Demixing Sines and Spikes Using Multiple Measurement Vectors

Hoomaan Maskan[1,1], Sajad Daei[1,2], Mohammad Hossein Kahaei[1,3,*]

Abstract

We address the line spectral estimation problem with multiple measurement corrupted vectors. Such scenarios appear in many practical applications such as radar, optics, and seismic imaging in which the measurements can be modeled as the sum of a spectrally sparse and a block-sparse signal known as outlier. Our aim is to demix the two components and for this purpose, we design a convex problem whose objective function promotes both of the structures. Using the Positive Trigonometric Polynomials (PTP) theory, we reformulate the dual problem as a Semidefinite Program (SDP). Our theoretical results state that for a fixed number of measurements N and constant number of outliers, up to $\mathcal{O}(N)$ spectral lines can be recovered using our SDP problem as long as a minimum frequency separation condition is satisfied. Our simulation results also show that increasing the number of samples per measurement vectors reduces the minimum required frequency separation for successful recovery.

Keywords: Spectral super resolution, demixing, multiple measurement vector, atomic norm, convex optimization.

1. Introduction

Spectral super resolution is the problem of estimating the spectrum of a signal composed of sinusoids using finite number of samples. This problem, also known as line spectral estimation, is of great importance in signal processing applications such as radar [1, 2, 3, 4], multi-path channel estimation [5], seismic imaging [6], and magnetic resonance imaging [7].

There exist three main attitudes toward spectral super resolution problems: non-parametric methods, parametric approaches [8], and optimization-based methods [9, 10, 11]. Periodogram as a non-parametric method can localize sinusoids up to a limited resolution [12] in the noiseless case. Multiple Signal

---

[*]Corresponding author

Email addresses: `hoomaan.maskan@umu.se` (Hoomaan Maskan), `Sajad.Daei@eurecom.fr` (Sajad Daei), `kahaei@iust.ac.ir` (Mohammad Hossein Kahaei)

[1]Department of Mathemetics & Mathematical Statistics, Umeå University

[2]EURECOM- Communication Systems

[3]School of Electrical Engineering, Iran University of Science & Technology

Classification (MUSIC) is a parametric method which can recover sinusoids perfectly [13]. However, the performance of this method degrades in the presence of noise or outliers. Also, MUSIC needs the correlation matrix of the signal and lack of measurements can highly affect the performance of MUSIC. Other examples of parametric approaches are Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) [14] and Matrix Pencil method [15]. Optimization based approaches minimize the continuous counterpart of the ℓ_1 norm known as the Total Variation (TV) norm [9]. These methods are shown to be robust against Gaussian noise[9]. However, their performance degrades when outliers are present. Tang et.al. proposed a mathematical formulation for the spectral super resolution problem using Atomic Norm Minimization (ANM) [16]. For more illustration, consider a time dispersive multipath channel. The problem is to estimate channel delays and the corresponding complex coefficients using a limited number of pilots. This problem is studied using spectral super resolution and ANM [5, 17].

In most applications, an array of sensors is utilized to receive the signal. In real scenarios, the output of some sensors might be corrupted by perturbations and this makes it harder to super resolve the spectrum of the signal. Thus, the received signal can be described as a mixture of the transmitted signal and spiky noise. This noise can be due to the interference arising from other signals, lightning discharges, and sensor failures. The problem of estimating the transmitted signal from the latter mixture is known as demixing sines and spikes. The demixing problem using the single measurement vector (SMV) is studied in [18] and [19]. In some certain settings in applications, we are allowed to collect Multiple Measurement Vectors (MMVs). For example, in Direction Of Arrival (DOA) estimation in array processing [20], the aim is to estimate the DOAs of narrowband sources by observing the output of a sensor array (a group of sensors) during a time window. As each sensor collects a measurement vector (takes snapshot) at each time instance, we have access to MMVs in a time interval. As mentioned earlier, sensors might be exposed to perturbations which can lead to corrupted measurements (interpreted as outliers). To jointly estimate the sources and the ourliers, one could use multiple disjoint SMV demixing problems (corresponding to multiple snapshots) or a single large SMV problem by increasing the array size. However, these approaches do not seem to be reasonable due to cost limitations and array physical constraints. Therefore, it is necessary to exploit the temporal redundancy contained in the MMVs by assuming that the sources remain fixed in a time interval.

In this work, the benefits of using MMVs in the demixing problem are investigated. It is shown that using MMVs makes it possible to localize the sines with high precision. According to the fact that the measurement vectors share the same spectral characteristic of the signal of interest, it is possible to use this joint spectral sparsity and distinguish the signal of interest from the outliers. According to the applied signal model, a new method for spectral super resolution in the presence of outliers is proposed. Also, due to the infinite dimensionality of the TV norm minimization problem, the dual problem is investigated. Using positive trigonometric polynomials (PTP) theory [21], a tractable SDP is pro-

2

posed. A vector dual polynomial is formed using the dual variables of the latter SDP. Also, a sufficient condition for the exact recovery of the proposed method is provided.

The rest of the paper is as follows: In Section 2 the demixing problem for the MMV case is formulated, in Section 3 the TV norm minimization is applied to distinguish the signal of interest from the outliers, in Section 4 the dual problem is investigated and a new SDP is proposed, in Section 5 dense Gaussian perturbation is added to the model and the corresponding SDP is proposed. Section 6 presents the numerical results, Section 7 provides the proof for the main theorem, and Section 8 is devoted to the conclusion and future work discussions. Also, the proof of the main theorem can be found in Section 9.

Notation. Throughout this paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The ith element of the vector $\boldsymbol{x}$ is given by $\boldsymbol{x}_i$. $|.|$ denotes cardinality for sets and absolute value for scalars. $f^{(i)}(t)$ denotes the ith derivative of $f(t)$ with respect to t. Transpose, conjugate, and hermitian of a matrix or vector are given by $(.)^T$, $(.)^*$, and $(.)^H$ respectively.

2. Problem Formulation

Suppose that the signal of interest is composed of K complex exponentials

$$s_{jl} = \sum_{k=1}^{K} a_{kl} e^{i2\pi j f_k} \quad , j \in \mathcal{N}, l \in \mathcal{L}, \tag{1}$$

where $\mathcal{N} = \{0, \ldots, N-1\}$, $\mathcal{L} = \{1, \ldots, L\}$, $a_{kl} \in \mathbb{C}$ is the complex amplitude corresponding to the kth frequency, $i = \sqrt{-1}$, N is the length of the sinusoids, L is the number of measurements or snapshots taken over time, and $f_k \in \mathbb{T}$ where $\mathbb{T} := \{f_1, \ldots, f_K\} \subset [0,1]$ is the support set of the signal. In the Fourier domain, (1) can be expressed as

$$G_l(f) = \sum_{k=1}^{K} a_{kl} \delta(f - f_k), \tag{2}$$

where $\delta(f - f_k)$ is Dirac delta function located at f_k. The signal can be expressed in a matrix form $\boldsymbol{S}$ whose columns denote the measurements for one snapshot and the rows correspond to the output of each sensor for different snapshots. Note that we can write

$$s_{jl} = \sum_{k=1}^{K} a_{kl} e^{i2\pi j f_k} = \int_0^1 e^{i2\pi j f} G_l(df) = (\mathcal{F}_N G_l)_j,$$

where $\mathcal{F}_N$ maps the measure G_l to its first N Fourier series coefficients. Here, we study the full measurement case. The results can be extended to the random sampling case [22].

3

As stated in Section 1, outliers degrade the performance of recent optimization-based spectral super resolution methods. In order to overcome this problem, the effect of the outliers should be considered in the initial model used for the received signal. Following the same approach of [18], the outliers are added to the received signal as a matrix $\boldsymbol{Z}$

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{Z} = [(\mathcal{F}_N G_1), \ldots, (\mathcal{F}_N G_L)] + \boldsymbol{Z}, \qquad (3)$$

where $\boldsymbol{Y}_{jl}$ and $\boldsymbol{Z}_{jl}$ are the received signal and the outliers at jth sensor and lth snapshot respectively. Note that the outliers affect few number of sensors such that the outlier matrix $\boldsymbol{Z}_{N \times L}$ is considered to be row-sparse and $\boldsymbol{\Omega} \subset \{0, \ldots, N-1\}$ denotes the overall support set of the outliers which means the rows of $\boldsymbol{Z}_{N \times L}$ with nonzero ℓ_2 norms.

3. Total Variation Norm Minimization

Without any prior assumption, the demixing problem is ill-posed. Sparse assumption on the signal structure is proved to be helpful in solving linear inverse problems. In compressed sensing theory, *Restricted-Isometry Property (RIP)* guaranteed that a random sampling operator would preserve most of the signal's energy with high probability (see e.g. [23, 24, 25, 26] for more details). However, in spectral super resolution, it is possible that the non-zero spectral information of the signal lies in the null space of the sampling operator. Thus, an additional condition called the *minimum separation condition* should be met[27].

Definition 1. (Minimum separation) Consider the set $\mathbb{T}$ as the set of support. The minimum separation is defined as the minimum wrap-around distance between any elements of $\mathbb{T}$,

$$\Delta := \Delta(\mathbb{T}) = \inf_{(f_1, f_2) \in \mathbb{T}: f_1 \neq f_2} |f_2 - f_1|.$$

For clarification, the wrap-around distance between $f_1 = \frac{1}{5}$ and $f_2 = \frac{4}{5}$ is equal to $\frac{2}{5}$.

In compressed sensing theory, the $\ell_{1,2}$ norm was used to promote group sparsity of the received signals sharing the same support sets (see *e.g.* [28, 29]). The continuous counterpart of $\ell_{1,2}$ norm is the group Total Variation (gTV) norm

$$\|\boldsymbol{X}\|_{\mathrm{gTV}} := \sup_{\substack{\|\boldsymbol{F}(t)\|_2 \leq 1, t \in \mathbb{T} \\ \boldsymbol{F}: \mathbb{T} \to \mathbb{C}^L}} \sum_{l=1}^{L} Re\{ \int_{\mathbb{T}} \boldsymbol{F}_l^H(t) \boldsymbol{X}_l(dt)\}.$$

Fernandez proved that a minimum separation of $\frac{2.52}{N-1}$ has to be met so that the gTV norm minimization achieves exact recovery [27]. Following the same

insight of [18], we propose the following optimization problem for demixing in the MMV case

$$\min_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{Z}}} \|\tilde{\boldsymbol{G}}\|_{gTV} + \lambda\|\tilde{\boldsymbol{Z}}\|_{1,2} \quad s.t. \ \boldsymbol{Y} = [\mathcal{F}_N\tilde{G_1},\dots,\mathcal{F}_N\tilde{G_L}] + \tilde{\boldsymbol{Z}}, \qquad (4)$$

where $\lambda > 0$ is a regularization parameter, $\|.\|_{1,2}$ denotes the matrix $l_{1/2}$ norm and $\mathcal{F}_N$ is the linear operator mapping a vector to its N lowest Fourier Coefficients. The main contribution of this paper is to show that under certain assumptions, the above problem has a unique solution.

Theorem 1. *Consider N measurements with L snapshots and suppose that the ℓ_2 norm of each row in the outlier matrix $\boldsymbol{Z}$ is non-zero with probability $\frac{s}{N}$. Let the elements of the support set $\mathbb{T}$ satisfy the minimum separation condition of $\Delta \geq \Delta_{min} = \frac{2.52}{N-1}$. If the phases $\boldsymbol{a}_l$ and the nonzero entries of $\boldsymbol{Z}$ are i.i.d uniformly distributed in $[0, 2\pi]$, then (4) with $\lambda = 1/\sqrt{N}$ provides the exact solution with probability at least $1 - \epsilon$ for any $\epsilon > 0$ as long as*

$$K < C_K N \left(\log \frac{N}{\epsilon} \right)^{-1} \left(1 + \frac{1}{L} \log \frac{\sqrt{L}N^3}{\epsilon} \right)^{-1},$$

$$s < C_s N \left(\log \frac{N}{\epsilon} \right)^{-1} \left(1 + \frac{1}{L} \log \frac{\sqrt{L}N^3}{\epsilon} \right)^{-1},$$

for some constants C_K, C_s, and $N \geq 2 \times 10^3$.

Remark. Using MMVs leads to an increased probability of successful recovery. To see this, consider demixing the corresponding columns of $\boldsymbol{S}$ and $\boldsymbol{Z}$ in the SMV case [18]. For a fixed N, each column of $\boldsymbol{S}$ and $\boldsymbol{Z}$ can be recovered from the corresponding column of $\boldsymbol{Y}$ with a probability of at least $1 - \epsilon$. Thus, in order to recover all the columns of $\boldsymbol{S}$, the probability of successful recovery would be at least $1 - L\epsilon$. However, in order to solve the problem with a single optimization, as proposed in Theorem 1, the probability of successful recovery for weaker conditions on N and K, is at least $1 - \sqrt{L}\epsilon$. This explicitly certifies that the proposed method outperforms L individual SMVs in terms of the success probability. It is also worth mentioning that the simulation results in Section 6 indicate that the MMV performance can actually be better than even a single SMV. This issue can also be captured by Theorem 1 since by increasing L, the conditions on K, s would be weaker. This in turn shows that for fixed K and s, the performance of our method enhances by increasing the number L of snapshots. It is also worth noting that in the SMV case ($L = 1$), the bound in Theorem 1 reduces to the required sample complexity of SMV case obtained in [18, Theorem 2.2].

The proof of Theorem 1 appears in Section 7. In Section 4 we look at the dual of (4) and reformulate it as an SDP.

4. Dual Problem

According to the infinite dimensionality of gTV norm in (4), we look at its dual formulation and analyze it. The proposed demixing problem (4) is closely related to the atomic norm minimization problem introduced in [9]. Using the fact that our signal of interest is composed of K complex exponentials, we can present it sparsely with an atomic set containing N-dimensional sinusoids. The measurements of each snapshot or time sample form a measurement matrix as in (3). As a consequence, it is crucial that we use matrix form atoms to build up our signal. Consider the following atomic set with $\|.\|_2$ denoting l_2 norm

$$\mathcal{A} = \{\boldsymbol{a}(f,\phi)\boldsymbol{b}^H : f \in [0,1], \phi \in [0,2\pi], \|\boldsymbol{b}\|_2 = 1\}$$

for any $\boldsymbol{b} \in \mathbb{C}^{L \times 1}$ and

$$\boldsymbol{a}(f,\phi) = \frac{1}{\sqrt{N}}e^{i\phi}[1, e^{i2\pi f}, \ldots, e^{i2\pi(N-1)f}]^T \in \mathbb{C}^N.$$

125 Using the above definition of the atomic set, we can define the matrix $\boldsymbol{S}$ as

$$\boldsymbol{S} = \sqrt{N}\sum_{k=1}^{K} \boldsymbol{a}(f_k,\phi)\boldsymbol{\psi}_k^H = \sum_{k=1}^{K} c_k \boldsymbol{a}(f_k,\phi)\boldsymbol{b}_k^H, \tag{5}$$

where $c_k = \sqrt{N}\|\boldsymbol{\psi}\|_2 > 0$ and $\boldsymbol{b}_k = c_k^{-1}\boldsymbol{\psi}_k\sqrt{N}$ with $\|\boldsymbol{b}_k\|_2 = 1$. According to [16, 30], spectral super resolution problem can be treated using atomic norm minimization. This attitude arises from the fact that in spectral super resolution problem the spectrum of the signal of interest is sparse. The atomic norm is defined as

$$\|\boldsymbol{X}\|_{\mathcal{A}} := \inf\{t > 0 : \boldsymbol{X} \in tconv(\mathcal{A})\}, \tag{6}$$

where $conv(\mathcal{A})$ denotes the convex hull of the atomic set $\mathcal{A}$.

Using the definition of the atomic norm, (4) can be represented as

$$\min_{\tilde{\boldsymbol{S}},\tilde{\boldsymbol{Z}}} \|\tilde{\boldsymbol{S}}\|_{\mathcal{A}} + \lambda\|\tilde{\boldsymbol{Z}}\|_{1,2} \quad s.t. \quad \boldsymbol{Y} = \tilde{\boldsymbol{S}} + \tilde{\boldsymbol{Z}}. \tag{7}$$

In order to formulate the dual problem, we need the definition of dual atomic norm as

$$\begin{aligned}
\|\boldsymbol{\Gamma}\|_{\mathcal{A}}^* &= \sup_{\|\tilde{\boldsymbol{S}}\|_{\mathcal{A}} \leq 1} <\boldsymbol{\Gamma}, \tilde{\boldsymbol{S}}>_{\mathbb{F}}, \\
&= \sup_{\substack{f \in [0,1] \\ \phi \in [0,2\pi] \\ \|\boldsymbol{b}\|_2=1}} <\boldsymbol{\Gamma}, e^{i\phi}\boldsymbol{a}(f,0)\boldsymbol{b}^H>_{\mathbb{F}}, \\
&= \sup_{\substack{f \in [0,1] \\ \|\boldsymbol{b}\|_2=1}} | <\boldsymbol{\Gamma}, \boldsymbol{a}(f,0)\boldsymbol{b}^H>_{\mathbb{F}} |, \\
&= \sup_{f \in [0,1]} \|\boldsymbol{\Gamma}^H\boldsymbol{a}(f)\|_2,
\end{aligned}$$

6

where $< . >_{\mathbb{F}}$ shows the Frobenius inner product. Using the above definition, the dual of (7) is written as

$$\max_{\boldsymbol{\Gamma} \in \mathbb{C}^{N \times L}} Re < \boldsymbol{Y}, \boldsymbol{\Gamma} >_{\mathbb{F}} \ s.t. \qquad \sup_{f \in [0,1]} \|\boldsymbol{\Gamma}^H \boldsymbol{a}(f,0)\|_2 \le 1,$$
$$\|\boldsymbol{\Gamma}\|_{\infty,2} \le \lambda, \tag{8}$$

where $Re < . >$ denotes the real part of the inner product and $\|.\|_{\infty,2}$ is the matrix infinity/2 norm defined as

$$\|\boldsymbol{\Gamma}\|_{\infty,2} = \max_i \|\boldsymbol{\Gamma}_{i,:}\|_2.$$

By applying the PTP theory [21], the maximization constraint in (8) can be reformulated as a Linear Matrix Inequality (LMI) given by

$$\max_{\boldsymbol{\Gamma} \in \mathbb{C}^{N \times L}, \boldsymbol{\Lambda} \in \mathbb{C}^{N \times N}} Re < \boldsymbol{Y}, \boldsymbol{\Gamma} >_{\mathbb{F}} \ s.t. \qquad \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}^H & \boldsymbol{I}_L \end{bmatrix} \succeq 0,$$
$$\mathcal{T}^*(\boldsymbol{\Lambda}) = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix},$$
$$\|\boldsymbol{\Gamma}\|_{\infty,2} \le \lambda, \tag{9}$$

where $\mathcal{T}^*$ is defined as

$$\mathcal{T}^*(\boldsymbol{\Lambda})_j = \sum_{i=1}^{N-j+1} \boldsymbol{\Lambda}_{i,i+j-1},$$

$\boldsymbol{I}_L$ denotes the identity matrix of size $L \times L$, $\mathbf{0} \in \mathbb{C}^{N-1}$ is a zero vector, and $\succeq 0$ denotes positive semi-definiteness.

In order to localize the frequencies of the signal of interest and the noisy spikes, Lemma1 is presented.

Lemma 1. *The solution to (7) is unique if for $\boldsymbol{\Gamma} \in \mathbb{C}^{N \times L}$ and the vector-valued dual polynomial $\boldsymbol{Q} = \boldsymbol{a}(f,0)^H \boldsymbol{\Gamma}$, we have*

$$\boldsymbol{Q}(f_k) = \frac{c_k}{|c_k|} \boldsymbol{b}_k^H \quad for \ k \ s.t. \ f_k \in \mathbb{T}, \tag{10a}$$

$$\|\boldsymbol{Q}(f_j)\|_2 < 1 \qquad \forall f_j \in [0,1] \backslash \mathbb{T}, \tag{10b}$$

and for any $d \in \boldsymbol{\Omega}$ and $l \in \boldsymbol{\Omega}^c$,

$$\boldsymbol{\Gamma}_{d,:} = \lambda \frac{\boldsymbol{Z}_{d,:}}{\|\boldsymbol{Z}_{d,:}\|_2}, \tag{10c}$$

$$\|\boldsymbol{\Gamma}_{l,:}\|_2 < \lambda. \tag{10d}$$

Proof. If we find a $\boldsymbol{\Gamma}$ satisfying the above conditions, it is dual feasible. Consider

$\hat{S}$ and $\hat{Z}$ as the solutions to (7). Then, we would have

$$
\begin{aligned}
\|\hat{S}\|_{\mathcal{A}} &\geq \|\hat{S}\|_{\mathcal{A}}\|\Gamma\|_{\mathcal{A}}^* \\
&\geq <\Gamma, \hat{S}>_{\mathbb{R}} \\
&= <\Gamma, \sum_{k=1}^{K} c_k \boldsymbol{a}(f_k, \phi_k) \boldsymbol{b}_k^H >_{\mathbb{R}} \\
&= \sum_{k=1}^{K} Re\{c_k^* <\Gamma, \boldsymbol{a}(f_k, \phi_k)\boldsymbol{b}_k^H >\} \\
&= \sum_{k=1}^{K} Re\{c_k^* <\boldsymbol{b}_k, \boldsymbol{Q}(f_k)^H >\} \\
&= \sum_{k=1}^{K} Re\{c_k^* \frac{c_k}{|c_k|}\} \geq \|\hat{S}\|_{\mathcal{A}}.
\end{aligned}
$$

140 Also,

$$
\begin{aligned}
Re <\hat{Y}, \Gamma> &= Re <\hat{S}, \Gamma> + Re <\hat{Z}, \Gamma> \\
&= \|\hat{S}\|_{\mathcal{A}} + \sum_{d \in \boldsymbol{\Omega}} Re\{\hat{\boldsymbol{Z}}_{d,:}^*, \Gamma_{d,:}\} \\
&= \|\hat{S}\|_{\mathcal{A}} + \lambda \sum_{d \in \boldsymbol{\Omega}} Re\{\frac{\boldsymbol{Z}_{d,:}^* \boldsymbol{Z}_{d,:}}{\|\boldsymbol{Z}_{d,:}\|_2}\} = \lambda\|\boldsymbol{Z}\|_{1,2},
\end{aligned}
$$

where the last equality is derived using 10c. Therefore, we must have $<\Gamma, \hat{S}>_{\mathbb{R}} = \|\hat{S}\|_{\mathcal{A}} + \lambda\|\hat{Z}\|_{1,2}$. Thus, by strong duality, $\hat{S}$ and $\hat{Z}$ are primal optimal and Γ is dual optimal. To investigate uniqueness, we consider $\tilde{S} = \sum_{k \in \widetilde{\mathbb{T}}} \tilde{c}_k \boldsymbol{a}(\tilde{f}_k, \tilde{\phi}_k) \tilde{\boldsymbol{b}}_k^H$ and $\tilde{Z}$ with supports $\widetilde{\mathbb{T}}$ and $\widetilde{\Omega}$, respectively as the other

145 optimal solutions to (7). Then, we get

$$
< \tilde{\boldsymbol{Y}}, \boldsymbol{\Gamma} >_{\mathbb{R}} = < \tilde{\boldsymbol{S}}, \boldsymbol{\Gamma} >_{\mathbb{R}} + < \tilde{\boldsymbol{Z}}, \boldsymbol{\Gamma} >_{\mathbb{R}}
$$

$$
= \sum_{\tilde{f}_k \in \mathbb{T} \cap \widetilde{\mathbb{T}}} Re\{\tilde{c}_k < \tilde{b}_k, \boldsymbol{Q}(\tilde{f}_k)^H >\}
$$

$$
+ \sum_{\tilde{f}_j \in \mathbb{T}^c \cap \widetilde{\mathbb{T}}} Re\{\tilde{c}_j < \tilde{b}_j, \boldsymbol{Q}(\tilde{f}_j)^H >\}
$$

$$
+ \sum_{d \in \Omega \cap \widetilde{\Omega}} Re\{< \tilde{\boldsymbol{Z}}_{d,:}^*, \boldsymbol{\Gamma}_{d,:} >\} + \sum_{l \in \Omega^c \cap \widetilde{\Omega}} Re\{< \tilde{\boldsymbol{Z}}_{l,:}^*, \boldsymbol{\Gamma}_{l,:} >\}
$$

$$
\leq \sum_{\tilde{f}_k \in \mathbb{T} \cap \widetilde{\mathbb{T}}} Re\{\tilde{c}_k \|\tilde{\boldsymbol{b}}_k\|_2 \|\boldsymbol{Q}(\tilde{f}_k)\|_2 + \sum_{\tilde{f}_j \in \mathbb{T}^c \cap \widetilde{\mathbb{T}}} Re\{\tilde{c}_j \|\tilde{\boldsymbol{b}}_j\|_2 \|\boldsymbol{Q}(\tilde{f}_j)\|_2
$$

$$
+ \lambda \sum_{d \in \Omega \cap \widetilde{\Omega}} Re\{\|\tilde{\boldsymbol{Z}}_{d,:}\|_2\} + \|\boldsymbol{\Gamma}_{\Omega^c,:}\|_{\infty,2} \sum_{l \in \Omega^c \cap \widetilde{\Omega}} Re\{\|\tilde{\boldsymbol{Z}}_{l,:}\|_2\}
$$

$$
< \sum_{\tilde{f}_k \in \mathbb{T} \cap \widetilde{\mathbb{T}}} \tilde{c}_k \|\tilde{\boldsymbol{b}}_k\|_2 + \sum_{\tilde{f}_j \in \mathbb{T}^c \cap \widetilde{\mathbb{T}}} \tilde{c}_j \|\tilde{\boldsymbol{b}}_j\|_2 + \lambda \sum_{d \in \Omega \cap \widetilde{\Omega}} Re\{\|\tilde{\boldsymbol{Z}}_d\|_2\}
$$

$$
+ \lambda \sum_{l \in \Omega^c \cap \widetilde{\Omega}} Re\{\|\tilde{\boldsymbol{Z}}_{l,:}\|_2\} = \|\tilde{\boldsymbol{S}}\|_{\mathcal{A}} + \lambda \|\tilde{\boldsymbol{Z}}\|_{1,2},
$$

which contradicts the strong duality. Thus, all optimal solutions are solely supported on $\mathbb{T}$ and Ω. Since the atoms in $\mathbb{T}$ and Ω are linearly independent, the pair $(\hat{\boldsymbol{S}}, \hat{\boldsymbol{Z}})$ is the unique optimal solution to (7). $\square$

5. Demixing in Presence of Dense Perturbation

In many practical scenarios (*e.g.* Direction of Arrival (DOA) estimation), the presence of dense perturbations is unavoidable [31]. When the received signal is perturbed with dense noise, one can modify (4) as a new optimization problem and the corresponding SDP. In this case, the received data is in the form of

$$
\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{Z} + \boldsymbol{W} = [(\mathcal{F}_N G_1), \ldots, (\mathcal{F}_N G_L)] + \boldsymbol{Z} + \boldsymbol{W},
$$

150 where $\boldsymbol{W} \in \mathbb{C}^{N \times L}$ is the additive noise matrix with *i.i.d.* elements distributed as zero mean Gaussian distribution with standard deviation σ. Now, by modifying (4), we reach

$$
\min_{\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{Z}}} \|\tilde{\boldsymbol{G}}\|_{gTV} + \lambda \|\tilde{\boldsymbol{Z}}\|_{1,2} \quad s.t. \ \|\boldsymbol{Y} - [\mathcal{F}_N \tilde{G}_1, \ldots, \mathcal{F}_N \tilde{G}_L] - \tilde{\boldsymbol{Z}}\|_F \leq \eta, \quad (11)
$$

where η is an upper-bound of $\|\boldsymbol{W}\|_F$. In what follows, we derive the dual problem of (11) and its corresponding semidefinite relaxation.

9

Lemma 2. *The dual problem of (11) is*

$$\max_{\boldsymbol{\Gamma}\in\mathbb{C}^{N\times L}} Re <\boldsymbol{Y},\boldsymbol{\Gamma}>_{\mathbb{F}} -\eta\|\boldsymbol{\Gamma}\|_F \; s.t. \qquad \sup_{f\in[0,1]}\|\boldsymbol{\Gamma}^H\boldsymbol{a}(f,0)\|_2 \leq 1,$$
$$\|\boldsymbol{\Gamma}\|_{\infty,2} \leq \lambda, \tag{12}$$

which is equivalent to the following SDP,

$$\max_{\boldsymbol{\Gamma}\in\mathbb{C}^{N\times L},\boldsymbol{\Lambda}\in\mathbb{C}^{N\times N}} Re <\boldsymbol{Y},\boldsymbol{\Gamma}>_F -\eta\|\boldsymbol{\Gamma}\|_F \; s.t. \qquad \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}^H & \boldsymbol{I}_L \end{bmatrix} \succeq 0,$$
$$\mathcal{T}^*(\boldsymbol{\Lambda}) = \begin{bmatrix} 1 \\ \boldsymbol{0} \end{bmatrix},$$
$$\|\boldsymbol{\Gamma}\|_{\infty,2} \leq \lambda, \tag{13}$$

with $\boldsymbol{0} \in \mathbb{C}^{n-1}$ being a vector of zeros.

Proof. Problem (11) can be reformulated as

$$\min_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{Z}}}\|\tilde{\boldsymbol{G}}\|_{gTV} + \lambda\|\tilde{\boldsymbol{Z}}\|_{1,2} \; s.t. \; \|\boldsymbol{Y}-\boldsymbol{U}\|_{\mathbb{F}}^2 \leq \eta^2$$
$$\boldsymbol{U} = [\mathcal{F}_N\tilde{G}_1,\ldots,\mathcal{F}_N\tilde{G}_L] + \tilde{\boldsymbol{Z}}.$$

The Lagrangian of the above problem is

$$\mathcal{L}(\tilde{\boldsymbol{G}},\tilde{\boldsymbol{Z}},\boldsymbol{\Gamma}) = \|\tilde{\boldsymbol{G}}\|_{gTV} - \langle[\tilde{G}_1,\ldots,\tilde{G}_L],\mathcal{F}_N^*\boldsymbol{\Gamma}\rangle_F + \lambda\|\tilde{\boldsymbol{Z}}\|_{1,2} - \langle\tilde{\boldsymbol{Z}},\boldsymbol{\Gamma}\rangle_F$$
$$+ \langle\boldsymbol{U},\boldsymbol{\Gamma}\rangle_F + \nu\left(\|\boldsymbol{Y}-\boldsymbol{U}\|_F^2 - \eta^2\right). \tag{14}$$

Due to the first constraint in (12), $\|\tilde{\boldsymbol{G}}\|_{gTV} - \langle[\tilde{G}_1,\ldots,\tilde{G}_L],\mathcal{F}_N^*\boldsymbol{\Gamma}\rangle_F$ is minimized for $\tilde{\boldsymbol{G}} = 0$ and because of the second constraint in (12), $\lambda\|\tilde{\boldsymbol{Z}}\|_{1,2} - \langle\tilde{\boldsymbol{Z}},\boldsymbol{\Gamma}\rangle_F$ is minimized for $\tilde{\boldsymbol{Z}} = 0$. Next, considering the convexity of (14), we evaluate its gradient w.r.t $\boldsymbol{U}$ and set the result to zero to get,

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{U}} = \boldsymbol{\Gamma} - 2\nu(\boldsymbol{Y}-\boldsymbol{U}) = 0 \rightarrow \boldsymbol{U} = \boldsymbol{Y} - \frac{\boldsymbol{\Gamma}}{2\nu}.$$

Therefore,

$$\langle\boldsymbol{Y},\boldsymbol{\Gamma}\rangle_F - \frac{1}{2\nu}\|\Gamma\|_F^2 + \nu\left(\|\frac{\Gamma}{2\nu}\|_F^2 - \eta^2\right) = \langle\boldsymbol{Y},\boldsymbol{\Gamma}\rangle_F - \frac{1}{4\nu}\|\Gamma\|_F^2 - \nu\eta^2. \tag{15}$$

Since ν is positive, taking derivative w.r.t ν, setting the result equal to zero, and plugging the result back leads to (12) with the constraints used. Thus, (12) is the dual problem of (11) and using the PTP theory [21], (13) is concluded. $\square$
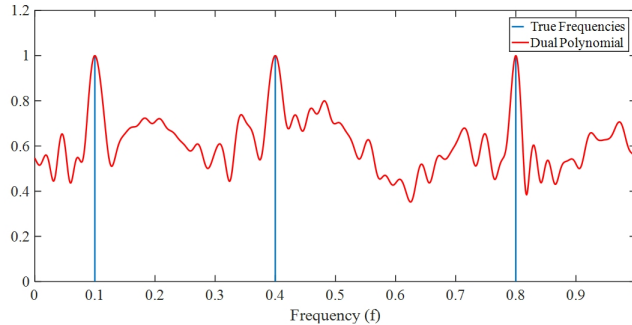
Figure 1: ℓ_2 norm of dual polynomial and true frequencies.

6. Numerical Results

6.1. Without Perturbation

In this subsection, numerical experiments are presented to evaluate the performance of the method proposed in Section 4. First, we investigate the constraints (10a) and (10b) on the dual polynomial and the constraints (10c) and (10d) on the dual variable. Using these constraints, one can localize the signal frequencies and the outliers' spikes. Next, the minimum required frequency separation for successful recovery in the MMV case is compared with the one needed in the SMV case. In all simulations of this subsection, the number of sensors or the signal length is set to $N = 50$. In the first part of the simulations, the signal of interest $\boldsymbol{S} \in \mathbb{C}^{N \times L}$ has $K = 3$ frequencies and the coefficients a_{kl} are always drawn from a standard i.i.d complex Gaussian distribution. The outliers' spikes are considered to be in $s = 3$ different random positions in each snapshot. For better visualization, it is assumed that outliers happen in each sensor only once. Figure (1) depicts $\|\boldsymbol{Q}(f)\|_2$ for $L = 5$ snapshots and $\mathbb{T} = \{0.1, 0.4, 0.8\}$. As it can be observed, the signal frequencies can be estimated by solving $\|\boldsymbol{Q}(f)\|_2 = 1$ for all $f \in [0, 1]$. The outliers are localized in each receiving sensor using (10c). We considered $s = 3$ noisy spikes occurring randomly in each measurement without replacement. Thus, with $L = 5$ we expect to detect 15 outliers in the receiver. Figure (2) depicts the result. As it turns out, Figure (2) verifies the conclusion of Lemma 1.

Next, we investigate the minimum separation condition. To do this, we consider two frequencies slowly taking distance. The first frequency is fixed at $f_1 = 0.2$ and the second one has a distance of $f_\delta = \{0.1/N : 0.1/N : 1.5/N\}$ from f_1. During this experiment, $s = 10$ outliers out of $N = 50$ are considered in the overall measurement process. Also, we define $\boldsymbol{f}_{est} = [f_1^{est}, f_2^{est}]$ as the estimated frequencies vector. A successful estimation is defined as when

$$\max\{|\boldsymbol{f}_{est} - \boldsymbol{f}_{true}|\} \leq 10^{-4} \tag{16}$$

where $\boldsymbol{f}_{true}$ denotes the true frequencies. With this definition, Figure (3) illustrates the probability of successful recovery for $L = \{1, 3, 5\}$ over 100 Monte-
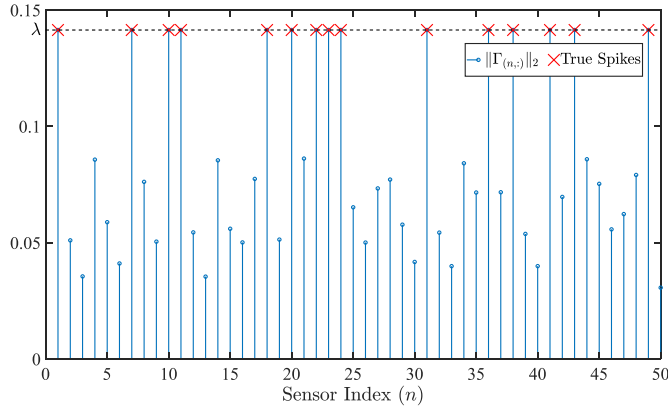
11

Figure 2: ℓ_2 norm of $\boldsymbol{\Gamma}$ rows in Lemma 1 in terms of sensor indices. The estimated spike locations are found by identifying indices where the ℓ_2 norm of Γ rows achieves λ.
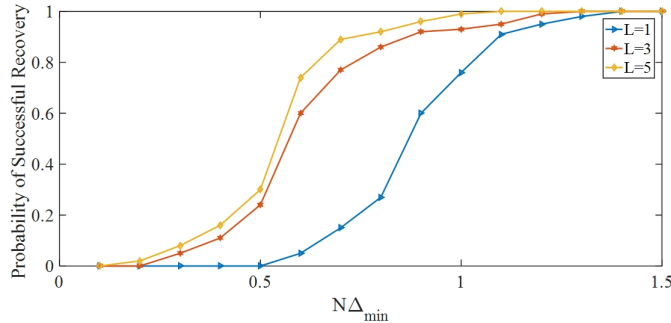

Figure 3: Probability of successful recovery in terms of minimum separation for various number of snapshots.

Carlo simulations. As seen, the minimum required frequency separation is decreased with an increase in the number of snapshots. For more illustration on Theorem 1, phase transition diagrams for s and K are plotted in Figures 4 and 5 with parameters $N = 50, K = 5$, and $N = 50, s = 5$, respectively. In each Monte-Carlo simulation, the frequencies are randomly generated satisfying the separation condition and the total number of Monte-Carlo iterations is considered to be 100. As Figure (5) shows, for fixed N and K, the probability of successful recovery increases when the number of snapshots (L) rises. Also, increasing the number of sources K leads to recovery corruption for fixed L and N. Moreover, Figure (4) shows that increasing the number of snapshots can affect the maximum number of possible spikes to recover. However, this behaviour is up to a limit point. This is aligned with our bound in Theorem 1.
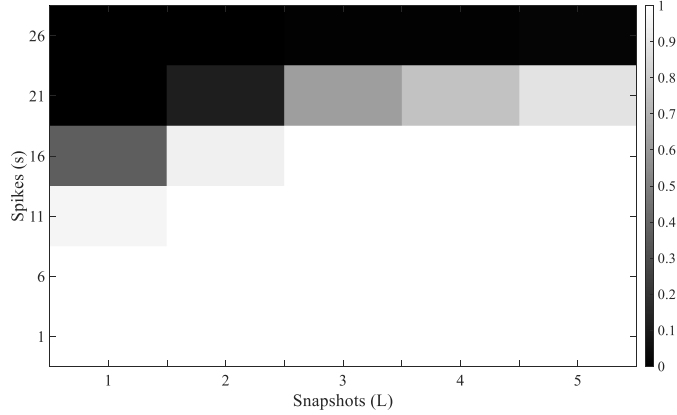
Figure 4: Performance of our atomic norm minimization method for diverse values of the number of snapshots (L) and outliers' spikes (s). The heat-map shows the probability of successful recovery (white: success black: failure).
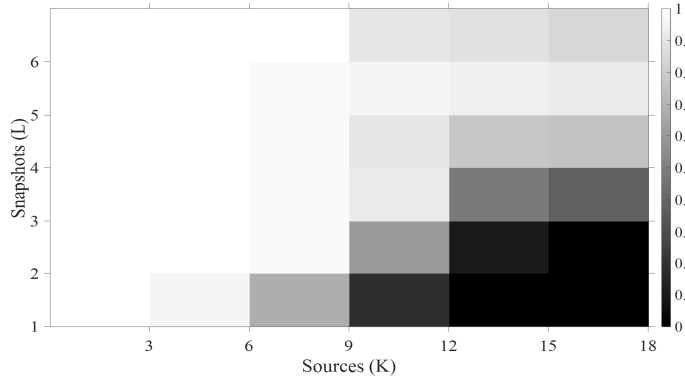

Figure 5: Performance of our atomic norm minimization method for diverse values of the number of snapshots (L) and sources (K). The heat-map shows the probability of successful recovery (white: success black: failure).

6.2. With Perturbation

In this subsection, various scenarios are considered and the results are com-
pared with the state of the art SPA [32] method. Note that comparison with
conventional SPICE method [33] is not implemented as this method was de-
signed to estimate on-grid frequencies leading to basis mismatch issues and
estimation inaccuracies. Therefore, we only compare our method with SPA [32]
which is often regarded as a continuous version of SPICE [33]. In order to get
closer to a more realistic scenario, consider the DOA estimation problem with
$K = 3$ sources where the first and third sources are considered coherent. Take
the incoming directions to be $\boldsymbol{f}_{true} = [0.1, 0.4, 0.8]$ and the corresponding esti-
mation vector to be $\boldsymbol{f}_{est}$. In what follows, we compare our proposed method
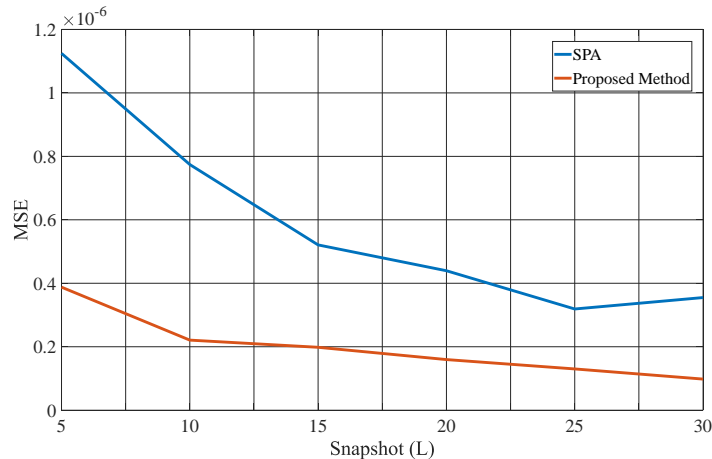
13

Figure 6: Comparison of the performance of proposed method over SPA when the number of snapshots increases and the spikes energy is low.

with the SPA [32] when both methods are perturbed with impulsive spiky and Gaussian dense noises.

6.2.1. Effect of Taking Snapshots

To compare the performance of the two methods for different number of snapshots, we set the number of DOA sensors, N, to 50 and the number of spikes, s, to 10. The elements of the noise matrix $\boldsymbol{W}$ are considered to be *i.i.d.* and distributed as Gaussian with zero mean and variance 0.5. Also, the spectral norm of the spiky noise $\|\boldsymbol{Z}\|_F$ is set to 25. This value does not affect the performance of the proposed method but devastates the performance of SPA. The number of snapshots, L, is ranged from 5 to 30 and the result is presented in Figure 6 for 100 Monte-Carlo simulations. The error is measured in terms of the Mean Square Error (MSE) defined as

$$MSE = \frac{1}{K}\|\boldsymbol{f}_{est} - \boldsymbol{f}_{true}\|_2^2.$$

As Figure (6) depicts, both methods show improvements as the number of snapshots increases. However, it is apparent that the proposed method has higher accuracy than the SPA method.

6.2.2. Effect of Impulsive Noise Energy

Here, we would like to investigate the performance of two methods for various levels of $\|\boldsymbol{Z}\|_F$ which is the Frobenius norm of $\boldsymbol{Z}$. The setting is similar to that of the previous subsection except that the number of snapshots is fixed to 10. The result is shown in Figure (7). As expected, the performance of the proposed method does not change considerably compared to the SPA in this
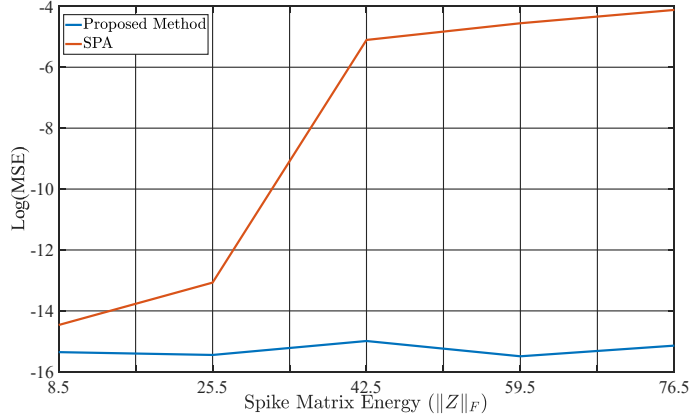
14

Figure 7: Comparison of the performance of proposed method over SPA in terms of the energy of spikes *i.e.* $\|Z\|_F$.

scenario. However, it can be observed that by increasing the energy of spikes, the performance of SPA is corrupted in such a way that after some threshold a successful recovery will be impossible.

6.2.3. SNR

₂₂₅   This subsection focuses on the effect of Gaussian noise variance employed in W on the performance of both methods in terms of MSE. The simulation is based on the same setting as the first subsection except that the variance of each element in W *i.e.* σ^2 varies from 0.5 to 6.5, $L = 5$ and $s = 5$ with the spikes energy level of 30. The result is shown in Figure (8). As seen, σ^2 can ₂₃₀ have a huge impact on the performance. Both methods tend to experience a threshold after which their MSEs increase dramatically. However, this threshold for the proposed method is significantly higher than the SPA, which indicates the greater robustness of the proposed method against Gaussian noise energy.

7. Proof of Theorem 1

₂₃₅   In order to prove that problem (7) achieves exact demixing, we construct a trigonometric dual polynomial. Following the same line of [18], we apply the following kernel to build up the dual polynomial,

$$\bar{K}(f) \quad := \quad \mathcal{D}_{0.247m}(f)\mathcal{D}_{0.339m}(f)\mathcal{D}_{0.414m}(f) = \sum_{l=-m}^{m} c_l e^{i2\pi lf},$$
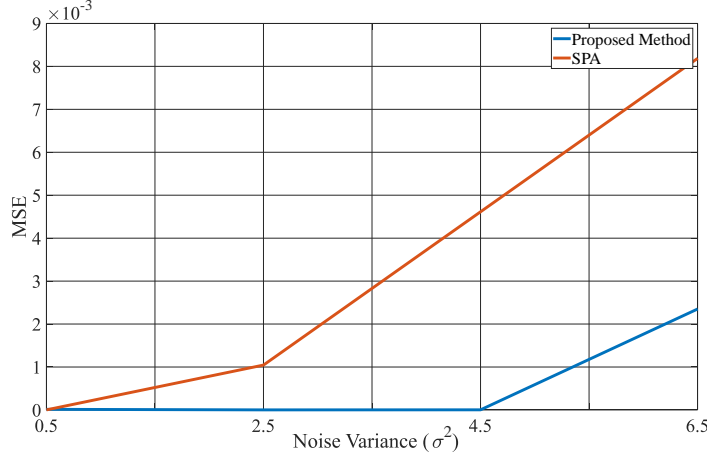
15

Figure 8: Performance of the proposed method over SPA when the dense noise energy increases.

where $N = 2m + 1$, $\boldsymbol{c} \in \mathbb{C}^N$ is the convolution of the Fourier coefficients of the above kernels, and $\mathcal{D}_m$ is the Dirichlet kernel of order $m > 0$ defined as

$$\mathcal{D}_m(f) := \frac{1}{N} \sum_{l=-m}^{m} e^{i2\pi l f}.$$

According to the presence of outliers, conventional forms of dual polynomial can not be applied since the constraints (10d) and (10c) will not be met. Therefore, we use the randomized vector form of the dual polynomial presented in [18] as

$$\boldsymbol{Q}(f) = \boldsymbol{Q}_{aux}(f) + \boldsymbol{R}(f), \tag{17}$$

where

$$\boldsymbol{Q}_{aux}(f) = \sum_{l \in \boldsymbol{\Omega}^c} \boldsymbol{\Gamma}_{l,:} e^{-i2\pi l f},$$

$$\boldsymbol{R}(f) = \frac{1}{\sqrt{N}} \sum_{d \in \boldsymbol{\Omega}} \boldsymbol{r}_{d,:} e^{-i2\pi d f},$$

where $\boldsymbol{r}_{d,:} = \frac{\boldsymbol{Z}_{d,:}}{\|\boldsymbol{Z}_{d,:}\|_2}$ and $\boldsymbol{r} \in \mathbb{C}^{s \times L}$. Note that (10c) is immediately satisfied since $\lambda = 1/\sqrt{N}$. Now we should build up the dual polynomial so that the other constraints in Lemma 1 are met. Using the same interpolation technique of [27], we set the value of the dual polynomial equal to $\frac{c_k}{|c_k|} \boldsymbol{b}_k^H = h_k \boldsymbol{b}_k^H$ at $f_k \in \mathbb{T}$ and set the derivative of the dual polynomial equal to zero at the same points. Setting the derivative to zero forces the dual polynomial to shape such that f_k be a local extremum and bounds the value of the dual polynomial at these points.

16

Thus, the following set of equations is formed for any $f_k \in \mathbb{T}$

$$\boldsymbol{Q}(f_k) = h_k \boldsymbol{b}_k^H, \tag{18a}$$

$$\boldsymbol{Q}_R^{(1)}(f_k) + i\boldsymbol{Q}_I^{(1)} = 0, \tag{18b}$$

where $\boldsymbol{Q}_R^{(1)}$ denotes the real part of the first derivative of $\boldsymbol{Q}$ and $\boldsymbol{Q}_I$ is the imaginary part of $\boldsymbol{Q}$. Using (17) in the above equations yields

$$\boldsymbol{Q}_{aux}(f_k) = h_k \boldsymbol{b}_k^H - \boldsymbol{R}(f_k), \tag{19a}$$

$$(\boldsymbol{Q}_{aux})_R^{(1)}(f_k) + i(\boldsymbol{Q}_{aux})_I^{(1)}(f_k) = -\boldsymbol{R}_R^{(1)}(f_k) - i\boldsymbol{R}_I^{(1)}(f_k). \tag{19b}$$

250 To interpolate $\boldsymbol{Q}(f)$ with $\bar{K}(f)$, we need to confine the kernel to $\boldsymbol{\Omega}^c$, as discussed for the missing data case in [16]. Thus,

$$K(f) := \sum_{l \in \boldsymbol{\Omega}^c} c_l e^{i2\pi l f} = \sum_{l=-m}^{m} \delta_{\boldsymbol{\Omega}^c}(l) c_l e^{i2\pi l f}, \tag{20}$$

where $\delta_{\boldsymbol{\Omega}^c}(l)$ are Bernoulli random variables with parameter $\frac{N-s}{N}$. Therefore, $\mathbb{E}K$ is an scaled version of $\bar{K}$

$$\mathbb{E}K(f) = \frac{N-s}{N} \sum_{l=-m}^{m} c_l e^{i2\pi l f} = \frac{N-s}{N} \bar{K}(f). \tag{21}$$

The asymptotic behaviour of $K(f)$, $\bar{K}(f)$, and their derivatives is investigated 255 in [27]. With $K(f)$ restricted to $\boldsymbol{\Omega}^c$ we can express $\boldsymbol{Q}_{aux}$ in terms of $K(f)$ and its first derivative as

$$\boldsymbol{Q}_{aux} = \sum_{k=1}^{K} \boldsymbol{\alpha}_k K(f - f_k) + \kappa \boldsymbol{\beta}_k K^{(1)}(f - f_k), \tag{22}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^{K \times L}$ and $\boldsymbol{\beta} \in \mathbb{C}^{K \times L}$ are such that (19a) and (19b) are satisfied and $\kappa := 1/\sqrt{\bar{K}^{(2)}(0)}$. $\boldsymbol{\alpha}_k$ is the kth row of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_k$ is the kth row of $\boldsymbol{\beta}$. The system of equations is then represented as

$$\begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_1 \\ \boldsymbol{D}_1^T & \boldsymbol{D}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix} - \frac{1}{\sqrt{N}} \boldsymbol{B_\Omega r}, \tag{23}$$

260 where $\boldsymbol{0} \in \mathbb{C}^{K \times L}$ is a zero matrix, $\boldsymbol{\Phi}_{k,:} = h_k \boldsymbol{b}_k^H$,

$$(D_0)_{jl} = K(f_j - f_l), (D_1)_{jl} = \kappa K^{(1)}(f_j - f_l), (D_2)_{jl} = -\kappa^2 K^{(2)}(f_j - f_l),$$

$$\frac{1}{\sqrt{N}} \boldsymbol{B_\Omega r} = [\boldsymbol{R}(f_1), \dots, \boldsymbol{R}(f_k), \boldsymbol{R}^{(1)}(f_1), \dots, \boldsymbol{R}^{(1)}(f_k)]^T,$$

$$\boldsymbol{B_\Omega} = [\boldsymbol{\nu}(d_1), \dots, \boldsymbol{\nu}(d_s)],$$

17

$$\boldsymbol{\nu}(g) := [e^{-i2\pi gf_1}, \ldots \quad , \quad e^{-i2\pi gf_k},$$
$$i2\pi g\kappa e^{-i2\pi gf_1} \quad , \quad \ldots, i2\pi g\kappa e^{-i2\pi gf_k}]^T.$$

By solving (23), one can find $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and define $\boldsymbol{Q}(f)$ as

$$\boldsymbol{Q}(f) = \sum_{k=1}^{K} \boldsymbol{\alpha}_k K(f - f_k) + \kappa \boldsymbol{\beta}_k K^{(1)}(f - f_k) + \boldsymbol{R}(f) \qquad (24a)$$

$$= \boldsymbol{G}_0^T(f) \boldsymbol{D}^{-1} \left(\begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix} - \frac{1}{\sqrt{N}} \boldsymbol{B_\Omega r} \right) + \boldsymbol{R}(f), \qquad (24b)$$

where $\boldsymbol{G}_p(f)$ is defined as

$$\boldsymbol{G}_p(f) := \kappa^p [K^{(p)}(f - f_1), \ldots, K^{(p)}(f - f_k),$$
$$\kappa K^{(p+1)}(f - f_1), \ldots, \kappa K^{(P+1)}(f - f_k)]^T, \qquad (25)$$

for $p = 0, 1, 2, \ldots$ Now we should verify that the polynomial we formed above is guaranteed to be valid with high probability. If one can prove that $\boldsymbol{D}^{-1}$ exists, then (23) can be solved and (10a) holds. Consider $\bar{\boldsymbol{D}}$ as the deterministic version of $\boldsymbol{D}$. Lemma 8 helps defining a condition under which $\boldsymbol{D}^{-1}$ exists and its deviation is bounded. We consider $\varepsilon_{\boldsymbol{D}}^c$ as the event in which $\boldsymbol{D}^{-1}$ exists with probability $1 - \epsilon/5$ for $\epsilon > 0$ under the assumption of Theorem 1. With this Lemma, one can conclude that in $\varepsilon_{\boldsymbol{D}}^c$ (10a) holds. Note that (10c) holds according to the definition of $\boldsymbol{Q}(f)$. All that remains is to prove (10b) and (10d). We use the results of Lemma 7,9 below and Lemma 3.5 from [18] which put bounds on the deviations of $\boldsymbol{B_\Omega}$, $\boldsymbol{\nu}(d)$, and $\boldsymbol{G}_p(f)$, respectively. We use $\varepsilon_{\boldsymbol{B}}^c$ and $\varepsilon_{\boldsymbol{\nu}}^c$ as the events in which $\boldsymbol{B_\Omega}$ and $\boldsymbol{\nu}(d)$ are bounded with probability at least $1 - \epsilon/5$ under the assumption of Theorem 1, respectively.

Proposition 1. *Under the assumption of Theorem 1 and conditioned on $\varepsilon_{\boldsymbol{B}}^c \cap \varepsilon_{\boldsymbol{D}}^c \cap \varepsilon_{\boldsymbol{\nu}}^c$, (10b) holds with probability at least $1 - \epsilon/5$.*

Proof. Consider $\bar{\boldsymbol{Q}}(f)$ as the dual polynomial constructed using $\bar{\boldsymbol{K}}(f)$. We can

rewrite (24b) in a more general form for $\boldsymbol{K}(f)$ and $\bar{\boldsymbol{K}}(f)$ as

$$\kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f) := \kappa^\iota \sum_{j=1}^{K} \bar{\boldsymbol{\alpha}}_j \bar{\boldsymbol{K}}^{(\iota)}(f - f_j) +$$

$$\kappa^{\iota+1} \sum_{j=1}^{K} \bar{\boldsymbol{\beta}}_j \bar{\boldsymbol{K}}^{(\iota+1)}(f - f_j) = \bar{\boldsymbol{G}}_\iota(f)^T \bar{\boldsymbol{D}}^{-1} \begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix} \qquad (26)$$

$$\kappa^\iota \boldsymbol{Q}^{(\iota)}(f) := \kappa^\iota \sum_{j=1}^{K} \boldsymbol{\alpha}_j \boldsymbol{K}^{(\iota)}(f - f_j) +$$

$$\kappa^{\iota+1} \sum_{j=1}^{K} \boldsymbol{\beta}_j \boldsymbol{K}^{(\iota+1)}(f - f_j) + \kappa^\iota \boldsymbol{R}^{(\iota)}(f)$$

$$= \boldsymbol{G}_\iota(f)^T \boldsymbol{D}^{-1} \left(\begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix} - \frac{1}{\sqrt{N}} \boldsymbol{B}_\Omega \boldsymbol{r} \right) + \kappa^\iota \boldsymbol{R}^{(\iota)}(f). \qquad (27)$$

We can also express (27) as

$$\kappa^\iota \boldsymbol{Q}^{(\iota)}(f) := \kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f) + \kappa^\iota \boldsymbol{R}^{(\iota)}(f) - \frac{1}{\sqrt{N}} \boldsymbol{G}_\iota(f)^T \boldsymbol{D}^{-1} \boldsymbol{B}_\Omega \boldsymbol{r}$$

$$+ (\boldsymbol{G}_\iota(f) - \frac{N-s}{N} \bar{\boldsymbol{G}}_\iota(f))^T \boldsymbol{D}^{-1} \begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix}$$

$$+ \frac{N-s}{N} \bar{\boldsymbol{G}}_\iota(f)^T (\boldsymbol{D}^{-1} - \frac{N}{N-s} \bar{\boldsymbol{D}}^{-1}) \begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix}.$$

Noting that $\|\boldsymbol{Q}(f)\|_2 \le \|\bar{\boldsymbol{Q}}(f)\|_2 + \|\boldsymbol{Q}(f) - \bar{\boldsymbol{Q}}(f)\|_2$, for (10b) to hold, we should have $\|\bar{\boldsymbol{Q}}(f)\|_2 + \|\boldsymbol{Q}(f) - \bar{\boldsymbol{Q}}(f)\|_2 \le 1$. The following lemmas complete the proof.

Lemma 3. *Under the assumptions of Proposition 1,* $\|\boldsymbol{Q}(f) - \bar{\boldsymbol{Q}}(f)\|_2 \le 10^{-2}$.

Lemma 4. *Under the assumptions of Proposition 1,* $\|\bar{\boldsymbol{Q}}(f)\|_2 < 0.99$. *Also*

$$\frac{1}{2} \frac{d^2 \|\boldsymbol{Q}(f)\|_2}{df^2} = \|\boldsymbol{Q}'\|_2^2 + Re\{\boldsymbol{Q}''\boldsymbol{Q}^H(d)\} < 0, \qquad (28)$$

$$\forall f \in A_{near} := \{f \,|\, |f - f_j| \le 0.09 \; for f_j \in \mathbb{T}\}.$$

The proof of the above lemmas appear in Section 9. $\qquad \square$

Now, we prove (10d) as the last step to prove Theorem 1.

Proposition 2. *Under the assumption of Theorem 1 and conditioned on* $\varepsilon_{\boldsymbol{B}}^c \cap \varepsilon_{\boldsymbol{D}}^c \cap \varepsilon_{\boldsymbol{\nu}}^c$, *(10d) holds with probability at least* $1 - \epsilon/5$.

Proof. We can express $\mathbf{\Gamma}_{l,:}$ as

$$\mathbf{\Gamma}_{l,:} = \sum_{j=1}^{K} c_l \boldsymbol{\alpha}_j e^{i2\pi l f_j} + i2\pi l\kappa \sum_{j=1}^{K}\sum_{j=1}^{K} \boldsymbol{\beta}_j e^{i2\pi l f_j}$$
$$= c_l \boldsymbol{\nu}(l)^H \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = c_l \boldsymbol{\nu}(l)^H \boldsymbol{D}^{-1} \left(\begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix} - \frac{1}{\sqrt{N}} \boldsymbol{B}_\Omega \boldsymbol{r} \right)$$
$$= c_l \left(< P\boldsymbol{D}^{-1}\boldsymbol{\nu}(l), \boldsymbol{\Phi} > + \frac{1}{\sqrt{N}} < \boldsymbol{B}_\Omega^H \boldsymbol{D}^{-1}\boldsymbol{\nu}(l), \boldsymbol{r} > \right). \qquad (29)$$

We use the results from [18] to bound $\|P\boldsymbol{D}^{-1}\boldsymbol{\nu}(l)\|_2$ and $\boldsymbol{B}_\Omega^H\boldsymbol{D}^{-1}\boldsymbol{\nu}(l)$,

$$\|P\boldsymbol{D}^{-1}\boldsymbol{\nu}(l)\|_2^2 \le 640K \le \frac{0.18^2 N}{log40/\epsilon} \quad in \; \varepsilon_D^c \qquad (30a)$$

$$\|\boldsymbol{B}_\Omega^H\boldsymbol{D}^{-1}\boldsymbol{\nu}(l)\|_2^2 \le 640C_{\boldsymbol{B}}^2 KN \le \frac{0.18^2 N^2}{log40/\epsilon} \quad in \; \varepsilon_D^c \cap \varepsilon_B^c. \qquad (30b)$$

By applying the vector form of the Hoeffding's inequality [22] with $t = 0.18\sqrt{N}$ for (30a) and $t = 0.18N$ for (30b), we can conclude that each term in (29) is greater than its corresponding t with probability $\epsilon/10$. Thus,

$$\|\mathbf{\Gamma}_{l,:}\|_{\infty,2} \le$$
$$\|\boldsymbol{c}\|_\infty \left(\|\boldsymbol{\nu}(l)^H \boldsymbol{D}^{-1} P^T \boldsymbol{\Phi}\|_2 + \frac{1}{\sqrt{N}} \|\boldsymbol{\nu}(l)^H \boldsymbol{D}^{-1} \boldsymbol{B}_\Omega \boldsymbol{r}\|_2 \right)$$
$$\le \frac{2.6}{N}(0.36\sqrt{N}) = \frac{0.936}{\sqrt{N}} < \frac{1}{\sqrt{N}},$$

with probability at least $1 - \epsilon/5$. $\qquad \square$

8. Conclusion and Future Work

The problem of demixing exponential form signals and outliers using MMVs was discussed. A new convex optimization problem was proposed to solve the demixing problem. It was shown that with the minimum frequency separation condition satisfied, there exists a dual polynomial which interpolates the sign pattern of the signal and helps estimating the signal frequencies. Also, the dual variable was utilized to localize the outliers in the receiver.

As an extension to this work, one can investigate the demixing problem using an arbitrary sampling scheme. This is the case when integer sampling is not possible. Also, the computational complexity of the available SDPs is high. For practical purposes, it is mandatory to reduce the computational complexity of the proposed method.

9. Proofs

9.1. Proof of Lemma3

First, we bound $\|\kappa^\iota \boldsymbol{Q}^{(\iota)}(f) - \kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f)\|_2$ on a grid. Then, the result is extended to the continuous domain $[0, 1]$ and then (10b) is proved. In order to bound $\|\kappa^\iota \boldsymbol{Q}^{(\iota)}(f) - \kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f)\|_2$, we can bound each term in

$$
\|\kappa^\iota \boldsymbol{R}^{(\iota)}(f)\|_2 + \|\frac{1}{\sqrt{N}}\boldsymbol{G}_\iota(f)^T \boldsymbol{D}^{-1}\boldsymbol{B}_{\boldsymbol{\Omega}}\boldsymbol{r}\|_2
$$

$$
+ \|(\boldsymbol{G}_\iota(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f))^T \boldsymbol{D}^{-1} \left[\begin{array}{c} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{array}\right] \|_2
$$

$$
+ \|\frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)^T (\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1}) \left[\begin{array}{c} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{array}\right] \|_2 \tag{31}
$$

on a grid $\mathcal{G}$ such that $|\mathcal{G}| = 200\sqrt{L}N^3$ where $|\mathcal{G}|$ is the cardinality of $\mathcal{G}$. Since, $\iota \in \{0, 1, 2, 3\}$, we are dealing with $|\mathcal{U}| = 4|\mathcal{G}|$ points. To bound each term in (31), we leverage Lemma 4 of [34], which is stated as follows.

Lemma 5 ([34] Lemma 4). *Consider a matrix $\boldsymbol{\Psi} \in \mathbb{C}^{K \times L}$ with rows $\{\boldsymbol{\Psi}_k\}_{k=1}^K$ and the vector $\boldsymbol{0} \neq \boldsymbol{\omega} \in \mathbb{C}^K$. If the rows of $\boldsymbol{\Psi}$ are independently distributed on the complex hyper-sphere $\mathbb{S}^{2L-1}$, then for all $t > \|\boldsymbol{w}\|_2$, we have*

$$
\mathbb{P}\{\|\sum_{k=1}^K \boldsymbol{\omega}_k \boldsymbol{\Psi}_k\|_2 \geq t\} \leq e^{-L\left(\frac{t^2}{\|\omega\|_2^2} - log\frac{t^2}{\|\omega\|_2^2} - 1\right)} \quad \forall \boldsymbol{\omega} \in \mathbb{C}^K, \boldsymbol{\omega} \neq 0, t > 0. \tag{32}
$$

Each term in (31) is associated with an event ε_q and $q = \{1, 2, 3, 4\}$. For the ease of reading, we separate the proof of the bounds on each term.

9.1.1. Bound on ε_1

The first term in (31) can be expressed as

$$
\kappa^\iota \boldsymbol{R}^{(\iota)}(f) = \frac{\kappa^\iota}{\sqrt{N}}\sum_{d \in \boldsymbol{\Omega}} \boldsymbol{r}_{d,:}(i2\pi d)^{(\iota)}e^{-i2\pi df} \quad (\iota) = \{0, 1, 2, 3\}.
$$

Therefore, we define

$$
\varepsilon_1 := \{\|\kappa^\iota \boldsymbol{R}^{(\iota)}(f)\|_2 \geq t \ \ for \ \ all \ \ f \in |\mathbb{T}_{grid}|\}.
$$

By setting $\boldsymbol{\Psi} = \boldsymbol{r}$ and

$$
\boldsymbol{\omega} = \frac{\kappa^\iota}{\sqrt{N}}\left[(i2\pi l_1)^{(\iota)}e^{i2\pi l_1 f}, \ldots, (i2\pi l_s)^\iota e^{i2\pi l_s f}\right]^T,
$$

in (32) and using the union bound, we can conclude that

$$
\mathbb{P}\{\sup_{f \in \mathcal{U}}\|\kappa^\iota \boldsymbol{R}^{(\iota)}(f)\|_2 \geq t\} \leq |\mathbb{T}_{grid}|e^{-L\left(\frac{t^2}{\|\omega\|_2^2} - log\frac{t^2}{\|\omega\|_2^2} - 1\right)}. \tag{33}
$$

If we set

$$\frac{t^2}{\|\omega\|_2^2} - \log \frac{t^2}{\|\omega\|_2^2} - 1 \geq \frac{1}{L} \log \frac{|\mathbb{T}_{grid}|}{\epsilon/20}, \tag{34}$$

we get at most $\epsilon/20$ probability of occurrence for (33). By leveraging [34, Lemma 5], a sufficient condition for (34) to hold is

$$\frac{t^2}{\|\omega\|_2^2} \geq 2(1 + \frac{1}{L} \log \frac{|\mathbb{T}_{grid}|}{\epsilon/20})$$

$$\rightarrow \|\omega\|_2^2 \leq t^2 \left(2(1 + \frac{1}{L} \log \frac{|\mathbb{T}_{grid}|}{\epsilon/20})\right)^{-1} \leq \frac{t^2}{2} \left(1 + \frac{1}{L} \log \frac{|\mathbb{T}_{grid}|}{\epsilon/20}\right)^{-1}. \tag{35}$$

The above result combined with the bound [18]

$$\|\boldsymbol{\omega}\|_2^2 \leq \frac{\kappa^{2\iota}}{N} (2\pi m)^{2\iota} s \leq \frac{\pi^6 s}{N}, \tag{36}$$

leads to the sufficient condition,

$$s \leq \frac{N}{\pi^6} (1 + \frac{1}{L} \log \frac{|\mathbb{T}_{grid}|}{\epsilon})^{-1},$$

which is actually satisfied by the second sufficient condition in Theorem 1 after setting $t = \frac{10^{-2}}{8}$ and C_s small enough. Thus, one can conclude that the event ε_1 happens with probability at most $\epsilon/20$ under the assumptions of Proposition 1.

9.1.2. Bound on ε_2

Following the same procedure as for ε_1, one can bound the second term in (31). Consider $\boldsymbol{\Psi} = \boldsymbol{r}$ and

$$\boldsymbol{\omega} = \frac{1}{\sqrt{N}} \boldsymbol{G}_\iota^T(f) \boldsymbol{D}^{-1} \boldsymbol{B_\Omega}.$$

Note that we can write

$$\|\frac{1}{\sqrt{N}} \boldsymbol{G}_\iota^T(f) \boldsymbol{D}^{-1} \boldsymbol{B_\Omega}\|_2 \leq \frac{1}{\sqrt{N}} \|\boldsymbol{B_\Omega}\| \|\boldsymbol{D}^{-1}\| \|\boldsymbol{G}_\iota(f)\|_2. \tag{37}$$

where $\|.\|$ denotes the operator norm. Now, we should find the sufficient conditions for bounding each term of (37). The bound for the terms $\|\boldsymbol{D}^{-1}\|$ and $\|\boldsymbol{G}_\iota(f)\|_2$ can be found below in Lemmas 8 and 9, respectively. Lemma 6 will provide a new bound for $\|\boldsymbol{B_\Omega}\|$.

Lemma 6. *Under the assumptions of Theorem 1, the event*

$$\varepsilon_{\boldsymbol{B}} = \left\{ \|\boldsymbol{B_\Omega}\| > C_{\boldsymbol{B}} \sqrt{N} \left(\log(\frac{N}{\epsilon})(1 + \frac{1}{L} \log \frac{N^3 \sqrt{L}}{\epsilon}) \right)^{-1/2} \right\}$$

will occur with probability at most $\epsilon/5$ for some constant $C_{\boldsymbol{B}}$.

22

Proof. Define $H := \boldsymbol{B_\Omega} \boldsymbol{B_\Omega^H}$ which is

$$H = \sum_{l \in \boldsymbol{\Omega}} \nu(l) \nu^*(l).$$

The matrix H is dissipated around $\bar{H} = \sum_{l=-m}^{m} \nu(l) \nu^*(l)$. Using the result of Lemma E.1 in [18], we have

$$\|\bar{H}\| \leq 260 \pi^2 N \log K. \tag{38}$$

Using the bound on s from Theorem 1, we can write

$$s \leq C_s N \left(\left(\log \frac{N}{\epsilon} \right)^{-1} \left(1 + \frac{1}{L} \log(\frac{\sqrt{L} N^3}{\epsilon}) \right)^{-1} \leq C_s N \left(\log K \right)^{-1} \left(1 + \frac{1}{L} \log(\frac{\sqrt{L} N^3}{\epsilon}) \right)^{-1}.$$

Then from (38) and the above bound, we can bound $\|\frac{s}{N} \bar{H}\|$ as

$$\|\frac{s}{N} \bar{H}\| \leq \frac{260 \pi^2 C_s N}{1 + \frac{1}{L} \log \frac{\sqrt{L} N^3}{\epsilon}} = \frac{C_{\boldsymbol{B}}^2}{2} N (1 + \frac{1}{L} \log \frac{\sqrt{L} N^3}{\epsilon})^{-1}.$$

Now, we can control the deviation of H from $\bar{H}$ using the following Lemma.

Lemma 7. *Under the assumptions of Theorem 1*

$$\|H - \frac{s}{N} \bar{H}\| \leq \frac{C_{\boldsymbol{B}}^2}{2} N (1 + \frac{1}{L} \log \frac{\sqrt{L} N^3}{\epsilon})^{-1}$$

with probability at least $1 - \frac{\epsilon}{5}$.

The proof of the above Lemma is given after the current proof. Using the result from Lemma 7, we have

$$\|\boldsymbol{B_\Omega}\| \leq \sqrt{\|H\|} \leq \sqrt{\frac{s}{N} \|\bar{H}\| + \|H - \frac{s}{N} \bar{H}\|}$$

$$\leq \sqrt{\frac{C_{\boldsymbol{B}}^2}{2} N (1 + \frac{1}{L} \log \frac{\sqrt{L} N^3}{\epsilon})^{-1} + \frac{C_{\boldsymbol{B}}^2}{2} N (1 + \frac{1}{L} \log \frac{\sqrt{L} N^3}{\epsilon})^{-1}}$$

$$= C_{\boldsymbol{B}} \sqrt{N} (1 + \frac{1}{L} \log \frac{\sqrt{L} N^3}{\epsilon})^{-1/2} \tag{39}$$

with probability at least $1 - \frac{\epsilon}{5}$. This concludes the proof of Lemma 6. $\qquad \square$

Proof of Lemma 7. Under the assumptions of Theorem 1, one can write

$$H = \sum_{l=-m}^{m} \delta_\Omega(l) \nu(l) \nu^*(l),$$

where $\delta_\Omega(l), l = -m, ..., m$ are i.i.d. Bernoulli random variables with parameter $\frac{s}{n}$. Next we can build zero-mean self adjoint matrices from H as

$$X_l := (\delta_\Omega - \frac{s}{N}) \nu(l) \nu^*(l),$$

so that we can apply Matrix Bernstein inequality [35].

23

Theorem 2 (Matrix Bernstein inequality [35]). *Let $\{X_l\}$ be a finite sequence of independent zero-mean self-adjoint random matrices of dimension d such that $\|X_l\| \leq B$ almost surely for a certain constant B. For all $t \geq 0$ and a positive constant σ^2*

$$\mathbb{P}\{\|\sum_{l=-m}^{m} X_l\| \geq t\} \leq d e^{-\frac{t^2/2}{\sigma^2 + Bt/3}} \tag{40}$$

for $\|\sum_{l=-m}^{m} \mathbb{E}(X_l^2)\| \leq \sigma^2$.

In order to be able to apply the recent theorem on X_l, we need a bound on $\|X_l\|$. Using Lemma 3.5 in [18], we have

$$\|X_l\| \leq \sup_{-m \leq l \leq m} \|\nu(l)\|_2^2 \leq B := 10K.$$

Also, to find the value for σ^2, we can write

$$\sigma^2 := \|\sum_{l=-m}^{m} \mathbb{E}(X_l^2)\| = \|\sum_{l=-m}^{m} \mathbb{E}((\bar{\delta}(l) - \frac{s}{N}))^2 \|\nu(l)\|_2^2 \nu(l)\nu^*(l)\|$$

$$\leq 10K \frac{s}{N}\|\bar{H}\| \leq 10C_{\boldsymbol{B}}^2 NK(\log\frac{N}{\epsilon}(1 + \frac{1}{L}\log\frac{\sqrt{L}N^3}{\epsilon}))^{-1}.$$

Thus, if we set $t := \frac{C_{\boldsymbol{B}}^2 N}{2}(\log\frac{N}{\epsilon}(1 + \frac{1}{L}\log\frac{\sqrt{L}N^3}{\epsilon}))^{-1}$, we can take $\sigma^2 = 20Kt$ in Theorem 2. This will yield

$$\mathbb{P}\{\|H - \frac{s}{n}\bar{H}\| \geq t\} \leq 2K e^{\frac{-t^2/2}{\sigma^2 + Bt/3}} = 2K e^{\frac{-3t}{140K}}.$$

The above inequality will lead to the conclusion that to get the maximum probability of occurrence of $\epsilon/5$, we should have

$$K \leq \frac{3C_{\boldsymbol{B}}^2 N}{280}(\log\frac{10K}{\epsilon}(1 + \frac{1}{L}\log\frac{N^3\sqrt{L}}{\epsilon}))^{-1},$$

which is satisfied by the bound on K in Theorem 1 if we set C_K small enough. $\square$

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

$$\varepsilon_D := \left\{\|D - \frac{N-s}{N}\bar{D}\| \geq \frac{N-s}{4N}C_D\left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-\frac{1}{2}}\right\}$$

occurs with probability at most $\epsilon/5$. Also, in the complement event ε_D^c, D^{-1} exists and

$$\|D^{-1}\| \leq 8, \|D^{-1} - \frac{N}{N-s}\bar{D}^{-1}\| \leq C_D\left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-\frac{1}{2}}$$

where C_D is a constant.

24

Proof. In order to prove, we need lemmas G.1 and G.3 from [18] which both hold in our setting. Next, using Lemma G.1 and triangle inequality, we can bound the smallest singular value of D by a positive number which ensures invertibility of D. Then, setting $A = D, B = p\bar{D}$ in Lemma G.3 we get $\|D^{-1}\| \leq 8$ (note that $N/(N-s) \leq 2$ with $s \leq N/2$). Next, define

$$X_l := (p - \delta_{\Omega^c}(l))c_l\nu(l)\nu^*(l)$$

for any $-m \leq l \leq m$ with $p = (N-s)/N$. Note that $\bar{D} = \sum_{l=-m}^{m} c_l\nu(l)\nu^*(l)$ and $D = \sum_{l=-m}^{m} \delta_{\Omega^c}(l)c_l\nu(l)\nu^*(l)$. Thus, $\mathbb{E}(X_l) = 0$. Using the same calculation in [18](Lemmas 3.4, 3.5), $\|X_l\| \leq B := \frac{12.6K}{m}$. Also,

$$\mathbb{E}(X_l^2) = p(1-p)c_l^2\|\nu(l)\|_2^2\nu(l)\nu^*(l)$$

which leads to $\sum_{l=-m}^{m} \mathbb{E}(X_l^2) \leq \sigma^2 := \frac{18.5pk}{m}$ as in [18]. Now, set $t = \frac{p}{4}C'_D(1 + \frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})^{-\frac{1}{2}}$ with $C'_D = \min\{1, C_D/4\}$ in Theorem 2 with B and σ^2 as defined above. Then,

$$\frac{t^2/2}{\sigma^2 + Bt/3} = \frac{pm}{32K}C'^2_D \left[18.5(1 + \frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon}) + 1.05C'_D(1 + \frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})^{\frac{1}{2}}\right]^{-1}$$

$$> \frac{pm}{32K}C'_D \left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-1} [18.05 + 1.05C'_D]^{-1}$$

$$= \frac{N-s}{K}C''_D \left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-1}$$

Theorem 2 implies that

$$\mathbb{P}\{\|D - p\bar{D}\| > t\} \leq 2Ke^{-\frac{N-s}{K}C''_D\left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-1}}$$

This probability is smaller than $\epsilon/5$ as long as

$$K < NC''_D \left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-1} \left(\log\frac{10K}{\epsilon}\right)^{-1}, s < N/2$$

which is the case by assumptions in Theorem 1 if C_K and C_s are small enough. This concludes the proof. $\square$

Lemma 9. *Consider the equispaced grid $\mathcal{G} \subset [0,1]$ with cardinality $|\mathbb{T}_{grid}| = 200\sqrt{L}N^3$. Then, the event*

$$\varepsilon_G := \left\{\|G_\iota(f) - \frac{N-s}{N}\bar{G}_\iota(f)\|_2 > C_G \left(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon})\right)^{-\frac{1}{2}}\right\}$$

for any $f \in \mathcal{G}$, $\iota \in \{0, 1, 2, 3\}$ and constant C_G, has probability bounded by $\epsilon/5$.

Proof. We need vector Bernstein inequality to prove this lemma.

Theorem 3 (Vector Bernstein inequality [36]). *Let $\mathcal{P} \subset \mathbb{R}^d$ be a finite sequence of independent zero-mean random vectors with $\|\boldsymbol{p}\|_2 \leq B$ a.s. and $\sum_{\boldsymbol{p} \in \mathcal{P}} \mathbb{E}\|\boldsymbol{p}\|_2^2 \leq \sigma^2$ for all $\boldsymbol{p} \in \mathcal{P}$, where B and σ^2 are both positive constants. Then,*

$$\mathbb{P}\left\{\|\sum_{\boldsymbol{p} \in \mathcal{P}} \boldsymbol{p}\|_2 \geq t\right\} \leq e^{-\frac{t^2}{8\sigma^2} + \frac{1}{4}}$$

for $0 \leq t \leq \frac{\sigma^2}{B}$.

Using the definition of K and $\bar{K}$ we can rewrite $\boldsymbol{G}_\iota(f)$ and $\bar{\boldsymbol{G}}_\iota(f)$ as

$$\boldsymbol{G}_\iota(f) = \sum_{l=-m}^{m} \delta_{\Omega^c}(l)(i2\pi\kappa l)^\iota c_l e^{i2\pi lf}\boldsymbol{\nu}(l), \bar{G}_\iota(f) = \sum_{l=-m}^{m} (i2\pi\kappa l)^\iota c_l e^{i2\pi lf}\boldsymbol{\nu}(l).$$

Note that by defining

$$\boldsymbol{p}(\iota, l) := (\delta_{\Omega^c}(l) - p)(i2\pi\kappa l)^\iota c_l e^{i2\pi lf}\boldsymbol{\nu}(l),$$

where $p = \frac{N-s}{N}$ (parameter of i.i.d Bernoulli random variables $\delta_{\Omega^c}(-m), ..., \delta_{\Omega^c}(m)$) we have $\boldsymbol{G}_\iota(f) - p\bar{\boldsymbol{G}}_\iota(f) = \sum_{l=-m}^{m} \boldsymbol{p}(l)$. Also using lemmas 3.3, 3.4, 3.5 from [18], leads to

$$\|\boldsymbol{p}(\iota, l)\|_2 \leq B := \frac{128\sqrt{K}}{m}, \quad \sum_{l=-m}^{m} \mathbb{E}(\|\boldsymbol{p}(\iota, l)\|_2^2) \leq \sigma^2 := \frac{3.2510^4 K}{m}.$$

Now, using Theorem 3 we have

$$\mathbb{P}\left\{\sup_{f \in \mathcal{G}} \|\boldsymbol{G}_\iota(f) - p\bar{\boldsymbol{G}}_\iota(f)\|_2 \geq t, \quad \iota = \{0, 1, 2, 3\}\right\} \leq 4|\mathbb{T}_{grid}|e^{-\frac{t^2}{8\sigma^2} + \frac{1}{4}}$$

To make the r.h.s smaller than $\epsilon/5$, take

$$t := \sqrt{\frac{26 \times 10^4 K}{m}\left(\frac{1}{4} + \log(\frac{20|\mathbb{T}_{grid}|}{\epsilon})\right)}$$

This choice is valid since

$$\frac{t}{\sigma} = \sqrt{8(\frac{1}{4} + \log(\frac{20 * |\mathbb{T}_{grid}|}{\epsilon}))} \leq \sqrt{74 + 24\log(N) + 4\log(L) + 8\log(\frac{1}{\epsilon})}$$

$$\leq \sqrt{74 + 44\log(N)} + \sqrt{8\log(\frac{1}{\epsilon})} \leq 0.452\sqrt{N}\sqrt{8\log(\frac{1}{\epsilon})} \leq 0.46\sqrt{N}$$

where we have used $\sqrt{74 + 44\log(N)} \leq 0.452\sqrt{N}$ and assumed $N \geq 2 \times 10^3$, $L \leq N^5$ and either $K \geq 1$ or $s \geq 1$. Thus, $t/\sigma \leq 0.46\sqrt{N} \leq \sigma/B$. The desired result holds as long as

$$C_G(1 + \frac{1}{L}\log(\frac{|\mathbb{T}_{grid}|}{\epsilon}))^{-\frac{1}{2}} \geq t \geq \sqrt{\frac{2 \times 10^3 K}{N}\left(\frac{1}{4} + \log(\frac{8 \times 10^3 \sqrt{L}N^3}{\epsilon})\right)}$$

with C_K small enough. $\qquad\square$

Using Lemmas 8, H.8 and Corollary H.9 from [18] with respect to the new bounds for s and K and the first condition of Theorem 1 combined with Lemma 6, we find tight bounds for (37) as

$$\frac{1}{\sqrt{N}}\|\boldsymbol{B_\Omega}\|\|\boldsymbol{D}^{-1}\|\|\boldsymbol{G_\iota}(f)\|_2 \leq \frac{8(C_{\bar{\boldsymbol{\nu}}}+C_{\boldsymbol{\nu}})\|\boldsymbol{B_\Omega}\|}{\sqrt{N}}$$

$$\leq \frac{8(C_{\bar{\boldsymbol{\nu}}}+C_{\boldsymbol{\nu}})C_{\boldsymbol{B}}\left(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon}\right)^{-\frac{1}{2}}\sqrt{N}}{\sqrt{N}} \quad , C_{\boldsymbol{B}} = \frac{C_{\mathcal{U}}}{8\left(C_{\bar{\boldsymbol{\nu}}}+C_{\boldsymbol{\nu}}\right)}.$$

Thus, by setting $t = \frac{10^{-2}}{8}$ and using Lemma 5 and the union bound, we obtain

$$\mathbb{P}\{\sup_{f\in\mathcal{U}}\|\frac{1}{\sqrt{N}}\boldsymbol{G}_\iota^T(f)\boldsymbol{D}^{-1}\boldsymbol{B_\Omega}\|_2 \geq \frac{10^{-2}}{8}\}$$

$$\leq |\mathbb{T}_{grid}|e^{-L\left(\frac{\frac{10^{-2}}{8}^2}{C_{\mathcal{U}}^2(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})-1}-\log\frac{\frac{10^{-2}}{8}^2}{C_{\mathcal{U}}^2(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})-1}-1\right)}. \tag{41}$$

With the same reasoning for ε_1 and small enough $C_{\mathcal{U}}$, the event

$$\varepsilon_2 := \{\|\frac{1}{\sqrt{N}}\boldsymbol{G}_\iota^T(f)\boldsymbol{D}^{-1}\boldsymbol{B_\Omega}\|_2 \geq \frac{10^{-2}}{8} \quad for\ all\ f\in\mathcal{U}\}$$

holds with probability at most $\epsilon/20$ under the assumptions of Proposition1.

9.1.3. Bound on ε_3

For the third term, we consider $\boldsymbol{\Psi} = \boldsymbol{\Phi}$ and

$$\boldsymbol{\omega} = \boldsymbol{P}\boldsymbol{D}^{-1}\left(\boldsymbol{G_\iota}(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)\right),$$

where $\boldsymbol{P} \in \mathbb{R}^{K\times 2K}$ is a projection matrix, which selects the first K elements in a vector and $\|\boldsymbol{P}\| = 1$. According to Lemmas 8 and 9 we can write

$$\|\boldsymbol{P}\boldsymbol{D}^{-1}\left(\boldsymbol{G_\iota}(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)\right)\|_2 \leq \|\boldsymbol{P}\|\|\boldsymbol{D}^{-1}\|\|\boldsymbol{G_\iota}(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)\|_2$$

$$\leq 8\|\boldsymbol{G_\iota}(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)\|_2 \leq C_{\mathcal{U}}\left(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon}\right)^{-\frac{1}{2}}.$$

By setting $t = \frac{10^{-2}}{8}$ and applying (32) and the union bound, we have

$$\mathbb{P}\{\sup_{f\in\mathcal{U}}\|(\boldsymbol{G_\iota}(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f))^T\boldsymbol{D}^{-1}\begin{bmatrix}\boldsymbol{\Phi}\\\boldsymbol{0}\end{bmatrix}\|_2 \geq \frac{10^{-2}}{8}\}$$

$$\leq |\mathbb{T}_{grid}|e^{-L\left(\frac{\frac{10^{-2}}{8}^2}{C_{\mathcal{U}}^2(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})-1}-\log\frac{\frac{10^{-2}}{8}^2}{C_{\mathcal{U}}^2(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})-1}-1\right)}. \tag{42}$$

27

Therefore, with the same reasoning for ε_1 and ε_2, the event

$$\varepsilon_3 := \{\|(\boldsymbol{G}_\iota(f) - \frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f))^T \boldsymbol{D}^{-1} \begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix}\|_2 \geq \frac{10^{-2}}{8} \ \ for \ all \ f \in \mathcal{U}\}$$

holds with probability at most $\epsilon/20$ under the assumptions of Proposition1.

9.1.4. Bound on ε_4

At last, one can bound the fourth term in (31) by considering $\boldsymbol{\Psi} = \boldsymbol{\Phi}$ and

$$\boldsymbol{\omega} = \frac{N-s}{N}\boldsymbol{P}\left(\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1}\right)\bar{\boldsymbol{G}}_\iota(f).$$

Using the Lemmas 8,9 we get

$$\|\frac{N-s}{N}\boldsymbol{P}\left(\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1}\right)\bar{\boldsymbol{G}}_\iota(f)\|_2 \leq$$

$$\|\boldsymbol{P}\|\|\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1}\|\|\bar{\boldsymbol{G}}_\iota(f)\|_2 \leq$$

$$C_{\bar{\boldsymbol{\nu}}}\|\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1}\| \leq C_\mathcal{U}\left(1 + \frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon}\right)^{-\frac{1}{2}}.$$

By applying (32) and the union bound and setting $t = \frac{10^{-2}}{8}$, one can write

$$\mathbb{P}\{\sup_{f\in\mathcal{U}}\|\frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)^T(\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1})\begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix}\|_2 \geq \frac{10^{-2}}{8}\}$$

$$\leq |\mathbb{T}_{grid}|e^{-L\left(\frac{\frac{10^{-2}}{8}^2}{C_\mathcal{U}^2(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})^{-1}} - \log\frac{\frac{10^{-2}}{8}^2}{C_\mathcal{U}^2(1+\frac{1}{L}\log\frac{|\mathbb{T}_{grid}|}{\epsilon})^{-1}} - 1\right)}. \tag{43}$$

Therefore, with the same reasoning for ε_1, ε_2, and ε_3 the event

$$\varepsilon_4 := \{\|\frac{N-s}{N}\bar{\boldsymbol{G}}_\iota(f)^T(\boldsymbol{D}^{-1} - \frac{N}{N-s}\bar{\boldsymbol{D}}^{-1})\begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{0} \end{bmatrix}\|_2 \geq \frac{10^{-2}}{8}$$
$$for \ all \ f \in \mathcal{U}\}$$

holds with probability at most $\epsilon/20$ under the assumptions of Proposition1. Thus, using (33),(41),(42),(43), and the triangle inequality, we conclude that

$$\sup_{f\in\mathcal{U}}\|\kappa^\iota\boldsymbol{Q}^{(\iota)}(f) - \kappa^\iota\bar{\boldsymbol{Q}}^{(\iota)}(f)\|_2 \leq \frac{10^{-2}}{2} \tag{44}$$

holds with probability at least $1 - \epsilon/5$ under the assumptions of Proposition1. Next, using Bernstein polynomial inequality[37], we extend the results to the continuous domain $[0,1]$. Considering $f \in [0,1]$ and $f_g \in \mathcal{G}$, we have

$$\|\kappa^\iota\boldsymbol{Q}^{(\iota)}(f) - \kappa^\iota\bar{\boldsymbol{Q}}^{(\iota)}(f)\|_2 \leq \|\kappa^\iota\boldsymbol{Q}^{(\iota)}(f_g) - \kappa^\iota\boldsymbol{Q}^{(\iota)}(f)\|_2$$
$$+ \|\kappa^\iota\bar{\boldsymbol{Q}}^{(\iota)}(f_g) - \kappa^\iota\boldsymbol{Q}^{(\iota)}(f_g)\|_2 + \|\kappa^\iota\bar{\boldsymbol{Q}}^{(\iota)}(f) - \kappa^\iota\bar{\boldsymbol{Q}}^{(\iota)}(f_g)\|_2.$$

Then, consider the third term in the right side of the above inequality. We had $\bar{\boldsymbol{Q}}^{(\iota)}(f) \in \mathbb{C}^{1\times L}$ and for any $\boldsymbol{v} \in \mathbb{C}^{1\times L}, \|\boldsymbol{v}\| \leq \sqrt{L}\|\boldsymbol{v}\|_\infty$. The jth entry of $\bar{\boldsymbol{Q}}^{(\iota)}(f)$ is

$$|\kappa^\iota \bar{\boldsymbol{Q}}_j^{(\iota)}(f)| \leq | < \bar{\boldsymbol{D}}^{-1}\bar{\boldsymbol{G}}_\iota(f), \boldsymbol{\Phi}_{:,j} > | \leq 8\sqrt{K}\left(256\sqrt{K}\right) = CK \leq CN^2.$$

Next, take $\kappa^\iota \bar{\boldsymbol{Q}}_j^{(\iota)}(f)$ as a polynomial of $z = e^{-i2\pi f}$ with degree m and apply the Bernstein polynomial inequality as

$$|\kappa^\iota \bar{\boldsymbol{Q}}_j^{(\iota)}(f) - \kappa^\iota \bar{\boldsymbol{Q}}_j^{(\iota)}(f_g)| \leq |e^{-i2\pi f} - e^{-i2\pi f_g}|\sup_z \left|\frac{d\kappa^\iota \bar{\boldsymbol{Q}}_j^{(\iota)}(z)}{dz}\right|$$

$$\leq |e^{-i\pi(f+f_g)}2sin(\pi(-f+f_g))|m\sup_f|\kappa^\iota \bar{\boldsymbol{Q}}_j^{(\iota)}(f)| \leq CN^3|f-f_g|.$$

Thus,

$$\|\kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f) - \kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f_g)\|_2 \leq \sqrt{L}\|\kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f) - \kappa^\iota \bar{\boldsymbol{Q}}^{(\iota)}(f_g)\|_\infty \leq C\sqrt{L}N^3|f-f_g|.$$

The above calculations reveal that the grid size $|\mathbb{T}_{grid}| = 1/|f-f_g|$ should be such that $|f-f_g| \leq \frac{10^{-2}}{4C\sqrt{L}N^3}$. Using the same arguments, one can obtain the same bound for $\|\kappa^\iota \boldsymbol{Q}^{(\iota)}(f) - \kappa^\iota \boldsymbol{Q}^{(\iota)}(f_g)\|_2$. Combining the above results with (44) proves the lemma.

9.2. Proof of Lemma4

Consider $A_{far} = [0,1]\backslash A_{near}$, where A_{near} is defined in Lemma 4. We prove that $\|\bar{\boldsymbol{Q}}(f)\|_2 < 0.99$ in A_{far}. Next, it is shown that $\|\boldsymbol{Q}(f)\|_2 < 1$ in A_{near}. For $\|\bar{\boldsymbol{Q}}(f)\|_2$, we write

$$\|\bar{\boldsymbol{Q}}(f)\|_2 \leq \sum_{f_k \in \mathbb{T}} \|\boldsymbol{\alpha}_k\|_2 |\bar{K}(f-f_k)| + \sum_{f_k \in \mathbb{T}} \kappa\|\boldsymbol{\beta}_k\|_2|\bar{K}'(f-f_k)|$$

$$\leq \|\boldsymbol{\alpha}\|_{\infty,2} \sum_{f_k \in \mathbb{T}} |\bar{K}(f-f_k)| + \|\boldsymbol{\beta}\|_{\infty,2} \sum_{f_k \in \mathbb{T}} \kappa|\bar{K}'(f-f_k)|.$$

Using Lemma H.10 from [18], we have $\sum_{j=1}^{K} \kappa^\iota|\bar{K}^{(\iota)}(f-f_j)| \leq 127C_1 + 2.42C_2$ for some properly chosen C_1 and C_2. Thus,

$$\|\bar{\boldsymbol{Q}}(f)\|_2 \leq (\|\boldsymbol{\alpha}\|_{\infty,2} + \|\boldsymbol{\beta}\|_{\infty,2})(127C_1 + 2.42C_2).$$

In the following, we calculate the upper bounds for $\|\boldsymbol{\alpha}\|_{\infty,2}$ and $\|\boldsymbol{\beta}\|_{\infty,2}$. Recall (23) for the deterministic case. Using this equation, we have

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \bar{\boldsymbol{D}}_2^{-1}\bar{\boldsymbol{D}}_1 \end{bmatrix} \bar{\boldsymbol{D}}_3^{-1}\boldsymbol{\Phi},$$

where $\bar{\boldsymbol{D}}_3 \triangleq \bar{\boldsymbol{D}}_0 + \bar{\boldsymbol{D}}_1\bar{\boldsymbol{D}}_2^{-1}\bar{\boldsymbol{D}}_1$. According to Lemma 4.1 from [27] and the fact that $\|\boldsymbol{\Phi}\|_{\infty,2} = 1$, we have$\|\boldsymbol{\alpha}\|_{\infty,2} = \|\bar{\boldsymbol{D}}_3^{-1}\boldsymbol{\Phi}\|_{\infty,2} \leq 1 + 2.37 \times 10^{-2}$ and

$\|\boldsymbol{\beta}\|_{\infty,2} \leq \|\bar{\boldsymbol{D}}_2^{-1}\bar{\boldsymbol{D}}_1\bar{\boldsymbol{D}}_3^{-1}\boldsymbol{\Phi}\|_{\infty,2} \leq \frac{4.247}{m} \times 10^{-2}$. Therefore, by the proper choices of C_1 and C_2 we get

$$\|\bar{\boldsymbol{Q}}(f)\|_2 < 0.99 \quad for \ f \in A_{far}.$$

In order to show that $\|\boldsymbol{Q}(f) < 1\|_2$ in A_{near}, it is enough to show that the second derivative of $\|\boldsymbol{Q}(f) < 1\|_2$ is negative in A_{near}. In a mathematical fashion, it is enough to prove the following inequality,

$$\frac{1}{2}\frac{d^2\|\boldsymbol{Q}(f)\|_2}{df^2} = \|\boldsymbol{Q}'\|_2^2 + Re\{\boldsymbol{Q}''\boldsymbol{Q}^H(d)\} < 0. \tag{45}$$

Now, we investigate each term in the above inequality. For the first term, we can write

$$\|\kappa\boldsymbol{Q}'(f)\|_2^2 = \|\kappa\boldsymbol{Q}'(f) - \kappa\bar{\boldsymbol{Q}}'(f) + \kappa\bar{\boldsymbol{Q}}'(f)\|_2^2$$
$$\leq 10^{-4} + 2 \times 10^{-2}\|\kappa\bar{\boldsymbol{Q}}'(f)\|_2 + \|\kappa\bar{\boldsymbol{Q}}'(f)\|_2^2,$$

which by applying the kernel bounds of [27] leads to

$$\|\kappa\bar{\boldsymbol{Q}}'(f)\|_2 \leq \|\boldsymbol{\alpha}\|_{\infty,2}\sum_{k=1}^{K}\kappa|\bar{K}'(f-f_k)| + \|\boldsymbol{\beta}\|_{\infty,2}\sum_{k=1}^{K}\kappa^2|\bar{K}''(f-f_k)|$$

$$\leq 1.0237 \times 2.409 \times 10^{-2} + \frac{4.247 \times 10^{-2}}{m}(0.087) \leq 0.0247,$$

where the last inequality is achieved using $m \geq 10^3$. The second term of (45) can be represented as

$$Re\left\{\kappa^2\boldsymbol{Q}''(f)\boldsymbol{Q}^H(f)\right\} = Re\left\{\kappa^2(\boldsymbol{Q}''(f) - \bar{\boldsymbol{Q}}''(f))\boldsymbol{Q}^H(f)\right\}$$
$$+ Re\left\{\kappa^2\bar{\boldsymbol{Q}}''(f)(\boldsymbol{Q}(f) - \bar{\boldsymbol{Q}}(f))^H\right\} + Re\left\{\kappa^2\bar{\boldsymbol{Q}}''(f)\bar{\boldsymbol{Q}}^H(f)\right\}$$
$$\leq 0.0101 + 0.01 + Re\left\{\kappa^2\bar{\boldsymbol{Q}}''(f)\bar{\boldsymbol{Q}}^H(f)\right\}.$$

Next, we inspect the term $\kappa^2\bar{\boldsymbol{Q}}''(f)\bar{\boldsymbol{Q}}^H(f)$. According to (26), we get

$$\kappa^2\bar{\boldsymbol{Q}}''(f)\bar{\boldsymbol{Q}}^H(f) = \kappa^2\bar{Q}''(f)\boldsymbol{b}^H\boldsymbol{b}\bar{Q}^*(f) = \kappa^2\bar{Q}''(f)\bar{Q}^*(f),$$

which is a scalar value. Also, note that

$$\kappa^2 Re\left\{\bar{Q}''(f)\bar{Q}^*(f)\right\} = \kappa^2\left(\bar{Q}_R''(f)\bar{Q}_R(f) + |\bar{Q}_I''(f)||\bar{Q}_I|\right)$$
$$\leq (-0.8915 \times 2.015 + 0.0474 \times 2.555) \leq -1.6752.$$

Thus,

$$\frac{\kappa^2}{2}\frac{d^2\|\boldsymbol{Q}(f)\|_2}{df^2} \leq -1.6752 + 0.0201 + 12.01 \times 10^{-4} < 0$$

and the proof is complete.

References

[1] R. Heckel, M. Soltanolkotabi, Generalized Line Spectral Estimation via Convex Optimization, in: IEEE Transactions on Information Theory, 2018. `doi:10.1109/TIT.2017.2757003`.

[2] S. Sayyari, S. Daei, F. Haddadi, Blind two-dimensional super resolution in multiple input single output linear systems, arXiv preprint arXiv:2005.10882 (2020).

[3] M. J. Jirhandeh, H. Hezaveh, M. H. Kahaei, Super-resolution doa estimation for wideband signals using non-uniform linear arrays with no focusing matrix, IEEE Wireless Communications Letters 11 (3) (2021) 641–644.

[4] S. Daei, M. Kountouris, Blind goal-oriented massive access for future wireless networks, arXiv preprint arXiv:2205.07092 (2022).

[5] H. Hezaveh, I. Valiulahi, M. Kahaei, Ofdm based sparse time dispersive channel estimation with prior information, IET Communications (2020).

[6] L. Borcea, G. Papanicolaou, C. Tsogka, Imaging and time reversal in random media, Inverse Problems 18 (5) (2002) 1247–1279. `doi:10.1088/0266-5611/18/5/303`.
URL `https://doi.org/10.1088%2F0266-5611%2F18%2F5%2F303`

[7] A. Koochakzadeh, P. Pal, E. T. Ahrens, Spike localization in Zero Time of Echo (ZTE) magnetic resonance imaging, in: Conference Record of 51st Asilomar Conference on Signals, Systems and Computers, ACSSC 2017, 2018. `doi:10.1109/ACSSC.2017.8335625`.

[8] P. Stoica, R. L. Moses, et al., Spectral analysis of signals (2005).

[9] B. N. Bhaskar, G. Tang, B. Recht, Atomic norm denoising with applications to line spectral estimation, IEEE Transactions on Signal Processing 61 (23) (2013) 5987–5999.

[10] S. Razavikia, A. Amini, S. Daei, Reconstruction of binary shapes from blurred images via hankel-structured low-rank matrix recovery, IEEE Transactions on Image Processing 29 (2019) 2452–2462.

[11] I. Valiulahi, S. Daei, F. Haddadi, F. Parvaresh, Two-dimensional super-resolution via convex relaxation, IEEE Transactions on Signal Processing 67 (13) (2019) 3372–3382.

[12] F. J. Harris, On the use of windows for harmonic analysis with the discrete fourier transform, Proceedings of the IEEE 66 (1) (1978) 51–83.

[13] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE transactions on antennas and propagation 34 (3) (1986) 276–280.

[14] R. Roy, T. Kailath, Esprit-estimation of signal parameters via rotational invariance techniques, IEEE Transactions on acoustics, speech, and signal processing 37 (7) (1989) 984–995.

[15] T. K. Sarkar, O. Pereira, Using the matrix pencil method to estimate the parameters of a sum of complex exponentials, IEEE Antennas and Propagation Magazine 37 (1) (1995) 48–55.

[16] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE transactions on information theory 59 (11) (2013) 7465–7490.

[17] H. Hezave, M. Javadzadeh, M. H. Kahaei, Sparse signal reconstruction using blind super-resolution with arbitrary sampling, IEEE Signal Processing Letters 27 (2020) 615–619. `doi:10.1109/LSP.2020.2986133`.

[18] C. Fernandez-Granda, G. Tang, X. Wang, L. Zheng, Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers, Information and Inference: A Journal of the IMA 7 (1) (2017) 105–168.

[19] S. Bayat, S. Daei, Separating radar signals from impulsive noise using atomic norm minimization, IEEE Transactions on Circuits and Systems II: Express Briefs (2020).

[20] Y. Park, Y. Choo, W. Seong, Multiple snapshot grid free compressive beamforming, The Journal of the Acoustical Society of America 143 (6) (2018) 3849–3859.

[21] B. Dumitrescu, Positive trigonometric polynomials and signal processing applications, Vol. 103, Springer, 2007.

[22] Z. Yang, L. Xie, Exact joint sparse frequency recovery via optimization methods, IEEE Transactions on Signal Processing 64 (19) (2016) 5145–5157.

[23] S. Daei, A. Amini, F. Haddadi, Optimal weighted low-rank matrix recovery with subspace prior information, arXiv preprint arXiv:1809.10356 (2018).

[24] S. Daei, F. Haddadi, A. Amini, Sample complexity of total variation minimization, IEEE Signal Processing Letters 25 (8) (2018) 1151–1155.

[25] S. Daei, F. Haddadi, A. Amini, M. Lotz, On the error in phase transition computations for compressed sensing, IEEE Transactions on Information Theory 65 (10) (2019) 6620–6632.

[26] S. Daei, F. Haddadi, A. Amini, Living near the edge: A lower-bound on the phase transition of total variation minimization, IEEE Transactions on Information Theory 66 (5) (2019) 3261–3267.

[27] C. Fernandez-Granda, Super-resolution of point sources via convex programming, Information and Inference: A Journal of the IMA 5 (3) (2016) 251–303.

[28] S. Daei, F. Haddadi, A. Amini, Distribution-aware block-sparse recovery via convex optimization, IEEE Signal Processing Letters 26 (4) (2019) 528–532.

[29] S. Daei, F. Haddadi, A. Amini, Exploiting prior information in block-sparse signals, IEEE Transactions on Signal Processing 67 (19) (2019) 5093–5102.

[30] V. Chandrasekaran, B. Recht, P. A. Parrilo, A. S. Willsky, The convex geometry of linear inverse problems, Foundations of Computational mathematics 12 (6) (2012) 805–849.

[31] Z. Yang, J. Li, P. Stoica, L. Xie, Sparse methods for direction-of-arrival estimation (2017). arXiv:1609.09596.

[32] Z. Yang, L. Xie, C. Zhang, A discretization-free sparse and parametric approach for linear array signal processing, IEEE Transactions on Signal Processing 62 (19) (2014) 4959–4973. doi:10.1109/TSP.2014.2339792.

[33] P. Stoica, P. Babu, J. Li, Spice: A sparse covariance-based estimation method for array processing, IEEE Transactions on Signal Processing 59 (2) (2011) 629–638. doi:10.1109/TSP.2010.2090525.

[34] Z. Yang, J. Tang, Y. C. Eldar, L. Xie, On the sample complexity of multi-channel frequency estimation via convex optimization, IEEE Transactions on Information Theory 65 (4) (2018) 2302–2315.

[35] J. A. Tropp, User-friendly tail bounds for sums of random matrices, Foundations of Computational Mathematics 12 (2012) 389–434.

[36] E. J. Candes, Y. Plan, A probabilistic and ripless theory of compressed sensing, IEEE transactions on information theory 57 (11) (2011) 7235–7254.

[37] A. Schaeffer, Inequalities of a. markoff and s. bernstein for polynomials and related functions, Bulletin of the American Mathematical Society 47 (8) (1941) 565–579.