Robustness of Two-Dimensional Line Spectral Estimation Against Spiky Noise

Iman Valiulahi, Farzan Haddadi, and Arash Amini

***Abstract*—The aim of two-dimensional (2D) line spectral estimation is to super-resolve spectral point sources from finite time-domain samples. In many applications such as radar and sonar, spiky noise occurs when electronic instruments temporarily enter cut-off or saturation region. To overcome this problem, a new convex program is presented to simultaneously estimate the spectral point sources and spiky noise in the 2D case. To prove uniqueness of the solution, it is sufficient to show that a dual certificate exists. Construction of the dual certificate imposes a mild condition on the separation of the spectral point sources. Also, number of spikes and detectable spectral sources are shown to be a logarithmic function of number of time samples. Simulation results confirm the validity of our theoretical results.**

***Index Terms*—Two-dimensional line spectral estimation, total variation norm, continuous domain dictionary, convex optimization.**

I. Introduction

Two-dimensional (2D) line spectral estimation (LSE) has received much attention in signal processing in recent years. It is a fundamental concern of many applications such as orthogonal frequency-division multiplexing (OFDM) passive radar [1], super-resolution imaging [2], and conventional radar [3]. In the LSE problem, a sparse combination of sinusoids is observed. A subset of observations is often corrupted by spiky noise because of various reasons. In the OFDM passive radar [1] for instance, the transmitted symbols should be estimated at the receiver side. In some cases, large estimation errors occur leading to the demodulation error at the receiver side. This results in significant errors in the observed signal which is modeled as spiky noise, hence increasing side lobes of the clutter and strong targets. Due to the fact that the reflection of targets is usually small, this noise makes heavy interference with the real targets [1]. Another application of the LSE problem is recovering a high-resolution image of stars from a low-resolution telescope. Meanwhile, spiky noise happens in astronomical images because of thermal generation in pixels or collision of cosmic rays with image sensors [2]. Moreover, this type of perturbation may occur in the super-resolution radar imaging [3] by lightning discharges or telephone switching transients [4]. Recovering delay and Doppler shifts in the super-resolution radar imaging [3] can be significantly affected by spiky noise. The goal of this paper is to investigate the 2D LSE problem in the presence of spiky noise.

Parametric approaches in the LSE problem are based on dividing the observation space into signal and noise subspaces by singular value decomposition [5], [6]. Although the computational complexity of these approaches is low, the sensitivity to additive perturbations is high. Also, they cannot determine the spiky noise location and number of spectral sources.

In another field of study known as compressive sensing, an exact solution can be achieved in an underdetermined linear system of equations by assuming that the signal of interest is sparse in a known discrete dictionary [7]. Furthermore, ℓ_1 minimization can be used to recover support of the signal in the DFT basis. In many practical applications such as radar and sonar, however, the spectral sources belong to a continuous dictionary. Thereby, mismatch between the actual and reconstructed sources in the DFT basis is inevitable [8]. In [9], it is shown that refining the grid greatly increases the computational complexity. Further, ℓ_1 minimization does not achieve the exact solution in this ultrahigh-dimensional setting.

The LSE model incorporates a 2D spectral matrix $\boldsymbol{X} \in \mathbb{C}^{n_1 \times n_2}$ given by

$$X_{\boldsymbol{k}} = \sum_{i=1}^{r} d_i \mathrm{e}^{j2\pi \boldsymbol{f}_i^T \boldsymbol{k}}, \qquad \boldsymbol{k} \in J, \tag{1}$$

where $\boldsymbol{d}$ is the complex amplitude vector, $d_i = |d_i| e^{j\phi_i},\ \phi_i \sim \mathcal{U}(0, 2\pi)$, $|d_i| \sim \delta_{0.5} + \mathcal{X}^2(1)$, r is the level of sparsity, $\boldsymbol{f}_i \in [0,1]^2$. Without loss of generality, assume that $n_1 = n_2 = n$, so $J = \{1, \dots, n\} \times \{1, \dots, n\}$ is the 2D integer square that indicates time indices.

In [10], Candes and Fernandez proposed a non-parametric approach to super-resolve the inherent frequencies in (1) for the 1D and 2D cases. Their approach is based on Total-Variation (TV) norm –the continuous version of the ℓ_1 norm– minimization which promotes the sparsity of a continuous function. It is proven that a linear combination of the fourth power of the Dirichlet kernel and its derivatives can be used to construct a valid dual certificate for this problem. This construction imposes a minimum separation of $4/n$ and $4.76/n$ between the frequency sources in dimensions one and two, respectively. To achieve a sharper bound on minimum separation in the 1D case, [11] constructed the dual certificate using three Dirichlet kernels with different cut-off frequencies. Also, [12] extended this approach to the 2D case. The required minimum separations for the 1D and 2D cases are $2.52/n$ and $3.36/n$, respectively. In [13], 1D signals are observed at a random subset of time instances. Under mild assumptions on the minimum separation and with $\mathcal{O}(r \log r \log n)$ partial time samples, Tang et.al. proved that one can always find a random trigonometric polynomial that estimates the point sources. This approach is extended to the 2D case in [14].

The inherent infinite dimensionality of the LSE problem is generally an obstacle. One can approximate the problem on a

I. Valiulahi and F. Haddadi are with the School of Electrical Engineering, Iran University of Science & Technology (IUST), Tehran, Iran. A. Amini is with the Electrical Engineering Department, Sharif University, Tehran, Iran.

fine grid [15] or solve the TV-norm minimization directly by linear programming [16]. In [10], the associated dual problem of TV-regularization is converted into linear matrix inequalities (LMI) using positive trigonometric polynomial (PTP) theory [17]. However, in higher dimensions, a hierarchy of sum of squares relaxation is required [18]. Magnitude of the trigonometric polynomial in each subband of its frequency domain is controllable by the coefficients that are obtained from PTP theory [19]–[21].

Additive noise is inevitable in most applications. Recently, a significant line of research is focused on the support stability of TV-norm minimization when measurements are corrupted by dense perturbations [22]. Establishing a trade-off between noise power and TV-norm, known as burling least angle regression (BLASSO), is a common technique in such problems [23], [24]. In [25], a precise comparison between robustness of optimization-based methods and conventional approaches against additive noise is provided. Spiky noise is another kind of corruption that may appear due to hardware failure. Subspace decomposition approaches are not able to estimate the sources when a subset of time samples is corrupted by spiky noise. They are indeed designed to overcome Gaussian-type noises and are relatively inefficient against spiky noise.

In [26], a convex optimization problem that incorporates the sparsity feature of spiky noise in the cost function is suggested in order to estimate simultaneously both the spectral sources and spiky noise in the 1D setting. Spiky noise in the 1D case is also studied in [27] and [28]. While many works in the literature study the 1D case, multi-dimensional analysis is required in many applications such as super-resolution imaging [2] and MIMO radar [29]. [30] proposed an innovative approach based on matrix completion that super-resolves the spectral sources and detects spiky noise in the 2D case. The proposed Enhanced Matrix Completion (EMaC) is also shown to find the exact solution if the sample complexity exceeds $\mathcal{O}(r^2 \log^3 n^2)$, under some incoherence conditions.

In this paper, the main contribution is to construct a valid dual certificate for penalized TV-regularization in the 2D LSE problem. The proposed certificate is a 2D trigonometric low-pass polynomial that can interpolate any sign pattern of the signal. This feature is used to localize the support of the spectral sources. Further, its coefficients belong to the interior of the sub-differential of ℓ_1 norm. One can take advantage of this feature to detect the locations of spiky noise. Under mild conditions on the separation between the spectral sources, it has shown that if the numbers of spectral sources and spiky noise occurrences are restricted by a logarithmic function of the number of samples, the suggested semidefinite programming achieves, with high probability, the exact solution. A semidefinite programming using PTP theory is also proposed for the associated dual problem of penalized TV-regularization. Finally, we provide numerical phase transition graphs certifying the fact that both the number of spectral sources and spiky noise are controlled by a logarithmic function of the number of time samples.

The rest of the paper is organized as follows. The problem is formulated in Section II. Penalized TV-norm minimization and our main theorem are presented in Section III. Construction of the dual certificate and implementation of the dual problem are provided in Sections IV and V, respectively. Section VI is devoted to numerical experiments. Finally, conclusions are discussed in Section VII.

Throughout the paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The ith element of the vector $\boldsymbol{x}$ and the $\boldsymbol{k} = [k_1, k_2]^T$ element of the matrix $\boldsymbol{X}$ are given by x_i and $X_{\boldsymbol{k}}$, respectively. The operator $|\cdot|$ denotes the cardinality of sets, absolute value for scalars and element-wise absolute value for vectors and matrices; also $\|\boldsymbol{z}\|_\infty = \max_i |z_i|$. For a function f and a matrix $\boldsymbol{A}$, $\|f\|_\infty$, $\|\boldsymbol{A}\|_\infty$, $\|\boldsymbol{A}\|$ and $\|\boldsymbol{A}\|_1$ are defined as $\sup_t |f(t)|$, $\sup_{\|\boldsymbol{x}\|_\infty \le 1} \|\boldsymbol{A}\boldsymbol{x}\|_\infty = \max_i \sum_j |A_{i,j}|$, $\sup_{\|\boldsymbol{z}\|_2 \le 1} \|\boldsymbol{A}\boldsymbol{z}\|_2$ and $\sum_{i,j} |A(i,j)|$, respectively. The function $\mathrm{relint}(C)$ denotes the relative interior of a set C. The sub-differential of a function f at point x is shown by $\partial f(\cdot)(x)$. ith derivate and i_1, i_2 partial derivatives of a 1D function $f(t)$ and a 2D function $f(\boldsymbol{t} := [t_1, t_2]^T)$ are denoted by $f^i(t)$ and $f^{i_1 i_2}(\boldsymbol{t})$, respectively. The operators $(\cdot)^T$ and $(\cdot)^*$ represent transpose and Hermitian transpose of a vector, respectively. The function $\mathrm{sgn}(\boldsymbol{x})$ is reserved for the element-wise sign of the vector $\boldsymbol{x}$. Also, $\mathrm{vec}(\boldsymbol{X})$ denotes the columns of $\boldsymbol{X}$ being stacked on top of each other. The inner product between two functions f and g is defined as $\langle f, g \rangle := \int f(t) g(t) dt$. The operator $\otimes$ is the Kronecker product and the adjoint of a linear operator $\boldsymbol{\mathcal{F}}$ is denoted by $\boldsymbol{\mathcal{F}}^*$. To show that $\boldsymbol{A}$ is a semidefinite matrix we write $\boldsymbol{A} \succeq 0$.

II. Problem Formulation

In the spectral domain, the signal in (1) is a linear combination of Dirac delta functions:

$$\boldsymbol{\mu} = \sum_{\boldsymbol{f}_i \in T} d_i \delta(\boldsymbol{f} - \boldsymbol{f}_i), \tag{2}$$

where T is the support of the signal and $\delta(\boldsymbol{f} - \boldsymbol{f}_i)$ denotes the Dirac delta function located at $\boldsymbol{f}_i$. The main goal in the LSE problem is to recover the location and amplitude of each delta by finite time-domain samples. As mentioned in the previous section, many practical applications such as radar and sonar suffer from spiky noise due to the electrical instruments failure. Assume that spiky noise is added to the signal (1) as:

$$Y_{\boldsymbol{k}} = X_{\boldsymbol{k}} + Z_{\boldsymbol{k}}, \qquad \boldsymbol{k} \in J, \tag{3}$$

where $Z_{\boldsymbol{k}}$ is an element of the sparse noise matrix $\boldsymbol{Z} \in \mathbb{C}^{n \times n}$ with s non-zero entries. The observation model can be written in the matrix form:

$$\boldsymbol{Y} = \boldsymbol{\mathcal{F}}\boldsymbol{\mu} + \boldsymbol{Z}, \tag{4}$$

where $\boldsymbol{\mathcal{F}}(\cdot)$ is a linear operator that maps any measure in the frequency domain to J in the time domain. The problem is to estimate the spectral sources and locations of spiky noise from $\boldsymbol{Y}$.

III. Robust Total Variation Minimization

Spectral sparsity is not single-handedly sufficient to solve this problem. In fact, if two sources are located too close to each other, it would be impossible to resolve them [10].

Definition 3.1: Let $\mathbb{T}^2$ be the 2D torus obtained by identifying the endpoints on $[0,1]^2$. For each set of points $T \subset \mathbb{T}^2$, the minimum separation is defined as:

$$\begin{aligned}\Delta(T) &:= \inf_{\boldsymbol{t}_i,\boldsymbol{t}_j\in T,\ \boldsymbol{t}_i\neq\boldsymbol{t}_j} \|\boldsymbol{t}_i-\boldsymbol{t}_j\|_\infty \\ &= \inf_{i\neq j} \max\{|t_{1i}-t_{1j}|, |t_{2i}-t_{2j}|\},\end{aligned} \tag{5}$$

where $|t_{1i}-t_{1j}|, |t_{2i}-t_{2j}|$ denote the wrap-around distances on the unit circle.

The ℓ_1 minimization is not suitable for the LSE problem, due to the discretization of the spectral domain. Whereas, the TV-norm defined as $\|\boldsymbol{\nu}\|_{\mathrm{TV}} := \sup_\rho \sum_{E\in\rho} |\boldsymbol{\nu}(E)|$ can promote sparsity in continuous functions. Indeed, TV-norm maximizes the disjoint sum of positive measures $|\boldsymbol{\nu}(\cdot)|$ over all partitions ρ of the signal domain. In the special case for (2), $\|\boldsymbol{\mu}\|_{\mathrm{TV}} = \sum_{i=1}^r |d_i|$.

Most inverse problems are solved by minimizing a cost function that promotes an inherent structure of the signal of interest. This concept emerges in compressed sensing [7]. The cost function may also include a specific penalty term to perform side tasks. For instance, [31] has shown that penalizing ℓ_1 norm with an ℓ_2 error term is an efficient way for denoising. The observation model (4) is the sum of two sparse signals in different domains. [26] balances the TV-norm of spectral sources and the ℓ_1 norm of spiky noise in the 1D case. We generalize this approach to the 2D case by introducing the following optimization problem:

$$\mathrm{P_{TN}}: \quad \min_{\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{Z}}} \ \|\tilde{\boldsymbol{\mu}}\|_{\mathrm{TV}} + \lambda\|\tilde{\boldsymbol{Z}}\|_1 \ \text{ subject to } \ \boldsymbol{Y} = \boldsymbol{\mathcal{F}}\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{Z}},$$

where $\lambda > 0$ is a regularization parameter that makes a trade-off between TV-norm of spectral spikes and ℓ_1 norm of spiky noise. The first goal of this paper is to prove that $\mathrm{P_{TN}}$ achieves exact recovery. The following theorem states that the solution of $\mathrm{P_{TN}}$ is exact, under some specific conditions.

Theorem 3.1: Let $J = \{1,\dots,n\}\times\{1,\dots,n\}$ be the set of indices of observed entries in the matrix, $\boldsymbol{Y} = \boldsymbol{\mathcal{F}}\boldsymbol{\mu} + \boldsymbol{Z}$, where each element of the noise matrix is independently non-zero with probability $\frac{s}{n^2}$ supported on the set $\Omega \subset J$ ($|\Omega| = s$). Also, the support of $\boldsymbol{\mu} = \sum_{\boldsymbol{f}_i\in T} d_i\delta(\boldsymbol{f}-\boldsymbol{f}_i)$, obeys $\Delta(T) \geq \frac{3.36}{n-1}$, where $\boldsymbol{f}_i \in [0,1]^2$ and $|T| = r$. If $r+s \leq n^2$,

$$\begin{aligned} &r \leq C_r\big(\log\frac{n^2}{\epsilon}\big)^{-2}n^2, \qquad && s \leq C_s\big(\log\frac{n^2}{\epsilon}\big)^{-2}n^2, \\ &n \geq 4\times 10^3, && \lambda = \frac{1}{n}, \end{aligned} \tag{6}$$

which C_r and C_s are numerical constants, then, the exact solution of $\mathrm{P_{TN}}$ is $(\boldsymbol{\mu},\boldsymbol{Z})$ with probability $1-\epsilon$ for any $\epsilon \in [0,1]$.

Therefore, one can estimate the spectral sources and spiky noise samples up to a logarithmic function of the number of time-domain samples under a mild condition on the separation of the spectral sources. Note that the large amount of $n \geq 4\times10^3$ in Theorem 3.1 should not be considered as a sampling budget. This is only required for mathematical proofs and in the phase transition graphs (see Fig. 2) we show that $\mathrm{P_{TN}}$ works for even small amounts of n.

IV. Construction of the Dual Certificate

The following Proposition justifies that if there exists a low-pass trigonometric polynomial that belongs to the interior of the sub-differential of TV-norm and its coefficient lies on the interior of the sub-differential of ℓ_1 norm, it will be a valid certificate for $\mathrm{P_{TN}}$.

Proposition 4.1: (Proof in Appendix C). Under the conditions of Theorem 3.1, for any sign patterns $\boldsymbol{h} \in \mathbb{C}^{|T|}$ and $\boldsymbol{r} \in \mathbb{C}^{|\Omega|}$ such that $|h_i| = 1$ and $|r_l| = 1$, for all i and l, if there exists the following 2D low-pass trigonometric polynomial

$$\boldsymbol{\mathcal{F}}^*\boldsymbol{C} = Q(\boldsymbol{f}) = \sum_{\boldsymbol{k}\in J} C_{\boldsymbol{k}} e^{-j2\pi\boldsymbol{f}^T\boldsymbol{k}}, \tag{7}$$

such that

$$Q(\boldsymbol{f}_i) = h_i, \qquad \forall \boldsymbol{f}_i \in T, \tag{8}$$

$$|Q(\boldsymbol{f})| < 1, \qquad \forall \boldsymbol{f} \notin T, \tag{9}$$

$$\frac{C_{\boldsymbol{k}_l}}{\lambda} = r_l, \qquad \forall \boldsymbol{k}_l \in \Omega, \tag{10}$$

$$|C_{\boldsymbol{k}}| < \lambda, \qquad \forall \boldsymbol{k} \notin \Omega, \tag{11}$$

where $\boldsymbol{f}_i = [f_{1i}, f_{2i}]^T$ and $\boldsymbol{k}_l = [k_{1l}, k_{2l}]^T$, then $(\boldsymbol{\mu},\boldsymbol{Z})$ is the unique solution of $\mathrm{P_{TN}}$.

The conditions (8) and (9) state that $Q(\boldsymbol{f}) \in \mathrm{relint}(\partial(\|\cdot\|_{\mathrm{TV}}(\boldsymbol{\mu})))$, so for any measure $\hat{\boldsymbol{\mu}}$, we have $\|\boldsymbol{\mu}+\hat{\boldsymbol{\mu}}\|_{\mathrm{TV}} \geq \|\boldsymbol{\mu}\|_{\mathrm{TV}} + \langle Q, \hat{\boldsymbol{\mu}}\rangle$. Similarly, it can be deduced from (10) and (11) that $\frac{\boldsymbol{C}}{\lambda} \in \mathrm{relint}(\partial(\|\cdot\|_1(\boldsymbol{Z})))$, so for any $\hat{\boldsymbol{Z}}$, we have $\|\boldsymbol{Z}+\hat{\boldsymbol{Z}}\|_1 \geq \|\boldsymbol{Z}\|_1 + \langle\frac{\boldsymbol{C}}{\lambda}, \hat{\boldsymbol{Z}}\rangle$. Let $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}+\hat{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{Z}} = \boldsymbol{Z}+\hat{\boldsymbol{Z}}$ as a feasible point such that $\boldsymbol{Y} = \boldsymbol{\mathcal{F}}\bar{\boldsymbol{\mu}} + \bar{\boldsymbol{Z}}$:

$$\begin{aligned} &\|\bar{\boldsymbol{\mu}}\|_{\mathrm{TV}} + \lambda\|\bar{\boldsymbol{Z}}\|_1 \geq \|\boldsymbol{\mu}\|_{\mathrm{TV}} + \lambda\|\boldsymbol{Z}\|_1 + \langle Q, \bar{\boldsymbol{\mu}}-\boldsymbol{\mu}\rangle \\ &+ \lambda\langle\frac{\boldsymbol{C}}{\lambda}, \bar{\boldsymbol{Z}}-\boldsymbol{Z}\rangle_F \geq \|\boldsymbol{\mu}\|_{\mathrm{TV}} + \lambda\|\boldsymbol{Z}\|_1 \\ &+ \langle\boldsymbol{C}, \boldsymbol{\mathcal{F}}^*(\bar{\boldsymbol{\mu}}-\boldsymbol{\mu}) + \bar{\boldsymbol{Z}}-\boldsymbol{Z}\rangle_F \geq \|\boldsymbol{\mu}\|_{\mathrm{TV}} + \lambda\|\boldsymbol{Z}\|_1. \end{aligned} \tag{12}$$

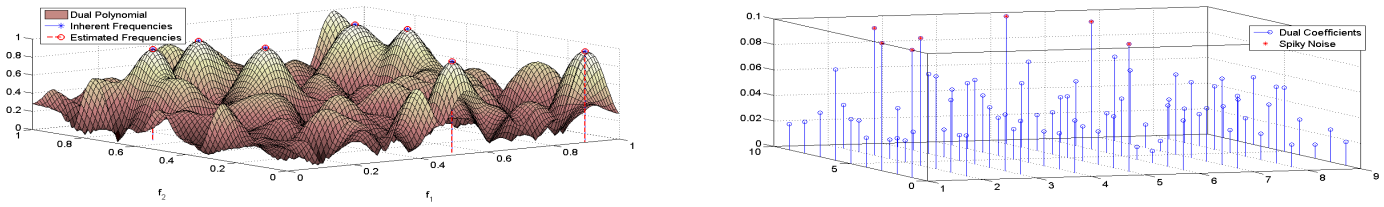
Hence, this concludes the existence and uniqueness of the solution $(\boldsymbol{\mu},\boldsymbol{Z})$. The coefficient of $Q(\boldsymbol{f})$ can be obtained by solving the dual problem of $\mathrm{P_{TN}}$, the following section claims this issue.

V. The Dual Problem

Though the primal problem $\mathrm{P_{TN}}$ can be implemented using Vandermonde decomposition of Toeplitz matrices [32], we are more interested in the dual approach because it gives us the chance to simultaneously detect both the ground-truth spectral sources and the location of spiky noise. The infinite dimensional $\mathrm{P_{TN}}$ is indeed converted into a tractable problem. At the first step, the dual problem of $\mathrm{P_{TN}}$ is obtained by Lagrange theorem. The results of PTP theory is then applied to convert the explicit constraint of the dual problem into LMIs [17].

The associated dual problem of $\mathrm{P_{TN}}$ is given by

$$\begin{aligned} &\max_{\boldsymbol{C}\in\mathbb{C}^{n\times n}} \langle\boldsymbol{C}, \boldsymbol{Y}\rangle \\ &\text{subject to } \ \|\boldsymbol{\mathcal{F}}^*\boldsymbol{C}\|_\infty \leq 1, \ \ \|\boldsymbol{C}\|_\infty \leq \lambda, \end{aligned} \tag{13}$$



(a) (b)

Fig. 1: Frequency and spiky noise localization. In Fig.1(a), the support of the spectral and estimated sources are represented by the blue and red lines, respectively. The magnitude of dual solution and support of spiky noise are shown in the blue lines and red stars in Fig. 1(b), respectively.

where $\boldsymbol{C}$ is the dual variable. Due to the establishment of the Slater's conditions, there is no gap between the objective value of the dual and primal problem. Therefore, $\langle \hat{\boldsymbol{C}}, \boldsymbol{Y} \rangle = \langle \hat{\boldsymbol{C}}, \boldsymbol{\mathcal{F}}\boldsymbol{\mu} + \boldsymbol{Z} \rangle = \|\boldsymbol{\mu}\|_{\mathrm{TV}} + \lambda \|\boldsymbol{Z}\|_1$, $\langle \boldsymbol{\mathcal{F}}^* \hat{\boldsymbol{C}}, \boldsymbol{\mu} \rangle + \langle \hat{\boldsymbol{C}}, \boldsymbol{Z} \rangle = \langle \mathrm{sgn}(\boldsymbol{\mu}), \boldsymbol{\mu} \rangle + \lambda \langle \mathrm{sgn}(\boldsymbol{Z}), \boldsymbol{Z} \rangle$. Consequently, $|\boldsymbol{\mathcal{F}}^* \hat{\boldsymbol{C}}| = 1$ and $|\hat{\boldsymbol{C}}_{\boldsymbol{k}}| = \lambda$ if $\boldsymbol{f} \in T$ and $\boldsymbol{k} \in \Omega$, respectively. This provides a strategy to recover the support of the spectral sources and spiky noise (see Fig 1).

The magnitude of the trigonometric polynomial can be controlled by LMIs using the results of the PTP theory [17]. In dimension two, we need to consider the hierarchy of sum of squares relaxations [18]. Let us define sum-of-squares relaxation degree vector $\boldsymbol{n'} := [n'_1, n'_2]^T$. Without loss of generality, assume that $n' = n'_1 = n'_2$. One can define a zero-padded extension $\tilde{\boldsymbol{C}}$ of $\boldsymbol{C}$ under $n' \geq n$, $\tilde{J} = \{1, \cdots, n'\} \times \{1, \cdots, n'\}$, such that $\tilde{c}_{\boldsymbol{k}} = \begin{cases} c_{\boldsymbol{k}} & \text{if } \boldsymbol{k} \in J \\ 0 & \text{otherwise} \end{cases}$. Regarding the results of [18], problem (13) can be written as:

$$\begin{aligned} \max_{\boldsymbol{C}, \boldsymbol{Q}_0} \quad & \langle \boldsymbol{Y}, \boldsymbol{C} \rangle \\ & \text{subject to } \delta_{\boldsymbol{k}} = \mathrm{tr}[\boldsymbol{\Theta}_{\boldsymbol{k}} \boldsymbol{Q}_0], \quad \boldsymbol{k} \in \tilde{J}, \\ & \begin{bmatrix} \boldsymbol{Q}_0 & \mathrm{vec}(\tilde{\boldsymbol{C}}) \\ (\mathrm{vec}(\tilde{\boldsymbol{C}}))^H & \mathbf{1} \end{bmatrix} \succeq 0, \quad \|\boldsymbol{C}\|_\infty \leq \lambda, \end{aligned} \tag{14}$$

where $\boldsymbol{Q}_0 \in \mathbb{C}^{n'^2 \times n'^2}$ is a Hermitian matrix such that $\boldsymbol{Q}_0 \succeq 0$, $\boldsymbol{\Theta}_{\boldsymbol{k}} = \boldsymbol{\Theta}_{k_2} \otimes \boldsymbol{\Theta}_{k_1}$, $\boldsymbol{\Theta}_k \in \mathbb{C}^{n' \times n'}$ is an elementary Toeplitz matrix with ones on it's k-th diagonal and zeros elsewhere. Notice that $k = 0$ is associated with the main diagonal, positive and negative values are reserved for the upper and lower diagonals, respectively. Also, $\delta_{\mathbf{0}} = 1$ and $\delta_{\boldsymbol{k}} = 0$ if $\boldsymbol{k} \neq \mathbf{0}$. To support the theoretical results, in the next section, numerical experiments are presented.

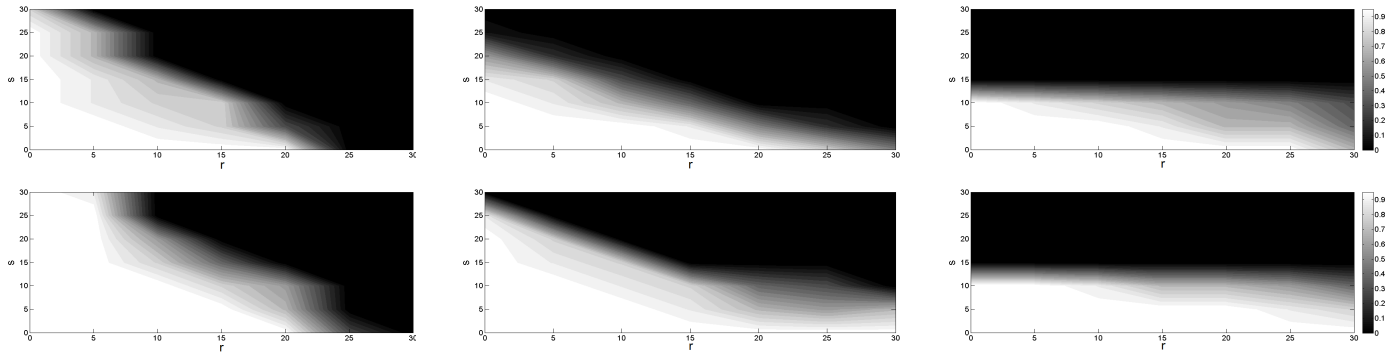
VI. Experiments

In this section, numerical experiments for observation model (4) are provided to investigate performance of the proposed positive semidefinite programing (14). In the first experiment, $r = 7$ frequency sources, without any separation condition, in the domain $[0, 1]^2$ and $s = 7$ spiky noises in the set J are randomly generated. The magnitude and phase of each sinusoid is randomly generated from $\delta_{0.5} + \mathcal{X}^2(1)$ and $\mathcal{U}(0, 2\pi)$, respectively. Problem (14) is implemented by CVX [33] under $n' = n$ and the technique described in Section V is leveraged in order to localize the support of the frequency sources and spiky noise. Fig. 1 demonstrates that the local extremums of $|\boldsymbol{\mathcal{F}}^* \hat{\boldsymbol{C}}|$ that achieve one and the locations of $|\hat{\boldsymbol{C}}|$ that achieve λ are associated with the inherent frequencies of (1) and the locations of spiky noise, respectively.

In the second experiment, under the fixed minimum separation condition $\frac{3}{n-1}$ and $n' = n$ in (14), the phase transitions of the proposed approach for different amounts of the regularization parameter λ are depicted. The regularization parameter is indeed varied when varying k and s (Fig. 2). As mentioned, λ makes a balance between the structure of two different components. Small λ more strongly penalizes TV-norm of the spectral sources than ℓ_1 norm of spiky noise. This leads to a cost function which is more appropriate to manifest the time domain sparsity. Fig. 2 verifies this claim when λ is changed from small to large values. The first and second rows in Fig. 2 are respectively related to different numbers of measurements $n^2 = 64$ and $n^2 = 81$. The estimation is considered successful if the normalized mean squared error $\|\bar{\boldsymbol{X}} - \hat{\boldsymbol{X}}\|_2 / \|\bar{\boldsymbol{X}}\|_2 \leq 10^{-3}$, where $\bar{\boldsymbol{X}}$ and $\hat{\boldsymbol{X}}$ are respectively associated with $\boldsymbol{X}$ and reconstructed data when the corresponding indices of spiky noise are eliminated. The grayscale images show the empirical success rate and each point is related to a specific (r, s, n^2). Under the minimum separation condition, one can verify that the number of detectable ground-truth spectral sources and spiky noise are controlled by a logarithmic function of the number of samples as discussed in Theorem 3.1.

It is beneficial for attentive readers to interpret phase transition graphs in the OFDM passive radar [1]. Explicitly, suppose that we have additional knowledge that the environment is highly corrupted by demodulation error (spiky noise). Our simulation results reveal that if we choose small λ, the optimization problem (14) is able to detect a more number of noise spikes. The larger λ is the solution for the case of knowing in advance that the number of delay-Doppler sources is large.

It is worth mentioning that we adopt a brute-force search method with the step size $1/200$ along each axis and find the locations at which the magnitude of the dual polynomial above $1 - 10^{-5}$ to recover the spectral sources. This search method is problematic in two aspects. First, the brute-force approach resorts back to discretization. Note that the main goal of using TV norm instead of ℓ_1 norm in the compressed sensing is to avoid this discretization [22]. Second, user-parameters are required for the above implementation for which no role of thumb exists. Unlike the search method



(d) (e) (f)

Fig. 2: Grayscale images, under the fixed minimum separation condition $\frac{3}{n-1}$, show the empirical rate of success of (14) over 10 trials. First and second rows are corresponding to $n^2 = 64$ and $n^2 = 81$, respectively. Also, each column from left to right respectively show the results for $\lambda = 0.1$, $\lambda = 0.125$ and $\lambda = 0.2$.

proposed here, authors in [32] presented an analytical approach in which the frequencies are computed by the numerical solution of the SDP. The presented approach in [32] can be leveraged in our problem. Once the SDP in (14) is solved using CVX [33], one is able to obtain the solution of the 2-level Toeplitz matrix associated with the primal problem $\mathrm{P_{TN}}$ for free. The frequencies can then be computed by performing a Vandermonde decomposition on the 2-level Toeplitz matrix [32]. This approach seems to be more appropriate in practical implementation since it involves much easier user parameter tuning and can be generalized to high dimensions in a straightforward manner.

VII. Conclusion and Future Direction

In this paper, the 2D LSE problem, when a subset of time domain samples is corrupted by spiky noise, is investigated. In addition, a semidefinite program that achieves the exact solution under mild conditions on number of spike noise number and separation of spectral sources is proposed. One can extend the suggested approach to arbitrary dimensions which is the fundamental concern of many applications such as MIMO radar [34]. Also, it would be nice if one studies model (3) in compressed sensing regime in which a random subset of time-domain samples which some of them are corrupted by spike noise is available. Analyzing 2D LSE problem in the presence of both dense and spiky perturbations would be an interesting future research direction. One only needs to restrict the affine constraint of $\mathrm{P_{TN}}$ to the power of dense noise though deriving theoretical justifications might be very sophisticated.

VIII. Acknowledgment

Iman Valiulahi is thankful to Sajad Daei and Hamid Fathi for their helpful discussions of this problem during his master degree at IUST.

Appendix A
Deterministic Certificate

Without loss of generality, assume that $J = \{-m, \cdots, m\} \times \{-m, \cdots, m\}$ where $m = \frac{n-1}{2}$ or $m = \frac{n}{2} - 1$ when n is odd or even, respectively. To construct the dual certificate in the presence of spiky noise, it is necessary to consider the LSE problem in the noiseless case. In [10], it is shown that the following polynomial can estimate the frequency sources under a mild condition on their separations:

$$\bar{Q}(\boldsymbol{f}) = \sum_{\boldsymbol{k} \in J} \bar{C}_{\boldsymbol{k}} e^{-j2\pi \boldsymbol{f}^T \boldsymbol{k}}, \tag{15}$$

such that

$$\bar{Q}(\boldsymbol{f}_i) = h_i, \qquad \forall \boldsymbol{f}_i \in T, \tag{16}$$
$$|\bar{Q}(\boldsymbol{f})| < 1, \qquad \forall \boldsymbol{f} \notin T. \tag{17}$$

The same approach in [10] and [12] is followed to construct a deterministic dual certificate as below:

$$\bar{Q}(\boldsymbol{t}) = \sum_{\boldsymbol{f}_i \in T} \bar{\alpha}_i \bar{K}(\boldsymbol{f} - \boldsymbol{f}_i) + \bar{\beta}_{1i} \bar{K}^{10}(\boldsymbol{f} - \boldsymbol{f}_i) + \bar{\beta}_{2i} \bar{K}^{01}(\boldsymbol{f} - \boldsymbol{f}_i), \tag{18}$$

where $\bar{\boldsymbol{\alpha}}$, $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are interpolation vectors. To meet (16) and (17), the following conditions are sufficient

$$\bar{Q}(\boldsymbol{f}_i) = h_i, \qquad \boldsymbol{f}_i \in T, \tag{19}$$
$$\nabla \bar{Q}(\boldsymbol{f}_i) = 0, \qquad \boldsymbol{f}_i \in T. \tag{20}$$

In [12], authors suggested $\bar{K}(\boldsymbol{f}) = \bar{K}_{\boldsymbol{\gamma}}(f_1)\bar{K}_{\boldsymbol{\gamma}}(f_2)$, to construct the dual certificate, in which $\bar{K}_{\boldsymbol{\gamma}}(f) = \prod_{i=1}^{3} K(\gamma_i m, f) = \sum_{k=-m}^{m} c_k e^{j2\pi k f}$, where $K(\bar{m}, f) = \frac{1}{2\bar{m}+1} \sum_{k=-\bar{m}}^{\bar{m}} e^{j2\pi k f}$ is known as the Dirichlet kernel, $\gamma_1 = 0.247$, $\gamma_2 = 0.339$, $\gamma_3 = 0.414$, and $\boldsymbol{c} \in \mathbb{C}^n$ is the convolution of the Fourier coefficients of $k(\gamma_1 m, f)$, $k(\gamma_2 m, f)$, and $k(\gamma_3 m, f)$. Conditions (19) and (20) can be reformulated as a matrix equation:

$$\underbrace{\begin{bmatrix} \bar{\boldsymbol{E}}_{00} & \kappa \bar{\boldsymbol{E}}_{10} & \kappa \bar{\boldsymbol{E}}_{01} \\ -\kappa \bar{\boldsymbol{E}}_{10} & -\kappa^2 \bar{\boldsymbol{E}}_{20} & -\kappa^2 \bar{\boldsymbol{E}}_{11} \\ -\kappa \bar{\boldsymbol{E}}_{01} & -\kappa^2 \bar{\boldsymbol{E}}_{11} & -\kappa^2 \bar{\boldsymbol{E}}_{02} \end{bmatrix}}_{\bar{\boldsymbol{E}}} \begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \kappa^{-1} \bar{\boldsymbol{\beta}}_1 \\ \kappa^{-1} \bar{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \tag{21}$$

where $(\bar{\boldsymbol{E}}_{i_1 i_2})_{\ell, j} = \bar{K}^{i_1 i_2}(\boldsymbol{f}_\ell - \boldsymbol{f}_j)$ and $\kappa := \frac{1}{\sqrt{|K^2|(0)}}$. We borrow bounds on $\|\boldsymbol{c}\|_\infty$ and κ from [11] which are useful in

our proof

$$\|\boldsymbol{c}\|_\infty \leq \frac{1.3}{m}, \tag{22}$$

$$\frac{0.467}{m} \leq \kappa \leq \frac{0.468}{m}, \quad \text{for } m \geq 2\times 10^3. \tag{23}$$

APPENDIX B
RANDOM CERTIFICATE

Spiky noise randomly corrupts a subset of time samples. It is similar to random sampling in compressed sensing literature [13]. [26] used the same technique to incorporate the randomness of spiky noise into the dual certificate construction. We follow the proposed approach to construct a valid certificate for the 2D case when a subset of time samples does not follow the exponential structure as considered. First, $Q(\boldsymbol{f})$ is divided into two terms $Q(\boldsymbol{f}) := Q_{\text{aux}}(\boldsymbol{f}) + R(\boldsymbol{f})$, where $Q_{\text{aux}}(\boldsymbol{f}) := \sum_{\boldsymbol{k}\in\Omega^c} C_{\boldsymbol{k}} e^{-j2\pi \boldsymbol{f}^T\boldsymbol{k}}$, and $R(\boldsymbol{f}) := \frac{1}{n}\sum_{\boldsymbol{k}_l\in\Omega} r_l e^{-j2\pi \boldsymbol{f}^T\boldsymbol{k}_l}$. The coefficients of the first and second terms are restricted to Ω^c and Ω, respectively. From the definitions, it is obvious that (10) is satisfied and $R(\boldsymbol{f})$ has no degrees of freedom. So, it is necessary to construct $Q_{\text{aux}}(\boldsymbol{f})$ to guarantee other conditions in Proposition 4.1. The inequality $|Q(\boldsymbol{f})| < 1,\ \forall \boldsymbol{f} \notin T$ is satisfied by setting to zero the partial derivatives of $Q(\boldsymbol{f})$ at T. Thus,

$$Q_{\text{aux}}(\boldsymbol{f}_i) = h_i - R(\boldsymbol{f}_i), \qquad \forall \boldsymbol{f}_i \in T, \tag{24}$$

$$\nabla Q_{\text{aux}}(\boldsymbol{f}_i) = -\nabla R(\boldsymbol{f}_i), \qquad \forall \boldsymbol{f}_i \in T. \tag{25}$$

Let us define a restricted version of $\bar{K}$ in Ω^c as below:

$$K(\boldsymbol{f}) := \sum_{\boldsymbol{k}\in\Omega^c} c_{k_1}c_{k_2}e^{j2\pi \boldsymbol{f}^T\boldsymbol{k}} = \sum_{\boldsymbol{k}\in J} \delta_{\Omega^c}(\boldsymbol{k}) c_{k_1}c_{k_2}e^{j2\pi \boldsymbol{f}^T\boldsymbol{k}},$$

where $\delta_{\Omega^c}(\boldsymbol{k}) = 1$ when $\boldsymbol{k} = \Omega^c$, $\delta_{\Omega^c}(\boldsymbol{k}) = 0$ otherwise. Under the noise condition of Theorem 3.1, these are independent Bernoulli random variables with parameter $\frac{n^2-s}{n^2}$. Hence, the expectation of $K(\boldsymbol{f})$ can be written as:

$$\mathbb{E}(K(\boldsymbol{f})) = \frac{n^2-s}{n^2}\sum_{k\in J} c_{k_1}c_{k_2}e^{j2\pi \boldsymbol{f}^T\boldsymbol{k}} = \frac{n^2-s}{n^2}\bar{K}(\boldsymbol{f}).$$

The mean of partial derivatives of $K(\boldsymbol{f})$ can be obtained by the same technique. The function $Q_{\text{aux}}(\boldsymbol{f})$ is constructed by a linear combination of $K(\boldsymbol{f})$ and its partial derivatives as:

$$Q_{\text{aux}}(\boldsymbol{t}) = \sum_{\boldsymbol{f}_i\in T} \alpha_i K(\boldsymbol{f}-\boldsymbol{f}_i) + \beta_{1i}K^{10}(\boldsymbol{f}-\boldsymbol{f}_i) + \beta_{2i}K^{01}(\boldsymbol{f}-\boldsymbol{f}_i), \tag{26}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2 \in \mathbb{C}^{|T|}$ are interpolation coefficient vectors. The conditions (24) and (25) can be recast in the matrix form as follows:

$$\underbrace{\begin{bmatrix} \boldsymbol{E}_{00} & \kappa\boldsymbol{E}_{10} & \kappa\boldsymbol{E}_{01} \\ -\kappa\boldsymbol{E}_{10} & -\kappa^2\boldsymbol{E}_{20} & -\kappa^2\boldsymbol{E}_{11} \\ -\kappa\boldsymbol{E}_{01} & -\kappa^2\boldsymbol{E}_{11} & -\kappa^2\boldsymbol{E}_{02} \end{bmatrix}}_{\boldsymbol{E}} \begin{bmatrix} \boldsymbol{\alpha} \\ \kappa^{-1}\boldsymbol{\beta}_1 \\ \kappa^{-1}\boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix} - \frac{1}{n}B_\Omega \boldsymbol{r} \tag{27}$$

where $(\boldsymbol{E}_{i_1i_2})_{\ell,j} = K^{(i_1i_2)}(\boldsymbol{f}_\ell - \boldsymbol{f}_j)$ and $\frac{1}{n}B_\Omega\boldsymbol{r}$ can be written in terms of the components of $R(\boldsymbol{f})$ and their partial derivatives

$$\frac{1}{n}B_\Omega\boldsymbol{r} = \big[R(\boldsymbol{f}_1)\cdots R(\boldsymbol{f}_r),\ R^{10}(\boldsymbol{f}_1)\cdots R^{10}(\boldsymbol{f}_r), R^{01}(\boldsymbol{f}_1)\cdots R^{01}(\boldsymbol{f}_r)\big]^T. \tag{28}$$

Define $B_\Omega := \big[\boldsymbol{b}(\boldsymbol{k}_{i_1}),\cdots,\boldsymbol{b}(\boldsymbol{k}_{i_s})\big]$, $\Omega = \{i_1,\cdots,i_s\}$, and $\boldsymbol{b}(\boldsymbol{k}) = \begin{bmatrix} 1 \\ -j2\pi\kappa k_1 \\ -j2\pi\kappa k_2 \end{bmatrix} \otimes \begin{bmatrix} e^{-j2\pi \boldsymbol{f}_1^T\boldsymbol{k}} \\ \cdot \\ e^{-j2\pi \boldsymbol{f}_r^T\boldsymbol{k}} \end{bmatrix}$, where $\boldsymbol{k}_{i_1},\cdots,\boldsymbol{k}_{i_s}$ are associated with $\boldsymbol{k}\in\Omega$. Interpolation vectors can be computed by solving the linear system in (27), so

$$Q(\boldsymbol{f}) = w^{00}(\boldsymbol{f})^T E^{-1}\left(\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix} - \frac{1}{\sqrt{n}}B_\Omega\boldsymbol{r}\right) + R(\boldsymbol{f}), \tag{29}$$

where $w^{i_1i_2}(\boldsymbol{f})$ for $i_1, i_2 \in \{0,1,2\}$ is defined as:

$$w^{i_1i_2}(\boldsymbol{f}) := \kappa^{i_1+i_2}\Big[K^{i_1i_2}(\boldsymbol{f}-\boldsymbol{f}_1),\cdots,K^{i_1i_2}(\boldsymbol{f}-\boldsymbol{f}_r), \kappa K^{i_1+1i_2}(\boldsymbol{f}-\boldsymbol{f}_1),\cdots,\kappa K^{i_1+1i_2}(\boldsymbol{f}-\boldsymbol{f}_r) \kappa K^{i_1i_2+1}(\boldsymbol{f}-\boldsymbol{f}_1),\cdots,\kappa K^{i_1i_2+1}(\boldsymbol{f}-\boldsymbol{f}_r)\Big]^T. \tag{30}$$

The following Lemma establishes an upper bound on ℓ_2 norm of $\boldsymbol{b}(\boldsymbol{k})$.

Lemma B.1: If $m \geq 2\times 10^3$, then $\|\boldsymbol{b}(\boldsymbol{k})\|_2^2 \leq 21r$, for $\boldsymbol{k}\in J$.

Proof.

$$\|\boldsymbol{b}(\boldsymbol{k})\|_2^2 \leq r\Big(1 + \max_{|k_1|\leq m}(2\pi k_1\kappa)^2 + \max_{|k_2|\leq m}(2\pi k_2\kappa)^2\Big) \leq 21r, \tag{31}$$

where the result of (23) is used. The following lemma suggests an upper bound on the operator norm of $\boldsymbol{B}_\Omega$ with certain probability.

Lemma B.2: (Proof in Appendix D). If the conditions of Theorem 3.1 hold, the event

$$\varepsilon_B := \left\{\|\boldsymbol{B}_\Omega\| > C_B(\log\frac{n^2}{\epsilon})^{-1/2}n\right\}, \tag{32}$$

happens with probability $\frac{\epsilon}{5}$ where C_B is a numerical constant. The following lemma states that $w^{i_1i_2}(\boldsymbol{f})$ is concentrated around the scaled version of

$$\bar{w}^{i_1i_2}(\boldsymbol{f}) := \kappa^{i_1+i_2}\Big[\bar{K}^{i_1i_2}(\boldsymbol{f}-\boldsymbol{f}_1),\cdots,\bar{K}^{i_1i_2}(\boldsymbol{f}-\boldsymbol{f}_r) \kappa\bar{K}^{i_1+1i_2}(\boldsymbol{f}-\boldsymbol{f}_1),\cdots,\kappa\bar{K}^{i_1+1i_2}(\boldsymbol{f}-\boldsymbol{f}_r) \kappa\bar{K}^{i_1i_2+1}(\boldsymbol{f}-\boldsymbol{f}_1),\cdots,\kappa\bar{K}^{i_1i_2+1}(\boldsymbol{f}-\boldsymbol{f}_r)\Big]^T, \tag{33}$$

on a fine grid with high probability.

Lemma B.3: (Proof in Appendix E). Let $\mathcal{G}\subset[0,1]^2$ be a 2D equispaced $800n^4$-point grid that discretizes $[0,1]^2$. If the conditions of Theorem 3.1 hold, then the event

$$\varepsilon_v := \left\{\left\|w^{i_1i_2}(\boldsymbol{f}) - \frac{n^2-s}{n^2}\bar{w}^{i_1i_2}(\boldsymbol{f})\right\|_2 > C_v(\log\frac{n^2}{\epsilon})^{-1/2}\right\}, \tag{34}$$

happens with probability $\epsilon/5$ for all $\boldsymbol{f} \in \mathcal{G}$, $i_1, i_2 \in \{0,1,2,3\}$ and numerical constant C_v.

APPENDIX C
PROOF OF PROPOSITION 4.1

First, it is necessary to determine the uniqueness of the solution of the linear system (27). The following lemma shows that $\boldsymbol{E}$ is concentrated around $\bar{\boldsymbol{E}}$ with high probability. Consequently, $\boldsymbol{E}$ is invertible and one can bound the operator norm of its inverse.

Lemma C.1: (Proof in Appendix F). If the conditions of Theorem 3.1 hold, then the event

$$\varepsilon_E := \Bigg\{ \Big\| \boldsymbol{E} - \frac{n^2-s}{n^2}\bar{\boldsymbol{E}} \Big\| \geq \frac{n^2-s}{4n^2} \min\left\{1, \frac{C_E}{4}\right\} \Big(\log \frac{n^2}{\epsilon}\Big)^{\frac{-1}{2}} \Bigg\}, \quad (35)$$

happens with probability $\epsilon/5$. Also, under the event ε_E^c, $\boldsymbol{E}$ is invertible and

$$\begin{aligned} &\big\|\boldsymbol{E}^{-1}\big\| \leq 8, \\ &\Big\|\boldsymbol{E}^{-1} - \frac{n^2}{n^2-s}\bar{\boldsymbol{E}}^{-1}\Big\| \leq C_E \big(\log \frac{n^2}{\epsilon}\big)^{\frac{-1}{2}}, \end{aligned} \quad (36)$$

where C_E is the numerical constant.

Indeed, this lemma claims that, under the event ε_E^c, the linear system (27) has an stable solution. So, $Q(\boldsymbol{f})$ is well defined and (8) holds. In order to meet (9), it is sufficient to show that $Q(\boldsymbol{f})$ is concentrated around $\bar{Q}(\boldsymbol{f})$ on a fine gird. Then, using Bernstein inequality, one can demonstrate that this property holds on the whole $[0,1]^2$. Finally, some bounds on $\bar{Q}(\boldsymbol{f})$ and its partial derivatives from [12] are borrowed to complete the proof.

Lemma C.2: (Proof in Appendix G). If the conditions of theorem 3.1 hold, then $|Q(\boldsymbol{f})| < 1, \quad \text{for} \quad \boldsymbol{f} \notin T$, with probability $1-\epsilon/5$ under the event $\varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c$.

The last task is to show that (11) is met. The following lemma asserts that under the event $\varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c$, one can control the magnitude of dual polynomial's coefficients with high probability.

Lemma C.3: (Proof in Appendix H). If the conditions of Theorem 3.1 hold, then $|C_{\boldsymbol{k}}| < \frac{1}{n}, \quad \text{for} \quad \boldsymbol{k} \in \Omega^c$, under the event $\varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c$.

Finally, the same technique in [26] is used to complete the proof. Consider ε_Q and ε_q as the events, such that (9) and (11) hold, respectively. By De Morgan's laws and union bound we get

$$\begin{aligned} \mathbb{P}((\varepsilon_Q \cap \varepsilon_q)^c) &= \mathbb{P}(\varepsilon_Q^c \cup \varepsilon_q^c) \\ &\leq \mathbb{P}(\varepsilon_Q^c \cup \varepsilon_q^c | \varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c) + \mathbb{P}(\varepsilon_B \cap \varepsilon_E \cap \varepsilon_v) \\ &\leq \mathbb{P}(\varepsilon_Q^c | \varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c) + \mathbb{P}(\varepsilon_q^c | \varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c) \\ &\quad + \mathbb{P}(\varepsilon_B) + \mathbb{P}(\varepsilon_E) + \mathbb{P}(\varepsilon_v) \\ &\leq \epsilon, \end{aligned} \quad (37)$$

which holds by the fact that for any pair of events ε_A and ε_B, $\mathbb{P}(\varepsilon_A) \leq \mathbb{P}(\varepsilon_A|\varepsilon_B^c) + \mathbb{P}(\varepsilon_B)$. On the other hand, via Lemmas C.3, C.3, C.1, B.3 and B.2, it is shown that the construction is valid with probability at least $1-\epsilon$.

APPENDIX D
PROOF OF LEMMA B.2

In order to obtain an upper bound on the operator norm of B_Ω, under the assumptions of theorem 3.1, one can show that

$$\boldsymbol{H} := B_\Omega B_\Omega^* = \sum_{\boldsymbol{k}\in\Omega} \boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k}), \quad (38)$$

is concentrated around $\frac{s}{n^2}\bar{\boldsymbol{H}} = \frac{s}{n^2}\sum_{\boldsymbol{k}\in J} \boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k})$. Using the following lemma, an upper bound on the operator norm of $\bar{\boldsymbol{H}}$ is computed.

Lemma D.1: (Proof in Section XIII in [35]). If the conditions of Theorem 3.1 hold, then $\|\bar{\boldsymbol{H}}\| \leq 223707\ n^2 \log^2 r$.

Regarding the fact that $s \leq C_s n^2 \big(\log^2 r \ \log \frac{n^2}{\epsilon}\big)^{-1}$ in Theorem 3.1, $\|\frac{s}{n^2}\bar{\boldsymbol{H}}\| \leq \frac{C_B^2}{2} n^2 \big(\log \frac{n^2}{\epsilon}\big)^{-1}$, if C_B is set small enough. One can demonstrate that $\boldsymbol{H}$ concentrates around the scaled version of $\bar{\boldsymbol{H}}$ using matrix Bernstein inequality.

Lemma D.2: (Proof in Section XIV in [35]). If the conditions of Theorem 3.1 hold, then $\|\boldsymbol{H} - \frac{s}{n^2}\hat{\boldsymbol{H}}\| \leq \frac{C_B^2}{2} n^2 \big(\log \frac{n^2}{\epsilon}\big)^{-1}$ with probability at least $1-\epsilon/5$.

Consequently, one can bound the operator norm of B_Ω by triangle inequality

$$\begin{aligned} \|B_\Omega\| \leq \sqrt{\|\boldsymbol{H}\|} &\leq \sqrt{\|\frac{s}{n^2}\bar{\boldsymbol{H}}\| + \|\boldsymbol{H} - \frac{s}{n^2}\bar{\boldsymbol{H}}\|} \\ &\leq C_B n \big(\log \frac{n^2}{\epsilon}\big)^{-1/2}, \end{aligned} \quad (39)$$

which happens with probability at least $1-\epsilon/5$.

APPENDIX E
PROOF OF LEMMA B.3

Lemma E.1: (Vector Bernstein inequality [36]). Let $\boldsymbol{u}(1), \cdots, \boldsymbol{u}(L)$ be independent zero-mean random vectors of dimension d. If $\|\boldsymbol{u}(k)\|_2 \leq B \quad \forall k$, we have

$$\mathbb{P}\Big\{ \sum_{k=1}^{L} \|\boldsymbol{u}(k)\|_2 \geq t \Big\} \leq \exp\big(-\frac{t^2}{8\sigma^2} + \frac{1}{4}\big), \quad (40)$$

for any $0 \leq t \leq \sigma^2$ where $\sum_{k=1}^{L} \mathbb{E}\big[\|\boldsymbol{u}(k)\|_2^2\big] \leq \sigma^2$.

Let us write $\bar{w}^{i_1 i_2}(\boldsymbol{f})$ and $w^{i_1 i_2}(\boldsymbol{f})$ in term of $\boldsymbol{b}$ using the definition of $\bar{K}(\boldsymbol{f})$ and $K(\boldsymbol{f})$

$$\begin{aligned} \bar{w}^{i_1 i_2}(\boldsymbol{f}) &= \sum_{\boldsymbol{k}\in J} (j2\pi\kappa)^{i_1+i_2} k_1^{i_1} k_2^{i_2} c_{k_1} c_{k_2} e^{j2\pi \boldsymbol{f}^T \boldsymbol{k}} \boldsymbol{b}(\boldsymbol{k}), \\ w^{i_1 i_2}(\boldsymbol{f}) &= \sum_{\boldsymbol{k}\in J} \delta_{\Omega^c}(\boldsymbol{k}) (j2\pi\kappa)^{i_1+i_2} k_1^{i_1} k_2^{i_2} c_{k_1} c_{k_2} e^{j2\pi \boldsymbol{f}^T \boldsymbol{k}} \boldsymbol{b}(\boldsymbol{k}), \end{aligned}$$

where δ_{Ω^c} is i.i.d. Bernoulli random variable with parameter $p := \frac{n^2-s}{n^2}$. We apply the result of vector Bernstein inequality in Lemma E.1 to finite sequences of zero-mean random vectors of the form

$$\boldsymbol{u}^{i_1 i_2}(\boldsymbol{k}) := (\delta_{\Omega^c}(\boldsymbol{k}) - p)(j2\pi\kappa)^{i_1+i_2} k_1^{i_1} k_2^{i_2} c_{k_1} c_{k_2} e^{j2\pi \boldsymbol{f}^T \boldsymbol{k}} \boldsymbol{b}(\boldsymbol{k}),$$

to demonstrate that the deviation between $w^{i_1 i_2}(\boldsymbol{f})$ and the scaled version of $\bar{w}^{i_1 i_2}(\boldsymbol{f})$ is small enough with high probability. For $i_1, i_2 \in \{0,1,2,3\}$, one can write

$$w^{i_1 i_2}(\boldsymbol{f}) - p\bar{w}^{i_1 i_2}(\boldsymbol{f}) = \sum_{\boldsymbol{k}\in J} \boldsymbol{u}(\boldsymbol{k}). \quad (41)$$

To calculate B in Lemma E.1, one can obtain an upper bound on ℓ_2 norm of $\boldsymbol{u}$

$$\|\boldsymbol{u}(\boldsymbol{k})\|_2 \leq \pi^{i_1+i_2}\|\boldsymbol{c}\|_\infty^2 \sup_{\boldsymbol{k}\in J}\|\boldsymbol{b}(\boldsymbol{k})\|_2 \leq B := \frac{7745}{m^2}\sqrt{r},$$

where the last inequality comes from (23), (22), and choosing $i_1 = i_2 = 3$. Also, To compute σ^2 in Lemma E.1,

$$\begin{aligned}&\sum_{\boldsymbol{k}\in J}\mathbb{E}\|\boldsymbol{u}^{i_1 i_2}(\boldsymbol{k})\|_2^2 = \sum_{\boldsymbol{k}\in J} c_{k_1}^2 c_{k_2}^2\|\boldsymbol{b}(\boldsymbol{k})\|_2^2\\ &\cdot (2\pi\kappa)^{2i_1+2i_2}k_1^{2i_1}k_2^{2i_2}\mathbb{E}\big[(\delta_{\Omega^c}(\boldsymbol{l})-p)^2\big]\\ &\leq 21r(2m+1)^2\pi^{2i_1+2i_2}\|\boldsymbol{c}\|_\infty^4 \leq \sigma^2 := \frac{240\times 10^6 r}{m^2}, \quad (42)\end{aligned}$$

where the first inequality is obtained from Lemma (B.1), (23), and the fact that variance of Bernoulli model is equal to $p(1-p) \leq 1$. The second inequality comes from (22) for $i_1 = i_2 = 3$. By leveraging the result of vector Bernstein inequality in Lemma E.1, we have

$$\begin{aligned}&\mathbb{P}\Big[\sup_{\boldsymbol{f}\in\mathcal{G}}\|w^{i_1 i_2}(\boldsymbol{f}) - \bar{w}^{i_1 i_2}(\boldsymbol{f})\|_2 \geq t, \quad i_1, i_2 \in \{0,1,2,3\}\Big]\\ &\leq 9|\mathcal{G}|\exp\left(\frac{-t^2}{8\sigma^2}+\frac{1}{4}\right), \qquad \text{for } 0\leq t\leq \frac{\sigma^2}{B}, \quad (43)\end{aligned}$$

by the union bound. The lower bound of probability is equal to $\epsilon/5$, if $t = \sigma\sqrt{8(\frac{1}{4}+\log\frac{45|\mathcal{G}|}{\epsilon})}$. In the following, it is shown that this choice of t satisfies $0 \leq t \leq \frac{\sigma^2}{B}$,

$$\begin{aligned}\frac{t}{\sigma} &= \sqrt{8\left(\frac{1}{4}+\log\frac{45|\mathcal{G}|}{\epsilon}\right)} \leq \sqrt{86+32\log n+8\log\frac{1}{\epsilon}}\\ &\leq 0.4\sqrt{n}+\sqrt{8\log\frac{1}{\epsilon}}, \quad (44)\end{aligned}$$

where the last inequality comes from the fact that $\sqrt{86+32\log n} \leq n$ for $n \geq 13$. Consequently, $t \leq \frac{\sigma^2}{B}$ if C_r and C_s are set small enough in Theorem 3.1. The desired result is obtained for

$$\sqrt{\frac{768\times 10^7 r}{n^2}\left(\frac{1}{4}+\log\frac{36\times 10^3 n^4}{\epsilon}\right)} \leq t \leq C_v\big(\log(\frac{n^2}{\epsilon})\big)^{\frac{-1}{2}}, \quad (45)$$

if C_r is set small enough in Theorem 3.1.

Appendix F
Proof of Lemma C.1

The proof involves the same techniques which were first proposed in [13]. In the following Lemma, it is demonstrated that matrix $\bar{\boldsymbol{E}}$ is close to the identity matrix, as a result, it is invertible.

Lemma F.1: (Proof in Section XVII in [35]). If the conditions of Theorem 3.1 hold, then $\|\boldsymbol{I}-\bar{\boldsymbol{E}}\| \leq 0.24$, $\|\bar{\boldsymbol{E}}\| \leq 1.24$, $\|\bar{\boldsymbol{E}}^{-1}\| \leq 1.32$.

It is possible to write $\bar{\boldsymbol{E}}$ and $\boldsymbol{E}$ in the terms of $\boldsymbol{b}\boldsymbol{b}^*$:

$$\bar{\boldsymbol{E}} = \sum_{\boldsymbol{k}\in J} c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k}), \quad (46)$$

$$\boldsymbol{E} = \sum_{\boldsymbol{k}\in J}\delta_{\Omega^c}(\boldsymbol{k})c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k}). \quad (47)$$

One can show that $\boldsymbol{E}$ is concentrated around $p\bar{\boldsymbol{E}}$, which $p := \frac{n^2-s}{n^2}$ with high probability. the self-adjoint zero mean matrix $\boldsymbol{X}$ is defined as below:

$$\boldsymbol{X}(\boldsymbol{k}) := (p-\delta_{\Omega^c}(\boldsymbol{k}))c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k}), \quad (48)$$

then$\mathbb{E}(\boldsymbol{X}(\boldsymbol{k})) = (p-\mathbb{E}(\delta_{\Omega^c}(\boldsymbol{k})))c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k}) = 0$. It is possible to bound the operator norm of $\boldsymbol{X}$ using Lemma B.1 as follows:

$$\begin{aligned}\|\boldsymbol{X}(\boldsymbol{k})\| &\leq \max_{\boldsymbol{k}\in J}\|c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k})\|\\ &\leq \|\boldsymbol{c}\|_\infty^2\max_{\boldsymbol{k}\in J}\|\boldsymbol{b}(\boldsymbol{k})\|_2^2 \leq B := \frac{36r}{m^2}. \quad (49)\end{aligned}$$

Also,

$$\begin{aligned}&\sum_{\boldsymbol{k}\in J}\mathbb{E}\big(\boldsymbol{X}(\boldsymbol{k})\boldsymbol{X}^T(\boldsymbol{k})\big) =\\ &\qquad\Big\|\sum_{\boldsymbol{k}\in J}c_{k_1}^2c_{k_2}^2\|\boldsymbol{b}(\boldsymbol{k})\|_2^2\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k})\mathbb{E}\big[(\delta_{\Omega^c}-p)^2\big]\Big\|\\ &\qquad\leq 21rp(1-p)\|\boldsymbol{c}\|_\infty^2\sum_{\boldsymbol{k}\in J}c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})\boldsymbol{b}^*(\boldsymbol{k})\\ &\qquad\leq \frac{36rp}{m^2}\|\bar{\boldsymbol{E}}\| \leq \sigma^2 := \frac{45rp}{m^2}, \quad (50)\end{aligned}$$

where the first inequality uses the variance of the Bernoulli model with parameter p and Lemma B.1, the second stems from (22) and the definition of $\bar{\boldsymbol{E}}$ and the last one is the result of Lemma F.1. For an ease notation, $t = \frac{p}{4}C_{\min}\big(\log\frac{n^2}{\epsilon}\big)^{-\frac{1}{2}}$ where $C_{\min} := \min\{1, \frac{C_E}{4}\}$. By matrix Bernstein inequality [36], one can write

$$\begin{aligned}&\mathbb{E}\left\{\left\|\boldsymbol{E}^{-1}-p\bar{\boldsymbol{E}}^{-1}\right\| \geq t\right\} \leq\\ &6r\exp\left(\frac{-pC_{\min}^2m^2}{32r}\left(45\log\frac{n^2}{\epsilon}+3C_{\min}\sqrt{\log\frac{n^2}{\epsilon}}\right)^{-1}\right)\\ &\leq 6r\exp\left(\frac{-C'_E(n^2-s)}{r\log\frac{n^2}{\epsilon}}\right), \quad (51)\end{aligned}$$

for a numerical constant C'_E. The lower bound on this probability is $\epsilon/5$ when

$$r \leq \frac{C'_D n^2}{2}\left(\log\frac{30r}{\epsilon}\log\frac{n^2}{\epsilon}\right)^{-1}, \quad s \leq \frac{n^2}{2}, \quad (52)$$

which hold under the assumption of Theorem 3.1, if C_r and C_s are set small enough.

Consequently, a lower bound on the smallest singular value of $\boldsymbol{E}$ can be obtained by triangle inequality,

$$\frac{\sigma_{\min}(\boldsymbol{E})}{p} \geq \sigma_{\min}(\boldsymbol{I}) - \|\boldsymbol{I}-\bar{\boldsymbol{E}}\| - \frac{1}{p}\|\boldsymbol{E}-p\bar{\boldsymbol{E}}\| \geq 0.51, \quad (53)$$

with high probability. Therefore, $\boldsymbol{E}$ is invertible. [13, Appendix E] states that for any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ so that $\boldsymbol{B}$ is invertible and $\|\boldsymbol{A}-\boldsymbol{B}\|\|\boldsymbol{B}^{-1}\| \leq \frac{1}{2}$ one can write

$$\begin{aligned}&\|\boldsymbol{A}^1\| \leq 2\|\boldsymbol{B}^{-1}\|,\\ &\|\boldsymbol{A}^{-1}-\boldsymbol{B}^{-1}\| \leq 2\|\boldsymbol{B}^{-1}\|^2\|\boldsymbol{A}-\boldsymbol{B}\|. \quad (54)\end{aligned}$$

Consider $\boldsymbol{A} := \boldsymbol{E}$ and $\boldsymbol{B} := p\bar{\boldsymbol{E}}$. Using Lemma (F.1) and conditioned on (51), we have $\|\boldsymbol{E}-p\bar{\boldsymbol{E}}\|\big\|(p\boldsymbol{E})^{-1}\big\| \leq \frac{1}{2}$, with probability at least $1-\epsilon/5$. Based on this and Lemma F.1, and event (51) we also have

$$\begin{aligned}
&\|\boldsymbol{E}^{-1}\| \leq 2\big\|(p\bar{\boldsymbol{E}})^{-1}\big\| \leq \frac{4}{p},\\
&\big\|\boldsymbol{E}^{-1}-(p\bar{\boldsymbol{E}})^{-1}\big\| \leq 2\big\|(p\bar{\boldsymbol{E}}^{-1})\big\|^2\|\boldsymbol{E}-p\bar{\boldsymbol{E}}\|\\
&\leq \frac{C_E}{2p}\big(\log\frac{n^2}{\epsilon}\big)^{\frac{-1}{2}},
\end{aligned} \tag{55}$$

with the same probability. Regarding the conditions of Theorem 3.1 $s \leq \frac{n^2}{2}$, therefore $\frac{1}{p} \leq 2$. This concludes the proof.

Appendix G
Proof of Lemma C.2

It is possible to express $Q^{i_1i_2}(\boldsymbol{f})$ and $\bar{Q}^{i_1i_2}(\boldsymbol{f})$ in the terms of $\boldsymbol{h}$ and $\boldsymbol{r}$:

$$\kappa^{i_1+i_2}\bar{Q}^{i_1i_2}(\boldsymbol{f}) = \bar{w}^{i_1i_2}(\boldsymbol{f})^T\bar{\boldsymbol{E}}^{-1}\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \tag{56}$$

$$\begin{aligned}
\kappa^{i_1+i_2}Q(\boldsymbol{f}) &= w^{i_1i_2}(\boldsymbol{f})^T\boldsymbol{E}^{-1}\bigg(\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix} - \frac{1}{n}B_\Omega\boldsymbol{r}\bigg)\\
&\quad + \kappa^{i_1+i_2}R^{i_1i_2}(\boldsymbol{f}).
\end{aligned} \tag{57}$$

$Q^{i_1i_2}(\boldsymbol{f})$ and $\bar{Q}^{i_1i_2}(\boldsymbol{f})$ are related to each other as

$$\begin{aligned}
&\kappa^{i_1+i_2}Q^{i_1i_2}(\boldsymbol{f}) = \kappa^{i_1+i_2}\bar{Q}^{i_1i_2}(\boldsymbol{f}) + \kappa^{i_1+i_2}R^{i_1i_2}(\boldsymbol{f})\\
&+ I_1^{i_1i_2}(\boldsymbol{f}) + I_2^{i_1i_2}(\boldsymbol{f}) + I_3^{i_1i_2}(\boldsymbol{f}),
\end{aligned} \tag{58}$$

in which

$$I_1^{i_1i_2}(\boldsymbol{f}) := \frac{-1}{n}w^{i_1i_2}(\boldsymbol{f})^T\boldsymbol{E}^{-1}B_\Omega\boldsymbol{r}, \tag{59}$$

$$I_2^{i_1i_2}(\boldsymbol{f}) := \bigg(w^{i_1i_2}(\boldsymbol{f}) - \frac{n^2-s}{n^2}\bar{w}^{i_1i_2}(\boldsymbol{f})\bigg)^T\boldsymbol{E}^{-1}\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \tag{60}$$

$$I_3^{i_1i_2}(\boldsymbol{f}) := \frac{n^2-s}{n^2}\bar{w}^{i_1i_2}(\boldsymbol{f})^T\bigg(\boldsymbol{E}^{-1} - \frac{n^2}{n^2-s}\bar{\boldsymbol{E}}\bigg)\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}. \tag{61}$$

In the following Lemma, it is shown that there exist upper bounds on these terms in the 2D grid $\mathcal{G}$ with high probability.

Lemma G.1: (Proof in Section XIX in [35]). If the conditions of Theorem 3.1 hold, under the condition $\varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c$, the events

$$\varepsilon_R := \bigg\{\sup_{\boldsymbol{f}\in\mathcal{G}}\big|\kappa^{i_1+i_2}R^{i_1i_2}(\boldsymbol{f})\big| \geq \frac{10^{-2}}{8}, \quad i_1,i_2\in\{0,1,2,3\}\bigg\}, \tag{62}$$

and

$$\varepsilon_i := \bigg\{\sup_{\boldsymbol{f}\in\mathcal{G}}\big|I_i^{i_1i_2}(\boldsymbol{f})\big| \geq \frac{10^{-2}}{8}, \quad i_1,i_2\in\{0,1,2,3\}\bigg\}, \tag{63}$$

for $i\in\{0,1,2,3\}$ and the 2D equispaced grid $\mathcal{G}$ with set size $800n^4$, happen with probability at most $\epsilon/5$. Consequently, by triangle inequality

$$\sup_{\boldsymbol{f}\in\mathcal{G}}\big|\kappa^{i_1+i_2}Q^{i_1i_2}(\boldsymbol{f}) - \kappa^{i_1+i_2}\bar{Q}^{i_1i_2}(\boldsymbol{f})\big| \leq \frac{10^{-2}}{2}, \tag{64}$$

with probability at least $1-\epsilon/5$ under the condition $\varepsilon_B^c \cap \varepsilon_E^c \cap \varepsilon_v^c$.

It has already shown that the deviation between $Q^{i_1i_2}(\boldsymbol{f})$ and $\bar{Q}^{i_1i_2}(\boldsymbol{f})$ is small on a fine grid. In the following, this approach is extended to the whole $[0,1]^2$.

Lemma G.2: (Proof in Section XXI in [35]). If the conditions of Theorem 3.1 hold, then

$$\kappa^{i_1+i_2}|Q^{i_1i_2}(\boldsymbol{f}) - \bar{Q}^{i_1i_2}(\boldsymbol{f})| \leq 10^{-2}, \quad i_1,i_2\in\{0,1,2,3\}. \tag{65}$$

$[0,1]^2$ is divided into two domains

$$\begin{aligned}
\mathcal{S}_{\text{near}} &= \big\{\boldsymbol{f}|\ \|\boldsymbol{f}-\boldsymbol{f}_i\|_\infty \leq 0.09\big\},\\
\mathcal{S}_{\text{far}} &= [0,1]^2 \setminus \mathcal{S}_{\text{near}}.
\end{aligned} \tag{66}$$

[12] proved that $|\bar{Q}(\boldsymbol{f})| \leq 0.9866$ for $\boldsymbol{f}\in\mathcal{S}_{\text{far}}$. One can leverage the result of Lemma G.2 and triangle inequality to obtain $|Q(\boldsymbol{f})| \leq |\bar{Q}(\boldsymbol{f})| + 10^{-2} \leq 1$, for $\boldsymbol{f}\in\mathcal{S}_{\text{far}}$.

Also, [12] demonstrated that the following Hessian matrix is negative definite in domain $\boldsymbol{f}\in\mathcal{S}_{\text{near}}$, so $|\bar{Q}(\boldsymbol{f})| \leq 1$ in this domain, $\bar{\boldsymbol{H}} = \begin{bmatrix}\bar{Q}^{20}(\boldsymbol{t}) & \bar{Q}^{11}(\boldsymbol{t})\\ \bar{Q}^{11}(\boldsymbol{t}) & \bar{Q}^{02}(\boldsymbol{t})\end{bmatrix}$. More precisely, $\bar{Q}^{20} \leq -1.4809m^2$, $\bar{Q}^{02} \leq -1.4809m^2$ and $|\bar{Q}^{11}| \leq 1.4743m^2$. It is possible to rewrite the elements of the matrix $\bar{\boldsymbol{H}}$ for $Q(\boldsymbol{f})$ and bring matrix $\boldsymbol{H}$, then using the result of Lemma G.2. One can write

$$Q^{20}(\boldsymbol{f}) \leq -1.5209m^2, \quad |Q^{11}(\boldsymbol{f})| \leq 1.5143m^2, \tag{67}$$

by (23). If the matrix $\boldsymbol{H}$ is concave, then $Q(\boldsymbol{f}) < 1$. A sufficient condition for concavity of this matrix is $\mathrm{Tr}(\boldsymbol{H}) < 0$ and $\det(\boldsymbol{H}) > 0$. We can write

$$\begin{aligned}
\mathrm{Tr}(\boldsymbol{H}) &= Q^{20}(\boldsymbol{f}) + Q^{02}(\boldsymbol{f}),\\
\det(\boldsymbol{H}) &= |Q^{20}(\boldsymbol{f})||Q^{02}(\boldsymbol{f})| - |Q^{11}(\boldsymbol{f})|^2.
\end{aligned} \tag{68}$$

By (67), it is easy to see that $\mathrm{Tr}(\boldsymbol{H}) < 0$ and $\det(\boldsymbol{H}) > 0$, so the Hessian matrix $\boldsymbol{H}$ is negative definite in the domain $\mathcal{S}_{near}$. This concludes the proof.

Appendix H
Proof of Lemma C.3

One can recast the coefficient $C_{\boldsymbol{k}}$ in terms of $\boldsymbol{h}$ and $\boldsymbol{r}$. Let $\boldsymbol{k}$ be an arbitrary element of Ω^c

$$\begin{aligned}
C_{\boldsymbol{k}} &= c_{k_1}c_{k_2}\bigg(\sum_{i=1}^r \alpha_i e^{j2\pi\boldsymbol{f}_i^T\boldsymbol{k}}\\
&\quad + i2\pi\kappa k_1\sum_{i=1}^r\beta_{1i}e^{i2\pi\boldsymbol{f}_i^T\boldsymbol{k}} + i2\pi\kappa k_2\sum_{i=1}^r\beta_{2i}e^{i2\pi\boldsymbol{f}_i^T\boldsymbol{k}}\bigg)\\
&= c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})^*\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} = c_{k_1}c_{k_2}\boldsymbol{b}(\boldsymbol{k})^*E^{-1}\bigg(\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix} - \frac{1}{n}B_\Omega\boldsymbol{r}\bigg)\\
&= c_{k_1}c_{k_2}\bigg(\langle\boldsymbol{p}\boldsymbol{E}^{-1}\boldsymbol{b}(\boldsymbol{k}),\boldsymbol{h}\rangle + \frac{1}{n}\langle B_\Omega^*\boldsymbol{E}^{-1}\boldsymbol{b}(\boldsymbol{k}),\boldsymbol{r}\rangle\bigg),
\end{aligned}$$

where $\boldsymbol{P} \in \mathbb{R}^{r\times 3r}$ is the projection matrix that retains the first r entries of a vector. To bound $|C_{\boldsymbol{k}}|$, a bound on $\boldsymbol{P}\boldsymbol{E}^{-1}\boldsymbol{b}$ is obtained as below

$$\|\boldsymbol{P}\boldsymbol{E}^{-1}\boldsymbol{b}(\boldsymbol{k})\|_2^2 \leq \|\boldsymbol{P}\|^2\|\boldsymbol{E}^{-1}\|^2\|\boldsymbol{b}(\boldsymbol{k})\|_2^2 \leq 1344r \leq \frac{0.07^2n^2}{\log\frac{40}{\epsilon}},$$

where the last inequality is achieved under the conditions of Theorem 3.1 when C_k is set small enough and the second one is a combination of Lemmas B.1 , C.1 and the fact that $\|\boldsymbol{P}\|_2 = 1$. Also, we have

$$\begin{aligned}\|\boldsymbol{B}_\Omega^*\boldsymbol{E}^{-1}\boldsymbol{b}(\boldsymbol{k})\|_2^2 &\leq \|\boldsymbol{B}_\Omega\|^2\|\boldsymbol{E}^{-1}\|^2\|\boldsymbol{b}(\boldsymbol{k})\|_2^2\\ &\leq 1344rn^2C_B^2 \leq \frac{0.07^2n^2}{\log\frac{40}{\epsilon}}, \qquad (69)\end{aligned}$$

where the second inequality stems from Lemmas B.1 and B.2 and the last one comes from the assumption of Theorem 3.1 if C_r is set small enough.

One can obtain $\epsilon/10$ for the minimum probability of the following events by Hoeffding's inequality

$$|\langle\boldsymbol{P}\boldsymbol{E}^{-1}\boldsymbol{b}(\boldsymbol{k}),\boldsymbol{h}\rangle| > 0.07n, \quad |\langle\boldsymbol{B}_\Omega^*\boldsymbol{E}^{-1}\boldsymbol{b}(\boldsymbol{k}),\boldsymbol{r}\rangle| > 0.07n^2.$$

Using $\|\boldsymbol{c}\|_\infty \leq \frac{1.3}{m}$ and the union bound, we have $|C_{\boldsymbol{k}}| \leq \frac{2.6^2}{n^2}\big(0.07n + 0.07n\big) \leq \frac{1}{n}$, with probability at least $1-\epsilon/5$. This concludes the proof.

REFERENCES

[1] L. Zheng and X. Wang, "Super-resolution delay-doppler estimation for ofdm passive radar," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2197–2210, 2017.

[2] A. Popowicz, A. Kurek, T. Blachowicz, V. Orlov, and B. Smolka, "On the efficiency of techniques for the reduction of impulsive noise in astronomical images," *Monthly Notices of the Royal Astronomical Society*, vol. 463, no. 2, pp. 2172–2189, 2016.

[3] R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi, "Super-resolution radar," *Information and Inference: A Journal of the IMA*, vol. 5, no. 1, pp. 22–75, 2016.

[4] X. Lu, J. Wang, A. Ponsford, and R. Kirlin, "Impulsive noise excision and performance analysis," in *2010 IEEE Radar Conference*, pp. 1295–1300, IEEE, 2010.

[5] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2267–2280, 1992.

[6] Y. Hua, "A pencil-music algorithm for finding two-dimensional angles and polarizations using crossed dipoles," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 3, pp. 370–376, 1993.

[7] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[9] N. Verzelen *et al.*, "Minimax risks for sparse regressions: Ultra-high dimensional phenomenons," *Electronic Journal of Statistics*, vol. 6, pp. 38–90, 2012.

[10] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[11] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Information and Inference: A Journal of the IMA*, vol. 5, no. 3, pp. 251–303, 2016.

[12] I. Valiulahi, S. Daei, F. Haddadi, and F. Parvaresh, "Two-dimensional super-resolution via convex relaxation," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3372–3382, 2019.

[13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE transactions on information theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[14] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, 2015.

[15] V. Duval and G. Peyré, "Sparse spikes super-resolution on thin grids ii: the continuous basis pursuit," *Inverse Problems*, vol. 33, no. 9, p. 095008, 2017.

[16] Y. De Castro, F. Gamboa, D. Henrion, and J.-B. Lasserre, "Exact solutions to super resolution on semi-algebraic domains in higher dimensions," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 621–630, 2017.

[17] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*. Springer, 2017.

[18] W. Xu, J.-F. Cai, K. V. Mishra, M. Cho, and A. Kruger, "Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional (d 2) off-the-grid frequencies," in *2014 Information Theory and Applications Workshop (ITA)*, pp. 1–4, IEEE, 2014.

[19] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Spectral super-resolution with prior knowledge," *IEEE transactions on signal processing*, vol. 63, no. 20, pp. 5342–5357, 2015.

[20] I. Valiulahi, H. Fathi, S. Daei, and F. Haddadi, "Off-the-grid two-dimensional line spectral estimation with prior information," *arXiv preprint arXiv:1704.06922*, 2017.

[21] Z. Yang and L. Xie, "Fast convex optimization method for frequency estimation with prior knowledge in all dimensions," *Signal Processing*, vol. 142, pp. 271–280, 2018.

[22] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[23] C. Poon and G. Peyré, "Multidimensional sparse super-resolution," *SIAM Journal on Mathematical Analysis*, vol. 51, no. 1, pp. 1–44, 2019.

[24] Y. De Castro and F. Gamboa, "Exact reconstruction using beurling minimal extrapolation," *Journal of Mathematical Analysis and applications*, vol. 395, no. 1, pp. 336–354, 2012.

[25] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[26] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *Information and Inference: A Journal of the IMA*, 2016.

[27] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.

[28] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35–47, 2010.

[29] Z. Yang, J. Tang, Y. C. Eldar, and L. Xie, "On the sample complexity of multichannel frequency estimation via convex optimization," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2302–2315, 2018.

[30] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.

[31] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[32] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, 2016.

[33] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.

[34] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *IEEE Transactions on Information Theory*, 2017.

[35] I. Valiulahi, F. Haddadi, and A. Amini, "Robustness of two-dimensional line spectral estimation against spiky noise," *arXiv preprint arXiv:1807.01185*, 2018.

[36] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, vol. 81, no. 1, pp. 73–205, 1995.