The Annals of Statistics
2024, Vol. 52, No. 1, 52–77
https://doi.org/10.1214/23-AOS2292

MAXIMUM LIKELIHOOD FOR HIGH-NOISE GROUP ORBIT ESTIMATION AND SINGLE-PARTICLE CRYO-EM

BY ZHOU FAN[1,a], ROY R. LEDERMAN[1,b], YI SUN[2,e], TIANHAO WANG[1,c] AND SHENG XU[1,d]

[1]*Department of Statistics and Data Science, Yale University,* [a]*zhou.fan@yale.edu,* [b]*roy.lederman@yale.edu,* [c]*tianhao.wang@yale.edu,* [d]*sheng.xu@yale.edu*

[2]*Department of Statistics, University of Chicago,* [e]*yisun@statistics.uchicago.edu*

Motivated by applications to single-particle cryo-electron microscopy (cryo-EM), we study several problems of function estimation in a high noise regime, where samples are observed after random rotation and possible linear projection of the function domain. We describe a stratification of the Fisher information eigenvalues according to transcendence degrees of graded pieces of the algebra of group invariants, and we relate critical points of the log-likelihood landscape to a sequence of moment optimization problems, extending previous results for a discrete rotation group without projections.

We then compute the transcendence degrees and forms of these optimization problems for several examples of function estimation under $\mathsf{SO}(2)$ and $\mathsf{SO}(3)$ rotations, including a simplified model of cryo-EM as introduced by Bandeira, Blum-Smith, Kileel, Niles-Weed, Perry and Wein. We affirmatively resolve conjectures that third-order moments are sufficient to locally identify a generic signal up to its rotational orbit in these examples.

For low-dimensional approximations of the electric potential maps of two small protein molecules, we empirically verify that the noise scalings of the Fisher information eigenvalues conform with our theoretical predictions over a range of SNR, in a model of $\mathsf{SO}(3)$ rotations without projections.

1. Introduction. We study several problems of function estimation in low dimensions, where the function is observed under random and unknown rotations of its domain. Let $f : \mathcal{X} \to \mathbb{R}$ be a function on the unit circle $\mathcal{X} = \mathcal{S}^1$, the unit sphere $\mathcal{X} = \mathcal{S}^2$ or $\mathcal{X} = \mathbb{R}^3$. Let G be a rotation group acting on $\mathcal{X}$. We seek to estimate f from samples of the form

$$f_{\mathfrak{g}}(x) + \text{white noise},$$

where each sample consists of the function $f_{\mathfrak{g}}(x) = f(\mathfrak{g}^{-1} \cdot x)$ rotated by a uniformly random element $\mathfrak{g} \in \mathsf{G}$ and observed with continuous Gaussian white noise on $\mathcal{X}$. Equivalently, choosing an orthonormal basis for $L_2(\mathcal{X})$, the basis coefficients of $f_{\mathfrak{g}}$ are observed with i.i.d. Gaussian noise, having some entrywise noise variance $\sigma^2 > 0$. We focus on a regime of this problem where each sample has high noise $\sigma^2 \gtrsim \|f\|_{L_2}^2$, and the information from many rotated samples must be combined to obtain an accurate estimate of f. We study also a variant of this model where samples are observed under an additional linear projection.

Our primary motivation is a formulation of this problem that models molecular reconstruction in single-particle cryo-electron microscopy (cryo-EM) [13, 16, 21]. In this application, $f : \mathbb{R}^3 \to \mathbb{R}$ is the electric potential of an unknown molecular structure. Two-dimensional projections of this potential are measured for many samples of the molecule, each in a different and unknown rotated orientation, typically with a high level of measurement noise. The molecular structure is determined by estimating this electric potential f from the rotated and projected samples, and then fitting an atomic model [6, 41]. A brief introduction to

Received July 2021; revised May 2023.
MSC2020 subject classifications. 62G08.
Key words and phrases. Mixture models, nonparametric function estimation, group invariance.

cryo-EM and a discussion of its relation to the problems studied in this work are presented in Appendix F [14].

Among computational procedures for solving this reconstruction problem, regularized versions of maximum likelihood estimation (MLE), as implemented via expectation-maximization or stochastic gradient descent, are commonly used [35, 37, 38, 40]. However, many theoretical properties of the optimization landscape and reconstruction errors of these procedures are not fully understood in cryo-EM applications [6].

In this work, we study the Fisher information matrix and log-likelihood function landscape associated with maximum likelihood estimation for a basic model of the cryo-EM reconstruction problem, as well as several simpler statistical models with qualitative similarities. These models may be of independent interest while building up to the complexity of cryo-EM:

- (Continuous multireference alignment, Section 3.) Estimating a function on the unit circle $\mathcal{X} = \mathcal{S}^1$, under $\mathsf{SO}(2)$ rotations of the circle [5, 9].
- (Spherical registration, Section 4.1.) Estimating a function on the unit sphere $\mathcal{X} = \mathcal{S}^2$, under $\mathsf{SO}(3)$ rotations of the sphere [4], Section 4.5.
- (Unprojected cryo-EM, Section 4.2.) Estimating a function on $\mathbb{R}^3$ under $\mathsf{SO}(3)$ rotations about the origin, without tomographic projection [4], Section 4.7. Such a problem arises in a related application of cryo-ET, discussed in Appendix F.1.
- (Cryo-EM, Section 4.3.) Estimating a function on $\mathbb{R}^3$ under $\mathsf{SO}(3)$ rotations about the origin, with tomographic projection [4], Section 4.7.

1.1. *Group orbit recovery and related literature.* Classical literature on function estimation has explored the rich interplay between the complexity of infinite-dimensional function classes, the statistical difficulty of estimation and the role of regularization [22, 24, 42]. We restrict attention instead to a finite-dimensional function space for each of the above models, with the goal of understanding connections between estimation in these models having latent rotations and the algebraic structure of the underlying rotational group.

Choosing a d-dimensional function basis, each of the above function estimation problems may be restated as an *orbit recovery problem* [3, 4] of estimating the coefficients $\theta_* \in \mathbb{R}^d$ of f in this basis, from noisy observations of θ_* that are rotated by elements of a subgroup $\mathsf{G} \subset \mathsf{O}(d)$. This subgroup G represents the transformation of basis coefficients under rotations of the function domain $\mathcal{X}$. A body of recent literature has studied both specific and general instances of this orbit recovery problem [2–5, 12, 15, 33, 34, 36]. When G is the group of cyclic rotations of coordinates (a.k.a. discrete multireference alignment), [33] first proved that the optimal squared error for estimating generic signals $\theta_* \in \mathbb{R}^d$ in high noise is significantly larger than that in the model without latent rotations, scaling as σ^6 rather than as just the noise variance σ^2. This analysis was extended to nongeneric signals for continuous multireference alignment in [5] and to general group actions in [3, 4]. Many authors have studied further extensions of the multireference alignment and cryo-EM models, including models with nonuniform distributions of group elements [1, 39], with a dihedral group [8], with down-sampled observations [10] and with sparse signals [11, 18].

Our current work is inspired, in particular, by results of [4], which placed cryo-EM and other examples of function estimation in this context, and connected statistical properties of method-of-moments estimators in these problems to properties of the invariant algebra of the group action. Further connections between the algebraic structure of invariants and the geometry of the log-likelihood function landscape were developed in [15, 26]. As a central technical ingredient, these papers derived a series expansion of the log-likelihood function in powers of σ^{-1}, in [15] for orbit recovery models without linear projection, and in [26] for more general Gaussian mixture models that include the models we study in this work. We discuss below several relevant results of [4, 15, 26] in further detail.

1.2. *Overview of results.* In Section 2, we introduce the general orbit recovery model both with and without a linear projection, and describe results that connect geometric properties of the log-likelihood function to properties of the invariant algebra of the group action. In Sections 3 and 4, we apply this connection to study the preceding problems of function estimation, including continuous multireference alignment (MRA) and cryo-EM. In Section 5, we report results of numerical simulations for estimating the electric potential functions of two small protein molecules in an unprojected cryo-EM model, which corroborate predictions of our theory for the spectra of the Fisher information matrices.

Here, we provide a brief overview of these results in the context of related literature.

Fisher information and log-likelihood function landscape. For general orbit recovery problems with $\theta_* \in \mathbb{R}^d$, in a high-noise regime $\sigma^2 \gtrsim \|\theta_*\|^2$, results of [15, 26] demonstrated that it is informative to study properties of the (negative) population log-likelihood function $R(\theta)$ via a series expansion in powers of σ^{-1} of the form

$$R(\theta) = \sum_{k=0}^{\infty} \frac{1}{\sigma^{2k}} R_k(\theta). \tag{1.1}$$

Each term $R_k(\theta)$ is a G-invariant polynomial function of θ, which locally around θ_*, may depend on a number of "degrees-of-freedom" of θ (indicated by the rank of $\nabla^2 R_k(\theta_*)$) that may be strictly smaller than the total dimension d. For a model with discrete group G and no linear projection, [15] showed this number exactly coincides with $\operatorname{trdeg} \mathcal{R}^{\mathsf{G}}_{\le k}$, the transcendence degree of the invariant subalgebra generated by all G-invariant polynomials of degree $\le k$. This implies a graded structure of the Fisher information matrix $I(\theta_*)$ for generic signal vectors θ_*, where eigenvalues corresponding to different degrees-of-freedom have different scalings with σ^{-1}. Furthermore, local minimizers of $R(\theta)$ have a certain correspondence with successive local minimizers of each function $R_k(\theta)$.

In this work, we first extend these results to a model where $\mathsf{G} \subseteq \mathsf{O}(d)$ may be continuous, and samples may be observed with an additional linear projection. This extension encompasses a basic formulation of the molecular reconstruction problem in cryo-EM. The main result of [26] implies that a series expansion analogous to (1.1) continues to hold for the population log-likelihood function $R(\theta)$ in such a model. However, as anticipated from the structure of the expansion in [26], the algebraic properties of its terms differ from the unprojected setting studied in [15]. We show here that the number of degrees-of-freedom associated to each function $R_k(\theta)$ coincides with the transcendence degree of a possibly reduced subalgebra generated by order-k moments of the projected signal (Theorem 2.7). In addition, a version of the correspondence between local minimizers of $R(\theta)$ and of successive local minimizers of $R_k(\theta)$ remains true over a bounded domain of $\mathbb{R}^d$ (Theorems 2.11 and 2.13). When the group G is continuous, we extend the arguments of [15] to address technical issues arising from the Fisher information matrix $I(\theta_*)$ being singular, and the locus of minimizers of $R(\theta)$ being a manifold of positive rather than zero dimension.

Multireference alignment and cryo-EM. These general results enable our study of maximum likelihood procedures in specific function estimation problems, the main focus of our work. In high-noise regimes of these problems, statistical properties of the MLE are related to the structures of the subalgebras $\mathcal{R}^{\mathsf{G}}_{\le k}$ and to their transcendence degrees. In particular, the squared-error risk of the MLE is dictated by the smallest nonzero eigenvalue of the Fisher information matrix $I(\theta_*)$, and scales as σ^{2K} for generic signals θ_* where K is the smallest integer for which $\operatorname{trdeg} \mathcal{R}^{\mathsf{G}}_{\le K}$ equals $\operatorname{trdeg} \mathcal{R}^{\mathsf{G}}$, the transcendence degree of the full G-invariant algebra. This connects with a central result of [4], which showed that K is the lowest-order

moment needed to identify θ_* up to a finite list of group orbits, and that σ^{2K} is also the scaling of the sample complexity required for estimating θ_* up to such a finite list. We apply our general results to determine the explicit value of K in several function estimation examples.

For our model of continuous MRA on $\mathcal{S}^1$, we verify that $K = 3$ (Theorem 3.1). This is expected from known results about estimation using third-order moments in similar observation models for both discrete and continuous MRA in [5, 33]. We also show that the optimization landscape of $R(\theta)$ may possess spurious local minimizers even for generic Fourier coefficient vectors $\theta_* \in \mathbb{R}^d$ (Theorem 3.4) when the number of Fourier basis functions d exceeds a small constant. This statement is analogous to results shown for discrete MRA in [15], although our construction here in the continuous setting has a different structure.

For spherical registration and projected and unprojected cryo-EM under an $\mathsf{SO}(3)$-action, a primary contribution of our work is proving also that $\operatorname{trdeg} \mathcal{R}^{\mathsf{G}}_{\leq K} = \operatorname{trdeg} \mathcal{R}^{\mathsf{G}}$ for $K = 3$. An iterative algorithm for estimating θ_* from third-order moments in cryo-EM was first proposed by Kam in [25], which implicitly assumed that these moments are sufficient to identify θ_* (up to symmetries such as chirality). Formal conjectures that $K = 3$ were stated in [4] and verified numerically for small values of the basis dimension d in exact-precision arithmetic. We prove that $K = 3$ for d exceeding small absolute constants (Theorems 4.1, 4.6 and 4.9), hence resolving several of these conjectures that third-order moments are sufficient to locally identify the orbit of θ_*.

Writing the terms $R_k(\theta)$ of (1.1) as

$$R_k(\theta) = s_k(\theta) + q_k(\theta),$$

where $s_k(\theta)$ depends on the additional degrees-of-freedom of θ beyond those which define $R_{k-1}(\theta)$, our proofs of $K = 3$ leverage a connection between $\operatorname{trdeg} \mathcal{R}^{\mathsf{G}}_{\leq k}$ and the generic ranks of the Hessians $\nabla^2 s_k(\theta)$ (Lemma 2.8). We show that $\nabla^2 s_3(\theta)$ is generically of full rank by using an inductive "frequency marching" argument on the dimension d and explicitly analyzing $\operatorname{rank}(\nabla^2 s_3(\theta))$ for special choices of $\theta \in \mathbb{R}^d$. As a byproduct of these analyses, we derive explicit forms for $s_3(\theta)$, which define optimization problems analogous to bispectrum inversion problems studied in MRA models [7].

In an unprojected spherical registration model over $\mathcal{S}^2$, recent independent work of [29] provides a more quantitative version of this inductive frequency marching argument. The result of [29], Lemma 5.6, implies that for some absolute constant $\underline{d}$ and any dimension $d > \underline{d}$, the last $d - \underline{d}$ columns of $\nabla^2 s_3(\theta)$ are of full rank for generic $\theta \in \mathbb{R}^d$, and [29] obtained a quantitative lower bound on the smallest singular value in a smoothed analysis over θ. Our proofs show versions of this statement that are less quantitative but more explicit about the value of $\underline{d}$, holding down to $\underline{d}$ small enough where the full-rank condition for the entire matrix $\nabla^2 s_3(\theta)$ may be explicitly checked. We carry this out for both spherical registration and cryo-EM, and in particular, our inductive argument in the projected cryo-EM model is more complex than in the unprojected models and uses different ideas.

Simulations of the Fisher information for small proteins. To empirically investigate the predictions of this body of theory in a cryo-EM example, we computed in simulation the observed Fisher information matrices for the electric potential functions of two small protein molecules—a rotavirus VP6 trimer and hemoglobin—in a model without tomographic projection.

We developed and employed a procedure of adaptively constructing a radial function basis in the Fourier domain (Appendix E) so as to reduce the dimension of the function space needed to approximate the true potential. Applying this construction, we obtained function bases of dimension $d \approx 400$ that capture the coarse trimer structure of the rotavirus example, and of dimension $d \approx 4000$ that capture the secondary structures of both proteins up to spatial

resolutions of 7–8 Angstroms. At these dimensions and spatial resolutions, the theoretically predicted σ^{-2}, σ^{-4} and σ^{-6} scalings of the Fisher information eigenvalues were apparent in simulation for sufficiently high noise. We observe deviations from these predictions at lower levels of noise, and also in higher-dimensional function spaces that may be necessary to approximate the potentials to better spatial resolutions.

Notation. We use the conventions $\langle u, v\rangle = \sum_i \overline{u_i} v_i$ for the complex inner product, $\|u\|$ for the (real or complex) ℓ_2-norm, $\|M\|$ for the $\ell_2 \to \ell_2$ operator norm for matrices and $\mathbf{i} = \sqrt{-1}$ for the imaginary unit.

For a measure space (X, μ), $L_2(X, \mathbb{C})$ is the L_2-space of functions $f : X \to \mathbb{C}$ with inner-product $\int_X \overline{f(x)} g(x) \mu(\mathrm{d}x)$. We write $L_2(X) = L_2(X, \mathbb{R})$ for the analogous L_2-space of real-valued functions. $\mathcal{S}^1$ and $\mathcal{S}^2$ are the unit circle and unit sphere.

For differentiable $f : \mathbb{R}^d \to \mathbb{R}^k$, $\mathrm{d}f(x) \in \mathbb{R}^{k \times d}$ is its derivative or Jacobian at x. For twice-differentiable $f : \mathbb{R}^d \to \mathbb{R}$, $\nabla f(x) = \mathrm{d}f(x)^\top \in \mathbb{R}^d$ is its gradient, and $\nabla^2 f(x) \in \mathbb{R}^{d \times d}$ is its Hessian. We will write d_x, ∇_x, ∇_x^2 to clarify that the variable of differentiation is x. For a subset of coordinates y, we write $\nabla_y f(x)$ and $\nabla_y^2 f(x)$ as the components of this gradient and Hessian in y.

For a smooth manifold $\mathcal{M}$ and twice-differentiable $f : \mathcal{M} \to \mathbb{R}$, we write $\nabla f(x)|_{\mathcal{M}}$ and $\nabla^2 f(x)|_{\mathcal{M}}$ for its gradient and Hessian evaluated in any choice of local chart at $x \in \mathcal{M}$. We will often not make the choice of chart explicit when referring to properties of $\nabla f(x)|_{\mathcal{M}}$ and $\nabla^2 f(x)|_{\mathcal{M}}$ that do not depend on the specific choice of chart.

2. The general orbit recovery model in high noise.

2.1. *Model and likelihood.* Let $\theta_* \in \mathbb{R}^d$ be an unknown signal of interest. Let $\mathsf{G} \subseteq \mathsf{O}(d)$ be a known compact subgroup of the orthogonal group of dimension d. We denote by Λ the unique Haar probability measure on G, satisfying

$$\Lambda(\mathsf{G}) = 1, \qquad \Lambda(g \cdot S) = \Lambda(S \cdot g) = \Lambda(S),$$

for any $g \in \mathsf{G}$ and Borel measurable subset $S \subseteq \mathsf{G}$. In the *unprojected orbit recovery model*, we observe n noisy and rotated samples of θ_*, given by

$$Y_i = g_i \cdot \theta_* + \sigma \varepsilon_i \in \mathbb{R}^d, \quad i = 1, \dots, n, \tag{2.1}$$

where $g_1, \dots, g_n \overset{\text{i.i.d.}}{\sim} \Lambda$ are Haar-uniform random elements of G, and $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{Id}_{d \times d})$ are Gaussian noise vectors independent of $g_1, \dots, g_n$. The signal θ_* is identifiable only up to an arbitrary rotation in G, that is, it is identifiable up to its orbit

$$\mathcal{O}_{\theta_*} = \{g \cdot \theta_* : g \in \mathsf{G}\}.$$

Our goal is to estimate $\mathcal{O}_{\theta_*}$ from the observed rotated samples $Y_1, \dots, Y_n$.

In the *projected orbit recovery model*, we consider an additional known linear map $\Pi : \mathbb{R}^d \to \mathbb{R}^{\tilde{d}}$. (Note that Π may not necessarily be an orthogonal projection; our terminology is borrowed from the example of tomographic projection in cryo-EM.) We observe n samples

$$Y_i = \Pi(g_i \cdot \theta_*) + \sigma \varepsilon_i \in \mathbb{R}^{\tilde{d}}, \quad i = 1, \dots, n, \tag{2.2}$$

where $g_1, \dots, g_n \overset{\text{i.i.d.}}{\sim} \Lambda$ as before, and $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{Id}_{\tilde{d} \times \tilde{d}})$ are Gaussian noise vectors in the projected dimension $\tilde{d}$. Our goal is again to estimate $\mathcal{O}_{\theta_*}$ from $Y_1, \dots, Y_n$.

The unprojected and projected orbit recovery models are both Gaussian mixture models, where the distribution of mixture centers is the law of $g \cdot \theta_* \in \mathbb{R}^d$ or of $\Pi(g \cdot \theta_*) \in \mathbb{R}^{\tilde{d}}$ induced

by the uniform law $g \sim \Lambda$ over G. This mixture distribution may be continuous if $\mathsf{G} \subseteq \mathsf{O}(d)$ is a continuous subgroup. In both models, we denote the negative sample log-likelihood as

$$R_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log p_\theta(Y_i),$$

where $p_\theta(Y_i)$ is the Gaussian mixture density for Y_i, marginalizing over the unknown rotation $g_i \sim \Lambda$. This density is given in the projected setting by

$$p_\theta(y) = \int_{\mathsf{G}} \frac{1}{(2\pi\sigma^2)^{\tilde{d}/2}} \exp\left(-\frac{\|y - \Pi(g\cdot\theta)\|^2}{2\sigma^2}\right) \mathrm{d}\Lambda(g), \tag{2.3}$$

and in the unprojected setting by the same expression with $\Pi = \mathrm{Id}$ and $\tilde{d} = d$. The maximum likelihood estimator (MLE) of θ_* is $\hat{\theta}_n = \arg\min_{\theta\in\mathbb{R}^d} R_n(\theta)$. Since R_n satisfies the invariance $R_n(\theta) = R_n(g\cdot\theta)$ for all $g \in \mathsf{G}$, the MLE is also only defined up to its orbit $\mathcal{O}_{\hat{\theta}_n}$.

REMARK 2.1 (Identifiability of the orbit). The parameter θ_* is identifiable up to the distribution of the mixture centers $g\cdot\theta_*$ or $\Pi(g\cdot\theta_*)$. In the unprojected model, the equality in law $g\cdot\theta \overset{L}{=} g\cdot\theta'$ over $g \sim \Lambda$ holds if and only if $\mathcal{O}_\theta = \mathcal{O}_{\theta'}$, so θ_* is identifiable exactly up to its orbit.

In projected models, there may be further nonidentifiability. For instance, under the tomographic projection arising in cryo-EM, we have $\Pi(g\cdot\theta) \overset{L}{=} \Pi(g\cdot\theta')$ when θ' represents the mirror reflection of θ [6]. Thus, in this setting there may be two distinct orbits, which cannot be further identified, and θ_* is recovered only up to chirality.

In general, the number of distinct orbits with the same image under Π depends on the interaction between the structures of G and Π, and can be infinite. For example, for the trivial group $\mathsf{G} = \{\mathrm{Id}\}$ and the projection $\Pi : \mathbb{R}^d \to \mathbb{R}^{d-k}$ that removes the last k coordinates of θ, $\mathcal{O}_\theta = \{\theta\}$ and $\Pi(\mathcal{O}_\theta) = \Pi(\mathcal{O}_{\theta_*})$ for any θ sharing the same first $d-k$ coordinates as θ_*.

We use the equivalence notation

$$\Pi(\mathcal{O}_\theta) \equiv \Pi(\mathcal{O}_{\theta_*})$$

to mean that $\Pi(\mathcal{O}_\theta) = \Pi(\mathcal{O}_{\theta_*})$ as subsets of $\mathbb{R}^{\tilde{d}}$, and in addition, $\Pi(g\cdot\theta) \overset{L}{=} \Pi(g\cdot\theta_*)$ under the Haar-uniform law $g \sim \Lambda$. Thus, θ_* is identifiable up to this equivalence. We will restrict attention to projected models where

(2.4) there are a finite number of orbits $\mathcal{O}_\theta$ such that $\Pi(\mathcal{O}_\theta) \equiv \Pi(\mathcal{O}_{\theta_*})$, for generic $\theta_* \in \mathbb{R}^d$.

An equivalent algebraic characterization is provided in Proposition 2.6(b) below.

We denote the negative *population* log-likelihood function by

$$R(\theta) = \mathbb{E}\big[R_n(\theta)\big] = -\mathbb{E}\big[\log p_\theta(Y)\big], \tag{2.5}$$

where the expectation is taken under the true model $Y \sim p_{\theta_*}$. $R(\theta)$ depends implicitly on θ_*, but we will omit this dependence in the notation. This population log-likelihood is minimized at $\theta \in \mathcal{O}_{\theta_*}$ in the unprojected model, and at $\{\theta : \Pi(\mathcal{O}_\theta) \equiv \Pi(\mathcal{O}_{\theta_*})\}$ in projected models.

2.2. *Invariant polynomials and the high-noise expansion.* For sufficiently high noise σ^2, it is informative to study $R(\theta)$ via a series expansion of the Gaussian density of (2.3) in powers of σ^{-1}, as developed in [15, 26]. We review this expansion in this section.

Let $\mathcal{R}^{\mathsf{G}}$ be the (real) algebra of all G-invariant polynomial functions $p : \mathbb{R}^d \to \mathbb{R}$. These are the polynomials p that satisfy

$$p(\theta) = p(g\cdot\theta) \quad \text{for all } \theta \in \mathbb{R}^d \text{ and } g \in \mathsf{G}.$$

For each integer $k \geq 0$, let $\mathcal{R}^{\mathsf{G}}_{\leq k}$ be the subalgebra generated by the G-invariant polynomials having total degree at most k. This subalgebra consists of the polynomials $p \in \mathcal{R}^{\mathsf{G}}$ that may be expressed as $p(\theta) = q(p_1(\theta), \ldots, p_j(\theta))$ for some polynomial q and some $p_1, \ldots, p_j \in \mathcal{R}^{\mathsf{G}}$ each having degree $\leq k$ (where p itself may have degree larger than k).

Examples of polynomials in $\mathcal{R}^{\mathsf{G}}_{\leq k}$ include the entries of the symmetric moment tensors

$$T_k(\theta) = \int_{\mathsf{G}} (g \cdot \theta)^{\otimes k} \, \mathrm{d}\Lambda(g) \in \mathbb{R}^{d \times \cdots \times d}, \tag{2.6}$$

where $T_k(\theta)$ is a tensor of order k. The entries of $T_k(\theta)$ are the kth-order mixed moments of the distribution of Gaussian mixture centers $g \cdot \theta$. Conversely, any G-invariant polynomial $p(\theta)$ of degree $\leq k$ satisfies the identity

$$p(\theta) = \int_{\mathsf{G}} p(g \cdot \theta) \, \mathrm{d}\Lambda(g),$$

and decomposing p on the right-hand side into a sum of monomials shows that $p(\theta)$ is an affine linear combination of entries of $T_1, \ldots, T_k$. Hence, $\mathcal{R}^{\mathsf{G}}_{\leq k}$ is generated by $T_1, \ldots, T_k$, and the subalgebra $\mathcal{R}^{\mathsf{G}}_{\leq k}$ may be intuitively understood as containing all information in the moments of orders 1 to k for the Gaussian mixture defined by θ.

For the projected model with projection Π, we define analogously the projected moment tensors

$$\widetilde{T}_k(\theta) = \int_{\mathsf{G}} (\Pi \cdot g \cdot \theta)^{\otimes k} \, \mathrm{d}\Lambda(g) \in \mathbb{R}^{\tilde{d} \times \cdots \times \tilde{d}},$$

which are again the mixed moments of the Gaussian mixture centers $\Pi \cdot g \cdot \theta$. We then define

$$\widetilde{\mathcal{R}}^{\mathsf{G}}_{\leq k} = \text{subalgebra of } \mathcal{R}^{\mathsf{G}} \text{ generated by the entries of } \widetilde{T}_1, \ldots, \widetilde{T}_k.$$

Since each entry of $\widetilde{T}_k$ is a G-invariant polynomial of degree k, we have $\widetilde{\mathcal{R}}^{\mathsf{G}}_{\leq k} \subseteq \mathcal{R}^{\mathsf{G}}_{\leq k}$, but equality does not necessarily hold.

We denote by $\langle \cdot, \cdot \rangle$ the Euclidean inner product in the vectorization of these tensor spaces $\mathbb{R}^{d \times \cdots \times d}$ and $\mathbb{R}^{\tilde{d} \times \cdots \times \tilde{d}}$, and by $\| \cdot \|^2_{\mathrm{HS}}$ the corresponding squared Euclidean norm. We will use the following general form of the large-σ series expansion of the population log-likelihood $R(\theta)$. We explain how the results of [26] yield this form in Appendix A.

THEOREM 2.2. *Let $\mathsf{G} \subseteq \mathsf{O}(d)$ be any compact subgroup. Fix any $\theta_* \in \mathbb{R}^d$ and any integer $K \geq 0$.*

(a) *In the unprojected orbit recovery model, $R(\theta)$ admits an expansion*

$$R(\theta) = C_0 + \sum_{k=1}^{K} \frac{1}{\sigma^{2k}} \big(s_k(\theta) + q_k(\theta)\big) + q(\theta). \tag{2.7}$$

Here, $C_0 \in \mathbb{R}$, $q_k \in \mathcal{R}^{\mathsf{G}}_{\leq k-1}$ is a polynomial of degree at most $2k$, and $s_k \in \mathcal{R}^{\mathsf{G}}_{\leq k}$ is the polynomial

$$s_k(\theta) = \frac{1}{2(k!)} \big\| T_k(\theta) - T_k(\theta_*) \big\|^2_{\mathrm{HS}}. \tag{2.8}$$

The remainder $q(\theta)$ is G-invariant and satisfies, for all $\theta \in \mathbb{R}^d$ with $\|\theta\| \leq \sigma$,

$$\begin{aligned} |q(\theta)| &\leq \frac{C_K (1 \vee \|\theta\|)^{2K+2}}{\sigma^{2K+2}}, \\ \big\|\nabla q(\theta)\big\| &\leq \frac{C_K (1 \vee \|\theta\|)^{2K+1}}{\sigma^{2K+2}}, \\ \big\|\nabla^2 q(\theta)\big\| &\leq \frac{C_K (1 \vee \|\theta\|)^{2K}}{\sigma^{2K+2}}. \end{aligned} \tag{2.9}$$

(b) *In the projected orbit recovery model, $R(\theta)$ admits an expansion*

$$R(\theta) = C_0 + \sum_{k=1}^{K} \frac{1}{\sigma^{2k}} \big(\tilde{s}_k(\theta) + \langle \widetilde{T}_k(\theta), P_k(\theta)\rangle + q_k(\theta)\big) + q(\theta). \tag{2.10}$$

Here, $C_0 \in \mathbb{R}$, $q_k \in \widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k-1}$ is a polynomial of degree at most $2k$, all entries of P_k are polynomials of degree at most k belonging to $\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k-1}$, P_k satisfies $P_k(\theta_*) = 0$ and $\tilde{s}_k \in \widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k}$ is the polynomial,

$$\tilde{s}_k(\theta) = \frac{1}{2(k!)} \big\| \widetilde{T}_k(\theta) - \widetilde{T}_k(\theta_*) \big\|^2_{\mathrm{HS}}. \tag{2.11}$$

The remainder $q(\theta)$ is G-invariant and satisfies (2.9) for all $\theta \in \mathbb{R}^d$ with $\|\theta\| \le \sigma$.

The above constants C_0, C_K, the coefficients of the polynomials $q_k(\theta)$ and $P_k(\theta)$, and the forms of the functions $q(\theta)$ may all depend on θ_*, G, d, $\tilde{d}$ and the projection Π.

The exact forms of $q_k(\theta)$ and $P_k(\theta)$ can be explicitly derived (see [15], Section 4.2, for these derivations in the unprojected setting) but we will not require them in what follows. Our arguments will only require the forms of the "leading" terms $s_k(\theta)$ and $\tilde{s}_k(\theta)$ defined in (2.8) and (2.11).

2.3. *Fisher information in high noise.* Consider the Fisher information matrix

$$I(\theta_*) = \nabla^2 R(\theta)|_{\theta=\theta_*}.$$

In this section, we characterize the eigenvalues and eigenvectors of $I(\theta_*)$ for high noise and generic $\theta_* \in \mathbb{R}^d$. This generalizes [15], Theorem 4.14, for the unprojected model and a discrete group.

DEFINITION 2.3. A subset $S \subseteq \mathbb{R}^d$ is *generic* if $\mathbb{R}^d \setminus S$ is contained in the zero set of some nonzero analytic function $\psi : \mathbb{R}^d \to \mathbb{R}^k$, for some $k \ge 1$.

If $S \subseteq \mathbb{R}^d$ is generic, then $\mathbb{R}^d \setminus S$ has zero Lebesgue measure [32]. We say that a statement holds for generic $\theta_* \in \mathbb{R}^d$ if it holds for all θ_* in some generic subset of $\mathbb{R}^d$.

Our characterization of $I(\theta_*)$ is in terms of the number of distinct "degrees-of-freedom" captured by the moments of the Gaussian mixture model up to each order k. This is formalized by the notion of the transcendence degrees of the subalgebras $\mathcal{R}^{\mathsf{G}}_{\le k}$ and $\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k}$.

DEFINITION 2.4. Polynomials $p_1, \dots, p_k : \mathbb{R}^d \to \mathbb{R}$ are *algebraically independent* (over $\mathbb{R}$) if there is no nonzero polynomial $q : \mathbb{R}^k \to \mathbb{R}$ such that $q(p_1(\theta), \dots, p_k(\theta))$ is identically 0 for all $\theta \in \mathbb{R}^d$.

For any $\mathcal{A} \subseteq \mathcal{R}^{\mathsf{G}}$, its *transcendence degree* $\mathrm{trdeg}(\mathcal{A})$ is the maximum cardinality of any algebraically independent subset $A \subseteq \mathcal{A}$. Any maximal such subset $A \subseteq \mathcal{A}$ is a *transcendence basis* for $\mathcal{A}$.

Geometrically, by the Jacobian criterion for algebraic independence (cf. Lemma A.2), the transcendence degree coincides with the maximum number of linearly independent gradient vectors of the polynomials in $\mathcal{A}$, evaluated at any generic point $\theta \in \mathbb{R}^d$.

As a simple example, if G is the symmetric group of all permutations of d coordinates, then $\mathcal{R}^{\mathsf{G}}$ is the algebra of all symmetric polynomials in d variables. Each subalgebra $\mathcal{R}^{\mathsf{G}}_{\le k}$ for $k \le d$ has transcendence degree exactly equal to k, and one choice of a transcendence basis for $\mathcal{R}^{\mathsf{G}}_{\le k}$ is the set of symmetric power sums $\{\theta_1^j + \cdots + \theta_d^j : j = 1, \dots, k\}$.

For the full invariant algebra $\mathcal{R}^{\mathsf{G}}$, if $\mathsf{G} \subset \mathsf{O}(d)$ is any discrete subgroup as studied in [15], then $\mathrm{trdeg}(\mathcal{R}^{\mathsf{G}}) = d$. More generally, we have the following.

PROPOSITION 2.5. *Let* G *be a compact subgroup of* $\mathsf{O}(d)$. *Then*

$$\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d - \max_{\theta\in\mathbb{R}^d} \dim(\mathcal{O}_\theta),$$

where $\dim(\mathcal{O}_\theta)$ *is the dimension of the orbit* $\mathcal{O}_\theta$ *as a submanifold of* $\mathbb{R}^d$. *Here, the maximum orbit dimension* $\max_{\theta\in\mathbb{R}^d}\dim(\mathcal{O}_\theta)$ *is also the orbit dimension for generic points* $\theta\in\mathbb{R}^d$.

We will mostly consider group actions where this generic orbit dimension equals the group dimension $\dim(\mathsf{G})$, so that $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d - \dim(\mathsf{G})$. In particular, for the function estimation examples to be discussed in Sections 3 and 4, we will have $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d-1$ for an action of G that is isomorphic to $\mathsf{SO}(2)$, and $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d-3$ for an action of G that is isomorphic to $\mathsf{SO}(3)$.

It was shown in [4], Theorem 3.15, for generic signals $\theta_*\in\mathbb{R}^d$, that the values of the moment tensors $T_1(\theta_*),\dots,T_k(\theta_*)$ are sufficient to identify θ_* up to a finite list of possible orbits if and only if $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\le k}) = \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}})$. More informally, the order of moments needed to "locally" identify the orbit of θ_* coincides with the order of moments needed to capture all $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}})$ degrees-of-freedom of the invariant algebra. Throughout this paper, we will denote this number as K in the unprojected model and as $\widetilde{K}$ in the projected model, which are well defined by the following proposition. We defer proofs of Propositions 2.5 and 2.6 to Appendix A.

PROPOSITION 2.6. *For any compact subgroup* $\mathsf{G}\subseteq\mathsf{O}(d)$,

(a) *there is a smallest integer* $K<\infty$ *for which* $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\le K}) = \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}})$.

(b) Π *satisfies* (2.4) *if and only if there is a smallest integer* $\widetilde{K}<\infty$ *for which* $\operatorname{trdeg}(\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le \widetilde{K}}) = \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}})$.

In the unprojected model, let us now denote

$$d_0 = \max_{\theta\in\mathbb{R}^d}\dim(\mathcal{O}_\theta), \qquad d_k = \operatorname{trdeg}\mathcal{R}^{\mathsf{G}}_{\le k} - \operatorname{trdeg}\mathcal{R}^{\mathsf{G}}_{\le k-1} \quad \text{for } k=1,\dots,K \tag{2.12}$$

to decompose the total dimension of θ_* as $d = d_0 + d_1 + \cdots + d_K$. In the projected model, assuming the condition (2.4), let us similarly denote

$$\tilde{d}_0 = \max_{\theta\in\mathbb{R}^d}\dim(\mathcal{O}_\theta), \qquad \tilde{d}_k = \operatorname{trdeg}\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k} - \operatorname{trdeg}\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k-1} \quad \text{for } k=1,\dots,\widetilde{K} \tag{2.13}$$

to decompose the total dimension as $d = \tilde{d}_0 + \tilde{d}_1 + \cdots + \tilde{d}_{\widetilde{K}}$. The following result expresses the spectral properties of the Fisher information matrix in terms of these decompositions.

THEOREM 2.7. *For generic* $\theta_*\in\mathbb{R}^d$, *some* $(\theta_*, \mathsf{G}, \Pi)$*-dependent constants* $\sigma_0, C, c>0$ *and function* $\varepsilon(\sigma)$ *satisfying* $\varepsilon(\sigma)\to 0$ *as* $\sigma\to\infty$, *and all* $\sigma>\sigma_0$:

(a) *In the unprojected orbit recovery model*:

1. *The Fisher information matrix* $I(\theta_*)$ *has rank exactly* $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d - d_0$. *Defining* K *by Proposition* 2.6(*a*), *for each* $k=1,\dots,K$,

$$\text{exactly } d_k \text{ eigenvalues of } I(\theta_*) \text{ belong to } \left[c\sigma^{-2k}, C\sigma^{-2k}\right].$$

2. *For each* $k=1,\dots,K$, *let* V_k *be the subspace spanned by the leading* $d_1+\cdots+d_k$ *eigenvectors of* $I(\theta_*)$, *and let* W_k *be the subspace spanned by the gradient vectors* $\{\nabla p(\theta_*) : p\in\mathcal{R}^{\mathsf{G}}_{\le k}\}$. *Then the sin-theta distance between* V_k *and* W_k *is bounded as*

$$\left\|\sin\Theta(V_k, W_k)\right\| < \varepsilon(\sigma).$$

3. *For any $k = 1, \dots, K$ and any polynomial $p \in \mathcal{R}^{\mathsf{G}}_{\le k}$, the gradient $\nabla p(\theta_*) \in \mathbb{R}^d$ is orthogonal to the null space of $I(\theta_*)$ and satisfies*

$$\nabla p(\theta_*)^\top I(\theta_*)^\dagger \nabla p(\theta_*) \le C\sigma^{2k},$$

where $I(\theta_*)^\dagger$ is the Moore–Penrose pseudo-inverse.

(b) *In the projected orbit recovery model satisfying condition (2.4), the same statements hold with $\mathcal{R}^{\mathsf{G}}_{\le k}$, K and d_k replaced by $\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k}$, $\widetilde{K}$ and $\tilde{d}_k$, where $\widetilde{K}$ is defined by Proposition 2.6(b).*

REMARK. Theorem 2.7(a1) states that $I(\theta_*)$ has eigenvalues on differing scales of σ^{-2} in high noise, with d_k such eigenvalues scaling as σ^{-2k}, and $d_0 = \dim(\mathcal{O}_{\theta_*})$ eigenvalues of 0 representing the nonidentifiable degrees-of-freedom tangent to $\mathcal{O}_{\theta_*}$. Thus there are d_k degrees-of-freedom in θ_* that are estimated with asymptotic variance $O(\sigma^{2k}/n)$ by the MLE. The largest such variance is $O(\sigma^{2K}/n)$, which is in accordance with results about list-recovery of generic signals in [4] and with the σ^6 sample complexity established in [33] for multireference alignment, where $K = 3$.

REMARK. Theorem 2.7(a2) describes also the associated spaces of eigenvectors of $I(\theta_*)$, where the eigenspaces corresponding to eigenvalues at scales $\sigma^{-2}, \dots, \sigma^{-2k}$ coincide approximately with the span of the gradients of G-invariant polynomials up to degree k. Theorem 2.7(a3) then implies that the functional $p(\theta_*)$ for any $p \in \mathcal{R}^{\mathsf{G}}_{\le k}$ is estimated by the plug-in MLE $p(\hat{\theta}_n)$ with asymptotic variance $O(\sigma^{2k}/n)$. Similar statements hold for projected models by Theorem 2.7(b).

The following result connects the above sequences of transcendence degrees and gradients $\{\nabla p(\theta_*) : p \in \mathcal{R}^{\mathsf{G}}_{\le k}\}$ to the terms $s_k(\theta)$ and $\tilde{s}_k(\theta)$ in the series expansions of $R(\theta)$ in Theorem 2.2. We will use this to deduce the values of these transcendence degrees for the function estimation examples of Sections 3 and 4.

LEMMA 2.8.

(a) *In the unprojected orbit recovery model, let $s_k(\theta)$ be defined by (2.8). Then each matrix $\nabla^2 s_k(\theta)|_{\theta=\theta_*}$ is positive semidefinite. For any $k \ge 1$, at generic $\theta_* \in \mathbb{R}^d$,*

$$\operatorname{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\le k}\big) = \operatorname{rank}\big(\nabla^2 s_1(\theta) + \cdots + \nabla^2 s_k(\theta)|_{\theta=\theta_*}\big), \tag{2.14}$$

and the span of $\{\nabla p(\theta_*) : p \in \mathcal{R}^{\mathsf{G}}_{\le k}\}$ is the column span of $\nabla^2 s_1(\theta) + \cdots + \nabla^2 s_k(\theta)|_{\theta=\theta_*}$.

(b) *In the projected orbit recovery model, the same holds for $\widetilde{\mathcal{R}}^{\mathsf{G}}_{\le k}$ and $\tilde{s}_k(\theta)$ as defined by (2.11).*

REMARK 2.9. We restrict attention to generic signals $\theta_* \in \mathbb{R}^d$ in this work. The specific condition for θ_* that we use in Theorem 2.7 and in Theorems 2.11 and 2.13 to follow is that the gradient vectors $\{\nabla p(\theta_*) : p \in \mathcal{R}^{\mathsf{G}}_{\le K}\}$ or $\{\nabla p(\theta_*) : p \in \widetilde{\mathcal{R}}^{\mathsf{G}}_{\le \widetilde{K}}\}$ span a subspace of dimension $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}})$ or $\operatorname{trdeg}(\widetilde{\mathcal{R}}^{\mathsf{G}})$, respectively.

Different behavior may be observed for nongeneric signals: For $\mathsf{G} = \{+\mathrm{Id}, -\mathrm{Id}\}$, which has been studied in [43, 44], the Fisher information $I(\theta_*)$ is singular at $\theta_* = 0$ (even though $d_0 = 0$, as the group is discrete). This leads to a $n^{-1/4}$ rate of estimation error near $\theta_* = 0$, instead of the $n^{-1/2}$ parametric rate. This $n^{-1/4}$ rate holds more generally for any discrete group G at signals θ_* whose orbit points are not pairwise distinct, which are precisely those signals where the Fisher information $I(\theta_*)$ is singular [12].

A different distinction between generic and nongeneric signals was highlighted in [33] when G is the group of cyclic rotations of coordinates in $\mathbb{R}^d$. There, orbits of generic signals are uniquely identified by moments up to the order $K = 3$, but identification of nongeneric signals having zero power in certain Fourier frequencies may require moments up to the order $d - 1$. For such nongeneric signals, we expect $I(\theta_*)$ to be nonsingular and the MLE to attain the parametric rate, but with asymptotic variance scaling as $\sigma^{2(d-1)}/n$ rather than $\sigma^{2K}/n = \sigma^6/n$. In a related model of continuous MRA, this asymptotic scaling is implied by the results of [5].

2.4. *Global likelihood landscape.* In this section, we establish correspondences between global and local minimizers of the population negative log-likelihood $R(\theta)$ with those of a sequence of moment optimization problems. These results are similar to results of [15], Sections 4.3 and 4.5, for discrete groups G, with a distinction that when G is continuous, these minimizers are not isolated points but rather manifolds of positive dimension.

We recall the following structural property for smooth nonconvex optimization landscapes, under which convergence to the global optimum from a random initialization is guaranteed for various descent-based optimization algorithms [17, 23, 28].

DEFINITION 2.10. The problem of minimizing a twice-continuously differentiable function $f : \mathcal{V} \to \mathbb{R}$ over a smooth manifold $\mathcal{V}$ is *globally benign* if each point $x \in \mathcal{V}$ where $\nabla f(x)|_{\mathcal{V}} = 0$ is either a global minimizer of f over $\mathcal{V}$, or has a direction of strictly negative curvature, $\lambda_{\min}(\nabla^2 f(x)|_{\mathcal{V}}) < 0$.

Here, $\nabla f(x)|_{\mathcal{V}}$ and $\nabla^2 f(x)|_{\mathcal{V}}$ denote the gradient and Hessian of f on $\mathcal{V}$, which may be taken in any choice of a smooth local chart around $x \in \mathcal{V}$.

Minimizing $R(\theta)$ in high noise may be viewed as successively solving a sequence of moment optimizations defined by the terms of its expansion in Theorem 2.2. To ease notation, let us collect the vectorized moment tensors up to order k as

$$M_k(\theta) = \operatorname{vec}\big(T_1(\theta), \dots, T_k(\theta)\big) \in \mathbb{R}^{d+d^2+\cdots+d^k}, \tag{2.15}$$

$$\widetilde{M}_k(\theta) = \operatorname{vec}\big(\widetilde{T}_1(\theta), \dots, \widetilde{T}_k(\theta)\big) \in \mathbb{R}^{\tilde{d}+\tilde{d}^2+\cdots+\tilde{d}^k}. \tag{2.16}$$

Fixing the true signal $\theta_* \in \mathbb{R}^d$, we define the moment varieties

$$\mathcal{V}_k(\theta_*) = \big\{\theta \in \mathbb{R}^d : M_k(\theta) = M_k(\theta_*)\big\}, \qquad \mathcal{V}_0(\theta_*) = \mathbb{R}^d, \tag{2.17}$$

$$\widetilde{\mathcal{V}}_k(\theta_*) = \big\{\theta \in \mathbb{R}^d : \widetilde{M}_k(\theta) = \widetilde{M}_k(\theta_*)\big\}, \qquad \widetilde{\mathcal{V}}_0(\theta_*) = \mathbb{R}^d. \tag{2.18}$$

These are the points $\theta \in \mathbb{R}^d$ for which the mixed moments of the Gaussian mixture model defined by θ match those of the true signal θ_* up to order k.

We state a general result on the optimization landscape, assuming that the Jacobian matrices $\mathrm{d}M_k$ and $\mathrm{d}\widetilde{M}_k$ have constant rank over $\mathcal{V}_k(\theta_*)$ and $\widetilde{\mathcal{V}}_k(\theta_*)$, so that $\mathcal{V}_k(\theta_*)$ and $\widetilde{\mathcal{V}}_k(\theta_*)$ are smooth manifolds. Then, recalling $s_k(\theta)$ and $\tilde{s}_k(\theta)$ from (2.8) and (2.11), we consider the optimization problem

$$\text{minimize } s_k(\theta) \text{ over } \theta \in \mathcal{V}_{k-1}(\theta_*) \tag{2.19}$$

in the unprojected setting, and

$$\text{minimize } \tilde{s}_k(\theta) \text{ over } \theta \in \widetilde{\mathcal{V}}_{k-1}(\theta_*) \tag{2.20}$$

in the projected setting. These are polynomial optimization problems in θ that are defined independently of the noise level σ^2. The following theorem guarantees that the landscape

of $R(\theta)$ is globally benign in high noise, as long as the landscape of each problem (2.19) or (2.20) is globally benign, and the final moment variety $\mathcal{V}_K(\theta_*)$ or $\widetilde{\mathcal{V}}_{\widetilde{K}}(\theta_*)$ contains only the points which globally minimizer $R(\theta)$. We illustrate part (a) of this result using a simple example of orthogonal Procrustes alignment at the conclusion of this section.

THEOREM 2.11. *For generic* $\theta_* \in \mathbb{R}^d$:

(a) *In the unprojected model, define* K *by Proposition* 2.6(*a*). *Suppose that* $\mathcal{V}_K(\theta_*) = \mathcal{O}_{\theta_*}$. *Suppose also that for each* $k = 1, \dots, K$, *the derivative matrix* $\mathrm{d}M_k(\theta)$ *has constant rank over* $\mathcal{V}_k(\theta_*)$, *and the minimization of* $s_k(\theta)$ *over* $\mathcal{V}_{k-1}(\theta_*)$ *is globally benign. Then for some* $\sigma_0 \equiv \sigma_0(\theta_*, \mathsf{G})$ *and any* $\sigma > \sigma_0$, *the minimization of* $R(\theta)$ *is also globally benign.*

(b) *In the projected model satisfying* (2.4), *define* $\widetilde{K}$ *by Proposition* 2.6(*b*). *Suppose that* $\widetilde{\mathcal{V}}_{\widetilde{K}}(\theta_*) = \{\theta : \Pi(\mathcal{O}_\theta) \equiv \Pi(\mathcal{O}_{\theta_*})\}$. *Suppose also that for each* $k = 1, \dots, \widetilde{K}$, *the derivative matrix* $\mathrm{d}\widetilde{M}_k(\theta)$ *has constant rank over* $\widetilde{\mathcal{V}}_k(\theta_*)$, *and the minimization of* $\tilde{s}_k(\theta)$ *over* $\widetilde{\mathcal{V}}_{k-1}(\theta_*)$ *is globally benign. Then for any constant* $B > 0$, *some* $\sigma_0 \equiv \sigma_0(\theta_*, \mathsf{G}, \Pi, B)$, *and any* $\sigma > \sigma_0$, *the minimization of* $R(\theta)$ *is globally benign over the domain* $\{\theta \in \mathbb{R}^d : \|\theta\| < B(\|\theta_*\| + \sigma)\}$.

In Theorem 2.11(b), we have restricted to a ball $\{\theta \in \mathbb{R}^d : \|\theta\| < B(\|\theta_*\| + \sigma)\}$, as the landscape of $R(\theta)$ outside this ball may depend on the specific interaction between G and Π. In practice, such a bound for $\|\theta\|$ may be known a priori, so that optimization may indeed be restricted to this ball. (In unprojected models, we show that $R(\theta)$ cannot have critical points outside this ball for any group G, allowing us to remove such a restriction in part (a).)

Whether the conditions of Theorem 2.11 hold depends on the specific model, and both positive and negative examples for discrete groups were exhibited in [15]. In models where they do not hold, $R(\theta)$ may in fact have spurious local minimizers in high noise, and Theorem 2.2 can be used to further establish a correspondence between the local minimizers of $R(\theta)$ and those of the above moment optimizations. We formalize one such result—not fully general, but sufficient to study many examples of interest—as follows.

DEFINITION 2.12. Suppose $\mathcal{V}_{K-1}(\theta_*)$ is a smooth manifold. A critical point θ of $s_K(\theta)|_{\mathcal{V}_{K-1}(\theta_*)}$ is *nondegenerate up to orbit* if $\mathcal{O}_\theta$ is a smooth manifold of dimension d_0 in a local neighborhood of θ, and

$$\operatorname{rank}\big(\nabla^2 s_K(\theta)|_{\mathcal{V}_{K-1}(\theta_*)}\big) = \dim\big(\mathcal{V}_{K-1}(\theta_*)\big) - d_0.$$

Note that $\nabla^2 s_K(\theta)|_{\mathcal{V}_{K-1}(\theta_*)}$ is a symmetric matrix of dimension $\dim(\mathcal{V}_{K-1}(\theta_*))$. For any critical point θ of $s_K|_{\mathcal{V}_{K-1}(\theta_*)}$, the null space of this Hessian must contain the tangent space to $\mathcal{O}_\theta$, and Definition 2.12 ensures that this Hessian has no further rank degeneracy.

THEOREM 2.13. *For generic* $\theta_* \in \mathbb{R}^d$:

(a) *In the unprojected model, suppose that* $\mathrm{d}M_k(\theta)$ *has constant rank over* $\mathcal{V}_k(\theta_*)$, *and the minimization of* $s_k(\theta)$ *over* $\mathcal{V}_{k-1}(\theta_*)$ *is globally benign for each* $k = 1, \dots, K-1$. *Then for some* (θ_*, G)*-dependent constant* $\sigma_0 > 0$ *and function* $\varepsilon(\sigma)$ *satisfying* $\varepsilon(\sigma) \to 0$ *as* $\sigma \to \infty$, *and for all* $\sigma > \sigma_0$:

1. *Let* θ_+ *be any* local *minimizer of* $s_K(\theta)$ *over* $\mathcal{V}_{K-1}(\theta_*)$ *that is nondegenerate up to orbit. Then there exists a local minimizer* θ'_+ *of* $R(\theta)$ *where* $\|\theta_+ - \theta'_+\| < \varepsilon(\sigma)$.

2. *Conversely, suppose that all critical points of* $s_K(\theta)$ *over* $\mathcal{V}_{K-1}(\theta_*)$ *are nondegenerate up to orbit. Let* θ_+ *be any local minimizer of* $R(\theta)$. *Then there exists a local minimizer* θ'_+ *of* $s_K(\theta)$ *over* $\mathcal{V}_{K-1}(\theta_*)$ *where* $\|\theta_+ - \theta'_+\| < \varepsilon(\sigma)$.

(b) *In the projected model satisfying* (2.4), *statement* (1.) *holds with* K, M_k, $\mathcal{V}_k$ *and* s_k *replaced by* $\widetilde{K}$, $\widetilde{M}_k$, $\widetilde{\mathcal{V}}_k$ *and* $\tilde{s}_k$, *where* σ_0 *and* $\varepsilon(\sigma)$ *may depend also on the projection* Π. *Statement* (2) *holds for local minimizers* θ_+ *of* $R(\theta)$ *satisfying* $\|\theta_+\| < B(\|\theta_*\| + \sigma)$ *for any constant* $B > 0$, *where* σ_0 *and* $\varepsilon(\sigma)$ *may depend also on* Π *and* B.

The guarantees of Theorems 2.11 and 2.13 may be translated to the sample log-likelihood $R_n(\theta)$ by establishing concentration of $\nabla R_n(\theta)$ and $\nabla^2 R_n(\theta)$ around $\nabla R(\theta)$ and $\nabla^2 R(\theta)$ [31]. For orbit recovery models in the high-noise regime, we believe that it may be possible to obtain sharp concentration bounds by deriving a series expansion also of the *empirical* log-likelihood function $R_n(\theta)$ in powers of σ^{-1}, and analyzing the concentration term-by-term. Some results of this form were obtained for models without linear projection in [15], Lemma 4.11 and Corollary 4.18, and we leave the analysis of the empirical log-likelihood function and landscape for more general models as an open problem for future work.

EXAMPLE 2.14 (Landscape of orthogonal Procrustes alignment). We illustrate Theorems 2.11 and 2.13 using a simple example of *orthogonal Procrustes alignment* [19, 20, 34].

In this problem, samples of an object consisting of $m \geq 3$ atoms in $\mathbb{R}^3$ are observed under random orthogonal rotations and reflections. We represent the object as $\theta_* \in \mathbb{R}^{3\times m} \cong \mathbb{R}^d$ where $d = 3m$. The rotational group is $\mathsf{G} = \mathsf{O}(3) \otimes \mathrm{Id}_m \subset \mathsf{O}(d)$, where a common orthogonal matrix in three dimensions is applied to all m atoms. Assuming the generic condition that $\mathrm{rank}(\theta_*) = 3$, that is, these m atoms do not lie on a common 2-dimensional subspace, we study the likelihood landscape for estimating θ_* from many independently rotated samples.

In this model, we check in Appendix B that $K = 2$, $(d_0, d_1, d_2) = (3, 0, d-3)$, $\mathcal{V}_1(\theta_*) = \mathbb{R}^d$, and $\mathcal{V}_2(\theta_*) = \{g \cdot \theta_* : g \in \mathsf{G}\} = \mathcal{O}_{\theta_*}$. The first two moment tensors $T_1(\theta)$ and $T_2(\theta)$ are given by $T_1(\theta) = 0$ and $T_2(\theta) = \frac{1}{3}\,\mathrm{Id}_{3\times 3} \otimes (\theta^\top \theta) \in \mathbb{R}^{d\times d}$, and the terms $s_1(\theta)$ and $s_2(\theta)$ in (2.7) are given by $s_1(\theta) = 0$ and

$$s_2(\theta) = \frac{1}{12}\|\theta^\top \theta - \theta_*^\top \theta_*\|_{\mathrm{HS}}^2,$$

where $\theta^\top\theta, \theta_*^\top\theta_* \in \mathbb{R}^{m\times m}$. The minimization of $s_1(\theta)$ over $\mathcal{V}_0(\theta_*) = \mathbb{R}^d$ is trivially globally benign. We show in Appendix B that $\mathrm{d}M_2(\theta)$ has constant rank over $\mathcal{V}_2(\theta_*)$, and that the minimization of $s_2(\theta)$ over $\mathcal{V}_1(\theta_*) = \mathbb{R}^d$ is also globally benign, with minimizers given exactly by $\mathcal{V}_2(\theta_*) = \mathcal{O}_{\theta_*}$. Thus, Theorem 2.11(a) implies that the landscape of $R(\theta)$ is also globally benign for sufficiently high noise, and the only local minimizers of $R(\theta)$ are rotations and reflections of the true object.

A variation of this problem is the rotation-only variant, where we observe 3-dimensional rotations (but not reflections) of the object. Then the rotational group is instead $\mathsf{G} = \mathsf{SO}(3) \otimes \mathrm{Id}_m \subset \mathsf{O}(d)$. We show in Appendix B that still $K = 2$, $(d_0, d_1, d_2) = (3, 0, d-3)$, and the forms of $T_1(\theta)$, $T_2(\theta)$, $\mathcal{V}_1(\theta_*)$, $\mathcal{V}_2(\theta_*)$, $s_1(\theta)$, $s_2(\theta)$ are identical to the above (even though the full log-likelihood $R(\theta)$ is not). Thus the minimization of $s_2(\theta)$ over $\mathcal{V}_1(\theta_*) = \mathbb{R}^d$ is still globally benign, with minimizers $\mathcal{V}_2(\theta_*)$. However, this set of minimizers is now written as

$$\mathcal{V}_2(\theta_*) = \{g \cdot \theta_* : g \in \mathsf{G}\} \cup \{-g \cdot \theta_* : g \in \mathsf{G}\} = \mathcal{O}_{\theta_*} \cup \mathcal{O}_{-\theta_*}$$

constituting two distinct orbits under this more restrictive group action. The first orbit $\mathcal{O}_{\theta_*}$ are the global minimizers of $R(\theta)$. The second orbit corresponds to the mirror reflection $-\theta_*$, which does not globally minimize $R(\theta)$, but the difference between $R(\theta_*)$ and $R(-\theta_*)$ lies in the remainder term of the expansion (2.7). Theorem 2.13(a) shows that for high noise, $R(\theta)$ will have spurious local minimizers near (but not exactly equal to) this second orbit $\mathcal{O}_{-\theta_*}$.

3. Continuous multireference alignment. We now specialize the preceding general results to the problem of estimating a periodic function on the circle, observed under $\mathsf{SO}(2)$ rotations of its domain. We will refer to this as the *continuous MRA* model. This provides a simpler 1-dimensional analogue of the 2-dimensional and 3-dimensional problems that we will discuss in Section 4.

To describe the model, let $f : \mathcal{S}^1 \to \mathbb{R}$ be a periodic function on the unit circle $\mathcal{S}^1 \cong [0, 1)$. We identify the rotational group $\mathsf{SO}(2)$ also with $[0, 1)$, and represent the rotation of f by an element $\mathfrak{g} \in \mathsf{SO}(2) \cong [0, 1)$ as $f_{\mathfrak{g}}(t) = f(t + \mathfrak{g} \bmod 1)$. Each sample is an observation of the rotated function $f_{\mathfrak{g}}$ with additive white noise,

$$f_{\mathfrak{g}}(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t),$$

where $\mathfrak{g} \sim \mathrm{Unif}([0, 1))$ and $\mathrm{d}W(t)$ denotes a standard Gaussian white noise process on $\mathcal{S}^1$. This may be understood as observing a realization of the Gaussian process $\{F(h)\}_{h\in L_2(\mathcal{S}^1)} = \{\int h(t)[f_{\mathfrak{g}}(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)]\}_{h\in L_2(\mathcal{S}^1)}$ with mean and covariance functions

$$\mathbb{E}\big[F(h)\big] = \int_0^1 h(t) f_{\mathfrak{g}}(t)\,\mathrm{d}t, \qquad \mathrm{Cov}\big[F(h_1), F(h_2)\big] = \sigma^2 \int_0^1 h_1(t)h_2(t)\,\mathrm{d}t, \tag{3.1}$$

or equivalently as observing all coefficients of $f_{\mathfrak{g}}$ in a complete orthonormal basis $\{h_j(t)\}_{j=1}^{\infty}$ of $L_2(\mathcal{S}^1)$ with independent $\mathcal{N}(0, \sigma^2)$ noise for each basis coefficient.

We consider the real Fourier basis on $\mathcal{S}^1$, given by

$$h_0(t) = 1, \qquad h_{l1}(t) = \sqrt{2}\cos 2\pi l t, \qquad h_{l2}(t) = \sqrt{2}\sin 2\pi l t \quad \text{for } l = 1, 2, 3, \ldots. \tag{3.2}$$

We then restrict our model to the finite-dimensional space of functions $f : \mathcal{S}^1 \to \mathbb{R}$ that have finite bandlimit $L \geq 1$ in this basis, that is, f admits a representation

$$f(t) = \theta^{(0)} h_0(t) + \sum_{l=1}^{L} \theta_1^{(l)} h_{l1}(t) + \sum_{l=1}^{L} \theta_2^{(l)} h_{l2}(t). \tag{3.3}$$

Importantly, the space of such bandlimited functions is closed under rotations of $\mathcal{S}^1$. Writing

$$\theta = \big(\theta^{(0)}, \theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_1^{(L)}, \theta_2^{(L)}\big) \in \mathbb{R}^d, \quad d = 2L + 1$$

for the vector of Fourier coefficients, the rotation $f \mapsto f_{\mathfrak{g}}$ corresponds to $\theta \mapsto g \cdot \theta$, where g belongs to the block-diagonal representation

$$\mathsf{G} = \left\{ \mathrm{diag}\left(1, \begin{pmatrix} \cos 2\pi\mathfrak{g} & \sin 2\pi\mathfrak{g} \\ -\sin 2\pi\mathfrak{g} & \cos 2\pi\mathfrak{g} \end{pmatrix}, \ldots, \begin{pmatrix} \cos 2\pi L\mathfrak{g} & \sin 2\pi L\mathfrak{g} \\ -\sin 2\pi L\mathfrak{g} & \cos 2\pi L\mathfrak{g} \end{pmatrix}\right) : \mathfrak{g} \in [0, 1) \right\} \tag{3.4}$$

of $\mathsf{SO}(2)$. The observation model for the Fourier coefficients of f then takes the form of (2.1), where we observe coordinates of $g \cdot \theta$ with entrywise i.i.d. $\mathcal{N}(0, \sigma^2)$ noise.

Theorem 3.1 below first characterizes, for this model, the decomposition of total dimension described in Theorem 2.7. As a direct consequence of this result, we state Corollary 3.2 which summarizes the implications for identifying θ_* based on its low-order moments, and for the spectral structure of the Fisher information matrix $I(\theta_*)$.

THEOREM 3.1. *For any $L \geq 1$, we have*

$$\mathrm{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\leq 1}\big) = 1, \qquad \mathrm{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\leq 2}\big) = L + 1, \qquad \mathrm{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\leq 3}\big) = \mathrm{trdeg}\big(\mathcal{R}^{\mathsf{G}}\big) = 2L = d - 1.$$

COROLLARY 3.2. *A generic signal $\theta_* \in \mathbb{R}^d$ in this continuous MRA model has the following properties*:

(a) *θ_* is identified up to a finite list of orbits by the moments of $g \cdot \theta_*$ up to order $K = 3$ when $L \geq 2$, and order $K = 2$ when $L = 1$.*

(b) *For (θ_*, G)-dependent constants $C, c > 0$ independent of σ, the Fisher information $I(\theta_*)$ has $d_0 = 1$ eigenvalue of 0 and d_k eigenvalues in $[c\sigma^{-2k}, C\sigma^{-2k}]$ for $k = 1, 2, 3$ and $(d_1, d_2, d_3) = (1, L, L-1)$.*

Part (a) of this corollary follows immediately from [4], Theorem 3.15, (which we review in Appendix A.2), and part (b) follows from Theorem 2.7.

In Appendix C.1, we provide a proof of Theorem 3.1 using our general result of Lemma 2.8, as a warm-up for our analyses of the $\mathsf{SO}(3)$-rotational models to follow. We note that for a similar observation model of continuous MRA studied in [5], a stronger form of Corollary 3.2(a) is already known, namely that third-order moments are sufficient to identify generic signals $\theta_* \in \mathbb{R}^d$ up to a single unique orbit.

Next, we study the moment optimization problems of (2.19), and we describe more explicit forms for these optimization problems in this continuous MRA example. Denote the Fourier coefficients of the true function f by $\theta_* \in \mathbb{R}^d$. Define the complex Fourier coefficients

$$u^{(0)}(\theta) = \theta^{(0)} \in \mathbb{R}, \qquad u^{(l)}(\theta) = \theta_1^{(l)} + \mathbf{i}\theta_2^{(l)} = r_l(\theta)e^{\mathbf{i}\lambda_l(\theta)} \in \mathbb{C},$$

where $(r_l(\theta), \lambda_l(\theta))$ for $l \geq 1$ are the magnitude and phase of $u^{(l)}(\theta)$. Write as shorthand

$$r_{l,l',l''}(\theta) = r_l(\theta)r_{l'}(\theta)r_{l''}(\theta), \qquad \lambda_{l,l',l''}(\theta) = \lambda_l(\theta) - \lambda_{l'}(\theta) - \lambda_{l''}(\theta).$$

Here, $\lambda_{l,l',l''}(\theta)$ are the elements of the Fourier bispectrum of θ.

THEOREM 3.3. *For any $L \geq 1$,*

$$\begin{aligned}
s_1(\theta) &= \frac{1}{2}\big(\theta^{(0)} - \theta_*^{(0)}\big)^2,\\
s_2(\theta) &= \frac{1}{4}\big(\big(\theta^{(0)}\big)^2 - \big(\theta_*^{(0)}\big)^2\big)^2 + \frac{1}{8}\sum_{l=1}^{L}\big(r_l(\theta)^2 - r_l(\theta_*)^2\big)^2,\\
s_3(\theta) &= \frac{1}{48}\big(\big(u^{(0)}(\theta)\big)^3 - \big(u^{(0)}(\theta_*)\big)^3\big)^2\\
&\quad + \frac{1}{16}\sum_{\substack{l,l',l''=0\\ l=l'+l''}}^{L}\big|u^{(l)}(\theta)\overline{u^{(l')}(\theta)u^{(l'')}(\theta)} - u^{(l)}(\theta_*)\overline{u^{(l')}(\theta_*)u^{(l'')}(\theta_*)}\big|^2\\
&= \frac{1}{12}\big(\big(\theta^{(0)}\big)^3 - \big(\theta_*^{(0)}\big)^3\big)^2 + \frac{1}{8}\sum_{l=1}^{L}\big(\theta^{(0)} \cdot r_l(\theta)^2 - \theta_*^{(0)} \cdot r_l(\theta_*)^2\big)^2\\
&\quad + \frac{1}{16}\sum_{\substack{l,l',l''=1\\ l=l'+l''}}^{L}\big(r_{l,l',l''}(\theta)^2 + r_{l,l',l''}(\theta_*)^2\\
&\quad - 2r_{l,l',l''}(\theta)r_{l,l',l''}(\theta_*)\cos\big(\lambda_{l,l',l''}(\theta_*) - \lambda_{l,l',l''}(\theta)\big)\big).
\end{aligned}$$

Since each moment variety $\mathcal{V}_k(\theta_*)$ in (2.17) is precisely the set of points $\{\theta \in \mathbb{R}^d : s_1(\theta) = 0, \ldots, s_k(\theta) = 0\}$, this implies also that

$$\begin{aligned}
&\mathcal{V}_0(\theta_*) = \mathbb{R}^d, \qquad \mathcal{V}_1(\theta_*) = \big\{\theta : \theta^{(0)} = \theta_*^{(0)}\big\},\\
&\mathcal{V}_2(\theta_*) = \big\{\theta : \theta^{(0)} = \theta_*^{(0)} \text{ and } r_l(\theta) = r_l(\theta_*) \text{ for each } l = 1, \ldots, L\big\}.
\end{aligned}$$

Thus the minimization of $s_1(\theta)$ on $\mathcal{V}_0(\theta_*)$ is over the global function mean $\theta^{(0)}$, the minimization of $s_2(\theta)$ on $\mathcal{V}_1(\theta_*)$ is over the Fourier power spectrum $\{r_l(\theta) : l = 1, \dots, L\}$ and the minimization of $s_3(\theta)$ on $\mathcal{V}_2(\theta_*)$ is over the Fourier bispectrum $\{\lambda_{l,l',l''}(\theta) : l = l' + l''\}$.

In high noise, minimizing the population log-likelihood function $R(\theta)$ becomes similar to successively minimizing $s_1(\theta)$, $s_2(\theta)$ and $s_3(\theta)$. The following result describes the nature of these three optimization landscapes.

THEOREM 3.4. *For any $L \geq 1$ and generic $\theta_* \in \mathbb{R}^d$, the minimizations of $s_1(\theta)$ over $\mathcal{V}_0(\theta_*)$ and of $s_2(\theta)$ over $\mathcal{V}_1(\theta_*)$ are globally benign. However, for any $L \geq 30$, there exists a nonempty open subset $U \subset \mathbb{R}^d$ such that for any $\theta_* \in U$, the minimization of $s_3(\theta)$ over $\mathcal{V}_2(\theta_*)$ has a local minimizer outside $\mathcal{O}_{\theta_*}$ that is nondegenerate up to orbit.*

The correspondence between optimization landscapes shown in Theorem 2.13(a) then implies that, for the class of signals $\theta_* \in U$ described in Theorem 3.4 and in sufficiently high noise, the landscape of the population negative log-likelihood function $R(\theta)$ must also have spurious local minimizers near those of $s_3(\theta)$. The particular local minimizers of $s_3(\theta)$ that we exhibit in the proof of Theorem 3.4 correspond to certain Fourier phase shifts of the true signal. This example is somewhat analogous to the spurious local minimizers discovered in dimensions $d \geq 53$ for the log-likelihood landscape of discrete MRA in [15], Section 4.6.

We conjecture, based on the algebraic similarities between these models, that spurious local minimizers of $R(\theta)$ may also exist for generic $\theta_* \in \mathbb{R}^d$ in the $\mathsf{SO}(3)$-rotational models to be discussed in Section 4, and we leave this as an open question.

4. Spherical registration and cryo-EM. We now describe examples of estimating a function in 2 or 3 dimensions, observed under $\mathsf{SO}(3)$ rotations of its domain. Section 4.1 studies estimation on the sphere, Section 4.2 studies estimation in $\mathbb{R}^3$ and Section 4.3 studies a simplified "cryo-EM model" of estimation in $\mathbb{R}^3$ with a tomographic projection onto a 2-dimensional plane.

4.1. *Spherical registration.* Let $\mathcal{S}^2 \subset \mathbb{R}^3$ be the unit sphere, and let $f : \mathcal{S}^2 \to \mathbb{R}$ be a function on this sphere. We parametrize $\mathcal{S}^2$ by the latitude $\phi_1 \in [0, \pi]$ and longitude $\phi_2 \in [0, 2\pi)$. Writing $f_{\mathfrak{g}}(\phi_1, \phi_2) = f(\mathfrak{g}^{-1} \cdot (\phi_1, \phi_2))$ for the rotation of the function f, we consider the observation model with samples

$$f_{\mathfrak{g}}(\phi_1, \phi_2)\, \mathrm{d}(\phi_1, \phi_2) + \sigma\, \mathrm{d}W(\phi_1, \phi_2),$$

where $\mathfrak{g} \in \mathsf{SO}(3)$ is a uniform random rotation for each sample, $\mathrm{d}(\phi_1, \phi_2) = \sin\phi_1\, \mathrm{d}\phi_1\, \mathrm{d}\phi_2$ denotes the surface area measure on $\mathcal{S}^2$, and $\mathrm{d}W(\phi_1, \phi_2)$ is a standard Gaussian white noise process on $\mathcal{S}^2$. This observation model may be understood as observing a realization of the Gaussian process $\{\int h(\phi_1, \phi_2)[f_{\mathfrak{g}}(\phi_1, \phi_2)\, \mathrm{d}(\phi_1, \phi_2) + \sigma\, \mathrm{d}W(\phi_1, \phi_2)]\}_{h \in L_2(\mathcal{S}^2)}$ defined analogously to (3.1), or equivalently, as observing each coefficient of $f_{\mathfrak{g}}$ in an orthonormal basis of $L_2(\mathcal{S}^2)$ with i.i.d. $\mathcal{N}(0, \sigma^2)$ noise.

We choose as our orthonormal basis the real spherical harmonics

$$h_{lm}(\phi_1, \phi_2) \quad \text{for } l = 0, 1, 2, \dots \text{ and } m = -l, -l+1, \dots, l-1, l.$$

We assume that $f : \mathcal{S}^2 \to \mathbb{R}$ has a finite bandlimit $L \geq 1$ in this basis, that is, it takes the form

$$f(\phi_1, \phi_2) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \theta_m^{(l)} h_{lm}(\phi_1, \phi_2). \tag{4.1}$$

We may then represent f by its vector of real spherical harmonic coefficients

$$\theta = \big(\theta_m^{(l)} : l = 0, \dots, L \text{ and } m = -l, \dots, l\big) \in \mathbb{R}^d, \qquad d = (L+1)^2.$$

This subspace of bandlimited functions is closed under $\mathsf{SO}(3)$-rotations of $\mathcal{S}^2$, and we review the forms of h_{lm} and of the rotational action on the basis coefficients in Appendix D.2.

The following result describes the decomposition of total dimension in Theorem 2.7(a) for bandlimits $L \geq 10$.

THEOREM 4.1. *For any $L \geq 10$, we have*

$$\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\leq 1}) = 1, \qquad \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\leq 2}) = L+1, \qquad \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\leq 3}) = \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d-3.$$

COROLLARY 4.2. *A generic signal $\theta_* \in \mathbb{R}^d$ in this spherical registration model for $L \geq 10$ has the following properties*:

(a) *θ_* may be identified up to a finite list of orbits by the moments of $g \cdot \theta_*$ up to order $K = 3$.*

(b) *For (θ_*, G)-dependent constants $C, c > 0$ independent of σ, the Fisher information $I(\theta_*)$ has $d_0 = 3$ eigenvalues of 0 and d_k eigenvalues in $[c\sigma^{-2k}, C\sigma^{-2k}]$ for $k = 1, 2, 3$ and $(d_1, d_2, d_3) = (1, L, L(L+1) - 3)$.*

REMARK 4.3. The result of Theorem 4.1 was conjectured for all bandlimits $L \geq 10$ in [4], Conjecture 4.11, and it was verified numerically in exact-precision arithmetic for $L \in \{10, \ldots, 16\}$. Our result resolves this conjecture for all $L \geq 10$. Conversely, for low bandlimits $L \leq 9$, it was shown in [4], Section 4.5, that $K > 3$ strictly, meaning that moments up to third order are insufficient to locally identify θ_* up to its orbit.

Turning to the forms of $s_k(\theta)$ in (2.19), let us denote the real spherical harmonic coefficients of the true function by $\theta_* \in \mathbb{R}^d$. We write as shorthand

$$u^{(l)}(\theta) = \big(u^{(l)}_m(\theta) : m = -l, \ldots, l\big) \in \mathbb{C}^{2l+1}$$

for the complex spherical harmonic coefficients at frequency l, which are defined from the real coefficients $(\theta^{(l)}_m : m = -l, \ldots, l)$ by a unitary transform described in (D.S11). We denote

$$B_{l,l',l''}(\theta) = \sum_{m=-l}^{l} \sum_{\substack{m'=-l' \\ m''=m+m'}}^{l'} \sum_{m''=-l''}^{l''} \langle l, m; l', m' | l'', m'' \rangle \overline{u^{(l)}_m(\theta) u^{(l')}_{m'}(\theta)} u^{(l'')}_{m''}(\theta), \tag{4.2}$$

where $\langle l, m; l', m' | l'', m'' \rangle \in \mathbb{R}$ is the Clebsch–Gordan coefficient. These quantities express the integrals of three-fold products of spherical harmonics over $\mathcal{S}^2$ and arise naturally in the computation of third-order moments of $g \cdot \theta$. We review their definition in Appendix D.1. The functions $B_{l,l',l''}(\theta)$ are analogous to the scaled components $r_{l,l',l''}(\theta)\lambda_{l,l',l''}(\theta)$ of the Fourier bispectrum that appeared in the 1-dimensional MRA example of Section 3. The minimizations of $s_1(\theta)$, $s_2(\theta)$ and $s_3(\theta)$ described in Theorem 2.13 may then be analogously understood as minimizing the global function mean, the power in each spherical harmonic frequency, and certain "bispectrum" variables for each frequency.

THEOREM 4.4. *For any $L \geq 1$,*

$$s_1(\theta) = \frac{1}{2}\big(u^{(0)}(\theta) - u^{(0)}(\theta_*)\big)^2,$$

$$s_2(\theta) = \frac{1}{4} \sum_{l=0}^{L} \frac{1}{2l+1} \big(\|u^{(l)}(\theta)\|^2 - \|u^{(l)}(\theta_*)\|^2\big)^2,$$

$$s_3(\theta) = \frac{1}{6} \sum_{\substack{l,l',l''=0 \\ |l-l'| \leq l'' \leq l+l'}}^{L} \frac{1}{2l''+1} \big|B_{l,l',l''}(\theta) - B_{l,l',l''}(\theta_*)\big|^2.$$

We prove Theorems 4.1 and 4.4 in Appendix D.2. Here, let us describe the high-level proof idea for Theorem 4.1, which is used also in our analyses of the cryo-EM models to follow. By Lemma 2.8, it suffices to analyze the ranks of the Hessians $\nabla^2 s_1(\theta_*)$, $\nabla^2 s_2(\theta_*)$ and $\nabla^2 s_3(\theta_*)$ at a generic point $\theta_* \in \mathbb{R}^d$. This analysis is straightforward for s_1, s_2 and the core of the proof is to show that $\nabla^2 s_3(\theta_*)$ has full rank $d-3$ (which accounts for the 3-dimensional orbit of θ_*) when $L \geq 10$.

Importantly, for any matrix $M(\theta)$ that is analytic in θ, we have $\operatorname{rank}(M(\theta)) < k$ if and only if every $k \times k$ submatrix of $M(\theta)$ has determinant 0. Because the $k \times k$ minors are themselves analytic in θ, this holds either for all $\theta \in \mathbb{R}^d$, or only for θ outside a generic subset of $\mathbb{R}^d$. So, if $\operatorname{rank}(M(\theta)) \geq k$ at some point $\theta \in \mathbb{R}^d$, then it must hold that $\operatorname{rank}(M(\theta)) \geq k$ at all generic $\theta \in \mathbb{R}^d$, and we record this for reference as the following fact.

FACT 4.5. For any $k \geq 1$ and matrix $M(\theta)$ whose entries are analytic in θ, we have $\operatorname{rank} M(\theta_*) \geq k$ for generic points $\theta_* \in \mathbb{R}^d$ if and only if there exists at least one point $\theta_* \in \mathbb{R}^d$ for which this inequality holds.

Thus, to show that $\operatorname{rank}(\nabla^2 s_3(\theta_*)) \geq d-3$ for generic $\theta_* \in \mathbb{R}^d$, it suffices to construct a single point $\theta_* \in \mathbb{R}^d$ where this holds. We do this by analyzing the explicit form of $\nabla^2 s_3(\theta_*)$ derived from Theorem 4.4. For $L = 10$, we exhibit such a point θ_* numerically. We then use this as a base case to inductively construct θ_* for all $L \geq 10$, by carefully choosing certain coordinates of θ_* to be 0 so that $\nabla^2 s_3(\theta_*)$ has a sparse structure and its rank may be explicitly deduced from the ranks of 2×2 submatrices.

4.2. *Unprojected cryo-EM.* Consider now a function $f : \mathbb{R}^3 \to \mathbb{R}$, and the action of $\mathsf{SO}(3)$ on $\mathbb{R}^3$ given by rotation about the origin. Write $f_{\mathfrak{g}}(x) = f(\mathfrak{g}^{-1} \cdot x)$ for the rotated function. We consider the observation model with samples

$$f_{\mathfrak{g}}(x)\,\mathrm{d}x + \sigma\,\mathrm{d}W(x),$$

where $\mathfrak{g} \in \mathsf{SO}(3)$ is uniformly random for each sample, and $\mathrm{d}W(x)$ is a standard Gaussian white noise process on $\mathbb{R}^3$. This is an unprojected model of the single-particle reconstruction problem in cryo-EM, to which we will add a tomographic projection in the next section. This model may be of independent interest for applications to cryo-ET, described in Appendix F.1.

We model f using a basis representation for its Fourier transform $\hat{f} : \mathbb{R}^3 \to \mathbb{C}$, similar to the approach of [4], Section 4.6. We parametrize the Fourier domain $\mathbb{R}^3$ by spherical coordinates (ρ, ϕ_1, ϕ_2) with radius $\rho \geq 0$, latitude $\phi_1 \in [0, \pi]$ and longitude $\phi_2 \in [0, 2\pi)$ and decompose $\hat{f}(\rho, \phi_1, \phi_2)$ in a complex basis $\{\hat{j}_{lsm}\}$ given by the product of the complex spherical harmonics $y_{lm}(\phi_1, \phi_2)$ (reviewed in Appendix D.1) with radial functions $z_s(\rho)$:

$$\hat{j}_{lsm}(\rho, \phi_1, \phi_2) = z_s(\rho) y_{lm}(\phi_1, \phi_2) \quad \text{for } s \geq 1, l \geq 0, m \in \{-l, \ldots, l\}. \tag{4.3}$$

Here, $\{z_s : s \geq 1\}$ may be any system of radial basis functions $z_s : [0, \infty) \to \mathbb{R}$ satisfying the orthogonality relation

$$\int_0^\infty \rho^2 z_s(\rho) z_{s'}(\rho)\, d\rho = \mathbf{1}\{s = s'\}, \tag{4.4}$$

so that $\{\hat{j}_{lsm}\}$ are orthonormal over $L_2(\mathbb{R}^3, \mathbb{C})$. The inverse Fourier transforms $\{j_{lsm}\}$ of $\{\hat{j}_{lsm}\}$ then provide a complex orthonormal basis in the original signal domain of f.

Fixing integer bandlimits $L \geq 1$ and $S_0, \ldots, S_L \geq 1$, we define the index set

$$\mathcal{I} = \{(l, s, m) : 0 \leq l \leq L, 1 \leq s \leq S_l, -l \leq m \leq l\}, \qquad d = |\mathcal{I}| = \sum_{l=0}^{L} (2l+1) S_l \tag{4.5}$$

and assume that f is $(L, S_0, \dots, S_L)$-*bandlimited* in the sense of admitting the finite basis representation

$$f = \sum_{(l,s,m)\in\mathcal{I}} u_m^{(ls)} \cdot j_{lsm}, \quad u = \big(u_m^{(ls)} : (l,s,m) \in \mathcal{I}\big) \in \mathbb{C}^d. \tag{4.6}$$

This corresponds to modeling the Fourier transform $\hat{f}$ up to the spherical frequency L, and up to the radial frequency S_l for each spherical component $l = 0, 1, \dots, L$. For real-valued functions f, writing $u = \hat{V}^*\theta$ for a unitary transform $\hat{V} \in \mathbb{C}^{d\times d}$ defined explicitly in (D.S31), we then obtain a real sequence representation

$$f = \sum_{(l,s,m)\in\mathcal{I}} \theta_m^{(ls)} \cdot h_{lsm}, \quad \theta = \big(\theta_m^{(ls)} : (l,s,m) \in \mathcal{I}\big) \in \mathbb{R}^d \tag{4.7}$$

for a real-valued orthonormal basis $\{h_{lsm}\}$. We describe the forms of h_{lsm} and the rotational action on the basis coefficients $\theta \in \mathbb{R}^d$ in Appendix D.3.

The following result describes the decomposition of total dimension in Theorem 2.7(a), assuming $L \geq 1$ and $S_l \geq 2$ for each $l = 0, \dots, L$. (Note that the case of $S_0 = \cdots = S_L = 1$ would be similar to the spherical registration example of Section 4.1, and a lower bound of $L \geq 10$ would be needed in this case to ensure $K = 3$.)

THEOREM 4.6. *For any $L \geq 1$ and $S_0, \dots, S_L \geq 2$, we have*

$$\operatorname{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\leq 1}\big) = S_0,$$

$$\operatorname{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\leq 2}\big) = \sum_{l=0}^{L} d(S_l), \quad d(S_l) \equiv \begin{cases} \dfrac{S_l(S_l+1)}{2} & \text{for } S_l < 2l+1, \\ (2l+1)(S_l - l) & \text{for } S_l \geq 2l+1, \end{cases}$$

$$\operatorname{trdeg}\big(\mathcal{R}^{\mathsf{G}}_{\leq 3}\big) = \operatorname{trdeg}\big(\mathcal{R}^{\mathsf{G}}\big) = d - 3.$$

COROLLARY 4.7. *In this unprojected cryo-EM model with $S_0, \dots, S_L \geq 2$, a generic signal $\theta_* \in \mathbb{R}^d$ may be identified up to a finite list of orbits by the moments of $g \cdot \theta_*$ up to order $K = 3$ if $L \geq 2$, and up to order $K = 2$ if $L = 1$.*

Turning to the forms of $s_k(\theta)$ that define the moment optimization (2.19), write $\theta_* \in \mathbb{R}^d$ for the true coefficients in the above real basis $\{h_{lsm}\}$. Let

$$u^{(ls)}(\theta) = \big(u_m^{(ls)}(\theta) : m = -l, \dots, l\big) \in \mathbb{C}^{2l+1} \tag{4.8}$$

be the components of the *complex* coefficients $u = \hat{V}^*\theta$ for the frequency pair (l, s), and define analogously to (4.2)

$$B_{(l,s),(l',s'),(l'',s'')}(\theta) = \sum_{m=-l}^{l} \sum_{\substack{m'=-l' \\ m''=m+m'}}^{l'} \sum_{m''=-l''}^{l''} \langle l, m; l', m' | l'', m'' \rangle \overline{u_m^{(ls)}(\theta) u_{m'}^{(l's')}(\theta)} u_{m''}^{(l''s'')}(\theta). \tag{4.9}$$

When the original function $f : \mathbb{R}^3 \to \mathbb{R}$ is real valued, we verify in the proof of Theorem 4.8 below that each $B_{(l,s),(l',s'),(l'',s'')}(\theta)$ is also real valued.

THEOREM 4.8. *For any $L \geq 1$ and $S_0, \dots, S_L \geq 1$,*

$$s_1(\theta) = \frac{1}{2} \sum_{s=1}^{S_0} \big(u^{(0s)}(\theta) - u^{(0s)}(\theta_*)\big)^2,$$

$$s_2(\theta) = \frac{1}{4}\sum_{l=0}^{L}\frac{1}{2l+1}\sum_{s,s'=1}^{S_l}\big(\langle u^{(ls)}(\theta), u^{(ls')}(\theta)\rangle - \langle u^{(ls)}(\theta_*), u^{(ls')}(\theta_*)\rangle\big)^2,$$

$$s_3(\theta) = \frac{1}{12}\sum_{\substack{l,l',l''=0\\ |l-l'|\le l''\le l+l'}}^{L}\frac{1}{2l''+1}\sum_{s=1}^{S_l}\sum_{s'=1}^{S_{l'}}\sum_{s''=1}^{S_{l''}}\big(B_{(l,s),(l',s'),(l'',s'')}(\theta) - B_{(l,s),(l',s'),(l'',s'')}(\theta_*)\big)^2.$$

In this model, the optimization of $s_1(\theta)$ is over the mean component $u^{(0s)}(\theta)$ corresponding to each radial frequency s. The optimization of $s_2(\theta)$ is over not just the power $\|u^{(ls)}(\theta)\|^2$ within each frequency pair (l, s), but also the cross-correlations between $u^{(ls)}$ and $u^{(ls')}$ for different radial frequencies s and s'.

The proofs of Theorems 4.6 and 4.8 are deferred to Appendix D.3. The argument for Theorem 4.6 is similar to that of Theorem 4.1: When $S_0, \dots, S_L \ge 2$, the claim that $\operatorname{rank}(\nabla^2 s_3(\theta_*)) \ge d - 3$ may be established by induction on L down to the base case of $L = 1$ rather than $L = 10$, using a different construction of the point $\theta_* \in \mathbb{R}^d$ that induces a sparse structure in $\nabla^2 s_3(\theta_*)$.

4.3. *Projected cryo-EM.* We now extend the model of the preceding section to include the tomographic projection arising in cryo-EM. In this projected model, we observe samples

$$(\Pi \cdot f_{\mathfrak{g}})(x)\,\mathrm{d}x + \sigma\,\mathrm{d}W(x) \tag{4.10}$$

on $\mathbb{R}^2$ where, for $x = (x_1, x_2) \in \mathbb{R}^2$, the tomographic projection Π is defined by

$$(\Pi \cdot f_{\mathfrak{g}})(x_1, x_2) = \int_{-\infty}^{\infty} f_{\mathfrak{g}}(x_1, x_2, x_3)\,\mathrm{d}x_3, \tag{4.11}$$

and $\mathrm{d}W(x)$ in (4.10) is a standard Gaussian white noise process on the projected domain $\mathbb{R}^2$.

Our model setup is similar to [4], Section 4.6. We again model the Fourier transform of f in a basis $\{\hat{j}_{lsm}\}$ given by the product of complex spherical harmonics with radial functions. We restrict f to a space of $(L, S_0, \dots, S_L)$-bandlimited functions with representation

$$f = \sum_{(l,s,m)\in\mathcal{I}} u_m^{(ls)} \cdot j_{lsm} = \sum_{(l,s,m)\in\mathcal{I}} \theta_m^{(ls)} \cdot h_{lsm} \tag{4.12}$$

for the index set $\mathcal{I}$ defined in (4.5), where $\{j_{lsm}\}$ are the inverse Fourier transforms of $\{\hat{j}_{lsm}\}$, and the second equality describes a parametrization by an equivalent real orthonormal basis $\{h_{lsm}\}$ as before. In Appendix D.4, we apply the Fourier slice theorem to derive basis representations for the tomographic projection $\Pi \cdot f$. These take the forms

$$\Pi \cdot f = \sum_{(s,m)\in\widetilde{\mathcal{I}}} \tilde{u}_m^{(s)} j_{sm} = \sum_{(s,m)\in\widetilde{\mathcal{I}}} \tilde{\theta}_m^{(s)} h_{sm}, \tag{4.13}$$

where $\{j_{sm}\}$ and $\{h_{sm}\}$ are (complex and real, resp.) basis functions over $\mathbb{R}^2$, and $\Pi \cdot f$ is bandlimited to an index set

$$\widetilde{\mathcal{I}} = \big\{(s,m) : 1 \le s \le S, -L \le m \le L\big\}, \qquad \tilde{d} = |\widetilde{\mathcal{I}}| = S(2L+1) \tag{4.14}$$

for $S = \max(S_0, \dots, S_L)$. This expresses Π as a linear map from $\theta \in \mathbb{R}^d$ to $\tilde{\theta} \in \mathbb{R}^{\tilde{d}}$, and we give its explicit form in (D.S57). We choose radial functions to ensure that the basis $\{h_{sm}\}$ is orthonormal in $L_2(\mathbb{R}^2)$, so that (4.10) is equivalent to observing the coefficients of $\Pi \cdot f$ in this basis with i.i.d. $\mathcal{N}(0, \sigma^2)$ noise. Further details of the setup are described in Appendix D.4.

The following result verifies that when the bandlimits satisfy $L \ge 1$ and $S_1, \dots, S_L \ge 4$, we have also $\operatorname{trdeg} \widetilde{\mathcal{R}}^{\mathsf{G}}_{\le \widetilde{K}} = \operatorname{trdeg} \mathcal{R}^{\mathsf{G}}$ for $\widetilde{K} = 3$.

THEOREM 4.9. *For any $L \geq 1$ and $S_0, \ldots, S_L \geq 4$, we have*

$$\operatorname{trdeg}(\widetilde{\mathcal{R}}^{\mathsf{G}}_{\leq 1}) = S_0,$$

$$\operatorname{trdeg}(\widetilde{\mathcal{R}}^{\mathsf{G}}_{\leq 2}) = \sum_{l=0}^{L} d(S_l), \qquad d(S_l) \equiv \begin{cases} \dfrac{S_l(S_l+1)}{2} & \text{for } S_l < 2l+1, \\ (2l+1)(S_l - l) & \text{for } S_l \geq 2l+1, \end{cases}$$

$$\operatorname{trdeg}(\widetilde{\mathcal{R}}^{\mathsf{G}}_{\leq 3}) = \operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}) = d - 3,$$

which matches the values of $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\leq 1})$, $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\leq 2})$, *and* $\operatorname{trdeg}(\mathcal{R}^{\mathsf{G}}_{\leq 3})$ *in the unprojected setting of Theorem 4.6.*

COROLLARY 4.10. *In this projected cryo-EM model with $S_0, \ldots, S_L \geq 4$, a generic signal $\theta_* \in \mathbb{R}^d$ may be identified up to a finite list of orbits by the moments of $\Pi(g \cdot \theta_*)$ up to order $\widetilde{K} = 3$ if $L \geq 2$, and order $\widetilde{K} = 2$ if $L = 1$.*

We prove Theorem 4.9 in Appendix D.4, where we also state an analogue of Theorem 4.8 that describes the explicit forms of $\tilde{s}_k(\theta)$ for $k = 1, 2, 3$ in this projected model.

Our proof of Theorem 4.9 again constructs a point $\theta_* \in \mathbb{R}^d$ where $\operatorname{rank}(\nabla^2 \tilde{s}_3(\theta_*)) \geq d - 3$. However, the form of $\nabla^2 \tilde{s}_3(\theta_*)$ now involves the precise form of the projection Π, and our choice of θ_* does not induce sparsity in this Hessian. Instead, we choose θ_* to have many coordinates equal to 0, and track the dependence of minors of $\nabla^2 \tilde{s}_3(\theta_*)$ on the nonzero coordinates of θ_* to show they are generically nonvanishing. We give this argument in the proof of Lemma D.10 in Appendix D.4.

REMARK 4.11. Taking $S_0 = \cdots = S_L = S$ yields a model equivalent to the projected cryo-EM model with S spherical shells in [4], Section 4.6. In [4], Conjecture 4.17, the authors conjectured that a generic signal θ_* may be identified up to a finite list of orbits by third-order moments if $S \geq 2$. Corollary 4.10(a) thus resolves this conjecture positively when $S \geq 4$. The constraint $S \geq 4$ is technical, and we believe that the conjecture holds as stated for $S \in \{2, 3\}$ as well, but we do not pursue these cases in this work.

5. Numerical evaluations of the Fisher information. We conclude with an empirical investigation of the spectrum of the Fisher information matrix in two simulated examples of the unprojected cryo-EM model described in Section 4.2.

In each example, we begin with a near-atomic-resolution electric potential map estimated from a cryo-EM experiment. We obtain a finite-dimensional approximation to this map by applying a low-pass filter to its Fourier transform, followed by a basis approximation for the filtered map. We simulate noisy and rotated samples using this finite-dimensional approximation as the underlying true signal, for various inverse-SNR parameters

$$\alpha \equiv \sigma^2 / \|\theta_*\|^2.$$

We then study the dependence of eigenvalues of the observed information matrix $\nabla^2 R_n(\theta_*)$ on α.

Rotavirus VP6 trimer. We consider a map of the VP6 trimer in bovine rotavirus, reported in [45] (EMDB:1461). A contour plot of this map is overlaid with the atomic structure previously obtained by [30] (PDB:1QHD), in Figure 1(a). We applied low-pass filters in the Fourier domain at two different cutoff frequencies, a "low-resolution" frequency of

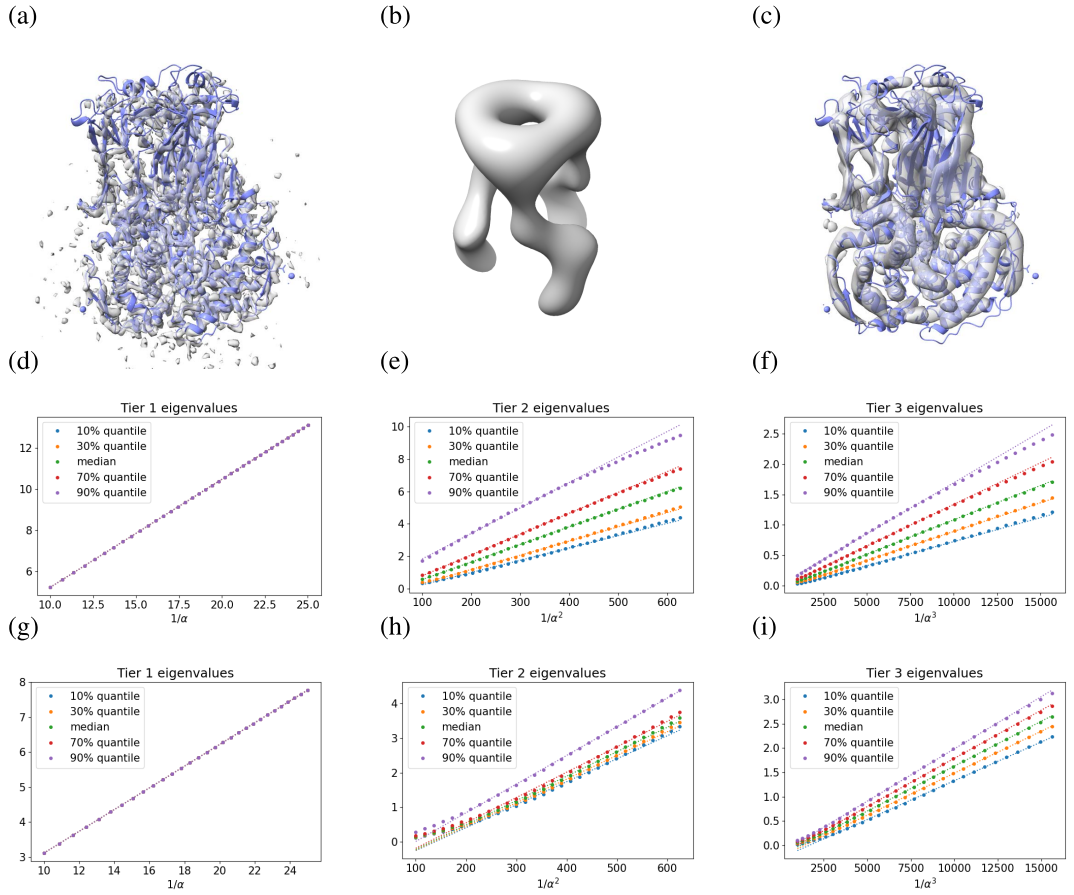

FIG. 1. (*a*) 3.8Å*-resolution cryo-EM map of the rotavirus VP*6 *trimer, overlaid with the atomic structure.* (*b*) *A finite-dimensional approximation using* 405 *basis functions at* 24.6Å*-resolution* (*displayed in a rotated orientation for clarity*). (*c*) *An approximation using* 4410 *basis functions at* 8.2Å*-resolution.* (*d–f*) *We stratify the eigenvalues of the* 405*-dimensional observed Fisher information corresponding to* (*b*) *into three "eigenvalue tiers" according to Theorem* 4.6*, and plot the scalings of the* 10*th,* 30*th,* 50*th,* 70*th and* 90*th percentiles of eigenvalues in each tier against* $1/\alpha \propto \sigma^{-2}$, $1/\alpha^2 \propto \sigma^{-4}$ *and* $1/\alpha^3 \propto \sigma^{-6}$. (*These quantiles nearly overlap for Tier* 1.) *Linear trends fitted using least squares are shown as dashed lines.* (*g–i*) *The same for the* 4410*-dimensional Fisher information matrix corresponding to* (*c*).

$(24.6\text{Å})^{-1}$ and a "medium-resolution" frequency of $(8.2\text{Å})^{-1}$. The corresponding smoothed maps in the spatial domain are depicted in Figure S1 of Appendix E.

We approximated each smoothed map using a finite basis of the form (4.3), with an adaptive construction of the radial functions $\{z_s\}$ to maximize the power captured by each successive radial frequency. Details of our numerical procedures are described in Appendix E. Choosing bandlimits $(S, L) = (5, 8)$ and total dimension $d = L(S+1)^2 = 405$ gave an accurate approximation to the 24.6Å-resolution map that reveals the trimer composition of the VP6 complex, as depicted in Figure 1(b). Choosing bandlimits $(S, L) = (10, 20)$ and total dimension $d = L(S+1)^2 = 4410$ gave an accurate approximation of the 8.2Å-resolution map that captures interesting aspects of the tertiary and secondary structure, as shown in Figure 1(c). We denote the basis coefficients of these approximated maps as $\theta_* \in \mathbb{R}^d$.

We computed the Hessians $\nabla^2 R_n(\theta_*)$ of the empirical negative log-likelihood functions from $n = 500{,}000$ simulated samples, with inverse-SNR $\alpha = \sigma^2/\|\theta_*\|^2 \in [0.04, 0.10]$. We then separated the largest $d-3$ eigenvalues of $\nabla^2 R_n(\theta_*)$ into three "tiers" with cardinalities (d_1, d_2, d_3) as implied by Theorem 4.6. Figure 1(d–f) depicts representative eigenvalues in

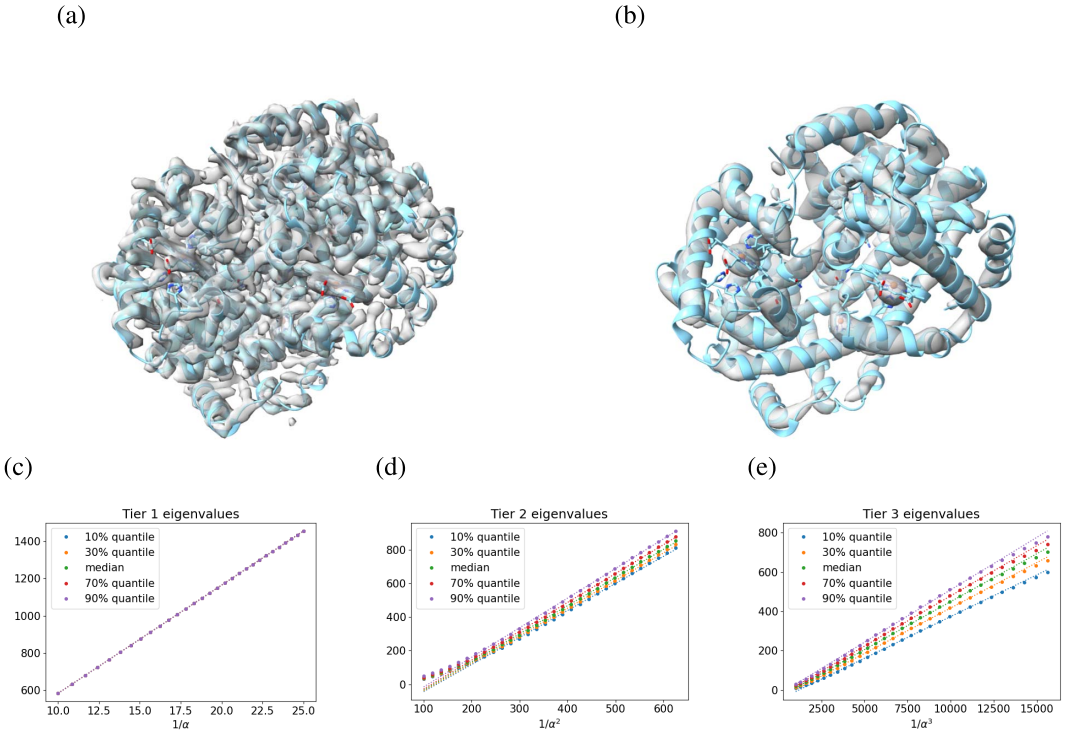

FIG. 2. (*a*) 3.4Å*-resolution cryo-EM map of hemoglobin, overlaid with the atomic structure.* (*b*) *A finite-dimensional approximation using* 3528 *basis functions at* 7.0Å*-resolution.* (*c–e*) *The* 10*th*, 30*th*, 50*th*, 70*th and* 90*th percentiles of eigenvalues within each "eigenvalue tier" of the* 3528*-dimensional observed Fisher information, plotted against* $1/\alpha \propto \sigma^{-2}$, $1/\alpha^2 \propto \sigma^{-4}$, $1/\alpha^3 \propto \sigma^{-6}$ *as in Figure* 1.

each tier, plotted against $1/\alpha \propto \sigma^{-2}$, $1/\alpha^2 \propto \sigma^{-4}$ and $1/\alpha^3 \propto \sigma^{-6}$. A linear trend is observed in all settings, in agreement with the prediction of Theorem 2.7. This may be contrasted with Figure S2 in Appendix E, which instead plots eigenvalues in all three tiers against $1/\alpha \propto \sigma^{-2}$, and where nonlinearity of the scaling is visually apparent for Tiers 2 and 3.

Hemoglobin. We consider a map of hemoglobin, reported in [27] (EMDB:3650, PDB:5NI1). A contour plot overlaid with the atomic structure is presented in Figure 2(a). We applied a low-pass filter with cutoff frequency $(7.0\text{Å})^{-1}$ in the Fourier domain, depicted in Figure S1. We then applied a basis approximation with bandlimits $(S, L) = (8, 20)$ and total dimension $d = 3528$. The approximated map is shown in Figure 2(b), and captures important aspects of the secondary structure including the locations of the α-helices and embedded prosthetic heme groups. We denote the basis coefficients of this approximation as θ_*.

Figure 2(c–e) again depicts the leading $d - 3$ eigenvalues of $\nabla^2 R_n(\theta_*)$ computed from $n = 500{,}000$ simulated samples, stratified into three tiers of sizes (d_1, d_2, d_3). Linear trends with $1/\alpha \propto \sigma^{-2}$, $1/\alpha^2 \propto \sigma^{-4}$ and $1/\alpha^3 \propto \sigma^{-6}$ are again observed, and may be contrasted with the nonlinear scalings of eigenvalues in Tiers 2 and 3 with $1/\alpha$ as depicted in Figure S2.

We note that although the eigenvalues of $\nabla^2 R_n(\theta_*)$ do scale with powers of the SNR $1/\alpha$ according to our theoretical predictions, at any fixed SNR and for basis dimensions exceeding $d \approx 100$, we do not observe a clear separation between the eigenvalues of Tier 2 and of Tier 3, due to the variation in magnitude of eigenvalues corresponding to differing radial frequencies within each tier.

In these examples, we also begin to observe some deviations from the predicted eigenvalue scalings at the higher and lower ends of tested SNR. Deviations in higher basis dimensions

d and at lower SNR $1/\alpha$ (seen in Figures 1(h–i) and 2(d)) are likely finite-sample effects due to differences between the observed information matrix $\nabla^2 R_n(\theta_*)$ and the (population) Fisher information $I(\theta_*) = \nabla^2 R(\theta_*)$. We believe that deviations at higher SNR $1/\alpha$ (seen in Figures 1(e–f) and 2(e)) reflect a departure of the behavior of the population Fisher information $I(\theta_*)$ from the predictions of the large-σ theoretical regime. Our largest tested SNR $1/\alpha = 25$ yields a spectral SNR (average power of signal / average power of noise at a fixed Fourier radius) of 0.2–0.4 near the origin of the Fourier domain, which we believe reflects a level of noise that may be slightly higher than that of modern cryo-EM experiments.

6. Conclusion. In this work, we characterized properties of the Fisher information matrix and log-likelihood function landscape for continuous group orbit estimation problems in a high noise regime, showing that they are related to the structure of the invariant algebra of the rotational group. We applied these results to study several models of function estimation in finite-dimensional function spaces, in particular establishing that third-order moment information is sufficient to locally identify generic signals in these models.

In many interesting applications including single-particle cryo-EM, the target function at full spatial resolution may not admit an accurate low-dimensional approximation. In such settings, our theoretical results may have relevance to estimating lower-dimensional smoothed approximations of the function. We demonstrated in simulation that this theory can accurately predict the noise scalings of the Fisher information eigenvalues for two small protein molecules over a range of sufficiently high noise, or low SNR. We highlight the theoretical understanding of likelihood-based estimation in high-dimensional and infinite-dimensional settings and over a broader range of SNR as a question for future work.

Acknowledgments. We would like to thank Fred Sigworth for helpful discussions about cryo-EM, and for suggesting to us the hemoglobin example. We would also like to thank two anonymous reviewers for their detailed feedback, which has helped us significantly improve the exposition of our manuscript.

Funding. ZF was supported in part by NSF Grants DMS-1916198 and DMS-2142476. RRL was supported in part by NIH/NIGMS 1R01GM136780-01. YS was supported in part by NSF Grants DMS-1701654, DMS-2039183 and DMS-2054838.

SUPPLEMENTARY MATERIAL

Supplementary appendices (DOI: 10.1214/23-AOS2292SUPP; .pdf). The Appendices contain proofs of our general results in Section 2, further details for Example 2.14 on the Procrustes alignment model, further details and proofs for the function estimation problems discussed in Sections 3 and 4, details on our numerical simulations in Section 5 and a brief introduction to cryo-EM and its relation the problems studied in this work.

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., 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

[3] 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, Vail, CO, USA.

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[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] BENDORY, T., MICKELIN, O. and SINGER, A. (2022). Sparse multi-reference alignment: Sample complexity and computational hardness. In *ICASSP* 2022-2022 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) 8977–8981. IEEE, Singapore, Singapore.

[12] BRUNEL, V.-E. (2019). Learning rates for Gaussian mixtures under group action. In *Conference on Learning Theory* 471–491.

[13] DUBOCHET, J., ADRIAN, M., CHANG, J.-J., HOMO, J.-C., LEPAULT, J., MCDOWALL, A. W. and SCHULTZ, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21** 129–228.

[14] FAN, Z., LEDERMAN, R. R, SUN, Y., WANG, T. and XU, S. (2024). Supplement to "Maximum likelihood for high-noise group orbit estimation and single-particle cryo-EM." https://doi.org/10.1214/23-AOS2292SUPP

[15] 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

[16] FRANK, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*: *Visualization of Biological Molecules in Their Native State*. Oxford Univ. Press, London.

[17] GE, R., HUANG, F., JIN, C. and YUAN, Y. (2015). Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Conference on Learning Theory* 797–842.

[18] GHOSH, S. and RIGOLLET, P. (2023). Sparse multi-reference alignment: Phase retrieval, uniform uncertainty principles and the beltway problem. *Found. Comput. Math.* **23** 1851–1898. MR4649436 https://doi.org/10.1007/s10208-022-09584-6

[19] GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B* **53** 285–339. MR1108330

[20] GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika* **40** 33–51. MR0405725 https://doi.org/10.1007/BF02291478

[21] HENDERSON, R., BALDWIN, J. M., CESKA, T. A., ZEMLIN, F., BECKMANN, E. A. and DOWNING, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213** 899–929.

[22] IBRAGIMOV, I. and HAS'MINSKII, R. (1981). *Statistical Estimation*: *Asymptotic Theory*. Springer, Berlin.

[23] JIN, C., GE, R., NETRAPALLI, P., KAKADE, S. M. and JORDAN, M. I. (2017). How to escape saddle points efficiently. In *Proceedings of the* 34*th International Conference on Machine Learning-Volume* 70 1724–1732. JMLR.org.

[24] JOHNSTONE, I. M. (2017). *Gaussian Estimation*: *Sequence and Wavelet Models*. Unpublished draft.

[25] 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

[26] 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

[27] KHOSHOUEI, M., RADJAINIA, M., BAUMEISTER, W. and DANEV, R. (2017). Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* **8** 1–6.

[28] LEE, J. D., SIMCHOWITZ, M., JORDAN, M. I. and RECHT, B. (2016). Gradient descent only converges to minimizers. In *Conference on Learning Theory* 1246–1257.

[29] LIU, A. and MOITRA, A. (2021). Algorithms from invariants: Smoothed analysis of orbit recovery over SO(3). arXiv e-prints arXiv–2106.

[30] MATHIEU, M., PETITPAS, I., NAVAZA, J., LEPAULT, J., KOHLI, E., POTHIER, P., PRASAD, B. V., COHEN, J. and REY, F. A. (2001). Atomic structure of the major capsid protein of rotavirus: Implications for the architecture of the virion. *EMBO J.* **20** 1485–1497. https://doi.org/10.1093/emboj/20.7.1485

[31] MEI, S., BAI, Y. and MONTANARI, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

[32] MITYAGIN, B. S. (2020). The zero set of a real analytic function. *Mat. Zametki* **107** 473–475. MR4070868 https://doi.org/10.4213/mzm12620

[33] 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

[34] 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

[35] PUNJANI, A., RUBINSTEIN, J. L., FLEET, D. J. and BRUBAKER, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14** 290–296. https://doi.org/10.1038/nmeth.4169

[36] 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

[37] SCHERES, S. H., GAO, H., VALLE, M., HERMAN, G. T., EGGERMONT, P. P., FRANK, J. and CARAZO, J.-M. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat. Methods* **4** 27–29.

[38] 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

[39] 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. MR4072349 https://doi.org/10.1088/1361-6420/ab6139

[40] 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

[41] 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

[42] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[43] WU, Y. and ZHOU, H. H. (2021). Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. *Math. Stat. Learn.* **4** 143–220. MR4383733 https://doi.org/10.4171/msl/29

[44] XU, J., HSU, D. and MALEKI, A. (2016). Global analysis of expectation maximization for mixtures of two Gaussians. In *Proceedings of the* 30*th International Conference on Neural Information Processing Systems* 2684–2692.

[45] ZHANG, X., SETTEMBRE, E., XU, C., DORMITZER, P. R., BELLAMY, R., HARRISON, S. C. and GRIGORIEFF, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA* **105** 1867–1872.