Bayesian Perspective for Orientation Determination in Cryo-EM with Application to Structural Heterogeneity Analysis

Sheng Xu*†
Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA

Amnon Balanov†
School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel

Amit Singer
Program in Applied and Computational Mathematics and Department of Mathematics, Princeton University, Princeton, NJ 08544, USA

Tamir Bendory
School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel

February 23, 2026

Abstract

Accurate orientation estimation is a crucial component of 3D molecular structure reconstruction, both in single-particle cryo-electron microscopy (cryo-EM) and in the increasingly popular field of cryo-electron tomography (cryo-ET). The dominant approach, which involves searching for the orientation that maximizes cross-correlation relative to given templates, is sub-optimal, particularly under low signal-to-noise conditions. In this work, we propose a Bayesian framework for more accurate and flexible orientation estimation, with the minimum mean square error (MMSE) estimator serving as a key example. Through simulations, we demonstrate that the MMSE estimator consistently outperforms the cross-correlation-based method, especially in challenging low signal-to-noise scenarios, and we provide a theoretical framework that supports these improvements.

When incorporated into iterative refinement algorithms in the 3D reconstruction pipeline, the MMSE estimator markedly improves reconstruction accuracy, reduces model bias, and enhances robustness to the "Einstein from Noise" artifact. Crucially, we demonstrate that orientation estimation accuracy has a decisive effect on downstream structural heterogeneity analysis. In particular, integrating the MMSE-based pose estimator into frameworks for continuous heterogeneity recovery yields accuracy improvements approaching those obtained with ground-truth poses, establishing MMSE-based pose estimation as a key enabler of high-fidelity conformational landscape reconstruction. These findings indicate that the proposed Bayesian framework could substantially advance cryo-EM and cryo-ET by enhancing the accuracy, robustness, and reliability of 3D molecular structure reconstruction, thereby facilitating deeper insights into complex biological systems.

1 Introduction

Determining the precise three-dimensional (3D) orientation of biological molecules from their noisy two-dimensional (2D) projection images is a fundamental challenge in cryo-electron microscopy (cryo-EM) [1, 36, 8]. This process, known as orientation estimation, is crucial for various cryo-EM

*Corresponding author; email: sxu21@princeton.edu
†These authors contributed equally.

applications, including 3D reconstruction algorithms [49, 44], heterogeneity analysis [54, 58, 11, 64], and beyond [38]. For example, Figure 1(a) illustrates the role of orientation estimation within the cryo-EM 3D reconstruction workflow.

In cryo-electron tomography (cryo-ET), orientation estimation presents an additional challenge in the form of subtomogram averaging. Notably, cryo-ET suffers from higher noise levels compared to single-particle cryo-EM due to the complex and heterogeneous nature of cellular samples and the challenges of capturing data from multiple angles within thicker specimens. Subtomogram averaging offers an effective approach to enhance the signal-to-noise ratio (SNR), ultimately resulting in the reconstruction of high-resolution structures. This technique often involves extracting multiple similar subtomograms containing the target protein complex or macromolecule from a large cryo-electron tomogram reconstructed from all available tilts (typically from $-60°$ to $+60°$), followed by aligning and averaging them [63, 60]. Unlike traditional cryo-EM, this process typically aligns 3D structures directly, without the direct use of 2D projections (see Figure 1(b) for an illustration).

Mathematically, the orientation estimation tasks in cryo-EM and cryo-ET are slightly different. In the process of single-particle cryo-EM, which involves 2D tomographic projections, the mathematical model can be formulated as:

$$y = \Pi(g \circ V) + \varepsilon, \tag{1.1}$$

where $y : \mathbb{R}^2 \to \mathbb{R}$ is the observed 2D projection image, $V : \mathbb{R}^3 \to \mathbb{R}$ is the underlying 3D molecular structure, Π is the tomographic projection operator, g is the unknown 3D rotation operator of interest, ε represents measurement noise, and $g \circ V(x) \equiv V(g^{-1}x)$, representing a rotation acting on a volume V with 3D coordinate x. Analogously, the mathematical model in cryo-ET subtomogram averaging, which aligns directly with the 3D structure, can be represented by:

$$y = g \circ V + \varepsilon, \tag{1.2}$$

where $y : \mathbb{R}^3 \to \mathbb{R}$ is the observed 3D subtomogram, $V : \mathbb{R}^3 \to \mathbb{R}$ is the underlying 3D molecular structure, g is the unknown 3D rotation operator of interest, and ε represents measurement noise. Then, the goal of orientation estimation is *to find the "best" 3D rotation g based on the 2D projection image (in the cryo-EM case (1.1)) or the 3D subtomogram (in the cryo-ET case (1.2)) with respect to the 3D reference V. Namely, we aim to estimate the rotation g given the sample y and the 3D structure V.*

The actual mathematical models used in cryo-EM and cryo-ET are more intricate than those presented in (1.1) and (1.2), incorporating additional factors such as the contrast transfer function (CTF) and in-plane translations, as elaborated in Appendix A. Although the framework introduced in this work is capable of addressing the full cryo-EM model, we adopt the simplified models in (1.1) and (1.2) to more effectively communicate the primary insights.

The gap. The common approach to estimating the rotation of an observation in the models above involves scanning through a pre-defined set of possible rotations and selecting the one that either maximizes the correlation or minimizes the distance to the given reference (2D projection image or 3D molecular structure, depending on the application). Typically, these metrics involve weighted correlations and distances, where the weights account for the noise characteristics [49].

From an estimation theory perspective, this process corresponds to the maximum likelihood estimator (MLE), which does not incorporate prior information about the *rotation distribution*, i.e., the distribution of the 3D rotation g appearing in models (1.1) or (1.2). A natural extension is the maximum a posteriori (MAP) estimator, which combines the data likelihood with a prior distribution on rotations. Notably, when a uniform prior is assumed, indicating that all orientations

are equally likely a priori, the MAP estimator reduces to the MLE estimator [55]. A more rigorous treatment can be found in Section 2.3.

However, Bayesian theory provides a much deeper and richer statistical framework that leads to improvement: replacing the MLE or MAP estimator with the *Bayes estimator*. The full potential of this estimator, that provides optimal accuracy according to a user-defined loss function and allows for integrating prior knowledge about *rotation distribution*, remained untapped so far.

The term *rotation distribution*, used interchangeably with orientation distribution in this paper, refers to the probabilistic law governing how orientations are distributed over all possible rotations in 3D space. In cryo-EM, this concept is critical because molecules are first suspended in a thin aqueous solution and subsequently vitrified into amorphous ice, at which point their orientations become fixed but remain uncontrolled. Ideally, all orientations would be equally likely, corresponding to a uniform distribution. In practice, however, molecules often adopt *preferred orientations*, typically due to interactions with the air–water interface or other sample-preparation factors [57, 36, 34]. As a result, certain orientations are disproportionately represented, introducing systematic biases that can degrade reconstruction quality and complicate orientation estimation. Accounting for these deviations from uniformity is therefore essential for both methodological development and the interpretation of cryo-EM data.

Currently, the most practical way to obtain orientation estimates is through modules of reconstruction software such as RELION [49] and cryoSPARC [44]. These packages are primarily designed for 3D structure reconstruction and therefore treat pose estimation as a latent-variable subproblem. In particular, during refinement they compute posterior weights over a discrete set of candidate poses for each particle and use these weights for weighted averaging in the reconstruction step (a soft-assignment over poses). At the same time, for downstream usage and reporting, the metadata typically provides a *single* orientation per particle, which is commonly the MLE estimator, that is, the maximizer of the pose posterior. Thus, while these pipelines do incorporate latent-variable pose weighting internally for reconstruction, they do not typically output a Bayesian rotation estimate for each particle.

In practice, researchers frequently use the reported per-particle orientation point estimates for downstream tasks such as heterogeneity analysis [13, 64] and structural validation [46]. However, relying solely on MLE-based estimates can be suboptimal, especially in low-SNR regimes where MLE estimators are known to perform poorly. This work therefore explores opportunities to improve per-particle orientation estimation by using Bayes estimators that explicitly average over pose uncertainty that incorporate prior knowledge of the rotation distribution.

The Bayesian framework. The Bayesian framework has become a powerful and widely adopted tool in cryo-EM, now recognized as the leading method for recovering 3D molecular structures [49, 44, 58, 13]. It effectively addresses challenges like overfitting and parameter tuning while enhancing interpretability [48]. By explicitly modeling uncertainties, Bayesian methods enable more accurate and robust reconstructions of molecular structures, driving significant advances in both resolution and structural flexibility. While the majority of these methods focus on the task of structure reconstruction (e.g., see Section 4) and aim to achieve the MAP estimator for the volume structure [49, 44], the Bayesian framework offers broader possibilities for addressing other problems.

This framework enables the use of any loss function over the rotational group tailored to users' requirements and accommodates a broad range of prior distributions, beyond the uniform distribution. While the Bayesian framework can be adapted to different loss functions, this work focuses on the mean-squared error loss, which is equivalent to the chordal distance between 3D rotations. The

primary reason for using this loss function is that its corresponding Bayes estimator has a closed-form analytical solution, making it computationally efficient and easy to interpret. We denote this estimator as $\hat{g}_{\mathrm{MMSE}}$, where MMSE stands for minimum mean square error. It is worth noting that for any given loss function and prior distribution, the Bayes estimator is optimal among all possible estimators, in the sense that it minimizes the posterior expected loss. Moreover, the computational cost of calculating our $\hat{g}_{\mathrm{MMSE}}$ estimator is at the same scale as that of the commonly used MLE and MAP orientation estimators (See Section 3 for more details).

Applications. Orientation estimation is not only integral to 3D structure reconstruction but also underpins a range of downstream tasks. Below, we highlight several representative applications that demonstrate its broader methodological significance.

One prominent example is heterogeneity analysis, where the goal is to capture structural variability that may arise from differences in composition, discrete states, or continuous conformational changes. A substantial portion of state-of-the-art methods [13, 64, 33, 35, 44, 42] make the simplifying assumption that particle poses are already known and fixed, typically obtained from a consensus refinement procedure; these pose estimates then form the basis for subsequent modeling of structural variability. In this work, we consider continuous conformational heterogeneity and demonstrate in Section 5 that the quality of orientation estimation has a direct impact on the fidelity of the recovered conformational landscapes.

Another important application is 3D volume alignment. Unlike the setting of (1.2), where an observation is aligned to a known reference volume, this task involves aligning two noisy volumes of the same structure with unknown relative rotations. This scenario arises, for instance, in the computation of the Fourier Shell Correlation (FSC) curve, a standard tool for estimating the spectral signal-to-noise ratio (SSNR) and resolution. As shown in Appendix C, this problem can be expressed within the same statistical framework as (1.2), differing only by an additional noise term, which highlights its close methodological connection to orientation estimation.

A further noteworthy application concerns validation. A classic example is the method of tilt pairs introduced by Rosenthal and Henderson [46], in which a small subset of images (e.g., around ten) is collected at known relative tilt angles. While the absolute particle orientations remain unknown, the relative orientations between tilt pairs are specified by the experimental geometry. The degree of agreement between estimated and known relative rotations then provides a direct validation of orientation assignment accuracy.

These examples are by no means exhaustive, but they underscore that orientation estimation is not merely a technical nuisance in reconstruction workflows, but rather a key methodological component that enables diverse downstream analyses and validation strategies across cryo-EM and cryo-ET.

Overview of results and contributions. In this work, we introduce a versatile Bayesian framework for orientation estimation with strong statistical guarantees and high flexibility. Section 2 formulates the problem and develops the Bayesian MMSE estimator. We show theoretically that the MMSE estimator coincides with the MLE estimator in high-SNR regimes (Proposition 2.2), while consistently outperforming it in low-SNR conditions (Figure 1), a typical scenario in cryo-EM and cryo-ET [8].

Section 3 evaluates the MMSE orientation estimator through simulations. We investigate the effect of different prior distributions and the discretization resolution L of the rotation group $\mathsf{SO}(3)$. Results indicate that non-uniform priors substantially improve estimation accuracy, underscoring the value of incorporating prior knowledge (Figure 2). We further show that in high-SNR regimes

the estimation error scales as $L^{1/3}$ (Figure 3), highlighting discretization as the dominant error source, whereas in low-SNR regimes noise dominates.

In Section 4, we connect the MMSE estimator to 2D image recovery and 3D structure reconstruction, showing that it naturally leads to the expectation-maximization (EM) framework for models without projections (Proposition 4.1). Empirical results confirm that the MMSE estimator consistently outperforms the MLE and MAP estimators in both tasks, providing higher accuracy and greater robustness against the "Einstein from Noise" artifact [19, 3] (Figures 4–5).

Finally, Section 5 integrates MMSE-based pose estimation into structural heterogeneity analysis. Following the fixed-pose framework of RECOVAR, we show that MMSE pose estimator consistently improves the recovery of conformational variability compared to the MLE counterpart. Section 6 concludes with a discussion of further applications of the Bayesian approach to orientation estimation and future research directions.

Main takeaways. Figure 1 illustrates the critical role of orientation estimation in the cryo-EM and cryo-ET reconstruction processes, showcasing the superior performance of the MMSE orientation estimator, $\hat{g}_{\mathrm{MMSE}}$, compared to the MLE orientation estimator, $\hat{g}_{\mathrm{MLE}}$. Specifically, the curves show that the MMSE estimator consistently produces more accurate estimates than the MLE estimator, with the performance gap widening as the SNR decreases. Moreover, incorporating prior knowledge of *rotation distribution* into our MMSE estimator allows for even better performance (See Figure 2). This performance gap becomes even more pronounced when the orientation estimators are incorporated into a reconstruction algorithm (see Section 4).

Beyond reconstruction accuracy, we demonstrate that orientation estimation quality has a direct and substantial impact on downstream structural heterogeneity analysis. Understanding structural variability is essential for characterizing the dynamic behavior of macromolecular complexes. Methods such as RECOVAR [13], designed to recover structural heterogeneity, typically assume known particle poses (i.e., fixed-pose methods). In reality, these poses must be inferred, and MLE-derived estimates are commonly used in practice. We show that replacing MLE-derived poses with MMSE estimates in RECOVAR substantially improves the recovery of the latent conformational manifold, bringing results closer to those obtained using ground-truth poses. These findings establish MMSE orientation estimation not only as a tool that enhances reconstruction fidelity, but also as a facilitator that advances state-of-the-art continuous heterogeneity analysis.

The main takeaway of this paper is the recommendation to adopt the Bayesian MMSE orientation estimator for determining the orientation of each observation, in place of the commonly used MLE estimator. The MMSE estimator demonstrates superior performance even under a uniform rotation distribution and offers further improvements when incorporating prior knowledge of the underlying rotation distribution. Importantly, while current software packages do not directly compute the MMSE estimator, they already calculate all the necessary components required for its implementation. Therefore, adopting the MMSE approach can be achieved with minimal additional computational cost. The detailed implementation and code are available at https://github.com/AmnonBa/bayesian-orientation-estimation.

2 Problem formulation and the MMSE orientation estimator

In this section, we present a particular Bayes estimator for orientation determination under a squared-error loss, commonly known as the *minimum mean-square error* (MMSE) estimator. The name reflects that this estimator minimizes the expected posterior mean-squared error, and it is given by the posterior (conditional) mean rather than a pointwise maximization [26].

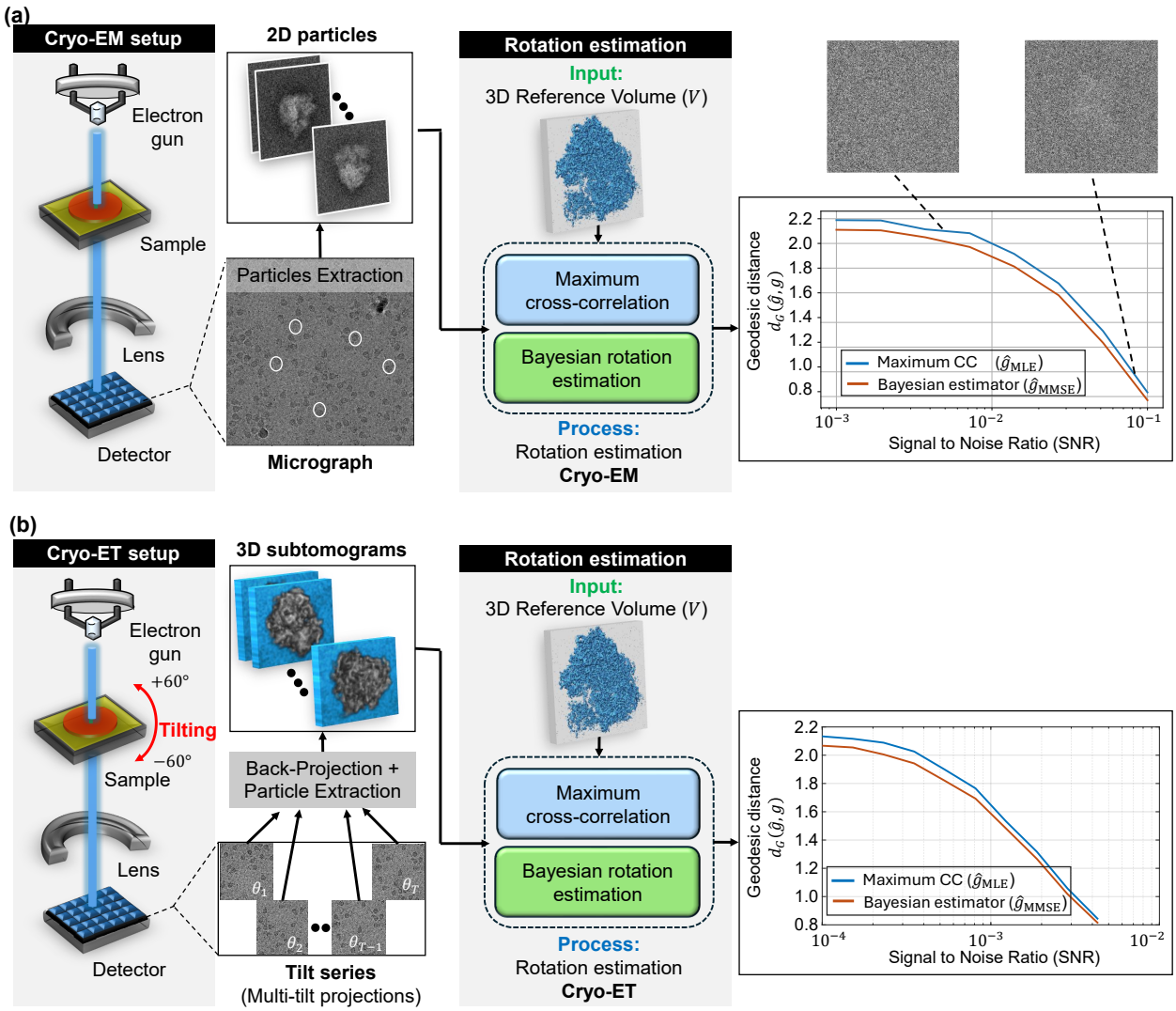


Figure 1: Comparison between maximum cross-correlation ($\hat{g}_{\mathrm{MLE}}$) and Bayesian ($\hat{g}_{\mathrm{MMSE}}$) orientation estimators in cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) applications. The figure illustrates the general workflow in cryo-EM and cryo-ET techniques, highlighting the role of orientation estimation in each technique. Panel **(a)** illustrates the model *with 2D projections* (single-particle cryo-EM model, (1.1)), while panel **(b)** shows the model of the subtomogram averaging in cryo-ET, (1.2)). **(a) Cryo-EM** involves imaging macromolecules embedded in a thin layer of vitreous ice using an electron beam in a transmission electron microscope (TEM). The process generates 2D projection images (micrographs) of particles in unknown 3D orientations. These 2D particles are then identified and extracted from the micrographs, forming the basis for subsequent steps of the macromolecule's 3D structure reconstruction. **(b) Cryo-ET** involves imaging a sample from multiple known tilt angles (typically from $-60°$ to $+60°$) to create 2D projections, which are then combined computationally to reconstruct 3D subtomograms. In this context, a subtomogram refers to a small volume containing an individual 3D particle. The subtomograms are extracted by a particle picker algorithm for further analysis. **(a+b):** The rotation estimation problem involves determining the relative orientation of a noisy 2D particle (in cryo-EM) or a noisy 3D subtomogram (in cryo-ET) relative to a reference volume V. The reference volume structure used in both setups is identical and corresponds to the 80S ribosome [61]. Under high SNR conditions, both rotation estimators closely approximate the true relative rotation. However, as the SNR decreases, the estimation accuracy deteriorates. Importantly, across all SNR levels, the geodesic angular distance between the MMSE orientation estimator and the true rotation consistently remains lower than that of the MLE orientation estimator. For **(a)**, the estimation was conducted using a grid size of $L = 3000$ samples of the rotation group $\mathsf{SO}(3)$, while for **(b)**, a grid size of $L = 300$ was used. Each point in the two curve plots represents the average error computed over 3000 trials.

We begin by introducing a flexible mathematical model that encompasses various typical applications involving orientation estimation. Following this, we present a couple of metrics designed to assess the quality of our estimators, providing a robust framework for evaluating performance. Finally, we introduce the class of Bayes estimators, with the MMSE estimator serving as a primary example.

Throughout this paper, we use g to denote both the rotation operator and its corresponding matrix representation. For instance, g can be represented by a three by three rotation matrix in 3D. The intended meaning will be clear from the context, and this slight abuse of notation should not cause confusion.

2.1 Mathematical model for orientation estimation

We consider a unified framework for modeling measurement processes, encompassing problems such as 2D template matching, orientation estimation in cryo-EM, and subtomogram averaging in cryo-ET. To better focus on the core aspects of our methodology, we omit certain physical effects, such as CTF and in-plane shifts, in this formulation. A complete mathematical model of the cryo-EM imaging process, incorporating these effects, is provided in Appendix A.

We begin with a continuous-domain formulation. Let $V(x) : \mathbb{R}^n \to \mathbb{R}$ denote a reference structure, where $n = 2$ for 2D images and $n = 3$ for 3D volumes. For example, when $n = 3$, V corresponds to the continuous 3D electron density, as commonly used in cryo-EM or cryo-ET. Let $\mathsf{G} \subset \mathsf{SO}(n)$ be a compact group of rotations correspondingly, and let $g \in \mathsf{G}$ be an unknown transformation drawn from a distribution Λ over G.

We denote the measurement as a function $y(x') : \mathbb{R}^m \to \mathbb{R}$, where $m \leq n$ accounts for possible dimension reduction due to projection (e.g., $m = 2$ in cryo-EM). Then, the continuous measurement model is given by

$$y(x') = \Pi\Big((g \circ V)(x)\Big) + \varepsilon(x'), \qquad x \in \mathbb{R}^n \text{ and } x' \in \mathbb{R}^m,$$

where $(g \circ V)(x) := V(g^{-1}x)$ is the rotated structure, Π is a known linear operator (such as a tomographic projection or identity), and ε denotes additive noise in the measurement domain.

In practice, measurements are only available in discretized form due to finite resolution. In the discrete setting, we assume that the operator Π also incorporates sampling onto a grid of size d, i.e., it includes both projection (if applicable) and discretization. The resulting discrete measurement model becomes

$$y_i = \Pi(g_i \circ V) + \varepsilon_i, \tag{2.1}$$

where $y_i \in \mathbb{R}^d$, and $\varepsilon_i \sim \mathcal{N}(0, \Sigma)$ is Gaussian noise with $d \times d$ positive-definite covariance matrix Σ. To simplify notation, we omit Π when it solely represents the discretization sampling operator and use the same symbol y to refer to the discretized measurement throughout the paper, unless stated otherwise.

Applications of the model. We present three typical examples of this model. In all cases, the goal is to estimate g given the sample y, the structure V, and the covaraince matrix Σ.

1. *2D template matching.* In this case, Π is solely the discretization sampling operator, $d = N \times N$ with N the grid size of the 2D image, $g \in \mathsf{G} = \mathsf{SO}(2)$ is a 2D in-plane rotation, and V is a given 2D template image.

2. *Rotation estimation in cryo-EM.* Here, we consider a special case of (2.1) where Π comprises both the sampling and tomographic projection operators, $d = N \times N$ with N the grid size of 2D projection images, $g \in \mathsf{G} = \mathsf{SO}(3)$ is a 3D rotation, and V is a given 3D volume representing a known reference 3D structure or a well-grounded structure from prior data analyses.

3. *3D structure alignment in cryo-ET.* In this scenario, we consider a special case of (2.1) where Π is solely the discretization sampling operator, d corresponds to the total dimension of 3D subtomograms, $g \in \mathsf{G} = \mathsf{SO}(3)$ is a 3D rotation, and V is a given 3D volume. Notably, the 3D alignment problem is also a critical step in the computational pipeline of cryo-EM, see, e.g., [53, 17].

2.2 Preliminaries

Before introducing the specific estimator for the model specified in (2.1), it is instructive to briefly revisit the general Bayesian framework and the concept of the Bayes estimator. This will provide the necessary foundation for understanding the development and analysis of the proposed MMSE estimator, as well as its statistical properties superior to those of the widely used MLE estimator.

Overview of the Bayesian framework and the Bayes estimators. Suppose that we aim to estimate a true rotation $g \in \mathsf{G}$ drawn from a known prior distribution Λ. Let $\hat{g}$ be an estimator of g based on a measurement y and let $\mathcal{L}(g, \hat{g})$ be a loss function. The *Bayes risk* of $\hat{g}$ is defined as $\mathbb{E}_\Lambda[\mathcal{L}(g, \hat{g})]$, where the expectation is taken over the data generation process of y given g and the prior distribution Λ of g. The *Bayes estimator* with respect to the loss $\mathcal{L}$ [32, Chapter 4, Theorem 1.1] is defined as the estimator that minimizes the Bayes risk among all possible estimators, i.e.,

$$\hat{g}_{\mathsf{Bayes}} = \arg\min_{\hat{g}} \mathbb{E}_\Lambda \left[\mathcal{L}(g, \hat{g})\right]. \tag{2.2}$$

Equivalently, it is the estimator that minimizes the posterior expected loss $\mathbb{E}_\Lambda[\mathcal{L}(g, \hat{g})\,|y]$, where the expectation is taken over the posterior distribution of g given the measurement y, and any other known parameter such as volume V, with the prior Λ. For the case of $G = \mathsf{SO}(3)$, it is given explicitly by

$$\begin{aligned}\hat{g}_{\mathsf{Bayes}} &= \arg\min_{\hat{g}\in\mathsf{SO}(3)} \mathbb{E}_\Lambda \left[\mathcal{L}(g, \hat{g})\,|y, V\right] \\ &= \arg\min_{\hat{g}\in\mathsf{SO}(3)} \int_{\mathsf{SO}(3)} \mathcal{L}(g, \hat{g})\, \mathbb{P}_\Lambda(g|y, V)\mathrm{d}g,\end{aligned} \tag{2.3}$$

where by Bayes' law, we have,

$$\mathbb{P}_\Lambda(g|y, V)\mathrm{d}g = \frac{\mathbb{P}(y|g, V)\mathrm{d}\Lambda(g)}{\int_{\mathsf{G}} \mathbb{P}(y|g, V)\mathrm{d}\Lambda(g)}. \tag{2.4}$$

The posterior distribution of the rotation g given an observation y under model (2.1). To introduce the MMSE estimator corresponding to the model (2.1), we first compute the posterior distribution of g given y and all the additional parameters V, and Σ. We obtain the conditional likelihood density

$$\mathbb{P}(y|g, V) = \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}} \exp\Big(-\frac{1}{2}(y - \Pi(g \circ V))^{\mathsf{T}}\Sigma^{-1}(y - \Pi(g \circ V))\Big). \tag{2.5}$$

Note that g follows the underlying prior distribution Λ. Applying Bayes' law, the posterior distribution of g given y is

$$\mathbb{P}_\Lambda(g|y,V)\mathrm{d}g = \frac{\exp\Big(-\frac{1}{2}(y-\Pi(g\circ V))^\mathsf{T}\Sigma^{-1}(y-\Pi(g\circ V))\Big)\mathrm{d}\Lambda(g)}{\int_\mathsf{G}\exp\Big(-\frac{1}{2}(y-\Pi(g\circ V))^\mathsf{T}\Sigma^{-1}(y-\Pi(g\circ V))\Big)\mathrm{d}\Lambda(g)}. \tag{2.6}$$

Metrics over SO(3). As the Bayes estimator is closely related to the given loss function $\mathcal{L}$, we present two candidate metrics on the 3D special orthogonal group $\mathsf{SO}(3)$. Here we represent any rotation $g \in \mathsf{SO}(3)$ in its natural three by three matrix representation.

1. *Chordal distance.* For any two rotations $g_1, g_2 \in \mathsf{SO}(3)$, the chordal distance is defined as

$$d_F(g_1,g_2) \triangleq \|g_1-g_2\|_\mathrm{F} = \sqrt{\mathrm{tr}\Big((g_1-g_2)^\mathsf{T}(g_1-g_2)\Big)}, \tag{2.7}$$

   where $\|\cdot\|_\mathrm{F}$ represents the matrix Frobenius norm, and $\mathrm{tr}(\cdot)$ is the trace of a matrix. This metric is easy to compute and analyze; however, it does not take into account the group structure of rotations.

2. *Geodesic distance.* For any two rotations $g_1, g_2 \in \mathsf{SO}(3)$, the geodesic distance is defined as

$$d_G(g_1,g_2) \triangleq \arccos\left(\frac{\mathrm{tr}(g_2 g_1^{-1})-1}{2}\right), \tag{2.8}$$

   where $\mathrm{tr}(\cdot)$ is the trace of a matrix. This metric reflects the shortest path between g_1 and g_2 in the 3D rotation manifold.

We note that the chordal and geodesic distances are closely related. In particular, if $\theta \triangleq d_G(g_1,g_2)$ denotes the geodesic rotation angle between g_1 and g_2 (2.8), then

$$d_F(g_1,g_2) = \|g_1-g_2\|_F = 2\sqrt{2}\,\sin(\theta/2) = 2\sqrt{2}\,\sin\big(d_G(g_1,g_2)/2\big). \tag{2.9}$$

Consequently, d_F is a strictly increasing function of d_G and, for small angular errors, the two are locally equivalent in the sense that $d_F(g_1,g_2) \approx \sqrt{2}\,d_G(g_1,g_2)$ (see, e.g., [18]). More generally, our framework is flexible and can accommodate other loss functions over any group G. For a comprehensive discussion of metrics on $\mathsf{SO}(3)$ and their relationships, we refer the reader to [18, 21]. Similar considerations apply to $\mathsf{SO}(2)$; we omit a detailed discussion for brevity.

2.3 The MLE, MAP and MMSE orientation estimators

In the following presentation, we focus on independent and identically distributed Gaussian noise with $\Sigma = \sigma^2 I_{d\times d}$. These noise statistics are commonly employed in modern software tools [29]. While the proposed framework can incorporate more advanced noise models, we restrict attention to this case for clarity of presentation and to simplify the theoretical analysis.

The MLE estimator. Recalling the maximum cross-correlation method we mentioned earlier, we now connect it with the mathematical model in (2.1) and introduce the corresponding MLE estimator. The rotation g that minimizes the distance between the corresponding projected rotated volume $\Pi(g\circ V)$ and the observation y is exactly the MLE estimator defined as

$$\hat{g}_\mathsf{MLE} := \arg\max_{g\in\mathsf{SO}(3)} \mathbb{P}(y|g,V)$$

$$
\begin{aligned}
&= \arg\max_{g \in \mathsf{SO}(3)} \exp\Big(- \|y - \Pi(g \circ V)\|^2 / 2\sigma^2 \Big) \\
&= \arg\min_{g \in \mathsf{SO}(3)} \|y - \Pi(g \circ V)\|^2, \qquad (2.10)
\end{aligned}
$$

where the first equality follows from (2.5), and the second equality follows from the assumption $\Sigma = \sigma^2 I_{d\times d}$. In other words, $\hat{g}_{\mathsf{MLE}}$ maximizes the conditional density (2.5) over all possible rotations in $\mathsf{SO}(3)$. In the absence of the tomographic projection (e.g., in cryo-ET), the MLE estimator further simplifies to

$$
\hat{g}_{\mathsf{MLE}} = \arg\max_{g \in \mathsf{SO}(3)} y^{\mathsf{T}} (g \circ V),
$$

which corresponds to the rotation g such that the rotated structure maximizes the correlation with y. This estimator is also frequently used in single-particle cryo-EM, under the assumption that the norm of $\Pi(g \circ V)$ is approximately constant for all g. This estimator can be approximated by performing a search over a pre-defined grid of 3D rotations, selecting the rotation that minimizes the distance between the measurement y and the projected rotated volume $\Pi(g \circ V)$. This approach forms the basis of standard practices in single-particle cryo-EM and cryo-ET.

The MAP estimator. The MAP estimator extends the MLE estimator, by incorporating prior knowledge on the distribution of the rotation in the special group $\mathsf{SO}(3)$, $\mathrm{d}\Lambda(g)$. Formally, the MAP estimator is defined as

$$
\begin{aligned}
\hat{g}_{\mathsf{MAP}} &= \arg\max_{g \in \mathsf{SO}(3)} \mathbb{P}_{\Lambda}(g|y, V) \\
&= \arg\max_{g \in \mathsf{SO}(3)} \frac{\mathbb{P}(y|g, V)\mathrm{d}\Lambda(g)}{\int_{\mathsf{G}} \mathbb{P}(y|g, V)\mathrm{d}\Lambda(g)} \\
&= \arg\max_{g \in \mathsf{SO}(3)} \mathbb{P}(y|g, V)\mathrm{d}\Lambda(g) \\
&= \arg\max_{g \in \mathsf{SO}(3)} \exp\Big(- \|y - \Pi(g \circ V)\|^2 / 2\sigma^2 \Big) \mathrm{d}\Lambda(g) \\
&= \arg\min_{g \in \mathsf{SO}(3)} \|y - \Pi(g \circ V)\|^2 / 2\sigma^2 - \log\left(\mathrm{d}\Lambda(g)\right), \qquad (2.11)
\end{aligned}
$$

where the second equality follows from Bayes' law (2.4), and the third equality from independence of the denominator on g. In other words, $\hat{g}_{\mathsf{MAP}}$ maximizes the posterior density (2.6) over all possible rotations in $\mathsf{SO}(3)$. It can be seen that the MAP estimator $\hat{g}_{\mathsf{MAP}}$ (2.11) coincide with the MLE estimator $\hat{g}_{\mathsf{MLE}}$ (2.10) when the prior distribution $\mathrm{d}\Lambda(g)$ is uniform over $\mathsf{SO}(3)$.

The MMSE estimator. For any rotation distribution Λ and loss function $\mathcal{L}(\cdot, \cdot)$ over $\mathsf{SO}(3)$, the MLE and MAP estimators can be further improved by the corresponding Bayes estimator which minimizes the posterior expected loss $\mathbb{E}_{\Lambda}[\mathcal{L}(g, \cdot) \mid y, V]$. In particular, following the definition (2.3) of the Bayes estimator, and for the chordal distance $d_F(\cdot, \cdot)$ (2.7), with its corresponding squared loss $\mathcal{L}_F(\cdot, \cdot) = d_F^2(\cdot, \cdot)$, the Bayes estimator takes the form

$$
\hat{g}_{\mathsf{MMSE}} := \arg\min_{\hat{g} \in \mathsf{SO}(3)} \mathbb{E}_{\Lambda}\left[d_F^2\left(g, \hat{g}\right) \mid y, V \right], \qquad (2.12)
$$

where the expectation is taken over the posterior density (2.6). Here we denote the estimator as $\hat{g}_{\mathsf{MMSE}}$, since minimizing the squared chordal distance is equivalent to minimizing the entrywise

mean-square error between the two 3×3 rotation matrices. Alternative distance measures, such as the geodesic distance on $\mathsf{SO}(3)$ discussed earlier, could certainly be considered. In this work, however, we focus on the chordal distance, primarily because it admits a particularly convenient derivation of the associated Bayes estimator, as shown below, thereby offering both computational efficiency and interpretability. Importantly, these Bayes estimators are generally distinct from both the MLE and MAP estimators—even under a uniform prior—since it minimizes the posterior expected loss rather than selecting the most probable rotation.

To compute $\hat{g}_{\mathsf{MMSE}}$, we begin by noting that it is defined over the manifold $\mathsf{SO}(3)$. In general, directly solving the optimization problem on a non-Euclidean manifold is challenging. To address this, we first relax the optimization domain to the ambient Euclidean space $\mathbb{R}^{3\times 3}$, which leads to the following intermediate estimator:

$$\begin{aligned}
\hat{g}_{\mathsf{relax}} &:= \operatorname*{arg\,min}_{\hat{g}\in\mathbb{R}^{3\times 3}} \mathbb{E}_{\Lambda}\left[d_F^2\left(g,\hat{g}\right) \,\middle|\, y, V\right] \\
&= \mathbb{E}_{\Lambda}[g \mid y, V] \\
&= \frac{\int_{\mathsf{SO}(3)} g \exp\Big(-\frac{1}{2}(y-\Pi(g\circ V))^{\mathsf{T}}\Sigma^{-1}(y-\Pi(g\circ V))\Big)\,\mathrm{d}\Lambda(g)}{\int_{\mathsf{SO}(3)} \exp\Big(-\frac{1}{2}(y-\Pi(g\circ V))^{\mathsf{T}}\Sigma^{-1}(y-\Pi(g\circ V))\Big)\,\mathrm{d}\Lambda(g)} \\
&= \frac{\int_{\mathsf{SO}(3)} g \exp\Big(-\frac{1}{2\sigma^2}\left\|y-\Pi(g\circ V)\right\|_2^2\Big)\,\mathrm{d}\Lambda(g)}{\int_{\mathsf{SO}(3)} \exp\Big(-\frac{1}{2\sigma^2}\left\|y-\Pi(g\circ V)\right\|_2^2\Big)\,\mathrm{d}\Lambda(g)}.
\end{aligned} \tag{2.13}$$

Here, since we naturally represent $g \in \mathsf{SO}(3)$ as a 3×3 rotation matrix, the expectation (or integral) can be interpreted directly in this matrix representation, yielding a 3×3 matrix as the posterior mean matrix[1]. However, $\hat{g}_{\mathsf{relax}}$ does not, in general, lie in $\mathsf{SO}(3)$ and hence may not be a valid rotation operator. To obtain a feasible solution on the manifold, we subsequently apply the *orthogonal Procrustes* procedure to project $\hat{g}_{\mathsf{relax}}$ back onto $\mathsf{SO}(3)$, as discussed next.

Given a matrix A, to obtain a valid rotation matrix we seek $\Omega \in \mathsf{SO}(3)$ that is closest to A in the Frobenius norm:

$$\operatorname*{minimize}_{\Omega} \quad \|\Omega - A\|_F \quad \text{subject to} \quad \Omega \in \mathsf{SO}(3). \tag{2.14}$$

This optimization problem is known as the *Orthogonal Procrustes* problem in the linear algebra literature. It can be solved efficiently using the singular value decomposition (SVD) of A [14]. Concretely, given a matrix $A \in \mathbb{R}^{3\times 3}$ (in our case $A = \hat{g}_{\text{relax}}$), compute its singular value decomposition

$$A = U\Sigma V^{\top}, \qquad U, V \in O(3),\ \Sigma \succeq 0. \tag{2.15}$$

The closest rotation matrix in Frobenius norm is then obtained by

$$\Omega = U\,\mathrm{diag}\big(1,\,1,\,\det(UV^{\top})\big)\,V^{\top} \ \in \mathsf{SO}(3). \tag{2.16}$$

Equivalently, if $\det(UV^{\top}) = -1$, one flips the sign of the last column of U (or V) before forming $UV^{\top}$, ensuring $\det(\Omega) = 1$ and yielding the Procrustes solution. Thus, applying the Orthogonal Procrustes procedure to $\hat{g}_{\mathsf{relax}}$ projects it back onto $\mathsf{SO}(3)$, yielding a valid rotation matrix. The following proposition establishes that this valid rotation matrix coincides with the MMSE estimator $\hat{g}_{\mathsf{MMSE}}$, thereby completing the procedure for obtaining the MMSE estimate. A detailed proof is provided in Appendix D.

[1] More generally, if one works with different representations of $\mathsf{SO}(3)$, or interprets the expectation as acting on operators (as in Proposition 4.1), the resulting estimator may take different forms.

Proposition 2.1. *The Bayes estimator of the loss function defined by the chordal distance* $\hat{g}_{\mathsf{MMSE}}$ (2.12) *is equal to the Orthogonal Procrustes solution* (2.14) *applied on the intermediate estimator* $\hat{g}_{\mathsf{relax}}$*, i.e.,*

$$\hat{g}_{\mathsf{MMSE}} = \arg\min_{\Omega \in \mathsf{SO}(3)} \ \|\hat{g}_{\mathsf{relax}} - \Omega\|_F \,. \tag{2.17}$$

The MMSE, MAP and MLE estimators in the high SNR regime. We conclude the section by showing the next proposition. It shows that the MMSE and MAP estimators converge in the high SNR regime (i.e., $\sigma \to 0$), as demonstrated empirically in Figures 3, 4, and 5. This implies that the MMSE estimator's superior statistical properties are most advantageous in low SNR conditions, which are common in structural biology applications like cryo-EM and cryo-ET [8]. In these low SNR environments, the MMSE estimator consistently outperforms its MAP counterpart.

Proposition 2.2. *Let* Λ *be any distribution over* $\mathsf{SO}(3)$ *with a strictly positive density, bounded below by some constant* $c > 0$*. Let* $\hat{g}_{\mathsf{MLE}}$*,* $\hat{g}_{\mathsf{MAP}}$ *be the MLE and MAP rotation estimators as defined in* (2.10) *and* (2.11)*, respectively, and assume they are unique. Let* $\hat{g}_{\mathsf{MMSE}}$ *be the MMSE estimator as defined in* (2.12)*. Then, we have,*

$$\lim_{\sigma \to 0} \hat{g}_{\mathsf{MMSE}} = \lim_{\sigma \to 0} \hat{g}_{\mathsf{MAP}} = \hat{g}_{\mathsf{MLE}}. \tag{2.18}$$

The proof of the proposition is provided in Appendix E. The proof relies on the existing result, [45, Theorem 5.10], and is valid not only to rotations but also to other group operators, such as translations.

3 Numerical methods for MMSE orientation estimation

This section compares the numerical performance of the MMSE and MLE estimators. We also introduce various types of prior distributions for $\mathsf{SO}(3)$ beyond the uniform distribution and demonstrate how incorporating this prior knowledge can significantly improve the performance of the MMSE rotation estimator. Furthermore, we study the influence of the number of sampling points of the group $\mathsf{SO}(3)$ of 3D rotations (namely, the number of candidate rotations), denoted as L, on the quality of rotation estimation. For better illustration, we consider the simplified setting of model (2.1), where $\Sigma = \sigma^2 I_{d\times d}$, although our method applies to the general covariance setting as well.

Numerical procedure and sampling of $\mathsf{SO}(3)$. Since the posterior distribution of $g \in \mathsf{SO}(3)$ in (2.6) is continuous, a numerical discretization of the rotation group is required. In this work, the expectation is approximated using a quadrature rule over $\mathsf{SO}(3)$. Let $\{(g^{(\ell)}, w_\ell)\}_{\ell=0}^{L-1}$ denote a numerical quadrature on $\mathsf{SO}(3)$, constructed as a product of a spherical quadrature on S^2 and a circular rule over the Euler angles [16, 15]. The implementation employed here follows the publicly available MATLAB code described in [20].

We first consider the uniform case, where Λ is the Haar distribution on $\mathsf{SO}(3)$. For each quadrature node, define $x_\ell := \Pi\big(g^{(\ell)} \circ V\big)$, for $0 \le \ell \le L-1$, where $\{g^{(\ell)}\}_{\ell=0}^{L-1}$ are the quadrature nodes and $\{w_\ell\}_{\ell=0}^{L-1}$ are their associated weights. The MMSE estimator can then be expressed as the weighted average

$$\hat{g}_{\mathsf{MMSE}} = \mathbb{E}[g \mid y, V] \approx \frac{\displaystyle\sum_{\ell=0}^{L-1} g^{(\ell)}\, w_\ell \exp\Big(-\tfrac{\|y - x_\ell\|^2}{2\sigma^2}\Big)}{\displaystyle\sum_{\ell=0}^{L-1} w_\ell \exp\Big(-\tfrac{\|y - x_\ell\|^2}{2\sigma^2}\Big)} = \sum_{\ell=0}^{L-1} g^{(\ell)}\, p^{(\ell)}, \tag{3.1}$$

Algorithm 1 `Numerical solver for MMSE rotation estimator`

Input: Observation y following (2.1), and a set of candidate rotations $\{g^{(\ell)}\}_{\ell=0}^{L-1}$ of $\mathsf{SO}(3)$.
Output: MMSE estimator of the rotation, $\hat{g}_{\mathsf{MMSE}}$.

1. Compute:

$$\hat{g}_{\mathsf{relax}} = \frac{\sum_{\ell=0}^{L-1} g^{(\ell)} \exp\Big(- \left\| y - \Pi\left(g^{(\ell)} \circ V\right) \right\|^2 / 2\sigma^2 \Big)}{\sum_{\ell=0}^{L-1} \exp\Big(- \left\| y - \Pi\left(g^{(\ell)} \circ V\right) \right\|^2 / 2\sigma^2 \Big)}. \tag{3.5}$$

2. Perform the Orthogonal Procrustes algorithm (2.14) on $\hat{g}_{\mathsf{relax}}$ in (3.5) to get the proper rotation estimator $\hat{g}_{\mathsf{MMSE}}$.

where the posterior quadrature weights are defined as

$$p^{(\ell)} \triangleq \frac{w_\ell \exp\Big(- \frac{\|y - x_\ell\|^2}{2\sigma^2} \Big)}{\sum_{r=0}^{L-1} w_r \exp\Big(- \frac{\|y - x_r\|^2}{2\sigma^2} \Big)}, \qquad 0 \le \ell \le L-1. \tag{3.2}$$

For a general non-uniform prior $\Lambda(g)$, the same quadrature can be applied with modified posterior weights:

$$p^{(\ell)} \propto w_\ell \, \Lambda\big(g^{(\ell)}\big) \, \exp\Big(- \frac{\|y - x_\ell\|^2}{2\sigma^2} \Big), \tag{3.3}$$

followed by normalization to ensure $\sum_{\ell=0}^{L-1} p^{(\ell)} = 1$.

Similarly, the MLE estimator in (2.10) relies on the discretization of $\mathsf{SO}(3)$, which can be approximated through grid search using the uniform rotation samples or pre-defined grid $\{g^{(\ell)}\}_{\ell=0}^{L-1}$ as

$$\hat{g}_{\mathsf{MLE}} = \arg\min_{0 \le \ell \le L-1} \|y - x_\ell\|^2 . \tag{3.4}$$

Both estimators require evaluating $\|y - x_\ell\|$ for all candidate rotations, which naively costs $O(d)$ per candidate and thus $O(Ld)$ overall. This estimate, however, overlooks the structure of $\mathsf{SO}(3)$. Any 3D rotation can be decomposed into a *viewing direction* (two degrees of freedom, discretized with about $L^{2/3}$ samples) and an *in-plane rotation* about that direction (one degree of freedom, discretized with $L^{1/3}$ samples). By exploiting FFT-based methods, all $L^{1/3}$ in-plane rotations can be evaluated simultaneously for each viewing direction. Consequently, the effective complexity becomes $O(L^{2/3} d \log d)$, which improves over the naive bound by a factor of $L^{1/3}$ [30, 27]. This reduction is substantial in practice, since L often reaches tens or even hundreds of thousands in high-resolution cryo-EM.

The impact of non-uniform distributions of rotations. One key advantage of the Bayesian framework is its flexibility in incorporating different prior distributions for the rotations. In practical cryo-EM applications, the distribution of particle orientations is often non-uniform due to preferred particle orientations or sample preparation artifacts [57, 36]. If the rotation distribution can be

well estimated in an early stage, the information of the rotation distribution can then be integrated into the Bayes estimator to improve rotation estimation.

To illustrate rotation estimation under different prior distributions on $\mathsf{SO}(3)$, we replace the uniform distribution on $\mathsf{SO}(3)$ with an isotropic Gaussian (IG) distribution on $\mathsf{SO}(3)$, denoted by $g \sim \mathcal{IG}_{\mathsf{SO}(3)}(\eta)$, parameterized by a scalar variance η^2. This distribution is frequently used in machine learning probabilistic models on $\mathsf{SO}(3)$ [9, 31, 22]. It is worth noting that the IG distribution serves as a typical example, and similar phenomena as observed here extend to other non-uniform distributions on $\mathsf{SO}(3)$ as well.

The IG distribution $\mathcal{IG}_{\mathsf{SO}(3)}(\eta)$ can be represented in an axis-angle form, with uniformly sampled axes of rotation and a rotation angle $\omega \in [0, \pi]$. The scalar variance η^2 controls the distribution of the rotation angle ω: as $\eta \to \infty$, the $\mathsf{SO}(3)$ distribution approaches uniformity, whereas as $\eta \to 0$, ω becomes increasingly concentrated around 0, i.e., the rotation angle around the rotation axis is small. We apply the inverse sampling method to obtain i.i.d. samples from the IG distribution. Further details on this distribution are provided in Appendix F.1.

Figure 2 illustrates how incorporating a prior distribution over $\mathsf{SO}(3)$ rotations, governed by the variance parameter η of an isotropic Gaussian distribution affects, the accuracy of the MMSE rotation estimator. In all cases, the true underlying rotation distribution is modeled as an isotropic Gaussian distribution $\mathcal{IG}_{\mathsf{SO}(3)}(\eta = 0.1)$ over $\mathsf{SO}(3)$. The MLE estimator was computed according to (3.4). For the MMSE estimators, the estimation process used different prior distributions with variance parameters $\eta = 0.7, 0.5$, and 0.1, respectively. Specifically, the candidate rotations $g^{(\ell)}$ were generated according to these different priors (see (3.1)), highlighting the impact of prior mismatch on estimation accuracy. As the variance decreases (indicating a more concentrated and less uniform distribution closer to the true underlying distribution), the performance of the MMSE estimator improves, particularly at lower SNR conditions. These findings highlight the value of incorporating prior knowledge in rotation estimation, demonstrating its potential to enhance accuracy substantially. In stark contrast, the MLE estimator remains entirely unchanged regardless of the true underlying rotation distribution. As a result, its performance remains suboptimal, particularly in scenarios where the true distribution deviates significantly from uniformity.

We note that in experimental cryo-EM/ET datasets the viewing-angle distribution is typically unknown. Estimating this distribution, and understanding how *prior misspecification* impacts downstream reconstruction and heterogeneity analysis remains an important open problem. We discuss possible approaches for estimating the rotational prior in Appendix F.2. Recent work has begun to quantify misspecification effects under non-uniform group actions; see, e.g., [62]. In this light, Figure 2 can be viewed as a controlled *prior-mismatch* study: the data are generated from a concentrated isotropic-Gaussian prior, whereas the MMSE (and MAP) estimators are evaluated using an assumed prior with varying concentration. We observe a graceful degradation under mismatch, and a clear improvement as the assumed prior approaches the true distribution. Importantly, even under the common assumption of a uniform prior, the MMSE estimator improves over MLE in the low-SNR regime, while more accurate prior information yields additional gains when strong preferred orientations are present.

The impact of sampling grid size L and SNR on estimation accuracy. To study these effects, we restrict ourselves to the *uniform* rotation distribution setting. Figure 3 illustrates the impact of the sampling grid size of $\mathsf{SO}(3)$ together with different levels of SNR, on the geodesic distance between the MLE and MMSE rotation estimators relative to the true rotations. Several observations can be made. First, at high SNR, the MLE and MMSE estimators nearly coincide, consistent with Proposition 2.2. In this regime, the geodesic error decreases with the grid resolution,

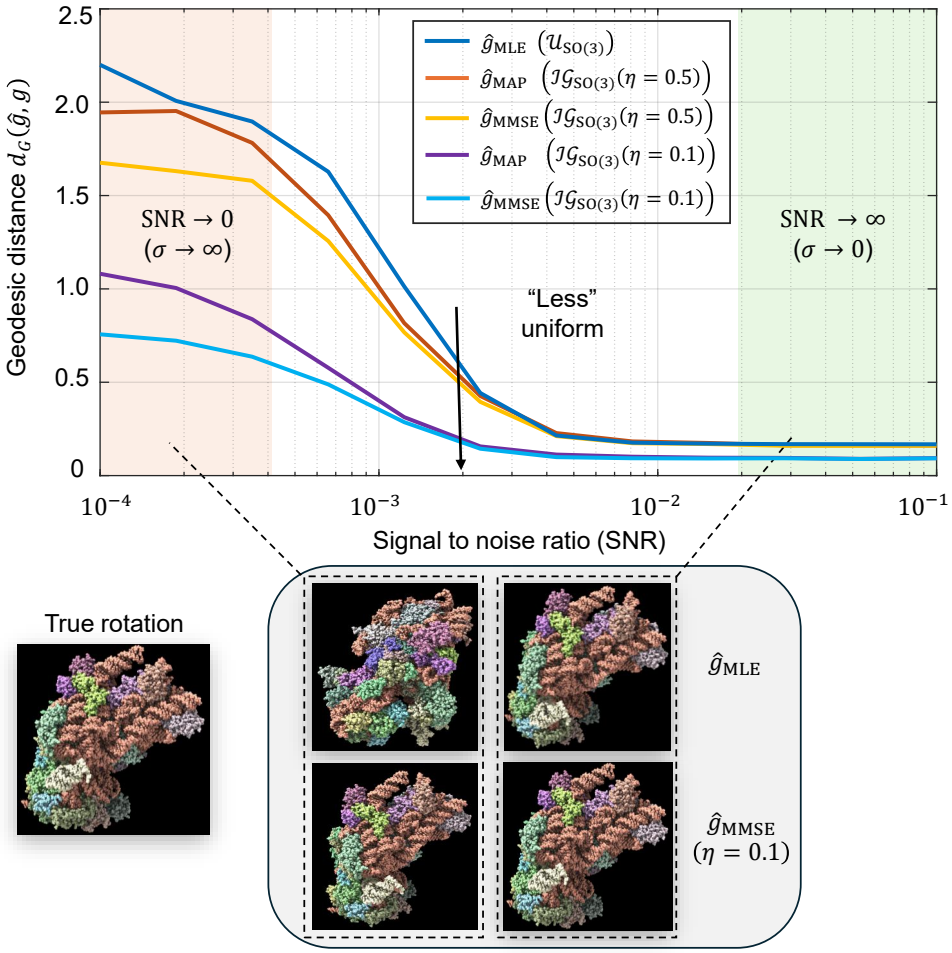


Figure 2: Impact of incorporating the prior rotation distribution on estimation accuracy. Simulations are performed for the cryo-ET model (1.2), excluding the projection step. The true rotation distribution is modeled as an isotropic Gaussian $\mathcal{IG}_{\mathsf{SO}(3)}(\eta = 0.1)$. Estimation performance is measured using the geodesic distance defined in (2.8). Here, g denotes the true rotation, $\hat{g}_{\mathrm{MLE}}$ is the maximum-likelihood estimator from (3.4), $\hat{g}_{\mathrm{MAP}}$ is the maximum a posteriori estimator from (2.11), and $\hat{g}_{\mathrm{MMSE}}$ denotes the Bayesian minimum mean square error estimator from (3.1). The MMSE estimators are computed assuming isotropic Gaussian priors on $\mathsf{SO}(3)$ with different concentration parameters $\eta \in \{0.5, 0.1\}$ (see Appendix F.1). As η decreases, the prior becomes more concentrated and closer to the true underlying distribution, leading to improved accuracy of both the MAP and MMSE estimators. Each data point in the plot is averaged over 3000 Monte Carlo trials using a rotation grid of size $L = 2976$. **Bottom 2×2 panel.** The four images compare denoised 3D volumes obtained using different rotation estimators at two representative noise regimes marked on the curve plot. Rows correspond to the estimator: top row uses $\hat{g}_{\mathrm{MLE}}$, bottom row uses $\hat{g}_{\mathrm{MMSE}}$ with the true prior $\mathcal{IG}_{\mathsf{SO}(3)}(\eta = 0.1)$. Columns correspond to SNR: left column is a low-SNR example, and right column is a high-SNR example (where the true rotation is correctly classified).

scaling empirically as $L^{1/3}$. This behavior arises because a discretization of $\mathsf{SO}(3)$ involves three angular parameters, so the resolution in each parameter direction grows like $L^{1/3}$. Second, as the SNR decreases, noise dominates and the advantage of refining the grid diminishes: both estimators approach similar performance, and the dependence on L becomes weaker. In the extreme limit $\sigma \to \infty$, the mean geodesic distance of the MLE and MMSE estimators is indistinguishable and shows no dependence on the grid size L.

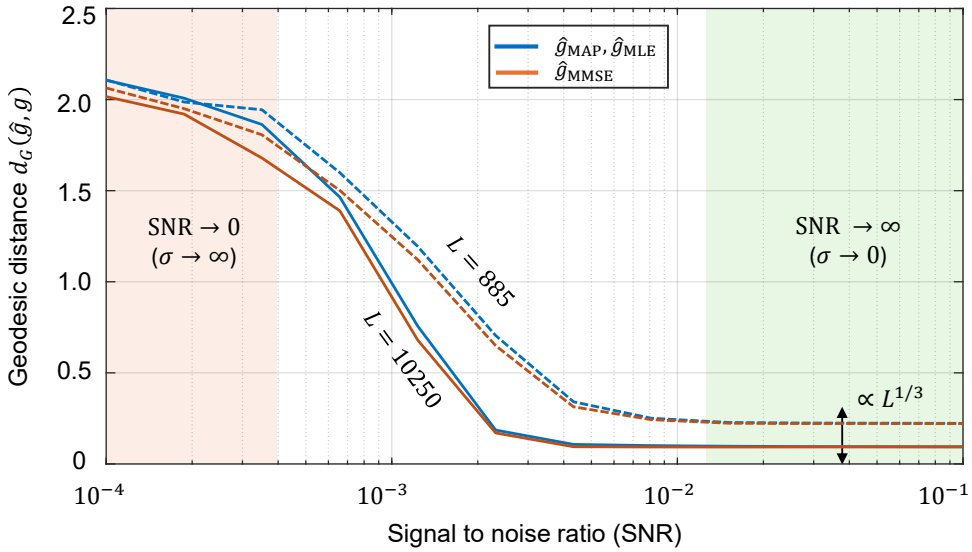


Figure 3: Impact of the sampling grid size of $\mathsf{SO}(3)$ (L) and the signal-to-noise ratio (SNR) on rotation estimation accuracy. This figure shows the accuracy of rotation estimation under varying sampling grid sizes L of the rotation group $\mathsf{SO}(3)$ and different SNR levels of the observed data y in the model (2.1). Simulations are performed for the cryo-ET model (1.2), excluding the projection step. The metric used for comparison is the geodesic distance, as defined in (2.8). Here, g denotes the true rotation, $\hat{g}_{\mathrm{MLE}}$ represents the MLE estimator from (3.4), and $\hat{g}_{\mathrm{MMSE}}$ denotes the Bayesian MMSE estimator from (3.1). In the high SNR regime ($\sigma \to 0$), the MLE and MMSE estimators converge, and the geodesic distance scales empirically as $\propto L^{1/3}$. This scaling reflects the three-parameter nature of $\mathsf{SO}(3)$ rotations, where the resolution of the sampling grid improves as L increases. The results shown are based on Monte Carlo simulations with 3000 trials per data point.

Sampling resolution and spectral information. Another factor that influences the discrepancy between the MAP and MMSE estimators is the effective spectral information available at a given sampling resolution. When comparing MAP and MMSE estimates across different image sizes, it is important to distinguish increasing the number of pixels from increasing usable spectral content. If the underlying signal is effectively bandlimited and the image is already sampled at (or above) the Nyquist rate for that bandlimit, then increasing the sampling resolution mainly amounts to oversampling and should not materially change the likelihood or the posterior over $\mathsf{SO}(3)$; consequently, the MAP–MMSE estimators are expected to remain essentially unchanged. In contrast, when higher sampling resolution is accompanied by additional usable high-frequency signal (i.e., non-negligible per-frequency SNR at higher bands), the posterior typically becomes more concentrated, and we expect the MAP and MMSE estimates to approach each other in geodesic distance (at a fixed SNR level).

4 Bayesian orientation estimation as part of the volume reconstruction problem

Thus far, we have introduced the MLE, MAP and MMSE orientation estimators for estimating the rotation between a single noisy observation y and a reference volume V, as defined in (2.1). An intriguing question that arises is how these estimators can be utilized and influence performance within a 3D volume reconstruction process, which constitutes the main computational challenge in

cryo-EM and cryo-ET.

Structure reconstruction typically follows two primary approaches: *hard-assignment* or *soft-assignment* methods, which are generally implemented through iterative refinement. In the hard-assignment approach, each observation is assigned a single orientation based on the highest correlation, and the 3D structure is then reconstructed given the rotations. In contrast, the soft-assignment method assigns probabilities across all possible orientations for each observation, enabling the 3D structure to be recovered as a weighted average of the observations, with weights determined by these probabilities. The iterative application of the soft-assignment procedure aligns with the EM algorithm [10, 51], which serves as the core computational method in modern cryo-EM [49, 44].

In the following, we demonstrate that incorporating the MMSE orientation estimator into the volume reconstruction process in the cryo-ET model (without projections) resembles the expectation-maximization (EM) algorithm, as it accounts for the full distribution of possible outcomes. In contrast, substituting the MLE estimator into the algorithm operates more like a hard-assignment reconstruction method, focusing exclusively on the most likely outcome. This distinction underscores the broader applicability and flexibility of the Bayes estimator in capturing uncertainty and delivering more accurate estimates for structure reconstruction.

4.1 Connection to the EM algorithm in volume reconstruction without projection

Unlike the previous orientation estimation model (2.1), we consider the following simplified model for volume reconstruction (see also [52, 6, 12]). We observe M i.i.d samples taking the form

$$y_i = g_i \circ V + \varepsilon_i, \qquad i = 0, \ldots, M-1, \tag{4.1}$$

where V is 3D volume structure of interest, $\{g_i\}_{i=0}^{M-1} \in \mathsf{SO}(3)$ are unknown latent variables following i.i.d. *uniform* distribution, satisfying $g_i \circ V(x) \equiv V(g_i^{-1}x)$, and $\{\varepsilon_i\}_{i=0}^{M-1}$ are i.i.d isotropic Gaussian noise with variance σ^2. For the case where V represents a 2D image, $\{g_i\}_{i=0}^{M-1} \in \mathsf{SO}(2)$ corresponds to in-plane rotations, and the model is used for image recovery (see more in Section 4.2) [37]. The goal is to recover V from the observations $\{y_i\}_{i=0}^{M-1}$, treating the rotations $\{g_i\}_{i=0}^{M-1}$ as latent variables.

To distinguish the model (4.1) from (2.1), we highlight the key differences as follows:

(i) The parameter of interest is the unknown structure V here, whereas it was the single rotation g in model (2.1);

(ii) We observe M i.i.d. samples instead of a single observation, meaning that all observed M samples are used collectively to estimate the underlying volume structure V;

(iii) Although the rotations $\{g_i\}_{i=0}^{M-1}$ are also unknown, they are treated as nuisance parameters, and we are not directly concerned with their estimation (though admittedly, more accurate estimation of $\{g_i\}_{i=0}^{M-1}$ could often contribute to better estimation of V).

The most common method for solving this reconstruction problem is the EM algorithm, which applies soft assignment iteratively, as outlined in Algorithm 2. In each iteration, the algorithm uses the volume estimate from the previous iteration, denoted as $\hat{V}^{(t)}$, to update the volume estimate $\hat{V}^{(t+1)}$, based on the observations $\{y_i\}_{i=0}^{M-1}$. The following proposition illustrates the relationship between the volume structure update rule at iteration $t+1$, and the MMSE orientation estimator introduced in (2.12).

Algorithm 2 `EM algorithm for volume reconstruction`

Input: An initial volume $\hat{V}^{(0)}$, number of iteration T and observations $\{y_i\}_{i=0}^{M-1}$ given by (4.1).
Output: Final volume estimation after T iteration, $\hat{V}^{(T)}$.
Each Iteration, for $t = 0 \ldots, T-1$**:**

1. Compute for every $0 \leq i \leq M-1$:

$$p_i^{(t)}(g) = \frac{\exp\left(-\left\|y_i - \left(g \circ \hat{V}^{(t)}\right)\right\|^2 / 2\sigma^2\right)}{\int_{g \in \mathsf{SO}(3)} \exp\left(-\left\|y_i - \left(g \circ \hat{V}^{(t)}\right)\right\|^2 / 2\sigma^2\right) \mathrm{d}g}, \tag{4.2}$$

2. Update the volume estimate:

$$\begin{aligned} \hat{V}^{(t+1)} =& \arg\min_V \frac{1}{M} \sum_{i=0}^{M-1} \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \left\|y_i - (g \circ V)\right\|^2 \mathrm{d}g \\ =& \frac{1}{M} \sum_{i=0}^{M-1} \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g)(g^{-1} \circ y_i) \mathrm{d}g. \end{aligned} \tag{4.3}$$

To state the next result, we first introduce the *MMSE back-rotation operator*. Given an observation y_i and the current reference volume $\hat{V}^{(t)}$, let $p_i^{(t)}(g)$ denote the posterior weights over $g \in \mathsf{SO}(3)$ as defined in (4.2). We define $\widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t}$ as the posterior average of the *inverse action* g^{-1}:

$$\widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t} := \mathbb{E}\left[g^{-1} \,\middle|\, y_i, \hat{V}^{(t)}\right], \tag{4.4}$$

which should be interpreted as a linear operator acting on functions (or volumes). Namely, for any test function f on which rotations act,

$$\left(\widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t} \circ f\right) := \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \left(g^{-1} \circ f\right) \mathrm{d}g. \tag{4.5}$$

Proposition 4.1 (MMSE operator form of the EM M-step)**.** *Let $\hat{V}^{(t+1)}$ be the $(t+1)$-th volume estimator in the EM algorithm as described in Algorithm 2, and let $\widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t}$ be defined as in (4.4)–(4.5). Then the M-step update can be written as*

$$\hat{V}^{(t+1)} = \frac{1}{M} \sum_{i=0}^{M-1} \widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t} \circ y_i. \tag{4.6}$$

The proof of this proposition is presented in Appendix G. In words, the proposition shows that the update rule for the volume structure estimation at iteration $t+1$, given the volume $\hat{V}^{(t)}$, is equivalent to aligning each observation $\{y_i\}_{i=0}^{M-1}$ using the associated back-rotation MMSE operator $\widehat{\mathfrak{g}}_{\mathrm{MMSE},i,t}$, computed based on the reference volume $\hat{V}^{(t)}$, and then averaging the aligned observations (4.6). Thus, the MMSE back-rotation operator is a key ingredient in the EM algorithm for volume reconstruction.

Remark 4.2. *The* MMSE back-rotation operator $\widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t}$ (4.4) *should be distinguished from the MMSE rotation* matrix estimator $\hat{g}_{\mathsf{MMSE}}$ *obtained in* (2.12) *by taking a posterior mean in the* 3×3 *matrix embedding and then projecting onto* $\mathsf{SO}(3)$ *(via Orthogonal Procrustes). In the matrix formulation, one forms* $\hat{g}_{\mathsf{relax}}$ *as in* (2.13) *and then projects to obtain a single* $\mathsf{SO}(3)$ *element. In contrast, in* (4.5) *the inversion is handled* inside *the posterior average: the expectation is taken over the inverse* actions $g^{-1} \circ (\cdot)$*, yielding a linear operator that acts directly on* y_i *(and is not, in general, itself a member of* $\mathsf{SO}(3)$*). While these two objects differ in representation, they are both posterior-weighted aggregations over rotations; the operator form is the one that naturally appears in the EM M-step for reconstructing the volume.*

Despite the subtle differences arising from rotation estimation versus structure reconstruction, the previously introduced MMSE rotation estimator (2.12) can still serve as a practical tool for assessing the performance of soft assignments. Concretely, we take the posterior rotation matrix mean as in (4.4), and then perform an Orthogonal Procrustes projection to replace $\widehat{\mathfrak{g}}_{\mathsf{MMSE},i,t}$. In this way, the MMSE rotation estimator provides a concrete approximation to the soft assignment in structure reconstruction. In numerical experiments, it is compared with the hard assignment, demonstrating superior performance relative to the maximum-likelihood estimator, as discussed below.

The MLE estimator as part of volume reconstruction. In contrast to the soft assignment procedure, if we replace the MMSE operator with the corresponding MLE operator, we obtain the structure reconstruction algorithm by applying *hard assignment* iteratively. To be more specific, the MLE operator applied to the observations $\{y_i\}_{i=0}^{M-1}$ can be viewed as a hard assignment among all possible rotations. In practice, this procedure involves making a hard decision where a single rotation is selected from the rotation grid according to the closest alignment. In the $(t+1)$-th iteration, similarly to (4.6), the hard-assignment process can be expressed as follows:

$$\hat{V}^{(t+1)} = \frac{1}{M} \sum_{i=0}^{M-1} \hat{\mathfrak{g}}_{\mathsf{MLE},i,t} \circ y_i, \tag{4.7}$$

where $\hat{\mathfrak{g}}_{\mathsf{MLE},i,t}$ is defined by

$$\hat{\mathfrak{g}}_{\mathsf{MLE},i,t} = \underset{g^{-1} \in \mathsf{SO}(3)}{\arg\max}\ \mathbb{P}(g | y_i, \hat{V}^{(t)}) \tag{4.8}$$

$$= \underset{g^{-1} \in \mathsf{SO}(3)}{\arg\max}\ p_i^{(t)}(g) = \underset{g^{-1} \in \mathsf{SO}(3)}{\arg\min}\ \|y_i - (g \circ \hat{V}^{(t)})\|^2. \tag{4.9}$$

In other words, $\hat{\mathfrak{g}}_{\mathsf{MLE},i,t} = (\hat{g}_{\mathsf{MLE},i,t})^{-1}$ where

$$\hat{g}_{\mathsf{MLE},i,t} := \underset{g \in \mathsf{SO}(3)}{\arg\min}\ \|y_i - (g \circ \hat{V}^{(t)})\|^2.$$

Hence, the hard assignment can be viewed as using the exact inverse of the MLE estimator introduced earlier.

4.2 Empirical results for volume reconstruction and the "Einstein from Noise" phenomenon

We demonstrate empirically volume reconstruction by applying the MLE estimator $\hat{g}_{\mathsf{MLE}}$ and the MMSE estimator $\hat{g}_{\mathsf{MMSE}}$ as part of the reconstruction problem, as specified in Section 4.1.

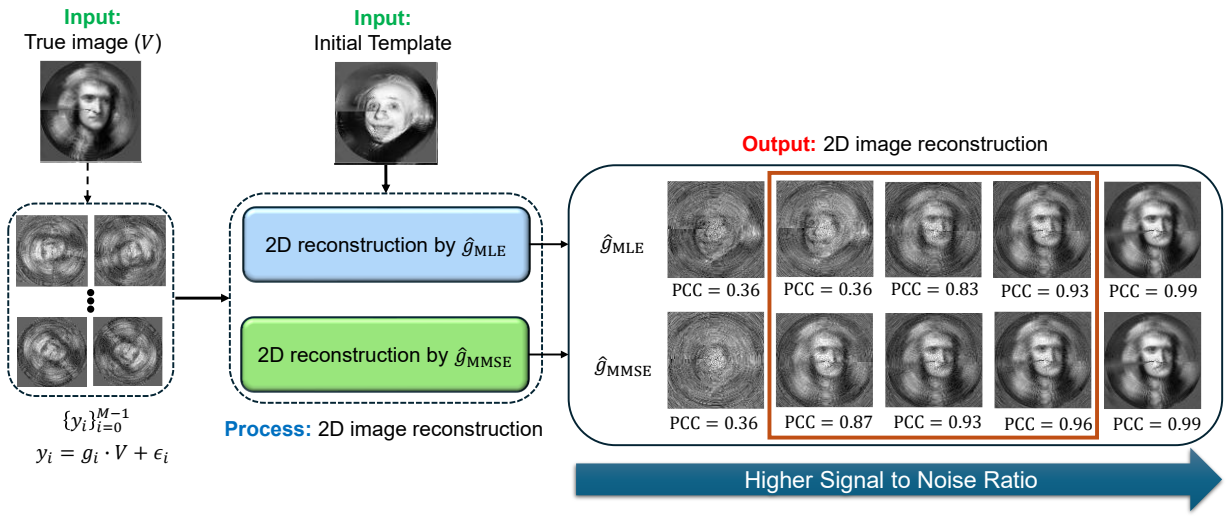


Figure 4: Comparison of 2D image recovery using the MMSE and the MLE rotation estimators. Iterative image recovery procedures with the MMSE estimator ($\hat{g}_{\mathrm{MMSE}}$) and MLE estimator ($\hat{g}_{\mathrm{MLE}}$) are defined in (4.6) and (4.7), respectively. The experiment employs a template image of Einstein and a ground truth image of Newton, both rotated in 2D over a uniform polar grid with $L = 30$ samples. Each image is of size 100×100 pixels, and the radial direction is discretized using $R = 300$ points. The reconstructed images within the dark-orange rectangle (right panel) show superior performance with the MMSE rotation estimator, with Pearson cross-correlation (PCC) provided for each reconstructed image. The MLE and MMSE reconstructions are nearly identical at high SNR ($\sigma \to 0$), as predicted by Proposition 2.2. The SNR values used for the panels (from right to left) are 10^{-2}, 4×10^{-3}, 2×10^{-3}, 7×10^{-4}, and 2×10^{-4}. At very low SNR ($\sigma \to \infty$), the "Einstein from Noise" effect appears, where the estimator resembles the template image of Einstein rather than the underlying truth of Newton. In intermediate SNR ranges, using the MMSE estimator in the iterative step clearly outperforms the MLE estimator.

Description of the experiments. We demonstrate the reconstruction processes, which integrates MLE and MMSE rotation estimators as intermediate steps, using the two examples: 2D image recovery (Figure 4) and 3D volume reconstruction without projection (Figure 5). The iterative reconstruction process, as outlined in Algorithm 2, was performed until convergence (i.e., when the relative difference between consecutive iterations fell below a predefined threshold of 10^{-3} or until reaching a maximum of 100 iterations.

Figures 4 and 5 were generated using slightly different methods. For the 2D experiment presented in Figure 4, we used polar coordinates, while in the 3D experiment, shown in Figure 5, we used a standard Cartesian basis. The primary difference lies in the interpolation required for producing Figure 5, which utilizes cubic interpolation for each observation based on the estimated rotation. This introduces certain "quantization" errors. Additionally, as detailed in Section 2, the 3D reconstruction process, presented in Figure 5, requires a "rounding" step which amounts to solving the Orthogonal Procrustes procedure. In Figure 4, the true and template structures were generated in a polar representation with $d = 300$ radial points and $L = 30$ polar angle points. The reconstruction process was performed with $M = 5 \times 10^4$ observations. The additive noise was added in the polar representation. In Figure 5, $M = 3000$ observations were used, with the rotation group $\mathsf{SO}(3)$ grid size of $L = 300$, and a volume size of $32 \times 32 \times 32$.

Empirical observations. A few observations can be made from Figures 4 and 5. First, in the case of high SNR (i.e., as $\sigma \to 0$), the volume reconstruction is similar whether using the MLE

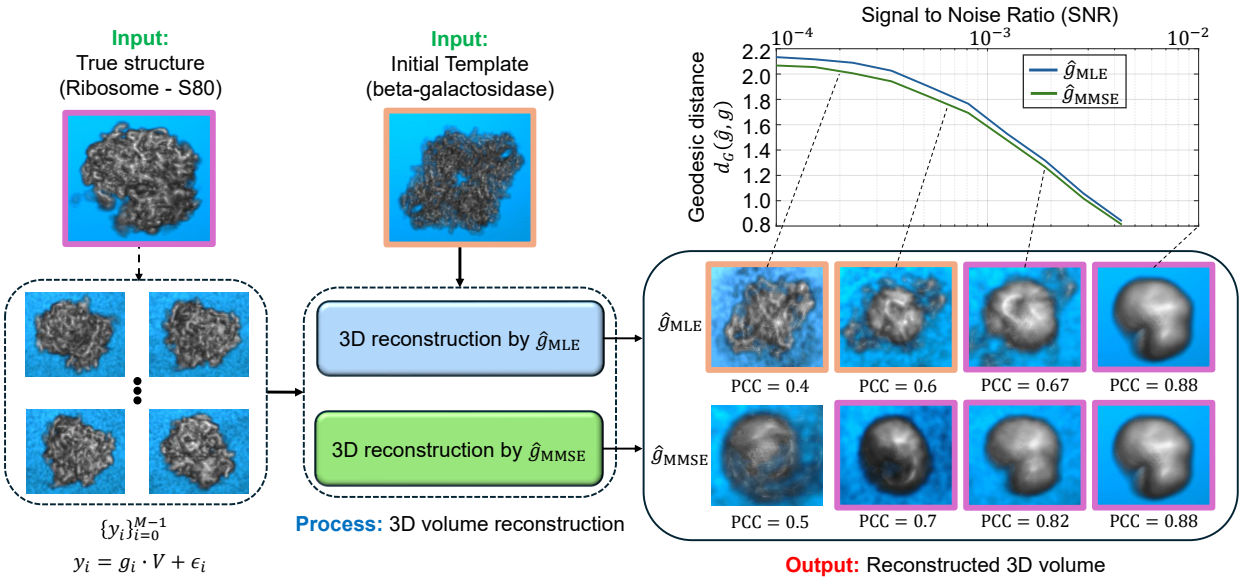


Figure 5: Comparison of 3D structure reconstruction in cryo-ET subtomograms averaging using the MMSE and MLE rotation estimators. Iterative structure reconstruction procedures with the MMSE estimator ($\hat{g}_{\mathrm{MMSE}}$) and MLE estimator ($\hat{g}_{\mathrm{MLE}}$) are defined in (4.6) and (4.7), respectively. At high SNR levels, a low-resolution structure emerges due to finite grid sampling of the rotation group $\mathsf{SO}(3)$, effectively acting as a low-pass filter. The 3D reconstruction using the MMSE estimator consistently outperforms the reconstruction using the MLE estimator. For high SNR conditions (i.e., $\sigma \to 0$), both estimators yield similar 3D structures, as expected from Proposition 2.2. The SNR values used for the panels (from right to left) are 10^{-2}, 2×10^{-3}, 7×10^{-4}, and 2×10^{-4}, with a volume size of $32 \times 32 \times 32$. The boxes highlighted in purple resemble the true input structure (Ribosome-S80 [61]), while the orange-highlighted boxes are more similar to the initial template (beta-galactosidase [7]), illustrating the "Einstein from Noise" phenomenon. Notably, at very low SNRs, the "Einstein from Noise" effect is evident with the MLE estimator but not with the MMSE estimator.

estimator or the MMSE estimator. This similarity is theoretically supported by Proposition 2.2. However, as the SNR decreases, the reconstructions diverge, with the volumes reconstructed using the MMSE estimator showing a better correlation with the true volume. Second, in scenarios of extremely low SNR, where the structural signal is nearly nonexistent, the phenomenon known as "Einstein from Noise" manifests in both 2D and 3D contexts. This phenomenon pertains to the inherent model bias within the reconstruction procedure, specifically in relation to the initial templates. In such cases, the reconstructed volume exhibits structural similarities to the initial template, even though the observations do not substantiate this outcome. The generation of a structured image from entirely noisy data has attracted considerable attention, particularly during a significant scientific debate regarding the structure of an HIV molecule [40, 19, 59, 56, 39]; for a comprehensive description and statistical analysis, see [3, 2, 4, 5]. Notably, our empirical evidence suggests that the "Einstein from Noise" phenomenon is more pronounced when adopting the MLE estimator compared to the MMSE estimator, implying that the MMSE approach is less vulnerable to the choice of the initial template. Furthermore, our experiment suggests that the advantage of using the Bayesian MMSE estimator over the MLE estimator is more significant in 3D structure reconstruction tasks compared to 2D image recovery, where the 3D setting is a problem of greater interest to researchers in structural biology.

5 Structural heterogeneity analysis

Understanding heterogeneity is central to revealing the dynamic behavior of macromolecular complexes in cryo-EM. Unlike traditional 3D classification methods that assume a small number of discrete states, modern approaches aim to capture continuous structural variability by embedding projection images into a latent conformational space. Here, we adopt the RECOVAR framework [13], which performs heterogeneity analysis using principal component analysis (PCA) based on a regularized estimate of the conformational covariance matrix. The key idea is to infer the covariance of 3D volumes directly from noisy 2D projection images, leveraging the fact that although the volumes themselves are unobserved, their projections contain sufficient statistical information.

5.1 Fixed-pose methods

A central computational assumption in RECOVAR is that the poses $\{\phi_i\}$ of all particle images are known and fixed, typically obtained from a consensus refinement procedure. This "fixed-pose" assumption underlies many modern heterogeneity analysis pipelines, including 3D Classification in cryoSPARC [44], 3D Variability Analysis (3DVA) [42], 3DFlex [43], cryoDRGN [64], CryoDRGN-AI-fixed [33], Opus-DSD [35], and RECOVAR [13]. These methods are collectively referred to as *fixed-pose methods*, since they treat poses as known inputs during downstream inference. Among them, RECOVAR has recently emerged as one of the most effective approaches. According to the CryoBench evaluation [24], which benchmarks multiple methods across diverse synthetic and experimental datasets, RECOVAR consistently achieves state-of-the-art performance in both structural resolution and latent space recovery. Notably, CryoBench evaluations were performed by external users, indicating RECOVAR's strong out-of-the-box performance and robustness to hyperparameter choices. This motivates our choice to build upon the RECOVAR framework.

Despite its strong empirical performance, the fixed-pose assumption merits careful scrutiny. In practice, ground-truth poses are never directly accessible and cannot be recovered exactly, even if the underlying volume is known, due to noise and the intrinsic ill-posedness of the inverse problem. Most existing methods address this by using MLE pose estimates in real data. This gap between assumed and achievable pose accuracy motivates a central question: to what extent can more accurate pose estimation improve downstream heterogeneity recovery? In this work, we address this question by replacing MLE poses with our MMSE estimator within RECOVAR.

5.2 The mathematical model

The forward model [13, Eq. (1)] used in RECOVAR is expressed in the Fourier domain as

$$\tilde{y}_i = C_i \hat{P}(\phi_i) \tilde{V}_i + \tilde{\varepsilon}_i, \tag{5.1}$$

where $\tilde{y}_i \in \mathbb{C}^{N^2}$ is the observed Fourier-transformed image (with N denoting the grid size along each spatial axis), $\tilde{V}_i \in \mathbb{C}^{N^3}$ is the (unknown) 3D Fourier volume corresponding to a particular conformation, $\hat{P}(\phi_i)$ is the tomographic projection operator from 3D to 2D after a rigid-body motion $\phi_i = (g_i, t_i)$ with $g_i \in \mathsf{SO}(3)$ a rotation and $t_i \in \mathbb{R}^2$ an in-plane shift, C_i is the contrast transfer function (CTF), and $\tilde{\varepsilon}_i$ is additive noise.

This formulation is a discretized and sampled version of our continuous forward model in (A.2). Specifically, the operator $\hat{P}(\phi_i)$ acting on $\tilde{V}_i$ corresponds directly to $e^{-2\pi i \boldsymbol{\omega} \cdot t_i} \cdot (g_i \circ \tilde{V})(\omega_1, \omega_2, 0)$, where each component reflects the same physical process, namely, in-plane shift, 3D rotation, and evaluation on the central slice in Fourier space.

To model structural variability, we treat each underlying volume $\tilde{V}_i$ as a random sample from a distribution over 3D conformations. We assume this distribution has a well-defined mean $\mu \in \mathbb{C}^{N^3}$ and covariance $\Sigma \in \mathbb{C}^{N^3 \times N^3}$ given by

$$\mu = \mathbb{E}[\tilde{V}], \quad \Sigma = \mathbb{E}[(\tilde{V} - \mu)(\tilde{V} - \mu)^*], \tag{5.2}$$

where the expectation is taken over the conformational distribution. The objective of RECOVAR is to estimate both μ and Σ from the observed projections $\tilde{y}_i$ with given poses ϕ_i, using regularized least-squares minimization [13, Eqns. (2)–(3)].

5.3 Numerical experiments

In our experiments, we consider a synthetic one-dimensional *conformational transition*, a standard benchmark for continuous heterogeneity analysis methods such as CRYODRGN [64] and RECOVAR [13]. The corresponding ground-truth density maps along the conformational coordinate are shown in Figure 6(a). To generate the projection images, the underlying conformational states were uniformly sampled along the one-dimensional transition, and molecular orientations were sampled from a uniform distribution over $\mathsf{SO}(3)$. Figure 6(b) further displays a representative noisy projection under high-noise conditions.

To avoid storing and manipulating the full high-dimensional covariance matrix $\Sigma \in \mathbb{C}^{N^3 \times N^3}$, RECOVAR first estimates a low-dimensional subspace of rank $r \ll N^3$ in which the covariance is approximated. Following [13, Appendix A.2], we estimate $\hat{\Sigma}$ via regularized least-squares, then form a rank-d approximation $\hat{\Sigma}_{\text{col}}$ by selecting a subset of columns using a greedy SNR-based criterion that ensures both high SNR and low inter-column correlation. An orthonormal basis $\tilde{U} \in \mathbb{C}^{N^3 \times r}$ spanning $\hat{\Sigma}_{\text{col}}$ is obtained via randomized SVD. The conformational mean $\hat{\mu}$ is also estimated at this stage. Each conformational volume is then represented as

$$\tilde{V}_i = \tilde{U} z_i + \hat{\mu}, \tag{5.3}$$

where $\hat{\mu}$ is the estimated conformational mean from the first stage, and $z_i \in \mathbb{C}^r$ are low-dimensional coordinates. In RECOVAR, these coordinates are not explicitly recovered; their covariance matrix $\hat{\Sigma}_{\tilde{U}} \in \mathbb{C}^{r \times r}$ is estimated directly via a reduced least-squares problem (Eq. (14) in [13]), yielding a low-rank covariance approximation that characterizes the variability of z_i. The eigen-decomposition of $\hat{\Sigma}_{\tilde{U}}$ then provides the estimated principal components and eigenvalues in the reduced space. The complete procedure is detailed in [13, Algorithm 1].

We perform our analyses on a controlled synthetic dataset of 30,000 projection images ($N = 128$), simulated from a one-dimensional conformational transition discretized into 50 equally spaced states to approximate continuous heterogeneity [13]. We first evaluate the effect of orientation estimation on covariance recovery. The study is conducted in the high-noise regime of RECOVAR (Figure 6(b)), where even with ground-truth poses, the top 30 principal components capture only about 40% of the variance due to finite-sample and high-noise effects. We set $r = 50$ and compare three pose sources: ground truth, MMSE estimates, and MLE estimates, keeping all other steps identical. For fixed-pose methods, poses are treated as given, so computational cost remains unchanged.

The accuracy of the estimated subspace is quantified by the percentage of total variance captured:

$$\|U_k^* \Sigma U_k\|_S / \|\Sigma\|_S,$$

where U_k contains the first k estimated principal components, Σ is the ground-truth covariance matrix, and $\Sigma^{1/2}$ is its matrix square root. Here, $\|A\|_S = \sum \sigma_i(A)$ is the Schatten 1-norm of matrix A with σ_i denoting the singular values of A.

Figures 6(e) and 6(f) present the variance-capture results as described and the corresponding eigenvalue recovery, respectively. Figure 6(d) visualizes the first five recovered principal components. Overall, MMSE pose estimation yields principal components and eigenvalue spectra that are consistently closer to the ground truth than those obtained from MLE pose estimation.

We further perform a more granular evaluation of structure reconstruction at the level of individual conformational states. For this analysis, we select all projection images corresponding to a specific conformational state and embed them into the latent conformational space (z-space, see (5.3)) using RECOVAR, conditioned on different pose priors: ground-truth, MMSE estimates, and MLE estimates. For each subset, the reconstruction corresponding to the mean embedding serves as an estimate of that specific conformational state. These reconstructions provide a direct measure of how different orientation estimates affect the accuracy of heterogeneity structure reconstruction. Figure 6(c) demonstrates that MMSE orientation estimates produce reconstructed structures closer to the ground-truth conformation than those obtained from MLE, further highlighting the advantage of MMSE pose estimation in high-noise heterogeneity analysis. We further compared the reconstructions obtained using MMSE and MLE orientation estimates against the ground-truth conformational state. The reported local FSC resolution scores (in Å) show that the 50% quantile resolutions are 11.45 Å and 11.55 Å for MMSE and MLE, respectively, while the 90% quantiles are 18.83 Å and 21.43 Å (also see [13] Supplementary S.J.).

6 Discussion and conclusions

In this work, we have introduced the Bayesian framework for enhancing orientation estimation for various applications in structural biology. The proposed approach offers greater flexibility and improved accuracy compared to existing methods, with the MMSE estimator as a prime example. This technique handles diverse structural conformations and arbitrary rotation distributions across sample sets. Our empirical results establish that the proposed MMSE estimator consistently surpasses the performance of the current methods, particularly in challenging low SNR environments, as well as when prior information on rotation distribution is available or approximately known. We provide a theoretical foundation to explain these performance gains. As rotation determination is crucial for both 2D and 3D reconstruction processes, we further illustrate how utilizing the MMSE estimator as a soft-assignment step in iterative refinement leads to significant improvements over hard-assignment methods. Moreover, the proposed Bayesian approach empirically offers enhanced resilience against the "Einstein from Noise" phenomenon, effectively reducing model bias and improving the overall reliability of structural reconstructions. Thus, our main recommendation is to adopt the Bayesian MMSE rotation estimator over the MLE estimator in related application scenarios. Already integrated into most software, the Bayesian rotation estimator can be easily implemented with minimal computational cost, offering improved accuracy and resilience for orientation determination and structure reconstruction.

Future work. The MMSE orientation estimator represents the most natural Bayes estimator in the context of cryo-EM, as it coincides with the M-step reconstruction of the EM algorithm. Nevertheless, alternative Bayes estimators associated with different loss functions could also be explored in future studies. A particularly promising direction is the direct estimation of the rotation distribution from observed images and the incorporation of this information into rotation estimation or EM-based 3D reconstruction algorithms [23, 62]. Such approaches could substantially improve rotation accuracy, especially in low-SNR regimes, as indicated by our results in Figure 2, and consequently enhance volume reconstruction quality. Further extensions may involve modeling

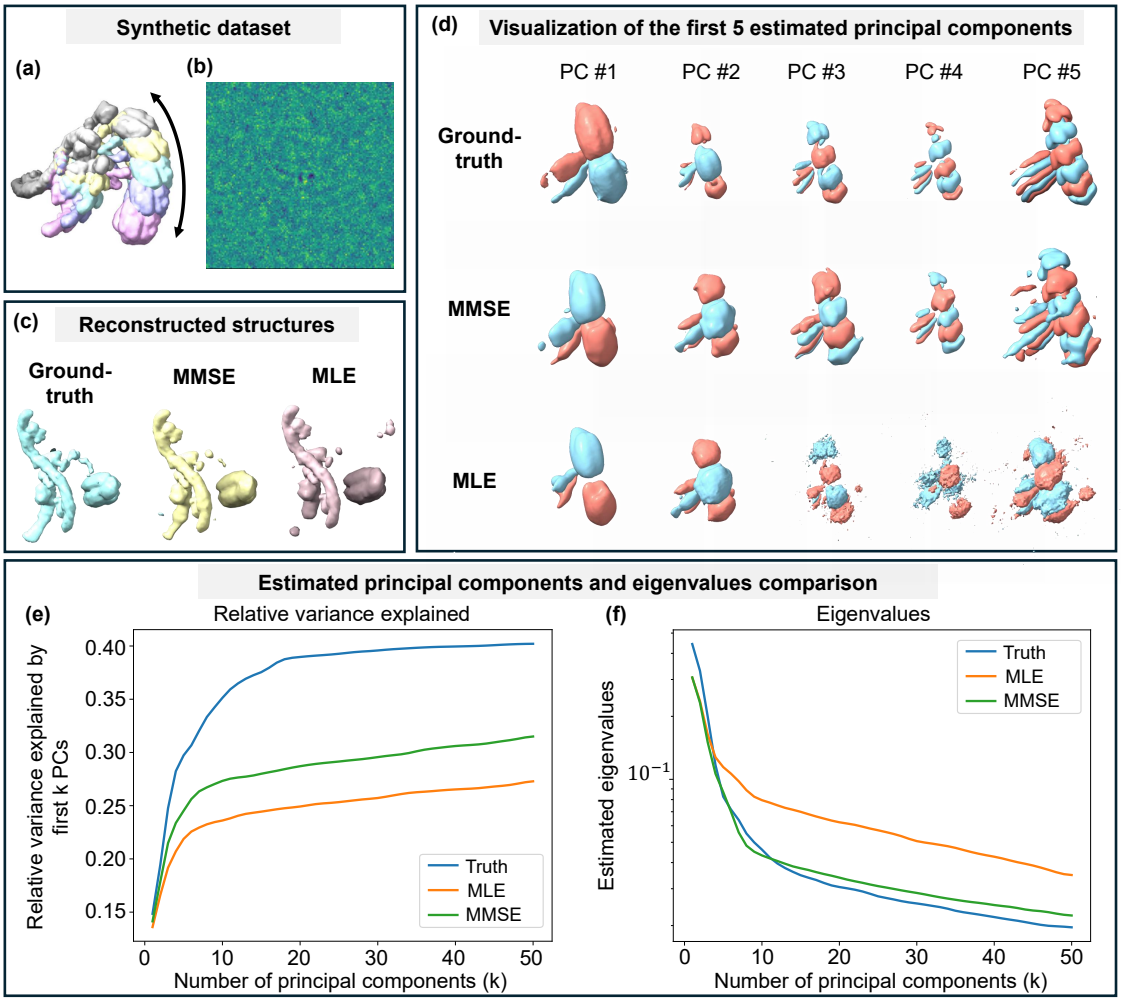


Figure 6: MMSE pose estimation improves heterogeneity reconstruction over MLE. (**a**) Ground-truth density maps from a synthetic dataset simulating a one-dimensional conformational transition, a standard benchmark for methods such as cryoDRGN [64] and RECOVAR [13]. Fifty equally spaced states approximate the continuous trajectory, with colors indicating positions along the pathway (only five states are shown in the figure for clarity). (**b**) Example of a noisy projection image from the high-noise regime used in this study, where even ground-truth poses recover only $\sim 40\%$ of the variance for the top 30 components due to finite-sample and high-noise effects. (**c**) Comparison of reconstructed structures using different pose priors (ground-truth, MMSE, MLE), showing MMSE yields structures closer to the true conformation state. (**d**) First five principal components estimated using ground-truth poses, MMSE pose estimates, and MLE pose estimates. MMSE results closely match the ground truth in structural detail, whereas MLE reconstructions degrade beyond the first modes, becoming blurrier and less representative of the underlying variability. (**e**) Subspace accuracy, measured as the percentage of total variance captured (as defined in (5.3)). Ground truth achieves $\sim 40\%$, MMSE $\sim 30\%$, and MLE $\sim 25\%$. (**f**) Eigenvalue recovery: both methods estimate the largest true eigenvalues well, but MMSE remains accurate for smaller eigenvalues, while MLE substantially overestimates them. These results show that MMSE pose estimation consistently yields more accurate recovery of conformational variability than MLE.

structural uncertainties, particularly relevant for flexible or heterogeneous proteins, and addressing general pose estimation problems that jointly consider rotations and translations. Finally, leveraging the Bayesian framework to derive confidence regions for individual rotations presents an exciting opportunity to improve the interpretability and reliability of rotation estimation in cryo-EM.

Data Availability

The detailed implementation and code are available at https://github.com/AmnonBa/bayesian-orientation-estimation.

Acknowledgment

A.S. and S.X. are supported in part by AFOSR under Grant FA9550-23-1-0249, the Simons Foundation Math+X Investigator Award, NSF under Grant DMS 2510039, and NIH/NIGMS under Grant R01GM136780-01. T.B. is supported in part by BSF under Grant 2020159, in part by NSF-BSF under Grant 2024791, in part by ISF under Grant 1924/21, and in part by a grant from The Center for AI and Data Science at Tel Aviv University (TAD). We thank Marc Auréle Gilles for providing guidance on RECOVAR software. We also thank Marc Auréle Gilles, Eric Verbeke, and Ruiyi Yang for their helpful discussions.

References

[1] Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015.

[2] Amnon Balanov, Tamir Bendory, and Wasim Huleihel. Confirmation bias in Gaussian mixture models. *IEEE Transactions on Information Theory*, 2025.

[3] Amnon Balanov, Wasim Huleihel, and Tamir Bendory. Einstein from noise: Statistical analysis. *arXiv preprint arXiv:2407.05277*, 2024.

[4] Amnon Balanov, Wasim Huleihel, and Tamir Bendory. Expectation-maximization for multi-reference alignment: Two pitfalls and one remedy. *arXiv preprint arXiv:2505.21435*, 2025.

[5] Amnon Balanov, Alon Zabatani, and Tamir Bendory. Structure from noise: Confirmation bias in particle picking in structural biology. *arXiv preprint arXiv:2507.03951*, 2025.

[6] Afonso S Bandeira, Ben Blum-Smith, Joe Kileel, Jonathan Niles-Weed, Amelia Perry, and Alexander S Wein. Estimation under group actions: recovering orbits from invariants. *Applied and Computational Harmonic Analysis*, 66:236–319, 2023.

[7] Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. Structure of β-galactosidase at 3.2-å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences*, 111(32):11709–11714, 2014.

[8] Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020.

[9] Gabriele Corso, Bowen Jing, Regina Barzilay, Tommi Jaakkola, et al. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *International Conference on Learning Representations (ICLR 2023)*, 2023.

[10] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.

[11] Claire Donnat, Axel Levy, Frederic Poitevin, Ellen D Zhong, and Nina Miolane. Deep generative modeling for volume reconstruction in cryo-electron microscopy. *Journal of structural biology*, 214(4):107920, 2022.

[12] Zhou Fan, Roy R Lederman, Yi Sun, Tianhao Wang, and Sheng Xu. Maximum likelihood for high-noise group orbit estimation and single-particle cryo-EM. *The Annals of Statistics*, 52(1):52–77, 2024.

[13] Marc Aurèle Gilles and Amit Singer. Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proceedings of the National Academy of Sciences*, 122(9):e2419140122, 2025.

[14] John C Gower and Garmt B Dijksterhuis. *Procrustes problems*, volume 30. OUP Oxford, 2004.

[15] Manuel Gräf. A unified approach to scattered data approximation on and SO (3). *Advances in Computational Mathematics*, 37(3):379–392, 2012.

[16] Manuel Gräf and Daniel Potts. Sampling sets and quadrature formulae on the rotation group. *Numerical Functional Analysis and Optimization*, 30(7-8):665–688, 2009.

[17] Yael Harpaz and Yoel Shkolnisky. Three-dimensional alignment of density maps in cryo-electron microscopy. *Biological Imaging*, 3:e8, 2023.

[18] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International journal of computer vision*, 103:267–305, 2013.

[19] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.

[20] Jeremy Hoskins, Yuehaw Khoo, Oscar Mickelin, Amit Singer, and Yuguan Wang. Subspace method of moments for ab initio 3-D single-particle cryo-EM reconstruction. *arXiv preprint arXiv:2410.06889*, 2024.

[21] Du Q Huynh. Metrics for 3D rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35:155–164, 2009.

[22] Yesukhei Jagvaral, Francois Lanusse, and Rachel Mandelbaum. Unified framework for diffusion generative models in SO(3): Applications in computer vision and astrophysics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12754–12762, 2024.

[23] Noam Janco and Tamir Bendory. An accelerated expectation-maximization algorithm for multi-reference alignment. *IEEE Transactions on Signal Processing*, 70:3237–3248, 2022.

[24] Minkyu Jeon, Rishwanth Raghu, Miro Astore, Geoffrey Woollard, J Feathers, Alkin Kaz, Sonya Hanson, Pilar Cossio, and Ellen Zhong. Cryobench: Diverse and challenging datasets for the heterogeneity problem in cryo-em. *Advances in Neural Information Processing Systems*, 37:89468–89512, 2024.

[25] Norman L Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous univariate distributions, volume 2*, volume 2. John wiley & sons, 1995.

[26] Steven M Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[27] Joe Kileel, Nicholas F Marshall, Oscar Mickelin, and Amit Singer. Fast expansion into harmonics on the ball. *SIAM Journal on Scientific Computing*, 47(2):A1117–A1144, 2025.

[28] Joe Kileel, Oscar Mickelin, Amit Singer, and Sheng Xu. Two datasets are better than one: Method of double moments for 3-d reconstruction in cryo-em. *arXiv preprint arXiv:2511.07438*, 2025.

[29] Dari Kimanius, Liyi Dong, Grigory Sharov, Takanori Nakane, and Sjors HW Scheres. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal*, 478(24):4169–4185, 2021.

[30] Peter J Kostelec and Daniel N Rockmore. FFTs on the rotation group. *Journal of Fourier analysis and applications*, 14(2):145–179, 2008.

[31] Adam Leach, Sebastian M Schmon, Matteo T Degiacomi, and Chris G Willcocks. Denoising diffusion probabilistic models on SO(3) for rotational alignment. In *ICLR 2022 workshop on geometrical and topological representation learning*, 2022.

[32] Erich L Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

[33] Axel Levy, Rishwanth Raghu, J Ryan Feathers, Michal Grzadkowski, Frédéric Poitevin, Jake D Johnston, Francesca Vallese, Oliver Biggs Clarke, Gordon Wetzstein, and Ellen D Zhong. End-to-end neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *bioRxiv*, 10(2024.05):30–596729, 2024.

[34] Bufan Li, Dongjie Zhu, Huigang Shi, and Xinzheng Zhang. Effect of charge on protein preferred orientation at the air–water interface in cryo-electron microscopy. *Journal of Structural Biology*, 213(4):107783, 2021.

[35] Zhenwei Luo, Fengyun Ni, Qinghua Wang, and Jianpeng Ma. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nature Methods*, 20(11):1729–1738, 2023.

[36] Dmitry Lyumkis. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*, 294(13):5181–5197, 2019.

[37] Chao Ma, Tamir Bendory, Nicolas Boumal, Fred Sigworth, and Amit Singer. Heterogeneous multireference alignment for images with application to 2D classification in single particle reconstruction. *IEEE Transactions on Image Processing*, 29:1699–1710, 2019.

[38] M-E Mäeots and Radoslav I Enchev. Structural dynamics: Review of time-resolved cryo-EM. *Acta Crystallographica Section D: Structural Biology*, 78(8):927–935, 2022.

[39] Youdong Mao, Luis R Castillo-Menendez, and Joseph G Sodroski. Reply to Subramaniam, van Heel, and Henderson: Validity of the cryo-electron microscopy structures of the HIV-1 envelope glycoprotein complex. *Proceedings of the National Academy of Sciences*, 110(45):E4178–E4182, 2013.

[40] Youdong Mao, Liping Wang, Christopher Gu, Alon Herschhorn, Anik Désormeaux, Andrés Finzi, Shi-Hua Xiang, and Joseph G Sodroski. Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proceedings of the National Academy of Sciences*, 110(30):12438–12443, 2013.

[41] Siegfried Matthies, J Muller, and GW Vinel. On the normal distribution in the orientation space. *Texture, Stress, and Microstructure*, 10(1):77–96, 1988.

[42] Ali Punjani and David J Fleet. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of structural biology*, 213(2):107702, 2021.

[43] Ali Punjani and David J Fleet. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nature Methods*, 20(6):860–870, 2023.

[44] 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.

[45] CP Robert. Monte carlo statistical methods, 1999.

[46] Peter B Rosenthal and Richard Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of molecular biology*, 333(4):721–745, 2003.

[47] TI Savjolova. Preface to novye metody issledovanija tekstury polikristalliceskich materialov. *Metallurgija, Moscow*, 4(2):6–2, 1985.

[48] Sjors HW Scheres. A Bayesian view on cryo-EM structure determination. *Journal of molecular biology*, 415(2):406–418, 2012.

[49] Sjors HW Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, 180(3):519–530, 2012.

[50] Nir Sharon, Joe Kileel, Yuehaw Khoo, Boris Landa, and Amit Singer. Method of moments for 3d single particle ab initio modeling with non-uniform distribution of viewing angles. *Inverse Problems*, 36(4):044003, 2020.

[51] Fred J Sigworth, Peter C Doerschuk, Jose-Maria Carazo, and Sjors HW Scheres. An introduction to maximum-likelihood methods in cryo-EM. In *Methods in enzymology*, volume 482, pages 263–294. Elsevier, 2010.

[52] Amit Singer. Mathematics for cryo-electron microscopy. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3995–4014. World Scientific, 2018.

[53] Amit Singer and Ruiyi Yang. Alignment of density maps in Wasserstein distance. *Biological Imaging*, 4:e5, 2024.

[54] Carlos Oscar S Sorzano, A Jimenez, Javier Mota, Jose Luis Vilas, David Maluenda, Marta Martinez, Erney Ramirez-Aportela, Tomas Majtner, J Segura, Ruben Sanchez-Garcia, et al. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. *Acta Crystallographica Section F: Structural Biology Communications*, 75(1):19–32, 2019.

[55] M Kay Steven. Fundamentals of statistical signal processing. *PTR Prentice-Hall, Englewood Cliffs, NJ*, 10(151045):148, 1993.

[56] Sriram Subramaniam. Structure of trimeric HIV-1 envelope glycoproteins. *Proceedings of the National Academy of Sciences*, 110(45):E4172–E4174, 2013.

[57] 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.

[58] Bogdan Toader, Fred J Sigworth, and Roy R Lederman. Methods for cryo-EM single particle reconstruction of macromolecules having continuous heterogeneity. *Journal of molecular biology*, 435(9):168020, 2023.

[59] Marin van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177, 2013.

[60] Abigail JI Watson and Alberto Bartesaghi. Advances in cryo-ET data processing: meeting the demands of visual proteomics. *Current Opinion in Structural Biology*, 87:102861, 2024.

[61] Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors HW Scheres. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife*, 3:e03080, 2014.

[62] Sheng Xu, Anderson Ye Zhang, and Amit Singer. Misspecified maximum likelihood estimation for non-uniform group orbit recovery. *arXiv preprint arXiv:2509.22945*, 2025.

[63] Peijun Zhang. Advances in cryo-electron tomography and subtomogram averaging and classification. *Current opinion in structural biology*, 58:249–258, 2019.

[64] Ellen D Zhong, Tristan Bepler, Bonnie Berger, and Joseph H Davis. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature methods*, 18(2):176–185, 2021.

Appendix

A Full cryo-EM model

The (homogeneous) cryo-EM imaging process can be described by a comprehensive generative model that incorporates the key physical components of image formation, including random orientations, in-plane shifts, tomographic projection, and contrast transfer function (CTF).

Formally, each observed 2D image y_i is modeled as

$$y_i(x_1, x_2) = h_i * T_{t_i}\Big(\Pi\big((g_i \circ V)(x_1, x_2, x_3)\big)\Big) + \varepsilon_i(x_1, x_2), \quad \text{for } (x_1, x_2, x_3) \in \mathbb{R}^3, \tag{A.1}$$

where V denotes the underlying 3D volume, g_i denotes the random rotation, Π is the tomographic projection operator (i.e., line integration along x_3 axis), T_{t_i} denotes the in-plane shift operator with shift $t_i \in \mathbb{R}^2$, $*$ denotes the 2D convolution in the image plane, h_i is the point spread function (whose Fourier transform is commonly known as CTF), and ε_i denotes the measurement noise. Equivalently, by applying the Fourier transform and using the Fourier Slice Theorem, we obtain the forward model in the Fourier domain:

$$\tilde{y}_i(\omega_1, \omega_2) = \tilde{h}_i \cdot e^{-2\pi \mathbf{i} \boldsymbol{\omega} \cdot t_i} \cdot \Big((g_i \circ \tilde{V})(\omega_1, \omega_2, 0)\Big) + \tilde{\varepsilon}_i(\omega_1, \omega_2), \quad \text{for } \boldsymbol{\omega} = (\omega_1, \omega_2) \in \mathbb{R}^2, \tag{A.2}$$

where $\tilde{y}_i, \tilde{h}_i, \tilde{\varepsilon}_i$ are the 2D Fourier transforms of y_i, h_i, ε_i, respectively, and $\tilde{V}$ is the 3D Fourier transform of V.

In practice, both the images and the underlying volume are discretized and sampled on finite grids. Depending on whether the model is implemented in the spatial or Fourier domain, different but equivalent discretized versions arise. Specifically:

- In the **spatial domain**, each image y_i is represented as a vector in $\mathbb{R}^{N^2}$, and the 3D volume V is discretized as an array in $\mathbb{R}^{N^3}$.
- In the **Fourier domain**, we work with complex-valued data, where $\tilde{y}_i \in \mathbb{C}^{N^2}$ and $\tilde{V} \in \mathbb{C}^{N^3}$ are the discrete Fourier representations of the image and volume, respectively. Since the images and the volume are real-valued, it follows that $\tilde{y}_i$ and $\tilde{V}$ are conjugate symmetric, i.e., $\overline{\tilde{y}_i(\omega_1, \omega_2)} = \tilde{y}_i(-\omega_1, -\omega_2)$ and $\overline{\tilde{V}(\omega_1, \omega_2, \omega_3)} = \tilde{V}(-\omega_1, -\omega_2, -\omega_3)$.

Here, N denotes the resolution of the reconstructed grid: each image consists of $N \times N$ pixels, and the volume is represented on an $N \times N \times N$ voxel grid.

In the heterogeneous case, the underlying volume V may itself be a realization drawn from a distribution over conformational states $V_i \sim \mathcal{P}_V$ where $\mathcal{P}_V$ represents a (possibly low-dimensional) distribution capturing the variability in molecular conformations, such as discrete states, continuous manifolds. This setting gives rise to the heterogeneous cryo-EM problem, where the goal is to infer not just a single structure, but the distribution $\mathcal{P}_V$ from noisy 2D projections.

B Bayes pose estimation with rotations and shifts

To account for in-plane shifts, we model the latent pose as $\phi = (g, t) \in \mathsf{SO}(3) \times \mathbb{R}^2$, where g denotes the 3D rotation and t denotes the in-plane translation. Given a volume V, the forward model can be written as

$$p(y \mid g, t) \propto \exp\Big(-\tfrac{1}{2} \, \|y - \mathcal{A}(g, t)V\|^2_{\Sigma^{-1}} \Big), \tag{B.1}$$

where $\mathcal{A}(g, t)$ is the cryo-EM/ET forward operator including both rotation and shift (e.g., in the Fourier domain the shift acts via the phase factor $e^{-2\pi i \, \omega \cdot t}$; cf. Appendix A). We assume priors $\Lambda(g)$ on rotations and $\pi(t)$ on shifts. The joint posterior then takes the form

$$p(g, t \mid y) \propto p(y \mid g, t) \, \Lambda(g) \, \pi(t). \tag{B.2}$$

Additive loss function. To define a Bayes estimator, we specify a loss function on $\mathsf{SO}(3) \times \mathbb{R}^2$. Throughout, we adopt an *additive* pose loss,

$$\mathcal{L}\big((g, t), (\widehat{g}, \widehat{t})\big) = \mathcal{L}_g(g, \widehat{g}) + \mathcal{L}_t(t, \widehat{t}), \tag{B.3}$$

which penalizes rotation and translation errors separately (optionally with application-dependent weights to reconcile units). Under (B.3), the Bayes pose estimator is defined by

$$(\widehat{g}_{\text{Bayes}}, \widehat{t}_{\text{Bayes}}) \in \underset{\widehat{g} \in \mathsf{SO}(3), \, \widehat{t} \in \mathbb{R}^2}{\arg\min} \; \mathbb{E}\big[\mathcal{L}\big((g, t), (\widehat{g}, \widehat{t})\big) \,\big|\, y\big] \,, \tag{B.4}$$

where the expectation is taken with respect to the *joint* posterior (B.2). Because the loss is additive, the minimization in (B.4) decomposes into two marginal Bayes estimators:

$$\widehat{g}_{\text{Bayes}} = \underset{\widehat{g} \in \mathsf{SO}(3)}{\arg\min} \; \mathbb{E}[\mathcal{L}_g(g, \widehat{g}) \mid y] \,, \tag{B.5}$$

$$\widehat{t}_{\text{Bayes}} = \underset{\widehat{t} \in \mathbb{R}^2}{\arg\min} \; \mathbb{E}\big[\mathcal{L}_t(t, \widehat{t}) \mid y\big] \,. \tag{B.6}$$

Closed-form MMSE rules under squared loss. For translations, choosing Euclidean squared loss $\mathcal{L}_t(t, \widehat{t}) = \|t - \widehat{t}\|_2^2$ yields the posterior mean (similarly to (2.13))

$$\widehat{t}_{\text{MMSE}} = \mathbb{E}[t \mid y] = \int_{\mathsf{SO}(3)} \int_{\mathbb{R}^2} t\, p(g, t \mid y)\, \mathrm{d}t\, \mathrm{d}g, \tag{B.7}$$

where $p(g, t \mid y)$ is defined through (B.2). For rotations, when g is represented by its 3×3 matrix and we take the squared chordal/Frobenius loss $\mathcal{L}_g(g, \widehat{g}) = \|g - \widehat{g}\|_F^2$, the corresponding Bayes estimator is obtained by first forming the relaxed posterior mean matrix (similarly to (2.13))

$$\widehat{g}_{\text{relax}} = \mathbb{E}[g \mid y] = \int_{\mathsf{SO}(3)} \int_{\mathbb{R}^2} g\, p(g, t \mid y)\, \mathrm{d}t\, \mathrm{d}g, \tag{B.8}$$

and then projecting it onto $\mathsf{SO}(3)$ via the Orthogonal Procrustes map (similarly to (2.17)),

$$\widehat{g}_{\text{MMSE}} = \underset{\Omega \in \mathsf{SO}(3)}{\arg\min} \; \left\| \widehat{g}_{\text{relax}} - \Omega \right\|_F. \tag{B.9}$$

which returns the closest valid rotation matrix (in Frobenius norm) to $\widehat{g}_{\text{relax}}$.

C Alignment between two noisy volumes

The model described in (2.1) can be generalized to address the alignment problem between two noisy volumes. In (2.1), a single observation y was assumed to follow the model, with the volume V considered deterministic and known. For aligning two noisy volumes, we now consider two observations y_i for $i = 1, 2$, given by:

$$y_i = g_i \circ V + \epsilon_i, \tag{C.1}$$

where $\epsilon_i \sim \mathcal{N}(0, \Sigma_{d\times d})$ for a positive-definite cocaraince matrix $\Sigma_{d\times d}$, and g_i are drawn from a distribution Λ over SO(3). We assume that $\epsilon_1, \epsilon_2, g_1, g_2$ are independent, and that the volume V is deterministic but unknown. Furthermore, we assume that the noise statistics are invariant under the group actions of SO(3), meaning the distributions of ϵ and $g \circ \epsilon$ are identical for any $g \in$ SO(3). The objective is to determine the "best" alignment $\hat{g} \in$ SO(3) between y_1 and y_2, according to criteria such as the MLE, MAP, or MMSE estimators.

From (C.1), the unknown volume V can be expressed as:

$$V = g_1^{-1} \circ (y_1 - \epsilon_1). \tag{C.2}$$

Substituting (C.2) into (C.1), yields:

$$\begin{aligned} y_2 &= g_2 \circ V + \epsilon_2 && \text{(C.3)} \\ &= g_2 \circ \left(g_1^{-1} \circ (y_1 - \epsilon_1)\right) + \epsilon_2 && \text{(C.4)} \\ &= (g_2 \cdot g_1^{-1}) \circ y_1 - (g_2 \cdot g_1^{-1}) \circ \epsilon_1 + \epsilon_2 && \text{(C.5)} \\ &= (g_2 \cdot g_1^{-1}) \circ y_1 + \epsilon, && \text{(C.6)} \end{aligned}$$

where (C.5) follows from the associativity of group actions, and in(C.6), we have used $\epsilon \triangleq (g_2 \cdot g_1^{-1}) \circ \epsilon_1 + \epsilon_2$. Since the noise statistics are invariant under group actions, it follows that $\epsilon \sim \mathcal{N}(0, 2\Sigma_{d\times d})$.

Thus, the problem of estimating the relative rotation $g_2 \cdot g_1^{-1}$ between y_1 and y_2 is similar to estimating y_2 under the assumption that V is known. However, the covariance of the additive noise ϵ in this estimation is doubled compared to the case when V is known.

D Proof of Proposition 2.1

By the definition of $\hat{g}_{\mathsf{MMSE}}$ (2.12), we have,

$$\hat{g}_{\mathsf{MMSE}} = \arg\min_{\hat{g}\in\mathsf{SO}(3)} \int_{\mathsf{SO}(3)} \|g-\hat{g}\|_F^2 \, \mathbb{P}_\Lambda(g|y,V)\mathrm{d}g. \tag{D.1}$$

Then, as $\|g\|_F, \|\hat{g}\|_F$ are fixed for any $g, \hat{g} \in \mathsf{SO}(3)$, we have,

$$\hat{g}_{\mathsf{MMSE}} = \arg\max_{\hat{g}\in\mathsf{SO}(3)} \int_{\mathsf{SO}(3)} \langle \hat{g}, g\rangle \mathbb{P}_\Lambda(g|y,V)\mathrm{d}g \tag{D.2}$$

$$= \arg\max_{\hat{g}\in\mathsf{SO}(3)} \langle \hat{g}, \hat{g}_{\mathsf{relax}}\rangle. \tag{D.3}$$

As a consequence, (D.3) is equal to the rotation $\hat{g} \in \mathsf{SO}(3)$ that has the highest correlation with the relexed estimator $\hat{g}_{\mathsf{relax}}$ as in (2.13). It follows that

$$\hat{g}_{\mathsf{MMSE}} = \arg\max_{\hat{g}\in\mathsf{SO}(3)} \langle \hat{g}, \hat{g}_{\mathsf{relax}}\rangle = \arg\min_{\hat{g}\in\mathsf{SO}(3)} \|\hat{g}-\hat{g}_{\mathsf{relax}}\|_F^2 \tag{D.4}$$

Thus, $\hat{g}_{\mathsf{MMSE}}$ is exactly the Procrustes procedure performed on the intermediate estimator $\hat{g}_{\mathsf{relax}}$.

E Proof of Proposition 2.2

We split the proof into two parts. First, we show that

$$\lim_{\sigma\to 0} \hat{g}_{\mathsf{MAP}} = \hat{g}_{\mathsf{MLE}}. \tag{E.1}$$

Second, we show that

$$\lim_{\sigma\to 0} \hat{g}_{\mathsf{MMSE}} = \hat{g}_{\mathsf{MLE}}. \tag{E.2}$$

For (E.1), the argument is straightforward. Since the prior density $\Lambda(g)$ is bounded below by some positive constant, the term $\log(\mathrm{d}\Lambda(g))$ becomes negligible as $\sigma \to 0$ compared to the other term. Consequently, the MAP estimator converges to the MLE in this limit.

For (E.2), we apply the following theorem [45, Corollary 5.11]:

Theorem E.1. *Consider h a real-valued function defined on a closed and bounded set Θ, of $\mathbb{R}^p$. Let π be a positive density on Θ. If there exists a unique solution θ^* satisfying*

$$\theta^* = \arg\max_{\theta\in\Theta} h(\theta), \tag{E.3}$$

then

$$\lim_{\lambda\to\infty} \frac{\int_\Theta \theta e^{\lambda h(\theta)}\pi(\theta)\mathrm{d}\theta}{\int_\Theta e^{\lambda h(\theta)}\pi(\theta)\mathrm{d}\theta} = \theta^*, \tag{E.4}$$

provided h is continuous at θ^*.

Let $h(g) = -\left\|y - \Pi(g \circ V)\right\|^2$ and $\pi(g)\mathrm{d}g = \mathrm{d}\Lambda(g)$. We immediately get

$$\begin{aligned}
\lim_{\sigma\to 0} \hat{g}_{\text{relax}} &= \lim_{\sigma\to 0} \frac{\int_{\mathsf{SO}(3)} g \exp\Big(-\left\|y - \Pi(g\circ V)\right\|^2 / 2\sigma^2\Big)\mathrm{d}\Lambda\left(g\right)}{\int_{\mathsf{SO}(3)} \exp\Big(-\left\|y - \Pi(g\circ V)\right\|^2 / 2\sigma^2\Big)\mathrm{d}\Lambda\left(g\right)} \\
&= \lim_{\lambda\to\infty} \frac{\int_{\mathsf{SO}(3)} g \exp\Big(-\lambda\left\|y - \Pi(g\circ V)\right\|^2\Big)\mathrm{d}\Lambda(g)}{\int_{\mathsf{SO}(3)} \exp\Big(-\lambda\left\|y - \Pi(g\circ V)\right\|^2\Big)\mathrm{d}\Lambda(g)} \\
&= \arg\max_{g\in\mathsf{SO}(3)} h(g) = \hat{g}_{\mathsf{MLE}}.
\end{aligned} \tag{E.5}$$

Finally, we note that $\hat{g}_{\mathsf{MMSE}}$ is obtained by applying the Orthogonal Procrustes procedure to project $\hat{g}_{\mathsf{relax}}$ back onto $\mathsf{SO}(3)$. Since $\hat{g}_{\mathsf{MLE}}$ is itself an element of $\mathsf{SO}(3)$, it follows naturally that

$$\lim_{\sigma\to 0} \hat{g}_{\mathsf{MMSE}} = \hat{g}_{\mathsf{MLE}}. \tag{E.6}$$

F Non-uniform distribution of rotations

F.1 Isotropic Gaussian distribution of SO(3) rotations

The IG distribution $\mathcal{IG}_{\mathsf{SO}(3)}(\eta)$, parameterized by the scalar variance η^2, can be represented in an axis-angle form, with uniformly sampled axes and a rotation angle $\omega \in [0, \pi]$, having the following probability density function [47]:

$$f(\omega) = \frac{1-\cos\omega}{\pi} \sum_{\ell=0}^{\infty} (2\ell+1) e^{-\ell(\ell+1)\eta^2} \frac{\sin\left((\ell+1/2)\omega\right)}{\sin(\omega/2)}. \tag{F.1}$$

Notably, the uniform distribution on $\mathsf{SO}(3)$, denoted $\mathcal{U}_{\mathsf{SO}(3)}$, corresponds to uniformly sampled axes and is characterized by the density function $f(\omega) = \frac{1-\cos\omega}{\pi}$, which represents the limiting case of $\eta \to \infty$ in (F.1).

To provide intuition for the spread induced by η, consider the concentrated regime $\eta \ll 1$. In this regime, one may use the standard small-angle approximation in which a rotation is represented by a rotation vector $r \in \mathbb{R}^3$ with

$$r \sim \mathcal{N}(0, \eta^2 I_3), \qquad \omega \approx \|r\|, \tag{F.2}$$

so that ω is approximately Maxwell distributed with scale parameter η [25]. Consequently, its standard deviation is

$$\mathrm{Std}(\omega) \approx \eta\sqrt{3 - \frac{8}{\pi}} = \eta\sqrt{\frac{3\pi - 8}{\pi}} \approx 0.67\,\eta \quad \text{(radians)}. \tag{F.3}$$

For example, $\eta = 0.1$ corresponds to $\mathrm{Std}(\omega) \approx 0.067$ rad $\approx 3.8°$.

For $\eta > 1$, the series converges quickly, and $\ell_{\max} = 5$ is typically sufficient for sub-percent accuracy. However, as η decreases, convergence becomes slower, making it inefficient for modeling concentrated distributions. Fortunately, this series has been well studied, and an excellent approximation for $\eta < 1$ has been proposed in the literature [41], given by the following closed-form approximation:

$$f(\omega) \simeq \frac{1-\cos\omega}{\sqrt{\pi}} \eta^{-\frac{3}{2}} e^{\left(\frac{\eta}{4} - \frac{(\omega/2)^2}{\eta}\right)} \left(\frac{\omega - e^{-\pi^2/\eta}\left((\omega-2\pi)\,e^{\pi\omega/\eta} + (\omega+2\pi)\,e^{-\pi\omega/\eta}\right)}{2\sin\left(\frac{\omega}{2}\right)}\right). \tag{F.4}$$

The method for generating samples according to the isotropic Gaussian distribution on $\mathsf{SO}(3)$ relies on the inverse sampling theorem. This technique enables sampling from a specified probability distribution by utilizing its cumulative distribution function (CDF). Specifically, it involves generating a uniform random number between 0 and 1 and then applying the inverse CDF to obtain a sample from the target distribution [47].

F.2 Estimating the rotational prior

A key practical ingredient in Bayesian pose estimation is the prior over rotations. From an algorithmic standpoint, two complementary approaches are currently being developed to handle *unknown* viewing-angle distributions. First, building on EM-type refinement, recent work proposes to introduce an explicit parameterization of the orientation distribution over $\mathsf{SO}(3)$ and to jointly optimize the structure and the viewing-angle distribution in the M-step, effectively recovering the rotational prior from the data [62]. This perspective generalizes Bayesian EM refinement pipelines such as RELION by treating the pose prior as an estimable object that can be learned jointly with the structure during refinement [49]. Second, *method-of-moments* approaches relax the uniformity assumption and aim to recover both the structure and the unknown viewing-angle distribution from low-order statistics, typically via alternating optimization between the two components [50, 28].

In practice, we recommend starting with a uniform (or weak) prior when little is known about the rotational prior, and then optionally estimating a smoothed viewing-angle distribution from an initial refinement and plugging it into the MMSE estimator; simple sensitivity checks (uniform vs. learned priors, and varying regularization) can help diagnose the effect of preferred orientations. We emphasize, however, that principled estimation of the rotational prior, and a full theory of robustness to prior misspecification in cryo-EM/ET, remains an important open direction for future work.

G Proof of Proposition 4.1

Following (4.3), the update law (M-step) of the $t+1$ iteration of the expectation-maximization algorithm is given by:

$$\hat{V}^{(t+1)} = \arg\min_{V} \sum_{i=0}^{M-1} \int_{g\in\mathsf{SO}(3)} p_i^{(t)}(g)\, \|y_i - (g\circ V)\|^2 \,\mathrm{d}g, \tag{G.1}$$

where

$$p_i^{(t)}(g) = \frac{\exp\Big(-\Big\|y_i - \Big(g\circ \hat{V}^{(t)}\Big)\Big\|^2/2\sigma^2\Big)}{\int_{g\in\mathsf{SO}(3)} \exp\Big(-\Big\|y_i - \Big(g\circ \hat{V}^{(t)}\Big)\Big\|^2/2\sigma^2\Big)\mathrm{d}g}. \tag{G.2}$$

As the action by the group element g preserves the norm, we have,

$$\|y_i - (g\circ V)\|^2 = \left\|g^{-1}\circ y_i - V\right\|^2 = \left\|g^{-1}\circ y_i\right\|^2 + \|V\|^2 - 2\left\langle g^{-1}\circ y_i, V\right\rangle. \tag{G.3}$$

Thus, substituting (G.3) into (G.1), leads to:

$$\hat{V}^{(t+1)} = \arg\min_{V} \sum_{i=0}^{M-1} \int_{g\in\mathsf{SO}(3)} p_i^{(t)}(g) \left\{\left\|g^{-1}\circ y_i\right\|^2 + \|V\|^2 - 2\left\langle g^{-1}\circ y_i, V\right\rangle\right\} \mathrm{d}g. \tag{G.4}$$

As the term $\|g \circ y_i\|^2$ is independent of V, (G.4) can be simplified to:

$$\hat{V}^{(t+1)} = \arg\min_{V} \sum_{i=0}^{M-1} \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \left\{ \|V\|^2 - 2 \langle g^{-1} \circ y_i, V \rangle \right\} \mathrm{d}g. \tag{G.5}$$

Let us denote by $h(V)$ the right hand side of (G.5):

$$h(V) \triangleq \sum_{i=0}^{M-1} \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \left\{ \|V\|^2 - 2 \langle g^{-1} \circ y_i, V \rangle \right\} \mathrm{d}g. \tag{G.6}$$

Taking the first-order condition with respect to V gives

$$\frac{\partial}{\partial V} h(V) = 2 \sum_{i=0}^{M-1} \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \left(V - g^{-1} \circ y_i \right) \mathrm{d}g = 0. \tag{G.7}$$

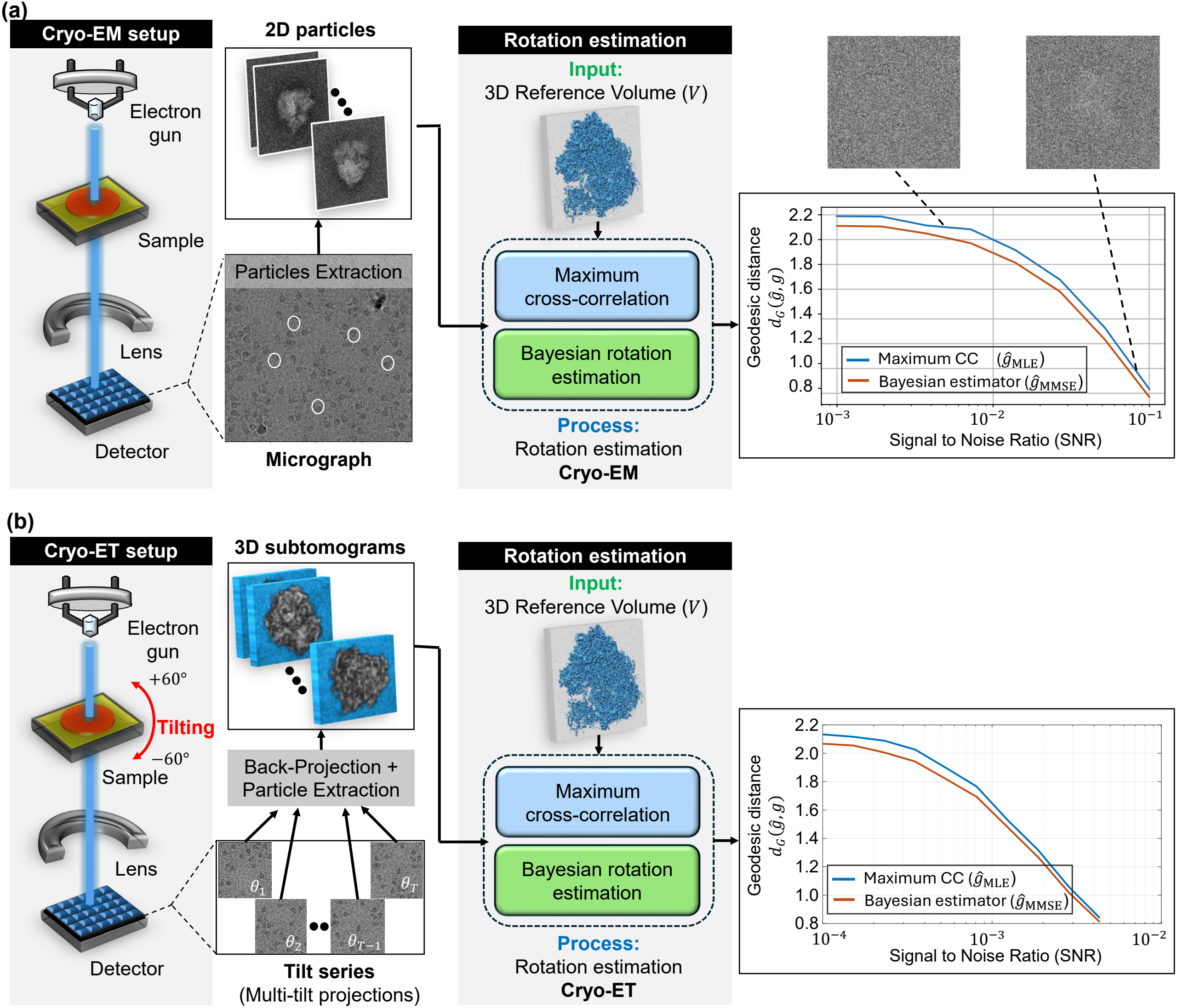
As $\int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) = 1$ for every i, we can simplify (G.7) to obtain:

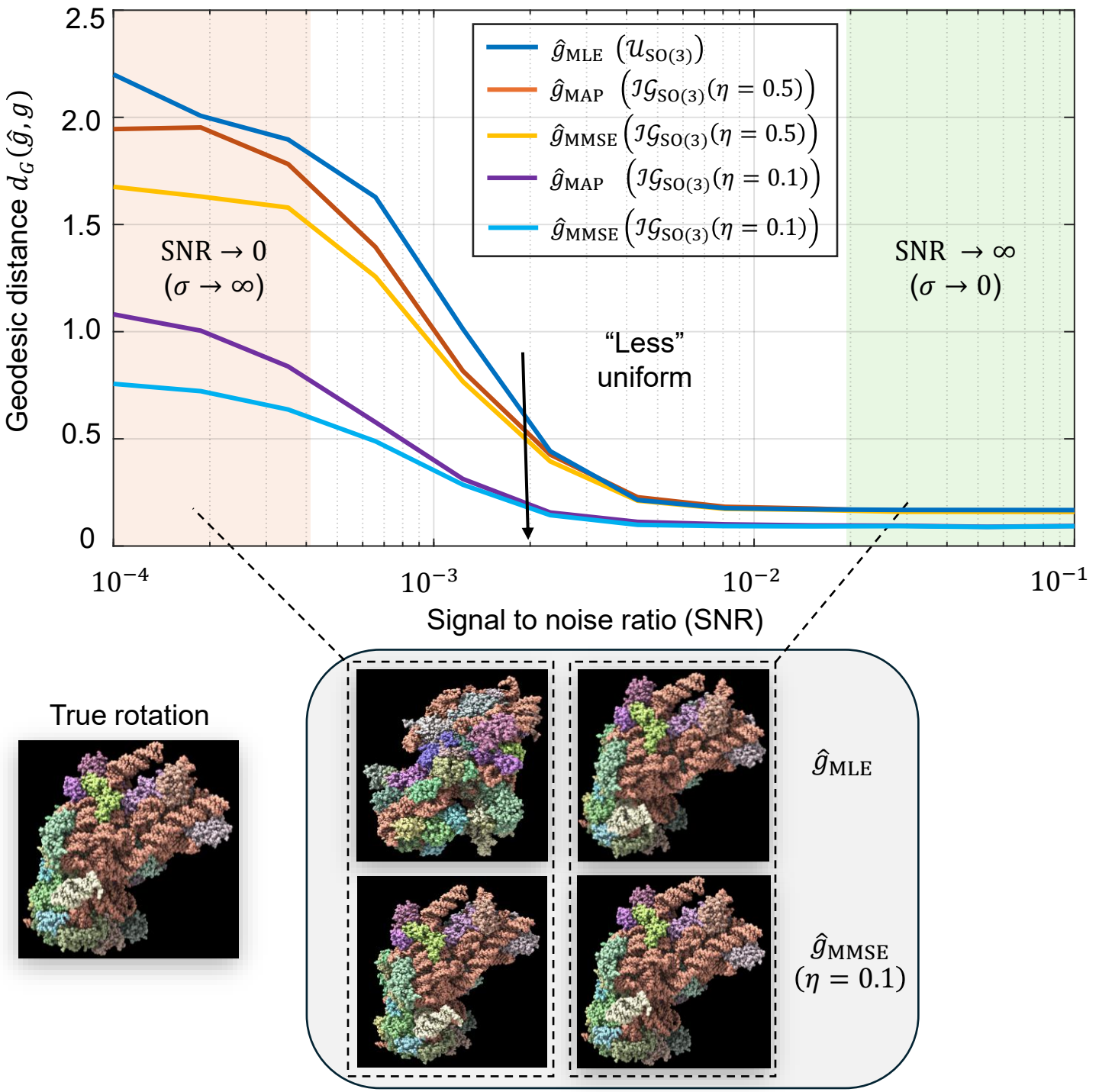
$$\begin{aligned} V =& \frac{1}{M} \sum_{i=0}^{M-1} \int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \cdot (g^{-1} \circ y_i) \\ =& \frac{1}{M} \sum_{i=0}^{M-1} \Big(\int_{g \in \mathsf{SO}(3)} p_i^{(t)}(g) \cdot g^{-1} \mathrm{d}g \Big) \circ y_i. \end{aligned} \tag{G.8}$$

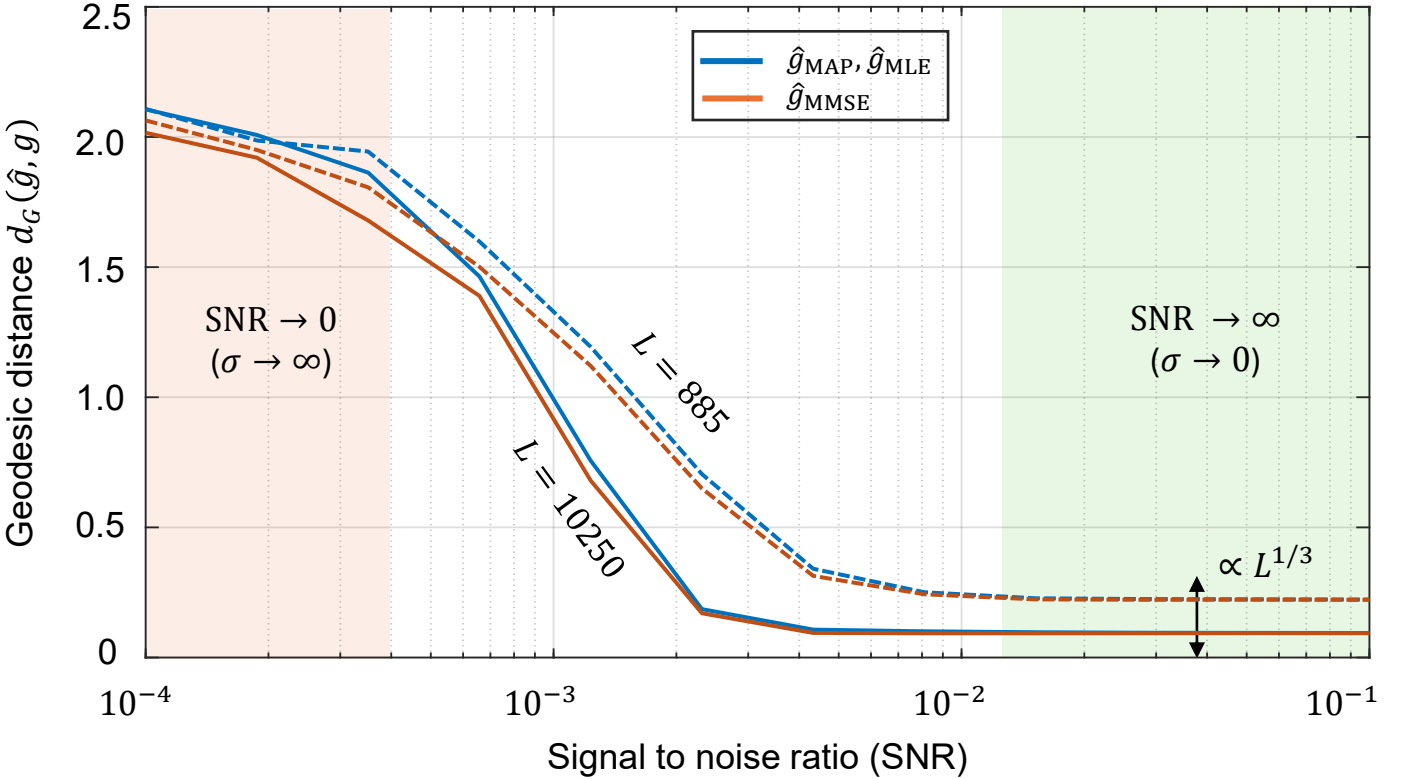
Thus, using the definition (4.5) we obtain

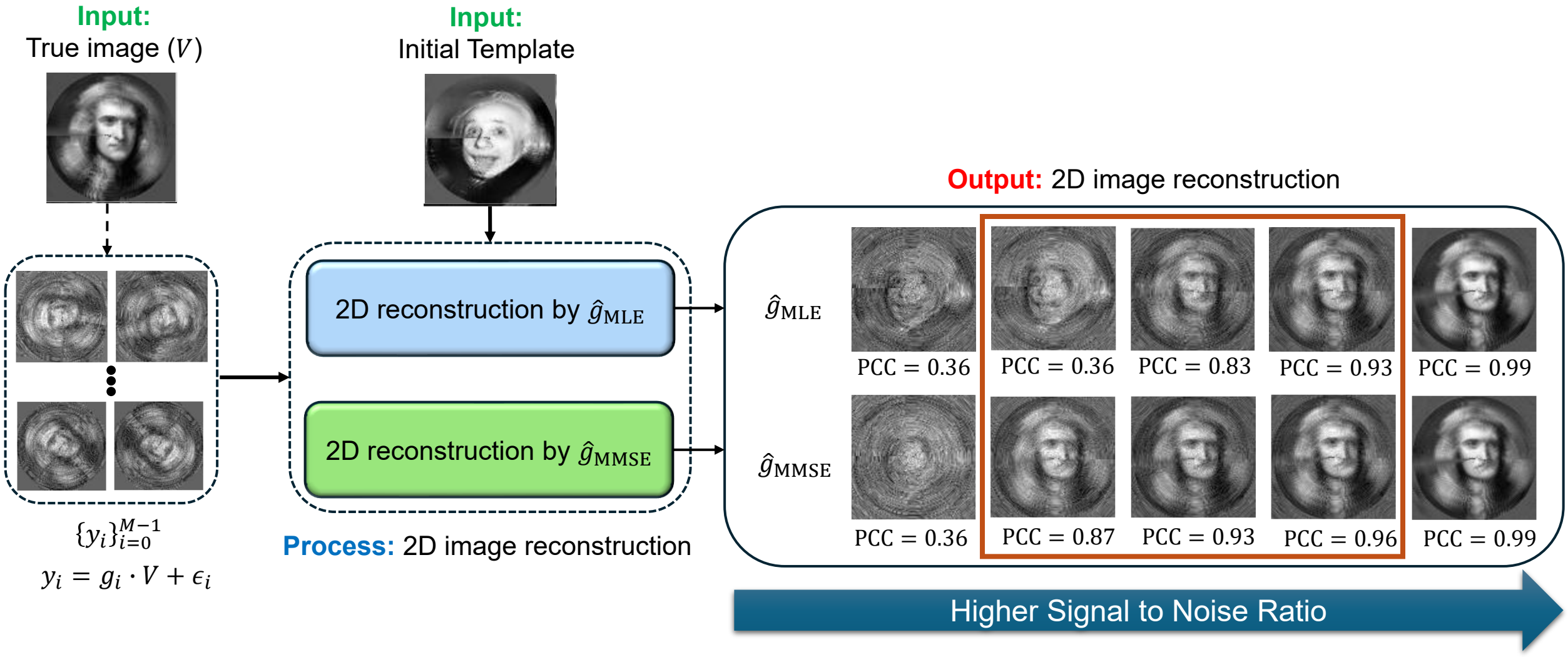
$$\hat{V}^{(t+1)} = \frac{1}{M} \sum_{i=0}^{M-1} \hat{\mathfrak{g}}_{\mathsf{MMSE},i,t} \circ y_i, \tag{G.9}$$

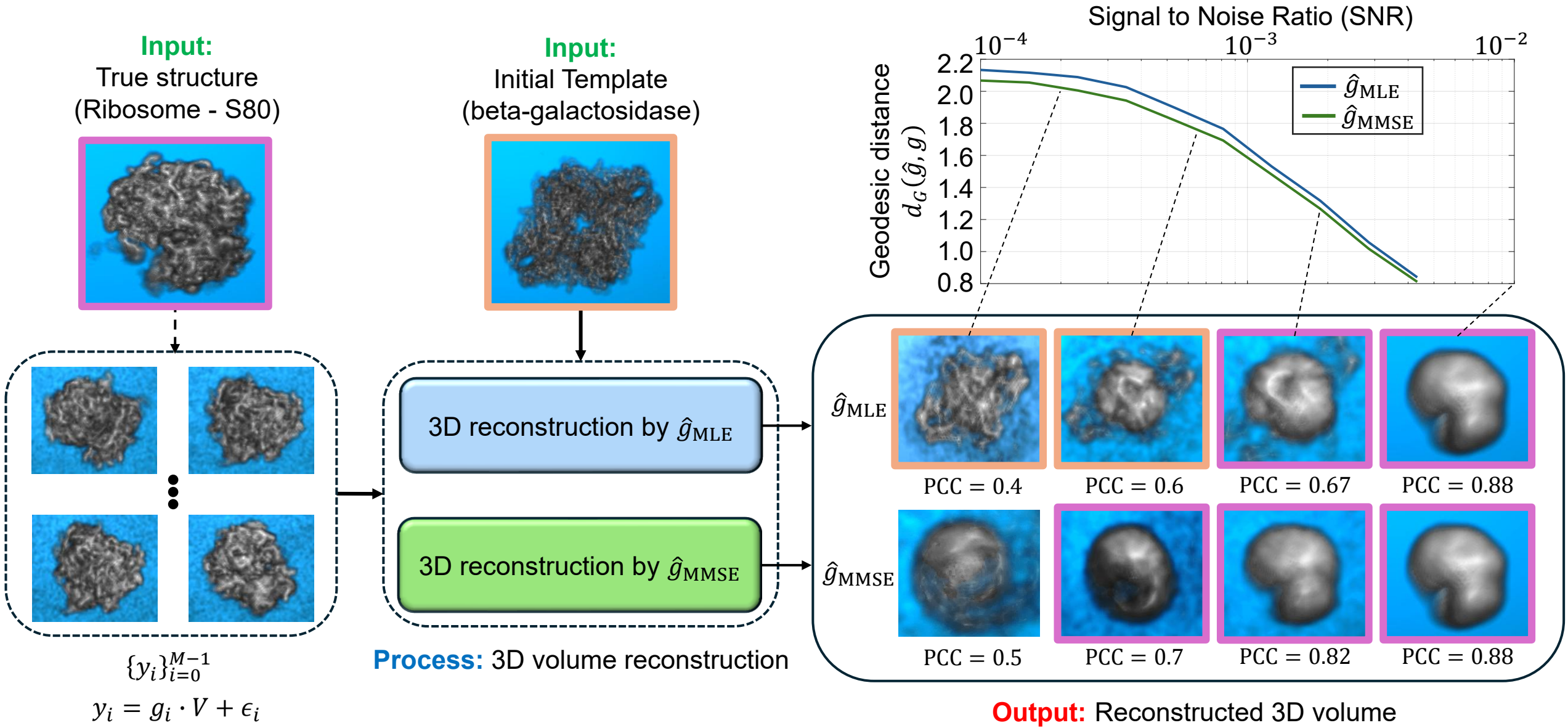
which proves the proposition.


(a)
Cryo-EM setup
Electron gun
Sample
Lens
Detector
2D particles
Particles Extraction
Micrograph
Rotation estimation
Input:
3D Reference Volume (V)
Maximum cross-correlation
Bayesian rotation estimation
Process:
Rotation estimation
Cryo-EM
Geodesic distance $d_G(\hat{g}, g)$
Maximum CC ($\hat{g}_{\mathrm{MLE}}$)
Bayesian estimator ($\hat{g}_{\mathrm{MMSE}}$)
Signal to Noise Ratio (SNR)
(b)
Cryo-ET setup
Electron gun
+60°
Tilting
−60°
Sample
Lens
Detector
3D subtomograms
Back-Projection + Particle Extraction
θ_1
θ_2
θ_{T-1}
θ_T
Tilt series
(Multi-tilt projections)
Rotation estimation
Input:
3D Reference Volume (V)
Maximum cross-correlation
Bayesian rotation estimation
Process:
Rotation estimation
Cryo-ET
Geodesic distance $d_G(\hat{g}, g)$
Maximum CC ($\hat{g}_{\mathrm{MLE}}$)
Bayesian estimator ($\hat{g}_{\mathrm{MMSE}}$)
Signal to Noise Ratio (SNR)


Geodesic distance $d_G(\hat{g}, g)$
2.5
2.0
1.5
1.0
0.5
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
Signal to noise ratio (SNR)
$\hat{g}_{\mathrm{MLE}}$ $(\mathcal{U}_{\mathrm{SO}(3)})$
$\hat{g}_{\mathrm{MAP}}$ $(\mathcal{IG}_{\mathrm{SO}(3)}(\eta = 0.5))$
$\hat{g}_{\mathrm{MMSE}}$ $(\mathcal{IG}_{\mathrm{SO}(3)}(\eta = 0.5))$
$\hat{g}_{\mathrm{MAP}}$ $(\mathcal{IG}_{\mathrm{SO}(3)}(\eta = 0.1))$
$\hat{g}_{\mathrm{MMSE}}$ $(\mathcal{IG}_{\mathrm{SO}(3)}(\eta = 0.1))$
SNR $\to 0$
$(\sigma \to \infty)$
SNR $\to \infty$
$(\sigma \to 0)$
"Less"
uniform
True rotation
$\hat{g}_{\mathrm{MLE}}$
$\hat{g}_{\mathrm{MMSE}}$
$(\eta = 0.1)$

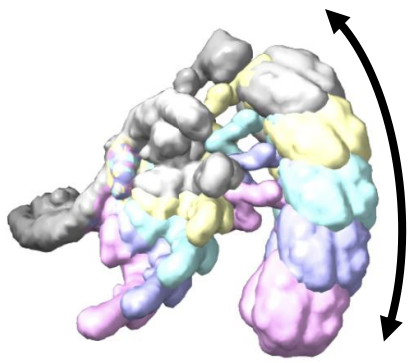

$\hat{g}_{\mathrm{MAP}}, \hat{g}_{\mathrm{MLE}}$
$\hat{g}_{\mathrm{MMSE}}$
SNR → 0
($\sigma \to \infty$)
SNR → ∞
($\sigma \to 0$)
$L = 885$
$L = 10250$
$\propto L^{1/3}$
Geodesic distance $d_G(\hat{g}, g)$
Signal to noise ratio (SNR)
2.5
2.0
1.5
1.0
0.5
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}


Input:
True image (V)
Input:
Initial Template
$\{y_i\}_{i=0}^{M-1}$
$y_i = g_i \cdot V + \epsilon_i$
2D reconstruction by $\hat{g}_{\mathrm{MLE}}$
2D reconstruction by $\hat{g}_{\mathrm{MMSE}}$
Process: 2D image reconstruction
Output: 2D image reconstruction
$\hat{g}_{\mathrm{MLE}}$
$\hat{g}_{\mathrm{MMSE}}$
PCC = 0.36
PCC = 0.36
PCC = 0.83
PCC = 0.93
PCC = 0.99
PCC = 0.36
PCC = 0.87
PCC = 0.93
PCC = 0.96
PCC = 0.99
Higher Signal to Noise Ratio

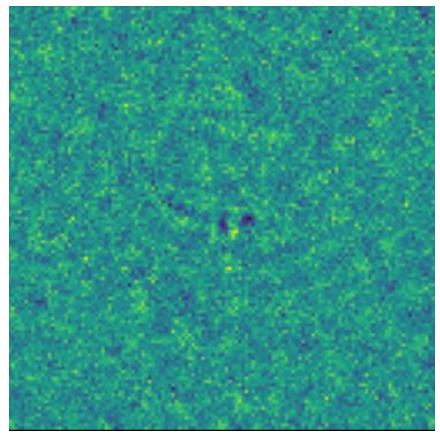

Input:
True structure
(Ribosome - S80)
Input:
Initial Template
(beta-galactosidase)
3D reconstruction by $\hat{g}_{\mathrm{MLE}}$
3D reconstruction by $\hat{g}_{\mathrm{MMSE}}$
$\{y_i\}_{i=0}^{M-1}$
$y_i = g_i \cdot V + \epsilon_i$
Process: 3D volume reconstruction
Signal to Noise Ratio (SNR)
10^{-4}
10^{-3}
10^{-2}
Geodesic distance
$d_G(\hat{g}, g)$
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
$\hat{g}_{\mathrm{MLE}}$
$\hat{g}_{\mathrm{MMSE}}$
$\hat{g}_{\mathrm{MLE}}$
PCC = 0.4
PCC = 0.6
PCC = 0.67
PCC = 0.88
$\hat{g}_{\mathrm{MMSE}}$
PCC = 0.5
PCC = 0.7
PCC = 0.82
PCC = 0.88
Output: Reconstructed 3D volume

Synthetic dataset

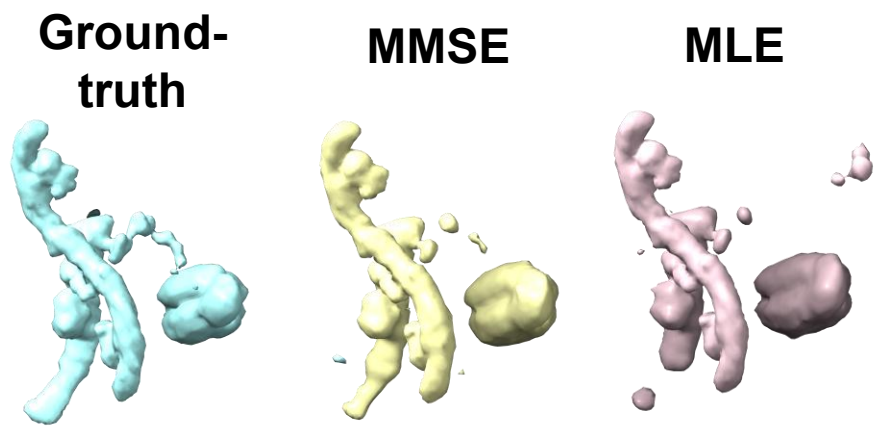
(a)



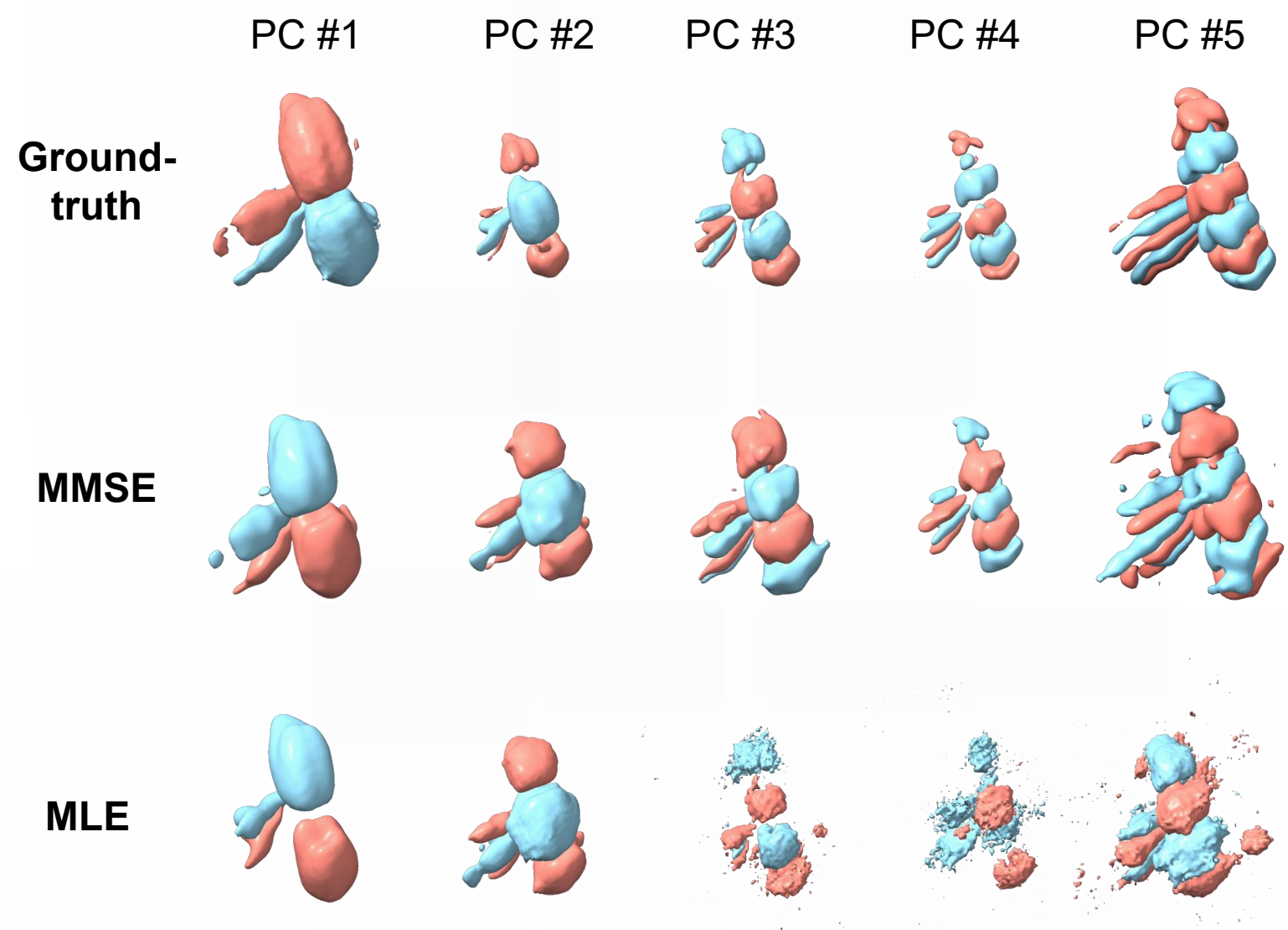
(b)



(c) **Reconstructed structures**

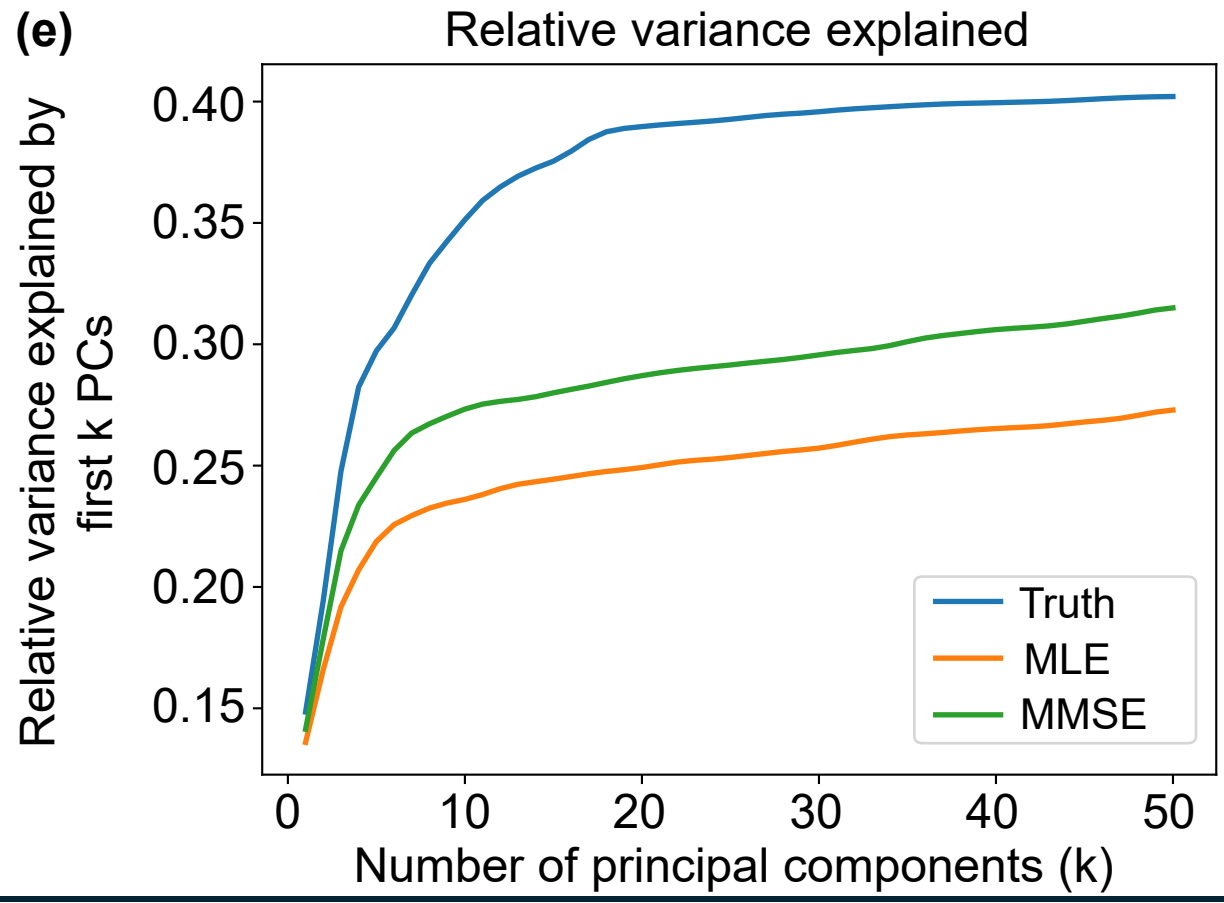
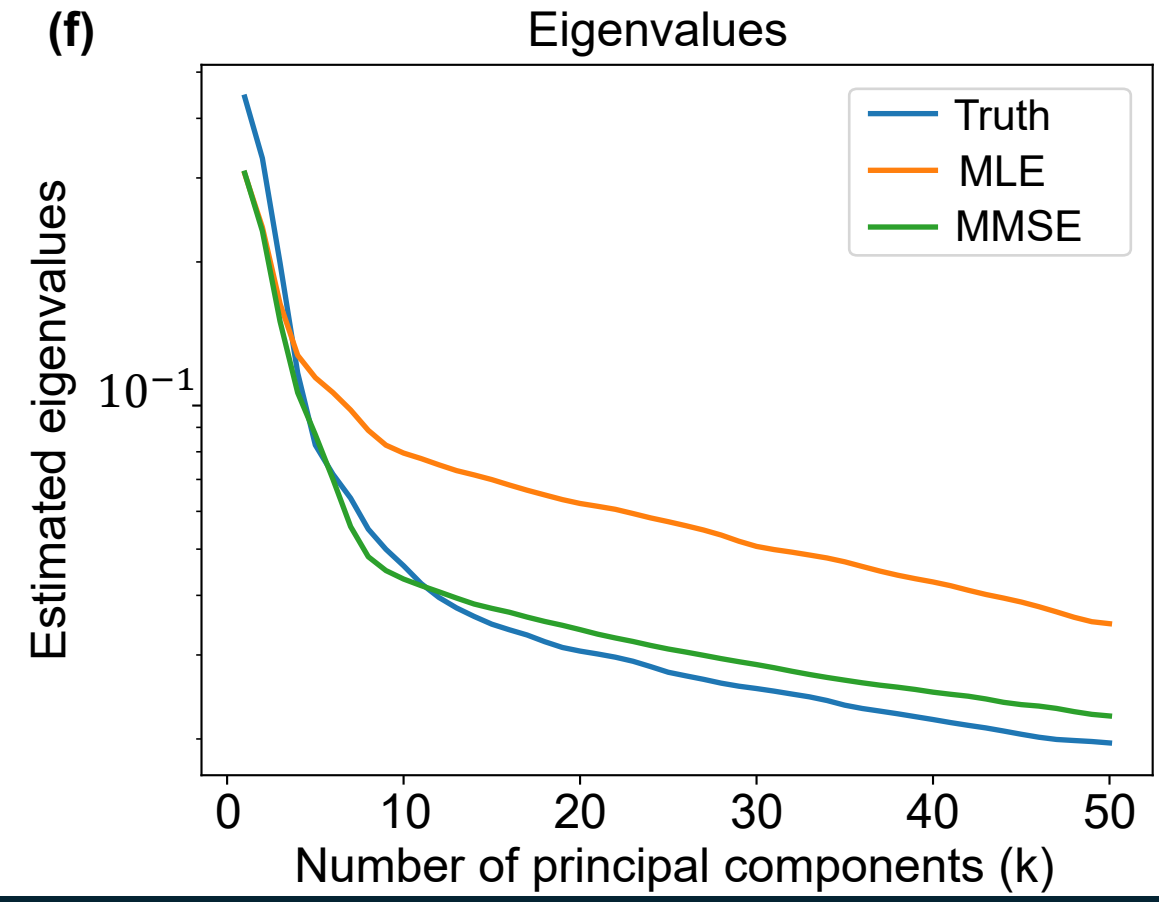


(d) **Visualization of the first 5 estimated principal components**



Estimated principal components and eigenvalues comparison

(e)



(f)