Rates of estimation for high-dimensional multi-reference alignment

Zehao Dou, Zhou Fan, Harrison Zhou

May 5, 2022

Abstract

We study the continuous multi-reference alignment model of estimating a periodic function on the circle from noisy and circularly-rotated observations. Motivated by analogous high-dimensional problems that arise in cryo-electron microscopy, we establish minimax rates for estimating generic signals that are explicit in the dimension K. In a high-noise regime with noise variance $\sigma^2 \gtrsim K$, the rate scales as σ^6 and has no further dependence on the dimension. This rate is achieved by a bispectrum inversion procedure, and our analyses provide new stability bounds for bispectrum inversion that may be of independent interest. In a low-noise regime where $\sigma^2 \lesssim K/\log K$, the rate scales instead as $K\sigma^2$, and we establish this rate by a sharp analysis of the maximum likelihood estimator that marginalizes over latent rotations. A complementary lower bound that interpolates between these two regimes is obtained using Assouad's hypercube lemma.

Contents

1 Introduction

Multi-reference alignment (MRA) refers to the problem of estimating an unknown signal from noisy samples that are subject to latent rotational transformations (Ritov, 1989; Bandeira et al., 2014). This problem has seen renewed interest in recent years, as a simplified model for molecular reconstruction in cryo-electron microscopy (cryo-EM) and related methods of molecular imaging (Bendory et al., 2020a; Singer and Sigworth, 2020). It arises also in various other applications in structural biology and image registration (Sadler and Giannakis, 1992; Brown, 1992; Diamond, 1992). Recent literature has established rates of estimation for MRA in fixed

Department of Statistics and Data Science, Yale University, USA.
zehao.dou@yale.edu, zhou.fan@yale.edu, huibin.zhou@yale.edu

dimensions (Perry et al., 2019; Bandeira et al., 2020; Abbe et al., 2018a; Ghosh and Rigollet, 2021), describing a rich picture of how these rates may depend on the signal-to-noise ratio and properties of the underlying signal. However, many applications of MRA involve high-dimensional signals, and there is currently limited understanding of optimal rates of estimation in high-dimensional settings.

In the continuous MRA model—the focus of this work—the signal is a smooth periodic function f on the circular domain $[-\pi, \pi)$. We observe independent samples of f in additive white noise, where each sample has a uniformly random latent rotation of its domain (Bandeira et al., 2020; Fan et al., 2021). The true function f is identifiable only up to rotation, and we will study its estimation under the rotation-invariant squared-error loss

$$L(\hat{f}, f) = \min_{\alpha \in [-\pi,\pi)} \int_{-\pi}^{\pi} \left(\hat{f}(t) - f(t - \alpha \bmod 2\pi)\right)^2 \mathrm{d}t. \quad (1)$$

In the closely related discrete MRA model, the signal is instead a vector $x \in \mathbb{R}^K$, observed in additive Gaussian noise with cyclic permutations of its coordinates (Bandeira et al., 2014; Perry et al., 2019). The continuous and discrete models are similar, in that both rotational actions are diagonalized in the (continuous or discrete, resp.) Fourier basis, and these diagonal actions have similar forms.

A recent line of work has studied rates of estimation for MRA in "low dimensions", treating as constant the dimension K for discrete MRA, or the maximum Fourier frequency K for continuous MRA. Many such results have specifically focused on a regime of high noise: In this regime, Perry et al. (2019) showed that the squared-error risk for estimating "generic" signals scales with the noise standard deviation as σ^6. Bandeira et al. (2020) showed that this scaling for estimating a "non-generic" signal depends on its pattern of zero and non-zero Fourier coefficients, and derived rate-optimal upper and lower bounds over minimax classes of such signals. Rates of estimation for MRA with non-uniform rotations were studied in Abbe et al. (2018a), with a dihedral group of both rotations and reflections in Bendory et al. (2022), with sparse signals in Ghosh and Rigollet (2021), and with down-sampled observations in a super-resolution context in Bendory et al. (2020b).

It is empirically observed, for example in Fan et al. (2021, Section 5), that electric potential functions of protein molecules in cryo-EM applications may require basis representations with dimensions in the thousands to capture secondary structure, and even higher dimensions to achieve near-atomic resolution. Motivated by this observation, in this paper, we extend the above line of work to study the continuous MRA model in potentially high dimensions, in both high-noise and low-noise regimes. Our main results are described informally as follows: Let

$$\theta(f) = (\theta_{1,1}, \theta_{1,2}, \theta_{2,1}, \theta_{2,2}, \theta_{3,1}, \theta_{3,2}, \ldots)$$

be the coefficients of f in the real Fourier basis over $[-\pi, \pi)$, and let

$$(r_k \cos \phi_k, r_k \sin \phi_k) = (\theta_{k,1}, \theta_{k,2}) \quad (2)$$

be the representation of the coefficients in the k^{th} Fourier frequency in terms of the magnitude r_k and phase ϕ_k. Consider a $2K$-dimensional class of signals f represented by

$$\Theta = \Big\{ f : \; r_k \asymp 1 \text{ for } k = 1, \ldots, K, \; r_k = 0 \text{ for all } k \geq K+1 \Big\}.$$

Our results distinguish two separate signal-to-noise regimes for estimating f, based on the size of the entry-wise noise variance σ^2 in the Fourier basis. We establish sharp minimax rates of estimation in both regimes, under mild conditions for the sample size N, that are explicit in their dependence on the dimension K.

Theorem (Informal)**.** *(a) (High noise) If $\sigma^2 \gtrsim K$ and $N \gtrsim \sigma^6 \log K$, then*

$$\inf_{\hat{f}} \sup_{f \in \Theta} \mathbb{E}[L(\hat{f}, f)] \asymp \frac{\sigma^6}{N}.$$

(b) (Low noise) If $\sigma^2 \lesssim K/\log K$ and $N \gtrsim K\sigma^2 \log K$, then

$$\inf_{\hat{f}} \sup_{f \in \Theta} \mathbb{E}[L(\hat{f}, f)] \asymp \frac{K\sigma^2}{N}.$$

We refer to Theorems 2.1 and 2.2 to follow for precise statements of these results.

In the high-noise regime where $\sigma^2 \gtrsim K$, we show that the minimax rate is achieved by a variant of a third-order method-of-moments (MoM) procedure. The scaling with σ^6 matches previous results of Perry et al. (2019), but a notable new feature of the rate is that it does not have an explicit dependence on the dimension K. In the MRA model, for functions having the Fourier coefficients (2) and bandlimited to maximum Fourier frequency K, second-order moments correspond to the power spectrum

$$\Big\{r_k^2 : k = 1, \ldots, K\Big\}$$

and third-order moments to the Fourier bispectrum

$$\Big\{\phi_{k+l} - \phi_k - \phi_l : k, l \in \{1, \ldots, K\} \text{ and } k + l \leq K\Big\}$$

Method-of-moments in this context is also known as bispectrum inversion (Sadler and Giannakis, 1992; Bendory et al., 2017), which aims to estimate the Fourier phases $\{\phi_k\}$ from an estimate of the bispectrum. Results of Bendory et al. (2017); Perry et al. (2019) imply that for signals where $r_k \neq 0$ for every $k = 1, \ldots, K$, these phases are uniquely determined by the bispectrum. Our analyses quantify this statement in two ways: First, we show that the linear system relating the bispectrum to the phases is well-conditioned, having all non-trivial singular values on the order of $1/\sqrt{K}$ (cf. Lemma 4.7). Thus, each phase ϕ_k may be estimated with squared error σ^6/K, leading to an overall estimation rate that is independent of K. Second, to resolve phase ambiguities before solving this linear system, we show stability for bispectrum inversion also in a periodic ℓ_∞ sense (cf. Lemma 4.10). These results combine to give an estimator that achieves the minimax rate of estimation in the regime $\sigma^2 \gtrsim K$.

Our definition of the low-noise regime $\sigma^2 \lesssim K/\log K$ and minimax rate in this regime are most comparable to recent results of Romanov et al. (2021), which studied instead the discrete MRA model in the asymptotic limit $K \to \infty$ and $(\sigma^2 \log K)/K \to 1/\alpha \in (0, \infty)$. This work studied a Bayesian setting where each Fourier coefficient of f has a standard Gaussian prior, and showed a transition in the Bayes risk and associated sample complexity at the sharp threshold $\alpha = 2$. The analysis in Romanov et al. (2021) for the low-noise regime relies on the discreteness of the rotational model, analyzing a template matching procedure that can *exactly* recover the latent rotation for each sample if the true signal f is known. For continuous MRA, it is clear that this estimation of each rotation is possible only up to some per-sample error that is independent of the sample size N. Averaging the correspondingly rotated samples would yield an estimation bias that does not vanish with N. Direct application of third-order method-of-moments also does not yield the optimal estimation rate across the entire low-noise regime. We instead analyze the maximum-likelihood estimator (MLE) that marginalizes over latent rotations, to obtain the minimax rate in this regime. Our proof uses the classical idea of second-order Taylor expansion of the empirical log-likelihood function, although establishing the estimation rate for sample sizes $N \gtrsim K\sigma^2 \log K$ requires a delicate argument to bound the Hessian of the empirical log-likelihood. We describe this argument in Section 5.

1.1 Further related literature

A body work on MRA and related models focuses on the synchronization approach, which seeks to first estimate the latent rotation of each sample based on the relative rotational alignments between pairs of samples (Singer, 2011). In the context of cryo-EM, this is known also as the "common lines" method (Singer et al., 2010; Singer and Shkolnisky, 2011). Algorithms developed and studied for estimating these pairwise alignments include spectral procedures (Singer, 2011; Singer and Shkolnisky, 2011; Ling, 2022), semidefinite relaxations (Singer, 2011; Singer and Shkolnisky, 2011; Bandeira et al., 2014, 2015), and iterative power method or approximate message passing approaches (Boumal, 2016; Perry et al., 2018).

In high-noise regimes, synchronization-based estimation may fail to recover the latent rotations, or may lead to a biased and inconsistent estimate of the underlying signal. A separate line of work has studied alternative method-of-moments or maximum likelihood procedures for the MRA problem, which marginalize over the latent rotations (Abbe et al., 2018a; Boumal et al., 2018; Perry et al., 2019; Bandeira et al., 2020; Ghosh and Rigollet, 2021; Bendory et al., 2022). These papers relate the rate of estimation in high noise to the order of moments needed to identify the true signal, which may differ depending on the sparsity pattern of its Fourier coefficients and the distribution of the latent random rotations.

Related analyses have been performed for three-dimensional rotational actions, as arising in Procrustes alignment problems (Pumir et al., 2021) and cryo-EM (Sharon et al., 2020). For cryo-EM, these methods encompass invariant-features approaches (Kam, 1980) and expectation-maximization algorithms (Sigworth, 1998; Scheres et al., 2005; Scheres, 2012). The works Bandeira et al. (2017); Abbe et al. (2018b) studied method-of-moments estimators in problems with general rotational groups, where Bandeira et al. (2017) related the rates of estimation and numbers of moments needed to identify the true signal to the structure of the invariant polynomial algebra of the group action. In these general settings, Brunel (2019); Fan et al. (2020); Katsevich and Bandeira (2020); Fan et al. (2021) studied also properties of the log-likelihood function, its optimization landscape, and the Fisher information matrix, relating the structure of the invariant algebra to asymptotic rates of estimation for the MLE.

1.2 Outline

Section 2 provides a formal statement of the continuous MRA model and of our main results. Section 3 provides some preliminaries that relate the loss function to the Fourier magnitudes and phases. Section 4 proposes and analyzes a third-order method-of-moments estimator, which determines the phases by inverting the Fourier bispectrum. This estimator attains the minimax upper bound for squared-error risk in the high-noise regime. Section 5 analyzes the maximum likelihood estimator that attains the minimax upper bound for squared-error risk in the low-noise regime. Finally, Section 6 gives a minimax lower bound using Assouad's lemma, which matches the upper bounds of Sections 4 and 5 while also interpolating between these two signal-to-noise regimes.

1.3 Notation

For a complex number $z = re^{i\theta} \in \mathbb{C}$, $\overline{z} = re^{-i\theta}$ is its complex conjugate. $\operatorname{Arg} z = \theta$ is its principal argument in the range $[-\pi, \pi)$. $\langle u, v\rangle = \sum_k u_k \overline{v_k}$ is the ℓ_2 inner-product for real or complex vectors, and $\|u\| = \sqrt{\langle u, u\rangle}$ is the ℓ_2 norm. $I_K \in \mathbb{R}^{K\times K}$ is the identity matrix in dimension K. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ is the complex mean-zero Gaussian distribution, with independent real and imaginary parts having real Gaussian distribution $\mathcal{N}(0, \frac{\sigma^2}{2})$. We write $a \wedge b = \min(a, b)$. For a function $F : \mathbb{R}^k \to \mathbb{R}$, we denote its gradient and Hessian by $\nabla F \in \mathbb{R}^k$ and $\nabla^2 F \in \mathbb{R}^{k\times k}$. For two distributions P and Q, $D_{\mathrm{KL}}(P\|Q) = \int \log(\frac{P}{Q})dP$ is their Kullback-Leibler (KL) divergence.

2 Model and main results

Let $\mathcal{A} = [-\pi, \pi)$ be identified with the unit circle, with addition modulo 2π. Let $f : \mathcal{A} \to \mathbb{R}$ be a smooth periodic function on $\mathcal{A}$. We represent rotations of the circle by angles $\alpha \in \mathcal{A}$, and denote the function f with domain rotated by α as

$$f_\alpha(t) = f(t - \alpha \bmod 2\pi).$$

We study estimation of f from N i.i.d. samples of the form

$$f_\alpha(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t), \qquad \alpha \sim \mathrm{Unif}([-\pi, \pi)).$$

In each sample, α represents a different latent and uniformly random rotation of the domain of f, and the entire rotated function f_α is observed with additive continuous white noise $\sigma\,\mathrm{d}W(t)$ on the circle. We assume that $\sigma > 0$ is a fixed and known noise level. As f is identifiable only up to rotation, we consider the rotation-invariant loss (1).

Note that we may alternatively study a model where each rotated function $f_\alpha(t)$ is observed with Gaussian noise only at a discrete set of points $t \in \mathcal{A}$ fixed or sampled from its domain (Bandeira et al., 2020; Bendory et al., 2020b). We study the above continuous observation model so as to abstract away aspects of the problem that are related to this discrete sampling.

The mean value of f over the circle is invariant to rotations, and is easily estimated by averaging across samples. Thus, let us assume for simplicity and without loss of generality that f has known mean 0. Passing to the Fourier domain, we assume that f is bandlimited to K Fourier frequencies, i.e. f admits the Fourier sequence representation

$$f(t) = \sum_{k=1}^{K} \theta_{k,1} f_{k,1}(t) + \theta_{k,2} f_{k,2}(t), \qquad f_{k,1}(t) = \frac{1}{\sqrt{\pi}} \cos kt, \quad f_{k,2}(t) = \frac{1}{\sqrt{\pi}} \sin kt,$$

where $\{f_{k,1}, f_{k,2} : k = 1, \ldots, K\}$ are orthonormal Fourier basis functions over $[-\pi, \pi)$, and

$$\theta = (\theta_{1,1}, \theta_{1,2}, \ldots, \theta_{K,1}, \theta_{K,2}) \in \mathbb{R}^{2K}$$

are the Fourier coefficients of f. We assume implicitly throughout the paper that $K \geq 2$, and we are interested in applications with potentially large values of this bandlimit K.

Importantly, due to the choice of Fourier basis, the $2K$-dimensional space of such bandlimited functions is closed under rotations of the circle. The rotation $f \mapsto f_\alpha$ induces a map from the Fourier coefficients of f to those of f_α, which we denote as $\theta \mapsto g(\alpha) \cdot \theta$ for an orthogonal matrix $g(\alpha) \in \mathbb{R}^{2K \times 2K}$. Explicitly, this map $\theta \mapsto g(\alpha) \cdot \theta$ is given separately for each Fourier frequency $k = 1, \ldots, K$ by

$$\begin{pmatrix} \theta_{k,1} \\ \theta_{k,2} \end{pmatrix} \mapsto \begin{pmatrix} \cos k\alpha & -\sin k\alpha \\ \sin k\alpha & \cos k\alpha \end{pmatrix} \begin{pmatrix} \theta_{k,1} \\ \theta_{k,2} \end{pmatrix}, \tag{3}$$

and $g(\alpha)$ is the block-diagonal matrix with these 2×2 blocks. Equivalently, writing

$$(\theta_{k,1}, \theta_{k,2}) = (r_k \cos \phi_k, r_k \sin \phi_k)$$

where $r_k \geq 0$ is the magnitude and $\phi_k \in \mathcal{A}$ is the phase (identified modulo 2π), this map is given for each $k = 1, \ldots, K$ by

$$(r_k, \phi_k) \mapsto (r_k, \phi_k + k\alpha). \tag{4}$$

The samples $f_\alpha(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)$ represented in this Fourier sequence space take the form

$$y^{(m)} = g(\alpha^{(m)}) \cdot \theta + \sigma \varepsilon^{(m)} \in \mathbb{R}^{2K} \text{ for } m = 1, \ldots, N \tag{5}$$

where $\alpha^{(1)}, \ldots, \alpha^{(N)} \overset{\text{iid}}{\sim} \text{Unif}([-\pi, \pi))$, $\varepsilon^{(1)}, \ldots, \varepsilon^{(N)} \overset{\text{iid}}{\sim} \mathcal{N}(0, I_{2K})$, and these are independent. Writing $\hat{\theta} \in \mathbb{R}^{2K}$ for the Fourier coefficients of the estimated function $\hat{f}$ (which should likewise be bandlimited to K Fourier frequencies), the loss (1) is equivalent to

$$L(\hat{\theta}, \theta) = \min_{\alpha \in \mathcal{A}} \|\hat{\theta} - g(\alpha) \cdot \theta\|^2. \tag{6}$$

In the remainder of this paper, we will consider the problem in this sequence form.

We reserve the notation θ^* for the Fourier coefficients of the true unknown function. Fixing constants $\underline{c}, \bar{c} > 0$, for some $r > 0$, we consider the parameter space of "generic" Fourier coefficient vectors

$$\Theta(r) = \Big\{\theta^* \in \mathbb{R}^{2K} : \underline{c} r \leq r_k(\theta^*) \leq \bar{c} r \text{ for all } k = 1, \ldots, K\Big\}. \tag{7}$$

Our main results are the following two theorems, which characterize the minimax rates of estimation over $\Theta(r)$ in high-noise and low-noise regimes.

Theorem 2.1 (Minimax risk in high noise). *Fix any constant $c_0 > 0$. If $\sigma^2 \geq c_0 K r^2$, then for a constant $C_0 > 0$ depending only on $\underline{c}, \bar{c}, c_0$ and for any $N \geq C_0 \frac{\sigma^6}{r^6} \log K$,*

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \Theta(r)} \mathbb{E}_{\theta^*}[L(\hat{\theta}, \theta^*)] \asymp \frac{\sigma^6}{r^4 N}.$$

Theorem 2.2 (Minimax risk in low noise). *There exist constants $C_0, C_1 > 0$ depending only on $\underline{c}, \bar{c}$ such that if $\sigma^2 \leq \frac{K r^2}{C_1 \log K}$ and $N \geq C_0 \frac{K\sigma^2}{r^2} \log K$, then*

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \Theta(r)} \mathbb{E}_{\theta^*}[L(\hat{\theta}, \theta^*)] \asymp \frac{K\sigma^2}{N}.$$

In both statements, $\mathbb{E}_{\theta^*}$ is the expectation over N samples $y^{(1)}, \ldots, y^{(N)}$ from the model (5) with true parameter θ^*. The infimum $\inf_{\hat{\theta}}$ is over all estimators $\hat{\theta}$ based on these samples, and $\asymp$ denotes upper and lower bounds up to constant multiplicative factors that depend only on $\underline{c}, \bar{c}$ and c_0 for Theorem 2.1.

We observe that the problem is invariant under simultaneous rescaling of r^2, σ^2, and the loss function $L(\hat{\theta}, \theta^*)$. Hence, for interpretation, we may restrict without loss of generality to $r = 1$, and this yields the informal statements previously described in the introduction.

3 Preliminaries

3.1 Bounds for the loss

For $\phi, \phi' \in \mathcal{A} = [-\pi, \pi)$, identifying $\mathcal{A}$ with the unit circle, we define the circular distance

$$|\phi - \phi'|_{\mathcal{A}} = \min_{j \in \mathbb{Z}} |\phi - \phi' + 2\pi j|. \tag{8}$$

It is direct to check that $(\phi, \phi') \mapsto |\phi - \phi'|_{\mathcal{A}}$ is a metric on $\mathcal{A}$, satisfying the triangle inequality and the upper bound

$$|\phi - \phi'|_{\mathcal{A}} \leq \min(\pi, |\phi - \phi'|). \tag{9}$$

We may express and bound the loss (6) in terms of the Fourier magnitudes and phases.

Proposition 3.1. *Let $\theta = (r_k \cos \phi_k, r_k \sin \phi_k)_{k=1}^K$ and $\theta' = (r'_k \cos \phi'_k, r'_k \sin \phi'_k)_{k=1}^K$. Then*

$$L(\theta, \theta') = \sum_{k=1}^K (r_k - r'_k)^2 + \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K 2 r_k r'_k \Big[1 - \cos(\phi_k - \phi'_k + k\alpha)\Big]. \tag{10}$$

Consequently, for universal constants $C, c > 0$,

$$\begin{aligned}\sum_{k=1}^K (r_k - r'_k)^2 + c \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K r_k r'_k |\phi_k - \phi'_k + k\alpha|_{\mathcal{A}}^2 \\ \leq L(\theta, \theta') \leq \sum_{k=1}^K (r_k - r'_k)^2 + C \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K r_k r'_k |\phi_k - \phi'_k + k\alpha|_{\mathcal{A}}^2.\end{aligned}$$

Proof. For any $\alpha \in \mathbb{R}$, we have

$$\begin{aligned}\|\theta' - g(\alpha) \cdot \theta\|^2 &= \sum_{k=1}^K \left[(r'_k \cos \phi'_k - r_k \cos(\phi_k + k\alpha))^2 + (r'_k \sin \phi'_k - r_k \sin(\phi_k + k\alpha))^2\right] \\ &= \sum_{k=1}^K (r_k - r'_k)^2 + 2 r_k r'_k \left[1 - \cos(\phi_k - \phi'_k + k\alpha)\right].\end{aligned}$$

Taking the infimum over α gives (10). The consequent inequalities follow from the bounds $c|t|_{\mathcal{A}}^2 \leq 1 - \cos(t) \leq C|t|_{\mathcal{A}}^2$ for universal constants $C, c > 0$, applied with $t = \phi_k - \phi'_k + k\alpha$ for each k. ☐

3.2 Complex representation

It will be notationally and conceptually convenient to pass between $\theta \in \mathbb{R}^{2K}$ and a complex representation by $\tilde{\theta} \in \mathbb{C}^K$. We use throughout

$$\operatorname{Arg} z \in [-\pi, \pi) \tag{11}$$

for the principal complex argument of $z \in \mathbb{C}$. Recalling the k^{th} Fourier coefficient pair $(\theta_{k,1}, \theta_{k,2}) = (r_k \cos \phi_k, r_k \sin \phi_k)$, we set

$$\tilde{\theta}_k = \theta_{k,1} + i\theta_{k,2} = r_k e^{i\phi_k} \in \mathbb{C}. \tag{12}$$

For $\theta, \theta' \in \mathbb{R}^{2K}$, note then that

$$\langle \theta, \theta' \rangle = \sum_{k=1}^{K} \theta_{k,1}\theta'_{k,1} + \theta_{k,2}\theta'_{k,2} = \sum_{k=1}^{K} \operatorname{Re} \tilde{\theta}_k \overline{\tilde{\theta}'_k} = \frac{\langle \tilde{\theta}, \tilde{\theta}' \rangle + \langle \tilde{\theta}', \tilde{\theta} \rangle}{2} \tag{13}$$

where the left side is the real inner-product, and the right side is the complex inner-product $\langle u, v \rangle = \sum_k u_k \overline{v_k}$.

Similarly, we may represent the sample $y^{(m)} \in \mathbb{R}^{2K}$ from (5) by $\tilde{y}^{(m)} \in \mathbb{C}^K$ where

$$\tilde{y}_k^{(m)} = y_{k,1}^{(m)} + i y_{k,2}^{(m)} \in \mathbb{C}.$$

Then, recalling the form of the rotational action (4), we have

$$\tilde{y}_k^{(m)} = r_k e^{i(\phi_k + k\alpha^{(m)})} + \sigma \tilde{\varepsilon}_k^{(m)} \in \mathbb{C} \tag{14}$$

where $\tilde{\varepsilon}_k^{(m)} = \varepsilon_{k,1}^{(m)} + i\varepsilon_{k,2}^{(m)} \sim \mathcal{N}_{\mathbb{C}}(0, 2)$ is complex Gaussian noise, independent across both frequencies $k = 1, \ldots, K$ and samples $m = 1, \ldots, N$.

4 Method-of-moments estimator

In this section, we develop an estimator based on a third-order method-of-moments idea. We analyze the risk of this estimator for any noise level $\sigma^2 > 0$, and show in particular that it attains the minimax upper bound of Theorem 2.1 in the high-noise regime.

Throughout this section, let us denote the Fourier magnitudes and phases of the true parameter as $\theta^* = (r_k \cos \phi_k, r_k \sin \phi_k)_{k=1}^K$ and write $\mathbb{E}$ for $\mathbb{E}_{\theta^*}$. Observe from (14) that for every $k = 1, \ldots, K$,

$$\mathbb{E}\big[|\tilde{y}_k^{(m)}|^2\big] = r_k^2 + 2\sigma^2.$$

Then $N^{-1} \sum_{m=1}^N |\tilde{y}_k^{(m)}|^2 - 2\sigma^2$ provides an unbiased estimate of r_k^2. Furthermore, denote

$$\mathcal{I} = \Big\{(k, l) : k, l \in \{1, \ldots, K\} \text{ and } k + l \leq K\Big\}. \tag{15}$$

Applying that $\{\tilde{\varepsilon}_k^{(m)} : k = 1, \ldots, K\}$ are independent with mean 0, and also $\mathbb{E}[(\tilde{\varepsilon}_k^{(m)})^2] = 0$ (cf. Proposition A.1 of Appendix A), for any $(k, l) \in \mathcal{I}$ including the case $k = l$ we have

$$\begin{aligned} \mathbb{E}\left[\tilde{y}_{k+l}^{(m)} \cdot \overline{\tilde{y}_k^{(m)}} \cdot \overline{\tilde{y}_l^{(m)}}\right] &= \mathbb{E}\left[r_{k+l} e^{i(\phi_{k+l} + (k+l)\alpha^{(m)})} \cdot r_k e^{i(-\phi_k - k\alpha^{(m)})} \cdot r_l e^{i(-\phi_l - l\alpha^{(m)})}\right] \\ &= r_{k+l} r_k r_l e^{i(\phi_{k+l} - \phi_k - \phi_l)}. \end{aligned}$$

Thus the complex argument of $N^{-1} \sum_{m=1}^N \tilde{y}_{k+l}^{(m)} \cdot \overline{\tilde{y}_k^{(m)}} \cdot \overline{\tilde{y}_l^{(m)}}$ provides an estimate of the Fourier bispectrum component $\phi_{k+l} - \phi_k - \phi_l$ modulo 2π, from which we may hope to recover the individual phases ϕ_k.

This motivates the following class of method-of-moments procedures:

1. For each $k = 1, \ldots, K$, estimate r_k by

$$\hat{r}_k = \left(\frac{1}{N} \sum_{m=1}^N |\tilde{y}_k^{(m)}|^2 - 2\sigma^2\right)_+^{1/2}. \tag{16}$$

2. For each $(k, l) \in \mathcal{I}$, compute

$$\hat{B}_{k,l} = \frac{1}{N} \sum_{m=1}^N \tilde{y}_{k+l}^{(m)} \cdot \overline{\tilde{y}_k^{(m)}} \cdot \overline{\tilde{y}_l^{(m)}}, \tag{17}$$

and choose a version of its complex argument $\hat{\Phi}_{k,l}$ in $\mathbb{R}$ such that $\hat{\Phi}_{k,l} - \operatorname{Arg} \hat{B}_{k,l} = 0 \bmod 2\pi$.

3. Estimate $\phi = (\phi_k : k = 1, \ldots, K)$ by the least-squares estimator

$$\hat{\phi} = \operatorname*{arg\,min}_{\phi \in \mathbb{R}^K} \sum_{(k,l) \in \mathcal{I}} \left(\hat{\Phi}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l) \right)^2. \tag{18}$$

Then estimate θ by $\hat{\theta} = (\hat{r}_k \cos \hat{\phi}_k, \hat{r}_k \sin \hat{\phi}_k)_{k=1}^K$.

To enable a theoretical analysis, we have defined the least-squares procedure (18) in Step 3 using the usual squared difference $(\hat{\Phi}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l))^2$ over $\mathbb{R}$, rather than the squared distance $|\hat{\Phi}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}}^2$ over the circle $\mathcal{A}$. Hence the estimate will depend on the specific choice of argument $\hat{\Phi}_{k,l}$ in Step 2, which we have left ambiguous above.

We proceed by first studying in Section 4.1 an "oracle" version of this estimator, where $\hat{\Phi}_{k,l}$ is chosen in Step 2 using knowledge of the true phases $\phi_1, \ldots, \phi_K$ as the unique version of the argument of $\hat{B}_{k,l}$ for which $\hat{\Phi}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l) \in [-\pi, \pi)$. This choice satisfies an exact distributional symmetry in sign. We leverage this symmetry to provide a risk bound for this oracle procedure, which matches the minimax upper bound in the high-noise regime of Theorem 2.1.

To develop an actual estimator based on this oracle idea, we propose in Section 4.2 a method of mimicking this oracle using a pilot estimate of $\phi_1, \ldots, \phi_K$ that is obtained by first minimizing an ℓ_∞-type optimization objective. We prove an ℓ_∞-stability bound for bispectrum inversion, which implies that the resulting choice of $\hat{\Phi}_{k,l}$ coincides with the oracle choice with high probability as long as $N \gtrsim \frac{\sigma^6}{r^6} \log K$. Consequently, this estimator attains the same estimation rate without oracle knowledge.

4.1 The oracle procedure

Let us identify each entry of the true Fourier phase vector as a real value $\phi_k \in [-\pi, \pi)$, and set

$$\Phi_{k,l} = \phi_{k+l} - \phi_k - \phi_l \in \mathbb{R}. \tag{19}$$

We emphasize that this arithmetic is carried out in $\mathbb{R}$, not modulo 2π. We consider an oracle version of the above method-of-moments procedure, where $\hat{\Phi}_{k,l}^{\text{oracle}} \in [\Phi_{k,l} - \pi, \Phi_{k,l} + \pi)$ is chosen in Step 2 as the unique version of the complex argument of $\hat{B}_{k,l}$ that belongs to this range. Recalling the complex representation of θ in (12) and defining

$$B_{k,l} = \tilde{\theta}_{k+l} \cdot \overline{\tilde{\theta}_k} \cdot \overline{\tilde{\theta}_l} = r_{k+l} r_k r_l e^{i(\phi_{k+l} - \phi_k - \phi_l)} \in \mathbb{C}, \tag{20}$$

note that this means, for the principal argument specified in (11),

$$\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l} = \operatorname{Arg}(\hat{B}_{k,l} / B_{k,l}) \in [-\pi, \pi). \tag{21}$$

We will write $\hat{\Phi}^{\text{oracle}} = \hat{\Phi}^{\text{oracle}}(\phi)$ if we wish to make explicit the dependence of this definition on the phase vector ϕ of the true signal.

We denote by $\hat{\phi}^{\text{oracle}}$ the resulting least-squares estimate of ϕ in (18), and by $\hat{\theta}^{\text{oracle}}$ the corresponding estimate of θ. In the remainder of this subsection, we describe an argument showing the following squared-error risk bound for $\hat{\theta}^{\text{oracle}}$, deferring detailed proofs to Appendix A.

Lemma 4.1. *Suppose $\underline{c} r \leq r_k \leq \bar{c} r$ for each $k = 1, \ldots, K$. There exist constants $C, C_0 > 0$ depending only on $\underline{c}, \bar{c}$ such that if $\frac{Nr^6}{\sigma^6} \wedge \frac{N^{2/3} r^2}{\sigma^2} \geq C_0 \log K$, then*

$$\mathbb{E}[L(\hat{\theta}^{\text{oracle}}, \theta^*)] \leq C \left(\frac{K\sigma^2}{N} + \frac{K\sigma^4}{Nr^2} + \frac{\sigma^6}{Nr^4} \right). \tag{22}$$

We divide the argument for Lemma 4.1 into the analysis of Step 1 of the MoM procedure for estimating the Fourier magnitudes $\{r_k\}_{k=1}^K$, Step 2 for estimating the bispectrum components $\{\Phi_{k,l}\}_{(k,l) \in \mathcal{I}}$, and Step 3 for recovering the phases $\{\phi_k\}_{k=1}^K$ from the bispectrum.

Estimating r_k. Standard Gaussian and chi-squared tail bounds show the following guarantee for estimating the Fourier magnitudes r_k via $\hat{r}_k$, defined in (16).

Lemma 4.2. *For each $k = 1, \dots, K$ and a universal constant $c > 0$,*

$$\mathbb{P}[\hat{r}_k \geq r_k(1+s)] \leq 2\exp\left(-cNs^2\left(\frac{r_k^2}{\sigma^2} \wedge \frac{r_k^4}{\sigma^4}\right)\right) \text{ for all } s \geq 0, \tag{23}$$

$$\mathbb{P}[\hat{r}_k \leq r_k(1-s)] \leq 2\exp\left(-cNs^2\left(\frac{r_k^2}{\sigma^2} \wedge \frac{r_k^4}{\sigma^4}\right)\right) \text{ for all } s \in [0,1). \tag{24}$$

Integrating these tail bounds yields the following immediate corollary.

Corollary 4.3. *For each $k = 1, \dots, K$ and a universal constant $C > 0$,*

$$\mathbb{E}[(\hat{r}_k - r_k)^2] \leq C\left(\frac{\sigma^2}{N} + \frac{\sigma^4}{Nr_k^2}\right).$$

Estimating $\Phi_{k,l}$. Applying a concentration inequality for cubic polynomials in independent Gaussian random variables, derived from Latala (2006), we obtain the following tail bounds for estimating $B_{k,l}$ by $\hat{B}_{k,l}$ in Step 2, and for estimating the bispectrum component $\Phi_{k,l}$ by the oracle estimator $\hat{\Phi}_{k,l}^{\text{oracle}}$.

Lemma 4.4. *Consider any $(k,l) \in \mathcal{I}$ and suppose $r_{k+l}, r_k, r_l \geq \underline{r}$. Then for universal constants $C, c > 0$ and any $s > 0$,*

$$\mathbb{P}\left[|\hat{B}_{k,l}/B_{k,l} - 1| \geq s\right] \leq C\exp\left(-c\left(\frac{Ns^2\underline{r}^2}{\sigma^2} \wedge \frac{Ns^2\underline{r}^6}{\sigma^6} \wedge \frac{(Ns)^{2/3}\underline{r}^2}{\sigma^2}\right)\right). \tag{25}$$

Furthermore, for universal constants $C, c > 0$ and any $s \in (0, \pi/2)$,

$$\mathbb{P}\left[|\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l}| \geq s\right] \leq C\exp\left(-c\left(\frac{Ns^2\underline{r}^2}{\sigma^2} \wedge \frac{Ns^2\underline{r}^6}{\sigma^6} \wedge \frac{(Ns)^{2/3}\underline{r}^2}{\sigma^2}\right)\right). \tag{26}$$

Corollary 4.5. *Consider any $(k,l) \in \mathcal{I}$ and suppose $r_{k+l}, r_k, r_k \geq \underline{r}$. Then for a universal constant $C > 0$,*

$$\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})^2] \leq C\left(\frac{\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6}\right)$$

A key property of the oracle estimator $\hat{\Phi}_{k,l}^{\text{oracle}}$ is an exact distributional symmetry in sign,

$$\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l} \overset{L}{=} -\hat{\Phi}_{k,l}^{\text{oracle}} + \Phi_{k,l}. \tag{27}$$

This implies, for example, that $\mathbb{E}[\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l}] = 0$ so that $\hat{\Phi}^{\text{oracle}}$ is an unbiased estimate of Φ. For bispectral components $\Phi_{k,l}$ and $\Phi_{x,y}$ that have an overlapping index, the corresponding estimates $\hat{\Phi}_{k,l}^{\text{oracle}}$ and $\hat{\Phi}_{x,y}^{\text{oracle}}$ are not independent. Our proof of Lemma 4.1 requires a sharper bound on the expected product of their errors than what is naively obtained from the preceding Corollary 4.5 and Cauchy-Schwarz. Part (b) of the following key lemma establishes this bound using the representation (21), a Taylor expansion of the function $\operatorname{Arg} z = \operatorname{Im} \operatorname{Ln} z$ around $z = 1$, and analysis of high-order moments of complex Gaussian variables.

Lemma 4.6. *Let $(k,l), (x,y) \in \mathcal{I}$, and suppose $r_k, r_l, r_{k+l}, r_x, r_y, r_{x+y} \geq \underline{r}$. For some universal constants $C, c > 0$,*

(a) If $\{k, l, k+l\}$ is disjoint from $\{x, y, x+y\}$, then

$$\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})] = 0.$$

(b) If $\{k, l, k+l\} \cap \{x, y, x+y\}$ has cardinality 1, then

$$\left|\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})]\right| \leq C\left(\frac{\sigma^2}{N\underline{r}^2} + e^{-c\left(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2}\right)}\right). \tag{28}$$

(c) *For any* $(k,l),(x,y) \in \mathcal{I}$,

$$\left|\mathbb{E}[(\hat{\Phi}^{\text{oracle}}_{k,l} - \Phi_{k,l})(\hat{\Phi}^{\text{oracle}}_{x,y} - \Phi_{x,y})]\right| \leq C\left(\frac{\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6}\right).$$

Estimating ϕ_k**.** We now translate the preceding bounds for estimating the Fourier bispectrum $\{\Phi_{k,l}\}$ to estimating the phases $\{\phi_k\}$ using the least squares procedure (18).

Define the matrix $M \in \mathbb{R}^{\mathcal{I}\times K}$ with rows indexed by the bispectrum index set $\mathcal{I}$ from (15), such that the linear system (19) may be expressed as $\Phi = M\phi$. That is, row (k,l) of M is given by $e_{k+l} - e_k - e_l$ where $e_k \in \mathbb{R}^K$ is the k^{th} standard basis vector. Then (18) is given explicitly by

$$\hat{\phi} = M^\dagger \hat{\Phi}$$

where $M^\dagger$ is the Moore-Penrose pseudo-inverse.

Recall that a rotation of the circular domain of f induces the map (4), which does not change the bispectral components $\Phi_{k,l}$. This is reflected by the property that $(1,2,3,\ldots,K)$ belongs to the kernel of M. The following lemma shows that this is the unique vector in the kernel. Furthermore, M is well-conditioned on the subspace orthogonal to this kernel, with all remaining $K-1$ singular values on the same order of $\sqrt{K}$.

Lemma 4.7. *M has rank exactly $K-1$, and the kernel of M is the span of $(1,2,3,\ldots,K) \in \mathbb{R}^K$. All $K-1$ non-zero eigenvalues of $M^\top M \in \mathbb{R}^{K\times K}$ are integers in the interval $[K+1, 2K+1]$.*

This yields the following corollary for estimation of the Fourier phases $\{\phi_k\}$, up to a global rotation that is represented by an additive shift in the direction of $(1,2,3,\ldots,K)$.

Corollary 4.8. *Suppose $r_k \geq \underline{r}$ for each $k = 1,\ldots,K$. Then for universal constants $C, c > 0$,*

$$\mathbb{E}\left[\inf_{\alpha\in\mathbb{R}} \sum_{k=1}^K |\hat{\phi}^{\text{oracle}}_k - \phi_k + k\alpha|^2_{\mathcal{A}}\right] \leq C\left(\frac{K\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6} + Ke^{-c(\frac{N\underline{r}^2}{\sigma^2}\wedge\frac{N\underline{r}^6}{\sigma^6}\wedge\frac{N^{2/3}\underline{r}^2}{\sigma^2})}\right). \tag{29}$$

Proof. By adding a multiple of $(1,2,3,\ldots,K)$ to ϕ and absorbing this shift into α, we may assume without loss of generality that ϕ is orthogonal to $(1,2,3,\ldots,K)$. Under this assumption, we will then upper-bound the left side by choosing $\alpha = 0$. Since $\Phi = M\phi$, this implies $M^\dagger\Phi = M^\dagger M\phi = \phi$, the last equality holding because Lemma 4.7 implies that $M^\dagger M$ is the projection orthogonal to $(1,2,3,\ldots,K)$. Applying $\operatorname{Tr} AB \leq \operatorname{Tr} B \cdot \|A\|_{\text{op}}$ for positive semidefinite A, B, where $\|\cdot\|_{\text{op}}$ is the $\ell_2 \to \ell_2$ operator norm,

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^K |\hat{\phi}^{\text{oracle}}_k - \phi_k|^2_{\mathcal{A}}\right] &\leq \mathbb{E}[\|\hat{\phi}^{\text{oracle}} - \phi\|^2] = \mathbb{E}[\|M^\dagger\hat{\Phi}^{\text{oracle}} - M^\dagger\Phi\|^2] \\
&= \mathbb{E}[(\hat{\Phi}^{\text{oracle}} - \Phi)^\top M^{\dagger\top} M^\dagger (\hat{\Phi}^{\text{oracle}} - \Phi)] \\
&= \operatorname{Tr}\left(\mathbb{E}[(\hat{\Phi}^{\text{oracle}} - \Phi)(\hat{\Phi}^{\text{oracle}} - \Phi)^\top] M^{\dagger\top} M^\dagger\right) \\
&\leq \operatorname{Tr}(M^{\dagger\top}M^\dagger) \cdot \|\mathbb{E}[(\hat{\Phi}^{\text{oracle}} - \Phi)(\hat{\Phi}^{\text{oracle}} - \Phi)^\top]\|_{\text{op}}.
\end{aligned}$$

Note that $\operatorname{Tr} M^{\dagger\top}M^\dagger = \operatorname{Tr} M^\dagger M^{\dagger\top}$, where $M^\dagger M^{\dagger\top} = (M^\top M)^\dagger \in \mathbb{R}^{K\times K}$. Lemma 4.7 implies that $M^\dagger M^{\dagger\top}$ has 1 eigenvalue of 0 and $K-1$ eigenvalues belonging to $[1/(2K+1), 1/(K+1)]$. Thus

$$\operatorname{Tr} M^{\dagger\top} M^\dagger \leq \frac{K-1}{K+1} < 1. \tag{30}$$

We have $\|A\|_{\text{op}} \leq \|A\|_\infty$ for positive semidefinite A, where $\|A\|_\infty$ is the $\ell_\infty \to \ell_\infty$ operator norm given by the maximum absolute row sum. For a universal constant $C > 0$ and each $(k,l) \in \mathcal{I}$, there are at most C pairs $(x,y) \in \mathcal{I}$ for which $\{k, l, k+l\} \cap \{x, y, x+y\}$ has cardinality 2 or 3, and at most CK pairs $(x,y) \in \mathcal{I}$ for which $\{k, l, k+l\} \cap \{x, y, x+y\}$ has cardinality 1. Applying Lemma 4.6(b) for those pairs for which this cardinality is 1, Lemma 4.6(c) for those pairs for which this cardinality is 2 or 3, and Lemma 4.6(a) for all remaining pairs, we obtain for different universal constants $C, c > 0$ that

$$\|\mathbb{E}[(\hat{\Phi}^{\text{oracle}} - \Phi)(\hat{\Phi}^{\text{oracle}} - \Phi)^\top]\|_\infty \leq C\left(\frac{K\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6} + Ke^{-c(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2})}\right).$$

Combining this with (30) yields the lemma. □

Let us remark that using Lemma 4.6(b) in place of Lemma 4.6(c) for the pairs where $\{k, l, k+l\}$ and $\{x, y, x+y\}$ overlap in one index is important for removing a factor of K in the $\sigma^6/(N\underline{r}^6)$ component of the error, which will be the leading contribution to the overall estimation error in the high-noise regime.

Lemma 4.1 now follows from the loss upper bound in Proposition 3.1 in terms of the separate estimation errors for magnitude and phase, together with Corollaries 4.3 and 4.8.

4.2 Mimicking the oracle

We now consider the method-of-moments procedure where the choice of $\hat{\Phi}_{k,l}$ in Step 2 is determined instead by the following method: Compute a "pilot" estimate of ϕ as any minimizer of the ℓ_∞-type objective

$$\tilde{\phi} = \operatorname*{arg\,min}_{\phi \in \mathcal{A}^K} \max_{(k,l) \in \mathcal{I}} |\operatorname{Arg} \hat{B}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}}, \tag{31}$$

where a minimizer exists because $\mathcal{A}$ is compact under $|\cdot|_{\mathcal{A}}$. Identify each entry $\tilde{\phi}_k \in [-\pi, \pi)$ of this estimate as a real value, and set $\tilde{\Phi}_{k,l} = \tilde{\phi}_{k+l} - \tilde{\phi}_k - \tilde{\phi}_l$ where arithmetic is again carried out in $\mathbb{R}$, not modulo 2π. Then choose $\hat{\Phi}^{\text{opt}}_{k,l} \in [\tilde{\Phi}_{k,l} - \pi, \tilde{\Phi}_{k,l} + \pi)$ as the unique version of the complex argument of $\hat{B}_{k,l}$ belonging to this range. Let $\hat{\phi}^{\text{opt}}$ be the resulting least-squares estimate of ϕ in (18), and let $\hat{\theta}^{\text{opt}}$ be the corresponding estimate of θ.

Unlike $\hat{\theta}^{\text{oracle}}$, here $\hat{\theta}^{\text{opt}}$ constitutes a true estimator of θ that does not require oracle knowledge. Note that to obtain an estimation guarantee that is rate-optimal in squared-error loss, we do not use the pilot estimate $\tilde{\phi}$ as the final estimator, but only to resolve the phase ambiguity of the estimated bispectrum before applying the least-squares procedure for bispectrum inversion in Step 3.

The guarantee for this method is summarized in the following lemma.

Lemma 4.9. *In the setting of Lemma 4.1, the risk bound (22) holds also with $\hat{\theta}^{\text{opt}}$ in place of $\hat{\theta}^{\text{oracle}}$.*

We prove Lemma 4.9 by showing that, with high probability, $\hat{\Phi}^{\text{opt}} = \hat{\Phi}^{\text{oracle}}(\phi')$ for some phase vector ϕ' that is equivalent to ϕ. Then on this event, $\hat{\Phi}^{\text{opt}}$ achieves the same loss as the oracle estimator. The main additional ingredient in the proof is a deterministic ℓ_∞-stability bound for recovery of the Fourier phases from the bispectrum, stated in the following result.

Lemma 4.10. *Fix any $\delta \in (0, \pi/3)$ and $\phi, \phi' \in \mathbb{R}^K$. Denote $\Phi_{k,l} = \phi_{k+l} - \phi_k - \phi_l$ and $\Phi'_{k,l} = \phi'_{k+l} - \phi'_k - \phi'_l$. If*

$$|\Phi_{k,l} - \Phi'_{k,l}|_{\mathcal{A}} \leq \delta \text{ for all } (k,l) \in \mathcal{I},$$

then there exists some $\alpha \in \mathbb{R}$ such that

$$|\phi_k - \phi'_k - k\alpha|_{\mathcal{A}} \leq \delta \text{ for all } k = 1, \ldots, K.$$

This guarantees that, if $\tilde{\phi}$ yields a bispectrum $\tilde{\Phi}$ which is elementwise close to the true bispectrum Φ in the circular distance modulo 2π, then $\tilde{\phi}$ must also be elementwise close to ϕ up to a rotation of the circular domain.

Let us call $\phi' \in \mathbb{R}^K$ equivalent to $\phi \in \mathbb{R}^K$ if there exists $\alpha \in \mathbb{R}$ for which

$$|\phi'_k - \phi_k + k\alpha|_{\mathcal{A}} = 0 \text{ for each } k = 1, \ldots, K. \tag{32}$$

This means that ϕ and ϕ' represent the same Fourier phases up to rotation of the circular domain. The above guarantee is sufficient to show that if each quantity $\operatorname{Arg} \hat{B}_{k,l}$ estimates the true bispectral component $\Phi_{k,l}$ up to a small constant error in the circular distance $|\cdot|_{\mathcal{A}}$, then its version $\hat{\Phi}^{\text{opt}}_{k,l}$ that is chosen using $\tilde{\phi}$ must coincide exactly with the oracle choice $\hat{\Phi}^{\text{oracle}}_{k,l}(\phi')$, based on a phase vector ϕ' that is equivalent to the true phase vector ϕ.

Corollary 4.11. *Let $\hat{B}_{k,l}$ be as defined in (17), and suppose $\phi \in \mathbb{R}^K$ is such that*

$$|\operatorname{Arg}\hat{B}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}} < \pi/12 \text{ for every } (k,l) \in \mathcal{I}. \tag{33}$$

Then there exists ϕ' equivalent to ϕ such that $\hat{\Phi}^{\mathrm{opt}} = \hat{\Phi}^{\mathrm{oracle}}(\phi')$.

Proof. By the definition of the optimization procedure which defines $\tilde{\phi}$ in (31),

$$\max_{(k,l)\in\mathcal{I}} |\operatorname{Arg}\hat{B}_{k,l} - (\tilde{\phi}_{k+l} - \tilde{\phi}_k - \tilde{\phi}_l)|_{\mathcal{A}} \le \max_{(k,l)\in\mathcal{I}} |\operatorname{Arg}\hat{B}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}}. \tag{34}$$

By assumption, the right side is at most $\pi/12$. Then by the triangle inequality for $|\cdot|_{\mathcal{A}}$, for every $(k,l) \in \mathcal{I}$, we have $|(\tilde{\phi}_{k+l} - \tilde{\phi}_k - \tilde{\phi}_l) - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}} < \pi/6$. Applying Lemma 4.10, we obtain for some $\alpha \in \mathbb{R}$ and all $k = 1, \ldots, K$ that $|\tilde{\phi}_k - \phi_k - k\alpha|_{\mathcal{A}} < \pi/6$. This means that there exists ϕ' equivalent to ϕ for which, for the usual absolute value,

$$|\tilde{\phi}_k - \phi'_k| < \pi/6 \text{ for all } k = 1, \ldots, K.$$

Then denoting $\Phi'_{k,l} = \phi'_{k+l} - \phi'_k - \phi'_l$, by the triangle inequality, $|\tilde{\Phi}_{k,l} - \Phi'_{k,l}| < \pi/2$ for all $(k,l) \in \mathcal{I}$. Since ϕ' is equivalent to ϕ, also

$$|\operatorname{Arg}\hat{B}_{k,l} - (\phi'_{k+l} - \phi'_k - \phi'_l)|_{\mathcal{A}} = |\operatorname{Arg}\hat{B}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}} < \pi/12.$$

So by the definition of $\hat{\Phi}^{\mathrm{oracle}}(\phi')$, we have $|\hat{\Phi}^{\mathrm{oracle}}_{k,l}(\phi') - \Phi'_{k,l}| < \pi/12$ for the usual absolute value. Then $|\hat{\Phi}^{\mathrm{oracle}}_{k,l}(\phi') - \tilde{\Phi}_{k,l}| < \pi/2 + \pi/12 < \pi$ for all $(k,l) \in \mathcal{I}$, meaning that $\hat{\Phi}^{\mathrm{oracle}}(\phi') = \hat{\Phi}^{\mathrm{opt}}$. □

The tail bounds of Lemma 4.4 may be used to show that the event (33) holds with high probability. On this event, the loss of $\hat{\theta}^{\mathrm{opt}}$ matches exactly that of $\hat{\theta}^{\mathrm{oracle}}$, which is controlled by Lemma 4.1. Combining with a crude bound for the loss on the complementary event, which has exponentially small probability in N, we obtain Lemma 4.9.

Finally, let us check that this estimation guarantee in Lemma 4.9 coincides with our stated minimax rate in Theorem 2.1 when restricted to the high-noise regime.

Proof of Theorem 2.1, upper bound. Let $\underline{r} = \underline{c}r$. When $\sigma^2 \ge c_0 K r^2$ and $K \ge 2$, there is a constant $c > 0$ for which

$$\frac{\sigma^6}{\underline{r}^6}\log K \ge \frac{c\sigma^3}{\underline{r}^3}(\log K)^{3/2}.$$

Then the required condition $\frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2} \ge C_0 \log K$ in Lemma 4.9 is implied by $N \ge C_0'(\sigma^6/r^6)\log K$ for a sufficiently large constant $C_0' > 0$. The upper bound (22) is also at most $C'\sigma^6/Nr^4$ in this regime of $\sigma^2 \ge c_0 K r^2$, for a constant $C' > 0$, and this yields the minimax upper bound of Theorem 2.1. □

We remark that Lemma 4.9 gives an estimation guarantee not just in the high-noise regime, but for any noise level σ^2. In a regime of *very* low noise $\sigma^2 \le r^2$, it also implies the upper bound of Theorem 2.2.

Proof of Theorem 2.2, upper bound, for $\sigma^2 \le r^2$. Let $\underline{r} = \underline{c}r$. When $\sigma^2 \le r^2$ and $K \ge 2$, the required condition $\frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2} \ge C_0 \log K$ is implied by $N \ge C_0' \frac{K\sigma^2}{r^2}\log K$ for a sufficiently large constant $C_0' > 0$. The upper bound (22) is also at most $C'K\sigma^2/N$ in this regime, and this yields the minimax upper bound of Theorem 2.2. □

In high dimensions K and the noise regime $r^2 \ll \sigma^2 \ll Kr^2/\log K$, (22) exhibits the rate $K\sigma^4/Nr^2$, which is larger than the minimax rate $K\sigma^2/N$. This rate arises from estimating the Fourier magnitudes $\{r_k\}$ without using phase information. In this regime, the above method-of-moments procedure becomes suboptimal. We will instead analyze in Section 5 the maximum likelihood estimator, to establish the minimax rate over the entire low-noise regime described by Theorem 2.2.

Remark 4.12. This proof of the minimax upper bounds is information-theoretic in nature, in that the pilot estimate used to mimic the oracle may require exponential time in K to compute. We describe in Appendix A.5 an alternative "frequency marching" method, as discussed also in (Bendory et al., 2017, Section IV), which provides a computationally efficient alternative to mimic the oracle at the expense of a larger requirement for the sample size N.

This method sets

$$\tilde{\phi}_1 = 0$$

and, for each $k = 2, \ldots, K$, sets

$$\tilde{\phi}_k = \operatorname{Arg} \hat{B}_{1,k-1} + \tilde{\phi}_{k-1} \bmod 2\pi$$

to define a pilot estimator $\tilde{\phi}$ for ϕ. We show that, resolving the phase ambiguity of $\hat{\Phi}$ using this pilot estimate and then re-estimating $\hat{\phi}$ by least squares, the resulting procedure achieves the same risk as described in Lemmas 4.1 and 4.9 under a requirement for N that is larger by a factor of K^2.

5 Maximum likelihood estimator

The method-of-moments procedure analyzed in the preceding section is not rate-optimal over the full low-noise regime described by Theorem 2.2. Motivated by this observation, and by the more common use of likelihood-based approaches in practice (Sigworth, 1998; Scheres, 2012), in this section we analyze the maximum likelihood estimator (MLE) in the setting of Theorem 2.2.

Define the log-likelihood function

$$l(\theta, y) = \log p_\theta(y) := \log \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{2K} \exp\left(-\frac{\|y - g(\alpha) \cdot \theta\|^2}{2\sigma^2} \right) d\alpha \right] \tag{35}$$

where $p_\theta(y)$ denotes the Gaussian mixture density that marginalizes over the unknown rotation. Then the MLE is given by

$$\hat{\theta}^{\mathrm{MLE}} = \arg\min_{\theta \in \mathbb{R}^{2K}} R_N(\theta), \qquad R_N(\theta) = -\frac{1}{N} \sum_{m=1}^{N} l(\theta, y^{(m)}),$$

where $R_N(\theta)$ denotes the negative empirical log-likelihood.

The main result of this section is the following risk bound for $\hat{\theta}^{\mathrm{MLE}}$ in the low-noise setting of Theorem 2.2.

Theorem 5.1. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for some $\underline{c}, \bar{c} > 0$ and every $k = 1, \ldots, K$. There exist constants $C, C_0, C_1 > 0$ depending only on $\underline{c}, \bar{c}$ such that if $\sigma^2 \leq \frac{Kr^2}{C_1 \log K}$ and $N \geq C_0 K (1 + \frac{\sigma^2}{r^2}) \log(K + \frac{Kr^2}{\sigma^2})$, then*

$$\mathbb{E}_{\theta^*}[L(\hat{\theta}^{\mathrm{MLE}}, \theta^*)] \leq \frac{CK\sigma^2}{N}.$$

For $\sigma^2 \geq r^2$, this requirement for N reduces to that of Theorem 2.2, up to a modified constant $C_0 > 0$. Combined with the argument for $\sigma^2 \leq r^2$ in Section 4.2, this proves the minimax upper bound of Theorem 2.2.

In the remainder of this section, we prove Theorem 5.1. The proof applies a classical idea of second-order Taylor expansion for the log-likelihood function. Observe first that the negative log-likelihood $R_N(\theta)$ satisfies the rotational invariance $R_N(\theta) = R_N(g(\alpha) \cdot \theta)$ for all $\alpha \in \mathcal{A}$. Thus $\hat{\theta}^{\mathrm{MLE}}$ is defined only up to rotation, and all rotations of $\hat{\theta}^{\mathrm{MLE}}$ incur the same loss. To fix this rotation and ease notation in the analysis, let us denote by $\hat{\theta}^{\mathrm{MLE}}$ the rotation of the MLE such that

$$\|\hat{\theta}^{\mathrm{MLE}} - \theta^*\|^2 = \min_{\alpha \in \mathcal{A}} \|g(\alpha) \cdot \hat{\theta}^{\mathrm{MLE}} - \theta^*\|^2 = L(\hat{\theta}^{\mathrm{MLE}}, \theta^*), \tag{36}$$

where θ^* is the true parameter. Since $\hat{\theta}^{\mathrm{MLE}}$ minimizes $R_N(\theta)$, we have $0 \geq R_N(\hat{\theta}^{\mathrm{MLE}}) - R_N(\theta^*)$. Then Taylor expansion (for this rotation of $\hat{\theta}^{\mathrm{MLE}}$ that satisfies (36)) gives

$$0 \geq R_N(\hat{\theta}^{\mathrm{MLE}}) - R_N(\theta^*)$$

$$= \nabla R_N(\theta^*)^\top(\hat{\theta}^{\text{MLE}} - \theta^*) + \frac{1}{2}(\hat{\theta}^{\text{MLE}} - \theta^*)^\top \nabla^2 R_N(\tilde{\theta})(\hat{\theta}^{\text{MLE}} - \theta^*) \tag{37}$$

where $\tilde{\theta} \in \mathbb{R}^{2K}$ is on the line segment between θ^* and $\hat{\theta}^{\text{MLE}}$. Heuristically, Theorem 5.1 will follow from the bounds

$$\left|\nabla R_N(\theta^*)^\top(\hat{\theta}^{\text{MLE}} - \theta^*)\right| \lesssim \sqrt{\frac{K}{N\sigma^2}} \cdot \|\hat{\theta}^{\text{MLE}} - \theta^*\|, \tag{38}$$

$$(\hat{\theta}^{\text{MLE}} - \theta^*)^\top \nabla^2 R_N(\tilde{\theta})(\hat{\theta}^{\text{MLE}} - \theta^*) \gtrsim \frac{1}{\sigma^2} \cdot \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2. \tag{39}$$

Applying these to (37) and rearranging yields the desired result $\|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \lesssim K\sigma^2/N$.

The bulk of the proof lies in establishing an appropriate version of (39). This requires a delicate argument for large K, as naive uniform concentration and Lipschitz bounds for $\nabla^2 R_N(\theta) \in \mathbb{R}^{2K\times 2K}$ fail to establish (39) in the full ranges of σ^2 and N that are specified by Theorem 5.1. In the remainder of this section, we describe the components of this argument, deferring detailed proofs to Appendix B.

5.1 Gradient and Hessian of the log-likelihood

To simply the model, observe that each sample $y^{(m)}$ satisfies the equality in law

$$y^{(m)} = g(\alpha^{(m)}) \cdot \theta^* + \sigma\varepsilon^{(m)} \overset{L}{=} g(\alpha^{(m)}) \cdot (\theta^* + \sigma\varepsilon^{(m)}).$$

Furthermore, $g(\alpha^{(m)})^{-1}g(\alpha) = g(\alpha - \alpha^{(m)})$ where, if $\alpha \sim \text{Unif}([-\pi, \pi))$ is a uniformly random rotation, then $\alpha - \alpha^{(m)}$ is also uniformly random for any fixed $\alpha^{(m)}$. Applying these observations to the form (35) of the log-likelihood function, we obtain the equality in law for the negative log-likelihood process

$$\left\{R_N(\theta) : \theta \in \mathbb{R}^{2K}\right\} \overset{L}{=} \left\{-\frac{1}{N}\sum_{m=1}^N l(\theta,\, \theta^* + \sigma\varepsilon^{(m)}) : \theta \in \mathbb{R}^{2K}\right\}. \tag{40}$$

That is to say, having defined the log-likelihood function to marginalize over a uniformly random latent rotation, the distribution of $\{R_N(\theta) : \theta \in \mathbb{R}^{2K}\}$ is the same under the model $y^{(m)} = g(\alpha^{(m)})\cdot\theta^* + \sigma\varepsilon^{(m)} \sim p_{\theta^*}$ as under a model $y^{(m)} = \theta^* + \sigma\varepsilon^{(m)}$ without latent rotations. Thus, in the analysis, we will henceforth assume the simpler model

$$y^{(m)} = \theta^* + \sigma\varepsilon^{(m)} \text{ for } m = 1, \ldots, N, \qquad \varepsilon^{(1)}, \ldots, \varepsilon^{(N)} \overset{\text{iid}}{\sim} \mathcal{N}(0, I_{2K}). \tag{41}$$

Under this model (41), expanding the square in the exponent of (35), $R_N(\theta)$ may be written as

$$\begin{aligned} R_N(\theta) = \frac{1}{N}\sum_{m=1}^N K\log 2\pi\sigma^2 + \frac{\|\theta\|^2}{2\sigma^2} + \frac{\|\theta^* + \sigma\varepsilon^{(m)}\|^2}{2\sigma^2} \\ - \log\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(\frac{\langle\theta^* + \sigma\varepsilon^{(m)}, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) d\alpha\right]. \end{aligned} \tag{42}$$

Given $\theta, \varepsilon \in \mathbb{R}^{2K}$, define $\mathcal{P}_{\theta,\varepsilon}$ to be the tilted probability law over angles $\alpha \in \mathcal{A}$ with density

$$\frac{d\mathcal{P}_{\theta,\varepsilon}(\alpha)}{d\alpha} = \exp\left(\frac{\langle\theta^* + \sigma\varepsilon,\, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) \bigg/ \int_{-\pi}^{\pi} \exp\left(\frac{\langle\theta^* + \sigma\varepsilon,\, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) d\alpha. \tag{43}$$

Then direct computation shows that the gradient and Hessian of $R_N(\theta)$ take the forms

$$\nabla R_N(\theta) = \frac{\theta}{\sigma^2} - \frac{1}{N}\sum_{m=1}^N \frac{1}{\sigma^2}\mathbb{E}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}\left[g(\alpha)^{-1}(\theta^* + \sigma\varepsilon^{(m)})\right] \tag{44}$$

$$\nabla^2 R_N(\theta) = \frac{1}{\sigma^2}I - \frac{1}{N}\sum_{m=1}^N \frac{1}{\sigma^4}\text{Cov}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}\left[g(\alpha)^{-1}(\theta^* + \sigma\varepsilon^{(m)})\right] \tag{45}$$

where the expectation and covariance are over the random rotation $\alpha \sim \mathcal{P}_{\theta,\varepsilon^{(m)}}$ (conditional on $\varepsilon^{(m)}$) following the above law.

5.2 Tail bound

As a first step of the proof, we fix a small constant $\delta_1 \in (0,1)$ to be determined based on the constants $\underline{c}, \bar{c} > 0$ of Theorem 5.1, and define the domain

$$\mathcal{B}(\delta_1) = \{\theta : \|\theta - \theta^*\| \leq \delta_1 \|\theta^*\|\} \subset \mathbb{R}^{2K}. \tag{46}$$

We establish the following lemma, which shows that $\hat{\theta}^{\text{MLE}}$ belongs to this domain $\mathcal{B}(\delta_1)$ with high probability, and provides also an upper bound for the fourth moment of $\hat{\theta}^{\text{MLE}}$.

Lemma 5.2. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for some $\underline{c}, \bar{c} > 0$ and each $k = 1, \ldots, K$. Fix any constant $\delta_1 > 0$, and define $\mathcal{B}(\delta_1)$ by (46). Then there exist constants $C_0, C_1, C', c' > 0$ depending only on $\underline{c}, \bar{c}, \delta_1$ such that if $\sigma^2 \leq \frac{Kr^2}{C_1 \log K}$ and $N \geq C_0 K$, then*

$$\mathbb{P}\left[\hat{\theta}^{\text{MLE}} \in \mathcal{B}(\delta_1)\right] \geq 1 - e^{-c' N (\log K)^2 / K}, \tag{47}$$

$$\mathbb{E}[\|\hat{\theta}^{\text{MLE}}\|^4] \leq C' \|\theta^*\|^4. \tag{48}$$

To show this lemma, define the population negative log-likelihood $R(\theta) = \mathbb{E}_{\theta^*}[R_N(\theta)]$, where the equality in law (40) allows us to evaluate the expectation under the simplified model (41). Then the KL-divergence between p_{θ^*} and p_θ is given by

$$D_{\text{KL}}(p_{\theta^*} \| p_\theta) = R(\theta) - R(\theta^*) = \mathbb{E}_{\theta^*}[R_N(\theta)] - \mathbb{E}_{\theta^*}[R_N(\theta^*)]. \tag{49}$$

Recalling the form (42) for the negative log-likelihood $R_N(\theta)$, we have

$$D_{\text{KL}}(p_{\theta^*} \| p_\theta) = \frac{\|\theta\|^2 - \|\theta^*\|^2}{2\sigma^2} + \text{I} - \text{II} \tag{50}$$

where

$$\text{I} = \mathbb{E} \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{\langle \theta^* + \sigma\varepsilon, g(\alpha) \cdot \theta^* \rangle}{\sigma^2}\right) d\alpha$$
$$\text{II} = \mathbb{E} \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{\langle \theta^* + \sigma\varepsilon, g(\alpha) \cdot \theta \rangle}{\sigma^2}\right) d\alpha$$

and both expectations are over $\varepsilon \sim \mathcal{N}(0, I_{2K})$.

We may apply a quadratic Taylor expansion of $\langle \theta^*,\, g(\alpha) \cdot \theta^* \rangle = \sum_k r_k(\theta^*)^2 \cos k\alpha$ around $\alpha = 0$, to write

$$\langle \theta^*,\, g(\alpha) \cdot \theta^* \rangle - \|\theta^*\|^2 \approx -\sum_{k=1}^{K} r_k(\theta^*)^2 \cdot \frac{k^2 \alpha^2}{2} \asymp -K^3 r^2 \alpha^2. \tag{51}$$

Then $\int \exp(\theta^*, g(\alpha) \cdot \theta^* / \sigma^2)\, d\alpha$ in I may be approximated by a Gaussian integral over $\alpha \in \mathbb{R}$. Upper bounding II by the supremum over α, and applying a standard covering net argument to control the suprema of the Gaussian processes $\langle \varepsilon,\, g(\alpha) \cdot \theta^* \rangle$ and $\langle \varepsilon,\, g(\alpha) \cdot \theta \rangle$, we obtain the following lower bound on the KL-divergence.

Lemma 5.3. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for each $k = 1, \ldots, K$, and $\sigma^2 \leq Kr^2$. Then there are constants $C_2, C_3 > 0$ depending only on $\underline{c}, \bar{c}$ such that for any $\theta \in \mathbb{R}^{2K}$,*

$$D_{\text{KL}}(p_{\theta^*} \| p_\theta) \geq \frac{\min_{\alpha \in \mathcal{A}} \|\theta^* - g(\alpha) \cdot \theta\|^2}{2\sigma^2} - \frac{1}{2} \log\left(\frac{C_2 K^3 r^2}{\sigma^2}\right) - \frac{C_3(\|\theta^*\| + \|\theta\|)}{\sigma} \cdot \sqrt{\log K}.$$

Comparing this with the rate of uniform concentration of the negative log-likelihood $R_N(\theta)$ around its mean $R(\theta)$ (cf. Lemma B.3), we obtain an exponential tail bound for the probability of the event

$$\|\theta^* - \hat{\theta}^{\text{MLE}}\| \in \left[n\delta_1 \|\theta^*\|, (n+1)\delta_1 \|\theta^*\|\right]$$

for each integer $n \geq 1$. Summing this bound over all $n \geq 1$ yields Lemma 5.2.

5.3 Lower bound for the information matrix

In light of Lemma 5.2, to show (39) with high probability, it suffices to establish a version of the lower bound

$$\nabla^2 R_N(\theta) \gtrsim \frac{1}{\sigma^2} \cdot I \qquad \text{uniformly over } \theta \in \mathcal{B}(\delta_1). \tag{52}$$

Denote the tangent vector to the rotational orbit $\{g(\alpha) \cdot \theta^* : \alpha \in \mathcal{A}\}$ at θ^* by

$$u^* = \frac{d}{d\alpha} g(\alpha) \cdot \theta^* \Big|_{\alpha=0} = g'(0) \cdot \theta^*. \tag{53}$$

From the rotational invariance of $R(\theta)$, it is easy to see that the expected (Fisher) information matrix $\mathbb{E}[\nabla^2 R_N(\theta^*)] = \nabla^2 R(\theta^*)$ must be singular, with u^* belonging to its kernel. Thus we cannot expect the bound (52) to hold in all directions of $\mathbb{R}^{2K}$, but only in those directions orthogonal to u^*. This will suffice to show (39), because we will check that choosing $\hat{\theta}^{\text{MLE}}$ to satisfy (36) also ensures $\hat{\theta}^{\text{MLE}} - \theta^*$ is orthogonal to u^*. The statement (52) restricted to directions orthogonal to u^* is formalized in the following lemma.

Lemma 5.4. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for each $k = 1, \ldots, K$. Fix any constant $\eta > 0$. There exist constants $C_0, C_1, \delta_1, c > 0$ depending only on $\underline{c}, \bar{c}, \eta$ such that if $\sigma^2 \leq \frac{Kr^2}{C_1 \log K}$ and $N \geq C_0 K(1 + \frac{\sigma^2}{r^2}) \log(K + \frac{Kr^2}{\sigma^2})$, then with probability at least $1 - e^{-\frac{cN}{K(1+\sigma^2/r^2)}}$, the following holds: For every $\theta \in \mathcal{B}(\delta_1)$ and every unit vector $v \in \mathbb{R}^{2K}$ satisfying $\langle u^*, v\rangle = 0$,*

$$v^\top \nabla^2 R_N(\theta) v \geq \frac{1-\eta}{\sigma^2}.$$

From the form of $\nabla^2 R_N(\theta)$ in (45), observe that

$$v^\top \nabla^2 R_N(\theta) v = \frac{1}{\sigma^2} - \frac{1}{N\sigma^4} \sum_{m=1}^{N} \text{Var}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon^{(m)}}} \Big[v^\top g(\alpha)^{-1}(\theta^* + \sigma\varepsilon^{(m)})\Big]. \tag{54}$$

The proof of Lemma 5.4 is based on a refinement of the argument in the preceding section, to approximate the distribution $\mathcal{P}_{\theta,\varepsilon}$ in the above variance by a Gaussian law over α. Here, applying a separate bound to control the Gaussian process $\sup_\alpha \langle \varepsilon, g(\alpha) \cdot \theta\rangle$ will be too loose to obtain the lemma. We instead perform a Taylor expansion of $\langle \theta^* + \sigma\varepsilon,\, g(\alpha) \cdot \theta\rangle$ around its (random, ε-dependent) mode

$$\alpha_0 = \arg\max_\alpha \langle \theta^* + \sigma\varepsilon,\, g(\alpha) \cdot \theta\rangle,$$

and combine this with the condition $\theta \in \mathcal{B}(\delta_1)$ to obtain a quadratic approximation

$$\frac{\langle \theta^* + \sigma\varepsilon,\, g(\alpha) \cdot \theta\rangle}{\sigma^2} - \text{constant} \asymp -\frac{K^3 r^2}{\sigma^2}(\alpha - \alpha_0)^2$$

where the constant is independent of α. Thus, $\mathcal{P}_{\theta,\varepsilon}$ for any $\theta \in \mathcal{B}(\delta_1)$ may be approximated by a Gaussian law with mean α_0 and variance on the order of $\sigma^2/(K^3 r^2)$. Applying a Taylor expansion also of $v^\top g(\alpha)^{-1}(\theta^* + \sigma\varepsilon)$ around $\alpha = \alpha_0$, and approximating the variance over $\alpha \sim \mathcal{P}_{\theta,\varepsilon}$ by the variance with respect to this Gaussian law, we obtain a bound

$$\text{Var}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}} \Big[v^\top g(\alpha)^{-1}(\theta^* + \sigma\varepsilon)\Big] \leq \eta\sigma^2$$

for a small constant $\eta > 0$, which is sufficient to show Lemma 5.4.

These Taylor expansion arguments may be formalized on a high-probability event for ε, where this event is dependent on θ and v. More precisely, let

$$\tilde{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{C}^K, \qquad \tilde{v} = (v_1, \ldots, v_K) \in \mathbb{C}^K, \qquad \tilde{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_K) \in \mathbb{C}^K$$

denote the complex representations of θ, v, ε as defined in Section 3.2. For each $\theta \in \mathcal{B}(\delta_1)$ and unit test vector $v \in \mathbb{R}^{2K}$ with $\langle u^*, v\rangle = 0$, we define a (θ, v)-dependent domain $\mathcal{E}(\theta, v, \delta_1) \subset \mathbb{R}^{2K}$ by the four conditions

$$\sup_{\alpha \in \mathcal{A}} |\langle \varepsilon, g(\alpha) \cdot \theta\rangle| \leq \frac{\delta_1 \|\theta^*\|^2}{\sigma}$$

$$\sup_{\alpha\in\mathcal{A}} \Big|\langle \varepsilon, g(\alpha)\cdot v\rangle\Big| \leq \sqrt{\frac{Kr^2}{\sigma^2}}$$

$$\sup_{\alpha,\beta\in[-\pi,\pi)} \left|\mathrm{Re}\sum_{k=1}^K \overline{\varepsilon_k} e^{ik\beta}\Big(e^{ik\alpha}-1-ik\alpha\Big)\theta_k\right| \leq \delta_1\alpha^2 K^{5/2} r\sqrt{\frac{Kr^2}{\sigma^2}}$$

$$\sup_{\alpha,\beta\in[-\pi,\pi)} \left|\mathrm{Re}\sum_{k=1}^K \overline{\varepsilon_k}\Big(e^{ik\alpha}-e^{ik\beta}\Big)v_k\right| \leq \delta_1|\alpha-\beta| K\sqrt{\frac{Kr^2}{\sigma^2}}$$

The following deterministic lemma holds on the event that $\varepsilon \in \mathcal{E}(\theta, v, \delta_1)$.

Lemma 5.5. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for each $k = 1, \ldots, K$. Fix any $\eta > 0$. There exist constants $C_1, \delta_1 > 0$ depending only on $\underline{c}, \bar{c}, \eta$ such that if $\sigma^2 \leq \frac{Kr^2}{C_1 \log K}$, then the following holds: For any $\theta \in \mathcal{B}(\delta_1)$, any unit vector $v \in \mathbb{R}^{2K}$ satisfying $\langle u^*, v\rangle = 0$, and any (deterministic) $\varepsilon \in \mathcal{E}(\theta, v, \delta_1)$,*

$$\mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon}}\Big[v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon)\Big] \leq \eta\sigma^2. \tag{55}$$

Each of the four conditions defining $\mathcal{E}(\theta, v, \delta_1)$ involves the supremum of a Gaussian process, which may be bounded using a standard covering net argument. We remark that each of these conditions is defined with the right side being a factor $\sqrt{Kr^2/\sigma^2}$ larger than the mean value of the left side, so that their failure probabilities are exponentially small in Kr^2/σ^2. This is summarized in the following result.

Lemma 5.6. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for each $k = 1, \ldots, K$. Fix any constant $\delta_1 > 0$, any $\theta \in \mathcal{B}(\delta_1)$, and any unit vector v satisfying $\langle u^*, v\rangle = 0$. For some constants $C_1, c > 0$ depending only on $\underline{c}, \bar{c}, \delta_1$, if $\sigma^2 \leq \frac{Kr^2}{C_1 \log K}$, then*

$$\mathbb{P}_{\varepsilon\sim\mathcal{N}(0,I)}\Big[\varepsilon \notin \mathcal{E}(\theta, v, \delta_1)\Big] \leq e^{-cKr^2/\sigma^2}.$$

Finally, we combine Lemmas 5.5 and 5.6 to conclude the proof of Lemma 5.4: We may write the second term of (54) as

$$\frac{1}{N\sigma^4}\sum_{m=1}^N \mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}\Big[v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon^{(m)})\Big]\cdot \mathbb{1}\{\varepsilon^{(m)} \in \mathcal{E}(\theta, v, \delta_1)\}$$
$$+\frac{1}{N\sigma^4}\sum_{m=1}^N \mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}\Big[v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon^{(m)})\Big]\cdot \mathbb{1}\{\varepsilon^{(m)} \notin \mathcal{E}(\theta, v, \delta_1)\}.$$

The first sum is bounded by Lemma 5.5, while the second sum is sparse by Lemma 5.6 and may be controlled using a Chernoff bound for binomial random variables. Taking a union bound over a covering net of pairs (θ, v) shows Lemma 5.4.

5.4 Proof of Theorem 5.1

We now combine the preceding lemmas to conclude the proof of Theorem 5.1. Let $C_0, C_1, \delta_1 > 0$ be such that the conclusions of Lemma 5.4 hold for $\eta = 1/2$. Define the event

$$\mathcal{E} = \left\{\hat{\theta}^{\mathrm{MLE}} \in \mathcal{B}(\delta_1) \text{ and } \sup_{\theta\in\mathcal{B}(\delta_1)}\ \sup_{v:\|v\|=1,\langle u^*,v\rangle=0} v^\top\nabla^2 R_N(\theta)v \geq \frac{1}{2\sigma^2}\right\}.$$

When $\mathcal{E}$ holds, we have also $\tilde{\theta} \in \mathcal{B}(\delta_1)$ in the Taylor expansion (37). Recall our choice of rotation (36) for $\hat{\theta}^{\mathrm{MLE}}$. Then the first-order condition for (36) gives

$$0 = \frac{d}{d\alpha}\|\hat{\theta}^{\mathrm{MLE}} - g(\alpha)\cdot\theta^*\|^2\Big|_{\alpha=0} = -2\langle u^*, \hat{\theta}^{\mathrm{MLE}} - \theta^*\rangle,$$

so that $\langle u^*, \hat{\theta}^{\text{MLE}} - \theta^* \rangle = 0$. Then (37) and the definition of $\mathcal{E}$ imply

$$0 \geq \mathbb{1}\{\mathcal{E}\} \left(\nabla R_N(\theta^*)^\top (\hat{\theta}^{\text{MLE}} - \theta^*) + \frac{1}{4\sigma^2} \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \right).$$

Rearranging, we get

$$\mathbb{1}\{\mathcal{E}\} \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \leq -\mathbb{1}\{\mathcal{E}\} \cdot 4\sigma^2 \cdot \nabla R_N(\theta^*)^\top (\hat{\theta}^{\text{MLE}} - \theta^*) \leq 4\sigma^2 \cdot \|\nabla R_N(\theta^*)\| \cdot \|\hat{\theta}^{\text{MLE}} - \theta^*\|.$$

Dividing by $\|\hat{\theta}^{\text{MLE}} - \theta^*\|$, squaring both sides, and taking expectation yields

$$\mathbb{E}\Big[\mathbb{1}\{\mathcal{E}\} \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \Big] \leq 16\sigma^4 \mathbb{E}\Big[\|\nabla R_N(\theta^*)\|^2 \Big]. \tag{56}$$

From (44), we have

$$\nabla R_N(\theta^*) = \frac{1}{N} \sum_{m=1}^{N} \left(\frac{\theta^*}{\sigma^2} - \frac{1}{\sigma^2} \mathbb{E}_{\alpha \sim \mathcal{P}_{\theta^*, \varepsilon^{(m)}}} \Big[g(\alpha)^{-1} (\theta^* + \sigma \varepsilon^{(m)}) \Big] \right).$$

These summands (the per-sample score vectors) are independent random vectors with mean 0, by the first-order condition for θ^* minimizing $R(\theta)$. So

$$\begin{aligned}
\mathbb{E}\Big[\|\nabla R_N(\theta^*)\|^2 \Big] &= \frac{1}{N} \mathbb{E}_{\varepsilon \sim \mathcal{N}(0,I)} \left[\left\| \frac{\theta^*}{\sigma^2} - \frac{1}{\sigma^2} \mathbb{E}_{\alpha \sim \mathcal{P}_{\theta, \varepsilon}} \Big[g(\alpha)^{-1} (\theta^* + \sigma \varepsilon) \Big] \right\|^2 \right] \\
&= \frac{1}{N\sigma^4} \mathbb{E}_{\varepsilon \sim \mathcal{N}(0,I)} \left[\left\| \mathbb{E}_{\alpha \sim \mathcal{P}_{\theta, \varepsilon}} \Big[g(\alpha)^{-1} (\theta^* + \sigma \varepsilon) \Big] \right\|^2 - \|\theta^*\|^2 \right] \\
&\leq \frac{1}{N\sigma^4} \mathbb{E}_{\varepsilon \sim \mathcal{N}(0,I)} \big[\|\theta^* + \sigma \varepsilon\|^2 - \|\theta^*\|^2 \big] = \frac{2K}{N\sigma^2}.
\end{aligned}$$

Combining with (56),

$$\mathbb{E}\Big[\mathbb{1}\{\mathcal{E}\} \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \Big] \leq \frac{32K\sigma^2}{N}.$$

By Lemmas 5.2 and 5.4, $\mathbb{P}[\mathcal{E}^c] \leq e^{-\frac{cN}{K(1+\sigma^2/r^2)}}$ for some constant $c > 0$. Then applying also (48), for some constant $C > 0$,

$$\mathbb{E}\Big[\mathbb{1}\{\mathcal{E}^c\} \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \Big] \leq \sqrt{\mathbb{E}[\|\hat{\theta}^{\text{MLE}} - \theta^*\|^4]} \cdot \sqrt{\mathbb{P}[\mathcal{E}^c]} \leq CKr^2 \cdot e^{-\frac{cN}{2K(1+\sigma^2/r^2)}}.$$

Under the given assumption $N \geq C_0 K(1 + \frac{\sigma^2}{r^2}) \log(K + \frac{Kr^2}{\sigma^2})$ for sufficiently large $C_0 > 0$, this implies also $N \geq C_0' K(1 + \frac{\sigma^2}{r^2}) \log N$ for a large constant $C_0' > 0$. (This is verified in the proof of Lemma 5.4, cf. (113) of Appendix B.) Then for a different constant $C' > 0$,

$$\mathbb{E}\Big[\mathbb{1}\{\mathcal{E}^c\} \|\hat{\theta}^{\text{MLE}} - \theta^*\|^2 \Big] \leq CKr^2 \cdot e^{-\frac{cN}{2K(1+\sigma^2/r^2)}} \leq \frac{C'\sigma^2}{N}.$$

Combining the above two risk bounds on $\mathcal{E}$ and $\mathcal{E}^c$ yields Theorem 5.1.

6 Minimax lower bounds

In this section, we show the minimax lower bounds of Theorems 2.1 and 2.2. The lower bounds will be implied by estimation of the Fourier phases $\phi_k(\theta^*)$ only, even when the Fourier magnitudes $r_k(\theta^*)$ are equal and known. Fixing any $r > 0$, consider the parameter space

$$\mathcal{P}(r) = \Big\{ \theta^* \in \mathbb{R}^{2K} : r_k(\theta^*) = r \text{ for all } k = 1, \ldots, K \Big\}.$$

The main result of this section is the following minimax lower bound over $\mathcal{P}(r)$, which is valid for any noise level $\sigma^2 > 0$ and interpolates between the low-noise and high-noise regimes.

Lemma 6.1. *For a universal constant $c > 0$ and any $r, \sigma^2 > 0$,*

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \mathcal{P}(r)} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq c \cdot \min\left(\max\left(\frac{K\sigma^2}{N}, \frac{\sigma^6}{Nr^4 e^{3Kr^2/2\sigma^2}}\right), Kr^2\right). \tag{57}$$

Let us check that this implies the minimax lower bounds of Theorems 2.1 and 2.2.

Proof of Theorems 2.1 and 2.2, lower bounds. Assuming $\sigma^2 \geq c_0 Kr^2$, choosing the second argument of $\max(\cdot)$ in (57) gives

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \Theta(r)} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq \inf_{\hat{\theta}} \sup_{\theta^* \in \mathcal{P}(\underline{c}r)} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq c \cdot \min\left(\frac{\sigma^6}{Nr^4}, Kr^2\right)$$

for a constant $c > 0$ depending on $\underline{c}, c_0$. When $N \geq C_0 \frac{\sigma^6}{r^6} \log K$ for sufficiently large $C_0 > 0$, we have $\sigma^6/(Nr^4) < Kr^2$, so this gives the lower bound of Theorem 2.1. For any $\sigma^2 > 0$, choosing the first argument of $\max(\cdot)$ in (57) also gives

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \Theta(r)} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq \inf_{\hat{\theta}} \sup_{\theta^* \in \mathcal{P}(\underline{c}r)} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq c \cdot \min\left(\frac{K\sigma^2}{N}, Kr^2\right).$$

When $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$ and $N \geq C_0 K$ for sufficiently large $C_0, C_1 > 0$, we have $K\sigma^2/N < Kr^2$, so this gives the lower bound of Theorem 2.2. ☐

Finally, we describe the arguments that show Lemma 6.1, deferring detailed proofs to Appendix C. Denote $p_\theta(y)$ as the Gaussian mixture density of y, as in (35). The proof will apply Assouad's hypercube construction together with an upper bound on the KL-divergence $D_{\mathrm{KL}}(p_\theta \| p_{\theta'})$. For the low-noise regime of Theorem 2.2, a tight upper bound is provided by (58) below, which is immediate from the data processing inequality. For the high-noise regime of Theorem 2.1, we apply an argument from Bandeira et al. (2020) for bounding the χ^2-divergence, and track carefully the dependence of this argument on the dimension K.

Lemma 6.2. *For any $\theta, \theta' \in \mathbb{R}^{2K}$,*

$$D_{\mathrm{KL}}(p_\theta \| p_{\theta'}) \leq \frac{\|\theta - \theta'\|^2}{2\sigma^2}. \tag{58}$$

Furthermore, let $\theta = (r_k \cos\phi_k, r_k \sin\phi_k)_{k=1}^K$ and $\theta' = (r'_k \cos\phi'_k, r'_k \sin\phi'_k)_{k=1}^K$. Denote $R^2 = \max(\sum_{k=1}^K r_k^2, \sum_{k=1}^K {r'_k}^2)$ and $\bar{r} = \max(\max_{k=1}^K r_k, \max_{k=1}^K r'_k)$. Then also

$$D_{\mathrm{KL}}(p_\theta \| p_{\theta'}) \leq \frac{e^{R^2/2\sigma^2}}{4\sigma^4} \sum_{k=1}^K (r_k^2 - {r'_k}^2)^2 + \frac{3\bar{r}^2 R^2 e^{3R^2/2\sigma^2}}{2\sigma^6} \cdot \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K \Big[(r_k - r'_k)^2 + r_k r'_k (\phi_k - \phi'_k + k\alpha)^2\Big]. \tag{59}$$

To prove Lemma 6.1, we restrict attention to a discrete space of 2^K parameters $\theta^\tau \in \mathcal{P}(r)$, indexed by the hypercube $\tau \in \{0,1\}^K$, where all Fourier magnitudes are equal to r and the Fourier phases $\phi^\tau = (\phi_1^\tau, \ldots, \phi_K^\tau)$ are given by

$$\phi_k^\tau = \mathbb{1}\{\tau_k = 1\} \cdot \phi.$$

Here, the value $\phi \in \mathbb{R}$ is chosen maximally while ensuring that $D_{\mathrm{KL}}(p_{\theta^\tau} \| p_{\theta^{\tau'}}) \leq H(\tau, \tau')/N$ by the bounds of Lemma 6.2, where $H(\tau, \tau')$ is the Hamming distance on the hypercube. Applying Proposition 3.1, we may show that the loss between such parameters is also lower bounded in terms of Hamming distance as $L(\theta^\tau, \theta^{\tau'}) \gtrsim r^2\phi^2 \cdot H(\tau, \tau')$. Assouad's lemma (see e.g. (Cai and Zhou, 2012, Lemma 2)) then implies a minimax lower bound over the discrete parameter space $\{\theta^\tau : \tau \in \{0,1\}^K\}$, which in turn implies the lower bound of Lemma 6.1 over $\mathcal{P}(r)$.

A Proofs for method-of-moments estimation

We prove the results of Section 4 on the method-of-moments estimator.

Proposition A.1. *Let $\eta \sim \mathcal{N}_{\mathbb{C}}(0,2)$. Then we have the equalities in law $\eta \overset{L}{=} \overline{\eta}$ and $\eta \overset{L}{=} e^{i\phi}\eta$ for any $\phi \in \mathbb{R}$. Furthermore,*

$$\mathbb{E}[\eta^j \overline{\eta}^k] = 0 \text{ for all integers } j \neq k, \tag{60}$$

$$\mathbb{E}[|\eta|^{2j}] \leq 4^j j! \text{ for all integers } j \geq 1. \tag{61}$$

Proof. We may represent $\eta = Re^{i\alpha}$ where $R^2 \sim \chi_2^2$ is independent of $\alpha \sim \text{Unif}([-\pi,\pi))$. Then $\bar{\eta} = Re^{-i\alpha}$, $e^{i\phi}\eta = Re^{i(\phi+\alpha)}$, and $\eta^j \bar{\eta}^k = R^{j+k}e^{i(j-k)\alpha}$, so $\eta \overset{L}{=} \overline{\eta}$, $\eta \overset{L}{=} e^{i\phi}\eta$, and (60) follow. For (61), write $|\eta|^2 = R^2 = Z^2 + Z'^2$ where $Z, Z' \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Then $\mathbb{E}[|\eta|^{2j}] = \mathbb{E}[(Z^2+Z'^2)^j] \leq 2^j \mathbb{E}[Z^{2j} + Z'^{2j}]$. We have $\mathbb{E}[Z^{2j}] = (2j-1)!! \leq 2^{j-1} \cdot j!$, showing (61). □

A.1 Estimation of r_k

Proof of Lemma 4.2. Write $\theta_k = (\theta_{k,1}, \theta_{k,2}) \in \mathbb{R}^2$ and $\varepsilon_k^{(m)} = (\varepsilon_{k,1}^{(m)}, \varepsilon_{k,2}^{(m)}) \in \mathbb{R}^2$. Since $|\tilde{y}_k^{(m)}|^2 = \|\theta_k + \sigma\varepsilon_k^{(m)}\|^2$ and $\|\theta_k\|^2 = r_k^2$, we have

$$\frac{1}{N}\sum_{m=1}^N |\tilde{y}_k^{(m)}|^2 - 2\sigma^2 = r_k^2 + \frac{1}{N}\sum_{m=1}^N 2\sigma\langle \varepsilon_k^{(m)}, \theta_k\rangle + \frac{\sigma^2}{N}\sum_{m=1}^N (\|\varepsilon_k^{(m)}\|^2 - 2).$$

Applying $N^{-1}\sum_m 2\sigma\langle \varepsilon_k^{(m)}, \theta_k\rangle \sim \mathcal{N}(0, 4\sigma^2 r_k^2/N)$, $\sum_m \|\varepsilon_k^{(m)}\|^2 \sim \chi_{2N}^2$, and standard Gaussian and chi-squared tail bounds, for a universal constant $c > 0$ and any $t > 0$ we have

$$\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N 2\sigma\langle \varepsilon_k^{(m)}, \theta_k\rangle \geq t\right] \leq e^{-cNt^2/\sigma^2 r_k^2}, \quad \mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N (\|\varepsilon_k^{(m)}\|^2 - 2) \geq t\right] \leq e^{-cN(t \wedge t^2)}.$$

Then for a universal constant $c' > 0$,

$$\begin{aligned}\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N |\tilde{y}_k^{(m)}|^2 - 2\sigma^2 \geq (1+t)r_k^2\right] &\leq \mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N 2\sigma\langle \varepsilon_k^{(m)}, \theta_k\rangle \geq \frac{tr_k^2}{2}\right] + \mathbb{P}\left[\frac{\sigma^2}{N}\sum_{m=1}^N (\|\varepsilon_k^{(m)}\|^2 - 2) \geq \frac{tr_k^2}{2}\right] \\ &\leq 2\exp\left(-c'N\left(\frac{t^2 r_k^2}{\sigma^2} \wedge \frac{tr_k^2}{\sigma^2} \wedge \frac{t^2 r_k^4}{\sigma^4}\right)\right).\end{aligned}$$

Applying this with $t = 2s + s^2$ and recalling the definition of $\hat{r}_k$ from (16), the left side is exactly $\mathbb{P}[\hat{r}_k \geq r_k(1+s)]$. Then, considering separately the cases $s \geq 1$ and $s \leq 1$, the right side reduces to the upper bound (23). For the lower bound, similarly for any $t > 0$, a lower chi-squared tail bound gives

$$\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N (\|\varepsilon_k^{(m)}\|^2 - 2) \leq -t\right] \leq e^{-cNt^2}.$$

Then we obtain analogously

$$\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N |\tilde{y}_k^{(m)}|^2 - 2\sigma^2 \leq r_k^2(1-t)\right] \leq 2\exp\left(-cN\left(\frac{t^2 r_k^2}{\sigma^2} \wedge \frac{t^2 r_k^4}{\sigma^4}\right)\right).$$

Applying this with $t = 2s - s^2 \geq s$ for $s \in [0,1)$, we obtain (24). □

Proof of Corollary 4.3. We apply $\mathbb{E}[X^2] = \mathbb{E}[\int_0^\infty \mathbb{1}\{|X| \geq s\} \cdot 2s\,ds] = \int_0^\infty \mathbb{P}[|X| \geq s] \cdot 2s\,ds$ with $X = \hat{r}_k/r_k - 1$, and $\int_0^\infty s\,e^{-\alpha s^2} ds = \alpha^{-1}\int_0^\infty t\,e^{-t^2} dt \leq C/\alpha$. Then Lemma 4.2 gives

$$\mathbb{E}[(\hat{r}_k - r_k)^2] = r_k^2 \cdot \mathbb{E}[(\hat{r}_k/r_k - 1)^2] \leq r_k^2 \cdot \int_0^\infty 8s\left(e^{-cNs^2 r_k^2/\sigma^2} + e^{-cNs^2 r_k^4/\sigma^4}\right) ds \leq C\left(\frac{\sigma^2}{N} + \frac{\sigma^4}{Nr_k^2}\right).$$

□

A.2 Oracle estimation of $\Phi_{k,l}$

Proof of Lemma 4.4. Recall $B_{k,l}$ from (20) and $\hat{B}_{k,l}$ from (17). We first show concentration of $\hat{B}_{k,l}$ around $B_{k,l}$. Let us write

$$\tilde{y}_k^{(m)} = r_k e^{i(\phi_k + k\alpha^{(m)})} + \sigma\tilde{\varepsilon}_k^{(m)} = e^{ik\alpha^{(m)}}\tilde{\theta}_k(1 + (\sigma/r_k)\eta_k^{(m)})$$

where $\tilde{\theta}_k = r_k e^{i\phi_k}$ is the complex representation of $(\theta_{k,1}, \theta_{k,2})$, and $\eta_k^{(m)} = e^{-ik\alpha^{(m)}}(r_k/\tilde{\theta}_k)\tilde{\varepsilon}_k^{(m)}$ is a rotation of the Gaussian noise. By Proposition A.1, we still have $\eta_k^{(m)} \sim \mathcal{N}_{\mathbb{C}}(0,2)$ where these remain independent across all $k = 1, \ldots, K$ and $m = 1, \ldots, N$. Applying this to (17), the factors $e^{ik\alpha^{(m)}}, e^{il\alpha^{(m)}}, e^{i(k+l)\alpha^{(m)}}$ cancel to yield

$$\hat{B}_{k,l} = \frac{1}{N}\sum_{m=1}^N \tilde{\theta}_{k+l}\overline{\tilde{\theta}_k\tilde{\theta}_l}\left(1 + (\sigma/r_{k+l})\eta_{k+l}^{(m)}\right)\left(1 + (\sigma/r_k)\overline{\eta_k^{(m)}}\right)\left(1 + (\sigma/r_l)\overline{\eta_l^{(m)}}\right) = B_{k,l}(1 + \mathrm{I} + \mathrm{II} + \mathrm{III}) \quad (62)$$

where

$$\mathrm{I} = \frac{\sigma}{N}\sum_{m=1}^N \frac{\overline{\eta_k^{(m)}}}{r_k} + \frac{\overline{\eta_l^{(m)}}}{r_l} + \frac{\eta_{k+l}^{(m)}}{r_{k+l}}$$

$$\mathrm{II} = \frac{\sigma^2}{N}\sum_{m=1}^N \frac{\overline{\eta_k^{(m)}\eta_l^{(m)}}}{r_k r_l} + \frac{\overline{\eta_k^{(m)}}\eta_{k+l}^{(m)}}{r_k r_{k+l}} + \frac{\overline{\eta_l^{(m)}}\eta_{k+l}^{(m)}}{r_l r_{k+l}}$$

$$\mathrm{III} = \frac{\sigma^3}{N}\sum_{m=1}^N \frac{\overline{\eta_k^{(m)}\eta_l^{(m)}}\eta_{k+l}^{(m)}}{r_k r_l r_{k+l}}.$$

To bound I, observe that

$$\frac{\sigma}{N}\sum_{m=1}^N \frac{\operatorname{Re}\eta_k^{(m)}}{r_k} \sim \mathcal{N}\left(0, \frac{\sigma^2}{N r_k^2}\right), \quad (63)$$

and similarly for the imaginary part and for the other two terms of I. Then by a Gaussian tail bound,

$$\mathbb{P}[|\mathrm{I}| \geq t] \leq C\exp(-cNt^2\underline{r}^2/\sigma^2). \quad (64)$$

To bound II, consider first $k \neq l$ and $\sum_m \operatorname{Re}\overline{\eta_k^{(m)}} \cdot \operatorname{Re}\overline{\eta_l^{(m)}}$. Each term $\operatorname{Re}\overline{\eta_k^{(m)}} \cdot \operatorname{Re}\overline{\eta_l^{(m)}}$ is the product of two independent standard Gaussian variables. Then applying (Latala, 2006, Corollary 1) with $d = 2$, $A = I$, $\|A\|_{\{1,2\}} = \sqrt{N}$, and $\|A\|_{\{1\},\{2\}} = 1$, we have

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{m=1}^N \operatorname{Re}\overline{\eta_k^{(m)}} \cdot \operatorname{Re}\overline{\eta_l^{(m)}}\right| \geq t\right] \leq Ce^{-cN(t\wedge t^2)}.$$

So

$$\mathbb{P}\left[\left|\frac{\sigma^2}{N}\sum_{m=1}^N \frac{\operatorname{Re}\overline{\eta_k^{(m)}}}{r_k} \cdot \frac{\operatorname{Re}\overline{\eta_l^{(m)}}}{r_l}\right| \geq t\right] \leq C\exp\left(-cN\left(\frac{t\underline{r}^2}{\sigma^2} \wedge \frac{t^2\underline{r}^4}{\sigma^4}\right)\right). \quad (65)$$

The same bound holds for all products of real and imaginary parts of $\eta_k^{(m)}$ and $\eta_l^{(m)}$, except for $\operatorname{Re}\eta_k^{(m)} \cdot \operatorname{Re}\eta_l^{(m)}$ and $\operatorname{Im}\eta_k^{(m)} \cdot \operatorname{Im}\eta_l^{(m)}$ when $k = l$. For these products, we may consider them together and apply

$$\frac{\sigma^2}{N}\sum_{m=1}^N \frac{(\operatorname{Re}\eta_k^{(m)})^2}{r_k^2} - \frac{(\operatorname{Im}\eta_k^{(m)})^2}{r_k^2} = \frac{2\sigma^2}{N r_k^2}\sum_{m=1}^N \frac{\operatorname{Re}\eta_k^{(m)} - \operatorname{Im}\eta_k^{(m)}}{\sqrt{2}} \cdot \frac{\operatorname{Re}\eta_k^{(m)} + \operatorname{Im}\eta_k^{(m)}}{\sqrt{2}}$$

where now $(\operatorname{Re}\eta_k^{(m)} - \operatorname{Im}\eta_k^{(m)})/\sqrt{2}$ and $(\operatorname{Re}\eta_k^{(m)} + \operatorname{Im}\eta_k^{(m)})/\sqrt{2}$ are independent standard Gaussian variables. The bound (65) then holds for this sum, and this shows

$$\mathbb{P}\left[\left|\frac{\sigma^2}{N}\sum_{m=1}^N \frac{\overline{\eta_k^{(m)}\eta_l^{(m)}}}{r_k r_l}\right| > t\right] \leq C\exp\left(-cN\left(\frac{t\underline{r}^2}{\sigma^2} \wedge \frac{t^2\underline{r}^4}{\sigma^4}\right)\right)$$

for the first term of II. Applying the same argument for the remaining two terms of II,

$$\mathbb{P}[|\mathrm{II}| \geq t] \leq C \exp\left(-cN\left(\frac{t\underline{r}^2}{\sigma^2} \wedge \frac{t^2\underline{r}^4}{\sigma^4}\right)\right). \tag{66}$$

We apply a similar argument to bound III. Consider first $k \neq l$ and $\sum_m \operatorname{Re}\eta_k^{(m)} \cdot \operatorname{Re}\eta_l^{(m)} \cdot \operatorname{Re}\eta_{k+l}^{(m)}$. Each term $\operatorname{Re}\eta_k^{(m)} \cdot \operatorname{Re}\eta_l^{(m)} \cdot \operatorname{Re}\eta_{k+l}^{(m)}$ is the product of three independent standard Gaussian variables. Then applying (Latala, 2006, Corollary 1) with $d = 3$, $A = \sum_{m=1}^N e_m \otimes e_m \otimes e_m$, $\|A\|_{\{1,2,3\}} = \sqrt{N}$, $\|A\|_{\{1,2\},\{3\}} = 1$, and $\|A\|_{\{1\},\{2\},\{3\}} = 1$,

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{m=1}^N \operatorname{Re}\eta_k^{(m)} \cdot \operatorname{Re}\eta_l^{(m)} \cdot \operatorname{Re}\eta_{k+l}^{(m)}\right| \geq t\right] \leq Ce^{-c(Nt^2 \wedge Nt \wedge (Nt)^{2/3})} \leq Ce^{-cN(t^2 \wedge \frac{t^{2/3}}{N^{1/3}})}.$$

(The second inequality applies $t \geq t^2 \wedge \frac{t^{2/3}}{N^{1/3}}$ for any $N \geq 1$ and $t \geq 0$.) The same bound holds for all combinations of real and imaginary parts of $\eta_k^{(m)}, \eta_l^{(m)}, \eta_{k+l}^{(m)}$, except again for products having $\operatorname{Re}\eta_k^{(m)} \cdot \operatorname{Re}\eta_l^{(m)}$ or $\operatorname{Im}\eta_k^{(m)} \cdot \operatorname{Im}\eta_l^{(m)}$ when $k = l$. These products may be bounded by applying

$$\frac{1}{2}\operatorname{Re}\eta_{2k}^{(m)} \cdot \left((\operatorname{Re}\eta_k^{(m)})^2 - (\operatorname{Im}\eta_k^{(m)})^2\right) = \operatorname{Re}\eta_{2k}^{(m)} \cdot \frac{\operatorname{Re}\eta_k^{(m)} - \operatorname{Im}\eta_k^{(m)}}{\sqrt{2}} \cdot \frac{\operatorname{Re}\eta_k^{(m)} + \operatorname{Im}\eta_k^{(m)}}{\sqrt{2}}$$

and similarly for $\operatorname{Im}\eta_{2k}^{(m)} \cdot ((\operatorname{Re}\eta_k^{(m)})^2 - (\operatorname{Im}\eta_k^{(m)})^2)$, where $\operatorname{Re}\eta_{2k}^{(m)}$, $\operatorname{Im}\eta_{2k}^{(m)}$, $(\operatorname{Re}\eta_k^{(m)} - \operatorname{Im}\eta_k^{(m)})/\sqrt{2}$, and $(\operatorname{Re}\eta_k^{(m)} + \operatorname{Im}\eta_k^{(m)})/\sqrt{2}$ are independent standard Gaussian variables. Thus

$$\mathbb{P}[|\mathrm{III}| \geq t] \leq C \exp\left(-cN\left(\frac{t^2\underline{r}^6}{\sigma^6} \wedge \frac{t^{2/3}\underline{r}^2}{N^{1/3}\sigma^2}\right)\right). \tag{67}$$

Combining (64), (66), and (67), for any $s > 0$ we obtain

$$\mathbb{P}\left[|\hat{B}_{k,l}/B_{k,l} - 1| \geq s\right] \leq C \exp\left(-cN\left(\frac{s^2\underline{r}^2}{\sigma^2} \wedge \frac{s\underline{r}^2}{\sigma^2} \wedge \frac{s^2\underline{r}^4}{\sigma^4} \wedge \frac{s^2\underline{r}^6}{\sigma^6} \wedge \frac{s^{2/3}\underline{r}^2}{N^{1/3}\sigma^2}\right)\right).$$

We have

$$\frac{s\underline{r}^2}{\sigma^2} \geq \frac{s^2\underline{r}^2}{\sigma^2} \wedge \frac{s^{2/3}\underline{r}^2}{N^{1/3}\sigma^2}, \qquad \frac{s^2\underline{r}^4}{\sigma^4} \geq \frac{s^2\underline{r}^2}{\sigma^2} \wedge \frac{s^2\underline{r}^6}{\sigma^6},$$

so this simplifies to (25).

Finally, for any $z \in \mathbb{C}$ and any $s \in (0,1)$, observe that $|z-1| < s$ implies $|\operatorname{Arg} z| < \arcsin s < \pi s/2$ for the principal argument (11). Then, recalling that $\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l} = \operatorname{Arg}(\hat{B}_{k,l}/B_{k,l})$ from (21), we obtain for any $s \in (0, \pi/2)$ that

$$\mathbb{P}\left[|\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l}| \geq s\right] \leq \mathbb{P}\left[|\hat{B}_{k,l}/B_{k,l} - 1| \geq \frac{2s}{\pi}\right]$$

and (26) follows. □

Proof of Corollary 4.5. We apply $\mathbb{E}[X^2] = \int_0^\infty \mathbb{P}[|X| \geq s] \cdot 2s\,ds$ and Lemma 4.4 to obtain, for universal constants $C, c > 0$,

$$\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})^2] \leq \int_0^{\pi/2} Cs\left(e^{-cNs^2\underline{r}^2/\sigma^2} + e^{-cNs^2\underline{r}^6/\sigma^6} + e^{-c(Ns)^{2/3}\underline{r}^2/\sigma^2}\right)ds \\ + C\left(e^{-cN\underline{r}^2/\sigma^2} + e^{-cN\underline{r}^6/\sigma^6} + e^{-cN^{2/3}\underline{r}^2/\sigma^2}\right),$$

where the second term bounds the integral from $s = \pi/2$ to $s = \pi$. The result then follows from applying $\int_0^\infty se^{-\alpha s^2}ds = \alpha^{-1}\int_0^\infty te^{-t^2}dt \leq C/\alpha$ and $\int_0^\infty se^{-\alpha s^{2/3}}ds = \alpha^{-3}\int_0^\infty t^3 e^{-t^2} \cdot 3t^2\,dt \leq C/\alpha^3$ for the first term, $e^{-cx} \leq C/x$ and $e^{-cx} \leq C/x^3$ for the second term, and $\sigma^6/N^2\underline{r}^6 \leq \sigma^6/N\underline{r}^6$. □

Proof of Lemma 4.6. Part (c) follows from Corollary 4.5 and Cauchy-Schwarz. For part (a), recall from (21) and the expression for $\hat{B}_{k,l}$ in (62) that

$$\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l} = \operatorname{Arg}(\hat{B}_{k,l}/B_{k,l}) = \operatorname{Arg} \frac{1}{N} \sum_{m=1}^{N} \Big(1 + (\sigma/r_{k+l})\eta_{k+l}^{(m)}\Big) \Big(1 + (\sigma/r_k)\overline{\eta_k^{(m)}}\Big) \Big(1 + (\sigma/r_l)\overline{\eta_l^{(m)}}\Big)$$

Since $\eta_k^{(m)}$ are independent across $k = 1, \ldots, K$ and $m = 1, \ldots, N$, we obtain in the setting of part (a) that $\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l}$ is independent of $\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y}$. Furthermore, applying the conjugation symmetry of Proposition A.1 to the variables $\eta_k^{(m)}$, we have the equality in law $\hat{B}_{k,l}/B_{k,l} \overset{L}{=} \overline{\hat{B}_{k,l}/B_{k,l}}$ for the quantity inside $\operatorname{Arg}(\cdot)$. Since $\operatorname{Arg} z = -\operatorname{Arg} \overline{z}$ whenever $\operatorname{Arg} z \neq -\pi$, and the probability is 0 that $\operatorname{Arg} \hat{B}_{k,l}/B_{k,l} = -\pi$ exactly, this equality in law implies the sign symmetry (27). Hence $\mathbb{E}[\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l}] = 0$. This shows part (a).

It remains to show part (b). Let Ln denote the principal value of the complex logarithm with branch cut on the negative real line, so that $\operatorname{Arg} z = \operatorname{Im} \operatorname{Ln} z$ whenever $\operatorname{Arg} z \neq -\pi$. Denote $\delta_{k,l} = \hat{B}_{k,l}/B_{k,l} - 1$. Then (with probability 1)

$$(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y}) = \operatorname{Arg} \frac{\hat{B}_{k,l}}{B_{k,l}} \cdot \operatorname{Arg} \frac{\hat{B}_{x,y}}{B_{x,y}} = \operatorname{Im} \operatorname{Ln}(1 + \delta_{k,l}) \cdot \operatorname{Im} \operatorname{Ln}(1 + \delta_{x,y}).$$

Let us fix an integer $J = J(N, \underline{r}^2, \sigma^2) \geq 1$ to be determined, and apply a Taylor expansion of $t \mapsto \operatorname{Ln}(1 + t\delta)$ around $t = 0$ to write

$$\operatorname{Im} \operatorname{Ln}(1 + \delta) = q(\delta) + r(\delta), \qquad q(\delta) = \operatorname{Im} \sum_{j=1}^{J} \frac{(-1)^{j-1}}{j} \delta^j, \qquad r(\delta) = \operatorname{Im} \int_0^1 \delta^{J+1} \cdot \frac{(-1)^J (1-t)^J}{(1 + t\delta)^{J+1}} dt.$$

Define the event $\mathcal{E} = \{|\delta_{k,l}| < 1/2 \text{ and } |\delta_{x,y}| < 1/2\}$. We may then apply the approximation

$$\mathbb{E}\Big[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})\Big] = \mathbb{E}\Big[q(\delta_{k,l}) \cdot q(\delta_{x,y})\Big] + \mathrm{I} + \mathrm{II} + \mathrm{III}$$

where we define the three error terms

$$\begin{aligned}
\mathrm{I} &= -\mathbb{E}\Big[\mathbb{1}\{\mathcal{E}^c\} \cdot q(\delta_{k,l}) \cdot q(\delta_{x,y})\Big] \\
\mathrm{II} &= \mathbb{E}\Big[\mathbb{1}\{\mathcal{E}\}\Big(q(\delta_{k,l}) \cdot r(\delta_{x,y}) + r(\delta_{k,l}) \cdot q(\delta_{x,y}) + r(\delta_{k,l}) \cdot r(\delta_{x,y})\Big)\Big] \\
\mathrm{III} &= \mathbb{E}\Big[\mathbb{1}\{\mathcal{E}^c\} \cdot (\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})\Big]
\end{aligned}$$

To bound these errors, let $C, C', c, c' > 0$ denote universal constants changing from instance to instance. Recall from (25) that

$$\mathbb{P}[\mathcal{E}^c] \leq \mathbb{P}[|\delta_{k,l}| \geq 1/2] + \mathbb{P}[|\delta_{x,y}| \geq 1/2] \leq C e^{-c(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2})}. \tag{68}$$

Also, for any $j \geq 1$, applying $\mathbb{E}[|X|^j] = \int_0^\infty \mathbb{P}[|X| \geq s] \cdot j s^{j-1} \, ds$ with $X = 2\delta_{k,l}$, and applying also for $Z \sim \mathcal{N}(0,1)$ that $\mathbb{E}[|Z|^j] \leq \mathbb{E}[Z^{2j}]^{1/2} = [(2j-1)!!]^{1/2} \leq (2j)^{(j-1)/2}$, we have from (25) that

$$\begin{aligned}
\mathbb{E}[|2\delta_{k,l}|^j] &\leq \int_0^\infty j s^{j-1} \cdot C(e^{-cNs^2\underline{r}^2/\sigma^2} + e^{-cNs^2\underline{r}^6/\sigma^6} + e^{-c(Ns)^{2/3}\underline{r}^2/\sigma^2}) \, ds \\
&= Cj \left(\left(\frac{\sigma^j}{\underline{r}^j N^{j/2}} + \frac{\sigma^{3j}}{\underline{r}^{3j} N^{j/2}} \right) \cdot \int_0^\infty t^{j-1} e^{-ct^2} dt + \left(\frac{\sigma^{3j}}{\underline{r}^{3j} N^j} \right) \cdot \int_0^\infty t^{3j-3} e^{-ct^2} \cdot 3t^2 \, dt \right) \\
&\leq (C_0 j)^{\frac{j}{2}} \left(\frac{\sigma^j}{\underline{r}^j N^{j/2}} + \frac{\sigma^{3j}}{\underline{r}^{3j} N^{j/2}} \right) + (C_0 j)^{\frac{3j}{2}} \left(\frac{\sigma^{3j}}{\underline{r}^{3j} N^j} \right)
\end{aligned}$$

where C_0 in the last line is a universal constant, which we will later assume satisfies $C_0 \geq 3$. Let us set

$$J = \left\lfloor \frac{1}{4C_0 e} \left(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2} \right) \right\rfloor \tag{69}$$

for this constant $C_0 > 0$. Note that if the quantity inside $\lfloor\cdot\rfloor$ is less than 1, then the statement of part (b) holds since the left side of (28) is at most π^2, and the right side is an arbitrarily large constant. Thus, we may assume henceforth that $J \geq 1$. The above gives

$$\mathbb{E}[|2\delta_{k,l}|^j] \leq 2\left(\frac{j}{4eJ}\right)^{j/2} + \left(\frac{j}{4eJ}\right)^{3j/2}.$$

Applying $|q(\delta)| \leq J\max(|\delta|, |\delta|^J)$, and also $|r(\delta)| \leq |2\delta|^{J+1}$ for $|\delta| < 1/2$, this shows

$$\begin{aligned}
\mathbb{E}[|q(\delta_{k,l})|^2] &\leq J^2\mathbb{E}[|\delta_{k,l}|^2] + J^2\mathbb{E}[|\delta|^{2J}] \leq CJ + J^2 e^{-cJ},\\
\mathbb{E}[|q(\delta_{k,l})|^3] &\leq J^3\mathbb{E}[|\delta_{k,l}|^3] + J^3\mathbb{E}[|\delta|^{3J}] \leq CJ^{3/2} + J^3 e^{-cJ},\\
\mathbb{E}[\mathbb{1}\{\mathcal{E}\}|r(\delta_{k,l})|^2] &\leq \mathbb{E}[|\delta_{k,l}|^{2J+2}] \leq Ce^{-cJ}.
\end{aligned}$$

Then, applying these bounds together with (68), Hölder's inequality, and Cauchy-Schwarz,

$$|\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}| \leq CJ^2 e^{-cJ} \leq C' e^{-c'J}.$$

This gives the second term on the right side of (28).

Finally, let us bound the dominant term $\mathbb{E}[q(\delta_{k,l})q(\delta_{x,y})]$ using the condition that $\{k, l, k+l\} \cap \{x, y, x+y\}$ has cardinality 1. Applying $2\operatorname{Im} u \cdot \operatorname{Im} v = \operatorname{Re} u\bar{v} - \operatorname{Re} uv$, we have

$$\mathbb{E}[q(\delta_{k,l})q(\delta_{x,y})] = \sum_{i,j=1}^{J} \frac{(-1)^{i+j}}{ij}\mathbb{E}[\operatorname{Im}\delta_{k,l}^i \cdot \operatorname{Im}\delta_{x,y}^j] = \sum_{i,j=1}^{J} \frac{(-1)^{i+j}}{2ij}\Big(\operatorname{Re}\mathbb{E}[\delta_{k,l}^i\overline{\delta_{x,y}^j}] - \operatorname{Re}\mathbb{E}[\delta_{k,l}^i\delta_{x,y}^j]\Big).$$

From the expression for $\hat{B}_{k,l}$ in (62), observe that

$$\delta_{k,l} = \frac{\hat{B}_{k,l}}{B_{k,l}} - 1 = \frac{1}{N}\sum_{m=1}^{N}\left(1 + \frac{\sigma}{r_{k+l}}\eta_{k+l}^{(m)}\right)\left(1 + \frac{\sigma}{r_k}\overline{\eta_k^{(m)}}\right)\left(1 + \frac{\sigma}{r_l}\overline{\eta_l^{(m)}}\right) - 1.$$

We view this as a polynomial in the variables $\{\eta_{k+l}^{(m)}, \overline{\eta_k^{(m)}}, \overline{\eta_l^{(m)}} : m = 1, \ldots, N\}$ where, after canceling $+1$ with -1, each monomial has total degree at least 1 in these variables. We consider three cases.

Case 1: $k + l = x + y$. This allows possibly $k = l$ and/or $x = y$, but ensures $\{k, l\} \cap \{x, y\} = \emptyset$ since $\{k, l, k+l\} \cap \{x, y, x+y\}$ has cardinality 1. We may expand $\delta_{k,l}^i\delta_{x,y}^j$ as a sum of monomials in $\eta_{k+l}^{(m)}, \overline{\eta_k^{(m)}}, \overline{\eta_l^{(m)}}, \overline{\eta_x^{(m)}}, \overline{\eta_y^{(m)}}$ with degree at least 1, and observe that $k + l = x + y$ is distinct from $\{k, l, x, y\}$ because it is strictly greater in value. Then (60) from Proposition A.1 implies $\mathbb{E}[\delta_{k,l}^i\delta_{x,y}^j] = 0$. We may also expand $\delta_{k,l}^i\overline{\delta_{x,y}^j}$ as a sum of monomials in $\eta_{k+l}^{(m)}, \overline{\eta_k^{(m)}}, \overline{\eta_l^{(m)}}, \overline{\eta_{k+l}^{(m)}}, \eta_x^{(m)}, \eta_y^{(m)}$. Since $\{k, l\}$ are distinct from $\{x, y, k+l\}$, any monomial involving $\overline{\eta_k^{(m)}}, \overline{\eta_l^{(m)}}$ has vanishing expectation. Similarly, any monomial involving $\eta_x^{(m)}, \eta_y^{(m)}$ has vanishing expectation. Thus the only non-vanishing terms are

$$\mathbb{E}[\delta_{k,l}^i\overline{\delta_{x,y}^j}] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{m=1}^{N}\frac{\sigma}{r_{k+l}}\eta_{k+l}^{(m)}\right)^i\left(\frac{1}{N}\sum_{m=1}^{N}\frac{\sigma}{r_{k+l}}\overline{\eta_{k+l}^{(m)}}\right)^j\right]. \tag{70}$$

Then applying the equality in law $N^{-1}\sum_{m=1}^{N}(\sigma/r_{k+l})\eta_{k+l}^{(m)} \overset{L}{=} \eta \cdot \sigma/(r_{k+l}\sqrt{N})$ where $\eta \sim \mathcal{N}_{\mathbb{C}}(0, 2)$, together with (60) and (61) and the bound $j! \leq j^j$,

$$\left|\mathbb{E}[\delta_{k,l}^i\overline{\delta_{x,y}^j}]\right| = \left(\frac{\sigma}{r_{k+l}\sqrt{N}}\right)^{i+j}\mathbb{E}[\eta^i\overline{\eta^j}] \leq \mathbb{1}\{i = j\}\left(\frac{4j\sigma^2}{N\underline{r}^2}\right)^j.$$

So, recalling the definition of J from (69) where $C_0 \geq 3$,

$$\left|\mathbb{E}[q(\delta_{k,l})q(\delta_{x,y})]\right| \leq \sum_{j=1}^{J}\frac{1}{2j^2}\left(\frac{4j\sigma^2}{N\underline{r}^2}\right)^j \leq \left(\frac{2\sigma^2}{N\underline{r}^2}\right)\sum_{j=1}^{J}\left(\frac{4J\sigma^2}{N\underline{r}^2}\right)^{j-1} \leq \left(\frac{2\sigma^2}{N\underline{r}^2}\right)\sum_{j=1}^{\infty}\left(\frac{1}{C_0 e}\right)^{j-1}.$$

Thus we obtain, for a universal constant $C > 0$,

$$\left|\mathbb{E}[q(\delta_{k,l})q(\delta_{x,y})]\right| \leq \frac{C\sigma^2}{N\underline{r}^2}. \tag{71}$$

This concludes the proof in Case 1.

Case 2: $k = x + y$. (By symmetry, this addresses also $l = x + y$, $x = k + l$, and $y = k + l$.) Then $\{k, l\} \cap \{x, y, k+l\} = \emptyset$ and $\{k+l\} \cap \{x, y\} = \emptyset$, because $k+l$ is greater than $\{k, l\}$, k is greater than $\{x, y\}$, and $l = x$ or $l = y$ would imply that $\{k, l, k+l\} \cap \{x, y, x+y\}$ has cardinality 2. We may expand $\delta^i_{k,l}\overline{\delta^j_{x,y}}$ as a sum of monomials in $\eta^{(m)}_{k+l}, \overline{\eta^{(m)}_k}, \overline{\eta^{(m)}_l}, \eta^{(m)}_x, \eta^{(m)}_y$. Since $\{k, l\}$ are distinct from $\{x, y, k+l\}$, (60) implies $\mathbb{E}[\delta^i_{k,l}\overline{\delta^j_{x,y}}] = 0$. We may also expand $\delta^i_{k,l}\delta^j_{x,y}$ as a sum of monomials in $\eta^{(m)}_{k+l}, \overline{\eta^{(m)}_k}, \overline{\eta^{(m)}_l}, \eta^{(m)}_k, \overline{\eta^{(m)}_x}, \overline{\eta^{(m)}_y}$. Here, $k + l$ is distinct from $\{k, l, x, y\}$, and $\{x, y\}$ are distinct from $\{k, k+l\}$, so any monomial involving $\eta^{(m)}_{k+l}, \overline{\eta^{(m)}_x}, \overline{\eta^{(m)}_y}$ has vanishing expectation. If $l \neq k$, then also l is distinct from $\{k, k+l\}$ so monomials involving $\overline{\eta^{(m)}_l}$ have vanishing expectation, yielding

$$\mathbb{E}[\delta^i_{k,l}\delta^j_{x,y}] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{m=1}^N \frac{\sigma}{r_k}\overline{\eta^{(m)}_k}\right)^i \left(\frac{1}{N}\sum_{m=1}^N \frac{\sigma}{r_k}\eta^{(m)}_k\right)^j\right].$$

This is analogous to (70), and the same argument as above gives (71).

If instead, $l = k$, then we obtain that the only non-zero terms of $\mathbb{E}[\delta^i_{k,l}\delta^j_{x,y}]$ are

$$\mathbb{E}\left[\delta^i_{k,l}\delta^j_{x,y}\right] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{m=1}^N \frac{2\sigma}{r_k}\overline{\eta^{(m)}_k} + \frac{\sigma^2}{r_k^2}(\overline{\eta^{(m)}_k})^2\right)^i \left(\frac{1}{N}\sum_{m=1}^N \frac{\sigma}{r_k}\eta^{(m)}_k\right)^j\right] \tag{72}$$

Let us distribute this product and then factor the expectations of the resulting terms, using independence of $\{\eta^{(m)}_k : m = 1, \ldots, N\}$. We write $\sum_{(a_1,\ldots,a_N)|a}$ for the sum over all tuples of nonnegative integers $(a_1, \ldots, a_N)$ that sum to a. Then the above may be rewritten as

$$\begin{aligned}
\mathbb{E}\left[\delta^i_{k,l}\delta^j_{x,y}\right] &= \sum_{\substack{a,b\geq 0\\ a+b=i}} \binom{i}{a}\mathbb{E}\left[\left(\frac{1}{N}\sum_{m=1}^N \frac{2\sigma}{r_k}\overline{\eta^{(m)}_k}\right)^a \left(\frac{1}{N}\sum_{m=1}^N \frac{\sigma^2}{r_k^2}(\overline{\eta^{(m)}_k})^2\right)^b \left(\frac{1}{N}\sum_{m=1}^N \frac{\sigma}{r_k}\eta^{(m)}_k\right)^j\right] \\
&= \sum_{\substack{a,b\geq 0\\ a+b=i}} \binom{i}{a} 2^a N^b \mathbb{E}\left[\left(\sum_{m=1}^N \frac{\sigma}{Nr_k}\overline{\eta^{(m)}_k}\right)^a \left(\sum_{m=1}^N \left(\frac{\sigma}{Nr_k}\overline{\eta^{(m)}_k}\right)^2\right)^b \left(\sum_{m=1}^N \frac{\sigma}{Nr_k}\eta^{(m)}_k\right)^j\right] \\
&= \sum_{\substack{a,b\geq 0\\ a+b=i}} \binom{i}{a} 2^a N^b \sum_{(a_1,\ldots,a_N)|a} \; \sum_{(b_1,\ldots,b_N)|b} \; \sum_{(j_1,\ldots,j_N)|j} \\
&\qquad \binom{a}{a_1,\ldots,a_N}\binom{b}{b_1,\ldots,b_N}\binom{j}{j_1,\ldots,j_N} \prod_{m=1}^N \mathbb{E}\left[\left(\frac{\sigma\overline{\eta^{(m)}_k}}{Nr_k}\right)^{a_m+2b_m} \left(\frac{\sigma\eta^{(m)}_k}{Nr_k}\right)^{j_m}\right]
\end{aligned}$$

where the last line uses that there are $\binom{a}{a_1,\ldots,a_N}$ ways to choose a_1 of the factors $(\sum_{m=1}^N \frac{\sigma}{Nr_k}\overline{\eta^{(m)}_k})^a$ to correspond to $m = 1$, a_2 to correspond to $m = 2$, etc., and similarly for b and j.

Then, applying (60) and (61),

$$\left|\mathbb{E}\left[\delta^i_{k,l}\delta^j_{x,y}\right]\right| \leq \sum_{\substack{a,b\geq 0\\ a+b=i}} \binom{i}{a} 2^a N^b \sum_{(a_1,\ldots,a_N)|a} \; \sum_{(b_1,\ldots,b_N)|b} \; \sum_{(j_1,\ldots,j_N)|j}$$

$$\binom{a}{a_1,\dots,a_N}\binom{b}{b_1,\dots,b_N}\binom{j}{j_1,\dots,j_N}\prod_{m=1}^N \mathbb{1}\{a_m+2b_m=j_m\}\left(\frac{4\sigma^2}{N^2r_k^2}\right)^{j_m} j_m!$$

Observe that $\binom{j}{j_1,\dots,j_N}\cdot\prod_{m=1}^N j_m! = j! \le j^j$. The condition $a_m+2b_m=j_m$ for every $m=1,\dots,N$ requires $a+2b=\sum_m a_m+2b_m=\sum_m j_m=j$. For (a,b,j) satisfying this requirement, fixing any partitions $(a_1,\dots,a_N)|a$ and $(b_1,\dots,b_N)|b$, there is exactly one partition $(j_1,\dots,j_N)|j$ for which $a_m+2b_m=j_m$ holds for every $m=1,\dots,N$. Thus, the above gives

$$\left|\mathbb{E}\left[\delta_{k,l}^i\delta_{x,y}^j\right]\right| \le \left(\frac{4j\sigma^2}{N^2\underline{r}^2}\right)^j \sum_{\substack{a,b\ge 0\\ a+b=i,\, a+2b=j}} \binom{i}{a}2^aN^b \sum_{(a_1,\dots,a_N)|a}\ \sum_{(b_1,\dots,b_N)|b}\binom{a}{a_1,\dots,a_N}\binom{b}{b_1,\dots,b_N}.$$

Observe now that $\sum_{(a_1,\dots,a_N)|a}\binom{a}{a_1,\dots,a_N}$ counts exactly the number of assignments of each of a labeled objects to N bins, by first determining the number of objects in each bin, followed by their identities. So $\sum_{(a_1,\dots,a_N)|a}\binom{a}{a_1,\dots,a_N}=N^a$. Applying the similar identity for b and $N^{a+2b}=N^j$ above,

$$\left|\mathbb{E}\left[\delta_{k,l}^i\delta_{x,y}^j\right]\right| \le \left(\frac{4j\sigma^2}{N\underline{r}^2}\right)^j \sum_{\substack{a,b\ge 0\\ a+b=i,\, a+2b=j}} \binom{i}{a}2^a.$$

Then, recalling that $\mathbb{E}\left[\delta_{k,l}^i\overline{\delta_{x,y}^j}\right]=0$,

$$\Big|\mathbb{E}[q(\delta_{k,l})q(\delta_{x,y})]\Big| \le \sum_{i,j=1}^J \frac{1}{2ij}\left|\mathbb{E}\left[\delta_{k,l}^i\delta_{x,y}^j\right]\right| \le \sum_{j=1}^J\frac{1}{2j}\left(\frac{4j\sigma^2}{N\underline{r}^2}\right)^j\sum_{\substack{a,b\ge 0\\ a+2b=j}}\frac{1}{a+b}\binom{a+b}{a}2^a.$$

We may apply

$$\sum_{\substack{a,b\ge 0\\ a+2b=j}}\frac{1}{a+b}\binom{a+b}{a}2^a \le \sum_{a=0}^j\binom{j}{a}2^a = 3^j.$$

Then, recalling the definition of J from (69) where $C_0\ge 3$,

$$\Big|\mathbb{E}[q(\delta_{k,l})q(\delta_{x,y})]\Big| \le \sum_{j=1}^J\frac{1}{2j}\left(\frac{12j\sigma^2}{N\underline{r}^2}\right)^j \le \frac{6\sigma^2}{N\underline{r}^2}\sum_{j=1}^J\left(\frac{12J\sigma^2}{N\underline{r}^2}\right)^{j-1} \le \frac{6\sigma^2}{N\underline{r}^2}\sum_{j=1}^\infty\left(\frac{3}{C_0e}\right)^{j-1}.$$

This again yields (71), and concludes the proof in Case 2.

Case 3: $k=x$. (By symmetry, this addresses also $k=y$, $l=x$, and $l=y$.) This ensures that $\{k+l,k+y\}\cap\{k,l,y\}=\emptyset$ and $k+l\neq k+y$, because $k+l$ is greater than $\{k,l\}$, $k+y$ is greater than $\{k,y\}$, and $k+l=y$ or $k+y=l$ or $k+l=k+y$ would lead to $\{k,l,k+l\}\cap\{x,y,x+y\}$ having cardinality 2. We may expand $\delta_{k,l}^i\delta_{x,y}^j$ as monomials in $\eta_{k+l}^{(m)},\overline{\eta_k^{(m)}},\overline{\eta_l^{(m)}},\eta_{k+y}^{(m)},\overline{\eta_y^{(m)}}$. Since $\{k+l,k+y\}$ are distinct from $\{k,l,y\}$, (60) implies $\mathbb{E}[\delta_{k,l}^i\delta_{x,y}^j]=0$. We may also expand $\delta_{k,l}^i\overline{\delta_{x,y}^j}$ as monomials in $\eta_{k+l}^{(m)},\overline{\eta_k^{(m)}},\overline{\eta_l^{(m)}},\overline{\eta_{k+y}^{(m)}},\eta_k^{(m)},\eta_y^{(m)}$. Since $k+l$ is distinct from $\{k,l,k+y\}$ and $k+y$ is distinct from $\{k,y,k+l\}$, (60) implies that any monomials involving $\eta_{k+l}^{(m)}$ or $\overline{\eta_{x+y}^{(m)}}$ have vanishing expectation. Note that since $k+l\neq k+y$, also $l\neq y$. If k is distinct from $\{l,y\}$, then monomials involving $\overline{\eta_l^{(m)}}$ or $\eta_y^{(m)}$ also have vanishing expectation, so

$$\mathbb{E}[\delta_{k,l}^i\overline{\delta_{x,y}^j}] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{m=1}^N\frac{\sigma}{r_k}\overline{\eta_k^{(m)}}\right)^i\left(\frac{1}{N}\sum_{m=1}^N\frac{\sigma}{r_k}\eta_k^{(m)}\right)^j\right].$$

This is analogous to (70), and the same argument as in Case 1 leads to (71). If instead $k=l$ (which by symmetry addresses also $k=y$), then

$$\mathbb{E}[\delta_{k,l}^i\overline{\delta_{x,y}^j}] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{m=1}^N\frac{2\sigma}{r_k}\overline{\eta_k^{(m)}}+\frac{\sigma^2}{r_k^2}(\overline{\eta_k^{(m)}})^2\right)^i\left(\frac{1}{N}\sum_{m=1}^N\frac{\sigma}{r_k}\eta_k^{(m)}\right)^j\right].$$

This is analogous to (72), and the same argument as in Case 2 leads to (71). This concludes the proof in Case 3. Combining these three cases shows (28). □

A.3 Oracle estimation of ϕ_k

Proof of Lemma 4.7. Suppose $M\phi = 0$. Then for all $(k,l) \in \mathcal{I}$, $\phi_{k+l} = \phi_k + \phi_l$. Then $\phi_2 = \phi_1 + \phi_1 = 2\phi_1$, $\phi_3 = \phi_2 + \phi_1 = 3\phi_1$, etc., and $\phi_K = \phi_{K-1} + \phi_1 = K\phi_1$, so ϕ is a multiple of $(1, 2, 3, \ldots, K)$. Conversely, any multiple of $(1, 2, 3, \ldots, K)$ satisfies $M\phi = 0$ by the definition of M. This shows the first statement.

Denote $T = M^\top M$. We explicitly compute T: Let $M_k \in \mathbb{R}^{\mathcal{I}}$ be the k^{th} column of M. The diagonal entries of T are $T_{kk} = \|M_k\|^2$. If $2k > K$, then the non-zero entries of M_k correspond to the (i,j) pairs

$$\begin{aligned}(i,j) = (1, k-1), \ldots, (k-1, 1) &: M_{(i,j),k} = 1\\ (i,j) = (1, k), \ldots, (K-k, k) &: M_{(i,j),k} = -1\\ (i,j) = (k, 1), \ldots, (k, K-k) &: M_{(i,j),k} = -1\end{aligned}$$

So $T_{kk} = (k-1) + 2(K-k) = 2K - 1 - k$. If $2k \le K$, then the non-zero entries of M_k correspond to the (i,j) pairs

$$\begin{aligned}(i,j) = (1, k-1), \ldots, (k-1, 1) &: M_{(i,j),k} = 1\\ (i,j) = (1, k), \ldots, (k-1, k), (k+1, k), \ldots, (K-k, k) &: M_{(i,j),k} = -1\\ (i,j) = (k, 1), \ldots, (k, k-1), (k, k+1), \ldots, (k, K-k) &: M_{(i,j),k} = -1\\ (i,j) = (k,k) &: M_{(i,j),k} = -2\end{aligned}$$

So $T_{kk} = (k-1) + 2(K-k-1) + 4 = 2K + 1 - k$. Thus, for all $k = 1, \ldots, K$,

$$T_{kk} = 2K + 1 - k - 2 \cdot \mathbb{1}\{2k > K\}.$$

For $1 \le j < k \le K$, when $j + k > K$, we have $T_{jk} = M_j^\top M_k = -2$ where the only non-zero contributions to this inner-product come from rows $(k-j, j), (j, k-j) \in \mathcal{I}$. When $j + k \le K$, the non-zero contributions come from rows $(k-j, j), (j, k-j), (j, k), (k, j) \in \mathcal{I}$, and these cancel exactly to yield $T_{jk} = M_j^\top M_k = 0$. Combining these diagonal and off-diagonal components, T has the form

$$T = \begin{pmatrix} 2K & & & \\ & 2K-1 & & \\ & & \ddots & \\ & & & K+1 \end{pmatrix} - \begin{pmatrix} & & & 2 \\ & & 2 & 2 \\ & \iddots & \vdots & \vdots \\ 2 & \ldots & 2 & 2 \end{pmatrix}$$

where the second matrix accounts also for the term $-2 \cdot \mathbb{1}\{2k > K\}$ of the diagonal entries.

Now let $\lambda > 0$ be a positive eigenvalue of T, with non-zero eigenvector $x = (x_1, x_2, \ldots, x_K)$. This must be orthogonal to the null vector $(1, 2, \ldots, K)$ so $x_1 + 2x_2 + \ldots + Kx_K = 0$. From the above form of T, the equation $Tx = \lambda x$ may be arranged as the linear system

$$\begin{aligned}(2K - \lambda)x_1 &= 2x_K\\ (2K - 1 - \lambda)x_2 &= 2(x_{K-1} + x_K)\\ &\vdots\\ (K + 2 - \lambda)x_{K-1} &= 2(x_2 + \ldots + x_K)\\ (K + 1 - \lambda)x_K &= 2(x_1 + x_2 + \ldots + x_K).\end{aligned}$$

Summing these equations and adding $x_1 + 2x_2 + \ldots + Kx_K$ to both sides, we obtain

$$(2K + 1 - \lambda)(x_1 + x_2 + \ldots + x_K) = 2(x_1 + 2x_2 + \ldots + Kx_K) + (x_1 + 2x_2 + \ldots + Kx_K) = 0.$$

If $x_1 + x_2 + \ldots + x_K \neq 0$, then this implies $\lambda = 2K+1$. If $x_1 + x_2 + \ldots + x_K = 0$, but $\lambda \notin \{K+1, K+2, \ldots, 2K\}$, then from the above linear system, we have the implications

$$(K + 1 - \lambda)x_K = 2(x_1 + x_2 + \ldots + x_K) = 0 \Rightarrow x_K = 0$$

$$(2K-\lambda)x_1 = 2x_K = 0 \Rightarrow x_1 = 0$$
$$(K+2-\lambda)x_{K-1} = 2(x_2+x_3+\ldots+x_K) = 2(0-x_1) = 0 \Rightarrow x_{K-1} = 0$$
$$(2K-1-\lambda)x_2 = 2(x_{K-1}+x_K) = 0 \Rightarrow x_2 = 0$$

and so forth. Then $x_1 = x_2 = \ldots = x_K = 0$, which contradicts $x \neq 0$. Thus any positive eigenvalue λ of T is one of the values $\{K+1, K+2, \ldots, 2K+1\}$. □

Proof of Lemma 4.1. Recall the loss upper bound from Proposition 3.1. For constants $C, C' > 0$, applying $ab \leq a^2 + b^2$,

$$\begin{aligned}
L(\hat{\theta}^{\text{oracle}}, \theta^*) &\leq \sum_{k=1}^{K}(\hat{r}_k - r_k)^2 + C \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K} \hat{r}_k r_k \left|\hat{\phi}_k^{\text{oracle}} - \phi_k + k\alpha\right|_{\mathcal{A}}^2 \\
&= \sum_{k=1}^{K}(\hat{r}_k - r_k)^2 + C \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K} [r_k^2 + r_k(\hat{r}_k - r_k)] \left|\hat{\phi}_k^{\text{oracle}} - \phi_k + k\alpha\right|_{\mathcal{A}}^2 \\
&\leq (C+1)\sum_{k=1}^{K}(\hat{r}_k - r_k)^2 + C \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K} r_k^2 \left(\left|\hat{\phi}_k^{\text{oracle}} - \phi_k + k\alpha\right|_{\mathcal{A}}^2 + \left|\hat{\phi}_k^{\text{oracle}} - \phi_k + k\alpha\right|_{\mathcal{A}}^4\right) \\
&\leq C' \left(\sum_{k=1}^{K}(\hat{r}_k - r_k)^2 + r^2 \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K} \left|\hat{\phi}_k^{\text{oracle}} - \phi_k + k\alpha\right|_{\mathcal{A}}^2\right).
\end{aligned}$$

The expectation may be bounded using Corollaries 4.3 and 4.8. For the bound (29) of Corollary 4.8, fixing $c > 0$ be the constant in the exponent, observe that the given condition for N with $C_0 > 0$ large enough implies

$$c\left(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2}\right) \geq \frac{c}{2}\left(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2}\right) + 3\log K.$$

We may then apply $e^{-(c/2)x}, e^{-(c/2)x^{2/3}} \leq C/x$ for a constant $C > 0$ to obtain

$$e^{-c(\frac{N\underline{r}^2}{\sigma^2} \wedge \frac{N\underline{r}^6}{\sigma^6} \wedge \frac{N^{2/3}\underline{r}^2}{\sigma^2})} \leq \frac{C}{K^3}\left(\frac{\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6} + \frac{\sigma^3}{N\underline{r}^3}\right) \leq \frac{2C}{K^3}\left(\frac{\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6}\right). \tag{73}$$

Then Corollary 4.8 gives simply

$$\mathbb{E}\left[\inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K} |\hat{\phi}_k^{\text{oracle}} - \phi_k + k\alpha|_{\mathcal{A}}^2\right] \leq C'\left(\frac{K\sigma^2}{Nr^2} + \frac{\sigma^6}{Nr^6}\right)$$

and combining this with Corollary 4.3 yields the lemma. □

A.4 Estimation of $\Phi_{k,l}$ by optimization

Proof of Lemma 4.10. Set $v = \phi - \phi'$. We must show: If $v \in \mathbb{R}^K$ is such that

$$|v_{k+l} - v_k - v_l|_{\mathcal{A}} \leq \delta \text{ for all } (k,l) \in \mathcal{I} \tag{74}$$

then there exists $\alpha \in \mathbb{R}$ with $|v_k - k\alpha|_{\mathcal{A}} \leq \delta$ for all $k = 1, \ldots, K$.

We induct on K. For $K = 1$ the result holds trivially by setting $\alpha = v_1$. Suppose the result holds for $K-1$. Consider $v \in \mathbb{R}^K$ that satisfies (74). By the induction hypothesis, there exists $\alpha \in \mathbb{R}$ such that $|v_k - k\alpha|_{\mathcal{A}} \leq \delta$ for $k = 1, \ldots, K-1$. Then by the triangle inequality, it is immediate to see that

$$|v_K - K\alpha|_{\mathcal{A}} \leq |v_K - v_1 - v_{K-1}|_{\mathcal{A}} + |v_1 - \alpha|_{\mathcal{A}} + |v_{K-1} - (K-1)\alpha|_{\mathcal{A}} \leq 3\delta.$$

To complete the induction, we must show the stronger bound of δ instead of 3δ.

For this, let

$$\alpha_* = \arg\min_{\alpha\in\mathbb{R}}\left(\max_{k=1}^{K-1} |v_k - k\alpha|_{\mathcal{A}}\right), \qquad \varepsilon_k = |v_k - k\alpha_*|_{\mathcal{A}} \text{ for } k = 1, \ldots, K-1, \qquad \varepsilon = \max_{k=1}^{K-1} \varepsilon_k.$$

Note that a minimizing α_* exists because the minimum may equivalently be restricted to the compact domain $[-\pi, \pi]$. The induction hypothesis implies $\varepsilon \le \delta$. By definition of $|\cdot|_{\mathcal{A}}$, there exists $j_k \in \mathbb{Z}$ for each $k = 1, \ldots, K-1$ such that

$$\varepsilon_k = |v_k - k\alpha_* + 2\pi j_k|.$$

Furthermore, we claim that there must exist two indices $k, l \in \{1, \ldots, K-1\}$ for which

$$\varepsilon = v_k - k\alpha_* + 2\pi j_k \quad \text{and} \quad -\varepsilon = v_l - l\alpha_* + 2\pi j_l.$$

This is because

$$\Big\{k \in \{1, \ldots, K-1\} : \varepsilon_k = \varepsilon\Big\} = \Big\{k \in \{1, \ldots, K-1\} : |v_k - k\alpha_* + 2\pi j_k| = \varepsilon\Big\}$$

is non-empty by definition of ε. If $v_k - k\alpha_* + 2\pi j_k = \varepsilon$ for every k belonging to this set, then we may decrease the value of $\max_{k=1}^{K-1} |v_k - k\alpha_*|_{\mathcal{A}}$ by slightly increasing α_*, which contradicts the optimality of α_*. Similarly if $v_k - k\alpha_* + 2\pi j_k = -\varepsilon$ for all k in this set, then we may decrease $\max_{k=1}^{K-1} |v_k - k\alpha_*|_{\mathcal{A}}$ by slightly decreasing α_*, again contradicting the optimality of α_*. Thus the claimed indices k, l exist.

Then, for this index $k \in \{1, \ldots, K-1\}$, we have

$$v_k - k\alpha_* \in 2\pi\mathbb{Z} + \varepsilon, \qquad v_{K-k} - (K-k)\alpha_* \in 2\pi\mathbb{Z} + [-\varepsilon, \varepsilon], \qquad v_K - v_k - v_{K-k} \in 2\pi\mathbb{Z} + [-\delta, \delta].$$

Adding these three conditions and applying $\varepsilon \le \delta$,

$$v_K - K\alpha_* \in 2\pi\mathbb{Z} + [-\delta, 2\varepsilon + \delta] \subseteq 2\pi\mathbb{Z} + [-\delta, 3\delta].$$

Similarly, for this index $l \in \{1, \ldots, K-1\}$, we have

$$v_l - l\alpha_* \in 2\pi\mathbb{Z} - \varepsilon, \qquad v_{K-l} - (K-l)\alpha_* \in 2\pi\mathbb{Z} + [-\varepsilon, \varepsilon], \qquad v_K - v_l - v_{K-l} \in 2\pi\mathbb{Z} + [-\delta, \delta].$$

Then adding these conditions, also

$$v_K - K\alpha_* \in 2\pi\mathbb{Z} + [-2\varepsilon - \delta, \delta] \subseteq 2\pi\mathbb{Z} + [-3\delta, \delta].$$

Since $3\delta < \pi$ strictly, the above two conditions combine to show that $v_K - K\alpha_* \in 2\pi\mathbb{Z} + [-\delta, \delta]$, i.e. $|v_K - K\alpha_*|_{\mathcal{A}} \le \delta$. This completes the induction. ☐

Proof of Lemma 4.9. Let $\mathcal{E}$ be the event where (33) holds. Note that this is exactly the event where $|\hat{\Phi}_{k,l}^{\text{oracle}}(\phi) - \Phi_{k,l}| < \pi/12$ for all $(k,l) \in \mathcal{I}$. Then by Lemma 4.4 and a union bound,

$$\mathbb{P}[\mathcal{E}^c] \le CK^2\left(e^{-cNr^2/\sigma^2} + e^{-cNr^6/\sigma^6} + e^{-N^{2/3}r^2/\sigma^2}\right) \le \frac{C'}{K}\left(\frac{\sigma^2}{Nr^2} + \frac{\sigma^6}{Nr^6}\right)$$

where the second inequality holds under the given condition for N as argued in (73). Applying Corollaries 4.11 and 4.8,

$$\begin{aligned}
&\mathbb{E}\left[\inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K |\hat{\phi}_k^{\text{opt}} - \phi_k + k\alpha|_{\mathcal{A}}^2\right] \\
&= \mathbb{E}\left[\mathbb{1}\{\mathcal{E}\} \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K |\hat{\phi}_k^{\text{opt}} - \phi_k + k\alpha|_{\mathcal{A}}^2\right] + \mathbb{E}\left[\mathbb{1}\{\mathcal{E}^c\} \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K |\hat{\phi}_k^{\text{opt}} - \phi_k + k\alpha|_{\mathcal{A}}^2\right] \\
&\le \mathbb{E}\left[\inf_{\alpha \in \mathbb{R}} \sum_{k=1}^K |\hat{\phi}_k^{\text{oracle}}(\phi') - \phi'_k + k\alpha|_{\mathcal{A}}^2\right] + CK \cdot \mathbb{P}[\mathcal{E}^c] \le C'\left(\frac{K\sigma^2}{Nr^2} + \frac{\sigma^6}{Nr^6}\right).
\end{aligned}$$

This is (up to a universal constant) the same risk bound as established for the oracle estimator itself in Corollary 4.8. The remainder of the proof is then the same as that of Lemma 4.1. ☐

A.5 Estimation of $\Phi_{k,l}$ by frequency marching

We describe in this section an alternative frequency marching method for mimicking the oracle estimator, which is more explicit and computationally efficient but requires a larger sample size N to succeed. Let

$$\tilde{\phi}_1 = 0$$

and, for each $k = 2, \ldots, K$, set

$$\tilde{\phi}_k = \operatorname{Arg} \hat{B}_{1,k-1} + \tilde{\phi}_{k-1} \bmod 2\pi.$$

This defines a vector $\tilde{\phi} \in [-\pi, \pi)^K$, which we use in place of (31). Then, as in Section 4.2, define $\tilde{\Phi}_{k,l} = \tilde{\phi}_{k+l} - \tilde{\phi}_k - \tilde{\phi}_l$ with arithmetic carried out over $\mathbb{R}$, and choose $\hat{\Phi}^{\mathrm{fm}}_{k,l} \in [\tilde{\Phi}_{k,l} - \pi, \tilde{\Phi}_{k,l} + \pi)$ as the unique version of the phase of $\hat{B}_{k,l}$ belonging to this range. Finally, let $\hat{\phi}^{\mathrm{fm}}$ be the resulting least-squares estimate of ϕ in (18), and let $\hat{\theta}^{\mathrm{fm}}$ be the resulting estimate of θ. Again, this procedure uses the frequency-marching estimate $\tilde{\phi}$ only as a pilot estimate to resolve the phase ambiguity of the estimated bispectrum, which is then inverted using a least-squares approach.

The following lemma is analogous to Corollary 4.11, but requires an improvement for the error of $\operatorname{Arg} \hat{B}_{k,l}$ by a factor of $1/K$.

Lemma A.2. *Suppose*

$$|\operatorname{Arg} \hat{B}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|_{\mathcal{A}} < \pi/(6K) \text{ for every } (k,l) \in \mathcal{I}. \tag{75}$$

Then there exists ϕ' equivalent to ϕ such that $\hat{\Phi}^{\mathrm{fm}} = \hat{\Phi}^{\mathrm{oracle}}(\phi')$.

Proof. For each $k = 2, \ldots, K$, by the definition of $\tilde{\phi}_k$ and the triangle inequality,

$$\left|\tilde{\phi}_k - \phi_k + k\phi_1\right|_{\mathcal{A}} \le \left|\operatorname{Arg} \hat{B}_{1,k-1} - (\phi_k - \phi_1 - \phi_{k-1})\right|_{\mathcal{A}} + \left|\tilde{\phi}_{k-1} - \phi_{k-1} + (k-1)\phi_1\right|_{\mathcal{A}}.$$

Under the given condition, recursively applying this bound and using $\tilde{\phi}_1 - \phi_1 + \phi_1 = 0$ for $k = 1$,

$$\left|\tilde{\phi}_k - \phi_k + k\alpha\right|_{\mathcal{A}} \le \frac{\pi(k-1)}{6K} < \frac{\pi}{6} \text{ for } \alpha = \phi_1 \text{ and all } k = 1, \ldots, K.$$

This means there exists ϕ' equivalent to ϕ for which $|\tilde{\phi}_k - \phi'_k| < \pi/6$ for all $k = 1, \ldots, K$, and the remainder of the argument is the same as in Corollary 4.11. □

The following guarantee is then analogous to Lemma 4.9, now describing the estimator $\hat{\theta}^{\mathrm{fm}}$ under a requirement for N that is larger by a factor of K^2.

Proposition A.3. *Suppose $\underline{c}r \le r_k \le \bar{c}r$ for each $k = 1, \ldots, K$. There exist constants $C, C_0 > 0$ depending only on $\underline{c}, \bar{c}$ such that if $\frac{Nr^6}{\sigma^6} \wedge \frac{N^{2/3}r^2}{\sigma^2} \ge C_0 K^2 \log K$, then the guarantee (22) holds also for $\hat{\theta}^{\mathrm{fm}}$.*

Proof. Let $\mathcal{E}$ be the event where (75) holds. This is exactly the event where $|\hat{\Phi}^{\mathrm{oracle}}_{k,l}(\phi) - \Phi_{k,l}| < \pi/(6K)$ for all $(k,l) \in \mathcal{I}$. Then by Lemma 4.4 and a union bound,

$$\mathbb{P}[\mathcal{E}^c] \le CK^2 \left(e^{-cNr^2/K^2\sigma^2} + e^{-cNr^6/K^2\sigma^6} + e^{-N^{2/3}r^2/K^{2/3}\sigma^2}\right).$$

Under the given condition for N, an argument similar to (73) shows that this implies

$$\mathbb{P}[\mathcal{E}^c] \le \frac{C'}{K}\left(\frac{\sigma^2}{Nr^2} + \frac{\sigma^6}{Nr^6}\right),$$

and the remainder of the proof is the same as that of Lemma 4.9. □

B Proofs for maximum likelihood estimation in low noise

B.1 KL divergence and tail bound for the MLE

The following lemma bounds the Gaussian process $\langle \varepsilon, g(\alpha)\cdot\theta\rangle$ which appears in (42).

Lemma B.1. *Let $\varepsilon \sim \mathcal{N}(0, I_{2K})$. For a universal constant $C > 0$, any $\theta \in \mathbb{R}^{2K}$, and any $s, t > 0$,*

$$\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > t \text{ and } \|\varepsilon\| \le s\right] \le \frac{8\pi\|\theta\|Ks}{t}\cdot e^{-\frac{t^2}{8\|\theta\|^2}} \tag{76}$$

$$\mathbb{E}\left[\sup_{\alpha\in\mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle|\right] \le C\|\theta\|\sqrt{\log K}. \tag{77}$$

Proof. For each fixed $\alpha \in \mathcal{A}$, we have $\langle \varepsilon, g(\alpha)\cdot\theta\rangle \sim \mathcal{N}(0, \|\theta\|^2)$. Thus by a Gaussian tail bound,

$$\mathbb{P}[|\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > t/2] \le 2e^{-\frac{t^2}{8\|\theta\|^2}}.$$

We set $\delta = t/(2\|\theta\|Ks)$ and take $N_\delta \subset \mathcal{A}$ as a δ-net of $\mathcal{A} = [-\pi, \pi)$ in the metric $|\cdot|_{\mathcal{A}}$, having cardinality

$$|N_\delta| = \frac{2\pi}{\delta} = \frac{4\pi\|\theta\|Ks}{t}.$$

For any $\alpha, \alpha' \in \mathcal{A}$ such that $|\alpha - \alpha'|_{\mathcal{A}} \le \delta$, from the definition (3) of the diagonal blocks of $g(\alpha)$, we have $\|g(\alpha) - g(\alpha')\|_{\text{op}} \le K\delta$. Thus, on the event $\{\|\varepsilon\| \le s\}$,

$$\big|\langle \varepsilon, g(\alpha)\cdot\theta\rangle - \langle \varepsilon, g(\alpha')\cdot\theta\rangle\big| \le K\delta s\|\theta\| = t/2.$$

So

$$\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > t \text{ and } \|\varepsilon\| \le s\right] \le \mathbb{P}\left[\sup_{\alpha\in N_\delta} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > t/2\right] \le |N_\delta|\cdot 2e^{-\frac{t^2}{8\|\theta\|^2}}$$

which yields (76). Applying (76) with $s = \sqrt{4K}$ and integrating from $t = 4\|\theta\|\sqrt{\log K}$ to $t = \infty$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\alpha\in\mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle|\cdot \mathbf{1}\{\|\varepsilon\| \le \sqrt{4K}\}\right] \\
&\le \mathbb{E}\left[4\|\theta\|\sqrt{\log K} + \int_{4\|\theta\|\sqrt{\log K}}^{\infty} \mathbf{1}\left\{\sup_{\alpha\in\mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > t \text{ and } \|\varepsilon\| \le \sqrt{4K}\right\} dt\right] \\
&\le 4\|\theta\|\sqrt{\log K} + \int_{4\|\theta\|\sqrt{\log K}}^{\infty} \frac{16\pi\|\theta\|K^{3/2}}{t} e^{-t^2/8\|\theta\|^2} dt \\
&= 4\|\theta\|\sqrt{\log K} + 16\pi\|\theta\|K^{3/2}\int_{4\sqrt{\log K}}^{\infty} e^{-t^2/8} dt \le C\|\theta\|\sqrt{\log K}
\end{aligned}$$

for a universal constant $C > 0$ and any $K \ge 2$. Applying a chi-squared tail bound, we have also

$$\begin{aligned}
\mathbb{E}\left[\sup_{\alpha\in\mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle|\cdot \mathbf{1}\{\|\varepsilon\| \ge \sqrt{4K}\}\right] &\le \|\theta\|\cdot\mathbb{E}\Big[\|\varepsilon\|\cdot\mathbf{1}\{\|\varepsilon\| \ge \sqrt{4K}\}\Big] \\
&\le \|\theta\|\cdot\mathbb{E}[\|\varepsilon\|^2]^{1/2}\mathbb{P}[\|\varepsilon\| \ge \sqrt{4K}]^{1/2} \le \|\theta\|\cdot\sqrt{2K}\cdot e^{-cK}
\end{aligned}$$

for a universal constant $c > 0$. Combining the above gives (77). □

The next lemma formalizes the statement (51) obtained by a Taylor expansion around $\alpha = 0$.

Lemma B.2. *Suppose $\underline{c}r \le r_k(\theta^*) \le \bar{c}r$ for each $k = 1, \ldots, K$. Fix any constant $\delta_0 \in [0, \underline{c}^2/\bar{c}^2]$. Then there are constants $C, c > 0$ depending only on $\underline{c}, \bar{c}$ (and independent of δ_0) such that for all $\alpha \in [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$,*

$$cK^3r^2\alpha^2 \le \|\theta^*\|^2 - \langle \theta^*, g(\alpha)\cdot\theta^*\rangle \le CK^3r^2\alpha^2. \tag{78}$$

Furthermore, there is a constant $\iota > 0$ depending only on $\underline{c}, \bar{c}, \delta_0$ such that for all $\alpha \in [-\pi, \pi) \setminus [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$,

$$\langle \theta^*, g(\alpha)\cdot\theta^*\rangle \le (1-\iota)\|\theta^*\|^2. \tag{79}$$

Proof. We write as shorthand $r_k = r_k(\theta^*)$. From (4), observe that

$$\langle \theta^*,\, g(\alpha)\cdot\theta^*\rangle = \sum_{k=1}^{K} r_k^2 \cos k\alpha = \|\theta^*\|^2 - \sum_{k=1}^{K} r_k^2(1-\cos k\alpha). \tag{80}$$

This is an even function of α, so it suffices to consider $\alpha \in [0,\pi]$. Suppose first that $0 \le \alpha \le \delta_0/K$. By Taylor expansion around $\alpha = 0$,

$$\sum_{k=1}^{K} r_k^2(1-\cos k\alpha) = \sum_{k=1}^{K} r_k^2 \cdot \frac{k^2\alpha^2}{2} + r(\alpha),$$

where $|r(\alpha)| \le \sum_{k=1}^{K} r_k^2 (k\alpha)^3/6$. Observe that

$$\frac{\sum_{k=1}^{K} k^2 r_k^2/4}{\sum_{k=1}^{K} k^3 r_k^2/6} \ge \frac{\underline{c}^2/4}{\bar{c}^2/6} \cdot \frac{\sum_{k=1}^{K} k^2}{\sum_{k=1}^{K} k^3} = \frac{\underline{c}^2}{\bar{c}^2} \cdot \frac{2K+1}{K(K+1)} \ge \frac{\underline{c}^2}{K\bar{c}^2} \ge \frac{\delta_0}{K}.$$

Then for $0 \le \alpha \le \delta_0/K$, we have $|r(\alpha)| \le \sum_{k=1}^{K} r_k^2 (k\alpha)^2/4$. Applying this and the above Taylor expansion to (80) gives

$$\sum_{k=1}^{K} r_k^2 \cdot \frac{k^2\alpha^2}{4} \le \|\theta^*\|^2 - \langle \theta^*, g(\alpha)\cdot\theta^*\rangle \le \sum_{k=1}^{K} r_k^2 \cdot \frac{3k^2\alpha^2}{4},$$

which implies (78).

Now consider $\delta_0/K < \alpha \le \pi$. In the sequence $(\cos\alpha, \cos 2\alpha, \ldots, \cos K\alpha)$, we claim that there are at most $\lceil K/2 \rceil$ items belonging to the interval $[\cos L, 1]$, where $L = \min(\delta_0/2, \pi/8)$:

- If $\frac{\delta_0}{K} < \alpha < \frac{\pi}{K}$, then $\alpha, 2\alpha, \ldots, K\alpha \in (0,\pi)$. So $\cos(k\alpha) \in [\cos L, 1]$ implies that $k\alpha \in [0, L]$, and the number of such items is at most $L/\alpha \le (\delta_0/2)/(\delta_0/K) = K/2$.
- If $\frac{t\pi}{K} \le \alpha < \frac{(t+1)\pi}{K}$ for some $1 \le t \le \frac{K}{4} - 1$, then $\alpha, 2\alpha, \ldots, K\alpha \in (0, (t+1)\pi)$. So $\cos(k\alpha) \in [\cos L, 1]$ implies that $k\alpha$ falls into one of $t+1$ closed intervals of width L, and the number of such items is at most
$$(t+1)\cdot\left\lceil \frac{L}{\alpha} \right\rceil \le (t+1)\cdot\left(\frac{\pi/8}{t\pi/K} + 1\right) = \frac{K}{8} + \frac{K}{8t} + t + 1 \le \frac{K}{8} + \frac{K}{8} + \frac{K}{4} = \frac{K}{2}.$$
- If $\frac{\pi}{4} < \alpha \le \pi$, then any two consecutive items $\cos k\alpha$ and $\cos(k+1)\alpha$ cannot both belong to $[\cos L, 1]$, since $\alpha > \frac{\pi}{4} \ge 2L$. Therefore, the number of items would not exceed $\lceil K/2 \rceil$.

Denoting $B = \{k : \cos k\alpha \notin [\cos L, 1]\}$, we then have $|B| \ge \lfloor K/2 \rfloor$ and $1 - \cos L \ge c$ a small constant depending on $\underline{c}, \bar{c}, \delta_0$, so

$$\sum_{k=1}^{K} r_k^2(1-\cos k\alpha) \ge \sum_{k\in B} r_k^2(1-\cos L) \ge c\lfloor K/2 \rfloor (\underline{c}r)^2 \ge \iota\|\theta^*\|^2$$

for a constant $\iota > 0$ depending only on $\underline{c}, \bar{c}, \delta_0$. Applying this to (80) gives (79). □

Proof of Lemma 5.3. Recall the form (50) for the KL divergence. For II, upper bounding the average over α by the maximum,

$$\begin{aligned} \mathrm{II} \le \mathbb{E} \log \sup_{\alpha\in\mathcal{A}} \exp\left(\frac{\langle \theta^* + \sigma\varepsilon, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) &\le \frac{\sup_{\alpha\in\mathcal{A}} \langle \theta^*, g(\alpha)\cdot\theta\rangle}{\sigma^2} + \mathbb{E}\left[\frac{\sup_{\alpha\in\mathcal{A}} \langle \varepsilon, g(\alpha)\cdot\theta\rangle}{\sigma}\right] \\ &\le \frac{\sup_{\alpha\in\mathcal{A}} \langle \theta^*, g(\alpha)\cdot\theta\rangle}{\sigma^2} + \frac{C\|\theta\|}{\sigma}\cdot\sqrt{\log K}, \end{aligned} \tag{81}$$

where the last inequality applies (77) from Lemma B.1. Similarly, to lower bound I, let us set $\delta_0 = \underline{c}^2/\bar{c}^2$ and apply

$$\mathrm{I} \ge \mathbb{E}\left[\log \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{\langle \theta^*, g(\alpha)\cdot\theta^*}{\sigma^2}\right) d\alpha \cdot \inf_{\alpha} \exp\left(\frac{\varepsilon, g(\alpha)\cdot\theta^*}{\sigma}\right)\right]$$

$$= \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{\langle \theta^*, g(\alpha) \cdot \theta^* \rangle}{\sigma^2}\right) d\alpha - \mathbb{E}\left[\frac{\sup_{\alpha \in \mathcal{A}} \langle \varepsilon, g(\alpha) \cdot \theta^* \rangle}{\sigma}\right]$$
$$\geq \log \frac{1}{2\pi} \int_{-\delta_0/K}^{\delta_0/K} \exp\left(\frac{\langle \theta^*, g(\alpha) \cdot \theta^* \rangle}{\sigma^2}\right) d\alpha - \frac{C\|\theta^*\|}{\sigma} \cdot \sqrt{\log K}$$

Applying the upper bound of (78) from Lemma B.2, we have for a constant $C > 0$ that

$$\int_{-\delta_0/K}^{\delta_0/K} \exp\left(\frac{\langle \theta^*, g(\alpha) \cdot \theta^* \rangle}{\sigma^2}\right) d\alpha \geq \exp\left(\frac{\|\theta^*\|^2}{\sigma^2}\right) \int_{-\delta_0/K}^{\delta_0/K} \exp\left(-\frac{CK^3 r^2}{\sigma^2}\alpha^2\right) d\alpha$$
$$= \exp\left(\frac{\|\theta^*\|^2}{\sigma^2}\right) \left(\frac{2CK^3r^2}{\sigma^2}\right)^{-1/2} \cdot \sqrt{2\pi}\left(1 - 2\widetilde{\Phi}\left(\sqrt{\frac{2CK^3r^2}{\sigma^2}} \cdot \frac{\delta_0}{K}\right)\right)$$

where $\widetilde{\Phi}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ is the right tail probability of the standard Gaussian law. Applying the given condition $Kr^2 \geq \sigma^2$, the input to $\widetilde{\Phi}$ is bounded below by a positive constant. Then the value for $\widetilde{\Phi}$ is bounded away from $1/2$, so for a constant $C_2 > 0$,

$$\frac{1}{2\pi} \int_{-\delta_0/K}^{\delta_0/K} \exp\left(\frac{\langle \theta^*, g(\alpha) \cdot \theta^* \rangle}{\sigma^2}\right) d\alpha \geq \exp\left(\frac{\|\theta^*\|^2}{\sigma^2}\right) \cdot \left(\frac{C_2 K^3 r^2}{\sigma^2}\right)^{-1/2}. \tag{82}$$

Thus

$$\mathrm{I} \geq \frac{\|\theta^*\|^2}{\sigma^2} - \frac{1}{2}\log\left(\frac{C_2K^3r^2}{\sigma^2}\right) - \frac{C\|\theta^*\|}{\sigma} \cdot \sqrt{\log K}. \tag{83}$$

Combining (50), (81), and (83) and applying

$$\min_{\alpha \in \mathcal{A}} \|\theta^* - g(\alpha) \cdot \theta\|^2 = \|\theta^*\|^2 + \|\theta\|^2 - 2 \sup_{\alpha \in \mathcal{A}} \langle \theta^*, g(\alpha) \cdot \theta \rangle$$

yields the lemma. □

The following lemma establishes concentration of $R_N(\theta)$ around its mean $R(\theta)$, uniformly over bounded domains of θ.

Lemma B.3. *For a universal constant $c > 0$, any $M > 0$, and any $t > 0$,*

$$\mathbb{P}\left[\sup_{\theta: \|\theta\| \leq M} |R_N(\theta) - R(\theta)| > 4t\right]$$
$$\leq 2\left(1 + \tfrac{2M\sqrt{2\|\theta^*\|^2 + (4K+4t)\sigma^2}}{t\sigma^2}\right)^{2K} e^{-\frac{cN\sigma^2t^2}{M^2}} + 4e^{-cN(t \wedge \frac{t^2}{K} \wedge \frac{t^2\sigma^2}{\|\theta^*\|^2})}. \tag{84}$$

Proof. Recalling the form of $R_N(\theta)$ from (42), we have

$$R_N(\theta) = \mathrm{I} - \mathrm{II}(\theta) + \mathrm{const}(\theta)$$

where

$$\mathrm{I} = \frac{1}{N}\sum_{m=1}^{N} \frac{\|\theta^* + \sigma\varepsilon^{(m)}\|^2}{2\sigma^2},$$
$$\mathrm{II}(\theta) = \frac{1}{N}\sum_{m=1}^{N} f(\varepsilon^{(m)}, \theta) := \frac{1}{N}\sum_{m=1}^{N} \log\frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(\frac{\langle \theta^* + \sigma\varepsilon^{(m)}, g(\alpha) \cdot \theta \rangle}{\sigma^2}\right) d\alpha,$$

and $\mathrm{const}(\theta)$ is a term not depending on the randomness $\{\varepsilon^{(m)}\}$. We analyze separately the concentration of the terms I and $\mathrm{II}(\theta)$.

For the given value $t > 0$, define the event $\mathcal{E} = \{|\mathrm{I} - \mathbb{E}[\mathrm{I}]| < 2t\}$. We have

$$\frac{\|\theta^* + \sigma\varepsilon^{(m)}\|^2}{2\sigma^2} = \frac{\|\theta^*\|^2}{2\sigma^2} + \frac{\langle \theta^*, \varepsilon^{(m)} \rangle}{\sigma} + \frac{\|\varepsilon^{(m)}\|^2}{2}.$$

Here the first term is deterministic, and the latter two terms satisfy $N^{-1}\sum_{m=1}^N \langle \theta^*, \varepsilon^{(m)} \rangle \sim \mathcal{N}(0, \|\theta^*\|^2/N)$ and $\sum_{m=1}^N \|\varepsilon^{(m)}\|^2 \sim \chi^2_{2KN}$. Then by standard Gaussian and chi-squared tail bounds, for a universal constant $c > 0$ and any $t > 0$,

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{m=1}^N \frac{\langle \theta^*, \varepsilon^{(m)} \rangle}{\sigma}\right| \geq t\right] \leq 2e^{-\frac{Nt^2\sigma^2}{2\|\theta^*\|^2}}, \qquad \mathbb{P}\left[\left|\frac{1}{N}\sum_{m=1}^N \frac{\|\varepsilon^{(m)}\|^2}{2} - K\right| \geq t\right] \leq 2e^{-cNK(\frac{t}{K} \wedge \frac{t^2}{K^2})}.$$

So

$$\mathbb{P}[\mathcal{E}^c] = \mathbb{P}\Big[|\mathrm{I} - \mathbb{E}[\mathrm{I}]| \geq 2t\Big] \leq 2e^{-\frac{Nt^2\sigma^2}{2\|\theta^*\|^2}} + 2e^{-cNK(\frac{t}{K} \wedge \frac{t^2}{K^2})} \leq 4e^{-c'N(t \wedge \frac{t^2}{K} \wedge \frac{t^2\sigma^2}{\|\theta^*\|^2})}. \tag{85}$$

On the event $\mathcal{E}$, we have $|R_N(\theta) - R(\theta)| \leq 2t + |\mathrm{II}(\theta) - \mathbb{E}[\mathrm{II}(\theta)]|$ as well as

$$\frac{1}{N}\sum_{m=1}^N \|\theta^* + \sigma\varepsilon^{(m)}\|^2 \leq \mathbb{E}[\|\theta^* + \sigma\varepsilon^{(m)}\|^2] + 4t\sigma^2 = \|\theta^*\|^2 + (2K + 4t)\sigma^2. \tag{86}$$

Recalling the probability law $\mathcal{P}_{\theta,\varepsilon}$ from (43), the ε-gradient of the function $f(\varepsilon, \theta)$ defining $\mathrm{II}(\theta)$ is bounded as

$$\|\nabla_\varepsilon f(\varepsilon, \theta)\| = \left\|\frac{1}{\sigma}\mathbb{E}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}\left[g(\alpha) \cdot \theta\right]\right\| \leq \frac{1}{\sigma}\mathbb{E}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}\left[\|g(\alpha) \cdot \theta\|\right] = \frac{\|\theta\|}{\sigma}.$$

Thus $f(\varepsilon, \theta)$ is $\frac{\|\theta\|}{\sigma}$-Lipschitz in ε. Then by Gaussian concentration of measure, for universal constants $C, c > 0$, we have that $f(\varepsilon, \theta) - \mathbb{E}[f(\varepsilon, \theta)]$ is $\frac{C\|\theta\|}{\sigma}$-subgaussian, and Hoeffding's inequality yields

$$\mathbb{P}\left[|\mathrm{II}(\theta) - \mathbb{E}[\mathrm{II}(\theta)]| > t\right] = \mathbb{P}\left[\left|\frac{1}{N}\sum_{m=1}^N f(\varepsilon^{(m)}, \theta) - \mathbb{E}[f(\varepsilon^{(m)}, \theta)]\right| > t\right] \leq 2e^{-\frac{cN\sigma^2t^2}{\|\theta\|^2}}.$$

Now to obtain uniform concentration over $\{\theta : \|\theta\| \leq M\}$, set $\delta = t\sigma^2/\sqrt{2\|\theta^*\|^2 + (4K + 4t)\sigma^2}$, and let N_δ be a δ-net of $\{\theta \in \mathbb{R}^{2K} : \|\theta\| \leq M\}$ having cardinality

$$|N_\delta| \leq \left(1 + \frac{2M}{\delta}\right)^{2K} = \left(1 + \frac{2M\sqrt{2\|\theta^*\|^2 + (4K + 4t)\sigma^2}}{t\sigma^2}\right)^{2K}.$$

The θ-gradient of $\mathbb{E}[\mathrm{II}(\theta)] = \mathbb{E}[f(\varepsilon^{(m)}, \theta)]$ is bounded as

$$\begin{aligned}
\|\nabla_\theta \mathbb{E}[\mathrm{II}(\theta)]\| &= \left\|\frac{1}{\sigma^2}\mathbb{E}_{\varepsilon \sim \mathcal{N}(0,I)}\left[\mathbb{E}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}[g(\alpha)^{-1}(\theta^* + \sigma\varepsilon)]\right]\right\| \\
&\leq \frac{1}{\sigma^2}\mathbb{E}[\|\theta^* + \sigma\varepsilon\|] \leq \frac{1}{\sigma^2}\sqrt{\mathbb{E}[\|\theta^* + \sigma\varepsilon\|^2]} = \frac{1}{\sigma^2}\sqrt{\|\theta^*\|^2 + 2K\sigma^2}.
\end{aligned}$$

Similarly, on the event $\mathcal{E}$, the θ-gradient of $\mathrm{II}(\theta)$ without expectation is bounded as

$$\begin{aligned}
\|\nabla_\theta \mathrm{II}(\theta)\| &= \left\|\frac{1}{\sigma^2} \cdot \frac{1}{N}\sum_{m=1}^N \mathbb{E}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon^{(m)}}}[g(\alpha)^{-1}(\theta^* + \sigma\varepsilon^{(m)})]\right\| \\
&\leq \frac{1}{\sigma^2} \cdot \frac{1}{N}\sum_{m=1}^N \|\theta^* + \sigma\varepsilon^{(m)}\| \leq \frac{1}{\sigma^2}\sqrt{\frac{1}{N}\sum_{m=1}^N \|\theta^* + \sigma\varepsilon^{(m)}\|^2} \leq \frac{1}{\sigma^2}\sqrt{\|\theta^*\|^2 + (2K + 4t)\sigma^2},
\end{aligned}$$

the last inequality applying (86). Therefore $\mathrm{II}(\theta) - \mathbb{E}[\mathrm{II}(\theta)]$ is Lipschitz in θ, with Lipschitz constant at most

$$\frac{1}{\sigma^2}\sqrt{\|\theta^*\|^2 + 2K\sigma^2} + \frac{1}{\sigma^2}\sqrt{\|\theta^*\|^2 + (2K + 4t)\sigma^2} \leq \frac{1}{\sigma^2}\sqrt{2\|\theta^*\|^2 + (4K + 4t)\sigma^2} = \frac{t}{\delta}.$$

Then

$$\mathbb{P}\left[\sup_{\theta:\|\theta\|\leq M}|R_N(\theta)-R(\theta)|>4t \text{ and } \mathcal{E}\right]\leq \mathbb{P}\left[\sup_{\theta:\|\theta\|\leq M}|\mathrm{II}(\theta)-\mathbb{E}[\mathrm{II}(\theta)]|>2t \text{ and } \mathcal{E}\right]$$
$$\leq \mathbb{P}\left[\sup_{\theta\in N_\delta}|\mathrm{II}(\theta)-\mathbb{E}[\mathrm{II}(\theta)]|>t\right]\leq 2|N_\delta|e^{-\frac{cN\sigma^2t^2}{M^2}}.$$

Combining this with (85) gives (84). □

Proof of Lemma 5.2. For the given value of δ_1 and each integer $n\geq 1$, define

$$\Gamma_n=\left\{\theta: n\delta_1\|\theta^*\|\leq \min_{\alpha\in\mathcal{A}}\|\theta^*-g(\alpha)\cdot\theta\|<(n+1)\delta_1\|\theta^*\|\right\}\subset\mathbb{R}^{2K}.$$

Observe that $\|\theta\|\leq[1+(n+1)\delta_1]\|\theta^*\|$ for $\theta\in\Gamma_n$, so Lemma 5.3 implies

$$D_{\mathrm{KL}}(p_{\theta^*}\|p_\theta)\geq\frac{n^2\delta_1^2\|\theta^*\|^2}{2\sigma^2}-\frac{1}{2}\log\left(\frac{C_2K^3r^2}{\sigma^2}\right)-\frac{[2+(n+1)\delta_1]C_3\|\theta^*\|}{\sigma}\sqrt{\log K}\text{ for all }\theta\in\Gamma_n.$$

When $\underline{c}r\leq r_k(\theta^*)\leq\bar{c}r$ for every $k=1,\ldots,K$, we have $\underline{c}^2Kr^2\leq\|\theta^*\|^2\leq\bar{c}^2Kr^2$. Then, under the given assumption that $\frac{Kr^2}{\sigma^2}\geq C_1\log K$ for a sufficiently large constant $C_1>0$ (depending on $\underline{c},\bar{c},\delta_1$), setting

$$t_n=c_0n^2Kr^2/\sigma^2$$

for a sufficiently small constant $c_0>0$, the above implies that

$$D_{\mathrm{KL}}(p_{\theta^*}\|p_\theta)\geq 10t_n\text{ for all }\theta\in\Gamma_n.$$

Applying (84) with $t=t_n$ and $M=[1+(n+1)\delta_1]\|\theta^*\|\leq Cn\sqrt{Kr^2}$ gives, for some constants $C,C',c,c'>0$ depending on δ_1,

$$\begin{aligned}&\mathbb{P}\left[\sup_{\theta:\|\theta\|\leq[1+(n+1)\delta_1]\|\theta^*\|}|R_N(\theta)-R(\theta)|>4t_n\right]\\&\leq 2\left(1+C\sqrt{1+\tfrac{\sigma^2}{n^2r^2}}\right)^{2K}e^{-cn^2N\cdot\frac{Kr^2}{\sigma^2}}+4e^{-cn^2N\cdot\frac{Kr^2}{\sigma^2}(1\wedge\frac{n^2r^2}{\sigma^2})}\\&\leq(C'K)^Ke^{-c'nN\log K}+e^{-c'nN(\log K)^2/K}\end{aligned}$$

where the second line applies $n\geq 1$, $\frac{Kr^2}{\sigma^2}\geq C_1\log K$, and $\frac{\sigma^2}{r^2}\leq\frac{K}{C_1\log K}$ to simplify the bound. On the event where $|R_N(\theta)-R(\theta)|\leq 4t_n$ and $|R_N(\theta^*)-R(\theta^*)|\leq 4t_n$, since $D_{\mathrm{KL}}(p_{\theta^*}\|p_\theta)=R(\theta)-R(\theta^*)\geq 10t_n$, we must then have $R_N(\theta)-R_N(\theta^*)\geq 2t_n>0$ so that θ is not the MLE. Thus,

$$\mathbb{P}\left[\hat{\theta}^{\mathrm{MLE}}\in\Gamma_n\right]\leq(C'K)^Ke^{-c'nN\log K}+e^{-c'nN(\log K)^2/K}. \tag{87}$$

Summing over all $n\geq 1$ and recalling our choice of rotation for $\hat{\theta}^{\mathrm{MLE}}$ that satisfies (36),

$$\mathbb{P}\left[\|\hat{\theta}^{\mathrm{MLE}}-\theta^*\|\geq\delta_1\|\theta^*\|\right]\leq\sum_{n=1}^\infty\mathbb{P}\left[\hat{\theta}^{\mathrm{MLE}}\in\Gamma_n\right]\leq(C'K)^K\sum_{n=1}^\infty e^{-c'nN\log K}+\sum_{n=1}^\infty e^{-c'nN(\log K)^2/K}.$$

Under the given assumption $N\geq C_0K$ for a sufficiently large constant $C_0>0$, both exponents $c'N\log K$ and $c'N(\log K)^2/K$ are bounded below by a constant. Then summing these geometric series gives, for some modified constants $C,c,c'>0$,

$$\mathbb{P}\left[\|\hat{\theta}^{\mathrm{MLE}}-\theta^*\|\geq\delta_1\|\theta^*\|\right]\leq(CK)^Ke^{-cN\log K}+e^{-cN(\log K)^2/K}\leq e^{-c'N(\log K)^2/K}$$

where the second inequality holds again under the assumption $N \geq C_0 K$. This shows (47).

To show (48), we apply $\|\theta\| \leq [1+(n+1)\delta_1]\|\theta^*\|$ for $\theta \in \Gamma_n$ and (87) to get, for some constants $C, C', c' > 0$ depending on δ_1,

$$\mathbb{E}[\|\hat{\theta}^{\text{MLE}}\|^4] \leq [(1+\delta_1)\|\theta^*\|]^4 + \sum_{n=1}^{\infty}[(1+(n+1)\delta_1)\|\theta^*\|]^4 \cdot \mathbb{P}[\hat{\theta}^{\text{MLE}} \in \Gamma_n]$$
$$\leq C\|\theta^*\|^4 \left(1 + (C'K)^K \sum_{n=1}^{\infty} n^4 e^{-c'nN\log K} + \sum_{n=1}^{\infty} n^4 e^{-c'nN(\log K)^2/K}\right).$$

For a sufficiently large constant $A > 0$, we have $n^4 e^{-An} < e^{-An/2}$ for all $n \geq 1$. Hence, under the condition $N \geq C_0 K$ for sufficiently large $C_0 > 0$, we have

$$\mathbb{E}[\|\hat{\theta}^{\text{MLE}}\|^4] \leq C\|\theta^*\|^4 \left(1 + (C'K)^K \sum_{n=1}^{\infty} e^{-(c'/2)nN\log K} + \sum_{n=1}^{\infty} e^{-(c'/2)nN(\log K)^2/K}\right)$$
$$\leq C\|\theta^*\|^4 \left(1 + (C'K)^K e^{-c''N\log K} + e^{-c''N(\log K)^2/K}\right) \leq C'\|\theta^*\|^4.$$

□

B.2 Lower bound for the information matrix

Define the domain

$$\mathcal{F}_1(\theta, \delta_1) = \left\{\varepsilon : \sup_{\alpha \in \mathcal{A}} |\langle \varepsilon, g(\alpha) \cdot \theta \rangle| \leq \frac{\delta_1 \|\theta^*\|^2}{\sigma}\right\} \subset \mathbb{R}^{2K}. \tag{88}$$

The following deterministic lemma guarantees that the law $\mathcal{P}_{\theta,\varepsilon}$ concentrates near 0 when $\varepsilon \in \mathcal{F}_1(\theta, \delta_1)$.

Lemma B.4. *Suppose $\underline{c}r \leq r_k(\theta^*) \leq \bar{c}r$ for each $k = 1, \ldots, K$. Fix any $\delta_0 > 0$. Then there exist constants $C_1, \delta_1 > 0$ depending only on $\underline{c}, \bar{c}, \delta_0$ such that if $\sigma^2 \leq \frac{Kr^2}{C_1 \log K}$, then the following holds: For any $\theta \in \mathcal{B}(\delta_1)$ and any (deterministic) $\varepsilon \in \mathcal{F}_1(\theta, \delta_1)$,*

$$\sup_{\alpha \in [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]} \langle \theta^* + \sigma\varepsilon,\ g(\alpha) \cdot \theta \rangle > \sup_{\alpha \in [-\pi,\pi) \setminus [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]} \langle \theta^* + \sigma\varepsilon,\ g(\alpha) \cdot \theta \rangle. \tag{89}$$

Furthermore, for a constant $c > 0$ depending only on $\underline{c}, \bar{c}, \delta_0$,

$$\mathbb{P}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}\Big[|\alpha|_{\mathcal{A}} > \delta_0/K\Big] \leq e^{-cKr^2/\sigma^2}. \tag{90}$$

Proof. Define $I_1 = [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$ and $I_2 = [-\pi, \pi) \setminus [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$. Let us write

$$\langle \theta^* + \sigma\varepsilon, g(\alpha) \cdot \theta \rangle = \langle \theta^*, g(\alpha)\theta^* \rangle + \langle \theta^*, g(\alpha)(\theta - \theta^*) \rangle + \sigma\langle \varepsilon, g(\alpha)\theta \rangle.$$

The conditions $\theta \in \mathcal{B}(\delta_1)$ and $\varepsilon \in \mathcal{F}_1(\theta, \delta_1)$ show for the second and third terms

$$|\langle \theta^*, g(\alpha)(\theta - \theta^*) \rangle| \leq \|\theta^*\| \cdot \|\theta - \theta^*\| \leq \delta_1 \|\theta^*\|^2, \qquad |\sigma\langle \varepsilon, g(\alpha)\theta \rangle| \leq \delta_1 \|\theta^*\|^2. \tag{91}$$

For the first term, Lemma B.2 implies that for constants $C > 0$ and $\iota > 0$,

$$\langle \theta^*, g(\alpha)\theta^* \rangle \leq (1 - \iota) \cdot \|\theta^*\|^2 \text{ if } \alpha \in I_2, \qquad \langle \theta^*, g(\alpha)\theta^* \rangle \geq \|\theta^*\|^2 - CK^3 r^2 \alpha^2 \text{ if } \alpha \in I_1. \tag{92}$$

Then for all $\alpha \in I_2$, we have $\langle \theta^* + \sigma\varepsilon, g(\alpha) \cdot \theta \rangle \leq (1 - \iota + 2\delta_1)\|\theta^*\|^2$, while for $\alpha = 0 \in I_1$, we have $\langle \theta^* + \sigma\varepsilon, g(\alpha) \cdot \theta \rangle \geq (1 - 2\delta_1)\|\theta^*\|^2$. Setting $\delta_1 < \iota/4$, this shows (89).

To show (90), we may correspondingly write the density (43) for the distribution $\mathcal{P}_{\theta,\varepsilon}$ as

$$\frac{d\mathcal{P}_{\theta,\varepsilon}(\alpha)}{d\alpha} \propto \exp\left(\frac{\langle \theta^* + \sigma\varepsilon, g(\alpha) \cdot \theta \rangle}{\sigma^2}\right) = \exp\left(\frac{\langle \theta^*, g(\alpha)\theta^* \rangle}{\sigma^2} + \frac{\langle \theta^*, g(\alpha)(\theta - \theta^*) \rangle}{\sigma^2} + \frac{\sigma\langle \varepsilon, g(\alpha)\theta \rangle}{\sigma^2}\right).$$

Then

$$\int_{I_2} \exp\left(\frac{\langle \theta^* + \sigma\varepsilon, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) d\alpha \leq 2\pi \exp\left(\frac{(1-\iota+2\delta_1)\|\theta^*\|^2}{\sigma^2}\right),$$
$$\int_{I_1} \exp\left(\frac{\langle \theta^* + \sigma\varepsilon, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) d\alpha \geq \exp\left(\frac{(1-2\delta_1)\|\theta^*\|^2}{\sigma^2}\right) \cdot \int_{-\delta_0/K}^{\delta_0/K} \exp\left(-\frac{CK^3r^2\alpha^2}{\sigma^2}\right) d\alpha.$$

Lower bounding this Gaussian integral using the same argument as (82), for a constant $C' > 0$,

$$\int_{-\delta_0/K}^{\delta_0/K} \exp\left(-\frac{CK^3r^2\alpha^2}{\sigma^2}\right) d\alpha \geq \left(\frac{C'K^3r^2}{\sigma^2}\right)^{-1/2}.$$

Combining these bounds and choosing $\delta_1 < \iota/4$, for a constant $c > 0$,

$$J := \frac{\int_{I_1} \exp\left(\frac{\langle \theta^*+\sigma\varepsilon, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) d\alpha}{\int_{I_2} \exp\left(\frac{\langle \theta^*+\sigma\varepsilon, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) d\alpha} \geq \left(\frac{C'K^3r^2}{\sigma^2}\right)^{-1/2} \exp\left(\frac{cKr^2}{\sigma^2}\right) \geq \exp\left(\frac{cKr^2}{2\sigma^2}\right)$$

where the last inequality applies $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$ for sufficiently large $C_1 > 0$, so that $\frac{K^3r^2}{\sigma^2} \leq K^2 e^{\frac{cKr^2}{2\sigma^2}} \leq e^{\frac{cKr^2}{\sigma^2}}$. Since $\mathbb{P}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon}}[|\alpha|_{\mathcal{A}} \geq \frac{\delta_0}{K}] = 1/(1+J) < 1/J$, this shows (90). □

Next, recall the complex representations

$$\tilde{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{C}^K, \qquad \tilde{v} = (v_1, \ldots, v_K) \in \mathbb{C}^K, \qquad \tilde{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_K) \in \mathbb{C}^K$$

and define $\mathcal{F}_2(\theta, v, \delta_1) \subset \mathbb{R}^{2K}$ as the set of vectors $\varepsilon \in \mathbb{R}^{2K}$ that satisfy the following three conditions:

$$\sup_{\alpha\in\mathcal{A}} \Big|\langle \varepsilon, g(\alpha)\cdot v\rangle\Big| \leq \sqrt{\frac{Kr^2}{\sigma^2}} \tag{93}$$
$$\sup_{\alpha,\beta\in[-\pi,\pi)} \left|\mathrm{Re}\sum_{k=1}^K \overline{\varepsilon_k} e^{ik\beta}\Big(e^{ik\alpha} - 1 - ik\alpha\Big)\theta_k\right| \leq \delta_1 \alpha^2 K^{5/2} r \sqrt{\frac{Kr^2}{\sigma^2}} \tag{94}$$
$$\sup_{\alpha,\beta\in[-\pi,\pi)} \left|\mathrm{Re}\sum_{k=1}^K \overline{\varepsilon_k}\Big(e^{ik\alpha} - e^{ik\beta}\Big)v_k\right| \leq \delta_1 |\alpha-\beta| K \sqrt{\frac{Kr^2}{\sigma^2}} \tag{95}$$

The domain $\mathcal{E}(\theta, v, \delta_1)$ in Lemma 5.5 is given by $\mathcal{F}_1(\theta, \delta_1) \cap \mathcal{F}_2(\theta, v, \delta_1)$.

Proof of Lemma 5.5. Let us denote

$$y = \theta^* + \sigma\varepsilon$$

The idea will be to approximate $\mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon}}$ by the variance with respect to a Gaussian law over α. We fix a small constant $\delta_0 > 0$ to be determined, and take $C_1 > 0$ large enough and $\delta_1 > 0$ small enough so that the conclusions of Lemma B.4 hold. Let

$$\alpha_0 = \arg\max_{\alpha\in\mathcal{A}} \langle y,\, g(\alpha)\cdot\theta\rangle \tag{96}$$

(where we may take any maximizer if it is not unique). Then (89) guarantees that $\alpha_0 \in [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$.

In the sense of Section 3.2, denote the complex representations of $\theta^*, \varepsilon, y, \theta, v \in \mathbb{R}^{2K}$ by

$$\tilde{\theta}^* = (\theta_1^*, \ldots, \theta_K^*) \in \mathbb{C}^K, \qquad \tilde{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_K) \in \mathbb{C}^K, \qquad \tilde{y} = (y_1, \ldots, y_K) \in \mathbb{C}^K,$$
$$\tilde{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{C}^K, \qquad \tilde{v} = (v_1, \ldots, v_K) \in \mathbb{C}^K.$$

Then the complex representation of $g(\alpha)\cdot\theta$ is $(e^{ik\alpha}\theta_k : k = 1, \ldots, K)$. By the inner-product relation (13), we have

$$\langle y, g(\alpha)\cdot\theta\rangle = \mathrm{Re}\sum_{k=1}^K \overline{y_k}\cdot e^{ik\alpha}\theta_k. \tag{97}$$

The first-order condition for optimality of α_0 in (96) yields

$$0 = \frac{d}{d\alpha}\langle y, g(\alpha)\cdot\theta\rangle\Big|_{\alpha=\alpha_0} = \operatorname{Re}\sum_{k=1}^{K}\overline{y_k}\cdot ike^{ik\alpha_0}\cdot\theta_k.$$

Applying this condition and the decomposition

$$e^{ik\alpha} = e^{ik\alpha_0}\Big[1 + ik(\alpha-\alpha_0) + \Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big)\Big]$$

to (97), we may write the density function (43) for the distribution $\mathcal{P}_{\theta,\varepsilon}$ as

$$\frac{d\mathcal{P}_{\theta,\varepsilon}(\alpha)}{d\alpha} \propto \exp\left(\frac{\langle y, g(\alpha)\cdot\theta\rangle}{\sigma^2}\right) \propto \exp\left(\frac{p(\alpha)}{\sigma^2}\right),$$

where (also dropping constant terms that do not depend on α)

$$p(\alpha) = \operatorname{Re}\sum_{k=1}^{K}\overline{y_k}\cdot e^{ik\alpha_0}\Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big)\theta_k.$$

For $\alpha \in [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$, we now establish a quadratic approximation for $p(\alpha)$. We have

$$p(\alpha) = \mathrm{I}(\alpha) + \mathrm{II}(\alpha) + \mathrm{III}(\alpha)$$

where

$$\begin{aligned}
\mathrm{I}(\alpha) &= \operatorname{Re}\sum_{k=1}^{K}\overline{\theta_k^*}\cdot e^{ik\alpha_0}\Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big)\theta_k^*\\
\mathrm{II}(\alpha) &= \operatorname{Re}\sum_{k=1}^{K}\overline{\theta_k^*}\cdot e^{ik\alpha_0}\Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big)(\theta_k - \theta_k^*)\\
\mathrm{III}(\alpha) &= \operatorname{Re}\sum_{k=1}^{K}\sigma\overline{\varepsilon_k}\cdot e^{ik\alpha_0}\Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big)\theta_k.
\end{aligned}$$

For $\mathrm{I}(\alpha)$, observe that $\overline{\theta_k^*}\theta_k^* = |\theta_k^*|^2$ is real, and

$$\begin{aligned}
\operatorname{Re} e^{ik\alpha_0}\Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big) &= \cos(k\alpha) - \cos(k\alpha_0) + k(\alpha-\alpha_0)\sin(k\alpha_0)\\
&= -\frac{k^2(\alpha-\alpha_0)^2}{2}\cos(k\tilde{\alpha}),
\end{aligned}$$

for some $\tilde{\alpha}$ between α and α_0. Since $\alpha, \alpha_0 \in [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$ and $k \le K$, for sufficiently small δ_0 this implies

$$-\frac{k^2(\alpha-\alpha_0)^2}{4} \ge \operatorname{Re} e^{ik\alpha_0}\Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha-\alpha_0)\Big) \ge -\frac{3k^2(\alpha-\alpha_0)^2}{4}.$$

Then, applying $\underline{c}^2 r^2 \le |\theta_k^*|^2 \le \bar{c}^2 r^2$, there are constants $C, c > 0$ (independent of δ_0) such that

$$-cK^3r^2(\alpha-\alpha_0)^2 \ge \mathrm{I}(\alpha) \ge -CK^3r^2(\alpha-\alpha_0)^2.$$

For $\mathrm{II}(\alpha)$, we apply $|e^{is} - 1 - is| \le s^2$ for all real values $s \in \mathbb{R}$, Cauchy-Schwarz, and the condition $\theta \in \mathcal{B}(\delta_1)$ to obtain

$$|\mathrm{II}(\alpha)| \le \sum_{k=1}^{K} k^2(\alpha-\alpha_0)^2|\overline{\theta_k^*}||\theta_k - \theta_k^*| \le C(\alpha-\alpha_0)^2 r\sqrt{\sum_{k=1}^{K}k^4}\sqrt{\sum_{k=1}^{K}|\theta_k-\theta_k^*|^2} \le C'\delta_1(\alpha-\alpha_0)^2K^3r^2,$$

for constants $C, C' > 0$ independent of δ_1. For $\mathrm{III}(\alpha)$, we apply the condition (94) for $\varepsilon \in \mathcal{F}_2(\theta, v, \delta_1)$ to obtain

$$|\mathrm{III}(\alpha)| \leq \delta_1(\alpha - \alpha_0)^2 K^3 r^2.$$

Combining these bounds, for sufficiently small $\delta_1 > 0$ and some constants $C_0, c_0 > 0$ which we may take independent of δ_0, δ_1, we arrive at the desired quadratic approximation

$$-c_0 K^3 r^2 (\alpha - \alpha_0)^2 \geq p(\alpha) \geq -C_0 K^3 r^2 (\alpha - \alpha_0)^2 \qquad \text{for } \alpha \in [-\tfrac{\delta_0}{K}, \tfrac{\delta_0}{K}]. \tag{98}$$

This implies the following variance bound: Denote $I_1 = [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$ and $I_2 = [-\pi, \pi) \setminus [-\frac{\delta_0}{K}, \frac{\delta_0}{K}]$. For any bounded function $f : [-\pi, \pi) \to \mathbb{R}$, denote $\|f\|_\infty = \sup_{\alpha \in [-\pi,\pi)} |f(\alpha)|$. Then

$$\begin{aligned}
\mathrm{Var}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}[f(\alpha)] &= \inf_{x \in \mathbb{R}} \frac{\int_{-\pi}^{\pi} (f(\alpha) - x)^2 e^{p(\alpha)/\sigma^2} d\alpha}{\int_{-\pi}^{\pi} e^{p(\alpha)/\sigma^2} d\alpha} \\
&\leq \inf_{x \in \mathbb{R}} \frac{\int_{I_1} (f(\alpha) - x)^2 e^{p(\alpha)/\sigma^2} d\alpha + 4\|f\|_\infty^2 \int_{I_2} e^{p(\alpha)/\sigma^2} d\alpha}{\int_{-\pi}^{\pi} e^{p(\alpha)/\sigma^2} d\alpha} \\
&\leq \inf_{x \in \mathbb{R}} \frac{\int_{I_1} (f(\alpha) - x)^2 e^{-c_0 K^3 r^2 (\alpha - \alpha_0)^2/\sigma^2} d\alpha}{\int_{I_1} e^{-C_0 K^3 r^2 (\alpha - \alpha_0)^2/\sigma^2} d\alpha} + 4\|f\|_\infty^2 e^{-c(\delta_0) K r^2/\sigma^2}
\end{aligned} \tag{99}$$

where, in the last line, we have used (98) as well as (90) to bound $\mathbb{P}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}[\alpha \in I_2] \leq e^{-c(\delta_0) K r^2/\sigma^2}$ for a constant $c(\delta_0) > 0$ depending on δ_0. For the denominator of the first term of (99), we may evaluate the Gaussian integral as

$$\begin{aligned}
&\int_{I_1} e^{-C_0 K^3 r^2 (\alpha - \alpha_0)^2/\sigma^2} d\alpha \\
&= \left(\frac{2C_0 K^3 r^2}{\sigma^2}\right)^{-1/2} \sqrt{2\pi} \left(1 - \tilde{\Phi}\left[\sqrt{\frac{2C_0 K^3 r^2}{\sigma^2}} \left(\frac{\delta_0}{K} + \alpha_0\right)\right] - \tilde{\Phi}\left[\sqrt{\frac{2C_0 K^3 r^2}{\sigma^2}} \left(\frac{\delta_0}{K} - \alpha_0\right)\right]\right).
\end{aligned}$$

Here, since $|\alpha_0| \leq \delta_0/K$, both values of $\tilde{\Phi}$ are at most $1/2$. Furthermore, under the condition $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$ for $C_1 > 0$ large enough depending on δ_0, at least one value of $\tilde{\Phi}$ is less than $1/4$. Thus, for a constant $C > 0$ independent of δ_0, δ_1, we have simply

$$\int_{I_1} e^{-C_0 K^3 r^2 (\alpha - \alpha_0)^2/\sigma^2} d\alpha \geq \left(\frac{CK^3 r^2}{\sigma^2}\right)^{-1/2}.$$

Combining this with the normalization constant for the Gaussian law in the numerator of the first term of (99), we obtain

$$\mathrm{Var}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}[f(\alpha)] \leq C\, \mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[f(\alpha)] + 4\|f\|_\infty^2 e^{-c(\delta_0) K r^2/\sigma^2}, \qquad \tau^2 := \frac{C' \sigma^2}{K^3 r^2}. \tag{100}$$

Here $C, C' > 0$ are some constants depending only on $\underline{c}, \bar{c}$ and independent of δ_0, δ_1, whereas $c(\delta_0)$ depends also on δ_0.

Finally, we apply this bound (100) to the function $f(\alpha) = v^\top g(\alpha)^{-1} y = \langle y, g(\alpha) v \rangle$. Observe that

$$\|f\|_\infty \leq \|\theta^*\| \|v\| + \sigma \sup_{\alpha \in [-\pi, \pi)} \langle \varepsilon, g(\alpha) v \rangle \leq C\sqrt{Kr^2}, \tag{101}$$

the last inequality using $\|v\| = 1$ and (93) for $\varepsilon \in \mathcal{F}_2(\theta, v, \delta_1)$. To bound $\mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}$, we apply again the inner-product relation (13) to write the complex representation of $f(\alpha)$ as

$$f(\alpha) = \mathrm{Re} \sum_{k=1}^{K} \overline{y_k} \cdot e^{ik\alpha} v_k = \mathrm{I}(\alpha) + \mathrm{II}(\alpha) + \mathrm{III}(\alpha)$$

where

$$\mathrm{I}(\alpha) = \mathrm{Re} \sum_{k=1}^{K} \overline{\theta_k^*} \cdot e^{ik\alpha_0} \Big(1 + ik(\alpha - \alpha_0)\Big) v_k$$

$$\mathrm{II}(\alpha) = \mathrm{Re} \sum_{k=1}^{K} \overline{\theta_k^*} \cdot e^{ik\alpha_0} \Big(e^{ik(\alpha-\alpha_0)} - 1 - ik(\alpha - \alpha_0)\Big) v_k$$

$$\mathrm{III}(\alpha) = \mathrm{Re} \sum_{k=1}^{K} \sigma \overline{\varepsilon_k} \cdot e^{ik\alpha} v_k$$

For $\mathrm{I}(\alpha)$, we may drop the constant term that is independent of α and write

$$\mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\mathrm{I}(\alpha)] = \mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)} \left[\mathrm{Re} \sum_{k=1}^{K} \overline{\theta_k^*} \cdot ik\alpha \cdot v_k + \mathrm{Re} \sum_{k=1}^{K} \overline{\theta_k^*} \cdot (e^{ik\alpha_0} - 1) ik\alpha \cdot v_k \right]. \tag{102}$$

Recalling the tangent vector u^* from (53), observe that its complex representation is

$$\tilde{u}^* = \frac{d}{d\alpha} \Big(e^{ik\alpha} \theta_k^* : k = 1, \ldots, K\Big)\Big|_{\alpha=0} = \Big(ik\theta_k^* : k = 1, \ldots, K\Big).$$

Then the inner-product relation (13) and the given orthogonality condition $\langle u^*, v \rangle = 0$ imply

$$\mathrm{Re} \sum_{k=1}^{K} -ik \cdot \overline{\theta_k^*} \cdot v_k = 0,$$

so the first term inside the variance of (102) is 0. Applying $|e^{ik\alpha_0} - 1| \leq k|\alpha_0| \leq \delta_0 k/K$ for the second term, for constants $C, C' > 0$ independent of δ_0,

$$\begin{aligned}
\mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\mathrm{I}(\alpha)] &= \mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\alpha] \cdot \left(\mathrm{Re} \sum_{k=1}^{K} \overline{\theta_k^*} \cdot (e^{ik\alpha_0} - 1) ik \cdot v_k \right)^2 \\
&\leq \mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\alpha] \cdot \left(\sum_{k=1}^{K} |\theta_k^*| \cdot \frac{\delta_0 k^2}{K} \cdot |v_k| \right)^2 \\
&\leq \mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\alpha] \cdot \frac{C \delta_0^2 r^2}{K^2} \sum_{k=1}^{K} k^4 \sum_{k=1}^{K} |v_k|^2 \leq \tau^2 \cdot C' \delta_0^2 K^3 r^2.
\end{aligned}$$

For $\mathrm{II}(\alpha)$, applying $|e^{is} - 1 - is| \leq s^2$ for any real value $s \in \mathbb{R}$, for constants $C, C' > 0$ independent of δ_0 we have

$$\begin{aligned}
\mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\mathrm{II}(\alpha)] &\leq \mathbb{E}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\mathrm{II}(\alpha)^2] \\
&\leq \mathbb{E}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)} \left[\left(\sum_{k=1}^{K} k^2 (\alpha - \alpha_0)^2 |\theta_k^*| |v_k| \right)^2 \right] \\
&\leq \mathbb{E}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[(\alpha - \alpha_0)^4] \cdot C r^2 \sum_{k=1}^{K} k^4 \sum_{k=1}^{K} |v_k|^2 \leq \tau^4 \cdot C' K^5 r^2.
\end{aligned}$$

For $\mathrm{III}(\alpha)$, we may center by a constant independent of α and apply (95) to obtain

$$\mathrm{Var}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)}[\mathrm{III}(\alpha)] \leq \mathbb{E}_{\alpha \sim \mathcal{N}(\alpha_0, \tau^2)} \left[\left(\mathrm{Re} \sum_{k=1}^{K} \sigma \overline{\varepsilon_k} \Big(e^{ik\alpha} - e^{ik\alpha_0}\Big) v_k \right)^2 \right]$$

$$\leq \sigma^2 \cdot \mathbb{E}_{\alpha \sim \mathcal{N}(\alpha_0,\tau^2)}[(\alpha-\alpha_0)^2] \cdot \frac{\delta_1^2 K^3 r^2}{\sigma^2} = \tau^2 \cdot \delta_1^2 K^3 r^2.$$

Combining all of the above,

$$\begin{aligned} \operatorname{Var}_{\alpha \sim \mathcal{N}(\alpha_0,\tau^2)}[f(\alpha)] &\leq 3\operatorname{Var}[\mathrm{I}(\alpha)] + 3\operatorname{Var}[\mathrm{II}(\alpha)] + 3\operatorname{Var}[\mathrm{III}(\alpha)] \\ &\leq \tau^2 \cdot C(\delta_0^2+\delta_1^2)K^3 r^2 + \tau^4 \cdot CK^5 r^2 \end{aligned} \tag{103}$$

for a constant $C > 0$ independent of δ_0, δ_1.

Applying (101) and (103) and the value of τ^2 to (100), for a constant $C' > 0$ independent of δ_0, δ_1,

$$\operatorname{Var}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}[f(\alpha)] \leq C' \left(\delta_0^2 + \delta_1^2 + \frac{\sigma^2}{Kr^2} + \frac{Kr^2}{\sigma^2} e^{-c(\delta_0)Kr^2/\sigma^2}\right)\sigma^2.$$

Then, choosing $\delta_0, \delta_1 > 0$ sufficiently small depending on η, and applying $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$ for a sufficiently large constant $C_1 > 0$ depending on δ_0 and η, we obtain $\operatorname{Var}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon}}[f(\alpha)] \leq \eta\sigma^2$ as desired. □

Proof of Lemma 5.6. Applying (76) with $t = \delta_1 \|\theta^*\|^2/\sigma$,

$$\mathbb{P}\left[\sup_{\alpha \in \mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > \frac{\delta_1\|\theta^*\|^2}{\sigma} \text{ and } \|\varepsilon\| \leq s\right] \leq C\sqrt{\frac{\sigma^2}{Kr^2}} \cdot Ks \cdot e^{-cKr^2/\sigma^2}$$

for constants $C, c > 0$ (depending on δ_1). Let us take

$$s = \max(\sqrt{4K}, \sqrt{Kr^2/\sigma^2}).$$

Then applying $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$ for sufficiently large $C_1 > 0$, this probability bound is at most $e^{-c'Kr^2/\sigma^2}$. By a chi-squared tail bound, since $\|\varepsilon\|^2 \sim \chi^2_{2K}$ and $s^2 \geq 4K$, we have $\mathbb{P}[\|\varepsilon\|^2 > s^2] \leq e^{-cs^2} \leq e^{-cKr^2/\sigma^2}$. Combining these bounds gives, for a constant $c > 0$,

$$\mathbb{P}\left[\sup_{\alpha \in \mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot\theta\rangle| > \frac{\delta_1\|\theta^*\|^2}{\sigma}\right] \leq e^{-cKr^2/\sigma^2},$$

so $\varepsilon \in \mathcal{F}_1(\theta,\delta_1)$ with probability at least $1 - e^{-cKr^2/\sigma^2}$. The same argument applied with the unit vector v in place of θ shows

$$\mathbb{P}\left[\sup_{\alpha \in \mathcal{A}} |\langle \varepsilon, g(\alpha)\cdot v\rangle| > \sqrt{\frac{Kr^2}{\sigma^2}}\right] \leq e^{-cKr^2/\sigma^2},$$

so (93) holds with probability at least $1 - e^{-cKr^2/\sigma^2}$.

For the condition (94), note that $\varepsilon_k \sim \mathcal{N}_{\mathbb{C}}(0,2)$ and these are independent for $k = 1, \ldots, K$. Then

$$f_{\alpha,\beta}(\varepsilon) := \alpha^{-2} \sum_{k=1}^K \overline{\varepsilon_k} e^{ik\beta}(e^{ik\alpha} - 1 - ik\alpha)\theta_k$$

has distribution $\mathcal{N}_{\mathbb{C}}(0, 2\tau^2)$ where

$$\tau^2 = \alpha^{-4} \sum_{k=1}^K |e^{ik\beta}(e^{ik\alpha} - 1 - ik\alpha)\theta_k|^2.$$

So $\operatorname{Re} f_{\alpha,\beta}(\varepsilon) \sim \mathcal{N}(0,\tau^2)$. Applying $|e^{is} - 1 - is| \leq s^2$ for all $s \in \mathbb{R}$, we have

$$\tau^2 \leq \sum_{k=1}^K k^4 |\theta_k|^2 \leq K^4 \sum_{k=1}^K |\theta_k|^2 \leq CK^5 r^2.$$

Then setting

$$t = \delta_1 K^{5/2} r \sqrt{\frac{Kr^2}{\sigma^2}},$$

a Gaussian tail bound yields, for a constant $c > 0$ (depending on δ_1),

$$\mathbb{P}[|\operatorname{Re} f_{\alpha,\beta}(\varepsilon)| > t/2] \le e^{-cKr^2/\sigma^2}.$$

Differentiating in α and β and applying $|e^{is} - 1 - is| \le s^2$ and $|e^{is} - 1 - is + s^2/2| \le |s|^3$ for $s \in \mathbb{R}$, we have

$$|\partial_\beta \operatorname{Re} f_{\alpha,\beta}(\varepsilon)| = \left|\alpha^{-2} \operatorname{Re} \sum_{k=1}^K \overline{\varepsilon_k} \cdot ike^{ik\beta}(e^{ik\alpha} - 1 - ik\alpha)\theta_k\right| \le \sum_{k=1}^K k^3 |\varepsilon_k||\theta_k| \le K^3 \|\varepsilon\|\|\theta\|,$$

$$|\partial_\alpha \operatorname{Re} f_{\alpha,\beta}(\varepsilon)| = \left|\alpha^{-3} \operatorname{Re} \sum_{k=1}^K \overline{\varepsilon_k} e^{ik\beta}(e^{ik\alpha}(ik\alpha - 2) + 2 + ik\alpha)\theta_k\right| \le C \sum_{k=1}^K k^3 |\varepsilon_k||\theta_k| \le CK^3 \|\varepsilon\|\|\theta\|.$$

(For the second line, we have explicitly evaluated the derivative, and then applied $e^{ik\alpha} = 1 + ik\alpha - k^2\alpha^2/2 + O(k^3\alpha^3)$ and canceled terms to yield the first inequality.) Then, on an event $\{\|\varepsilon\| < s\}$, $\operatorname{Re} f_{\alpha,\beta}(\varepsilon)$ is L-Lipschitz in both α and β, for $L = C_0 K^{7/2} rs$ and a constant $C_0 > 0$. We set $\delta = t/(4L)$ and let N_δ be a δ-net of $[-\pi, \pi)$ having cardinality

$$|N_\delta| = \frac{8\pi L}{t} \le C\sqrt{\frac{\sigma^2}{Kr^2}} \cdot Ks.$$

Then

$$\mathbb{P}\left[\sup_{\alpha,\beta \in [-\pi,\pi)} |\operatorname{Re} f_{\alpha,\beta}(\varepsilon)| > t \text{ and } \|\varepsilon\| < s\right] \le \mathbb{P}\left[\sup_{\alpha,\beta \in N_\delta} |\operatorname{Re} f_{\alpha,\beta}(\varepsilon)| > t/2\right] \le |N_\delta|^2 \cdot e^{-cKr^2/\sigma^2}.$$

Then, setting $s = \max(\sqrt{4K}, \sqrt{Kr^2/\sigma^2})$ and applying the same argument as above shows

$$\mathbb{P}\left[\sup_{\alpha,\beta \in [-\pi,\pi)} |\operatorname{Re} f_{\alpha,\beta}(\varepsilon)| > t\right] \le e^{-cKr^2/\sigma^2},$$

so (94) holds with probability at least $1 - e^{-cKr^2/\sigma^2}$.

The argument for (95) is analogous. We define $\gamma = \alpha - \beta$,

$$f_{\beta,\gamma}(\varepsilon) := \gamma^{-1} \sum_{k=1}^K \overline{\varepsilon_k} e^{ik\beta}(e^{ik\gamma} - 1)v_k \sim \mathcal{N}_{\mathbb{C}}(0, 2\tau^2), \qquad \tau^2 := \gamma^{-2} \sum_{k=1}^K |(e^{ik\beta}(e^{ik\gamma} - 1)v_k|^2$$

and set $t = \delta_1 K\sqrt{Kr^2/\sigma^2}$. Applying $|e^{ik\gamma} - 1| \le k|\gamma|$ and $\|v\| = 1$, we have $\tau^2 \le \sum_{k=1}^K k^2 |v_k|^2 \le K^2$, so that a Gaussian tail bound yields $\mathbb{P}[|\operatorname{Re} f_{\alpha,\beta}(\varepsilon)| > t/2] \le e^{-cKr^2/\sigma^2}$. On the event $\{\|\varepsilon\| < s\}$, we have the Lipschitz bounds

$$|\partial_\beta \operatorname{Re} f_{\beta,\gamma}(\varepsilon)| = \left|\gamma^{-1} \operatorname{Re} \sum_{k=1}^K \overline{\varepsilon_k} \cdot ike^{ik\beta}(e^{ik\gamma} - 1)v_k\right| \le \sum_{k=1}^K k^2 |\varepsilon_k||v_k| \le K^2 \|\varepsilon\|\|v\| \le K^2 s,$$

$$|\partial_\gamma \operatorname{Re} f_{\beta,\gamma}(\varepsilon)| = \left|\gamma^{-2} \operatorname{Re} \sum_{k=1}^K \overline{\varepsilon_k} \cdot e^{ik\beta}(e^{ik\gamma}(ik\gamma - 1) + 1)v_k\right| \le C \sum_{k=1}^K k^2 |\varepsilon_k||v_k| \le CK^2 \|\varepsilon\|\|v\| \le CK^2 s,$$

where we have applied $e^{ik\gamma} = 1 + ik\gamma + O(k^2\gamma^2)$. Then applying a covering net argument as above, whose details we omit for brevity, we obtain

$$\mathbb{P}\left[\sup_{\beta \in [-\pi,\pi),\ \gamma \in [-2\pi,2\pi)} |\operatorname{Re} f_{\beta,\gamma}(\varepsilon)| > t\right] \le e^{-cKr^2/\sigma^2},$$

so (95) holds with probability at least $1 - e^{-cKr^2/\sigma^2}$. Combining these bounds yields the lemma. □

Proof of Lemma 5.4. Throughout the proof, C, C', c, c' etc. are positive constants depending only on $\underline{c}, \bar{c}, \eta$ and changing from instance to instance. Recall the expression (54). Let $C_1, \delta_1 > 0$ be such that the conclusion of Lemma 5.5 holds with $\eta/6$ in place of η. For any $\theta \in \mathcal{B}(\delta_1)$ and unit vector v satisfying $\langle u^*, v\rangle = 0$, let us apply

$$\mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon}}\left[v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon)\right] \le \mathbb{E}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon}}\left[\left(v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon)\right)^2\right] \le \|\theta^*+\sigma\varepsilon\|^2$$

to upper-bound the second term of (54) as

$$\frac{1}{N}\sum_{m=1}^N \mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}\left[v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon^{(m)})\right] \le I_1(\theta,v)+I_2(\theta,v)+I_3 \tag{104}$$

where

$$\begin{aligned}
I_1(\theta,v) &= \frac{1}{N}\sum_{m=1}^N \mathrm{Var}_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}\left[v^\top g(\alpha)^{-1}(\theta^*+\sigma\varepsilon^{(m)})\right]\cdot \mathbf{1}[\varepsilon^{(m)}\in\mathcal{E}(\theta,v,\delta_1)],\\
I_2(\theta,v) &= \frac{1}{N}\sum_{m=1}^N \|\theta^*+\sigma\varepsilon^{(m)}\|^2\cdot\mathbf{1}\left[\varepsilon^{(m)}\notin\mathcal{E}(\theta,v,\delta_1) \text{ and } \|\varepsilon^{(m)}\|^2 \le 4K+\frac{Kr^2}{\sigma^2}\right],\\
I_3 &= \frac{1}{N}\sum_{m=1}^N \|\theta^*+\sigma\varepsilon^{(m)}\|^2\cdot\mathbf{1}\left[\|\varepsilon^{(m)}\|^2 > 4K+\frac{Kr^2}{\sigma^2}\right].
\end{aligned}$$

Here I_1, I_2 are dependent on (θ, v), whereas I_3 is independent of (θ, v).

Lemma 5.5 applied with $\eta/6$ immediately gives the deterministic bound

$$I_1(\theta,v) \le \eta\sigma^2/6. \tag{105}$$

For $I_2(\theta, v)$, on the event $\|\varepsilon\|^2 \le 4K + Kr^2/\sigma^2$, we have for a constant $C_2 > 0$ that

$$\|\theta^*+\sigma\varepsilon\|^2 \le 2\|\theta^*\|^2 + 2\sigma^2\|\varepsilon\|^2 \le C_2K(r^2+\sigma^2).$$

Thus

$$I_2(\theta,v) \le C_2K(r^2+\sigma^2)\cdot\frac{1}{N}\sum_{m=1}^N \mathbf{1}[\varepsilon^{(m)}\notin\mathcal{E}(\theta,v,\delta_1)].$$

Denote $p = \mathbb{P}[\varepsilon^{(m)} \notin \mathcal{E}(\theta, v, \delta_1)]$ and $q = \frac{\eta\sigma^2}{6C_2K(r^2+\sigma^2)}$. By Lemma 5.6,

$$p \le e^{-cKr^2/\sigma^2},$$

so in particular $p < q$ for $\frac{Kr^2}{\sigma^2} \ge C_1 \log K$ and sufficiently large $C_1 > 0$. Then by a Chernoff bound for binomial random variables (Vershynin, 2018, Theorem 2.3.1),

$$\mathbb{P}[I_2(\theta,v)\ge\eta\sigma^2/6] = \mathbb{P}\left[\sum_{m=1}^N \mathbf{1}[\varepsilon^{(m)}\notin\mathcal{E}(\theta,v,\delta_1)] \ge Nq\right] \le \left(\frac{ep}{q}\right)^{Nq}.$$

We have $(ep/q) \le e^{-c'Kr^2/\sigma^2}$ when $\frac{Kr^2}{\sigma^2} \ge C_1 \log K$ for sufficiently large $C_1 > 0$, so this yields

$$\mathbb{P}[I_2(\theta,v)\ge\eta\sigma^2/6] \le e^{-\frac{cN}{1+\sigma^2/r^2}}. \tag{106}$$

For I_3, we bound separately its mean and its concentration. Denote the summand of I_3 as

$$z^{(m)} = \|\theta^*+\sigma\varepsilon^{(m)}\|^2\cdot\mathbf{1}\left[\|\varepsilon^{(m)}\|^2 > 4K+\frac{Kr^2}{\sigma^2}\right].$$

Let $p' = \mathbb{P}[\|\varepsilon^{(m)}\|^2 > 4K + Kr^2/\sigma^2]$. Then applying $\|\varepsilon^{(m)}\|^2 \sim \chi^2_{2K}$ and a chi-squared tail bound, $p' \leq e^{-c(2K+Kr^2/\sigma^2)} \leq e^{-cKr^2/\sigma^2}$. So by Cauchy-Schwarz,

$$\mathbb{E}[z^{(m)}] \leq \sqrt{\mathbb{E}\left[\|\theta^* + \sigma\varepsilon^{(m)}\|^4\right]} \cdot \sqrt{p'} \leq \sigma^2 \cdot \sqrt{CK^2(1 + r^4/\sigma^4) \cdot e^{-cKr^2/\sigma^2}} \leq \eta\sigma^2/12 \tag{107}$$

the last inequality holding for $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$ and sufficiently large $C_1 > 0$. For the concentration, let $\|X\|_{\psi_1} = \inf\{s > 0 : \mathbb{E}[e^{|X|/s}] \leq 2\}$ denote the sub-exponential norm of a random variable X. Observe that similarly by Cauchy-Schwarz,

$$\begin{aligned}\mathbb{E}\Big[e^{\frac{|z^{(m)}|}{s}}\Big] &= 1 - p' + \mathbb{E}\left[\exp\left(\frac{\|\theta^* + \sigma\varepsilon^{(m)}\|^2}{s}\right) \cdot \mathbf{1}\left[\|\varepsilon^{(m)}\|^2 > 4K + \frac{Kr^2}{\sigma^2}\right]\right] \\ &\leq 1 + \sqrt{\mathbb{E}\Big[\exp\Big((2CKr^2 + 4\sigma^2\|\varepsilon^{(m)}\|^2)/s\Big)\Big]} \cdot \sqrt{p'}.\end{aligned}$$

Applying the moment generating function bound $\mathbb{E}[\exp(t\|\varepsilon^{(m)}\|^2)] = (1-2t)^{-K} \leq e^{4tK}$ for $t < 1/4$, we have

$$\mathbb{E}\Big[e^{\frac{|z^{(m)}|}{s}}\Big] \leq 1 + e^{CKr^2/s} \cdot e^{8K\sigma^2/s} \cdot e^{-cKr^2/2\sigma^2} \leq 2$$

when $s \geq C'\sigma^2(1 + \sigma^2/r^2)$ for a sufficiently large constant $C' > 0$. So $\|z^{(m)}\|_{\psi_1} \leq C\sigma^2(1 + \sigma^2/r^2)$, and Bernstein's inequality (Vershynin, 2018, Theorem 2.8.1) gives

$$\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N z^{(m)} - \mathbb{E}[z^{(m)}] > \eta\sigma^2/12\right] \leq e^{-\frac{cN}{(1+\sigma^2/r^2)^2}}.$$

Combining with (107),

$$\mathbb{P}[I_3 \geq \eta\sigma^2/6] = \mathbb{P}\left[\frac{1}{N}\sum_{m=1}^N z^{(m)} > \eta\sigma^2/6\right] \leq e^{-\frac{cN}{(1+\sigma^2/r^2)^2}} \leq e^{-\frac{cN}{K(1+\sigma^2/r^2)}}, \tag{108}$$

the second inequality using $1 + \sigma^2/r^2 \leq K$ under the given condition $\frac{Kr^2}{\sigma^2} \geq C_1 \log K$.

Applying (54), (104), (105), and (106), we have

$$\begin{aligned}\mathbb{P}\left[v^\top \nabla^2 R_N(\theta) v \leq \frac{1}{\sigma^2} - \frac{\eta}{2\sigma^2} \text{ and } I_3 \leq \frac{\eta\sigma^2}{6}\right] &\leq \mathbb{P}\left[I_1(\theta, v) + I_2(\theta, v) + I_3 \geq \frac{\eta\sigma^2}{2} \text{ and } I_3 \leq \frac{\eta\sigma^2}{6}\right] \\ &\leq \mathbb{P}\left[I_2(\theta, v) \geq \frac{\eta\sigma^2}{6}\right] \leq e^{-\frac{cN}{1+\sigma^2/r^2}}.\end{aligned} \tag{109}$$

Let us apply a covering net argument to take a union bound over (θ, v), and then combine with the bound (108) for I_3 which is independent of (θ, v). We compute the Lipschitz constant of $v^\top \nabla^2 R_N(\theta) v$ in both v and θ: Taking the gradient in v,

$$\Big\|\nabla_v\Big[v^\top \nabla^2 R_N(\theta) v\Big]\Big\| = \big\|2\nabla^2 R_N(\theta) v\big\| = 2 \sup_{u:\|u\|=1} u^\top \nabla^2 R_N(\theta) v.$$

Then denoting $y^{(m)} = \theta^* + \sigma\varepsilon^{(m)}$ and applying (45),

$$\begin{aligned}\Big\|\nabla_v\Big[v^\top \nabla^2 R_N(\theta) v\Big]\Big\| &\leq 2\left(\sup_{u:\|u\|=1} \frac{u^\top v}{\sigma^2} - \frac{1}{N\sigma^4}\sum_{m=1}^N \mathrm{Cov}_{\alpha \sim \mathcal{P}_{\theta,\varepsilon^{(m)}}}[u^\top g(\alpha)^{-1} y^{(m)}, v^\top g(\alpha)^{-1} y^{(m)}]\right) \\ &\leq \frac{2}{\sigma^2} + \frac{2}{N\sigma^4}\sum_{m=1}^N \|y^{(m)}\|^2 \leq \frac{2}{\sigma^2} + \frac{4\|\theta^*\|^2}{\sigma^4} + \frac{4}{N\sigma^2}\sum_{m=1}^N \|\varepsilon^{(m)}\|^2.\end{aligned} \tag{110}$$

Similarly, taking the gradient in θ,

$$\big\|\nabla_\theta\left[v^\top \nabla^2 R_N(\theta) v\right]\big\| = \sup_{u:\|u\|=1} \nabla^3 R_N(\theta)[u, v, v]$$

where $\nabla^3 R_N(\theta)[u,v,v]$ is the 3rd-derivative tensor of $R_N(\theta)$ evaluated at $u \otimes v \otimes v \in \mathbb{R}^{2K\times 2K\times 2K}$. We have

$$\nabla^3 R_N(\theta)[u,v,v] = -\frac{1}{N\sigma^6}\sum_{m=1}^N \kappa^3_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}[u^\top g(\alpha)^{-1}y^{(m)}, v^\top g(\alpha)^{-1}y^{(m)}, v^\top g(\alpha)^{-1}y^{(m)}]$$

where $\kappa^3_{\alpha\sim\mathcal{P}_{\theta,\varepsilon^{(m)}}}[\cdot,\cdot,\cdot]$ denotes the 3rd-order mixed cumulant with respect to $\alpha \sim \mathcal{P}_{\theta,\varepsilon^{(m)}}$. For any random variables X, Y, Z, the moment-cumulant relations and Hölder's inequality give

$$|\kappa^3[X,Y,Z]| \le C \cdot \mathbb{E}[|X|^3]^{1/3}\mathbb{E}[|Y|^3]^{1/3}\mathbb{E}[|Z|^3]^{1/3}.$$

Thus,

$$\left\|\nabla_\theta\left[v^\top\nabla^2 R_N(\theta)v\right]\right\| \le \frac{C}{N\sigma^6}\sum_{m=1}^N \|y^{(m)}\|^3 \le \frac{4C\|\theta^*\|^3}{\sigma^6} + \frac{4C}{N\sigma^3}\sum_{m=1}^N \|\varepsilon^{(m)}\|^3. \tag{111}$$

On an event

$$\mathcal{A} = \left\{\|\varepsilon^{(m)}\|^2 \le N \text{ for all } m = 1,\dots,N\right\},$$

(110) and (111) imply that $v^\top\nabla^2 R_N(\theta)v$ is L_v-Lipschitz in v and L_θ-Lipschitz in θ, for

$$L_v = C'\left(\frac{Kr^2}{\sigma^2}+N\right)\frac{1}{\sigma^2}, \qquad L_\theta = C'\left(\frac{Kr^2}{\sigma^2}+N\right)^{3/2}\frac{1}{\sigma^3}.$$

Let N_v be a δ_v-net of $\{v : \|v\| = 1, \langle u^*, v\rangle = 0\}$ and N_θ a δ_θ-net of $\{\theta : \theta \in \mathcal{B}(\delta_1)\}$, for $\delta_v = \eta/(4L_v\sigma^2)$ and $\delta_\theta = \eta/(4L_\theta\sigma^2)$. This guarantees, for each $\theta \in \mathcal{B}(\delta_1)$ and unit vector v with $\langle u^*, v\rangle = 0$, there exists $(\theta', v') \in N_\theta \times N_v$ such that

$$\left|v^\top\nabla^2 R_N(\theta)v - {v'}^\top\nabla^2 R_N(\theta')v'\right| \le L_\theta\|\theta-\theta'\| + L_v\|v-v'\| \le \eta/2\sigma^2.$$

Then, applying these Lipschitz bounds together with the pointwise bound (109),

$$\begin{aligned}&\mathbb{P}\left[\sup_{\theta\in\mathcal{B}(\delta_1)}\ \sup_{v:\|v\|=1,\,\langle u^*,v\rangle=0} v^\top\nabla^2 R_N(\theta)v \le \frac{1}{\sigma^2}-\frac{\eta}{\sigma^2} \text{ and } I_3 \le \frac{\eta\sigma^2}{6} \text{ and } \mathcal{A}\right]\\ &\quad\le \mathbb{P}\left[\sup_{\theta\in N_\theta}\ \sup_{v\in N_v} v^\top\nabla^2 R_N(\theta)v \le \frac{1}{\sigma^2}-\frac{\eta}{2\sigma^2} \text{ and } I_3 \le \frac{\eta\sigma^2}{6}\right] \le |N_v|\cdot|N_\theta|\cdot e^{-\frac{cN}{1+\sigma^2/r^2}}.\end{aligned} \tag{112}$$

We may take the above nets to have cardinalities

$$\begin{aligned}|N_v| &\le \left(1+\frac{2}{\delta_v}\right)^{2K} \le \left[C'\left(\frac{Kr^2}{\sigma^2}+N\right)\right]^{2K},\\ |N_\theta| &\le \left(1+\frac{C\sqrt{Kr^2}}{\delta_\theta}\right)^{2K} \le \left[C'\left(\frac{Kr^2}{\sigma^2}+N\right)^{3/2}\sqrt{\frac{Kr^2}{\sigma^2}}\right]^{2K}.\end{aligned}$$

Observe that under the given assumptions $N \ge C_0K(1+\frac{\sigma^2}{r^2})\log(K+\frac{Kr^2}{\sigma^2})$ and $\frac{Kr^2}{\sigma^2} \ge C_1\log K$, we have also

$$\frac{N}{\log N} \ge \frac{C_0K(1+\frac{\sigma^2}{r^2})\log(K+\frac{Kr^2}{\sigma^2})}{\log[C_0K(1+\frac{\sigma^2}{r^2})\log(K+\frac{Kr^2}{\sigma^2})]}.$$

If $\frac{\sigma^2}{r^2} \le 1$, then $\log[C_0K(1+\frac{\sigma^2}{r^2})\log(K+\frac{Kr^2}{\sigma^2})] \le \log[2C_0\frac{Kr^2}{\sigma^2}\log\frac{2Kr^2}{\sigma^2}] \le \log(2C_0) + 2\log\frac{Kr^2}{\sigma^2}$ for $\frac{Kr^2}{\sigma^2} \ge C_1\log K$ and sufficiently large $C_1 > 0$. Then

$$\frac{N}{\log N} \ge \frac{C_0K(1+\frac{\sigma^2}{r^2})\log\frac{Kr^2}{\sigma^2}}{\log(2C_0)+2\log\frac{Kr^2}{\sigma^2}} \ge C_0'K(1+\tfrac{\sigma^2}{r^2})$$

where we may take $C_0' = C_0/[2\log(2C_0)]$ for sufficiently large $C_0 > 0$. If instead $1 \le \frac{\sigma^2}{r^2} \le \frac{K}{C_1 \log K}$, then $\log[C_0 K(1+\frac{\sigma^2}{r^2})\log(K + \frac{Kr^2}{\sigma^2})] \le \log[C_0 K^2 \log 2K] \le \log C_0 + 3\log K$. Then

$$\frac{N}{\log N} \ge \frac{C_0 K(1+\frac{\sigma^2}{r^2})\log K}{\log C_0 + 3\log K} \ge C_0' K(1+\tfrac{\sigma^2}{r^2})$$

where we may take $C_0' = C_0/[3\log C_0]$. Thus, in both cases we have

$$N \ge C_0' K(1+\tfrac{\sigma^2}{r^2})\log N. \tag{113}$$

For sufficiently large $C_0, C_0' > 0$ and some constant $c' > 0$, the right side of (112) may then be bounded as $|N_v| \cdot |N_\theta| \cdot e^{-\frac{cN}{1+\sigma^2/r^2}} \le e^{-\frac{c'N}{1+\sigma^2/r^2}}$.

Combining this with (108) and the chi-squared tail bound $\mathbb{P}[\mathcal{A}^c] \le N \cdot \mathbb{P}[\|\varepsilon^{(m)}\|^2 > N] \le Ne^{-cN} \le e^{-c'N}$ for $N \ge C_0 K$, this gives

$$\mathbb{P}\left[\sup_{\theta\in\mathcal{B}(\delta_1)} \sup_{v:\|v\|=1,\, \langle u^*, v\rangle = 0} v^\top \nabla^2 R_N(\theta) v \le \frac{1-\eta}{\sigma^2}\right] \le e^{-\frac{c'N}{1+\sigma^2/r^2}} + \mathbb{P}[I_3 > \tfrac{\eta\sigma^2}{6}] + \mathbb{P}[\mathcal{A}^c] \le e^{-\frac{cN}{K(1+\sigma^2/r^2)}}$$

which implies the lemma. □

C Proofs for minimax lower bound

Proof of Lemma 6.2. The first bound (58) is basic and due to the data processing inequality: Let $q_\theta(\alpha, y)$ denote the joint density of $\alpha \sim \text{Unif}([-\pi,\pi))$ and $y = g(\alpha)\cdot\theta + \sigma\varepsilon$. Then the data processing inequality implies

$$\begin{aligned} D_{\mathrm{KL}}(p_\theta\|p_{\theta'}) \le D_{\mathrm{KL}}(q_\theta\|q_{\theta'}) &= \mathbb{E}_{(\alpha,y)\sim q_\theta}\left[-\frac{\|y-g(\alpha)\cdot\theta\|^2}{2\sigma^2} + \frac{\|y-g(\alpha)\cdot\theta'\|^2}{2\sigma^2}\right] \\ &= \mathbb{E}_{\substack{\alpha\sim\text{Unif}([-\pi,\pi))\\ \varepsilon\sim\mathcal{N}(0,I)}}\left[-\frac{\|\sigma\varepsilon\|^2}{2\sigma^2} + \frac{\|g(\alpha)\cdot(\theta-\theta')+\sigma\varepsilon\|^2}{2\sigma^2}\right] = \frac{\|\theta-\theta'\|^2}{2\sigma^2}. \end{aligned}$$

In the remainder of the proof, we show (59). Let us write $\alpha, \alpha' \sim \text{Unif}([-\pi,\pi))$ for independent random rotations, $\mathbb{E}$ for the expectation over only α, α' (fixing y and ε), and $g = g(\alpha)$ and $g' = g(\alpha')$. We have

$$D_{\mathrm{KL}}(p_\theta\|p_{\theta'}) \le \chi^2(p_\theta\|p_{\theta'}) := \int_{\mathbb{R}^{2K}} \frac{[p_\theta(y) - p_{\theta'}(y)]^2}{p_\theta(y)} dy$$

where the right side is the χ^2-divergence, see e.g. (Tsybakov, 2008, Lemma 2.7). We derive an upper bound for $\chi^2(p_\theta\|p_{\theta'})$ using the idea of (Bandeira et al., 2020, Theorem 9): Let $\varphi(z) = (2\pi\sigma^2)^{-K}\exp(-\|z\|^2/2\sigma^2)$ be the density of $\mathcal{N}(0, \sigma^2 I_{2K})$. Then

$$p_\theta(y) = \mathbb{E}[\varphi(y - g\theta)] = \mathbb{E}\Big[\varphi(y)\cdot e^{\frac{y^\top g\theta}{\sigma^2} - \frac{\|\theta\|^2}{2\sigma^2}}\Big]. \tag{114}$$

By Jensen's inequality and the condition $\mathbb{E}[g] = 0$,

$$p_\theta(y) \ge \varphi(y)\cdot e^{\frac{y^\top\mathbb{E}[g]\theta}{\sigma^2} - \frac{\|\theta\|^2}{2\sigma^2}} = \varphi(y)\cdot e^{-\frac{\|\theta\|^2}{2\sigma^2}}. \tag{115}$$

Then applying (114), (115), and the moment generating function

$$\int \varphi(y) e^{\frac{y^\top(g\theta + g'\theta')}{\sigma^2}} dy = e^{\frac{\|g\theta+g'\theta'\|^2}{2\sigma^2}} = e^{\frac{\|\theta\|^2+\|\theta'\|^2+2\langle g\theta, g'\theta'\rangle}{2\sigma^2}},$$

we get

$$\chi^2(p_\theta\|p_{\theta'}) \le \int \frac{\Big(\mathbb{E}\Big[\varphi(y)\cdot e^{\frac{y^\top g\theta}{\sigma^2} - \frac{\|\theta\|^2}{2\sigma^2}}\Big] - \mathbb{E}\Big[\varphi(y)\cdot e^{\frac{y^\top g\theta'}{\sigma^2} - \frac{\|\theta'\|^2}{2\sigma^2}}\Big]\Big)^2}{\varphi(y)\cdot e^{-\frac{\|\theta\|^2}{2\sigma^2}}} dy$$

$$= \int \varphi(y)\Big(e^{-\frac{\|\theta\|^2}{2\sigma^2}}\mathbb{E}\Big[e^{\frac{y^\top g\theta}{\sigma^2}+\frac{y^\top g'\theta}{\sigma^2}}\Big] - 2e^{-\frac{\|\theta'\|^2}{2\sigma^2}}\mathbb{E}\Big[e^{\frac{y^\top g\theta}{\sigma^2}+\frac{y^\top g'\theta'}{\sigma^2}}\Big] + e^{-\frac{\|\theta'\|^2}{\sigma^2}+\frac{\|\theta\|^2}{2\sigma^2}}\mathbb{E}\Big[e^{\frac{y^\top g\theta'}{\sigma^2}+\frac{y^\top g'\theta'}{\sigma^2}}\Big]\Big)dy$$

$$= e^{\frac{\|\theta\|^2}{2\sigma^2}}\mathbb{E}\left[e^{\frac{\langle g\theta, g'\theta\rangle}{\sigma^2}} - 2e^{\frac{\langle g\theta, g'\theta'\rangle}{\sigma^2}} + e^{\frac{\langle g\theta', g'\theta'\rangle}{\sigma^2}}\right]$$

$$= e^{\frac{\|\theta\|^2}{2\sigma^2}}\mathbb{E}\left[e^{\frac{\langle \theta, g\theta\rangle}{\sigma^2}} - 2e^{\frac{\langle \theta, g\theta'\rangle}{\sigma^2}} + e^{\frac{\langle \theta', g\theta'\rangle}{\sigma^2}}\right]$$

$$= e^{\frac{\|\theta\|^2}{2\sigma^2}}\sum_{m=0}^{\infty}\frac{1}{\sigma^{2m}m!}\mathbb{E}\big[\langle\theta, g\theta\rangle^m - 2\langle\theta, g\theta'\rangle^m + \langle\theta', g\theta'\rangle^m\big]. \tag{116}$$

For $m = 0$ and $m = 1$, the summand of (116) is 0. We evaluate the summand for $m = 2$, and upper bound it for $m \geq 3$. Let

$$\tilde{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{C}^K, \qquad \tilde{\theta}' = (\theta'_1, \ldots, \theta'_K) \in \mathbb{C}^K$$

be the complex representations of θ, θ' as defined in Section 3.2. For $m = 2$, applying (13),

$$\mathbb{E}[\langle\theta, g(\alpha)\theta'\rangle^2] = \sum_{k_1,k_2=1}^{K}\mathbb{E}\Big[\operatorname{Re}(\overline{\theta_{k_1}}e^{ik_1\alpha}\theta'_{k_1})\cdot\operatorname{Re}(\overline{\theta_{k_2}}e^{ik_2\alpha}\theta'_{k_2})\Big].$$

For any $k_1, k_2 \in \{1, \ldots, K\}$, applying $\operatorname{Re}\bar{x}y = (x\bar{y} + \bar{x}y)/2$ and $\mathbb{E}[e^{ik\alpha}] = 0$ for any non-zero integer k,

$$\begin{aligned}\mathbb{E}\Big[\operatorname{Re}(\overline{\theta_{k_1}}e^{ik_1\alpha}\theta'_{k_1})\cdot\operatorname{Re}(\overline{\theta_{k_2}}e^{ik_2\alpha}\theta'_{k_2})\Big] &= \frac{1}{4}\mathbb{E}\Big[(e^{-ik_1\alpha}\theta_{k_1}\overline{\theta'_{k_1}} + e^{ik_1\alpha}\overline{\theta_{k_1}}\theta'_{k_1})(e^{-ik_2\alpha}\theta_{k_2}\overline{\theta'_{k_2}} + e^{ik_2\alpha}\overline{\theta_{k_2}}\theta'_{k_2})\Big]\\ &= \frac{1}{2}\mathbb{1}\{k_1 = k_2\}|\theta_{k_1}|^2|\theta'_{k_1}|^2.\end{aligned}$$

Then $\mathbb{E}[\langle\theta, g\theta'\rangle^2] = \frac{1}{2}\sum_{k=1}^K r_k^2 {r'_k}^2$. This identity holds also with $\theta = \theta'$, so

$$\mathbb{E}\big[\langle\theta, g\theta\rangle^2 - 2\langle\theta, g\theta'\rangle^2 + \langle\theta', g\theta'\rangle^2\big] = \frac{1}{2}\sum_{k=1}^{K}(r_k^2 - {r'_k}^2)^2. \tag{117}$$

For any $m \geq 3$ and every $k_1, \ldots, k_m \in \{1, \ldots, K\}$, applying again $\mathbb{E}[e^{ik\alpha}] = 0$ for $k \neq 0$, we have similarly

$$\begin{aligned}\mathbb{E}\left[\prod_{\ell=1}^{m}\operatorname{Re}(\overline{\theta_{k_\ell}}e^{ik_\ell\alpha}\theta'_{k_\ell})\right] &= \frac{1}{2^m}\mathbb{E}\left[\prod_{\ell=1}^{m}(e^{-ik_\ell\alpha}\theta_{k_\ell}\overline{\theta'_{k_\ell}} + e^{ik_\ell\alpha}\overline{\theta_{k_\ell}}\theta'_{k_\ell})\right]\\ &= \frac{1}{2^m}\sum_{s_1,\ldots,s_m\in\{+1,-1\}}\mathbb{1}\{s_1k_1 + \ldots + s_mk_m = 0\}\cdot\prod_{\ell:s_\ell=+1}\theta_{k_\ell}\overline{\theta'_{k_\ell}}\cdot\prod_{\ell:s_\ell=-1}\overline{\theta_{k_\ell}}\theta'_{k_\ell}\end{aligned}$$

Noting that the left side is real and taking the real part on the right side, this is equal to

$$\frac{1}{2^m}\sum_{s_1,\ldots,s_m\in\{+1,-1\}}\mathbb{1}\{s_1k_1 + \ldots + s_mk_m = 0\}\left(\prod_{\ell=1}^{m}r_{k_\ell}r'_{k_\ell}\right)\cos\left(\sum_{\ell=1}^{m}s_\ell\phi_{k_\ell} - s_\ell\phi'_{k_\ell}\right).$$

Then, summing over all $k_1, \ldots, k_m \in \{1, \ldots, K\}$ and applying this also for $\theta = \theta'$,

$$\begin{aligned}&\mathbb{E}\big[\langle\theta, g\theta\rangle^m - 2\langle\theta, g\theta'\rangle^m + \langle\theta', g\theta'\rangle^m\big]\\ &= \frac{1}{2^m}\sum_{k_1,\ldots,k_m=1}^{K}\sum_{s_1,\ldots,s_m\in\{+1,-1\}}\mathbb{1}\{s_1k_1 + \ldots + s_mk_m = 0\}\cdot\\ &\qquad\left[\left(\prod_{\ell=1}^{m}r_{k_\ell} - \prod_{\ell=1}^{m}r'_{k_\ell}\right)^2 + 2\left(\prod_{\ell=1}^{m}r_{k_\ell}r'_{k_\ell}\right)\left(1 - \cos\left(\sum_{\ell=1}^{m}s_\ell\phi_{k_\ell} - s_\ell\phi'_{k_\ell}\right)\right)\right] =: \mathrm{I} + \mathrm{II},\end{aligned}$$

where I is the term involving $(\prod_\ell r_{k_\ell} - \prod_\ell r'_{k_\ell})^2$, and II is the term involving $\cos(\sum_\ell s_\ell \phi_{k_\ell} - s_\ell \phi'_{k_\ell})$.

To upper bound I, let us write

$$\prod_{\ell=1}^m r_{k_\ell} - \prod_{\ell=1}^m r'_{k_\ell} = \sum_{j=1}^m (r_{k_j} - r'_{k_j}) r_{k_1} \dots r_{k_{j-1}} r'_{k_{j+1}} \dots r'_{k_m}.$$

Then

$$\left(\prod_{\ell=1}^m r_{k_\ell} - \prod_{\ell=1}^m r'_{k_\ell}\right)^2 \leq m \cdot \sum_{j=1}^m (r_{k_j} - r'_{k_j})^2 \left(r_{k_1} \dots r_{k_{j-1}} r'_{k_{j+1}} \dots r'_{k_m}\right)^2.$$

So

$$\mathrm{I} \leq \sum_{j=1}^m \sum_{k_1,\dots,k_m=1}^K \sum_{s_1,\dots,s_m \in \{+1,-1\}} \mathbb{1}\{s_1k_1 + \dots + s_mk_m = 0\} \cdot \frac{m}{2^m} (r_{k_j} - r'_{k_j})^2 \left(r_{k_1} \dots r_{k_{j-1}} r'_{k_{j+1}} \dots r'_{k_m}\right)^2$$

Consider this summand for $j = 1$. Note that fixing $s_1, \dots, s_m$ and $k_1, \dots, k_{m-1}$, there is at most one choice for the remaining index $k_m \in \{1, \dots, K\}$ that satisfies $s_1k_1 + \dots + s_mk_m = 0$. Thus, the summand for $j = 1$ is at most

$$\max_{k_m=1}^K {r'_{k_m}}^2 \cdot \sum_{k_1,\dots,k_{m-1}=1}^K \sum_{s_1,\dots,s_m \in \{+1,-1\}} \frac{m}{2^m} (r_{k_1} - r'_{k_1})^2 \left(r'_{k_2} \dots r'_{k_{m-1}}\right)^2 \leq m \sum_{k=1}^K (r_k - r'_k)^2 \bar{r}^2 R^{2(m-2)}.$$

The same bound holds for each summand $j = 1, \dots, m$, yielding

$$\mathrm{I} \leq m^2 \sum_{k=1}^K (r_k - r'_k)^2 \bar{r}^2 R^{2(m-2)}.$$

To bound II, observe that when $s_1k_1 + \dots + s_mk_m = 0$, we have

$$\cos\left(\sum_{\ell=1}^m s_\ell \phi_{k_\ell} - s_\ell \phi'_{k_\ell}\right) = \cos\left(\sum_{\ell=1}^m s_\ell \phi_{k_\ell} - s_\ell \phi'_{k_\ell} + \alpha \cdot s_\ell k_\ell\right)$$

for any $\alpha \in \mathbb{R}$. Then applying $1 - \cos(x) \leq x^2/2$ for any $x \in \mathbb{R}$, we obtain

$$2\left(1 - \cos\left(\sum_{\ell=1}^m s_\ell \phi_{k_\ell} - s_\ell \phi'_{k_\ell}\right)\right) \leq \inf_{\alpha \in \mathbb{R}} \left(\sum_{\ell=1}^m s_\ell \phi_{k_\ell} - s_\ell \phi'_{k_\ell} + \alpha \cdot s_\ell k_\ell\right)^2 = \inf_{\alpha \in \mathbb{R}} m \sum_{j=1}^m (\phi_{k_j} - \phi'_{k_j} + \alpha k_j)^2.$$

So

$$\mathrm{II} \leq \inf_{\alpha \in \mathbb{R}} \sum_{j=1}^m \sum_{k_1,\dots,k_m=1}^K \sum_{s_1,\dots,s_m \in \{+1,-1\}} \mathbb{1}\{s_1k_1 + \dots + s_mk_m = 0\} \cdot \frac{m}{2^m} (\phi_{k_j} - \phi'_{k_j} + \alpha k_j)^2 \left(\prod_{\ell=1}^m r_{k_\ell} r'_{k_\ell}\right).$$

Applying $\sum_{k=1}^K r_k r'_k \leq \sum_{k=1}^K (r_k^2 + {r'_k}^2)/2 \leq R^2$ and a similar argument as above, for any fixed $\alpha \in \mathbb{R}$, this summand for $j = 1$ is at most

$$\begin{aligned} &\max_{k_m=1}^K r_{k_m} r'_{k_m} \cdot \sum_{k_1,\dots,k_{m-1}=1}^K \sum_{s_1,\dots,s_m \in \{+1,-1\}} \frac{m}{2^m} (\phi_{k_1} - \phi'_{k_1} + \alpha k_1)^2 \left(r_{k_1} \dots r_{k_{m-1}} r'_{k_1} \dots r'_{k_{m-1}}\right) \\ &\leq m \sum_{k=1}^K r_k r'_k (\phi_k - \phi'_k + \alpha k)^2 \bar{r}^2 R^{2(m-2)}. \end{aligned}$$

The same bound holds for each summand $j = 1, \dots, m$, yielding

$$\mathrm{II} \leq \inf_{\alpha \in \mathbb{R}} m^2 \sum_{k=1}^K r_k r'_k (\phi_k - \phi'_k + \alpha k)^2 \cdot \bar{r}^2 \cdot R^{2(m-2)}.$$

Combining these bounds for I and II, we arrive at

$$\mathbb{E}\big[\langle\theta, g\theta\rangle^m - 2\langle\theta, g\theta'\rangle^m + \langle\theta', g\theta'\rangle^m\big] \le m^2\bar{r}^2 \cdot R^{2(m-2)} \cdot \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K}(r_k - r_k')^2 + r_k r_k'(\phi_k - \phi_k' + \alpha k)^2.$$

Let us now apply this to (116) and sum over $m \ge 3$: We have

$$\sum_{m=3}^{\infty} \frac{m^2 R^{2(m-2)}}{\sigma^{2m} m!} = \sum_{m=3}^{\infty} \frac{mR^2}{(m-1)(m-2)\sigma^6} \cdot \frac{R^{2(m-3)}}{\sigma^{2(m-3)}(m-3)!} \le \frac{3R^2}{2\sigma^6} e^{R^2/\sigma^2}.$$

Then

$$\begin{aligned}&\sum_{m=3}^{\infty} \frac{e^{\|\theta\|^2/2\sigma^2}}{\sigma^{2m} m!} \mathbb{E}\big[\langle\theta, g\theta\rangle^m - 2\langle\theta, g\theta'\rangle^m + \langle\theta', g\theta'\rangle^m\big] \\ &\le \frac{3\bar{r}^2 R^2 e^{3R^2/2\sigma^2}}{2\sigma^6} \cdot \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K}(r_k - r_k')^2 + r_k r_k'(\phi_k - \phi_k' + \alpha k)^2.\end{aligned}$$

Applying this and (117) to (116) gives (59). ☐

For a specific regime of parameters $\theta, \theta' \in \mathbb{R}^{2K}$, we simplify the lower bound for the loss in Proposition 3.1 by expressing the squared distance $|\phi_k - \phi_k' + k\alpha|_{\mathcal{A}}^2$ on the circle $\mathcal{A}$ in terms of the usual squared distance $(\phi_k - \phi_k' + k\alpha)^2$ on $\mathbb{R}$.

Lemma C.1. *Fix any $\theta, \theta' \in \mathbb{R}^{2K}$ and let $\theta = (r_k \cos\phi_k, r_k \sin\phi_k)_{k=1}^K$ and $\theta' = (r_k' \cos\phi_k', r_k' \sin\phi_k')_{k=1}^K$. For each $\alpha \in \mathbb{R}$, let $K_0(\alpha) \in [0, K]$ be the largest integer for which $|K_0(\alpha) \cdot \alpha| \le \pi/2$. If $r_k, r_k' \ge \underline{r}$ and $|\phi_k - \phi_k'| \le \pi/3$ for each $k = 1, \ldots, K$, then for a universal constant $c > 0$,*

$$L(\theta, \theta') \ge \sum_{k=1}^{K}(r_k - r_k')^2 + c \inf_{\alpha\in\mathbb{R}} \left((K - K_0(\alpha))\underline{r}^2 + \sum_{k=1}^{K_0(\alpha)} r_k r_k'(\phi_k - \phi_k' + k\alpha)^2\right) \tag{118}$$

where the second summation is understood as 0 if $K_0(\alpha) = 0$.

Proof. Recall the form (10) of the loss from Proposition 3.1, where the infimum over α may be restricted to $[-\pi, \pi]$ by periodicity. We provide a lower bound for $\alpha \in [0, \pi]$, and the case $\alpha \in [-\pi, 0]$ is analogous.

If $\alpha = 0$, let $K_0 = K_1 = \ldots = K$. Otherwise if $\alpha \in (0, \pi]$, let $0 \le K_0 \le K_1 \le K_2 \le \ldots$ be such that each K_m is the largest integer in $[0, K]$ for which $K_m \cdot \alpha \le 2\pi m + \frac{\pi}{2}$. Note that if $K_m < K$ strictly, then we must have also $K_m < K_{m+1}$. If $K_0 \ge 1$, then for every $k \in [1, K_0]$, we have $k\alpha \in [0, \pi/2]$, so $\phi_k - \phi_k' + k\alpha \in [-\pi/3, 5\pi/6]$ and

$$1 - \cos(\phi_k - \phi_k' + k\alpha) \ge c(\phi_k - \phi_k' + k\alpha)^2$$

for a universal constant $c > 0$. Thus

$$\sum_{k=1}^{K_0} r_k r_k' \Big[1 - \cos(\phi_k - \phi_k' + k\alpha)\Big] \ge c \sum_{k=1}^{K_0} r_k r_k'(\phi_k - \phi_k' + k\alpha)^2. \tag{119}$$

This bound is also trivially true if $K_0 = 0$.

Now fix any $m \ge 0$ where $K_m < K$ strictly. Consider the values

$$k \in \{K_m + 1, \ldots, K_{m+1}\}.$$

For each such k, we have $k\alpha \in (2\pi m + \frac{\pi}{2}, 2\pi m + \frac{5\pi}{2}]$. Let a be the number of such values k where $k\alpha \in (2\pi m + \frac{\pi}{2}, 2\pi m + \frac{3\pi}{2}]$, and let b be the number of such values k where $k\alpha \in (2\pi m + \frac{3\pi}{2}, 2\pi m + \frac{5\pi}{2}]$. Then we must have $a \geq 1$ because $\alpha \in [0, \pi]$. Also, the number of multiples of α belonging to $(2\pi m + \frac{3\pi}{2}, 2\pi m + \frac{5\pi}{2}]$ is at most 1 more than the number of multiples of α belonging to $(2\pi m + \frac{\pi}{2}, 2\pi m + \frac{3\pi}{2}]$, so $b \leq a+1$. Thus

$$\frac{a}{K_{m+1} - K_m} = \frac{a}{a+b} \geq \frac{a}{2a+1} \geq \frac{1}{3}.$$

For $k = K_m + 1, \ldots, K_m + a$, we must have $\phi_k - \phi'_k + k\alpha \in (2\pi m + \frac{\pi}{6}, 2\pi m + \frac{11\pi}{6}]$, so $1 - \cos(\phi_k - \phi'_k + k\alpha) \geq c$ for a universal constant $c > 0$. Then

$$\sum_{k=K_m+1}^{K_{m+1}} r_k r'_k \Big[1 - \cos(\phi_k - \phi'_k + k\alpha)\Big] \geq \sum_{k=K_m+1}^{K_m+a} c \cdot r_k r'_k \geq c\underline{r}^2 \cdot \frac{K_{m+1} - K_m}{3}.$$

Now summing over all $m \geq 0$ where $K_m < K$,

$$\sum_{k=K_0+1}^{K} r_k r'_k \Big[1 - \cos(\phi_k - \phi'_k + k\alpha)\Big] \geq \frac{c\underline{r}^2}{3} \cdot (K - K_0). \tag{120}$$

Applying (119) and (120) and the analogous bounds for $\alpha \in [-\pi, 0]$ to (10), and taking the infimum over $\alpha \in [-\pi, \pi]$, we obtain (118). □

We conclude the proof of Lemma 6.1 using the following version of Assouad's hypercube lower bound from (Cai and Zhou, 2012, Lemma 2).

Lemma C.2. *Fix $m \geq 1$, let $\{P_\tau : \tau \in \{0,1\}^m\}$ be any 2^m probability distributions, and let $\psi(P_\tau)$ take values in a metric space with metric d. Then for any $s > 0$ and any estimator $\hat{\psi}(X)$ based on $X \sim P_\tau$,*

$$\begin{aligned}&\sup_{\tau \in \{0,1\}^m} \mathbb{E}_{X \sim P_\tau}\Big[d(\hat{\psi}(X), \psi(P_\tau))^s\Big] \\ &\geq \frac{m}{2^{s+1}} \cdot \min_{H(\tau,\tau') \geq 1} \frac{d(\psi(P_\tau), \psi(P_{\tau'}))^s}{H(\tau, \tau')} \cdot \min_{H(\tau,\tau')=1} \Big(1 - D_{\mathrm{TV}}(P_\tau, P_{\tau'})\Big).\end{aligned}$$

Here, $H(\tau, \tau') = \sum_{i=1}^m \mathbb{1}\{\tau_i \neq \tau'_i\}$ is the Hamming distance between τ and τ', and $D_{\mathrm{TV}}(P_\tau, P_{\tau'})$ is the total-variation distance between P_τ and $P_{\tau'}$.

Proof of Lemma 6.1. We define 2^K parameters $\theta^\tau \in \mathcal{P}(r)$ indexed by $\tau \in \{0,1\}^K$: Fix a value $\phi \in [0, \pi/3]$ to be determined. For each $\tau \in \{0,1\}^K$, set

$$\phi_k^\tau = \tau_k \phi = \begin{cases} \phi & \text{if } \tau_k = 1 \\ 0 & \text{if } \tau_k = 0. \end{cases} \tag{121}$$

Then let θ^τ be the vector where $r_k(\theta) = r$ and $\phi_k(\theta) = \phi_k^\tau$ for each $k = 1, \ldots, K$.

Let $P_\tau = p_{\theta^\tau}^N$ denote the law of N samples $y^{(1)}, \ldots, y^{(N)} \overset{iid}{\sim} p_{\theta^\tau}$. Let $\mathcal{O}_\theta = \{g(\alpha) \cdot \theta : \alpha \in \mathcal{A}\}$ be the rotational orbit of θ. Then $d(\mathcal{O}_\theta, \mathcal{O}_{\theta'}) := L(\theta, \theta')^{1/2} = \min_{\alpha \in \mathcal{A}} \|\theta' - g(\alpha) \cdot \theta\|$ defines a metric over the space of all such orbits. We apply Lemma C.2 with $m = K$, $\psi(P_\tau) = \mathcal{O}_{\theta^\tau}$, this metric $d(\mathcal{O}_\theta, \mathcal{O}_{\theta'})$, and $s = 2$. Applying (58) and $|e^{is} - e^{it}| \leq |s - t|$ for all $s, t \in \mathbb{R}$,

$$D_{\mathrm{KL}}(p_{\theta^\tau} \| p_{\theta^{\tau'}}) \leq \frac{\|\theta^\tau - \theta^{\tau'}\|^2}{2\sigma^2} = \frac{r^2}{2\sigma^2} \sum_{k=1}^K \big|e^{i\phi_k^\tau} - e^{i\phi_k^{\tau'}}\big|^2 \leq \frac{r^2\phi^2}{2\sigma^2} \cdot H(\tau, \tau')$$

where $H(\tau, \tau')$ is the Hamming distance. Applying (59) with the right side evaluated at $\alpha = 0$, also

$$D_{\mathrm{KL}}(p_{\theta^\tau} \| p_{\theta^{\tau'}}) \leq \frac{r^2\phi^2}{A} \cdot H(\tau, \tau'), \qquad A := \frac{2\sigma^6}{3Kr^4 e^{3Kr^2/2\sigma^2}}$$

Then, setting

$$\phi = \min\left(\frac{1}{r\sqrt{N}} \cdot \max(\sqrt{2\sigma^2}, \sqrt{A}), \frac{\pi}{3}\right), \tag{122}$$

these bounds imply for both cases of the max that $D_{\mathrm{KL}}(p_{\theta^\tau} \| p_{\theta^{\tau'}}) \le H(\tau,\tau')/N$. Then by Pinsker's inequality (see e.g. (Tsybakov, 2008, Lemma 2.5)),

$$D_{\mathrm{TV}}(P_\tau, P_{\tau'}) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P_\tau \| P_{\tau'})} = \sqrt{\frac{N}{2} D_{\mathrm{KL}}(p_{\theta^\tau} \| p_{\theta^{\tau'}})} \le \sqrt{\frac{1}{2} H(\tau,\tau')},$$

so

$$\min_{H(\tau,\tau')=1} \Big(1 - D_{\mathrm{TV}}(P_\tau, P_{\tau'})\Big) \ge 1 - \sqrt{1/2} > 0.$$

Since $\phi_k \in [0,\pi/3]$ for every k, we may apply Lemma C.1 to lower-bound the loss: For a universal constant $c > 0$, we have

$$L(\theta^\tau, \theta^{\tau'}) \ge cr^2 \inf_{K_0 \in [0,K]} \inf_{\alpha \in \mathbb{R}} \left(K - K_0 + \sum_{k=1}^{K_0} (\phi_k^\tau - \phi_k^{\tau'} + k\alpha)^2 \right). \tag{123}$$

For any fixed $K_0 \in [2,K]$, the inner infimum over α is attained at $\alpha = -\sum_{k=1}^{K_0} k(\phi_k^\tau - \phi_k^{\tau'}) / \sum_{k=1}^{K_0} k^2$, and we have

$$\begin{aligned} \inf_{\alpha \in \mathbb{R}} \sum_{k=1}^{K_0} (\phi_k^\tau - \phi_k^{\tau'} + k\alpha)^2 &= \sum_{k=1}^{K_0} (\phi_k^\tau - \phi_k^{\tau'})^2 - \frac{\left(\sum_{k=1}^{K_0} k(\phi_k^\tau - \phi_k^{\tau'})\right)^2}{\sum_{k=1}^{K_0} k^2} \\ &= \phi^2 \cdot \left(H(\tau^{K_0}, \tau'^{K_0}) - \frac{\left(\sum_{k=1}^{K_0} k(\tau_k - \tau_k')\right)^2}{\sum_{k=1}^{K_0} k^2} \right) \end{aligned}$$

where $\tau^{K_0} = (\tau_1, \ldots, \tau_{K_0})$, $\tau'^{K_0} = (\tau_1', \ldots, \tau_{K_0}')$, and $H(\tau^{K_0}, \tau'^{K_0})$ is the Hamming distance of these subvectors in $\{0,1\}^{K_0}$. Subject to a constraint that $H(\tau^{K_0}, \tau'^{K_0}) = h$, we have

$$\begin{aligned} \left(\sum_{k=1}^{K_0} k(\tau_k - \tau_k') \right)^2 &\le \Big(K_0 + (K_0 - 1) + \ldots + (K_0 - h + 1) \Big)^2 \\ &= \left(\frac{h(2K_0 - h + 1)}{2} \right)^2 = h \cdot \frac{h(2K_0 - h + 1)^2}{4} \le h \cdot \frac{(2K_0+1)^3}{27}, \end{aligned}$$

where the last inequality is tight at the maximizer $h = (2K_0+1)/3$. Then for any $K_0 \in [2,K]$,

$$H(\tau^{K_0}, \tau'^{K_0}) - \frac{\left(\sum_{k=1}^{K_0} k(\tau_k - \tau_k')\right)^2}{\sum_{k=1}^{K_0} k^2} \ge H(\tau^{K_0}, \tau'^{K_0}) \cdot \left(1 - \frac{(2K_0+1)^3/27}{K_0(K_0+1)(2K_0+1)/6} \right) \ge \frac{2H(\tau^{K_0}, \tau'^{K_0})}{27}.$$

Applying also $K - K_0 \ge H((\tau_{K_0+1}, \ldots, \tau_K), (\tau_{K_0+1}', \ldots, \tau_K'))$ and $\phi \le \pi/3$, we get

$$\inf_{\alpha \in \mathbb{R}} \left(K - K_0 + \sum_{k=1}^{K_0} (\phi_k^\tau - \phi_k^{\tau'} + k\alpha)^2 \right) \ge c\phi^2 H(\tau, \tau') \tag{124}$$

for a universal constant $c > 0$. For $K_0 = 0$ or $K_0 = 1$ (and any $K \ge 2$), we may instead lower bound the left side by $K - K_0 \ge K/2 \ge H(\tau,\tau')/2$, so that this bound (124) holds also. Thus, taking the infimum in (123) over $K_0 \in [0,K]$,

$$d(\psi(P_\tau), \psi(P_{\tau'}))^2 = L(\theta^\tau, \theta^{\tau'}) \ge c' r^2 \phi^2 \cdot H(\tau,\tau') \tag{125}$$

for a universal constant $c' > 0$. Applying Lemma C.2 with $m = K$ and $s = 2$, we obtain

$$\inf_{\hat\theta} \sup_{\theta^* \in \mathcal{P}(r)} \mathbb{E}_{\theta^*}[L(\theta^*, \hat\theta)] \ge \inf_{\hat\theta} \sup_{\theta^\tau : \tau \in \{0,1\}^K} \mathbb{E}_{\theta^\tau}[L(\theta^\tau, \hat\theta)] \ge c'' K r^2 \phi^2.$$

Applying the form of ϕ from (122) concludes the proof. □

Acknowledgements The authors would like to thank Yihong Wu for a helpful discussion about KL-divergence in mixture models.

Funding Z. Fan is supported in part by NSF DMS 1916198, DMS 2142476. H. H. Zhou is supported in part by NSF grants DMS 2112918, DMS 1918925, and NIH grant 1P50MH115716.

References

Emmanuel Abbe, Tamir Bendory, William Leeb, João M Pereira, Nir Sharon, and Amit Singer. Multireference alignment is easier with an aperiodic translation distribution. *IEEE Transactions on Information Theory*, 65(6):3565–3584, 2018a.

Emmanuel Abbe, Joao M Pereira, and Amit Singer. Estimation in the group action channel. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 561–565. IEEE, 2018b.

Afonso S Bandeira, Moses Charikar, Amit Singer, and Andy Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 459–470, 2014.

Afonso S Bandeira, Yutong Chen, and Amit Singer. Non-unique games over compact groups and orientation estimation in cryo-em. *arXiv preprint arXiv:1505.03840*, 2015.

Afonso S Bandeira, Ben Blum-Smith, Joe Kileel, Amelia Perry, Jonathan Weed, and Alexander S Wein. Estimation under group actions: recovering orbits from invariants. *arXiv preprint arXiv:1712.10163*, 2017.

Afonso S Bandeira, Jonathan Niles-Weed, and Philippe Rigollet. Optimal rates of estimation for multi-reference alignment. *Mathematical Statistics and Learning*, 2(1):25–75, 2020.

Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on signal processing*, 66(4):1037–1050, 2017.

Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020a.

Tamir Bendory, Ariel Jaffe, William Leeb, Nir Sharon, and Amit Singer. Super-resolution multi-reference alignment. *arXiv preprint arXiv:2006.15354*, 2020b.

Tamir Bendory, Dan Edidin, William Leeb, and Nir Sharon. Dihedral multi-reference alignment. *IEEE Transactions on Information Theory*, 2022.

Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

Nicolas Boumal, Tamir Bendory, Roy R Lederman, and Amit Singer. Heterogeneous multireference alignment: A single pass approach. In *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2018.

Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4): 325–376, 1992.

Victor-Emmanuel Brunel. Learning rates for gaussian mixtures under group action. In *Conference on Learning Theory*, pages 471–491. PMLR, 2019.

T Tony Cai and Harrison H Zhou. Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics*, 40(5):2389–2420, 2012.

Robert Diamond. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Science*, 1(10):1279–1287, 1992.

Zhou Fan, Yi Sun, Tianhao Wang, and Yihong Wu. Likelihood landscape and maximum likelihood estimation for the discrete orbit recovery model. *Communications on Pure and Applied Mathematics*, 2020.

Zhou Fan, Roy R Lederman, Yi Sun, Tianhao Wang, and Sheng Xu. Maximum likelihood for high-noise group orbit estimation and single-particle cryo-em. *arXiv preprint arXiv:2107.01305*, 2021.

Subhro Ghosh and Philippe Rigollet. Multi-reference alignment for sparse signals, uniform uncertainty principles and the beltway problem. *arXiv preprint arXiv:2106.12996*, 2021.

Zvi Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, 82(1):15–39, 1980.

Anya Katsevich and Afonso Bandeira. Likelihood maximization and moment matching in low snr gaussian mixture models. *arXiv preprint arXiv:2006.15202*, 2020.

Rafal Latala. Estimates of moments and tails of Gaussian chaoses. *The Annals of Probability*, 34(6):2315–2331, 2006.

Shuyang Ling. Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Applied and Computational Harmonic Analysis*, 2022.

Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Message-passing algorithms for synchronization problems over compact groups. *Communications on Pure and Applied Mathematics*, 71 (11):2275–2322, 2018.

Amelia Perry, Jonathan Weed, Afonso S Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multireference alignment. *SIAM Journal on Mathematics of Data Science*, 1(3):497–517, 2019.

Thomas Pumir, Amit Singer, and Nicolas Boumal. The generalized orthogonal procrustes problem in the high noise regime. *Information and Inference: A Journal of the IMA*, 10(3):921–954, 2021.

Ya'Acov Ritov. Estimating a signal with noisy nuisance parameters. *Biometrika*, 76(1):31–37, 1989.

Elad Romanov, Tamir Bendory, and Or Ordentlich. Multi-reference alignment in high dimensions: Sample complexity and phase transition. *SIAM Journal on Mathematics of Data Science*, 3(2):494–523, 2021.

Brian M Sadler and Georgios B Giannakis. Shift-and rotation-invariant object reconstruction using the bispectrum. *JOSA A*, 9(1):57–69, 1992.

Sjors HW Scheres. Relion: implementation of a bayesian approach to cryo-em structure determination. *Journal of structural biology*, 180(3):519–530, 2012.

Sjors HW Scheres, Mikel Valle, Rafael Nuñez, Carlos OS Sorzano, Roberto Marabini, Gabor T Herman, and Jose-Maria Carazo. Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of molecular biology*, 348(1):139–149, 2005.

Nir Sharon, Joe Kileel, Yuehaw Khoo, Boris Landa, and Amit Singer. Method of moments for 3d single particle ab initio modeling with non-uniform distribution of viewing angles. *Inverse Problems*, 36(4): 044003, 2020.

Fred J Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology*, 122(3):328–339, 1998.

Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.

Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM journal on imaging sciences*, 4(2):543–572, 2011.

Amit Singer and Fred J Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual Review of Biomedical Data Science*, 3, 2020.

Amit Singer, Ronald R Coifman, Fred J Sigworth, David W Chester, and Yoel Shkolnisky. Detecting consistent common lines in cryo-em by voting. *Journal of structural biology*, 169(3):312–322, 2010.

Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.