METHOD OF MOMENTS FOR 3-D SINGLE PARTICLE *AB INITIO* MODELING WITH NON-UNIFORM DISTRIBUTION OF VIEWING ANGLES

Nir Sharon*,†

School of Mathematical Sciences, Tel Aviv University

Joe Kileel†

Program in Applied and Computational Mathematics, Princeton University

Yuehaw Khoo

Department of Statistics and the College, The University of Chicago

Boris Landa

Applied Mathematics Program, Yale University.

Amit Singer

Program in Applied and Computational Mathematics and Department of Mathematics, Princeton University

Abstract. Single-particle reconstruction in cryo-electron microscopy (cryo-EM) is an increasingly popular technique for determining the 3-D structure of a molecule from several noisy 2-D projections images taken at unknown viewing angles. Most reconstruction algorithms require a low-resolution initialization for the 3-D structure, which is the goal of *ab initio* modeling. Suggested by Zvi Kam in 1980, the method of moments (MoM) offers one approach, wherein low-order statistics of the 2-D images are computed and a 3-D structure is estimated by solving a system of polynomial equations. Unfortunately, Kam's method suffers from restrictive assumptions, most notably that viewing angles should be distributed uniformly. Often unrealistic, uniformity entails the computation of higher-order correlations, as in this case first and second moments fail to determine the 3-D structure. In the present paper, we remove this hypothesis, by permitting an unknown, non-uniform distribution of viewing angles in MoM. Perhaps surprisingly, we show that this case is *statistically easier* than the uniform case, as now first and second moments generically suffice to determine low-resolution expansions of the molecule. In the idealized setting of a known, non-uniform distribution, we find an efficient provable algorithm inverting first and second moments. For unknown, non-uniform distributions, we use non-convex optimization methods to solve for both the molecule and distribution.

2010 *Mathematics Subject Classification.* Primary: 78M05, 90C26; Secondary: 14Q99.

Key words and phrases. cryo-EM, *ab initio* modeling, autocorrelation analysis, method of moments, spherical harmonics, Wigner matrices, polynomial equations, non-convex optimization.

*Corresponding author: Nir Sharon.

†The first two authors contributed equally.

1. **Introduction.** Single-particle cryo-electron microscopy (cryo-EM) is an imaging method for determining the high-resolution 3-D structure of biological macromolecules without crystallization [25, 35]. The reconstruction process in cryo-EM determines the 3-D structure of a molecule from its noisy 2-D tomographic projection images. By virtue of the experimental setup, each projection image is taken at an unknown viewing direction and has a very high level of noise, due to the small electron dose one can apply to the specimen before inflicting severe radiation damage, *e.g.*, [12, 24, 41]. The computational pipeline that leads from the raw data, given many large unsegmented micrographs of projections, to the 3-D model consists of the following stages. The first step is particle picking, in which 2-D projection images are selected from micrographs. The selected particle images typically undergo 2-D classification to assess data quality and further improve particle picking. At this point, the 3-D reconstruction process begins, where often it is divided into two substeps of low-resolution modeling and 3-D refinement. In this paper, we focus on the mathematical aspects of the former, namely the modeling part. In particular, we suggest using the method of moments (MoM) for *ab initio* modeling. We illustrate this workflow with an overview given in Figure 1.

The last step in the reconstruction, also known as the refinement step, aims to improve the resolution as much as possible. This refinement process is typically a variant of the expectation-maximization (EM) algorithm which seeks the maximum likelihood estimator (MLE) via an efficient implementation, *e.g.*, [52]. As such, 3-D refinement requires an initial structure that is close to the correct target structure [28, 51]. Serving this purpose, an *ab initio* model is the result of a reconstruction process which depends solely on the data at hand with no *a priori* assumptions about the 3-D structure of the molecule [49]. We remark that the two primary challenges for cryo-EM reconstruction are the high level of noise and the unknown viewing directions. Mathematically, without the presence of noise, the unknown viewing directions could be recovered using common lines [61, 62]. Then, the 3-D structure follows, for example, by tomographic inversion, see, *e.g.*, [2]. Reliable detection of common lines is limited however to high signal-to-noise (SNR) ratio. As a result, the application of common lines based approaches is often limited to 2-D class averages rather than the original raw images [56]. Other alternatives such as frequency marching [7] and optimization using stochastic gradient have been suggested [48]. As optimization processes are designed to minimize highly non-convex cost functions, methods like SGD are not guaranteed to succeed. In addition, as in the case of EM, it is not *a priori* clear how many images are required.

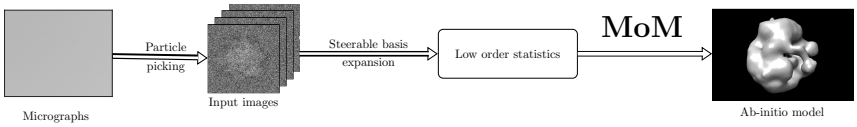


FIGURE 1. A schematic flowchart of 3-D reconstruction using method of moments (MoM).

Approximately forty years ago, Zvi Kam proposed a method for 3-D *ab initio* reconstruction based on computing the mean and covariance of the 2-D noisy images [33]. In order to uniquely determine the volume, the third moment (triple correlation) is also used besides the mean and covariance. In this approach, known

as Kam's method, the 3-D volume is reconstructed without estimating the viewing directions. In this sense, Kam's method is strikingly different from common lines based approaches and maximum likelihood and other optimization methods that rely on orientation estimation for each image. Crucially, Kam's method is effective at *arbitrary* levels of noise, given sufficiently many picked particles for accurate estimation of the moment statistics. Additionally, Kam's method does not require any starting model, and it requires only one pass through the data to compute moments (contrary to other approaches needing access to the measurements multiple times). Despite the aforementioned advantages, Kam's method relies on the restrictive assumption that the viewing directions for the images are distributed uniformly over the sphere. This hypothesis, alongside other technical issues, has so far prevented a direct application of Kam's method to experimental cryo-EM data, for which viewing angles are typically non-uniform [4, 26, 44, 59]. This situation motivates us to explore generalizations of Kam's method better suited to cryo-EM data.[1]

In this paper, we generalize Kam's theory to the case of non-uniform distribution of viewing directions. We regard Kam's original approach with uniform distribution of viewing angles as a degenerate instance of MoM. In our formulation, we estimate both the 3-D structure and the unknown distribution of viewing angles jointly from the first two moments of the Fourier transformed images. More precisely, for n images $I_j, j = 1, \dots, n$, the first and second empirical moments of the Fourier transformed images, given in polar coordinates, $\widehat{I}_j(r, \varphi), j = 1, \dots, n$, are

$$\widetilde{m}_1(r, \varphi) = \frac{1}{n} \sum_{j=1}^{n} \widehat{I}_j(r, \varphi), \quad \text{and} \quad \widetilde{m}_2(r, \varphi, r', \varphi') = \frac{1}{n} \sum_{j=1}^{n} \widehat{I}_j(r, \varphi) \widehat{I}_j(r', \varphi'), \quad (1)$$

which upon the above discretization become 2-D and 4-D tensors, respectively. Our basic rationale for trying to obtain the volume from the first two moments is as follows. Supposing the distribution of rotations of the *image plane* to be uniform, then in the limit $n \to \infty$ the first moment is radially symmetric, that is, it is only a function of r but is independent of φ. Therefore, $\widetilde{m}_1$ may be regarded as a 1-D vector. Similarly, the second moment is a 3-D tensor (rather than 4-D) since it will only depend on φ and φ' through $\varphi - \varphi'$ as $n \to \infty$. Also $I_j(r', \varphi')$ is linearly related to the molecule's volume via a tomographic projection. Thus, for images of size $N \times N$ pixels, the first and second moments should give rise to $\mathcal{O}(N^3)$ polynomial equations for the unknown volume and distribution. Assuming the volume is of size $N \times N \times N$ (and the distribution is of lower dimensionality), then the first and second moments have "just" the right number of equations (in terms of leading order) to determine the unknowns. Unfortunately, when the distribution of viewing directions is uniform, as noted by Kam [33], the information encoded in the second moment is algebraically redundant; essentially it is the autocorrelation function (or equivalently, the power spectrum), and this information is insufficient for determining the structure of the molecule. As we will see, a non-uniform distribution of viewing directions introduces additional terms in both the first and second moments, and extends the number of independent equations beyond the autocorrelation case. In particular, we will show that non-uniformity guarantees uniqueness from the analytical counterparts of $\widetilde{m}_1$ and $\widetilde{m}_2$ in cases of a known distribution, and

[1] We remark that Kam's method, assuming uniform rotations, is of significant current interest in X-ray free electron laser (XFEL) single molecule imaging, where the assumption of uniformity more closely matches experimental reality [21, 45, 65].

it guarantees finitely many solutions in other, more realistic, cases of an unknown distribution.

Our work is inspired by several earlier studies on simplified models in a setting called Multi-Reference Alignment (MRA). In MRA, a given group of transformations acts on a vector space of signals [5]. For example, the group SO(2) acts on the space of band-limited signals over the unit circle by rotating them counterclockwise (as a 1-D analog of cryo-EM). The task then is to estimate a ground truth signal from multiple noisy samples, corresponding to unknown group elements of a finite cyclic subgroup of SO(2) acting on the signal. The papers [6, 9] show that for a uniform distribution over the group, the signal can be estimated from the third moment, and the number of samples required scales like the third power of the noise variance. On the other hand, for a non-uniform and also aperiodic distributions over the group, the signal can be estimated from the second moment, and the required number of samples scales quadratically with the noise variance [1].

Despite the success of signal recovery in MRA from the first two moments under the action of the cyclic group, it is not apparent that such a strategy is still applicable in the case of cryo-EM. First, in cryo-EM, each image is obtained from the ground truth volume not just by applying a rotation in SO(3), but also a tomographic projection. Moreover, the studies mentioned above (of MRA) consider finite abelian groups, whereas, in the case of cryo-EM, the group under consideration is the continuous non-commutative group SO(3). The goal of this paper is then to investigate whether the first and second moment of the images is also sufficient for solving the inverse problem of structure determination in the cryo-EM setting.

1.1. **Our contribution.** We formulate the reconstruction problem in cryo-EM as an inverse problem of determining the volume and the distribution of viewing directions from the first two moments of the images. Assuming the volume and distribution are band-limited functions, they are discretized by Prolate Spheroidal Wave Functions (PSWFs) and Wigner matrices, respectively. The moments give rise to a polynomial system in which the unknowns are the coefficients of the volume and the distribution. Using computational algebraic geometry techniques [20, 23, 58], we exhibit a range of band limits for the volume and the distribution such that the polynomial system has only finitely many solutions, pointing to the possibility of exact recovery in these regimes. Additionally, we comment on numerical stability issues, by providing condition number formulas for moment inversion. In the setting where the rotational distribution is known, we prove that the number of solutions is generically 1 and present an efficient algorithm for recovering the volume using ideas from tensor decomposition [31]. For the practical case of an unknown distribution, we rely on methods from non-convex optimization and demonstrate, with synthetic data, successful *ab initio* model recovery of a molecule from the first two moments.

1.2. **Organization.** The paper is organized as follows. In Section 2, we present discretizations for the volume and distribution and derive the polynomial system obtained from the first two moments. In Section 3, we demonstrate that there exists a range of band limits where the polynomial system for the unknown molecule and distribution has only finitely many solutions. In Section 4, we discuss some implementation details on how the system is solved and present numerical and visual results. Proofs and background material are provided in appendices. For research reproducibility, MATLAB code is publicly available at GitHub.com.[2]

[2]The full address of the GitHub repository is https://github.com/nirsharon/nonuniformMoM.

2. **Method of moments.** We begin by introducing the image formation model. Then, convenient basis for discretizing various continuous objects, namely the images and the volume (in the Fourier domain) as well as the distribution for orientations, are introduced. From these, relationships between the moments of the 2-D images and the 3-D molecular volume can be derived, enabling us to fit the molecular structure to the empirical moments of the images.

2.1. **Image formation model and the 3-D reconstruction problem.** In cryo-EM, data is acquired by projecting particles embedded in ice along the direction of the beaming electrons, resulting in tomographic images of the particles. The particles orient themselves randomly with respect to the projection direction. More formally, let $\phi\colon \mathbb{R}^3 \to \mathbb{R}$ be the Coulomb potential of the 3-D volume, and the projection operator be denoted by $\mathcal{P}\colon \mathbb{R}^3 \to \mathbb{R}^2$, where

$$\mathcal{P}\phi(x_1, x_2) := \int_{-\infty}^{\infty} \phi(x_1, x_2, x_3)\, dx_3. \tag{2}$$

Assuming the j-th particle comes from the same volume ϕ but rotated by $R_j \in \mathrm{SO}(3)$, the image formation model is [10, 25]

$$I_j = h_j * \mathcal{P}\left(R_j^T \cdot \phi\right) + \varepsilon_j, \quad R_j \in \mathrm{SO}(3), \quad j = 1, \ldots, n\,, \tag{3}$$

where ε_j is a random field modeling the noise term and h_j is a point spread function, whose Fourier transform is known as the contrast transfer function (CTF) [42, 50, 60]. Each image is assumed to lie within the box $[-1, 1] \times [-1, 1]$. For size $N \times N$ discretized images, we assume the random field $\varepsilon_j \sim \mathcal{N}(0, \sigma^2 I_{N^2})$, $j = 1, \ldots, n$. Here R_j denotes an element in the group of 3×3 rotations $\mathrm{SO}(3)$, and we define the group action by[3]

$$R_j^T \cdot \phi(x_1, x_2, x_3) := \phi(R_j \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T). \tag{4}$$

The rotations R_j's are not known since the molecules can take any orientation with respect to projection direction. For the purpose of simplifying the exposition, we shall henceforth disregard the CTF, by assuming

$$I_j = \mathcal{P}\left(R_j^T \cdot \phi\right) + \varepsilon_j, \quad j = 1, \ldots, n. \tag{5}$$

The presence of CTF is not expected to have a major impact on our main results, and we will incorporate the CTF in a future work. Typically, it is convenient to consider Fourier transform of the images, since by projection slice theorem, the Fourier transform $\widehat{I_j}$ of I_j gives a slice of the Fourier coefficients $\widehat{\phi}$ of the volume ϕ:

$$\widehat{I_j}(x_1, x_2) = \widehat{\mathcal{P}(R_j^T \cdot \phi)}(x_1, x_2) + \widehat{\varepsilon_j} = (R_j^T \cdot \widehat{\phi})(x_1, x_2, x_3)|_{x_3=0} + \widehat{\varepsilon_j}. \tag{6}$$

The goal of cryo-EM is to recover $\widehat{\phi}$ from the Fourier coefficients of the projections $\widehat{I_j}(x_1, x_2)$. While reconstructing $\widehat{\phi}$ given estimated R_j's amounts to solving a standard computed tomography problem, we wish to reconstruct $\widehat{\phi}$ directly from the noisy images without estimating the rotations, for reasons detailed above. To this end, we assume the rotations are sampled from a distribution ρ on $\mathrm{SO}(3)$, where $\rho\colon \mathrm{SO}(3) \to \mathbb{R}$ is a smooth band-limited function. Then from the empirical moments of the images $\{\widehat{I_j}\}_{j=1}^n$, we jointly estimate the volume $\widehat{\phi}$ and the distribution ρ.

[3]Here we prefer to write the action of R^T and correspondingly later we use Wigner U-matrices, instead of R and Wigner D-matrices. While simply notational, these conventions allow us to cite identities from [19] verbatim, which are in terms of Wigner U-matrices and not Wigner D-matrices.

2.2. **Representation of the volume, the distribution of rotations and the images.** As mentioned previously, the proposed method of moments consists of fitting the analytical moments

$$m_1 = \mathbb{E}_{R\sim\rho}[\mathcal{P}\left(\widehat{R^T\cdot\phi}\right)], \quad \text{and} \quad m_2 = \mathbb{E}_{R\sim\rho}[\mathcal{P}\left(\widehat{R^T\cdot\phi}\right)\otimes\mathcal{P}\left(\widehat{R^T\cdot\phi}\right)]. \tag{7}$$

to their empirical counterparts $\widetilde{m}_1$ and $\widetilde{m}_2$ as appears in (1) after debiasing.[4] Through fitting to the empirical moments, we seek to determine the Fourier volume $\hat{\phi}$ and also the distribution ρ. In this section, we present discretizations of $\hat{\phi}$ and ρ by expanding them using convenient bases.

2.2.1. *Basis for the Fourier volume $\widehat{\phi}$.* Since the image formation model involves rotations of the Fourier volume $\widehat{\phi}$, it is convenient to represent $\widehat{\phi}$ as an element of a function space closed under rotations; in fact, this is the same as representing $\widehat{\phi}$ using spherical harmonics (see the Peter-Weyl theorem [19]):

$$\widehat{\phi}(\kappa,\theta,\varphi) = \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\sum_{s=1}^{S(\ell)} A_{\ell,m,s}F_{\ell,s}(\kappa)Y_\ell^m(\theta,\varphi). \tag{8}$$

Here Y_ℓ^m are the (complex) spherical harmonics:

$$Y_\ell^m(\theta,\varphi) = \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}\, P_\ell^m(\cos\theta)\, e^{im\varphi} \tag{9}$$

with associated Legendre polynomials P_ℓ^m defined by:

$$P_\ell^m(x) = \frac{(-1)^m}{2^\ell \ell!}(1-x^2)^{m/2}\frac{d^{\ell+m}}{dx^{\ell+m}}(x^2-1)^\ell. \tag{10}$$

In Cartesian coordinates, spherical harmonics are polynomials of degree ℓ. Without loss of generality, the radial frequency functions $F_{\ell,s}$ should form an orthonormal family (for each fixed ℓ) with respect to $\kappa^2 d\kappa$, where $s = 1,\dots,S(\ell)$ is referred to as the radial index. Choices of radial functions suitable for molecular densities include spherical Bessel functions [3], which are eigenfunctions of the Laplacian on a closed ball with Dirichlet boundary condition, as well as the radial components of 3-D prolate spheroidal wave functions [57].

We assume the volume is band-limited with Fourier coefficients supported within a radius of size $\pi N/2$, *i.e.*, the Nyquist cutoff frequency for the images I_j's discretized on a grid of size $N\times N$ (over the square $[-1,1]\times[-1,1]$). Under this assumption, the maximum degree and radial indices L and $S(\ell)$ in (8) are essentially finite. Further details on the particular basis functions $F_{\ell,s}$ and cutoffs L and $S(\ell)$ that we choose to use are deferred to Section A in the appendix. Note that in practice, as we target low-resolution modeling, one can choose to decrease either the cutoff or the grid size to obtain more compact settings. The coefficients $A_{\ell,m,s}\in\mathbb{C}$ furnish our representation of $\widehat{\phi}$ using spherical harmonics. Note that since ϕ is real valued, its Fourier transform is conjugate-symmetric, which imposes restrictions on the coefficients $A_{\ell,m,s}$. The specific constraints are presented in Section 4.1.

The advantage of expanding $\widehat{\phi}$ in terms of spherical harmonics is that the space of degree ℓ spherical harmonics is closed under rotation; in group-theoretic language,

[4] By the law of large numbers, $\widetilde{m}_1\to m_1$ and $\widetilde{m}_2\to m_2+\sigma^2 I$ almost surely as $n\to\infty$, so m_1 is fitted to $\widetilde{m}_1$ and m_2 to $\widetilde{m}_2-\sigma^2 I$. For notational convenience, we drop $\sigma^2 I$ in what follows, either assuming $\tilde{m}_2$ has been appropriately debiased already or $\sigma=0$.

this space forms a linear representation of SO(3).[5] Thus the action of a rotation on $\widehat{\phi}$ amounts to a linear transformation on the expansion coefficients $A_{\ell,m,s}$ (with a block structure according to ℓ and s). More precisely, fixing the vector space spanned by $\{Y_\ell^m(\theta,\varphi)\}_{m=-\ell}^{\ell}$ for a specific ℓ, the action of a rotation R on this vector space is represented by the *Wigner matrix* $U^\ell(R) \in \mathbb{C}^{(2\ell+1)\times(2\ell+1)}$ (see [19, p. 343]) so that:

$$R^T \cdot Y_\ell^m(x) = Y_\ell^m(Rx) = \sum_{m'=-\ell}^{\ell} U_{m,m'}^\ell(R) Y_\ell^{m'}(x), \quad x \in S^2. \tag{11}$$

In particular, the matrix $U^\ell(R)$ is unitary, with entries degree ℓ polynomials in the entries of R [19]. For all $R_1, R_2 \in \mathrm{SO}(3)$ and ℓ, the group homomorphism property reads $U^\ell(R_1R_2) = U^\ell(R_1)U^\ell(R_2)$. In light of (11), 3-D bases of the form $\{F_{\ell,s}(\kappa)Y_\ell^m(\theta,\phi)\}_{\ell,m,s}$ have been called *steerable bases*.

2.2.2. *Basis for the probability distribution of rotations ρ.* As we shall see, when expanding the volume in terms of spherical harmonics, the analytical moments (7) involve integrating different monomials of $\{U^l(R)\}_{\ell=0}^{L}$ with respect to the measure $\rho(R)dR$. To this end, we assume the probability density ρ over SO(3) is a smooth band-limited function (and in a function space closed under rotation) by expanding

$$\rho(R) = \sum_{p=0}^{P} \sum_{u,v=-p}^{p} B_{p,u,v} U_{u,v}^p(R), \quad R \in \mathrm{SO}(3). \tag{12}$$

By Peter-Weyl, these form an orthonormal basis for $L^2(\mathrm{SO}(3))$, and for higher p they are increasingly oscillatory functions on SO(3). Thus, expansion (12) is analogous to using spherical harmonics to expand a smooth function on the sphere, or using Fourier modes for a function on the circle. The cutoff $P \in \mathbb{N}$ is the band limit of the distribution ρ; we shall see in the next section that since we use only first and second moments it makes sense to assume $P \leq 2L$. Note that in the special case of a uniform distribution, the only nonzero coefficient is $B_{0,0,0} = 1$. Also, dR denotes the *Haar measure*, which is the unique volume form on the group of total mass one that is invariant under left action. Using the Euler angles parameterization of SO(3), the Haar measure is of the form

$$dR = \frac{1}{8\pi^2} \sin(\beta) d\alpha d\beta d\gamma, \tag{13}$$

where the normalizing constant ensures $\int_{\mathrm{SO}(3)} dR = \int_{\alpha=0}^{2\pi} \int_{\beta=0}^{\pi} \int_{\gamma=0}^{2\pi} dR = 1$.

2.2.3. *Basis for the 2-D images.* At this point, we discuss convenient representations for the images after Fourier transform, $\widehat{I}_j$. Similarly to volumes, it is desirable to represent images using a function space closed under in-plane rotations, *i.e.*, SO(2). By the Peter-Weyl theorem, this is the same as expanding using Fourier modes, in a 2-D steerable basis:

$$\widehat{I}_j(\kappa,\varphi) = \sum_{q=-Q}^{Q} \sum_{t=1}^{T(q)} a_{q,t}^j f_{q,t}(\kappa) e^{iq\varphi}. \tag{14}$$

Here the radial frequency functions $f_{q,t}$ (for fixed q) are taken to be an orthonormal basis with respect to $\kappa d\kappa$, with κ referred to as the radial frequency. Comparing

[5] In fact, this is an irreducible representation of SO(3) and varying ℓ these give all irreps.

to expansion (8) (see Section 2.2), it makes most sense to set $Q = L$. Again, owing to the Nyquist frequency for the discretized images I_j, we may bound the cutoffs $T(q)$. Typical choices for $f_{q,t}$ for representing tomographic images include Fourier-Bessel functions [66] and the radial components of 2-D prolate spheroidal wave functions [57]. Details on our specific choices are given in Section A.2 in the appendix.

2.2.4. *Choice of radial functions.* For the finite expansions in (8) and (14) to accurately represent the Fourier transforms of the electric potential and its slices, one should carefully choose the radial functions $F_{\ell,s}$ and $f_{q,t}$, together with the truncation-related quantities L, $S(\ell)$, Q, and $T(q)$. In this work, we consider $F_{\ell,s}$ and $f_{q,t}$ to be the radial parts of the three-dimensional and two-dimensional PSWFs [57], respectively. In Appendix A, we describe some of the key properties of the PSWFs, and propose upper bounds for setting L, $S(\ell)$, Q, and $T(q)$. In practice, band limits would be selected by balancing these expressivity considerations together with the well-posedness and conditioning considerations of Section 3.

2.3. **Low-order moments.** In this section, we derive the analytical relationship between the first two moments for the observed images $\{a^j_{q,t}\}_{j,q,t}$, and the coefficients $\{A_{\ell,m,s}\}_{\ell,m,s}$ and $\{B_{p,u,v}\}_{p,u,v}$ of the volume and distribution of rotations. These relationships will be used to determine $\{A_{\ell,m,s}\}_{\ell,m,s}$ and $\{B_{p,u,v}\}_{p,u,v}$ via solving a nonlinear least-squares problem.

To this end, we first register a crucial relationship between the coefficients of the 2-D images and the 3-D volume. By indexing the images in terms of $R \in \mathrm{SO}(3)$ (instead of j in (14)), we have:

$$\widehat{I_R}(\kappa,\varphi) = \sum_{q=-Q}^{Q} \sum_{t=1}^{T(q)} a^R_{q,t} f_{q,t}(\kappa) e^{iq\varphi}. \tag{15}$$

On the other hand, using the Fourier slice theorem and (11):

$$\widehat{I_R}(\kappa,\varphi) = R^T \cdot \widehat{\phi}(\kappa, \frac{\pi}{2}, \varphi) \tag{16}$$

$$= \sum_{\ell=0}^{L} \sum_{s=1}^{S(\ell)} \sum_{m=-\ell}^{\ell} A_{\ell,m,s}\, F_{\ell,s}(\kappa)\; R^T \cdot Y^m_\ell(\frac{\pi}{2}, \varphi) \tag{17}$$

$$= \sum_{\ell=0}^{L} \sum_{s=1}^{S(\ell)} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} A_{\ell,m,s}\, F_{\ell,s}(\kappa)\, U^\ell_{m,m'}(R)\, Y^{m'}_\ell(\frac{\pi}{2}, \varphi). \tag{18}$$

Multiplying (15) and (16) by $f_{q,t}(\kappa)e^{-iq\varphi}$ and integrating against $\frac{1}{2\pi}\kappa d\kappa d\varphi$, then combining the orthogonality relation

$$\frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} f_{q_1,t_1}(\kappa) e^{iq_1\varphi} f_{q_2,t_2}(\kappa) e^{-iq_2\varphi} d\varphi \kappa d\kappa \;=\; \mathbb{1}_{q_1=q_2}\, \mathbb{1}_{t_1=t_2}$$

with $Y^{m'}_\ell(\frac{\pi}{2},\varphi) \propto e^{im'\varphi}$, tells us

$$a^R_{q,t} = \sum_{\ell=|q|}^{L} \sum_{s=1}^{S(\ell)} \sum_{m=-\ell}^{\ell} A_{\ell,m,s}\, U^\ell_{m,q}(R)\, \gamma^{q,t}_{\ell,s}, \tag{19}$$

where $\gamma_{\ell,s}^{q,t}$ are constants depending on the radial functions:

$$\gamma_{\ell,s}^{q,t} := \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} Y_\ell^q(\frac{\pi}{2},\varphi)\, e^{-iq\varphi}\, F_{\ell,s}(\kappa)\, f_{q,t}(\kappa)\kappa d\kappa d\varphi \tag{20}$$

$$= \frac{1}{2\pi}\sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-q)!}{(\ell+q)!}}\, P_\ell^q(0)\int_0^\infty F_{\ell,s}(\kappa)\, f_{q,t}(\kappa)\kappa d\kappa. \tag{21}$$

From the term $P_\ell^q(0)$, we see $\gamma_{\ell,s}^{q,t} = 0$ if $q \not\equiv \ell \pmod 2$ (and if $|q| > \ell$ then $\gamma_{\ell,s}^{q,t} := 0$). Also one may check $\gamma_{\ell,s}^{-q,t} = (-1)^q \gamma_{\ell,s}^{q,t}$. Equation (19) connects 2-D image coefficients with 3-D volume coefficients. We note we may as well choose $Q = L$ in (15), since if $|q| > L$ then $a_{q,t}^R = 0$. In practice, the coefficients $\gamma_{\ell,s}^{q,t}$ are calculated via numerical integration over a closed segment, according to the localization property of the PSWFs, see Appendix A and [39].

2.3.1. *The first moment.* In this section, from (19) the relationship between the first moment of the images and the volume is derived. Taking the expectation over R, and using the distribution expansion (12), we get

$$\mathbb{E}_R[a_{q,t}^R] = \sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}\sum_{m=-\ell}^{\ell} A_{\ell,m,s}\gamma_{\ell,s}^{q,t}\int U_{m,q}^\ell(R)\rho(R)dR \tag{22}$$

$$= \sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}\sum_{m=-\ell}^{\ell}\sum_{p=0}^{P}\sum_{u,v=-p}^{p} A_{\ell,m,s}\, B_{p,u,v}\, \gamma_{\ell,s}^{q,t}\int U_{m,q}^\ell(R)U_{u,v}^p(R)dR \tag{23}$$

$$= \sum_{\ell=|q|}^{\min(L,P)}\sum_{s=1}^{S(\ell)}\sum_{m=-\ell}^{\ell} A_{\ell,m,s}\; B_{\ell,-m,-q}\; \gamma_{\ell,s}^{q,t}\; \frac{(-1)^{m+q}}{2\ell+1}. \tag{24}$$

The last equation follows from the orthogonality of the Wigner matrix entries [11, p. 68]

$$\int_{\mathrm{SO}(3)} \overline{U_{m,n}^\ell(R)}U_{u,v}^p(R)\, dR = \frac{1}{2\ell+1}\; \mathbb{1}_{\ell=p}\; \mathbb{1}_{u=m}\; \mathbb{1}_{v=n}, \tag{25}$$

and

$$\overline{U_{u,v}^p(R)} = (-1)^{u+v}\, U_{-u,-v}^p(R). \tag{26}$$

The first moment gives a set of bilinear forms in the unknowns $\{A_{\ell,m,s}\}_{\ell,m,s}$ and $\{B_{p,u,v}\}_{p,u,v}$, as seen in (24) for each (q,t) with $|q| \le \min(L,P)$ and $1 \le t \le T(q)$.

It is convenient to provide compact notation for the first moment formula. To this end, we introduce:

1. $\mathcal{A}_\ell$, a matrix of size $S(\ell)\times(2\ell+1)$ given by $(\mathcal{A}_\ell)_{s,m} = A_{\ell,m,s}$
2. β_ℓ^q, a vector of size $2\ell+1$ given by $(\beta_\ell^q)_m = \frac{(-1)^m}{2\ell+1}B_{\ell,-m,-q}$
3. Γ_ℓ^q, a matrix of size $T(q)\times S(\ell)$ given by $(\Gamma_\ell^q)_{t,s} = (-1)^q\gamma_{\ell,s}^{q,t}$.

Item 2 is zero if $\ell < |q|$ and item 3 is zero if either $\ell < |q|$ or $\ell \not\equiv q \,(\mathrm{mod}\ 2)$. In this notation, the first moment formula (24) (with fixed q and varying t) reads:

$$m_1(q) \;:=\; \left(\mathbb{E}[a_{q,t}^R]\right)_{t=1,\ldots,T(q)} \;=\; \sum_{\substack{\ell\,:\,|q|\le\ell\le L\\ \ell\equiv q\,(\mathrm{mod}\ 2)}} \Gamma_\ell^q\,\mathcal{A}_\ell\,\beta_\ell^q. \tag{27}$$

Here $m_1(q) \in \mathbb{C}^{T(q)}$ is nonzero only if $|q| \le \min(L,P)$.

2.3.2. *The second moment.* Higher moments require higher powers of the image coefficients, and so in the case of the second moment and for $|q_1|, |q_2| \leq L$, we have

$$\mathbb{E}_R\left[a^R_{q_1,t_1} a^R_{q_2,t_2}\right] = \sum_{\ell_1=|q_1|}^{L} \sum_{s_1=1}^{S(\ell_1)} \sum_{m_1=-\ell_1}^{\ell_1} \sum_{\ell_2=|q_2|}^{L} \sum_{s_2=1}^{S(\ell_2)} \sum_{m_2=-\ell_2}^{\ell_2} A_{\ell_1,m_1,s_1} \gamma^{q_1,t_1}_{\ell_1,s_1} \tag{28}$$

$$\times \; A_{\ell_2,m_2,s_2} \gamma^{q_2,t_2}_{\ell_2,s_2} \int U^{\ell_1}_{m_1,q_1}(R) U^{\ell_2}_{m_2,q_2}(R) \rho(R) dR \tag{29}$$

where

$$\int U^{\ell_1}_{m_1,q_1}(R) U^{\ell_2}_{m_2,q_2}(R) \rho(R) dR = \sum_{p=0}^{P} \sum_{u,v=-p}^{p} B_{p,u,v} \int U^{\ell_1}_{m_1,q_1}(R) U^{\ell_2}_{m_2,q_2}(R) U^{p}_{u,v}(R) dR. \tag{30}$$

The product of two Wigner matrix entries is expressed as a linear combination of Wigner matrix entries [19, p. 351],

$$U^{\ell_1}_{m_1,q_1}(R) U^{\ell_2}_{m_2,q_2}(R) = \sum_{\ell_3=|\ell_2-\ell_1|}^{\ell_1+\ell_2} \mathcal{C}_{\ell_3}(\ell_1,\ell_2,m_1,m_2,q_1,q_2)\, U^{\ell_3}_{m_1+m_2,n_1+n_2}(R), \tag{31}$$

where

$$\mathcal{C}_{\ell_3}(\ell_1,\ell_2,m_1,m_2,q_1,q_2) = C(\ell_1,m_1;\ell_2,m_2|\ell_3,m_1+m_2) C(\ell_1,q_1;\ell_2,q_2|\ell_3,q_1+q_2), \tag{32}$$

is the product of two Clebsch-Gordan coefficients. This product is nonzero only if (ℓ_1,ℓ_2,ℓ_3) satisfy the triangle inequalities. Substituting (31) into (30), and invoking (25) and (26), we obtain:

$$\int U^{\ell_1}_{m_1,q_1}(R) U^{\ell_2}_{m_2,q_2}(R) \rho(R) dR = \sum_{p} \mathcal{C}_p(\ell_1,\ell_2,m_1,m_2,q_1,q_2)$$

$$\times \; B_{p,-m_1-m_2,-q_1-q_2} \frac{(-1)^{m_1+m_2+q_1+q_2}}{2p+1} \tag{33}$$

where the sum is over p satisfying $\max(|\ell_1-\ell_2|, |m_1+m_2|, |q_1+q_2|) \leq p \leq \min(\ell_1+\ell_2, P)$. Now substituting into (28) gives:

$$\mathbb{E}_R\left[a^R_{q_1,t_1} a^R_{q_2,t_2}\right] = \sum_{\ell_1,s_1,m_1,\ell_2,s_2,m_2} A_{\ell_1,m_1,s_1} A_{\ell_2,m_2,s_2} \gamma^{q_1,t_1}_{\ell_1,s_1} \gamma^{q_2,t_2}_{\ell_2,t_2} (-1)^{q_1+q_2} \times$$

$$\sum_{p} B_{p,-m_1-m_2,-q_1-q_2} \mathcal{C}_p(\ell_1,\ell_2,m_1,m_2,q_1,q_2) \frac{(-1)^{m_1+m_2}}{2p+1} \tag{34}$$

where the first sum has the range of (28) and the second sum has range of (33). The second moment thus gives a set of polynomials in unknowns $\{A_{\ell,m,s}\}_{\ell,m,s}$ and $\{B_{p,u,v}\}_{p,u,v}$, quadratic in the volume coefficients and linear in the distribution coefficients, namely, the expression in (34) for each (q_1,t_1,q_2,t_2) with $|q_1| \leq L$, $|q_2| \leq L$, $|q_1+q_2| \leq P$, $1 \leq t_1 \leq T(q_1)$ and $1 \leq t_2 \leq T(q_2)$. Also, it may be assumed that $P \leq 2L$, since $B_{p,u,v}$ with $p > 2L$ does not contribute in either (34) or (24).

As for the first moment, it will be convenient to rewrite the second moment in compact notation. Let us further introduce:

4. $\mathcal{B}^{q_1,q_2}_{\ell_1,\ell_2}$, a matrix of size $(2\ell_1+1)\times(2\ell_2+1)$ given by

$$(\mathcal{B}^{q_1,q_2}_{\ell_1,\ell_2})_{m_1,m_2} = \sum_p B_{p,-m_1-m_2,-q_1-q_2}\mathcal{C}_p(\ell_1,\ell_2,m_1,m_2,q_1,q_2)\tfrac{(-1)^{m_1+m_2}}{2p+1},$$

where the sum is over $\max(|\ell_1-\ell_2|,|m_1+m_2|,|q_1+q_2|)\le p\le \min(\ell_1+\ell_2,P)$ and C_p denotes the product Clebsch-Gordan coefficients in (32).

Item 4 is zero if either $\ell_1<|q_1|$ or $\ell_2<|q_2|$ or $\max(|\ell_1-\ell_2|,|q_1+q_2|)>P$. Now for fixed q_1,q_2 and varying t_1,t_2, the second moment (34) neatly reads:

$$m_2(q_1,q_2) := \left(\mathbb{E}[a^R_{q_1,t_1}a^R_{q_2,t_2}]\right)_{\substack{t_1=1,\dots,T(q_1)\\ t_2=1,\dots,T(q_2)}} = \sum_{\substack{\ell_1,\ell_2\,:\,|q_1|\le\ell_1\le L\\ |q_2|\le\ell_2\le L\\ \ell_1\equiv q_1 \pmod 2\\ \ell_2\equiv q_2\pmod 2\\ |\ell_1-\ell_2|\le P}} \Gamma^{q_1}_{\ell_1}\,\mathcal{A}_{\ell_1}\,\mathcal{B}^{q_1,q_2}_{\ell_1,\ell_2}\,\mathcal{A}^T_{\ell_2}\,(\Gamma^{q_2}_{\ell_2})^T. \tag{35}$$

Here $m_2(q_1,q_2)\in\mathbb{C}^{T(q_1)\times T(q_2)}$ is nonzero only if $|q_1|,|q_2|\le L$ and $|q_1+q_2|\le P$.

3. **Uniqueness Guarantees and Conditioning.** Here, we derive uniqueness guarantees and comment on intrinsic conditioning for the polynomial system defined by the first and second moments, (27) and (35).

Analysis comes in four cases, according to assumptions on the distribution ρ: whether ρ is known or unknown; and if ρ is invariant to in-plane rotations, i.e., ρ depends only on the viewing directions up to rotations that retain the z-axis. This invariance restricts ρ to be a non-uniform distribution function over S^2, see subsection 4.2. If ρ is not invariant to in-plane rotations, we say ρ is totally non-uniform as a distribution on the entire SO(3). Throughout, our general finding is *well-posedness*, *i.e.*, the molecule is uniquely determined by first and second moments up to finitely many solutions, under genericity assumptions, for a range of band limits L and P. In the case of a known totally non-uniform distribution, we prove the number of solutions is 1, and give an efficient, explicit algorithm to solve for $\{A_{\ell,m,s}\}$. For all cases, sensitivity of the solution to errors in the moments is quantified by condition number formulas.

3.1. **Known, totally non-uniform ρ.** For this case, we have a provable algorithm that recovers $\{A_{\ell,m,s}\}$ from (27) and (35) (up to the satisfaction of technical genericity and band limit conditions). Remarkably, while the polynomial system is nonlinear (consisting of both quadratic and linear equations), our method is based only on linear algebra. The main technical idea is *simultaneous diagonalization* borrowed from Jennrich's well-known algorithm for third-order tensor decomposition [31], that was also used recently for signal recovery in MRA [46].

Theorem 1. *The molecule $\{A_{\ell,m,s}\}$ is uniquely determined by the analytical first and second moments, (27) and (35), in the case the distribution $\{B_{p,u,v}\}$ is* totally non-uniform, known *and $P\ge 2L$, provided it also holds:*

(i) The matrices $B_1 := \mathcal{B}^{L,L}_{L,L}$ and $B_2 = \mathcal{B}^{L,-L}_{L,L}$ of size $(2L+1)\times(2L+1)$ both have full rank, and $B_1B_2^{-1}$ has distinct eigenvalues. Likewise $B_3 := \mathcal{B}^{L-1,L-1}_{L-1,L-1}$ and $B_4 = \mathcal{B}^{L-1,1-L}_{L-1,L-1}$ of size $(2L-1)\times(2L-1)$ both have full rank, and $B_3B_4^{-1}$ has distinct eigenvalues.

(ii) *Writing* $B_1 B_2^{-1} =: Q_{12} D_{12} Q_{12}^{-1}$ *and* $B_3 B_4^{-1} =: Q_{34} D_{34} Q_{34}^{-1}$ *for eigendecompositions, the vectors* $b_{12} := Q_{12}^{-1} \beta_L^L$ *of size* $2L+1$ *and* $b_{34} := Q_{34}^{-1} \beta_{L-1}^{L-1}$ *of size* $2L-1$ *both have no zero entries.*

(iii) *For* $\ell \leq L-2$*, the matrix* $\mathcal{B}_{\ell,L}^{\ell,L}$ *of size* $(2\ell+1) \times (2L+1)$ *has full row rank.*

(iv) *For all* ℓ*, the matrix* $\mathcal{A}_\ell$ *of size* $S(\ell) \times (2\ell+1)$ *has full column rank.*

(v) *For* $\ell \geq |q|$ *with* $\ell \equiv q \pmod 2$*, the matrix* Γ_ℓ^q *of size* $T(q) \times S(\ell)$ *has full column rank.*

Moreover in this case, there is a provable algorithm inverting (27) *and* (35) *to get* $\{A_{\ell,m,s}\}$ *in time* $\mathcal{O}\left(L^2 \cdot T^3\right)$*, where* $T := \max_q T(q)$*.*

Proof. For this proof, we need some general properties of the *Moore-Penrose pseudo-inverse*, denoted by †, as in [8]. In particular, if $Y \in \mathbb{C}^{n_1 \times n_2}$ has full column rank and $Z \in \mathbb{C}^{n_2 \times n_3}$ has full row rank, then $Y^\dagger Y = I_{n_2}$, $ZZ^\dagger = I_{n_2}$, $(YZ)^\dagger = Z^\dagger Y^\dagger$, and also, pseudo-inversion and transposition commute.

Proceeding, the second moment with $q_1 = L, q_2 = L$ tells us:

$$m_2(L,L) \;=\; \Gamma_L^L \mathcal{A}_L B_1 (\mathcal{A}_L)^T (\Gamma_L^L)^T \;\in\; \mathbb{C}^{T(L) \times T(L)}, \tag{36}$$

and with $q_1 = L, q_2 = -L$:

$$m_2(L,-L) \;=\; \Gamma_L^L \mathcal{A}_L B_2 (\mathcal{A}_L)^T (\Gamma_L^{-L})^T \;\in\; \mathbb{C}^{T(L) \times T(L)}, \tag{37}$$

where $\Gamma_L^L = (-1)^L \Gamma_L^{-L}$. We compute $(-1)^L$ times the Moore-Penrose psuedo-inverse of (37) and then multiply this on the right of (36). Because Γ_L^L and $\mathcal{A}_L$ are each tall with full column rank by assumptions (v) and (iv), respectively, and B_2 is invertible by (i), properties of the pseudo-inverse imply:

$$\begin{aligned}
(-1)^L m_2(L,L) m_2(L,-L)^\dagger &= \left(\Gamma_L^L \mathcal{A}_L B_1 (\mathcal{A}_L)^T (\Gamma_L^L)^T\right) \left(\Gamma_L^L \mathcal{A}_L B_2 (\mathcal{A}_L)^T (\Gamma_L^L)^T\right)^\dagger \\
&= (\Gamma_L^L \mathcal{A}_L) B_1 (\mathcal{A}_L)^T (\Gamma_L^L)^T (\Gamma_L^L)^{T\dagger} (\mathcal{A}_L)^{T\dagger} B_2^{-1} (\Gamma_L^L \mathcal{A}_L)^\dagger \\
&= (\Gamma_L^L \mathcal{A}_L) B_1 B_2^{-1} (\Gamma_L^L \mathcal{A}_L)^\dagger \\
&= (\Gamma_L^L \mathcal{A}_L) Q_{12} D_{12} Q_{12}^{-1} (\Gamma_L^L \mathcal{A}_L)^\dagger \\
&= \left(\Gamma_L^L \mathcal{A}_L Q_{12}\right) D_{12} \left(\Gamma_L^L \mathcal{A}_L Q_{12}\right)^\dagger,
\end{aligned} \tag{38}$$

where we have substituted in an eigendecomposition $B_1 B_2^{-1} = Q_{12} D_{12} Q_{12}^{-1}$. As $B_1 B_2^{-1}$ has distinct eigenvalues by condition (i), we see that the eigenvectors of $(-1)^L m_2(L,L) m_2(L,-L)^\dagger$ are unique up to scale and given as the columns of $\Gamma_L^L \mathcal{A}_L Q_{12}$. Thus, $\Gamma_L^L \mathcal{A}_L Q_{12} = X\Lambda$, where X consists of eigenvectors of (38) and Λ is an unknown (as yet) diagonal matrix.

To disambiguate the scales Λ, we compare with the first moment for $q = L$:

$$m_1(L) = \Gamma_L^L \mathcal{A}_L \beta_L^L = X\Lambda Q_{12}^{-1} \beta_L^L = X\Lambda b_{12}. \tag{39}$$

Multiplying on the left by $X^\dagger$ gives $X^\dagger m_1(L) = \Lambda b_{12}$, an equality of matrix-vector products in which the only unknown is the diagonal matrix Λ. By the full support of b_{12} (assumption (ii)), this determines Λ. Substituting into $X\Lambda$, we now know $\Gamma_L^L \mathcal{A}_L Q_{12}$. Multiplying on the left by $\Gamma_L^{L\dagger}$ and on the right by Q_{12}^{-1} tells us $\mathcal{A}_L$.

Backward marching, the second moment with $q_1 = L-2$ and $q_2 = L$ reads:

$$m_2(L-2,L) = \Gamma_L^{L-2} \mathcal{A}_L B_{L,L}^{L-2,L} (\mathcal{A}_L)^T (\Gamma_L^L)^T \;+\; \Gamma_{L-2}^{L-2} \mathcal{A}_{L-2} B_{L-2,L}^{L-2,L} (\mathcal{A}_L)^T (\Gamma_L^L)^T. \tag{40}$$

At this point, we know the first term, and thus the second term gives us $\mathcal{A}_{L-2}$ by appropriately multiplying by pseudo-inverses ($B_{L,L}^{L-2,L}$ is right-invertible by (iii)).

Then, we may look at the second moments with $q_1 = L-4$ and $q_2 = L$ to similarly determine $\mathcal{A}_{L-4}$, and so on, to $\mathcal{A}_0$ or $\mathcal{A}_1$ (depending on the parity of L). Analogous reasoning and usage of the assumptions gives $\mathcal{A}_{L-1}, \mathcal{A}_{L-3}, \ldots$

We have provided an algorithm to solve for each $\mathcal{A}_\ell$, which proves uniqueness of $\mathcal{A}_\ell$ as a byproduct. The time complexity of the algorithm is $\mathcal{O}(L^2T^3)$ since it involves $\mathcal{O}(L^2)$ matrix operations –matrix multiplications, pseudo-inversions or eigendecompositions – of matrices whose dimensions are all bounded by T. (Note that back-substituting to solve for $\mathcal{A}_\ell$ involves $\mathcal{O}(L-\ell)$ such matrix operations.) □

We remark that condition (iv), which just involves the choice of radial bases, appears to always hold for PSWFs using the cutoffs proposed in Appendix A. Conditions (i), (ii) and (iii) just involve the distribution, and are full-rank, spectral and non-vanishing hypotheses. Condition (iv) just involves the molecule and in particular requires $S(L) \geq 2L+1$, which limits L to be less than the Nyquist frequency where $S(L_{\text{Nyquist}}) = 1$.

Our algorithm goes by *reverse*[6] frequency marching, as we solve for top-frequency coefficients from the second moment (35) where $q_1, q_2 = \pm L, \pm(L-1)$ via eigenvectors (similar to simultaneous diagonalization in Jennrich's algorithm), and then solving for lower-frequency coefficients via linear systems. While our conditions in Theorem 1 are certainly not necessary, fortunately for generic[7] (A, B), those conditions are satisfied, so that the method applies:

Lemma 2. *Condition (ii) in Theorem 1 holds for Zariski-generic* $\{B_{p,u,v}\}$. *If* $S(L) \geq 2L+1$, *then condition (iii) holds for Zariski-generic* $\{A_{\ell,m,s}\}$. *At least for* $L \leq 100$, *conditions (i) and (iii) hold for Zariski-generic* $\{B_{p,u,v}\}$.

Proof of Lemma 2. Conditions (i)-(iv) are all Zariski-open, *i.e.*, their failure implies $\{A_{\ell,m,s}\}$ or $\{B_{p,u,v}\}$ obey polynomial equations. As such, to conclude genericity, it suffices to exhibit a single point $\{A_{\ell,m,s}\}$ or $\{B_{p,u,v}\}$, where the conditions are met. For conditions (i), (iii), we verified the conditions hold at randomly selected points on computer up to $L \leq 100$. Conditions (ii) and (iv) are obviously generic. □

By uniqueness, A is a well-defined function of the first and second moments m_1 and m_2 almost everywhere. It is useful to quantify the "sensitivity" of A to errors in m_1, m_2, as, *e.g.*, in practice one can access only empirical estimates $\widetilde{m}_1$ and $\widetilde{m}_2$. An *a posteriori* (absolute) condition number for A is given by the reciprocal of the least singular value of the Jacobian matrix of the algebraic map:

$$m_B : \{A_{\ell,m,s}\} \mapsto \big\{m_1(q),\, m_2(q_1,q_2)\big\}. \tag{41}$$

Throughout this section, all condition formulas are in the sense of [16, Section 14.3], for which the domain and image of our moment maps are viewed as Riemannian manifolds. To this end, when ρ is unknown, dense open subsets of the orbit spaces $\{(A,B) \bmod \mathrm{SO}(3)\}$, $\{A \bmod \mathrm{SO}(3)\}$, $\{B \bmod \mathrm{SO}(3)\}$ naturally identify as Riemannian manifolds (for the construction, see [15]).

[6] Reverse frequency marching is natural given the sparsity structure of (35): only $\mathcal{A}_{\ell_1}$ and $\mathcal{A}_{\ell_2}$ with $\ell_1 \geq |q_1|$, $\ell_1 \equiv q_1 \pmod 2$ and $\ell_2 \geq |q_2|$, $\ell_2 \equiv q_2 \pmod 2$ appear in the moments $m_2(q_1,q_2)$.

[7] This means generic with respect to the *Zariski topology* [30]. Equivalently, there is a non-zero polynomial p in A, B such that $p(A,B) \neq 0$ implies the conditions in Theorem 1 are met.

3.2. **Known, in-plane uniform ρ.** For this case, given a particular image size (and other image parameters), together with band limits L and P, we have code[8] which decides if, for generic A and B, the molecule A is determined by (27) and (35), up to finitely many solutions. The basis for this code is the so-called *Jacobian test* for algebraic maps, see Appendix B. Below is an illustrative computation.

Computational Result 3. Consider 43×43 pixel images, and the following parameters for prolates (representative values): a bandlimit c (see Appendix A) chosen as the Nyquist frequency, 2-D prescribed accuracy (95) set to $\epsilon = 10^{-3}$ and 3-D truncation parameter (75) to be $\delta = 0.99$ [9]. We varied band limits L in (8) and P in (12), and randomly fixed (12) to give a known in-plane uniform distribution. For each (L, P), we computed the numerical rank of the Jacobian matrix of the polynomial map m_B of (41) at a randomly chosen A, with random B. The Jacobian was convincingly of full numerical rank for a variety of band limits, as seen in Table 1. Cases where the gap between the two least singular values of the Jacobian matrix exceeds a threshold of 10^6 are set as indecisive numerics, and appears in the table as **?**. Note that if the rank was calculated in exact arithmetic, this gives a proof that for generic (A, B) generic fibers of the map m_B consist of finitely many A; *i.e.*, first and second moments (with known in-plane uniform distribution) determine the molecule up to finitely many solutions. For fibers and related definitions, see Appendix B.

Again, the sensitivity of A as a locally defined function of (27) and (35) is quantified by the reciprocal of the least singular value of the Jacobian matrix of m_B.

3.3. **Unknown, totally non-uniform ρ.** In this case, it is important to note that solutions come in symmetry classes. If (A, B) have specified moments, then so too for $(R \cdot A, R \cdot B)$ for all $R \in \mathrm{SO}(3)$, that is, we may jointly rotate the molecule and probability distribution and the moments are left invariant. So, solutions come in 3-dimensional equivalence classes, and we are interested in solutions modulo $\mathrm{SO}(3)$.

That said, we have code which accepts a particular image size (and other image parameters), together with band limits L and P. The code then numerically decides which of the following situations occur: *i)* for generic (A, B), both A and B are determined by (27) and (35) up to finitely many solutions modulo $\mathrm{SO}(3)$; *ii)* for generic (A, B), the molecule A is determined by (27) and (35) up to finitely many solutions modulo $\mathrm{SO}(3)$, whereas the distribution B has infinitely many solutions; *iii)* for generic (A, B), both A and B have infinitely many solutions modulo $\mathrm{SO}(3)$. Note these cases are (essentially) exhaustive, since if B is determined so is A in the regime of Theorem 1. Moreover, we noticed the case *ii)* really does arise, *e.g.*, this seems to happen when $P = 2L$.

Computational Result 4. We keep the running example of 43×43 pixel images, and the prolates parameters of a bandlimit c chosen as the Nyquist frequency, 2-D prescribed accuracy (95) set to $\epsilon = 10^{-3}$ and 3-D truncation parameter (75) of $\delta = 0.99$. We varied band limits L in (8) and P in (12). For each (L, P), we computed the numerical rank of the Jacobian matrix of the polynomial map

$$m : \{A_{\ell,m,s}, B_{p,u,v}\} \mapsto \big\{m_1(q),\, m_2(q_1, q_2)\big\}. \tag{42}$$

[8] Available in GitHub: https://github.com/nirsharon/nonuniformMoM/JacobianTest.

[9] The value of δ means we allow only 1% of the energy to be outside the ball, and is chosen to best model a molecule structure which is assumed to be mostly supported inside a ball.

at a randomly chosen point in the domain. The numerical rank of the Jacobian convincingly equaled three less (that is $d_1 = 3$, see Appendix B) than full column rank for a variety of band limits, see Table 2. Cases where the gap between the third and fourth least singular values of the Jacobian matrix exceeds a threshold of 10^6 are set as indecisive numerics, and appears in the table as **?**. If the rank were calculated in exact arithmetic, this furnishes a proof that generic fibers of the map m consist of finitely many SO(3)-orbits; that is, first and second moments determine both the molecule and the totally non-uniform distribution up to finitely many solutions (modulo global rotation).

For band limits L and P such that generically there are only finitely many solutions for (A, B) mod SO(3), the sensitivity of (A, B) mod SO(3) as a (locally defined) function of (27) and (35) is quantified by the reciprocal of the fourth least singular of m. For band limits such that generically there are only finitely many solutions for A mod SO(3), the sensitivity of A mod SO(3) as a locally defined of (27) and (35) is quantified by the reciprocal of the fourth least singular value of

$$\mathcal{P}_A \operatorname{Jac}(m|_{(A,B)})^\dagger \tag{43}$$

where † denotes pseudo-inverse and $\mathcal{P}_A$ is the differential of $(A, B) \mapsto A$ mod SO(3). We compute (43) by analytically differentiating (27) and (35), evaluating at (A, B) and place as diagonal blocks of a matrix, and finally applying pseudo-inverse which is SVD-based.

3.4. **Unknown, in-plane uniform ρ.** Again in this case, solutions come in 3-symmetry classes, orbits under the action of global rotation, so we are interested in solutions modulo SO(3). We have code which accepts a particular image size (and other image parameters), together with band limits L and P, and numerically decides if for generic (A, B), both A and B are determined by (27) and (35) up to finitely many solutions modulo SO(3), or if there are infinitely many solutions. We did not find parameters giving a "mixed" result as in case *ii)* above.

Computational Result 5. For 43×43 pixel images, and the parameters for prolates (representative values): a bandlimit c chosen as the Nyquist frequency, 2-D prescribed accuracy (95) set to $\epsilon = 10^{-3}$ and 3-D truncation parameter (75) of $\delta = 0.99$. We varied band limits L in (8) and P in (12), restricting (12) to an in-plane uniform distribution. For each (L, P), we computed the numerical rank of the Jacobian matrix of the polynomial map:

$$m : \{A_{\ell,m,s}, B_{p,u,0}\} \mapsto \big\{m_1(q), m_2(q_1, q_2)\big\}. \tag{44}$$

at a randomly chosen point in the domain. The numerical rank of the Jacobian convincingly equaled three less than full column rank for a variety of band limits, see Table 3. Cases where the gap between the third and fourth least singular values of the Jacobian matrix exceeds a threshold of 10^6 are set as indecisive numerics, and appears in the table as **?**. If the rank was calculated in exact arithmetic, this furnishes a proof that generic fibers of the map m consist of finitely many SO(3)-orbits; that is, first and second moments determine both the molecule and the in-plane uniform distribution up to finitely many solutions (modulo global rotation).

For band limits L and P such that generically there are only finitely many solutions for (A, B) mod SO(3), the sensitivity of (A, B) mod SO(3) as a function of moments is quantified by the reciprocal of the fourth least singular of m. For

example, in the $P = 2$ row of Table 3, when evaluating at random (A, B), this worked out to:

$$1.98\times10^{15}, \quad 47.1, \quad 209, \quad 2700, \quad 4.66\times10^{4}, \quad 1.17\times10^{6}, \quad 6.02\times10^{7}, \quad 9.10\times10^{8}.$$

Further, in the $L = 4$ column of Table 1, evaluating at random (A, B) gave:

$$1.44 \times 10^{16}, \quad 2.15 \times 10^{15}, \quad 209, \quad 154, \quad 1360.$$

In practice, we run this refined Jacobian test (takes $<$ 1 minute on a standard laptop) to identify well-conditioned band limits L and P before we attempt non-convex optimization.

TABLE 1. *Uniqueness for inverting the first two moments in the case of a known, in-plane uniform ρ, according to band limits.* Generically finitely many solutions for A is denoted by ✓, infinitely many solutions for A is denoted by ✗, and indecisive numerics is denoted by ❓.

	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$	$L=7$	$L=8$	$L=9$	$L=10$
$P=0$	✗	✗	✗	✗	✗	✗	✗	✗	✗
$P=1$	✗	✗	❓	✓	✓	✓	✓	✓	✓
$P=2$	✗	✓	✓	✓	✓	✓	✓	✓	✓
$P=3$	✗	✓	✓	✓	✓	✓	✓	✓	✓
$P=4$	✗	✓	✓	✓	✓	✓	✓	✓	✓

TABLE 2. *Uniqueness for inverting the first two moments in the case of an unknown, totally non-uniform ρ, according to band limits.* Generically finitely many solutions for (A, B) mod SO(3) is denoted by ✓, finitely many solutions for A mod SO(3) but infinitely many solutions for B mod SO(3) is denoted by ∼, infinitely many solutions for A mod SO(3) is denoted by ✗, and indecisive numerics is denoted by ❓.

	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$	$L=7$	$L=8$	$L=9$	$L=10$
$P=0$	✗	✗	✗	✗	✗	✗	✗	✗	✗
$P=1$	✓	✓	✓	✓	✓	✓	✓	✓	❓
$P=2$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$P=3$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$P=4$	∼	✓	✓	✓	✓	✓	✓	✓	✓

4. **Numerical Optimization and First Visual Examples.** After studying the theoretical properties of the polynomial system which is defined by the first two moments, we discuss in this section aspects of numerically inverting the polynomial map via optimization.

TABLE 3. *Uniqueness for inverting the first two moments in the case of an unknown, in-plane uniform ρ, according to band limits.* Generically finitely many solutions for (A, B) mod SO(3) is denoted by ✓, infinitely many solutions for A mod SO(3) and B mod SO(3) is denoted by ✗, and indecisive numerics is denoted by **?**.

	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$	$L=7$	$L=8$	$L=9$	$L=10$
$P=0$	✗	✗	✗	✗	✗	✗	✗	✗	✗
$P=1$	✗	✗	✗	✗	✗	✗	✗	✗	✗
$P=2$	✗	✓	✓	✓	✓	✓	**?**	**?**	**?**
$P=3$	✗	✓	✓	✓	✓	✓	✓	**?**	**?**
$P=4$	✗	✓	✓	✓	✓	✓	✓	**?**	**?**

4.1. **Incorporating natural constraints in optimization.** When determining the coefficients $A = \{A_{\ell,m,s}\}_{\ell,m,s}$ and $B = \{B_{p,u,v}\}_{p,u,v}$, the search space has to be restricted in order to ensure the coefficients stem from some physical volume and density.

4.1.1. *Constraints on the volume.* To ensure the volume $\phi\colon \mathbb{R}^3 \to \mathbb{R}$ is a real-valued function, one has to ensure its Fourier transformation $\widehat{\phi} : \mathbb{R}^3 \to \mathbb{C}$ satisfies conjugate symmetry $\hat{\phi}(\kappa,\theta,\varphi) = \bar{\hat{\phi}}(\kappa,\pi-\theta,\pi+\varphi)$. That is, in spherical coordinates,

$$\sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\sum_{s=1}^{S(\ell)}\overline{A_{\ell,m,s}Y_\ell^m(\theta,\varphi)F_{\ell,s}(\kappa)} = \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\sum_{s=1}^{S(\ell)}A_{\ell,m,s}Y_\ell^m(\pi-\theta,\pi+\varphi)F_{\ell,s}(\kappa).$$

Assuming the basis $\{F_{\ell,s}\}$ is a set of real-valued functions, along with the facts that $\overline{Y_\ell^m(\theta,\varphi)} = (-1)^m Y_\ell^{-m}(\theta,\varphi)$ and $Y_\ell^m(\pi-\theta,\pi+\phi) = (-1)^\ell Y_\ell^m(\theta,\phi)$, we get

$$\sum_{\ell,m,s}\overline{A_{\ell,-m,s}}(-1)^{-m}Y_\ell^{-m}(\theta,\varphi)F_{\ell,s} = \sum_{\ell,m,s}A_{\ell,m,s}(-1)^\ell Y_\ell^m(\theta,\varphi)F_{\ell,s}$$

This further implies

$$\overline{A_{\ell,m,s}}(-1)^{-m} = A_{\ell,-m,s}(-1)^\ell. \tag{45}$$

Having such relationships, $\{A_{\ell,m,s}\}_{\ell,m,s}$ can thus be written in terms of some real coefficients $\{\alpha_{\ell,m,s}\}_{\ell,m,s}$ as:

$$A_{\ell,m,s} = \begin{cases} \alpha_{\ell,m,s} - i(-1)^{l+m}\alpha_{\ell,m,s}, & m>0, \\ i^l\alpha_{\ell,m,s} & m=0, \\ (-1)^{l+m}\alpha_{\ell,m,s} + i\alpha_{\ell,m,s}, & m<0. \end{cases} \tag{46}$$

The latter means that instead of solving a complex optimization problem in terms of the coefficients $A_{\ell,m,s}$, one can work with the real coefficients $\alpha_{\ell,m,s}$ of (46). Otherwise, the equality constraints (45) are required.

4.1.2. *Constraints on the density.* Similarly, to ensure the density ρ being a real-valued function, we need to ensure

$$\sum_{p=0}^{P}\sum_{u=-p}^{p}\sum_{v=-p}^{p} B_{p,u,v}U^p_{u,v}(R) = \sum_{p=0}^{P}\sum_{u=-p}^{p}\sum_{v=-p}^{p} \overline{B_{p,u,v}U^p_{u,v}(R)}. \tag{47}$$

The fact that $\overline{U^p_{u,v}(R)} = (-1)^{v-u}U^p_{-u,-v}(R)$ leads to

$$B_{p,u,v} = (-1)^{u-v}\overline{B_{p,-u,-v}}. \tag{48}$$

Again, from such relationships, it can be shown that an alternative to (48) can be written in terms of real coefficients $\beta_{p,u,v}$:

$$B_{p,u,v} = \begin{cases} \beta_{p,u,v} + (-1)^{u-v}i\beta_{p,-u,-v}, & (u,v) \succ_{\text{lex}} (0,0), \\ \beta_{p,0,0}, & (u,v) = 0, \\ \beta_{p,u,v} - (-1)^{u-v}i\beta_{p,-u,-v}, & (u,v) \prec_{\text{lex}} (0,0). \end{cases} \tag{49}$$

Here, $\prec_{\text{lex}}$ is the lexicographical order, that is $(u_1,v_1) \prec_{\text{lex}} (u_2,v_2)$ iff $u_1 < u_2$ or both $u_1 = u_2$ and $v_1 < v_2$.

Two additional constraints are required. First, the integral of any density function is one. To ensure such a correct normalization, we simply let

$$B_{0,0,0} = \int \sum_{p=0}^{P}\sum_{u=-p}^{p}\sum_{v=-p}^{p} B_{p,u,v}U^p_{u,v}(R)dR = 1, \tag{50}$$

which means it is no longer considered as unknown. Finally, the nonnegativity of the density is ensured via a collocation method, that is requiring

$$\rho(R_i) = \sum_{p,u,v} B_{p,u,v}U^p_{u,v}(R_i) \geq 0, \tag{51}$$

for R_i's on a near uniform, refined grid on SO(3). While (51) does not prevent the density from becoming negative off the SO(3) grid, requiring the density to be non-negative entirely on SO(3) leads to an optimization problem that is much more costly to solve in practice. Note that we do not enforce positivity of ρ by requiring it to be a sum-of-squares, as, *e.g.*, already in the case of an in-plane uniform distribution on the sphere $S^2 \subset \mathbb{R}^3$, not all nonnegative polynomials may be written as a sum-of-squares, see Motzkin's example when $P = 6$ [43].

4.2. **Accommodating invariance to in-plane rotations.** While molecules typically exhibit preferred orientations, there is no physical reason why molecules should have preferred *in-plane* orientations. In this section, we focus on the case of non-uniform rotational distributions invariant to in-plane rotations since these distributions better model real cryo-EM data sets.

For simplicity, we fix the image plane as perpendicular to the z-axis. We add the prior that the density for drawing R equals the density for drawing $Rz(\alpha)$, for all $R \in \text{SO}(3)$ and all rotations $z(\alpha)$ of $\alpha \in \mathbb{R}$ radians about the z-axis. This assumption reads

$$\rho(R) = \rho\left(Rz(\alpha)\right) \quad R \in \text{SO}(3), \quad \alpha \in \mathbb{R}. \tag{52}$$

Therefore,

$$\sum_{p,u,v} B_{p,u,v}\, U^p_{uv}(R) \;=\; \sum_{p,u,v} B_{p,u,v}\, U^p_{uv}(Rz(\alpha)) \tag{53}$$

$$= \sum_{p,u,v} B_{p,u,v} \left(U^p(R) U^p\big(z(\alpha)\big) \right)_{uv} \tag{54}$$

Here we used the group representation property of U^p. Checking explicitly the action of $z(\alpha)$ on degree p spherical harmonics,

$$U^p\big(z(\alpha)\big) = \mathrm{diag}(e^{-ip\alpha}, e^{-i(p-1)\alpha}, \ldots, e^{ip\alpha}). \tag{55}$$

So continuing the above,

$$\sum_{p,u,v} B_{p,u,v}\, U^p_{uv}(R) = \sum_{p,u,v} B_{p,u,v}\, U^p_{uv}(R) e^{iv\alpha}, \quad R \in \mathrm{SO}(3), \quad \alpha \in \mathbb{R}. \tag{56}$$

This is equivalent to $B_{p,u,v} = 0$ for $v \neq 0$ where v ranges over $-p, -p+1, \ldots, p$. To sum, we have found that in-plane invariance is captured by:

$$d\rho(R) = \sum_{p,u} B_{p,u,0}\, U^p_{u0}(R)\, dR \tag{57}$$

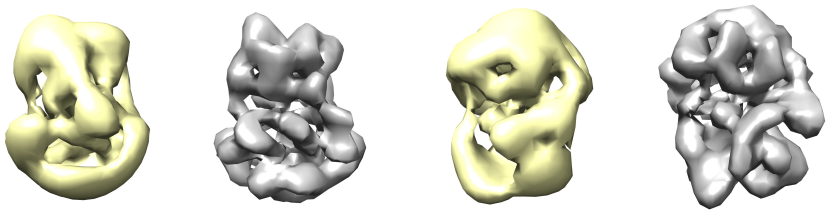
For a sanity check, a distribution with in-plane invariance should sample a rotation with density only depending on which point maps to the north pole. Namely, $\rho(R)$ should only depend on the last column of R, that is, $R(:,3) = R_{\bullet 3}$. Indeed, this holds as $U^p_{u0}(R) = (-1)^u \sqrt{\frac{4\pi}{2l+1}} \overline{Y^u_p}(R_{\bullet 3})$ [19, Eqn. 9.44, Pg. 342].

Restricting the expansion of ρ as above, we easily see the first moment is independent of φ. It is now merely a linear combination of basis functions $F_{\ell,s}(\kappa)$. Likewise, for the second moment, angular dependency is only on the difference $\varphi_1 - \varphi_2$, meaning it is a linear combination of basis functions $e^{im(\varphi_1-\varphi_2)} F_{\ell_1,s_1}(\kappa_1) F_{\ell_2,s_2}(\kappa_2)$. Thus, in subsection (3.4), we have the following polynomial map, now with fewer B variables and fewer invariants than in subsection (3.3)

$$m\colon \{A_{\ell,m,s}, B_{p,u,0}\} \mapsto \{\, \mathbb{E}_R[a^R_{0,t}], \quad \mathbb{E}_R[a^R_{q_1,t_1} a^R_{-q_1,t_2}] \,\}. \tag{58}$$

4.3. **Direct method – known totally non-uniform distribution.** For the "easy" case of a known, totally non-uniform distribution, we have implemented the provable algorithm in Theorem 1. The method's performance is illustrated by way of an example. As the ground truth volume, we use EMD-0409, that is, the catalytic subunit of protein kinase A bound to ATP and IP20 [32], as presented at the online cryo-EM data-bank [38]. The volumetric array's original dimension is 128 voxels in each direction, which we downsampled by a factor of three to 43. The volume was expanded using PSWFs with a band limit c chosen to be the Nyquist frequency and 3-D truncation parameter (75) of $\delta = 0.99$. Before downsampling, the full expansion consists of degree $L = 40$; with downsampling and proper truncation, we aim to recover the terms up to degree $L = 7$. For the known totally non-uniform distribution, we took $P = 14$ (per Theorem 1), and then formed a particular distribution using a sums-of-squares. Precisely, we formed a random linear combination of Wigner entries up to degree 7, multiplied this by its complex conjugate, invoked (26) and (32) to rewrite the result as a linear combination of Wigner entries up to degree 14, repeated for a second square, added, and finally normalized to satisfy (50). Then, with the distribution known as such, the volume contributes 1080 unknowns (without discounting for (45)). Providing the algorithm with m_1 and m_2, our method took 0.24 seconds on a standard laptop, and recovered

the unknowns A up to a relative error in $\mathcal{L}^2$ norm of 5.4×10^{-11}. Visual results are in Figure 2.



View 1 View 2

FIGURE 2. Two views of the reconstruction as provided by the algorithm of Theorem 1 to the case of known, totally non-uniform distribution. The ground truth volume appears on the right of each pair (in gray), whereas the lower degree estimation resulting from the downsampled volume appears on the left (in yellow). Note that the estimation is visually identical to the truncated volume, and it thus illustrates the effect of truncation.

4.4. **Setting up a least-squares formulation.** For the cases where we lack a direct method, we formulate the problem in terms of minimizing a least-squares cost function. First, we define the unknowns of our optimization process to be the coefficients of the volume $A = \{A_{l,m,s}\}$ and distribution $B = \{B_{p,u,v}\}$. The explicit formulas (27) and (35) provide means to write the low-order moments (7) as functions of our unknown coefficients, that is $m_1 = m_1(A, B)$ and $m_2 = m_2(A, B)$.

In practice, given data images, one estimates the low-order statistics using the empirical moments $\widetilde{m}_1$ and $\widetilde{m}_2$ of (1), but now given in PSWFs coordinates

$$(\widetilde{m}_1)_{q,t} = \frac{1}{n}\sum_{j=1}^{n} a^j_{q,t} \quad \text{and} \quad (\widetilde{m}_2)_{q_1,t_1,q_2,t_2} = \frac{1}{n}\sum_{j=1}^{n} a^j_{q_1,t_1} a^j_{q_2,t_2}, \tag{59}$$

The connection between the empirical moments and their analytical formulas as functions of our unknowns gives rise to a nonlinear least-squares

$$\min_{A,B} \sum_{q=-Q}^{Q} \sum_{t=0}^{T(q)} \left(m_1(A,B)_{q,t} - (\widetilde{m}_1)_{q,t} \right)^2 + \lambda \sum_{q_1,q_2=-Q}^{Q} \sum_{t_1,t_2=0}^{T(q)} \left(m_2(A,B)_{q_1,t_1,q_2,t_2} - (\widetilde{m}_2)_{q_1,t_1,q_2,t_2} \right)^2, \tag{60}$$

where λ is a parameter chosen to balance the errors from both terms. In particular, two main considerations determine the value of λ. First is the number of elements in each summand. Namely, the second moment includes many more entries than the first moment. Therefore, without the effect of noise, λ is set to be the ratio between the number of entries in first moment and the second moment. The second factor to balance is the different convergence rates of the empirical moments, see also [1]. The nonlinear least-squares (60) may be adjusted to incorporate the constraints on $\{A_{l,m,s}\}$ and $\{B_{p,u,v}\}$ that ensure ϕ is a real-valued volume and ρ a probability density.

We remark that it is interesting to consider pre-conditioners, or more intricate weighings, in the formation of the nonlinear least-squares cost (60). Such might alleviate high condition numbers observed in Section 3, and potentially accelerate optimization algorithms. While we have not tested a pre-conditioner in optimization experiments yet, one possibility would be to consider the following normalized cost:

$$\min_{A,B} \sum_{q=-Q}^{Q} \sum_{t=0}^{T(q)} \left(m_1(A,B)_{q,t} - (\widetilde{m}_1)_{q,t} \right)^2 \Big/ \, (\widetilde{m}_1)_{q,t}^2$$
$$+ \lambda \sum_{q_1,q_2=-Q}^{Q} \sum_{t_1,t_2=0}^{T(q)} \left(m_2(A,B)_{q_1,t_1,q_2,t_2} - (\widetilde{m}_2)_{q_1,t_1,q_2,t_2} \right)^2 \Big/ \, (\widetilde{m}_2)_{q_1,t_1,q_2,t_2}^2 \,. \quad (61)$$

Effectively, (61) scales each polynomial in (A,B) given by m_1 and m_2 to take value 1.

4.5. **Complexity analysis of inverting the moments via gradient-based optimization.** Before moving forward to further numerical examples, we state the computational load of minimizing the least-squares cost function (60). It is worth noting that in many modern *ab initio* algorithms, like SGD [48] and EM [52], the runtime of each iteration is measured with respect to the size of the set of data images, which can be huge. In our approach, we only carry out one pass over the data to collect the low-order statistics. In here, we assume the empirical moments are already given, and so the complexity of each iteration is merely a function of the size of the moments or equivalently depends on the size and resolution of the data images, as reflected by their PSWF representations.

Many possible algorithms exist to minimize the least squares problem (60), for example direct gradient descent methods, such as trust-region [47], or alternating approaches, including alternating stochastic gradient descent. Here, we present the complexity of evaluating the cost function and its gradient, regardless of the specific algorithm or implementation one wishes to exploit.

For simplicity, denote by S and T two bounds for the radial indices $S(\ell)$ and $T(q)$ of the 3-D and 2-D PSWF expansions, respectively. Typically, it is sufficient to take $S = S(0)$ and $T = T(0)$, as radial degree decreases as overall degree (ℓ) increases.

Starting from the first moment (27): with a fixed ℓ we have to apply two matrix-vector products in a row which requires an order of $\mathcal{O}\left(S\ell + TS\right)$ arithmetic operations. The variable ℓ increases up to L, which sums up to a total of $L \cdot \mathcal{O}\left(S\ell + TS\right) = \mathcal{O}\left(LS(L+T)\right)$. The gradient uses the precomputed remainder $m_1(A,B) - (\widetilde{m}_1)$ and is calculated by two terms with similar complexity as the above. Namely, the cost of both evaluation and gradient calculations is again $\mathcal{O}\left(LS(L+T)\right)$.

For the second moment, we follow (35): establishing $\Gamma_\ell^q \mathcal{A}_\ell$ is done in $\mathcal{O}\left(TSL\right)$ and applying the product in $\mathcal{O}\left(TL^2\right)$. Overall, the evaluation is bounded by

$$\mathcal{O}\left(L^2(TSL + TL^2)\right) = \mathcal{O}\left(TL^3(S+L)\right). \quad (62)$$

The gradient is a bit more complicated, in short, there are two terms for the volume derivatives and one term for the distribution part, with the precomputed remainder $m_2(A,B) - (\widetilde{m}_2)$ we get an overall complexity of $\mathcal{O}\left(L^2S(L^2+T^2+TL)\right)$. In summary, the first moment requires third-order complexity with respect to the different parameters where the second moment requires a total power of five.

Finally, the parameters T, S, and L can be described by the PSWF representation: the length L of the 3-D PSWF expansion and the bound on the radial indices S are related to the parameter c of sampling rate, and are bounded according to (78). Additional bound, now on the radial 2-D expansion T, uses the accuracy parameter ϵ of the 2-D images and the above L as given in (95). For more details on those parameters, see Appendix A.

4.6. **Remark on using semidefinite programming (SDP) relaxation.** Solving the nonlinear least-squares problem in Eq. (60) could suffer from slow convergence because the cost function is a polynomial of degree 6. We remark that in principle, it is possible to apply a semidefinite programming relaxation to facilitate the optimization. For convenience, let the second moments $m_2(A,B)_{q_1,t_1,q_2,t_2}$ be summarized as

$$m_2(A,B)_{q_1,t_1,q_2,t_2} := G_{q_1,t_1,q_2,t_2}(AA^T \otimes B) \tag{63}$$

where $G_{q_1,t_1,q_2,t_2}(\cdot)$ is a linear operator that captures the RHS of Eq. (34). If we define

$$\bar{A} = AA^T,$$

the optimization problem can be written as

$$\min_{\substack{A,\bar{A},B \\ \bar{A}=AA^T}} \sum_{q=-Q}^{Q} \sum_{t=0}^{T(q)} \left(m_1(A,B)_{q,t} - (\widetilde{m}_1)_{q,t} \right)^2 + \lambda \sum_{q_1,q_2=-Q}^{Q} \sum_{t_1,t_2=0}^{T(q)} \left(G_{q_1,t_1,q_2,t_2}(\bar{A} \otimes B) - (\widetilde{m}_2)_{q_1,t_1,q_2,t_2} \right)^2 .$$

To deal with the non-convex constraint $\bar{A} = AA^T$, we propose the following relaxed constraint

$$\bar{A} \succeq AA^T, \tag{64}$$

which gives the following non-linear least squares problem

$$\min_{\substack{A,\bar{A},B \\ \bar{A}\succeq AA^T}} \sum_{q=-Q}^{Q} \sum_{t=0}^{T(q)} \left(m_1(A,B)_{q,t} - (\widetilde{m}_1)_{q,t} \right)^2 + \lambda \sum_{q_1,q_2=-Q}^{Q} \sum_{t_1,t_2=0}^{T(q)} \left(G_{q_1,t_1,q_2,t_2}(\bar{A} \otimes B) - (\widetilde{m}_2)_{q_1,t_1,q_2,t_2} \right)^2 . \tag{65}$$

Comparing with (60), although (65) is still a non-convex problem, the degree of the polynomial in the cost function of (65) is 4 (instead of 6). Furthermore, one can solve (65) efficiently by minimizing $(A,\bar{A})$ and B in an alternating fashion. Therefore if at the optimum $\bar{A} \approx AA^T$ in spite of the relaxation (64), solving (65) can be advantageous.

We remark on the special case when the density coefficient B is given. In this situation, one can consider an SDP relaxation

$$\begin{aligned} \min_{\substack{A,\bar{A},\\ \bar{A}\succeq AA^T}} \quad & \mathrm{Tr}(\bar{A}) \qquad (66)\\ \text{subject to} \quad & \left|m_1(A,B)_{q,t} - (\widetilde{m}_1)_{q,t}\right| \le \epsilon_{q,t}, \quad 0 \le t \le T(q), \quad -Q \le q \le Q,\\ & \left|G_{q_1,t_1,q_2,t_2}(\bar{A}\otimes B) - (\widetilde{m}_2)_{q_1,t_1,q_2,t_2}\right| \le \epsilon_{q_1,t_1,q_2,t_2},\\ & 0 \le t_1 \le T(q_1), \quad 0 \le t_2 \le T(q_2), \quad -Q \le q_1, q_2 \le Q. \end{aligned}$$

The nuclear norm minimization strategy as in matrix completion [17] is used to promote $\bar{A}$ to be of rank-1. We test the SDP in (66) when given a fixed B_0. We generate B_0 for a non-uniform distribution from a 6-th degree nonnegative polynomial over the rotation group, i.e. letting $P = 6$. We generate a volume with random coefficients A_0 with $L = 3$. Noise is added to the moments in the following manner:

$$\begin{aligned} (\widetilde{m}_1)_{q,t} &= m_1(A_0,B_0)_{q,t} + \left|m_1(A_0,B_0)_{q,t}\right| z_{q,t},\\ (\widetilde{m}_2)_{q_1,t_1,q_2,t_2} &= G_{q_1,t_1,q_2,t_2}(A_0A_0^*\otimes B_0) + \left|G_{q_1,t_1,q_2,t_2}(A_0A_0^*\otimes B_0)\right| z_{q_1,t_1,q_2,t_2}. \end{aligned}$$

Where

$$z_{q,t}, z_{q_1,t_1,q_2,t_2} \sim \mathrm{Uniform}[-\epsilon,\epsilon],$$

and

$$0 \le t \le T(q),\ -Q \le q \le Q,\ 0 \le t_1 \le T(q_1),\ 0 \le t_2 \le T(q_2),\ -Q \le q_1, q_2 \le Q.$$

In this case, we set in (66),

$$\epsilon_{q,t} = \epsilon\left|m_1(A_0,B_0)_{q,t}\right| \quad \text{and} \quad \epsilon_{q_1,t_1,q_2,t_2} = \epsilon\left|G_{q_1,t_1,q_2,t_2}(A_0A_0^*\otimes B_0)\right|.$$

The stability results in recovering A_0 are shown in Figure 3. We ran five simulations for every ϵ and average the relative error

$$\mathtt{RE} = \frac{\|\bar{A} - A_0A_0^*\|_F}{\|A_0A_0^*\|_F}.$$

Results show an exact recovery in the noiseless case and slowly increasing in relative error as ϵ grows.

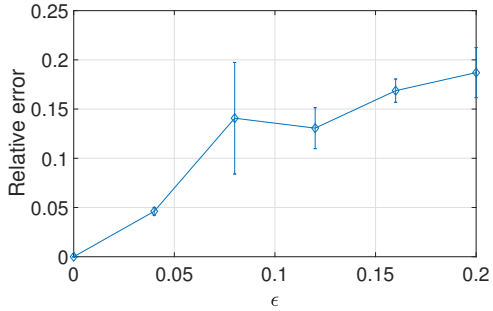


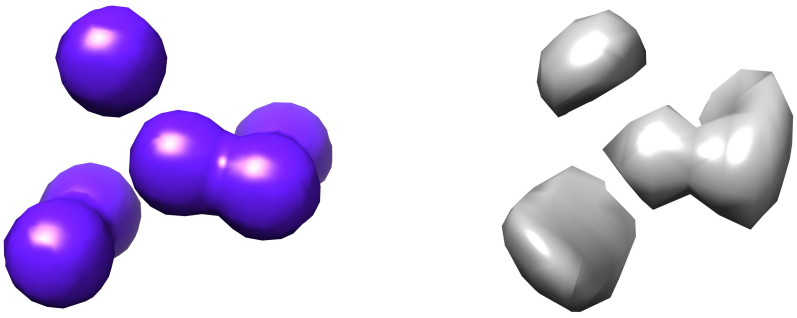
FIGURE 3. Stability of the SDP in (66) when fixing the density to be a non-uniform density.

4.7. **Volume from moments – non-uniform vs. uniform.** As a first numerical example, we present a recovery comparison between the cases of uniform and non-uniform distributions of rotations. In this example, we use as a ground truth a low degree approximation of a mixture of six Gaussians, given in a non-symmetric conformation. The approximation, which we ultimately use as our reference, is attained by discretizing the initial volume to $23 \times 23 \times 23$ and truncating the PSWFs expansion to $L = 4$. This expansion consists of 118 coefficients in total. The other PSWFs parameters that we use are a band limit c that corresponds to the Nyquist frequency and 3-D truncation parameter (75) of $\delta = 0.99$. The original volume and its approximation appear in Figure 4.

We divide the example into two scenarios of different distributions, uniform and non-uniform. In each case, we start from the analytic moments (7), calculated with respect to 2-D prescribed accuracy (95) of $\epsilon = 10^{-3}$, and obtain an estimation based on minimizing the least squares cost function (60). The optimization is carried with a gradient-based method, specifically we use an implementation of the *trust-region* algorithm, see *e.g.*, [47]. In the first case, we use as the distribution of rotations a quadratic expansion $P = 2$ which is in-plane uniform. Based on the in-plane invariance, we present this distribution as a function on the sphere in Figure 5. For the second case, we use a uniform distribution of rotations.

In both cases, we let the optimization reach numerical convergence, where the progress in minimization is minor. In this example, it is usually at about $100 - 150$ iterations. In the case of non-uniform distribution, we observe that choosing a random initial guess can have an effect on the speed of convergence but has almost no influence on the resulted volume. In other words, we gain numerical evidence for uniqueness. The estimated volume, in this case, is depicted on the left side of Figure 6.

On the other hand, in the case of a uniform distribution, while convergence was typically quicker than in the non-uniform case, the results vary between different initial guesses, indicating the richness of the space of possible solutions. One such solution appears on the right side of Figure 6. This behavior of the optimization solver agrees with our previous knowledge on the ill-posedness of Kam's method and also with the Jacobian test which shows degree deficiency of the polynomial system defined by the first and second moment under the uniform distribution.



(A) Mixture of Gaussians

(B) A low degree approximation using PSWF expansion with $L = 4$

FIGURE 4. Ground truth volumes

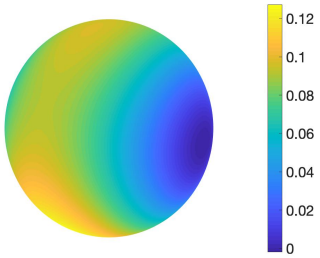
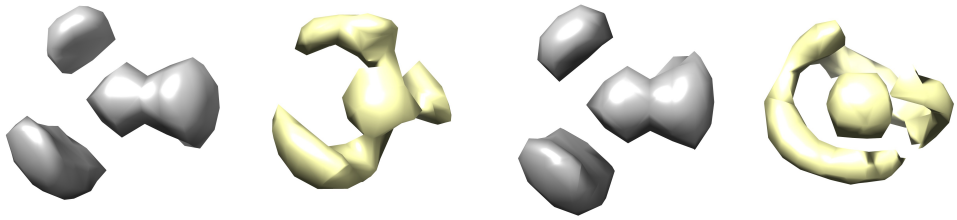


FIGURE 5. The non-uniform distribution of viewing angles which we use for Section 4.7. This distribution satisfies in-plane invariance and depicted as a function on the sphere



(A) Recovery under non-uniform distribution

(B) Recovery under uniform distribution

FIGURE 6. Comparison of reconstructions for two cases of non-uniform and uniform distribution of rotations: ground truth volume, as also seen in Figure 4b, appears on the left of each pair (in gray), where the estimation is on the right (in yellow)

4.8. **Comparing volumes using FSC.** A commonly used cryo-EM resolution measure is the Fourier shell correlation (FSC) [29]. The FSC measures cross-correlation coefficient between two 3-D volumes over each corresponding shell. That is, given two volumes ϕ_1 and ϕ_2, the FSC in a shell κ is calculated using all voxels $\boldsymbol{\kappa}$ on this κ-th shell:

$$\mathrm{FSC}(\kappa) = \frac{\sum_{\|\boldsymbol{\kappa}\|=\kappa} \phi_1(\boldsymbol{\kappa})\overline{\phi_2(\boldsymbol{\kappa})}}{\sqrt{\sum_{\|\boldsymbol{\kappa}\|=\kappa} \left|\phi_1(\boldsymbol{\kappa})\right|^2 \sum_{\|\boldsymbol{\kappa}\|=\kappa} \left|\phi_2(\boldsymbol{\kappa})\right|^2}} \tag{67}$$

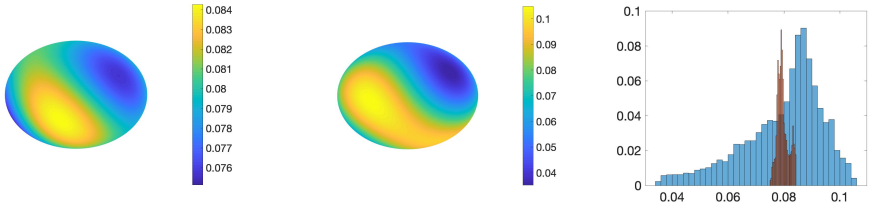
Customary, the resolution is determined by a cutoff value. The threshold question is discussed in [64], where in our case since we wish to compare a reconstructed volume against its ground truth, we use the 0.5 threshold. Since we focus on *ab initio* modeling, we aim to estimate a low-resolution version of the molecule from the first two moments. Thus, we expect the cutoff to reach a value which ensures a good starting point for a refinement procedure.

4.9. **Visual example and the effect of non-uniformity.** We next introduce an example for the most realistic scenario of an unknown, in-plane uniform distribution, by inverting the moment map of a real-world structure through minimization of

a least-squares cost function (60). In this example, we once again illustrate the feasibility of numerically approaching the solution, without any prior assumption on the volume.

The example is constructed as follows. As the ground truth volume, we once again use EMD-0409, the catalytic subunit of protein kinase A bound to ATP and IP20 [32], as presented at the online cryo-EM data-bank [38]. The map original dimension is $128 \times 128 \times 128$ voxels. Since we aim to recover a low-resolution model, we reduce complexity and downsample it by a factor of three to 43. We firstly expand this volume using PSWFs with a band limit c chosen as the Nyquist frequency and 3-D truncation parameter (75) of $\delta = 0.99$. The full expansion consists of degree $L = 40$, and after truncation to maximize conditioning, as done in Section 3.3, we aim to recover the low degree counterpart up to degree $L = 6$. The moments were calculated with respect to 2-D prescribed accuracy (95) of $\epsilon = 10^{-3}$ and in the absence of noise. The volume contributes 657 unknowns to be optimized.

As the ground truth distribution, we choose three different functions: uniform, highly non-uniform and a non-uniform case in-between. The two non-uniform cases are cubic spherical harmonics expansions ($P = 3$) and satisfy in-plane invariance and so we present them in Figure 7 as functions on the sphere, together with a histogram to compare and illustrate their "non-uniformness". The non-uniform distributions add extra 15 unknowns which means that, in total, we optimize 672 unknowns in the cases of non-uniform distribution and only 657 unknowns in the case of uniform distribution.



(A) The less non-uniform distribution function on the sphere

(B) The more non-uniform distribution function on the sphere

(C) The probability of each value to appear in the distribution: a comparison to illustrate the different non-uniformity levels of the two distributions

FIGURE 7. The two non-uniform distributions in use

In the optimization process, we use the limit of the empirical moments (59) ($n \to \infty$) as our input moments. As before, we use a *trust-region* algorithm, see *e.g.*, [47], which is a gradient-based method. To fix the initialization between the different cases, we start the search with the zero volume. In cases of non-uniform distribution, we provide a random non-uniform distribution to start with. Our method is implemented in MATLAB R2017b, and we calculated the example on a laptop with a 2.9 GHz Intel Core i5 processor and 16 GB 2133 MHz memory.

The result we present next is obtained after 60 iterations of trust-region, each iteration usually uses up to 30 inner iterations to estimate the most accurate step size. The runtime of this example is about 55 minutes for each model, where at this point, our naive implementation does not support any parallelization which

potentially can lead to a significant improvement in the total runtime. For example, the evaluation of the second moment and its associated gradient part are related to the leading complexity term as described in Section 4.5. Their implementation is based upon matrix product as seen in the form (35). This part can remarkably benefit from parallel execution. Note that evaluating the PSWF functions, as well as the product Clebsch-Gordan coefficients (which appears in the moments), are all calculated offline as a preprocessing step.

We present a comparison between the different FSC curves for the three cases. As implied by Figure 8, the resolution increases (lower FSC cut) as the non-uniformity becomes more significant. Specifically, with the uniform distribution we obtain merely 39.1Å, where for the two other non-uniform cases we get 22.5Å and 19.0Å as the non-uniformity increases in the examples of Figure 8.

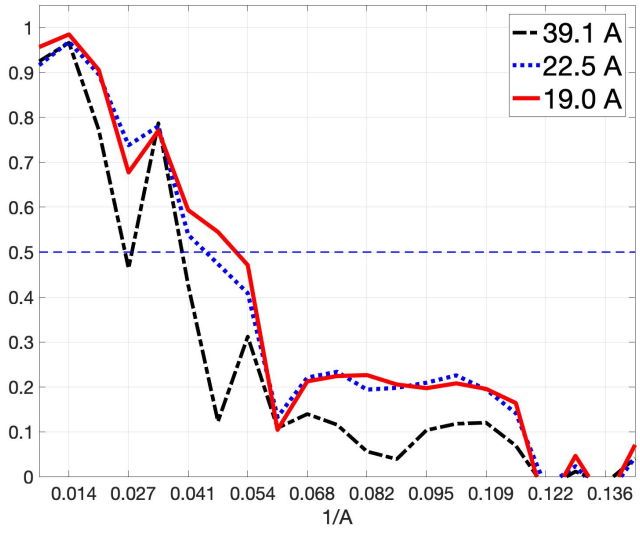


FIGURE 8. The FSC curves of the three test cases. The dashed curve (in black) is of the uniform distribution, the dot line (blue) is of the less radical non-uniform case, and the solid curve (red) is of the most non-uniform distribution case. As customary, we use the conventional FSC cutoff value of 0.5.

A visual demonstration of the output of the optimization is presented in Figure 9, where we plot side by side the ground truth and three models, from the uniform to the most non-uniform one.

4.10. **Recovery from noisy images.** We conclude this section with an example of recovering a volume from its noisy projection images. The volume is a mixture of six Gaussians, synthetically designed to have no spatial symmetry. The volume's size is $15 \times 15 \times 15$ and its full PSWF expansion is of length $L = 13$, with band limit c chosen as the Nyquist frequency and 3-D truncation parameter (75) of $\delta = 0.99$. We use an in-plane uniform distribution of rotations, very localized on a 45 degree cone, represented with an expansion length of $P = 3$. The distribution function is

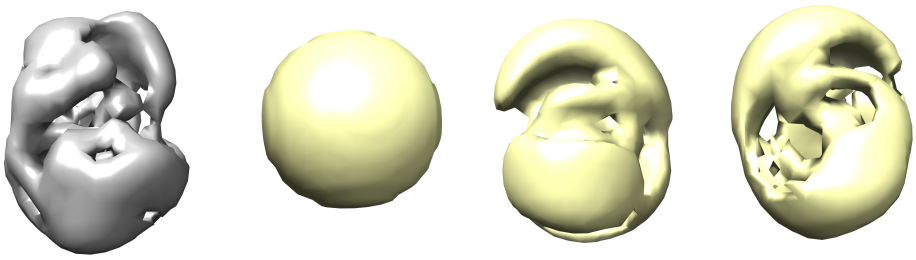


FIGURE 9. The estimations which were obtained by inverting the moments via optimization. The ground truth volume appears on the left (in gray), where the models are on the right (in yellow), ordered as associated with the different distributions, from uniform on the left to the most non-uniform on the right.

shown on Figure 10 and can model a realistic scenario of highly anisotropic viewing directions (see, e.g., [4]). Using the distribution, we generated projection images to obtain $200,000$ observations. These images were then contaminated with noise. The SNR of an image I_j with the noise term ε_j is $\mathrm{SNR}_j = \|I_j - \varepsilon_j\|^2 / \|\varepsilon_j\|^2$, using the Frobenius norm. The noise was chosen to achieve an average SNR value of $1/3$. Three examples of clean images and their noisy versions are depicted in Figure 11. As seen in Figure 11, the projections are hardly noticed in the noisy images for the naked eye.

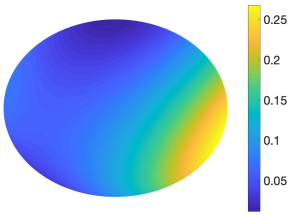


FIGURE 10. The distribution function on the sphere.

We expand the noisy images using a 2-D PSWF basis, as appears in (14). Then, the coefficients and their double-products are averaged to estimate the first and second moments as in (59). The reconstruction uses the empirical moments to estimate the volume and distribution. For the volume, our gradient-based least-squares algorithm targets its full expansion, which consists of 192 unknowns. The unknown distribution includes 8 unknowns spherical harmonics coefficients. We reached the result we present next very quickly, starting from a random initial guess. It took about 15 iterations of trust-region; each iteration could use up to 30 inner iterations to estimate the most accurate step size. The runtime of this example is less than 10 minutes.

A visual demonstration of the estimated volume is provided in Figure 12. We present the estimation, side by side, to the original volume. As seen in the various pictures, the reconstruction, while not perfect, captures most features and the general shape of the structure. This encouraging result indicates that inverting

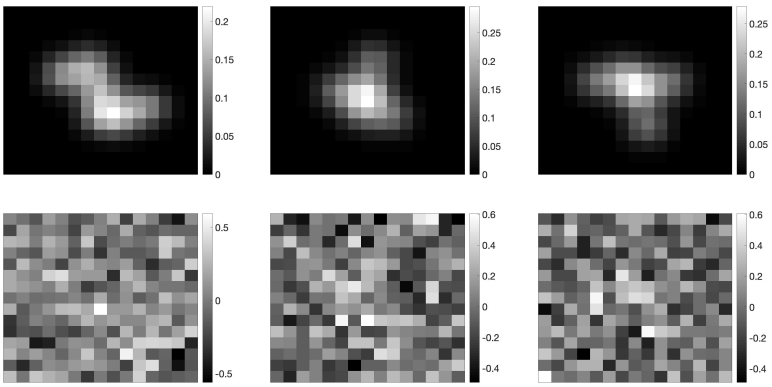


FIGURE 11. Three projection images: in upper row as clean and noisy images. The resulted SNR is about 1/3.

the moments is possible also from noisy moments and that the mapping has some robustness to small perturbations.

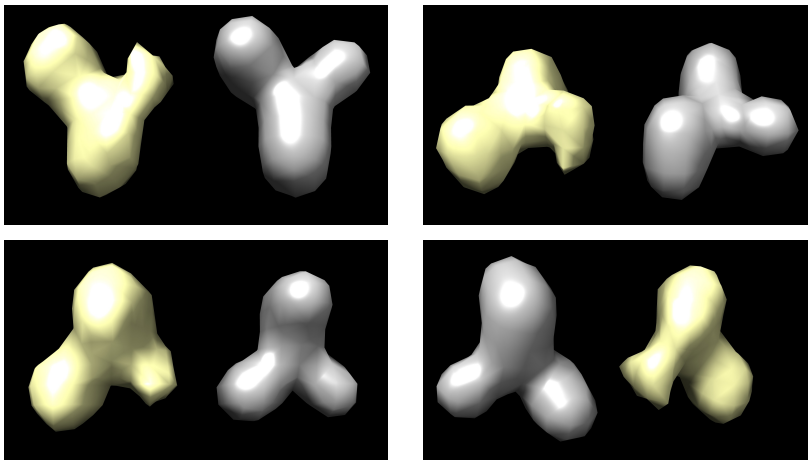


FIGURE 12. Reconstruction from moments of noisy images: an illustration taken from four different viewing angles. The estimation appears in yellow (left volume on the top left corner picture) and the original volume is in gray.

5. **Discussion and Conclusion.** The method of moments offers an attractive approach for modeling volumes in cryo-EM. This statistical method completely bypasses the estimation of viewing directions by treating them directly as nuisance parameters. The assumption of a non-uniform distribution of viewing angles enables in many cases volume estimation using only the first and second moments of the data. This phenomenon opens the door for fast, single-pass reconstruction algorithms, based on inverting the map from the volume and distribution to the low-order statistics of the projection images.

This paper extended Zvi Kam's original method of moments for cryo-EM to the setting of a non-uniform distribution of viewing directions. We formulated the reconstruction problem using appropriate discretizations for the images, the volume, and the distribution. Then, we derived moment formulas using properties of the spherical harmonic functions and Wigner matrix entries. Computational algebra was employed to analyze the resulting large-scale system of polynomial equations. The analysis shows the seeming complication of an unknown, non-uniform distribution renders 3-D reconstruction *easier* than in the uniform case, as now only first and second moments are required to determine a low-resolution expansion of the molecule, up to finitely many solutions. Intermediate cases were treated; remarkably, when the distribution is known and totally non-uniform over SO(3), there is an efficient, provable algorithm to invert the first and second moments non-linearly. Additionally, our work addressed several numerical and computational aspects of the method of moments. An implementation of a trust-region method was presented and used to illustrate the advantages of our approach over Kam's classical approach by numerical experiments involving synthetic volumes.

We regard our work as a definite, albeit initial step toward developing the method of moments for *ab initio* modeling from experimental datasets. Firstly, even in the synthetic cases considered here, further work on the optimization side is warranted. Variations on our nonlinear cost function that incorporate a pre-conditioner, e.g., (61), could be considered. Secondly, other techniques for large-scale nonlinear least squares optimization should be tried, such as Levenberg-Marquardt [40] or Variable Projection [18], where in the latter one can exploit the linearity in the moments with respect to the distribution, by eliminating out the distribution. Thirdly, to get our method working on images, further effects, such as the CTF and imperfect centering of picked particles, should be incorporated into the moment formulas. Fourthly, accurate covariance estimation in high dimensions requires eigenvalue shrinkage [22], the theory for which may call for a modification in the non-uniform setting.

To simplify our exposition, we have stuck to the asymmetric and homogeneous cases here, although both of these can be relaxed in the method of moments. Specifically, as already noted in Kam's original paper [33], point symmetries of molecules are reflected in the vanishing of certain expansion coefficients, see also [63]. Therefore, MoM can be reformulated using fewer coefficients for symmetric molecules. This fact, alongside with further improvement of the representation of the distribution, may pave the way for recovery under practical cases of very restricted viewing angles, as reported in literature [4, 26, 44, 59]. At the same time, heterogeneity, at least if it is finite and discrete, can be expressed using a mixture of volumes and a corresponding mixture of moments, see [14, 5]. In future work, computational algebra should be applied to these cases to check whether the first and second moments remain sufficient for unique recovery.

To conclude, we raise one further possibility, in some sense at odds with the message of this paper. In the non-uniform case, we have determined that the first and second moments are sufficient information-theoretically for volume recovery. Nonetheless, the resulting optimization landscape is potentially challenging, due to non-convexity or ill-conditioning. Thus, despite the increased *statistical* cost of estimating the third moment, it seems worthwhile to ask what can be gained *computationally* by reprising the third moment in MoM (or at least, using a carefully chosen slice of the third moment). Specifically, we would like to answer this question: can the third moment facilitate more efficient modeling at higher resolution?

Acknowledgments. The authors thank Nicolas Boumal, Peter Bürgisser, Eitan Levin, Dilano Saldin and Yoel Shkolnisky for stimulating conversations, and the anonymous referees for their valuable comments.

This research was supported in parts by Award Number R01GM090200 from the NIGMS, FA9550-17-1-0291 from AFOSR, Simons Foundation Math+X Investigator Award, the Simons Collaboration on Algorithms and Geometry, the Moore Foundation Data-Driven Discovery Investigator Award, and NSF BIGDATA Award IIS-1837992. BL's research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 723991 - CRYOMATH).

REFERENCES

[1] Emmanuel Abbe, Tamir Bendory, William Leeb, João M. Pereira, Nir Sharon, and Amit Singer. Multireference alignment is easier with an aperiodic translation distribution. *IEEE Transactions on Information Theory*, 65(6):3565–3584, 2019.

[2] Joakim Andén and Amit Singer. Structural variability from noisy tomographic projections. *SIAM Journal on Imaging Sciences*, 11(2):1441–1492, 2018.

[3] Larry C. Andrews. *Special Functions of Mathematics for Engineers*. McGraw-Hill New York, 1992.

[4] Philip R. Baldwin and Dmitry Lyumkis. Non-uniformity of projection distributions attenuates resolution in cryo-EM. *Progress in Biophysics and Molecular Biology*, 2019. In press, available online.

[5] Afonso S. Bandeira, Ben Blum-Smith, Joe Kileel, Amelia Perry, Jonathan Weed, and Alexander S. Wein. Estimation under group actions: recovering orbits from invariants. *arXiv preprint arXiv:1712.10163v2*, 2018.

[6] Afonso S. Bandeira, Philippe Rigollet, and Jonathan Weed. Optimal rates of estimation for multi-reference alignment. *arXiv preprint arXiv:1702.08546v2*, 2018.

[7] Alex Barnett, Leslie Greengard, Andras Pataki, and Marina Spivak. Rapid solution of the cryo-EM reconstruction problem by frequency marching. *SIAM Journal on Imaging Sciences*, 10(3):1170–1195, 2017.

[8] Adi Ben-Israel and Thomas N.E. Greville. *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics. Spring-Verlag New York, 2nd edition, 2003.

[9] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on Signal Processing*, 66(4):1037–1050, 2017.

[10] Tejal Bhamre, Teng Zhang, and Amit Singer. Denoising and covariance estimation of single particle cryo-EM images. *Journal of Structural Biology*, 195(1):72–81, 2016.

[11] L. C. Biedenharn and James D. Louck. *Angular Momentum in Quantum Physics : Theory and Application*. Number volume 8 in Section, Mathematics of Physics. Cambridge University Press, 1984.

[12] Nikhil Biyani, Ricardo D. Righetto, Robert McLeod, Daniel Caujolle-Bert, Daniel Castano-Diez, Kenneth N. Goldie, and Henning Stahlberg. Focus: the interface between data collection and data processing in cryo-EM. *Journal of Structural Biology*, 198(2):124–133, 2017.

[13] Mourad Boulsane and Abderrazek Karoui. The finite hankel transform operator: Some explicit and local estimates of the eigenfunctions and eigenvalues decay rates. *Journal of Fourier Analysis and Applications*, 24(6):1554–1578, 2018.

[14] Nicolas Boumal, Tamir Bendory, Roy R. Lederman, and Amit Singer. Heterogeneous multireference alignment: a single pass approach. In *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2018.

[15] Glen E. Bredon. *Introduction to Compact Transformation Groups*. Academic Press, New York-London, 1972. Pure and Applied Mathematics, volume 46.

[16] Peter Bürgisser and Felipe Cucker. *Condition: the Geometry of Numerical Algorithms*, volume 349 of *Fundamental Principles of Mathematical Sciences*. Springer, Heidelberg, 2013.

[17] Emmanuel J Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

[18] Guang-Yong Chen, Min Gan, C.L. Philip Chen, and Han-Xiong Li. A regularized variable projection algorithm for separable nonlinear least-squares problems. *IEEE Transactions on Automatic Control*, 64(2):526–537, 2018.

[19] Gregory S. Chirikjian and Alexander B. Kyatkin. *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition*. Courier Dover Publications, 2016.

[20] David A. Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms: an Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer, Cham, 4th edition, 2015.

[21] Jeffrey J. Donatelli, Peter H. Zwart, and James A. Sethian. Iterative phasing for fluctuation X-ray scattering. *Proceedings of the National Academy of Sciences*, 112(33):10286–10291, 2015.

[22] David L. Donoho, Matan Gavish, and Iain M. Johnstone. Optimal shrinkage of eigenvalues in the spiked covariance model. *Annals of statistics*, 46(4):1742, 2018.

[23] David Eisenbud. *Commutative Algebra: With a View Toward Algebraic Geometry*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[24] Niels Fischer, Piotr Neumann, Andrey L. Konevega, Lars V. Bock, Ralf Ficner, Marina V. Rodnina, and Holger Stark. Structure of the *E. coli* ribosome–EF-Tu complex at < 3 Å resolution by C_s-corrected cryo-EM. *Nature*, 520(7548):567–570, 2015.

[25] Joachim Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford University Press, 2nd edition, 2006.

[26] Robert M. Glaeser and Bong-Gyoon Han. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophysics Reports*, 3(1–3):1–7, 2017.

[27] Steven J Gortler, Alexander D Healy, and Dylan P Thurston. Characterizing generic global rigidity. *American Journal of Mathematics*, 132(4):897–939, 2010.

[28] Nikolaus Grigorieff. FREALIGN: high-resolution refinement of single particle structures. *Journal of Structural Biology*, 157(1):117–125, 2007.

[29] George Harauz and Marin van Heel. Exact filters for general geometry three dimensional reconstruction. *Optik*, 73(4):146–156, 1986.

[30] Joe Harris. *Algebraic Geometry: a First Course*, volume 133 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. Corrected reprint of the 1992 original.

[31] Richard A. Harshman. Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16:1 – 84, 1970.

[32] Mark A. Herzik, Mengyu Wu, and Gabriel C. Lander. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nature Communications*, 10(1):1032, 2019.

[33] Zvi Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, 82(1):15–39, 1980.

[34] Rami Katz and Yoel Shkolnisky. Sampling and approximation of bandlimited volumetric data. *Applied and Computational Harmonic Analysis*, 47(1):235–247, 2019.

[35] Werner Kühlbrandt. The resolution revolution. *Science*, 343(6178):1443–1444, 2014.

[36] Boris Landa and Yoel Shkolnisky. Approximation scheme for essentially bandlimited and space-concentrated functions on a disk. *Applied and Computational Harmonic Analysis*, 43(3):381–403, 2017.

[37] Boris Landa and Yoel Shkolnisky. Steerable principal components for space-frequency localized images. *SIAM Journal on Imaging Sciences*, 10(2):508–534, 2017.

[38] Catherine L. Lawson, Ardan Patwardhan, Matthew L. Baker, Corey Hryc, Eduardo Sanz Garcia, Brian P. Hudson, Ingvar Lagerstedt, Steven J. Ludtke, Grigore Pintilie, Raul Sala, et al. Emdatabank unified data resource for 3DEM. *Nucleic Acids Research*, 44(D1):D396–D403, 2015.

[39] Roy R. Lederman. Numerical algorithms for the computation of generalized prolate spheroidal functions. *arXiv preprint arXiv:1710.02874*, 2017.

[40] M.L.A. Lourakis and Antonis A. Argyros. Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 1526–1531. IEEE, 2005.

[41] Alan Merk, Alberto Bartesaghi, Soojay Banerjee, Veronica Falconieri, Prashant Rao, Mindy I. Davis, Rajan Pragani, Matthew B. Boxer, Lesley A. Earl, Jacqueline L.S. Milne, et al. Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell*, 165(7):1698–1707, 2016.

[42] Joseph A. Mindell and Nikolaus Grigorieff. Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, 142(3):334–347, 2003.

[43] T. S. Motzkin. The arithmetic-geometric inequality. In *Proc. Sympos. Wright-Patterson Air Force Base Ohio, 1965*, pages 205–224. Academic Press, New York, 1967.

[44] Katerina Naydenova and Christopher J. Russo. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature Communications*, 8(1):629, 2017.

[45] K. Pande, P. Schwander, M. Schmidt, and D.K. Saldin. Deducing fast electron density changes in randomly orientated uncrystallized biomolecules in a pump–probe experiment. *Philosophical Transactions of the Royal Society B*, 369(1647), 2014.

[46] Amelia Perry, Jonathan Weed, Afonso S Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multireference alignment. *SIAM Journal on Mathematics of Data Science*, 1(3):497–517, 2019.

[47] M.J.D. Powell and Y. Yuan. A trust region algorithm for equality constrained optimization. *Mathematical Programming*, 49(1):189–211, 1990.

[48] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 2017.

[49] Cyril F. Reboul, Michael Eager, Dominika Elmlund, and Hans Elmlund. Single-particle cryo-EM–improved *ab-initio* 3-D reconstruction with SIMPLE/PRIME. *Protein Science*, 27(1):51–61, 2018.

[50] Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015.

[51] Sjors H.W. Scheres. A Bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, 415(2):406–418, 2012.

[52] Sjors H.W. Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012.

[53] Kirill Serkh. On generalized prolate spheroidal functions. Technical Report TR-1519, Department of Mathematics, Yale University, 2015.

[54] Yoel Shkolnisky. Prolate spheroidal wave functions on a disc – integration and approximation of two-dimensional bandlimited functions. *Applied and Computational Harmonic Analysis*, 22(2):235–256, 2007.

[55] Amit Singer and Mihai Cucuringu. Uniqueness of low-rank matrix completion by rigidity theory. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1621–1641, 2010.

[56] Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. *SIAM Journal on Imaging Sciences*, 4(2):543–572, 2011.

[57] David Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty – iv: extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal*, 43(6):3009–3057, 1964.

[58] Andrew J. Sommese and Charles W. Wampler, II. *The Numerical Solution of Systems of Polynomials: Arising in Engineering and Science*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[59] Yong Zi Tan, Philip R. Baldwin, Joseph H. Davis, James R. Williamson, Clinton S. Potter, Bridget Carragher, and Dmitry Lyumkis. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods*, 14(8):793–796, 2017.

[60] Beata Turoňová, Florian K.M. Schur, William Wan, and John A.G. Briggs. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *Journal of Structural Biology*, 199(3):187–195, 2017.

[61] B.K. Vainshtein and A.B. Goncharov. Determination of the spatial orientation of arbitrarily arranged identical particles of unknown structure from their projections. In *Soviet Physics Doklady*, volume 31, pages 278–280, 1986.

[62] Marin van Heel. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111–123, 1987.

[63] Marin van Heel. Pointgroup symmetry of oligomeric macromolecules. *Structure*, 7:1575–1583, 1999.

[64] Marin van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005.

[65] Benjamin von Ardenne, Martin Mechelke, and Helmut Grubmüller. Structure determination from single molecule X-ray scattering with three photons per image. *Nature Communications*, 9(1):2375, 2018.

[66] Zhizhen Zhao and Amit Singer. Fourier–Bessel rotational invariant eigenimages. *Journal of the Optical Society of America A*, 30(5):871–877, 2013.

Appendix A. **Prolate Spheroidal Wave Functions.** Here we describe key properties of the PSWFs, and propose a method for setting the expansion parameters L, $S(\ell)$, Q, and $T(q)$. We begin with the three-dimensional PSWFs, where we describe important properties established in the literature [57, 34, 53], and outline our choice for setting L and $S(\ell)$, accordingly. Then, we proceed with a short analogous description for the two-dimensional PSWFs (summarizing results of [57]), and derive a method for choosing Q and $T(q)$ by directly exploiting the fact that the images to be expanded are tomographic projections of a bandlimited and localized volume function (employing our previous representation for the volume function).

A.1. **Volume function representation with three-dimensional PSWFs.** Let $\Phi : \mathbb{R}^3 \to \mathbb{R}$ be a square integrable (volume) function on $\mathbb{R}^3$, representing the true underlying electric potential of the molecule, and denote by $\hat{\Phi}$ its three-dimensional Fourier transform. It is common practice to assume that $\Phi(x)$ is bandlimited (*i.e.*, $\hat{\Phi}$ is restricted to a ball) while being localized in space. Functions satisfying this property are naturally represented by three-dimensional PSWFs, as detailed next.

We say that the function $\Phi(x)$ as *c-bandlimited* if $\hat{\Phi}(\omega)$ vanishes outside a ball of radius c. That is, Φ is c-bandlimited if

$$\Phi(x) = \left(\frac{1}{2\pi}\right)^3 \int_{c\mathbf{B}} \hat{\Phi}(\omega) e^{\imath \omega x} d\omega, \quad x \in \mathbb{R}^3, \tag{68}$$

where $\mathbf{B}$ is the unit ball. Among all c-bandlimited functions, the three-dimensional PSWFs on $\mathbf{B}$ [57] are the most energy concentrated in $\mathbf{B}$, while constituting an orthonormal system over $\mathcal{L}^2(\mathbf{B})$. Namely, they satisfy

$$\begin{aligned} \Psi_i = &\operatorname{argmin}_{\psi} \|\psi\|_{\mathcal{L}^2(\mathbb{R}^3)} \\ &\text{subject to } \|\psi\|_{\mathcal{L}^2(\mathbf{B})} = 1, \ \left\langle \psi, \Psi_j \right\rangle_{\mathcal{L}^2(\mathbf{B})} = 0, \ \forall j < i, \end{aligned} \tag{69}$$

for $i = 1, 2, \ldots$, *i.e.*, Ψ_1 is the most energy concentrated c-bandlimited function, Ψ_2 is the most energy concentrated c-bandlimited function orthogonal to Ψ_1, and so on. Three-dimensional PSWFs can be obtained as the solutions to the integral equation

$$\alpha \Psi(x) = \int_{\mathbf{B}} \Psi(\omega) e^{\imath c \omega x} d\omega, \quad x \in \mathbf{B}, \tag{70}$$

where we denote the solutions (the PSWFs with bandlimit c) as $\Psi^c_{\ell,m,s}(x)$ and their corresponding eigenvalues as $\alpha^c_{\ell,m,s}$, where the enumeration over i in (69) is replaced with an enumeration over the triplet ℓ, m, s described below, and the eigenvalues appear in non-increasing ordering with respect to the original enumerate i. $\Psi^c_{\ell,m,s}(x)$ and $\alpha^c_{\ell,m,s}$ together form the eigenfunctions and eigenvalues of (70), with $m \in \mathbb{Z}$, $\ell \in \mathbb{N} \cup \{0\}$, and $s \in \mathbb{N}$. Furthermore, the functions $\Psi^c_{\ell,m,s}(x)$ are orthogonal on both $\mathbf{B}$ and $\mathbb{R}^3$ using the standard $\mathcal{L}^2$ inner products on $\mathbf{B}$ and $\mathbb{R}^3$, respectively, and are dense in both the class of $\mathcal{L}^2(\mathbf{B})$ functions and in the class

of c-bandlimited functions on $\mathbb{R}^3$. In spherical coordinates, the functions $\Psi^c_{\ell,m,s}(x)$ agree with the form in the right-hand side of (8), and can be expressed as

$$\Psi^c_{\ell,m,s}(r,\theta,\varphi) = F^c_{\ell,s}(r)Y^m_\ell(\theta,\varphi), \tag{71}$$

where $Y^m_\ell(\theta,\varphi)$ are the spherical harmonics (see (9)). Numerical evaluation of the three-dimensional PSWFs (in particular of the radial part $F^c_{\ell,s}$) was considered in [39].

From the properties of the three-dimensional PSWFs mentioned above, any volume function $\Phi(x) \in \mathcal{L}^2(\mathbb{R}^3)$ can expanded in $\mathbf{B}$ as

$$\Phi(x) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\sum_{s=1}^{\infty}\widetilde{A}_{\ell,m,s}\Psi^c_{\ell,m,s}(x), \quad x\in\mathbf{B}, \qquad \widetilde{A}_{\ell,m,s} = \int_{\mathbf{B}}\Phi(x)\overline{\Psi^c_{\ell,m,s}(x)}dx, \tag{72}$$

where $\overline{(\cdot)}$ denotes complex conjugation. Next, we consider the truncation of the expansion in (72), where it is convenient to bound the resulting truncation error in terms of the assumed spatial localization of $\Phi(x)$. Towards this end, we say that the function $\Phi(x)$ is *ε-concentrated* if

$$\sqrt{\int_{x\notin\mathbf{B}}\left|\Phi(x)\right|^2 dx} \le \varepsilon. \tag{73}$$

Additionally, we define the normalized eigenvalues

$$\lambda^c_{\ell,m,s} = \left(\frac{c}{2\pi}\right)^3\left|\alpha_{\ell,m,s}\right|^2, \tag{74}$$

where we mention that $0\le\lambda^c_{\ell,m,s}\le 1$, $\lambda^c_{\ell,m,s} = \lambda^c_{\ell,0,s}$ for all triplets (ℓ,m,s), and $\lambda^c_{\ell,m,s}\underset{s\to\infty}{\longrightarrow}0$ for every ℓ. Now, we propose to set $S(\ell)$ according to

$$S(\ell) = \max_{s\in\mathbb{N}}\left\{s:\ \lambda^c_{\ell,0,s}\ge\delta\right\}, \tag{75}$$

where $\delta\in(0,1)$ is some constant, and set L to be the largest ℓ for which $S(\ell)$ is defined (*i.e.*, such that the set $\left\{s:\ \lambda^c_{\ell,0,s}\ge\delta\right\}$ is non-empty). Correspondingly, the volume function resulting from the truncating the expansion in (72), according to the chosen $S(\ell)$ and L, is

$$\phi(x) = \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\sum_{s=1}^{S(\ell)}\widetilde{A}_{\ell,m,s}\Psi^c_{\ell,m,s}(x). \tag{76}$$

The following proposition bounds the error of approximating $\Phi(x)$ by $\phi(x)$.

Proposition 1. *Let $\Phi(x)$ be c-bandlimited with a unit $\mathcal{L}^2(\mathbf{B})$ norm and assume it is ε-concentrated. Then,*

$$\|\Phi-\phi\|_{\mathcal{L}^2(\mathbf{B})} \le \varepsilon\sqrt{\frac{\delta}{1-\delta}}. \tag{77}$$

The proof follows immediately from Theorem 5 in [34] and from our choices of $S(\ell)$ and L. It is evident that the approximation error in the right-hand side of (77) can be made arbitrarily small by taking δ sufficiently small. Furthermore, in the case that $\Phi(x)$ is localized in space, *i.e.*, $\varepsilon\ll 1$, we can take δ to be large, possibly even close to 1, and still get approximation errors sufficiently small for our purposes.

A.1.1. *Length of the expansion.* Clearly, the number of basis functions taking part in the expansion (76), which is given explicitly by $\sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell} S(\ell)$, depends on the number of normalized eigenvalues $\lambda^c_{\ell,m,s}$ exceeding δ. In this respect, the normalized eigenvalues $\lambda^c_{\ell,m,s}$ are known to admit the following three distinct regions of behavior (when sorted in descending order). The first is called the "flat region", where $\lambda^c_{\ell,m,s}$ take values very close to 1, the second is called the "transitional region", where $\lambda^c_{\ell,m,s}$ shift rapidly from values close to 1 to values close to 0, and the third is called the "decay region", where $\lambda^c_{\ell,m,s}$ are very close to 0 and exhibit a super-exponential decay rate. As for the number of basis functions chosen according to (75), the following holds [53]:

$$\begin{aligned}\sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell} S(\ell) &= \left|\left\{(\ell,m,s):\ \lambda^c_{\ell,m,s}\geq\delta\right\}\right| \\ &= \frac{c^3}{4.5\pi}+\frac{c^2}{2\pi^2}\log(c)\log\left(\frac{1-\delta}{\delta}\right)+o(c^2\log(c)),\end{aligned} \tag{78}$$

where the first, second, and third terms on the right-hand side of (78) correspond to the number of normalized eigenvalues $\lambda^c_{\ell,m,s}$ exceeding δ from the flat region, the transitional region, and the decay region of the eigenvalues, respectively. Clearly, the asymptotically dominant term is $\mathcal{O}(c^3)$, which corresponds to the number of terms in the expansion chosen from the flat region. Additionally, we need an extra $\mathcal{O}(c^2\log(c))$ terms if we take δ to be small (note that the second term in the right hand-side of (78) is negative for $\delta > 0.5$, meaning that asymptotically we need less than $c^3/4.5\pi$ terms for values of δ close to 1). The remaining $o(c^2\log(c))$ terms from the decay region are negligible compared to the leading asymptotic terms.

A.1.2. *Fourier domain representation.* Up to this point, we have shown that three-dimensional PSWFs are naturally adapted for expanding a volume function $\Phi(x)$ which is bandlimited and localized in space, where we provided an appropriate error bound (77). However, note that in (8) we actually expand the Fourier transform of the molecular potential. We now connect our previous expansion of $\Phi(x)$ with the expansion of its Fourier transform, and show that in fact (and uniquely for PSWFs) the two coincide, in the sense that expanding a function in three-dimensional PSWFs is equivalent to expanding its Fourier transform in three-dimensional PSWFs (after an appropriate scaling and dilation). Let $\hat{\Psi}_{\ell,m,s}$ denote the three-dimensional Fourier transform of $\Psi_{\ell,m,s}$, then by (70) it is easy to verify that

$$\hat{\Psi}_{\ell,m,s}(\omega) = \frac{(2\pi)^3}{c^3\alpha_{\ell,m,s}}\Psi_{\ell,m,s}\left(\frac{\omega}{c}\right)\cdot\mathbf{1}_{c\mathbf{B}}(\omega), \tag{79}$$

where $\mathbf{1}_{c\mathbf{B}}(\omega)$ is the indicator function on $c\mathbf{B}$. It is evident that the Fourier transform of each three-dimensional PSWF is equal to itself up to a constant factor, a dilation by c, and a restriction to a ball of radius c. Consequently, by taking the Fourier transform of (76) we have

$$\hat{\phi}(\omega) = \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\sum_{s=0}^{S_\ell} A_{\ell,m,s}\Psi^c_{\ell,m,s}\left(\frac{\omega}{c}\right)/c^{3/2}, \qquad \omega\in c\mathbf{B}, \tag{80}$$

where

$$A_{\ell,m,s} = \frac{8\pi^3}{c^{3/2}\alpha_{\ell,m,s}}\widetilde{A}_{\ell,m,s}. \tag{81}$$

We conclude this part as follows. Given a bandlimit c (typically chosen as the Nyquist frequency corresponding to the projection images' resolution), we take the radial part $F_{\ell,m,s}(k)$ of (8) as $F^c_{\ell,m,s}(k/c)/c^{3/2} \cdot \mathbf{1}_c(k)$, where $\mathbf{1}_c(k)$ is the indicator function on $[0,c]$, and $F^c_{\ell,m,s}(r)$ is the radial part of the three-dimensional PSWFs on $\mathbf{B}$ (the factor $1/c^{3/2}$ ensures that $F_{\ell,m,s}(k)$ are orthonormal over $[0,\infty)$ w.r.t the measure k^2dk). Then, setting $S(\ell)$ according to (75) for a given parameter δ allows for the controlled approximation error (77).

A.2. **Projection image representation with two-dimensional PSWFs.** In the sequel, we are interested in providing a suitable representation for the projection images of the rotated copies of $\phi(x)$. By the Fourier slice theorem, the two-dimensional Fourier transforms of such projections are equal to slices from the three-dimensional Fourier transform of $\phi(x)$ (*i.e.*, of $\hat{\phi}(\omega)$). Therefore, if $\phi(x)$ is c-bandlimted, then the projection images are bandlimited to a disk of radius c. Additionally, we expect the projection images to be localized in the unit disk if $\phi(x)$ is sufficiently localized in the unit ball. For such projection images, two-dimensional PSWFs are expected to provide a natural representation (see [36]).

We briefly summarize properties of the two-dimensional PSWFs which are used in our context. In essence, the properties of the two-dimensional PSWFs are analogous to those of the three-dimensional PSWFs when replacing the unit ball $\mathbf{B}$ with the unit disk $\mathbf{D}$. Let $P : \mathbb{R}^2 \to \mathbb{R}$ be a square integrable function on $\mathbb{R}^2$, representing a tomographic projection of ϕ. We say that $P(x)$ as *c-bandlimited* if its two-dimensional Fourier transform, denoted by $\hat{P}(\omega)$, vanishes outside a disk of radius c. That is, P is c-bandlimited if

$$P(x) = \left(\frac{1}{2\pi}\right)^2 \int_{c\mathbf{D}} \hat{P}(\omega) e^{\imath\omega x} d\omega, \quad x \in \mathbb{R}^2. \tag{82}$$

Among all c-bandlimited functions, the two-dimensional PSWFs on $\mathbf{D}$ are the most energy concentrated in $\mathbf{D}$, that is, they satisfy (69) when replacing $\mathbf{B}$ with $\mathbf{D}$, while constituting an orthonormal system over $\mathcal{L}^2(\mathbf{D})$. The two-dimensional PSWFs were derived and analyzed in [57], and were shown to be the solutions to the integral equation

$$\beta\psi(x) = \int_{\mathbf{D}} \psi(\omega) e^{\imath c\omega x} d\omega, \quad x \in \mathbf{D}. \tag{83}$$

We denote the PSWFs with bandlimit c as $\psi^c_{q,t}(x)$, and their corresponding eigenvalues as $\beta^c_{q,t}$, which together form the eigenfunctions and eigenvalues of (83), with $q \in \mathbb{Z}$, and $t \in \mathbb{N}$. Furthermore, the functions $\psi^c_{q,t}(x)$ are orthogonal on both $\mathbf{D}$ and $\mathbb{R}^2$ using the standard $\mathcal{L}^2$ inner products on $\mathbf{D}$ and $\mathbb{R}^2$, respectively, and are dense in both the class of $\mathcal{L}^2(\mathbf{D})$ functions and in the class of c-bandlimited functions on $\mathbb{R}^2$. In polar coordinates, the functions $\psi^c_{q,t}(x)$ agree with the form in the right-hand side of (14), and can be expressed as

$$\psi^c_{q,t}(r,\varphi) = \frac{1}{\sqrt{2\pi}} f^c_{q,t}(r) e^{\imath q\varphi}, \tag{84}$$

where the eigenfunctions $\psi^c_{q,t}(x)$ are normalized to have an $\mathcal{L}^2(\mathbf{D})$ norm of 1. Numerical evaluation of the two-dimensional PSWFs was considered in [54].

From the properties of the two-dimensional PSWFs mentioned above, any function $P(x) \in \mathcal{L}^2(\mathbb{R}^2)$ can be expanded in $\mathbf{D}$ as

$$P(x) = \sum_{q=-\infty}^{\infty} \sum_{t=0}^{\infty} \widetilde{a}_{q,t} \psi_{q,t}^c(x), \quad x \in \mathbf{D}, \qquad \widetilde{a}_{q,t} = \int_{\mathbf{D}} P(x) \overline{\psi_{q,t}^c(x)} dx. \tag{85}$$

Now, considering the truncated expansion

$$I(x) := \sum_{q=-Q}^{Q} \sum_{t=0}^{T(q)} \widetilde{a}_{q,t} \psi_{q,t}^c(x), \tag{86}$$

we are interested in controlling the error

$$\|P - I\|_{\mathcal{L}^2_{\mathbf{D}}}^2 = \sum_{q=-Q}^{Q} \sum_{t>T(q)} \left|\widetilde{a}_{q,t}\right|^2 + \sum_{|q|>Q} \sum_{t=0}^{\infty} \left|\widetilde{a}_{q,t}\right|^2. \tag{87}$$

From (83), the Fourier transform of $\psi_{m,k}$ can be expressed as

$$\hat{\psi}_{m,k}(\omega) = \frac{4\pi^2}{c^2 \beta_{m,k}} \psi_{m,k}(\frac{\omega}{c}) \cdot \mathbf{1}_{c\mathbf{D}}(\omega), \tag{88}$$

where $\mathbf{1}_{c\mathbf{D}}(\omega)$ is the indicator function on $c\mathbf{D}$, which is analogous to the relation between the three-dimensional PSWFs $\Psi_{\ell,m,s}^c$ and their Fourier transforms $\hat{\Psi}_{\ell,m,s}^c$ in (79). Continuing, taking the Fourier transform of (86) gives

$$\hat{I}(\omega) = \sum_{q=-Q}^{Q} \sum_{t=0}^{T(q)} a_{q,t} \psi_{q,t}^c(\frac{\omega}{c}) \sqrt{2\pi}/c, \qquad \omega \in c\mathbf{D} \tag{89}$$

where

$$a_{q,t} = \frac{(2\pi)^{3/2}}{c\beta_{q,t}} \widetilde{a}_{q,t}. \tag{90}$$

We will now relate 2D basis representation error to that of the 3D basis functions. Comparing the 2D expansion (89) with the relation between 2-D and 3-D coefficients (19), while employing (90) and (81) we have

$$\widetilde{a}_{q,t} = \frac{c\beta_{q,t}}{(2\pi)^{3/2}} a_{q,t} = \sum_{\ell=|q|}^{L} \sum_{s=1}^{S(\ell)} \sum_{m=-\ell}^{\ell} \widetilde{A}_{\ell,m,s} \, U_{m,q}^{\ell}(R) \, \eta_{\ell,s}^{q,t}, \tag{91}$$

for $|q| \leq L$, where $\widetilde{a}_{q,t} = 0$ for $|q| > L$, and

$$\eta_{\ell,s}^{q,t} = \frac{c\beta_{q,t}}{(2\pi)^{3/2}} \frac{8\pi^3}{c^{3/2}\alpha_{\ell,m,s}} \gamma_{\ell,s}^{q,t} = \frac{(2\pi)^{3/2}\beta_{q,t}}{\sqrt{c}\alpha_{\ell,m,s}} \gamma_{\ell,s}^{q,t}, \tag{92}$$

where $\gamma_{\ell,s}^{q,t}$ is from (20). Using the Cauchy–Schwarz inequality, we can write

$$\begin{aligned}|\widetilde{a}_{q,t}| &\leq \sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}\sum_{m=-\ell}^{\ell}|\widetilde{A}_{\ell,m,s}\,U_{m,q}^{\ell}(R)\,\eta_{\ell,s}^{q,t}| \\ &\leq \left(\sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}\sum_{m=-\ell}^{\ell}\left|\widetilde{A}_{\ell,m,s}\right|^2\right)^{1/2}\left(\sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}\sum_{m=-\ell}^{\ell}\left|U_{m,q}^{\ell}(R)\,\eta_{\ell,s}^{q,t}\right|^2\right)^{1/2} \\ &\leq \|\phi\|_{\mathcal{L}^2(\mathbf{B})}\cdot\left(\sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}|\eta_{\ell,s}^{q,t}|^2\right)^{1/2}, \end{aligned} \tag{93}$$

where we also used the fact that $U^{\ell}(R)$ is a unitary matrix. Finally, taking $Q = L$ and assuming w.l.o.g that $\|\phi\|_{\mathcal{L}^2(\mathbf{B})} = 1$, we obtain from (87) and (93) that

$$\|P - I\|_{\mathcal{L}^2_{\mathbf{D}}}^2 \leq \sum_{q=-L}^{L}\sum_{t>T(q)}\sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}|\eta_{\ell,s}^{q,t}|^2. \tag{94}$$

Given a prescribed accuracy ϵ, for every $-L \leq q \leq L$ we choose $T(q)$ to be the smallest integer such that

$$\sum_{t>T(q)}\sum_{\ell=|q|}^{L}\sum_{s=1}^{S(\ell)}|\eta_{\ell,s}^{q,t}|^2 \leq \frac{\epsilon}{2L+1}, \tag{95}$$

which results in

$$\|P - I\|_{\mathcal{L}^2_{\mathbf{D}}}^2 \leq \epsilon, \tag{96}$$

where $\eta_{\ell,s}^{q,t}$ are computed by evaluating $\gamma_{\ell,s}^{q,t}$ of (20) via numerical integration (using Gauss-Legendre quadratures). Note that the right-hand side of (95) is determined by the decay rate of $\eta_{\ell,s}^{q,t}$ in t, which is dominated by the decay rate of the the eigenvalues of the two-dimensional PSWFs $\beta_{q,t}$. Those are known to admit a rapid decay in the form of a super-exponential decay rate following a certain transitional region (see [13, 53]). Hence, if $T(q)$ is sufficiently large then (95) can be satisfied for an arbitrarily small ϵ with a marginal increase in the number of required terms. Last, we mention that when provided with images sampled on a Cartesian grid, the coefficients $\widetilde{a}_{q,t}$ can be approximated accurately from the images by fast algorithms [36, 37].

Appendix B. **Linearizing polynomial maps with the Jacobian matrix.** In this section, we describe the linearization technique from computational algebraic geometry we used to obtain the uniqueness results in Tables 1, 2, 3 from Section 3. The first paper to apply algebraic geometry techniques to cryo-EM was [5]. Nevertheless, similar Jacobian tests have been used in other applications such as for testing rigidity in sensor network localization, see e.g., [27] and testing whether a matrix can be completed into a low-rank matrix [55].

To state the method, we fix $\mathbb{C}^N = \mathbb{C}^{N'} \oplus \mathbb{C}^{N''}$, let π' and π'' be projection onto the factors, and consider a *polynomial map* $F = (F_1, \ldots, F_M) : \mathbb{C}^N \to \mathbb{C}^M$ (that is, each coordinate function $F_i = F_i(x_1, \ldots, x_M)$ is a polynomial on $\mathbb{C}^N$). While F is

generally a nonlinear map, its first derivative at $q \in \mathbb{C}^N$ is a linear map represented by the Jacobian matrix

$$dF := \left(\frac{\partial F_j}{\partial x_i} \right)_{\substack{i=1,\dots,M \\ j=1,\dots,N}} \tag{97}$$

In addition, we define the *fiber* in $q \in \mathbb{C}^N$ by

$$F_q := \{\widetilde{q} \in \mathbb{C}^N \mid F(\widetilde{q}) = F(q)\} \subset \mathbb{C}^N,$$

and the *projected fiber* by

$$\overline{\pi'(F_q)} \subset \mathbb{C}^{N'}.$$

For $q' \in \mathbb{C}^{N'}$ and $q'' \in \mathbb{C}^{N''}$, define the *specialized fiber* by

$$\left(F|_{\mathbb{C}^{N'} \oplus q''}\right)_{q'} = \{\widetilde{q}' \in \mathbb{C}^{N'} \mid F(\widetilde{q}' \oplus q'') = F(q' \oplus q'')\} \subset \mathbb{C}^{N'}.$$

Because F is described by polynomials, there is a tight relationship between the dimension of fibers of F (as algebraic varieties) and the dimension of the kernels of dF (as linear spaces). This is summarized by the Jacobian tests below. Somewhat remarkably, the linear algebra tests are done at a *single point* in the domain of F, but imply algebraic geometric statements for *almost all* points in the domain of F.

Theorem 6. *Suppose it is known that, generically, the fiber, projected fiber and specialized fiber have dimensions $\geq d_1, d_2, d_3$, respectively (if we have no such knowledge, then $d_1 = d_2 = d_3 = 0$). Choose particular points $q_0 \in \mathbb{C}^N$, $q_0' \in \mathbb{C}^{N'}$ and $q_0'' \in \mathbb{C}^{N''}$.*

- Vanilla Jacobian test: *if* $\operatorname{rank} dF(q_0) = N - d_1$, *then generic fibers have dimension exactly* d_1.
- Projected Jacobian test: *if* $\dim \pi'(\ker dF(q_0)) = d_2$, *then generic projected fibers have dimension exactly* d_2.
- Specialized Jacobian test: *if* $\operatorname{rank}\ d\left(F|_{\mathbb{C}^{N'} \oplus q_0''}\right)(q_0') = N' - d_3$, *then generic specialized fibers have dimension exactly* d_3.

TABLE 4. Vanilla, projected and specialized Jacobian tests: these show that a system of polynomial equations generically has only finitely solutions. Notation: $F : \mathbb{C}^N \to \mathbb{C}^M$ is a polynomial map, $\mathbb{C}^N = \mathbb{C}^{N'} \oplus \mathbb{C}^{N''}$ where π', π'' are orthogonal projections onto the factors, and d_1, d_2, d_3 are the dimension bounds in Theorem 6.

	polynomial map	arbitrary choices	linearization	rank check
vanilla	$\mathbb{C}^N \xrightarrow{F} \mathbb{C}^M$	$q \in \mathbb{C}^N$	$\mathbb{C}^N \xrightarrow{dF(q)} \mathbb{C}^M$	$\operatorname{rank}(dF(q)) = N - d_1$
projected	$\mathbb{C}^{N'} \oplus \mathbb{C}^{N''} = \mathbb{C}^N \xrightarrow{F} \mathbb{C}^M$	$q \in \mathbb{C}^N$	$\mathbb{C}^{N'} \subset \mathbb{C}^N \xrightarrow{dF(q)} \mathbb{C}^M$	$\dim\left(\pi'(\ker dF(q))\right) = d_2$
specialized	$\mathbb{C}^{N'} \oplus \mathbb{C}^{N''} = \mathbb{C}^N \xrightarrow{F} \mathbb{C}^M$	$q' \in \mathbb{C}^{N'}$ $q'' \in \mathbb{C}^{N''}$	$\mathbb{C}^{N'} \oplus q'' \xrightarrow{dF(q' \oplus q'')} \mathbb{C}^M$	$\operatorname{rank}\ d\left(F|_{\mathbb{C}^{N'} \oplus q''}\right)(q') = N' - d_3$

Several technical remarks are in order. Firstly, in Theorem 6, the fiber, projected fiber and specialized fiber are *affine algebraic varieties* and hence a *dimension* is defined for each of their *irreducible components* according to [20]. The meaning of

the theorem is that each component has dimension exactly d_1, d_2, d_3, respectively. Crucially, affine algebraic varieties have finitely many components. Thus the theorem implies "finitely many solutions" up to symmetries, if the symmetries give d_1, d_2, d_3-dimensional ambiguities, respectively. Secondly, "generic" in Theorem 6 is with respect to the *Zariski topology*. Concretely, there exists some polynomial G on $\mathbb{C}^N$ such that for all $q \in \mathbb{C}^N$ with $G(q) \neq 0$ the implications in the theorem hold. In particular, any property that holds generically holds on a Lebesgue full measure subset of points. Thirdly, the Jacobian ranks in Theorem 6 take on generic values, as each minor of the relevant matrix is a polynomial in q, q', q''.

Theorem 6 states rigorous conclusions if the Jacobian rank tests are *passed*. On the other hand, if the tests *fail* for q_0, q_0', q_0'', and q_0, q_0', q_0'' were drawn randomly from *any* continuous distribution on $\mathbb{F}^N$, then by genericity of the Jacobian ranks, with probability 1, the generic fibers, projected fibers, or specialized fibers of F have dimension *strictly* more than d_1, d_2, or d_3,

We applied the specialized test in subsection 3.2 with $d_1 = 0$, the vanilla and projected tests in subsection 3.3 with $d_1 = d_2 = 3$ and the vanilla test in subsection 3.4 with $d_1 = 3$. The settings of 3 reflect the fact, in the latter two subsections, that the fibers are SO(3)-sets and we are interested in solutions modulo global rotation. The bounds may be seen as instances of the orbit-stabilizer theorem, see [5, Proposition 4.11]. When the Jacobian rank tests were passed, this meant that, generically, there are only finitely many solutions up to global ambiguities.

In practice, we ran the Jacobian tests in floating-point arithmetic and used SVD for robust rank estimation. Namely, we looked at multiplicative gaps between consecutive singular values, and regarded any gap exceeding a predefined threshold (10^6) as evidence that all lower singular values should be regarded as zero. While these computations fall short of a fully rigorous mathematical proof due to the possibility of rounding errors in floating-point arithmetic, it was typically evident which singular values ought to be counted as zero or non-zero.

E-mail address: nsharon@tauex.tau.ac.il
E-mail address: jkileel@math.princeton.edu
E-mail address: yuehaw.khoo@gmail.com
E-mail address: sboris20@gmail.com
E-mail address: amits@math.princeton.edu