

Approximate Support Recovery of Atomic Line Spectral Estimation: A Tale of Resolution and Precision

Qiuwei Li Gongguo Tang

Department of Electrical Engineering, Colorado School of Mines, CO, USA

Abstract

This work investigates the parameter estimation performance of super-resolution line spectral estimation using atomic norm minimization. The focus is on analyzing the algorithm's accuracy of inferring the frequencies and complex magnitudes from noisy observations. When the Signal-to-Noise Ratio is reasonably high and the true frequencies are separated by $O(\frac{1}{n})$, the atomic norm estimator is shown to localize the correct number of frequencies, each within a neighborhood of size $O(\sqrt{\log n/n^3\sigma})$ of one of the true frequencies. Here n is half the number of temporal samples and σ^2 is the Gaussian noise variance. The analysis is based on a primal-dual witness construction procedure. The obtained error bound matches the Cramér-Rao lower bound up to a logarithmic factor. The relationship between resolution (separation of frequencies) and precision or accuracy of the estimator is highlighted. Our analysis also reveals that the atomic norm minimization can be viewed as a convex way to solve a ℓ_1 -norm regularized, nonlinear and nonconvex least-squares problem to global optimality.

Keywords: atomic norm, line spectral estimation, primal-dual witness construction, super-resolution, support recovery

1 Introduction

Line spectral estimation, which aims at approximately inferring the frequency and coefficient parameters from a superposition of complex sinusoids embedded in white noise, is one of the fundamental problems in statistical signal processing. When the temporal and frequency domains are exchanged, this classical problem was reinterpreted as the problem of mathematical super-resolution recently [1–3]. This line of work promotes the use of a convex sparse regularizer to solve inverse problems involving spectrally sparse signals, distinguishing them from classical methods based on root finding and singular value decompositions (e.g., Prony's method, MUSIC, ESPRIT, Matrix Pencil, etc.). The convex regularizer, a particular instance of the general atomic norms, has been shown to achieve optimal performance in signal completion [4], denoising [5], and outlier removal [6, 7]. For these signal processing tasks, either one can recover the spectral signal exactly (and hence extract the true frequencies precisely), or the error metric is defined using the signal instead of the frequency parameters. The most relevant question of the accuracy of noisy frequency estimation has been elusive. This work investigates the parameter estimation performance of super-resolution line spectral estimation using atomic norm minimization. More precisely, given noisy observations

$$y(t) = x^*(t) + w(t), t = -n, \dots, n \quad (1.1)$$

of a spectrally sparse signal

$$x^*(t) = \sum_{\ell=1}^k c_\ell^* \exp(i2\pi f_\ell^* t), t = -n, \dots, n \quad (1.2)$$

Email addresses: qiuwei@mines.edu (Qiuwei Li), gtang@mines.edu (Gongguo Tang)
 This work was supported by the National Science Foundation [CCF-1464205, CCF-1704204].

with unknown frequencies $T^* = \{f_\ell^*\}_{\ell=1}^k$ and complex amplitudes $\{c_\ell^*\}_{\ell=1}^k$, we will derive conditions under which the atomic norm formulation will return the correct number of frequencies, and establish bounds on the frequency and coefficient estimation errors. An informal version of our main result is given in the following theorem, while a formal statement is presented in Theorem 2.1.

Theorem 1.1 (Informal). *Suppose we observe $2n+1$ noisy consecutive samples $y(t) = x^*(t) + w(t)$ of the signal (1.2) with $w(t)$ being i.i.d. complex Gaussian valuables of mean zero and variance σ^2 . If the unknown frequencies are well-separated, the Signal-to-Noise Ratio (SNR) is large, and the dynamic range of the coefficients is small, then with probability at least $1 - \frac{1}{n^2}$, solving an atomic norm regularized least-squares problem with a large enough regularization parameter will return exactly k estimated frequencies $\{f_\ell^{\text{glob}}\}_{\ell=1}^k$ and coefficients $\{c_\ell^{\text{glob}}\}_{\ell=1}^k$ that, when properly ordered, satisfy*

$$\max_{1 \leq \ell \leq k} |c_\ell^*| |f_\ell^{\text{glob}} - f_\ell^*| = O\left(\frac{\sqrt{\log n}}{n^{3/2}} \sigma\right), \quad (1.3)$$

$$\max_{1 \leq \ell \leq k} |c_\ell^{\text{glob}} - c_\ell^*| = O\left(\sqrt{\frac{\log n}{n}} \sigma\right). \quad (1.4)$$

We would like to first point out that this *frequency estimator* $\{f_i^{\text{glob}}\}$ given by the atomic norm regularized least-squares is asymptotically unbiased. The ℓ_1 norm minimization (atomic norm minimization is extension of it) is usually considered biased because it pushes down the solution using the ℓ_1 norm. In the context of atomic norm minimization, the estimator for the coefficient vector is indeed biased for the same reason. However, the frequency estimator, which is of more interest, might still be unbiased since it is not pushed down by the atomic norm formulation. Indeed, our result shows that the frequency estimator is at least asymptotically unbiased.

Corollary 1.1. *Under the same setup as in Theorem 1.1, with probability at least $1 - \frac{1}{n^2}$, the frequency estimator obtained by the atomic norm regularized minimization is asymptotic unbiased.*

Proof. To see this, we note that

$$\begin{aligned} \mathbb{E}[f_i^{\text{glob}}] - f_i^* &\leq \mathbb{E}\{|f_i^{\text{glob}} - f_i^*|\} = \int_{\Omega} |f_i^{\text{glob}}(\omega) - f_i^*(\omega)| d\omega + \int_{\Omega^c} |f_i^{\text{glob}}(\omega) - f_i^*(\omega)| d\omega \\ &\leq O\left(\frac{\sqrt{\log n}}{c_{\min}^* n^{3/2}} \sigma\right) + \frac{2}{n^2} \\ &= o\left(\frac{1}{n}\right). \end{aligned}$$

Here Ω is the high-probability sample space where our main result (1.3) holds, Ω^c is its complement space, and c_{\min}^* is defined as the smallest magnitude of $\{c_i^*\}$. The second inequality follows from Eq. (1.3), $\int_{\Omega} d\omega \leq 1$, $\int_{\Omega^c} d\omega \leq \frac{1}{n^2}$, and the fact that any frequency is defined in $\mathbb{T} = [0, 1]$. Therefore, the frequency estimator is at least asymptotically unbiased. \square

By the asymptotic unbiasedness of our *atomic frequency estimator* and considering that the Cramér-Rao bound (CRB) [8] can be viewed as the best squared error bound for any unbiased frequency estimators, we now compare our main result (1.3) (after taking the square) with the CRB, as well as the two most famous classical line spectral estimation methods, i.e., the MUSIC and Maximum Likelihood Estimation (MLE), in Table 1. We conclude that the squared error bound of the *atomic frequency estimator* matches the CRB up to a logarithmic factor. We also note that the MUSIC and the MLE only have asymptotic mean squared error in the sense that the number of snapshots T has to be infinitely large [8]. We emphasize that our results are non-asymptotic, which hold for finite-length, single-snapshot signals (i.e., $T = 1$), while classical methods such as MUSIC and MLE are not efficient (i.e., approaching CRB) even with an infinite number of snapshots, as long as the signal length n is finite.

Method	Squared-Error Bound
CRB [8]	$O(\frac{\sigma^2}{c_{\min}^{\star 2} n^3})$
MUSIC [8]	$O(\frac{\sigma^2}{T c_{\min}^{\star 2} n^3} + \frac{\sigma^4}{T c_{\min}^{\star 4} n^4})$
MLE [8]	$O(\frac{\sigma^2}{T c_{\min}^{\star 2} n^3} + \frac{\sigma^4}{T c_{\min}^{\star 4} n^4})$
This work (1.3)	$O(\frac{\sigma^2 \log n}{c_{\min}^{\star 2} n^3})$

Table 1: Comparison with the classical line spectral estimation methods.

2 Signal Model and Atomic Norm Regularization

This paper considers the spectral estimation problem: given noisy temporal samples, how well can we estimate the locations and determine the magnitudes of spectral lines? The signal of interest $x^\star(t)$ as expressed in (1.2) is composed of only a small number of spectral spikes located in a normalized interval $\mathbb{T} = [0, 1]$. We abuse notation and call $T^\star = \{f_\ell^\star\}_{\ell=1}^k$ the support of \mathbf{x}^\star . The number of frequencies, k , is referred to as the model order. The goal is to approximately localize these parameters from a small number $2n + 1$ of equispaced noisy samples given in (1.1). For technical simplicity, we assume $n = 2M$ is an even number. The noise components $w(t)$ are i.i.d. centrally symmetric complex Gaussian variables with variance σ^2 . To simplify notation, we stack the temporal samples into vectors and write the observation model as

$$\mathbf{y} = \mathbf{x}^\star + \mathbf{w}, \quad (2.1)$$

where $\mathbf{x}^\star := [x^\star(-n), \dots, x^\star(n)]^T$, $\mathbf{y} := [y(-n), \dots, y(n)]^T$ and $\mathbf{w}^\star := [w^\star(-n), \dots, w^\star(n)]^T$.

To estimate the frequency vector $\mathbf{f}^\star := [f_1^\star, \dots, f_k^\star]^T$ and the complex coefficient vector $\mathbf{c}^\star := [c_1^\star, \dots, c_k^\star]^T$, we assume k is small and treat \mathbf{x}^\star as a sparse combination of atoms $\mathbf{a}(f) := [e^{i2\pi(-n)f}, \dots, e^{i2\pi nf}]^T$ parameterized by frequency $f \in \mathbb{T}$, that is,

$$\mathbf{x}^\star = \sum_{\ell=1}^k c_\ell^\star \mathbf{a}(f_\ell^\star). \quad (2.2)$$

To exploit the structure of \mathbf{x}^\star encoded in the set of atoms $\mathcal{A} := \{\mathbf{a}(f), f \in \mathbb{T}\}$, we follow [4, 9] and define the associated atomic norm as

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf \left\{ \sum_{\ell} |c_\ell| : \mathbf{x} = \sum_{\ell} c_\ell \mathbf{a}(f_\ell), \forall f_\ell \in \mathbb{T}, c_\ell \in \mathbb{C} \right\}. \quad (2.3)$$

The dual norm of the atomic norm, which is useful both algorithmically and theoretically, is defined for any vector \mathbf{z} as $\|\mathbf{z}\|_{\mathcal{A}}^\star = \sup_{f \in \mathbb{T}} |\mathbf{a}(f)^H \mathbf{z}|$, where H denotes the Hermitian (conjugate transpose) operation. To solve atomic norm minimizations numerically, the authors of [5, 10] (see also [1]) first proposed to reformulate the atomic norm (2.3) as an equivalent semidefinite program. Other numerical schemes are studied in [11–14].

Given the noisy observation model (2.1), it is natural to denoise \mathbf{x}^\star by solving the atomic norm regularized minimization program [5, 15]:

$$\mathbf{x}^{\text{glob}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{x}\|_{\mathcal{A}}. \quad (2.4)$$

For technical reasons, we used a weighted ℓ_2 norm, $\|\mathbf{z}\|_{\mathbf{Z}} := \sqrt{\mathbf{z}^H \mathbf{Z} \mathbf{z}}$, to measure data fidelity. Here $\mathbf{Z} = \operatorname{diag}(\frac{g_M(\ell)}{M}) \in \mathbb{R}^{(4M+1) \times (4M+1)}$ with $g_M(\ell), \ell = -2M, \dots, 2M$ defined in [4] as the discrete convolution of two triangular functions. We remark that, in practice, both a standard ℓ_2 norm $\|\cdot\|_2$ and a weighted ℓ_2 norm $\|\cdot\|_{\mathbf{Z}}$ achieve similarly

satisfying performance. In this work, we use $\|\cdot\|_{\mathbf{Z}}$ with $\mathbf{Z} = \text{diag}(\frac{g_M(\ell)}{M})$ mainly for the purpose of introducing the Jackson kernel $K(f_2 - f_1) := \mathbf{a}(f_1)^H \mathbf{Z} \mathbf{a}(f_2)$ so that we can exploit the beautiful decaying properties of the Jackson kernel (see Section A.3 for more details). When we exchange the frequency and temporal domains, this weighting scheme trusts low-frequency samples more than high-frequency ones, even though the noise levels are the same. The second term is a regularization term that penalizes solutions with large atomic norms, which typically correspond to spectrally dense signals. The regularization parameter λ , whose value will be given later, controls the trade-off between data fidelity and sparsity.

Once \mathbf{x}^{glob} was solved, we can extract estimates of the frequencies either from the primal optimal solution \mathbf{x}^{glob} or from the corresponding dual optimal solution. Our goal is to characterize conditions such that i) we obtain exactly k estimated frequencies; ii) there is a natural correspondence between the estimated frequencies and the true frequencies, whose distances can be explicitly controlled; iii) the distances between the corresponding coefficients can also be explicitly bounded.

To formally present the main theorem, we need to define a few more quantities. It is known that there is a resolution limit of the atomic norm approach in resolving the atoms, or the frequency parameter \mathbf{f}^* , even from the noiseless data [16]. Therefore, to recover the support of the line spectral signal \mathbf{x}^* , we need to impose certain separation condition on the distances of the true frequencies. For this purpose, we define $\Delta(T) = \min_{\{f_\ell, f_m\} \subset T: f_\ell \neq f_m} |f_\ell - f_m|$, where $|\cdot|$ is understood as the wrap-around distance in \mathbb{T} . For example, $|0.1 - 0.9| = 0.2$ under this distance. We also define 1) the dynamic range of the coefficients $B^* := \frac{c_{\max}^*}{c_{\min}^*}$, where c_{\max}^* and c_{\min}^* denote the maximal and minimal modules of $\{c_\ell^*\}_{\ell=1}^k$; 2) the normalized noise level $\gamma_0 := \sigma \sqrt{\frac{\log n}{n}}$; 3) the Noise-to-Signal Ratio $\gamma := \gamma_0 / c_{\min}^*$ and 4) the regularization parameter $\lambda = 0.646 X^* \gamma_0$ for some positive constant X^* to be determined later. Now we are ready to present our main result.

Theorem 2.1. *Suppose we observe $2n + 1$ noisy consecutive samples $y_\ell = x_\ell^* + w_\ell$ of the signal (1.2) or (2.2) with w_ℓ being i.i.d. complex Gaussian valuables of mean zero and variance σ^2 . We assume $n \geq 130$ and*

$$\Delta(T^*) \geq 2.5009/n, \quad (2.5)$$

$$X^* B^* \gamma \leq 10^{-3} \text{ and } B^* / X^* \leq 10^{-4}. \quad (2.6)$$

Then with probability at least $1 - \frac{1}{n^2}$, the optimal solution of (2.4) has a decomposition $\mathbf{x}^{\text{glob}} = \sum_{\ell=1}^k c_\ell^{\text{glob}} \mathbf{a}(f_\ell^{\text{glob}})$ involving exactly k atoms, whose frequencies and coefficients, when properly ordered, satisfy

$$\max_{1 \leq \ell \leq k} |c_\ell^*| |f_\ell^{\text{glob}} - f_\ell^*| \leq 0.4(X^* + 35.2)\gamma_0/n, \quad (2.7)$$

$$\max_{1 \leq \ell \leq k} |c_\ell^{\text{glob}} - c_\ell^*| \leq (X^* + 35.2)\gamma_0. \quad (2.8)$$

Several remarks on the conditions follow. Because of the weighting scheme we use in (2.4), our choice of λ differs from the standard one in [10] by a factor $1/n$ and ensures that the weighted dual atomic norm of the noise, $\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*$, is less than λ with high probability. For technical reasons, our separation condition (2.5) is stronger compared with the previous works [1, 2, 5, 17–20]¹. The conditions (2.6) wrap several requirements on the problem parameters for the conclusions to hold: the dynamic range of the coefficients B^* , the Noise-to-Signal Ratio γ , and the normalized noise γ_0 should all be small while the regularization parameter λ should be large enough as measured by X^* .

It is worth noting that (2.6) implicitly imposes a strong assumption on the Noise-to-Signal Ratio

$$\gamma \leq 10^{-7} / B^{*2}$$

¹Note that our separation condition is a bit larger when comparing to these recent works in super-resolution, while there are two other things to be considered. One thing is that most of these works require strong assumptions on the noise in their models (e.g., the noise is bounded), while our work removes such assumptions and hence can deal with the more general Gaussian noise. To make this possible, we have to develop a new proof strategy involving the two-step construction process of the dual certificate. Another thing is that although some prior works achieve small resolution limit (even comparable to the Rayleigh diffraction limit [20]), they study a different problem. For example, [20] considers the signal denoising problem, that is, stable recovery of the whole signal \mathbf{x} rather than the parameter estimation (i.e., the source location recovery). While the focus of our work is the accuracy of parameter estimation in Gaussian noise, which might be more significant for practical applications such as Radar and single-molecule microscopy, where precisely locating each target/point source is extremely important. Since the parameter estimation problem is much harder than the denoising problem, we have to relax a bit the separation condition for ease of analysis.

implying a sufficiently large n (but still finite). For high-level ideas, there might be two reasons to account for this phenomenon. One is that the problem of line spectral estimation is known to be sensitive to noise. Another is inherently from our proof regime, which makes the constants in Eq. (2.6) a bit conservative. More precisely, the ultimate objective is to show the boundedness and interpolation property of the target polynomial (see Proposition 4.1 for more details). Our method is using an “existing” dual polynomial in [1] satisfying this property and showing the distance between these two polynomials is sufficiently small. So, we require the noise level to be small, since we will see in Lemma 4.2 that the noise level will influence this distance.

One more remark is that the quantity $35.2\gamma_0$ in our results is related to the expected dual atomic norm of the weighted Gaussian noise $\mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*$. By noting the definition $\lambda = 0.646X^*\gamma_0$, we can rewrite the error bounds (2.7) and (2.8) in a more concise way:

$$\max_{1 \leq \ell \leq k} |c_\ell^*| |f_\ell^{\text{glob}} - f_\ell^*| = O(\lambda + \mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*)/n, \quad (2.9)$$

$$\max_{1 \leq \ell \leq k} |c_\ell^{\text{glob}} - c_\ell^*| = O(\lambda + \mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*). \quad (2.10)$$

Eq. (2.9) and (2.10) imply that the error bounds are determined jointly by the regularization parameter λ and the expected dual atomic norm of the weighted Gaussian noise $\mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*$. Since the regularization parameter λ has the same order as $\mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}$, the estimated frequencies and coefficients are guaranteed to have errors of orders $O(\mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*/n)$ and $O(\mathbb{E}\|\mathbf{Z}\mathbf{w}\|_{\mathcal{A}}^*)$, respectively. Remarkably, using atomic dual norm strategy allows us dealing with the Gaussian noise, while most prior works [2, 17–19] in approximate support recovery have to build their theoretical foundations on the bounded-noise assumption, which dramatically narrow down the applications.

Now we summarize the above comparisons of our result with those state-of-the-art modern support recovery methods in the Table 2.

Paper	Bounded Noise	Positive Measure	Support Condition	SNR	Support Recovery
[1, Theorem 1.5]	Yes	No	$\Delta \geq \frac{2}{n}$	Finite	None
[2, Theorem 1.2]	No	No	$\Delta \geq \frac{2}{n}$	Finite	None
[20, Theorem 1]	No	Yes	RRC	Finite	None
[17, Theorem 1.2]	Yes	No	$\Delta \geq \frac{2}{n}$	Finite	Exist
[5, Theorem 2]	No	No	$\Delta \geq \frac{2}{n}$	Finite	Exist
[18, Theorem 2]	Yes	No	NDSC	Infinite	Unique
[19, Theorem 2]	Yes	Yes	NDSC	Infinite	Unique
Theorem 1.2	No	No	$\Delta \geq \frac{2.5009}{n}$	Finite	Unique

Table 2: Comparison with other modern line spectral estimation/super-resolution methods. The *Positive Measure* column refers to whether the result requires the ground-truth measure to be positive. RRC is short for Rayleigh Regularity condition [20, Definition 1.1], which generalizes the standard separation condition to clustered support. NDSC stands for the non-degenerate source condition [18, Definition 5]. In the *Support Recovery* column, *None* indicates that the work considers signal recovery instead of support recovery; *Existence* means that the work shows the existence of at least one recovered parameter around each ground-true parameter, but fails to theoretically eliminate the possibility of spurious recovered parameters; *Uniqueness* shows that around each true parameter there is one and only one recovered parameter.

Finally, our proof for Theorem 2.1 also reveals the connection between the atomic norm minimization (2.4) and the following ℓ_1 -norm regularized, nonlinear and nonconvex least-squares program:

$$\underset{\mathbf{f}, \mathbf{c}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1, \quad (2.11)$$

where $\mathbf{f} := [f_1, \dots, f_k]^T$, $\mathbf{c} := [c_1, \dots, c_k]^T$, and $\mathbf{A}(\mathbf{f}) := [\mathbf{a}(f_1), \dots, \mathbf{a}(f_k)]$. The program (2.11) is highly nonconvex, with numerous local minima and saddle points, so solving it to global optimality is very difficult. Our analysis

shows that, under the conditions of (2.1), the convex program (2.4) shares the same global optimum as the nonconvex program (2.11), implying that the atomic norm minimization provides a new convex way to solve the nonconvex program to global optimality. We summarize the result in the following corollary, with the formal proof listed in Appendix J.

Corollary 2.1. *Under the same setup as in Theorem 2.1, with probability at least $1 - \frac{1}{n^2}$, the frequencies and coefficients estimated by the atomic norm regularized minimization (2.4) constitute a global optimum of the ℓ_1 -regularized nonlinear least-squares program (2.11).*

3 Prior Art and Inspirations

Classical line spectral estimation techniques can be broadly classified into two camps: non-parametric and parametric methods. Non-parametric methods are mainly based on Fourier analysis [21, 22]. Such approaches have low computational complexities and no need for signal models. These methods have limited frequency resolution due to spectral leakage. Parametric methods, however, can achieve high resolution for parameter estimation. For example, Prony’s method based on polynomial root-finding [23, 24] can resolve arbitrarily close frequencies in the noiseless setting. Yet this method is highly sensitive to noise and would fail even in the small noise regime. As stable versions of Prony’s method, the subspace methods recast the noise-sensitive polynomial root-finding problem into more robust matrix eigenvalue problems. For instance, the matrix pencil method [25] arranges the observations into a matrix pencil whose generalized eigenvalues and eigenvectors contain information about the frequencies; the MUSIC algorithm [26] and the ESPRIT method [27] decompose the autocorrelation matrix into noise-subspace and signal subspace using eigenvalue decomposition and extract frequency estimates from the signal subspace. Both algorithms were shown to achieve CRB asymptotically [28, 29] when the signal length $2n + 1$ and the number of snapshots approach infinite. However, these classical methods are not efficient (i.e., approaching the CRB) even with an infinite number of snapshots, as long as the signal length is finite. Also, all classical parametric methods require knowledge of the model order.

Modern convex optimization based methods formulate line spectral estimation as a linear inverse problem and exploit signal sparsity using ℓ_1 -type regularizations. Such methods are modular, robust, and do not require knowledge of model orders. To apply the ℓ_1 regularization techniques, the continuous frequency domain is divided into a grid of discrete frequencies. When the true frequencies fall onto the discrete Fourier grid, work in compressive sensing guarantees optimal recovery performance [30–32]. When the frequencies do not fall onto the Fourier grid, however, the performance of ℓ_1 minimization degrades significantly due to basis mismatch [33]. The basis mismatch issue can be mitigated by employing finer grids [34, 35], which unfortunately often leads to numerical instability.

Atomic norm regularization avoids basis mismatch by enforcing sparsity directly in the continuous frequency domain. Given a set of atoms, possibly indexed by continuous parameters, one constructs an atomic norm in a principled way as a generalization of the ℓ_1 -norm to promote signals with parsimonious representations. Using the notion of descent cones, the authors of [9] argued that the atomic norm is the best possible convex proxy for recovering sparse models. For the special line spectral estimation problem, where the atomic norm is induced by the set of parameterized complex exponentials, atomic regularizations have been shown to achieve optimal performance for several signal processing tasks. For instance, atomic norm minimization recovers a spectrally sparse signal from a minimal number of random signal samples [4], identifies and removes a maximal number of outliers [6, 7], and performs denoising with an error approaching the minimax rate [5]. When multiple measurement vectors are available, a method of exploiting the joint sparsity pattern of different signals to further improve estimation accuracy is proposed in [36–38]. All these works draw inspirations from the dual polynomial construction strategy developed in the pioneer work [1]. This paper adds to this line of work by showing that the atomic framework produces optimal noisy frequency estimators.

Several closely related works also studied conditions for approximate support recovery from noisy observations. The work [17] developed error bounds on spectral support recovery for bounded noise. In [5], the authors derived suboptimal bounds for the Gaussian noise model. In [39], the authors extended this line of research to general measurement schemes beyond Fourier samples using the Beurling-LASSO (B-LASSO) program. The B-LASSO programs, which minimizes a least-squares term plus the measure total variation norm, is mathematically equivalent to the atomic norm formulation. All these works [5, 17, 39] cannot guarantee the recovery of exactly one frequency in each neighborhood of the true frequencies. In this regard, the work by Duval and Peyré [18] showed that as long as

the SNR is large enough and the sources are well-separated and satisfy a *non-degenerate source condition*, then total variation norm regularization can recover the correct number of the Diracs with both the coefficient error and the frequency error scale as the ℓ_2 norm of the noise. Compared with their work, our result uses the (weighted) dual atomic norm of the noise in place of the ℓ_2 norm, which differ by order of \sqrt{n} , allowing our bound to match the CRB up to a logarithmic factor. In addition, their work relies on a *non-degenerate source condition* [18, Definition 5] that is not proven to hold in the spectral super-resolution setting. In this sense, the present paper is the first to rigorously establish that in a high SNR regime this approach yields the right number of frequencies. Further our proof technique based on the primal-dual witness construction is also very different from that employed in [18] based on a perturbation analysis of the dual certificate in the noise-free case. In particular, our analysis reveals the connection between the convex approach and a natural nonlinear least-squares method for spectral estimation. More recently, [19] studies the support recovery for positive measures. For a comparison, there are several major differences worth remarking here: 1) in [19] more emphasis is put on the asymptotic analysis, while the presented work instead deals with non-asymptotic settings with finite signal length; 2) [19] requires the underlying noise to have finite ℓ_2 norm, which severely restricts the scope of noises satisfying such a property, excluding the well-known and most common Gaussian noise, while the presented results allow the underlying noise to be Gaussian; 3) in addition to requiring a sufficiently large signal-to-noise ratio, the main result in [19] also relies on a *non-degenerate source condition* that is not proven to hold in the spectral super-resolution setting.

4 Proof by Primal-Dual Witness Construction

Duality plays an important role in understanding atomic norm regularized line spectral estimation. Standard Lagrangian analysis shows that the dual problem of (2.4) has the following form:

$$\begin{aligned} \mathbf{q}^{\text{glob}} &= \underset{\mathbf{q}}{\operatorname{argmax}} \frac{1}{2} \|\mathbf{y}\|_{\mathbf{Z}}^2 - \frac{1}{2} \|\mathbf{y} - \lambda \mathbf{q}\|_{\mathbf{Z}}^2 \\ &\text{subject to } \|\mathbf{Z}\mathbf{q}\|_{\mathcal{A}}^* \leq 1. \end{aligned} \quad (4.1)$$

The complex trigonometric polynomial $Q(f) := \mathbf{a}(f)^H \mathbf{Z}\mathbf{q}$ corresponding to a dual feasible solution \mathbf{q} is called a dual polynomial. The dual polynomial associated with the unique dual optimal solution $Q^{\text{glob}}(f) := \mathbf{a}(f)^H \mathbf{Z}(\mathbf{q}^{\text{glob}})$ certifies the optimality of the unique primal optimal solution \mathbf{x}^{glob} , and vice versa. The uniqueness of primal and dual optimal solutions is a consequence of the strong convexity of the objective functions of (2.4) and (4.1), respectively. In particular, the primal-dual optimal solutions are related by $\mathbf{q}^{\text{glob}} = (\mathbf{y} - \mathbf{x}^{\text{glob}})/\lambda$. We summarize these in the following proposition, with the proof given in Appendix I:

Proposition 4.1. *Let the decomposition $\hat{\mathbf{x}} = \sum_{\ell=1}^{\hat{k}} \hat{c}_{\ell} \mathbf{a}(\hat{f}_{\ell})$ with distinct frequencies $\hat{T} = \{\hat{f}_{\ell}\} \subset \mathbb{T}$ and nonzero coefficients $\{\hat{c}_{\ell}\}$ and set $\hat{\mathbf{q}} = (\mathbf{y} - \hat{\mathbf{x}})/\lambda$. Suppose the corresponding dual polynomial $\hat{Q}(f) = \mathbf{a}(f)^H \mathbf{Z}\hat{\mathbf{q}}$ satisfies the following Bounded Interpolation Property (BIP):*

$$\begin{aligned} \hat{Q}(\hat{f}_{\ell}) &= \operatorname{sign}(\hat{c}_{\ell}), \ell = 1, \dots, \hat{k} \quad (\text{Interpolation}); \\ |\hat{Q}(f)| &< 1, \forall f \notin \hat{T} \quad (\text{Boundedness}); \end{aligned}$$

then $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$ are the unique primal-dual optimal solutions to (2.4) and (4.1), that is, $\hat{\mathbf{x}} = \mathbf{x}^{\text{glob}}$ and $\hat{\mathbf{q}} = \mathbf{q}^{\text{glob}}$. Here the operation $\operatorname{sign}(c) := c/|c|$ for a nonzero complex number and applies entry-wise to a vector.

Proposition 4.1 gives a way to extract the frequencies from the dual optimal solution – one can simply identify the frequencies where the dual polynomial corresponding to the dual optimal solution achieves magnitude 1. The uniqueness of the dual solution for (2.4) makes the construction of a dual certificate much harder compared with the line spectral signal completion problem [4] and demixing problem [6, 7]. For the latter two problems, while the primal optimal solution is unique, the dual optimal solutions are non-unique. One usually chooses one dual solution that is easier to analyze (e.g., the one with minimal energy). For the support recovery problem, we need to simultaneously construct the primal and dual solutions, which witness the optimality of each other. In the compressive sensing

literature, this construction process is called the *primal-dual witness construction* [40]. In sparse recovery problems, a candidate primal solution is relatively easy to find, since when the noise is relatively small, the support of the recovered signal would not change. So one only needs to solve a LASSO problem restricted to the true support to determine the candidate coefficients, as was done in [40]. For the optimization (2.4), due to the continuous nature of the atoms, even a bit of noise would drive the support away from the true one. So to construct a candidate primal solution (hence a candidate dual solution), we need to simultaneously seek for the candidate support $\{\hat{f}_\ell\}$ and the candidate coefficients $\{\hat{c}_\ell\}$.

4.1 Proof Outline

We use the ℓ_1 -regularized, nonlinear and nonconvex program (2.11), which we copy below, to find plausible candidates for $\{\hat{f}_\ell\}$ and $\{\hat{c}_\ell\}$:

$$\underset{\mathbf{f}, \mathbf{c}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1,$$

where $\mathbf{f} = [f_1, \dots, f_k]^T$, $\mathbf{c} = [c_1, \dots, c_k]^T$ and $\mathbf{A}(\mathbf{f}) = [\mathbf{a}(f_1), \dots, \mathbf{a}(f_k)]$. Note that we have effectively fixed the number of estimated frequencies \hat{k} in Proposition 4.1 to be k . But unlike in compressive sensing we cannot fix $\mathbf{f} = \mathbf{f}^*$ to solve for \mathbf{c} only as was done in [40]. The program (2.11) is highly nonconvex, with numerous local minima, local maxima, and saddle points. So solving it to global optimality is hard even in theory. We are primarily interested in its local minimum $(\{\hat{f}_\ell\}, \{\hat{c}_\ell\})$ in a neighborhood of the true frequencies and coefficients $(\mathbf{f}^*, \mathbf{c}^*)$. To find this local minimum, we will run gradient descent to (2.11) using $(\mathbf{f}^*, \mathbf{c}^*)$ as initialization. We will argue that under conditions presented in Theorem 2.1, each \hat{f}_ℓ and \hat{c}_ℓ stay close to f_ℓ^* and c_ℓ^* as given in (2.7) and (2.8), respectively. The major tool we use is the contraction mapping theorem. As shown in Corollary 2.1, the local minimum found in this manner is actually a global optimum of (2.11).

The rest of arguments consist of showing that $\hat{\mathbf{x}} = \sum_{\ell=1}^k \hat{c}_\ell \mathbf{a}(\hat{f}_\ell)$ with $\{\hat{f}_\ell\}$ and $\{\hat{c}_\ell\}$ constructed as described above satisfies the Bounded Interpolation Property of Proposition 4.1. The Interpolation property is automatically satisfied due to the construction process and the main challenge is to show the Boundedness property $|\hat{Q}(f)| < 1, \forall f \notin \hat{T}$. The harder part is showing the Boundedness property. For ease of interpretation we first collect the definitions of the most important variables that will be used throughout the proof, and then introduce the *main logic* and the *two-step construction process* of the proof.

Symbol	Definition
$(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$	The local minima of $\underset{\mathbf{f}, \mathbf{c}}{\text{minimize}} \frac{1}{2} \ \mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{x}^*\ _{\mathbf{Z}}^2 + \lambda \ \mathbf{c}\ _1$ that is closest to $(\mathbf{f}^*, \mathbf{c}^*)$
$(\hat{\mathbf{f}}, \hat{\mathbf{c}})$	The local minima of $\underset{\mathbf{f}, \mathbf{c}}{\text{minimize}} \frac{1}{2} \ \mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\ _{\mathbf{Z}}^2 + \lambda \ \mathbf{c}\ _1$ that is closest to $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$
\mathbf{x}^λ	The primal solution defined by the local minima $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$ via $\hat{\mathbf{x}} := \sum_{\ell=1}^k c_\ell^\lambda \mathbf{a}(f_\ell^\lambda)$
$\hat{\mathbf{x}}$	The primal solution defined by the local minima $(\hat{\mathbf{f}}, \hat{\mathbf{c}})$ via $\hat{\mathbf{x}} := \sum_{\ell=1}^k \hat{c}_\ell \mathbf{a}(\hat{f}_\ell)$
\mathbf{q}^λ	The dual solution corresponding to the primal solution \mathbf{x}^λ , that is, $\mathbf{q}^\lambda := (\mathbf{x}^* - \mathbf{x}^\lambda)/\lambda$
$\hat{\mathbf{q}}$	The dual solution corresponding to the primal solution $\hat{\mathbf{x}}$, that is, $\hat{\mathbf{q}} := (\mathbf{y} - \hat{\mathbf{x}})/\lambda$
\mathbf{q}^*	$\mathbf{q}^* := \lim_{\lambda \rightarrow 0} \mathbf{q}^\lambda$, satisfying the Boundedness and Interpolation property for $(\mathbf{f}^*, \mathbf{c}^*)$

Main Logic: Firstly, identify that $Q^*(f) := \mathbf{a}(f)^H \mathbf{Z} \mathbf{q}^*$ satisfies the Boundedness property with some similar arguments used in [1]. Secondly, establish that $\hat{\mathbf{q}}$ and \mathbf{q}^* are sufficiently close (so are $\hat{Q}(f) := \mathbf{a}(f)^H \mathbf{Z} \hat{\mathbf{q}}$ and $Q^*(f) = \mathbf{a}(f)^H \mathbf{Z} \mathbf{q}^*$). Therefore $\hat{Q}(f)$ also satisfies the Boundedness property. It turns out that directly showing the closeness $\hat{\mathbf{q}}$ and \mathbf{q}^* is difficult. That is why we introduce the intermediate dual variable \mathbf{q}^λ and use the *two-step construction process*, i.e., first showing \mathbf{q}^* is close to \mathbf{q}^λ and then showing \mathbf{q}^λ is close to $\hat{\mathbf{q}}$.

Two-step Construction Process: We will first find a local minimum $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$ of $\frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{x}^*\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1$ around $(\mathbf{f}^*, \mathbf{c}^*)$, where one should note we replaced the noisy signal \mathbf{y} in (2.11) by the noise-free signal \mathbf{x}^* . We will then run gradient descent to (2.11) using $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$ as initialization. The intermediate quantities $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$ will serve as a bridge between $(\mathbf{f}^*, \mathbf{c}^*)$ and $(\hat{\mathbf{f}}, \hat{\mathbf{c}})$ to make the proof easier. The key is noting that $\hat{Q}(f) = \mathbf{a}(f)^H \mathbf{Z} \hat{\mathbf{q}}$ is close to

$Q^\lambda(f) = \mathbf{a}(f)^H \mathbf{Z} \mathbf{q}^\lambda$, where $\mathbf{q}^\lambda = (\mathbf{x}^* - \mathbf{x}^\lambda)/\lambda$ and $\mathbf{x}^\lambda = \sum_{\ell=1}^k c_\ell^\lambda \mathbf{a}(f_\ell^\lambda)$, and $Q^\lambda(f)$ is close to $Q^*(f) = \mathbf{a}(f)^H \mathbf{Z} \mathbf{q}^*$. Here $\mathbf{q}^* = \lim_{\lambda \rightarrow 0} \mathbf{q}^\lambda$ is a dual certificate used to certify the atomic decomposition of \mathbf{x}^* . The former claim can be showed using the closeness of $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$ and $(\hat{\mathbf{f}}, \hat{\mathbf{c}})$. The later claim, however, must take advantage of the fact that $\mathbf{q}^* = \lim_{\lambda \rightarrow 0} \mathbf{q}^\lambda = -\frac{d}{d\lambda} \mathbf{x}^\lambda|_{\lambda=0}$ and apply the triangle inequality to

$$Q^\lambda(f) - Q^*(f) = \frac{1}{\lambda} \int_0^\lambda \mathbf{a}(f)^H \mathbf{Z} \left(\frac{d}{dt} \mathbf{x}^0 - \frac{d}{dt} \mathbf{x}^t \right) dt,$$

where $\frac{d}{dt} \mathbf{x}^0 = \lim_{\lambda \rightarrow 0} \frac{d}{d\lambda} \mathbf{x}^\lambda := \frac{d}{d\lambda} \mathbf{x}^*$. The closeness of $(\mathbf{f}^\lambda, \mathbf{c}^\lambda)$ and $(\mathbf{f}^*, \mathbf{c}^*)$ ensures that the derivatives in the integrand are also close. Finally, we exploit the properties of $Q^*(f)$ which are similar to those established in [1] to complete the proof.

4.2 A Formal Proof: Applying the Contraction Mapping Theorem

Theorem 4.1 (Contraction Mapping Theorem). *Given a Banach space \mathcal{B} equipped with a norm $\|\cdot\|$, a bounded closed set $\mathcal{N} \subset \mathcal{B}$ and a map $\Theta : \mathcal{N} \rightarrow \mathcal{B}$, if $\Theta(\mathcal{N}) \subset \mathcal{N}$ (the non-escaping property) and there exists $\rho \in (0, 1)$ such that $\|\Theta(\mathbf{v}) - \Theta(\mathbf{w})\| \leq \rho \|\mathbf{v} - \mathbf{w}\|$ for each $\mathbf{v}, \mathbf{w} \in \mathcal{N}$ (the contraction property), then there exists a unique $\mathbf{v}^* \in \mathcal{N}$ such that $\Theta(\mathbf{v}^*) = \mathbf{v}^*$.*

This classical result helps to find a candidate solution for the construction of a valid dual certificate. To see this we first choose the bounded closed set \mathcal{N} to be a small region around the target joint frequency-coefficient vector $\boldsymbol{\theta}^* := (\mathbf{f}^*, \mathbf{u}^*, \mathbf{v}^*)$ (where \mathbf{u}^* and \mathbf{v}^* denote respectively the real and imaginary parts of \mathbf{c}^*). Let the fixed point map Θ be the gradient map of (2.11). The key is to determine the size of \mathcal{N} in which the non-escaping and the contraction properties of the fixed point map Θ hold. Then, the contraction mapping theorem implies that iteratively performing the gradient map Θ from any initial point in \mathcal{N} would produce a candidate solution that still lies in \mathcal{N} (by the non-escaping property) and hence is close to $\boldsymbol{\theta}^*$ (since \mathcal{N} is small). Finally relating the fixed point equation to the BIP property shows that such a candidate solution generates a valid dual certificate.

In order to apply the contraction mapping theorem to our problem, we choose the norm in Theorem 4.1 to be a weighted ℓ_∞ norm $\|\cdot\|_\infty$ given by $\|(\mathbf{f}, \mathbf{u}, \mathbf{v})\|_\infty := \|(\mathbf{S}\mathbf{f}, \mathbf{u}, \mathbf{v})\|_\infty$ with $\mathbf{S} := \sqrt{|K''(0)|} \text{diag}(|\mathbf{c}^*|)$ and $K(\cdot)$ is the Jackson kernel (refer to Appendix A for an introduction). This weighted ℓ_∞ norm is used as a metric function to define the neighborhood \mathcal{N} around $\boldsymbol{\theta}^*$. The choice of the weighting matrix \mathbf{S} ensures that the larger a coefficient c_i^* is, the smaller the neighborhood in the direction of the frequency f_i . In addition, since $\sqrt{|K''(0)|}$ is of order $O(n)$, the frequency neighborhood is smaller than the coefficient neighborhood by the same order. Next, we choose the fixed point map Θ to be a weighted gradient map of (2.11)

$$\Theta(\boldsymbol{\theta}) := \boldsymbol{\theta} - \mathbf{W}^* \nabla \left(\frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{c}\|_1 \right), \quad (4.2)$$

where the gradient ∇ is taken with respect to the parameter $\boldsymbol{\theta} = (\mathbf{f}, \mathbf{u}, \mathbf{v})$ and the weighting matrix

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{S}^{-2} & & \\ & \mathbf{I}_k & \\ & & \mathbf{I}_k \end{bmatrix}. \quad (4.3)$$

Scaling the gradient vector by \mathbf{W}^* ensures that the Jacobian matrix of the second term in (4.2) is close to the identity matrix, which makes it easier to show the contraction property of Θ .

4.2.1 Two-step Construction Process

As discussed in Section 4.1, we divide the construction process into two steps. We first analyze the fixed point map Θ^λ obtained by replacing the noisy observation vector \mathbf{y} in (4.2) by the noise-free signal \mathbf{x}^* . We determine a region around $\boldsymbol{\theta}^*$, say \mathcal{N}^* , such that both the contraction and non-escaping properties of Θ^λ are satisfied in \mathcal{N}^* . Then by the contraction mapping theorem, iterating the gradient map Θ^λ in \mathcal{N}^* initialized by $\boldsymbol{\theta}^*$ generates a unique fixed point $\boldsymbol{\theta}^\lambda := (\mathbf{f}^\lambda, \mathbf{u}^\lambda, \mathbf{v}^\lambda)$. These results are summarized in the following lemma:

Lemma 4.1 (The First Fixed Point Map). *Let the first fixed point map be the weighted gradient map of the nonconvex program (2.11) with the noisy signal \mathbf{y} replaced by the noise-free signal \mathbf{x}^* :*

$$\Theta^\lambda(\boldsymbol{\theta}) := \boldsymbol{\theta} - \mathbf{W}^* \nabla \left(\frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{x}^*\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1 \right), \quad (4.4)$$

where the gradient ∇ is taken with respect to the parameter $\boldsymbol{\theta} = (\mathbf{f}, \mathbf{u}, \mathbf{v})$. Let the regularization parameter λ vary in $[0, 0.646X^*\gamma_0]$. Define a neighborhood $\mathcal{N}^* := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq X^*\gamma_0/\sqrt{2}\}$. Suppose that the separation condition (2.5) and the SNR condition (2.6) hold. Then the map Θ^λ has a unique fixed point $\boldsymbol{\theta}^\lambda \in \mathcal{N}^*$ satisfying $\Theta^\lambda(\boldsymbol{\theta}^\lambda) = \boldsymbol{\theta}^\lambda$. Furthermore, according to the implicit function theorem, $\boldsymbol{\theta}^\lambda$ is a continuously differentiable function of λ whose derivative is given by

$$\frac{d}{d\lambda} \boldsymbol{\theta}^\lambda = -(\nabla^2 G^\lambda(\boldsymbol{\theta}^\lambda))^{-1} \frac{\partial}{\partial \lambda} \nabla G^\lambda(\boldsymbol{\theta}^\lambda). \quad (4.5)$$

Finally, when λ turns to zero, the fixed point $\boldsymbol{\theta}^\lambda$ converges to $\boldsymbol{\theta}^*$, i.e., $\lim_{\lambda \rightarrow 0} \boldsymbol{\theta}^\lambda = \boldsymbol{\theta}^*$, and therefore $\lim_{\lambda \rightarrow 0} \mathbf{x}^\lambda = \mathbf{x}^*$.

Proof of Lemma 4.1. See Appendix D. □

We now turn to the gradient map Θ in (4.2) defined in a region \mathcal{N}^λ around $\boldsymbol{\theta}^\lambda$. Similar to the first step, we show the contraction and non-escaping properties of Θ in \mathcal{N}^λ , which imply that iterating the gradient map Θ initialized by $\boldsymbol{\theta}^\lambda$ produces a unique fixed point $\hat{\boldsymbol{\theta}} := (\hat{\mathbf{f}}, \hat{\mathbf{u}}, \hat{\mathbf{v}})$.

Lemma 4.2 (The Second Fixed Point Map). *Let the second fixed point map be the weighted gradient map of the nonconvex program (2.11):*

$$\Theta(\boldsymbol{\theta}) = \boldsymbol{\theta} - \mathbf{W}^* \nabla \left(\frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1 \right) \quad (4.6)$$

and the region $\mathcal{N}^\lambda := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^\lambda\|_\infty \leq 35.2\gamma_0/\sqrt{2}\}$. Set the regularization parameter λ as $0.646X^*\gamma_0$ in (4.6). Suppose that the separation condition (2.5) and the SNR condition (2.6) hold. Then with probability at least $1 - \frac{1}{n^2}$, $\Theta(\boldsymbol{\theta})$ has a unique fixed point $\hat{\boldsymbol{\theta}}$ living in \mathcal{N}^λ .

Proof of Lemma 4.2. See Appendix E. □

The radius of the second contraction region \mathcal{N}^λ is determined by a high probability bound on the dual atomic norm of the Gaussian noise and ensures that \mathcal{N}^λ is a non-escaping set for $\Theta(\boldsymbol{\theta})$. So far, we have identified the neighborhoods where the two fixed points $\boldsymbol{\theta}^\lambda$ and $\hat{\boldsymbol{\theta}}$ live in, which is the key to show the validity of the dual certificates later. Figure 1 illustrates the main results of Lemma 4.1 and Lemma 4.2.

Road Map: Define two pre-certificates using the two fixed points as $\mathbf{q}^\lambda := (\mathbf{x}^* - \mathbf{x}^\lambda)/\lambda$ and $\hat{\mathbf{q}} := (\mathbf{y} - \hat{\mathbf{x}})/\lambda$ with the corresponding pre-dual polynomials denoted by $Q^\lambda(f)$ and $\hat{Q}(f)$. Here $\mathbf{x}^\lambda = \sum_{\ell=1}^k c_\ell^\lambda \mathbf{a}(f_\ell^\lambda)$ and $\hat{\mathbf{x}} = \sum_{\ell=1}^k \hat{c}_\ell \mathbf{a}(\hat{f}_\ell)$. Let $\mathbf{q}^* = \lim_{\lambda \rightarrow 0} \mathbf{q}^\lambda$. The remaining steps are to:

- 1) Show that \mathbf{q}^* is a valid dual certificate that certifies the atomic decomposition of \mathbf{x}^* , i.e., $Q^*(f) = \mathbf{a}(f)^H \mathbf{Z} \mathbf{q}^*$ satisfies $Q^*(f_\ell^*) = \text{sign}(c_\ell^*)$, $\ell = 1, \dots, k$ and $|Q^*(f)| < 1, \forall f \notin T^*$;
- 2) Use Lemma 4.1 to bound the pointwise distance between $Q^*(f)$ and $Q^\lambda(f)$;
- 3) Use Lemma 4.2 to bound the pointwise distance between $Q^\lambda(f)$ and $\hat{Q}(f)$.

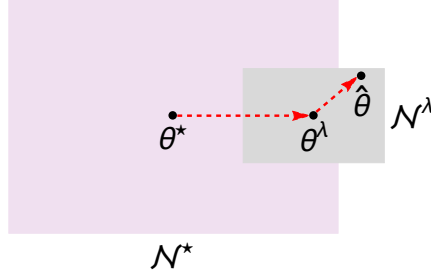


Figure 1: Use the true parameter vector θ^* as an initialization and run the first weighted gradient map (4.4) to obtain the first fixed point $\theta^\lambda \in \mathcal{N}^*$. Run the second weighted gradient map (4.6) initialized by θ^λ to get the second fixed point $\hat{\theta} \in \mathcal{N}^\lambda$. The closeness of $\hat{\theta}$ and θ^* is determined by the sizes of the two neighborhoods \mathcal{N}^* and \mathcal{N}^λ , whose precise forms are given in Lemma 4.1 and 4.2, respectively.

4.2.2 Showing \mathbf{q}^* is a Dual Certificate

To show that \mathbf{q}^* is a dual certificate, it is sufficient to show that $Q^*(f)$ satisfies the Bounded Interpolation Property of Proposition 4.1. The Interpolation property is automatically satisfied due to the construction process, and we will show the Boundedness property using the arguments of [1]. In particular, fix an arbitrary point $f_0^* \in T^*$ as the reference point, and let f_{-1}^* be the first frequency in T^* that lies on the left of f_0^* while f_1^* be the first frequency in T^* that lies on the right. Here “left” and “right” are directions on the complex circle \mathbb{T} . We remark that the analysis depends only on the relative locations of $\{f_\ell^*\}$. Hence, to simplify the arguments, we assume that the reference point f_0^* is at 0 by shifting the frequencies if necessary. Then we divide the region between $f_0^* = 0$ and $f_1^*/2$ into three parts: Near Region $\mathcal{N} := [0, 0.24/n]$, Middle Region $\mathcal{M} := [0.24/n, 0.75/n]$ and Far Region $\mathcal{F} := [0.75/n, f_1^*/2]$. Also their symmetric counterparts are defined as $-\mathcal{N} := [-0.24/n, 0]$, $-\mathcal{M} := [-0.75/n, -0.24/n]$, and $-\mathcal{F} := [f_{-1}^*/2, -0.75/n]$. We first show that the dual polynomial has strictly negative curvature $|Q^*(f)|'' < 0$ in $\mathcal{N} = [0, 0.24/n]$ and $|Q^*(f)| < 1$ in $\mathcal{M} \cup \mathcal{F} = [0.24/n, f_1^*/2]$, implying $|Q^*(f)| < 1$ in $\mathcal{N} \cup \mathcal{M} \cup \mathcal{F} \setminus \{f_0^*\}$ by exploiting $|Q^*(f_0^*)| = 1$ and $|Q^*(f_0^*)'| = 0$. Then using the same symmetric arguments as in [1], we claim that $|Q^*(f)| < 1$ in $(-\mathcal{N}) \cup (-\mathcal{M}) \cup (-\mathcal{F}) \setminus \{f_0^*\}$. Combining these two results with the fact that the reference point f_0^* is chosen arbitrarily from T^* (and shifted to 0), we establish that the Boundedness property of $Q^*(f)$ holds in the entire $\mathbb{T} \setminus T^*$.

Lemma 4.3 (\mathbf{q}^* is a dual certificate). *The dual polynomial $Q^*(f)$ satisfies both the Interpolation and Boundedness properties with respect to the coefficients $\{c_\ell^*\}$ and the frequencies $\{f_\ell^*\}$. In addition, $Q^*(f)$ satisfies first*

$$\begin{aligned} Q_R^*(f) &\geq 0.887594, & Q_R^{*''}(f) &\leq -2.24483n^2, \\ |Q_I^*(f)| &\leq 0.0183836, & |Q_I^{*''}(f)| &\leq 0.113197n^2, \\ |Q^{*'}(f)| &\leq 0.821039n, & |Q^{*''}(f)| &\leq 3.40320n^2, \end{aligned}$$

and

$$Q_R^*(f)Q_R^{*''}(f) + |Q^{*'}(f)|^2 + |Q_I^*(f)||Q_I^{*''}(f)| \leq -1.316313n^2 < 0$$

for $f \in \mathcal{N}$, implying $|Q^*(f)|'' < 0$ in \mathcal{N} , and second,

$$\begin{aligned} |Q^*(f)| &\leq 0.927615, & f \in \mathcal{M}, \\ |Q^*(f)| &\leq 0.734123, & f \in \mathcal{F}. \end{aligned}$$

Here the subscripts R and I denote respectively the real and complex parts of $Q^*(f)$. Thus \mathbf{q}^* is a valid dual certificate to certify the atomic decomposition $\mathbf{x}^* = \sum_{\ell=1}^k c_\ell^* \mathbf{a}(f_\ell^*)$ such that $\|\mathbf{x}^*\|_{\mathcal{A}} = \sum_{\ell=1}^k |c_\ell^*|$.

Proof of Lemma 4.3. See Appendix F. □

Next lemma, with the proof given in Appendix G, exploits the closeness of θ^* and θ^λ shown in Lemma 4.1 to bound the pointwise distance between $Q^*(f)$ and $Q^\lambda(f)$.

Lemma 4.4 ($Q^\lambda(f)$ is close to $Q^*(f)$). *Under the settings of Lemma 4.1, let $Q^\lambda(f)$ and $Q^*(f)$ be the dual polynomials corresponding to θ^λ and θ^* , respectively. Then the distances between $Q^\lambda(f)$ and $Q^*(f)$ and their various derivatives are uniformly bounded:*

$$\begin{aligned} |Q^*(f) - Q^\lambda(f)| &\leq 28.7343X^*B^*\gamma, \quad f \in \mathcal{N}, & |Q^*(f) - Q^\lambda(f)| &\leq 39.3557X^*B^*\gamma, \quad f \in \mathcal{M}, \\ |Q^{*\prime}(f) - Q^{\lambda\prime}(f)| &\leq 44.4648nX^*B^*\gamma, \quad f \in \mathcal{N}, & |Q^*(f) - Q^\lambda(f)| &\leq 66.1596X^*B^*\gamma, \quad f \in \mathcal{F}, \\ |Q^{*\prime\prime}(f) - Q^{\lambda\prime\prime}(f)| &\leq 140.808n^2X^*B^*\gamma, \quad f \in \mathcal{N}. \end{aligned}$$

In the following, we will control the pointwise distance between $Q^\lambda(f)$ and $\hat{Q}(f)$ by taking advantage of the closeness of θ^λ and $\hat{\theta}$ given by Lemma 4.2. The key is to observe that

$$\hat{\mathbf{q}} - \mathbf{q}^\lambda = \frac{(\mathbf{y} - \hat{\mathbf{x}}) - (\mathbf{x}^* - \mathbf{x}^\lambda)}{\lambda} = \frac{\mathbf{w}}{\lambda} + \frac{\mathbf{x}^\lambda - \hat{\mathbf{x}}}{\lambda}$$

implying

$$|Q^\lambda(f) - \hat{Q}(f)| \leq \frac{|\mathbf{a}(f)^H \mathbf{Z} \mathbf{w}|}{\lambda} + \frac{|\mathbf{a}(f)^H \mathbf{Z}(\mathbf{x}^\lambda - \hat{\mathbf{x}})|}{\lambda}. \quad (4.7)$$

This separates the distance between $Q^\lambda(f)$ and $\hat{Q}(f)$ into two parts: one is $|\mathbf{a}(f)^H \mathbf{Z} \mathbf{w}|/\lambda$ determined by the dual atomic norm of the Gaussian noise \mathbf{w} , which is upper bounded in Appendix B; the other is $|\mathbf{a}(f)^H \mathbf{Z}(\mathbf{x}^\lambda - \hat{\mathbf{x}})|/\lambda$ that can be upper bounded by the dual atomic norm of $\mathbf{x}^\lambda - \hat{\mathbf{x}}$. We summarize the final result in Lemma 4.5, where the proof is given in Appendix H.

Lemma 4.5 ($\hat{Q}(f)$ is close to $Q^\lambda(f)$). *Under the settings of Lemma 4.2, let \hat{Q} and Q^λ be the dual polynomials corresponding to $\hat{\theta}$ and θ^λ , respectively. Then the pointwise distances between $Q^\lambda(f)$ and $\hat{Q}(f)$ and their derivatives are bounded:*

$$\begin{aligned} |\hat{Q}(f) - Q^\lambda(f)| &\leq 82.5975B^*/X^*, \quad f \in \mathcal{N}, & |\hat{Q}(f) - Q^\lambda(f)| &\leq 114.323B^*/X^*, \quad f \in \mathcal{M}, \\ |\hat{Q}(f)' - Q^{\lambda\prime}(f)| &\leq 180.283nB^*/X^*, \quad f \in \mathcal{N}, & |\hat{Q}(f) - Q^\lambda(f)| &\leq 162.903B^*/X^*, \quad f \in \mathcal{F}, \\ |\hat{Q}(f)'' - Q^{\lambda\prime\prime}(f)| &\leq 758.404n^2B^*/X^*, \quad f \in \mathcal{N}. \end{aligned}$$

4.2.3 Proof of Theorem 2.1

By combining Lemmas 4.3, 4.4, and 4.5, we are now ready to prove Theorem 2.1.

Basically, we will show that $\hat{\theta}$ constructed from the two-step gradient descent procedure and $\theta^{\text{glob}} := (\mathbf{f}^{\text{glob}}, \mathbf{u}^{\text{glob}}, \mathbf{v}^{\text{glob}})$ are the same point. Then the error bounds follow from the closeness of $\hat{\theta}$ and θ^* . First, we show that the signal $\hat{\mathbf{x}} = \sum_{\ell=1}^k \hat{c}_\ell \mathbf{a}(\hat{f}_\ell)$ and $\hat{\mathbf{q}} = (\mathbf{y} - \hat{\mathbf{x}})/\lambda$ constructed from the second fixed point $\hat{\theta}$ form primal and dual optimal solutions of (2.4). It suffices to show that the dual polynomial $\hat{Q}(f) = \mathbf{a}(f)^H \mathbf{Z} \hat{\mathbf{q}}$ satisfies the Bounded Interpolation Property of Proposition 4.1.

1) Showing the Interpolation property.

The Interpolation property has the following equivalences:

$$\begin{aligned} \hat{Q}(\hat{f}_\ell) = \text{sign}(\hat{c}_\ell), \ell = 1, \dots, k &\iff \mathbf{a}(\hat{f}_\ell)^H \mathbf{Z}(\mathbf{y} - \hat{\mathbf{x}}) = \lambda \text{sign}(\hat{c}_\ell), \ell = 1, \dots, k \\ &\iff \mathbf{a}(\hat{f}_\ell)^H \mathbf{Z}(\mathbf{y} - \mathbf{A}(\hat{\mathbf{f}})\hat{\mathbf{c}}) = \lambda \text{sign}(\hat{c}_\ell), \ell = 1, \dots, k \\ &\iff \mathbf{A}(\hat{\mathbf{f}})^H \mathbf{Z}(\mathbf{y} - \mathbf{A}(\hat{\mathbf{f}})\hat{\mathbf{c}}) = \lambda \hat{\mathbf{c}}./|\hat{\mathbf{c}}|. \end{aligned} \quad (4.8)$$

From Lemma 4.2, $\hat{\theta}$ is the fixed point solution of the map $\Theta(\theta) = \theta - \mathbf{W}^* \nabla G(\theta)$, i.e., $\Theta(\hat{\theta}) = \hat{\theta}$, implying $\nabla G(\hat{\theta}) = \mathbf{0}$ due to the invertibility of \mathbf{W}^* . Invoking the explicit expression for $\nabla G(\theta)$ developed in Appendix C, we get

$$\nabla G(\hat{\theta}) = \begin{bmatrix} \mathbb{R}\{(\mathbf{A}'(\hat{\mathbf{f}}) \text{diag}(\hat{\mathbf{c}}))^H \mathbf{Z}(\mathbf{A}(\hat{\mathbf{f}})\hat{\mathbf{c}} - \mathbf{y})\} \\ \mathbb{R}\{\mathbf{A}(\hat{\mathbf{f}})^H \mathbf{Z}(\mathbf{A}(\hat{\mathbf{f}})\hat{\mathbf{c}} - \mathbf{y}) + \lambda \hat{\mathbf{c}}./|\hat{\mathbf{c}}|\} \\ \mathbb{I}\{\mathbf{A}(\hat{\mathbf{f}})^H \mathbf{Z}(\mathbf{A}(\hat{\mathbf{f}})\hat{\mathbf{c}} - \mathbf{y}) + \lambda \hat{\mathbf{c}}./|\hat{\mathbf{c}}|\} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \quad (4.9)$$

Then the Interpolation property (4.8) follows from the last two row blocks of (4.9).

2) Showing the Boundedness property.

Following the same arguments preceding Lemma 4.3, it is sufficient to show $|\hat{Q}(f)| < 1$ in $\mathcal{N} \cup \mathcal{M} \cup \mathcal{F} \setminus \{\hat{f}_0\}$.

First, since \hat{f}_0 might be located in $-\mathcal{N}$ or \mathcal{N} , we bound $|\hat{Q}(f)|$ for $f \in (-\mathcal{N}) \cup \mathcal{N}$. The second-order Taylor expansion of $|\hat{Q}(f)|$ at $f = \hat{f}_0$ states

$$\begin{aligned} |\hat{Q}(f)| &= |\hat{Q}(\hat{f}_0)| + (f - \hat{f}_0)|\hat{Q}(\hat{f}_0)|' + \frac{1}{2}(f - \hat{f}_0)^2|\hat{Q}(\xi)|'' \\ &= 1 + (f - \hat{f}_0)|\hat{Q}(\hat{f}_0)|' + \frac{1}{2}(f - \hat{f}_0)^2|\hat{Q}(\xi)|'' \text{ for some } \xi \in (-\mathcal{N}) \cup \mathcal{N}, \end{aligned} \quad (4.10)$$

where for the second line we used a consequence of the interpolation property. We argue that

$$|\hat{Q}(\hat{f}_0)|' = \frac{\hat{Q}_R(\hat{f}_0)\hat{Q}_R(\hat{f}_0)' + \hat{Q}_I(\hat{f}_0)\hat{Q}_I(\hat{f}_0)'}{|\hat{Q}(\hat{f}_0)|} = \frac{\mathbb{R}\{\hat{c}_0\}\hat{Q}_R(\hat{f}_0)' + \mathbb{I}\{\hat{c}_0\}\hat{Q}_I(\hat{f}_0)'}{|\hat{c}_0||\hat{Q}(\hat{f}_0)|} = 0.$$

The last equality is a consequence of the first row block of (4.9) since $\mathbb{R}\{\hat{c}_0\}\hat{Q}_R(\hat{f}_0)' + \mathbb{I}\{\hat{c}_0\}\hat{Q}_I(\hat{f}_0)' = \mathbb{R}\{\hat{c}_0^H \mathbf{a}(\hat{f}_0)^H \mathbf{Z}(\mathbf{y} - \mathbf{A}(\hat{\mathbf{f}})\hat{\mathbf{c}})\}$. Therefore, it suffices to show that $|\hat{Q}(f)|'$ has strictly negative derivative in the symmetric Near Region $f \in (-\mathcal{N}) \cup \mathcal{N}$. By the symmetric arguments, it suffices to show this in \mathcal{N} . Since

$$|\hat{Q}(f)|'' = -\frac{(\hat{Q}_R(f)\hat{Q}_R(f)' + \hat{Q}_I(f)\hat{Q}_I(f)')^2}{|\hat{Q}(f)|^3} + \frac{\hat{Q}_R(f)\hat{Q}_R(f)'' + |\hat{Q}(f)'|^2 + |\hat{Q}_I(f)||\hat{Q}_I(f)''|}{|\hat{Q}(f)|},$$

we only need to show that

$$\hat{Q}_R(f)\hat{Q}_R(f)'' + |\hat{Q}(f)'|^2 + |\hat{Q}_I(f)||\hat{Q}_I(f)''| < 0,$$

which can be obtained by applying Lemma 4.3, Lemma 4.4, Lemma 4.5 and the triangle inequality to control these three terms $\hat{Q}_R(f)\hat{Q}_R(f)''$, $|\hat{Q}(f)'|^2$ and $|\hat{Q}_I(f)||\hat{Q}_I(f)''|$, respectively.

More precisely, the first term can be bounded by

$$\begin{aligned} &\hat{Q}_R(f)\hat{Q}_R(f)'' \\ &\leq Q_R^*(f)Q_R^*(f)'' + |\hat{Q}_R(f) - Q_R^*(f)||\hat{Q}_R(f)'' - Q_R^*(f)''| + |Q_R^*(f)||\hat{Q}_R(f)'' - Q_R^*(f)''| + |\hat{Q}_R(f) - Q_R^*(f)||Q_R^*(f)''| \\ &\leq (0.887594)(-2.24483n^2) + (28.7343X^*B^*\gamma + 82.5975B^*/X^*)(140.808n^2X^*B^*\gamma + 758.404n^2B^*/X^*) \\ &\quad + (1)(140.808n^2X^*B^*\gamma + 758.404n^2B^*/X^*) + (28.7343X^*B^*\gamma + 82.5975B^*/X^*)3.40320n^2 \\ &\leq -1.64194n^2, \end{aligned} \quad (4.11)$$

where we used the SNR condition (2.6): $X^*B^*\gamma \leq 10^{-3}$, $B^*/X^* \leq 10^{-4}$ in the last line. We now bound the second term

$$\begin{aligned} |\hat{Q}(f)'|^2 &= |\hat{Q}(f)' - Q^{*'}(f)|^2 + |Q^{*'}(f)|^2 + 2|Q^{*'}(f)||\hat{Q}(f)' - Q^{*'}(f)| \\ &\leq (44.4648nX^*B^*\gamma + 180.283nB^*/X^*)^2 + (0.821039n)^2 + 2(0.821039n)(44.4648nX^*B^*\gamma + 180.283nB^*/X^*) \\ &\leq 0.780629n^2. \end{aligned} \quad (4.12)$$

Finally, the third term can be bounded by

$$\begin{aligned} &|\hat{Q}_I(f)||\hat{Q}_I(f)''| \\ &\leq (|Q_I^*(f)| + |\hat{Q}(f) - Q^*(f)|)(|Q_I^{*''}(f)| + |\hat{Q}(f)'' - Q^{*''}(f)|) \\ &\leq (0.0183836 + (28.7343X^*B^*\gamma + 82.5975B^*/X^*))0.113197n^2 + (140.808n^2X^*B^*\gamma + 758.404n^2B^*/X^*) \\ &\leq 0.222917n^2. \end{aligned} \quad (4.13)$$

From (4.11), (4.12) and (4.13), we have

$$\hat{Q}_R(f)\hat{Q}_R(f)'' + |\hat{Q}(f)'|^2 + |\hat{Q}_I(f)| |\hat{Q}_I(f)''| \leq (-1.64194 + 0.780629 + 0.222917)n^2 < 0,$$

implying that $|\hat{Q}(f)|'' < 0$ in \mathcal{N} . This completes showing $|\hat{Q}(f)|'' < 0$ in $(-\mathcal{N}) \cup \mathcal{N}$ and

$$|\hat{Q}(f)| < 1, \quad \text{for } f \in (-\mathcal{N}) \cup \mathcal{N} \setminus \{f_0\}. \quad (4.14)$$

Next, we bound $|\hat{Q}(f)|$ in Middle Region

$$\begin{aligned} |\hat{Q}(f)| &\leq |Q^*(f)| + |Q^*(f) - Q^\lambda(f)| + |\hat{Q}(f) - Q^\lambda(f)| \\ &\leq 0.927615 + (39.3557X^*B^*\gamma + 114.323B^*/X^*) \\ &\leq 0.978403 < 1, \quad \text{for } f \in \mathcal{M}. \end{aligned} \quad (4.15)$$

Finally, we arrive at an upper bound of $|\hat{Q}(f)|$ in Far Region:

$$\begin{aligned} |\hat{Q}(f)| &\leq |Q^*(f)| + |Q^*(f) - Q^\lambda(f)| + |\hat{Q}(f) - Q^\lambda(f)| \\ &\leq 0.734123 + (66.1596X^*B^*\gamma + 162.903B^*/X^*) \\ &\leq 0.81658 < 1, \quad \text{for } f \in \mathcal{F}. \end{aligned} \quad (4.16)$$

From (4.14), (4.15) and (4.16), we obtain that $\hat{Q}(f)$ satisfies the BIP property and hence $\hat{\mathbf{q}}$ is a valid dual certificate that certifies the optimality of $\hat{\mathbf{x}} = \sum_{\ell=1}^k \hat{c}_\ell \mathbf{a}(\hat{f}_\ell)$. The uniqueness of the decomposition is also certified by $\hat{\mathbf{q}}$ implies that $\{\hat{f}_\ell\}_{\ell=1}^k = \{f_\ell^{\text{glob}}\}_{\ell=1}^k$ and $\{\hat{c}_\ell\}_{\ell=1}^k = \{c_\ell^{\text{glob}}\}_{\ell=1}^k$, i.e., $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{\text{glob}}$ are the same point.

As the final step, using Lemma 4.1, Lemma 4.2 and the triangle inequality, we have

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \leq \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\lambda\|_\infty + \|\boldsymbol{\theta}^\lambda - \boldsymbol{\theta}^*\|_\infty \leq (X^* + 35.2)\gamma_0/\sqrt{2}.$$

Then the desired results follow from the definition of the norm $\|\cdot\|_\infty$ and the fact that $\sqrt{|K''(0)|} \geq 3.289n^2$ for $n \geq 130$ by (A.2) and hence $1/\sqrt{2|K''(0)|} \leq 1/\sqrt{2(3.289)/n} \leq 0.3899/n \leq 0.4/n$. \square

5 Numerical Experiments

We present numerical results to support our theoretical findings. In particular, we first examine the phase transition curve of the rate of success in Figure 2. In preparing Figure 2, k complex coefficients c_1^*, \dots, c_k^* were generated uniformly from the unit complex circle such that $c_{\min}^* = c_{\max}^* = 1$ hence $B^* = 1$. We also generated k normalized frequencies f_1^*, \dots, f_k^* uniformly chosen from $[0, 1]$ such that every pair of frequencies are separated by at least $2.5/n$. Then the signal \mathbf{x}^* was formed according to (2.2). We created our observation \mathbf{y} by adding Gaussian noise of mean zero and variance σ^2 to the target signal \mathbf{x}^* . Let $\lambda = x\gamma_0$ (recall that $\lambda = 0.646X^*\gamma_0$ in Theorem 2.1 and hence $x = 0.646X^*$). We varied x and the Noise-to-Signal Ratio γ . For each fixed (x, γ) pair, 20 instances of the spectral line signals were generated. We then solved (2.4) for each instance and extracted the frequencies and coefficients. We declared success for an instance if i) the recovered frequency vector is within $\gamma/2n \ell_\infty$ distance of the true frequency vector \mathbf{f}^* , and ii) the recovered coefficient vector is within $2\lambda \ell_\infty$ distance of the true frequency vector \mathbf{c}^* . The rate of success for each algorithm is the proportion of successful instances.

From Figure 2, we observe that solving (2.4) is unable to identify the sinusoidal parameters if $x \leq 1$ and the performance of the method is unstable when x is around 1. When x is set to be slightly larger than 1, however, we almost always succeed in finding good estimates of the sinusoidal parameters as long as $x\gamma \leq c$ for some small constant c . This matches the findings in Theorem 2.1. Figure 2 also shows the constants in Theorem 2.1 are a bit conservative.

We also run simulations to compare the mean-squared error for our frequency estimate with those for MUSIC and the MLE, as well as the CRB. The simulation results are listed in Figure 3. We emphasize that the MLE is initialized using the true frequencies and coefficients, which are not available in practice. We focus on the case of two unknown frequencies and examine the effect of separation. We observe that the atomic norm minimization method always outperforms MUSIC, with increased performance gap when the frequencies become closer. While the MLE performs the best, its initialization is not practical.

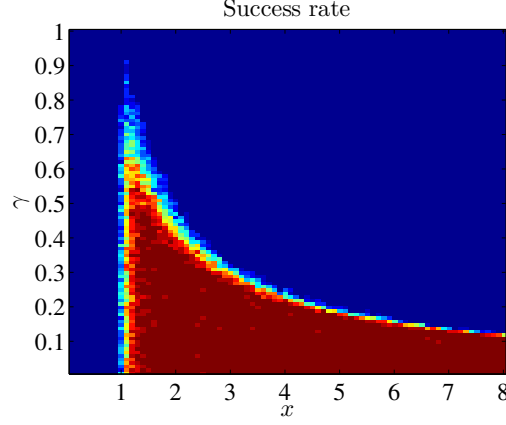


Figure 2: Rate of success for line spectral estimation by solving the atomic norm regularized program (2.4).

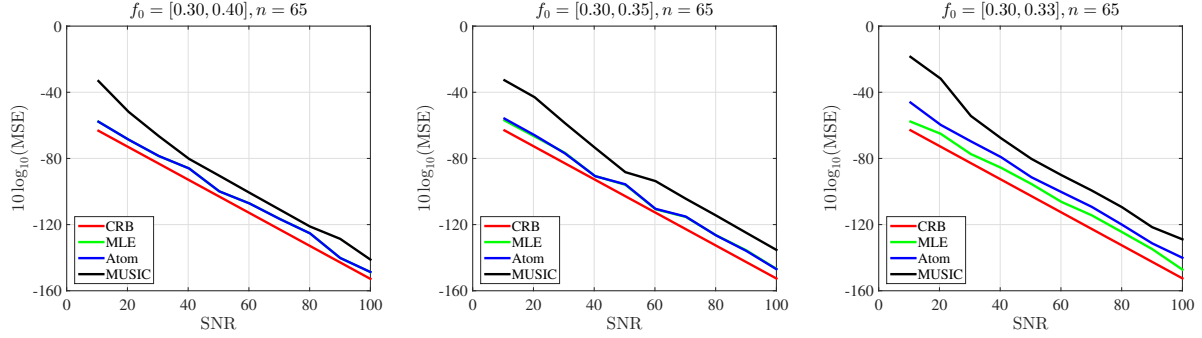


Figure 3: Performance comparison: Atomic norm minimization (2.4) (labeled as “Atom”), MUSIC, MLE initialized by the true parameters, and the CRB.

6 Conclusions

This work considers the problem of approximately estimating the frequencies and coefficients of a superposition of complex sinusoids in white noise. By using a primal-dual witness construction, we established theoretical performance guarantees for atomic norm minimization algorithm in line spectral parameter estimation. The obtained error bounds match the Cramér-Rao lower bound up to a logarithmic factor. The relationship between resolution (separation of frequencies) and precision or accuracy of the estimator is highlighted. Our analysis also reveals that the atomic norm minimization can be viewed as a convex way to solve a ℓ_1 -norm regularized, nonlinear and nonconvex least-squares problem to global optimality.

Appendices

Appendix A Jackson Kernel

For any integer $M > 0$, the Jackson kernel, also known as the squared Fejér kernel, is defined by [4, Eq. (IV.2)] or [1, Eq. (2.3) with $M = f_c/2 + 1$]

$$K(f) = \left[\frac{\sin(\pi M f)}{M \sin(\pi f)} \right]^4. \quad (\text{A.1})$$

The Jackson kernel shows up in the construction of dual polynomials that satisfy the Boundedness and Interpolation properties. The choice of the Jackson kernel is due to its nice properties as easily seen from its graph: it attains one at the peak, and quickly decrease to zero. Candès and Fernandez-Granda showed in [1] that as long as the frequencies composing a signal satisfy certain separation condition, then a dual polynomial can be constructed as a linear combination of shifted copies of the Jackson kernel and its first-order derivative to certify that the decomposition achieves the signal's atomic norm.

We use $K'(f)$, $K''(f)$, $K'''(f)$ to denote respectively the first, second, and third order derivatives of the Jackson kernel and more generally $K^{(\ell)}(f)$ the ℓ th order derivative. We will frequently use the second order derivative of the Jackson kernel evaluated at zero $K''(0)$, whose value is [4, Above Eq. (IV.5)]

$$K''(0) = -\frac{4\pi^2(M^2 - 1)}{3} = -\frac{\pi^2(n^2 - 4)}{3}.$$

Here we used the convention that $n = 2M$. Then its absolute value $|K''(0)|$ (denoted by τ) falls into the interval

$$|K''(0)| \in \left[\left(\frac{\pi^2}{3} - \frac{4\pi^2}{3(130)^2} \right) n^2, \left(\frac{\pi^2}{3} \right) n^2 \right], \quad \text{for } n \geq 130.$$

For ease of exposition, we give an explicit lower bound on $|K''(0)|$ (which is valid for any $n \geq 130$):

$$\tau := |K''(0)| \geq \left(\frac{\pi^2}{3} - \frac{4\pi^2}{3(130)^2} \right) n^2 \geq 3.289n^2, \quad \text{for } n \geq 130. \quad (\text{A.2})$$

At a high-level, the purpose of introducing $\tau = |K''(0)|$ is to normalize the second order derivative of the Jackson kernel to 1 at $f = 0$.

A.1 Decomposing the Jackson Kernel

The Jackson kernel admits the following decomposition [4]

$$K(f_2 - f_1) = \left[\frac{\sin(\pi M(f_2 - f_1))}{M \sin(\pi(f_2 - f_1))} \right]^4 = \mathbf{a}(f_1)^H \mathbf{Z} \mathbf{a}(f_2) = \frac{1}{M} \sum_{j=-2M}^{2M} g_M(j) e^{-i2\pi j(f_2 - f_1)},$$

where $M = n/2$ and \mathbf{Z} is an $n \times n$ diagonal matrix whose diagonal entries are given by $[\mathbf{Z}]_{\ell\ell} = \frac{g_M(\ell)}{M}$ with

$$g_M(\ell) = \frac{1}{M} \sum_{k=\max(\ell-M, -M)}^{\min(\ell+M, M)} \left(1 - \left| \frac{k}{M} \right| \right) \left(1 - \left| \frac{\ell - k}{M} \right| \right) \geq 0, \ell = -2M, \dots, 2M, \quad (\text{A.3})$$

the convolution of two discrete triangle functions scaled by $1/M$. The weighting function $g_M(\ell)$ attains its peak at zero and

$$g_M(0) = \frac{1}{M} \sum_{k=-M}^M \left(1 - \left| \frac{k}{M} \right| \right)^2 = \frac{2}{3} + \frac{2}{M^2} \stackrel{\textcircled{1}}{\leq} \frac{2}{3} + \frac{2}{65^2} \leq 0.667,$$

where ① holds for $M \geq 65$ (or $n \geq 130$) by noting that $2/M^2$ is a decreasing function of M . Using the definition of \mathbf{Z} , we bound $\|\mathbf{Z}\|_{\infty, \infty}$ as

$$\|\mathbf{Z}\|_{\infty, \infty} = \max_{-2M \leq j \leq 2M} \frac{g_M(j)}{M} = \frac{g_M(0)}{M} \leq \frac{0.667}{M}, \quad \text{for } n \geq 130. \quad (\text{A.4})$$

A.2 Decomposing the Jackson Kernel Matrices

We frequently use matrices formed by sampling the Jackson kernel and its derivatives at appropriate frequencies. Given a finite set of frequencies $T = \{f_\ell\}_{\ell=1}^k$ (or its vector form $\mathbf{f} \in \mathbb{R}^k$), we define

$$\begin{aligned} \mathbf{D}_0(\mathbf{f}) &:= [K(f_m - f_n)]_{1 \leq n \leq k, 1 \leq m \leq k} = \mathbf{A}(\mathbf{f})^H \mathbf{Z} \mathbf{A}(\mathbf{f}); \\ \mathbf{D}_1(\mathbf{f}) &:= [K'(f_m - f_n)]_{1 \leq n \leq k, 1 \leq m \leq k} = \mathbf{A}(\mathbf{f})^H \mathbf{Z} \mathbf{A}'(\mathbf{f}) = -\mathbf{A}'(\mathbf{f})^H \mathbf{Z} \mathbf{A}(\mathbf{f}); \\ \mathbf{D}_2(\mathbf{f}) &:= [K''(f_m - f_n)]_{1 \leq n \leq k, 1 \leq m \leq k} = -\mathbf{A}'(\mathbf{f})^H \mathbf{Z} \mathbf{A}'(\mathbf{f}) = \mathbf{A}''(\mathbf{f})^H \mathbf{Z} \mathbf{A}(\mathbf{f}) = \mathbf{A}(\mathbf{f})^H \mathbf{Z} \mathbf{A}''(\mathbf{f}), \end{aligned} \quad (\text{A.5})$$

where

$$\mathbf{A}(\mathbf{f}) := [\mathbf{a}(f_1), \dots, \mathbf{a}(f_k)], \quad \mathbf{A}'(\mathbf{f}) := i2\pi \text{diag}(\mathbf{n}) \mathbf{A}(\mathbf{f}), \quad \mathbf{A}''(\mathbf{f}) := (i2\pi \text{diag}(\mathbf{n}))^2 \mathbf{A}(\mathbf{f})$$

with $\mathbf{n} = [-n, -n+1, \dots, 0, \dots, n-1, n]^T$. More generally, the kernel matrix $\mathbf{D}_\ell(\mathbf{f}) := [K^{(\ell)}(f_m - f_n)]_{1 \leq n \leq k, 1 \leq m \leq k}$ satisfies the factorization

$$\mathbf{D}_\ell(\mathbf{f}) = (-1)^j \mathbf{A}^{(j)}(\mathbf{f})^H \mathbf{Z} \mathbf{A}^{(\ell-j)}(\mathbf{f}), \quad \text{for } j \leq \ell, \quad (\text{A.6})$$

where $\mathbf{A}^{(\ell)}(\mathbf{f})$ represents the ℓ th order derivative of the matrix $\mathbf{A}(\mathbf{f})$:

$$\mathbf{A}^{(\ell)}(\mathbf{f}) = (i2\pi \text{diag}(\mathbf{n}))^\ell \mathbf{A}(\mathbf{f}).$$

Similarly, we define the cross kernel matrices with respect to the frequency pair $(\mathbf{f}^1, \mathbf{f}^2)$ or $(\{f_\ell^1\}, \{f_\ell^2\})$ as

$$\mathbf{D}_\ell(\mathbf{f}^1, \mathbf{f}^2) = [K^{(\ell)}(f_m^2 - f_n^1)]_{1 \leq n \leq k, 1 \leq m \leq k}, \quad \text{for } \ell = 0, 1, 2.$$

We can also express $\mathbf{D}_\ell(\mathbf{f}^1, \mathbf{f}^2)$ in factorization forms

$$\mathbf{D}_\ell(\mathbf{f}^1, \mathbf{f}^2) = (-1)^j \mathbf{A}^{(j)}(\mathbf{f}^1)^H \mathbf{Z} \mathbf{A}^{(\ell-j)}(\mathbf{f}^2), \quad \text{for } j \leq \ell. \quad (\text{A.7})$$

A.3 Bounding the Jackson Kernel

The following lemma provides a set of bounds on the ℓ th derivative of the Jackson kernel for $\ell \in \{0, 1, 2, 3, 4\}$.

Lemma A.1 (Bounds on $|K^{(\ell)}|$). *For $\ell \in \{0, 1, 2, 3, 4\}$, let $K^{(\ell)}$ be the ℓ th derivative of K ($K = K^{(0)}$). Define $s(f)$ as a symmetric and periodic function with period 1 and $s(f) = \frac{1}{Mf(3-4f^2)}$ for $f \in (0, 1/2]$. Then for $f \in (0, 1/2]$, we have*

$$\begin{aligned} |K^{(0)}(f)| &\leq B_0(f) := s^4(f), \\ |K^{(1)}(f)| &\leq B_1(f) := 2\pi M s^4(f) \left(\frac{3\sqrt{3}}{8} + 2s(f) \right), \\ |K^{(2)}(f)| &\leq B_2(f) := (2\pi M)^2 s^4(f) \left(1 + \frac{3\sqrt{3}}{2} s(f) + 5s^2(f) \right), \\ |K^{(3)}(f)| &\leq B_3(f) := (2\pi M)^3 s^4(f) \left(c_1 + 6s(f) + \frac{45\sqrt{3}}{8} s^2(f) + 15s^3(f) \right), \\ |K^{(4)}(f)| &\leq B_4(f) := (2\pi M)^4 s^4(f) \left(\frac{5}{2} + c_2 s(f) + 30s^2(f) + \frac{45\sqrt{3}}{2} s^3(f) + \frac{105}{2} s^4(f) \right), \end{aligned}$$

where

$$c_1 = \frac{1}{2} \left(\sin \left(2 \tan^{-1} \left(\sqrt{\frac{1}{5} (\sqrt{129} + 12)} \right) \right) - 2 \sin \left(4 \tan^{-1} \left(\sqrt{\frac{1}{5} (\sqrt{129} + 12)} \right) \right) \right),$$

$$c_2 = -4 \sin \left(2 \tan^{-1} \left(\sqrt{\frac{1}{5} (\sqrt{129} + 12)} \right) \right) \left(4 \cos \left(2 \tan^{-1} \left(\sqrt{\frac{1}{5} (\sqrt{129} + 12)} \right) \right) - 1 \right).$$

Furthermore, $B_\ell(f)$ is decreasing in f on $(0, 1/2]$ and $B_\ell(\Omega - f) + B_\ell(\Omega + f)$ is increasing in f for any positive Ω such that $\Omega > f$ and $\Omega + f \leq 1/2$.

Proof. We need the following elementary bound on the sine function for $f \in [0, \frac{1}{2}]$:

$$\sin(\pi f) \geq f(3 - 4f^2). \quad (\text{A.8})$$

Clearly, a consequence is $\frac{1}{M|\sin(\pi f)|} \leq s(f)$, $f \in [-\frac{1}{2}, \frac{1}{2}] \setminus \{0\}$. We use this fact together with explicit expressions for $K^{(\ell)}(f)$ to develop upper bounds.

When $\ell = 0$,

$$|K(f)| = \left| \frac{\sin(\pi M f)}{M \sin(\pi f)} \right|^4 \leq s^4(f).$$

When $\ell = 1$,

$$K^{(1)}(f) = \frac{2\pi M}{(M \sin(\pi f))^4} \frac{1}{M} (-2 \cot(\pi f) \sin^4(\pi f M) + 2 \sin^3(\pi f M) \cos(\pi f M) M)$$

implying

$$|K^{(1)}(f)| \leq 2\pi M s^4(f) \left(\frac{3\sqrt{3}}{8} + 2s(f) \right),$$

since $\max_f |2 \cos(\pi f M) \sin(\pi f M)^3| \leq \frac{3\sqrt{3}}{8}$.

When $\ell = 2$,

$$K^{(2)}(f) = \frac{(2\pi M)^2}{(M \sin(\pi f))^4} \frac{1}{M^2} \times$$

$$\left((2 \cos(2\pi f) + 3) \csc^2(\pi f) \sin^4(\pi f M) - 8 \cot(\pi f) \sin^3(\pi f M) \cos(\pi f M) M + \sin^2(\pi f M) (2 \cos(2\pi f M) + 1) M^2 \right)$$

implying

$$|K^{(2)}(f)| \leq (2\pi M)^2 s^4(f) \left(1 + \frac{3\sqrt{3}}{2} s(f) + 5s^2(f) \right),$$

where we used $\max_f |8 \sin^3(\pi f M) \cos(\pi f M)| = \frac{3\sqrt{3}}{2}$ and $\max_f |\sin^2(\pi f M) (2 \cos(2\pi f M) + 1)| = 1$.

When $\ell = 3$,

$$K^{(3)}(f) = \frac{(2\pi M)^3}{(M \sin(\pi f))^4} \frac{1}{M^3} \times$$

$$\begin{aligned} & - (4 \cos(2\pi f) + 11) \cot(\pi f) \csc^2(\pi f) \sin^4(\pi f M) \\ & + 6(2 \cos(2\pi f) + 3) \csc^2(\pi f) \sin^3(\pi f M) \cos(\pi f M) M \\ & - 6 \cot(\pi f) \sin^2(\pi f M) (2 \cos(2\pi f M) + 1) \sin(4\pi f M) M^2 - \frac{1}{2} \sin(2\pi f M) M^3 \end{aligned}$$

implying

$$|K^{(3)}(f)| \leq (2\pi M)^3 s^4(f) \left(c_1 + 6s(f) + \frac{45\sqrt{3}}{8} s^2(f) + 15s^3(f) \right),$$

by recognizing the following upper bounds:

$$\begin{aligned} \max_{f \in [0, 1/2]} |(4 \cos(2\pi f) + 11) \cos(\pi f)| &= 15, & \max_f |6 \sin^2(\pi f M)(2 \cos(2\pi f M) + 1)| &= 6, \\ \max_{f \in [0, 1/2]} |6(2 \cos(2\pi f) + 3)| &= 30, & \max_f |\sin(4\pi f M) - (1/2) \sin(2\pi f M)| &= c_1, \\ \max_f |\sin^3(\pi f M) \cos(\pi f M)| &= 3\sqrt{3}/16. \end{aligned}$$

When $\ell = 4$,

$$\begin{aligned} K^{(4)}(f) &= \frac{(2\pi M)^4}{(M \sin(\pi f))^4} \frac{1}{M^4} \times \left(\frac{1}{2} (49 \cos(2\pi f) + 4 \cos(4\pi f) + 52) \csc^4(\pi f) \sin^4(\pi f M) \right. \\ &\quad - 8(4 \cos(2\pi f) + 11) \cot(\pi f) \csc^2(\pi f) \sin^3(\pi f M) \cos(\pi f M) M \\ &\quad + 6(2 \cos(2\pi f) + 3) \csc^2(\pi f) \sin^2(\pi f M)(2 \cos(2\pi f M) + 1) M^2 \\ &\quad - 4 \cot(\pi f) \sin(2\pi f M)(4 \cos(2\pi f M) - 1) M^3 \\ &\quad \left. + (2 \cos(4\pi f M) - \frac{1}{2} \cos(2\pi f M)) M^4 \right) \end{aligned}$$

implying

$$|K^{(4)}(f)| \leq (2\pi M)^4 s^4(f) \left(\frac{5}{2} + c_2 s(f) + 30 s^2(f) + \frac{45\sqrt{3}}{2} s^3(f) + \frac{105}{2} s^4(f) \right),$$

which follows from the following upper bounds:

$$\begin{aligned} \max_{f \in [0, 1/2]} \frac{1}{2} (49 \cos(2\pi f) + 4 \cos(4\pi f) + 52) &= 105/2, & \max_f |\sin^2(\pi f M)(2 \cos(2\pi f M) + 1)| &= 1, \\ \max_{f \in [0, 1/2]} |8(4 \cos(2\pi f) + 11) \cos(\pi f)| &= 120, & \max_f |4 \sin(2\pi f M)(4 \cos(2\pi f M) - 1)| &= c_2, \\ \max_f |\sin^3(\pi f M) \cos(\pi f M)| &= 3\sqrt{3}/16, & \max_f |2 \cos(4\pi f M) - 1/2 \cos(2\pi f M)| &= 5/2, \\ \max_{f \in [0, 1/2]} |6(2 \cos(2\pi f) + 3)| &= 30. \end{aligned}$$

Finally, $s(f)$ is nonnegative and is decreasing in $(0, 1/2]$ since $s'(f)$ is negative on $(0, 1/2)$. Therefore, the k th power $s^k(f)$ is decreasing in $(0, 1/2]$, which further implies that $B_\ell(f)$, $\ell = 0, 1, 2, 3, 4$ is decreasing in $(0, 1/2]$. In addition, since $s(f)$ is convex in $(0, 1/2]$, $s^k(f)$ is also convex as a consequence of the composition rule of convex and monotonic functions. Combining the convex and decreasing property of $s^k(f)$ on $(0, 1/2]$ and then applying arguments similar to those in [1, Lemma 2.6], we conclude that $B_\ell(\Omega - f) + B_\ell(\Omega + f)$ is increasing in f for any positive Ω such that $\Omega > f$ and $\Omega + f \leq 1/2$. \square

A.4 Bounding the Sums of the Jackson Kernel

Without loss of generality, we assume $0 \in T$ and develop bounds on $\sum_{f_i \in T \setminus \{0\}} |K^{(\ell)}(f - f_i)|$, $\ell \in \{0, 1, 2, 3, 4\}$ when f lives in a neighborhood around 0. It is easy to verify the following lemma based on the properties of $|K^{(\ell)}(f)|$, $\ell = 0, 1, 2, 3, 4$ in Lemma A.1. The proof parallels that of [1, Lemma 2.7] and is omitted here.

Lemma A.2. *Suppose $0 \in T$ and f_+ is the smallest positive frequency in T . Let $\Delta := \Delta(T) \geq \Delta_{\min}$ and $f \in [0, \bar{f}]$ where $\bar{f} \leq f_+/2$. Then for $\ell \in \{0, 1, 2, 3, 4\}$,*

$$\sum_{f_i \in T \setminus \{0\}} |K^{(\ell)}(f - f_i)| \leq F_\ell(\Delta, f) := F_\ell^+(\Delta, f) + F_\ell^-(\Delta, f) \leq F_\ell(\Delta_{\min}, \bar{f})$$

with

$$\begin{aligned} F_\ell^+(\Delta, f) &= \max \left\{ \max_{\Delta \leq \xi \leq 3\Delta_{\min}} |K^{(\ell)}(f - \xi)|, B_\ell(3\Delta_{\min} - f) \right\} + \sum_{j=2}^{\lfloor \frac{1}{2\Delta_{\min}} \rfloor} B_\ell(j\Delta_{\min} - f), \\ F_\ell^-(\Delta, f) &= \max \left\{ \max_{\Delta \leq \xi \leq 3\Delta_{\min}} |K^{(\ell)}(\xi)|, B_\ell(3\Delta_{\min}) \right\} + \sum_{j=2}^{\lfloor \frac{1}{2\Delta_{\min}} \rfloor} B_\ell(j\Delta_{\min} + f). \end{aligned}$$

$F_\ell(\Delta, f)$ is decreasing in Δ . When Δ is fixed as Δ_{\min} , $F_\ell(\Delta_{\min}, f)$ is increasing in f .

The following lemma provides bounds on $\sum_{f_i \in T} |K^{(\ell)}(f - f_i)|$ for $\ell \in \{0, 1, 2, 3, 4\}$ and is a direct consequence of the decreasing property of $B_\ell(\cdot)$.

Lemma A.3. *Suppose $0 \in T$, f_+ is the smallest positive frequency in T and $f \in [\underline{f}, f_+ - \bar{f}]$. Then for $\ell \in \{0, 1, 2, 3, 4\}$,*

$$\sum_{f_i \in T} |K^{(\ell)}(f - f_i)| \leq W_\ell(\underline{f}, \bar{f}) := \sum_{j=0}^{\lfloor \frac{1}{2\Delta_{\min}} \rfloor} B_\ell(j\Delta_{\min} + \underline{f}) + \sum_{j=0}^{\lfloor \frac{1}{2\Delta_{\min}} \rfloor} B_\ell(j\Delta_{\min} + \bar{f}).$$

A.5 Numerical Bounds on the Jackson Kernel Sums

Suppose $0 \in T$. Then by Lemma A.2 we can bound $\sum_{f_j \in T \setminus \{0\}} |K^{(\ell)}(f - f_j)|$ for $f \in [0, \bar{f}]$ as:

$$\sum_{f_j \in T \setminus \{0\}} |K^{(\ell)}(f - f_j)| \leq F_\ell(\Delta_{\min}, \bar{f}).$$

We list the values of $F_\ell(\Delta_{\min}, \bar{f})$ for different \bar{f} in Table A.5.

We can use Lemma A.3 to bound $\sum_{f_j \in T} |K^{(\ell)}(f - f_j)|$ for $f \in [\underline{f}, f_+ - \bar{f}]$ as

$$\sum_{f_j \in T} |K^{(\ell)}(f - f_j)| \leq W_\ell(\underline{f}, \bar{f}).$$

We list the values of $W_\ell(\underline{f}, \bar{f})$ for different \underline{f}, \bar{f} in Table A.5.

Finally, we list several numerical upper bounds on $|K^{(\ell)}(f)|$ and $K''(f)$ over different intervals in Table A.5, which directly follow from [1, equations (2.21)-(2.24)] and numerical computations.

f	$F_0(2.5/n, f)$	$F_1(2.5/n, f)$	$F_2(2.5/n, f)$	$F_3(2.5/n, f)$	$F_4(2.5/n, f)$
0	0.00755	$0.01236n$	$0.05610n^2$	$0.28687n^3$	
$0.002/n$	0.00755	$0.01236n$	$0.05610n^2$	$0.28687n^3$	
$0.24/n$	0.00757	$0.01241n$	$0.05637n^2$	$0.28838n^3$	
$0.2404/n$	0.00757	$0.01241n$	$0.05637n^2$	$0.28838n^3$	$1.671n^4$
$0.75/n$	0.00772	$0.01450n$			
$0.7504/n$	0.00772	$0.01454n$	$0.12675n^2$		

Table 3: Numerical upper bounds on $F_\ell(2.5/n, f)$.

f_1	f_2	$W_0(f_1, f_2)$	$W_1(f_1, f_2)$	$W_2(f_1, f_2)$
$0.7496/n$	$1.25/n$	0.71059	$5.2265n$	$48.033n^2$
$0.75/n$	$1.25/n$	0.70859	$5.2084n$	

Table 4: Numerical upper bounds on $W_\ell(f_1, f_2)$.

A.6 Controlling the Jackson Kernel Matrices

In this section, we derive several consequences of the joint frequency-coefficient vector $\boldsymbol{\theta} = (\mathbf{f}, \mathbf{u}, \mathbf{v})$ living in the neighborhood \mathcal{N}^* of the true joint frequency-coefficient vector $\boldsymbol{\theta}^* = (\mathbf{f}^*, \mathbf{u}^*, \mathbf{v}^*)$. Recall that \mathcal{N}^* contains all $\boldsymbol{\theta}$ that is close to $\boldsymbol{\theta}^*$ in the ℓ_∞ norm:

$$\mathcal{N}^* = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq X^* \gamma_0 / \sqrt{2}\}. \quad (\text{A.9})$$

f	$ K(f) $	$ K'(f) $	$ K''(f) $	$ K'''(f) $	$ K''''(f) $	$K''(f)$
$[0, 0.002/n]$	1	$0.00658n$	$3.290n^2$	$0.0649394n^3$		
$[0, 0.24/n]$	1	$0.789569n$	$3.290n^2$	$7.79273n^3$		$-2.35084n^2$
$[0, 0.2404/n]$	1	$0.790885n$	$3.290n^2$	$7.80572n^3$	$29.2227n^4$	
$[0.2396/n, 0.7504/n]$	0.90951	$2.46872n$	$3.290n^2$			

Table 5: Numerical upper bounds on $|K^{(\ell)}(f)|$ and $K''(f)$.

Recall that the weighted ℓ_∞ norm $\|\cdot\|_\infty$ is defined by $\|(\mathbf{f}, \mathbf{u}, \mathbf{v})\|_\infty := \|(\mathbf{S}\mathbf{f}, \mathbf{u}, \mathbf{v})\|_\infty$ with $\mathbf{S} := \sqrt{\tau} \text{diag}(|\mathbf{c}^*|)$. We remark that all the results in this section still hold for the bigger neighborhood $\hat{\mathcal{N}}$ defined by replacing X^* with $\hat{X} = X^* + 35.2$. Indeed, for the results to hold, the key requirement on $\boldsymbol{\theta}$ is $\|\mathbf{f} - \mathbf{f}^*\|_\infty \leq 0.002/n$. This condition holds for both regions because as we will show later

$$\|\mathbf{f} - \mathbf{f}^*\|_\infty \leq \begin{cases} 0.4X^*\gamma & \text{for } \boldsymbol{\theta} \in \mathcal{N}^*, \\ 0.4\hat{X}\gamma & \text{for } \boldsymbol{\theta} \in \hat{\mathcal{N}}. \end{cases}$$

Invoking the SNR condition (2.6), we conclude that the two upper bounds are much smaller than $0.002/n$ in both cases.

Our first set of results bound the distances between the parameters in $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}$: for each $j = 1, \dots, k$

$$\begin{aligned} \frac{|c_j - c_j^*|}{|c_j^*|} &\stackrel{\textcircled{1}}{\leq} X^*\gamma, \\ \frac{|c_j|}{|c_j^*|} &\stackrel{\textcircled{2}}{\leq} 1 + X^*\gamma, \\ \left| (|c_j|/|c_j^*|)^2 - 1 \right| &\stackrel{\textcircled{3}}{\leq} X^*\gamma(2 + X^*\gamma), \\ |f_j - f_j^*| &\stackrel{\textcircled{4}}{\leq} X^*\gamma/\sqrt{2\tau} \stackrel{\textcircled{5}}{\leq} 0.4X^*\gamma/n. \end{aligned} \tag{A.10}$$

For $\textcircled{1}$ to hold, first note $\|\mathbf{u} - \mathbf{u}^*\|_\infty \leq X^*\gamma_0/\sqrt{2}$ and $\|\mathbf{v} - \mathbf{v}^*\|_\infty \leq X^*\gamma_0/\sqrt{2}$ by (A.9). Also note

$$\begin{aligned} \|\mathbf{c} - \mathbf{c}^*\|_\infty^2 &= \max_\ell [(u_\ell - u_\ell^*)^2 + (v_\ell - v_\ell^*)^2] \\ &\leq \max_\ell (u_\ell - u_\ell^*)^2 + \max_\ell (v_\ell - v_\ell^*)^2 \\ &= \|\mathbf{u} - \mathbf{u}^*\|_\infty^2 + \|\mathbf{v} - \mathbf{v}^*\|_\infty^2 \leq 2(X^*\gamma_0/\sqrt{2})^2 = (X^*\gamma_0)^2. \end{aligned}$$

Finally $\textcircled{1}$ follows since $\max_j |c_j - c_j^*|/|c_j^*| \leq \|\mathbf{c} - \mathbf{c}^*\|_\infty/c_{\min}^*$ and $\gamma = \gamma_0/c_{\min}^*$. After we show $\textcircled{1}$, $\textcircled{2}$ follows from $|c_j|/|c_j^*| = |c_j - c_j^* + c_j^*|/|c_j^*|$ and the triangle inequality. $\textcircled{3}$ follows from $|(|c_j|/|c_j^*|)^2 - 1| = |(|c_j|/|c_j^*| + 1)(|c_j|/|c_j^*| - 1)|$. $\textcircled{4}$ follows from the definition of the ℓ_∞ norm:

$$\begin{aligned} \|S(\mathbf{f} - \mathbf{f}^*)\|_\infty &\leq X^*\gamma_0/\sqrt{2} \\ \implies \|\sqrt{\tau} \text{diag}(|\mathbf{c}^*|)(\mathbf{f} - \mathbf{f}^*)\|_\infty &\leq X^*\gamma_0/\sqrt{2} \\ \implies |f_j - f_j^*| &\leq X^*\gamma_0/|c_j^*|/\sqrt{2\tau}, \forall j \\ \implies |f_j - f_j^*| &\leq X^*\gamma_0/c_{\min}^*/\sqrt{2\tau} = X^*\gamma/\sqrt{2\tau}, \forall j. \end{aligned}$$

Finally $\textcircled{5}$ holds due to the fact that $\tau \geq 3.289n^2$ for $n \geq 130$ by (A.2) and hence

$$1/\sqrt{2\tau} \leq 1/\sqrt{2(3.289)}/n \leq 0.3899/n \leq 0.4/n.$$

Next, we present the second class of results that quantify the well-conditionedness of the Jackson kernel matrices $\mathbf{D}_\ell(\mathbf{f})$, $\ell = 0, 1, 2$. Such results are instrumental to dual certificate construction [1]. Since the minimal separation $\Delta(T)$ is a key quantity affecting the well-conditionedness, we first show that those frequencies $T^\lambda := \{f_\ell^\lambda\}$ and $\hat{T} := \{\hat{f}_\ell\}$ in Lemma 4.1 and Lemma 4.2 satisfy a separation condition, provided $T^* = \{f_\ell^*\}$ satisfy a slightly stronger separation condition. The proof is given in Appendix K.

Lemma A.4. *Let the separation condition (2.5) and the SNR condition (2.6) hold. Then both the frequencies $T^\lambda = \{f_\ell^\lambda\}$ returned by the first fixed point map (4.4) and the frequencies $\hat{T} = \{\hat{f}_\ell\}$ generated by the second fixed point map (4.6) have minimal separations at least $2.5/n$. Furthermore, the intermediate frequencies defined by $\tilde{T} = \{\tilde{f}_\ell\}_{\ell=1}^k$ with each $\tilde{f}_\ell \in [f_\ell^*, f_\ell^\lambda]$ or $[f_\ell^\lambda, f_\ell^*]$ and the second intermediate frequencies $\tilde{T}^\lambda := \{\tilde{f}_\ell\}_{\ell=1}^k$ with each $\tilde{f}_\ell \in [f_\ell^\lambda, \hat{f}_\ell]$ or $[\hat{f}_\ell, f_\ell^\lambda]$ also have minimal separations at least $2.5/n$:*

$$\min\{\Delta(T^\lambda), \Delta(\tilde{T}), \Delta(\hat{T}), \Delta(\tilde{T}^\lambda)\} \geq 2.5/n.$$

Now we are ready to provide numerical bounds related to the well-conditionedness of the Jackson kernel matrices $\mathbf{D}_\ell(\mathbf{f})$, $\ell = 0, 1, 2$:

$$\begin{aligned} \|\mathbf{I} - \mathbf{D}_0(\mathbf{f})\|_{\infty, \infty} &\stackrel{\textcircled{1}}{\leq} F_0(2.5/n, 0) \stackrel{\textcircled{4}}{\leq} 0.00755, \\ \|\mathbf{D}_1(\mathbf{f})/\sqrt{\tau}\|_{\infty, \infty} &\stackrel{\textcircled{2}}{\leq} F_1(2.5/n, 0)/\sqrt{\tau} \stackrel{\textcircled{5}}{\leq} 0.01236n/\sqrt{\tau} \leq 0.00682, \\ \|\mathbf{I} - (-\mathbf{D}_2(\mathbf{f})/\tau)\|_{\infty, \infty} &\stackrel{\textcircled{3}}{\leq} F_2(2.5/n, 0)/\tau \stackrel{\textcircled{6}}{\leq} 0.05610n^2/\tau \leq 0.0171, \end{aligned} \quad (\text{A.11})$$

where ①, ② and ③ follow because the diagonal entries of these kernel matrices are given by $[\mathbf{D}_0(\mathbf{f}, \mathbf{f})]_{\ell, \ell} = K(0) = 1$, $[\mathbf{D}_1(\mathbf{f}, \mathbf{f})]_{\ell, \ell} = K'(0) = 0$ and $[\mathbf{D}_2(\mathbf{f}, \mathbf{f})]_{\ell, \ell} = K''(0) = -\tau$ [4, Section IV.A]. Hence, it suffices to compute $\sum_{f_i \in T \setminus \{\zeta\}} |K^{(\ell)}(\zeta - f_i)|$ for $\zeta \in T$ which can be bounded by $F_\ell(2.5/n, 0)$ according to Lemma A.2 since $\Delta(T) \geq 2.5/n$ by Lemma A.4. The inequalities ④, ⑤ and ⑥ follow from the upper bounds on $F_\ell(2.5/n, 0)$ in Table A.5 and the fact that $\tau \geq 3.289n^2$ for $n \geq 130$ by (A.2).

To control the $\ell_{\infty, \infty}$ distance between two kernel matrices, say $\mathbf{D}_0(\mathbf{f})$ and $\mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)$, we apply the mean value theorem and Lemma A.2:

$$\begin{aligned} \|\mathbf{D}_0(\mathbf{f}) - \mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)\|_{\infty, \infty} &\stackrel{\textcircled{1}}{=} \|\mathbf{D}_0(f_1, \mathbf{f}) - \mathbf{D}_0(f_1, \mathbf{f}^*)\|_1 \\ &\leq \sum_{\ell} |K(f_\ell - f_1) - K(f_\ell^* - f_1)| \\ &\stackrel{\textcircled{2}}{=} \sum_{\ell} |K'(\tilde{f}_\ell - f_1)(f_\ell - f_\ell^*)| \\ &\leq (|K'(\tilde{f}_1 - f_1)| + \sum_{\ell \neq 1} |K'(\tilde{f}_\ell - f_1)|) \|\mathbf{f} - \mathbf{f}^*\|_\infty \\ &\stackrel{\textcircled{3}}{\leq} (F_1(2.5/n, 0.002/n) + \max_{f \in [0, 0.002/n]} |K'(f)|) \|\mathbf{f} - \mathbf{f}^*\|_\infty \\ &\stackrel{\textcircled{4}}{\leq} (0.01236n + 0.00658n)(0.4X^*\gamma/n) = 0.00758X^*\gamma, \end{aligned} \quad (\text{A.12})$$

where ① follows since by rearranging indices if necessary, we can assume without loss of generality that the maximum absolute row sum of $\mathbf{D}_0(\mathbf{f}) - \mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)$ happens at the first row; ② holds because we applied the mean value theorem for some \tilde{f}_ℓ between f_ℓ and f_ℓ^* ; ③ follows from the monotonic property of $F_\ell(2.5/n, f)$ in Lemma A.2 by taking into account that $\Delta(\tilde{T}) \geq 2.5/n$ (by Lemma A.4) and $\|\tilde{\mathbf{f}} - \mathbf{f}\|_\infty \leq \|\mathbf{f}^* - \mathbf{f}\|_\infty \leq 0.4X^*\gamma/n \leq 0.002/n$. ④ follows from the upper bounds on $F_1(2.5/n, 0.002/n)$ in Table A.5 and $\max_{f \in [0, 0.002/n]} |K'(f)|$ in Table A.5.

Applying the similar arguments as the step ③, we can get a more general result as follows

Lemma A.5. Let an arbitrary cluster of points $T := \{f_j\}$ satisfy the separation condition of $\Delta(T) \geq 2.5/n$. Assume $\underline{f} \leq |f - f_r| \leq \bar{f}$ for an arbitrary $f_r \in T$. Then,

$$\sum_j |K^{(\ell)}(f_j - f)| \leq F_\ell(2.5/n, \bar{f}) + \max_{f \in [\underline{f}, \bar{f}]} |K^{(\ell)}(f)|. \quad (\text{A.13})$$

To control $\|\mathbf{D}_\ell(\mathbf{f}, \mathbf{f}^*) - \mathbf{D}_\ell(\mathbf{f})\|_{\infty, \infty}$ in a similar manner for $\ell = 1, 2$, we note that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq X^* \gamma_0 / \sqrt{2}$ and both T and \tilde{T} are well-separated: $\Delta(T) \geq 2.5/n$ and $\Delta(\tilde{T}) \geq 2.5/n$ with \tilde{T} composed of certain “middle” frequencies $\tilde{f}_\ell \in [f_\ell, f_\ell^*]$ or $[f_\ell^*, f_\ell]$. Then using Lemma A.5, we upperbound $\|\mathbf{D}_\ell(\mathbf{f}, \mathbf{f}^*) - \mathbf{D}_\ell(\mathbf{f})\|_{\infty, \infty}$ as follows

$$\begin{aligned} \frac{1}{\sqrt{\tau}} \|\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) - \mathbf{D}_1(\mathbf{f})\|_{\infty, \infty} &\stackrel{\textcircled{1}}{\leq} 1/\sqrt{\tau} (F_2(2.5/n, 0.002/n) + \max_{f \in [0, 0.002/n]} |K''(f)|) \|\mathbf{f} - \mathbf{f}^*\|_\infty \\ &\stackrel{\textcircled{2}}{\leq} (1/\sqrt{3.289n^2})(0.05610n^2 + 3.290n^2)(0.4X^*\gamma/n) \leq 0.73802X^*\gamma, \end{aligned} \quad (\text{A.14})$$

where $\textcircled{1}$ follows by Lemma A.5 and $\textcircled{2}$ follows from the fact that $\tau \geq 3.289n^2$ for $n \geq 130$ in (A.2) and by combining the upper bound on $F_2(2.5/n, 0.002/n)$ in Table A.5 and the upper bound on $\max_{f \in [0, 0.002/n]} |K''(f)|$ in Table A.5. Similarly, following from Lemma A.5 and the mean value theorem, by combining the upper bound on $F_3(2.5/n, 0.002/n)$ in Table A.5 and the upper bound on $\max_{f \in [0, 0.002/n]} |K^{(3)}(f)|$ in Table A.5, we have

$$\begin{aligned} \frac{1}{\tau} \|\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*) - \mathbf{D}_2(\mathbf{f})\|_{\infty, \infty} &\leq \frac{1}{\tau} (F_3(2.5/n, 0.002/n) + \max_{f \in [0, 0.002/n]} |K'''(f)|) \|\mathbf{f} - \mathbf{f}^*\|_\infty \\ &\leq (1/3.289n^2)(0.28687n^3 + 0.0649394n^3)(0.4X^*\gamma/n) = 0.04279X^*\gamma. \end{aligned} \quad (\text{A.15})$$

To control $\|\mathbf{D}_\ell(\mathbf{f}^*) - \mathbf{D}_\ell(\mathbf{f})\|_{\infty, \infty}$, we rewrite $\mathbf{D}_\ell(\mathbf{f}^*) - \mathbf{D}_\ell(\mathbf{f})$ as

$$\mathbf{D}_\ell(\mathbf{f}^*) - \mathbf{D}_\ell(\mathbf{f}) = \mathbf{D}_\ell(\mathbf{f}^*) - \mathbf{D}_\ell(\mathbf{f}^*, \mathbf{f}) + \mathbf{D}_\ell(\mathbf{f}^*, \mathbf{f}) - \mathbf{D}_\ell(\mathbf{f}).$$

Then, the desired results follow from the triangle inequality of the $\ell_{\infty, \infty}$ norm:

$$\begin{aligned} \|\mathbf{D}_0(\mathbf{f}^*) - \mathbf{D}_0(\mathbf{f})\|_{\infty, \infty} &\leq \|\mathbf{D}_0(\mathbf{f}^*) - \mathbf{D}_0(\mathbf{f}^*, \mathbf{f})\|_{\infty, \infty} + \|\mathbf{D}_0(\mathbf{f}^*, \mathbf{f}) - \mathbf{D}_0(\mathbf{f})\|_{\infty, \infty} \\ &\stackrel{\textcircled{1}}{\leq} 2(0.00758X^*\gamma) = 0.01516X^*\gamma, \end{aligned} \quad (\text{A.16})$$

where $\textcircled{1}$ follows from (A.12) and an exchange of the roles of \mathbf{f} and \mathbf{f}^* ;

$$\begin{aligned} \frac{1}{\sqrt{\tau}} \|\mathbf{D}_1(\mathbf{f}^*) - \mathbf{D}_1(\mathbf{f})\|_{\infty, \infty} &\leq \frac{1}{\sqrt{\tau}} \|\mathbf{D}_1(\mathbf{f}^*) - \mathbf{D}_1(\mathbf{f}^*, \mathbf{f})\|_{\infty, \infty} + \frac{1}{\sqrt{\tau}} \|\mathbf{D}_1(\mathbf{f}^*, \mathbf{f}) - \mathbf{D}_1(\mathbf{f})\|_{\infty, \infty} \\ &\stackrel{\textcircled{1}}{\leq} 2(0.73802X^*\gamma) = 1.47604X^*\gamma, \end{aligned} \quad (\text{A.17})$$

where $\textcircled{1}$ follows from (A.14);

$$\begin{aligned} \frac{1}{\tau} \|\mathbf{D}_2(\mathbf{f}^*) - \mathbf{D}_2(\mathbf{f})\|_{\infty, \infty} &\leq \frac{1}{\tau} \|\mathbf{D}_2(\mathbf{f}^*) - \mathbf{D}_2(\mathbf{f}^*, \mathbf{f})\|_{\infty, \infty} + \frac{1}{\tau} \|\mathbf{D}_2(\mathbf{f}^*, \mathbf{f}) - \mathbf{D}_2(\mathbf{f})\|_{\infty, \infty} \\ &\stackrel{\textcircled{1}}{\leq} 2(0.04279X^*\gamma) = 0.08558X^*\gamma, \end{aligned} \quad (\text{A.18})$$

where $\textcircled{1}$ follows from (A.15).

Then following from Eq. (A.10), (A.14)-(A.15) and (A.16)-(A.18), and together with the sub-multiplicative property of the $\ell_{\infty, \infty}$ norm, we have

$$\begin{aligned} \left\| \frac{1}{\sqrt{\tau}} \text{diag}(1./|\mathbf{c}^*|) [\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_1(\mathbf{f}) \mathbf{c}] \right\|_\infty &\leq \frac{1}{\sqrt{\tau}} \|\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) - \mathbf{D}_1(\mathbf{f})\|_{\infty, \infty} \|1./\mathbf{c}^*\|_\infty \|\mathbf{c}^*\|_\infty \\ &\quad + \frac{1}{\sqrt{\tau}} \|\mathbf{D}_1(\mathbf{f})\|_{\infty, \infty} \|1./\mathbf{c}^*\|_\infty \|\mathbf{c} - \mathbf{c}^*\|_\infty \\ &\leq (0.73802X^*\gamma)B^* + (0.01236)B^*(X^*\gamma) = 0.75038B^*X^*\gamma, \end{aligned} \quad (\text{A.19})$$

where the last but one line follows from $\|1./\mathbf{c}^*\|_\infty \leq 1/c_{\min}^*$ and $\gamma = \gamma_0/c_{\min}^*$. Here and throughout the rest of the paper, we use $1./\mathbf{x}$, $1./|\mathbf{x}|$, \mathbf{y}/\mathbf{x} , $|\mathbf{y}|./|\mathbf{x}|$, $\mathbf{x} \odot \mathbf{y}$ and $\frac{1}{\mathbf{x}}$, $\frac{1}{|\mathbf{x}|}$, $\frac{\mathbf{y}}{\mathbf{x}}$, $\frac{|\mathbf{y}|}{|\mathbf{x}|}$ in the sense of pointwise arithmetic operations, here \mathbf{x}, \mathbf{y} stand for any vectors of the same length.

We apply similar arguments to develop the following bound

$$\begin{aligned} \left\| \frac{1}{\tau} \text{diag}(1./\mathbf{c}^*) [\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_2(\mathbf{f}) \mathbf{c}] \right\|_\infty &\leq \frac{1}{\tau} \|\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*) - \mathbf{D}_2(\mathbf{f})\|_{\infty, \infty} \|1./\mathbf{c}^*\|_\infty \|\mathbf{c}^*\|_\infty \\ &\quad + \frac{1}{\tau} \|\mathbf{D}_2(\mathbf{f})\|_{\infty, \infty} \|1./\mathbf{c}^*\|_\infty \|\mathbf{c} - \mathbf{c}^*\|_\infty \\ &\leq (0.08558X^*\gamma)B^* + (1.05610)B^*(X^*\gamma) \leq 1.14168B^*X^*\gamma. \end{aligned} \quad (\text{A.20})$$

Appendix B Bounding the Dual Atomic Norm of Gaussian Noise

In this section, we develop an upper bound on the dual atomic norm of the weighted Gaussian noise $\mathbf{Z}\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{Z}^2)$ for the positive definite diagonal matrix \mathbf{Z} with $[\mathbf{Z}]_{\ell, \ell} = \frac{g_M(\ell)}{M}$. First following [15, C.4 with $N \geq 4\pi(2n+1)$], we get

$$\sup_{f \in \mathbb{T}} |\mathbf{a}(f)^H \mathbf{Z}\mathbf{w}| \leq 2 \max_{m=0, \dots, N-1} |S_m|, \quad (\text{B.1})$$

where $\{S_m\}_{m=0}^{N-1}$ are N equispaced samples of the continuous function $\mathbf{a}(f)^H \mathbf{Z}\mathbf{w}$ defined on $\mathbb{T} = [0, 1]$:

$$S_m := \mathbf{a}\left(\frac{m}{N}\right)^H \mathbf{Z}\mathbf{w} = \sum_{\ell=-n}^n \frac{g_M(\ell)}{M} w_\ell e^{-i2\pi \ell \frac{m}{N}}.$$

Since $\{w_\ell\}$ are i.i.d Gaussian variables with mean zero and variance σ^2 , we have that each S_m is a Gaussian variable with mean zero and variance given by $\text{Var}(S_m) := \sum_{\ell=-n}^n \left(\frac{g_M(\ell)}{M}\right)^2 \sigma^2$. The main idea next is first to compute an upper bound (denoted by $\bar{\Pi}$) on the variance $\text{Var}(S_m)$ and then apply the Gaussian upper deviation inequality [41, Eq. (7.8)]

$$\mathbb{P}\left[|S_m| \geq t\sqrt{\bar{\Pi}}\right] \leq e^{-t^2/2} \quad (\text{B.2})$$

to get a high-probability upper bound on $|S_m|$. To evaluate $\bar{\Pi}$, it is instructive to first note

$$g_M(\ell) = \frac{1}{M} \sum_{k=\max(\ell-M, -M)}^{\min(\ell+M, M)} \left(1 - \left|\frac{k}{M}\right|\right) \left(1 - \left|\frac{\ell-k}{M}\right|\right),$$

with $\ell = -2M, \dots, 2M$, which is the convolution of two triangle functions:

$$g_M(\ell) = \frac{1}{M} \text{Tri}_M(\ell) * \text{Tri}_M(\ell), \quad \ell = -2M, \dots, 2M. \quad (\text{B.3})$$

Here the triangle function is defined by $\text{Tri}_M(\ell) := 1 - \frac{|\ell|}{M}$, $\ell = -M, \dots, M$ and $*$ represents the convolution operator. Apparently $\text{Var}(S_m)$ is the squared ℓ_2 norm of the vector $\mathbf{g}_M := [g_M(-2M), \dots, g_M(2M)]^T$ scaled by σ^2/M^2 . Since by Eq. (B.3), \mathbf{g}_M is the convolution of two (the same) triangular vectors $\mathbf{h}_M := [\text{Tri}_M(-M), \dots, \text{Tri}_M(M)]$ and then scaled by $1/M$, we obtain an upper bound on $\text{Var}(S_m)$ by applying Young's inequality $\|\mathbf{f} * \mathbf{g}\|_r \leq \|\mathbf{f}\|_p \|\mathbf{g}\|_q$ where $r^{-1} = p^{-1} + q^{-1} - 1$ and setting $r = 2, p = 2, q = 1$:

$$\text{Var}(S_m) = \sum_{\ell=-n}^n \left(\frac{g_M(\ell)}{M}\right)^2 \sigma^2 = \frac{\sigma^2}{M^2} \|\mathbf{g}_M\|_2^2 = \frac{\sigma^2}{M^4} \|\mathbf{h}_M * \mathbf{h}_M\|_2^2 \leq \frac{\sigma^2}{M^4} \|\mathbf{h}_M\|_1^2 \|\mathbf{h}_M\|_2^2. \quad (\text{B.4})$$

Therefore, to bound $\text{Var}(S_m)$, it remains to bound $\|\mathbf{h}_M\|_1^2$ and $\|\mathbf{h}_M\|_2^2$:

$$\|\mathbf{h}_M\|_1^2 = \left(\sum_{\ell=-M}^M \left(1 - \frac{|\ell|}{M} \right) \right)^2 = M^2 \quad \text{and} \quad \|\mathbf{h}_M\|_2^2 = \sum_{\ell=-M}^M \left(1 - \frac{|\ell|}{M} \right)^2 = \frac{2M}{3} + \frac{1}{3M}. \quad (\text{B.5})$$

Then plug (B.5) into (B.4), we obtain an upper bound on $\text{Var}(S_m)$ as

$$\text{Var}(S_m) \leq \frac{\sigma^2}{M^4} M^2 \left(\frac{2M}{3} + \frac{1}{3M} \right) = \sigma^2 \left(\frac{2}{3M} + \frac{1}{M^3} \right) \stackrel{\textcircled{1}}{=} \sigma^2 \left(\frac{4}{3n} + \frac{8}{n^3} \right) \stackrel{\textcircled{2}}{\leq} 1.334\sigma^2/n, \quad \text{for } n \geq 130, \quad (\text{B.6})$$

where ① follows from $n = 2M$ and ② follows since $8/n^2$ is a decreasing sequence of n implying the maximal happens at $n = 130$. Thus we can choose $\bar{\Pi} = 1.334\sigma^2/n$. Plugging $\bar{\Pi} = 1.334\sigma^2/n$ into the Gaussian tail bound (B.2), we get

$$\mathbb{P} \left[|S_m| \geq t\sqrt{1.334}\sigma/\sqrt{n} \right] \leq e^{-t^2/2} \quad (\text{B.7})$$

for all $m = 0, \dots, N-1$.

Applying the union bound yields

$$\mathbb{P} \left[\sup_{f \in \mathbb{T}} |\mathbf{a}(f)^H \mathbf{w}| \geq 2t\sqrt{1.334}\sigma/\sqrt{n} \right] \leq \mathbb{P} \left[\max_{m=0 \dots N-1} |S_m| \geq t\sqrt{1.334}\sigma/\sqrt{n} \right] \leq Ne^{-t^2/2}, \quad (\text{B.8})$$

where the first inequality follows from (B.1). Setting $t = \sqrt{8 \log n}$ in the above gives

$$\mathbb{P} \left[\sup_{f \in \mathbb{T}} |\mathbf{a}(f)^H \mathbf{w}| \geq \underbrace{4\sqrt{2}\sqrt{1.334}}_{\leq 6.534} \sqrt{\log n/n} \sigma \right] \leq \frac{8\pi(2n+1)}{n^4} \leq \frac{1}{n^2}, \quad (\text{B.9})$$

where the last inequality holds for $n \geq 130$. Therefore, we obtain that

$$\mathbb{P} \left[\sup_{f \in \mathbb{T}} |\mathbf{a}(f)^H \mathbf{w}| \leq 6.534 \sqrt{\log n/n} \sigma \right] \geq 1 - \frac{1}{n^2}, \quad \text{for } n \geq 130. \quad (\text{B.10})$$

To bound $\sup_{f \in \mathbb{T}} |\mathbf{a}'(f)^H \mathbf{Z} \mathbf{w}|$ and $\sup_{f \in \mathbb{T}} |\mathbf{a}''(f)^H \mathbf{Z} \mathbf{w}|$, a natural approach is to exploit the relations between $\mathbf{a}(\mathbf{f})$ and its derivatives $\mathbf{a}'(\mathbf{f})$, $\mathbf{a}''(\mathbf{f})$:

$$\mathbf{a}'(f) = (i2\pi \text{diag}(\mathbf{n}))\mathbf{a}(f) \text{ and } \mathbf{a}''(f) = (i2\pi \text{diag}(\mathbf{n}))^2 \mathbf{a}(f).$$

Similarly, define S'_m and S''_m as the m th equispaced sample of $\mathbf{a}'(f)^H \mathbf{Z} \mathbf{w}$ and $\mathbf{a}''(f)^H \mathbf{Z} \mathbf{w}$, respectively:

$$\begin{aligned} S'_m &= \mathbf{a}'(m/N)^H \mathbf{Z} \mathbf{w} = \mathbf{a}(m/N)^H (-i2\pi \text{diag}(\mathbf{n})) \mathbf{Z} \mathbf{w}, \\ S''_m &= \mathbf{a}''(m/N)^H \mathbf{Z} \mathbf{w} = \mathbf{a}(m/N)^H (-i2\pi \text{diag}(\mathbf{n}))^2 \mathbf{Z} \mathbf{w}. \end{aligned}$$

Hence $S'_m \sim \mathcal{N}(0, \text{Var}(S'_m))$ and $S''_m \sim \mathcal{N}(0, \text{Var}(S''_m))$ with

$$\begin{aligned} \text{Var}(S'_m) &= \sum_{\ell=-n}^n (2\pi\ell g_M(\ell)/M)^2 \sigma^2 \leq (2\pi n)^2 \left(\sum_{\ell=-n}^n (g_M(\ell)/M)^2 \sigma^2 \right) \stackrel{\textcircled{1}}{\leq} (2\pi n)^2 1.334\sigma^2/n, \\ \text{Var}(S''_m) &= \sum_{\ell=-n}^n ((2\pi\ell)^2 g_M(\ell)/M)^2 \sigma^2 \leq (2\pi n)^4 \left(\sum_{\ell=-n}^n (g_M(\ell)/M)^2 \sigma^2 \right) \stackrel{\textcircled{2}}{\leq} (2\pi n)^4 1.334\sigma^2/n, \end{aligned}$$

where ① and ② follow from (B.6). Applying the Gaussian deviation inequality to S'_m, S''_m yields

$$\mathbb{P} \left[|S'_m| \geq t2\pi\sqrt{1.334}\sqrt{n}\sigma \right] \leq 2e^{-t^2/2} \quad \text{and} \quad \mathbb{P} \left[|S''_m| \geq t4\pi^2\sqrt{1.334}n\sqrt{n}\sigma \right] \leq 2e^{-t^2/2}.$$

Then applying the same arguments as (B.8), we get for $n \geq 130$,

$$\begin{aligned} \mathbb{P} \left[\sup_{f \in \mathbb{T}} |\mathbf{a}'(f)^H \mathbf{w}| \leq \underbrace{8\sqrt{2\pi}\sqrt{1.334}}_{\leq 41.052} \sqrt{n \log n} \sigma \right] &\geq 1 - \frac{1}{n^2}, \\ \mathbb{P} \left[\sup_{f \in \mathbb{T}} |\mathbf{a}''(f)^H \mathbf{w}| \leq \underbrace{16\sqrt{2\pi^2}\sqrt{1.334}n}_{\leq 257.94} \sqrt{n \log n} \sigma \right] &\geq 1 - \frac{1}{n^2}. \end{aligned} \quad (\text{B.11})$$

Finally, we invoke that

$$\mathbf{A}(\mathbf{f}) = [\mathbf{a}(f_1), \dots, \mathbf{a}(f_k)], \quad \mathbf{A}'(\mathbf{f}) = [\mathbf{a}'(f_1), \dots, \mathbf{a}'(f_k)], \quad \mathbf{A}''(\mathbf{f}) = [\mathbf{a}''(f_1), \dots, \mathbf{a}''(f_k)],$$

and recognize that $\sup_{f \in \mathbb{T}} |\mathbf{a}^{(\ell)}(f)^H \mathbf{w}|$ is an upper bound on $\|\mathbf{A}^{(\ell)}(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty$ to get

$$\|\mathbf{A}^{(\ell)}(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty = \max_{f \in \{f_j\}} |\mathbf{a}^{(\ell)}(f)^H \mathbf{Z} \mathbf{w}| \leq \sup_{f \in \mathbb{T}} |\mathbf{a}^{(\ell)}(f)^H \mathbf{Z} \mathbf{w}|.$$

Together with (B.10), (B.11) and the definition $\gamma_0 = \sigma \sqrt{\frac{\log n}{n}}$, we obtain that the following inequalities hold for $n \geq 130$ with probability at least $1 - \frac{1}{n^2}$:

$$\begin{aligned} \|\mathbf{A}(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty &\leq \sup_{f \in \mathbb{T}} |\mathbf{a}(f)^H \mathbf{w}| \leq 6.534\gamma_0, \\ \|\mathbf{A}'(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty &\leq \sup_{f \in \mathbb{T}} |\mathbf{a}'(f)^H \mathbf{w}| \leq 41.052n\gamma_0, \\ \|\mathbf{A}''(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty &\leq \sup_{f \in \mathbb{T}} |\mathbf{a}''(f)^H \mathbf{w}| \leq 257.94n^2\gamma_0. \end{aligned} \quad (\text{B.12})$$

As a consequence, we claim that the following inequalities hold for $n \geq 130$ with probability at least $1 - \frac{1}{n^2}$:

$$\begin{aligned} \|\text{diag}(1./|\mathbf{c}^*|) \mathbf{A}'(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty / \sqrt{\tau} &\stackrel{\text{①}}{\leq} \|\text{diag}(1./|\mathbf{c}^*|)\|_{\infty, \infty} \|\mathbf{A}(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty / \sqrt{\tau} \\ &\leq \frac{1}{\sqrt{3.289n^2} c_{\min}^*} (41.052n\gamma_0) \leq 22.64\gamma, \end{aligned} \quad (\text{B.13})$$

$$\begin{aligned} \|\text{diag}(\mathbf{c}./|\mathbf{c}^*|^2) \mathbf{A}''(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty / \tau &\stackrel{\text{②}}{\leq} \|\text{diag}(\mathbf{c}./|\mathbf{c}^*|)\|_{\infty, \infty} \|\text{diag}(1./|\mathbf{c}^*|)\|_{\infty, \infty} \|\mathbf{A}''(\mathbf{f})^H \mathbf{Z} \mathbf{w}\|_\infty / \tau \\ &\leq \frac{1}{3.289n^2} (1 + X^*\gamma) \frac{1}{c_{\min}^*} (257.94n^2\gamma_0) \leq 78.43(1 + X^*\gamma)\gamma, \end{aligned} \quad (\text{B.14})$$

where ① follows from that $\|\mathbf{A}\mathbf{x}\|_\infty \leq \|\mathbf{A}\|_{\infty, \infty} \|\mathbf{x}\|_\infty$ by the definition of the $\ell_{\infty, \infty}$ norm and the fact $\tau \geq 3.289n^2$ for $n \geq 130$ by (A.2). ② follows from the sub-multiplicative property of the $\ell_{\infty, \infty}$ norm that $\|\mathbf{A}\mathbf{B}\mathbf{x}\|_\infty \leq \|\mathbf{A}\|_{\infty, \infty} \|\mathbf{B}\|_{\infty, \infty} \|\mathbf{x}\|_\infty$ and $\|\text{diag}(\mathbf{c}./|\mathbf{c}^*|)\|_{\infty, \infty} = \max_\ell |c_\ell|/|c_\ell^*| \leq (1 + X^*\gamma)$ which follows from the assumption $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq X^*\gamma_0/\sqrt{2}$ and the derived results (A.10).

Appendix C Gradient and Hessian for the Nonconvex Program (2.11)

Recall that the objective function G of the program (2.11) is

$$G(\mathbf{f}, \mathbf{c}) = \frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\|_Z^2 + \lambda \|\mathbf{c}\|_1.$$

We denote $\mathbf{c} = \mathbf{u} + i\mathbf{v}$ for $\mathbf{u} \in \mathbb{R}^k$ and $\mathbf{v} \in \mathbb{R}^k$.

C.1 Gradient

Let the operators $\Re\{\cdot\}$ and $\Im\{\cdot\}$ take respectively the real and imaginary parts of a complex number or vector. The gradient of $G(\mathbf{f}, \mathbf{c})$ with respect to $\boldsymbol{\theta} := (\mathbf{f}, \mathbf{u}, \mathbf{v}) \in \mathbb{R}^{3k}$ is defined by

$$\begin{aligned} \nabla G(\boldsymbol{\theta}) &= \begin{bmatrix} \partial G / \partial \mathbf{f} \\ \partial G / \partial \mathbf{u} \\ \partial G / \partial \mathbf{v} \end{bmatrix} \stackrel{\textcircled{1}}{=} \begin{bmatrix} \partial G / \partial \mathbf{f} \\ 2\Re\{\partial G / \partial \bar{\mathbf{c}}\} \\ 2\Im\{\partial G / \partial \bar{\mathbf{c}}\} \end{bmatrix} \stackrel{\textcircled{2}}{=} \begin{bmatrix} \Re\{(\mathbf{A}'(\mathbf{f}) \text{diag}(\mathbf{c}))^H \mathbf{Z}(\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y})\} \\ \Re\{\mathbf{A}(\mathbf{f})^H \mathbf{Z}(\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}) + \lambda \mathbf{c} / |\mathbf{c}|\} \\ \Im\{\mathbf{A}(\mathbf{f})^H \mathbf{Z}(\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}) + \lambda \mathbf{c} / |\mathbf{c}|\} \end{bmatrix} \\ &\stackrel{\textcircled{3}}{=} \begin{bmatrix} \Re\{\text{diag}(\mathbf{c})^H (-\mathbf{D}_1(\mathbf{f})\mathbf{c} + \mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{A}'(\mathbf{f})^H \mathbf{Z}\mathbf{w})\} \\ \Re\{\mathbf{D}_0(\mathbf{f})\mathbf{c} - \mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{A}(\mathbf{f})^H \mathbf{Z}\mathbf{w} + \lambda \mathbf{c} / |\mathbf{c}|\} \\ \Im\{\mathbf{D}_0(\mathbf{f})\mathbf{c} - \mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{A}(\mathbf{f})^H \mathbf{Z}\mathbf{w} + \lambda \mathbf{c} / |\mathbf{c}|\} \end{bmatrix}, \quad (\text{C.1}) \end{aligned}$$

where $\textcircled{1}$ holds for $G \in \mathbb{R}$. $\textcircled{2}$ follows from $\text{diag}(\mathbf{d}\mathbf{f})\mathbf{c} = \text{diag}(\mathbf{c})\mathbf{d}\mathbf{f}$ and $\mathbf{d}|\mathbf{c}| = \frac{\bar{\mathbf{c}}\mathbf{d}\mathbf{c} + \mathbf{c}\mathbf{d}\bar{\mathbf{c}}}{2|\mathbf{c}|}$. $\textcircled{3}$ follows from the kernel matrix factorization formulas (A.6)-(A.7) and by taking into account that $\mathbf{y} = \mathbf{x}^* + \mathbf{w} = \mathbf{A}(\mathbf{f}^*)\mathbf{c}^* + \mathbf{w}$.

C.2 Hessian

The symmetric Hessian matrix $\nabla^2 G(\boldsymbol{\theta})$ is given by

$$\nabla^2 G(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 G}{\partial \mathbf{f}^T \partial \mathbf{f}} & \frac{\partial^2 G}{\partial \mathbf{f}^T \partial \mathbf{u}} & \frac{\partial^2 G}{\partial \mathbf{f}^T \partial \mathbf{v}} \\ \frac{\partial^2 G}{\partial \mathbf{u}^T \partial \mathbf{f}} & \frac{\partial^2 G}{\partial \mathbf{u}^T \partial \mathbf{u}} & \frac{\partial^2 G}{\partial \mathbf{u}^T \partial \mathbf{v}} \\ \frac{\partial^2 G}{\partial \mathbf{v}^T \partial \mathbf{f}} & \frac{\partial^2 G}{\partial \mathbf{v}^T \partial \mathbf{u}} & \frac{\partial^2 G}{\partial \mathbf{v}^T \partial \mathbf{v}} \end{bmatrix} := \begin{bmatrix} \mathbf{H}_{\mathbf{ff}} & \mathbf{H}_{\mathbf{fu}} & \mathbf{H}_{\mathbf{fv}} \\ \mathbf{H}_{\mathbf{uf}} & \mathbf{H}_{\mathbf{uu}} & \mathbf{H}_{\mathbf{uv}} \\ \mathbf{H}_{\mathbf{vf}} & \mathbf{H}_{\mathbf{vu}} & \mathbf{H}_{\mathbf{vv}} \end{bmatrix}$$

with

$$\begin{aligned} \begin{bmatrix} \mathbf{H}_{\mathbf{ff}} \\ \mathbf{H}_{\mathbf{fu}} \\ \mathbf{H}_{\mathbf{fv}} \\ \mathbf{H}_{\mathbf{uu}} \\ \mathbf{H}_{\mathbf{vv}} \\ \mathbf{H}_{\mathbf{uv}} \end{bmatrix} &\stackrel{\textcircled{1}}{=} \begin{bmatrix} \Re\{(\mathbf{A}'(\mathbf{f})\boldsymbol{\Lambda})^H \mathbf{Z}\mathbf{A}'(\mathbf{f})\boldsymbol{\Lambda} + \text{diag}((\mathbf{A}''(\mathbf{f})\boldsymbol{\Lambda})^H \mathbf{Z}(\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}))\} \\ \Re\{(\mathbf{A}'(\mathbf{f})\boldsymbol{\Lambda})^H \mathbf{Z}\mathbf{A}(\mathbf{f}) + \text{diag}(\mathbf{A}'(\mathbf{f})^H \mathbf{Z}(\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}))\} \\ \Im\{-(\mathbf{A}'(\mathbf{f})\boldsymbol{\Lambda})^H \mathbf{Z}\mathbf{A}(\mathbf{f}) + \text{diag}(\mathbf{A}'(\mathbf{f})^H \mathbf{Z}(\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}))\} \\ \mathbf{A}(\mathbf{f})^H \mathbf{Z}\mathbf{A}(\mathbf{f}) + \lambda \text{diag}(\mathbf{v}^2 \odot |\mathbf{c}|^{-3}) \\ \mathbf{A}(\mathbf{f})^H \mathbf{Z}\mathbf{A}(\mathbf{f}) + \lambda \text{diag}(\mathbf{u}^2 \odot |\mathbf{c}|^{-3}) \\ -\lambda \text{diag}(\mathbf{u} \odot \mathbf{v} \odot |\mathbf{c}|^{-3}) \end{bmatrix} \\ &\stackrel{\textcircled{2}}{=} \begin{bmatrix} \Re\{-\boldsymbol{\Lambda}^H \mathbf{D}_2(\mathbf{f})\boldsymbol{\Lambda} - \text{diag}(\boldsymbol{\Lambda}^H \mathbf{A}''(\mathbf{f})^H \mathbf{Z}\mathbf{w}) - \text{diag}(\boldsymbol{\Lambda}^H (\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_2(\mathbf{f})\mathbf{c}))\} \\ \Re\{-\boldsymbol{\Lambda}^H \mathbf{D}_1(\mathbf{f}) - \text{diag}(\mathbf{A}'(\mathbf{f})^H \mathbf{Z}\mathbf{w}) + \text{diag}(\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^*) - \text{diag}(\mathbf{D}_1(\mathbf{f})\mathbf{c})\} \\ \Im\{\boldsymbol{\Lambda}^H \mathbf{D}_1(\mathbf{f}) - \text{diag}(\mathbf{A}'(\mathbf{f})^H \mathbf{Z}\mathbf{w}) + \text{diag}(\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^*) - \text{diag}(\mathbf{D}_1(\mathbf{f})\mathbf{c})\} \\ \mathbf{D}_0(\mathbf{f}) + \lambda \text{diag}(\mathbf{v}^2 \odot |\mathbf{c}|^{-3}) \\ \mathbf{D}_0(\mathbf{f}) + \lambda \text{diag}(\mathbf{u}^2 \odot |\mathbf{c}|^{-3}) \\ -\lambda \text{diag}(\mathbf{u} \odot \mathbf{v} \odot |\mathbf{c}|^{-3}) \end{bmatrix}, \quad (\text{C.2}) \end{aligned}$$

where we denoted $\boldsymbol{\Lambda} := \text{diag}(\mathbf{c})$ to simplify notation. $\textcircled{1}$ follows from direct computation and $\textcircled{2}$ follows from the matrix decomposition formulas (A.6)-(A.7) and by taking into account that $\mathbf{y} = \mathbf{x}^* + \mathbf{w} = \mathbf{A}(\mathbf{f}^*)\mathbf{c}^* + \mathbf{w}$.

Remarkably, if we replace the noisy signal \mathbf{y} in the objective function of the nonconvex program (2.11) with the noise-free signal \mathbf{x}^* to get

$$G^\lambda(\mathbf{f}, \mathbf{c}) = \frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{x}^*\|_Z^2 + \lambda \|\mathbf{c}\|_1,$$

then its gradient and Hessian matrix can be obtained from those of $G(\mathbf{f}, \mathbf{c})$ by setting the noise \mathbf{w} to zero.

Appendix D Proof of Lemma 4.1

Proof. The underlying fixed point map is

$$\Theta^\lambda(\boldsymbol{\theta}) = \boldsymbol{\theta} - \mathbf{W}^* \nabla G^\lambda(\boldsymbol{\theta}),$$

where G^λ is defined as the objective function of the nonconvex program (2.11) with the noisy signal \mathbf{y} replaced by the noise-free signal \mathbf{x}^* :

$$G^\lambda(\boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{x}^*\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1.$$

By Theorem 4.1, to show the existence and uniqueness of a point $\boldsymbol{\theta}^\lambda \in \mathcal{N}^*$ such that $\Theta^\lambda(\boldsymbol{\theta}^\lambda) = \boldsymbol{\theta}^\lambda$, the key is to show that Θ^λ satisfies the non-escaping condition and the contraction condition:

- (i) $\Theta^\lambda(\mathcal{N}^*) \subset \mathcal{N}^*$;
- (ii) There exists $\rho \in (0, 1)$ such that $\|\Theta^\lambda(\mathbf{v}) - \Theta^\lambda(\mathbf{w})\|_\infty \leq \rho \|\mathbf{v} - \mathbf{w}\|_\infty$ for any $\mathbf{v}, \mathbf{w} \in \mathcal{N}^*$.

D.1 Showing the Contraction Property

For $\mathbf{v}, \mathbf{w} \in \mathcal{N}^*$, we have

$$\|\Theta^\lambda(\mathbf{v}) - \Theta^\lambda(\mathbf{w})\|_\infty \stackrel{\textcircled{1}}{=} \left\| \int_0^1 [\nabla \Theta^\lambda(t\mathbf{v} + (1-t)\mathbf{w})](\mathbf{v} - \mathbf{w}) dt \right\|_\infty \stackrel{\textcircled{2}}{\leq} \max_{\boldsymbol{\theta} \in \mathcal{N}^*} \|\nabla \Theta^\lambda(\boldsymbol{\theta})\|_{\infty, \infty} \|\mathbf{v} - \mathbf{w}\|_\infty,$$

where ① follows from the integral form of the mean value theorem for vector-valued functions (see [42, Eq. (A.57)]); ② follows from the sub-multiplicative property of $\|\cdot\|_{\infty, \infty}$ and the fact that $t\mathbf{v} + (1-t)\mathbf{w} \in \mathcal{N}^*$ for $t \in [0, 1]$. Thus, it suffices to show

$$\max_{\boldsymbol{\theta} \in \mathcal{N}^*} \|\nabla \Theta^\lambda(\boldsymbol{\theta})\|_{\infty, \infty} < 1,$$

where the matrix $\ell_{\infty, \infty}$ norm is defined by (following from the definition of the ℓ_∞ norm)

$$\|\mathbf{A}\|_{\infty, \infty} = \left\| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{bmatrix} \right\|_{\infty, \infty} := \left\| \begin{bmatrix} \mathbf{S}\mathbf{A}_{11}\mathbf{S}^{-1} & \mathbf{S}\mathbf{A}_{12} & \mathbf{S}\mathbf{A}_{13} \\ \mathbf{A}_{21}\mathbf{S}^{-1} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31}\mathbf{S}^{-1} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{bmatrix} \right\|_{\infty, \infty},$$

with $\mathbf{S} = \sqrt{\tau} \text{diag}(|\mathbf{c}^*|)$. Together with

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{S}^{-2} & & \\ & \mathbf{I}_k & \\ & & \mathbf{I}_k \end{bmatrix},$$

we therefore obtain that

$$\|\mathbf{W}^* \mathbf{A}\|_{\infty, \infty} = \left\| \begin{bmatrix} \mathbf{S}^{-1} \mathbf{A}_{11} \mathbf{S}^{-1} & \mathbf{S}^{-1} \mathbf{A}_{12} & \mathbf{S}^{-1} \mathbf{A}_{13} \\ \mathbf{A}_{21} \mathbf{S}^{-1} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31} \mathbf{S}^{-1} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{bmatrix} \right\|_{\infty, \infty} = \|\mathbf{W}^{*\frac{1}{2}} \mathbf{A} \mathbf{W}^{*\frac{1}{2}}\|_{\infty, \infty} := \|\Upsilon(\mathbf{A})\|_{\infty, \infty}, \quad (\text{D.1})$$

where the linear operator $\Upsilon(\cdot) := \mathbf{W}^{*\frac{1}{2}}(\cdot)\mathbf{W}^{*\frac{1}{2}}$. The Jacobian of the fixed point map Θ^λ is given by

$$\nabla \Theta^\lambda(\boldsymbol{\theta}) = \mathbf{I} - \mathbf{W}^* \nabla^2 G^\lambda(\boldsymbol{\theta}), \quad (\text{D.2})$$

where the symmetric Hessian matrix $\nabla^2 G^\lambda(\boldsymbol{\theta})$ can be obtained from $\nabla^2 G(\boldsymbol{\theta})$ by setting the noise \mathbf{w} to zero:

$$\nabla^2 G^\lambda(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}_{\mathbf{ff}} & \mathbf{H}_{\mathbf{fu}} & \mathbf{H}_{\mathbf{fv}} \\ \mathbf{H}_{\mathbf{uf}} & \mathbf{H}_{\mathbf{uu}} & \mathbf{H}_{\mathbf{uv}} \\ \mathbf{H}_{\mathbf{vf}} & \mathbf{H}_{\mathbf{vu}} & \mathbf{H}_{\mathbf{vv}} \end{bmatrix}.$$

Due to the symmetric structure of the Hessian matrix, it suffices to know the expressions for the following block matrices (see Eq. (C.2)):

$$\begin{aligned} \mathbf{H}_{\mathbf{ff}} &= \mathbb{R}\{-\boldsymbol{\Lambda}^H \mathbf{D}_2(\mathbf{f}) \boldsymbol{\Lambda} - \text{diag}(\boldsymbol{\Lambda}^H (\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_2(\mathbf{f}) \mathbf{c}))\}; & \mathbf{H}_{\mathbf{uu}} &= \mathbf{D}_0(\mathbf{f}) + \lambda \text{diag}(\mathbf{v} \odot \mathbf{v} ./ |\mathbf{c}|^3); \\ \mathbf{H}_{\mathbf{fu}} &= \mathbb{R}\{-\boldsymbol{\Lambda}^H \mathbf{D}_1(\mathbf{f}) + \text{diag}(\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^*) - \text{diag}(\mathbf{D}_1(\mathbf{f}) \mathbf{c})\}; & \mathbf{H}_{\mathbf{vv}} &= \mathbf{D}_0(\mathbf{f}) + \lambda \text{diag}(\mathbf{u} \odot \mathbf{u} ./ |\mathbf{c}|^3); \\ \mathbf{H}_{\mathbf{fv}} &= \mathbb{I}\{\boldsymbol{\Lambda}^H \mathbf{D}_1(\mathbf{f}) + \text{diag}(\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^*) - \text{diag}(\mathbf{D}_1(\mathbf{f}) \mathbf{c})\}; & \mathbf{H}_{\mathbf{uv}} &= -\lambda \text{diag}(\mathbf{u} \odot \mathbf{v} ./ |\mathbf{c}|^3), \end{aligned}$$

where $\boldsymbol{\Lambda} = \text{diag}(\mathbf{c})$.

Next we compute the weighed $\ell_{\infty, \infty}$ norm of the Jacobian of the fixed point map Θ^λ :

$$\|\nabla \Theta^\lambda(\boldsymbol{\theta})\|_{\infty, \infty} \stackrel{\textcircled{1}}{=} \|\mathbf{W}^* \nabla^2 G^\lambda(\boldsymbol{\theta}) - \mathbf{I}\|_{\infty, \infty} \stackrel{\textcircled{2}}{=} \|\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta}) - \mathbf{W}^{*-1})\|_{\infty, \infty} \stackrel{\textcircled{3}}{=} \|\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I}\|_{\infty, \infty},$$

where ① follows from (D.2), ② follows from (D.1) by noting that $\mathbf{W}^* \nabla^2 G^\lambda(\boldsymbol{\theta}) - \mathbf{I} = \mathbf{W}^* (\nabla^2 G^\lambda(\boldsymbol{\theta}) - \mathbf{W}^{*-1})$ and ③ from the linearity of $\Upsilon(\cdot)$ and $\Upsilon(\mathbf{W}^{*-1}) = \mathbf{W}^{*\frac{1}{2}} \mathbf{W}^{*-1} \mathbf{W}^{*\frac{1}{2}} = \mathbf{I}$. Direct computation gives

$$\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I} = \begin{bmatrix} \frac{-1}{\tau} \mathbb{R}\{\boldsymbol{\Phi}^H \mathbf{D}_2(\mathbf{f}) \boldsymbol{\Phi}\} - \mathbf{I} & \frac{-1}{\sqrt{\tau}} \mathbb{R}\{\boldsymbol{\Phi}\} \mathbf{D}_1(\mathbf{f}) & \frac{-1}{\sqrt{\tau}} \mathbb{I}\{\boldsymbol{\Phi}\} \mathbf{D}_1(\mathbf{f}) \\ \frac{1}{\sqrt{\tau}} \mathbf{D}_1(\mathbf{f}) \mathbb{R}\{\boldsymbol{\Phi}\} & \mathbf{D}_0(\mathbf{f}) - \mathbf{I} & \\ \frac{1}{\sqrt{\tau}} \mathbf{D}_1(\mathbf{f}) \mathbb{I}\{\boldsymbol{\Phi}\} & & \mathbf{D}_0(\mathbf{f}) - \mathbf{I} \end{bmatrix} + \begin{bmatrix} \text{diag}(\mathbf{d}_{\mathbf{ff}}) & \text{diag}(\mathbf{d}_{\mathbf{fu}}) & \text{diag}(\mathbf{d}_{\mathbf{fv}}) \\ \text{diag}(\mathbf{d}_{\mathbf{fu}}) & \text{diag}(\mathbf{d}_{\mathbf{uu}}) & \text{diag}(\mathbf{d}_{\mathbf{uv}}) \\ \text{diag}(\mathbf{d}_{\mathbf{fv}}) & \text{diag}(\mathbf{d}_{\mathbf{uv}}) & \text{diag}(\mathbf{d}_{\mathbf{vv}}) \end{bmatrix}$$

where $\boldsymbol{\Phi} := \text{diag}(\mathbf{c} ./ |\mathbf{c}^*|)$ and

$$\begin{aligned} \mathbf{d}_{\mathbf{ff}} &= -\mathbb{R}\{\text{diag}(\mathbf{c} ./ |\mathbf{c}^*|^2)^H [\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_2(\mathbf{f}) \mathbf{c}] / \tau\}; & \mathbf{d}_{\mathbf{uu}} &= \lambda \text{diag}(\mathbf{u} \odot \mathbf{u} ./ |\mathbf{c}|^3); \\ \mathbf{d}_{\mathbf{fu}} &= \mathbb{R}\{\text{diag}(1 ./ |\mathbf{c}^*|) [\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_1(\mathbf{f}) \mathbf{c}] / \sqrt{\tau}\}; & \mathbf{d}_{\mathbf{uv}} &= \lambda \text{diag}(\mathbf{u} \odot \mathbf{v} ./ |\mathbf{c}|^3); \\ \mathbf{d}_{\mathbf{fv}} &= \mathbb{I}\{\text{diag}(1 ./ |\mathbf{c}^*|) [\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_1(\mathbf{f}) \mathbf{c}] / \sqrt{\tau}\}; & \mathbf{d}_{\mathbf{vv}} &= \lambda \text{diag}(\mathbf{v} \odot \mathbf{v} ./ |\mathbf{c}|^3). \end{aligned}$$

Clearly,

$$\|\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I}\|_{\infty, \infty} = \max \{\Pi_1^\lambda, \Pi_2^\lambda, \Pi_3^\lambda\}$$

with $\Pi_1^\lambda, \Pi_2^\lambda, \Pi_3^\lambda$ being the first, second and third absolute row sums of $\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I}$, respectively.

Bounding Π_1^λ .

$$\begin{aligned} \Pi_1^\lambda &\leq \left\| -\mathbb{R}\{\text{diag}(\mathbf{c} ./ |\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}) / \tau \text{diag}(\mathbf{c} ./ |\mathbf{c}^*|)\} - \mathbf{I} \right\|_\infty + 2 \left\| -\mathbb{R}\{\text{diag}(\mathbf{c} ./ |\mathbf{c}^*|)\} \mathbf{D}_1(\mathbf{f}) / \sqrt{\tau} \right\|_\infty \\ &\quad + 2 \left\| \text{diag}(1 ./ |\mathbf{c}^*|) [\mathbf{D}_1(\mathbf{f}^\lambda, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_1(\mathbf{f}) \mathbf{c}] / \sqrt{\tau} \right\|_\infty + \left\| \text{diag}(\mathbf{c} ./ |\mathbf{c}^*|^2) [\mathbf{D}_2(\mathbf{f}^\lambda, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_2(\mathbf{f}) \mathbf{c}] / \tau \right\|_\infty \\ &\stackrel{\textcircled{1}}{\leq} (0.05610 + 2.12X^*\gamma) + 2(1 + X^*\gamma)(0.01236) + 2(0.75038B^*X^*\gamma) + 1.14168B^*X^*\gamma \\ &\leq 0.08561, \end{aligned} \tag{D.3}$$

where ① follows from Eq. (A.11), (A.19), (A.20) and the following bound

$$\begin{aligned} \left\| -\mathbb{R}\{\text{diag}(\mathbf{c} ./ |\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}) / \tau \text{diag}(\mathbf{c} ./ |\mathbf{c}^*|)\} - \mathbf{I} \right\|_\infty &\leq \max_i \left| \frac{|c_i|^2}{|c_i^*|^2} - 1 \right| + 0.05610 \max_{i,j} \frac{|c_i| |c_j|}{|c_i^*| |c_j^*|} \\ &\leq X^*\gamma(2 + X^*\gamma) + 0.05610(1 + X^*\gamma)^2 \\ &\leq 1.05610(X^*\gamma)^2 + 2.113X^*\gamma + 0.05610 \leq 0.05610 + 2.12X^*\gamma. \end{aligned} \tag{D.4}$$

Bounding Π_2^λ and Π_3^λ .

Note Π_2^λ and Π_3^λ are of the same form. Thus we can bound them together:

$$\begin{aligned} \max\{\Pi_2^\lambda, \Pi_3^\lambda\} &\leq \|\mathbf{D}_1(\mathbf{f})\mathbb{R}\{\text{diag}(\mathbf{c}./|\mathbf{c}^*|)\}\|_\infty / \sqrt{\tau} + \|\text{diag}(1./|\mathbf{c}^*|)[\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f})\mathbf{c}]\|_\infty / \sqrt{\tau} \\ &\quad + \|\mathbf{D}_0(\mathbf{f}) - \mathbf{I}\|_\infty + 2\|\lambda \text{diag}(\mathbf{u} \odot \mathbf{v}./|\mathbf{c}|^3)\|_\infty \\ &\stackrel{\textcircled{1}}{\leq} (1 + X^*\gamma)(0.01236) + (0.75038B^*X^*\gamma) + (0.00755) + 2(0.646X^*\gamma) \\ &< \Pi_1^\lambda \text{ (since } B^*X^*\gamma \leq 10^{-3}\text{),} \end{aligned}$$

where $\textcircled{1}$ follows from Eq. (A.11), (A.19)-(A.20) and $\lambda \leq 0.646X^*\gamma_0$. Therefore,

$$\maximize_{\boldsymbol{\theta} \in \mathcal{N}^*} \|\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I}\|_{\infty, \infty} \leq 0.08561 < 1, \quad (\text{D.5})$$

implying the contraction property of $\Theta^\lambda(\boldsymbol{\theta})$.

D.2 Showing the Non-escaping Property

By the definition of the neighborhood \mathcal{N}^* , it suffices to bound the distance between $\Theta^\lambda(\boldsymbol{\theta})$ and $\boldsymbol{\theta}^*$:

$$\begin{aligned} \|\Theta^\lambda(\boldsymbol{\theta}) - \boldsymbol{\theta}^*\|_\infty &\stackrel{\textcircled{1}}{\leq} \|\Theta^\lambda(\boldsymbol{\theta}) - \Theta^\lambda(\boldsymbol{\theta}^*)\|_\infty + \|\Theta^\lambda(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^*\|_\infty \\ &\stackrel{\textcircled{2}}{=} \left\| \int_0^1 [\nabla_{\boldsymbol{\theta}} \Theta^\lambda((1-t)\boldsymbol{\theta}^* + t\boldsymbol{\theta})](\boldsymbol{\theta} - \boldsymbol{\theta}^*) dt \right\|_\infty + \|\Theta^\lambda(\boldsymbol{\theta}^*) - \boldsymbol{\theta}^*\|_\infty \\ &\stackrel{\textcircled{3}}{\leq} \maximize_{\mathbf{z} \in \mathcal{N}^*} \|\nabla_{\boldsymbol{\theta}} \Theta^\lambda(\mathbf{z})\|_{\infty, \infty} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty + \|\mathbf{W}^* \nabla G^\lambda(\boldsymbol{\theta}^*)\|_\infty \\ &\stackrel{\textcircled{4}}{\leq} (0.08561)(X^*\gamma_0/\sqrt{2}) + \lambda \stackrel{\textcircled{5}}{\leq} X^*\gamma_0/\sqrt{2}, \end{aligned}$$

where $\textcircled{1}$ follows from the triangle inequality, $\textcircled{2}$ follows from the integral form of the mean value theorem for vector-valued functions (see [42, Eq. (A.57)]), $\textcircled{3}$ follows from sub-multiplicative property of $\|\cdot\|_{\infty, \infty}$ and the fact that $(1-t)\boldsymbol{\theta}^* + t\boldsymbol{\theta} \in \mathcal{N}^*$ for $t \in [0, 1]$, $\textcircled{4}$ follows from

$$\|\mathbf{W}^* \nabla G^\lambda(\boldsymbol{\theta}^*)\|_\infty = \left\| \begin{bmatrix} 0 \\ \mathbb{R}\{\lambda \mathbf{c}^*./|\mathbf{c}^*|\} \\ \mathbb{I}\{\lambda \mathbf{c}^*./|\mathbf{c}^*|\} \end{bmatrix} \right\|_\infty \leq \lambda,$$

and $\textcircled{5}$ holds for $\lambda \leq 0.646X^*\gamma_0$ since $(0.08561)(X^*\gamma_0/\sqrt{2}) + 0.646X^*\gamma_0 \leq 0.9992X^*\gamma_0/\sqrt{2}$.

In sum, Θ^λ satisfies both the contraction and the non-escaping properties in \mathcal{N}^* . Therefore, by the contraction mapping theorem, the map Θ^λ has a unique fixed point $\boldsymbol{\theta}^\lambda \in \mathcal{N}^*$ satisfying $\Theta^\lambda(\boldsymbol{\theta}^\lambda) = \boldsymbol{\theta}^\lambda$.

We continue to show that $\boldsymbol{\theta}^\lambda$ is a differentiable function of λ . Define a function $F : \mathbb{R}^{3k} \times \mathbb{R} \mapsto \mathbb{R}^{3k}$ as $F(\boldsymbol{\theta}, \lambda) = \nabla G^\lambda(\boldsymbol{\theta})$ and recognize $F(\boldsymbol{\theta}, \lambda)$ is continuously differentiable since it has a continuous Jacobian given by

$$\partial F(\boldsymbol{\theta}, \lambda) = \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\theta}} F(\boldsymbol{\theta}, \lambda) & \frac{\partial}{\partial \lambda} F(\boldsymbol{\theta}, \lambda) \end{bmatrix} = \begin{bmatrix} \nabla^2 G^\lambda(\boldsymbol{\theta}) & \begin{bmatrix} \mathbf{0} \\ \mathbb{R}\{\mathbf{c}./|\mathbf{c}|\} \\ \mathbb{I}\{\mathbf{c}./|\mathbf{c}|\} \end{bmatrix} \end{bmatrix},$$

with $\frac{\partial}{\partial \boldsymbol{\theta}} F(\boldsymbol{\theta}, \lambda)$ nonsingular in \mathcal{N}^* by (D.5). Then according to the implicit function theorem (see [43, Proposition A.25]), there is a continuously differentiable function $\mathbf{g}(\cdot)$ such that $F(\mathbf{g}(\lambda), \lambda) = \nabla G^\lambda(\mathbf{g}(\lambda)) = \mathbf{0}$ and

$$\frac{d}{d\lambda} \mathbf{g}(\lambda) = -\left(\frac{\partial}{\partial \boldsymbol{\theta}} F(\mathbf{g}(\lambda), \lambda)\right)^{-1} \frac{\partial}{\partial \lambda} F(\mathbf{g}(\lambda), \lambda) = -(\nabla^2 G^\lambda(\mathbf{g}(\lambda)))^{-1} \frac{\partial}{\partial \lambda} \nabla G^\lambda(\mathbf{g}(\lambda)). \quad (\text{D.6})$$

Since $\nabla G^\lambda(\mathbf{g}(\lambda)) = \mathbf{0}$ is equivalent to $\Theta^\lambda(\mathbf{g}(\lambda)) = \mathbf{g}(\lambda)$, we conclude that $\boldsymbol{\theta}^\lambda = \mathbf{g}(\lambda)$ due to the uniqueness of the fixed point of Θ^λ . Therefore, $\boldsymbol{\theta}^\lambda$ is a differentiable function of λ and

$$\frac{d}{d\lambda}\boldsymbol{\theta}^\lambda = -(\nabla^2 G^\lambda(\boldsymbol{\theta}^\lambda))^{-1} \frac{\partial}{\partial \lambda} \nabla G^\lambda(\boldsymbol{\theta}^\lambda). \quad (\text{D.7})$$

Finally, let $\lim_{\lambda \rightarrow 0} \boldsymbol{\theta}^\lambda = \boldsymbol{\theta}^0$. Taking limit as λ goes to 0 in the equation $\nabla G^\lambda(\boldsymbol{\theta}^\lambda) = \mathbf{0}$ yields $\nabla G^0(\boldsymbol{\theta}^0) = \mathbf{0}$ due to the continuity of $\nabla G^\lambda(\boldsymbol{\theta})$ in λ and $\boldsymbol{\theta}$ and the continuity of $\boldsymbol{\theta}^\lambda$. Since $\nabla G^0(\boldsymbol{\theta}^*) = \mathbf{0}$ by direct computation and the solution is unique in \mathcal{N}^* , we conclude that $\lim_{\lambda \rightarrow 0} \boldsymbol{\theta}^\lambda = \boldsymbol{\theta}^0 = \boldsymbol{\theta}^*$. \square

Appendix E Proof of Lemma 4.2

Proof. The main idea is again to apply the contraction mapping theorem 4.1 to the fixed point map:

$$\Theta(\boldsymbol{\theta}) = \boldsymbol{\theta} - \mathbf{W}^* \nabla G(\boldsymbol{\theta}),$$

where G is the objective function of (2.11):

$$G(\boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{A}(\mathbf{f})\mathbf{c} - \mathbf{y}\|_{\mathbf{Z}}^2 + \lambda \|\mathbf{c}\|_1$$

with $\lambda = 0.646X^*\gamma_0$. By Theorem 4.1, showing the existence of a unique point $\hat{\boldsymbol{\theta}} \in \mathcal{N}^\lambda$ such that $\Theta(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}}$ can be reduced to showing that Θ satisfies both the non-escaping property and the contraction properties:

- (i) $\Theta(\mathcal{N}^\lambda) \subset \mathcal{N}^\lambda$;
- (ii) There exists $\rho \in (0, 1)$ such that $\|\Theta(\mathbf{v}) - \Theta(\mathbf{w})\|_\infty \leq \rho \|\mathbf{v} - \mathbf{w}\|_\infty$ for any $\mathbf{v}, \mathbf{w} \in \mathcal{N}^\lambda$.

E.1 Showing the Contraction Property

Recall that \mathcal{N}^* is a neighborhood centered at $\boldsymbol{\theta}^*$ and \mathcal{N}^λ is a neighborhood centered at $\boldsymbol{\theta}^\lambda$ defined respectively via

$$\mathcal{N}^* = \left\{ \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq \frac{X^*}{\sqrt{2}} \gamma_0 \right\} \quad \text{and} \quad \mathcal{N}^\lambda = \left\{ \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^\lambda\|_\infty \leq \frac{35.2}{\sqrt{2}} \gamma_0 \right\}.$$

Keep in mind that $\boldsymbol{\theta}^\lambda$ is the unique point in \mathcal{N}^* that satisfies $\nabla G^\lambda(\boldsymbol{\theta}^\lambda) = \mathbf{0}$. To show the contraction of Θ in \mathcal{N}^λ , our strategy is to show Θ is contractive in a larger set $\hat{\mathcal{N}}$ that contains \mathcal{N}^λ :

$$\hat{\mathcal{N}} = \left\{ \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq \frac{X^* + 35.2}{\sqrt{2}} \gamma_0 := \frac{\hat{X}}{\sqrt{2}} \gamma_0 \right\}.$$

Recognize that $\hat{\mathcal{N}}$ is a neighborhood centered at $\boldsymbol{\theta}^*$ but with a radius $35.2\gamma_0/\sqrt{2}$ larger than that of \mathcal{N}^* . Such a choice is made for the purpose of showing the closeness between the final fixed point solution $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$. We remark that the quantity $35.2\gamma_0/\sqrt{2}$ corresponds to the dual atomic norm of the weighted Gaussian noise. Adding such a noise norm term to the radius of the original neighborhood \mathcal{N}^* ensures that the region $\hat{\mathcal{N}}$ is large enough for $\Theta(\boldsymbol{\theta})$ to be non-escaping. This is reasonable because the second fixed point map (4.6) involves an additive Gaussian noise and we have shown that the first fixed point map (4.4) (the one constructed in the noise-free setting) satisfies the non-escaping property in \mathcal{N}^* .

Next, we apply arguments similar to those of showing the contraction of Θ^λ in \mathcal{N}^* . In particular, we first compute the expression of $\Upsilon(\nabla^2 G(\boldsymbol{\theta})) - \mathbf{I}$:

$$\Upsilon(\nabla^2 G(\boldsymbol{\theta})) - \mathbf{I} = \begin{bmatrix} \frac{-1}{\tau} \mathbb{R}\{\boldsymbol{\Phi}^H \mathbf{D}_2(\mathbf{f}) \boldsymbol{\Phi}\} - \mathbf{I} & \frac{-1}{\sqrt{\tau}} \mathbb{R}\{\boldsymbol{\Phi}\} \mathbf{D}_1(\mathbf{f}) & \frac{-1}{\sqrt{\tau}} \mathbb{I}\{\boldsymbol{\Phi}\} \mathbf{D}_1(\mathbf{f}) \\ \frac{1}{\sqrt{\tau}} \mathbf{D}_1(\mathbf{f}) \mathbb{R}\{\boldsymbol{\Phi}\} & \mathbf{D}_0(\mathbf{f}) - \mathbf{I} & \mathbf{0} \\ \frac{1}{\sqrt{\tau}} \mathbf{D}_1(\mathbf{f}) \mathbb{I}\{\boldsymbol{\Phi}\} & \mathbf{0} & \mathbf{D}_0(\mathbf{f}) - \mathbf{I} \end{bmatrix} + \begin{bmatrix} \text{diag}(\hat{\mathbf{d}}_{\mathbf{ff}}) & \text{diag}(\hat{\mathbf{d}}_{\mathbf{fu}}) & \text{diag}(\hat{\mathbf{d}}_{\mathbf{fv}}) \\ \text{diag}(\hat{\mathbf{d}}_{\mathbf{fu}}) & \text{diag}(\hat{\mathbf{d}}_{\mathbf{uu}}) & \text{diag}(\hat{\mathbf{d}}_{\mathbf{uv}}) \\ \text{diag}(\hat{\mathbf{d}}_{\mathbf{fv}}) & \text{diag}(\hat{\mathbf{d}}_{\mathbf{uv}}) & \text{diag}(\hat{\mathbf{d}}_{\mathbf{vv}}) \end{bmatrix}$$

with $\boldsymbol{\Phi} = \text{diag}(\mathbf{c}/|\mathbf{c}^*|)$ and

$$\begin{aligned}
\hat{\mathbf{d}}_{\mathbf{ff}} &= -\mathbb{R}\{\text{diag}(\mathbf{c}/|\mathbf{c}^*|^2)^H[\mathbf{A}''(\mathbf{f})^H\mathbf{Z}\mathbf{w} + \mathbf{D}_2(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_2(\mathbf{f})\mathbf{c}]/\tau\}; & \hat{\mathbf{d}}_{\mathbf{uu}} &= \lambda \text{diag}(\mathbf{u} \odot \mathbf{u}/|\mathbf{c}|^3); \\
\hat{\mathbf{d}}_{\mathbf{fu}} &= \mathbb{R}\{\text{diag}(1./|\mathbf{c}^*|)[- \mathbf{A}'(\mathbf{f})^H\mathbf{Z}\mathbf{w} + \mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f})\mathbf{c}]/\sqrt{\tau}\}; & \hat{\mathbf{d}}_{\mathbf{uv}} &= \lambda \text{diag}(\mathbf{u} \odot \mathbf{v}/|\mathbf{c}|^3); \\
\hat{\mathbf{d}}_{\mathbf{fv}} &= \mathbb{I}\{\text{diag}(1./|\mathbf{c}^*|)[- \mathbf{A}'(\mathbf{f})^H\mathbf{Z}\mathbf{w} + \mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f})\mathbf{c}]/\sqrt{\tau}\}; & \hat{\mathbf{d}}_{\mathbf{vv}} &= \lambda \text{diag}(\mathbf{v} \odot \mathbf{v}/|\mathbf{c}|^3).
\end{aligned}$$

Comparing the expressions for $[\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I}]$ and $[\Upsilon(\nabla^2 G(\boldsymbol{\theta})) - \mathbf{I}]$ shows that the latter differs in have additional noise terms in the first row and the first column blocks. We have shown that the first absolute row sum Π_1^λ of $[\Upsilon(\nabla^2 G^\lambda(\boldsymbol{\theta})) - \mathbf{I}]$ dominates the other row sums. Having additional noise terms will only increase the final bounds due to the application of the triangle inequality. Therefore, the first absolute row sum (denoted by $\hat{\Pi}_1$) of $[\Upsilon(\nabla^2 G(\boldsymbol{\theta})) - \mathbf{I}]$ also dominates and hence achieves the $\ell_{\infty, \infty}$ norm. Direct computation gives

$$\begin{aligned}
\hat{\Pi}_1 &\leq \Pi_1^\lambda + 2\|\text{diag}(1./|\mathbf{c}^*|)\mathbf{A}'(\mathbf{f})^H\mathbf{Z}\mathbf{w}\|_\infty/\sqrt{\tau} + \|\text{diag}(\mathbf{c}/|\mathbf{c}^*|^2)\mathbf{A}''(\mathbf{f})^H\mathbf{Z}\mathbf{w}\|_\infty/\tau \\
&\stackrel{\textcircled{1}}{\leq} 0.08561 + 2(22.64\gamma) + 78.43(1 + \hat{X}\gamma)\gamma \\
&\stackrel{\textcircled{2}}{\leq} 0.08563,
\end{aligned}$$

where $\textcircled{1}$ follows from $\Pi_1^\lambda \leq 0.08561$ and Eq. (B.12)-(B.14), $\textcircled{2}$ follows from $\hat{X} = X^* + 35.2$ and the SNR condition (2.6) that $X^*B^*\gamma \leq 10^{-3}$ and $B^*/X^* \leq 10^{-4}$ hence $2(22.64\gamma) + 78.43(1 + \hat{X}\gamma)\gamma \leq 0.00002$. Hence,

$$\max_{\boldsymbol{\theta} \in \mathcal{N}} \|\nabla \Theta(\boldsymbol{\theta})\|_{\infty, \infty} \leq 0.08563 < 1. \quad (\text{E.1})$$

This implies the contraction of Θ in \mathcal{N}^λ , since

$$\max_{\boldsymbol{\theta} \in \mathcal{N}^\lambda} \|\nabla \Theta(\boldsymbol{\theta})\|_{\infty, \infty} \leq \max_{\boldsymbol{\theta} \in \mathcal{N}} \|\nabla \Theta(\boldsymbol{\theta})\|_{\infty, \infty}.$$

E.2 Showing the Non-escaping Property

$$\begin{aligned}
\|\Theta(\boldsymbol{\theta}) - \boldsymbol{\theta}^\lambda\|_\infty &= \|(\Theta(\boldsymbol{\theta}) - \Theta(\boldsymbol{\theta}^\lambda)) + (\Theta(\boldsymbol{\theta}^\lambda) - \boldsymbol{\theta}^\lambda)\|_\infty \\
&\stackrel{\textcircled{1}}{\leq} \|\nabla \Theta(\tilde{\boldsymbol{\theta}})^T(\boldsymbol{\theta} - \boldsymbol{\theta}^\lambda)\|_\infty + \|\mathbf{W}^*\nabla G(\boldsymbol{\theta}^\lambda)\|_\infty \\
&\leq \max_{\tilde{\boldsymbol{\theta}} \in \mathcal{N}} \|\nabla \Theta(\tilde{\boldsymbol{\theta}})\|_{\infty, \infty} \|\boldsymbol{\theta} - \boldsymbol{\theta}^\lambda\|_\infty + \|\mathbf{W}^*\nabla G(\boldsymbol{\theta}^\lambda)\|_\infty \\
&\stackrel{\textcircled{2}}{\leq} (0.08563) \left(35.2\gamma_0/\sqrt{2}\right) + 22.7\gamma_0 \\
&\leq 35.117\gamma_0/\sqrt{2} < 35.2\gamma_0/\sqrt{2},
\end{aligned}$$

where $\textcircled{1}$ follows from the mean value theorem for some $\tilde{\boldsymbol{\theta}}$ on the line segment joining $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^\lambda$ and $\textcircled{2}$ follows from (E.1) and (E.2). Eq. (E.2) is given as follows

$$\begin{aligned}
\|\mathbf{W}^*\nabla G(\boldsymbol{\theta}^\lambda)\|_\infty &= \left\| \mathbf{W}^* \begin{bmatrix} \mathbb{R}\{-\text{diag}(\mathbf{c}^\lambda)^H(\mathbf{A}'(\mathbf{f}^\lambda)^H\mathbf{Z}\mathbf{w} + \mathbf{D}_1(\mathbf{f}^\lambda, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f}^\lambda)\mathbf{c}^\lambda)\} \\ \mathbb{R}\{-\mathbf{A}(\mathbf{f}^\lambda)^H\mathbf{Z}\mathbf{w} - \mathbf{D}_0(\mathbf{f}^\lambda, \mathbf{f}^*)\mathbf{c}^* + \mathbf{D}_0(\mathbf{f}^\lambda)\mathbf{c}^\lambda + \lambda\mathbf{c}^\lambda/|\mathbf{c}^\lambda|\} \\ \mathbb{I}\{-\mathbf{A}(\mathbf{f}^\lambda)^H\mathbf{Z}(\lambda\mathbf{w} - \mathbf{D}_0(\mathbf{f}^\lambda, \mathbf{f}^*)\mathbf{c}^* + \mathbf{D}_0(\mathbf{f}^\lambda)\mathbf{c}^\lambda + \lambda\mathbf{c}^\lambda/|\mathbf{c}^\lambda|)\} \end{bmatrix} \right\|_\infty \\
&\stackrel{\textcircled{1}}{=} \left\| \begin{bmatrix} \mathbb{R}\{-\text{diag}(\mathbf{c}^\lambda/|\mathbf{c}^*|)^H\mathbf{A}'(\mathbf{f}^\lambda)^H\mathbf{Z}\mathbf{w}\}/\sqrt{\tau} \\ \mathbb{R}\{-\mathbf{A}(\mathbf{f}^\lambda)^H\mathbf{Z}\mathbf{w}\} \\ \mathbb{I}\{-\mathbf{A}(\mathbf{f}^\lambda)^H\mathbf{Z}\mathbf{w}\} \end{bmatrix} \right\|_\infty \\
&\stackrel{\textcircled{2}}{\leq} \left\| \begin{bmatrix} 41.052n/\sqrt{\tau}(1 + X^*\gamma)\gamma_0 \\ 6.534\gamma_0 \\ 6.534\gamma_0 \end{bmatrix} \right\|_\infty \\
&\leq 22.7\gamma_0,
\end{aligned} \quad (\text{E.2})$$

where ① holds since $\nabla G^\lambda(\boldsymbol{\theta})$ vanishes at $\boldsymbol{\theta}^\lambda$:

$$\nabla G^\lambda(\boldsymbol{\theta}^\lambda) = \begin{bmatrix} \mathbb{R}\{-\text{diag}(\mathbf{c}^\lambda)^H(\mathbf{D}_1(\mathbf{f}^\lambda, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f}^\lambda)\mathbf{c}^\lambda)\} \\ \mathbb{R}\{-\mathbf{D}_0(\mathbf{f}^\lambda, \mathbf{f}^*)\mathbf{c}^* + \mathbf{D}_0(\mathbf{f}^\lambda)\mathbf{c}^\lambda + \lambda\mathbf{c}^\lambda./|\mathbf{c}^\lambda|\} \\ \mathbb{I}\{-\mathbf{D}_0(\mathbf{f}^\lambda, \mathbf{f}^*)\mathbf{c}^* + \mathbf{D}_0(\mathbf{f}^\lambda)\mathbf{c}^\lambda + \lambda\mathbf{c}^\lambda./|\mathbf{c}^\lambda|\} \end{bmatrix} = \mathbf{0}.$$

② holds with probability at least $1 - \frac{1}{n^2}$ by (B.12)-(B.14).

Hence both the contraction and the non-escaping properties are satisfied by Θ in \mathcal{N}^λ . Then by the contraction mapping theorem, we conclude the proof of Lemma 4.2. \square

Appendix F Proof of Lemma 4.3

Proof. To show that \mathbf{q}^* is a valid dual certificate, it is instructive to first relate \mathbf{q}^* to the derivative of \mathbf{x}^λ with respect to λ (where we treat \mathbf{x}^λ as a function of λ):

$$\mathbf{q}^* = \lim_{\lambda \rightarrow 0} \mathbf{q}^\lambda = \lim_{\lambda \rightarrow 0} \frac{\mathbf{x}^* - \mathbf{x}^\lambda}{\lambda} = -\frac{d}{d\lambda} \mathbf{x}^\lambda|_{\lambda=0}, \quad (\text{F.1})$$

where we used the fact that $\lim_{\lambda \rightarrow 0} \mathbf{x}^\lambda = \lim_{\lambda \rightarrow 0} \mathbf{A}(\mathbf{f}^\lambda)\mathbf{c}^\lambda = \mathbf{A}(\mathbf{f}^*)\mathbf{c}^* = \mathbf{x}^*$ by Lemma 4.1. Since $\mathbf{x}^\lambda = \mathbf{A}(\mathbf{f}^\lambda)\mathbf{c}^\lambda = \sum_\ell c_\ell^\lambda \mathbf{a}(f_\ell^\lambda)$, we compute the derivative $\frac{d}{d\lambda} \mathbf{x}^\lambda$ using the chain rule as:

$$\frac{d}{d\lambda} \mathbf{x}^\lambda = \sum_\ell \left(\frac{d}{d\lambda} u_\ell^\lambda + i \frac{d}{d\lambda} v_\ell^\lambda \right) \mathbf{a}(f_\ell^\lambda) + \sum_\ell c_\ell^\lambda \left(\frac{df_\ell^\lambda}{d\lambda} \mathbf{a}'(f_\ell^\lambda) \right) = [\mathbf{A}'(\mathbf{f}^\lambda) \text{diag}(\mathbf{c}^\lambda) \quad \mathbf{A}(\mathbf{f}^\lambda) \quad i\mathbf{A}(\mathbf{f}^\lambda)] \frac{d}{d\lambda} \boldsymbol{\theta}^\lambda, \quad (\text{F.2})$$

where $\mathbf{A}'(\mathbf{f}) = [\mathbf{a}'(f_1) \quad \cdots \quad \mathbf{a}'(f_k)]$. Therefore, using Eq. (F.1) and (F.2) we obtain:

$$\begin{aligned} \mathbf{q}^* &= -\lim_{\lambda \rightarrow 0} [\mathbf{A}'(\mathbf{f}^\lambda) \text{diag}(\mathbf{c}^\lambda) \quad \mathbf{A}(\mathbf{f}^\lambda) \quad i\mathbf{A}(\mathbf{f}^\lambda)] \frac{d}{d\lambda} \boldsymbol{\theta}^\lambda \\ &= -[\mathbf{A}'(\mathbf{f}^*) \text{diag}(\mathbf{c}^*) \quad \mathbf{A}(\mathbf{f}^*) \quad i\mathbf{A}(\mathbf{f}^*)] \lim_{\lambda \rightarrow 0} \frac{d}{d\lambda} \boldsymbol{\theta}^\lambda \\ &= [\mathbf{A}'(\mathbf{f}^*) \text{diag}(\mathbf{c}^*) \quad \mathbf{A}(\mathbf{f}^*) \quad i\mathbf{A}(\mathbf{f}^*)] (\nabla^2 G^0(\boldsymbol{\theta}^*))^{-1} \frac{\partial}{\partial \lambda} \nabla G^0(\boldsymbol{\theta}^*), \end{aligned} \quad (\text{F.3})$$

where in the second line we again used the fact that $\lim_{\lambda \rightarrow 0} \boldsymbol{\theta}^\lambda = \boldsymbol{\theta}^*$ by Lemma 4.1, and in the last line we used the expression for $d\boldsymbol{\theta}^\lambda/d\lambda$ given in (4.5).

We next compute $\frac{\partial}{\partial \lambda} \nabla G^0(\boldsymbol{\theta}^*)$ explicitly. Let $K^{(\ell)}(\cdot)$ denote the ℓ -order derivative of the Jackson kernel $K(\cdot)$ (see Appendix A for more details). Recall that $\mathbf{D}_\ell(\mathbf{f}^1, \mathbf{f}^2) := [K^{(\ell)}(f_m^2 - f_n^1)]_{1 \leq n \leq k, 1 \leq m \leq k}$ and $\mathbf{D}_\ell(\mathbf{f}) := \mathbf{D}_\ell(\mathbf{f}, \mathbf{f})$ are matrices formed by sampling the Jackson kernel and its derivatives. Then we have the following expression for $\nabla G^\lambda(\boldsymbol{\theta})$ (see Appendix C for more details)

$$\nabla G^\lambda(\boldsymbol{\theta}) = \begin{bmatrix} \mathbb{R}\{\text{diag}(\mathbf{c})(\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f})\mathbf{c})\} \\ \mathbb{R}\{-\mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* + \mathbf{D}_0(\mathbf{f})\mathbf{c} + \lambda\mathbf{c}./|\mathbf{c}|\} \\ \mathbb{I}\{-\mathbf{D}_0(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* + \mathbf{D}_0(\mathbf{f})\mathbf{c} + \lambda\mathbf{c}./|\mathbf{c}|\} \end{bmatrix}. \quad (\text{F.4})$$

Therefore, the partial derivative of (F.4) with respect to λ is the expanded complex sign vector:

$$\frac{\partial}{\partial \lambda} \nabla G^\lambda(\boldsymbol{\theta}^\lambda) = \begin{bmatrix} \mathbf{0} \\ \mathbb{R}\{\text{sign}(\mathbf{c}^\lambda)\} \\ \mathbb{I}\{\text{sign}(\mathbf{c}^\lambda)\} \end{bmatrix} := \begin{bmatrix} \mathbf{0} \\ \mathbf{s}_R^\lambda \\ \mathbf{s}_I^\lambda \end{bmatrix} \implies \frac{\partial}{\partial \lambda} \nabla G^0(\boldsymbol{\theta}^*) = \begin{bmatrix} \mathbf{0} \\ \mathbb{R}\{\text{sign}(\mathbf{c}^*)\} \\ \mathbb{I}\{\text{sign}(\mathbf{c}^*)\} \end{bmatrix} := \begin{bmatrix} \mathbf{0} \\ \mathbf{s}_R^* \\ \mathbf{s}_I^* \end{bmatrix}. \quad (\text{F.5})$$

Here $\mathbf{s}^\lambda = \mathbf{c}^\lambda./|\mathbf{c}^\lambda|$, $\mathbf{s}^* = \mathbf{c}^*./|\mathbf{c}^*|$ and the subscript R and I indicate the real and imaginary parts respectively.

Combining Eq. (F.3) and (F.5), we get

$$\mathbf{q}^* = [\mathbf{A}'(\mathbf{f}^*) \quad \mathbf{A}(\mathbf{f}^*) \quad i\mathbf{A}(\mathbf{f}^*)] \underbrace{\begin{bmatrix} \text{diag}(\mathbf{c}^*) & & \\ & \mathbf{I} & \\ & & \mathbf{I} \end{bmatrix}}_{:= [\boldsymbol{\beta}^T \quad \boldsymbol{\alpha}_R^T \quad \boldsymbol{\alpha}_I^T]^T} (\nabla^2 G^0(\boldsymbol{\theta}^*))^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{s}_R^* \\ \mathbf{s}_I^* \end{bmatrix}, \quad (\text{F.6})$$

where we have defined the coefficient vectors $\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}$ in (F.6). These coefficient vectors satisfy

$$\nabla^2 G^0(\boldsymbol{\theta}^*) \begin{bmatrix} \text{diag}(\mathbf{c}^*)^{-1} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_R \\ \boldsymbol{\alpha}_I \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{s}_R^* \\ \mathbf{s}_I^* \end{bmatrix}. \quad (\text{F.7})$$

By denoting $\boldsymbol{\alpha} = \boldsymbol{\alpha}_R + i\boldsymbol{\alpha}_I$ and $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_k]^T$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_k]^T$, we obtain an explicit form for the dual polynomial $Q^*(f)$:

$$Q^*(f) = \mathbf{a}(f)^H \mathbf{Z} \mathbf{q}^* = \sum_{\ell=1}^k \alpha_\ell K(f_\ell^* - f) + \sum_{\ell=1}^k \beta_\ell K'(f_\ell^* - f). \quad (\text{F.8})$$

To show that \mathbf{q}^* certifies the atomic decomposition $\mathbf{x}^* = \sum_{\ell=1}^k c_\ell^* \mathbf{a}(f_\ell^*)$, we need to establish that

- 1) $Q^*(f)$ satisfies $Q^*(f_\ell^*) = \text{sign}(c_\ell^*)$, $\ell = 1, \dots, k$ (Interpolation);
- 2) $|Q^*(f)| < 1, \forall f \notin T^*$ (Boundedness).

F.1 Showing the Interpolation Property

The Interpolation property follows from the construction process and is also easy to verify directly by noting

$$\nabla^2 G^0(\boldsymbol{\theta}^*) = \begin{bmatrix} -\mathbb{R}\{\text{diag}(\mathbf{c}^*)^H \mathbf{D}_2(\mathbf{f}^*) \text{diag}(\mathbf{c}^*)\} & \mathbb{R}\{-\text{diag}(\mathbf{c}^*)^H \mathbf{D}_1(\mathbf{f}^*)\} & \mathbb{I}\{\text{diag}(\mathbf{c}^*)^H \mathbf{D}_1(\mathbf{f}^*)\} \\ -\mathbb{R}\{\mathbf{D}_1(\mathbf{f}^*)^H \text{diag}(\mathbf{c}^*)\} & \mathbf{D}_0(\mathbf{f}^*) & 0 \\ -\mathbb{I}\{\mathbf{D}_1(\mathbf{f}^*)^H \text{diag}(\mathbf{c}^*)\} & 0 & \mathbf{D}_0(\mathbf{f}^*) \end{bmatrix}.$$

Indeed, the Interpolation property is a result of (F.7): since $\mathbf{D}_1(\mathbf{f}^*) \in \mathbb{R}^{k \times k}$ and $\mathbf{D}_1(\mathbf{f}^*)^T = -\mathbf{D}_1(\mathbf{f}^*)$ (see Appendix A), the last two row blocks in (F.7) read

$$\begin{aligned} & \begin{bmatrix} \mathbf{D}_1(\mathbf{f}^*) \mathbb{R}\{\text{diag}(\mathbf{c}^*)\} & \mathbf{D}_0(\mathbf{f}^*) & 0 \\ \mathbf{D}_1(\mathbf{f}^*) \mathbb{I}\{\text{diag}(\mathbf{c}^*)\} & 0 & \mathbf{D}_0(\mathbf{f}^*) \end{bmatrix} \begin{bmatrix} \text{diag}(\mathbf{c}^*)^{-1} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_R \\ \boldsymbol{\alpha}_I \end{bmatrix} = \begin{bmatrix} \mathbf{s}_R^* \\ \mathbf{s}_I^* \end{bmatrix} \\ \iff & \mathbf{D}_1(\mathbf{f}^*) (\mathbb{R}\{\text{diag}(\mathbf{c}^*)\} + i\mathbb{I}\{\text{diag}(\mathbf{c}^*)\}) \text{diag}(\mathbf{c}^*)^{-1} \boldsymbol{\beta} + \mathbf{D}_0(\mathbf{f}^*) (\boldsymbol{\alpha}_R + i\boldsymbol{\alpha}_I) = \mathbb{R}\{\text{sign}(\mathbf{c}^*)\} + i\mathbb{I}\{\text{sign}(\mathbf{c}^*)\} \\ \iff & \mathbf{D}_1(\mathbf{f}^*) \boldsymbol{\beta} + \mathbf{D}_0(\mathbf{f}^*) \boldsymbol{\alpha} = \text{sign}(\mathbf{c}^*) \\ \iff & Q^*(f_\ell^*) = \text{sign}(c_\ell^*), \ell = 1, \dots, k. \end{aligned} \quad (\text{F.9})$$

Furthermore, the first row block of (F.7) is equivalent to

$$\begin{aligned} & -\mathbb{R}\{\text{diag}(\mathbf{c}^*)^H \mathbf{D}_2(\mathbf{f}^*) \text{diag}(\mathbf{c}^*)\} \text{diag}(\mathbf{c}^*)^{-1} \boldsymbol{\beta} + \mathbb{R}\{-\text{diag}(\mathbf{c}^*)^H \mathbf{D}_1(\mathbf{f}^*)\} \boldsymbol{\alpha}_R + \mathbb{I}\{\text{diag}(\mathbf{c}^*)^H \mathbf{D}_1(\mathbf{f}^*)\} \boldsymbol{\alpha}_I = \mathbf{0} \\ \iff & \mathbb{R}\{\text{diag}(\mathbf{c}^*)^H (\mathbf{D}_2(\mathbf{f}^*) \boldsymbol{\beta} + \mathbf{D}_1(\mathbf{f}^*) \boldsymbol{\alpha})\} = \mathbf{0} \\ \iff & \mathbb{R}\{c_\ell^{*H} Q^*(f_\ell^*)'\} = 0, \ell = 1, \dots, k. \end{aligned} \quad (\text{F.10})$$

F.2 Showing the Boundedness Property

It remains to show that $Q^*(f)$ satisfies the Boundedness property, for which we follow the arguments of [1]. We start with estimating the coefficient vectors α and β by rewriting (F.6) as

$$\begin{bmatrix} \text{diag}(\mathbf{c}^*) & & \\ & \mathbf{I} & \\ & & \mathbf{I} \end{bmatrix} \Phi (\Phi \nabla^2 G^0(\boldsymbol{\theta}^*) \Phi)^{-1} \Phi \begin{bmatrix} \mathbf{0} \\ \mathbf{s}_R^* \\ \mathbf{s}_I^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_R \\ \boldsymbol{\alpha}_I \end{bmatrix}, \quad (\text{F.11})$$

where $\Phi = \text{diag} \left(\left[\text{diag} \left(\frac{1}{|\mathbf{c}^*|} \right), \mathbf{I}, \mathbf{I} \right] \right)$. Denoting $\Phi := \text{diag}(\mathbf{s}^*)$, we further simplify (F.11) as

$$\begin{bmatrix} -\mathbb{R}\{\Phi^H \mathbf{D}_2(\mathbf{f}^*) \Phi\} & \mathbb{R}\{-\Phi^H \mathbf{D}_1(\mathbf{f}^*)\} & \mathbb{I}\{\Phi^H \mathbf{D}_1(\mathbf{f}^*)\} \\ -\mathbb{R}\{\mathbf{D}_1(\mathbf{f}^*)^H \Phi\} & \mathbf{D}_0(\mathbf{f}^*) & 0 \\ -\mathbb{I}\{\mathbf{D}_1(\mathbf{f}^*)^H \Phi\} & 0 & \mathbf{D}_0(\mathbf{f}^*) \end{bmatrix} \begin{bmatrix} \Phi^{-1} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_R \\ \boldsymbol{\alpha}_I \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{s}_R^* \\ \mathbf{s}_I^* \end{bmatrix}. \quad (\text{F.12})$$

Denote

$$\begin{aligned} \tilde{\mathbf{D}}_2 &= -\mathbb{R}\{\text{diag}(\mathbf{s}^*)^H \mathbf{D}_2(\mathbf{f}^*) \text{diag}(\mathbf{s}^*)\}; \\ \tilde{\mathbf{D}}_1 &= \text{diag}(\mathbf{s}^*)^H \mathbf{D}_1(\mathbf{f}^*); \\ \tilde{\boldsymbol{\beta}} &= \text{diag}(\mathbf{s}^*)^{-1} \boldsymbol{\beta}. \end{aligned}$$

The last two row blocks of (F.12) give

$$\begin{aligned} \boldsymbol{\alpha}_R &= \mathbf{D}_0(\mathbf{f}^*)^{-1} [\mathbf{s}_R^* + \mathbb{R}\{\mathbf{D}_1(\mathbf{f}^*)^H \text{diag}(\mathbf{s}^*)\} \tilde{\boldsymbol{\beta}}]; \\ \boldsymbol{\alpha}_I &= \mathbf{D}_0(\mathbf{f}^*)^{-1} [\mathbf{s}_I^* + \mathbb{I}\{\mathbf{D}_1(\mathbf{f}^*)^H \text{diag}(\mathbf{s}^*)\} \tilde{\boldsymbol{\beta}}] \end{aligned}$$

implying

$$\begin{aligned} \boldsymbol{\alpha} &= \mathbf{D}_0(\mathbf{f}^*)^{-1} [\mathbf{s}^* + \mathbf{D}_1(\mathbf{f}^*)^H \text{diag}(\mathbf{s}^*) \tilde{\boldsymbol{\beta}}] \\ &= \mathbf{D}_0(\mathbf{f}^*)^{-1} [\mathbf{s}^* + \mathbf{D}_1(\mathbf{f}^*)^H \boldsymbol{\beta}] \\ &= \mathbf{s}^* - (\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)^{-1}) \mathbf{s}^* + \mathbf{D}_0(\mathbf{f}^*)^{-1} \mathbf{D}_1(\mathbf{f}^*)^H \boldsymbol{\beta}. \end{aligned} \quad (\text{F.13})$$

Without loss of generality, we assume $\mathbf{e}_1^T \mathbf{s}^* = 1$. Then

$$\alpha_1 = 1 - \left[(\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)^{-1}) \mathbf{s}^* - \mathbf{D}_0(\mathbf{f}^*)^{-1} \mathbf{D}_1(\mathbf{f}^*)^H \boldsymbol{\beta} \right]_1, \quad (\text{F.14})$$

where $[\cdot]_1$ stands for the first entry of a vector. The first row block of (F.12) leads to

$$\tilde{\mathbf{D}}_2 \tilde{\boldsymbol{\beta}} = \mathbb{R}\{\tilde{\mathbf{D}}_1 \boldsymbol{\alpha}_R\} - \mathbb{I}\{\tilde{\mathbf{D}}_1 \boldsymbol{\alpha}_I\} = \mathbb{R}\{\tilde{\mathbf{D}}_1 (\boldsymbol{\alpha}_R + i \boldsymbol{\alpha}_I)\} = \mathbb{R}\{\tilde{\mathbf{D}}_1 \boldsymbol{\alpha}\}.$$

Combining this with (F.13), we get

$$\begin{aligned} \tilde{\mathbf{D}}_2 \tilde{\boldsymbol{\beta}} &= \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1} [\mathbf{s}^* + \mathbf{D}_1^H \boldsymbol{\beta}]\} \\ &= \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1} \mathbf{s}^*\} + \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1}\} \mathbf{D}_1^H \boldsymbol{\beta} \\ &= \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1} \mathbf{s}^*\} + \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1}\} \mathbf{D}_1^H \text{diag}(\mathbf{s}^*) \tilde{\boldsymbol{\beta}} \\ &= \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1} \mathbf{s}^*\} + \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1}\} \tilde{\mathbf{D}}_1^H \tilde{\boldsymbol{\beta}}. \end{aligned}$$

This implies

$$(\tilde{\mathbf{D}}_2 - \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1}\} \tilde{\mathbf{D}}_1^H) \tilde{\boldsymbol{\beta}} = \mathbb{R}\{\tilde{\mathbf{D}}_1 \mathbf{D}_0(\mathbf{f}^*)^{-1} \mathbf{s}^*\}. \quad (\text{F.15})$$

F.2.1 Bounding $\|\tilde{\beta}\|_\infty$

First invoke (A.11) to get

$$\begin{aligned} \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} &\leq \frac{1}{1 - 0.00755}, \\ \{\|\mathbf{D}_1(\mathbf{f}^*)\|_{\infty,\infty}, \|\tilde{\mathbf{D}}_1\|_{\infty,\infty}\}/\sqrt{\tau} &\leq 0.01236n/\sqrt{\tau} \leq 0.00682, \\ \|\mathbf{I} - \tilde{\mathbf{D}}_2/\tau\|_{\infty,\infty} &\leq 0.0171. \end{aligned} \quad (\text{F.16})$$

These inequalities (F.16) immediately lead to

$$\begin{aligned} \|\tau\mathbf{I} - \tilde{\mathbf{D}}_2 + \mathbb{R}\{\tilde{\mathbf{D}}_1\mathbf{D}_0(\mathbf{f}^*)^{-1}\}\tilde{\mathbf{D}}_1^H\|_{\infty,\infty} &\stackrel{\textcircled{1}}{\leq} \tau \left(\|\mathbf{I} - \tilde{\mathbf{D}}_2/\tau\|_{\infty,\infty} + \|\tilde{\mathbf{D}}_1/\sqrt{\tau}\|_{\infty,\infty}^2 \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} \right) \\ &\stackrel{\textcircled{2}}{\leq} \tau (0.0171 + 0.00682^2/(1 - 0.00755)) \leq 0.01715\tau < \tau, \end{aligned} \quad (\text{F.17})$$

where $\textcircled{1}$ follows from the triangle inequality and the sub-multiplicative property of $\ell_{\infty,\infty}$ norm and $\textcircled{2}$ follows from (F.16). This implies that $\tilde{\mathbf{D}}_2 - \mathbb{R}\{\tilde{\mathbf{D}}_1\mathbf{D}_0(\mathbf{f}^*)^{-1}\}\tilde{\mathbf{D}}_1^H$ is nonsingular and well-conditioned. In particular,

$$\|(\tilde{\mathbf{D}}_2 - \mathbb{R}\{\tilde{\mathbf{D}}_1\mathbf{D}_0(\mathbf{f}^*)^{-1}\}\tilde{\mathbf{D}}_1^H)^{-1}\|_{\infty,\infty} \stackrel{\textcircled{1}}{\leq} \frac{1}{\tau(1 - 0.01715)} \leq \frac{1.0175}{\tau},$$

where $\textcircled{1}$ follows from (F.17). Then from (F.15), we have

$$\begin{aligned} \|\tilde{\beta}\|_\infty &\leq \|(\tilde{\mathbf{D}}_2 - \mathbb{R}\{\tilde{\mathbf{D}}_1\mathbf{D}_0(\mathbf{f}^*)^{-1}\}\tilde{\mathbf{D}}_1^H)^{-1}\|_{\infty,\infty} \|\mathbb{R}\{\tilde{\mathbf{D}}_1\mathbf{D}_0(\mathbf{f}^*)^{-1}\mathbf{s}^*\}\|_\infty \\ &\stackrel{\textcircled{1}}{\leq} \|(\tilde{\mathbf{D}}_2 - \mathbb{R}\{\tilde{\mathbf{D}}_1\mathbf{D}_0(\mathbf{f}^*)^{-1}\}\tilde{\mathbf{D}}_1^H)^{-1}\|_{\infty,\infty} \|\tilde{\mathbf{D}}_1\|_{\infty,\infty} \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} \|\mathbf{s}^*\|_\infty \\ &\stackrel{\textcircled{2}}{\leq} \frac{1.0175}{\tau} \frac{0.00682\sqrt{\tau}}{1 - 0.00755} \leq \frac{0.00700}{\sqrt{\tau}}, \end{aligned} \quad (\text{F.18})$$

where $\textcircled{1}$ follows from sub-multiplicative property of the operator norm $\|\cdot\|_{\infty,\infty}$, and $\textcircled{2}$ follows from Eq. (F.16) and $\|\mathbf{s}^*\|_\infty = 1$. This indicates that

$$\|\beta\|_\infty \leq \|\text{diag}(\mathbf{s}^*)\|_{\infty,\infty} \|\tilde{\beta}\|_\infty \leq 0.00700/\sqrt{\tau} \leq 0.00386/n := \beta^\infty, \quad (\text{F.19})$$

where the last inequality follows because $\tau \geq 3.289n^2$ for $n \geq 130$ by (A.2).

F.2.2 Bounding $\|\alpha\|_\infty$ and $\mathbb{R}\{\alpha_1\}$ and $|\mathbb{I}\{\alpha_1\}|$

From (F.13), we have

$$\begin{aligned} \|\alpha\|_\infty &\stackrel{\textcircled{1}}{\leq} \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} \|\mathbf{s}^*\|_\infty + \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} \|\mathbf{D}_1(\mathbf{f}^*)\|_{\infty,\infty} \|\beta\|_\infty \\ &\stackrel{\textcircled{2}}{\leq} \frac{1}{1 - 0.00755} + \frac{0.00682\sqrt{\tau}}{1 - 0.00755} \frac{0.00700}{\sqrt{\tau}} \\ &\leq 1.00766 := \alpha^\infty, \end{aligned} \quad (\text{F.20})$$

where $\textcircled{1}$ follows from the triangle inequality and the fact that $\|\mathbf{A}\mathbf{B}\mathbf{x}\|_\infty \leq \|\mathbf{A}\|_{\infty,\infty} \|\mathbf{B}\|_{\infty,\infty} \|\mathbf{x}\|_\infty$. $\textcircled{2}$ holds since $\|\mathbf{s}^*\|_\infty = 1$.

Second, recognizing that $\alpha_1 = 1 - [(\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)^{-1})\mathbf{s}^* - \mathbf{D}_0(\mathbf{f}^*)^{-1}\mathbf{D}_1(\mathbf{f}^*)^H\beta]_1$ by Eq. (F.14), we have $\mathbb{R}\{\alpha_1\} =$

$1 - [\mathbb{R}\{(\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)^{-1})\mathbf{s}^* - \mathbf{D}_0(\mathbf{f}^*)^{-1}\mathbf{D}_1(\mathbf{f}^*)^H\boldsymbol{\beta}\}]_1$. We further get an upper bound as follows

$$\begin{aligned}
& |[\mathbb{R}\{(\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)^{-1})\mathbf{s}^* - \mathbf{D}_0(\mathbf{f}^*)^{-1}\mathbf{D}_1(\mathbf{f}^*)^H\boldsymbol{\beta}\}]_1| \\
& \stackrel{\textcircled{1}}{\leq} \|(\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)^{-1})\mathbf{s}^* - \mathbf{D}_0(\mathbf{f}^*)^{-1}\mathbf{D}_1(\mathbf{f}^*)^H\boldsymbol{\beta}\|_\infty \\
& \stackrel{\textcircled{2}}{\leq} \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} \|\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)\|_{\infty,\infty} \|\mathbf{s}^*\|_\infty + \|\mathbf{D}_0(\mathbf{f}^*)^{-1}\|_{\infty,\infty} \|\mathbf{D}_1(\mathbf{f}^*)\|_{\infty,\infty} \|\boldsymbol{\beta}\|_\infty \\
& \leq \frac{0.00755}{1 - 0.00755} + \frac{0.00682\sqrt{\tau}}{1 - 0.00755} \frac{0.00700}{\sqrt{\tau}} \\
& \leq 0.00766,
\end{aligned}$$

where $\textcircled{1}$ follows from the real part of the first entry of a vector is no larger than the infinity norm of this vector and $\textcircled{2}$ follows from the triangle inequality and the sub-multiplicative property of infinity operator norm that $\|\mathbf{A}\mathbf{B}\mathbf{x}\|_\infty \leq \|\mathbf{A}\|_{\infty,\infty} \|\mathbf{B}\|_{\infty,\infty} \|\mathbf{x}\|_\infty$. The last inequality follows from Eq. (F.16) and (F.19). Combining the above arguments yields

$$\mathbb{R}\{\alpha_1\} \geq 1 - 0.00766 \quad \text{and} \quad |\mathbb{I}\{\alpha_1\}| \leq 0.00766. \quad (\text{F.21})$$

We are ready to show the Boundedness property following the simplifications used in [1]. In particular, fix an arbitrary point $f_0^* \in T^*$ as the reference point and let f_{-1}^* be the first frequency in T^* that lies on the left of f_0^* while f_1^* be the first frequency in T^* that lies on the right. Here “left” and “right” are directions on the complex circle \mathbb{T} . We remark that the analysis depends only on the relative locations of $\{f_\ell^*\}$. Hence, to simplify the arguments, we assume that the reference point f_0^* is at 0 by shifting the frequencies if necessary. Then we divide the region between $f_0^* = 0$ and $f_1^*/2$ into three parts: Near Region $\mathcal{N} := [0, 0.24/n]$, Middle Region $\mathcal{M} := [0.24/n, 0.75/n]$ and Far Region $\mathcal{F} := [0.75/n, f_1^*/2]$. Also their symmetric counterparts: $-\mathcal{N} := [-0.24/n, 0]$, $-\mathcal{M} := [-0.75/n, -0.24/n]$, and $-\mathcal{F} := [f_{-1}^*/2, -0.75/n]$. We first show that the dual polynomial has strictly negative curvature $|Q^*(f)|'' < 0$ in $\mathcal{N} = [0, 0.24/n]$ and $|Q^*(f)| < 1$ in $\mathcal{M} \cup \mathcal{F} = [0.24/n, f_1^*/2]$, implying $|Q^*(f)| < 1$ in $\mathcal{N} \cup \mathcal{M} \cup \mathcal{F} \setminus \{f_0^*\}$ by exploiting $|Q^*(f_0^*)| = 1$ and $|Q^*(f_0^*)'| = 0$. Then using the same symmetric arguments in [1], we claim that $|Q^*(f)| < 1$ in $(-\mathcal{N}) \cup (-\mathcal{M}) \cup (-\mathcal{F}) \setminus \{f_0^*\}$. Combining these two results with the fact that the reference point f_0^* is chosen arbitrarily from T^* (and shifted to 0), we establish that the Boundedness property of $Q^*(f)$ holds in the entire $\mathbb{T} \setminus T^*$.

F.2.3 Controlling $Q^*(f)$ in Near Region

For $f \in \mathcal{N}$, the second-order Taylor expansion of $|Q^*(f)|$ at $f_0^* = 0$ states

$$\begin{aligned}
|Q^*(f)| &= |Q^*(f_0^*)| + (f - f_0^*)|Q^*(f_0^*)'| + \frac{1}{2}(f - f_0^*)^2|Q^*(\xi)|'' \\
&= 1 + (f - f_0^*)|Q^*(f_0^*)'| + \frac{1}{2}(f - f_0^*)^2|Q^*(\xi)|'' \text{ for some } \xi \in \mathcal{N},
\end{aligned} \quad (\text{F.22})$$

with the second line following from the Interpolation property. We argue that

$$|Q^*(f_0^*)'| = \frac{Q_R^*(f_0^*)Q_R^*(f_0^*)' + Q_I^*(f_0^*)Q_I^*(f_0^*)'}{|Q^*(f_0^*)|} = \frac{\mathbb{R}\{c_0^*\}Q_R^*(f_0^*)' + \mathbb{I}\{c_0^*\}Q_I^*(f_0^*)'}{|c_0^*||Q^*(f_0^*)|} = \frac{\mathbb{R}\{c_0^{*H}Q^*(f_0^*)'\}}{|c_0^*||Q^*(f_0^*)|} = 0.$$

The last equality is due to (F.10). Together with (F.22), to bound $|Q^*(f)|$ strictly below 1, we only need to show the concavity of $|Q^*(f)|$ in Near Region (i.e., $|Q^*(f)|'' < 0$ for $f \in \mathcal{N}$). Since

$$|Q^*(f)|'' = -\frac{(Q_R^*(f)Q_R^*(f)') + Q_I^*(f)Q_I^*(f)')^2}{|Q^*(f)|^3} + \frac{Q_R^*(f)Q_R^*(f)'' + |Q^*(f)|'^2 + |Q_I^*(f)|Q_I^*(f)''}{|Q^*(f)|},$$

we only need to show that

$$Q_R^*(f)Q_R^*(f)'' + |Q^*(f)|'^2 + |Q_I^*(f)|Q_I^*(f)'' < 0.$$

Recall the expression for $Q^*(f)$ given in Eq. (F.8)

$$Q^*(f) = \sum_{f_\ell^* \in T^*} \alpha_\ell K(f_\ell^* - f) + \sum_{f_\ell^* \in T^*} \beta_\ell K'(f_\ell^* - f).$$

To bound the real part of $Q^*(f)$ in $\mathcal{N} = [0, 0.24/n]$, we observe

$$\begin{aligned} Q_R^*(f) &\geq \mathbb{R}\{\alpha_1 K(f)\} - \alpha^\infty \sum_{f_\ell^* \in T^* \setminus \{0\}} |K(f - f_\ell^*)| - \beta^\infty |K'(f)| - \beta^\infty \sum_{f_\ell^* \in T^* \setminus \{0\}} |K'(f - f_\ell^*)| \\ &\geq \mathbb{R}\{\alpha_1\} \min_{f \in \mathcal{N}} K(f) - \alpha^\infty F_0(2.5/n, f) - \beta^\infty (\max_{f \in \mathcal{N}} |K'(f)| + F_1(2.5/n, f)) \\ &\geq (1 - 0.00766)(0.905252) - (1.00766)0.00757 - (0.00386/n)(0.789569n + 0.01241n) \\ &\geq 0.887594, \end{aligned}$$

where the first inequality follows from an application of the triangle inequality, and the second is from Lemma A.2. The third inequality follows from evaluating $F_0(2.5/n, f)$ and $F_1(2.5/n, f)$ at $f = 0.24/n$, the numerical bounds in Tables A.5 and A.5 of Appendix A.5 and Eq. (F.19), (F.20), (F.21), as well as $\min_{f \in \mathcal{N}} K(f) \geq 0.905252$. This last bound follows from [1, Eq. (2.20), set $f_c = n - 2$] that $K(f) \geq 1 - \frac{\pi^2}{6}(n - 2)(n + 2)f^2$. Hence

$$\min_{f \in \mathcal{N}} K(f) \geq \min_{f \in \mathcal{N}} 1 - \frac{\pi^2}{6}(n - 2)(n + 2)f^2 \geq 1 - \frac{\pi^2}{6}(n - 2)(n + 2)(0.24/n)^2 \geq 0.905252.$$

Similarly, Combining Eq. (F.19), (F.20), (F.21), the upper bounds on $F_\ell(2.5/n, 0.24/n)$ in Table A.5 and the upper bounds for $\max_{f \in \mathcal{N}} |K^{(\ell)}(f)|$ and $\max_{f \in \mathcal{N}} K''(f)$ in Table A.5, we get

$$\begin{aligned} Q_R^{*''}(f) &\leq \mathbb{R}\{\alpha_1\} \max_{f \in \mathcal{N}} K''(f) + \alpha^\infty F_2(2.5/n, 0.24/n) + \beta^\infty (\max_{f \in \mathcal{N}} |K'''(f)| + F_3(2.5/n, 0.24/n)) \\ &\leq (1 - 0.00766)(-2.35084n^2) + (1.00766)(0.05637n^2) + (0.00386/n)(7.79273n^3 + 0.28838n^3) \\ &\leq -2.24483n^2; \\ |Q_I^*(f)| &\leq |\mathbb{I}\{\alpha_1\}| \max_{f \in \mathcal{N}} K(f) + \alpha^\infty F_0(2.5/n, 0.24/n) + \beta^\infty (\max_{f \in \mathcal{N}} |K'(f)| + F_1(2.5/n, 0.24/n)) \\ &\leq (0.00766) \times 1 + (1.00766)0.00757 + (0.00386/n)(0.789569n + 0.01241n) \\ &\leq 0.0183836; \\ |Q_I^{*''}(f)| &\leq |\mathbb{I}\{\alpha_1\}| \max_{f \in \mathcal{N}} |K''(f)| + \alpha^\infty F_2(2.5/n, 0.24/n) + \beta^\infty (\max_{f \in \mathcal{N}} |K'''(f)| + F_3(2.5/n, 0.24/n)) \\ &\leq (0.00766)(3.290n^2) + (1.00766)(0.05637n^2) + (0.00386/n)(7.79273n^3 + 0.28838n^3) \\ &\leq 0.113197n^2; \\ |Q^{*'}(f)| &\leq \alpha^\infty (\max_{f \in \mathcal{N}} |K'(f)| + F_1(2.5/n, 0.24/n)) + \beta^\infty (\max_{f \in \mathcal{N}} |K''(f)| + F_2(2.5/n, 0.24/n)) \\ &\leq (1.00766)(0.789569n + 0.01241n) + (0.00386/n)(3.290n^2 + 0.05637n^2) \\ &\leq 0.821039n; \\ |Q^{*''}(f)| &\leq \alpha^\infty (\max_{f \in \mathcal{N}} |K''(f)| + F_2(2.5/n, 0.24/n)) + \beta^\infty (\max_{f \in \mathcal{N}} |K'''(f)| + F_3(2.5/n, 0.24/n)) \\ &\leq (1.00766)(3.290n^2 + 0.05637n^2) + (0.00386/n)(7.79273n^3 + 0.28838n^3) \\ &\leq 3.40320n^2. \end{aligned}$$

Combining the lower bound on $Q_R^*(f)$ and the upper bounds on $Q_R^*(f)''$, $|Q^*(f)'|$, $|Q_I^*(f)|$ and $|Q_I^*(f)''|$, we arrive at

$$|Q^*(f)|'' = Q_R^*(f)Q_R^*(f)'' + |Q^*(f)'|^2 + |Q_I^*(f)||Q_I^*(f)''| \leq -1.316313n^2 < 0 \text{ in } \mathcal{N}.$$

F.2.4 Bounding $|Q^*(f)|$ in Middle Region

For upperbounding $|Q^*(f)|$ for $f \in \mathcal{M} = [0.24/n, 0.75/n]$, we firstly apply the triangle inequality

$$\begin{aligned}
|Q^*(f)| &= \left| \sum_{f_\ell^* \in T^*} \alpha_\ell K(f_\ell^* - f) + \sum_{f_\ell^* \in T^*} \beta_\ell K'(f_\ell^* - f) \right| \\
&\leq \|\alpha\|_\infty \left(|K(f)| + \sum_{f_\ell^* \in T^* \setminus \{0\}} |K(f - f_\ell^*)| \right) + \|\beta\|_\infty \left(|K'(f)| + \sum_{f_\ell^* \in T^* \setminus \{0\}} |K'(f - f_\ell^*)| \right) \\
&\leq \alpha^\infty |K(f)| + \beta^\infty |K'(f)| + \alpha^\infty F_0(2.5/n, f) + \beta^\infty F_1(2.5/n, f),
\end{aligned} \tag{F.23}$$

where the last inequality is from Lemma A.2. We then follow [1, Eq. (2.29)] to upperbound the first two terms in the last line

$$|K(f)| \leq 1 - \frac{\pi^2(n^2 - 4)f^2}{6} + \frac{\pi^4 n^4 f^4}{72} \quad \text{and} \quad |K'(f)| \leq \frac{\pi^2(n^2 - 4)f}{3}, \quad \text{for } f \in [-1/2, 1/2].$$

The rest of argument consists of defining

$$\begin{aligned}
L_1(f) &= \alpha^\infty \left(1 - \frac{1}{6} \pi^2(n^2 - 4)f^2 + \frac{1}{72} \pi^4 n^4 f^4 \right) + \beta^\infty \frac{1}{3} \pi^2(n^2 - 4)f; \\
L_2(f) &= \alpha^\infty F_0(2.5/n, f) + \beta^\infty F_1(2.5/n, f)
\end{aligned}$$

with the derivative of $L_1(f)$ given by

$$L_1'(f) = -\alpha^\infty \left(\frac{\pi^2(n^2 - 4)f}{3} - \frac{\pi^4 n^4 f^3}{18} \right) + \beta^\infty \frac{\pi^2(n^2 - 4)}{3} < 0, \quad \text{for } f \in \mathcal{M},$$

implying that $L_1(f)$ is decreasing. Also, $L_2(f)$ is increasing in \mathcal{M} by Lemma A.2. Hence, by the monotonic property, we have

$$|Q^*(f)| \leq L_1(0.24/n) + L_2(0.75/n) \leq 0.919779 + 0.007836 = 0.927615 < 1.$$

F.2.5 Bounding $|Q^*(f)|$ in Far Region.

Recall that $f_0^* = 0$ is the reference point. To simplify notation, we re-index the frequencies such that $\dots \leq f_{-1}^* < f_0^* = 0 < f_1^* < \dots$. For $f \in \mathcal{F} = [0.75/n, f_1^*/2] = [0.75/n, f_1^* - f_1^*/2]$, by Lemma A.3, we have

$$\begin{aligned}
\sum_j |K^{(\ell)}(f - f_j^*)| &\leq W_\ell(0.75/n, f_1^*/2) \stackrel{\textcircled{1}}{=} \sum_{j \geq 0} B_\ell(j(2.5/n) + 0.75/n) + \sum_{j \geq 0} B_\ell(j(2.5/n) + f_1^*/2) \\
&\stackrel{\textcircled{2}}{\leq} \sum_{j \geq 0} B_\ell(j(2.5/n) + 0.75/n) + \sum_{j \geq 0} B_\ell(j(2.5/n) + 1.25/n) \\
&\stackrel{\textcircled{3}}{=} W_\ell(0.75/n, 1.25/n),
\end{aligned} \tag{F.24}$$

where ① follow from the definition of $W(\underline{f}, \bar{f})$ in Lemma A.3, ② follows $f_1^*/2 = (f_1^* - f_0^*)/2 \geq \Delta_{\min}/2 = 1.25/n$ and decreasing property of $B_\ell(\cdot)$, and ③ follows from the definition of $W(\underline{f}, \bar{f})$.

Finally, applying (F.24), (F.19) and (F.20) to (F.8), we arrive at

$$\begin{aligned}
|Q^*(f)| &\leq \alpha^\infty \sum_\ell |K(f - f_\ell^*)| + \beta^\infty \sum_\ell |K'(f - f_\ell^*)| \\
&\leq 1.00766 W_0(0.75/n, 1.25/n) + (0.00386/n) W_1(0.75/n, 1.25/n) \\
&\leq 1.00766(0.70859) + (0.00386/n)(5.2084n) \\
&= 0.734123.
\end{aligned}$$

This concludes the proof of Lemma 4.3. □

Appendix G Proof of Lemma 4.4

Proof. We exploit the closeness of θ^* and θ^λ (see Lemma 4.1) to bound the pointwise distance between $Q^*(f)$ and $Q^\lambda(f)$. Note

$$Q^\lambda(f) - Q^*(f) = \mathbf{a}(f)^H \mathbf{Z}(\mathbf{q}^\lambda - \mathbf{q}^*) = \mathbf{a}(f)^H \mathbf{Z} \left(\frac{\mathbf{x}^* - \mathbf{x}^\lambda}{\lambda} + \frac{d}{d\lambda} \mathbf{x}^\lambda \Big|_{\lambda=0} \right) = \frac{1}{\lambda} \int_0^\lambda \mathbf{a}(f)^H \mathbf{Z} \left(\frac{d}{dt} \mathbf{x}^* - \frac{d}{dt} \mathbf{x}^t \right) dt,$$

which implies that

$$|Q^\lambda(f) - Q^*(f)| \leq \max_{0 \leq t \leq \lambda} \left| \mathbf{a}(f)^H \mathbf{Z} \left(\frac{d}{dt} \mathbf{x}^* - \mathbf{a}(f)^H \mathbf{Z} \left(\frac{d}{dt} \mathbf{x}^t \right) \right) \right|. \quad (\text{G.1})$$

We can also obtain similar bounds on the pointwise distances between derivatives of $Q^\lambda(f)$ and $Q^*(f)$.

Recall from Eq. (F.2), (4.5), and (F.5) that

$$\frac{d}{d\lambda} \mathbf{x}^\lambda = -[\mathbf{A}'(\mathbf{f}^\lambda) \text{diag}(\mathbf{c}^\lambda) \quad \mathbf{A}(\mathbf{f}^\lambda) \quad i\mathbf{A}(\mathbf{f}^\lambda)] (\nabla^2 G^\lambda(\theta^\lambda))^{-1} \boldsymbol{\rho}^\lambda, \quad (\text{G.2})$$

$$\frac{d}{d\lambda} \mathbf{x}^* = -[\mathbf{A}'(\mathbf{f}^*) \text{diag}(\mathbf{c}^*) \quad \mathbf{A}(\mathbf{f}^*) \quad i\mathbf{A}(\mathbf{f}^*)] (\nabla^2 G^0(\theta^*))^{-1} \boldsymbol{\rho}^*, \quad (\text{G.3})$$

where $\boldsymbol{\rho}^* = [\mathbf{0}^T \quad \mathbb{R}\{\text{sign}(\mathbf{c}^*)\}^T \quad \mathbb{I}\{\text{sign}(\mathbf{c}^*)\}^T]^T$ and $\boldsymbol{\rho}^\lambda = [\mathbf{0}^T \quad \mathbb{R}\{\text{sign}(\mathbf{c}^\lambda)\}^T \quad \mathbb{I}\{\text{sign}(\mathbf{c}^\lambda)\}^T]^T$.

Multiplying both sides of Eq. (G.2) and (G.3) by $-\mathbf{a}(f)^H \mathbf{Z}$ (and then inserting $\mathbf{W}^{*\frac{1}{2}} \mathbf{W}^{*\frac{1}{2}}$ (which equals \mathbf{I}) into the spaces before and after $(\nabla^2 G^0(\theta^*))^{-1}$ (and $(\nabla^2 G^\lambda(\theta^\lambda))^{-1}$) yield

$$\begin{aligned} -\mathbf{a}(f)^H \mathbf{Z} \left(\frac{d}{d\lambda} \mathbf{x}^\lambda \right) &= \boldsymbol{\nu}^\lambda(f) \Xi^\lambda \boldsymbol{\rho}^\lambda, \\ -\mathbf{a}(f)^H \mathbf{Z} \left(\frac{d}{d\lambda} \mathbf{x}^* \right) &= \boldsymbol{\nu}^*(f) \Xi^* \boldsymbol{\rho}^*. \end{aligned}$$

Here

$$\begin{aligned} \boldsymbol{\nu}^\lambda(f) &:= [\mathbf{D}_1(f, \mathbf{f}^\lambda) \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1} \quad \mathbf{D}_0(f, \mathbf{f}^\lambda) \quad i\mathbf{D}_0(f, \mathbf{f}^\lambda)], \\ \boldsymbol{\nu}^*(f) &:= [\mathbf{D}_1(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} \quad \mathbf{D}_0(f, \mathbf{f}^*) \quad i\mathbf{D}_0(f, \mathbf{f}^*)], \end{aligned} \quad (\text{G.4})$$

with $\mathbf{D}_\ell(f, \mathbf{f}^\lambda)$ a row vector defined by $\mathbf{D}_\ell(f, \mathbf{f}^\lambda) := [K_\ell(f_1^\lambda - f), \dots, K_\ell(f_k^\lambda - f)]$, and

$$\begin{aligned} \Xi^\lambda &:= \Upsilon(\nabla^2 G^\lambda(\theta^\lambda))^{-1}, \\ \Xi^* &:= \Upsilon(\nabla^2 G^0(\theta^*))^{-1}, \end{aligned}$$

where $\Upsilon(\cdot) := \mathbf{W}^{*\frac{1}{2}}(\cdot)\mathbf{W}^{*\frac{1}{2}}$ is a linear operator that normalizes the Hessian matrix so that it is close to the identity. As a consequence, we bound the integrand of (G.1) as follows

$$\begin{aligned} &|\boldsymbol{\nu}^*(f) \Xi^* \boldsymbol{\rho}^* - \boldsymbol{\nu}^\lambda(f) \Xi^\lambda \boldsymbol{\rho}^\lambda| \\ &\leq |\boldsymbol{\nu}^*(f) \Xi^* (\boldsymbol{\rho}^* - \boldsymbol{\rho}^\lambda)| + |\boldsymbol{\nu}^*(f) (\Xi^* - \Xi^\lambda) \boldsymbol{\rho}^\lambda| + |(\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)) \Xi^\lambda \boldsymbol{\rho}^\lambda| \\ &\leq \|\boldsymbol{\nu}^*(f)\|_1 \|\Xi^*\|_{\infty, \infty} \|\boldsymbol{\rho}^* - \boldsymbol{\rho}^\lambda\|_\infty + \|\boldsymbol{\nu}^*(f)\|_1 \|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \|\boldsymbol{\rho}^\lambda\|_\infty + \|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1 \|\Xi^\lambda\|_{\infty, \infty} \|\boldsymbol{\rho}^\lambda\|_\infty, \end{aligned} \quad (\text{G.5})$$

where the first line follows from the triangle inequality and the second line follows from Hölder's inequality and the sub-multiplicative property of the $\ell_{\infty, \infty}$ norm. We next develop upper bounds on $\|\boldsymbol{\nu}^*(f)\|_1$, $\|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1$, $\|\Xi^*\|_{\infty, \infty}$, $\|\Xi^* - \Xi^\lambda\|_{\infty, \infty}$, $\|\Xi^\lambda\|_{\infty, \infty}$, $\|\boldsymbol{\rho}^* - \boldsymbol{\rho}^\lambda\|_1$ and $\|\boldsymbol{\rho}^\lambda\|_\infty$.

Bounding $\|\Xi^*\|_{\infty, \infty}$ and $\|\Xi^\lambda\|_{\infty, \infty}$ and $\|\Xi^* - \Xi^\lambda\|_{\infty, \infty}$.

We note that both $\Xi^{*-1} = \Upsilon(\nabla^2 G^0(\theta^*))$ and $\Xi^{\lambda-1} = \Upsilon(\nabla^2 G^\lambda(\theta^\lambda))$ are close to the identity matrix. More

precisely, we have

$$\begin{aligned}
\|\mathbf{I} - \Xi^{\star-1}\|_{\infty} &\stackrel{\textcircled{1}}{=} \|\mathbf{I} - \Upsilon(\nabla^2 G^0(\boldsymbol{\theta}^*))\|_{\infty, \infty} \\
&\stackrel{\textcircled{2}}{\leq} [\|\mathbf{I} - \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)^H (-\mathbf{D}_2(\mathbf{f}^*)/\tau) \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\|_{\infty, \infty} + 2\|\text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\mathbf{D}_1(\mathbf{f}^*)/\sqrt{\tau}\|_{\infty, \infty}] \\
&\quad \vee (\|\text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\mathbf{D}_1(\mathbf{f}^*)/\sqrt{\tau}\|_{\infty, \infty} + \|\mathbf{I} - \mathbf{D}_0(\mathbf{f}^*)\|_{\infty, \infty}) \\
&\stackrel{\textcircled{3}}{\leq} \|\mathbf{I} - (-\mathbf{D}_2(\mathbf{f}^*)/\tau)\|_{\infty, \infty} + 2\|\mathbf{D}_1(\mathbf{f}^*)/\sqrt{\tau}\|_{\infty, \infty} \\
&\stackrel{\textcircled{4}}{\leq} 0.0171 + 2 \times 0.00682 \\
&\leq 0.03074,
\end{aligned}$$

where $a \vee b := \max(a, b)$. $\textcircled{1}$ follows from definition of Ξ^{\star} and $\textcircled{2}$ follows from applying the triangle inequality to the expression of $[\mathbf{I} - \Upsilon(\nabla^2 G^0(\boldsymbol{\theta}^*))]$. $\textcircled{3}$ follows since the infinity norm of any sign vector is 1, bounding $\|\text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\mathbf{D}_1(\mathbf{f}^*)/\sqrt{\tau}\|_{\infty, \infty}$ is equivalent to bound $\|\mathbf{D}_1(\mathbf{f}^*)/\sqrt{\tau}\|_{\infty, \infty}$. Finally, $\textcircled{4}$ follows from Eq. (A.11). This leads to

$$\|\Xi^{\star}\|_{\infty, \infty} \leq \frac{1}{1 - \|\mathbf{I} - \Xi^{\star-1}\|_{\infty, \infty}} \leq \frac{1}{1 - 0.03074} \leq 1.03172. \quad (\text{G.6})$$

According to (D.5), we have

$$\|\mathbf{I} - \Xi^{\lambda-1}\|_{\infty, \infty} = \|\mathbf{I} - \Upsilon(\nabla^2 G^{\lambda}(\boldsymbol{\theta}^{\lambda}))\|_{\infty, \infty} \leq 0.08561,$$

yielding

$$\|\Xi^{\lambda}\|_{\infty, \infty} \leq \frac{1}{1 - \|\mathbf{I} - \Xi^{\lambda-1}\|_{\infty, \infty}} \leq \frac{1}{1 - 0.08561} \leq 1.09363. \quad (\text{G.7})$$

Next, note

$$\|\Xi^{\star-1} - \Xi^{\lambda-1}\|_{\infty, \infty} = \|\Upsilon(\nabla^2 G^0(\boldsymbol{\theta}^*)) - \Upsilon(\nabla^2 G^{\lambda}(\boldsymbol{\theta}^{\lambda}))\|_{\infty, \infty} \leq \max\{\Pi_1, \Pi_2, \Pi_3\},$$

where Π_1, Π_2, Π_3 denote the first, second and third absolute row block sums of $[\Upsilon(\nabla^2 G^0(\boldsymbol{\theta}^*)) - \Upsilon(\nabla^2 G^{\lambda}(\boldsymbol{\theta}^{\lambda}))]$. We first bound Π_1 as follows

$$\begin{aligned}
\Pi_1 &= \|\text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}) \text{diag}(\mathbf{c}/|\mathbf{c}^*|) - \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}^*) \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\|_{\infty, \infty}/\tau \\
&\quad + 2\|\text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H \mathbf{D}_1(\mathbf{f}) - \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)^H \mathbf{D}_1(\mathbf{f}^*)\|_{\infty, \infty}/\sqrt{\tau} \\
&\quad + 2\|\text{diag}(\mathbf{1}/|\mathbf{c}^*|) [\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_1(\mathbf{f})\mathbf{c}]\|_{\infty}/\sqrt{\tau} \\
&\quad + \|\text{diag}(\mathbf{c}/|\mathbf{c}^*|^2)^H [\mathbf{D}_2(\mathbf{f}, \mathbf{f}^*)\mathbf{c}^* - \mathbf{D}_2(\mathbf{f})\mathbf{c}]\|_{\infty}/\tau \\
&\stackrel{\textcircled{1}}{\leq} [2.19778X^{\star}\gamma + 1.14168(X^{\star}\gamma)^2] + 2[1.48286X^{\star}\gamma + 1.47604(X^{\star}\gamma)^2] + 2(0.75038)X^{\star}B^{\star}\gamma + 1.14168X^{\star}B^{\star}\gamma \\
&\leq 7.81004X^{\star}B^{\star}\gamma \text{ (by } B^{\star}X^{\star}\gamma \leq 10^{-3}\text{)}, \quad (\text{G.8})
\end{aligned}$$

where $\textcircled{1}$ follows from combining Eq. (A.19)-(A.20) and (G.9)-(G.10), where (G.9)-(G.10) are given by

$$\begin{aligned}
&\|\text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}) \text{diag}(\mathbf{c}/|\mathbf{c}^*|) - \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}^*) \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\|_{\infty, \infty}/\tau \\
&\leq \|\text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}) \text{diag}((\mathbf{c} - \mathbf{c}^*)/|\mathbf{c}^*|)\|_{\infty, \infty}/\tau \\
&\quad + \|\text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H (\mathbf{D}_2(\mathbf{f}) - \mathbf{D}_2(\mathbf{f}^*)) \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\|_{\infty, \infty}/\tau \\
&\quad + \|\text{diag}((\mathbf{c} - \mathbf{c}^*)/|\mathbf{c}^*|)^H \mathbf{D}_2(\mathbf{f}^*) \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)\|_{\infty, \infty}/\tau \\
&\leq (1 + X^{\star}\gamma)(1.05610)(X^{\star}\gamma) + (1 + X^{\star}\gamma)(0.08558X^{\star}\gamma) + (X^{\star}\gamma)(1.05610) \quad (\text{by (A.18) and (A.11)}) \\
&\leq 2.19778X^{\star}\gamma + 1.14168(X^{\star}\gamma)^2 \quad (\text{G.9})
\end{aligned}$$

and

$$\begin{aligned}
& \left\| \text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H \mathbf{D}_1(\mathbf{f}) - \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|)^H \mathbf{D}_1(\mathbf{f}^*) \right\|_{\infty, \infty} / \sqrt{\tau} \\
& \leq \left\| \text{diag}(\mathbf{c}/|\mathbf{c}^*|)^H (\mathbf{D}_1(\mathbf{f}) - \mathbf{D}_1(\mathbf{f}^*)) \right\|_{\infty, \infty} / \sqrt{\tau} + \left\| \text{diag}((\mathbf{c} - \mathbf{c}^*)/|\mathbf{c}^*|)^H \mathbf{D}_1(\mathbf{f}^*) \right\|_{\infty, \infty} / \sqrt{\tau} \\
& \leq (1 + X^* \gamma)(1.47604 X^* \gamma) + (X^* \gamma)(0.00682) \quad (\text{by Eq. (A.17) and (A.11)}) \\
& \leq 1.48286 X^* \gamma + 1.47604 (X^* \gamma)^2. \quad (\text{G.10})
\end{aligned}$$

We next bound Π_2 and Π_3 :

$$\begin{aligned}
\{\Pi_2, \Pi_3\} & \stackrel{\textcircled{1}}{\leq} \left\| \mathbf{D}_1(\mathbf{f}) \text{diag}(\mathbf{c}/|\mathbf{c}^*|) - \mathbf{D}_1(\mathbf{f}^*) \text{diag}(\mathbf{c}^*/|\mathbf{c}^*|) \right\|_{\infty, \infty} / \sqrt{\tau} \\
& \quad + \left\| \mathbf{D}_0(\mathbf{f}) - \mathbf{D}_0(\mathbf{f}^*) \right\|_{\infty, \infty} + \left\| \text{diag}(1/|\mathbf{c}^*|) [\mathbf{D}_1(\mathbf{f}, \mathbf{f}^*) \mathbf{c}^* - \mathbf{D}_1(\mathbf{f}) \mathbf{c}] \right\|_{\infty} / \sqrt{\tau} \\
& \quad + \lambda \left\| \mathbf{u} \odot \mathbf{u} / |\mathbf{c}|^3 - \mathbf{u}^* \odot \mathbf{u}^* / |\mathbf{c}^*|^3 \right\|_{\infty} + \lambda \left\| \mathbf{u} \odot \mathbf{v} / |\mathbf{c}|^3 - \mathbf{u}^* \odot \mathbf{v}^* / |\mathbf{c}^*|^3 \right\|_{\infty} \\
& \stackrel{\textcircled{2}}{\leq} [1.48286 X^* \gamma + 1.47604 (X^* \gamma)^2] + 0.01516 X^* \gamma + 0.75038 X^* B^* \gamma + 2(0.646 X^* \gamma)(5.00701) X^* \gamma \\
& \leq 2.25636 X^* B^* \gamma \\
& < \Pi_1 \quad (\text{by } B^* X^* \gamma \leq 10^{-3}),
\end{aligned}$$

where $\textcircled{1}$ follows from the triangle inequality and $\textcircled{2}$ follows by combining Eq. (G.10), (A.16), (A.19), and (G.11). To show (G.11), we assume the norm $\left\| \mathbf{u} \odot \mathbf{u} / |\mathbf{c}|^3 - \mathbf{u}^* \odot \mathbf{u}^* / |\mathbf{c}^*|^3 \right\|_{\infty}$ is achieved by the ℓ th entry and proceed as

$$\begin{aligned}
\left| \frac{u_{\ell}^2}{|c_{\ell}|^3} - \frac{u_{\ell}^{*2}}{|c_{\ell}^*|^3} \right| & \stackrel{\textcircled{1}}{\leq} \left| \frac{c_{\ell}^2}{|c_{\ell}|^3} - \frac{c_{\ell}^{*2}}{|c_{\ell}^*|^3} \right| \\
& \stackrel{\textcircled{2}}{\leq} \frac{|c_{\ell}^2 - c_{\ell}^{*2}|}{|c_{\ell}^*|^3} + |c_{\ell}|^2 \left| \frac{1}{|c_{\ell}|^3} - \frac{1}{|c_{\ell}^*|^3} \right| \\
& \stackrel{\textcircled{3}}{\leq} \frac{X^* \gamma}{c_{\min}^*} \left((2 + X^* \gamma) + \frac{(X^* \gamma)^2 + 3(X^* \gamma) + 3}{1 - X^* \gamma} \right) \leq \frac{X^* \gamma}{c_{\min}^*} (5.00701),
\end{aligned} \quad (\text{G.11})$$

where $\textcircled{1}$ follows from $|\mathbb{R}\{a\}| \leq |a|$ for all $a \in \mathbb{C}$ and $\textcircled{2}$ follows from the triangle inequality. $\textcircled{3}$ follows from Eq. (G.12) and (G.13) given below:

$$\frac{|c_{\ell}^2 - c_{\ell}^{*2}|}{|c_{\ell}^*|^3} \leq \frac{1}{|c_{\ell}^*|} \frac{|c_{\ell} - c_{\ell}^*|}{|c_{\ell}^*|} \frac{|c_{\ell} + c_{\ell}^*|}{|c_{\ell}^*|} \leq \frac{X^* \gamma (2 + X^* \gamma)}{c_{\min}^*} \quad (\text{G.12})$$

and

$$\begin{aligned}
|c_{\ell}|^2 \left| \frac{1}{|c_{\ell}|^3} - \frac{1}{|c_{\ell}^*|^3} \right| & = |c_{\ell}|^2 \left| \frac{1}{|c_{\ell}|} - \frac{1}{|c_{\ell}^*|} \right| \left(\frac{1}{|c_{\ell}|^2} + \frac{1}{|c_{\ell}^*|^2} + \frac{1}{|c_{\ell}^*||c_{\ell}|} \right) \leq \frac{|c_{\ell} - c_{\ell}^*|}{|c_{\ell}||c_{\ell}^*|} \left(1 + \frac{|c_{\ell}|^2}{|c_{\ell}^*|^2} + \frac{|c_{\ell}|}{|c_{\ell}^*|} \right) \\
& \leq \frac{1}{|c_{\ell}|} X^* \gamma \left(1 + \frac{|c_{\ell}|^2}{|c_{\ell}^*|^2} + \frac{|c_{\ell}|}{|c_{\ell}^*|} \right) \\
& \leq \frac{1}{c_{\min}^* (1 - X^* \gamma)} X^* \gamma \left(1 + \frac{|c_{\ell}|^2}{|c_{\ell}^*|^2} + \frac{|c_{\ell}|}{|c_{\ell}^*|} \right) \\
& \leq \frac{X^* \gamma (X^* \gamma)^2 + 3(X^* \gamma) + 3}{c_{\min}^* (1 - X^* \gamma)}
\end{aligned} \quad (\text{G.13})$$

where the first line follows from $|a^3 - b^3| = |(a - b)(a^2 + ab + b^2)| = |a - b|(a^2 + ab + b^2)$ for any positive a, b . The second line holds since $\left| \frac{1}{|c_{\ell}|} - \frac{1}{|c_{\ell}^*|} \right| = \frac{||c_{\ell}| - |c_{\ell}^*||}{|c_{\ell}||c_{\ell}^*|} \leq \frac{|c_{\ell} - c_{\ell}^*|}{|c_{\ell}||c_{\ell}^*|}$ by the triangle inequality. The third line follows from $|c_{\ell} - c_{\ell}^*|/|c_{\ell}^*| \leq X^* \gamma$ by (A.10). For the fourth line to hold, note that by (A.10), $\frac{|c_{\ell} - c_{\ell}^*|}{|c_{\ell}^*|} \leq X^* \gamma$, which implies that

$|c_i| \geq |c_i^*| - |c_i - c_i^*| \geq (1 - X^* \gamma) |c_{\min}^*|$. The last line follows from $|c_\ell|/|c_\ell^*| \leq (1 + X^* \gamma)$. Finally, we get the bound

$$\|\Xi^{*-1} - \Xi^{\lambda-1}\|_{\infty, \infty} = \Pi_1 \leq 7.81004 X^* B^* \gamma$$

implying

$$\|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \leq \|\Xi^*\|_{\infty, \infty} \|\Xi^{*-1} - \Xi^{\lambda-1}\|_{\infty, \infty} \|\Xi^\lambda\|_{\infty, \infty} \leq (1.03172)(1.09363)(7.81004 X^* B^* \gamma) = 8.81222 X^* B^* \gamma. \quad (\text{G.14})$$

Bounding $\|\rho^* - \rho^\lambda\|_\infty$ and $\|\rho^\lambda\|_\infty$.

First recognize that $\|\rho^\lambda\|_\infty = 1$ since ρ^λ contains either signs or zeros. Assume the ℓ_∞ norm of $(\rho^* - \rho^\lambda)$ is achieved by $|\text{sign}(c_\ell^\lambda) - \text{sign}(c_\ell^*)|$, then applying triangle inequalities gives

$$\begin{aligned} \|\rho^* - \rho^\lambda\|_\infty &= \left| \frac{c_\ell^\lambda}{|c_\ell^\lambda|} - \frac{c_\ell^*}{|c_\ell^*|} \right| = \left| \frac{c_\ell^\lambda}{|c_\ell^\lambda|} - \frac{c_\ell^\lambda}{|c_\ell^*|} + \frac{c_\ell^\lambda}{|c_\ell^*|} - \frac{c_\ell^*}{|c_\ell^*|} \right| \leq \left| \frac{c_\ell^\lambda}{|c_\ell^\lambda|} - \frac{c_\ell^\lambda}{|c_\ell^*|} \right| + \frac{|c_\ell^* - c_\ell^\lambda|}{|c_\ell^*|} \\ &= |c_\ell^\lambda| \left| \frac{1}{|c_\ell^\lambda|} - \frac{1}{|c_\ell^*|} \right| + \frac{|c_\ell^* - c_\ell^\lambda|}{|c_\ell^*|} \\ &= |c_\ell^\lambda| \left| \frac{|c_\ell^\lambda| - |c_\ell^*|}{|c_\ell^\lambda| |c_\ell^*|} \right| + \frac{|c_\ell^* - c_\ell^\lambda|}{|c_\ell^*|} \\ &\leq 2 \frac{|c_\ell^* - c_\ell^\lambda|}{|c_\ell^*|} \leq 2 X^* \gamma. \end{aligned} \quad (\text{G.15})$$

Bounding $\nu^*(f)$, $\nu^*(f)'$, $\nu^*(f)''$ and $(\nu^*(f) - \nu^\lambda(f))$, $(\nu^*(f) - \nu^\lambda(f))'$, $(\nu^*(f) - \nu^\lambda(f))''$.

Applying the triangle inequality and the sub-multiplicative property of the norm to (G.5) and (G.4), we get

$$\begin{aligned} \|\nu^*(f) - \nu^\lambda(f)\|_1 &\leq \|[\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)]^T \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1} - \mathbf{D}_1(f, \mathbf{f}^\lambda)^T \Phi\|_1 + 2 \|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 \\ &\leq \|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_{1,1} + \|\mathbf{D}_1(f, \mathbf{f}^\lambda)\|_1 \|\Phi\|_{1,1} + 2 \|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1; \\ \|\nu^*(f)\|_1 &\leq \|\mathbf{D}_1(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_1 + 2 \|\mathbf{D}_0(f, \mathbf{f}^*)\|_1 \leq \|\mathbf{D}_1(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_{1,1} + 2 \|\mathbf{D}_0(f, \mathbf{f}^*)\|_1, \end{aligned} \quad (\text{G.16})$$

where $\Phi := \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}$ and $\mathbf{D}_\ell(f, \mathbf{f}) := [K^{(\ell)}(f_1 - f), \dots, K^{(\ell)}(f_k - f)]$. Similar bounds also apply to various derivatives of $\nu^*(f)$ and $\nu^\lambda(f)$, which we need in order to bound the distances between derivatives of $Q^*(f)$ and $Q^\lambda(f)$. Using (A.10) and $\tau \geq 3.289n^2$, we have

$$\begin{aligned} \|(\text{diag}(\mathbf{c}^\lambda) - \text{diag}(\mathbf{c}^*)) \mathbf{S}^{-1}\|_{1,1} &\leq (\max_i |c_i^\lambda - c_i^*|/|c_i^*|)/\sqrt{\tau} \leq X^* \gamma / \sqrt{\tau} \leq 0.552 X^* \gamma / n; \\ \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_{1,1} &\leq (1 + X \gamma) / \sqrt{\tau} \leq 0.552 / n, \end{aligned} \quad (\text{G.17})$$

which we need to continue the bounds in (G.16).

Since f may lie in different regions: Near Region, Middle Region, and Far Region, we next organize our analysis into three parts based on what region f is located in.

G.1 Near Region Analysis

We start with controlling $\|\mathbf{D}_\ell(f, \mathbf{f}^\lambda) - \mathbf{D}_\ell(f, \mathbf{f}^*)\|_1$ and $\|\mathbf{D}_\ell(f, \mathbf{f}^*)\|_1$ for $\ell = 0, 1, 2, 3$ in Near Region. When $\ell = 0$, we have

$$\begin{aligned} \|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 &= \sum_\ell |K(f_\ell^\lambda - f) - K(f_\ell^* - f)| \stackrel{\textcircled{1}}{\leq} \sum_\ell |K'(\tilde{f}_\ell - f)| \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\stackrel{\textcircled{2}}{\leq} \left(F_1(2.5/n, 0.2404/n) + \max_{f \in \mathcal{N}} |K'(f)| \right) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\stackrel{\textcircled{3}}{\leq} (0.01241n + 0.790885n)(0.4 X^* \gamma / n) \\ &\leq 0.321318 X^* \gamma, \end{aligned} \quad (\text{G.18})$$

where ① is due to the mean value theorem with \tilde{f}_ℓ located between f_ℓ^* and f_ℓ^λ . ② follows from Lemma A.5. To see this, first note that $\Delta(\{\tilde{f}_\ell\}) \geq 2.5/n$ by Lemma A.4. Second, $f \in \mathcal{N} = [0, 0.24/n]$ implies that

$$0 \leq |f - \tilde{f}_0| \leq |f - f_0^*| + |f_0^* - \tilde{f}_0| \leq 0.24/n + 0.4(10^{-3})/n = 0.2404/n.$$

We also used the definition $\hat{\mathcal{N}} = [0, 0.2404/n]$ in ②. ③ follows from the upper bound on $F_1(2.5/n, 0.2404/n)$ in Table A.5, the upper bound on $\max_{f \in \hat{\mathcal{N}}} |K'(f)|$ in Table A.5, as well as the upper bound on $\|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty$ in Lemma 4.1.

Applying arguments similar to those for (G.18), we can control $\|\mathbf{D}_\ell(f, \mathbf{f}^\lambda) - \mathbf{D}_\ell(f, \mathbf{f}^*)\|_1$ as

$$\|\mathbf{D}_\ell(f, \mathbf{f}^\lambda) - \mathbf{D}_\ell(f, \mathbf{f}^*)\|_1 \leq (F_{\ell+1}(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K^{(\ell+1)}(f)|) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty. \quad (\text{G.19})$$

We specialize the above inequality to $\ell = 1, 2, 3$ using the upper bounds on $F_\ell(2.5/n, 0.2404/n)$ in Table A.5 and those on $\max_{f \in \hat{\mathcal{N}}} |K^{(\ell)}(f)|$ in Table A.5 to obtain

$$\begin{aligned} \|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 &\leq (F_2(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K''(f)|) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\leq (0.05637n^2 + 3.290n^2)(0.4X^*\gamma/n) = 1.338548nX^*\gamma; \end{aligned} \quad (\text{G.20})$$

$$\begin{aligned} \|\mathbf{D}_2(f, \mathbf{f}^\lambda) - \mathbf{D}_2(f, \mathbf{f}^*)\|_1 &\leq (F_3(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K'''(f)|) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\leq (0.28838n^3 + 7.80572n^3)(0.4X^*\gamma/n) = 3.23764n^2X^*\gamma; \end{aligned} \quad (\text{G.21})$$

$$\begin{aligned} \|\mathbf{D}_3(f, \mathbf{f}^\lambda) - \mathbf{D}_3(f, \mathbf{f}^*)\|_1 &\leq (F_4(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K''''(f)|) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\leq (1.671n^4 + 29.2227n^4)(0.4X^*\gamma/n) = 12.3575n^3X^*\gamma. \end{aligned} \quad (\text{G.22})$$

Furthermore, we can use similar arguments and Lemma A.5 to control $\|\mathbf{D}_\ell(f, \mathbf{f})\|_1$ for $f \in \mathcal{N}$:

$$\|\mathbf{D}_\ell(f, \mathbf{f}^*)\|_1 \leq F_\ell(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K^{(\ell)}(f)|, \quad (\text{G.23})$$

which specializes to

$$\|\mathbf{D}_0(f, \mathbf{f}^*)\|_1 \leq F_0(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K(f)| \leq 0.00757 + 1 = 1.00757; \quad (\text{G.24})$$

$$\|\mathbf{D}_1(f, \mathbf{f}^*)\|_1 \leq F_1(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K'(f)| \leq 0.01241n + 0.790885n = 0.803295n; \quad (\text{G.25})$$

$$\|\mathbf{D}_2(f, \mathbf{f}^*)\|_1 \leq F_2(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K''(f)| \leq 0.05637n^2 + 3.290n^2 = 3.34637n^2; \quad (\text{G.26})$$

$$\|\mathbf{D}_3(f, \mathbf{f}^*)\|_1 \leq F_3(2.5/n, 0.2404/n) + \max_{f \in \hat{\mathcal{N}}} |K'''(f)| \leq 0.28838n^3 + 7.80572n^3 = 8.0941n^3. \quad (\text{G.27})$$

With these preparations, we are ready to control $\|\boldsymbol{\nu}^*(f)^{(\ell)} - \boldsymbol{\nu}^\lambda(f)^{(\ell)}\|_1$ and $\|\boldsymbol{\nu}^*(f)^{(\ell)}\|_1$ for $\ell = 0, 1, 2$ in Near Region. Generalizing (G.16) to the ℓ th derivative of $\boldsymbol{\nu}^*(f)$ and $\boldsymbol{\nu}^\lambda(f)$ to get

$$\begin{aligned} \|\boldsymbol{\nu}^*(f)^{(\ell)} - \boldsymbol{\nu}^\lambda(f)^{(\ell)}\|_1 &\leq \|\mathbf{D}_{\ell+1}(f, \mathbf{f}^\lambda) - \mathbf{D}_{\ell+1}(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 \\ &\quad + \|\mathbf{D}_{\ell+1}(f, \mathbf{f}^\lambda)\|_1 \|\text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_\ell(f, \mathbf{f}^\lambda) - \mathbf{D}_\ell(f, \mathbf{f}^*)\|_1; \\ \|\boldsymbol{\nu}^*(f)^{(\ell)}\|_1 &\leq \|\mathbf{D}_{\ell+1}(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_\ell(f, \mathbf{f}^*)\|_1. \end{aligned} \quad (\text{G.28})$$

Plugging Eq. (G.18), (G.20), (G.25) and (G.17) into (G.28), we obtain

$$\begin{aligned} \|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1 &\leq \|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 \\ &\quad + \|\mathbf{D}_1(f, \mathbf{f}^\lambda)\|_1 \|\text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 \\ &\leq 1.338548nX^*\gamma \frac{0.552}{n} + (0.803295n) \frac{0.552X^*\gamma}{n} + 2(0.321318X^*\gamma) \leq 1.82494X^*\gamma. \end{aligned} \quad (\text{G.29})$$

Plugging Eq. (G.20)-(G.21), (G.26) and (G.17) into (G.28), we obtain

$$\begin{aligned}
\|\nu^*(f)' - \nu^\lambda(f)'\|_1 &\leq \|\mathbf{D}_2(f, \mathbf{f}^\lambda) - \mathbf{D}_2(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 \\
&\quad + \|\mathbf{D}_2(f, \mathbf{f}^\lambda)\|_1 \|\text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 \\
&\leq 3.23764n^2 X^* \gamma \frac{0.552}{n} + (3.34637n^2) \frac{0.552 X^* \gamma}{n} + 2(1.338548n X^* \gamma) \\
&\leq 6.31147n X^* \gamma.
\end{aligned} \tag{G.30}$$

Plugging Eq. (G.21)-(G.22), (G.27) and (G.17) into (G.28), we obtain

$$\begin{aligned}
\|\nu^*(f)'' - \nu^\lambda(f)''\|_1 &\leq \|\mathbf{D}_3(f, \mathbf{f}^\lambda) - \mathbf{D}_3(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 \\
&\quad + \|\mathbf{D}_3(f, \mathbf{f}^\lambda)\|_1 \|\text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_2(f, \mathbf{f}^\lambda) - \mathbf{D}_2(f, \mathbf{f}^*)\|_1 \\
&\leq 12.3575n^3 X^* \gamma \frac{0.552}{n} + (8.0941n^3) \frac{0.552 X^* \gamma}{n} + 2(3.23764n^2 X^* \gamma) \\
&\leq 17.7646n^2 X^* \gamma.
\end{aligned} \tag{G.31}$$

Similarly, plugging Eq. (G.24)-(G.25) and (G.17) into (G.28), we have

$$\|\nu^*(f)\|_1 \leq \|\mathbf{D}_1(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_0(f, \mathbf{f}^*)\|_1 \leq (0.803295n) \frac{0.552}{n} + 2(1.00757) \leq 2.45856. \tag{G.32}$$

Plugging Eq. (G.25)-(G.26) and (G.17) into (G.28), we obtain

$$\|\nu^*(f)'\|_1 \leq \|\mathbf{D}_2(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_1(f, \mathbf{f}^*)\|_1 \leq (3.34637n^2) \frac{0.552}{n} + 2(0.803295n) \leq 3.4538n. \tag{G.33}$$

Finally, plugging Eq. (G.26)-(G.27) and (G.17) into (G.28), we arrive at

$$\|\nu^*(f)''\|_1 \leq \|\mathbf{D}_3(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_2(f, \mathbf{f}^*)\|_1 \leq (8.0941n^3) \frac{0.552}{n} + 2(3.34637n^2) \leq 11.1607n^2. \tag{G.34}$$

We are now ready to control the pointwise distance between $Q^{*(\ell)}(f)$ and $Q^{\lambda(\ell)}(f)$ using

$$|Q^{*(\ell)}(f) - Q^{\lambda(\ell)}(f)| \leq |\nu^*(f)^{(\ell)} \Xi^\lambda \rho^\lambda - \nu^*(f)^{(\ell)} \Xi^* \rho^*|, \quad \ell = 0, 1, 2. \tag{G.35}$$

Plugging Eq. (G.29)-(G.30), (G.6)-(G.14) and (G.15) to (G.35) with $\ell = 0$, we obtain for $f \in \mathcal{N}$

$$\begin{aligned}
&|Q^*(f) - Q^\lambda(f)| \\
&\leq \|\nu^*(f)\|_1 \|\Xi^*\|_{\infty, \infty} \|\rho^* - \rho^\lambda\|_1 + \|\nu^*(f)\|_1 \|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \|\rho^\lambda\|_\infty + \|\nu^*(f) - \nu^\lambda(f)\|_1 \|\Xi^\lambda\|_{\infty, \infty} \|\rho^\lambda\|_\infty \\
&\leq (2.45856)(1.03172)(2X^* \gamma) + (2.45856)(8.81222X^* B^* \gamma) + (1.82494X^* \gamma)(1.09363) \leq 28.7343X^* B^* \gamma.
\end{aligned}$$

Plugging Eq. (G.31)-(G.32), (G.6)-(G.14) and (G.15) to (G.35) with $\ell = 1$, we obtain for $f \in \mathcal{N}$

$$\begin{aligned}
&|Q^*(f)' - Q^\lambda(f)'| \\
&\leq \|\nu^*(f)'\|_1 \|\Xi^*\|_{\infty, \infty} \|\rho^* - \rho^\lambda\|_1 + \|\nu^*(f)'\|_1 \|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \|\rho^\lambda\|_\infty + \|\nu^*(f)' - \nu^\lambda(f)'\|_1 \|\Xi^\lambda\|_{\infty, \infty} \|\rho^\lambda\|_\infty \\
&\leq (3.4538n)(1.03172)(2X^* \gamma) + (3.4538n)(8.81222X^* B^* \gamma) + (6.31147n X^* \gamma)(1.09363) \leq 44.4648n X^* B^* \gamma.
\end{aligned}$$

Finally, plugging Eq. (G.33)-(G.34), (G.6)-(G.14) and (G.15) to (G.35) with $\ell = 2$, we get for $f \in \mathcal{N}$

$$\begin{aligned}
&|Q^*(f)'' - Q^\lambda(f)''| \\
&\leq \|\nu^*(f)''\|_1 \|\Xi^*\|_{\infty, \infty} \|\rho^* - \rho^\lambda\|_1 + \|\nu^*(f)''\|_1 \|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \|\rho^\lambda\|_\infty + \|\nu^*(f)'' - \nu^\lambda(f)''\|_1 \|\Xi^\lambda\|_{\infty, \infty} \|\rho^\lambda\|_\infty \\
&\leq (11.1607n^2)(1.03172)(2X^* \gamma) + (11.1607n^2)(8.81222X^* B^* \gamma) + (17.7646n^2 X^* \gamma)(1.09363) \leq 140.808n^2 X^* B^* \gamma.
\end{aligned}$$

G.2 Middle Region Analysis

We continue with bounding the pointwise distance between $Q^*(f)$ and $Q^\lambda(f)$ in Middle Region $\mathcal{M} = [0.24/n, 0.75/n]$. We start with controlling $\|\mathbf{D}_\ell(f, \mathbf{f}^\lambda) - \mathbf{D}_\ell(f, \mathbf{f}^*)\|_1$ and $|\mathbf{D}_\ell(f, \mathbf{f}^*)|_1$ for $\ell = 0, 1$. First note when $f \in \mathcal{M} = [0.24/n, 0.75/n]$, we have

$$\begin{aligned} (a) \quad & |f - \tilde{f}_0| \leq |f - f_0^*| + |f_0^* - \tilde{f}_0| \leq 0.75/n + 0.0004/n = 0.7504/n, \\ (b) \quad & |f - \tilde{f}_0| \geq |f - f_0^*| - |f_0^* - \tilde{f}_0| \geq 0.24/n - 0.0004/n = 0.2396/n. \end{aligned}$$

Denote $\hat{\mathcal{M}} = [0.2396/n, 0.7504/n]$. We combine the upper bounds on $F_\ell(2.5/n, 0.7504/n)$ in Table A.5 and the upper bounds on $\max_{f \in \hat{\mathcal{M}}} |K^{(\ell)}(f)|$ in Table A.5 to get

$$\begin{aligned} \|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 &\leq (F_1(2.5/n, 0.7504/n) + \max_{f \in \hat{\mathcal{M}}} |K'(f)|) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\leq (0.01454n + 2.46872n)(0.4X^*\gamma/n) = 0.993304X^*\gamma; \end{aligned} \quad (\text{G.36})$$

$$\begin{aligned} \|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 &\leq (F_2(2.5/n, 0.7504/n) + \max_{f \in \hat{\mathcal{M}}} |K''(f)|) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\leq (0.12675n^2 + 3.290n^2)(0.4X^*\gamma/n) = 1.36670nX^*\gamma. \end{aligned} \quad (\text{G.37})$$

In a similar manner, we use Lemma A.5 to control $\|\mathbf{D}_\ell(f, \mathbf{f})\|_1$ as follows

$$\|\mathbf{D}_0(f, \mathbf{f})\|_1 \leq F_0(2.5/n, 0.7504/n) + \max_{f \in \hat{\mathcal{M}}} |K(f)| \leq 0.00772 + 0.90951 = 0.91723; \quad (\text{G.38})$$

$$\|\mathbf{D}_1(f, \mathbf{f})\|_1 \leq F_1(2.5/n, 0.7504/n) + \max_{f \in \hat{\mathcal{M}}} |K'(f)| \leq 0.01454n + 2.46872n = 2.48326n. \quad (\text{G.39})$$

To control $\|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1$ and $\|\boldsymbol{\nu}^*(f)\|_1$ in the Middle Region, we plug Eq. (G.36)-(G.39) into (G.28) to get

$$\begin{aligned} \|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1 &\leq \|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 + \|\mathbf{D}_1(f, \mathbf{f}^\lambda)\|_1 \|\text{diag}(\mathbf{c}^*) \mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda) \mathbf{S}^{-1}\|_1 \\ &\quad + 2\|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 \\ &\leq 1.36670nX^*\gamma \frac{0.552}{n} + (2.48326n) \frac{0.552X^*\gamma}{n} + 2(0.993304X^*\gamma) \leq 4.11179X^*\gamma; \end{aligned} \quad (\text{G.40})$$

$$\|\boldsymbol{\nu}^*(f)\|_1 \leq \|\mathbf{D}_1(f, \mathbf{f}^*) \text{diag}(\mathbf{c}^*) \mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_0(f, \mathbf{f}^*)\|_1 \leq (2.48326n) \frac{0.552}{n} + 2(0.91723) \leq 3.20522. \quad (\text{G.41})$$

Finally, we control $|Q^*(f) - Q^\lambda(f)|$ in Middle Region by plugging Eq. (G.40)-(G.41), (G.6)-(G.14) and (G.15) to (G.35) with $\ell = 0$ to get

$$\begin{aligned} |Q^*(f) - Q^\lambda(f)| &\leq \|\boldsymbol{\nu}^*(f)\|_1 \|\Xi^*\|_{\infty, \infty} \|\boldsymbol{\rho}^* - \boldsymbol{\rho}^\lambda\|_1 + \|\boldsymbol{\nu}^*(f)\|_1 \|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \|\boldsymbol{\rho}^\lambda\|_\infty + \|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1 \|\Xi^\lambda\|_{\infty, \infty} \|\boldsymbol{\rho}^\lambda\|_\infty \\ &\leq (3.20522)(1.03172)(2X^*\gamma) + (3.20522)(8.81222X^*B^*\gamma) + (4.11179X^*\gamma)1.09363 \\ &\leq 39.3557X^*B^*\gamma, \quad f \in \mathcal{M}. \end{aligned}$$

G.3 Far Region Analysis

Lastly, we bound the pointwise distance between $Q^*(f)$ and $Q^\lambda(f)$ in Far Region $\mathcal{F} = [0.75/n, f_1^*/2]$. Again, we start with controlling $\|\mathbf{D}_\ell(f, \mathbf{f}^\lambda) - \mathbf{D}_\ell(f, \mathbf{f}^*)\|_1$ and $|\mathbf{D}_\ell(f, \mathbf{f}^*)|_1$ for $\ell = 0, 1$. First note when $f \in \mathcal{F} = [0.75/n, f_1^*/2]$, we have

$$\begin{aligned} (a) \quad & f - \tilde{f}_0 \geq f - f_0^* - |f_0^* - \tilde{f}_0| \geq 0.75/n - 0.0004/n = 0.74996/n, \\ (b) \quad & \tilde{f}_1 - f \geq -|f_1^* - f_1^*| + f_1^* - f \geq -0.0004n + f_1^*/2 \geq -0.0004/n + (2.5009/n)/2 \geq 1.25/n. \end{aligned}$$

Further note that $\{\tilde{f}_\ell\}$ satisfies the separation condition that $\Delta(\{\tilde{f}_\ell\}) \geq 2.5/n$. Then, following from Lemma A.3 and the upper bounds on $W_\ell(0.74996/n, 1.25/n)$ in Table A.5, we have

$$\begin{aligned} \|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 &\leq \|\mathbf{D}_1(\tilde{\mathbf{f}}, f)\|_1 \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \\ &\leq W_1(0.74996/n, 1.25/n) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \leq 5.2265n(0.4X^*\gamma/n) = 2.0906X^*\gamma; \end{aligned} \quad (\text{G.42})$$

$$\|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 \leq W_2(0.74996/n, 1.25/n) \|\mathbf{f}^\lambda - \mathbf{f}^*\|_\infty \leq 48.033n^2(0.4X^*\gamma/n) = 19.2132nX^*\gamma. \quad (\text{G.43})$$

Similarly, we can use Lemma A.3 to control $\|\mathbf{D}_\ell(f, \mathbf{f})\|_1$ for $\ell = 0, 1$ and $f \in \mathcal{F}$:

$$\begin{aligned} \|\mathbf{D}_0(f, \mathbf{f}^*)\|_1 &\leq W_0(0.74996/n, 1.25/n) \leq 0.71059; \\ \|\mathbf{D}_1(f, \mathbf{f}^*)\|_1 &\leq W_1(0.74996/n, 1.25/n) \leq 5.2265n. \end{aligned} \quad (\text{G.44})$$

Directly plugging Eq. (G.42)-(G.44) into (G.28), we arrive at

$$\begin{aligned} \|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1 &\leq \|\mathbf{D}_1(f, \mathbf{f}^\lambda) - \mathbf{D}_1(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^\lambda)\mathbf{S}^{-1}\|_1 + \|\mathbf{D}_1(f, \mathbf{f}^\lambda)\|_1 \|\text{diag}(\mathbf{c}^*)\mathbf{S}^{-1} - \text{diag}(\mathbf{c}^\lambda)\mathbf{S}^{-1}\|_1 \\ &\quad + 2\|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \mathbf{f}^*)\|_1 \\ &\leq 19.2132nX^*\gamma \frac{0.552}{n} + (5.2265n) \frac{0.552X^*\gamma}{n} + 2(2.0906X^*\gamma) \leq 17.6720X^*\gamma; \end{aligned} \quad (\text{G.45})$$

$$\|\boldsymbol{\nu}^*(f)\|_1 \leq \|\mathbf{D}_1(f, \mathbf{f}^*)\|_1 \|\text{diag}(\mathbf{c}^*)\mathbf{S}^{-1}\|_1 + 2\|\mathbf{D}_0(f, \mathbf{f}^*)\|_1 \leq (5.2265n) \frac{0.552}{n} + 2(0.71059) \leq 4.30621. \quad (\text{G.46})$$

As a final step, we control $|Q^*(f) - Q^\lambda(f)|$ in Far Region by plugging Eq. (G.45)-(G.46) and (G.6)-(G.15) to (G.35) to get

$$\begin{aligned} |Q^*(f) - Q^\lambda(f)| &\leq \|\boldsymbol{\nu}^*(f)\|_1 \|\Xi^*\|_{\infty, \infty} \|\boldsymbol{\rho}^* - \boldsymbol{\rho}^\lambda\|_1 + \|\boldsymbol{\nu}^*(f)\|_1 \|\Xi^* - \Xi^\lambda\|_{\infty, \infty} \|\boldsymbol{\rho}^\lambda\|_\infty + \|\boldsymbol{\nu}^*(f) - \boldsymbol{\nu}^\lambda(f)\|_1 \|\Xi^\lambda\|_{\infty, \infty} \|\boldsymbol{\rho}^\lambda\|_\infty \\ &\leq (4.30621)(1.03172)(2X^*\gamma) + (4.30621)(8.81222X^*B^*\gamma) + (17.6720X^*\gamma)(1.09363) \\ &\leq 66.1596X^*B^*\gamma, \quad f \in \mathcal{F}. \end{aligned}$$

This concludes the proof of Lemma 4.4. □

Appendix H Proof of Lemma 4.5

Proof. The expressions $\mathbf{q}^\lambda = \frac{\mathbf{x}^* - \mathbf{x}^\lambda}{\lambda}$ and $\hat{\mathbf{q}} = \frac{\mathbf{y} - \hat{\mathbf{x}}}{\lambda}$ lead to

$$\hat{\mathbf{q}} - \mathbf{q}^\lambda = \frac{(\mathbf{y} - \hat{\mathbf{x}}) - (\mathbf{x}^* - \mathbf{x}^\lambda)}{\lambda} = \frac{\mathbf{w}}{\lambda} + \frac{\mathbf{x}^\lambda - \hat{\mathbf{x}}}{\lambda},$$

implying

$$|Q^\lambda(f) - \hat{Q}(f)| \leq \underbrace{\frac{|\mathbf{a}(f)^H \mathbf{Z} \mathbf{w}|}{\lambda}}_{\Pi_1(f)} + \underbrace{\frac{|\mathbf{a}(f)^H \mathbf{Z} (\mathbf{x}^\lambda - \hat{\mathbf{x}})|}{\lambda}}_{\Pi_2(f)}. \quad (\text{H.1})$$

This separates the distance between $Q^\lambda(f)$ and $\hat{Q}(f)$ into two parts: one is $\Pi_1(f)$ associated with the dual atomic norm of the Gaussian noise \mathbf{w} whose upper bounds were developed in Appendix B; the other is $\Pi_2(f)$ corresponding to the dual atomic norm of $\mathbf{x}^\lambda - \hat{\mathbf{x}}$. The latter quantity can be bounded by similar arguments as controlling $|\mathbf{a}(f)^H \mathbf{Z} (\mathbf{x}^\lambda - \hat{\mathbf{x}})|$ in Lemma 4.4.

Bounding $\Pi_1(f)$.

Combining Eq. (B.12)-(B.14), we can upper bound $\Pi_1(f)$, $\Pi(f)'$ and $\Pi_1(f)''$ with high probability (at least $1 - 1/n^2$) for all $f \in \mathbb{T}$:

$$\begin{aligned}\Pi_1(f) &\leq 6.534\gamma_0/\lambda \leq 10.115/X^*; \\ \Pi_1(f)' &\leq 41.052n\gamma_0/\lambda \leq 63.458n/X^*; \\ \Pi_1(f)'' &\leq 257.94n^2\gamma_0/\lambda \leq 399.288n^2/X^*,\end{aligned}\tag{H.2}$$

where we used $\lambda = 0.646X^*\gamma_0$.

Bounding $\Pi_2(f)$.

$$\begin{aligned}\Pi_2(f) &= \frac{1}{\lambda} |\mathbf{D}_0(f, \mathbf{f}^\lambda) \mathbf{c}^\lambda - \mathbf{D}_0(f, \hat{\mathbf{f}}) \hat{\mathbf{c}}| \\ &\leq \frac{1}{\lambda} \|\mathbf{D}_0(f, \mathbf{f}^\lambda) - \mathbf{D}_0(f, \hat{\mathbf{f}})\|_1 \|\mathbf{c}^\lambda\|_\infty + \frac{1}{\lambda} \|\mathbf{D}_0(f, \hat{\mathbf{f}})\|_1 \|\hat{\mathbf{c}} - \mathbf{c}^\lambda\|_\infty \\ &\stackrel{\textcircled{1}}{\leq} \frac{1}{\lambda} \|\mathbf{D}_1(f, \tilde{\mathbf{f}})\|_1 \|\mathbf{f}^\lambda - \hat{\mathbf{f}}\|_\infty \|\mathbf{c}^\lambda\|_\infty + \frac{1}{\lambda} \|\mathbf{D}_0(f, \hat{\mathbf{f}})\|_1 \|\hat{\mathbf{c}} - \mathbf{c}^\lambda\|_\infty \\ &\stackrel{\textcircled{2}}{\leq} \frac{c_{\max}^\lambda}{0.646X^*\gamma_0} \left(\frac{0.4(35.2)\gamma}{n} \|\mathbf{D}_1(f, \tilde{\mathbf{f}})\|_1 + 35.2\gamma \|\mathbf{D}_0(f, \hat{\mathbf{f}})\|_1 \right) \\ &\stackrel{\textcircled{3}}{\leq} \frac{B^*(1 + X^*\gamma)}{0.646X^*} \left(\frac{14.08}{n} \|\mathbf{D}_1(f, \tilde{\mathbf{f}})\|_1 + 35.2 \|\mathbf{D}_0(f, \hat{\mathbf{f}})\|_1 \right),\end{aligned}$$

where $\textcircled{1}$ follows from the mean value theorem. For $\textcircled{2}$ to hold, first note that $\lambda = 0.646X^*\gamma_0$ and $\hat{\boldsymbol{\theta}} \in N^\lambda$ by Lemma 4.2. Then, we can upper bound $\|\hat{\mathbf{c}} - \mathbf{c}^\lambda\|_\infty$ as

$$\|\hat{\mathbf{c}} - \mathbf{c}^\lambda\|_\infty = \frac{|\hat{c}_j - c_j^\lambda|}{|c_j^\lambda|} |c_j^\lambda| \leq (35.2\gamma) c_{\max}^\lambda,$$

where the equality follows by assuming the ℓ_∞ norm is achieved by the j th row and the inequality follows by changing X^* to 35.2 in (A.10) and defining $c_{\max}^\lambda := \max_j |c_j^\lambda|$. $\textcircled{3}$ follows from $\gamma_0 = \gamma c_{\min}^*$ and

$$\frac{c_{\max}^\lambda}{c_{\min}^*} = B^* \frac{|c_j^\lambda|}{c_{\max}^*} \leq B^* \frac{|c_j^\lambda|}{|c_j^*|} \leq B^*(1 + X^*\gamma).$$

As a consequence, to control $\Pi_2(f)$, it reduces to bounding $\|\mathbf{D}_\ell(f, \tilde{\mathbf{f}})\|_1$ and $\|\mathbf{D}_\ell(f, \hat{\mathbf{f}})\|_1$. For this purpose, we first note that $\{\Delta(\tilde{T}), \Delta(\hat{T})\} \geq 2.5/n$ by Lemma A.4. Second, by

$$\|\tilde{\mathbf{f}} - \mathbf{f}^*\|_\infty \stackrel{\textcircled{1}}{\leq} \|\hat{\mathbf{f}} - \mathbf{f}^*\|_\infty \stackrel{\textcircled{2}}{\leq} 0.4(X^* + 35.2)\gamma \stackrel{\textcircled{3}}{\leq} 0.0004/n + 1.408 \times 10^{-6}/n = 0.000401408/n,$$

where $\textcircled{1}$ follows from the length of subinterval is smaller. $\textcircled{2}$ follows from Eq. (A.10) and $\textcircled{3}$ follows from the SNR condition (2.6). Thus, we can follow the same arguments that lead to Eq. (G.24)-(G.25) for Near Region, Eq. (G.38)-(G.39) for Middle Region, and Eq. (G.44) for Far Region to develop bounds on $\|\mathbf{D}_\ell(f, \hat{\mathbf{f}})\|_1$.

To have a concrete idea, we first show how to control $\|\mathbf{D}_\ell(f, \hat{\mathbf{f}})\|_1$ since the upper bounds for $\|\mathbf{D}_\ell(f, \tilde{\mathbf{f}})\|_1$ then follows by $\|\tilde{\mathbf{f}} - \mathbf{f}^*\|_\infty \leq \|\hat{\mathbf{f}} - \mathbf{f}^*\|_\infty$. First, consider $f \in \mathcal{N}$. Then we have

$$0 \leq |\hat{f}_0 - f| \leq |\hat{f}_0 - f_0^*| + |f_0^* - f| \leq 0.000401408/n + 0.24/n \leq 0.240401408/n.$$

With some abuse of notation, we denote $\hat{\mathcal{N}} := [0, 0.240401408/n]$. Second, consider $f \in \mathcal{M}$. Then we have

$$\begin{aligned}(a) \quad &|f - \hat{f}_0| \leq |f - f_0^*| + |f_0^* - \hat{f}_0| \leq 0.75/n + 0.000401408/n = 0.750401408/n; \\ (b) \quad &|f - \hat{f}_0| \geq |f - f_0^*| - |f_0^* - \hat{f}_0| \geq 0.24/n - 0.000401408/n = 0.239598592/n.\end{aligned}$$

Denote $\hat{\mathcal{M}} = [0.2396/n, 0.7504/n]$. At last, we consider $f \in \mathcal{F} = [0.75/n, f_1^*/2]$:

- (a) $f - \hat{f}_0 \geq f - f_0^* - |f_0^* - \hat{f}_0| \geq 0.75/n - 0.000401408/n = 0.749598592/n$;
- (b) $\hat{f}_1 - f \geq -|\hat{f}_1 - f_1^*| + f_1^* - f \geq -0.000401408/n + f_1^*/2 \geq -0.000401408/n + (2.5009/n)/2 \geq 1.25/n$.

Hence we can define $\hat{\mathcal{F}} := [0.749598592/n, 1.25/n]$. Furthermore, we remark that those numerical upper bounds in Table A.5-A.5 dose not change when evaluated for the newly defined intervals $\hat{\mathcal{N}}$, $\hat{\mathcal{M}}$ and $\hat{\mathcal{F}}$.

Finally, by directly plugging the upper bounds of $\|\mathbf{D}_\ell(f, \mathbf{f})\|_\infty$ in (G.24)-(G.25) for Near Region, (G.38)-(G.39) for Middle Region, and equation(G.44) for Far Region, it follows that

$$\Pi_2(f) \leq \frac{B^*}{X^*} \begin{cases} \frac{1.001}{0.646} \left(\frac{14.08}{n} (0.803295n) + 35.2(1.00757) \right) \leq 72.4825 \frac{B^*}{X^*}, & f \in \mathcal{N}; \\ \frac{1.001}{0.646} \left(\frac{14.08}{n} (2.48326n) + 35.2(0.91723) \right) \leq 104.208 \frac{B^*}{X^*}, & f \in \mathcal{M}; \\ \frac{1.001}{0.646} \left(\frac{14.08}{n} (5.2265n) + 35.2(0.71059) \right) \leq 152.788 \frac{B^*}{X^*}, & f \in \mathcal{F}. \end{cases} \quad (\text{H.3})$$

Similarly, from (G.25)-(G.27), we have an upper bound on $\Pi_2(f)'$ and $\Pi_2(f)''$ as follows

$$\begin{aligned} \Pi_2(f)' &\leq \frac{B^*(1 + X^*\gamma)}{0.646X^*} \left(\frac{14.08}{n} \|\mathbf{D}_2(f, \tilde{\mathbf{f}})\|_1 + 35.2 \|\mathbf{D}_1(f, \hat{\mathbf{f}})\|_1 \right) \\ &\leq \frac{B^*}{X^*} \frac{1.001}{0.646} \left(\frac{14.08}{n} (3.34637n^2) + 35.2(0.803295n) \right) \leq 116.825n \frac{B^*}{X^*}, & f \in \mathcal{N}; \end{aligned} \quad (\text{H.4})$$

$$\begin{aligned} \Pi_2(f)'' &\leq \frac{B^*(1 + X^*\gamma)}{0.646X^*} \left(\frac{14.08}{n} \|\mathbf{D}_3(f, \tilde{\mathbf{f}})\|_1 + 35.2 \|\mathbf{D}_2(f, \hat{\mathbf{f}})\|_1 \right) \\ &\leq \frac{B^*}{X^*} \frac{1.001}{0.646} \left(\frac{14.08}{n} (8.0941n^3) + 35.2(3.34637n^2) \right) \leq 359.116n^2 \frac{B^*}{X^*}, & f \in \mathcal{N}. \end{aligned} \quad (\text{H.5})$$

Combining (H.2)-(H.5) for $\Pi_1(f)$ and $\Pi_2(f)$, we can control $|\hat{Q}^{(\ell)}(f) - Q^{\lambda^{(\ell)}}(f)|$ in Near region $f \in \mathcal{N}$ as follows

$$\begin{aligned} |\hat{Q}(f) - Q^\lambda(f)| &\leq (10.115 + 72.4825)B^*/X^* = 82.5975B^*/X^*, & f \in \mathcal{N}; \\ |\hat{Q}(f)' - Q^{\lambda'}(f)| &\leq (63.458n + 116.825n)B^*/X^* = 180.283nB^*/X^*, & f \in \mathcal{N}; \\ |\hat{Q}(f)'' - Q^{\lambda''}(f)| &\leq (399.288n^2 + 359.116n^2)B^*/X^* = 758.404n^2B^*/X^*, & f \in \mathcal{N}. \end{aligned}$$

For the case of Middle Region and Far Region, we can upper bound them as:

$$\begin{aligned} |\hat{Q}(f) - Q^\lambda(f)| &\leq (10.115 + 104.208)B^*/X^* = 114.323B^*/X^*, & f \in \mathcal{M}; \\ |\hat{Q}(f) - Q^\lambda(f)| &\leq (10.115 + 152.788)B^*/X^* = 162.903B^*/X^*, & f \in \mathcal{F}. \end{aligned}$$

This completes the proof of Lemma 4.5. □

Appendix I Proof of Proposition 4.1

Proof. The uniqueness follows from the strongly convex quadratic term in (2.4). We next show the primal optimality of $\hat{\mathbf{x}}$ and the dual optimality of $\hat{\mathbf{q}}$ by establishing strong duality. First, $\hat{\mathbf{q}}$ is feasible to the dual program (4.1) because

of the BIP property. Second, we have the following chain of inequalities:

$$\begin{aligned}
\text{value of (4.1)} &= \frac{1}{2} \|\mathbf{y}\|_{\mathbf{Z}}^2 - \frac{1}{2} \|\mathbf{y} - \lambda \hat{\mathbf{q}}\|_{\mathbf{Z}}^2 \\
&= \frac{1}{2} \|\lambda \hat{\mathbf{q}}\|_{\mathbf{Z}}^2 + \lambda \mathbb{R}\{\hat{\mathbf{x}}^H \mathbf{Z} \hat{\mathbf{q}}\} \\
&= \frac{1}{2} \|\mathbf{y} - \hat{\mathbf{x}}\|_{\mathbf{Z}}^2 + \lambda \|\hat{\mathbf{c}}\|_1 \\
&\geq \frac{1}{2} \|\mathbf{y} - \hat{\mathbf{x}}\|_{\mathbf{Z}}^2 + \lambda \|\hat{\mathbf{x}}\|_{\mathcal{A}} = \text{value of (2.4)},
\end{aligned}$$

where the second line follows by plugging $\mathbf{y} = \hat{\mathbf{x}} + \lambda \hat{\mathbf{q}}$; the third line holds due to the Interpolation property; and the last line holds since $\|\hat{\mathbf{x}}\|_{\mathcal{A}} \leq \|\hat{\mathbf{c}}\|_1$ by (2.3). Since the weak duality theorem ensures that the other direction of the inequality always holds, we obtain strong duality. As a consequence, $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$ achieve primal optimality and dual optimality, respectively. This means $\hat{\mathbf{x}} = \mathbf{x}^{\text{glob}}$, $\hat{\mathbf{q}} = \mathbf{q}^{\text{glob}}$ due to uniqueness of the solutions. \square

Appendix J Proof of Corollary 2.1

Proof. Denote by $F(\mathbf{x})$ the objective functions for (2.4) and $G(\mathbf{f}, \mathbf{c})$ for (2.11), respectively. Assume $(\mathbf{f}^{\text{non}}, \mathbf{c}^{\text{non}})$ is a global optimum for (2.11) with $\mathbf{x}^{\text{non}} = \mathbf{A}(\mathbf{f}^{\text{non}})\mathbf{c}^{\text{non}}$, and $\mathbf{x}^{\text{glob}} = \mathbf{A}(\mathbf{f}^{\text{glob}})\mathbf{c}^{\text{glob}}$ is the global optimum of (2.4). Then

$$F(\mathbf{x}^{\text{glob}}) \leq F(\mathbf{x}^{\text{non}}) \leq G(\mathbf{f}^{\text{non}}, \mathbf{c}^{\text{non}}) \leq G(\mathbf{f}^{\text{glob}}, \mathbf{c}^{\text{glob}}), \quad (\text{J.1})$$

where the first inequality uses the optimality of \mathbf{x}^{glob} to (2.4); the second inequality follows from $\|\mathbf{x}^{\text{non}}\|_{\mathcal{A}} \leq \|\mathbf{c}^{\text{non}}\|_1$ by (2.3); and the last inequality follows from the optimality of $(\mathbf{f}^{\text{non}}, \mathbf{c}^{\text{non}})$ to (2.11). On the other hand, recognize that $\|\mathbf{x}^{\text{glob}}\|_{\mathcal{A}} = \|\mathbf{c}^{\text{glob}}\|_1$ since $\{f_{\ell}^{\text{glob}}\}$ satisfies the separation condition (revealed by Lemma A.4 in Appendix A). This leads to $G(\mathbf{f}^{\text{glob}}, \mathbf{c}^{\text{glob}}) = F(\mathbf{x}^{\text{glob}})$. Therefore, all inequalities in (J.1) become equalities and hence $G(\mathbf{f}^{\text{non}}, \mathbf{c}^{\text{non}}) = G(\mathbf{f}^{\text{glob}}, \mathbf{c}^{\text{glob}})$. This implies the global optimality of $(\mathbf{f}^{\text{glob}}, \mathbf{c}^{\text{glob}})$ for the nonconvex program (2.11). \square

Appendix K Proof of Lemma A.4

Proof. First of all, from Lemma 4.1, we have $\theta^{\lambda} \in \mathcal{N}^*$, implying $\|\mathbf{f}^{\lambda} - \mathbf{f}^*\|_{\infty} \leq 0.4X^*B^*\gamma/n$ by Eq. (A.10) and by Lemma 4.2, we obtain that $\hat{\theta} \in \mathcal{N}^{\lambda}$, which implies $\|\hat{\mathbf{f}} - \mathbf{f}^{\lambda}\|_{\infty} \leq 0.4(35.2)B^*\gamma/n$ by Eq. (A.10). More precisely, we bound $\Delta(T^{\lambda})$ as

$$\begin{aligned}
\Delta(T^{\lambda}) &\stackrel{\textcircled{1}}{=} \min_{i \neq j} |f_i^{\lambda} - f_j^{\lambda}| \\
&= \min_{i \neq j} |f_i^{\lambda} - f_i^* + f_i^* - f_j^* + f_j^* - f_j^{\lambda}| \\
&\stackrel{\textcircled{2}}{\geq} \min_{i \neq j} |f_i^* - f_j^*| - \max_i |f_i^{\lambda} - f_i^*| - \max_j |f_j^* - f_j^{\lambda}| \\
&\stackrel{\textcircled{3}}{\geq} \Delta(T^*) - 0.8X^*B^*\gamma/n \\
&\stackrel{\textcircled{4}}{\geq} 2.5009/n - 0.0008/n = 2.5001/n > 2.5/n,
\end{aligned}$$

where $\textcircled{1}$ follows from the definition of the separation distance and $\textcircled{2}$ follows from the triangle inequality. $\textcircled{3}$ follows from that θ^{λ} is the fixed point solution of the contraction map (4.4). Thus, $\theta^{\lambda} \in \mathcal{N}^*$ following from the non-escaping property by the contraction mapping theorem. This further implies that $\|\mathbf{f}^{\lambda} - \mathbf{f}^*\|_{\infty} \leq 0.4X^*B^*\gamma/n$ by (A.10). Finally, $\textcircled{4}$ follows from that T^* satisfies the separation condition (2.5): $\Delta(T^*) \geq 2.5009/n$.

For bounding $\Delta(\tilde{T})$, first identify that $-\max_i |\tilde{f}_i - f_i^*| \geq -\max_i |f_i^\lambda - f_i^*|$, since the inner point \tilde{f}_i is included in the interval $[f_i^*, f_i^\lambda]$ and hence the length of the subinterval $[\tilde{f}_i, f_i^*]$ is less than the entire interval $[f_i^*, f_i^\lambda]$. Then we immediately arrive at $\Delta(\tilde{T}) > 2.5/n$.

For $\Delta(\hat{T})$, we have

$$\begin{aligned}
\Delta(\hat{T}) &= \min_{i \neq j} |\hat{f}_i - \hat{f}_j| \\
&= \min_{i \neq j} |\hat{f}_i - f_i^\lambda + f_i^\lambda - f_j^\lambda + f_j^\lambda - \hat{f}_j| \\
&\stackrel{\textcircled{1}}{\geq} \min_{i \neq j} |f_i^\lambda - f_j^\lambda| - \max_i |\hat{f}_i - f_i^\lambda| - \max_j |\hat{f}_j - f_j^\lambda| \\
&\stackrel{\textcircled{2}}{\geq} \Delta(T^\lambda) - 2\|\hat{\mathbf{f}} - \mathbf{f}^\lambda\|_\infty \\
&\stackrel{\textcircled{2}}{\geq} \Delta(T^\lambda) - 2(14.08)B^*\gamma/n,
\end{aligned}$$

where $\textcircled{1}$ follows from the triangle inequality and $\textcircled{2}$ follows from the definition of $\|\hat{\mathbf{f}} - \mathbf{f}^\lambda\|_\infty$. $\textcircled{3}$ follows from that $\|\hat{\mathbf{f}} - \mathbf{f}^\lambda\|_\infty \leq 0.4(35.2)B^*\gamma/n = 14.08B^*\gamma/n$ by (A.10). Finally following from the SNR condition (2.6) and $\Delta(T^\lambda) \geq 2.5001/n$, we then have $\Delta(\hat{T}) \geq 2.5001/n - 2(14.08) \times 10^{-7}/n > 2.5/n$.

$\Delta(\tilde{T}^\lambda) \geq 2.5/n$ holds by the same strategy as $\Delta(\tilde{T}) > 2.5/n$. \square

References

- [1] E. J. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, *Communications on Pure and Applied Mathematics* 67 (6) (2014) 906–956.
- [2] E. J. Candès, C. Fernandez-Granda, Super-resolution from noisy data, *Journal of Fourier Analysis and Applications* 19 (6) (2013) 1229–1254.
- [3] C. Fernandez-Granda, Super-resolution of point sources via convex programming, *Information and Inference: A Journal of the IMA* 5 (3) (2016) 251–303.
- [4] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed Sensing Off the Grid, *Information Theory, IEEE Transactions on* 59 (11) (2013) 7465–7490.
- [5] G. Tang, B. N. Bhaskar, B. Recht, Near minimax line spectral estimation., *IEEE Transactions on Information Theory* 61 (1) (2015) 499–512.
- [6] G. Tang, P. Shah, B. N. Bhaskar, B. Recht, Robust line spectral estimation, in: 2014 48th Asilomar Conference on Signals, Systems and Computers, 2014, pp. 301–305. doi:10.1109/ACSSC.2014.7094450.
- [7] 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* (2016) iax005.
- [8] P. Stoica, N. Arye, MUSIC, maximum likelihood, and Cramer-Rao bound, *Acoustics, Speech and Signal Processing, IEEE Transactions on* 37 (5) (1989) 720–741.
- [9] 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.
- [10] 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.
- [11] N. Rao, P. Shah, S. Wright, Forward–backward greedy algorithms for atomic norm regularization, *IEEE Transactions on Signal Processing* 63 (21) (2015) 5798–5811.

- [12] A. Tewari, P. K. Ravikumar, I. S. Dhillon, Greedy algorithms for structurally constrained high dimensional problems, in: *Advances in Neural Information Processing Systems*, 2011, pp. 882–890.
- [13] N. Boyd, G. Schiebinger, B. Recht, The alternating descent conditional gradient method for sparse inverse problems, in: *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015 IEEE 6th International Workshop on, IEEE, 2015, pp. 57–60.
- [14] A. Eftekhari, M. B. Wakin, Greed is super: A new iterative method for super-resolution, in: *Global Conference on Signal and Information Processing (GlobalSIP)*, 2013 IEEE, IEEE, 2013, pp. 631–631.
- [15] B. N. Bhaskar, G. Tang, B. Recht, Atomic norm denoising with applications to line spectral estimation, *Signal Processing*, IEEE Transactions on 61 (23) (2013) 5987–5999.
- [16] G. Tang, Resolution limits for atomic decompositions via markov-bernstein type inequalities, in: *Sampling Theory and Applications (SampTA)*, 2015 International Conference on, IEEE, 2015, pp. 548–552.
- [17] C. Fernandez-Granda, Support detection in super-resolution, in: *Sampling Theory and Applications (SampTA)*, 2013 International Conference on, IEEE, 2013, pp. 145–148.
- [18] V. Duval, G. Peyré, Exact support recovery for sparse spikes deconvolution, *Foundations of Computational Mathematics* 15 (5) (2015) 1315–1355.
- [19] Q. Denoyelle, V. Duval, G. Peyré, Support recovery for sparse super-resolution of positive measures, *Journal of Fourier Analysis and Applications* 23 (5) (2017) 1153–1194.
- [20] V. I. Morgenshtern, E. J. Candes, Super-resolution of positive sources: The discrete setup, *SIAM Journal on Imaging Sciences* 9 (1) (2016) 412–444.
- [21] S. M. Kay, *Modern spectral estimation*, Pearson Education India, 1988.
- [22] P. Stoica, R. L. Moses, *Introduction to spectral analysis*, Vol. 1, Prentice hall Upper Saddle River, 1997.
- [23] R. de Prony, *Essai experimental et analytique*, *J. Ec. Polytech.(Paris)* 2 (1795) 24–76.
- [24] M. Kahn, M. Mackisack, M. Osborne, G. Smyth, On the consistency of prony’s method and related algorithms, *Journal of Computational and Graphical Statistics* 1 (4) (1992) 329–349.
- [25] Y. Hua, T. K. Sarkar, Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise, *Acoustics, Speech and Signal Processing*, IEEE Transactions on 38 (5) (1990) 814–824.
- [26] R. Schmidt, Multiple emitter location and signal parameter estimation, *IEEE Trans. on Antennas and Propagation* 34 (3) (1986) 276–280.
- [27] R. Roy, T. Kailath, ESPRIT - estimation of signal parameters via rotational invariance techniques, *IEEE Trans. on Acoustics, Speech and Signal Processing* 37 (7) (1989) 984–995.
- [28] P. Stoica, N. Arye, MUSIC, maximum likelihood, and Cramér-Rao bound, *IEEE Trans. on Acoustics, Speech and Signal Processing* 37 (5) (1989) 720–741.
- [29] M. Vetterli, P. Marziliano, T. Blu, Sampling signals with finite rate of innovation, *IEEE Trans. on Signal Processing* 50 (6) (2002) 1417–1428.
- [30] D. L. Donoho, Compressed sensing, *IEEE Transactions on information theory* 52 (4) (2006) 1289–1306.
- [31] E. J. Candès, et al., Compressive sampling, in: *Proceedings of the international congress of mathematicians*, Vol. 3, Madrid, Spain, 2006, pp. 1433–1452.
- [32] R. G. Baraniuk, Compressive sensing, *IEEE signal processing magazine* 24 (4).

- [33] Y. Chi, L. L. Scharf, A. Pezeshki, A. R. Calderbank, Sensitivity to basis mismatch in compressed sensing, *IEEE Transactions on Signal Processing* 59 (5) (2011) 2182–2195.
- [34] G. Tang, B. N. Bhaskar, B. Recht, Sparse recovery over continuous dictionaries-just discretize, in: 2013 Asilomar Conference on Signals, Systems and Computers, IEEE, 2013, pp. 1043–1047.
- [35] V. Duval, G. Peyré, Sparse spikes deconvolution on thin grids, *arXiv preprint arXiv:1503.08577*.
- [36] Y. Li, Y. Chi, Off-the-grid line spectrum denoising and estimation with multiple measurement vectors, *IEEE Transactions on Signal Processing* 64 (5) (2016) 1257–1269.
- [37] Z. Yang, L. Xie, Exact joint sparse frequency recovery via optimization methods, *IEEE Transactions on Signal Processing* 64 (19) (2014) 5145–5157.
- [38] S. Li, D. Yang, G. Tang, M. B. Wakin, Atomic norm minimization for modal analysis from random and compressed samples, *IEEE Transactions on Signal Processing* 66 (7) (2018) 1817–1831.
- [39] J.-M. Azais, Y. De Castro, F. Gamboa, Spike detection from inaccurate samplings, *Applied and Computational Harmonic Analysis* 38 (2) (2015) 177–195.
- [40] M. J. Wainwright, Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (lasso), *IEEE transactions on information theory* 55 (5) (2009) 2183–2202.
- [41] S. Foucart, H. Rauhut, *A mathematical introduction to compressive sensing*, Springer, 2013.
- [42] J. Nocedal, S. Wright, *Numerical optimization*, Springer Science & Business Media, 2006.
- [43] D. P. Bertsekas, *Nonlinear programming*, Athena scientific Belmont, 1999.