The Annals of Statistics
2024, Vol. 52, No. 1, 261–284
https://doi.org/10.1214/23-AOS2346

RATES OF ESTIMATION FOR HIGH-DIMENSIONAL MULTIREFERENCE ALIGNMENT

BY ZEHAO DOU[a], ZHOU FAN[b] AND HARRISON H. ZHOU[c]

Department of Statistics and Data Science, Yale University, [a]*zehao.dou@yale.edu,* [b]*zhou.fan@yale.edu,* [c]*huibin.zhou@yale.edu*

We study the continuous multireference 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 $\sigma^2 \gtrsim K$, for signals with Fourier coefficients of roughly uniform magnitude, 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. We extend these analyses also to signals whose Fourier coefficients have a slow power law decay.

1. Introduction. Multireference alignment (MRA) refers to the problem of estimating an unknown signal from noisy samples that are subject to latent rotational transformations [4, 28]. 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 [7, 38]. It arises also in various other applications in structural biology and image registration [13, 16, 30]. Recent literature has established rates of estimation for MRA in fixed dimensions [1, 6, 20, 25], 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 [6, 18]. The true function f is identifiable only up to rotation, and we will study its estimation under the rotation-invariant squared-error loss

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

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 [4, 25]. 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

Received August 2023.
MSC2020 subject classifications. Primary 62G05; secondary 62C20.
Key words and phrases. Function estimation, group invariance, bispectrum inversion.

for continuous MRA. Many such results have specifically focused on a regime of high noise: In this regime, [25] showed that the squared-error risk for estimating "generic" signals scales with the noise standard deviation as σ^6. Bandeira et al. [6] showed that this scaling for estimating a "nongeneric" signal depends on its pattern of zero and nonzero Fourier coefficients, and derived rate-optimal upper and lower bounds over minimax classes of such signals. Rates of estimation for MRA with nonuniform rotations were studied in [1], with a dihedral group of both rotations and reflections in [9], with sparse signals in [20], and with down-sampled observations in a superresolution context in [10].

It is empirically observed, for example, in [18], 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

$$\theta^* = (\theta^*_{1,1}, \theta^*_{1,2}, \theta^*_{2,1}, \theta^*_{2,2}, \theta^*_{3,1}, \theta^*_{3,2}, \dots)$$

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

$$f(t) = \sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi}} \theta^*_{k,1} \cos kt + \frac{1}{\sqrt{\pi}} \theta^*_{k,2} \sin kt,$$

and let

$$(r_k \cos \phi_k, r_k \sin \phi_k) = (\theta^*_{k,1}, \theta^*_{k,2}) \tag{2}$$

be the representation of the kth Fourier frequency in terms of the magnitude r_k and phase ϕ_k. Fixing a decay parameter $\beta \in [0, \frac{1}{2})$, we consider a class of signals f represented by

$$\Theta_\beta = \{ f : r_k \asymp k^{-\beta} \text{ for } k = 1, \dots, K, r_k = 0 \text{ for all } k \geq K+1 \},$$

where we bandlimit f to its first K Fourier frequencies. Our results distinguish two separate signal-to-noise regimes for estimating f, based on the size of the entrywise noise variance σ^2 in the Fourier basis. We establish sharp minimax rates of estimation in both regimes, for sufficiently large sample size N, that are explicit in their dependence on the dimension K.

THEOREM (Informal). *Let $\beta \in [0, \frac{1}{2})$.*

(a) (*High noise*) *If $\sigma^2 \gtrsim K^{1-2\beta}$ and $N \gtrsim K^{6\beta}\sigma^6 \log K$, then*

$$\inf_{\hat{f}} \sup_{f \in \Theta_\beta} \mathbb{E}[L(\hat{f}, f)] \asymp \frac{K^{4\beta}\sigma^6}{N}.$$

(b) (*Low noise*) *If $\sigma^2 \lesssim K^{1-2\beta}/\log K$ and $N \gtrsim K^{1+2\beta}\sigma^2 \log K$, then*

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

We refer to Theorems 2.1 and 2.2 for precise statements of these results. Our signal class with power law decay $\beta < 1/2$ is representative of a setting where the average power per Fourier frequency, $\|\theta^*\|^2/K \asymp K^{-2\beta}$, is of comparable magnitude to the power r_k^2 at a typical frequency $k \in \{1, \dots, K\}$. Our analyses of the estimators that achieve these minimax rates apply more generally to signals of this form (cf. Theorems 4.1 and 5.2).

For large N, this result implies that there is a sharp transition in the minimax estimation rate near the noise level $\sigma^2 \asymp K^{1-2\beta} \asymp \|\theta^*\|^2$, which separates the two signal-to-noise regimes of the problem. Such a transition may be anticipated by the results of [3], where $\sigma^2 \gtrsim \|\theta^*\|^2$ is the condition required to carry out the high-noise Taylor expansion of the chi-squared divergence and of [29], which provided a sharp analysis of the sample complexity in the low-noise regime for an analogous discrete MRA model (see below). As σ^2 varies in the small parameter window from $K^{1-2\beta}/\log K$ to $K^{1-2\beta}$ between these "low-noise" and "high-noise" regimes, our result confirms that there must be a rapid increase in the minimax risk, from roughly the order $K^{2-2\beta}/N$ to $K^{3-2\beta}/N$.

In the high-noise regime where $\sigma^2 \gtrsim \|\theta^*\|^2$, 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 [25], and a notable new feature of the rate is its scaling with the dimension K, for example, when $\beta = 0$, the rate has no explicit dependence on K. In the MRA model, for functions having the Fourier coefficients (2), second-order moments correspond to the power spectrum

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

and third-order moments to the Fourier bispectrum

$$\{\phi_{k+l} - \phi_k - \phi_l : k, l \in \{1, \dots, K\} \text{ and } k + l \le K\}.$$

Method-of-moments in this context is also known as bispectrum inversion [8, 30], which aims to estimate the Fourier phases $\{\phi_k\}$ from an estimate of the bispectrum. Results of [8, 25] imply that for signals where $r_k \neq 0$ for every $k = 1, \dots, K$, these phases are uniquely determined by the bispectrum. Our analyses quantify the conditioning of the linear system relating the bispectrum to the Fourier phases, which gives rise to the quantitative dependence of the estimation rate on K. To resolve phase ambiguities before solving this linear system, we prove also an important ℓ_∞ stability property of bispectrum inversion (cf. Lemma 4.9), which is of independent interest.

Our definition of the low-noise regime $\sigma^2 \lesssim K^{1-2\beta}/\log K \asymp \|\theta^*\|^2/\log K$ and minimax rate in this regime are related to the work of [29], which studied instead the discrete MRA model in the asymptotic limit $K \to \infty$ and $(\sigma^2 \log K)/K \to 1/\alpha \in (0, \infty)$, for a Bayesian setting where θ^* has a standard Gaussian prior. This work showed a transition in the Bayes risk and associated sample complexity at the sharp threshold $\alpha = 2$. The analysis in [29] relied on the discreteness of the rotational model, analyzing a template matching procedure that exactly recovers the latent rotation for each sample. For continuous MRA, this estimation of each rotation is possible only up to a per-sample error that is independent of the sample size N, and averaging the correspondingly rotated samples would yield an estimation bias that does not vanish with N. Our analysis shows that 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 upper bound in this regime.

1.1. *Further related literature.* A body of 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 [35]. In the context of cryo-EM, this is known also as the "common lines" method [36, 37]. Algorithms developed and studied for estimating these pairwise alignments include spectral procedures [24, 35, 37], semidefinite relaxations [4, 5, 35, 37] and iterative power method or approximate message passing approaches [11, 26].

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 [1, 6, 9, 12, 20, 25]. 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 [27] and cryo-EM [33]. For cryo-EM, these methods encompass invariant-features approaches [21] and expectation-maximization algorithms [31, 32, 34]. The works [2, 3] studied method-of-moments estimators in problems with general rotational groups, where [3] 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, [14, 18, 19, 22] 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. 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, $D_{\mathrm{KL}}(P\|Q) = \int \log(\frac{P}{Q})\, \mathrm{d}P$ is their Kullback–Leibler (KL) divergence.

2. Model and main results. Let $\mathcal{S}^1 = [-\pi, \pi)$ be identified with the unit circle, with addition modulo 2π. Let $f : \mathcal{S}^1 \to \mathbb{R}$ be a smooth periodic function on $\mathcal{S}^1$. We represent rotations of the circle by angles $\alpha \in \mathcal{A} = [-\pi, \pi)$, 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

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

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. An equivalent Gaussian sequence formulation of the model is discussed below. 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 $t \in \mathcal{S}^1$ that are fixed or randomly sampled [6, 10]. 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, that is, f admits the Fourier sequence representation

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

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

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

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 $k = 1, \dots, K$ 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 $k = 1, \dots, K$ 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

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

where $\alpha^{(1)}, \dots, \alpha^{(N)} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}([-\pi, \pi))$, $\varepsilon^{(1)}, \dots, \varepsilon^{(N)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_{2K})$, 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

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

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 $\beta \in [0, \frac{1}{2})$ and $\underline{c}, \bar{c} > 0$, we consider a parameter space of "generic" Fourier coefficient vectors with power law decay rate β, given by

$$\Theta_\beta = \big\{ \theta^* \in \mathbb{R}^{2K} : \underline{c} k^{-\beta} \leq r_k\big(\theta^*\big) \leq \bar{c} k^{-\beta} \text{ for all } k = 1, \dots, K \big\}. \tag{7}$$

Here, "generic" refers to the quantitative lower bound for each value $r_k(\theta^*)$ that matches the assumed upper bound up to a constant factor. This condition may be viewed as an analogue of the genericity condition in [25] that all Fourier magnitudes are bounded above and below by a constant, in our high-dimensional setting of interest with potentially large K and decaying Fourier magnitudes.

Our main results are the following two theorems, which characterize the minimax rates of estimation over Θ_β in high-noise and low-noise regimes.

THEOREM 2.1 (Minimax risk in high noise). *Fix any $\beta \in [0, \frac{1}{2})$ and any constant $c_0 > 0$. If $\sigma^2 \geq c_0 K^{1-2\beta}$, then for a constant $C_0 > 0$ depending only on β, $\underline{c}$, $\bar{c}$, c_0 and for any $N \geq C_0 K^{6\beta}\sigma^6 \log K$,*

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

THEOREM 2.2 (Minimax risk in low noise). *Fix any $\beta \in [0, \frac{1}{2})$. There exist constants $C_0, C_1 > 0$ depending only on β, $\underline{c}$, $\bar{c}$ such that if $\sigma^2 \leq \frac{K^{1-2\beta}}{C_1 \log K}$ and $N \geq C_0 K^{1+2\beta}\sigma^2 \log K$, then*

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

In both statements, $\mathbb{E}_{\theta^*}$ is the expectation over N samples $y^{(1)}, \dots, y^{(N)}$ 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 β, $\underline{c}$, $\bar{c}$, c_0.

3. Preliminaries.

3.1. *Bounds for the loss.* For $\phi, \phi' \in \mathcal{A} = [-\pi, \pi)$, 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

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

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 $C, c > 0$,

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

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

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

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

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 kth 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-hand side is the real inner product, and the right-hand 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 analyze an estimator based on a third-order method-of-moments idea. We prove a general risk bound that depends on the smallest nonzero Fourier magnitude $\underline{r} = \min_k r_k(\theta^*)$ of the true signal, valid for any noise level $\sigma^2 > 0$, and we show in particular that this achieves the minimax upper bound of Theorem 2.1 for signals $\theta^* \in \Theta_\beta$ 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 $k = 1, \ldots, K$,

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

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

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

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 in the Supplementary Material [17]), for any $(k, l) \in \mathcal{I}$ including the case $k = l$ we have

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

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 $k = 1, \dots, K$, 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 $\phi = (\phi_k : k = 1, \dots, K)$ by the least-squares estimator

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

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

Here, (18) is defined using the squared difference over $\mathbb{R}$ rather than over the periodic domain $\mathcal{A}$. Hence, the final estimate $\hat{\theta}$ depends 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 $\phi_1, \dots, \phi_K$ 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.

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 $\phi_1, \dots, \phi_K$ 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 $N \gtrsim \frac{\sigma^6}{\underline{r}^6}\log K$. Consequently, this estimator attains the same estimation rate without oracle knowledge. We summarize these results as the following theorem.

THEOREM 4.1. *Let $\hat{\theta} \in \{\hat{\theta}^{\mathrm{oracle}}, \hat{\theta}^{\mathrm{opt}}\}$ be the above method-of-moments estimator, where $\hat{\Phi}_{k,l}$ is chosen either using the oracle of Section 4.1 or the optimization procedure of Section 4.2. Suppose $r_k \geq \underline{r} > 0$ for each $k = 1, \dots, K$. There exist universal constants $C, C_0 > 0$ such that if $N \geq C_0(\frac{\sigma^6}{\underline{r}^6}\log K + \frac{\sigma^3}{\underline{r}^3}(\log K)^{3/2})$, then*

$$\mathbb{E}\big[L(\hat{\theta}, \theta^*)\big] \leq CK\left(\frac{\sigma^2}{N} + \frac{\sigma^4}{N\underline{r}^2}\right) + \frac{C\|\theta^*\|^2}{K}\left(\frac{K\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6}\right). \tag{19}$$

We remark that for signals where $\|\theta^*\|^2/K \asymp \underline{r}^2$, as is the case for our signal class Θ_β of interest, this risk bound reduces to

$$\mathbb{E}\big[L(\hat{\theta}, \theta^*)\big] \leq \frac{C}{N}\left(K\sigma^2 + \frac{K\sigma^4}{\underline{r}^2} + \frac{\sigma^6}{\underline{r}^4}\right).$$

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

$$\Phi_{k,l} = \phi_{k+l} - \phi_k - \phi_l \in \mathbb{R}. \tag{20}$$

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

$$B_{k,l} = \tilde{\theta}_{k+l} \cdot \overline{\tilde{\theta}_k} \cdot \overline{\tilde{\theta}_l} = r_{k+l} r_k r_l e^{i(\phi_{k+l} - \phi_k - \phi_l)} \in \mathbb{C}, \tag{21}$$

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

$$\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l} = \operatorname{Arg}(\hat{B}_{k,l}/B_{k,l}) \in [-\pi, \pi). \tag{22}$$

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 that Theorem 4.1 holds for $\hat{\theta}^{\text{oracle}}$, deferring detailed proofs to Appendix A in the Supplementary Material [17]. We divide the argument 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 $c > 0$,*

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

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

Integrating these tail bounds yields the following immediate corollary.

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

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

Estimating $\Phi_{k,l}$. Applying a concentration inequality for cubic polynomials in independent Gaussian random variables, derived from [23], 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 $s > 0$,*

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

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

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

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* $C > 0$,

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

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

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

This implies that $\mathbb{E}[\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l}] = 0$, and hence $\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})] = 0$ when these bispectral components do not have any overlapping index, as stated in part (a) of the following lemma.

For $\Phi_{k,l}$ 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 Theorem 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. Indeed, applying the representation (22) and a first-order Taylor approximation $\operatorname{Arg} z = \operatorname{Im} \operatorname{Ln} z \approx \operatorname{Im}(z-1)$ around $z = 1$, we obtain $\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})] \approx \mathbb{E}[\operatorname{Im}(\hat{B}_{k,l}/B_{k,l} - 1)\operatorname{Im}(\hat{B}_{x,y}/B_{x,y} - 1)]$, and it is easily checked that this latter expectation is of size $O(\sigma^2/N\underline{r}^2)$, exhibiting a cancellation of the $O(\sigma^6/N\underline{r}^6)$ error. However, a naive bound for the error of this Taylor approximation remains of size $O(\sigma^6/N\underline{r}^6)$. Part (b) of the following lemma establishes a sharp bound for $\mathbb{E}[(\hat{\Phi}_{k,l}^{\text{oracle}} - \Phi_{k,l})(\hat{\Phi}_{x,y}^{\text{oracle}} - \Phi_{x,y})]$ by carrying out the Taylor expansion to a higher order $J \asymp N\underline{r}^6/\sigma^6$ with a remainder that is exponentially small in $N\underline{r}^6/\sigma^6$, and exhibiting a similar cancellation in expectation for all terms of the Taylor expansion up to this order J.

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*

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

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

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

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

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

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 (20) 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 kth standard basis vector. Then (18) is given explicitly by

$$\hat{\phi} = M^\dagger \hat{\Phi}, \tag{29}$$

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 $(1, 2, 3, \dots, K)$ 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 $(1, 2, 3, \dots, K) \in \mathbb{R}^K$. All $K-1$ nonzero 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 $(1, 2, 3, \dots, K)$.

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

$$\mathbb{E}\left[\inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K} r_k^2 |\hat{\phi}_k^{\mathrm{oracle}} - \phi_k + k\alpha|_{\mathcal{A}}^2\right] \leq \frac{C\|\theta^*\|^2}{K}\left(\frac{K\sigma^2}{N\underline{r}^2} + \frac{\sigma^6}{N\underline{r}^6} + Ke^{-c(\frac{N\underline{r}^6}{\sigma^6}\wedge\frac{N^{2/3}\underline{r}^2}{\sigma^2})}\right). \tag{30}$$

PROOF. By adding a multiple of $(1, 2, 3, \dots, K)$ to ϕ and absorbing this shift into α, we may assume without loss of generality that ϕ is orthogonal to $(1, 2, 3, \dots, K)$. Under this assumption, we will then upper bound the left-hand 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 $(1, 2, 3, \dots, K)$. Set $D = \operatorname{diag}(r_k^2)_{k=1}^K \in \mathbb{R}^{K\times K}$. Then applying $\operatorname{Tr} AB \leq \operatorname{Tr} B \cdot \|A\|_{\mathrm{op}}$ for positive semidefinite A, B, where $\|\cdot\|_{\mathrm{op}}$ is the $\ell_2 \to \ell_2$ operator norm,

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

Here, $\operatorname{Tr} D = \sum_{k=1}^K r_k^2 = \|\theta^*\|^2$, and Lemma 4.7 implies $\|M^\dagger M^{\dagger\top}\|_{\mathrm{op}} = \|(M^\top M)^\dagger\|_{\mathrm{op}} \leq 1/(K+1)$.

We have $\|A\|_{\mathrm{op}} \leq \|A\|_\infty$ 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

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

Combining the above concludes the proof. $\square$

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.

Theorem 4.1 for $\hat{\theta}^{\text{oracle}}$ 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

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

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 $\hat{\Phi}_{k,l}^{\text{opt}} \in [\tilde{\Phi}_{k,l} - \pi, \tilde{\Phi}_{k,l} + \pi)$ 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.

We prove Theorem 4.1 for $\hat{\theta}^{\text{opt}}$ 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. By "equivalent," we mean that ϕ and ϕ' represent the same Fourier phases up to rotation of the circular domain, that is, there exists $\alpha \in \mathbb{R}$ for which

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

Then using $\hat{\Phi}^{\text{opt}}$ achieves the same loss as using $\hat{\Phi}^{\text{oracle}}(\phi)$. 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 4.9. *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*

$$\left|\Phi_{k,l} - \Phi'_{k,l}\right|_{\mathcal{A}} \le \delta \quad \textit{for all } (k,l) \in \mathcal{I},$$

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

$$\left|\phi_k - \phi'_k - k\alpha\right|_{\mathcal{A}} \le \delta \quad \textit{for all } k = 1, \dots, K.$$

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. In other words, this is an $\ell_\infty \to \ell_\infty$ operator-norm bound for the matrix $M^\dagger$ from (29), where the ℓ_∞ norms are defined using the circular distance per coordinate and modulo the equivalence relation (32).

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 $\hat{\Phi}_{k,l}^{\text{opt}}$ that is chosen using $\tilde{\phi}$ must coincide exactly with the oracle choice $\hat{\Phi}_{k,l}^{\text{oracle}}(\phi')$, based on a phase vector ϕ' that is equivalent to the true phase vector ϕ.

COROLLARY 4.10. *Let $\hat{B}_{k,l}$ be as defined in (17), and suppose $\phi \in \mathbb{R}^K$ is such that*

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

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

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

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

By assumption, the right-hand 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 4.9, we obtain for some $\alpha\in\mathbb{R}$ and all $k = 1,\dots,K$ 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,

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

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

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

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

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 $\hat{\theta}^{\text{opt}}$ matches exactly that of $\hat{\theta}^{\text{oracle}}$. Combining with a crude bound for the loss on the complementary event, which has exponentially small probability in N, we obtain Theorem 4.1 for $\hat{\theta}^{\text{opt}}$.

REMARK 4.11. We study this two-stage estimation procedure primarily to enable a theoretical analysis of its risk. One may alternatively consider a more direct procedure where the least-squares objective (18) is defined using the squared distance $|\hat{\Phi}_{k,l} - (\phi_{k+l} - \phi_k - \phi_l)|^2_{\mathcal{A}}$ over the periodic domain $\mathcal{A}$, which would avoid the need to identify a version of $\hat{\Phi}_{k,l}$. However, analyzing the risk of such a procedure may require an ℓ_2-analogue of the stability guarantee of Lemma 4.9, which seems more challenging to obtain. Here, stability in the ℓ_∞ sense allows us to circumvent this issue by first estimating the oracle choices of $\hat{\Phi}_{k,l}$ using the ℓ_∞-objective (31).

Finally, let us check that this estimation guarantee in Theorem 4.1 coincides with our stated minimax rate in Theorem 2.1 when restricted to parameters $\theta^*\in\Theta_\beta$ and to the high-noise regime.

PROOF OF THEOREM 2.1, UPPER BOUND. For $\theta^*\in\Theta_\beta$, we have $\underline{r}^2 \ge cK^{-2\beta}$ and $\|\theta^*\|^2 \le CK^{1-2\beta}$, for ($\beta$-dependent) constants $C, c > 0$. Thus, the risk bound of Theorem 4.1 reduces to

$$\mathbb{E}\left[L(\hat{\theta}^{\text{opt}}, \theta^*)\right] \le \frac{C}{N}\left(K\sigma^2 + K^{1+2\beta}\sigma^4 + K^{4\beta}\sigma^6\right) \le \frac{C'K^{4\beta}\sigma^6}{N}$$

for constants $C, C' > 0$, the last inequality holding in the high-noise setting $\sigma^2 \ge c_0 K^{1-2\beta}$. In this setting, 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 for N in Theorem 4.1 is implied by $N \ge C_0' K^{6\beta}\sigma^6 \log K$ for a sufficiently large constant $C_0' > 0$, and this yields the minimax upper bound of Theorem 2.1. $\square$

We remark that Theorem 4.1 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 $\sigma^2 \lesssim K^{-2\beta}$, it also implies the upper bound of Theorem 2.2.

PROOF OF THEOREM 2.2, UPPER BOUND, FOR $\sigma^2 \leq K^{-2\beta}$. For $\sigma^2 \leq K^{-2\beta}$, the risk bound of Theorem 4.1 reduces instead to

$$\mathbb{E}\big[L(\hat{\theta}^{\text{opt}}, \theta^*)\big] \leq \frac{C}{N}\big(K\sigma^2 + K^{1+2\beta}\sigma^4 + K^{4\beta}\sigma^6\big) \leq \frac{C' K\sigma^2}{N}.$$

The required condition for N is implied by $N \geq C_0' K^{1+2\beta}\sigma^2 \log K$ for a sufficiently large constant $C_0' > 0$, and this yields the minimax upper bound of Theorem 2.2. $\square$

In high dimensions K and the noise regime $K^{-2\beta} \ll \sigma^2 \ll K^{1-2\beta}/\log K$, (19) exhibits the rate $K^{1+2\beta}\sigma^4/N$, which is larger than the minimax rate $K\sigma^2/N$. This 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 bound 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 in the Supplementary Material [17] an alternative "frequency marching" method, as discussed also in [8], 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 reestimating $\hat{\phi}$ by least squares, the resulting procedure achieves the same risk as described in Theorem 4.1 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 [31, 34], in this section we analyze the maximum likelihood estimator (MLE) in the setting of Theorem 2.2.

Define the log-likelihood function

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

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

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

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

For the results of this section, we isolate the following general condition for the Fourier magnitudes of θ^*.

ASSUMPTION 5.1. There exists a constant $c_{\text{gen}} > 0$ such that for any $B \subseteq \{1, \dots, K\}$ with $|B| \geq K/2$,

$$\sum_{k \in B} r_k(\theta^*)^2 \geq c_{\text{gen}} \|\theta^*\|^2.$$

It is clear that this condition holds for our signal class Θ_β of interest. Our main result is then the following general risk bound for $\hat{\theta}^{\text{MLE}}$ in the low-noise setting of Theorem 2.2.

THEOREM 5.2. *Suppose Assumption 5.1 holds. Then there exist constants $C, C_0, C_1 > 0$ depending only on c_{gen} such that if $\sigma^2 \leq \frac{K}{C_1 \log K}$ and $N \geq C_0 K(1 + \frac{K\sigma^2}{\|\theta^*\|^2}) \log(K + \frac{\|\theta^*\|^2}{\sigma^2})$, then*

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

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

In the remainder of this section, we prove Theorem 5.2. 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, $\hat{\theta}^{\text{MLE}}$ is defined only up to rotation, and all rotations of $\hat{\theta}^{\text{MLE}}$ incur the same loss. To fix this rotation and ease notation in the analysis, let us denote by $\hat{\theta}^{\text{MLE}}$ the rotation of the MLE such that

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

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

$$\begin{aligned} 0 &\geq R_N(\hat{\theta}^{\text{MLE}}) - R_N(\theta^*) \\ &= \nabla R_N(\theta^*)^\top (\hat{\theta}^{\text{MLE}} - \theta^*) + \frac{1}{2}(\hat{\theta}^{\text{MLE}} - \theta^*)^\top \nabla^2 R_N(\tilde{\theta})(\hat{\theta}^{\text{MLE}} - \theta^*), \end{aligned} \tag{37}$$

where $\tilde{\theta} \in \mathbb{R}^{2K}$ is on the line segment between θ^* and $\hat{\theta}^{\text{MLE}}$. Heuristically, Theorem 5.2 will follow from the bounds

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

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

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 5.2. In the remainder of this section, we describe the components of this argument, deferring detailed proofs to Appendix B in the Supplementary Material [17].

5.1. *Gradient and Hessian of the log-likelihood.* To simplify the model, observe that each sample $y^{(m)}$ satisfies the equality in law

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

Furthermore, $g(\alpha^{(m)})^{-1} g(\alpha) = g(\alpha - \alpha^{(m)})$ where, if $\alpha \sim \mathrm{Unif}([-\pi, \pi))$ 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

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

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

$$y^{(m)} = \theta^* + \sigma\varepsilon^{(m)} \quad \text{for } m = 1, \dots, N, \qquad \varepsilon^{(1)}, \dots, \varepsilon^{(N)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_{2K}). \tag{41}$$

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

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

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

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

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

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

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

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, and define the domain

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

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

LEMMA 5.3. *Suppose that Assumption 5.1 holds. 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* c_{gen}, δ_1 *such that if* $\sigma^2 \le \frac{\|\theta^*\|^2}{C_1 \log K}$ *and* $N \ge C_0 K$, *then*

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

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

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 $\varepsilon \sim \mathcal{N}(0, I_{2K})$.

For sufficiently small $|\alpha|$, 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

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

where this last approximation holds under Assumption 5.1. 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 5.4. *Suppose Assumption 5.1 holds, and* $\sigma^2 \le \|\theta^*\|^2$. *Then there are constants* $C_2, C_3 > 0$ *depending only on* c_{gen} *such that for any* $\theta \in \mathbb{R}^{2K}$,

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

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

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

for each integer $n \ge 1$. Summing this bound over all $n \ge 1$ yields Lemma 5.3.

5.3. *Lower bound for the information matrix.* In light of Lemma 5.3, to show (39) with high probability, it suffices to establish a version of the lower bound

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

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

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

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 $\hat{\theta}^{\mathrm{MLE}}$ to satisfy (36) also ensures $\hat{\theta}^{\mathrm{MLE}} - \theta^*$ is orthogonal to u^*. The statement (52) restricted to directions orthogonal to u^* is formalized in the following lemma.

LEMMA 5.5. *Suppose Assumption 5.1 holds. Fix any constant $\eta > 0$. There exist constants $C_0, C_1, \delta_1, c > 0$ depending only on c_{gen}, η such that if $\sigma^2 \le \frac{\|\theta^*\|^2}{C_1 \log K}$ and $N \ge C_0 K(1 + \frac{K\sigma^2}{\|\theta^*\|^2}) \log(K + \frac{\|\theta^*\|^2}{\sigma^2})$, then with probability at least $1 - e^{-\frac{cN}{(1+K\sigma^2/\|\theta^*\|^2)^2}}$, the following holds: For every $\theta \in \mathcal{B}(\delta_1)$ and every unit vector $v \in \mathbb{R}^{2K}$ satisfying $\langle u^*, v\rangle = 0$,*

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

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

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

The proof of Lemma 5.5 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

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

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 $\frac{\sigma^2}{K^2\|\theta^*\|^2}$. 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

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

for a small constant $\eta > 0$, which is sufficient to show Lemma 5.5.

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

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

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

$$\sup_{\alpha\in\mathcal{A}} \big|\langle \varepsilon, g(\alpha)\cdot\theta\rangle\big| \le \frac{\delta_1\|\theta^*\|^2}{\sigma},$$
$$\sup_{\alpha\in\mathcal{A}} \big|\langle \varepsilon, g(\alpha)\cdot v\rangle\big| \le \frac{\|\theta^*\|}{\sigma},$$
$$\sup_{\alpha,\alpha'\in[-\pi,\pi)} \frac{1}{\alpha^2}\left|\operatorname{Re}\sum_{k=1}^K \overline{\varepsilon_k} e^{ik\alpha'}\big(e^{ik\alpha} - 1 - ik\alpha\big)\theta_k\right| \le \frac{\delta_1 K^2\|\theta^*\|^2}{\sigma},$$
$$\sup_{\alpha,\alpha'\in[-\pi,\pi)} \frac{1}{|\alpha-\alpha'|}\left|\operatorname{Re}\sum_{k=1}^K \overline{\varepsilon_k}\big(e^{ik\alpha} - e^{ik\alpha'}\big)v_k\right| \le \frac{\delta_1 K\|\theta^*\|}{\sigma}.$$

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

LEMMA 5.6. *Suppose Assumption 5.1 holds. Fix any $\eta > 0$. There exist constants $C_1, \delta_1 > 0$ depending only on c_{gen}, η such that if $\sigma^2 \le \frac{\|\theta^*\|^2}{C_1 \log K}$, 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) $\varepsilon \in \mathcal{E}(\theta, v, \delta_1)$,*

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

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-hand side being a factor $\|\theta^*\|/\sigma$ larger than the mean value of the left-hand side, so that their failure probabilities are exponentially small in $\|\theta^*\|^2/\sigma^2$. This is summarized in the following result.

LEMMA 5.7. *Suppose Assumption 5.1 holds. 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 $c_{\mathrm{gen}}, \delta_1$, if $\sigma^2 \le \frac{\|\theta^*\|^2}{C_1\log K}$, then*

$$\mathbb{P}_{\varepsilon\sim\mathcal{N}(0,I)}\big[\varepsilon \notin \mathcal{E}(\theta, v, \delta_1)\big] \le e^{-c\|\theta^*\|^2/\sigma^2}.$$

Finally, we combine Lemmas 5.6 and 5.7 to conclude the proof of Lemma 5.5: We may write the second term of (54) as

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

The first sum is bounded by Lemma 5.6, while the second sum is sparse by Lemma 5.7 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 5.5.

5.4. *Proof of Theorem* 5.2. We now combine the preceding lemmas to conclude the proof of Theorem 5.2. Let $C_0, C_1, \delta_1 > 0$ be such that the conclusions of Lemma 5.5 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

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

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

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

Rearranging, we get

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

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

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

From (44), we have

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

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,

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

Combining with (56),

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

By Lemmas 5.3 and 5.5, $\mathbb{P}[\mathcal{E}^c] \leq e^{-\frac{cN}{(1+K\sigma^2/\|\theta^*\|^2)^2}}$ for some constant $c > 0$. Then applying also (48), for some constant $C > 0$,

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

Under the given assumption $N \geq C_0 K(1 + \frac{K\sigma^2}{\|\theta^*\|^2}) \log(K + \frac{\|\theta^*\|^2}{\sigma^2})$ for sufficiently large $C_0 > 0$, this implies also $N \geq C_0' K(1 + \frac{K\sigma^2}{\|\theta^*\|^2}) \log N$ for a large constant $C_0' > 0$. (This is verified in the proof of Lemma 5.5, cf. (S55) of Appendix B in the Supplementary Material [17].) Then

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

Combining the above two risk bounds on $\mathcal{E}$ and $\mathcal{E}^c$ yields Theorem 5.2.

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 known. Fix any $\beta \in [0, \frac{1}{2})$, and consider the parameter space

$$\mathcal{P}_\beta = \{\theta^* \in \mathbb{R}^{2K} : r_k(\theta^*) = k^{-\beta} \text{ for all } k = 1, \dots, K\}.$$

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

LEMMA 6.1. *Fix any $\beta \in [0, \frac{1}{2})$. Then for some β-dependent constants $C, c > 0$ and any $\sigma^2 > 0$,*

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \mathcal{P}_\beta} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq c \cdot \min\left(\frac{1}{N} \cdot \max\left(K\sigma^2, \frac{K^{4\beta}\sigma^6}{e^{CK^{1-2\beta}/\sigma^2}}\right), K^{1-2\beta}\right). \tag{57}$$

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. By rescaling, we may assume without loss of generality that $\underline{c} \leq 1 \leq \bar{c}$, and hence $\mathcal{P}_\beta \subset \Theta_\beta$. Assuming $\sigma^2 \geq c_0 K^{1-2\beta}$, choosing the second argument of $\max(\cdot)$ in (57) gives

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \Theta_\beta} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq \inf_{\hat{\theta}} \sup_{\theta^* \in \mathcal{P}_\beta} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq c \cdot \min\left(\frac{K^{4\beta}\sigma^6}{N}, K^{1-2\beta}\right)$$

for a constant $c > 0$ depending on c_0. When $N \geq C_0 K^{6\beta}\sigma^6 \log K$ for sufficiently large $C_0 > 0$, we have $K^{4\beta}\sigma^6/N < K^{1-2\beta}$, 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

$$\inf_{\hat{\theta}} \sup_{\theta^* \in \Theta_\beta} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq \inf_{\hat{\theta}} \sup_{\theta^* \in \mathcal{P}_\beta} \mathbb{E}_{\theta^*}[L(\theta^*, \hat{\theta})] \geq c \cdot \min\left(\frac{K\sigma^2}{N}, K^{1-2\beta}\right).$$

When $N \geq C_0 K^{1+2\beta}\sigma^2 \log K$ for sufficiently large $C_0 > 0$, we have $K\sigma^2/N < K^{1-2\beta}$, 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 in the Supplementary Material [17]. 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 [6] 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*

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

The upper bound (59) is sufficient to prove Lemma 6.1 in the setting $\beta = 0$, where the argument is as follows: We restrict attention to a discrete space of 2^K parameters $\theta^\tau \in \mathcal{P}_0$, indexed by the hypercube $\tau \in \{0,1\}^K$, where all Fourier magnitudes are equal to 1 and the Fourier phases $\phi^\tau = (\phi_1^\tau, \dots, \phi_K^\tau)$ are given by

$$\phi_k^\tau = \tau_k \cdot \phi.$$

Here, the value $\phi \in \mathbb{R}$ is chosen maximally while ensuring that $D_{\mathrm{KL}}(p_{\theta^\tau} \| p_{\theta^{\tau'}}) \le H(\tau, \tau')/N$ 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., [15], 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 $\mathcal{P}_0$. For more general decay parameters $\beta \in [0, \frac{1}{2})$, we apply a variation of this argument where the parameters θ^τ are defined such that only the Fourier phases ϕ_k^τ for $k > K/2$ are nonzero. We establish a modified version of (59) for the corresponding vectors θ^τ, where $\bar{r}$ may be replaced by the maximum of (r_k, r'_k) over $k > K/2$. The remainder of the proof is then similar to the $\beta = 0$ setting.

Acknowledgments. 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 Grants DMS-1916198, DMS-2142476.
H. H. Zhou is supported in part by NSF Grants DMS-2112918, DMS-1918925 and NIH Grant 1P50MH115716.

SUPPLEMENTARY MATERIAL

Supplement to "Rates of estimation for high-dimensional multireference alignment" (DOI: 10.1214/23-AOS2346SUPP; .pdf). The supplementary appendices contain proofs of the theorems and supporting lemmas in the main text.

REFERENCES

[1] ABBE, E., BENDORY, T., LEEB, W., PEREIRA, J. M., SHARON, N. and SINGER, A. (2019). Multireference alignment is easier with an aperiodic translation distribution. *IEEE Trans. Inf. Theory* **65** 3565–3584. MR3959006 https://doi.org/10.1109/TIT.2018.2889674

[2] ABBE, E., PEREIRA, J. M. and SINGER, A. (2018). Estimation in the group action channel. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 561–565. IEEE, New York.

[3] BANDEIRA, A. S., BLUM-SMITH, B., KILEEL, J., NILES-WEED, J., PERRY, A. and WEIN, A. S. (2023). Estimation under group actions: Recovering orbits from invariants. *Appl. Comput. Harmon. Anal.* **66** 236–319. MR4609472 https://doi.org/10.1016/j.acha.2023.06.001

[4] BANDEIRA, A. S., CHARIKAR, M., SINGER, A. and ZHU, A. (2014). Multireference alignment using semidefinite programming. In *ITCS'*14—*Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 459–470. ACM, New York. MR3359498

[5] BANDEIRA, A. S., CHEN, Y. and SINGER, A. (2015). Non-unique games over compact groups and orientation estimation in cryo-EM. Preprint. Available at arXiv:1505.03840.
[6] BANDEIRA, A. S., NILES-WEED, J. and RIGOLLET, P. (2019). Optimal rates of estimation for multi-reference alignment. *Math. Stat. Learn.* **2** 25–75. MR4073147 https://doi.org/10.4171/msl/11
[7] BENDORY, T., BARTESAGHI, A. and SINGER, A. (2020). Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Process. Mag.* **37** 58–76. https://doi.org/10.1109/msp.2019.2957822
[8] BENDORY, T., BOUMAL, N., MA, C., ZHAO, Z. and SINGER, A. (2018). Bispectrum inversion with application to multireference alignment. *IEEE Trans. Signal Process.* **66** 1037–1050. MR3771661 https://doi.org/10.1109/TSP.2017.2775591
[9] BENDORY, T., EDIDIN, D., LEEB, W. and SHARON, N. (2022). Dihedral multi-reference alignment. *IEEE Trans. Inf. Theory* **68** 3489–3499. MR4433234 https://doi.org/10.1109/tit.2022.3146488
[10] BENDORY, T., JAFFE, A., LEEB, W., SHARON, N. and SINGER, A. (2022). Super-resolution multi-reference alignment. *Inf. Inference* **11** 533–555. MR4440674 https://doi.org/10.1093/imaiai/iaab003
[11] BOUMAL, N. (2016). Nonconvex phase synchronization. *SIAM J. Optim.* **26** 2355–2377. MR3566919 https://doi.org/10.1137/16M105808X
[12] BOUMAL, N., BENDORY, T., LEDERMAN, R. R. and SINGER, A. (2018). Heterogeneous multireference alignment: A single pass approach. In 2018 52*nd Annual Conference on Information Sciences and Systems* (*CISS*) 1–6. IEEE, New York.
[13] BROWN, L. G. (1992). A survey of image registration techniques. *ACM Comput. Surv.* **24** 325–376.
[14] BRUNEL, V.-E. (2019). Learning rates for Gaussian mixtures under group action. In *Conference on Learning Theory* 471–491. PMLR.
[15] CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607 https://doi.org/10.1214/12-AOS998
[16] DIAMOND, R. (1992). On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci.* **1** 1279–1287. https://doi.org/10.1002/pro.5560011006
[17] DOU, Z., FAN, Z. and ZHOU, H. (2024). Supplement to "Rates of estimation for high-dimensional multireference alignment." https://doi.org/10.1214/23-AOS2346SUPP
[18] FAN, Z., LEDERMAN, R. R., SUN, Y., WANG, T. and XU, S. (2021). Maximum likelihood for high-noise group orbit estimation and single-particle cryo-EM. Preprint. Available at arXiv:2107.01305.
[19] FAN, Z., SUN, Y., WANG, T. and WU, Y. (2023). Likelihood landscape and maximum likelihood estimation for the discrete orbit recovery model. *Comm. Pure Appl. Math.* **76** 1208–1302. MR4582297
[20] GHOSH, S. and RIGOLLET, P. (2021). Multi-reference alignment for sparse signals, uniform uncertainty principles and the beltway problem. Preprint. Available at arXiv:2106.12996.
[21] KAM, Z. (1980). The reconstruction of structure from electron micrographs of randomly oriented particles. *J. Theoret. Biol.* **82** 15–39. https://doi.org/10.1016/0022-5193(80)90088-0
[22] KATSEVICH, A. E. and BANDEIRA, A. S. (2023). Likelihood maximization and moment matching in low SNR Gaussian mixture models. *Comm. Pure Appl. Math.* **76** 788–842. MR4569606 https://doi.org/10.1002/cpa.22051
[23] LATAŁA, R. (2006). Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34** 2315–2331. MR2294983 https://doi.org/10.1214/009117906000000421
[24] LING, S. (2022). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Appl. Comput. Harmon. Anal.* **60** 20–52. MR4387245 https://doi.org/10.1016/j.acha.2022.02.003
[25] PERRY, A., WEED, J., BANDEIRA, A. S., RIGOLLET, P. and SINGER, A. (2019). The sample complexity of multireference alignment. *SIAM J. Math. Data Sci.* **1** 497–517. MR4002723 https://doi.org/10.1137/18M1214317
[26] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.* **71** 2275–2322. MR3862091 https://doi.org/10.1002/cpa.21750
[27] PUMIR, T., SINGER, A. and BOUMAL, N. (2021). The generalized orthogonal Procrustes problem in the high noise regime. *Inf. Inference* **10** 921–954. MR4312088 https://doi.org/10.1093/imaiai/iaaa035
[28] RITOV, Y. (1989). Estimating a signal with noisy nuisance parameters. *Biometrika* **76** 31–37. MR0991420 https://doi.org/10.1093/biomet/76.1.31
[29] ROMANOV, E., BENDORY, T. and ORDENTLICH, O. (2021). Multi-reference alignment in high dimensions: Sample complexity and phase transition. *SIAM J. Math. Data Sci.* **3** 494–523. MR4245326 https://doi.org/10.1137/20M1354994
[30] SADLER, B. M. and GIANNAKIS, G. B. (1992). Shift-and rotation-invariant object reconstruction using the bispectrum. *J. Opt. Soc. Amer. A* **9** 57–69.

[31] SCHERES, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180** 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

[32] SCHERES, S. H. W., VALLE, M., NUÑEZ, R., SORZANO, C. O. S., MARABINI, R., HERMAN, G. T. and CARAZO, J.-M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348** 139–149. https://doi.org/10.1016/j.jmb.2005.02.031

[33] SHARON, N., KILEEL, J., KHOO, Y., LANDA, B. and SINGER, A. (2020). Method of moments for 3D single particle *ab initio* modeling with non-uniform distribution of viewing angles. *Inverse Probl.* **36** 044003, 40 pp. MR4072349 https://doi.org/10.1088/1361-6420/ab6139

[34] SIGWORTH, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122** 328–339. https://doi.org/10.1006/jsbi.1998.4014

[35] SINGER, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.* **30** 20–36. MR2737931 https://doi.org/10.1016/j.acha.2010.02.001

[36] SINGER, A., COIFMAN, R. R., SIGWORTH, F. J., CHESTER, D. W. and SHKOLNISKY, Y. (2010). Detecting consistent common lines in cryo-EM by voting. *J. Struct. Biol.* **169** 312–322. https://doi.org/10.1016/j.jsb.2009.11.003

[37] SINGER, A. and SHKOLNISKY, Y. (2011). Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. *SIAM J. Imaging Sci.* **4** 543–572. MR2810897 https://doi.org/10.1137/090767777

[38] SINGER, A. and SIGWORTH, F. J. (2020). Computational methods for single-particle electron cryomicroscopy. *Annu. Rev. Biomed. Data Sci.* **3** 163–190. https://doi.org/10.1146/annurev-biodatasci-021020-093826