{L}\mathbf{u}. \end{aligned} \tag{8.46}$$

Here, we used (8.35) and the shorthand $\mathbf{v}^{(m,n)}(\mathbf{r})$ defined by

$$\begin{aligned}\left(\mathbf{v}^{(m,n)}\right)^H(\mathbf{r}) := \frac{1}{\kappa^{m+n}} \Bigg[&G_{(0,0)}^{(m,n)}(\mathbf{r},\mathbf{r}_1), ..., G_{(0,0)}^{(m,n)}(\mathbf{r},\mathbf{r}_S),\ \frac{1}{\kappa} G_{(1,0)}^{(m,n)}(\mathbf{r},\mathbf{r}_1), ..., \frac{1}{\kappa} G_{(1,0)}^{(m,n)}(\mathbf{r},\mathbf{r}_S), \\ &\frac{1}{\kappa} G_{(0,1)}^{(m,n)}(\mathbf{r},\mathbf{r}_1), ..., \frac{1}{\kappa} G_{(0,1)}^{(m,n)}(\mathbf{r},\mathbf{r}_S)\Bigg].\end{aligned}$$

Since $\mathbb{E}[G_{(m',n')}^{(m,n)}(\mathbf{r},\mathbf{r}_j)] = \bar{G}^{(m+m',n+n')}(\mathbf{r}-\mathbf{r}_j)$ (cf. (8.13)), we have

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

where

$$\begin{aligned}\left(\bar{\mathbf{v}}^{(m,n)}\right)^H(\mathbf{r}) := \frac{1}{\kappa^{m+n}} \Bigg[&\bar{G}^{(m,n)}(\mathbf{r}-\mathbf{r}_1), ..., \bar{G}^{(m,n)}(\mathbf{r}-\mathbf{r}_S),\ \frac{1}{\kappa} \bar{G}^{(m+1,n)}(\mathbf{r}-\mathbf{r}_1), ..., \frac{1}{\kappa} \bar{G}^{(m+1,n)}(\mathbf{r}-\mathbf{r}_S), \\ &\frac{1}{\kappa} \bar{G}^{(m,n+1)}(\mathbf{r}-\mathbf{r}_1), ..., \frac{1}{\kappa} \bar{G}^{(m,n+1)}(\mathbf{r}-\mathbf{r}_S)\Bigg].\end{aligned}$$

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

$$\begin{aligned}\frac{1}{\kappa^{m+n}}Q^{(m,n)}(\mathbf{r}) &= \left\langle \mathbf{u}, \mathbf{L}^H \mathbf{v}^{(m,n)}(\mathbf{r}) \right\rangle \\ &= \left\langle \mathbf{u}, \bar{\mathbf{L}}^H \bar{\mathbf{v}}^{(m,n)}(\mathbf{r}) \right\rangle + \underbrace{\left\langle \mathbf{u}, \mathbf{L}^H (\mathbf{v}^{(m,n)}(\mathbf{r}) - \bar{\mathbf{v}}^{(m,n)}(\mathbf{r})) \right\rangle}_{I_1^{(m,n)}(\mathbf{r})} + \underbrace{\left\langle \mathbf{u}, (\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(m,n)}(\mathbf{r}) \right\rangle}_{I_2^{(m,n)}(\mathbf{r})} \\ &= \frac{1}{\kappa^{m+n}}\bar{Q}^{(m,n)}(\mathbf{r}) + I_1^{(m,n)}(\mathbf{r}) + I_2^{(m,n)}(\mathbf{r}), \end{aligned} \tag{8.47}$$

where $\bar{\mathbf{L}}$ was defined below (8.30). 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 8.8 Let $\Omega \subset [0,1]^2$ be a finite set of points and suppose that $m+n \leqslant 2$. Then, for all $0 < \varepsilon \leqslant 1$ and for all $\delta > 0$,

$$\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_1^{(m,n)}(\mathbf{r})| \geqslant \varepsilon\right] \leqslant \delta + \mathbb{P}\left[\overline{\mathscr{E}}_{1/4}\right]$$

provided that

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

LEMMA 8.9 Let $\Omega \subset [0,1]^2$ be a finite set of points. Suppose that $m+n \leqslant 2$. Then, for all $\varepsilon, \delta > 0$, and for all $\xi > 0$ with

$$\xi \leqslant \frac{\varepsilon c_6}{\sqrt{\log\left(\frac{4|\Omega|}{\delta}\right)}}, \tag{8.48}$$

where $c_6 \leqslant 1/4$ is a numerical constant, it follows

$$\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_2^{(m,n)}(\mathbf{r})| \geqslant \varepsilon \middle| \mathscr{E}_\xi\right] \leqslant \delta.$$

Lemma 8.8 and 8.9 are proven in Section 8.3.1 and 8.3.2, respectively.

We are now ready to complete the proof of Lemma 8.7. From (8.47), we obtain for all $\xi > 0$ satisfying (8.48)

$$\begin{aligned}\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \frac{1}{\kappa^{n+m}} \left|Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r})\right| \geqslant 2\varepsilon\right] &= \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \left|I_1^{(m,n)}(\mathbf{r}) + I_2^{(m,n)}(\mathbf{r})\right| \geqslant 2\varepsilon\right] \\ &\leqslant \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \left|I_1^{(m,n)}(\mathbf{r})\right| \geqslant \varepsilon\right] + \mathbb{P}\left[\overline{\mathscr{E}}_\xi\right] + \mathbb{P}\left[\max_{\mathbf{r}\in\Omega} \left|I_2^{(m,n)}(\mathbf{r})\right| \geqslant \varepsilon | \mathscr{E}_\xi\right] \qquad (8.49) \\ &\leqslant 4\delta, \end{aligned}$$

where (8.49) follows from the union bound and the fact that $\mathbb{P}[A] = \mathbb{P}\left[A \cap \overline{B}\right] + \mathbb{P}[A \cap B] \leqslant \mathbb{P}\left[\overline{B}\right] + \mathbb{P}[A|B]$ by setting $B = \mathscr{E}_\xi$ and $A = \left\{\max_{\mathbf{r}\in\Omega} \left|I_2^{(m,n)}(\mathbf{r})\right| \geqslant \varepsilon\right\}$, and the last inequality follows from Lemmas 8.6,

8.8 and 8.9, respectively. In more details, we choose $\xi = \varepsilon c_6 \log^{-1/2}\left(\frac{4|\Omega|}{\delta}\right)$. It then follows from Lemma 8.9 that the third probability in (8.49) is smaller than δ. With this choice of ξ, the condition in Lemma 8.6 becomes $L \geqslant S\frac{c_4}{\varepsilon^2 c_6^2}\log\left(\frac{4|\Omega|}{\delta}\right)\log\left(\frac{18S^2}{\delta}\right)$, which is satisfied by choosing $c = \frac{c_4}{c_6^2}$. Moreover, $\xi \leqslant 1/4$ since $\varepsilon \leqslant 1$ and $c_6 \leqslant 1/4$. Thus, Lemma 8.6 yields $\mathbb{P}\left[\overline{\mathscr{E}}_\xi\right] \leqslant \delta$ and $\mathbb{P}\left[\overline{\mathscr{E}}_{1/4}\right] \leqslant \delta$. Finally, observe that the conditions of Lemma 8.8 are satisfied, thus the first probability in (8.49) can be upper-bounded by

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

This concludes the proof.

8.3.1 *Proof of Lemma 8.8* Set $\Delta\mathbf{v}^{(m,n)} := \mathbf{v}^{(m,n)}(\mathbf{r}) - \bar{\mathbf{v}}^{(m,n)}(\mathbf{r})$ for notational convenience. By the union bound, we have for all $a, b \geqslant 0$,

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{r}\in\Omega}|I_1^{(m,n)}(\mathbf{r})| \geqslant 2.5ab\right] &= \mathbb{P}\left[\max_{\mathbf{r}\in\Omega}\left|\left\langle \mathbf{u}, \mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\rangle\right| \geqslant 2.5ab\right] \\
&\leqslant \mathbb{P}\left[\bigcup_{\mathbf{r}\in\Omega}\left\{\left|\left\langle \mathbf{u}, \mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\rangle\right| \geqslant \left\|\mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\|_2 b\right\} \cup \left\{\left\|\mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\|_2 \geqslant 2.5a\right\}\right] \\
&\leqslant \mathbb{P}\left[\bigcup_{\mathbf{r}\in\Omega}\left\{\left|\left\langle \mathbf{u}, \mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\rangle\right| \geqslant \left\|\mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\|_2 b\right\} \cup \left\{\left\|\Delta\mathbf{v}^{(m,n)}\right\|_2 \geqslant a\right\} \cup \{\|\mathbf{L}\| \geqslant 2.5\}\right] \\
&\leqslant \mathbb{P}[\|\mathbf{L}\| \geqslant 2.5] + \sum_{\mathbf{r}\in\Omega}\left(\mathbb{P}\left[\left|\left\langle \mathbf{u}, \mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\rangle\right| \geqslant \left\|\mathbf{L}^H\Delta\mathbf{v}^{(m,n)}\right\|_2 b\right] + \mathbb{P}\left[\left\|\Delta\mathbf{v}^{(m,n)}\right\|_2 \geqslant a\right]\right) \\
&\leqslant \mathbb{P}\left[\overline{\mathscr{E}}_{1/4}\right] + |\Omega|4e^{-\frac{b^2}{4}} + \sum_{\mathbf{r}\in\Omega}\mathbb{P}\left[\left\|\Delta\mathbf{v}^{(m,n)}\right\|_2 \geqslant a\right] \qquad (8.50) \\
&\leqslant \mathbb{P}\left[\overline{\mathscr{E}}_{1/4}\right] + \frac{\delta}{2} + \sum_{\mathbf{r}\in\Omega}\mathbb{P}\left[\left\|\Delta\mathbf{v}^{(m,n)}\right\|_2 \geqslant a\right], \qquad (8.51)
\end{aligned}$$

where (8.50) follows from application of Hoeffding's inequality (stated below) and from $\{\|\mathbf{L}\| \geqslant 2.5\} \subseteq \overline{\mathscr{E}}_{1/4}$ according to (8.36). For (8.51), we used $|\Omega|4e^{-\frac{b^2}{4}} \leqslant \frac{\delta}{2}$ ensured by choosing $b = 2\sqrt{\log(8|\Omega|/\delta)}$.

LEMMA 8.10 (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 $t \geqslant 0$, and for all $\mathbf{v} \in \mathbb{C}^S$

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

We next upper-bound $\mathbb{P}\left[\left\|\Delta\mathbf{v}^{(m,n)}\right\|_2 \geqslant a\right]$ in (8.51). For all $\alpha \geqslant 0$, using that $12^{\frac{n+m+1}{2}} \leqslant 12^{\frac{3}{2}}$, we

have

$$\begin{aligned}
\mathbb{P}\left[\left\|\Delta \mathbf{v}^{(m,n)}\right\|_2 \geqslant \frac{\sqrt{3S}}{\sqrt{L}} 12^{\frac{3}{2}} c_1 \alpha\right] &\leqslant \mathbb{P}\left[\left\|\Delta \mathbf{v}^{(m,n)}\right\|_2 \geqslant \frac{\sqrt{3S}}{\sqrt{L}} 12^{\frac{n+m+1}{2}} c_1 \alpha\right] \\
&= \mathbb{P}\left[\left\|\Delta \mathbf{v}^{(m,n)}\right\|_2^2 \geqslant \frac{3S}{L} 12^{n+m+1} c_1^2 \alpha^2\right] \\
&\leqslant \sum_{k=1}^{3S} \mathbb{P}\left[|[\Delta \mathbf{v}^{(m,n)}]_k|^2 \geqslant \frac{1}{L} 12^{n+m+1} c_1^2 \alpha^2\right] \qquad (8.52) \\
&= \sum_{k=1}^{3S} \mathbb{P}\left[|[\Delta \mathbf{v}^{(m,n)}]_k| \geqslant \frac{1}{\sqrt{L}} 12^{\frac{n+m+1}{2}} c_1 \alpha\right] \\
&\leqslant 3S \cdot 2 \exp\left(-c \min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right)\right) \qquad (8.53) \\
&\leqslant \frac{\delta}{2|\Omega|}, \qquad (8.54)
\end{aligned}$$

where (8.52) follows from the union bound, (8.53) follows from Lemma 8.1. Finally, (8.54) follows by choosing $\alpha = \frac{c_2^2}{c} \log\left(\frac{12S|\Omega|}{\delta}\right)$ and using the fact that for $\alpha \geqslant c_2^2$ (since $c \geqslant 1$) we have $\min\left(\frac{\alpha^2}{c_2^4}, \frac{\alpha}{c_2^2}\right) = \frac{\alpha}{c_2^2}$. We have established that $\mathbb{P}\left[\left\|\Delta \mathbf{v}^{(m,n)}\right\|_2 \geqslant a\right] \leqslant \frac{\delta}{2|\Omega|}$ 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 (8.54) into (8.51) we get

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

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.

8.3.2 *Proof of Lemma 8.9* By the union bound

$$\begin{aligned}
\mathbb{P}\left[\max_{\mathbf{r}\in\Omega} |I_2^{(m,n)}(\mathbf{r})| \geqslant \varepsilon \middle| \mathscr{E}_\xi\right] &\leqslant \sum_{\mathbf{r}\in\Omega} \mathbb{P}\left[\left|\left\langle \mathbf{u}, (\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(m,n)}(\mathbf{r})\right\rangle\right| \geqslant \varepsilon \middle| \mathscr{E}_\xi\right] \\
&\leqslant \sum_{\mathbf{r}\in\Omega} \mathbb{P}\left[\left|\left\langle \mathbf{u}, (\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(m,n)}(\mathbf{r})\right\rangle\right| \geqslant \left\|(\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(m,n)}(\mathbf{r})\right\|_2 \frac{\varepsilon}{c_5 \xi}\right] \qquad (8.55) \\
&\leqslant |\Omega| 4 e^{-\frac{(\varepsilon/(c_5\xi))^2}{4}} \qquad (8.56) \\
&\leqslant \delta \qquad (8.57)
\end{aligned}$$

where (8.55) follows from (8.58) below, (8.56) follows by Hoeffding's inequality (cf. Lemma 8.10), and to obtain (8.57) we used the assumption (8.48) with $c_6 = 1/(2c_5)$.

To complete the proof, note that by (8.37) we have $\left\|\mathbf{L}-\bar{\mathbf{L}}\right\| \leqslant 2.5\xi$ on $\mathscr{E}_\xi$. Thus, conditioned on $\mathscr{E}_\xi$,

$$\left\|(\mathbf{L}-\bar{\mathbf{L}})^H \bar{\mathbf{v}}^{(m,n)}(\mathbf{r})\right\|_2 \leqslant \left\|\mathbf{L}-\bar{\mathbf{L}}\right\| \left\|\bar{\mathbf{v}}^{(m,n)}(\mathbf{r})\right\|_2 \leqslant 2.5\xi \left\|\bar{\mathbf{v}}^{(m,n)}(\mathbf{r})\right\|_1 \leqslant c_5 \xi, \qquad (8.58)$$

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

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

Here, c_5 is a numerical constant, and we used [9, C.12, Table 6] and $N/\kappa \leqslant 0.5514$.

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 8.7 to establish that $Q^{(m,n)}(\mathbf{r})$ is close to $\bar{Q}^{(m,n)}(\mathbf{r})$ with high probability uniformly for all $\mathbf{r} \in [0,1]^2$.

LEMMA 8.11 Let $\varepsilon, \delta > 0$. If

$$L \geqslant S\frac{c}{\varepsilon^2}\log^3\left(\frac{c'L^6}{\delta\varepsilon^2}\right) \tag{8.59}$$

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

$$\max_{\mathbf{r}\in[0,1]^2,(m,n):\ m+n\leqslant 2} \frac{1}{\kappa^{n+m}} \left|Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r})\right| \leqslant \varepsilon. \tag{8.60}$$

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

$$\max_{\mathbf{r}\in[0,1]^2} \min_{\mathbf{r}_g\in\Omega} \left|\mathbf{r}-\mathbf{r}_g\right| \leqslant \frac{\varepsilon}{3\tilde{c}L^{5/2}}. \tag{8.61}$$

The cardinality of the set Ω is

$$|\Omega| = \left(\frac{3\tilde{c}L^{5/2}}{\varepsilon}\right)^2 = c'L^5/\varepsilon^2. \tag{8.62}$$

First, we use Lemma 8.7 to show that $\left|Q^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r}_g)\right|$ is small for all points $\mathbf{r}_g \in \Omega$. Note that the condition of Lemma 8.7 is satisfied by assumption (8.59). Using the union bound over all 6 pairs (m,n) obeying $m+n \leqslant 2$, it now follows from Lemma 8.7, that

$$\left\{ \max_{\mathbf{r}_g\in\Omega, m+n\leqslant 2} \frac{1}{\kappa^{m+n}} \left|Q^{(m,n)}(\mathbf{r}_g) - \bar{Q}^{(m,n)}(\mathbf{r}_g)\right| \leqslant \frac{\varepsilon}{3} \right\} \tag{8.63}$$

holds with probability at least $1-6\delta' = 1-\frac{\delta}{2}$. Here, δ' is the original δ in Lemma 8.7. Next, we will prove that this result continues to hold uniformly for all $\mathbf{r} \in [0,1]^2$. To this end, we will also need that the event

$$\left\{ \max_{\mathbf{r}\in[0,1]^2, m+n\leqslant 2} \frac{1}{\kappa^{m+n}} \left|Q^{(m,n)}(\mathbf{r})\right| \leqslant \frac{\tilde{c}}{2}L^{3/2} \right\} \tag{8.64}$$

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 (8.63) and (8.64) hold simultaneously with probability at least $1-\delta$. The proof is concluded by noting that (8.63) and (8.64) imply (8.60) (see Section 8.4.2).

8.4.1 *Proof of the fact that* (8.64) *holds with probability at least* $1-\frac{\delta}{2}$*:* In order to show that (8.64) holds with probability at least $1-\frac{\delta}{2}$, we first upper-bound $|Q^{(m,n)}(\mathbf{r})|$. By (8.46),

$$\begin{aligned}\frac{1}{\kappa^{m+n}}\left|Q^{(m,n)}(\mathbf{r})\right| &= \left|\left\langle \mathbf{L}\mathbf{u}, \mathbf{v}^{(m,n)}(\mathbf{r})\right\rangle\right| \\ &\leqslant \|\mathbf{L}\|\,\|\mathbf{u}\|_2\left\|\mathbf{v}^{(m,n)}(\mathbf{r})\right\|_2 \\ &\leqslant \|\mathbf{L}\|\sqrt{S}\left\|\mathbf{v}^{(m,n)}(\mathbf{r})\right\|_2 \\ &\leqslant \|\mathbf{L}\|\sqrt{S}\sqrt{3S}\left\|\mathbf{v}^{(m,n)}(\mathbf{r})\right\|_\infty \\ &= \|\mathbf{L}\|\sqrt{3}S\max_{j,(m',n')\in\{(0,0),(1,0),(0,1)\}}\frac{1}{\kappa^{m+m'+n+n'}}\left|G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right|, \end{aligned} \tag{8.65}$$

where we used $\|\mathbf{u}\|_2=\sqrt{S}$, since the entries of $\mathbf{u}$ are ± 1. Next, note that, for all $\mathbf{r}$ and all $\mathbf{r}_j$, we have, by (8.16)

$$\begin{aligned}\frac{1}{\kappa^{m+m'+n+n'}}\left|G^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right| &= \frac{1}{\kappa^{m+m'+n+n'}}\mathbf{x}^H\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\mathbf{x} \leqslant \frac{1}{\kappa^{m+m'+n+n'}}\|\mathbf{x}\|_2^2\left\|\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right\| \\ &\leqslant c_1\frac{(2\pi N)^{m+m'+n+n'}}{\kappa^{m+m'+n+n'}}\sqrt{L}\|\mathbf{x}\|_2^2 \leqslant c_1 12^{\frac{m+m'+n+n'}{2}}\sqrt{L}\|\mathbf{x}\|_2^2 \\ &\leqslant c_1 12^{\frac{3}{2}}\sqrt{L}\|\mathbf{x}\|_2^2, \end{aligned} \tag{8.66}$$

where we used Lemma 8.3 to conclude $\left\|\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right\| \leqslant \left\|\mathbf{V}^{(m,n)}_{(m',n')}(\mathbf{r},\mathbf{r}_j)\right\|_F \leqslant c_1(2\pi N)^{m+m'+n+n'}\sqrt{L}$ and (8.66) follows from $m+m'+n+n'\leqslant 3$ (recall that $m+n\leqslant 2$). Substituting (8.66) into (8.65) and using that $S\leqslant L$ (by assumption (8.59)) yields

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

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

$$\begin{aligned}\mathbb{P}\left[\max_{\mathbf{r}\in[0,1]^2,m+n\leqslant 2}\frac{1}{\kappa^{m+n}}\left|Q^{(m,n)}(\mathbf{r})\right| \geqslant \frac{\tilde{c}}{2}L^{3/2}\right] &\leqslant \mathbb{P}\left[\|\mathbf{L}\|\,\|\mathbf{x}\|_2^2 \geqslant 2.5\cdot 3\right] \\ &\leqslant \mathbb{P}[\|\mathbf{L}\|\geqslant 2.5]+\mathbb{P}\left[\|\mathbf{x}\|_2^2\geqslant 3\right] \qquad (8.67)\\ &\leqslant \frac{\delta}{2} \qquad (8.68)\end{aligned}$$

as desired. Here, (8.67) follows from the union bound and (8.68) follows from $\mathbb{P}[\|\mathbf{L}\|\geqslant 2.5]\leqslant\mathbb{P}\left[\overline{\mathscr{E}}_{1/4}\right]\leqslant\frac{\delta}{4}$ (by (8.36) and application of Lemma 8.6; note that the condition of Lemma 8.6 is satisfied by (8.59)) and $\mathbb{P}\left[\|\mathbf{x}\|_2^2\geqslant 3\right]\leqslant\frac{\delta}{4}$, shown below. Using that $4\log(4/\delta)\leqslant L$ (by (8.59)), we obtain

$$\begin{aligned}\mathbb{P}\left[\|\mathbf{x}\|_2^2\geqslant 3\right] &\leqslant \mathbb{P}\left[\|\mathbf{x}\|_2^2\geqslant 2\left(1+\frac{2\log(4/\delta)}{L}\right)\right] \\ &\leqslant \mathbb{P}\left[\|\mathbf{x}\|_2\geqslant\left(1+\frac{\sqrt{2\log(4/\delta)}}{\sqrt{L}}\right)\right] \leqslant e^{-\frac{2\log(4/\delta)}{2}}=\frac{\delta}{4}, \end{aligned} \tag{8.69}$$

where we used $\sqrt{2(1+\beta^2)} \geqslant (1+\beta)$, for all β, and a standard concentration inequality for the norm of a Gaussian random vector, e.g., [32, Eq. 1.6]. This concludes the proof of (8.64) holding with probability at least $1-\frac{\delta}{2}$.

8.4.2 *Proof of the fact that* (8.63) *and* (8.64) *imply* (8.60)*:* Consider a point $\mathbf{r} \in [0,1]^2$ and let $\mathbf{r}_g$ be the point in Ω closest to $\mathbf{r}$ in ∞-distance. By the triangle inequality,

$$\frac{1}{\kappa^{n+m}}\left|Q^{(m,n)}(\mathbf{r})-\bar{Q}^{(m,n)}(\mathbf{r})\right| \leqslant \frac{1}{\kappa^{n+m}}\left[\left|Q^{(m,n)}(\mathbf{r})-Q^{(m,n)}(\mathbf{r}_g)\right|+\left|Q^{(m,n)}(\mathbf{r}_g)-\bar{Q}^{(m,n)}(\mathbf{r}_g)\right|+\left|\bar{Q}^{(m,n)}(\mathbf{r}_g)-\bar{Q}^{(m,n)}(\mathbf{r})\right|\right]. \quad (8.70)$$

We next upper-bound the terms in (8.70) 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 (8.70) can be upper-bounded according to

$$\begin{aligned}\left|Q^{(m,n)}(\mathbf{r})-Q^{(m,n)}(\mathbf{r}_g)\right| &= \left|Q^{(m,n)}(\tau,\nu)-Q^{(m,n)}(\tau,\nu_g)+Q^{(m,n)}(\tau,\nu_g)-Q^{(m,n)}(\tau_g,\nu_g)\right| \\ &\leqslant \left|Q^{(m,n)}(\tau,\nu)-Q^{(m,n)}(\tau,\nu_g)\right|+\left|Q^{(m,n)}(\tau,\nu_g)-Q^{(m,n)}(\tau_g,\nu_g)\right| \\ &\leqslant |\nu-\nu_g|\sup_z\left|Q^{(m,n+1)}(\tau,z)\right|+|\tau-\tau_g|\sup_z\left|Q^{(m+1,n)}(z,\nu_g)\right| \\ &\leqslant |\nu-\nu_g|2\pi N\sup_z\left|Q^{(m,n)}(\tau,z)\right|+|\tau-\tau_g|2\pi N\sup_z\left|Q^{(m,n)}(z,\nu_g)\right|, \quad (8.71)\end{aligned}$$

where (8.71) 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 8.3 (Bernstein's polynomial inequality [23, Cor. 8]) Let $p(\theta)$ be a trigonometric polynomial of degree N with complex coefficients p_k, i.e., $p(\theta)=\sum_{k=-N}^{N} p_k e^{i2\pi\theta k}$. Then

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

Substituting (8.64) into (8.71) yields

$$\frac{1}{\kappa^{m+n}}\left|Q^{(m,n)}(\mathbf{r})-Q^{(m,n)}(\mathbf{r}_g)\right| \leqslant \frac{\tilde{c}}{2}L^{5/2}(|\tau-\tau_g|+|\nu-\nu_g|) \leqslant \tilde{c}L^{5/2}\left|\mathbf{r}-\mathbf{r}_g\right| \leqslant \frac{\varepsilon}{3}, \quad (8.72)$$

where the last inequality follows from (8.61).

We next upper-bound the third absolute value in (8.70). Using steps analogous to those leading to (8.72), we obtain

$$\frac{1}{\kappa^{m+n}}\left|\bar{Q}^{(m,n)}(\mathbf{r}_g)-\bar{Q}^{(m,n)}(\mathbf{r})\right| \leqslant \frac{\varepsilon}{3}. \quad (8.73)$$

Substituting (8.63), (8.72), and (8.73) into (8.70) yields that

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

This concludes the proof of Lemma 8.11. □

8.5 *Step 3c: Ensuring that* $|Q(\mathbf{r})| < 1$ *for all* $\mathbf{r} \notin \mathscr{S}$

LEMMA 8.12 Suppose that

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

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

1. For all $\mathbf{r}$, that satisfy $\min_{\mathbf{r}_j \in \mathscr{S}} \left|\mathbf{r}-\mathbf{r}_j\right| \geqslant 0.2447/N$ we have that $|Q(\mathbf{r})| < 0.9963$.
2. For all $\mathbf{r} \notin \mathscr{S}$ that satisfy $0 < \left|\mathbf{r}-\mathbf{r}_j\right| \leqslant 0.2447/N$ for some $\mathbf{r}_j \in \mathscr{S}$, we have that $|Q(\mathbf{r})| < 1$.

Proof. Choose $\varepsilon = 0.0005$. It follows from Lemma 8.11 that

$$\frac{1}{\kappa^{n+m}}\left|Q^{(m,n)}(\mathbf{r}) - \bar{Q}^{(m,n)}(\mathbf{r})\right| \leqslant 0.0005 \tag{8.74}$$

for all $(m,n)\colon m+n \leqslant 2$, and for all $\mathbf{r}$ with probability at least $1-\delta$. To prove the lemma we will show that statements 1 and 2 follow from (8.74) and certain properties of $\bar{Q}^{(m,n)}(\mathbf{r})$ established in [9].

Statement 1 follows directly by combining (8.74) with the following result via the triangle inequality.

PROPOSITION 8.4 ([9, Lem. C.4]) For all $\mathbf{r}$, that satisfy $\min_{\mathbf{r}_j \in \mathscr{S}} |\mathbf{r}-\mathbf{r}_j| \geqslant 0.2447/N$ we have that $|Q(\mathbf{r})| < 0.9958$.

In order to prove statement 2, assume without loss of generality that $\mathbf{0} \in \mathscr{S}$, and consider $\mathbf{r}$ with $|\mathbf{r}| \leqslant 0.2447/N$. Statement 2 is established by showing that the Hessian matrix of $\tilde{Q}(\mathbf{r}) \coloneqq |Q(\mathbf{r})|$, i.e.,

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

is negative definite. This is done by showing that

$$\operatorname{trace}(\mathbf{H}) = \tilde{Q}^{(2,0)} + \tilde{Q}^{(0,2)} < 0 \tag{8.75}$$

$$\det(\mathbf{H}) = \tilde{Q}^{(2,0)}\tilde{Q}^{(0,2)} - (\tilde{Q}^{(1,1)})^2 > 0, \tag{8.76}$$

which implies that both eigenvalues of $\mathbf{H}$ are strictly negative. To prove (8.75) and (8.76), we will need the following result.

PROPOSITION 8.5 ([9, Sec. C.2]) For $|\mathbf{r}| \leqslant 0.2447/N$ and for $N \geqslant 512$,

$$1 \geqslant \bar{Q}(\mathbf{r}) \geqslant 0.6447 \tag{8.77}$$

$$\frac{1}{\kappa^2}\bar{Q}^{(2,0)}(\mathbf{r}) \leqslant -0.3550 \tag{8.78}$$

$$\frac{1}{\kappa^2}|\bar{Q}^{(1,1)}(\mathbf{r})| \leqslant 0.3251 \tag{8.79}$$

$$\frac{1}{\kappa^2}|\bar{Q}^{(1,0)}(\mathbf{r})| \leqslant 0.3344. \tag{8.80}$$

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

$$\begin{aligned}
\frac{1}{\kappa^2}\tilde{Q}^{(2,0)} &= -\frac{(Q_RQ_R^{(1,0)} + Q_IQ_I^{(1,0)})^2}{|Q|^3} + \frac{|Q^{(1,0)}|^2 + Q_RQ_R^{(2,0)} + Q_IQ_I^{(2,0)}}{|Q|} \\
&= -\frac{Q_R^2{Q_R^{(1,0)}}^2 + 2Q_RQ_R^{(1,0)}Q_IQ_I^{(1,0)} + Q_I^2{Q_I^{(1,0)}}^2}{|Q|^3} + \frac{{Q_R^{(1,0)}}^2 + {Q_I^{(1,0)}}^2 + Q_RQ_R^{(2,0)} + Q_IQ_I^{(2,0)}}{|Q|} \\
&= \left(1 - \frac{Q_R^2}{|Q|^2}\right)\frac{{Q_R^{(1,0)}}^2}{|Q|} - \frac{2Q_RQ_R^{(1,0)}Q_IQ_I^{(1,0)} + Q_I^2{Q_I^{(1,0)}}^2}{|Q|^3} + \frac{{Q_I^{(1,0)}}^2 + Q_IQ_I^{(2,0)}}{|Q|} + \frac{Q_R}{|Q|}Q_R^{(2,0)}.
\end{aligned} \tag{8.81}$$

By Proposition 8.5, using the triangle inequality, and using the fact that $\bar{Q}^{(m,n)}(\mathbf{r})$ is real, the following bounds are in force:

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

Using these bounds in (8.81) with $\varepsilon = 0.0005$ we obtain $\tilde{Q}^{(2,0)} < -0.3539$, which implies that (8.75) is satisfied.

It remains to verify (8.76). First note that

$$\begin{aligned}
&\frac{1}{\kappa^2}\tilde{Q}^{(1,1)} \\
&= \frac{Q_R^{(1,1)}Q_R + Q_R^{(1,0)}Q_R^{(0,1)} + Q_I^{(1,1)}Q_I + Q_I^{(1,0)}Q_I^{(0,1)}}{|Q|} - \frac{(Q_R^{(0,1)}Q_R + Q_I^{(0,1)}Q_I)(Q_R^{(1,0)}Q_R + Q_I^{(1,0)}Q_I)}{|Q|^3} \\
&= Q_R^{(1,1)}\frac{Q_R}{|Q|} + \frac{Q_R^{(1,0)}Q_R^{(0,1)}}{|Q|}\left(1 - \frac{Q_R^2}{|Q|^2}\right) + \frac{Q_I^{(1,1)}Q_I + Q_I^{(1,0)}Q_I^{(0,1)}}{|Q|} \\
&- \frac{Q_R^{(0,1)}Q_RQ_I^{(1,0)}Q_I + Q_I^{(0,1)}Q_I(Q_R^{(1,0)}Q_R + Q_I^{(1,0)}Q_I)}{|Q|^3}.
\end{aligned} \tag{8.82}$$

Using the bounds above in (8.82) yields, with $\varepsilon = 0.0005$, that $\frac{1}{\kappa^2}|\tilde{Q}^{(1,1)}| \leqslant 0.3267$. With $\frac{1}{\kappa^2}\tilde{Q}^{(2,0)} < -0.3539$, it follows that the RHS of (8.75) can be lower-bounded by

$$\frac{1}{\kappa^2}(0.3539^2 - 0.3267^2) = \frac{1}{\kappa^2}0.01855 > 0,$$

i.e., (8.76) holds. This concludes the proof of Statement 2. □

Funding

VM was supported by the Swiss National Science Foundation fellowship for advanced researchers under grant PA00P2_139678, and RH was supported by the Swiss National Science Foundation under grant P2EZP2_159065.

Acknowledgments

RH would like to thank Céline Aubel, Helmut Bölcskei, Emmanuel Candès, and Nora Loose for helpful discussions. RH would also like to thank Emmanuel Candès for his hospitality during a visit to the Statistics Department at Stanford, and Helmut Bölcskei for his support and for initiating this visit. We would also like to thank the referees for helpful comments and suggestions, which greatly improved the manuscript.

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, 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. *Math. Prog. Comp.*, 3(3):165–218, 2011.

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

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

[6] A. Beurling and L. Carleson. *The collected works of Arne Beurling: Complex analysis*, volume 1. Birkhauser, 1989.

[7] B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Trans. Signal Process.*, 61(23):5987–5999, 2013.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, 2004.

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

[10] 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.

[11] C. Carathéodory. Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen. *Rend. Circ. Mat.*, 32:193–217, 1911.

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

[13] L. Demanet, D. Needell, and N. Nguyen. Super-resolution via superset selection and pruning. In *Proc. 10th International Conf. on Sampling Theory and Applications*, Bremen, Germany, 2013.

[14] L. Demanet and N. Nguyen. The recoverability limit for superresolution via sparsity. *arXiv:1502.01385*, 2014.

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

[16] B. Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer, Dordrecht, The Netherlands, 2007.

[17] G. Durisi, V. I. Morgenshtern, and H. Bölcskei. On the sensitivity of continuous-time noncoherent fading channel capacity. *IEEE Trans. Inf. Theory*, 58(10):6372–6391, 2012.

[18] A. C. Fannjiang. The MUSIC algorithm for sparse objects: A compressed sensing analysis. *Inverse Probl.*, 27(3):035013, 2011.
[19] C. Fernandez-Granda. Super-Resolution of point sources via convex programming. *arXiv:1507.07034*, 2015.
[20] J. J. Fuchs. Sparsity and uniqueness for some specific under-determined linear systems. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, volume 5, pages 729–732, 2005.
[21] A. Gershman and N. Sidiropoulos, editors. *Space-Time Processing for MIMO Communications*. Wiley, Chichester; Hoboken, NJ, 2005.
[22] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.
[23] L. A. Harris. Bernstein's polynomial inequalities and functional analysis. *Irish Math. Soc. Bull.*, 36:19–33, 1996.
[24] R. Heckel and H. Bölcskei. Identification of sparse linear operators. *IEEE Trans. Inf. Theory*, 59(12):7985–8000, 2013.
[25] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.*, 57(6):2275–2284, 2009.
[26] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 2nd edition, 2012.
[27] A. Jakobsson, A. L. Swindlehurst, and P. Stoica. Subspace-based estimation of time delays and Doppler shifts. *IEEE Trans. Signal Process.*, 46(9):2472–2483, 1998.
[28] T. Kailath. Measurements on time-variant communication channels. *IRE Trans. Inf. Theory*, 8(5):229–236, 1962.
[29] W. Kozek and G. E. Pfander. Identification of operators with bandlimited symbols. *SIAM J. Math. Anal.*, 37(3):867–888, 2005.
[30] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Commun. Pur. Appl. Math.*, 67(11):1877–1904, 2014.
[31] A. Lapidoth. *A Foundation in Digital Communication*. Cambridge University Press, 2009.
[32] M. Ledoux and M. Talagrand. *Probability in Banach spaces: Isoperimetry and processes*. Springer, New York, NY, 1991.
[33] W. Liao and A. Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *Appl. Comput. Harmon. Anal.*, 2014.
[34] A. Moitra. The threshold for super-resolution via extremal functions. *arXiv:1408.1681*, 2014.
[35] V. I. Morgenshtern and E. J. Candès. Stable super-resolution of positive sources: The discrete setup. *arXiv:1504.00717*, 2014.
[36] G. E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.
[37] G. E. Pfander and D. F. Walnut. Measurement of time-variant linear channels. *IEEE Trans. Inf. Theory*, 52(11):4808–4820, 2006.
[38] A. Quinquis, E. Radoi, and F. C. Totir. Some radar imagery results using superresolution techniques. *IEEE Trans. Antennas Propag.*, 52(5):1230–1244, 2004.
[39] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18(0), 2013.
[40] G. Schiebinger, E. Robeva, and B. Recht. Superresolution without separation. *arXiv:1506.03144*, 2015.
[41] D. Slepian. On bandwidth. *Proc. IEEE*, 64(3):292–300, 1976.
[42] M. Soltanolkotabi. *Algorithms and Theory for Clustering and Nonconvex Quadratic Programming*. Stanford Ph.D. Dissertation, Standord, CA, 2014.
[43] P. Stoica and R. L. Moses. *Spectral Analysis of Signals*. Prentice Hall, Upper Saddle River, N.J., 2005.
[44] T. Strohmer. Pseudodifferential operators and Banach algebras in mobile communications. *Appl. Comput. Harmon. Anal.*, 20(2):237–249, 2006.
[45] G. Tang, B. N. Bhaskar, and B. Recht. Sparse recovery over continuous dictionaries-just discretize. In *Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047, Pacific Grove, CA, 2013.

[46] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Inform. Theory*, 59(11):7465–7490, 2013.
[47] 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.
[48] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.
[49] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing: Theory and applications*, pages 210–268. Cambridge University Press, New York, NY, 2012.
[50] Z. Yang, L. Xie, and P. Stoica. Vandermonde Decomposition of Multilevel Toeplitz Matrices with Application to Multidimensional Super-Resolution. *arXiv:1505.02510*, 2015.

A. Equivalence of (1.2) and (2.1)

Note that with

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

we obtain

$$\begin{aligned}\sum_{r=-N}^{N} e^{i2\pi\frac{rp}{L}} D_N\Big(\frac{r}{L}-\nu_j\Big) &= \sum_{r=-N}^{N} e^{i2\pi\frac{rp}{L}} \frac{1}{L}\sum_{k=-N}^{N} e^{i2\pi k\left(\frac{r}{L}-\nu_j\right)} \\ &= \sum_{k=-N}^{N} e^{-i2\pi k\nu_j} \frac{1}{L}\sum_{r=-N}^{N} e^{i2\pi\frac{r}{L}(p+k)} \\ &= e^{i2\pi p\nu_j}. \end{aligned} \tag{A.2}$$

Using (A.2) in (2.1) and using again (A.1), we get

$$\begin{aligned} y_p &= \sum_{j=1}^{S} b_j e^{i2\pi p\nu_j} \sum_{\ell=-N}^{N} \frac{1}{L}\sum_{k=-N}^{N} e^{i2\pi k\left(\frac{\ell}{L}-\tau_j\right)} x_{p-\ell} \\ &= \sum_{j=1}^{S} b_j e^{i2\pi p\nu_j} \frac{1}{L}\sum_{\ell,k=-N}^{N} e^{-i2\pi k\tau_j} e^{i2\pi(p-\ell)\frac{k}{L}} x_\ell, \end{aligned}$$

where we used that x_ℓ is L-periodic. This is (1.2), as desired.

B. Proof of (4.3)

First, write (4.1) in its equivalent form

$$y(t) = \int L_H(t,f) X(f) e^{i2\pi ft} df, \tag{B.1}$$

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

$$L_H(t,f) \coloneqq \iint s_H(\tau,\nu) e^{i2\pi(\nu t-\tau f)} d\tau d\nu. \tag{B.2}$$

Since $x(t)$ is band-limited to $[-B/2, B/2]$, we may write

$$X(f) = X(f)H_I(f), \quad H_I(f) \coloneqq \begin{cases} 1, & |f| \leqslant B/2 \\ 0, & \text{else.} \end{cases}$$

For $y(t)$ on $[-T/2, T/2]$ we may write

$$y(t) = y(t)h_O(t), \quad h_O(t) \coloneqq \begin{cases} 1, & |t| \leqslant T/2 \\ 0, & \text{else.} \end{cases}$$

With the input band-limitation and the output time-limitation, (B.1) becomes

$$y(t) = \int \overline{L_H}(t,f)X(f)e^{i2\pi ft}df, \tag{B.3}$$

where

$$\overline{L_H}(t,f) \coloneqq L_H(t,f)h_O(t)H_I(f) \tag{B.4}$$

i.e., the effect of input band-limitation and output time-limitation is accounted for by passing the probing signal through a system with time varying transfer function given by $\overline{L_H}$. The spreading function $\overline{s_H}$ of the system (B.3) and the time-varying transfer function $\overline{L_H}$ are related by the two-dimensional Fourier transform in (B.2). We see that $\overline{L_H}(t,f)$ is "band-limited" with respect to t and f, and hence, by the sampling theorem, $\overline{s_H}$ can be expressed in terms of its samples as

$$\overline{s_H}(\tau,\nu) = \sum_{m,\ell\in\mathbb{Z}} \overline{s_H}\left(\frac{m}{B}, \frac{l}{T}\right) \operatorname{sinc}\left(\left(\tau - \frac{m}{B}\right)B\right) \operatorname{sinc}\left(\left(\nu - \frac{\ell}{T}\right)T\right). \tag{B.5}$$

In terms of $\overline{s_H}(\tau,\nu)$ (B.3) can be written as

$$y(t) = \iint \overline{s_H}(\tau,\nu)x(t-\tau)e^{i2\pi\nu t}d\nu d\tau \tag{B.6}$$

and with (B.5)

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

According to (B.2) and (B.4), $\overline{s_H}(\tau,\nu)$ and $s_H(\tau,\nu)$ are related as

$$\overline{s_H}(\tau,\nu) = \int\int s_H(\tau',\nu')\operatorname{sinc}((\tau-\tau')B)\operatorname{sinc}((\nu-\nu')T)d\tau' d\nu'. \tag{B.7}$$

This concludes the proof of (4.3).

C. Proof of (4.11) and (4.12)

Starting with (4.8), we first use that $BT = L$ and $\tau_j = \bar{\tau}_j \frac{B}{L}$ and $\nu_j = \bar{\nu}_j \frac{T}{L}$ to obtain

$$\tilde{y}_p = \sum_{j=1}^{S} b_j \left(\sum_{\ell \in \mathbb{Z}} \operatorname{sinc}(\ell - \tau_j L) \tilde{x}_{p-\ell} \right) \left(\sum_{k \in \mathbb{Z}} \operatorname{sinc}(k - \nu_j L) e^{i2\pi \frac{kp}{BT}} \right). \tag{C.1}$$

Next, changing the order of summation according to $k = r + Lq$ with $r = N, ..., N$ and $q \in \mathbb{Z}$ yields

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \operatorname{sinc}(k - \nu_j L) e^{i2\pi \frac{kp}{BT}} &= \sum_{r=-N}^{N} \sum_{q \in \mathbb{Z}} \operatorname{sinc}\Big(\Big(\frac{r}{L} - \nu_j + q \Big) L \Big) e^{i2\pi \frac{(r+Lq)p}{L}} \\ &= \sum_{r=-N}^{N} e^{i2\pi \frac{rp}{L}} \sum_{q \in \mathbb{Z}} \operatorname{sinc}\Big(\Big(\frac{r}{L} - \nu_j + q \Big) L \Big) \\ &= \sum_{r=-N}^{N} e^{i2\pi \frac{rp}{L}} D_N\Big(\frac{r}{L} - \nu_j \Big). \end{aligned} \tag{C.2}$$

Here, (C.2) follows from the definition of the Dirichlet kernel in (4.9). Next, note that

$$\begin{aligned} \sum_{\ell \in \mathbb{Z}} \tilde{x}_{p-\ell} \operatorname{sinc}(\ell - L\tau_j) &= \sum_{\ell \in \mathbb{Z}} \tilde{x}_\ell \operatorname{sinc}(p - \ell - L\tau_j) \\ &= \sum_{\tilde{\ell}=-N}^{N} x_{\tilde{\ell}} \sum_{k=-1}^{1} \operatorname{sinc}\big(p - \tilde{\ell} - Lk - L\tau_j\big) \qquad \text{(C.3)} \\ &= \sum_{\ell=-N}^{N} x_\ell \tilde{D}_N \left(\frac{p-\ell}{L} - \tau_j \right), \qquad \text{(C.4)} \end{aligned}$$

where (C.3) follows from $\tilde{x}(\ell/B) = \tilde{x}_\ell = x_\ell$ for $\ell \in [-N-L, L+N]$ and $\tilde{x}(\ell/B) = 0$ for all other ℓ, and from L-periodicity of x_ℓ; (C.4) follows by the definition of $\tilde{D}_N(t)$ in (4.10). Substituting (C.2) and (C.4) into (C.1) yields (4.11), as desired.

Next, consider the case where $x(t)$ is T-periodic. In this case, the samples $x_\ell = x(\ell/B)$ are L-periodic. Similarly to the derivation above, changing the order of summation of the sum in the second brackets in (C.1) according to $\ell = \tilde{\ell} + Lk$ yields

$$\sum_{\ell \in \mathbb{Z}} \operatorname{sinc}(\ell - \tau_j L) x_{p-\ell} = \sum_{\tilde{\ell}=-N}^{N} x_{p-\tilde{\ell}} \sum_{k \in \mathbb{Z}} \operatorname{sinc}\left(\left(\frac{\tilde{\ell}}{L} - \tau_j + k \right) L \right) = \sum_{\ell=-N}^{N} D_N\left(\frac{\ell}{L} - \tau_j \right) x_{p-\ell}, \tag{C.5}$$

where we used that x_ℓ is L-periodic in the first equality, and the last equality follows by definition of the Dirichlet kernel in (4.9). Using (C.2) and (C.5) in (C.1) yields (4.12).

D. Proof of Proposition 4.1

By (A.2), (C.2) is equal to $e^{i2\pi p \nu_j}$. Using this in (4.11) and (4.12) yields

$$\tilde{y}_p - y_p = \sum_{j=1}^{S} b_j e^{i2\pi \nu_j p} \underbrace{\sum_{\ell=-N}^{N} \varepsilon_{j,p,\ell} x_\ell}_{\varepsilon_{j,p}},$$

where we defined the error

$$\varepsilon_{j,p,\ell} \coloneqq \tilde{D}_N\left(\frac{p-\ell}{L}-\tau_j\right) - D_N\left(\frac{p-\ell}{L}-\tau_j\right).$$

For $t \in [-1.5,..,1.5]$, we have

$$\tilde{D}_N(t) - D_N(t) \leqslant c/L.$$

Using that $\frac{p-\ell}{L}-\tau_j \in [-1.5,1.5]$ it follows that $\varepsilon_{j,p,\ell} \leqslant c/L$.

Recall that the b_j have random sign and the x_ℓ are i.i.d. $\mathcal{N}(0,1/L)$ distributed. By the union bound, we have, for all $\beta > 0$,

$$\mathbb{P}\left[|\tilde{y}_p - y_p| \geqslant \beta\|\mathbf{b}\|_2 c\frac{\beta}{L}\right] \leqslant \mathbb{P}\left[\left|\sum_{j=1}^{S} b_j e^{i2\pi p\nu_j}\varepsilon_{j,p}\right| \geqslant \beta\|\mathbf{b}\|_2 \max_{j,p}|\varepsilon_{j,p}|\right] + \mathbb{P}\left[\max_{j,p}|\varepsilon_{j,p}| \geqslant c\frac{\beta}{L}\right] \quad \text{(D.1)}$$

$$\leqslant (4+2L^2)e^{-\frac{\beta^2}{L}}, \quad \text{(D.2)}$$

where (D.2) is established immediately below. Setting $\alpha = \beta^2$ completes the proof of Proposition 4.1.

It is left to establish (D.2). With

$$\left(\sum_{j=1}^{S}\left(|b_j|e^{i2\pi p\nu_j}\varepsilon_{j,p}\right)^2\right)^{1/2} \leqslant \beta\|\mathbf{b}\|_2 \max_{j,p}|\varepsilon_{j,p}|$$

the first probability in (D.1) can be upper-bounded by

$$\mathbb{P}\left[\left|\sum_{j=1}^{S}\operatorname{sign}(b_j)|b_j|e^{i2\pi p\nu_j}\varepsilon_{j,p}\right| \geqslant \beta\left(\sum_{j=1}^{S}\left(|b_j|e^{i2\pi p\nu_j}\varepsilon_{j,p}\right)^2\right)^{1/2}\right] \leqslant 4e^{-\frac{\beta^2}{2}}$$

where we applied Hoeffding's inequality, i.e., Lemma 8.10 (recall that the b_j have random sign). By the union bound, the second probability in (D.1) is upper-bounded by

$$\mathbb{P}\left[\max_{j,p}\left|\varepsilon_{j,p}\right| \geqslant c\frac{\beta}{L}\right] \leqslant \sum_{j,p}\mathbb{P}\left[\left|\varepsilon_{j,p}\right| \geqslant c\frac{\beta}{L}\right] \leqslant SL2e^{-\frac{\alpha^2}{2}}, \quad \text{(D.3)}$$

where the last inequality is proven as follows. Since the x_ℓ are i.i.d. $\mathcal{N}(0,1/L)$, we have that $\varepsilon_{j,p} = \sum_{\ell=-N}^{N}\varepsilon_{\ell,p,j}x_\ell$ is Gaussian with variance $\frac{1}{L}\sum_{\ell=-N}^{N}\varepsilon_{\ell,p,j}^2 \leqslant \frac{c^2}{L^2}$, where we used that $\varepsilon_{\ell,p,j} \leqslant c/L$. Eq. (D.3) now follows from a standard bound on the tail probability of a Gaussian [31, Prop. 19.4.2] random variable.

E. Proof of Theorem 5.1

The following proposition—standard in the theory of compressed sensing (see e.g., [10])—shows that the existence of a certain dual polynomial guarantees that $\mathrm{L1}(\mathbf{y})$ in (5.3) succeeds in reconstructing $\mathbf{s}$.

PROPOSITION E.1 Let $\mathbf{y} = \mathbf{R}\mathbf{s}$ and let $\mathscr{S}$ denote the support of $\mathbf{s}$. Assume that $\mathbf{R}_{\mathscr{S}}$ has full column rank. If there exists a vector $\mathbf{v}$ in the row space of $\mathbf{R}$ with

$$\mathbf{v}_{\mathscr{S}} = \operatorname{sign}(\mathbf{s}_{\mathscr{S}}) \quad \text{and} \quad \left\|\mathbf{v}_{\overline{\mathscr{S}}}\right\|_\infty < 1 \quad \text{(E.1)}$$

then $\mathbf{s}$ is the unique minimizer of $\mathrm{L1}(\mathbf{y})$ in (5.3).

The proof now follows directly from Proposition 8.1. To see this, set $\mathbf{u} = \operatorname{sign}(\mathbf{s}_{\mathscr{S}})$ in Proposition 8.1 and consider the polynomial $Q(\mathbf{r})$ from Proposition 8.1. Define $\mathbf{v}$ as $[\mathbf{v}]_{(m,n)} = Q([m/K, n/K])$ and note that $\mathbf{v}$ satisfies (E.1) since $Q([m/K,n/K]) = \operatorname{sign}(\mathbf{s}_{(m,n)})$ for $(m,n) \in \mathscr{S}$ and $|Q([m/K,n/K])| < 1$ for $(m,n) \notin \mathscr{S}$.

F. Bound on U

We have that

$$U(t) = \hat{c}(2\pi N)^{-m} \sum_{p=-N}^{N} \min\left(1, \frac{1}{p^4}\right) P^{(m)}(p/L - t) \tag{F.1}$$

with

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

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 $t \in [0, 1/2]$.

For $m = 0$, we have that

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

Next, consider the case $m = 1$, and assume that $t \geqslant 1/L$. 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 $\sin(\pi t) \geqslant 2t \geqslant 2/L$ for $1/L \leqslant t \leqslant 1/2$ we get

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

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 $\sin(\pi t) \geqslant 2t \geqslant 2/L$ for $1/L \leqslant t \leqslant 1/2$ we get

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

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

$$|P^{(m)}(t)| \leqslant (L\pi)^m \frac{c_1}{M|\sin(\pi t)|} \leqslant \underbrace{1.0039 c_1}_{c_2 :=} (2\pi N)^m \frac{1}{M|\sin(\pi t)|}, \tag{F.2}$$

where c_1 is a numerical constant and where we used that $(L/(2N))^m \leqslant 1.0039$ for $N \geqslant 512$ and $m \leqslant 4$. Regarding the range $0 \leqslant |t| \leqslant 1/L$, simply note that by Bernstein's polynomial inequality (cf. Proposition 8.3) we have, for all t, from $|P^{(0)}(t)| \leqslant 4$, that

$$|P^{(m)}(t)| \leqslant 4(2\pi N)^m. \tag{F.3}$$

With $c_2 \geqslant 1$, application of (F.2) and (F.3) on (F.1) yields

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

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

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

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

G. Proof of Proposition 3.2

The argument is standard, see e.g., [46, Prop. 2.4]. By definition, $\mathbf{q}$ is dual feasible. To see this, note that

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

where the last inequality holds by assumption. By (3.4), we obtain

$$\langle \mathbf{q}, \mathbf{y}\rangle = \left\langle \mathbf{q}, \mathbf{G} \sum_{\mathbf{r}_n \in \mathscr{S}} b_n \mathbf{a}(\mathbf{r}_n) \right\rangle = \sum_{\mathbf{r}_n \in \mathscr{S}} b_n^* \langle \mathbf{q}, \mathbf{G}\mathbf{a}(\mathbf{r}_n)\rangle = \sum_{\mathbf{r}_n \in \mathscr{S}} b_n^* \text{sign}(b_n) = \sum_{\mathbf{r}_n \in \mathscr{S}} |b_n| \geqslant \|\mathbf{z}\|_{\mathscr{A}}, \tag{G.2}$$

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

$$\text{Re}\,\langle \mathbf{q}, \mathbf{y}\rangle = \text{Re}\,\langle \mathbf{q}, \mathbf{G}\mathbf{z}\rangle = \text{Re}\left\langle \mathbf{G}^H\mathbf{q}, \mathbf{z}\right\rangle \leqslant \left\|\mathbf{G}^H\mathbf{q}\right\|_{\mathscr{A}^*} \|\mathbf{z}\|_{\mathscr{A}} \leqslant \|\mathbf{z}\|_{\mathscr{A}}$$

where we used (3.4) for the last inequality. We thus have established that $\text{Re}\,\langle \mathbf{q}, \mathbf{y}\rangle = \|\mathbf{z}\|_{\mathscr{A}}$. 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 $\hat{\mathbf{z}} = \sum_{\mathbf{r}_n \in \hat{\mathscr{S}}} \hat{b}_n \mathbf{a}(\mathbf{r}_n)$ with $\|\hat{\mathbf{z}}\|_{\mathscr{A}} = \sum_{\mathbf{r}_n \in \hat{\mathscr{S}}} |\hat{b}_n|$ and $\hat{\mathscr{S}} \neq \mathscr{S}$ is another optimal solution. We then have

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

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

H. Comparison to MUSIC

As mentioned previously, the MUSIC algorithm (and related methods) can not be applied directly to the super-resolution radar problem, since MUSIC in general requires multiple sets of measurements (snapshots). However, as pointed out by Peter Stoica (personal communication, June 2015), multiple measurements can be obtained from a single measurement by sending a periodic input signal. Specifically, let $L = M^2$, and suppose that the entries of $\mathbf{x} \in \mathbb{C}^L$ are M-periodic, i.e., $\mathbf{x} = [\tilde{\mathbf{x}}^T, ..., \tilde{\mathbf{x}}^T]^T$, where $\tilde{\mathbf{x}} \in \mathbb{C}^M$. We next define a version of the time and frequency shift operators defined previously in (1.3)

$$[\mathscr{T}_\tau^{(L)}\mathbf{x}]_p := \frac{1}{L}\sum_{k=0}^{L-1}\left[\left(\sum_{\ell=0}^{L-1} x_\ell e^{-i2\pi\frac{\ell k}{L}}\right) e^{-i2\pi k\tau}\right] e^{i2\pi\frac{pk}{L}}, \quad [\mathscr{F}_\nu^{(L)}\mathbf{x}]_p := x_p e^{i2\pi p\nu}, \quad p = 0, ..., L-1,$$

where we have made the dependence on the length of $\mathbf{x}$ (i.e., L), explicit. By utilizing the fact that the entries of $\mathbf{x}$ are M-periodic, successively applying these two operators gives

$$\mathscr{F}_\nu^{(L)}\mathscr{T}_\tau^{(L)}\mathbf{x} = \begin{bmatrix} \mathscr{F}_\nu^{(M)}\mathscr{T}_{M\tau}^{(M)}\tilde{\mathbf{x}} \\ e^{i2\pi\nu M}\mathscr{F}_\nu^{(M)}\mathscr{T}_{M\tau}^{(M)}\tilde{\mathbf{x}} \\ \vdots \\ e^{i2\pi\nu M(M-1)}\mathscr{F}_\nu^{(M)}\mathscr{T}_{M\tau}^{(M)}\tilde{\mathbf{x}} \end{bmatrix}. \tag{H.1}$$

Let $\mathbf{y}$ be the measurement from (1.2), specifically $\mathbf{y} = \sum_{j=1}^S b_j \mathscr{F}_\nu^{(L)}\mathscr{T}_\tau^{(L)}\mathbf{x}$. We decompose $\mathbf{y}$ into blocks of size M according to $\mathbf{y} = [\mathbf{y}_0, ..., \mathbf{y}_{M-1}]^T$, where $\mathbf{y}_i \in \mathbb{C}^M$ for $i = 0, 1, \ldots, M-1$. Using (H.1) we conclude that $\mathbf{y}_p$ obeys the input-output relationship

$$\mathbf{y}_p = \sum_{j=1}^S \underbrace{b_j e^{i2\pi\nu_j M p}}_{b'_{j,p}} \mathscr{F}_{\nu_j}^{(M)}\mathscr{T}_{M\tau_j}^{(M)}\tilde{\mathbf{x}}, \quad p = 0, ..., M-1. \tag{H.2}$$

The MUSIC algorithm can be applied to the measurements in (H.2) in order to extract the time-frequency shifts (τ_j, ν_j). However, the following three conditions are necessary for MUSIC to succeed:

1. The number of measurements must be quadratic in the number of time-frequency shifts, i.e. $S^2 < M^2 = L$. In comparison, Theorem 3.1 shows that the convex program AN($\mathbf{y}$) in (2.5) succeeds even when S is linear (up to a log-factor) in L.

2. The delays τ are required to be in the interval $\in [0, \frac{1}{M}]$, as $\mathscr{T}^{(M)}_{M\tau}$ is 1-periodic in $M\tau$. In comparison, the atomic norm minimization approach of (2.5) only requires $\tau \in [0,1]$.

3. The frequency shifts ν_j need to be distinct, while the atomic norm minimization approach (2.5) only requires the minimum separation condition to be satisfied either in time, or in frequency.

In the absence of noise and as long conditions 1, 2, and 3 are satisfied MUSIC succeeds in recovering the time-frequency shifts provided the probing sequence $\tilde{\mathbf{x}}$ is chosen appropriately (e.g., by drawing the entries $\tilde{\mathbf{x}}$ i.i.d. uniform from the complex unit disk). However, the numerical experiment below suggests that MUSIC may be significantly more sensitive to noise compared to our convex programming approach, which appears to be a consequence of periodizing the input signal $\mathbf{x}$.

For our numerical experiments we use a setup similar to that in Section 5. We assume that the time-frequency shifts $(\tau_j, \nu_j), j = 1, ..., S$, lie on a fine grid with grid constant $\left(\frac{1}{L\cdot \mathrm{SRF}}, \frac{1}{L\cdot \mathrm{SRF}}\right)$, and super-resolution factor SRF $= 6$. Furthermore, we assume that $(\tau_j, \nu_j) \in [0, 1/M]^2$. We set $M = 17$, choose the time-frequency shifts as $\tau_j = \frac{j}{L}, \nu_j = \frac{j}{L}$, with $j = 1, ..., S$, $S \in \{1, 4, 16\}$, and draw the corresponding attenuation factors b_j i.i.d. uniformly at random from the complex unit disc. Note that by construction, the time-frequency shifts lie on the grid, and Conditions 1, 2, and 3 are satisfied.

We generate the MUSIC-compatible measurements $\mathbf{y}_p$ according to (H.2), where the entries of the probing signal $\tilde{\mathbf{x}} \in \mathbb{C}^M$ are drawn i.i.d. at random from the complex unit disc, and consider the following variant of the MUSIC algorithm. Given the measurements $\mathbf{y}_p, p = 0, ..., M-1$, we compute the matrix of eigenvectors $\mathbf{U} \in \mathbb{C}^{M\times M-S}$ corresponding to the $M - S$ smallest eigenvalues of $\mathbf{Z} = \sum_{p=0}^{M-1} \mathbf{y}_p \mathbf{y}_p^H$. The time-frequency shifts are identified as the S time-frequency shifts $(\tau, \nu) \in \{\left(\frac{k}{L\cdot \mathrm{SRF}}, \frac{\ell}{L\cdot \mathrm{SRF}}\right) : (k,\ell) \in \{0, ..., L\cdot \mathrm{SRF}/M - 1\}^2\}$ that minimize $\left\| \mathbf{U}^H \mathscr{F}^{(M)}_{\nu} \mathscr{T}^{(M)}_{M\tau} \tilde{\mathbf{x}} \right\|_2$.

We compare the MUSIC algorithm to our convex programming approach. Specifically, we generate the probing signal $\mathbf{x} \in \mathbb{C}^L$ by drawing its entries i.i.d. from the complex unit disc, and generate the measurement $\mathbf{y} \in \mathbb{C}^L$ according to (1.2). We then recover the time-frequency shifts with L1-ERR defined in (5.5), where the matrix $\mathbf{R}$ in (5.5) has columns $\mathscr{F}_{\nu}\mathscr{T}_{\tau}\mathbf{x}$, $(\nu, \tau) \in \{\left(\frac{k}{L\cdot \mathrm{SRF}}, \frac{\ell}{L\cdot \mathrm{SRF}}\right) : (k,\ell) \in \{0, ..., L\cdot \mathrm{SRF}/M - 1\}^2\}$.

The results are plotted in Figure 7. While L1-ERR perfectly recovers the position of the time-frequency shifts on the grid as long as SNR $\leqslant$ 15dB, the resolution error of MUSIC is quite large if there are many time-frequency shifts, even for large SNRs (e.g. for SNRs as high as 50dB).

This is not surprising, as MUSIC due to the periodization of the input signal, has to deal with a significantly worse conditioned matrix. To see this, set the time shifts to $\tau_j = 0$ and note that with this choice, writing (H.2) in matrix-vector form yields

$$\mathbf{y}_p = \tilde{\mathbf{F}}\mathbf{b}_p, \quad \tilde{\mathbf{F}} := [\mathscr{F}^{(M)}_{\nu_1}\tilde{\mathbf{x}}, ..., \mathscr{F}^{(M)}_{\nu_S}\tilde{\mathbf{x}}],$$

where $[\mathbf{b}_p]_j = b'_{j,p}$. The matrix $\tilde{\mathbf{F}} \in \mathbb{R}^{M\times S}$ is equivalent to the upper left corner of a $M^2 \times M^2$ DFT matrix with scaled rows. In particular, the rows are scaled by $[\tilde{\mathbf{x}}]_p$, that is, $[\tilde{\mathbf{F}}]_{p,j} = [\tilde{\mathbf{x}}]_p e^{i2\pi \frac{pj}{M^2}}$, for $p = 0, .., M-1$ and $j = 1, ..., S$. As a result this matrix is ill-conditioned, specially when M and S are large.

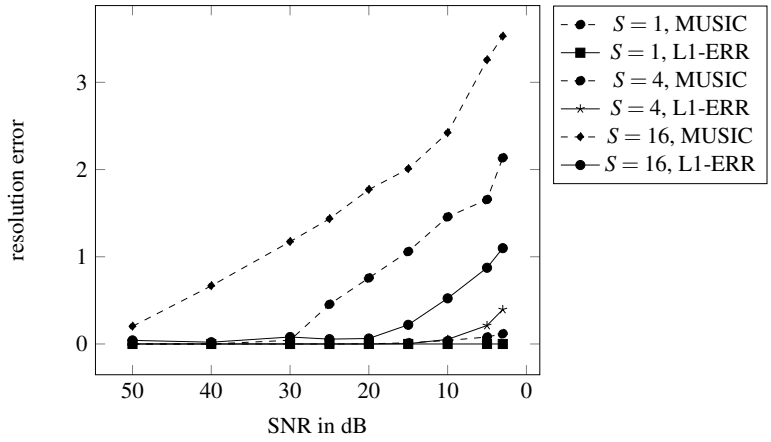

FIG. 7. This figure depicts the resolution error of L1-ERR and MUSIC as a function of the signal-to-noise ratio for a variety of sparsity levels. As in Section 5, the resolution error is $\frac{1}{S}\sum_{j=1}^{S} L\sqrt{(\hat{\tau}_j-\tau_j)^2+(\hat{\nu}_j-\nu_j)^2}$, where $(\hat{\tau}_j,\hat{\nu}_j)$ are the time-frequency shifts obtained by i) applying the MUSIC algorithm to the samples $\mathbf{y}_p+\mathbf{n}_p, p=0,...,M-1$ in (H.2), and ii) applying L1-ERR to the sample $\mathbf{y}+\mathbf{n}$ in (1.2). Here, $\mathbf{n}_p \in \mathbb{C}^M$ and $\mathbf{n} \in \mathbb{C}^L$ are additive Gaussian noise vectors, and the signal-to-noise ratio is calculated via $\mathrm{SNR} := \left(\sum_{p=0}^{M-1} \|\mathbf{y}_p\|_2^2\right) / \left(\sum_{p=0}^{M-1} \|\mathbf{n}_p\|_2^2\right)$ for MUSIC and via $\mathrm{SNR} := \|\mathbf{y}\|_2^2/\|\mathbf{n}\|_2^2$ for L1-ERR.

List of Figures

1 Illustration of the spreading function $s_H(\tau,\nu)$ and the corresponding smeared spreading function $\overline{s_H}(\tau,\nu)$. 11

2 The probing signal $\tilde{x}(t)$ and the real part of its Fourier transform $\tilde{X}(f)$ for $B=1, T=61, L=61$: $\tilde{x}(t)$ is essentially time-limited on an interval of length $3T$ and band-limited to $[-B/2, B/2]$. 12

3 Resolution error $L\sqrt{(\hat{\tau}_j-\tau_j)^2+(\hat{\nu}_j-\nu_j)^2}$ for the recovery of $S=10$ time-frequency shifts from the samples $y_p, p=-N,...,N$ in (4.12) (periodic input signal $x(t)$), and identification from the samples $\tilde{y}_p, p=-N,...,N$ in (4.11) (essentially time-limited input signal $\tilde{x}(t)$) with and without additive Gaussian noise n_p of a certain signal-to-noise ratio $\mathrm{SNR} := \|[\tilde{y}_{-N},...,\tilde{y}_N]\|_2^2/\|[n_{-N},...,n_N]\|_2^2$, by solving L1-ERR. 16

4 Localization of the time-frequency shifts via the estimated dual polynomial $Q(\tau,\nu)$ obtained by solving (6.3) with noiseless measurements. The red lines show the actual positions of the time-frequency shifts located at $(0.2, 0.5)$ and $(0.8, 0.5)$. Note that the estimated dual polynomial satisfies $|Q(\tau,\nu)|=1$ if $(\tau,\nu) \in \{(0.2,0.5),(0.8,0.5)\}$ and $|Q(\tau,\nu)|<1$ otherwise, thereby providing accurate identification of the time-frequency shifts. 18

5 Localization of the time-frequency shifts via the estimated dual polynomial $Q(\tau,\nu)$ obtained by solving (6.6) using noisy measurements (10dB noise). The estimated dual polynomial satisfies $|Q(\tau,\nu)|=1$ for $(\tau,\nu)=(0.4942, 0.7986)$ (marked by $\times$); this is very close to the original time-frequency shift $(0.5, 0.8)$ (marked by $\oplus$). 19

6 Plots of the random kernel $G_{(0,0)}(\mathbf{r},\mathbf{0})/G_{(0,0)}(\mathbf{0},\mathbf{0})$ along with the deterministic kernel $\bar{G}(\mathbf{r})$. 24

7 This figure depicts the resolution error of L1-ERR and MUSIC as a function of the signal-to-noise ratio for a variety of sparsity levels. As in Section 5, the resolution error is $\frac{1}{S}\sum_{j=1}^{S} L\sqrt{(\hat{\tau}_j-\tau_j)^2+(\hat{\nu}_j-\nu_j)^2}$, where $(\hat{\tau}_j,\hat{\nu}_j)$ are the time-frequency shifts obtained by i) applying the MUSIC algorithm to the samples $\mathbf{y}_p+\mathbf{n}_p, p=0,...,M-1$ in (H.2), and ii) applying L1-ERR to the sample $\mathbf{y}+\mathbf{n}$ in (1.2). Here, $\mathbf{n}_p \in \mathbb{C}^M$ and $\mathbf{n} \in \mathbb{C}^L$ are additive Gaussian noise vectors, and the signal-to-noise ratio is calculated via $\mathrm{SNR} := \left(\sum_{p=0}^{M-1} \|\mathbf{y}_p\|_2^2\right) / \left(\sum_{p=0}^{M-1} \|\mathbf{n}_p\|_2^2\right)$ for MUSIC and via $\mathrm{SNR} := \|\mathbf{y}\|_2^2/\|\mathbf{n}\|_2^2$ for L1-ERR. 52