Local Geometry of Nonconvex Spike Deconvolution from Low-Pass Measurements

Maxime Ferreira Da Costa, Yuejie Chi

Local Geometry of Nonconvex Spike Deconvolution from Low-Pass Measurements

Maxime Ferreira Da Costa*
CentraleSupélec | Université Paris-Saclay

Yuejie Chi†
Carnegie Mellon University

August 2022; Revised February 2023

Abstract

Spike deconvolution is the problem of recovering the point sources from their convolution with a known point spread function, which plays a fundamental role in many sensing and imaging applications. In this paper, we investigate the local geometry of recovering the parameters of point sources—including both amplitudes and locations—by minimizing a natural nonconvex least-squares loss function measuring the observation residuals. We propose preconditioned variants of gradient descent (GD), where the search direction is scaled via some carefully designed preconditioning matrices. We begin with a simple fixed preconditioner design, which adjusts the learning rates of the locations at a different scale from those of the amplitudes, and show it achieves a linear rate of convergence—in terms of *entrywise* errors—when initialized close to the ground truth, as long as the separation between the true spikes is sufficiently large. However, the convergence rate slows down significantly when the dynamic range of the source amplitudes is large. To bridge this issue, we introduce an adaptive preconditioner design, which compensates for the learning rates of different sources in an iteration-varying manner based on the current estimate. The adaptive design provably leads to an accelerated convergence rate that is independent of the dynamic range, highlighting the benefit of adaptive preconditioning in nonconvex spike deconvolution. Numerical experiments are provided to corroborate the theoretical findings.

Keywords: nonconvex spike deconvolution, preconditioned gradient descent, local geometry

Contents

*M. Ferreira Da Costa is with the Laboratory of Signals and Systems (L2S) at CentraleSupélec, Université Paris-Saclay. Part of this work was done while he was with University of Southern California and Carnegie Mellon University. Email: maxime.ferreira@centralesupelec.fr.

†Y. Chi is with Department of Electrical and Computer Engineering, Carnegie Mellon University. The work of Y. Chi is supported in part by Office of Naval Research under N00014-19-1-2404, and by National Science Foundation under CAREER ECCS-1818571 and ECCS-2126634. Email: yuejiechi@cmu.edu.

1 Introduction

Spike deconvolution, also known as super-resolution [1], is the task of recovering a stream of point sources from their convolution with a point spread function (PSF). This classical problem is at the core of many sensing and imaging modalities, including but not limited to radar, sonar, optical imaging, neuroimaging, and communication systems [2, 3, 4, 5]. The PSF, which models the physical limitations of the imaging device involved in the experimental process, is commonly assumed to act as a band-limited and shift-invariant low-pass filter on the point sources [6, 7]. The sharpness of the original sources, modeled by Dirac impulses, is degraded through the convolution process, introducing undesirable ambiguity on the complex amplitudes and locations of the sources. The spike deconvolution task amounts to inverting the low-passing effects of the PSF, and to recovering the original sources as precisely as possible.

There is a rich literature on algorithmic investigations of the spike deconvolution problem, ranging from classical root-finding methods such as Prony's method, subspace methods such as MUSIC [8, 9], ESPRIT [10] and matrix pencil [11], to more recent optimization methods such as atomic norm minimization (*a.k.a. total variation minimization*) [12, 13, 14, 15, 16] and basis pursuit [17, 18]. While classical approaches harness the algebraic properties of complex exponentials by mapping the observations onto a low-dimensional linear subspace to recover the parameters of interest, optimization methods, on the other hand, attempt to recover the parameters via minimizing some carefully-designed loss function. As such, optimization methods tend to be more versatile in adapting to different imaging modalities, as well as amenable to modern advances in large-scale optimization. Inspired by the development of compressive sensing [19, 20], initial approaches for spike deconvolution relies on a discretization of the spike locations, and then attempts to recover a sparse solution using sparsity-promoting convex relaxations such as the LASSO [21, 22]. However, the fundamental issue of basis mismatch [23] inherent to the discretization process may significantly hinder the localization performance, and increasing the grid size to reach finer precision levels leads to higher computational cost. Therefore, there has been a surge of interest in developing provably correct convex programs—such as the atomic norm minimization framework mentioned earlier—for spike deconvolution *over the continuum* in recent years, with strong performance guarantees developed under sufficient separations between the point sources [24, 25, 26, 27]. Nonetheless, the atomic norm framework requires solving a semidefinite program whose complexity scales at least cubically with respect to the signal length; and therefore is computationally expensive and memory inefficient. In addition, although it is in principle possible to examine the so-called dual polynomial to localize the sources [14], it often boils down to an additional post-processing step on the output of the convex program to recover the source parameters, which may hamper the guarantee of the overall procedure.

Motivated by the recent success of nonconvex methods, especially simple first-order methods, in various signal estimation and machine learning tasks [28, 29], we are interested in understanding the efficacy of first-order methods in nonconvex spike deconvolution. In fact, first-order methods have already been popular

empirically for spike deconvolution, but little is known about their theoretical underpinnings [30]. As a first step, this paper focuses on the local geometry and performance guarantees of recovering the parameters of point sources—including both amplitudes and locations—by minimizing a natural nonconvex least-squares loss function measuring the observation residuals.

1.1 Observation model

Formally, we formulate the spike deconvolution problem as follows. Consider a vector of $2r$ parameters $\boldsymbol{\theta}^\star = [a_1^\star, \ldots, a_r^\star, \tau_1^\star, \ldots, \tau_r^\star]^\mathsf{T}$, where $a_\ell^\star \in \mathbb{C}$ and $\tau_\ell^\star \in \mathbb{R}$ correspond to the complex amplitude and location of the ℓ-th spike, respectively, $\ell = 1, \ldots, r$. Denoting by $\mathcal{M}$ the set of Radon measures over the reals, we assume that the point source signal $\mu^\star \in \mathcal{M}$ to resolve is of the form

$$\mu^\star = \mu(\boldsymbol{\theta}^\star) = \sum_{\ell=1}^{r} a_\ell^\star \delta_{\tau_\ell^\star}, \tag{1}$$

where δ_τ stands for the Dirac function located at $\tau \in \mathbb{R}$. Let us further denote the largest and the smallest amplitude of the spikes as

$$a_{\max}^\star = \max_{1\le\ell\le r} |a_\ell^\star| = \|\boldsymbol{a}^\star\|_\infty, \qquad a_{\min}^\star = \min_{1\le\ell\le r} |a_\ell^\star|.$$

The dynamic range of the measure $\mu^\star$, an important quantity that will be used repetitively later, is thus defined as $a_{\max}^\star / a_{\min}^\star$. Denoting by $g \in L_1(\mathbb{R})$ the PSF, the temporal signal $x \in L_1(\mathbb{R})$ resulting from the convolution of the point source signal $\mu^\star$ and the PSF g reads

$$x(\tau) = (g * \mu^\star)(\tau) = \sum_{\ell=1}^{r} a_\ell^\star g(\tau - \tau_\ell^\star), \tag{2}$$

where $*$ denotes the convolution product.

A versatile observation model commonly encountered in practice considers the measurements to be taken from a uniform sampling of the Fourier transform of the temporal signal x. Denote by $\mathcal{F}(\cdot)$ the Fourier transform of a measure μ lying in $\mathcal{M}$, given by

$$\mathcal{F}(\mu)(f) = \int_{\mathbb{R}} e^{-i2\pi f\tau} \mathrm{d}\mu(\tau), \quad \forall \mu \in \mathcal{M},\ \forall f \in \mathbb{R}. \tag{3}$$

Denote by $G = \mathcal{F}(g)$ and $X = \mathcal{F}(x)$ the Fourier transform of g and x, respectively. We assume that the PSF is band-limited within the bandwidth $B = 1$ so that it constrains no frequency greater than $1/2$, *i.e.* $G(f) = 0$ for $|f| > \frac{1}{2}$.[1] For convenience, we assume that an odd number $N = 2n + 1$ of measurements are taken in the Fourier domain, uniformly spaced in the bandwidth $[-\frac{1}{2}, \frac{1}{2}]$. The sampled signal $\boldsymbol{x} \in \mathbb{C}^N$ writes

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{\Phi}(\mu^\star) \\ &= \mathrm{diag}(\boldsymbol{g}) \left[\mathcal{F}(\mu)\left(-\frac{n}{N}\right), \ldots, \mathcal{F}(\mu)\left(\frac{n}{N}\right)\right]^\top, \end{aligned} \tag{4}$$

where $\boldsymbol{\Phi} : \mathcal{M} \to \mathbb{C}^N$ represents the observation operator and $\boldsymbol{g} \in \mathbb{C}^N$ is a vector with generic term $g_k = G\left(\frac{k}{N}\right)$ for $k = -n, \ldots, n$. Up to a scaling, we assume $\boldsymbol{g}$ to have unit Euclidean norm, *i.e.* $\|\boldsymbol{g}\|_2 = 1$. The goal of spike deconvolution is thus to recover the measure $\mu^\star$, or equivalently, the parameter $\boldsymbol{\theta}^\star$, from $\boldsymbol{x}$.

1.2 Our contributions

In the rest of this paper, we assume that the model order r is known, and consider a natural nonconvex loss function, which aims to minimize the quadratic loss of the parameters $\boldsymbol{\theta} = [a_1, \ldots, a_r, \tau_1, \ldots, \tau_r]^\mathsf{T}$ of the Radon measures, given by

$$\min_{\boldsymbol{\theta}} \ \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2} \|\boldsymbol{\Phi}(\mu(\boldsymbol{\theta})) - \boldsymbol{x}\|_2^2. \tag{5}$$

[1] The normalization $B = 1$ is made without loss of generality, up to a rescaling of the source locations $\tau_\ell^\star$. Assuming that $\tau_\ell^\star \in [-\frac{T}{2}, \frac{T}{2}]$, where T is the length of the observation window, the deconvolution problem only depends on the time-bandwidth product $T \cdot B$ [31].

Due to the nonlinear form of the parameters, the loss function $\mathcal{L}(\boldsymbol{\theta})$ is clearly nonconvex. As a first step towards nonconvex spike deconvolution, we are interested in understanding the local geometry of the loss function (5) and its implications on the computation efficacy of first-order methods. Without loss of generality, we assume the autocorrelation function h of the PSF g to be a triangular low pass function, *i.e.*

$$h(\tau) = \int_{-\infty}^{\infty} f(u+\tau)\overline{f(u)}\mathrm{d}u = C\frac{\sin^2\left(\frac{\pi\tau}{2}\right)}{\left(\frac{\pi\tau}{2}\right)^2}, \tag{6}$$

where $C > 0$ is a constant, and that their respective Fourier transform $H(f)$ and $G(f)$ are linked through the relation $G(f) = \sqrt{H(f)}$ for all $f \in \mathbb{R}$. It comes that

$$g_k = G\left(\frac{k}{N}\right) = \sqrt{H\left(\frac{k}{N}\right)} = \sqrt{\frac{1}{n+1}\left(1 - \frac{|k|}{n+1}\right)}, \tag{7}$$

for all $k = -n, \dots n$ after rescaling with the constraint $\|\boldsymbol{g}\|_2 = 1$. Note that the triangular low-pass function and the Fejér kernel—its discrete counterpart—play an important role in the deconvolution literature and have been extensively proposed as a convolution kernel to evaluate the norm and distance between Radon measures [14]. Additionally, our main results can be re-derived with any other bandlimited PSF g by following analogous reasoning, as long as it autocorrelation h is an *absolutely integrable* function.

Concretely, we propose and analyze preconditioned variants of gradient descent (GD), where the search direction is scaled via some carefully designed preconditioning matrices. Our contributions are summarized as follows.

- We begin with a simple fixed preconditioner design, which adjusts the learning rates of the locations at a different scale from those of the amplitudes, and show it achieves a linear rate of convergence—in terms of *entrywise* errors—when initialized close to the ground truth, as long as the separation between the true spikes is sufficiently large. However, the convergence rate slows down significantly when the dynamic range of the source amplitudes is large.
- To bridge this issue, we introduce an adaptive preconditioner design, which compensates the learning rates of different sources in an iteration-varying manner based on the current estimate. The adaptive design provably leads to an accelerated convergence rate that is independent of the dynamic range, highlighting the benefit of adaptive preconditioning in nonconvex spike deconvolution.

Our result is based on understanding the geometric properties of scaled Hessian matrices via a set of novel summation bounds on the absolute sums of sampled Fejér kernels and higher-order derivatives, which might be of independent interest in other contexts.

1.3 Related work

The closest work to ours on recovering spike signals from low-pass observations using nonconvex optimization is [32]. The radius of the basin of attraction for gradient descent is characterized whenever the observation operator satisfies the restricted isometry property over the set of well-separated sparse measures. Although the problem setup is versatile, specializing this result in our context of low-pass measurements yields a convergence region whose size scales inversely with the number of sources, which is pessimistic when the number of sources is large. Moreover, the analysis in [32] focuses on the Euclidean error of the parameters, while we focus on the entrywise error, which is more meaningful for gauging the recovery quality of the point sources. Projected gradient methods, which merge pairs of colliding spikes at each iteration, have been proposed in [33, 34] but without theoretical convergence guarantees.

Our work can be viewed as falling into a growing line of research on developing provably efficient nonconvex methods—especially first-order methods—for high-dimensional signal estimation, examples including phase retrieval [35, 36], low-rank matrix estimation [37, 28], blind (sparse) deconvolution [38, 39, 40], dictionary learning [41, 42], multi-channel sparse deconvolution [43, 44], and so on. In particular, the preconditioned gradient methods considered in this paper are motivated by [45, 46, 47], which demonstrated that preconditioning can efficiently accelerate the convergence of gradient descent in ill-conditioned low-rank estimation.

1.4 Notation and paper organization

Vectors and matrices are denoted by boldface and capital boldface letters, respectively. Vectors $\boldsymbol{x} \in \mathbb{C}^N$ with odd dimension $N = 2n+1$ are indexed between $-n$ and n, so that $\boldsymbol{x} = [x_{-n}, \dots, x_n]^\top$ for convenience. Transpose and Hermitian transpose of a vector or a matrix $\boldsymbol{A}$ are denoted by $\boldsymbol{A}^\top$ and $\boldsymbol{A}^\mathsf{H}$, respectively. Furthermore, the adjoint of the operator $\boldsymbol{\Phi}$ is denoted by $\boldsymbol{\Phi}^*$. We write $\mathbf{1}_d$ and $\mathbf{0}_d$ the all-one and null vector (or matrix) in dimension d, respectively. With a slight abuse of notation, we denote by $|\boldsymbol{a}|$, $|\boldsymbol{a}|^2$, $\boldsymbol{a}^{-1}$ the vector with entries equal to the modulus, the squared modulus, and the inverse of the entries of $\boldsymbol{a}$, respectively. The element-wise product between two vectors $\boldsymbol{a}$ and $\boldsymbol{a}'$ is written as $\boldsymbol{a} \odot \boldsymbol{a}'$. We denote by $\mathcal{D}_\ell$ the space of ℓ-times differentiable functions of the real variable. For any function $h \in \mathcal{D}_\ell$, we write its ℓth derivative $h^{(\ell)}$. We denote by $\langle \cdot\,, \cdot \rangle$ and $\langle \cdot\,, \cdot \rangle_\mathbb{R} = \Re\left(\langle \cdot\,, \cdot \rangle\right)$ the usual inner product and real inner product between Radon measure, respectively. Additionally, we let $\delta^{(\ell)} \in \mathcal{M}$ be the functional which satisfies

$$g^{(\ell)}(\tau) = \left\langle \delta_\tau^{(\ell)}, g \right\rangle, \quad \forall g \in \mathcal{D}_\ell, \forall \tau \in \mathbb{R}. \tag{8}$$

Fejér kernel We denote by $F_N(\cdot)$ the normalized Fejér kernel of order $N = 2n+1$ defined by

$$\begin{aligned} F_N(t) &= \frac{1}{n+1} \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e^{i2\pi kt} \\ &= \begin{cases} \frac{\sin^2(\pi(n+1)t)}{(n+1)^2 \sin^2(\pi t)} & \text{if } t \notin \mathbb{Z} \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{9}$$

The Fejér kernel is a trigonometric polynomial, hence it is infinitely differentiable. We point out that the second derivative of $F_N(\cdot)$ at the origin satisfies

$$\begin{aligned} F_N''(0) &= \frac{1}{n+1} \sum_{k=-n}^{n} -4\pi^2 k^2 \left(1 - \frac{|k|}{n+1}\right) \\ &= -\frac{2}{3}\pi^2 n(n+2) < 0. \end{aligned} \tag{10}$$

Some of its properties, key to this paper, are derived and discussed in Appendix A. The Fejér kernel plays an important role in the sequel as the Gramian $\boldsymbol{\Phi}^*\boldsymbol{\Phi} : \mathcal{M} \to \mathcal{M}$ of the observation operator $\boldsymbol{\Phi}$ is a convolution product with F_N, *i.e.*

$$\boldsymbol{\Phi}^*\boldsymbol{\Phi}(\mu) = F_N * \mu, \quad \forall \mu \in \mathcal{M}. \tag{11}$$

Wrap-around distance For any set of r points $\boldsymbol{\tau} = \{\tau_1, \dots, \tau_r\} \subset \mathbb{T}$, we denote by $\Delta(\boldsymbol{\tau})$ is minimal *wrap-around distance*, defined by

$$\Delta(\boldsymbol{\tau}) \triangleq \min_{\ell \neq \ell'} \inf_{p \in \mathbb{Z}} |\tau_\ell - \tau_{\ell'} + p|\,. \tag{12}$$

The rest of this paper is organized as follows. Section 2 starts by defining the preconditioned gradient methods and two of its designs with provable local convergence guarantees, using a fixed preconditioner and an adaptive preconditioner in Section 2.2 and Section 2.3, respectively. Section 3 provides the analysis of the main theorems by controlling the conditioning of the scaled Hessian matrix of the loss function in a neighborhood of the ground truth. Numerical experiments are provided in Section 4 to corroborate our findings. Finally, a brief conclusion is drawn in Section 5.

2 How does preconditioning help local convergence?

2.1 Preconditioned gradient descent

Recognizing that the parameters corresponding to the amplitudes and locations may require different treatments, we consider iterates of preconditioned gradient descent (GD) to recover the ground truth parameters, where the preconditioner can possibly be iteration-varying. Given an initialization point $\boldsymbol{\theta}_0 \in \mathbb{C}^{2r}$,

the update sequence of preconditioned GD is obtained by successively moving oppositely along the direction of a *linear transform* of the gradient. More specifically, the update rule reads

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \boldsymbol{P}_k \nabla \mathcal{L}(\boldsymbol{\theta}_k), \tag{13}$$

where $\boldsymbol{\theta}_k = [\boldsymbol{a}_k^\top, \boldsymbol{\tau}_k^\top]^\top$ is the k-th iterate, $\boldsymbol{P}_k \in \mathbb{C}^{2r\times 2r}$ is a preconditioning matrix (also called preconditioner) that can vary at each iteration; the choice of $\boldsymbol{P}_k$ will be detailed momentarily. Here, it is worth noticing that there are no additional learning rates in (13), which can be thought of as already absorbed and set within the preconditioner $\boldsymbol{P}_k$. By analogy with the celebrated Newton-Raphson method, which selects $\boldsymbol{P}_k = \nabla^2 \mathcal{L}(\boldsymbol{\theta}_k)^{-1}$ (which might however be computationally expensive), the role of the preconditioning matrix $\boldsymbol{P}_k$ is to balance the local optimization landscape towards a quadratic function to improve the convergence rate towards a local minimum over the vanilla gradient method. By basic calculation, the gradient $\nabla \mathcal{L}(\boldsymbol{\theta})$ at point $\boldsymbol{\theta} = [a_1, \ldots, a_r, \tau_1, \ldots, \tau_r]^\mathsf{T}$ is given by

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{L}(\boldsymbol{\theta})}{\mathrm{d}a_j} &= \big\langle \boldsymbol{\Phi}(\delta_{\tau_j}), \boldsymbol{\Phi}(\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \big\rangle \\
&= \big\langle \delta_{\tau_j}, \boldsymbol{\Phi}^*\boldsymbol{\Phi}(\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \big\rangle \\
&= \big\langle \delta_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \big\rangle \\
&= \sum_{\ell=1}^r a_\ell F_N(\tau_j - \tau_\ell) - \sum_{\ell=1}^r a_\ell^\star F_N(\tau_j - \tau_\ell^\star)
\end{aligned} \tag{14a}$$

for $j = 1, \ldots, r$, and similarly,

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{L}(\boldsymbol{\theta})}{\mathrm{d}\tau_j} &= \Big\langle \boldsymbol{\Phi}(a_j \delta'_{\tau_j}), \boldsymbol{\Phi}(\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \Big\rangle_{\mathbb{R}} \\
&= \Big\langle a_j \delta'_{\tau_j}, \boldsymbol{\Phi}^*\boldsymbol{\Phi}(\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \Big\rangle_{\mathbb{R}} \\
&= \Big\langle a_j \delta'_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \Big\rangle_{\mathbb{R}} \\
&= \Re\Big(\overline{a_j} \Big(\sum_{\ell=1}^r a_\ell F'_N(\tau_j - \tau_\ell) - \sum_{\ell=1}^r a_\ell^\star F'_N(\tau_j - \tau_\ell^\star) \Big) \Big)
\end{aligned} \tag{14b}$$

for $j = 1, \ldots, r$.

In this paper, we are particularly interested in preconditioning matrices $\boldsymbol{P}_k$ that are diagonally structured, so they do not add computation overhead compared with vanilla gradient methods. In the sequel, we study the basin of attraction and the convergence rate of preconditioned GD for two different preconditioning strategies. The first consists of selecting a time-invariant, diagonal preconditioning matrix $\boldsymbol{P} = \boldsymbol{P}_k$ whose role is to judiciously renormalize the learning rates between the amplitudes $\boldsymbol{a} \in \mathbb{C}^r$ and the locations $\boldsymbol{\tau} \in \mathbb{R}^r$ which are of different units. The second strategy seeks to dynamically update the preconditioning matrix $\boldsymbol{P}_k$ based on the current iterate to better approximate the inverse of the Hessian matrix around the point $\boldsymbol{\theta}_k$ and accelerate convergence.

2.2 Invariant preconditioning

In this section, we seek to recover the ground truth parameter $\boldsymbol{\theta}^\star$ from an instance of the preconditioned GD algorithm (13) where the sequence of preconditioning matrices is constant, *i.e.* $\boldsymbol{P} = \boldsymbol{P}_k$ for all $k \in \mathbb{N}$. We fix

$$\boldsymbol{P} = \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_r \\ -F_N''(0)^{-1} A^{-2} \mathbf{1}_r \end{bmatrix} \right), \tag{15}$$

where $A > 0$ is an input parameter that controls the ratio between the learning rate applied to the amplitudes $\boldsymbol{a}_k$ of the sources and that applied to the locations $\boldsymbol{\tau}_k$ of the sources throughout the iterative process.

Performance metric To gauge the performance, we define by $\boldsymbol{S} \in \mathbb{C}^{2r \times 2r}$ the weighting matrix

$$\boldsymbol{S} = \operatorname{diag}\left(\begin{bmatrix} \boldsymbol{a}^{\star -1} \\ \sqrt{-F_N''(0)}\mathbf{1}_r, \end{bmatrix}\right), \tag{16}$$

where $F_N''(0)$ is given in (10), and study the convergence properties of preconditioned GD in terms of the infinity norm weighted by the matrix $\boldsymbol{S}$, *i.e.*

$$\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\|_\infty = \max_j \left\{ \frac{\left|a_{k,j} - a_j^\star\right|}{\left|a_j^\star\right|}, \sqrt{-F_N''(0)}\left|\tau_{k,j} - \tau_j^\star\right| \right\}. \tag{17}$$

Intuitively, the role of this scaling is to analyze a unitless metric that decorrelates the error with the dynamic range of the sources and with the problem dimension, as we have $\sqrt{-F_N''(0)} = \mathcal{O}(n)$ and that the error on the source locations is expected to be inversely proportional to the number of observation: $\|\boldsymbol{\tau}_k - \boldsymbol{\tau}^\star\|_\infty = \mathcal{O}\left(n^{-1}\right)$.

The following theorem establishes the linear convergence of preconditioned GD with a fixed preconditioning matrix $\boldsymbol{P}$ whenever the input parameter A is properly set, and the initial point $\boldsymbol{\theta}_0$ is close enough to the ground truth $\boldsymbol{\theta}^\star$, as long as the true spikes are sufficiently separated.

Theorem 1 (Linear convergence with invariant preconditioner)**.** *Suppose that $n \geq 2$ and that the input parameter A satisfies $\|\boldsymbol{a}^\star\|_\infty \leq \frac{3}{2}A$. Moreover, assume that*

$$\eta := 276.21 \frac{A^2 \|\boldsymbol{a}^\star\|_\infty}{\left(a_{\min}^\star\right)^3} \left((n+1)\,\Delta(\boldsymbol{\tau}^\star)\right)^{-2} < 1, \tag{18}$$

then if the initial point $\boldsymbol{\theta}_0 = [\boldsymbol{a}_0^\top, \boldsymbol{\tau}_0^\top]^\top$ satisfies

$$\|\boldsymbol{S}\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}^\star\right)\|_\infty \leq \frac{1}{2}, \tag{19}$$

the iterates $\{\boldsymbol{\theta}_k\}$ of preconditioned GD (13) with a fixed preconditioner (15) converge towards $\boldsymbol{\theta}^\star$ according to

$$\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\|_\infty \leq \left(1 - \frac{1}{4}\frac{\left(a_{\min}^\star\right)^2}{A^2}\left(1 - \eta\right)\right)^k \|\boldsymbol{S}\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}^\star\right)\|_\infty \tag{20}$$

for all $k \in \mathbb{N}$.

Theorem 1 indicates that preconditioned GD admits a linear rate of convergence as long as the separation condition $\Delta(\boldsymbol{\tau}^\star)$ is sufficiently large with respect to the dynamic range, *i.e.*

$$(n+1)\,\Delta(\boldsymbol{\tau}^\star) \gtrsim \left(\frac{\|\boldsymbol{a}^\star\|_\infty}{a_{\min}^\star}\right)^{3/2}. \tag{21}$$

Additionally, our finding is independent of the number of sources r of the input measure, both in terms of the size of the basin of attraction (cf. (19)) and the convergence rate. Faster convergence rate are achieved for smaller values of the parameter $0 < \eta < 1$, when the separation $\Delta(\boldsymbol{\tau}^\star)$ of the true spikes is larger or the dynamic range $\frac{\|\boldsymbol{a}^\star\|_\infty}{a_{\min}^\star}$ of the amplitudes is smaller. However, even for small values of η, the convergence rate ρ predicted by Theorem 1 is lower bounded by

$$\rho \geq 1 - \frac{1}{4}\left(\frac{a_{\min}^\star}{A}\right)^2 \geq 1 - \frac{9}{16}\left(\frac{a_{\min}^\star}{\|\boldsymbol{a}^\star\|_\infty}\right)^2. \tag{22}$$

This suggests that a high dynamic range will lead to a slow convergence rate, independently of the separation $\Delta(\boldsymbol{\tau}^\star)$. Additionally, the convergence guarantees established in Theorem 1 demand to adjust the input parameter A as a function of $\|\boldsymbol{a}^\star\|_\infty$, which can be impractical in scenarios with no postulate on the norm of the source amplitudes.

2.3 Adaptive preconditioning

In order to mitigate the limitations of the fixed preconditioning strategy presented in Section 2.2, we propose to study an instance of preconditioned GD where the preconditioner $\boldsymbol{P}_k$ varies at each iteration and is selected as a function of the current iterate $\boldsymbol{\theta}_k$. In particular, we fix

$$\boldsymbol{P}_k = \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_r \\ -F_N''(0)^{-1}\,|\boldsymbol{a}_k|^{-2} \end{bmatrix}\right). \tag{23}$$

Similar to Theorem 1, the next theorem guarantees a linear convergence rate of the iterates towards the ground truth $\boldsymbol{\theta}^\star$, provided a good enough initialization point $\boldsymbol{\theta}_0$, as long as the true spikes are sufficiently separated.

Theorem 2 (Linear convergence with adaptive preconditioner)**.** *Suppose that $n \geq 2$, and assume that*

$$\gamma := 11.60 \frac{\|\boldsymbol{a}\|_\infty}{a^\star_{\min}} \left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2} < \frac{1}{2}, \tag{24}$$

then if the initial point $\boldsymbol{\theta}_0 = [\boldsymbol{a}_0^\top, \boldsymbol{\tau}_0^\top]^\top$ satisfies

$$\|\boldsymbol{S}\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}^\star\right)\|_\infty \leq 1 - \sqrt{\frac{2}{3}}, \tag{25}$$

the iterates $\{\boldsymbol{\theta}_k\}$ of preconditioned GD (13) with an adaptive preconditioner (23) converge towards $\boldsymbol{\theta}^\star$ according to

$$\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\|_\infty \leq \left(\frac{1}{2} + \gamma\right)^k \|\boldsymbol{S}\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}^\star\right)\|_\infty \tag{26}$$

for all $k \in \mathbb{N}$.

Theorem 2 guarantees that preconditioned GD with an adaptive preconditioner achieves a *constant* linear rate of convergence in a similar basin of attraction, provided that the separation condition $\Delta(\boldsymbol{\tau}^\star)$ is sufficiently large with respect to the dynamic range, *i.e.*

$$(n+1)\,\Delta(\boldsymbol{\tau}^\star) \gtrsim \left(\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\right)^{1/2}, \tag{27}$$

which is much weaker than the requirement for the case using a fixed preconditioner, as indicated in Theorem 1. Consequently, this highlights the benefit of adaptive preconditioning in accelerating the convergence in the presence of high dynamic ranges for nonconvex spike deconvolution.

Remark. *We have not attempted to fully optimize the constants in the above theorems. Therefore, their values are set in a quite pessimistic fashion; see Section 4 for numerical experiments.*

3 Analysis

This section is devoted to proving the two main results of this paper comprised in Theorem 1 and Theorem 2. Before entering the core of the proofs, we first provide some warm-up analysis that will be required in the latter proofs.

3.1 Preliminaries

3.1.1 Contraction of entrywise errors

The convergence analysis of the preconditioned gradient method presented in Theorem 1 and Theorem 2 calls for understanding of the contraction properties of the sequence $\{\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\|_\infty\}$. Starting from the update rule (13), leveraging $\nabla\mathcal{L}(\boldsymbol{\theta}_\star) = \mathbf{0}$, and applying the fundamental theorem of calculus, we have

$$\boldsymbol{S}(\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_\star) = \boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{P}_k \nabla\mathcal{L}(\boldsymbol{\theta}_k) - \boldsymbol{\theta}_\star\right)$$

$$
\begin{aligned}
&= \boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star) - \boldsymbol{S}\boldsymbol{P}_k \left(\nabla\mathcal{L}(\boldsymbol{\theta}_k) - \nabla\mathcal{L}(\boldsymbol{\theta}_\star)\right) \\
&= \boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star) - \boldsymbol{S}\boldsymbol{P}_k \left(\int_0^1 \nabla^2\mathcal{L}\left(\boldsymbol{\theta}^\star + u\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right) \mathrm{d}u\right) \left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right) \\
&= \left[\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}_k \left(\int_0^1 \nabla^2\mathcal{L}\left(\boldsymbol{\theta}^\star + u\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right) \mathrm{d}u\right) \boldsymbol{S}^{-1}\right] \boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right). \qquad (28)
\end{aligned}
$$

Let $\mathcal{S}_k = \{\boldsymbol{\theta}^\star + u\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right) | u \in [0,1]\}$ be the line segment that connects $\boldsymbol{\theta}^\star$ and $\boldsymbol{\theta}_k$ in $\mathbb{C}^{2r}$, and denote by ρ_k the quantity

$$
\rho_k \triangleq \max_{\boldsymbol{\theta}\in\mathcal{S}_k} \left\|\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty, \qquad (29)
$$

where $\boldsymbol{H}(\boldsymbol{\theta}) = \nabla^2\mathcal{L}(\boldsymbol{\theta})$ is the Hessian of the loss function $\mathcal{L}$ at point $\boldsymbol{\theta}$. Continuing to bound (28) yields

$$
\begin{aligned}
\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_\star\right)\right\|_\infty &\le \left\|\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}_k \left(\int_0^1 \boldsymbol{H}\left(\boldsymbol{\theta}^\star + u\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right) \mathrm{d}u\right) \boldsymbol{S}^{-1}\right\|_\infty \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right)\right\|_\infty \\
&= \left\|\int_0^1 \left(\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}\left(\boldsymbol{\theta}^\star + u\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right)\boldsymbol{S}^{-1}\right) \mathrm{d}u\right\|_\infty \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right)\right\|_\infty \\
&\le \left(\int_0^1 \left\|\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}\left(\boldsymbol{\theta}^\star + u\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right)\boldsymbol{S}^{-1}\right\|_\infty \mathrm{d}u\right) \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right)\right\|_\infty \\
&\le \max_{\boldsymbol{\theta}\in\mathcal{S}_k} \left\{\left\|\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty\right\} \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right)\right\|_\infty \\
&= \rho_k \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_\star\right)\right\|_\infty. \qquad (30)
\end{aligned}
$$

Hence, the crux of the convergence analysis is to show that $\rho_k < 1$ (and control the size of ρ_k) uniformly over the segment $\mathcal{S}_k$ whenever the point $\boldsymbol{\theta}_k$ lies in an appropriate region centered around the ground truth $\boldsymbol{\theta}^\star$. Further analysis towards that goal requires an explicit derivation of the Hessian matrix $\boldsymbol{H}(\boldsymbol{\theta})$, which is done next.

3.1.2 Hessian decomposition

Recall that $\boldsymbol{H}(\boldsymbol{\theta}) = \nabla^2\mathcal{L}(\boldsymbol{\theta}) \in \mathbb{C}^{2r\times 2r}$ is the Hessian matrix of the loss function $\mathcal{L}$ in (5) at the point $\boldsymbol{\theta} = [\boldsymbol{a}^\top, \boldsymbol{\tau}^\top]^\top$. We decompose $\boldsymbol{H}(\boldsymbol{\theta})$ as

$$
\boldsymbol{H}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{H}_{a,a}(\boldsymbol{\theta}) & \boldsymbol{H}_{a,\tau}(\boldsymbol{\theta}) \\ \boldsymbol{H}_{a,\tau}^{\mathsf{H}}(\boldsymbol{\theta}) & \boldsymbol{H}_{\tau,\tau}(\boldsymbol{\theta}) \end{bmatrix}, \qquad (31)
$$

where each block is of size $r \times r$, with generic terms

$$
[\boldsymbol{H}_{a,a}(\boldsymbol{\theta})]_{(i,j)} = \frac{\mathrm{d}^2 f(\boldsymbol{\theta})}{\mathrm{d}a_i \mathrm{d}a_j}, \qquad [\boldsymbol{H}_{a,\tau}(\boldsymbol{\theta})]_{(i,j)} = \frac{\mathrm{d}^2 f(\boldsymbol{\theta})}{\mathrm{d}a_i \mathrm{d}\tau_j}, \qquad [\boldsymbol{H}_{\tau,\tau}(\boldsymbol{\theta})]_{(i,j)} = \frac{\mathrm{d}^2 f(\boldsymbol{\theta})}{\mathrm{d}\tau_i \mathrm{d}\tau_j}
$$

for $i, j = 1\ldots, r$. A direct calculation of the Hessian matrix $\boldsymbol{H}(\boldsymbol{\theta})$ (see, *e.g.* [32]) yields a decomposition of the form

$$
\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{G}(\boldsymbol{\theta}) + \boldsymbol{E}(\boldsymbol{\theta}), \qquad (32)
$$

where the terms $\boldsymbol{G}(\boldsymbol{\theta}) \in \mathbb{C}^{2r\times 2r}$ and $\boldsymbol{E}(\boldsymbol{\theta}) \in \mathbb{C}^{2r\times 2r}$ are described in the sequel.

Structure of $\boldsymbol{G}(\boldsymbol{\theta})$ The matrix $\boldsymbol{G}(\boldsymbol{\theta})$ can be written as

$$
\boldsymbol{G}(\boldsymbol{\theta}) = \mathrm{diag}\left(\begin{bmatrix} \mathbf{1}_r \\ \sqrt{-F_N''(0)}\boldsymbol{a} \end{bmatrix}\right)^{\mathsf{H}} \boldsymbol{D}(\boldsymbol{\tau}) \, \mathrm{diag}\left(\begin{bmatrix} \mathbf{1}_r \\ \sqrt{-F_N''(0)}\boldsymbol{a} \end{bmatrix}\right), \qquad (33)
$$

with $\boldsymbol{D}(\boldsymbol{\tau}) \in \mathbb{C}^{2r\times 2r}$ given with a block structure

$$
\boldsymbol{D}(\boldsymbol{\tau}) = \begin{bmatrix} \boldsymbol{D}_0(\boldsymbol{\tau}) & \boldsymbol{D}_1(\boldsymbol{\tau}) \\ \boldsymbol{D}_1(\boldsymbol{\tau})^{\mathsf{H}} & \boldsymbol{D}_2(\boldsymbol{\tau}) \end{bmatrix}. \qquad (34)
$$

The entries of the blocks $\boldsymbol{D}_0(\boldsymbol{\tau})$, $\boldsymbol{D}_1(\boldsymbol{\tau})$, $\boldsymbol{D}_2(\boldsymbol{\tau}) \in \mathbb{C}^{r\times r}$ are composed of

$$[\boldsymbol{D}_0(\boldsymbol{\tau})]_{(i,j)} = F_N(\tau_i - \tau_j), \tag{35a}$$

$$[\boldsymbol{D}_1(\boldsymbol{\tau})]_{(i,j)} = -F_N'(\tau_i - \tau_j)/\sqrt{-F_N''(0)}, \tag{35b}$$

$$[\boldsymbol{D}_2(\boldsymbol{\tau})]_{(i,j)} = F_N''(\tau_i - \tau_j)/F_N''(0) \tag{35c}$$

for all $i, j = 1, \ldots, r$. As shall be seen, the matrix $\boldsymbol{G}(\boldsymbol{\theta})$ is a relatively well-conditioned matrix whose spectrum can be controlled as a function of the separation parameter between the spikes $(n+1)\Delta(\boldsymbol{\tau}^\star)$, the dynamic range of the amplitudes $\frac{\|\boldsymbol{a}\|_\infty}{a^\star_{\min}}$, and the distance of $\boldsymbol{\theta}$ to the ground truth parameter $\boldsymbol{\theta}^\star$.

Structure of $\boldsymbol{E}(\boldsymbol{\theta})$ The matrix $\boldsymbol{E}(\boldsymbol{\theta})$ is given by the block structure decomposition

$$\boldsymbol{E}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{0}_{r\times r} & \boldsymbol{E}_1(\boldsymbol{\theta}) \\ \boldsymbol{E}_1(\boldsymbol{\theta})^{\mathsf{H}} & \boldsymbol{E}_2(\boldsymbol{\theta}) \end{bmatrix}, \tag{36}$$

where the entries of $\boldsymbol{E}_1(\boldsymbol{\theta})$, $\boldsymbol{E}_2(\boldsymbol{\theta}) \in \mathbb{C}^{r\times r}$ are given as

$$[\boldsymbol{E}_1(\boldsymbol{\theta})]_{(i,j)} = \begin{cases} \left\langle \delta'_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \right\rangle & i = j; \\ 0 & i \neq j; \end{cases} \tag{37a}$$

$$[\boldsymbol{E}_2(\boldsymbol{\theta})]_{(i,j)} = \begin{cases} a_j \left\langle \delta''_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \right\rangle & i = j; \\ 0 & i \neq j. \end{cases} \tag{37b}$$

It is worth noting that $\boldsymbol{E}(\boldsymbol{\theta}_\star) = \boldsymbol{0}$, hence $\boldsymbol{E}$ can be interpreted as a perturbation term that grows as $\boldsymbol{\theta}$ deviates from $\boldsymbol{\theta}_\star$.

With the above preliminaries, we are now ready to prove the main results of this paper. The following two sections present our convergence analyses for preconditioned GD using a fixed preconditioner and an adaptive preconditioner, respectively.

3.2 Proof of Theorem 1

We recall that $\boldsymbol{P}_k = \boldsymbol{P} = \operatorname{diag}\left(\begin{bmatrix} \boldsymbol{1}_r \\ -F_N''(0)^{-1}A^{-2}\boldsymbol{1}_r \end{bmatrix}\right)$ for all $k \in \mathbb{N}$ when the preconditioner is fixed. The contraction analysis (30) presented in Section 3.1.1 suggests that the contraction rate in Theorem 1 is controlled by the quantity $\left\|\boldsymbol{S}\boldsymbol{P}\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty$ in neighborhood around the ground truth $\boldsymbol{\theta}^\star$. The next theorem, whose proof is deferred to Appendix B, provides a uniform bound on this quantity on a neighborhood of the ground truth.

Theorem 3 (Uniform bound of the Hessian)**.** *Suppose that $n \geq 2$, $(n+1)\Delta(\boldsymbol{\tau}^\star) \geq 16.5$ and $\|\boldsymbol{\tau} - \boldsymbol{\tau}^\star\|_\infty \leq \frac{1}{4}\Delta(\boldsymbol{\tau}^\star)$. Let $\boldsymbol{\theta}_k = [\boldsymbol{a}_k^\top, \boldsymbol{\tau}_k^\top]^\top$ and assume that $A \geq \max\left\{\frac{3}{2}\|\boldsymbol{a}^\star\|_\infty, \|\boldsymbol{a}_k\|_\infty\right\}$ then the exist two positive constants*

$$K_\Delta = 2.13, \tag{38a}$$

$$K_\theta = 44.42, \tag{38b}$$

such that for all $\boldsymbol{\theta} = [\boldsymbol{a}^\top, \boldsymbol{\tau}^\top]^\top \in \mathcal{S}_k$ with $\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty < 1$, we have that

$$\begin{aligned} \left\|\boldsymbol{S}\boldsymbol{P}\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty \leq 1 - \left(\frac{a^\star_{\min}}{A}\right)^2 (1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2 \\ + (4K_\Delta + K_\theta \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty) \frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}} ((n+1)\Delta(\boldsymbol{\tau}^\star))^{-2} (1 + \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2. \end{aligned} \tag{39}$$

Theorem 3 provides a bound on $\left\|\boldsymbol{SPH}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty$ depending on the quantity $\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}$, which can be made small enough under the hypothesis of Theorem 1. We proceed with the rest of the proof by induction.

For the base case, it is trivial that (20) holds for $k=0$. We start the induction by assuming that

$$\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty \leq \left(1-\frac{1}{4}\left(\frac{a^\star_{\min}}{A}\right)^2(1-\eta)\right)^k \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\right)\right\|_\infty \tag{40}$$

holds for some $k\in\mathbb{N}$. We begin by verifying the assumptions of Theorem 3.

- First, as the dynamic range $\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\geq 1$, it is easy to see that the hypothesis (18) immediately implies that $(n+1)\Delta(\boldsymbol{\tau}^\star)\geq 16.6$.
- By the definition (17) and the induction hypothesis (40), it follows that

$$\begin{aligned}\left\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\right\|_\infty &\leq \frac{1}{\sqrt{-F_N''(0)}}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty \\ &\leq \frac{1}{\sqrt{-F_N''(0)}}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\right)\right\|_\infty \\ &\leq \frac{\sqrt{3}}{2\pi}\frac{1}{n+1}\leq\frac{1}{n+1}\leq\frac{\Delta(\boldsymbol{\tau}^\star)}{4}.\end{aligned} \tag{41}$$

- Furthermore, we have that

$$\begin{aligned}\left\|\boldsymbol{a}_k\right\|_\infty &\leq \left\|\boldsymbol{a}^\star\right\|_\infty+\left\|\boldsymbol{a}_k-\boldsymbol{a}^\star\right\|_\infty \\ &\leq \left\|\boldsymbol{a}^\star\right\|_\infty\left(1+\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty\right) \\ &< \left\|\boldsymbol{a}^\star\right\|_\infty\left(1+\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\right)\right\|_\infty\right) \\ &\leq \frac{3}{2}\left\|\boldsymbol{a}^\star\right\|_\infty\leq A.\end{aligned} \tag{42}$$

Hence, the assumptions of Theorem 3 hold, which yields

$$\begin{aligned}\rho_k &= \max_{\boldsymbol{\theta}\in\mathcal{S}_k}\left\|\boldsymbol{SPH}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty \\ &\leq 1-\frac{\left(a^\star_{\min}\right)^2}{4A^2}+\frac{9}{4}\left(4K_\Delta+\frac{1}{2}K_\theta\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2} \\ &\leq 1-\frac{\left(a^\star_{\min}\right)^2}{4A^2}\left(1-9\left(4K_\Delta+\frac{1}{2}K_\theta\right)\frac{A^2\|\boldsymbol{a}^\star\|_\infty}{{a^\star_{\min}}^3}\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}\right) \\ &\leq 1-\frac{\left(a^\star_{\min}\right)^2}{4A^2}\left(1-276.21\frac{A^2\|\boldsymbol{a}^\star\|_\infty}{{a^\star_{\min}}^3}\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}\right) \\ &\leq 1-\frac{\left(a^\star_{\min}\right)^2}{4A^2}\left(1-\eta\right),\end{aligned} \tag{43}$$

where we substituted the definition (18) of η in the last line. It results from the iterative analysis (30) that the next update $\boldsymbol{\theta}_{k+1}$ obeys

$$\begin{aligned}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_{k+1}-\boldsymbol{\theta}^\star\right)\right\|_\infty &\leq \rho_k\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty \\ &\leq \left(1-\frac{\left(a^\star_{\min}\right)^2}{4A^2}\left(1-\eta\right)\right)\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty \\ &\leq \left(1-\frac{\left(a^\star_{\min}\right)^2}{4A^2}\left(1-\eta\right)\right)^{k+1}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\right)\right\|_\infty,\end{aligned} \tag{44}$$

which concludes the proof of Theorem 1.

3.3 Proof of Theorem 2

We proceed with the proof of Theorem 2 analogously to the proof of Theorem 1 presented in Section 3.2. First, we establish the following intermediate theorem that controls the conditioning of the scaled Hessian matrix $\boldsymbol{S}\boldsymbol{P}_k(\boldsymbol{\theta})\boldsymbol{H}\boldsymbol{S}^{-1}$ uniformly over the segment $\mathcal{S}_k$ as a function of the weighted infinity-norm distance $\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty$. The proof of Theorem 4 is deferred to Appendix B.3.

Theorem 4 (Uniform bound of the Hessian)**.** *Suppose that* $n\geq 2$, $(n+1)\Delta(\boldsymbol{\tau}^\star)\geq 4.7$ *and* $\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty \leq \frac{1}{4}\Delta(\boldsymbol{\tau}^\star)$ *then the exists two positive constants*

$$K_\Delta \leq 2.32, \tag{45a}$$
$$K_\theta \leq 75.80, \tag{45b}$$

such that for all $\boldsymbol{\theta}=[\boldsymbol{a}^\top,\boldsymbol{\tau}^\top]^\top\in\mathcal{S}_k$ *satisfying* $\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty<1$ *we have that*

$$\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty \leq \frac{1}{\left(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty\right)^2}-1 + \left(4K_\Delta+K_\theta\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\frac{\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}}{\left(1-\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2}. \tag{46}$$

The rest of the proof follows similarly by induction. For the base case, it is trivial that the initial point $\boldsymbol{\theta}_0$ verifies (26). We now assume that $\boldsymbol{\theta}_k$ satisfies

$$\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty \leq \left(\frac{1}{2}+\gamma\right)^k\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\right)\right\|_\infty \tag{47}$$

for some $k\in\mathbb{N}$. Let's verify the assumptions of Theorem 4.

- First, as the dynamic range $\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\geq 1$, the assumption (24) easily implies that $(n+1)\Delta(\boldsymbol{\tau}^\star)\geq 4.7$.
- By the definition (17) and the induction hypothesis (47), it follows that

$$\begin{aligned}\left\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\right\|_\infty &\leq \frac{1}{\sqrt{-F_N''(0)}}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\right\|_\infty\\ &\leq \frac{1}{\sqrt{-F_N''(0)}}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\right)\right\|_\infty\\ &\leq \frac{\sqrt{3}}{\pi}\left(1-\sqrt{\frac{2}{3}}\right)\frac{1}{n+1}\\ &\leq \frac{1}{n+1}\leq\frac{\Delta(\boldsymbol{\tau}^\star)}{4}.\end{aligned} \tag{48}$$

Hence, the assumptions of Theorem 4 hold. Noticing that the function $f(u):=\frac{1}{(1-u)^2}$ is increasing over $[0,1)$, we have that $\frac{1}{\left(1-\left\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)^2\right\|_\infty\right)}\leq\frac{1}{2}$. Together with Theorem 4, this yields the bound

$$\begin{aligned}\rho_k &= \max_{\boldsymbol{\theta}\in\mathcal{S}_k}\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty\\ &\leq \frac{1}{2}+\left(4K_\Delta+K_\theta\left(1-\sqrt{\frac{2}{3}}\right)\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{2a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}\\ &\leq \frac{1}{2}+11.60\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}\\ &\leq \frac{1}{2}+\gamma<1,\end{aligned} \tag{49}$$

where we substituted the definition (24) of γ in the third inequality. It results from the iterative analysis (30) that the next update $\boldsymbol{\theta}_{k+1}$ satisfies

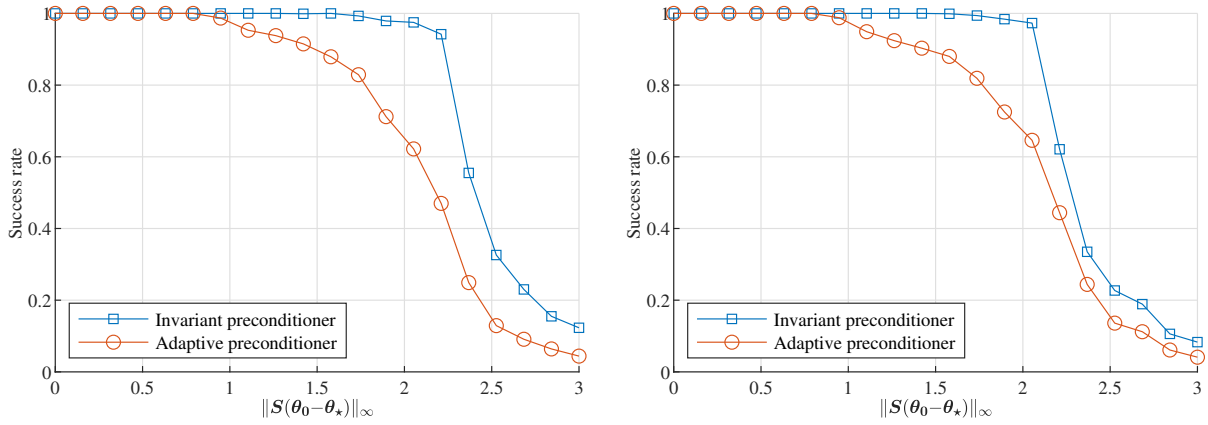
$$\begin{aligned}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_{k+1}-\boldsymbol{\theta}^{\star}\right)\right\|_{\infty} &\leq \rho_k \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_{k}-\boldsymbol{\theta}^{\star}\right)\right\|_{\infty} \\ &\leq \left(\frac{1}{2}+\gamma\right)\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_{k}-\boldsymbol{\theta}^{\star}\right)\right\|_{\infty} \\ &\leq \left(\frac{1}{2}+\gamma\right)^{k+1}\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_{0}-\boldsymbol{\theta}^{\star}\right)\right\|_{\infty}, \end{aligned} \tag{50}$$

which concludes the proof of Theorem 2.

4 Numerical experiments

This section provides a numerical validation of Theorem 1 and Theorem 2. In the following experiments, the signal length is set to $N = 65$ (*i.e.* $n = 32$). The ground truth signal is composed of $r = 6$ sources placed in the interval $[-\frac{1}{2}, \frac{1}{2})$ while ensuring that $(n+1)\Delta(\boldsymbol{\tau}^{\star}) \geq 2$, which is a more optimistic separation condition than what the theorems' statements suggest. Additionally, the dynamic range is denoted by $\kappa = \frac{\|\boldsymbol{a}^{\star}\|_{\infty}}{a^{\star}_{\min}}$. The complex amplitudes $\boldsymbol{a}^{\star} \in \mathbb{C}^r$ are selected independently and uniformly at random in a complex annulus with bounds $1 \leq |a^{\star}_{\ell}| \leq \kappa$. The input parameter of the invariant preconditioning scheme is set at $A = \frac{3}{2}\|\boldsymbol{a}^{\star}\|_{\infty}$.

Size of the basin of attraction Of critical importance in the analysis of Theorem 1 and Theorem 2 is the distance between the initial parameter $\boldsymbol{\theta}_0$ and the ground truth $\boldsymbol{\theta}_{\star}$. We start by comparing the success rates of both preconditioning schemes on reconstructing the ground truth as a function of the initialization distance $\|\boldsymbol{S}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_{\star})\|_{\infty}$. In each experiment, the starting point $\boldsymbol{\theta}_0$ is drawn uniformly over the set of points equidistant to $\boldsymbol{\theta}_{\star}$. An experiment is labeled as a success if $\|\boldsymbol{S}(\boldsymbol{\theta}_{200} - \boldsymbol{\theta}_{\star})\|_{\infty} \leq 10^{-2}$ after 200 iterations. Figure 1 suggests that, for both schemes, the size of the basin of attraction is independent of the dynamic range κ, and is around the order of magnitude $\|\boldsymbol{S}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_{\star})\|_{\infty} \simeq 1$. This suggests that the numerical constants ($1/2$ and $1 - \sqrt{2/3} \sim 0.184$, respectively) set forth in Theorem 1 and Theorem 2, respectively, are pessimistic and nonconvex spike deconvolution performs in a much more benign manner than predicted by our theory, indicating room for further refinements.



(a) $\kappa = 1$ (b) $\kappa = 6$

Figure 1: Success rate of the invariant and adaptive preconditioning schemes on reconstructing the ground truth $\boldsymbol{\theta}_{\star}$ as a function of the initialization distance $\|\boldsymbol{S}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_{\star})\|_{\infty}$ for different dynamic ranges κ. The results are averaged over 1000 randomized trials.

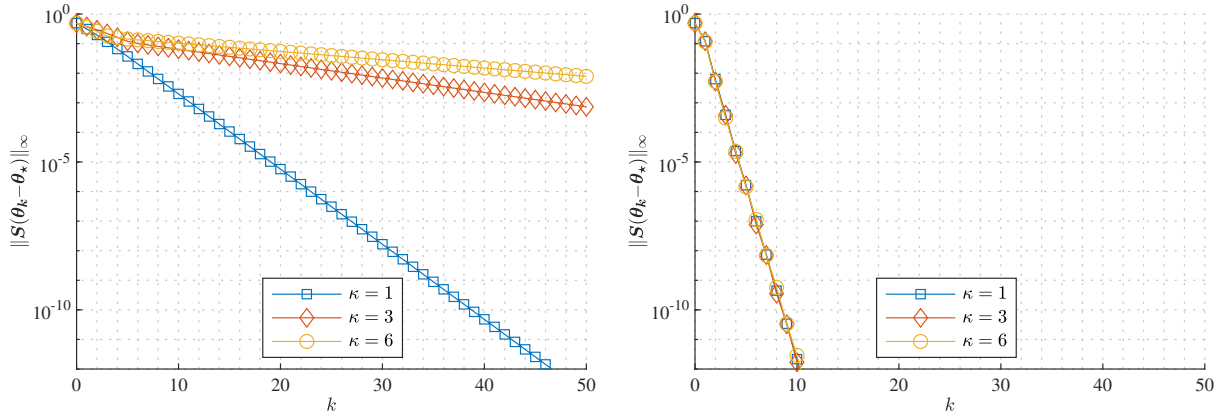
Linear convergence using a spectral initialization In practice, several ad hoc initialization methods could be envisaged to produce an initial point $\boldsymbol{\theta}_0$ that falls in the basin of attraction of the preconditioned

gradient descent methods. Herein, we proceed by uniformly discretizing the spectral domain over N elements. Given the knowledge of the ground truth model order r, the initial locations $\boldsymbol{\tau}_0$ are selected as the r elements of the discrete grid whose weighted Fourier transform best describes the observation $\boldsymbol{x}$. Mathematically, consider the following optimization problem

$$\boldsymbol{u}_0 = \arg\min_{\boldsymbol{u}} \ \frac{1}{2} \left\| \boldsymbol{x} - \operatorname{diag}(\boldsymbol{g}) \boldsymbol{F}_N \boldsymbol{u} \right\|_2^2 \ \text{ s.t. } \ \|\boldsymbol{u}\|_0 \leq r, \tag{51}$$

where $\boldsymbol{F}_N \in \mathbb{C}^{N \times N}$ is a discrete Fourier transform matrix, and the ℓ_0-norm denotes the cardinality of the support. Writing $\mathcal{I}_0 = \operatorname{supp}(\boldsymbol{u}_0) \subset [-n, \dots, n]$ the support of the solution $\boldsymbol{u}_0$ of (51), the parameter $\boldsymbol{\theta}_0$ is constructed in a second stage by selecting $\boldsymbol{\tau}_0 = [\frac{k_1}{N}, \dots, \frac{k_r}{N}]^\top$ where $k_\ell \in \mathcal{I}_0$, $\ell = 1, \dots, r$ and $\boldsymbol{a}_0 = \boldsymbol{u}_{|\mathcal{I}_0}$ as the restriction of $\boldsymbol{u}$ to the elements in $\mathcal{I}_0$. The program (51) is itself a non-convex sparse reconstruction problem, which we approximate the solution using the orthogonal matching pursuit algorithm [48]. The proposed initialization procedure offers several benefits over more classical methods: It is highly scalable, robust to high dynamic range, and does not involve any polynomial root finding subroutine.

Figure 2 pictures the convergence rate of preconditioned GD under the invariant and adaptive preconditioning schemes, respectively. For both schemes, $\boldsymbol{\theta}_0$ is selected according to the previously described initialization procedure. It can be seen that, although both preconditioning schemes ensure a linear converge of the iterate sequence, the convergence rate with a fixed preconditioner degrades as the dynamic range of the sources increases. In contrast, the one with an adaptive preconditioner remains unchanged. Additionally, the adaptive preconditioning scheme benefits from faster convergence rates for a given dynamic range. These experimental results corroborate the theoretical findings presented in Section 2.



(a) Invariant preconditioning

(b) Adaptive preconditioning

Figure 2: Convergence rates of the iterate sequence of preconditioned GD towards the ground truth as the dynamic range κ varies for: (a) the invariant preconditioning scheme; (b) the adaptive preconditioning scheme.

Noisy recovery We next examine the performance of preconditioned GD in the presence of noise. We assume observations of the form $\boldsymbol{x} = \boldsymbol{\Phi}(\mu^\star) + \boldsymbol{w}$, where $\boldsymbol{w}$ is white Gaussian noise, and estimate $\mu^\star$ by minimizing (5) starting from an initial point $\boldsymbol{\theta}_0$ obtained by the spectral initialization procedure described above. Figure 3 draws the statistical error $\|\boldsymbol{S}(\boldsymbol{\theta}_{200} - \boldsymbol{\theta}_\star)\|_\infty$ of both preconditioning schemes after 200 iterations — when convergence is reached — as a function of the signal-to-noise ratio (SNR), defined as $\mathsf{SNR} = \|\boldsymbol{\Phi}(\mu^\star)\|_2^2 / \|\boldsymbol{w}\|_2^2$. The results are benchmarked against the Cramér-Rao bound (CRB) [49]. Both statistical errors remain close to the CRB under a sufficiently large SNR, providing an empirical validation of the robustness of the proposed algorithms.

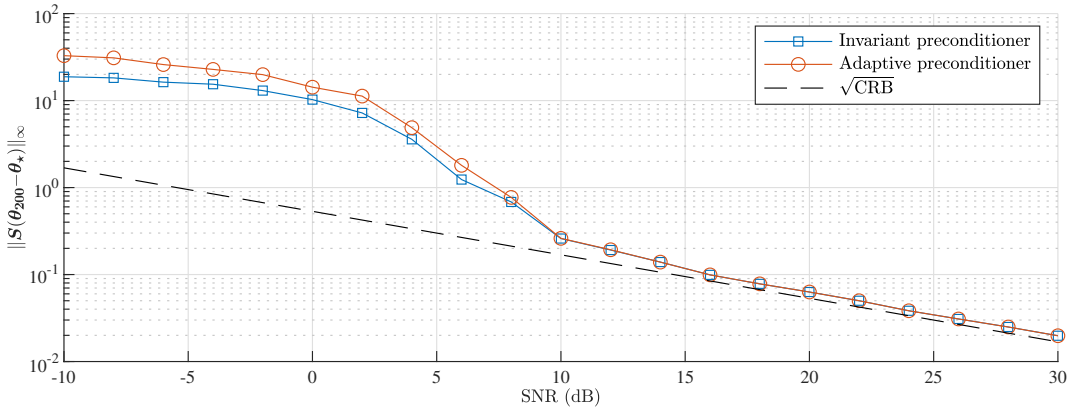


Figure 3: Statistical error $\|\boldsymbol{S}(\boldsymbol{\theta}_{200}-\boldsymbol{\theta}_\star)\|_\infty$ of both preconditioned GD schemes as a function of the SNR. The dynamic range is set to $\kappa = 3$, and the results are averaged over 1000 randomized trials.

5 Conclusion

This work proposed and analyzed preconditioned gradient methods for nonconvex spike deconvolution using both fixed and adaptive preconditioners, and demonstrated that for ground truth with sufficiently separated spikes, the proposed methods achieve a linear rate of convergence that is independent of the number of spikes, as long as a close enough initialization is provided near the ground truth. In particular, by designing the preconditioner to compensate adaptively for the amplitude profile of the spikes, it is possible to accelerate the convergence rate to be dimension-free and independent of the dynamic range, while the convergence using a fixed preconditioner slows down when the dynamic range is large. Our work thus highlights the importance of preconditioning in accelerating convergence in nonconvex spike deconvolution.

As a first step towards understanding the efficacy of first-order methods for spike deconvolution, this works opens up several interesting directions for further investigation.

- *Initialization schemes.* One immediate direction is to analyze initialization schemes that produce initial estimates that fall into the basin of attraction, which we suspect the procedure described in Section 4 is a good candidate.
- *Model order.* For simplicity, it is assumed that the model order r is known perfectly, which might not hold in practice. It is of great interest to develop modified algorithms when the model order is overspecified, which has recently been examined comprehensively in [50] for low-rank estimation from small random initializations.
- *General observations.* Another direction is to extend the analysis to more general observation operators, possibly including random sampling, missing data, as well as corruptions. This may necessarily require a reformulation of the loss function, such as a nonsmooth and nonconvex formulation using the least absolute deviation [47] to improve robustness.
- *Separation condition.* Last but not least, it is of great importance to study to what extent it is possible to relax the success condition in terms of the separation condition, possibly with additional positive constraints of the source amplitudes.

References

[1] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM Journal on Mathematical Analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.

[2] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[3] Z. Zhu, G. Tang, P. Setlur, S. Gogineni, M. B. Wakin, and M. Rangaswamy, "Super-resolution in SAR imaging: Analysis with the atomic norm," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2016, pp. 1–5.

[4] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nature methods*, vol. 9, no. 7, pp. 721–723, 2012.

[5] C. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.

[6] J. Lindberg, "Mathematical concepts of optical superresolution," *Journal of Optics - IOP Publishing*, vol. 14, no. 8, p. 83001, 2012.

[7] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[8] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[9] W. Liao and A. Fannjiang, "MUSIC for single-snapshot spectral estimation: Stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, 2016.

[10] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984 –995, Jul. 1989.

[11] A. Moitra, "Super-resolution, extremal functions and the condition number of Vandermonde matrices," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 821–830.

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

[13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions On Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

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

[15] Y. Chi and M. Ferreira Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 39–57, March 2020.

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

[17] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE transactions on signal processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[18] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[19] E. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406 –5425, Dec. 2006.

[20] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 1289–1306, Feb. 2006.

[21] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster storm using compressed sensing," *Nature methods*, vol. 9, no. 7, pp. 721–723, 2012.

[22] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries-just discretize," in *2013 Asilomar Conference on Signals, Systems and Computers*, Nov. 2013, pp. 1043–1047.

[23] Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity to basis mismatch of compressed sensing for spectrum analysis and beamforming," in *Proc. 6th U.S./Australia Joint Workshop on Defence Applications of Signal Processing (DASP)*, Sep. 2009, event-place: Lihue, HI.

[24] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Information and Inference*, vol. 5, no. 3, pp. 251–303, Sep. 2016.

[25] M. Ferreira Da Costa and W. Dai, "A tight converse to the spectral resolution limit via convex programming," in *2018 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2018, pp. 901–905.

[26] Q. Li and G. Tang, "Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision," *Applied and Computational Harmonic Analysis*, 2018.

[27] M. Ferreira Da Costa and Y. Chi, "On the stable resolution limit of total variation regularization for spike deconvolution," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 7237–7252, 2020.

[28] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5239–5269, 2019.

[29] L. Chizat and F. Bach, "On the global convergence of gradient descent for over-parameterized models using optimal transport," *Advances in neural information processing systems*, vol. 31, 2018.

[30] J. Huang, M. Sun, J. Ma, and Y. Chi, "Super-resolution image reconstruction for high-density three-dimensional single-molecule microscopy," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 763–773, 2017.

[31] Y. Chi, "Guaranteed Blind Sparse Spikes Deconvolution via Lifting and Convex Optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 782–794, 2016.

[32] Y. Traonmilin and J.-F. Aujol, "The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem," *Inverse Problems*, vol. 36, no. 4, p. 045003, 2020.

[33] Y. Traonmilin, J.-F. Aujol, and A. Leclaire, "Projected gradient descent for non-convex sparse spike estimation," *IEEE Signal Processing Letters*, vol. 27, pp. 1110–1114, 2020.

[34] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol, "Fast off-the-grid sparse recovery with over-parametrized projected gradient descent," *arXiv preprint arXiv:2202.13757*, 2022.

[35] E. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 1985–2007, 2015.

[36] Y. Chen and E. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Communications on Pure and Applied Mathematics*, vol. 70, no. 5, pp. 822–883, 2017.

[37] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6535–6579, 2016.

[38] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *Applied and computational harmonic analysis*, vol. 47, no. 3, pp. 893–934, 2019.

[39] Y. Chen, J. Fan, B. Wang, and Y. Yan, "Convex and nonconvex optimization are both minimax-optimal for noisy blind deconvolution under random designs," *Journal of the American Statistical Association*, pp. 1–11, 2021.

[40] K. Lee, Y. Li, M. Junge, and Y. Bresler, "Blind recovery of sparse signals from subsampled convolution," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 802–821, 2016.

[41] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere I: Overview and the geometric picture," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 853–884, 2016.

[42] D. Gilboa, S. Buchanan, and J. Wright, "Efficient dictionary learning with gradient descent," *ICML Workshop on Modern Trends in Nonconvex Optimization for Machine Learning*, 2018.

[43] L. Shi and Y. Chi, "Manifold gradient descent solves multi-channel sparse blind deconvolution provably and efficiently," *IEEE Transactions on Information Theory*, vol. 67, no. 7, pp. 4784–4811, 2021.

[44] Q. Qu, X. Li, and Z. Zhu, "Exact recovery of multichannel sparse blind deconvolution via gradient descent," *SIAM Journal on Imaging Sciences*, vol. 13, no. 3, pp. 1630–1652, 2020.

[45] T. Tong, C. Ma, and Y. Chi, "Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent," *Journal of Machine Learning Research*, vol. 22, pp. 1–63, 2021.

[46] T. Tong, C. Ma, A. Prater-Bennette, E. Tripp, and Y. Chi, "Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements," *Journal of Machine Learning Research*, vol. 23, no. 163, pp. 1–77, 2022.

[47] T. Tong, C. Ma, and Y. Chi, "Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2396–2409, 2021.

[48] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[49] L. L. Scharf and L. T. McWhorter, "Geometry of the Cramér-Rao bound," *Signal Processing*, vol. 31, no. 3, pp. 301–311, 1993.

[50] X. Xu, Y. Shen, Y. Chi, and C. Ma, "The power of preconditioning in overparameterized low-rank matrix sensing," *arXiv preprint arXiv:2302.01186*, 2023.

A Summation bounds of the Fejér kernel

The purpose of this section is to present Lemma 5, which delivers fundamental bounds on the absolute sum of the Fejér kernel and its derivatives at sampled points of interest. Although specific to the Fejér kernel, Lemma 5 could be adapted to any other absolutely integrable point spread function without a significant change in the proof structure.

Lemma 5 (Uniform bounds on the Fejér kernel)**.** *Suppose that $n \geq 2$. Let $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_r\} \subset \mathbb{T}$, and let $\alpha > 0$ be such that $(n+1)\Delta(\boldsymbol{\tau}) \geq \alpha$. Let $\boldsymbol{u} = \{u_{i,j}\}_{i \neq j} \subset \mathbb{R}$ be a set of $\frac{r(r-1)}{2}$ real numbers that are absolutely bounded by β such that*

$$(n+1) \max_{i \neq j} \{|u_{i,j}|\} \triangleq \beta < \frac{\alpha}{2}. \tag{52}$$

Then the inequalities

$$\max_i \sum_{j \neq i} \left| F_N^{(\ell)}(\tau_j - \tau_i + u_{i,j}) \right| \leq C_\ell (n+1)^\ell \left((n+1)\Delta(\boldsymbol{\tau}) \right)^{-2} \tag{53}$$

hold for $\ell = 0, 1, 2, 3$, where the constants C_ℓ only depend on α and β and are given by

$$C_0 = \frac{4}{\pi^2} (\alpha - 2\beta)^{-1} \alpha, \tag{54a}$$

$$C_1 = \left(\frac{4}{\pi} (\alpha - 2\beta)^{-1} + \frac{8}{\pi^2} (\alpha - 2\beta)^{-2} \right) \alpha, \tag{54b}$$

$$C_2 = \left(\frac{80}{9} (\alpha - 2\beta)^{-1} + \frac{16}{\pi} (\alpha - 2\beta)^{-2} + \frac{64}{3\pi^2} (\alpha - 2\beta)^{-3} \right) \alpha, \tag{54c}$$

$$C_3 = \left(\frac{16}{3} (\alpha - 2\beta)^{-1} \pi + \frac{1488}{27} (\alpha - 2\beta)^{-2} + \frac{192}{\pi} (\alpha - 2\beta)^{-3} + \frac{192}{\pi^2} (\alpha - 2\beta)^{-4} \right) \alpha. \tag{54d}$$

Proof. First, the function F_N is a trigonometric polynomial, hence infinitely differentiable. Let us begin by examining the general expression of $F_N(t)$ and its derivatives up to the third order. Assuming $t \notin \mathbb{Z}$, by basic calculation, we have that

$$F_N(t) = \frac{\sin^2(\pi(n+1)t)}{(n+1)^2 \sin^2(\pi t)}, \tag{55a}$$

$$\begin{aligned} F_N'(t) = {} & \frac{2\pi}{n+1} \cos((n+1)\pi t) \sin((n+1)\pi t) \csc^2(\pi t) \\ & - \frac{2\pi}{(n+1)^2} \sin((n+1)\pi t) \cot(\pi t) \csc^2(\pi t), \end{aligned} \tag{55b}$$

$$\begin{aligned} F_N''(t) = {} & 2\pi^2 \left(\cos^2((n+1)\pi t) - \sin^2((n+1)\pi t)\right) \csc^2(\pi t) \\ & - \frac{8\pi^2}{n+1} \cos((n+1)\pi t) \sin((n+1)\pi t) \cot(\pi t) \csc^2(\pi t) \\ & + \frac{2\pi^2}{(n+1)^2} \sin^2(\pi(n+1)t) \left(2\cot^2(\pi t) + 1\right) \csc^2(\pi t) \end{aligned} \tag{55c}$$

$$\begin{aligned} F_N'''(t) = {} & -(n+1)8\pi^3 \cos((n+1)\pi t) \sin((n+1)\pi t) \csc^2(\pi t) \\ & - 12\pi^3 \left(\cos^2((n+1)\pi t) - \sin^2((n+1)\pi t)\right) \cot(\pi t) \csc^2(\pi t) \\ & + \frac{12\pi^3}{n+1} \cos((n+1)\pi t) \sin((n+1)\pi t) \left(3\cot^2(\pi t) + 1\right) \csc^2(\pi t) \\ & - \frac{8\pi^3}{(n+1)^2} \sin^2((n+1)\pi t) \left(3\cot^3(\pi t) + 2\cot(\pi t)\right) \csc^2(\pi t). \end{aligned} \tag{55d}$$

Using the four trigonometric bounds $|\sin(a)| \le 1$, $|\cos(a)| \le 1$, $|\cos(a)\sin(a)| \le \frac{1}{2}$, and $\left|\cos^2(a) - \sin^2(a)\right| \le 1$ for $a \in \mathbb{R}$, and the triangle inequality, (55) can be further absolutely bounded as

$$|F_N(t)| \le \frac{1}{(n+1)^2} \csc^2(\pi t), \tag{56a}$$

$$|F_N'(t)| \le \frac{\pi}{n+1} \csc^2(\pi t) + \frac{2\pi}{(n+1)^2} |\cot(\pi t)| \csc^2(\pi t), \tag{56b}$$

$$|F_N''(t)| \le 2\pi^2 \csc^2(\pi t) + \frac{4\pi^2}{n+1} |\cot(\pi t)| \csc^2(\pi t) + \frac{2\pi^2}{(n+1)^2} \left(2\cot^2(\pi t) + 1\right) \csc^2(\pi t) \tag{56c}$$

$$\begin{aligned} |F_N'''(t)| \le {} & (n+1)4\pi^3 \csc^2(\pi t) + 12\pi^3 |\cot(t)| \csc^2(t) + \frac{12\pi^3}{n+1} \left(3\cot^2(\pi t) + 1\right) \csc^2(\pi t) \\ & + \frac{8\pi^3}{(n+1)^2} \left(3\left|\cot^3(\pi t)\right| + 2|\cot(\pi t)|\right) \csc^2(\pi t). \end{aligned} \tag{56d}$$

To continue, observe that a basic building block in the bound (56) takes the following function form, which is denoted by

$$h_\ell(t) = \cot^\ell(\pi t) \csc^2(\pi t), \qquad \ell = 0, 1, 2. \tag{57}$$

The functions h_ℓ, $\ell = 0, 1, 2$ are even, 1-periodic, and continuous, non-negative, decreasing and convex over the interval $(0, \frac{1}{2}]$. Define the three sums $S_\ell(n, \boldsymbol{\tau}, \boldsymbol{u}, i)$, $\ell = 0, 1, 2$ by

$$S_\ell(n, \boldsymbol{\tau}, \boldsymbol{u}, i) = \sum_{j \ne i} h_\ell\left(\tau_j - \tau_i + u_{i,j}\right). \tag{58}$$

It is straightforward to see that the terms of interest can be bounded in terms of $S_\ell(n, \boldsymbol{\tau}, \boldsymbol{u}, i)$ in view of (56). We shall claim the following bound of $S_\ell(n, \boldsymbol{\tau}, \boldsymbol{u}, i)$ holds, which will be proven at the end of proof:

$$S_\ell(n, \boldsymbol{\tau}, \boldsymbol{u}, i) \le \left(\frac{2}{\pi\Delta(\boldsymbol{\tau})}\right)^{\ell+2} \frac{1}{(\ell+1)\left(1 - 2\alpha^{-1}\beta\right)^{\ell+1}}, \qquad \ell = 0, 1, 2. \tag{59}$$

Interestingly, the above bound on $S_\ell\left(n, \boldsymbol{\tau}, \boldsymbol{u}, i\right)$ does not depend on n or i. We are now ready to establish the bound (53) of interest. First, we have

$$\begin{aligned}
\max_i \sum_{j\neq i} |F_N(\tau_j - \tau_i + u_{i,j})| &\leq \sum_{j\neq i} \frac{1}{(n+1)^2} h_0(\tau_j - \tau_i + u_{i,j}) \\
&= \frac{1}{(n+1)^2} S_0(n, \boldsymbol{\tau}, \boldsymbol{u}, i) \\
&\leq \frac{4}{\pi^2} \frac{\Delta(\boldsymbol{\tau})^{-2}}{(n+1)^2 \left(1 - 2\alpha^{-1}\beta\right)} \\
&= \frac{4}{\pi^2} \left(\alpha - 2\beta\right)^{-1} \alpha \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}. \qquad (60)
\end{aligned}$$

Similarly, for the case of the first derivative, we have

$$\begin{aligned}
\max_i \sum_{j\neq i} |F_N'(\tau_j - \tau_i + u_{i,j})| &\leq \sum_{j\neq i} \left(\frac{\pi}{n+1} h_0(\tau_j - \tau_i + u_{i,j}) + \frac{2\pi}{(n+1)^2} h_1(\tau_j - \tau_i + u_{i,j}) \right) \\
&= \frac{\pi}{n+1} S_0(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{2\pi}{(n+1)^2} S_1(n, \boldsymbol{\tau}, \boldsymbol{u}, i) \\
&\leq \frac{4\Delta(\boldsymbol{\tau})^{-2}}{\pi(n+1)\left(1 - 2\alpha^{-1}\beta\right)} + \frac{8\Delta(\boldsymbol{\tau})^{-3}}{\pi^2 (n+1)^2 \left(1 - 2\alpha^{-1}\beta\right)^2} \\
&\leq \left(\frac{4}{\pi} \left(\alpha - 2\beta\right)^{-1} + \frac{8}{\pi^2} \left(\alpha - 2\beta\right)^{-2} \right) \alpha(n+1) \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}, \qquad (61)
\end{aligned}$$

where the last line uses $(n+1)\Delta(\boldsymbol{\tau}) \geq \alpha$. Moving onto the second derivative, it follows

$$\begin{aligned}
&\max_i \sum_{j\neq i} |F_N''(\tau_j - \tau_i + u_{i,j})| \\
&\leq \sum_{j\neq i} \left(2\pi^2 h_0(\tau_j - \tau_i + u_{i,j}) + \frac{4\pi^2}{n+1} h_1(\tau_j - \tau_i + u_{i,j}) + \frac{2\pi^2}{(n+1)^2} \left(2h_2(\tau_j - \tau_i + u_{i,j}) + h_0(\tau_j - \tau_i + u_{i,j})\right) \right) \\
&= 2\pi^2 S_0(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{4\pi^2}{n+1} S_1(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{2\pi^2}{(n+1)^2} \left(2S_2(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + S_0(n, \boldsymbol{\tau}, \boldsymbol{u}, i)\right) \\
&\leq 2\pi^2 \left(1 + \frac{1}{(n+1)^2}\right) S_0(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{4\pi^2}{n+1} S_1(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{4\pi^2}{(n+1)^2} S_2(n, \boldsymbol{\tau}, \boldsymbol{u}, i) \\
&\leq \frac{20\pi^2}{9} S_0(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{4\pi^2}{n+1} S_1(n, \boldsymbol{\tau}, \boldsymbol{u}, i) + \frac{4\pi^2}{(n+1)^2} S_2(n, \boldsymbol{\tau}, \boldsymbol{u}, i) \\
&\leq \frac{80\Delta(\boldsymbol{\tau})^{-2}}{9\left(1 - 2\alpha^{-1}\beta\right)} + \frac{16\Delta(\boldsymbol{\tau})^{-3}}{\pi(n+1)\left(1 - 2\alpha^{-1}\beta\right)^2} + \frac{64\Delta(\boldsymbol{\tau})^{-4}}{3\pi^2 (n+1)^2 \left(1 - 2\alpha^{-1}\beta\right)^3} \\
&\leq \left(\frac{80}{9} \left(\alpha - 2\beta\right)^{-1} + \frac{16}{\pi} \left(\alpha - 2\beta\right)^{-2} + \frac{64}{3\pi^2} \left(\alpha - 2\beta\right)^{-3} \right) \alpha(n+1)^2 \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}. \qquad (62)
\end{aligned}$$

Finally, for the third derivative, it holds

$$\begin{aligned}
&\max_i \sum_{j\neq i} |F_N'''(\tau_j - \tau_i + u_{i,j})| \\
&\leq \sum_{j\neq i} \Big((n+1)4\pi^3 h_0(\tau_j - \tau_i + u_{i,j}) + 12\pi^3 h_1(\tau_j - \tau_i + u_{i,j}) \\
&\quad + \frac{12\pi^3}{n+1} \left(3h_2(\tau_j - \tau_i + u_{i,j}) + h_0(\tau_j - \tau_i + u_{i,j})\right) + \frac{8\pi^3}{(n+1)^2} \left(3h_3(\tau_j - \tau_i + u_{i,j}) + 2h_1(\tau_j - \tau_i + u_{i,j})\right) \Big)
\end{aligned}$$

$$
\begin{aligned}
&= (n+1)4\pi^3 S_0(n,\boldsymbol{\tau},\boldsymbol{u},i) + 12\pi^3 S_1(n,\boldsymbol{\tau},\boldsymbol{u},i) + \frac{12\pi^3}{n+1}\left(3S_2(n,\boldsymbol{\tau},\boldsymbol{u},i) + S_0(n,\boldsymbol{\tau},\boldsymbol{u},i)\right) \\
&\quad + \frac{8\pi^3}{(n+1)^2}\left(3S_3(n,\boldsymbol{\tau},\boldsymbol{u},i) + 2S_1(n,\boldsymbol{\tau},\boldsymbol{u},i)\right) \\
&\leq (n+1)4\pi^3\left(1+\frac{3}{(n+1)^2}\right) S_0(n,\boldsymbol{\tau},\boldsymbol{u},i) + 12\pi^3\left(1+\frac{4}{3\,(n+1)^2}\right) S_1(n,\boldsymbol{\tau},\boldsymbol{u},i) \\
&\quad + \frac{36\pi^3}{n+1} S_2(n,\boldsymbol{\tau},\boldsymbol{u},i) + \frac{24\pi^3}{(n+1)^2} S_3(n,\boldsymbol{\tau},\boldsymbol{u},i) \\
&\leq (n+1)\frac{64}{3}\pi\frac{\Delta(\boldsymbol{\tau})^{-2}}{1-2\alpha^{-1}\beta} + \frac{2488}{27}\frac{\Delta(\boldsymbol{\tau})^{-3}}{(1-2\alpha^{-1}\beta)^2} + \frac{192}{\pi(n+1)}\frac{\Delta(\boldsymbol{\tau})^{-4}}{(1-2\alpha^{-1}\beta)^3} + \frac{192}{\pi^2(n+1)^2}\frac{\Delta(\boldsymbol{\tau})^{-5}}{(1-2\alpha^{-1}\beta)^4} \\
&\leq \left(\frac{16}{3}(\alpha-2\beta)^{-1}\pi + \frac{1488}{27}(\alpha-2\beta)^{-2} + \frac{192}{\pi}(\alpha-2\beta)^{-3} + \frac{192}{\pi^2}(\alpha-2\beta)^{-4}\right)\alpha(n+1)^3\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}.
\end{aligned} \tag{63}
$$

The proof is thus completed if we can prove (59), which is the focus of the rest of the proof.

Proof of (59) We fix the index i and take the convention $u_{i,i} = 0$. As the functions $\{h_\ell\}$ are 1-periodic, the quantity $S_\ell(n,\boldsymbol{\tau},\beta,i)$ is invariant by integer translations of $\{\tau_j\}$'s. Therefore, one can make the assumption, up to a modulo considerations and a reordering of the indices that the sequence $\{\tau_j - \tau_i + u_{i,j}\}_j$ is within the range $[-\frac{1}{2},\frac{1}{2})$ and in an ascending order, so that

$$
-\frac{1}{2} \leq \tau_1 - \tau_i + u_{i,1} < \tau_2 - \tau_i + u_{i,2} < \cdots < \tau_r - \tau_i + u_{i,r} < \frac{1}{2}.
$$

Denote by r_+ and r_- the number of positive and negative elements in the set $\{\tau_j - \tau_i + u_{i,j}\}_j$, respectively. As $\tau_j - \tau_i + u_{i,j} = 0$ if and only if $i = j$, we have that $r_+ + r_- = r - 1$. Using the separation condition, and as h_ℓ is decreasing over $(0,\frac{1}{2}]$ and even, we have that

$$
\begin{aligned}
0 \leq h_\ell\left(\tau_{i+j} - \tau_i + u_{i,j}\right) &\leq h_\ell\left(j\Delta(\boldsymbol{\tau}) - \frac{\beta}{n+1}\right), \quad j = 1,\dots,r_+; \\
0 \leq h_\ell\left(\tau_{i-j} - \tau_i + u_{i,j}\right) \leq h_\ell\left(-j\Delta(\boldsymbol{\tau}) + \frac{\beta}{n+1}\right) &= h_\ell\left(j\Delta(\boldsymbol{\tau}) - \frac{\beta}{n+1}\right), \quad j = 1,\dots,r_-.
\end{aligned}
$$

We can subsequently bound the sum (58) as

$$
\begin{aligned}
S_\ell(n,\boldsymbol{\tau},\boldsymbol{u},i) &= \sum_{j=1}^{r_+} h_\ell\left(\tau_{i+j} - \tau_i + u_{i,(i+j)}\right) + \sum_{j=1}^{r_-} h_\ell\left(\tau_{i-j} - \tau_i + u_{i,(i-j)}\right) \\
&\leq \sum_{j=1}^{r_+} h_\ell\left(j\Delta(\boldsymbol{\tau}) - \frac{\beta}{n+1}\right) + \sum_{j=1}^{r_-} h_\ell\left(j\Delta(\boldsymbol{\tau}) - \frac{\beta}{n+1}\right) \\
&\leq 2\sum_{j=1}^{\lceil\frac{r-1}{2}\rceil} h_\ell\left(j\Delta(\boldsymbol{\tau}) - \frac{\beta}{n+1}\right).
\end{aligned} \tag{64}
$$

Identifying and associating the right-hand side of (64) to a Riemann sum with a mid-point rule and recalling that h_ℓ is decreasing and convex over $(0,\frac{1}{2}]$ leads to the majorant

$$
\begin{aligned}
S_\ell(n,\boldsymbol{\tau},\boldsymbol{u},i) &\leq 2\int_{\frac{\Delta(\boldsymbol{\tau})}{2}-\frac{\beta}{n+1}}^{(\lceil\frac{r-1}{2}\rceil+\frac{1}{2})\Delta(\boldsymbol{\tau})-\frac{\beta}{n+1}} h_\ell(u)\mathrm{d}u \\
&\leq 2\int_{\frac{\Delta(\boldsymbol{\tau})}{2}-\frac{\beta}{n+1}}^{\frac{1}{2}} h_\ell(u)\mathrm{d}u
\end{aligned}
$$

$$= 2\left(H_\ell\left(\frac{1}{2}\right) - H_\ell\left(\frac{\Delta(\boldsymbol{\tau})}{2} - \frac{\beta}{n+1}\right)\right), \tag{65}$$

where we used the non-negativity of h_ℓ and the inequality $\left(\left\lceil \frac{r-1}{2} \right\rceil + \frac{1}{2}\right)\Delta(\boldsymbol{\tau}) - \frac{\beta}{n+1} \leq \frac{1}{2}$ in the second inequality. Here, $\{H_\ell\}$'s are primitives of the functions $\{h_\ell\}$ over the interval $(0, \frac{1}{2}]$. Moreover, we have

$$H_\ell(t) = -\frac{1}{\pi\,(\ell+1)}\cot^{\ell+1}(\pi t) + c \tag{66}$$

for $\ell = 0, 1, 2$, where c is an arbitrary constant. This yields, with the inequality $0 \leq \cot(\pi u) \leq \frac{1}{\pi u}$ for all $0 < u < \frac{1}{2}$, a further simplification of (65):

$$\begin{aligned}
\Delta(\boldsymbol{\tau}) S_\ell(n, \boldsymbol{\tau}, \boldsymbol{u}, i) &\leq \frac{2}{\pi(\ell+1)}\cot^{\ell+1}\left(\pi\left(\frac{\Delta(\boldsymbol{\tau})}{2} - \frac{\beta}{n+1}\right)\right) \\
&\leq \left(\frac{2}{\pi}\right)^{\ell+2} \Delta(\boldsymbol{\tau})^{-\ell-1} \frac{1}{\left(1 - \frac{2\beta\Delta(\boldsymbol{\tau})^{-1}}{n+1}\right)^{\ell+1}} \\
&\leq \left(\frac{2}{\pi}\right)^{\ell+2} \Delta(\boldsymbol{\tau})^{-\ell-1} \frac{1}{(1 - 2\alpha^{-1}\beta)^{\ell+1}},
\end{aligned} \tag{67}$$

which immediately leads to the claimed bound (59). □

B Proof of the uniform Hessian bounds

This section is dedicated to establish Theorem 3 and Theorem 4.

B.1 Technical lemmas

Lemma 6 and Lemma 7 provide bounds on core quantities that are involved in the decomposition of the Hessian matrix $\boldsymbol{H}(\boldsymbol{\theta})$ characterized in Section 3.1.2. Their proofs are presented in Appendix B.4 and Appendix B.5, respectively.

Lemma 6. *Suppose that $n \geq 2$ and let $\boldsymbol{\tau} \subset \mathbb{T}$ be such that $(n+1)\Delta(\boldsymbol{\tau}) \geq \alpha$ for some $\alpha > 0$, then there exists a constant K_Δ with*

$$K_\Delta := \max\left\{C_0 + \frac{3\sqrt{3}}{4\pi}C_1, \frac{3\sqrt{3}}{4\pi}C_1 + \frac{27}{16\pi^2}C_2\right\} \tag{68}$$

where the constants $C_0, C_1, C_2 > 0$ are defined in in (54) with parameters α and $\beta = 0$ such that

$$\|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\|_\infty \leq K_\Delta((n+1)\Delta(\boldsymbol{\tau}))^{-2}. \tag{69}$$

Lemma 7. *Suppose that $n \geq 2$ and let $\boldsymbol{\tau}_\star = [\tau_1^\star, \ldots, \tau_r^\star]^\top$ and $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_r]^\top$ be two vectors of points around the torus. Assume that $(n+1)\Delta(\boldsymbol{\tau}^\star) \geq \alpha$. As long as $(n+1)\left\|\boldsymbol{\tau} - \boldsymbol{\tau}^\star\right\|_\infty \leq \beta < \frac{\alpha}{2}$, we have*

$$\left|\left\langle \Phi(\delta'_{\tau_j}), \Phi\left(\mu(\boldsymbol{\theta})\right) - \mu(\boldsymbol{\theta}_\star)\right\rangle\right| \leq \left(C_1\left\|\boldsymbol{a} - \boldsymbol{a}^\star\right\|_\infty + C_2\left\|\boldsymbol{a}^\star\right\|_\infty (n+1)\left\|\boldsymbol{\tau} - \boldsymbol{\tau}^\star\right\|_\infty\right)(n+1)\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}, \tag{70a}$$

$$\left|\left\langle \Phi(\delta''_{\tau_j}), \Phi\left(\mu(\boldsymbol{\theta})\right) - \mu(\boldsymbol{\theta}_\star)\right\rangle\right| \leq \left(C_2\left\|\boldsymbol{a} - \boldsymbol{a}^\star\right\|_\infty + C_3\left\|\boldsymbol{a}^\star\right\|_\infty (n+1)\left\|\boldsymbol{\tau} - \boldsymbol{\tau}^\star\right\|_\infty\right)(n+1)^2\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2} \tag{70b}$$

for all $j = 1, \ldots, r$, where the constants $C_1, C_2, C_3 > 0$ are defined in (54) with parameters (α, β).

B.2 Proof of Theorem 3

Recalling the expression of the Hessian in (32), it follows that

$$\boldsymbol{SPH}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I} = \boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I} + \boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}. \tag{71}$$

We proceed to bound $\left\|\boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty$ and $\left\|\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty$ separately, and then combine them via the triangle inequality.

Step 1: bound $\left\|\boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty$ From (16) and (33), we have that

$$\begin{aligned}
\boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I} &= \boldsymbol{SP}\,\mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ \sqrt{-F_N''(0)}\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}} \boldsymbol{D}(\boldsymbol{\tau})\,\mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ \sqrt{-F_N''(0)}\boldsymbol{a}\end{bmatrix}\right)\boldsymbol{S}^{-1} - \boldsymbol{I} \\
&= \mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ A^{-2}\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}} \boldsymbol{D}(\boldsymbol{\tau})\,\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star} \\ \boldsymbol{a}\end{bmatrix}\right) - \boldsymbol{I} \\
&= \mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ A^{-2}\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}} (\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I})\,\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star} \\ \boldsymbol{a}\end{bmatrix}\right) + \mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ A^{-2}\left|\boldsymbol{a}\right|^2\end{bmatrix}\right) - \boldsymbol{I}.
\end{aligned} \tag{72}$$

This immediately yields from the triangle inequality that

$$\begin{aligned}
&\left\|\boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty \\
&\quad\le \left\|\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ A^{-2}\odot\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}} (\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I})\,\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star} \\ \boldsymbol{a}\end{bmatrix}\right)\right\|_\infty + \left\|\mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ A^{-2}\odot\left|\boldsymbol{a}\right|^2\end{bmatrix}\right) - \boldsymbol{I}\right\|_\infty \\
&\quad\le \left\|\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ A^{-2}\odot\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}}\right\|_\infty \left\|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty \left\|\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star} \\ \boldsymbol{a}\end{bmatrix}\right)\right\|_\infty + \left\|\mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ A^{-2}\odot\left|\boldsymbol{a}\right|^2\end{bmatrix}\right) - \boldsymbol{I}\right\|_\infty \\
&\quad\le \max_j\left\{\frac{1}{\left|a_j^\star\right|}, \frac{\left|a_j\right|}{A^2}\right\} \max_j\left\{\left|a_j^\star\right|, \left|a_j\right|\right\} \left\|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty + \max_j\left\{\left|1 - \frac{\left|a_j\right|^2}{A^2}\right|\right\}.
\end{aligned} \tag{73}$$

The three maxima in (73) can be controlled using the basic relation

$$1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty \le \frac{\left|a_j^\star\right| - \left|a_{k,j} - a_j^\star\right|}{\left|a_j^\star\right|} \le \frac{\left|a_j\right|}{\left|a_j^\star\right|} \le \frac{\left|a_j^\star\right| + \left|a_{k,j} - a_j^\star\right|}{\left|a_j^\star\right|} \le 1 + \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty, \tag{74}$$

for $j = 1, \ldots, r$ and by exploiting the assumption $\|\boldsymbol{a}^\star\|_\infty \le A$ as follows

$$\max_j\left\{\frac{1}{\left|a_j^\star\right|}, \frac{\left|a_j\right|}{A^2}\right\} \le \max_j\left\{\frac{1}{\left|a_j^\star\right|}, \frac{\left|a_j^\star\right|}{A^2}\left(1 + \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)\right\} \le \frac{1}{a_{\min}^\star}\left(1 + \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right), \tag{75a}$$

$$\max_j\left\{\left|a_j^\star\right|, \left|a_j\right|\right\} \le \max_j\left\{\left|a_j^\star\right|, \left|a_j^\star\right|\left(1 + \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)\right\} \le \left\|\boldsymbol{a}^\star\right\|_\infty\left(1 + \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right), \tag{75b}$$

$$\max_j\left\{\left|1 - \frac{\left|a_j\right|^2}{A^2}\right|\right\} \le 1 - \min_j\left\{\frac{\left|a_j\right|^2}{A^2}\right\} \le 1 - \frac{\left(a_{\min}^\star\right)^2}{A^2}\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right). \tag{75c}$$

The bounds (75) and Lemma 6 imply with (73) that

$$\begin{aligned}
&\left\|\boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty \\
&\quad\le \frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a_{\min}^\star}\left(1 + \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2 \left\|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty + 1 - \frac{\left(a_{\min}^\star\right)^2\left(1 - \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2}{A^2} \\
&\quad\le \frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a_{\min}^\star}\left(1 + \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2 K_\Delta\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2} + 1 - \frac{\left(a_{\min}^\star\right)^2\left(1 - \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2}{A^2},
\end{aligned} \tag{76}$$

where the constant K_Δ defined in Lemma 6 is numerically evaluated to $K_\Delta \le 2.32$ by setting the parameter $\alpha = 16.5$.

Step 2: bound $\left\|\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty$ Using the block diagonal structure of the matrix $\boldsymbol{E}(\boldsymbol{\theta})$ defined in (36), we have that

$$\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1} = \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{-F_N''(0)}}\operatorname{diag}\left(\boldsymbol{a}^{\star -1}\right)\boldsymbol{E}_1(\boldsymbol{\theta}) \\ \frac{1}{\sqrt{-F_N''(0)}}\operatorname{diag}\left(A^{-2}\boldsymbol{a}^\star\right)^{\mathsf{H}}\boldsymbol{E}_1(\boldsymbol{\theta}) & -\frac{1}{F_N''(0)}A^{-2}\boldsymbol{E}_2(\boldsymbol{\theta}) \end{bmatrix}. \tag{77}$$

Therefore, it follows

$$\begin{aligned}\left\|\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty &\leq \frac{1}{\sqrt{-F_N''(0)}}\max_j\left\{\max\left\{\frac{1}{|a_j^\star|},\frac{|a_j^\star|}{A^2}\right\}\left|\left\langle\delta'_{\tau_j},F_N*(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}^\star))\right\rangle\right|\right\} \\ &\quad -\frac{1}{F_N''(0)}\max_j\left\{\frac{|a_j|}{A^2}\left|\left\langle\delta''_{\tau_j},F_N*(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}^\star))\right\rangle\right|\right\} \\ &\leq (a_{\min}^\star)^{-1}\left(1+\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty\right)\left(\frac{1}{\sqrt{-F_N''(0)}}\max_j\left\{\left|\left\langle\delta'_{\tau_j},F_N*(\mu(\boldsymbol{\theta})-\mu\left(\boldsymbol{\theta}^\star\right))\right\rangle\right|\right\}\right. \\ &\quad \left.-\frac{1}{F_N''(0)}\max_j\left\{\left|\left\langle\delta''_{\tau_j},F_N*(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}^\star))\right\rangle\right|\right\}\right),\end{aligned} \tag{78}$$

where we used the bounds (75) in the second line. From (78), it can be seen that bounding the quantity of interest amounts to controlling $\left|\left\langle\delta'_{\tau_j},F_N*(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}^\star))\right\rangle\right|$ and $\left|\left\langle\delta''_{\tau_j},F_N*(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}^\star))\right\rangle\right|$, which can be achieved by applying Lemma 7. Substituting the expression provided by Lemma 7 into (78) leads to

$$\begin{aligned}\left\|\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty &\leq \left(\left(C_1\frac{(n+1)}{\sqrt{-F_N''(0)}}+C_2\frac{(n+1)^2}{-F_N''(0)}\right)\frac{\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty}{\|\boldsymbol{a}^\star\|_\infty}\right. \\ &\quad \left.+\left(C_2\frac{(n+1)^2}{-F_N''(0)}+C_3\frac{(n+1)^3}{(-F_N''(0))^{3/2}}\right)\sqrt{-F_N''(0)}\,\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\|_\infty\right) \\ &\quad \cdot\frac{\|\boldsymbol{a}^\star\|_\infty}{a_{\min}^\star}\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}\left(1+\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty\right).\end{aligned} \tag{79}$$

Evaluating the constants $C_1\leq 2.75$, $C_2\leq 19.08$ and $C_3\leq 48.74$ defined in (54) with parameters $\alpha=16.5$ and $\beta=\frac{\alpha}{4}=4.125$, altogether with the inequality $\frac{(n+1)^2}{-F_N''(0)}\leq\frac{27}{16\pi^2}$ under the assumption $n\geq 2$ yields

$$\begin{aligned}\left\|\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty &\leq \left(K_a\frac{\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty}{\|\boldsymbol{a}^\star\|_\infty}+K_\tau\sqrt{-F_N''(0)}\,\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\|_\infty\right) \\ &\quad \cdot\frac{\|\boldsymbol{a}^\star\|_\infty}{a_{\min}^\star}\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}\left(1+\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty\right),\end{aligned} \tag{80}$$

where the constants K_a and K_τ are given by

$$K_a = C_1\sqrt{\frac{27}{16\pi^2}}+C_2\frac{27}{16\pi^2}\leq 4.40, \tag{81a}$$

$$K_\tau = C_2\frac{27}{16\pi^2}+C_3\left(\frac{27}{16\pi^2}\right)^{3/2}\leq 6.71. \tag{81b}$$

Step 3: combine the bounds The scaled Hessian matrix $\boldsymbol{SPH}(\boldsymbol{\theta})\boldsymbol{S}^{-1}$ can be controlled over $\mathcal{S}_k$ by the triangle inequality as follows

$$\begin{aligned}\left\|\boldsymbol{SPH}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty &\leq \left\|\boldsymbol{SP}\left(\boldsymbol{G}(\boldsymbol{\theta})+\boldsymbol{E}(\boldsymbol{\theta})\right)\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty \\ &\leq \left\|\boldsymbol{SPG}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty+\left\|\boldsymbol{SPE}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty.\end{aligned} \tag{82}$$

Furthermore, we note that the inequality $|\Delta(\boldsymbol{\tau}) - \Delta(\boldsymbol{\tau}^\star)| \leq 2\left\|\boldsymbol{\tau} - \boldsymbol{\tau}^\star\right\|_\infty$ holds for every $\boldsymbol{\tau}, \boldsymbol{\tau}^\star \subset \mathbb{R}$. This implies, with the assumption $\left\|\boldsymbol{\tau} - \boldsymbol{\tau}^\star\right\|_\infty \leq \frac{1}{4}\Delta(\boldsymbol{\tau}^\star)$, that $\Delta(\boldsymbol{\tau}) \geq \frac{1}{2}\Delta(\boldsymbol{\tau}^\star)$. Substituting the bounds given in (76) and (80) yields

$$
\begin{aligned}
&\left\|\boldsymbol{S}\boldsymbol{P}\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty \\
&\leq 1 - \frac{{a^\star_{\min}}^2}{A^2}\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)^2 \\
&\quad + \left(K_\Delta + K_a\frac{\left\|\boldsymbol{a} - \boldsymbol{a}^\star\right\|_\infty}{\left\|\boldsymbol{a}^\star\right\|_\infty} + K_\tau\sqrt{-F_N''(0)}\left\|\boldsymbol{\tau}_k - \boldsymbol{\tau}^\star\right\|_\infty\right)\frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}\left(1 + \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2 \\
&\leq 1 - \frac{{a^\star_{\min}}^2}{A^2}\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)^2 \\
&\quad + \left(K_\Delta + (K_a + K_\tau)\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)\frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}\left(1 + \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2 \\
&\leq 1 - \frac{{a^\star_{\min}}^2}{A^2}\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)^2 \\
&\quad + \left(K_\Delta + (K_a + K_\tau)\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)\frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a^\star_{\min}}4\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}\left(1 + \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2 \\
&\leq 1 - \frac{{a^\star_{\min}}^2}{A^2}\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)^2 \\
&\quad + \left(4K_\Delta + K_\theta\left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)\frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a^\star_{\min}}\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}\left(1 + \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2,
\end{aligned}
\tag{83}
$$

where we defined in the last line the constant K_θ as

$$
K_\theta = 4\left(K_a + K_\tau\right) \leq 44.42. \tag{84}
$$

This concludes the proof of the theorem. □

B.3 Proof of Theorem 4

We proceed analogously to the proof of Theorem 3 presented in Appendix B.2. We start from the expansion (32) of the Hessian matrix to get

$$
\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I} = \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I} + \boldsymbol{S}\boldsymbol{P}_k\boldsymbol{E}(\boldsymbol{\theta})\boldsymbol{S}^{-1}, \tag{85}
$$

and proceed to bound $\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty$ and $\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{E}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty$ individually before recombining them via the triangle inequality.

Step 1: bound $\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty$ From (16) and (33), we have that

$$
\begin{aligned}
\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I} &= \boldsymbol{S}\boldsymbol{P}_k\,\mathrm{diag}\left(\begin{bmatrix}\frac{\mathbf{1}_r}{\sqrt{-F_N''(0)}\boldsymbol{a}}\end{bmatrix}\right)^{\mathsf{H}}\boldsymbol{D}(\boldsymbol{\tau})\,\mathrm{diag}\left(\begin{bmatrix}\frac{\mathbf{1}_r}{\sqrt{-F_N''(0)}\boldsymbol{a}^{\mathsf{H}}}\end{bmatrix}\right)\boldsymbol{S}^{-1} - \boldsymbol{I} \\
&= \mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ |\boldsymbol{a}_k|^{-2}\odot\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}}\boldsymbol{D}(\boldsymbol{\tau})\,\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^\star \\ \boldsymbol{a}\end{bmatrix}\right) - \boldsymbol{I} \\
&= \mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ |\boldsymbol{a}_k|^{-2}\odot\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}}(\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I})\,\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^\star \\ \boldsymbol{a}\end{bmatrix}\right) + \mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ |\boldsymbol{a}_k|^{-2}\odot|\boldsymbol{a}|^2\end{bmatrix}\right) - \boldsymbol{I}.
\end{aligned}
\tag{86}
$$

This immediately yields from the triangle inequality

$$
\begin{aligned}
&\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{S}^{-1} - \boldsymbol{I}\right\|_\infty \\
&\quad\leq \left\|\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^{\star -1} \\ |\boldsymbol{a}_k|^{-2}\odot\boldsymbol{a}\end{bmatrix}\right)^{\mathsf{H}}(\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I})\,\mathrm{diag}\left(\begin{bmatrix}\boldsymbol{a}^\star \\ \boldsymbol{a}\end{bmatrix}\right)\right\|_\infty + \left\|\mathrm{diag}\left(\begin{bmatrix}\mathbf{1}_r \\ |\boldsymbol{a}_k|^{-2}\odot|\boldsymbol{a}|^2\end{bmatrix}\right) - \boldsymbol{I}\right\|_\infty
\end{aligned}
$$

$$\leq \left\| \operatorname{diag}\left(\begin{bmatrix} \boldsymbol{a}^{\star -1} \\ |\boldsymbol{a}_k|^{-2} \odot \boldsymbol{a} \end{bmatrix} \right)^{\mathsf{H}} \right\|_{\infty} \|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\|_{\infty} \left\| \operatorname{diag}\left(\begin{bmatrix} \boldsymbol{a}^{\star} \\ \boldsymbol{a} \end{bmatrix} \right) \right\|_{\infty} + \left\| \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_r \\ |\boldsymbol{a}_k|^{-2} \odot |\boldsymbol{a}|^2 \end{bmatrix} \right) - \boldsymbol{I} \right\|_{\infty}$$

$$\leq \max_j \left\{ \frac{1}{|a_j^\star|}, \frac{|a_j^\star|}{|a_{k,j}|^2} \right\} \max_j \left\{ |a_j^\star|, |a_j| \right\} \|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\|_{\infty} + \max_j \left\{ \left| \frac{|a_j^\star|^2}{|a_{k,j}|^2} - 1 \right| \right\} \tag{87}$$

The three maxima in (87) can be controlled using the basic relation

$$\frac{|a_j^\star|}{|a_{k,j}|} \leq \frac{|a_j^\star|}{|a_j^\star| - |a_{k,j} - a_j^\star|} \leq \frac{1}{1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty}, \tag{88}$$

as follows

$$\max_j \left\{ \frac{1}{|a_j^\star|}, \frac{|a_j^\star|}{|a_{k,j}|^2} \right\} \leq \max_j \left\{ \frac{1}{|a_j^\star|}, \frac{1}{|a_j^\star|} \frac{1}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} \right\} \leq \frac{1}{a_{\min}^\star} \frac{1}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} \tag{89a}$$

$$\max_j \left\{ |a_j^\star|, |a_j| \right\} \leq \max_j \left\{ |a_j^\star|, |a_j^\star| (1 + \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty) \right\} \leq \|\boldsymbol{a}^\star\|_\infty (1 + \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty) \tag{89b}$$

$$\max_j \left\{ \left| \frac{|a_j^\star|^2}{|a_{k,j}|^2} - 1 \right| \right\} \leq \max_j \left\{ 1 - \frac{|a_j^\star|^2}{|a_{k,j}|^2}, \frac{|a_j^\star|^2}{|a_{k,j}|^2} - 1 \right\}$$

$$\leq \max_j \left\{ 1, \frac{|a_j^\star|^2}{|a_{k,j}|^2} - 1 \right\} \leq \frac{1}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} - 1. \tag{89c}$$

The bounds (89) and Lemma 6 conclude with (87) on

$$\begin{aligned} &\left\| \boldsymbol{S}\boldsymbol{P}_k \boldsymbol{G}(\boldsymbol{\theta}) \boldsymbol{S}^{-1} - \boldsymbol{I} \right\|_\infty \\ &\quad \leq \frac{\|\boldsymbol{a}^\star\|_\infty}{a_{\min}^\star} \frac{1 + \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} \|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\|_\infty + \frac{1}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} - 1 \\ &\quad \leq \frac{\|\boldsymbol{a}^\star\|_\infty}{a_{\min}^\star} \frac{1 + \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} K_\Delta \left((n+1)\Delta(\boldsymbol{\tau}) \right)^{-2} + \frac{1}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} - 1, \end{aligned} \tag{90}$$

where the constant K_Δ defined in Lemma 6 is numerically evaluated to $K_\Delta \leq 2.32$ by setting the parameter $\alpha = 4.7$.

Step 2: bound $\left\| \boldsymbol{S}\boldsymbol{P}_k \boldsymbol{E}(\boldsymbol{\theta}) \boldsymbol{S}^{-1} \right\|_\infty$ Again, using the block diagonal structure of the matrix $\boldsymbol{E}(\boldsymbol{\theta})$ defined in (36), we similarly have that

$$\boldsymbol{S}\boldsymbol{P}_k \boldsymbol{E}(\boldsymbol{\theta}) \boldsymbol{S}^{-1} = \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{-F_N''(0)}} \operatorname{diag}\left(\boldsymbol{a}^{\star -1} \right) \boldsymbol{E}_1(\boldsymbol{\theta}) \\ \frac{1}{\sqrt{-F_N''(0)}} \operatorname{diag}\left(|\boldsymbol{a}_k|^{-2} \odot \boldsymbol{a}^\star \right)^{\mathsf{H}} \boldsymbol{E}_1(\boldsymbol{\theta}) & -\frac{1}{F_N''(0)} |\boldsymbol{a}_k|^{-2} \boldsymbol{E}_2(\boldsymbol{\theta}) \end{bmatrix}. \tag{91}$$

Therefore, the quantity of interest can be bounded as follows

$$\begin{aligned} \left\| \boldsymbol{S}\boldsymbol{P}_k \boldsymbol{E}(\boldsymbol{\theta}) \boldsymbol{S}^{-1} \right\|_\infty &\leq \frac{1}{\sqrt{-F_N''(0)}} \max_j \left\{ \max \left\{ \frac{1}{|a_j^\star|}, \frac{|a_j^\star|}{|a_{j,k}|^2} \right\} \left| \left\langle \delta'_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \right\rangle \right| \right\} \\ &\quad - \frac{1}{F_N''(0)} \max_j \left\{ \frac{|a_j|}{|a_{j,k}|^2} \left| \left\langle \delta''_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \right\rangle \right| \right\} \\ &\leq \frac{(a_{\min}^\star)^{-1}}{(1 - \|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\|_\infty)^2} \left(\frac{1}{\sqrt{-F_N''(0)}} \max_j \left\{ \left| \left\langle \delta'_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \right\rangle \right| \right\} \right. \\ &\quad \left. - \frac{1}{F_N''(0)} \max_j \left\{ \left| \left\langle \delta''_{\tau_j}, F_N * (\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}^\star)) \right\rangle \right| \right\} \right), \end{aligned} \tag{92}$$

where we used the inequalities (89) on the second line. Substituting the expression provided by Lemma 7 into (92) leads to

$$\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{E}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty \leq \left(\left(C_1\frac{n+1}{\sqrt{-F_N''(0)}}+C_2\frac{(n+1)^2}{-F_N''(0)}\right)\frac{\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty}{\|\boldsymbol{a}^\star\|_\infty}\right.$$
$$\left.+\left(C_2\frac{(n+1)^2}{-F_N''(0)}+C_3\frac{(n+1)^3}{(-F_N''(0))^{3/2}}\right)\sqrt{-F_N''(0)}\,\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\|_\infty\right)$$
$$\cdot\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\frac{((n+1)\Delta(\boldsymbol{\tau}))^{-2}}{(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty)^2}. \quad (93)$$

Evaluating the constants $C_1 \leq 3.24$, $C_2 \leq 22.90$ and $C_3 \leq 105.55$ defined in (54) with parameters $\alpha = 4.7$ and $\beta = \frac{\alpha}{4} = 1.2$, altogether with the inequality $\frac{(n+1)^2}{-F_N''(0)} \leq \frac{27}{16\pi^2}$ under the assumption $n \geq 2$ yields

$$\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{E}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty \leq \left(K_a\frac{\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty}{\|\boldsymbol{a}^\star\|_\infty}+K_\tau\sqrt{-F_N''(0)}\,\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\|_\infty\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\frac{((n+1)\Delta(\boldsymbol{\tau}))^{-2}}{(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty)^2}, \quad (94)$$

where the constants K_a and K_τ are given by

$$K_a = C_1\sqrt{\frac{27}{16\pi^2}}+C_2\frac{27}{16\pi^2} \leq 5.26, \quad (95a)$$
$$K_\tau = C_2\frac{27}{16\pi^2}+C_3\left(\frac{27}{16\pi^2}\right)^{3/2} \leq 11.38. \quad (95b)$$

Step 3: combine the bounds The scaled Hessian matrix $\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}$ can be controlled over $\mathcal{S}_k$ by the triangle inequality as follows

$$\begin{aligned}\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty &\leq \left\|\boldsymbol{S}\boldsymbol{P}_k\left(\boldsymbol{G}(\boldsymbol{\theta})+\boldsymbol{E}(\boldsymbol{\theta})\right)\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty \\ &\leq \left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty+\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{E}(\boldsymbol{\theta})\boldsymbol{S}^{-1}\right\|_\infty. \quad (96)\end{aligned}$$

Finally, the inequality $|\Delta(\boldsymbol{\tau})-\Delta(\boldsymbol{\tau}^\star)| \leq 2\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty$ holds for every $\boldsymbol{\tau},\boldsymbol{\tau}^\star \subset \mathbb{R}$. This implies, with the assumption $\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty \leq \frac{1}{4}\Delta(\boldsymbol{\tau}^\star)$, that $\Delta(\boldsymbol{\tau}) \geq \frac{1}{2}\Delta(\boldsymbol{\tau}^\star)$. Substituting the bounds given in (90) and (94) yields

$$\begin{aligned}&\left\|\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{S}^{-1}-\boldsymbol{I}\right\|_\infty \\ &\leq \frac{1}{(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty)^2}-1 \\ &+\left(K_\Delta\left(1+\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty\right)+K_a\frac{\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty}{\|\boldsymbol{a}^\star\|_\infty}+K_\tau\sqrt{-F_N''(0)}\,\|\boldsymbol{\tau}_k-\boldsymbol{\tau}^\star\|_\infty\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\frac{((n+1)\Delta(\boldsymbol{\tau}))^{-2}}{(1-\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty)^2} \\ &\leq \frac{1}{(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty)^2}-1 \\ &\quad+\left(K_\Delta\left(1+\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty\right)+\left(K_a+K_\tau\right)\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\frac{((n+1)\Delta(\boldsymbol{\tau}))^{-2}}{(1-\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty)^2} \\ &\leq \frac{1}{(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty)^2}-1 \\ &\quad+\left(K_\Delta+\left(K_\Delta+K_a+K_\tau\right)\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty\right)\frac{\|\boldsymbol{a}^\star\|_\infty}{a^\star_{\min}}\frac{((n+1)\Delta(\boldsymbol{\tau}))^{-2}}{(1-\|\boldsymbol{S}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star\right)\|_\infty)^2} \\ &\leq \frac{1}{(1-\|\boldsymbol{S}(\boldsymbol{\theta}_k-\boldsymbol{\theta}^\star)\|_\infty)^2}-1\end{aligned}$$

$$+ \left(K_\Delta + \left(K_\Delta + K_a + K_\tau\right) \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right) \frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a^\star_{\min}} \frac{4\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}}{\left(1 - \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right)^2}$$

$$\leq \frac{1}{\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)^2} - 1 + \left(4K_\Delta + K_\theta \left\|\boldsymbol{S}\left(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star\right)\right\|_\infty\right) \frac{\left\|\boldsymbol{a}^\star\right\|_\infty}{a^\star_{\min}} \frac{\left((n+1)\Delta(\boldsymbol{\tau}^\star)\right)^{-2}}{\left(1 - \left\|\boldsymbol{S}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^\star)\right\|_\infty\right)^2}, \tag{97}$$

where defined the constant K_θ as

$$K_\theta = 4\left(K_\Delta + K_a + K_\tau\right) \leq 75.80. \tag{98}$$

This concludes the proof of the theorem. □

B.4 Proof of Lemma 6

Leveraging the block structure (34) of the matrix $\boldsymbol{D}(\boldsymbol{\tau})$, it boils down to controlling

$$\left\|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty \leq \max\left\{\left\|\boldsymbol{D}_0(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty + \left\|\boldsymbol{D}_1(\boldsymbol{\tau})\right\|_\infty, \left\|\boldsymbol{D}_2(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty + \left\|\boldsymbol{D}_1(\boldsymbol{\tau})\right\|_\infty\right\}, \tag{99}$$

which can be accomplished with the aid of Lemma 5. Specifically, recalling the expressions in (35), applying Lemma 5 with parameters α and $\beta = 0$, and noticing that the inequality $\frac{(n+1)^2}{-F_N''(0)} \leq \frac{27}{16\pi^2}$ holds whenever $n \geq 2$, the quantities of interest in (99) can be controlled as follows

$$\left\|\boldsymbol{D}_0(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty = \max_i \sum_{j \neq i} \left|F_N(\tau_j - \tau_i)\right| \leq C_0 \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}, \tag{100a}$$

$$\begin{aligned}\left\|\boldsymbol{D}_1(\boldsymbol{\tau})\right\|_\infty &= \frac{1}{\sqrt{-F_N''(0)}} \max_i \sum_j \left|F_N'(\tau_j - \tau_i)\right| \leq C_1 \frac{n+1}{\sqrt{-F_N''(0)}} \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2} \\ &\leq \frac{3\sqrt{3}}{4\pi} C_1 \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2},\end{aligned} \tag{100b}$$

$$\begin{aligned}\left\|\boldsymbol{D}_2(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty &= \frac{1}{-F_N''(0)} \max_i \sum_{j \neq i} \left|F_N''(\tau_j - \tau_i)\right| \leq C_2 \frac{(n+1)^2}{-F_N''(0)} \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2} \\ &\leq \frac{27}{16\pi^2} C_2 \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2}.\end{aligned} \tag{100c}$$

Further substituting (100) into (99) leads to

$$\left\|\boldsymbol{D}(\boldsymbol{\tau}) - \boldsymbol{I}\right\|_\infty \leq \max\left\{C_0 + \frac{3\sqrt{3}}{4\pi} C_1, \frac{3\sqrt{3}}{4\pi} C_1 + \frac{27}{16\pi^2} C_2\right\} \left((n+1)\Delta(\boldsymbol{\tau})\right)^{-2} \tag{101}$$

which leads to the desired statement. □

B.5 Proof of Lemma 7

For the first inequality, following the definition, we have that

$$\begin{aligned}\left|\left\langle \Phi(\delta'_{\tau_j}), \Phi\left(\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}_\star)\right)\right\rangle\right| &= \left|\left\langle \delta'_{\tau_j}, \Phi^*\left(\Phi\left(\mu(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta}_\star)\right)\right)\right\rangle\right| \\ &= \left|\sum_{\ell=1}^{r} a_\ell F_N'(\tau_j - \tau_\ell) - \sum_{\ell=1}^{r^\star} a^\star_\ell F_N'(\tau_j - \tau^\star_\ell)\right| \\ &= \left|\sum_{\ell=1}^{r} (a_\ell - a^\star_\ell) F_N'(\tau_j - \tau_\ell) + \sum_{\ell=1}^{r} a^\star_\ell \left(F_N'(\tau_j - \tau_\ell) - F_N'(\tau_j - \tau^\star_\ell)\right)\right| \\ &\leq \left|\sum_{\ell=1}^{r} (a_\ell - a^\star_\ell) F_N'(\tau_j - \tau_\ell)\right| + \left|\sum_{\ell=1}^{r} a^\star_\ell \left(F_N'(\tau_j - \tau_\ell) - F_N'(\tau_j - \tau^\star_\ell)\right)\right|,\end{aligned} \tag{102}$$

where the last line used the triangle inequality. To proceed, we control the two terms separately. Using Hölder's inequality and Lemma 5, we obtain

$$\left|\sum_{\ell=1}^{r}(a_\ell - a_\ell^\star)F_N'(\tau_j - \tau_\ell)\right| \leq \|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty \sum_{\ell=1}^{r}|F_N'(\tau_j-\tau_\ell)| \\ \leq C_1(n+1)\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty ((n+1)\Delta(\boldsymbol{\tau}))^{-2}. \tag{103}$$

Next, the second term can be bounded by applying Hölder's inequality, the mean-value theorem and Lemma 5 with parameter $\beta \geq (n+1)\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty$ as follows

$$\begin{aligned}\left|\sum_{\ell=1}^{r} a_\ell^\star\left(F_N'(\tau_j-\tau_\ell) - F_N'(\tau_j-\tau_\ell^\star)\right)\right| &\leq \|\boldsymbol{a}^\star\|_\infty \sum_{\ell=1}^{r}|F_N'(\tau_j-\tau_\ell) - F_N'(\tau_j-\tau_\ell^\star)| \\ &\leq \|\boldsymbol{a}^\star\|_\infty \|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty \sum_{\ell=1}^{r}\sup_{|u_\ell|\leq\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty}|F_N''(\tau_j-\tau_\ell+u_\ell)| \\ &\leq \|\boldsymbol{a}^\star\|_\infty \|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty C_2(n+1)^2((n+1)\Delta(\boldsymbol{\tau}))^{-2}. \end{aligned}\tag{104}$$

Plugging the previous two bounds into the inequality (102) reduces to

$$\left|\left\langle \Phi(\delta_{\tau_j}'), \Phi\left(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}_\star)\right)\right\rangle\right| \\ \leq \left(C_1\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty + C_2\|\boldsymbol{a}^\star\|_\infty(n+1)\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty\right)(n+1)((n+1)\Delta(\boldsymbol{\tau}))^{-2}. \tag{105}$$

Moreover, we may show that through analogous reasoning that

$$\left|\left\langle \Phi(\delta_{\tau_j}''), \Phi\left(\mu(\boldsymbol{\theta})-\mu(\boldsymbol{\theta}_\star)\right)\right\rangle\right| \\ \leq \left(C_2\|\boldsymbol{a}-\boldsymbol{a}^\star\|_\infty + C_3\|\boldsymbol{a}^\star\|_\infty(n+1)\|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty\right)(n+1)^2((n+1)\Delta(\boldsymbol{\tau}))^{-2}, \tag{106}$$

which concludes the proof. □