

Representation theoretic patterns in multi-frequency class averaging for three-dimensional cryo-electron microscopy

YIFENG FAN

Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 W Main St, Urbana, IL 61801

TINGRAN GAO

Committee on Computational and Applied Mathematics, Department of Statistics, University of Chicago, 5747 S Ellis Ave, Chicago, IL 60637

AND

ZHIZHEN ZHAO[†]

Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 W Main St, Urbana, IL 61801

[†]Corresponding author. Email: zhizhenz@illinois.edu

[Received on 25 June 2019; revised on 1 January 2021; accepted on 11 April 2021]

We develop in this paper a novel intrinsic classification algorithm—*multi-frequency class averaging* (MFCA)—for classifying noisy projection images obtained from three-dimensional cryo-electron microscopy by the similarity among their viewing directions. This new algorithm leverages multiple irreducible representations of the unitary group to introduce additional redundancy into the representation of the optimal in-plane rotational alignment, extending and outperforming the existing class averaging algorithm that uses only a single representation. The formal algebraic model and representation theoretic patterns of the proposed MFCA algorithm extend the framework of Hadani and Singer to arbitrary irreducible representations of the unitary group. We conceptually establish the consistency and stability of MFCA by inspecting the spectral properties of a generalized local parallel transport operator through the lens of Wigner D -matrices. We demonstrate the efficacy of the proposed algorithm with numerical experiments.

Keywords: representation theory; spectral theory; differential geometry; Wigner matrices; cryo-electron microscopy; mathematical biology.

1. Introduction

The past decades have witnessed an emerging and continued impact of cryo-electron microscopy (cryo-EM), the Nobel Prize winning imaging technology for determining three-dimensional structures of macromolecules, on a wide range of natural scientific fields [13, 18, 26, 40, 48, 52]. Compared with its predecessor, X-ray crystallography, of which the success builds upon the potentially difficult procedure of crystallization, cryo-EM is able to image the macromolecules in their native states and produces large numbers of projection images for samples of molecules rapidly frozen in a thin layer of vitreous ice. The projection images can be thought of as tomographic projections of many copies of an identical molecule at unknown and random orientations. A major computational challenge in reconstructing the three-dimensional molecular structure from these projection images is the extremely

low signal-to-noise ratio (SNR) caused by the limited allowable electron dose (so as to avoid damaging the molecule before the imaging completes). It is thus customary to improve the SNR by performing *class averaging*—the procedure of aligning and then averaging out projection images taken along nearby viewing directions—from rotationally invariant pairwise comparisons of the projection images [26, 53], before the downstream reconstruction workflow such as angular reconstitution [39, 56, 65]. In addition to its scientific value, the rich geometric structure in the cryo-EM imaging model has also inspired many mathematical and algorithmic investigations [3, 4, 9, 29–31, 35, 59, 61, 62, 67, 71, 75].

1.1 Background: the mathematical model of cryo-electron microscopy and class averaging

Following [38, 63], we view the collection of projection images $\{I_i \in \mathbb{R}^{L \times L} \mid i = 1, \dots, N\}$ as tomographic projection images for the same three-dimensional object along projection directions uniformly sampled from the two-sphere S^2 , as it is more convenient to consider the imaging model in the molecule’s own lab frame, where the molecule is fixed and observed by an electron microscope at various orientations. For simplicity, we assume the projection images are all centered, i.e., the center of mass of the clean projection images is at the center of the images. The goal is to identify and classify projection images produced from similar projection directions, hereafter referred to as *viewing directions*.

A point $x \in \text{SO}(3)$ is identified with an orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ of $\mathbb{R}^3$, with orientation compatible with the canonical orthonormal coordinate frame of $\mathbb{R}^3$. We identify $\mathbf{e}_3 \in S^2$ with the viewing direction and denote it for $\pi(x)$ for the ease of notations. The two-dimensional image obtained by the microscope observed at a spatial orientation x is a real valued function $I : \mathbb{R}^2 \rightarrow \mathbb{R}$, given by the X-ray transform along the viewing direction:

$$I(s, t) = \int_{\mathbb{R}} \phi(s\mathbf{e}_1 + t\mathbf{e}_2 + r\mathbf{e}_3) dr \quad \text{for all } (s, t) \in \mathbb{R}^2 \quad (1.1)$$

where $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ is a real-valued function modeling the electromagnetic potential induced from the charges of the molecule. We assume the images $I(s, t)$ are all supported on a bounded set of $\mathbb{R}^2$ which fits into the size of the projection images.

To measure the similarity between any two projection images I_i and I_j , obtained by the tomographic projection along viewing directions $\pi(x_i) \in S^2$ and $\pi(x_j) \in S^2$, respectively, we compute a rotationally invariant distance between I_i and I_j defined as

$$d_{\text{RID}}(I_i, I_j) = \min_{\theta \in [0, 2\pi)} \|I_i - R_\theta(I_j)\|_{\text{F}}, \quad (1.2)$$

where $R_\theta(I_j)$ stands for the operation of rotating image I_j by an angle $\theta \in [0, 2\pi)$ in the counterclockwise orientation and $\|\cdot\|_{\text{F}}$ is the matrix Frobenius norm. The optimal alignment angle between I_i and I_j will be denoted as

$$\theta_{ij} = \arg \min_{\theta \in [0, 2\pi)} \|I_i - R_\theta(I_j)\|_{\text{F}}. \quad (1.3)$$

For images I_x and I_y obtained from viewing directions $\pi(x)$ and $\pi(y)$ for $x, y \in \text{SO}(3)$ and without noise contamination, [38] models the optimal alignment angle as the *transport data* encoding the angle of in-plane rotation needed to align frames x, y after one of them is parallel transported to the fiber

of the other using the canonical Levi–Civita connection on the unit sphere equipped with an induced Riemannian structure from the ambient space $\mathbb{R}^3$. A rough idea for filtering out far-apart viewing directions is through thresholding the rotationally invariant distances between pairs of projection images against a preset threshold parameter $\epsilon > 0$ that should be tuned to reflect the confidence in the accuracy of the imaging process. The pairwise comparison information after thresholding can be conveniently encoded into an *observation graph* $G = (V, E)$, where each vertex of G stands for one of the projection images and an edge (i, j) belongs to the edge set E if and only if the rotationally invariant distance $d_{\text{RID}}(I_i, I_j)$ is smaller than the threshold. In an ideal noiseless world, the geometry of the graph G is a neighborhood graph on the unit sphere S^2 , namely, two images are connected if and only if their viewing directions $\pi(x_i)$ and $\pi(x_j)$ are close on the unit sphere, $\langle \pi(x_i), \pi(x_j) \rangle \geq 1 - h$, for $h \ll 1$. From the noisy cryo-EM images, the rotationally invariant distances d_{RID} are affected by noise and d_{RID} -based similarity measure will connect images of very different views, introducing short-cut edges on the unit sphere. The main problem here is thus to distinguish the ‘good’ edges from the ‘bad’ ones in the graph G , or, in other words, to distinguish the true neighbors from the outliers. The existence of outliers makes the classification problem non-trivial. Without excluding the outliers, averaging rotationally aligned images with small invariant distance (1.2) yields a poor estimate of the true signal, rendering infeasible the three-dimensional *ab initio* reconstruction from denoised images. We refer interested readers to [21, 47] for more detailed statistical analysis of the rotationally invariant distance (1.2). The focus of this paper is to rectify the noise-contaminated empirical transport data using the spectral information of an integral operator constructed from the initial local transport data.

1.1.1 The class averaging algorithm One of the most natural ideas for performing class averaging is through the eigenvectors of the class averaging matrix constructed from the empirical transport data $\{e^{i\theta_{ij}}\}_{(i,j) \in E}$ [38, 63]. We briefly recapture the main steps in the class averaging algorithm below. Detailed discussions and the analysis of representation theoretical patterns can be found in [38, 63]. In this section, we use notation $[N] = \{1, 2, \dots, N\}$ for $N \in \mathbb{N}$.

The algorithm begins with computing rotationally invariant distances d_{ij} between all pairs of projection images I_i and I_j , along with the corresponding optimal alignment angles θ_{ij} . After that, construct an N -by- N Hermitian matrix H by

$$H_{ij} = \begin{cases} e^{i\theta_{ij}} & \text{if } (i, j) \in E, \\ 0 & \text{otherwise,} \end{cases} \quad (1.4)$$

where the edge set $E \subset [N] \times [N]$ is obtained by thresholding the pairwise distances $\{d_{ij} : 1 \leq i, j \leq N\}$, i.e., $(i, j) \in E$ if and only if d_{ij} is below a preset threshold $\epsilon > 0$, i.e.,

$$E := \left\{ (i, j) \in [N] \times [N] : d_{\text{RID}}(I_i, I_j) < \epsilon \right\}. \quad (1.5)$$

Set D as the diagonal matrix with diagonal entries

$$D_{ii} = \sum_{j=1}^N |H_{ij}|, \quad 1 \leq i \leq N \quad (1.6)$$

and compute the top three eigenvectors $\psi_1, \psi_2, \psi_3 \in \mathbb{C}^N$ of the normalized Hermitian matrix

$$\tilde{H} := D^{-1/2} H D^{-1/2}.$$

Each projection image is then associated with a point in $\mathbb{C}^3$ by means of the embedding map

$$\begin{aligned} \Psi : \{I_i\}_{i=1}^N &\longrightarrow \mathbb{C}^3 \\ I_i &\longmapsto (\psi_1(i), \psi_2(i), \psi_3(i)) \end{aligned}$$

where $\psi_1(i), \psi_2(i), \psi_3(i)$ denotes for the i th entries of ψ_1, ψ_2, ψ_3 , respectively. The measure of affinity between I_i and I_j is then computed using the embedding map Ψ :

$$A_{ij} := \frac{\left| \langle \Psi(I_i), \Psi(I_j) \rangle \right|}{\|\Psi(I_i)\| \|\Psi(I_j)\|}, \quad 1 \leq i \neq j \leq N. \quad (1.7)$$

Finally, the neighbors of a projection image I_i are determined by thresholding the affinity measures A_{ij} :

$$\text{Neighbors of } I_i := \{I_j \mid A_{ij} > 1 - \gamma\}$$

where $0 < \gamma < 1$ is another preset threshold parameter that controls the size of the neighborhoods.

1.2 Main contributions

The main contributions of this paper are (1) the introduction of the multi-frequency class averaging (MFCA) algorithm to improve the viewing direction classification of cryo-EM single particle images and (2) a complete characterization of the spectral information of a *generalized local parallel transport operator* underlying the geometric relation in MFCA.

Specifically, motivated by recent works [3, 24, 25, 32], which incorporate multiple representations of the pairwise comparison information into the synchronization problem, we propose in this paper an *MFCA* algorithm using the extended empirical transport data $\{e^{ik\theta_{ij}}\}_{(i,j) \in E}$ for $k = 1, 2, \dots, k_{\max}$. It creates more than one copy of the class averaging matrix—one for each ‘frequency channel’ corresponding to one irreducible representation of $\text{SO}(2)$ group element. Those matrices can be viewed as the discretization of the generalized local parallel transport operators $T_h^{(k)}$. A formal definition of $T_h^{(k)}$ can be found in (4.2). The new algorithm uses the top $2k + 1$ eigenvectors of the class averaging matrix at frequency k to embed the images into $2k + 1$ -dimensional complex space. The new frequency k -affinity measure is defined as the absolute normalized cross correlation of the embedded vectors. We also propose to aggregate the affinity measures across the frequency channels to enforce the consistency of the nearest neighbor identification. Since the performance of the algorithm depends on the properties and stability of the top eigenvectors, we perform the spectral analysis of the corresponding integral operator $T_h^{(k)}$. We show in Theorem 4.1 and Theorem 4.2 that the top eigenspace of $T_h^{(k)}$, denoted as $\mathbb{W}^{(k)}$, is $(2k + 1)$ -dimensional. In addition, we show that the top eigenvalue of $T_h^{(k)}$ decreases as k increases and the top spectral gap increases as k increases up to a threshold determined by the local neighborhood size. The increasing spectral gap implies the advantage of using higher frequency

information for class averaging, as the *numerical stability* of the eigen-decomposition step in MFCA depends on the magnitude of the spectral gap.

In addition to the characterization of the dimensionality of the top eigenspace $\mathbb{W}^{(k)}$ of $T_h^{(k)}$, we also demonstrate in Theorem 4.3 and Theorem 4.4 the existence of a canonical identification of $\mathbb{W}^{(k)}$ with a complex $(2k + 1)$ -dimensional linear space spanned by $(2k + 1)$ linearly independent entry functions in the *Wigner D-matrix* associated with the unique $(2k + 1)$ -dimensional unitary irreducible representation of $\text{SO}(3)$. A direct corollary of this canonical identification is the equality between the frequency- k affinity measure and the viewing angle, thus generalizing the result in [38] for the affinity measure (1.7). These facts establish the admissibility (consistency) of the proposed MFCA algorithm.

We emphasize that these theoretical results are not straightforward extensions of the techniques in [38] to the generalized localized parallel transport operator $T_h^{(k)}$. The generating-function-based approach in [38] is not easy to generalize to our setting without heavy notation and lengthy mathematical inductions. Instead, we observed that the constructions in [38] can be greatly simplified using an alternative construction by means of the Wigner D -matrices, which has been widely used in studies in mathematical physics concerning the irreducible representation of $\text{SO}(3)$.

In the clean, noiseless scenario, the MFCA matrices certainly carry identical information for exactly recovering the affinity among view directions of the projection images; the real advantage, as argued and demonstrated in the theoretical analysis of [32] and the experimental results of [24, 25], lies at the low SNR region where utilizing higher-moment information becomes particularly beneficial even without introducing additional independent measurements for those higher moments. Empirically, we observe that the algorithm can tolerate higher level of noise than what is allowed according to the traditional Davis–Kahan theorem [17]. In addition, the performance of the single frequency- k class averaging algorithm improves as k increases up to a critical frequency index determined by the spectral gap, magnitudes of the top eigenvalues and the noise level.

Besides the improved numerical stability due to increased spectral gap, using higher frequency information for class averaging can also be interpreted as leveraging the additional redundancy encoded in the consistency of the ‘higher order moments’, which is in line with our continued exploration for a ‘geometric harmonic retrieval’ initiated in [24, 25, 32]. Moreover, in contrast with the computationally demanding SDP approach in [3] or the noise-type-dependent approximate message passing approach in [54], the proposed MFCA algorithm is easily parallelizable as the eigen-decompositions for the class averaging matrices in each frequency channel are completely independent.

1.3 Organization of the paper

The rest of this paper is organized as follows. Section 2 introduces the MFCA algorithms; Section 3 introduces the basic mathematical set-up and notations for the spectral analysis in the remainder of this paper; Section 4 presents the main theoretical contributions; Section 5 interprets the admissibility of MFCA using the theoretical results; Section 6 discusses the noise robustness for the algorithm under two probabilistic models; Section 7 illustrates the efficacy of MFCA through some numerical experiments; Section 8 concludes and discusses potential future directions. The basics on group and representation theory and technical proofs are deferred to the Appendix.

2. Multi-frequency class averaging algorithms

Throughout our discussion involving multiple frequency channels, we will fix an integer $k_{\max} \geq 1$ for the total number of frequency channels considered. For each frequency $k = 1, \dots, k_{\max}$, we construct a

separate class averaging matrix by

$$H_{ij}^{(k)} = \begin{cases} e^{ik\theta_{ij}} & \text{if } (i, j) \in E \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

2.1 Single frequency- k affinity measure

The Hermitian matrix $H^{(k)}$ stores the empirical transport data under the k th irreducible representation of $\text{SO}(2)$. We then normalize each $H^{(k)}$ using the same degree matrix D as in (1.6); note that all matrices $H^{(k)}$ share the same sparsity pattern determined by E . After performing eigen-decomposition for $\tilde{H} = D^{-1/2}H^{(k)}D^{-1/2}$, we keep the top $(2k + 1)$ eigenvectors $\psi_1^{(k)}, \dots, \psi_{2k+1}^{(k)} \in \mathbb{C}^N$ and define the embedding

$$\begin{aligned} \Psi^{(k)} : \{I_i\}_{i=1}^N &\longrightarrow \mathbb{C}^{2k+1} \\ I_i &\longmapsto \left(\psi_1^{(k)}(i), \dots, \psi_{2k+1}^{(k)}(i) \right). \end{aligned} \quad (2.2)$$

We compute the affinity measure between I_i and I_j at frequency k as

$$A_{ij}^{(k)} := \frac{\left| \left\langle \Psi^{(k)}(I_i), \Psi^{(k)}(I_j) \right\rangle \right|}{\left\| \Psi^{(k)}(I_i) \right\| \left\| \Psi^{(k)}(I_j) \right\|}, \quad 1 \leq i \neq j \leq N. \quad (2.3)$$

Obviously, $\Psi^{(1)} = \Psi$ and $A_{ij}^{(1)} = A_{ij}$ in the traditional class averaging. We can perform κ -nearest neighbor search using the affinity measure $A_{ij}^{(k)}$ computed from an individual frequency k . The rationale behind the specific forms of (2.2) and (2.3) is the core of this paper. In a nutshell, we use a $(2k + 1)$ -dimensional embedding because by Theorem 4.1 and Theorem 4.2 we expect a spectral gap occurring between the $(2k + 1)^{\text{th}}$ and $(2k + 2)^{\text{th}}$ eigenvector of $H^{(k)}$ (counting multiplicities). The affinity measure (2.3) is related to the closeness of two viewing directions by the relation (4.15) in Theorem 4.4.

2.2 Combining information from multiple frequencies

Since each affinity measure in (2.3) reflects the closeness of two viewing directions, combining those scores together can enforce the consistency of the classification results at each frequency and improve the overall accuracy. We propose one way to aggregate the single frequency affinity measure as

$$A_{ij}^{\text{All}} := \prod_{k=1}^{k_{\max}} A_{ij}^{(k)}. \quad (2.4)$$

We choose aggregation (2.4) because the affinity measure (2.3) is related to the viewing angle by the relation (4.15) in Theorem 4.4. In particular, comparing (4.15) and [38, Theorem 6) tells us that

$$A_{ij}^{(k)} = A_{ij}^k \quad \text{for all } 1 \leq i \neq j \leq N.$$

We defer more detailed discussions of the geometric relation of this algorithm to Section 4.3 and Section 5.

REMARK 2.1 Note that (2.3) and (2.4) are not the only ways to distill and aggregate the affinity information from multiple irreducible representations. Other natural alternatives include

$$S_{ij}^{(k)} := 2 \left(\frac{\left| \left\langle \Psi^{(k)}(I_i), \Psi^{(k)}(I_j) \right\rangle \right|}{\left\| \Psi^{(k)}(I_i) \right\| \left\| \Psi^{(k)}(I_j) \right\|} \right)^{\frac{1}{k}} - 1, \quad 1 \leq i \neq j \leq N \quad (2.5)$$

which in the noiseless scenario satisfies

$$S_{ij}^{(k)} = S_{ij}^{(1)} = 2A_{ij} - 1, \quad \text{for all } k \geq 1.$$

Therefore, it is natural to combining all $G_{ij}^{(k)}$ by arithmetic averaging

$$S_{ij}^{\text{All}} := \frac{1}{k_{\max}} \sum_{k=1}^{k_{\max}} S_{ij}^{(k)}. \quad (2.6)$$

However, our empirical experiments suggest that it is numerically much more stable to avoid taking k th roots for large values of k . We provide a brief interpretation of this phenomenon in Section 5.

There can be other approaches to combine the affinity scores from multiple frequencies, such as weighted average among different frequencies or majority voting. We will explore other ways to integrate multi-frequency information in the future.

3. Preliminaries for the spectral analysis of MFCA

In this section, we introduce our set-up and notations for the spectral analysis of MFCA. For additional concepts in the relevant group and representation theory and Wigner D -matrix, please refer to A.

3.1 Set-up

Throughout this paper, we view $\text{SO}(3)$ as a $\text{SO}(2)$ -bundle over the two-dimensional sphere S^2 in $\mathbb{R}^3$. For any $d \in \mathbb{N}_+$, we view $\mathbb{C}^d$ as a Hilbert product space equipped with the canonical Hermitian inner product induced from the standard Euclidean inner product on $\mathbb{R}^d$. We will distinguish two different types of group actions on $\text{SO}(3)$: if $g \in \text{SO}(3)$, g acts on elements of $\text{SO}(3)$ by left multiplication, denoted as

$$g \triangleright x := gx, \quad \forall g, x \in \text{SO}(3).$$

If $w \in \text{SO}(2) \subset \text{SO}(3)$, unless otherwise specified, w is assumed to be uniquely identified with an $\text{SO}(3)$ element by

$$w = w(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{for } \theta \in [0, 2\pi), \quad (3.1)$$

and acts on elements of $\text{SO}(3)$ by right multiplication, i.e.,

$$x \triangleleft w := xw, \quad \forall x \in \text{SO}(3), w \in \text{SO}(2).$$

Unless confusions arise, we will also denote $x \triangleleft g =: xg$, $g, x \in \text{SO}(3)$ for the right action of $\text{SO}(3)$ on itself, when the context is clear.

Following the convention of [38], we denote the transport data between $x, y \in \text{SO}(3)$ by $T(x, y)$, the unique $\text{SO}(2)$ element satisfying

$$x \triangleleft T(x, y) = t_{\pi(x), \pi(y)} y, \quad (3.2)$$

where $t_{\pi(x), \pi(y)}$ is the parallel transport along the unique geodesic on S^2 connecting $\pi(y)$ to $\pi(x)$. The optimal alignment angle θ_{ij} computed from (1.3) can be used to construct an approximation of the transport data between x_i and x_j (the observation frames of I_i and I_j , respectively), at the presence of measurement and discretization error, by

$$\tilde{T}(x_i, x_j) := e^{t\theta_{ij}}. \quad (3.3)$$

We refer to the $\tilde{T}(x_i, x_j)$'s as the *empirical transport data*. As shown in [38], $T(x, y)$ satisfy the following properties:

$$T(x, y) = T(y, x)^{-1}, \quad \forall x, y \in \text{SO}(3) \quad (\text{Symmetry})$$

$$T(g \triangleright x, g \triangleright y) = T(x, y), \quad \forall x, y \in \text{SO}(3), \forall g \in \text{SO}(3) \quad (\text{Invariance})$$

$$T(x \triangleleft w_1, y \triangleleft w_2) = w_1^{-1} T(x, y) w_2, \quad \forall x, y \in \text{SO}(3), \forall w_1, w_2 \in \text{SO}(2). \quad (\text{Equivariance})$$

If $\rho : \text{SO}(2) \rightarrow \mathbb{C}$ is any unitary representation of $\text{SO}(2)$ on $\mathbb{C}$, then the three properties above can also be cast into

$$\rho(T(x, y)) = \overline{\rho(T(y, x))}, \quad \forall x, y \in \text{SO}(3) \quad (\text{Symmetry})$$

$$\rho(T(g \triangleright x, g \triangleright y)) = \rho(T(x, y)), \quad \forall x, y \in \text{SO}(3), \forall g \in \text{SO}(3) \quad (\text{Invariance})$$

$$\rho(T(x \triangleleft w_1, y \triangleleft w_2)) = \overline{\rho(w_1)} \rho(T(x, y)) \rho(w_2), \quad \forall x, y \in \text{SO}(3), \forall w_1, w_2 \in \text{SO}(2). \quad (\text{Equivariance})$$

We shall only assume the symmetry to be strictly satisfied by the empirical transport data; the other properties will be assumed to hold only approximately. To simplify notations, we denote for any $k \in \mathbb{Z}$

$$T^{(k)}(x, y) := \rho_k(T(x, y)), \quad \forall x, y \in \text{SO}(3) \quad (3.4)$$

where $\rho_k : \text{SO}(2) \rightarrow \mathbb{C}$ is the unique unitary irreducible representation of $\text{SO}(2)$ with character $k \in \mathbb{Z}$. The corresponding notation for the empirical transport data is $\tilde{T}^{(k)}(x_i, x_j)$.

In any of these irreducible representations, the empirical transport data $\{\tilde{T}^{(k)}(x_i, x_j) \mid 1 \leq i, j \leq N\}$ approximate the ground truth transport data $\{T^{(k)}(x_i, x_j) \mid 1 \leq i, j \leq N\}$ only when the viewing directions $\pi(x_i)$ and $\pi(x_j)$ are close to each other, in the sense that the vectors $\pi(x_i)$ and $\pi(x_j)$ belong to some small spherical cap of opening angle $\alpha \in [0, 2\pi)$.

3.2 Function on $\text{SO}(3)$ and isotypic decomposition

We will use the shorthand notation $\mathcal{H} = \mathbb{C}(\text{SO}(3))$ for the Hilbert space of smooth complex-valued functions on $\text{SO}(3)$, with standard Hermitian inner product

$$\langle f_1, f_2 \rangle_{\mathcal{H}} = \int_{\text{SO}(3)} f_1(x) \overline{f_2(x)} \, dx, \quad f_1, f_2 \in \mathcal{H}. \quad (3.5)$$

Here dx denotes the normalized Haar measure on $\text{SO}(3)$.

The left and right actions of the group elements induce corresponding actions on the Hilbert space $\mathcal{H}$ of complex-valued functions over $\text{SO}(3)$:

$$\begin{aligned} g \cdot s(x) &:= s(g^{-1} \triangleright x), \quad \forall f \in \mathcal{H}, x \in \text{SO}(3), g \in \text{SO}(3). \\ w \cdot s(x) &:= s(x \triangleleft w), \quad \forall s \in \mathcal{H}, x \in \text{SO}(3), w \in \text{SO}(2). \end{aligned} \quad (3.6)$$

The Hilbert space $\mathcal{H}$ can also be considered as a unitary representation of $\text{SO}(2)$. Let $\rho_k : \text{SO}(2) \rightarrow \mathbb{C}$ be the unique irreducible unitary representation of $\text{SO}(2)$ of character $k \in \mathbb{Z}$. $\mathcal{H}$ admits an isotypic decomposition

$$\mathcal{H} = \bigoplus_{k \in \mathbb{Z}} \mathcal{H}_k, \quad (3.7)$$

where

$$\mathcal{H}_k := \{s \in \mathcal{H} \mid s(x \triangleleft w) = \rho_k(w)s(x) \text{ for all } x \in \text{SO}(3) \text{ and } w \in \text{SO}(2)\}. \quad (3.8)$$

Note that $\text{SO}(3)$ acts on $\mathcal{H}_k$ unitarily from the left by

$$g \cdot s(x) := s(g^{-1} \triangleright x), \quad \forall g \in \text{SO}(3), s \in \mathcal{H}_k, x \in \text{SO}(3).$$

Each $\mathcal{H}_k$ thus admits an isotypic decomposition with respect to $\text{SO}(3)$, written as

$$\mathcal{H}_k = \bigoplus_{n \in \mathbb{N}_{\geq 0}} \mathcal{H}_{n,k} \quad (3.9)$$

where $\mathcal{H}_{n,k}$ denotes the isotypic component corresponding to the unique irreducible representation of $\text{SO}(3)$ of dimension $(2n + 1)$, for $n = 0, 1, \dots$. An important observation is that each $\mathcal{H}_{n,k}$ in (3.8) is of multiplicity 0 or 1 in $\mathcal{H}_k$:

THEOREM 3.1 [38, Theorem 7]. If $n < |k|$ then $\mathcal{H}_{n,k} = 0$. Otherwise, $\mathcal{H}_{n,k}$ is isomorphic to the unique irreducible representation of $\text{SO}(3)$ of dimension $(2n + 1)$.

4. Main theoretical results

4.1 Generalized parallel transport operators

The motivation for considering these isotypic decompositions is to study the top eigenspace of the *generalized parallel transport operator* $T^{(k)} : \mathcal{H} \rightarrow \mathcal{H}$, defined as

$$\left(T^{(k)}s\right)(x) := \int_{\text{SO}(3)} \rho_k(T(x, y)) s(y) \, dy = \int_{\text{SO}(3)} T^{(k)}(x, y) s(y) \, dy, \quad \forall s \in \mathcal{H}, x \in \text{SO}(3), \quad (4.1)$$

for all $k \in \mathbb{Z}$. When $k = 1$, $T^{(k)}$ reduces to the *parallel transport operator* $T : \mathcal{H} \rightarrow \mathcal{H}$ defined in [38, Section 2.3]. Similar to [38, Section 2.3.1], we can localize the generalized parallel transport operator $T^{(k)}$ for any $k \in \mathbb{Z}$ as

$$\left(T_h^{(k)}s\right)(x) := \int_{B(x, \alpha)} \rho_k(T(x, y)) s(y) \, dy = \int_{B(x, \alpha)} T^{(k)}(x, y) s(y) \, dy, \quad \forall s \in \mathcal{H}, x \in \text{SO}(3), \quad (4.2)$$

where $B(x, \alpha) = \{y \in \text{SO}(3) \mid (\pi(x), \pi(y)) > \cos \alpha =: 1 - h\}$. Using the symmetry, invariance and equivariance of the transport data (Section 3.1), we establish the following basic properties of $T^{(k)}$ for any $k \in \mathbb{Z}$:

- (1) $T^{(k)}$ is self-adjoint. This can be seen from the symmetry of transport data: for all $s, w \in \mathcal{H}$, we have

$$\begin{aligned} \left\langle T^{(k)}s, w \right\rangle_{\mathcal{H}} &= \int_{\text{SO}(3)} \int_{\text{SO}(3)} \rho_k(T(x, y)) s(y) \overline{w(x)} \, dy dx \\ &= \int_{\text{SO}(3)} \int_{\text{SO}(3)} s(y) \overline{\rho_k(T(y, x)) w(x)} \, dy dx = \left\langle s, T^{(k)}w \right\rangle_{\mathcal{H}}. \end{aligned}$$

- (2) $T^{(k)}$ commutes with the action of $\text{SO}(3)$ on $\mathcal{H}$: by the invariance of transport data we have for all $g \in \text{SO}(3)$ and $s \in \mathcal{H}$, $x \in \text{SO}(3)$,

$$\begin{aligned} \left(T^{(k)}(g \cdot s)\right)(x) &= \int_{\text{SO}(3)} \rho_k(T(x, y)) s(g^{-1} \triangleright y) \, dy \\ &\stackrel{z:=g^{-1} \triangleright y}{=} \int_{\text{SO}(3)} \rho_k\left(T\left(g \triangleright \left(g^{-1} \triangleright x\right), g \triangleright z\right)\right) s(z) \, dz \\ &= \int_{\text{SO}(3)} \rho_k\left(T\left(g^{-1} \triangleright x, z\right)\right) s(z) \, dz = \left(T^{(k)}s\right)(g^{-1} \triangleright x) = \left(g \cdot \left(T^{(k)}s\right)\right)(x). \end{aligned}$$

- (3) $\bigoplus_{\ell \neq -k} \mathcal{H}_\ell \subset \ker T^{(k)}$, and $T^{(k)}$ can be viewed as an operator from $\mathcal{H}_{-k}$ to itself. This can be verified using the equivariance of $T^{(k)}$. First, note that for any $s \in \mathcal{H}$ we have $T^{(k)}s \in \mathcal{H}_{-k}$, since for any $w \in \text{SO}(2)$ we have

$$\begin{aligned} w \cdot \left(T^{(k)}s\right)(x) &= \left(T^{(k)}s\right)(x \triangleleft w) \\ &= \int_{B(x, \alpha)} \rho_k(T(x \triangleleft w, y)) s(y) \, dy = \int_{B(x, \alpha)} \overline{\rho_k(w)} \rho_k(T(x, y)) s(y) \, dy \\ &= \rho_{-k}(w) \int_{B(x, \alpha)} \rho_k(T(x, y)) s(y) \, dy = \rho_{-k}(w) \left(T^{(k)}s\right)(x). \end{aligned}$$

This proves that $T^{(k)}$ maps $\mathcal{H}$ into $\mathcal{H}_{-k}$, by the definition of isotypic decomposition (3.8) with respect to the $\text{SO}(2)$ action. The conclusion that $\bigoplus_{\ell \neq -k} \mathcal{H}_\ell \subset \ker T^{(k)}$ then follows from Schur's lemma [11, Theorem 2.1].

The arguments above can be applied to $T_h^{(k)}$, *mutatis mutandis*, and thus the same properties hold for the local generalized parallel transport operator. Invoking Schur's lemma for a second time, we know that $T_h^{(k)}$ acts on $\mathcal{H}_{n, -k}$ as a scalar, i.e.,

$$T_h^{(k)}|_{\mathcal{H}_{n, -k}} = \lambda_n^{(k)}(h) \text{Id}|_{\mathcal{H}_{n, -k}}. \quad (4.3)$$

The multiplicity-one theorem (Theorem 3.1) tells us that $\lambda_n^{(k)} = 0$ for all $0 \leq n < |k|$. In order to calculate the remaining $\lambda_n^{(k)}$'s ($n \geq |k|$) explicitly, it suffices to fix a point $x_0 \in \text{SO}(3)$, and pick an arbitrary function $u \in \mathcal{H}_{n, -k}$ with $u(x_0)$, and use relation $\lambda_n^{(k)} = \left(T_h^{(k)}u\right)(x_0)/u(x_0)$. We will defer such computations for $0 < h \ll 1$ to Section 4.2. Next subsection summarizes these properties, in preparation for the discussion on the main algebraic structure of the *generalized intrinsic model* in Section 4.3.

In [38, 63], it was argued that the Hermitian matrix H in (1.4) should be understood as the discretization (under uniform random sampling on $\text{SO}(3)$) of an integral operator $T_h^{(1)}$. Consequently, many properties of the local transport data matrix H can be studied through its 'continuous limit' T_h , especially the eigenvalues and eigenvectors, which converge to the eigenvalues and eigenfunctions of T_h in an appropriate sense [44]; this perspective is common in the manifold learning literature

[5, 6, 15, 30, 62]. In the class averaging setting, the integral operator T_h enjoys many useful invariance and equivariance properties, which makes it relatively straightforward to study its spectral data using representation theoretic tools. Hadani and Singer noticed that T_h acts on the subspace $\mathcal{H}_{-1}$ of $\mathcal{H}$. The space $\mathcal{H}_{-1}$ is also canonically identified with the linear space of sections of a complex line bundle over $\text{SO}(3)$ induced by the unitary irreducible representation of $\text{U}(1)$ with character $k = 1$ [10, 12, 20, 34, 49, 50]. Furthermore, T_h commutes with the induced left action of $\text{SO}(3)$ on $\mathcal{H}_{-1}$, which by Schur's theorem indicates that the eigenspaces of T_h coincides with the isotypic components of $\mathcal{H}_{-1}$ under the left $\text{SO}(3)$ action. In particular, this mechanism can be used to show that the top eigenspace of T_h is the unique isotypic component of $\mathcal{H}_{-1}$ corresponding to the unique three-dimensional unitary irreducible representation of $\text{SO}(3)$ for all sufficiently small $h > 0$ and that the affinity measure $2A_{ij} - 1$ is exactly identical with the cosine value of the viewing angle between I_i and I_j in the noise-free setting.

4.2 Spectral properties of the local parallel transport operator

In this subsection, we summarize the spectral properties of $T_h^{(k)}$ for $h \ll 1$ (which is the relevant regime for class averaging). Proofs for the main theorems discussed in this subsection are deferred to B. These proofs essentially follow the proof ideas of [38, Theroem 3 and Theorem 4], with technical modification due to the complication of Jacobi polynomials—unlike the case for the Legendre polynomials involved in the analysis of single-frequency class averaging, no sharp Bernstein-type inequality is known for Jacobi polynomials arising from the Wigner d -matrices. We refer interested readers to discussions and conjectures in [14, 36, 45] for Bernstein-type inequalities for Jacobi polynomials.

THEOREM 4.1 (Eigenvalues of $T_h^{(k)}$ for small $h \ll 1$). The operator $T_h^{(k)}$ has a discrete spectrum $\lambda_n^{(k)}(h)$ for all $n \in \mathbb{N}$, and $\lambda_n^{(k)} = 0$ for all $0 \leq n < |k|$. For $n \geq |k|$ and $h \in (0, 2]$, the dimension of the eigenspace of $T_h^{(k)}$ corresponding to $\lambda_n^{(k)}$ is $2n + 1$. In addition, $\lambda_k^{(k)}$ and $\lambda_{k+1}^{(k)}$ have the following expressions:

$$\lambda_k^{(k)}(h) = \frac{1 - (1 - h/2)^{k+1}}{k + 1}, \quad (4.4)$$

$$\lambda_{k+1}^{(k)}(h) = \frac{2(k+1)(1 - (1 - h/2)^{k+2})}{k + 2} - \frac{(2k+1)(1 - (1 - h/2)^{k+1})}{k + 1}. \quad (4.5)$$

In the regime $h \ll 1$, the eigenvalue $\lambda_n^{(k)}(h)$ ($n \geq |k|$) adopts asymptotic expansion

$$\lambda_n^{(k)}(h) = \frac{1}{2}h - \frac{1}{8}(n^2 + n - k^2)h^2 + O(h^3). \quad (4.6)$$

REMARK 4.1 When $k = 1$, Theorem 4.1 reduces to [38, Theorem 3].

The proof of Theorem 4.1 in B.1 actually proves the stronger conclusion that each eigenvalue $\lambda_n^{(k)}(h)$ is a polynomial in $h > 0$ of degree $(n + 1)$ whenever $n \geq |k|$. The key step in the proof is identifying that the $(-k, -k)$ entry of the Wigner D -matrix $D_{-k, -k}^n(x) \in \mathcal{H}_{n, k}$ for $n \geq |k|$ and is an appropriate function u for calculating the eigenvalues. The largest three eigenvalues for cases $k = 1$ and $k = 2$ can

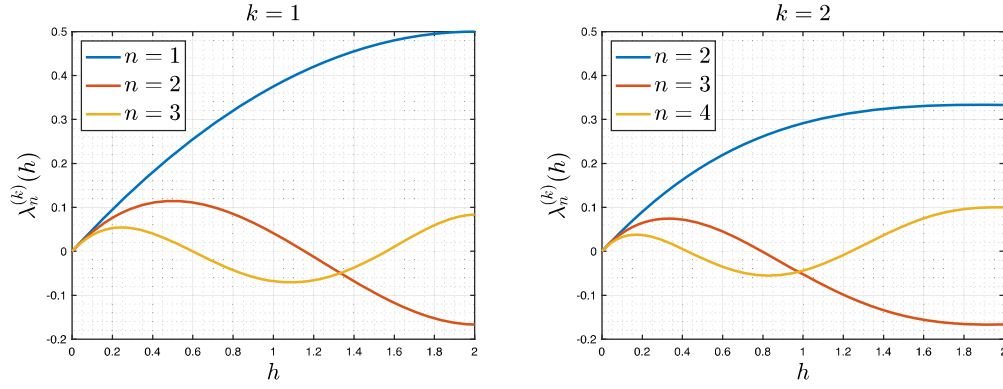


FIG. 1. The top three eigenvalues $\lambda_n^{(k)}(h)$ of operator $T_h^{(k)}$, for $k = 1$ (left) and $k = 2$ (right) over interval $h \in (0, 2]$.

be explicitly written out as

$$\begin{aligned}
 \lambda_1^{(1)}(h) &= \frac{1}{2}h - \frac{1}{8}h^2, \\
 \lambda_2^{(1)}(h) &= \frac{1}{2}h - \frac{5}{8}h^2 + \frac{1}{6}h^3, \\
 \lambda_3^{(1)}(h) &= \frac{1}{2}h - \frac{11}{8}h^2 + \frac{25}{24}h^3 - \frac{15}{64}h^4,
 \end{aligned} \tag{4.7}$$

and

$$\begin{aligned}
 \lambda_2^{(2)}(h) &= \frac{1}{2}h - \frac{1}{4}h^2 + \frac{1}{24}h^3, \\
 \lambda_3^{(2)}(h) &= \frac{1}{2}h - h^2 + \frac{13}{24}h^3 - \frac{3}{32}h^4, \\
 \lambda_4^{(2)}(h) &= \frac{1}{2}h - 2h^2 + \frac{57}{24}h^3 - \frac{70}{64}h^4 + \frac{7}{40}h^5.
 \end{aligned} \tag{4.8}$$

Plots of $\lambda_{k+i}^{(k)}$, for $i = 0, 1, 2$ are provided in Fig. 1.

COROLLARY 4.1 As k increases, the eigenvalue $\lambda_k^{(k)}$ decreases and $\lim_{k \rightarrow \infty} \lambda_k^{(k)} = 0$.

Proof. of Corollary 4.1 Based on Theorem 4.1, the difference between $\lambda_{k+1}^{(k+1)}$ and $\lambda_k^{(k)}$ for $k \geq 1$ is

$$\begin{aligned}
 \lambda_{k+1}^{(k+1)} - \lambda_k^{(k)} &= \frac{1 - (1 - h/2)^{k+2}}{k+2} - \frac{1 - (1 - h/2)^{k+1}}{k+1} = \frac{-1 + (1 - h/2)^{k+1}(1 + \frac{h}{2}(k+1))}{(k+1)(k+2)} \\
 &\stackrel{(a)}{<} \frac{-1 + (1 - h/2)^{k+1}(1 + h/2)^{k+1}}{(k+1)(k+2)} = \frac{-1 + (1 - h^2/4)^{k+1}}{(k+1)(k+2)} < 0,
 \end{aligned} \tag{4.9}$$

where (a) is based on the fact that $(1 + \frac{h}{2})^{k+1} > 1 + (k+1)\frac{h}{2}$ for $h \in (0, 2]$ via Taylor expansion. In addition, since $0 \leq 1 - h/2 < 1$, $\lim_{k \rightarrow \infty} \lambda_k^{(k)}(h) = \lim_{k \rightarrow \infty} \frac{1 - (1 - h/2)^{k+1}}{k+1} = 0$. $\square$

This is an important observation for determining the maximum frequency cutoff, which will be further discussed in Section 6.

It is natural to conjecture that the top eigenspace of $T_h^{(k)}$ is the $(2k+1)$ -dimensional space corresponding to eigenvalue $\lambda_k^{(k)}(h)$ for sufficiently small $h > 0$. Moreover, denote

$$\Delta_k := \arg \max_{h \in (0, 2]} \lambda_{k+1}^{(k)}(h) = \frac{1}{k+1}, \quad (4.10)$$

we have the following characterization of the *spectral gap* for $T_h^{(k)}$ in the regime $0 < h \ll 1$.

THEOREM 4.2 For every value of $h \in (0, 2]$, the largest eigenvalue of $T_h^{(k)}$ is $\lambda_k^{(k)}(h)$. In addition, for every value of $h \in (0, \Delta_k]$, the spectral gap $G^{(k)}(h)$ between the largest and the second largest eigenvalue of $T_h^{(k)}$ is

$$G^{(k)}(h) = \lambda_k^{(k)} - \lambda_{k+1}^{(k)} = \frac{2 - (1 - h/2)^{k+1}((k+1)h + 2)}{k+2}. \quad (4.11)$$

Again, when $k = 1$, Theorem 4.2 reduces to [38, Theorem 4]. The main technicality of the proof of Theorem 4.2, which is deferred to B.2, is to show that $\lambda_n^{(k)}(h) \leq \lambda_{k+1}^{(k)}(h)$ for every $h \in (0, \Delta_k]$ and $n \geq k+1$, which appears evident from Fig. 1. For small $0 < h \ll \Delta_k$, the spectral gap is approximately

$$G^{(k)}(h) \sim \frac{1+k}{4}h^2, \quad (4.12)$$

which gets larger as the ‘angular frequency’ $k \in \mathbb{N}$ increases. More generally, we have the following corollary.

COROLLARY 4.2 The spectral gap $G^{(k)}(h)$ increases as k increases from 1 to $k_{\max} = \left\lfloor \frac{1}{h} \right\rfloor - 1$.

Proof of Corollary 4.2. We show that for any $k \geq 2$, the difference $G^{(k)}(h) - G^{(k-1)}(h)$ is always positive for any $h \in (0, \Delta_k]$. To begin with, we explicitly write out the difference as

$$\begin{aligned} G^{(k)}(h) - G^{(k-1)}(h) &= \frac{2 - (1 - h/2)^{k+1}((k+1)h + 2)}{k+2} - \frac{2 - (1 - h/2)^k(kh + 2)}{k+1} \\ &= \frac{(1 - h/2)^k((k+1)^2h^2 + 2kh + 4) - 4}{2(k+1)(k+2)} \\ &= \frac{(1 - h/2)^k}{2(k+1)(k+2)} \underbrace{\left((k+1)^2h^2 + 2kh + 4 - 4(1 - h/2)^{-k} \right)}_{=:\xi(h)} \\ &= \frac{(1 - h/2)^k}{2(k+1)(k+2)} \xi(h), \end{aligned}$$

where $\xi(h)$ is defined as a function of h . Since the term in front of $\xi(h)$ is always positive for $h \in (0, \Delta_k]$, it suffices to show $\xi(h) > 0$ for any $k \geq 2$ and $h \in (0, \Delta_k]$. To this end, clearly $\xi(h) = 0$ when $h = 0$ then we can instead show the derivative of $\xi(h)$ is positive for any $h \in (0, \Delta_k]$. That is,

$$\frac{d\xi(h)}{dh} = 2(k+1)^2h + 2k - \frac{2k}{(1-h/2)^{k+1}}.$$

Again, when $h = 0$ we observe that $\frac{d\xi(h)}{dh}|_{h=0} = 0$. So in order to show $\frac{d\xi(h)}{dh} > 0$, for all $h \in (0, \Delta_k]$ we can instead check if the second-order derivative of $\xi(h)$ is positive for any $h \in (0, \Delta_k]$. Indeed, we have

$$\begin{aligned} \frac{d^2\xi(h)}{dh^2} &= 2(k+1)^2 - \frac{k(k+1)}{(1-h/2)^{k+2}} = \frac{(k+1)}{(1-h/2)^{k+2}} (2(k+1)(1-h/2)^{k+2} - k) \\ &\stackrel{(a)}{>} \frac{(k+1)}{(1-h/2)^{k+2}} (2(k+1)(1-(k+2)h/2) - k) \stackrel{(b)}{\geq} 0 \end{aligned}$$

where (a) comes from the inequality that $(1-x)^a > 1-xa$ for any $x \in (0, 1)$ and $a > 2$, (b) is satisfied since $2(k+1)(1-(k+2)h/2) - k$ is linear and monotonically decreasing for h and the equality only holds when $h = \Delta_k = \frac{1}{k+1}$. Therefore, we obtain that $\frac{d^2\xi(h)}{dh^2} > 0$, $\forall h \in [0, \Delta_k]$ and it follows that $\frac{d\xi(h)}{dh} > 0$, $\forall h \in (0, \Delta_k]$; furthermore, we can conclude that $G^{(k)}(h) - G^{(k-1)}(h) > 0$ for any $h \in (0, \Delta_k]$. $\square$

This justifies one benefit of setting $k > 1$ for class averaging, as larger spectral gaps provide more robustness to noise corruption for k satisfying $k < \frac{1}{h} - 1$. More detailed discussion on the performance of the algorithm under noise perturbation is in Section 6. In practice, the choice of frequency cutoff depends on the neighborhood size, noise type and noise level and may need to be empirically identified.

4.3 The main algebraic structure: generalized intrinsic model

Just as the intrinsic model established in [38] equates the ‘extrinsic model’ S^2 with the ‘intrinsic model’ of the top eigenspace $\mathbb{W}$ of $T = T^{(1)}$, we will generalize this correspondence to the setting for general complex irreducible unitary representations of $\text{SO}(2)$. More specifically, we establish the correspondence between the following two generalized models:

- *Generalized extrinsic model*: for every point $x = x(\varphi, \vartheta, \psi) \in \text{SO}(3)$, denote by $\delta_x^{(k)} : \mathbb{C} \rightarrow \mathbb{C}^{2k+1}$ for the unique complex morphism sending $1 \in \mathbb{C}$ to the first (index- $(-k)$) column of the Wigner D -matrix D^k (detailed in A), i.e.,

$$D_{\cdot, -k}^k(x) = \left(D_{-k, -k}^k(x), D_{-k+1, -k}^k(x), \dots, D_{k-1, -k}^k(x), D_{k, -k}^k(x) \right)^\top \in \mathbb{C}^{2k+1}.$$

- *Generalized intrinsic model*: define $\mathbb{W}^{(k)}$ as the top eigenspace of $T_h^{(k)}$, which by Theorem 4.1 and 4.2, is $(2k+1)$ -dimensional. Set for every point $x \in \text{SO}(3)$ the map

$$\varphi_x^{(k)} = \sqrt{1/(2k+1)} \cdot (\text{ev}_x | \mathbb{W}^{(k)})^* : \mathbb{C} \rightarrow \mathbb{W}^{(k)}, \quad (4.13)$$

where $\text{ev}_x : \mathcal{H} \rightarrow \mathbb{C}$ is the evaluation morphism at the point $x \in \text{SO}(3)$.

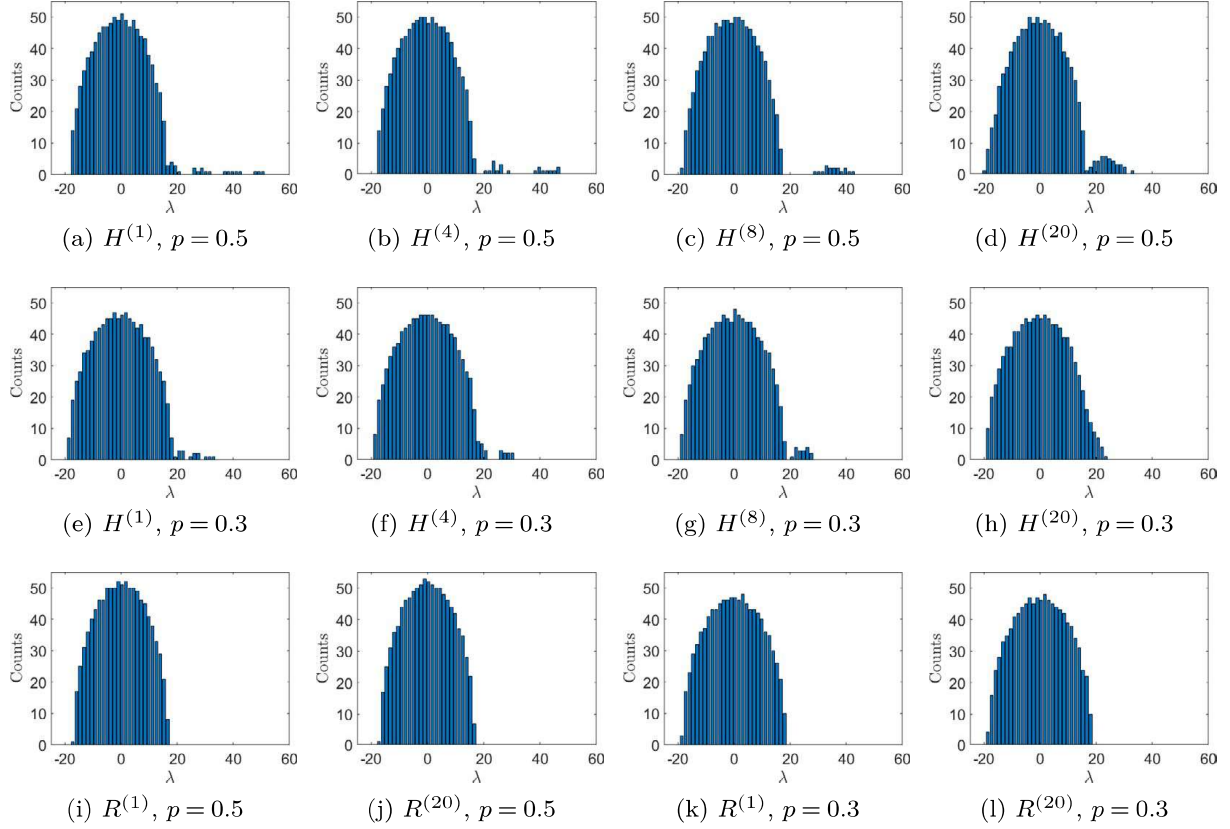


FIG. 2. Histograms of the eigenvalues of $H^{(k)}$ and $R^{(k)}$ in (6.4) for data generated from random rewiring model with $N = 1000$, $p = 0.5$ and $p = 0.3$.

The main algebraic structure of the multi-frequency intrinsic classification algorithm is summarized in the following main theorem of this section.

THEOREM 4.3 The morphism $\tau : \mathbb{C}^{2k+1} \rightarrow \mathcal{H}$ defined by

$$\begin{aligned} \tau : \mathbb{C}^{2k+1} &\longrightarrow \mathcal{H} \\ v &\longmapsto \left(x \mapsto \sqrt{2k+1} \cdot \left(\delta_x^{(k)} \right)^* (v) \right) \end{aligned}$$

is an isomorphism between $\mathbb{C}^{2k+1}$ and $\mathbb{W}^{(k)} \subset \mathcal{H}$ (as Hermitian vector spaces). Moreover, for every $x \in \text{SO}(3)$ and $k = 0, 1, \dots$ there holds

$$\tau \circ \delta_x^{(k)} = \varphi_x^{(k)}. \quad (4.14)$$

The proof of Theorem 4.3 is deferred to B.3. Our proof extends the arguments in the proof of [38, Theorem 5). A key observation is that the top eigenvector $\mathcal{H} \left(\lambda_k^{(k)}(h) \right)$ coincides with the isotypic subspace $\mathcal{H}_{k,-k}$ (see Section 4.1). Furthermore, Theorem 4.3 reveals the correspondence between the generalized extrinsic and intrinsic models, in terms of the viewing angle information they encode. This is summarized in the following result.

THEOREM 4.4 For every pair of frames $x, y \in \text{SO}(3)$, we have

$$\left| \langle \varphi_x^{(k)}(v), \varphi_y^{(k)}(u) \rangle_{\mathbb{W}^{(k)}} \right| = \left(\frac{\langle \pi(x), \pi(y) \rangle + 1}{2} \right)^k \quad (4.15)$$

for any choice of unit-norm complex numbers $v, u \in \mathbb{C}$.

The proof of Theorem 4.4 is deferred to B.4.

REMARK 4.2 When $k = 1$, Theorem 4.3 and Theorem 4.4 reduce to [38, Theorem 5] and [38, Theorem 6], respectively, up to a different scaling constant for τ . The difference arises from our alternative, explicit construction of the isomorphism τ using Wigner D -matrices.

5. Interpretation of the theoretical results for multi-frequency class averaging

In this section, we interpret the MFCA algorithm stated in Section 2 using the theoretical results established in Section 4 and provide conceptual explanations for the admissibility of MFCA in the noiseless regime.

First, under the assumption that the projection images $\{I_i \mid 1 \leq i \leq N\}$ are produced from orthonormal frames $\{x_i \mid 1 \leq i \leq N\}$ sampled i.i.d. uniformly on $\text{SO}(3)$ with respect to the normalized Haar measure, we view $\frac{1}{N}H^{(k)}$, the scaled class averaging matrix at frequency k defined in Section 2, as the discretization of the local parallel transport operator $T_h^{(k)}$. We know from standard results [44, Theorem 3.1] that the eigenvalues of $\frac{1}{N}H^{(k)}$ converges to the eigenvalues of the generalized localized parallel transport operator $T_h^{(k)}$ defined in (4.2) as the number of samples N goes to infinity and the opening angle α is sufficiently small. In particular, this implies that for large sample size N , the spectral gap of $\frac{1}{N}H^{(k)}$ converges to the spectral gap of $T_h^{(k)}$, which, by Theorem 4.1 and Theorem 4.2, is roughly of size $(1+k)h^2/4$ for $h \ll \frac{1}{k+1}$ and occurs between the $(2k+1)^{\text{th}}$ and the $(2k+2)^{\text{th}}$ eigenvalues of $H^{(k)}$ (ranked in decreasing order).

Moreover, as argued in [38, Theorem 2], the MFCA embedding $\Psi^{(k)}$ defined in (2.2) corresponds to the morphism (4.13) in the following form:

$$\frac{\Psi^{(k)}(x_i)}{\|\Psi^{(k)}(x_i)\|} \approx \varphi_{x_i}^{(k)}(1), \quad \text{for all } x_i \in \text{SO}(3), \quad (5.1)$$

where $\|\cdot\|$ stands for the standard norm on $\mathbb{C}^{2k+1}$. Combining (5.1) with Theorem 4.4 provides the justification for using $A^{(k)}$ to identify similar viewing angles,

$$A_{ij}^{(k)} = \frac{\left| \langle \Psi^{(k)}(I_i), \Psi^{(k)}(I_j) \rangle \right|}{\|\Psi^{(k)}(I_i)\| \|\Psi^{(k)}(I_j)\|} \approx \left| \langle \varphi_x^{(k)}(1), \varphi_y^{(k)}(1) \rangle_{\mathbb{W}^{(k)}} \right| = \left(\frac{\langle \pi(x_i), \pi(x_j) \rangle + 1}{2} \right)^k. \quad (5.2)$$

This relation is demonstrated in the top rows of Figs 5 and 13. In fact, Theorem 4.4 tells us that the affinity measure $S_{ij}^{(k)}$ defined in (2.5) coincides with the cosine value for the angle between the two viewing directions in the noiseless regime. The form of the approximation identity (5.2) also suggests

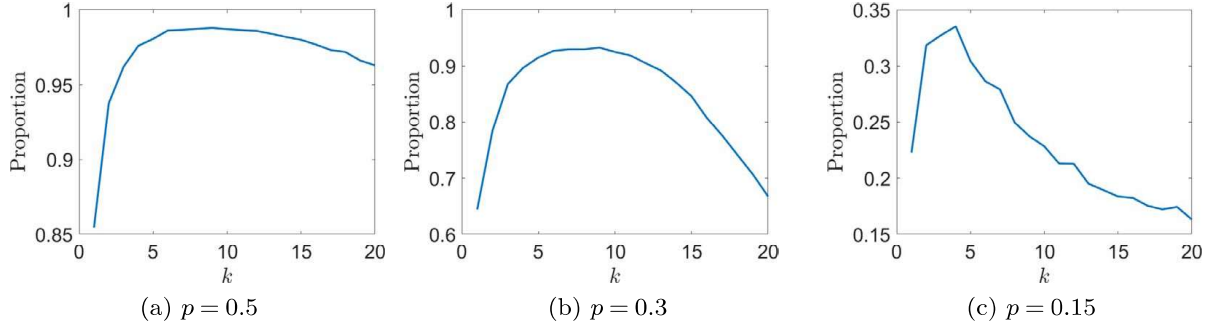


FIG. 3. Proportion of the estimated nearest neighbors that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.85$ for $p = 0.5, 0.3$ and 0.15 . The number of frames $N = 1000$ and the number of nearest neighbors is 50.

avoiding directly taking the k th root of the correlation between $\Psi^{(k)}(I_i)$ and $\Psi^{(k)}(I_j)$ as in (2.5) and (2.6) since this approach loses control of the numerical relative error when $A_{ij}^{(k)}$ is close to 0. In contrast, it is advantageous to use the multiplicative forms (2.3) and (2.4) which do not worsen the relative error. The logarithm of the combined affinity A^{All} has the following relation with the viewing angles,

$$\log(A_{ij}^{\text{All}}) = \sum_{k=1}^{k_{\max}} \log(A_{ij}^{(k)}) \approx \frac{k_{\max}(k_{\max} + 1)}{2} \log\left(\frac{\langle \pi(x_i), \pi(x_j) \rangle + 1}{2}\right). \quad (5.3)$$

Using A^{All} or $\log(A^{\text{All}})$ makes small viewing angles much more prominent in the numerical procedures. One may well expect other linear combinations of the $A_{ij}^{(k)}$ s, which are degree- $k_{\max}$ polynomials of the (cosine value of the) viewing angle. We leave these further explorations to future work.

6. Analysis under probabilistic models

In this section, we discuss the benefit of using $A^{(k)}$ with $k > 1$ to identify nearest neighbors when the measurement graph is perturbed by noise. To this end, we use the random rewiring model [63] for the entries of $H^{(1)}$ in Section 6.1 and extend it to incorporate small angular perturbation in Section 6.2. We start by randomly generating N orthonormal frames $x_1, x_2, \dots, x_N$ uniformly sampled from $\text{SO}(3)$ according to the Haar measure. Each frame x_i can be represented by a 3×3 orthogonal matrix $R_i = [R_i^1, R_i^2, R_i^3]$ and $\det(R_i) = 1$. We identify the third column R_i^3 as the *viewing angle* $\pi(x_i)$ of the molecule. The first two columns R_i^1 and R_i^2 form an orthonormal basis for the plane in $\mathbb{R}^3$ perpendicular to the viewing angle $\pi(x_i)$. If the viewing angles for two projection images belong to a small spherical cap with opening angle α , then we connect the two points in the graph (i.e. $(i, j) \in E$ if $\langle \pi(x_i), \pi(x_j) \rangle > \cos \alpha$). If x_i and x_j are two frames with the same viewing angle, $\pi(x_i) = \pi(x_j)$, then R_i^1, R_i^2 and R_j^1, R_j^2 are two orthogonal bases for the same plane and the rotation matrix $R_i^{-1}R_j$ has the following form:

$$R_i^{-1}R_j = \begin{pmatrix} \cos \theta_{ij} & -\sin \theta_{ij} & 0 \\ \sin \theta_{ij} & \cos \theta_{ij} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.1)$$

When the viewing angles are slightly different, (6.1) holds approximately. The optimal in-plane rotational angle θ_{ij} provides a good approximation to the angle θ_{ij} that ‘aligns’ the orthonormal bases for the planes $\pi(x_i)^\perp$ and $\pi(x_j)^\perp$. Therefore, if $\langle \pi(x_i), \pi(x_j) \rangle$ is close to 1, the angle θ_{ij} is given by

$$\theta_{ij} = \arg \min_{\theta \in [0, 2\pi)} \|R_i \rho(\theta) - R_j\|_F, \quad \text{with } \rho(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.2)$$

In other words, the ground truth local parallel transport data are computed by aligning the local frames within the connected neighborhood, determined by the entries of the matrix $R_i^{-1}R_j$:

$$\begin{aligned} \cos \theta_{ij} &= \frac{(R_i^{-1}R_j)_{11} + (R_i^{-1}R_j)_{22}}{\sqrt{[(R_i^{-1}R_j)_{11} + (R_i^{-1}R_j)_{22}]^2 + [(R_i^{-1}R_j)_{21} - (R_i^{-1}R_j)_{12}]^2}}, \\ \sin \theta_{ij} &= \frac{(R_i^{-1}R_j)_{21} - (R_i^{-1}R_j)_{12}}{\sqrt{[(R_i^{-1}R_j)_{11} + (R_i^{-1}R_j)_{22}]^2 + [(R_i^{-1}R_j)_{21} - (R_i^{-1}R_j)_{12}]^2}}. \end{aligned} \quad (6.3)$$

6.1 Random rewiring model

Starting from the clean neighborhood graph constructed above, we perturb the graph based on the following process: with probability p , we keep the clean edge and the associated transport data θ_{ij} , and with probability $1 - p$, we remove the edge (i, j) and randomly rewire i or j with a vertex drawn uniformly at random from the remaining vertices that are not already connected to i or j . We assume that if the link between i and j is a random link, then $\theta_{ij} = \phi_{ij}$, which is uniformly distributed over $[0, 2\pi)$. Our model assumes that the underlying graph of links between noisy data points is a small-world graph [68] on the sphere, with edges being randomly rewired with probability $1 - p$. The alignments take their correct values for true links and random values for the rewired edges. The parameter p controls the signal to noise ratio of the graph connection where $p = 1$ indicates the clean graph.

The matrix $H^{(k)}$ is a random matrix under this model with

$$H_{ij}^{(k)} = \begin{cases} e^{tk\theta_{ij}}, & \text{if } (i, j) \in E \text{ and with probability } p, \\ e^{tk\phi_{ij}}, & \text{if } (i, j) \notin E \text{ and with probability } \frac{(1-p)\bar{D}}{N-\bar{D}}. \end{cases} \quad (6.4)$$

Since the expected value of the random variable $e^{tk\phi}$ vanishes for $\phi \sim \text{Uniform}[0, 2\pi)$, the expected value of the matrix $H^{(k)}$ is

$$\mathbb{E}H^{(k)} = pH_{\text{clean}}^{(k)}, \quad (6.5)$$

where $H_{\text{clean}}^{(k)}$ is the clean matrix that corresponds to $p = 1$ obtained in the case that all links and angles are set up correctly. At each frequency k , the matrix $H^{(k)}$ can be decomposed into

$$H^{(k)} = pH_{\text{clean}}^{(k)} + R^{(k)}, \quad (6.6)$$

where $R^{(k)}$ is a random matrix whose elements are

$$R_{ij}^{(k)} = \begin{cases} (1-p)e^{ik\theta_{ij}}, & \text{if } (i,j) \in E \text{ and with probability } p, \\ -pe^{ik\theta_{ij}}, & \text{if } (i,j) \in E \text{ and with probability } 1-p, \\ e^{ik\phi_{ij}}, & \text{if } (i,j) \notin E \text{ and with probability } \frac{(1-p)\bar{D}}{N-\bar{D}}, \end{cases} \quad (6.7)$$

and $\bar{D}$ is the average degree of the clean neighborhood graph. The elements in $R^{(k)}$ are independent zero mean random variables with finite moments, since the elements of $R^{(k)}$ are bounded for $1 \leq k \leq k_{\max}$.

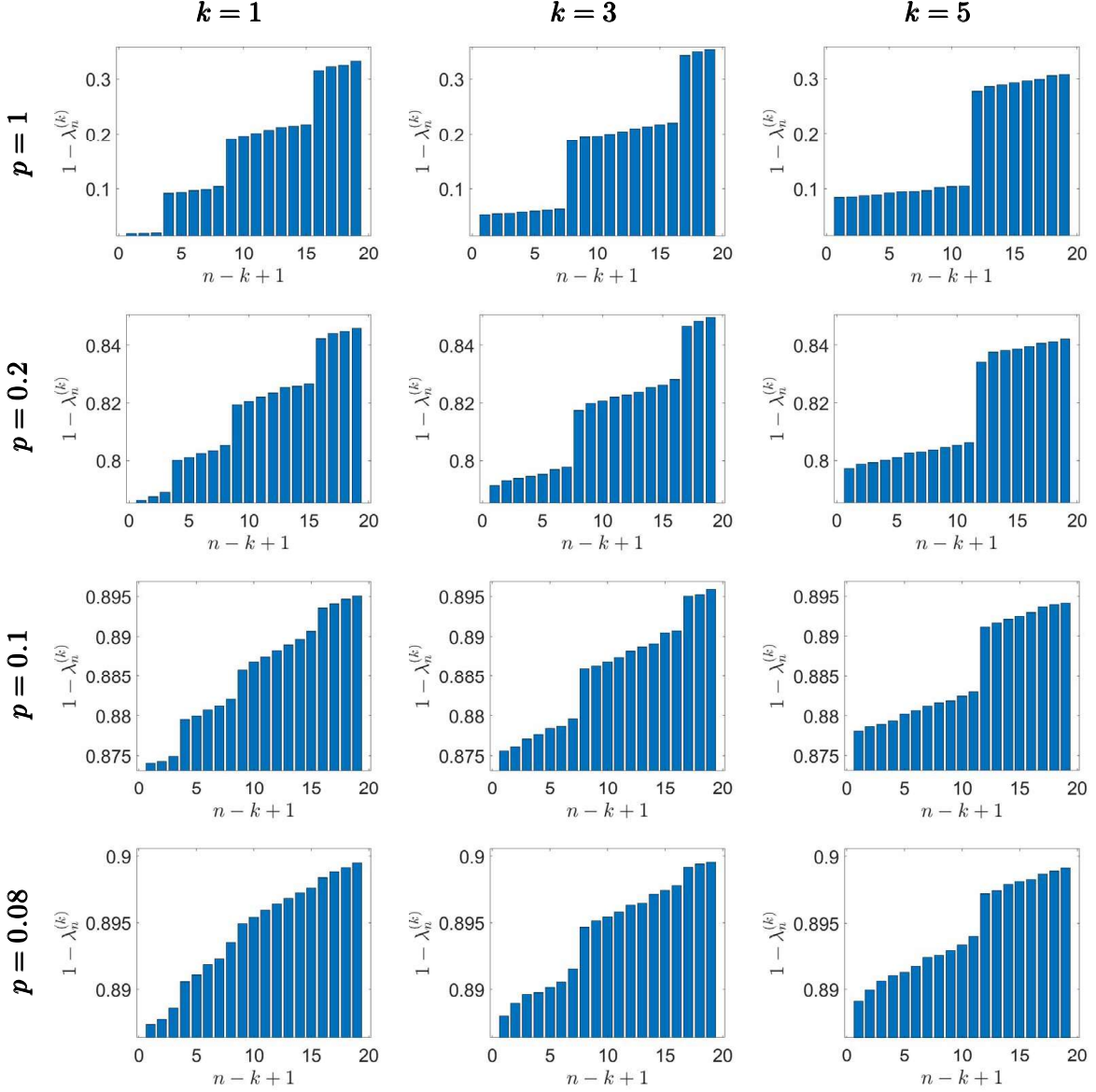
We use $\|M\|$ to denote the spectral norm of a matrix M . Since the underlying graph connectivity for all $R^{(k)}$ is identical and the mean and variance of $R_{ij}^{(k)}$ are identical across k , the quantity $\|R^{(k)}\|$ does not change over frequency index k . To find an upper bound on $\|R^{(k)}\|$, we take $p = 0$, where the matrix $R^{(k)}$ represents a sparse random graph. Since the surface area of a spherical cap with opening angle α is $4\pi \sin^2 \frac{\alpha}{2}$ and N points are uniformly distributed over the sphere, the average degree of the random graph is $N \sin^2 \frac{\alpha}{2}$. Adapting [42, Theorem 2.1] to our case, we can show that $\|R^{(k)}\| \leq 2\sqrt{N} \sin \frac{\alpha}{2}$ with high probability. In Fig. 2, we can see that the eigenvalues of $R^{(k)}$ follows Wigner's semi-circle law [69, 70].

For the following discussion, we denote $k_{\max} = \lfloor \frac{1}{h} - 1 \rfloor$. The ordered eigenvalues for $pH_{\text{clean}}^{(k)}$ are $\ell_1^{(k)} \geq \ell_2^{(k)} \geq \dots \geq \ell_N^{(k)}$, and the ordered eigenvalues for $H^{(k)}$ are $\tilde{\ell}_1^{(k)} \geq \tilde{\ell}_2^{(k)} \geq \dots \geq \tilde{\ell}_{2k+2}^{(k)} \geq \dots \geq \tilde{\ell}_N^{(k)}$, for $k = 1, \dots, k_{\max}$. The spectral gap after the $(2k+1)^{\text{th}}$ eigenvalue for $pH_{\text{clean}}^{(k)}$ is denoted as $\delta_k = \ell_{2k+1}^{(k)} - \ell_{2k+2}^{(k)}$. We note that $\{\ell_i^{(k)}\}_{i=1}^{2k+1} \approx pN\lambda_k^{(k)}$ and $\{\ell_i^{(k)}\}_{i=2k+2}^{4k+3} \approx pN\lambda_{k+2}^{(k)}$, and $\delta_k \approx pNG^{(k)}$, since $\frac{1}{N}H_{\text{clean}}^{(k)}$ is a discretization of the operator $T_h^{(k)}$. We consider the following three scenarios for the discussion of the stability of the algorithm under noise perturbation: (1) small noise regime ($\delta_1 \geq 2\|R^{(1)}\|$), (2) medium noise regime ($\delta_1 < 2\|R^{(1)}\| \leq \delta_{k_{\max}}$) and (3) large noise regime ($\delta_{k_{\max}} < 2\|R^{(1)}\|$).

- *Small noise regime.* This noise regime was previously considered in [63] to determine the threshold probability p_c for the approximation of the top three eigenvectors of $H^{(1)}$ and the top three eigenvectors of $H_{\text{clean}}^{(1)}$ under the random rewiring model. According to Corollary 4.2, the spectral gap gets larger for higher frequency index k . This implies that the linear space spanned by the first $(2k+1)$ eigenvectors of $H^{(k)}$ is closer to the top eigenspace of $T_h^{(k)}$, since the approximation error is inversely proportional to the spectral gap according to the renowned Davis–Kahan theorem [17, 72]. This also explains the choice of extracting the top $(2k+1)$ eigenvectors of $H^{(k)}$ in single frequency- k class averaging.

- *Medium noise regime.* In this situation, we can find a $\tilde{k}$ such that for all $\tilde{k} \leq k \leq k_{\max}$, $\delta_k > 2\|R^{(1)}\| = 2\|R^{(k)}\|$. In addition, we can show that $\ell_{4k+3}^{(k)} > \|R^{(k)}\|$ for $k = 1, \dots, k_{\max}$. This is because we have $\lambda_{k+1}^{(k)} > G^{(k)}$ for $k \leq k_{\max}$ and $\lambda_{k+1}^{(k)}$ decreases as k increases according to Theorem 4.1 and Theorem 4.2. Using the same argument as in the small noise regime, we can justify the benefit of using the top $(2k+1)$ eigenvectors of $H^{(k)}$ at $k > 1$.

- *Large noise regime.* If we further decrease p , the spectral norm of $R^{(k)}$ becomes larger than the spectral gap $\delta_{k_{\max}}$. According to Davis–Kahan theorem, it seems impossible to recover the eigenvectors


 FIG. 4. Bar plots of the 19 largest eigenvalues of the $\tilde{H}^{(k)}$ at different k and p values.

if the eigenvalue perturbation is too large. However, we observe that under this situation, the subspace spanned by the top $2k + 1$ eigenvectors of $H^{(k)}$ still has non-trivial correlation with the subspace spanned by the top $2k + 1$ eigenvectors of $H_{\text{clean}}^{(k)}$, if $\tilde{\ell}_{2k+1} > \|R_k^{(k)}\|$ and in other words $\ell_{2k+1} > \frac{1}{2}\|R_k^{(k)}\|$. This phenomenon is similar to the phase transition for eigenvalues and eigenvectors of a low rank matrix under the additive perturbation of a Gaussian Wigner matrix in [7, Section 3.1], although our underlying clean matrices $H_{\text{clean}}^{(k)}$ are full rank. It seems that the eigenvectors of the unperturbed matrix are possible to recover even when the spectral gap is much smaller than that required by Davis–Kahan. In this case, the Davis–Kahan theorem is insufficient to bound the distance between the subspaces since it does not consider the nature of the perturbation. It is useful to use perturbation bounds that take into account the nature of the perturbation such as the upper bound on the entry-wise deviation of the eigenvector in [23,

Theorem 8]. The theorem only applies to the situation with $\ell_{2k+1} > \|R^{(k)}\|$. According to the theorem, both δ_k and $\ell_{2k+1} - \|R^{(k)}\|$ appear in the denominators of the terms in the upper bound for the entry-wise deviation of the eigenvector. As k increases, the spectral gap δ_k increases, while the term $\ell_{2k+1} - \|R^{(k)}\|$ decreases based on Corollaries 4.1, 4.2 and Theorem 4.2. This implies that the upper bound of the deviation in [23, Theorem 8] will decrease initially as k increases from 1 because the reduction in the term that contains δ_k dominates, and then it will increase when the increments in the terms containing $\ell_{2k+1} - \|R^{(k)}\|$ becomes dominant. We empirically observe that the accuracy of the affinity measure $A^{(k)}$ increases with increasing k up to a critical cutoff k_c as detailed in Section 7.1. We identify k_c as the point when $\ell_{2k_c+2}^{k_c} \geq \frac{1}{2}\|R^{(k_c)}\|$ and $\ell_{2k_c+4}^{(k_c+1)} < \frac{1}{2}\|R^{(k_c)}\|$, which corresponds to when $\tilde{\ell}_{2k+2}^{(k)}$ becomes very close to $\|R^{(k)}\|$. The estimation of the top eigenspace gets less accurate when k increases beyond k_c , which will result in worse classification results using $A^{(k)}$ (see Fig. 8). The eigenvector perturbation of a full-rank matrix with additive random matrix is still an open problem and we will provide theoretical justification for our observations in the future.

Based on the discussions above, we see the benefit of using $A^{(k)}$ for $k > 1$ to select nearest neighbors because the underlying embedding $\Psi^{(k)}$ can be more stable than $\Psi^{(1)}$. Under additive noise perturbation in (6.4), each embedding $\Psi^{(k)}$ is perturbed randomly, but they have non-trivial correlation with the corresponding true eigenspace when the noise is not too large. In addition, $\Psi^{(k)}(i)$ encodes the viewing direction information in terms of the degree- k polynomial of the frame x_i and the underlying information on x_i is perturbed differently at different k even though the noise is not independent. The combined score is able to identify pairs that have consistently high affinities across k and filter out pairs that only have a couple of high scores across k .

6.2 Random rewiring model with small angular errors

We extend the random rewiring model in Section 6.1 to incorporate small angular errors in the pairwise alignment angles for the correctly connected pairs. Specifically, we consider additive errors in the angle,

$$\tilde{\theta}_{ij} = \theta_{ij} + \varepsilon_{ij}, \quad (6.8)$$

where ε_{ij} are independently drawn from a distribution γ on the interval $[0, 2\pi)$. We also assume that $\mathbb{E}(\varepsilon_{ij}) = 0 \bmod 2\pi$. We can evaluate $c_k = \mathbb{E}(e^{\iota k \varepsilon})$ for $\varepsilon \sim \gamma([0, 2\pi))$. The matrix $H^{(k)}$ is a random matrix under this model with

$$H_{ij}^{(k)} = \begin{cases} e^{\iota k(\theta_{ij} + \varepsilon_{ij})}, & \text{if } (i, j) \in E \text{ and with probability } p, \\ e^{\iota k \phi_{ij}}, & \text{if } (i, j) \notin E \text{ and with probability } \frac{(1-p)\bar{D}}{N-\bar{D}}. \end{cases} \quad (6.9)$$

Since the expected value of the random variable $e^{\iota k \phi}$ vanishes for $\phi \sim \text{Uniform}[0, 2\pi)$, the expected value of the matrix $H^{(k)}$ is

$$\mathbb{E}H^{(k)} = c_k p H_{\text{clean}}^{(k)}, \quad (6.10)$$

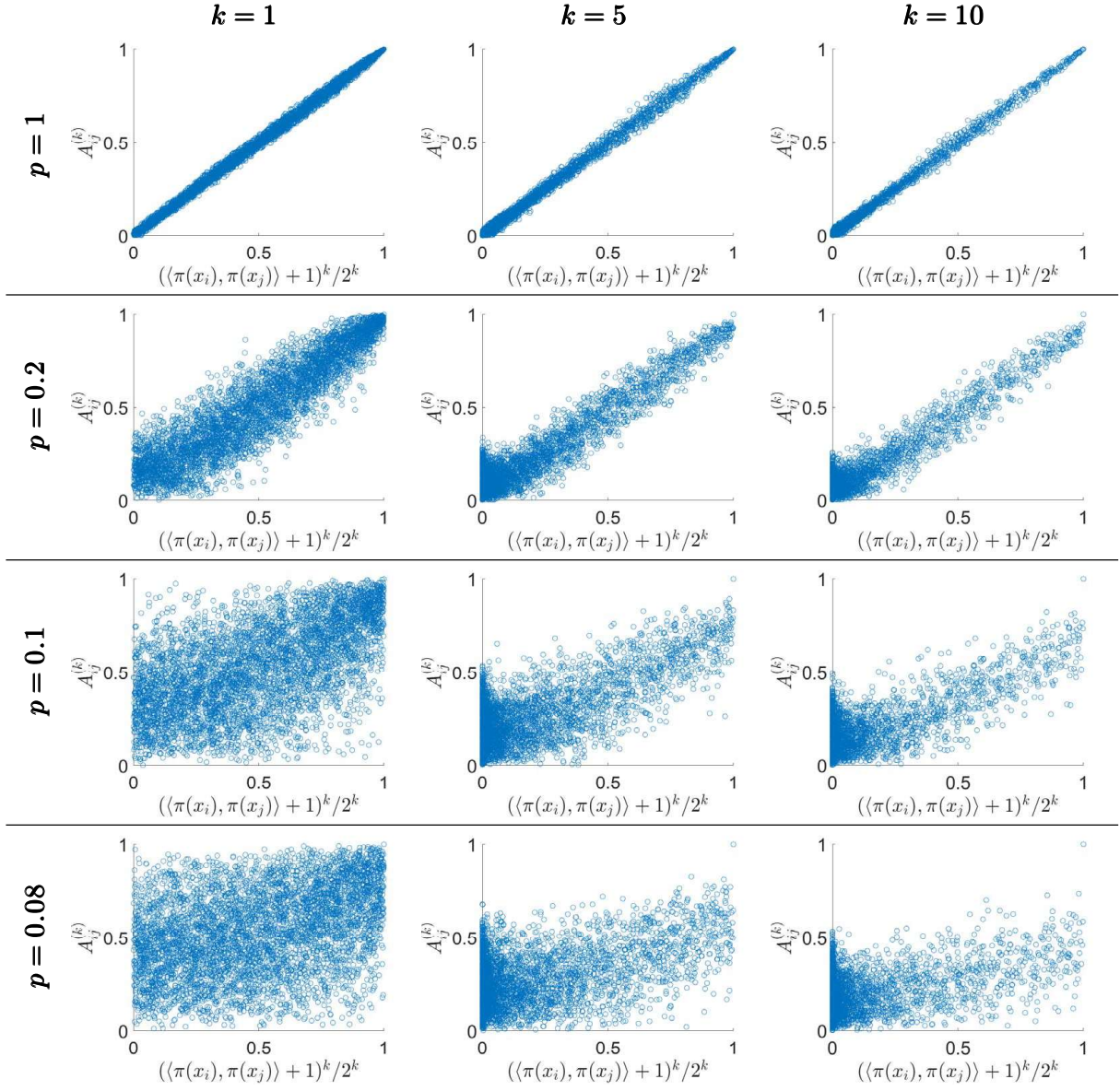


FIG. 5. Scatter plots of $A_{ij}^{(k)}$ against $(\langle \pi(x_i), \pi(x_j) \rangle + 1)^k / 2^k$ at $p = 1, 0.2, 0.1$ and 0.08 and $k = 1, 5$ and 10 . The robustness of the approximation (5.2) is considerably more robust for larger values of k .

where $H_{\text{clean}}^{(k)}$ is the clean matrix that corresponds to $p = 1$ obtained in the case that all links and angles are set up correctly. At each frequency k , the matrix $H^{(k)}$ can be decomposed into

$$H^{(k)} = c_k p H_{\text{clean}}^{(k)} + R^{(k)}, \quad (6.11)$$

where $R^{(k)}$ is a random matrix whose elements are

$$R_{ij}^{(k)} = \begin{cases} (e^{ik\varepsilon_{ij}} - c_k p) e^{ik\theta_{ij}}, & \text{if } (i, j) \in E \text{ and with probability } p, \\ -c_k p e^{ik\theta_{ij}}, & \text{if } (i, j) \in E \text{ and with probability } 1 - p, \\ e^{ik\phi_{ij}}, & \text{if } (i, j) \notin E \text{ and with probability } \frac{(1-p)\bar{D}}{N-\bar{D}}. \end{cases} \quad (6.12)$$

The analysis follows the steps in Section 6.1 and $\|R^{(k)}\| \leq 2\sqrt{N} \sin \frac{\alpha}{2}$ with high probability. Comparing Equation (6.10) with Equation (6.5), we find that the main difference is that the eigenvalues of $\mathbb{E}H^{(k)}$ are scaled by c_k at frequency k . The condition for the spectral algorithm to work is that the spectral gap $c_k p N G^{(k)}$ and the top eigenvalue $c_k p N \lambda_k^{(k)}$ are sufficiently large compared with $\|R^{(k)}\|$. For a well-concentrated distribution γ , we can first evaluate c_k and then determine the critical cutoff frequency k_c that satisfy the condition. With the same p in the random rewiring model, k_c gets smaller in the presence of additional small angular errors since $c_k < 1$. In Section 7.1, we show the performance of the algorithms on a couple of examples where the angular noise follows a von Mises distribution.

6.3 Discussions

In the previous two models, we only consider independent edge noise, i.e., the entries in $R^{(k)}$ for a fixed k are independent. Across different frequencies, the entries R_k are dependent through the relations of the irreducible representations of the angles $(\theta_{ij}, \phi_{ij}, \text{ and } \varepsilon_{ij})$ and the graph connectivity. We note that these are simplified models for illustrating the benefits of using $A^{(k)}$ for $k > 1$. In the application to cryo-EM 2-D image analysis, the edge perturbations are induced by the independent noise from each image. In this case, for fixed frequency, the entries in $R^{(k)}$ become dependent since the edge connections and alignments are affected by the noise in each image node. Still we observe similar benefits of using $A^{(k)}$ for $k > 1$ with the cryo-EM class averaging experiments detailed in Section 7.2. We leave the analysis of node level noise to future work. In addition, the current analysis focuses on data points that are uniformly distributed on the manifold. For non-uniformly distributed data points, different normalization techniques introduced in diffusion maps [15] are needed to compensate for the non-uniform sampling density.

7. Numerical results

We conducted two sets of numerical experiments. The first set involves simulations of the probabilistic model introduced in [63]. The second set applies the proposed algorithm on the noisy simulated projection images of a three-dimensional volume of 70S ribosome. We point out that there is no direct way to compare the performance of classification algorithms on real microscope images, since their viewing directions are unknown. The only way to compare classification algorithms on real data is indirectly, by evaluating the resulting three-dimensional reconstructions. Here we conduct only numerical experiments from which conclusions can be drawn directly for two-dimensional images. All experiments in this section were executed on a Linux machine with 16 Intel Xeon 2.5GHz cores and 512GB of RAM.

7.1 Experiments with random rewiring model

We generate $N = 10,000$ orthonormal frames $x_1, \dots, x_N$ in $\mathbb{R}^3$, uniformly sampled from $\text{SO}(3)$ with respect to the normalized Haar measure. To generate the noisy graph under the probabilistic model introduced in [63], we keep the correct edge in the neighborhood graph with probability p and use the ground truth local parallel transport data $e^{ik\theta_{ij}}$ in (6.3). With probability $1 - p$, we rewire the edge such that the node i is connected to a randomly selected node that is not connected with i . For the rewired edge, the optimal in-plane rotational alignment angle is replaced with an angle uniformly sampled from 0 to 2π .

In the first experiment, we use a small dataset with $N = 1000$ frames in order to visualize all eigenvalues of $H^{(k)}$. The clean geometric neighborhood is constructed by connecting points where $\langle \pi(x_i), \pi(x_j) \rangle > 0.8$ (the opening angle $\alpha = 36.9^\circ$) to make sure that the graph is well connected. We vary p and compute all the eigenvalues of $H^{(k)}$ to illustrate the analysis in Section 6. Figure 2 shows the histograms of the eigenvalues of the matrices $H^{(k)}$ and $R^{(k)}$. We observe that the top eigenvalue of $H^{(k)}$ decreases as k decreases which is consistent with Corollary 4.1. The upper bound for $\|R^{(k)}\|$ as discussed in Section 6 is $2\sqrt{N} \sin \frac{\alpha}{2} = 20$, which is consistent with the results shown in the bottom row of Fig. 2. In addition, the same figure shows that $\|R^{(k)}\|$ does not vary with frequency index k under the random rewiring model. Comparing Fig. 2a with Fig. 2b, we see that the spectral gap between $(2k + 1)$ th and $(2k + 2)$ th eigenvalues increases. Increasing k further, we observe that the $(2k + 2)$ th eigenvalue of $H^{(k)}$, i.e., $\tilde{\ell}_{2k+2}^{(k)}$, becomes very close to the right edge of the semi-circle as shown in Fig. 2c. Figure 3 shows the proportion of the estimated 50 nearest neighbors for each frame that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.85$. The proportion reaches the maximum at $k = 9$ for $p = 0.5$ and $p = 0.3$.

In the second experiment, we use 10,000 frames to show the spectral properties and the performance of the MFCA algorithm for large sample size. The clean geometric neighborhood graph is constructed by connecting points where $\langle \pi(x_i), \pi(x_j) \rangle > 0.92$ (within 23.1° opening angle). We compute the eigenvalues and eigenvectors of the normalized Hermitian matrix, $\tilde{H}^{(k)} = D^{-1/2} H D^{-1/2}$. Figure 4 shows the top eigenvalues of $\tilde{H}^{(k)}$. The multiplicities $2k + 1, 2k + 3, 2k + 5, \dots$ of the top eigenvalues are clearly demonstrated in the bar plots for $p = 1$ (the first row in Fig. 4). As p decreases, the top spectral gap gets smaller and when $p = 0.1$, it is hard to identify the spectral gap for $k = 1$, whereas the top spectral gap at $k = 5$ is still noticeable. This is consistent with our expectation for improved spectral stability for larger k .

The estimated $A_{ij}^{(k)}$'s provide good approximations to $\left(\langle \pi(x_i), \pi(x_j) \rangle + 1 \right)^k / 2^k$ (see the top row of Fig. 5). This approximation deteriorates as p decreases. The lower left sub-figure of Fig. 5 shows that the original single frequency class averaging nearest neighbor search algorithm fails at $p = 0.08$. Figure 6 shows the scatter plots of the combined affinity against the dot products $\langle \pi(x_i), \pi(x_j) \rangle$ between the true viewing angles at varying p . Even at $p = 0.08$, the combined affinity A_{ij}^{All} is still able to identify frames of similar viewing directions.

We evaluate the performance of the proposed algorithms on the nearest neighbor search by inspecting the magnitudes of the angles between the viewing directions of frames identified as neighbors by the algorithm. We identify for each frame 50 nearest neighbors with respect to the affinity measure and plot in Fig. 7 the histogram of the angles between the viewing directions of neighboring frames for varying rewiring probabilities $p = 1, 0.2, 0.1, 0.08$. From Fig. 7, we observe that using the affinity $A^{(k)}$ in (2.3) at higher frequency helps improve the performance of the single-frequency class averaging nearest neighbor search algorithm, especially for the noisy graph at $p = 0.08$ (i.e., 92% of the true edges are corrupted). Moreover, combining the measures at different k s according to (2.4) further improves the

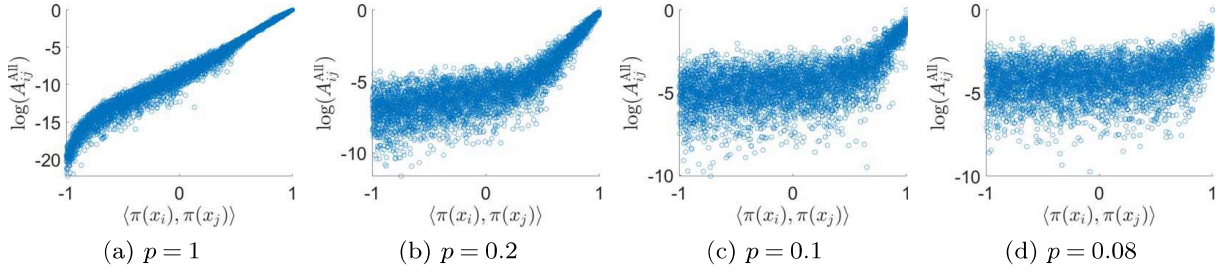


FIG. 6. Scatter plots for log MFCFA affinity $\log A_{ij}^{\text{All}}$ against $\langle \pi(x_i), \pi(x_j) \rangle$ at $p = 1, 0.2, 0.1$ and 0.08 .

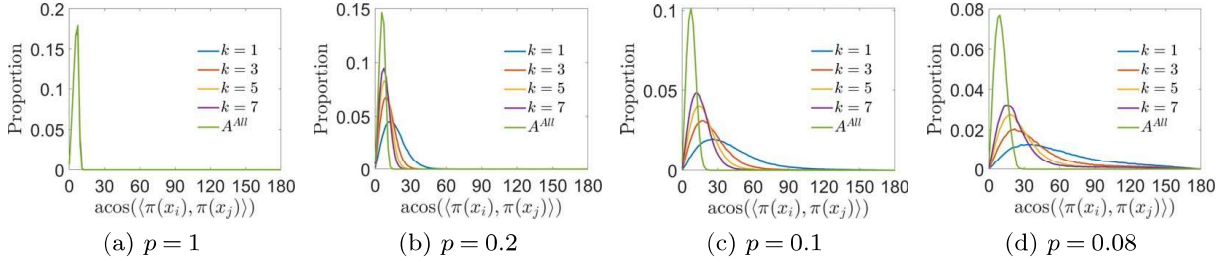


FIG. 7. Histogram of the angles (x-axis, in degrees) between the viewing directions of 10,000 simulated frames and its 50 neighboring points at $p = 1, 0.2, 0.1$ and 0.08 . For A^{All} , we use $k_{\text{max}} = 20$.

classification results with significant reduction of outliers at $p = 0.08$ compared to the single frequency nearest neighbor identification results.

Singer *et al.* proposed to use more than top three eigenvectors from $\tilde{H}^{(1)}$ for nearest neighbor classification in [63, Section 7]. We include it as an additional baseline for comparison here to illustrate the benefit of using the eigenvectors of $\tilde{H}^{(k)}$ for $k > 1$. Specifically, using the top $2k + 1$ eigenvectors of $\tilde{H}^{(1)}$, we define the affinity $B^{(k)}$ as

$$\tilde{\Psi}_k^{(1)}(i) = (\psi_1^{(1)}(i), \psi_2^{(1)}(i), \dots, \psi_{2k+1}^{(1)}(i)), \quad B^{(k)} = \frac{|\langle \tilde{\Psi}_k^{(1)}(i), \tilde{\Psi}_k^{(1)}(j) \rangle|}{\|\tilde{\Psi}_k^{(1)}(i)\| \|\tilde{\Psi}_k^{(1)}(j)\|}. \quad (7.1)$$

We compare the performance of the algorithms in terms of the proportion of estimated nearest neighbors that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.95$. Figure 8b shows that under large noise regimes, where 90% of the clean edges are randomly rewired, using $A^{(k)}$ at $k = 16$ outperforms the previous class averaging algorithm that uses only the eigenvectors from $\tilde{H}^{(1)}$. As shown in Fig. 8c, combining the information from different frequency channels can significantly boost the performance in finding true nearest neighbors. For S^{All} , the proportion reaches the maximum value 0.90 at $k = 30$. For A^{All} , the proportion reaches the maximum value 0.94 at $k = 20$. Because the higher-order terms $A^{(k)}$ get much smaller than 1 and become less informative, incorporating more $A^{(k)}$ components deteriorates the performance of the combined score A^{All} when $k > 20$. The combined affinity S^{All} is more stable at large k .

To understand why the combined affinities can significantly improve the classification results at $p = 0.08$, we check the values of $A^{(k)}$ for $k = 1, \dots, 25$ for pairs of frames x_i and x_j that satisfy $\langle \pi(x_i), \pi(x_j) \rangle < 0.95$, but are still identified as nearest neighbors by $A^{(1)}$. We observe that although the

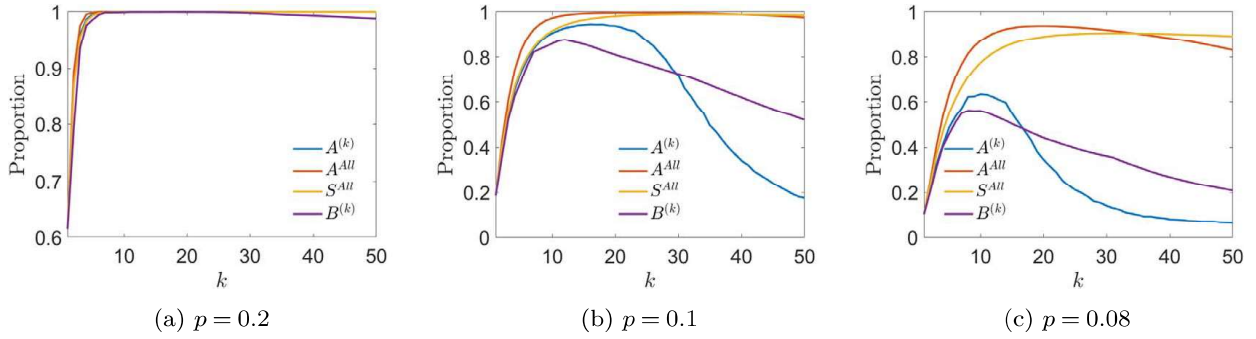


FIG. 8. Comparing the performance of different affinities according to $A^{(k)}$, A^{All} , S^{All} and $B^{(k)}$. We evaluate the proportion of the estimated nearest neighbors that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.95$.

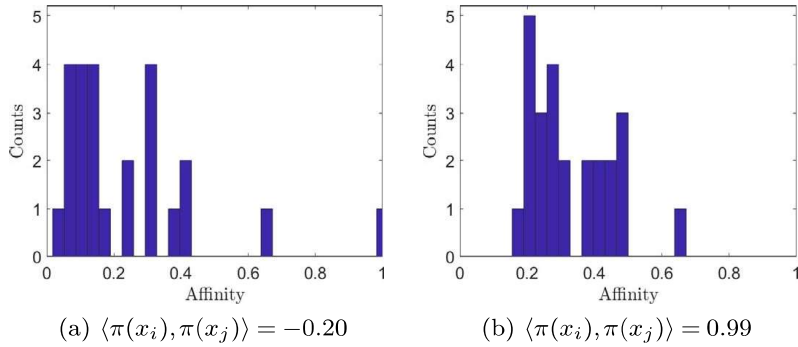


FIG. 9. Histograms of the affinities $A_{ij}^{(k)}$ with $k = 1, \dots, 25$ for (a) a pair of wrongly identified nearest neighbors by $A^{(1)}$ and (b) a good nearest neighbor pair identified by A^{All} , but not by any $A^{(k)}$. The data are generated under random rewiring model with $p = 0.08$.

corresponding affinities at frequency 1 are above 0.97, $A_{ij}^{(k)}$ at other frequency indices are below 0.7 and concentrated on the interval $(0, 0.2]$ (see the example in Fig. 9a). Therefore, the combined affinity A^{All} is very small and such pair will be removed from the nearest neighbor list. In contrast, for a pair of true nearest neighbors that does not appear in any nearest neighbor list by $A^{(k)}$ for $k = 1, \dots, 25$, we observe that although the affinities are lower than 0.7, all individual affinities lie between 0.2 and 0.5 (see Fig. 9b). Thus, the combined affinity A^{All} is higher for the pair in Fig. 9b than the pair in Fig. 9a. In summary, A^{All} is able to not only reject wrongly identified nearest neighbors by $A^{(k)}$, but also find new correct nearest neighbors that are missed by $A^{(k)}$.

In the third experiment, we incorporate the small angular perturbation into the random rewiring model according to Equation (6.9). Specifically, we assume that the distribution of the angular error follows the von Mises distribution,

$$\gamma(\varepsilon) = \frac{e^{\kappa \cos(\varepsilon)}}{2\pi I_0(\kappa)}, \quad (7.2)$$

where $I_0(\kappa)$ is the modified Bessel function of order 0. The parameter κ controls the concentration of the distribution. For this particular distribution, $c_k = \mathbb{E}(e^{I_k \varepsilon}) = \frac{I_k(\kappa)}{I_0(\kappa)}$, where $I_k(\kappa)$ is the modified Bessel function of order k for $k > 0$. The clean geometric neighborhood graph is constructed by connecting

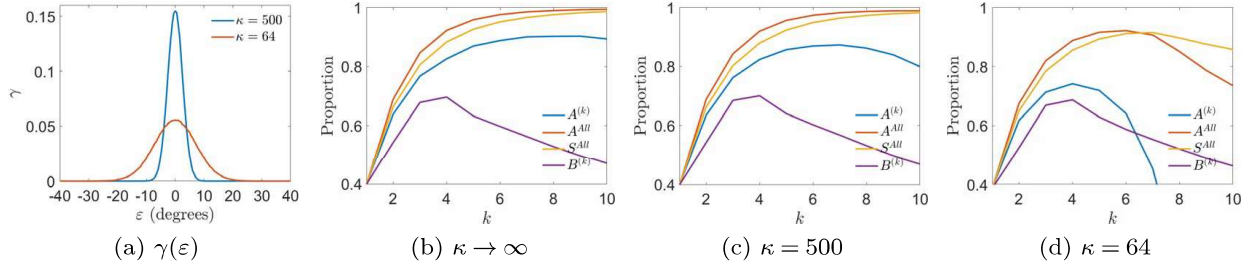


FIG. 10. Comparing the performance of different affinities according to $A^{(k)}$, A^{All} , S^{All} and $B^{(k)}$. 10b–10d The proportion of the estimated nearest neighbors that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.95$ with various κ . 10a Distributions of ε_{ij} for $\kappa = 500$ and $\kappa = 64$.

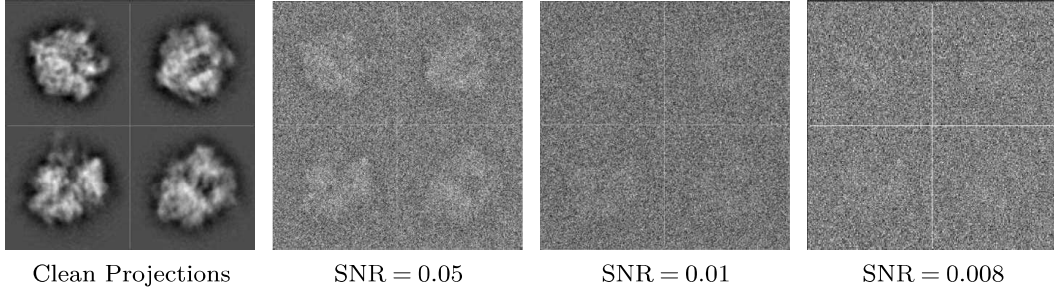


FIG. 11. Samples of simulated projection images on 70S ribosome. From left to right: clean projection images, images contaminated by additive white Gaussian noise with signal to noise ratio SNR= 0.05, 0.01 and 0.008.

points where $\langle \pi(x_i), \pi(x_j) \rangle > 0.7$ with 10,000 frames. We fix $p = 0.08$ (92% of the clean edges are randomly rewired) and vary the parameter κ in von Mises distribution. We show the accuracy of the 50-nearest neighbor identification in Fig. 10. Figure 10a depicts the distribution of the angle ε with $\kappa = 500$ and $\kappa = 64$.

Figure 10b shows the results for random rewiring model without angular perturbation and the performance of $A^{(k)}$ is consistently better than $B^{(k)}$. Comparing Fig. 10b with Fig. 8c, we find that we achieve higher accuracy in the nearest neighbor identification from a more densely connected graph in all approaches. From Figs 10b–10d, we find the performance of $B^{(k)}$ is stable over small angular perturbation. In comparison, the performance of single frequency affinity $A^{(k)}$ deteriorates as κ increases. This is due to the fact that c_k gets smaller as κ increases and both the top spectral gap and top eigenvalue of $\mathbb{E}H^{(k)}$ depend on c_k . Despite this, the combined scores still achieve higher accuracy than $A^{(k)}$ and $B^{(k)}$.

7.2 Experiments with simulated cryo-EM images

In this section, we apply MFCA on simulated cryo-EM projection images. For each image, the goal is to identify projection images with similar viewing directions. We simulate $N = 10,000$ clean projection images of size 129×129 pixels from a three-dimensional electron density map of the 70S ribosome. The orientations for the projection images are uniformly distributed over $\text{SO}(3)$. The clean images are contaminated by additive white Gaussian noise with different SNRs. Sample images are presented in Fig. 11. Here, we do not consider the effects of contrast transfer functions on the images. In order to initially identify similar images and the corresponding rotational alignments, we first expand each image on steerable basis, and denoise the images by using *steerable principal component analysis*

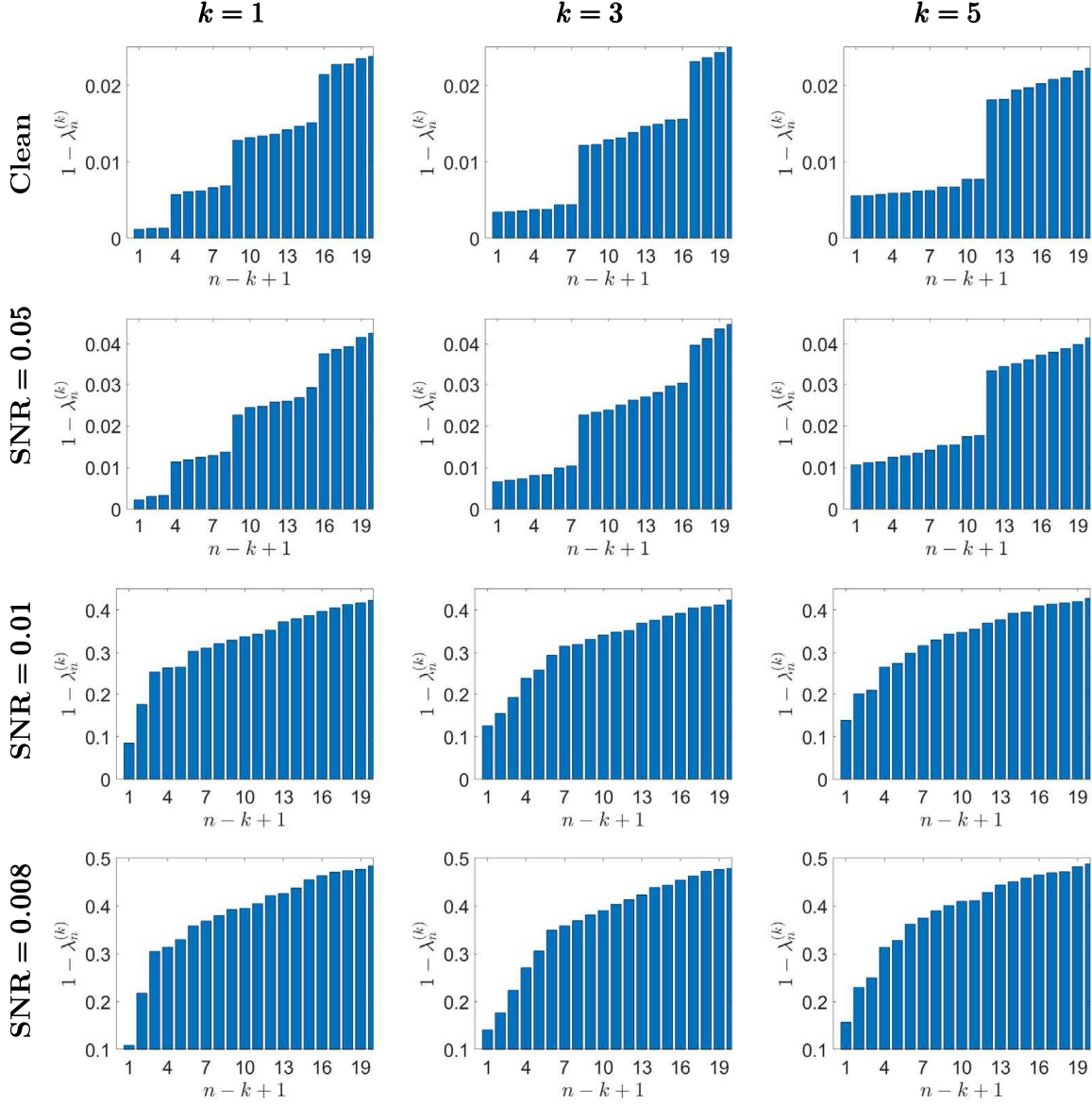


FIG. 12. Bar plots of the top 20 eigenvalues at different frequency k and SNR for simulated cryo-EM projection images.

(sPCA) [73]. Then we generate the rotationally invariant features [75] from the filtered expansion coefficients to efficiently identify nearest neighbors without performing all pairwise alignments. The optimal alignment parameters are estimated between initial nearest neighbor pairs. The initial nearest neighbor list and alignment parameters are used to construct the initial graph. For clean images, the initial graph corresponds to the true neighborhood graph. For the extremely noisy images illustrated in Fig. 11, the initial similarity measure is corrupted by noise and images of totally different views can be misidentified as nearest neighbors.

In Fig. 12, we present the spectral patterns of the top eigenvalues of $\tilde{H}^{(k)}$ built from our initial neighborhood identification and rotational alignment. At high SNR, such as $\text{SNR} \geq 0.05$, we can clearly

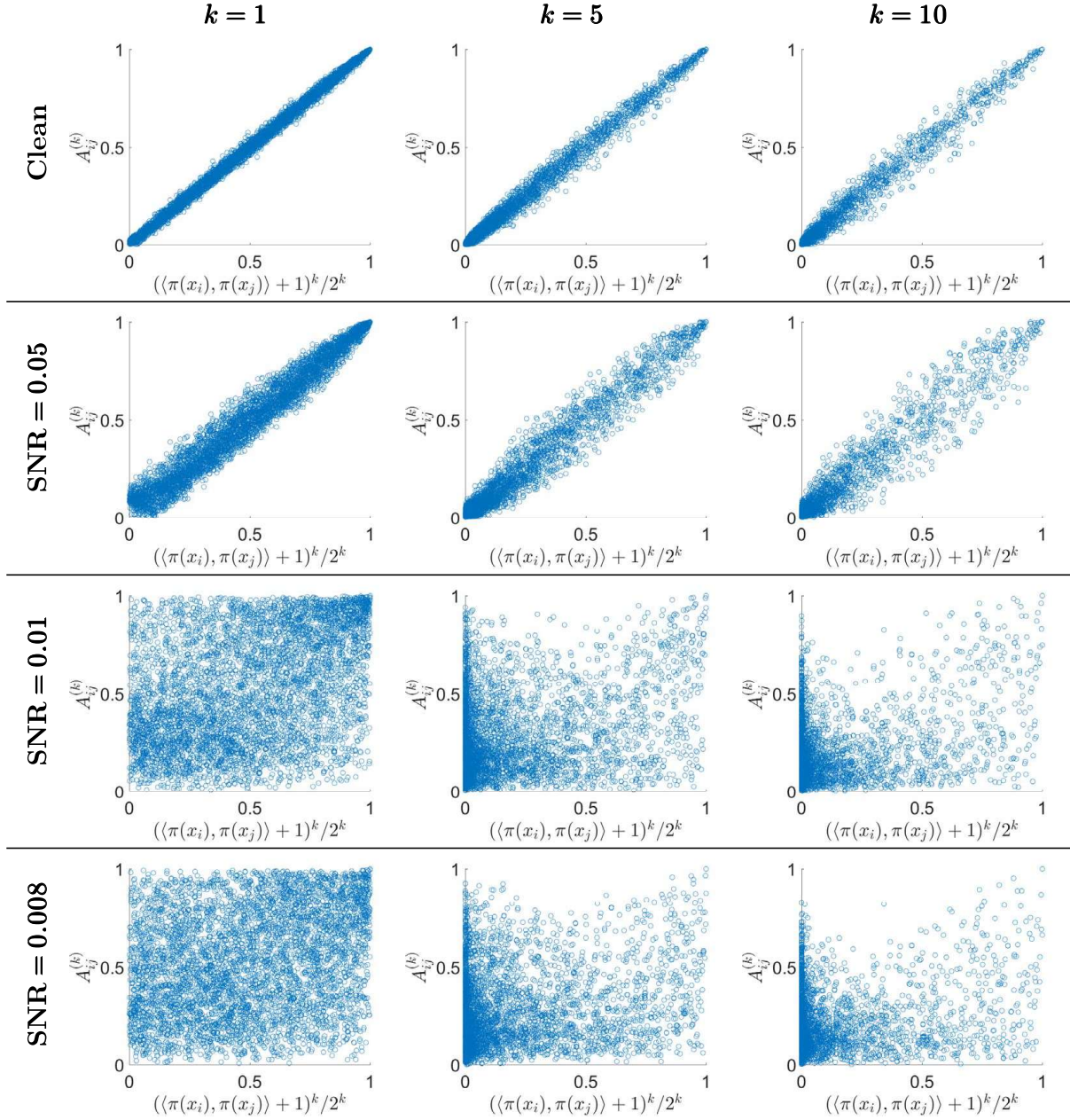


FIG. 13. Scatter plots of $A_{ij}^{(k)}$ against $((\pi(x_i), \pi(x_j)) + 1)^k / 2^k$ for clean and noisy projection images with SNR = 0.05, 0.01, 0.008, at frequency $k = 1, 5$ and 10 .

observe the multiplicities $2k + 1, 2k + 3, 2k + 5, \dots$ and the spectral gaps. As the SNR decreases, such spectral patterns deteriorate.

In Fig. 13, we present the scatter plots of $A_{ij}^{(k)}$ against $((\pi(x_i), \pi(x_j)) + 1)^k / 2^k$, with different SNRs. Similar to the synthetic dataset, the $A_{ij}^{(k)}$ s at frequency $k = 1$ fail at low SNRs, such as SNR = 0.01, 0.008, while the $A_{ij}^{(k)}$ s at frequency $k = 5, 10$ are still able to distinguish the images

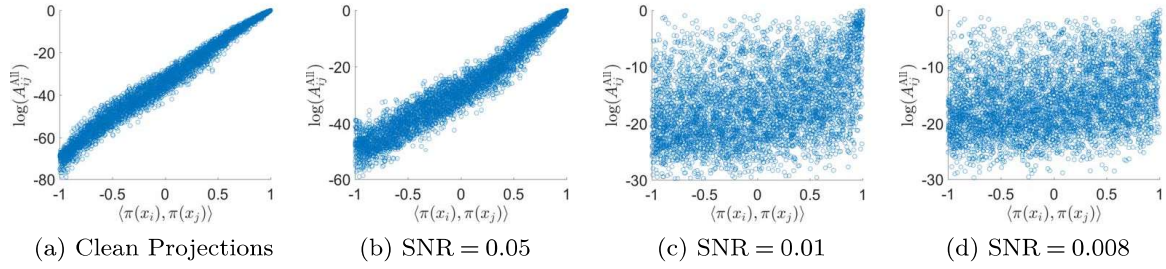


FIG. 14. Scatter plots for log MFCA affinity $\log A_{ij}^{\text{All}}$ against $\langle \pi(x_i), \pi(x_j) \rangle$ for clean and noisy projection images with SNR = 0.05, 0.01 and 0.008.

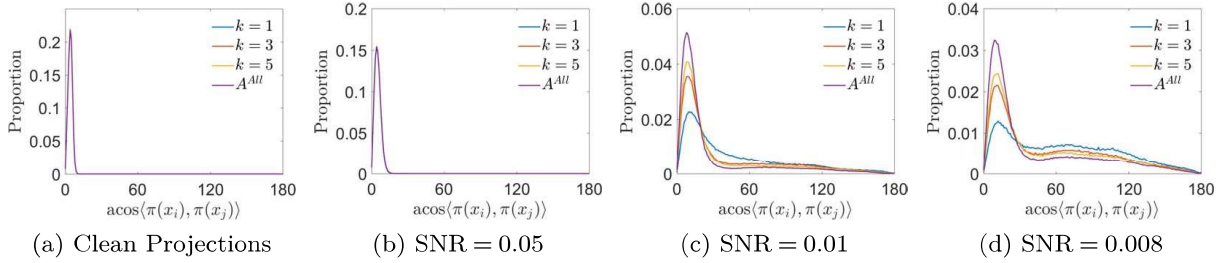


FIG. 15. Histogram of the angles (x-axis, in degrees) between the viewing directions of 10,000 simulated cryo-EM projection images and its 50 neighboring projection images, with different SNRs, from *left to right*: clean projection images, SNR = 0.05, 0.01, 0.008. Here we set the maximum frequency $k_{\max} = 20$.

with similar viewing directions (i.e., $(\langle \pi(x_i), \pi(x_j) \rangle + 1)^k / 2^k \approx 1$). This result indicates that better neighborhood image identification can be attained using higher frequency k . Moreover, Fig. 14 shows the scatter plots of the combined affinity against the dot products $\langle \pi(x_i), \pi(x_j) \rangle$ between the true viewing angles at varying SNRs. Even at SNR = 0.01, the combined affinity A_{ij}^{All} is still able to distinguish projection images that have similar views $\pi(x)$, in contrast to the approximation results in Fig. 13.

In Fig. 15, we evaluate the results by plotting the histogram of angles between viewing directions $\arccos(\pi(x_i), \pi(x_j))$ between all identified neighboring images I_i and I_j . At high SNR, such as SNR = 0.05, using single frequency information as $k = 1, 3, 5$ can achieve similar results as combining all the frequencies. At low SNRs, such as SNR = 0.01 and 0.008, A^{All} which uses all frequencies information up to $k = 20$ outperforms the results obtained from using only a single frequency at $k = 1, 3, 5$.

In Fig. 16, we compare the nearest neighbor classification results using affinities $A^{(k)}$, $B^{(k)}$, A^{All} and S^{All} at various frequency index k for noisy images with SNR = 0.05, 0.01 and 0.008. The latter two affinities combine $A^{(k')}$ for $k' = 1, \dots, k$. Each image is identified with 50 nearest neighbors and we evaluate the proportion of the estimated nearest neighbors that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.9$. At SNR = 0.05, all approaches achieve high accuracy (see Fig. 16a). At SNR = 0.01, $A^{(k)}$ is able to achieve better classification results than $B^{(k)}$ for k between 4 and 32 and the proportion reaches 67.7% for $A^{(k)}$ at $k = 22$. Using A^{All} can improve the results further at $k = 40$, where the proportion reaches 68.3%. The improvement of $A^{(k)}$ and A^{All} compared with $B^{(k)}$ gets more prominent at lower SNR (see Fig. 16c with SNR = 0.008).

We note that the construction of the initial graph structure relies on the evaluation of the rotational invariant distance based on the sPCA expansion coefficients of the projection images [73–75]. Thus, the noise model is different from the probabilistic models in Section 6 and the perturbation at each edge

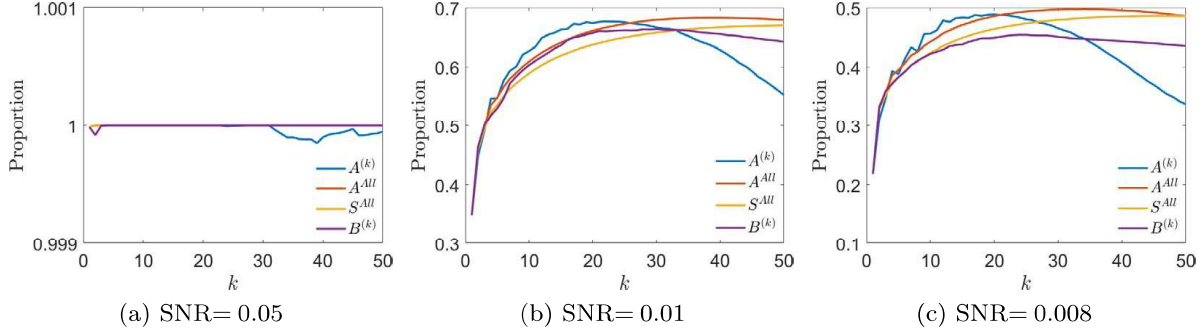


FIG. 16. Comparing the performance of different affinities according to $A^{(k)}$, A^{All} , S^{All} and $B^{(k)}$ for varying k on noisy cryo-EM images. We evaluate the proportion of the estimated nearest neighbors that satisfy $\langle \pi(x_i), \pi(x_j) \rangle > 0.9$.

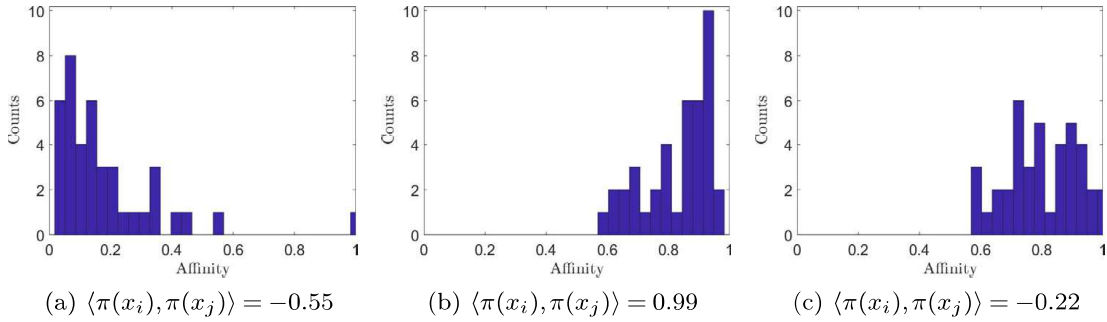


FIG. 17. Histograms of the affinities $A_{ij}^{(k)}$ with $k = 1, \dots, 40$ for (a) a pair of wrongly identified nearest neighbors by $A^{(1)}$, (b) a good nearest neighbor pair identified by A^{All} and (c) a wrongly identified nearest neighbor by A^{All} . The simulated cryo-EM images are of SNR=0.01.

is induced by the noise on the corresponding two nodes. Despite the difference in the noise model, we still observe the benefit of using $A^{(k)}$ with $k > 1$. However, the improvement of the combined affinity A^{All} is not as impressive as the examples shown in Fig. 8 and Fig. 10. Although we observe that certain misclassified nearest neighbors by $A^{(1)}$ can be corrected by A^{All} as shown in Fig. 17a, there are still some wrong nearest neighbors that enjoy consistently high affinities across different k s as shown in Fig. 17c.

8. Conclusion and future work

We propose in this paper a novel algorithm, referred to as MFCA, for classifying noisy projection images in three-dimensional cryo-electron microscopy by the similarity among viewing directions. The new algorithm is a generalization of the eigenvector-based approach of intrinsic classification first appeared in [38, 63]. We also extended the representation theoretical framework of [37, 38] by means of explicit constructions involving the Wigner D -matrices, which completely characterizes the spectral information of a generalized localized parallel transport operator acting on sections of certain complex line bundle over the two-dimensional unit sphere in $\mathbb{R}^3$; these theoretical results conceptually establish the admissibility and (improved) stability of the new MFCA algorithm.

One intriguing future direction is to investigate into refined and more systematic aggregations of the results obtained from each individual frequency channel. Potential candidates include (1) the harmonic-

retrieval-type transformations as in multi-frequency phase synchronization [32], (2) cross-frequency invariant features such as bispectrum [8, 41] and (3) tensor-based optimizations for multi-dimensional arrays [1, 43, 58]. The main idea is to further exploit the redundancy in the reconstructed information across different irreducible representations. A direct extension of the MFCA theoretical framework could be a refined geometric interpretation of the multi-frequency vector diffusion maps [24] in terms of aggregating invariant embeddings of the same underlying base manifold from multiple associated vector bundles of a fixed common principal bundle.

Another future direction of interest is to integrate the multi-frequency methodology into existing algorithmic approaches for tackling the *heterogeneity* problem in cryo-EM imaging analysis and comparative biology [2, 31, 46]. In the context of cryo-EM, this problem occurs when molecules in distinct conformations coexist in solution, and thus images collected in cryo-EM imaging from random orientations should typically be first clustered into subgroups (using, e.g., the maximum likelihood classification approaches [57, 60]) before single-particle reconstruction techniques can be applied to each individual subgroup. Recent studies [16, 27, 28] even provided evidence for a continuous distribution of conformation states to present in a solution, which is far beyond the capability of maximum likelihood classification methods. We expect significant performance boost and sharper theoretical results from extensions of the multi-frequency methodology in these problems.

Acknowledgements

The authors thank Vera Mikyoung Hur, Jared Bronski, Shmuel Weinberger and Shamgar Gurevich for useful discussions.

Funding

Strategic Research Initiatives in the University of Illinois at Urbana-Champaign (to Z. Z. and Y. F.); Alfred P. Sloan Foundation (to Z. Z. and Y. F.); National Science Foundation (DMS-1854791 to Z. Z. and Y. F., DMS-1854831 and IIS-1546413 to T. G.); AMS-Simons Travel Grant (to T. G.); DARPA (D15AP00109 to T.G.).

REFERENCES

1. ANKELE, M., LIM, L.-H., GROESCHEL, S. & SCHULTZ, T. (2017) Versatile, robust, and efficient tractography with constrained higher-order tensor fODFs. *Int. J. Comput. Assist. Radiol. Surg.*, **12**, 1257–1270.
2. BAJAJ, C., GAO, T., HE, Z., HUANG, Q. & LIANG, Z. (2018) SMAC: simultaneous mapping and clustering using spectral decompositions. *International Conference on Machine Learning*, Stockholm Sweden: PMLR. pp. 334–343.
3. BANDEIRA, A. S., CHEN, Y. & SINGER, A. (2020) Non-unique games over compact groups and orientation estimation in cryo-EM. *Inverse Problems*, **36**(6), 064002.
4. BANDEIRA, A. S., SINGER, A. & SPIELMAN, D. A. (2013) A Cheeger inequality for the graph connection Laplacian. *SIAM J. Matrix Anal. Appl.*, **34**, 1611–1630.
5. BELKIN, M. & NIYOGI, P. (2005) Towards a theoretical foundation for Laplacian-based manifold methods. *International Conference on Computational Learning Theory*, pp. 486–500. Springer.
6. BELKIN, M. & NIYOGI, P. (2007) Convergence of Laplacian eigenmaps. *Advances in Neural Information Processing Systems*, Vancouver, B.C., Canada: MIT Press. vol. 19, pp. 129.
7. BENAYCH-GEORGES, F. & NADAKUDITI, R. R. (2011) The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, **227**, 494–521.
8. BENDORY, T., BOUMAL, N., MA, C., ZHAO, Z. & SINGER, A. (2018) Bispectrum inversion with application to multireference alignment. *IEEE Trans. Signal Process.*, **66**, 1037–1050.

9. BOUMAL, N., SINGER, A., ABSIL, P.-A. & BLONDEL, V. D. (2014) Cramér–Rao bounds for synchronization of rotations. *Inf. Inference*, **3**, 1–39.
10. BOYLE, M. (2016) How should spin-weighted spherical functions be defined? *J. Math. Phys.*, **57**, 092504.
11. BRÖCKER, T. & TOM DIECK, T. (2013) *Representations of Compact Lie Groups*, vol. 98. Springer Science & Business Media.
12. CAMPBELL, W. B. (1971) Tensor and spinor spherical harmonics and the spin- s harmonics ${}_sY_{lm}(\theta, \varphi)$. *J. Math. Phys.*, **12**, 1763–1770.
13. CHEN, B. & FRANK, J. (2015) Two promising future developments of cryo-EM: capturing short-lived states and mapping a continuum of states of a macromolecule. *Microscopy*, **65**, 69–79.
14. CHOW, Y., GATTESCHI, L. & WONG, R. (1994) A Bernstein-type inequality for the Jacobi polynomial. *Proc. Amer. Math. Soc.*, **121**, 703–709.
15. COIFMAN, R. R. & LAFON, S. (2006) Diffusion maps. *Appl. Comput. Harmon. Anal.*, **21**, 5–30.
16. DASHTI, A., SCHWANDER, P., LANGLOIS, R., FUNG, R., LI, W., HOSSEINIZADEH, A., LIAO, H. Y., PALLESEN, J., SHARMA, G., STUPINA, V. A., SIMON, A. E., DINMAN, J. D., FRANK, J. & OURMAZD, A. (2014) Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 17492–17497.
17. DAVIS, C. & KAHAN, W. (1970) The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, **7**, 1–46.
18. DOYLE, D. A., CABRAL, J. M., PFUETZNER, R. A., KUO, A., GULBIS, J. M., COHEN, S. L., CHAIT, B. T. & MACKINNON, R. (1998) The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*, **280**, 69–77.
19. DRIVER, K. A. & JORDAAN, K. (2012) Bounds for extreme zeros of some classical orthogonal polynomials. *J. Approx. Theory*, **164**, 1200–1204.
20. EASTWOOD, M. & TOD, P. (1982) Edth—a differential operator on the sphere. *Math. Proc. Cambridge Philos. Soc.*, **92**, 317–330.
21. EL KAROUI, N. & WU, H.-T. (2016) Graph connection Laplacian methods can be made robust to noise. *Ann. Statist.*, **44**, 346–372.
22. ELBERT, A., LAFORGIA, A. & RODONÒ, L. G. (1994) On the zeros of Jacobi polynomials. *Acta Math. Hungar.*, **64**, 351–359.
23. ELDRIDGE, J., BELKIN, M. & WANG, Y. (2018) Unperturbed: spectral analysis beyond Davis–Kahan. *Algorithmic Learning Theory*, pp. 321–358. Lanzarote, Spain: PMLR.
24. FAN, Y. & ZHAO, Z. (2019) Cryo-electron microscopy image analysis using multi-frequency vector diffusion maps (in press). arXiv:1904.07772.
25. FAN, Y. & ZHAO, Z. (2019) Multi-frequency vector diffusion maps. *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri & R. Salakhutdinov eds), vol. 97. Long Beach, California, USA: PMLR, pp. 1843–1852.
26. FRANK, J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. New York, USA: Oxford University Press.
27. FRANK, J. (2018) New opportunities created by single-particle cryo-EM: the mapping of conformational space. *Biochemistry*, **57**, 888–888.
28. FRANK, J. & OURMAZD, A. (2016) Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, **100**, 61–67.
29. GAO, T. (2015) Hypoelliptic diffusion maps and their applications in automated geometric morphometrics. *Ph.D. Thesis*, Duke University.
30. GAO, T. (2021) The diffusion geometry of fibre bundles: Horizontal diffusion maps. *Applied and Computational Harmonic Analysis*, **50**, 147–215. Elsevier.
31. GAO, T., BRODZKI, J. & MUKHERJEE, S. (2021) The geometry of synchronization problems and learning group actions. *Discrete Comput. Geom.*, **65**, 150–211.
32. GAO, T. & ZHAO, Z. (2019) Multi-frequency phase synchronization. *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri & R. Salakhutdinov eds), vol. 97. Long Beach, California, USA: PMLR, pp. 2132–2141.

33. GELFAND, I. M., MINLOS, R. A. & SHAPIRO, Z. Y. (2018) *Representations of the Rotation and Lorentz Groups and Their Applications*. Mineola, New York, USA: Courier Dover Publications.
34. GOLDBERG, J. N., MACFARLANE, A. J., NEWMAN, E. T., ROHRLICH, F. & SUDARSHAN, E. G. (1967) Spin-spherical harmonics and ∂ . *J. Math. Phys.*, **8**, 2155–2161.
35. GUREVICH, S., HADANI, R. & SINGER, A. (2011) Representation theoretic patterns in three dimensional cryo-electron microscopy III—presence of point symmetries (in press).
36. HAAGERUP, U. & SCHLICHTKRULL, H. (2014) Inequalities for Jacobi polynomials. *Ramanujan J.*, **33**, 227–246.
37. HADANI, R. & SINGER, A. (2011) Representation theoretic patterns in three dimensional cryo-electron microscopy I: the intrinsic reconstitution algorithm. *Ann. of Math. (2)*, **174**, 1219–1241.
38. HADANI, R. & SINGER, A. (2011) Representation theoretic patterns in three-dimensional cryo-electron microscopy II—the class averaging problem. *Found. Comput. Math.*, **11**, 589–616.
39. HEEL, M. V. (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, **21**, 111–123.
40. HENDERSON, R. & McMULLAN, G. (2013) Problems in obtaining perfect images by single-particle electron cryomicroscopy of biological structures in amorphous ice. *Microscopy*, **62**, 43–50.
41. KAKARALA, R. (2012) The bispectrum as a source of phase-sensitive invariants for Fourier descriptors: a group-theoretic approach. *J. Math. Imaging Vision*, **44**, 341–353.
42. KHORUNZHY, A. (2001) Sparse random matrices: spectral edge and statistics of rooted trees. *Adv. in Appl. Probab.*, **33**, 124–140.
43. KOLDA, T. & BADER, B. (2009) Tensor decompositions and applications. *SIAM Rev.*, **51**, 455–500.
44. KOLTCHINSKII, V. & GINÉ, E. (2000) Random matrix approximation of spectra of integral operators. *Bernoulli*, **6**, 113–167.
45. KOORNWINDER, T., KOSTENKO, A. & TESCHL, G. (2018) Jacobi polynomials, Bernstein-type inequalities and dispersion estimates for the discrete Laguerre operator. *Adv. Math.*, **333**, 796–821.
46. LEDERMAN, R. R. & SINGER, A. (2020) A representation theory perspective on simultaneous alignment and classification. *Applied and Computational Harmonic Analysis*, **49**(3), 1001–1024. Elsevier.
47. LIN, C.-Y., MINASIAN, A., QI, X. J. & WU, H.-T. (2018) Manifold learning via the principle bundle approach. *Front. Appl. Math. Stat.*, **4**, 21.
48. MACKINNON, R. (2004) Potassium channels and the atomic basis of selective ion conduction (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.*, **43**, 4265–4277.
49. MALYARENKO, A. (2011) Invariant random fields in vector bundles and application to cosmology. *Ann. Inst. Henri Poincaré Probab. Stat.*, **47**, 1068–1095.
50. MARINUCCI, D. & PECCATI, G. (2011) *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications*, vol. 389. Cambridge, UK: Cambridge University Press.
51. NEWMAN, E. T. & PENROSE, R. (1966) Note on the Bondi–Metzner–Sachs group. *J. Math. Phys.*, **7**, 863–870.
52. OIKONOMOU, C. M. & JENSEN, G. J. (2017) The development of cryo-EM and how it has advanced microbiology. *Nat. Microbiol.*, **2**, 1577–1579.
53. PENCZEK, P. A., ZHU, J. & FRANK, J. (1996) A common-lines based method for determining orientations for $N > 3$ particle projections simultaneously. *Ultramicroscopy*, **63**, 205–218.
54. PERRY, A., WEIN, A. S., BANDEIRA, A. S. & MOITRA, A. (2018) Message-passing algorithms for synchronization problems over compact groups. *Commun. Pure Appl. Math.*, **71**(11), 2275–2322. Wiley Online Library.
55. REED, M. & SIMON, B. (1980) *Methods of Modern Mathematical Physics. Vol. 1. Functional Analysis*. San Diego, California, USA: Academic Press.
56. SALAS, D., LE GALL, A., FICHE, J.-B., VALERI, A., KE, Y., BRON, P., BELLOT, G. & NOLLMANN, M. (2017) Angular reconstitution-based 3D reconstructions of nanomolecular structures from superresolution light-microscopy images. *Proc. Natl. Acad. Sci. U. S. A.*
57. SCHERES, S. H. (2012) A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.*, **415**, 406–418.
58. SCHULTZ, T., FUSTER, A., GHOSH, A., DERICHE, R., FLORACK, L. & LIM, L.-H. (2014) Higher-order tensors in diffusion imaging. *Visualization and Processing of Tensors and Higher Order Descriptors*

- for *Multi-Valued Data* (C.-F. Westin, A. Vilanova & B. Burgeth eds). Berlin, Heidelberg: Springer, pp. 129–161.
59. SHKOLNISKY, Y. & SINGER, A. (2012) Viewing direction estimation in cryo-EM using synchronization. *SIAM J. Imag. Sci.*, **5**, 1088–1110.
 60. SIGWORTH, F. J., DOERSCHUK, P. C., CARAZO, J.-M. & SCHERES, S. H. W. (2010) An introduction to maximum-likelihood methods in cryo-EM. *Methods in enzymology*, vol. 482, pp. 263–294. San Diego, CA, USA: Academic Press Inc.
 61. SINGER, A. (2011) Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, **30**, 20–36.
 62. SINGER, A. & WU, H.-T. (2012) Vector diffusion maps and the connection Laplacian. *Commun. Pure Appl. Math.*, **65**, 1067–1144.
 63. SINGER, A., ZHAO, Z., SHKOLNISKY, Y. & HADANI, R. (2011) Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM J. Imag. Sci.*, **4**, 723–759.
 64. SZEGOŐ, G. (1939) *Orthogonal Polynomials*, vol. 23. Providence, Rhode Island, USA: American Mathematical Soc.
 65. VAINSHTEIN, B. & GONCHAROV, A. (1986) Determination of the spatial orientation of arbitrarily arranged identical particles of unknown structure from their projections. *Dokl. Phys.*, **31**, 278.
 66. VARSHALOVICH, D. A., MOSKALEV, A. N. & KHERSONSKII, V. K. (1988) *Quantum Theory of Angular Momentum*. Singapore: World Scientific.
 67. WANG, L. & SINGER, A. (2013) Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, **2**(2), 145–193. Oxford University Press.
 68. WATTS, D. J. & STROGATZ, S. H. (1998) Collective dynamics of ‘small-world’ networks. *Nature*, **393**, 440.
 69. WIGNER, E. P. (1955) Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, **62**, 548–564.
 70. WIGNER, E. P. (1958) Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, **67**, 325–327.
 71. YE, K. & LIM, L.-H. (2017) Cohomology of cryo-electron microscopy. *SIAM J. Appl. Algebra Geom.*, **1**, 507–535.
 72. YU, Y., WANG, T. & SAMWORTH, R. J. (2014) A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, **102**, 315–323.
 73. ZHAO, Z., SHKOLNISKY, Y. & SINGER, A. (2016) Fast steerable principal component analysis. *ans. Comput. Imaging*, **2**, 1–12.
 74. ZHAO, Z. & SINGER, A. (2013) Fourier–Bessel rotational invariant eigenimages. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, **30**, 871–877.
 75. ZHAO, Z. & SINGER, A. (2014) Rotationally invariant image representation for viewing direction classification in cryo-EM. *J. Struct. Biol.*, **186**, 153–166.

A. Basics on group and representation theory

A group $\mathcal{G}$ is a set with a multiplication operation: $\mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ obeying the following axioms:

1. For any $x, y \in \mathcal{G}$, $xy \in \mathcal{G}$ (closure);
2. For any $x, y, z \in \mathcal{G}$, $(xy)z = x(yz)$ (associativity);
3. There is a unique element of $\mathcal{G}$ denoted e and called the identity for which $ex = xe = x$ for any $x \in \mathcal{G}$;
4. For any $x \in \mathcal{G}$, there is a corresponding element $x^{-1} \in \mathcal{G}$ called the inverse of x , which satisfies $xx^{-1} = x^{-1}x = e$ for any $x \in \mathcal{G}$.

The group operations may not be commutative, i.e., xy is not necessarily equal to yx . This is crucial for our present purposes since three-dimensional rotations do not commute.

We have a group $\mathcal{G}$ acting on a set X . This means that each $g \in \mathcal{G}$ has the corresponding transformations based on a left (group) action $L_g : X \rightarrow X$ and a right (group) action $R_g : X \rightarrow X$. A left (group) action of $\mathcal{G}$ on X is a rule for combining elements $g \in \mathcal{G}$ and elements $x \in X$, denoted by $g \triangleright x$. We additionally require the following three axioms.

1. $g \triangleright x \in X$ for all $x \in X$ and $g \in \mathcal{G}$.
2. $e \triangleright x = x$ for all $x \in X$.
3. $g_2 \triangleright (g_1 \triangleright x) = (g_2 g_1) \triangleright x$ for all $x \in X$ and $g_1, g_2 \in \mathcal{G}$.

A right (group) action of $\mathcal{G}$ on X is a rule for combining elements $g \in \mathcal{G}$ and elements $x \in X$, denoted by $x \triangleleft g$. We additionally require the following three axioms.

1. $x \triangleleft g \in X$ for all $x \in X$ and $g \in \mathcal{G}$.
2. $x \triangleleft e = x$ for all $x \in X$.
3. $(x \triangleleft g_1) \triangleleft g_2 = x \triangleleft (g_1 g_2)$ for all $x \in X$ and $g_1, g_2 \in \mathcal{G}$.

The action of $\mathcal{G}$ on X extends to functions on X as shown in (3.6).

In the paper, we focus on two groups, namely $\text{SO}(2)$ and $\text{SO}(3)$. Both are compact Lie groups and admit irreducible representations. The group $\text{SO}(2)$ is commutative and thus its irreducible representations are one-dimensional complex numbers, $\rho_k(w(\theta)) = e^{ik\theta}$, for $w \in \text{SO}(2)$ with a rotational angle $\theta \in [0, 2\pi)$. The irreducible representations of $\text{SO}(3)$ are given by the Wigner D -matrices, which will be described in the subsection below.

A.1 Wigner's D - and d -matrices

In this section, we recall the definition and relevant properties of the Wigner's D - and d -matrices, which are used extensively in the paper for explicit computations related to the irreducible representations of $\text{SO}(3)$. Recall that elements of $\text{SO}(3)$ are realized as rotation matrices parameterized by *Euler angles* $(\varphi, \vartheta, \psi) \in [0, 2\pi) \times [0, \pi] \times [0, 2\pi)$: each $x \in \text{SO}(3)$ can be explicitly written as

$$x = x(\varphi, \vartheta, \psi) = \begin{pmatrix} \cos \varphi \cos \psi - \sin \varphi \sin \psi \cos \vartheta & -\cos \varphi \sin \psi - \sin \varphi \cos \psi \cos \vartheta & \sin \varphi \sin \vartheta \\ \sin \varphi \cos \psi + \cos \varphi \sin \psi \cos \vartheta & -\sin \varphi \sin \psi + \cos \varphi \cos \psi \cos \vartheta & -\cos \varphi \sin \vartheta \\ \sin \psi \sin \vartheta & \cos \psi \cos \vartheta & \cos \vartheta \end{pmatrix}. \quad (\text{A.1})$$

Note that this is equivalent to writing $x = R_1(\varphi) R_2(\vartheta) R_3(\psi)$, where

$$R_1(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix}, \quad R_2(\vartheta) = \begin{pmatrix} \cos \vartheta & 0 & \sin \vartheta \\ 0 & 1 & 0 \\ -\sin \vartheta & 0 & \cos \vartheta \end{pmatrix}, \quad R_3(\psi) = \begin{pmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The last column in the matrix representation (A.1) is exactly the *view direction* corresponding to $x \in \text{SO}(3)$. For the simplicity of statements, we denote the viewing direction of $x \in \text{SO}(3)$ as

$$\pi(x) = \pi(x(\varphi, \vartheta, \psi)) = (\sin \varphi \sin \vartheta, -\cos \varphi \sin \vartheta, \cos \vartheta)^\top \in \mathbb{R}^3.$$

For each integer $\ell = 0, 1, 2, \dots$, the Wigner's D -matrix $\text{SO}(3) \ni x \mapsto D^\ell(x) \in \mathbb{C}^{(2\ell+1) \times (2\ell+1)}$ is the unique (up to isomorphism) irreducible matrix representation of $\text{SO}(3)$ of index ℓ . For each $x \in \text{SO}(3)$, $D^\ell(x)$ is a $(2\ell+1)$ -by- $(2\ell+1)$ complex Hermitian matrix, of which the entries we denote by $D_{mn}^\ell(x)$ ($-\ell \leq m, n \leq \ell$). As group representations, we have for any $\ell = 0, 1, \dots$ and any $x, x' \in \text{SO}(3)$ the multiplicative formula

$$D^\ell(x') D^\ell(x) = D^\ell(x' \triangleright x). \quad (\text{A.2})$$

The $2\ell+1$ entries in the central column of D^ℓ , i.e., D_{m0}^ℓ ($-\ell \leq m \leq \ell$), give rise to the $2\ell+1$ independent spherical harmonics of degree ℓ . More generally, the $2\ell+1$ entries in the s th column ($-\ell \leq s \leq \ell$) of D^ℓ give rise to the $2\ell+1$ independent spin-weighted spherical harmonics of degree ℓ and weight s [20, 33]. Using the Euler angles, Wigner's D -matrices can be written explicitly as

$$D_{mn}^\ell(\varphi, \vartheta, \psi) := D_{mn}^\ell(x(\varphi, \vartheta, \psi)) = e^{-im\varphi} d_{mn}^\ell(\vartheta) e^{-in\psi}, \quad m, n = -\ell, \dots, \ell \quad (\text{A.3})$$

where matrices $d^\ell(\varphi)$ are known as Wigner's d -matrices. They are real $(2\ell+1)$ -by- $(2\ell+1)$ matrices with an explicit formula for its (m, n) th entry as

$$d_{mn}^\ell(\vartheta) = (-1)^{\ell-n} [(\ell+m)! (\ell-m)! (\ell+n)! (\ell-n)!]^{1/2} \sum_s (-1)^s \frac{\left(\cos \frac{\vartheta}{2}\right)^{m+n+2s} \left(\sin \frac{\vartheta}{2}\right)^{2\ell-m-n-2s}}{s! (\ell-m-s)! (\ell-n-s)! (m+n+s)!}$$

with the sum running over all $s \in \mathbb{Z}$ that make sense of the factorials [50, Section 3.3.2). We will only need the explicit form of d_{mn}^ℓ for the special case $m = n = -\ell$: in this case it is straightforward to verify that the summation consists of only one term $s = 2\ell$, and hence

$$d_{-\ell, -\ell}^\ell(\vartheta) = \left(\cos \frac{\vartheta}{2}\right)^{2\ell} = \left(\cos^2 \frac{\vartheta}{2}\right)^\ell = \left(\frac{1 + \cos \vartheta}{2}\right)^\ell. \quad (\text{A.4})$$

Alternatively, d_{mn}^ℓ can also be written explicitly in terms of Jacobi polynomials as (see, e.g., [50, Section 13.1.1))

$$d_{mn}^\ell(\vartheta) = 2^{-m} \left[\frac{(\ell-m)! (\ell+m)!}{(\ell-n)! (\ell+n)!} \right]^{\frac{1}{2}} (1 - \cos \vartheta)^{\frac{m-n}{2}} (1 + \cos \vartheta)^{\frac{m+n}{2}} P_{\ell-m}^{(m-n, m+n)}(\cos \vartheta) \quad (\text{A.5})$$

where $\{P_n^{(a,b)} : n = 0, 1, 2, \dots\}$ denote the sequence of Jacobi polynomials with parameters a, b [50, Section 13.1.1). This gives rise to the explicit formula for the diagonal entries of the Wigner d -matrices:

$$d_{mm}^\ell(\vartheta) = 2^{-m} (1 + \cos \vartheta)^m P_{\ell-m}^{(0, 2m)}(\cos \vartheta). \quad (\text{A.6})$$

In particular, we see directly from (A.5) that

$$d_{mn}^\ell(0) = \delta_{mn} P_{\ell-m}^{(0, 2m)}(1) = \delta_{mn} \cdot \binom{\ell-m}{\ell-m} = \delta_{mn} \quad (\text{A.7})$$

where δ_{mn} is the Kronecker delta notation

$$\delta_{mn} = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{otherwise.} \end{cases}$$

If the Euler angles of x' take the form $(0, 0, \psi)$, then by (A2) we have

$$D_{mn}^\ell((0, 0, \psi) \triangleright x) = \sum_{s=-\ell}^{\ell} D_{ms}^\ell(0, 0, \psi) D_{sn}^\ell(x) \stackrel{??}{=} \sum_{s=-\ell}^{\ell} d_{ms}^\ell(0) e^{-is\psi} D_{sn}^\ell(x) \stackrel{??}{=} e^{-im\psi} D_{mn}^\ell(x). \quad (\text{A.8})$$

We will need this relation in the proof of Theorem 4.1.

Recall from [66, pp. 21–22) that Euler angles admit physical interpretations for the rotation matrix: If we denote the canonical right-handed orthonormal basis in $\mathbb{R}^3$ by $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, and write $R_{\mathbf{e}_i}(\alpha) \in \text{SO}(3)$ for the rotation around axis $\mathbf{e}_i$ ($i = 1, 2, 3$) by angle α , then rotation by $x(\varphi, \vartheta, \psi) \in \text{SO}(3)$ is equivalent to (i) rotation by angle φ around $\mathbf{e}_3$, (ii) rotation by angle ϑ around the new axis $\mathbf{e}'_2 = R_{\mathbf{e}_3}(\varphi) \mathbf{e}_2$ and (iii) rotation by angle ψ around the new axis $\mathbf{e}'_3 = R_{\mathbf{e}_3}(\varphi) R_{\mathbf{e}_2}(\vartheta) \mathbf{e}_3$. From this geometric interpretation, it is clear that the action of $\text{SO}(2)$ on $\text{SO}(3)$ considered throughout this paper only affects the Euler angle ψ . In other words, under the canonical identification of $\text{SO}(2)$ with $\text{SO}(3)$ elements of the form

$$g = g(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \alpha \in [0, 2\pi) \quad (\text{A.9})$$

then $x(\varphi, \vartheta, \psi) \triangleleft g_r(\alpha) = x(\varphi, \vartheta, \psi + \alpha)$. Together with (A.3), this implies

$$\begin{aligned} D_{mn}^\ell(x(\varphi, \vartheta, \psi) \triangleleft g(\alpha)) &= D_{mn}^\ell(x(\varphi, \vartheta, \psi + \alpha)) \\ &= e^{-in\alpha} D_{mn}^\ell(x(\varphi, \vartheta, \psi)) = \rho_n(g^{-1}) D_{mn}^\ell(x(\varphi, \vartheta, \psi)) \end{aligned} \quad (\text{A.10})$$

where again ρ_n stands for the complex unitary irreducible representation of $\text{SO}(2)$ of character n .

B. Spectral analysis of the local generalized parallel transport operators

We prove Theorem 4.1 to Theorem 4.4 in this appendix.

B.1 Proof of Theorem 4.1

We begin with the isotypic decomposition (3.7) and (3.9). Following (4.3), our strategy is to find a ‘good point’ $x_0 \in \text{SO}(3)$ and a ‘good function’ $u \in \mathcal{H}_{n,-k}$ ($n \geq |k|$) such that $u(x_0) \neq 0$, and evaluate

$$\lambda_n^{(k)}(h) = \frac{(T_h^{(k)} u)(x_0)}{u(x_0)}. \quad (\text{B.1})$$

To this end, pick the following basis for the Lie algebra $\mathfrak{so}(3)$:

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is straightforward to check that these elements satisfy the commutator relations

$$[A_3, A_1] = A_2, \quad [A_2, A_3] = A_1, \quad [A_1, A_2] = A_3.$$

We fix $x_0 = I_3$, the canonical standard orthonormal frame in $\mathbb{R}^3$. We further equip $\text{SO}(3)$ with standard spherical coordinates—the Euler angles—of the form

$$x = x(\varphi, \vartheta, \psi) = x_0 \triangleleft e^{\varphi A_3} e^{\vartheta A_2} e^{\psi A_3}$$

where $(\varphi, \vartheta, \psi) \in (0, 2\pi) \times (0, \pi) \times (0, 2\pi)$, as in [38, Section 3.2.1]. The normalized Haar measure on $\text{SO}(3)$ is given by the density

$$\frac{\sin \theta}{8\pi^2} d\varphi d\vartheta d\psi.$$

Consider the subgroup T_{A_3} of $\text{SO}(3)$ generated by the infinitesimal element A_3 . For every $k \in \mathbb{Z}$ and $n \in \mathbb{N}$ with $n \geq |k|$, the Hilbert space $\mathcal{H}_{n,k}$ admits yet another isotypic decomposition with respect to the left action of T_{A_3} :

$$\mathcal{H}_{n,-k} = \bigoplus_{m=-n}^n \mathcal{H}_{n,-k}^m \quad (\text{B.2})$$

where $s \in \mathcal{H}_{n,-k}^m$ if and only if

$$s(e^{-tA_3} \triangleright x) = e^{imt} s(x) \quad \text{for every } x \in \text{SO}(3) \text{ and } t \in \mathbb{R}. \quad (\text{B.3})$$

As pointed out in [38, Section 3.3.1], elements of $\mathcal{H}_{n,-k}^m$ are often referred to as (generalized) *spherical functions*. In the physics literature, they are also known as *spin-weighted spherical functions*, which are closely related with Wigner D -matrices [10, 12, 20, 34, 51]. We extend the computation in [38, Section 3] to $k > 1$, by fully leveraging properties of the Wigner D -matrices. In fact, we are going to fix $m = -k$ and choose the ‘good function’ u as $D_{-k,-k}^n$, the $(-k, -k)$ th entry of the Wigner D -matrix of weight n , for any $n \geq |k|$ — it is clear from (A.10) that $D_{-k,-k}^n \in \mathcal{H}_{-k}$ for any $n \geq |k|$, and from (A.8) we know that $D_{-k,-k}^n$ satisfies (B3) with $m = -k$. Our goal is to evaluate

$$\lambda_n^{(k)}(h) = \frac{(T_h^{(k)} D_{-k,-k}^n)(x_0)}{D_{-k,-k}^n(x_0)}. \quad (\text{B.4})$$

Now, on the one hand, we have

$$D_{-k,-k}^n(x_0) = D_{-k,-k}^n(0, 0, 0) \stackrel{(A.3)}{=} d_{-k,-k}^n(0) \stackrel{(A.7)}{=} 1. \quad (\text{B.5})$$

On the other hand, note that by the invariance and equivariance of the transport data (3.4) we have for any $x = x(\varphi, \vartheta, \psi) \in \text{SO}(3)$

$$\begin{aligned} T^{(k)}(x_0, x) &= T^{(k)}(x_0, x_0 \triangleleft e^{\varphi A_3} e^{\vartheta A_2} e^{\psi A_3}) = T^{(k)}(x_0, e^{\varphi A_3} \triangleright x_0 \triangleleft e^{\vartheta A_2} e^{\psi A_3}) \\ &= T^{(k)}(e^{-\varphi A_3} \triangleright x_0, x_0 \triangleleft e^{\vartheta A_2} e^{\psi A_3}) = T^{(k)}(x_0 \triangleleft e^{-\varphi A_3}, x_0 \triangleleft e^{\vartheta A_2} e^{\psi A_3}) \\ &= e^{ik\varphi} T^{(k)}(x_0, x_0 \triangleleft e^{\vartheta A_2}) e^{ik\psi}, \end{aligned}$$

and

$$D_{-k,-k}^n(x) = D_{-k,-k}^n(\varphi, \vartheta, \psi) = e^{-ik\varphi} d_{-k,-k}^n(\vartheta) e^{-ik\psi} = e^{-ik\varphi} d_{-k,-k}^n(x_0 \triangleleft e^{\vartheta A_2}) e^{-ik\psi}.$$

Therefore,

$$\begin{aligned}
 \left(T_h^{(k)} D_{-k,-k}^n \right) (x_0) &= \int_{B(x,\alpha)} T^{(k)} (x_0, x) D_{-k,-k}^n (x) \, dx \\
 &= \int_{B(x,\alpha)} T^{(k)} \left(x_0, x_0 \triangleleft e^{\vartheta A_2} \right) D_{-k,-k}^n \left(x_0 \triangleleft e^{\vartheta A_2} \right) \, dx (\varphi, \vartheta, \psi) \\
 &= \int_{B(x,\alpha)} \rho_k \left(T \left(x_0, x_0 \triangleleft e^{\vartheta A_2} \right) \right) D_{-k,-k}^n \left(x_0 \triangleleft e^{\vartheta A_2} \right) \, dx (\varphi, \vartheta, \psi) \\
 &\stackrel{(*)}{=} \int_{B(x,\alpha)} D_{-k,-k}^n \left(\left(x_0 \triangleleft e^{\vartheta A_2} \right) \triangleleft T \left(x_0 \triangleleft e^{\vartheta A_2}, x_0 \right) \right) \, dx (\varphi, \vartheta, \psi) \\
 &\stackrel{(**)}{=} \int_{B(x,\alpha)} D_{-k,-k}^n \left(x_0 \triangleleft e^{\vartheta A_2} \right) \, dx (\varphi, \vartheta, \psi),
 \end{aligned}$$

where $(*)$ used the fact that $D_{-k,-k}^n \in \mathcal{H}_{-k}$, and $(**)$ follows from the definition (3.2) and the geometric fact that $x_0 \triangleleft e^{\vartheta A_2}$ is exactly the parallel transport of x_0 along the unique geodesic connecting $\pi(x_0)$ to $\pi(x_0 \triangleleft e^{\vartheta A_2})$:

$$\left(x_0 \triangleleft e^{\vartheta A_2} \right) \triangleleft T \left(x_0 \triangleleft e^{\vartheta A_2}, x_0 \right) = t_{\pi(x_0 \triangleleft e^{\vartheta A_2}), \pi(x_0)} x_0 = x_0 \triangleleft e^{\vartheta A_2}.$$

It follows that

$$\begin{aligned}
 \left(T_h^{(k)} D_{-k,-k}^n \right) (x_0) &= \int_{B(x,\alpha)} D_{-k,-k}^n \left(x_0 \triangleleft e^{\vartheta A_2} \right) \, dx (\varphi, \vartheta, \psi) \\
 &= \frac{1}{(2\pi)^2} \int_0^{2\pi} d\varphi \int_0^\alpha \frac{\sin \vartheta}{2} D_{-k,-k}^n (0, \vartheta, 0) \, d\vartheta = \int_0^\alpha \frac{\sin \vartheta}{2} d_{-k,-k}^n (\vartheta) \, d\vartheta.
 \end{aligned}$$

Since $d_{-k,-k}^n = d_{k,k}^n$ (see, e.g., [50, formula (3.16)]), this further implies

$$\begin{aligned}
 \left(T_h^{(k)} D_{-k,-k}^n \right) (x_0) &= \int_0^\alpha \frac{\sin \vartheta}{2} d_{k,k}^n (\vartheta) \, d\vartheta \stackrel{(A.6)}{=} 2^{-(k+1)} \int_0^\alpha \sin \vartheta (1 + \cos \vartheta)^k P_{n-k}^{(0,2k)} (\cos \vartheta) \, d\vartheta \\
 &= -2^{-(k+1)} \int_0^\alpha (1 + \cos \vartheta)^k P_{n-k}^{(0,2k)} (\cos \vartheta) \, d \cos \vartheta \\
 &\stackrel{z:=\cos \vartheta}{=} 2^{-(k+1)} \int_{1-h}^1 (1+z)^k P_{n-k}^{(0,2k)} (z) \, dz
 \end{aligned} \tag{B.6}$$

where in the last equality we used $h = 1 - \cos \alpha$. Using the explicit form of Jacobi polynomials (see, e.g., [64, Chapter IV, formula (4.2.1)])

$$\begin{aligned}
 P_{n-k}^{(0,2k)} (z) &= \sum_{\nu=0}^{n-k} \binom{n-k}{n-k-\nu} \binom{n+k}{\nu} \left(\frac{z-1}{2} \right)^\nu \left(\frac{z+1}{2} \right)^{n-k-\nu} \\
 &= \sum_{\nu=0}^{n-k} \binom{n-k}{\nu} \binom{n+k}{\nu} \left(\frac{z-1}{2} \right)^\nu \left(\frac{z+1}{2} \right)^{n-k-\nu}
 \end{aligned}$$

we have

$$\begin{aligned}
\left(T_h^{(k)} D_{-k,-k}^n\right)(x_0) &= 2^{-(k+1)} \cdot 2^k \sum_{\nu=0}^{n-k} \binom{n-k}{\nu} \binom{n+k}{\nu} \int_{1-h}^1 \left(\frac{z-1}{2}\right)^\nu \left(\frac{z+1}{2}\right)^{n-\nu} dz \\
&\stackrel{z:=1-2w}{=} \frac{1}{2} \sum_{\nu=0}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} \int_0^{\frac{h}{2}} w^\nu (1-w)^{n-\nu} \cdot 2 dw \\
&= \sum_{\nu=0}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} \int_0^{\frac{h}{2}} w^\nu (1-w)^{n-\nu} dw \\
&= \sum_{\nu=0}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} B\left(\frac{h}{2}; \nu+1, n-\nu+1\right)
\end{aligned}$$

where $B(x; a, b) = \int_0^x w^{a-1} (1-w)^{b-1} dw$ is the incomplete Beta function. It follows that for all $n \geq |k|$

$$\begin{aligned}
\lambda_n^{(k)}(h) &= \frac{\left(T_h^{(k)} D_{-k,-k}^n\right)(x_0)}{D_{-k,-k}^n(x_0)} = \left(T_h^{(k)} D_{-k,-k}^n\right)(x_0) \\
&= \sum_{\nu=0}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} B\left(\frac{h}{2}; \nu+1, n-\nu+1\right).
\end{aligned}$$

From the integral form of the incomplete Beta function, it is clear that $B(h/2; \nu+1, n-\nu+1)$ is a polynomial of degree $(n+1)$ in h . In particular, by repeatedly applying the recursive relation

$$B(x; a+1, b) = \frac{a}{b} B(x; a, b+1) - \frac{1}{b} x^a (1-x)^b,$$

we easily obtain

$$\lambda_k^{(k)}(h) = B\left(\frac{h}{2}; 1, k+1\right) = \frac{1 - (1-h/2)^{k+1}}{k+1}, \quad (\text{B.7})$$

$$\begin{aligned}
\lambda_{k+1}^{(k)}(h) &= B\left(\frac{h}{2}; 1, k+2\right) - (2k+1) B\left(\frac{h}{2}; 2, k+1\right) \\
&= B\left(\frac{h}{2}; 1, k+2\right) - \frac{2k+1}{k+1} B\left(\frac{h}{2}; 1, k+2\right) + \frac{2k+1}{k+1} \left(\frac{h}{2}\right) \left(1 - \frac{h}{2}\right)^{k+1} \\
&= \frac{2(k+1)(1 - (1-h/2)^{k+2})}{k+2} - \frac{(2k+1)(1 - (1-h/2)^{k+1})}{k+1}, \quad (\text{B.8})
\end{aligned}$$

$$\begin{aligned}
\lambda_{k+2}^{(k)}(h) &= B\left(\frac{h}{2}; 1, k+3\right) - 4(k+1) B\left(\frac{h}{2}; 2, k+2\right) + (k+1)(2k+1) B\left(\frac{h}{2}; 3, k+1\right) \\
&= B\left(\frac{h}{2}; 1, k+3\right) - 2B\left(\frac{h}{2}; 2, k+2\right) - (2k+1) \left(\frac{h}{2}\right)^2 \left(1 - \frac{h}{2}\right)^{k+1} \\
&= \frac{k}{k+2} B\left(\frac{h}{2}; 1, k+3\right) + \frac{2}{k+2} \left(\frac{h}{2}\right) \left(1 - \frac{h}{2}\right)^{k+2} - (2k+1) \left(\frac{h}{2}\right)^2 \left(1 - \frac{h}{2}\right)^{k+1} \\
&= \frac{k}{k+2} \cdot \frac{1 - (1 - h/2)^{k+3}}{k+3} + \frac{2}{k+2} \left(\frac{h}{2}\right) \left(1 - \frac{h}{2}\right)^{k+2} - (2k+1) \left(\frac{h}{2}\right)^2 \left(1 - \frac{h}{2}\right)^{k+1},
\end{aligned} \tag{B.9}$$

which give rise to (4.7) and (4.8).

It now remains to compute a quadratic approximation for $\lambda_n^{(k)}(h)$ for $h \rightarrow 0$, for all $n \geq |k|$. This can be done by direct computation using the integral form of the incomplete beta function: for all $n \geq |k|$,

$$\lambda_n^{(k)}(0) = 0, \tag{B.10}$$

$$\partial_h \lambda_n^{(k)}(0) = \sum_{\nu=0}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} \cdot \frac{1}{2} \left(\frac{h}{2}\right)^\nu \left(1 - \frac{h}{2}\right)^{n-\nu} \Big|_{h=0} = \frac{1}{2}, \tag{B.11}$$

$$\begin{aligned}
\partial_h^2 \lambda_n^{(k)}(0) &= \sum_{\nu=1}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} \cdot \frac{\nu}{4} \left(\frac{h}{2}\right)^{\nu-1} \left(1 - \frac{h}{2}\right)^{n-\nu} \Big|_{h=0} \\
&\quad + \sum_{\nu=0}^{n-k} (-1)^\nu \binom{n-k}{\nu} \binom{n+k}{\nu} \cdot \frac{-(n-\nu)}{4} \left(\frac{h}{2}\right)^\nu \left(1 - \frac{h}{2}\right)^{n-\nu-1} \Big|_{h=0} \\
&= -\frac{1}{4} (n-k)(n+k) - \frac{n}{4} = -\frac{1}{4} (n^2 + n - k^2)
\end{aligned} \tag{B.12}$$

and (4.6) follows from the Taylor expansion

$$\lambda_n^{(k)}(h) = \lambda_n^{(k)}(0) + h \partial_h \lambda_n^{(k)}(0) + \frac{h^2}{2} \partial_h^2 \lambda_n^{(k)}(0) + O(h^3).$$

This completes the entire proof of Theorem 4.1. □

B.2 Proof of Theorem 4.2

LEMMA B.1

1. There exists $h_1^{(k)} \in (0, 2]$ such that $\lambda_n^{(k)}(h) \leq \lambda_k^{(k)}(h)$ for every $n \geq k+1$ and $h \in (0, h_1^{(k)})$.
2. There exists $h_2^{(k)} \in (0, 2]$ such that $\lambda_n^{(k)}(h) \leq \lambda_{k+1}^{(k)}(h)$ for every $n \geq k+2$ and $h \in (0, h_2^{(k)})$.

Proof of Lemma B.1. Since $\lambda_n^{(k)}(0) = 0$ for all $k \in \mathbb{Z}$ and $n \geq |k|$, we will just compare the first-order derivatives $\partial_h \lambda_n^{(k)}(h)$ over an interval with 0 as the left end point. By (B.6), $\partial_h \lambda_n^{(k)}(h)$ admits a closed

form expression in terms of Jacobi polynomials:

$$\partial_h \lambda_n^{(k)}(h) = \frac{1}{2} \left(1 - \frac{h}{2}\right)^k P_{n-k}^{(0,2k)}(1-h) \stackrel{h=\cos \alpha}{=} \frac{1}{2} \left(\frac{1+\cos \alpha}{2}\right)^k P_{n-k}^{(0,2k)}(\cos \alpha).$$

In particular, under change-of-coordinates $h = 1 - \cos \alpha$, we have

$$\begin{aligned} \partial_h \lambda_k^{(k)}(h) &= \frac{1}{2} \left(\frac{1+\cos \alpha}{2}\right)^k P_0^{(0,2k)}(\cos \alpha) = \frac{1}{2} \left(\frac{1+\cos \alpha}{2}\right)^k, \\ \partial_h \lambda_{k+1}^{(k)}(h) &= \frac{1}{2} \left(\frac{1+\cos \alpha}{2}\right)^k P_1^{(0,2k)}(\cos \alpha) = \frac{1}{2} \left(\frac{1+\cos \alpha}{2}\right)^k [1 - (k+1)(1 - \cos \alpha)]. \end{aligned}$$

It is clear that $0 < \partial_h \lambda_{k+1}^{(k)}(h) < \partial_h \lambda_k^{(k)}(h)$ for all $h = 1 - \cos \alpha \in (0, 1/(k+1)]$, which together with $\lambda_k^{(k)}(0) = \lambda_{k+1}^{(k)}(0)$, gives rise to

$$\lambda_{k+1}^{(k)}(h) \leq \lambda_k^{(k)}(h) \quad \text{for all } 0 < h \leq \frac{1}{k+1}. \quad (\text{B.13})$$

With (B.13), the proof of both (1) and (2) of Lemma B.1 is reduced to only the part (2) of Lemma B.1. The remaining of this proof is devoted to establishing (2) of Lemma B.1.

By the classical result of Szegő [64, Theorem 8.21.13], there exists a fixed positive number $c > 0$ such that

$$P_{n-k}^{(0,2k)}(\cos \theta) = \frac{1}{\sqrt{n}} k(\theta) \left[\cos(N\theta + \gamma) + (n \sin \theta)^{-1} O(1) \right], \quad \text{for all } \frac{c}{n} \leq \theta \leq \pi - \frac{c}{n} \quad (\text{B.14})$$

where

$$\begin{aligned} k(\theta) &= \frac{1}{\sqrt{\pi \sin(\theta/2) \cos(\theta/2)} \cdot [\cos(\theta/2)]^{2k}} = \left(\frac{2}{1+\cos \theta}\right)^k \sqrt{\frac{2}{\pi \sin \theta}}, \\ N &= n + \frac{2k+1}{2}, \quad \lambda = -\frac{\pi}{4}. \end{aligned}$$

In particular, by making the $O(1)$ term in (B.14) explicit, we have for some absolute constant $C > 0$

$$\left(\frac{1+\cos \theta}{2}\right)^k \left| P_{n-k}^{(0,2k)}(\cos \theta) \right| \leq \sqrt{\frac{2}{n\pi}} \cdot \frac{1}{\sqrt{\sin \theta}} \left(1 + \frac{C}{n \sin \theta}\right) \quad \text{for all } \frac{c}{n} \leq \theta \leq \pi - \frac{c}{n}. \quad (\text{B.15})$$

Note that the left-hand side is precisely the absolute value of $2\partial_h \lambda_n^{(k)}(h) = 2\partial_h \lambda_n^{(k)}(1 - \cos \theta)$. We seek an upper bound for the right-hand side of (B.15) that holds uniformly for all sufficiently large n . To this end, consider the largest zero of $P_{n-k}^{(0,2k)}(x)$ for $x \in [-1, 1]$, denoted as $x_{n-k}^* = \cos \alpha_{n-k}^*$ (thus, α_{n-k}^* is the smallest zero of the function $\alpha \mapsto P_{n-k}^{(0,2k)}(\cos \alpha)$ on $\alpha \in [0, \pi]$). Well-known estimates for the extreme zero of Jacobi polynomials (see, e.g., [19, Section 2.2]) indicates

$$x_{n-k}^* > 1 - O\left(\frac{1}{n^2}\right) \text{ as } n \rightarrow \infty \quad \Rightarrow \quad \alpha_{n-k}^* \rightarrow 0 \text{ as } n \rightarrow \infty$$

thus for any $\epsilon_1 > 0$ there exists $N_1 > 0$ such that for all sufficiently large $n \geq N_1$ we have

$$\sin \alpha_{n-k}^* \geq (1 - \epsilon_1) \alpha_{n-k}^* \quad (\text{B.16})$$

since $\lim_{x \rightarrow 0} (\sin x) / x = 1$. In the meanwhile, [22, Theorem 3.1) bounds $x_{n-k}^* = \cos \alpha_{n-k}^*$ from above by

$$x_{n-k}^* < \frac{\left(2k + \frac{1}{2}\right)^2 + 4(n-k)\left(n + k + \frac{1}{2}\right)}{(2n - 2k + 1 + 2k)^2} = \frac{4n^2 + 2n + 1/4}{(2n + 1)^2} = 1 - \frac{2n + 3/4}{(2n + 1)^2}. \quad (\text{B.17})$$

Using the elementary inequality $1 - x^2/2 \leq \cos x$ for $x \in [0, 2]$, (B.17) leads to

$$1 - \frac{(\alpha_{n-k}^*)^2}{2} \leq \cos \alpha_{n-k}^* = x_{n-k}^* < 1 - \frac{2n + 3/4}{(2n + 1)^2} \Rightarrow \alpha_{n-k}^* > \sqrt{\frac{4n + 3/2}{(2n + 1)^2}} \rightarrow \frac{1}{\sqrt{n}} \quad \text{as } n \rightarrow \infty,$$

which further implies (1) for sufficiently large n , $\alpha_{n-k}^* \in [c/n, \pi - c/n]$, and (2) by choosing n sufficiently large we can ensure for the same arbitrary $\epsilon_1 > 0$ chosen for (B.16), that, in addition to (B.16), there holds

$$\alpha_{n-k}^* > \frac{1 - \epsilon_1}{\sqrt{n}}. \quad (\text{B.18})$$

Now consider the smallest local extremum μ_{n-k}^* of the function $P_{n-k}^{(0,2k)}(\cos \alpha)$ for $\alpha \in [0, \pi]$ that is larger than α_{n-k}^* , i.e.,

$$\mu_{n-k}^* := \min \left\{ \alpha \in [0, \pi] \mid \partial_\alpha P_{n-k}^{(0,2k)}(\cos \alpha) = 0 \text{ and } \alpha \geq \alpha_{n-k}^* \right\}$$

which by (B.18) is guaranteed to fall within $[c/n, \pi - c/n]$. For any $n \geq N_0$, by (B.15), (B.16) and (B.18), we have

$$\begin{aligned} \left(\frac{1 + \cos \mu_{n-k}^*}{2} \right)^k \left| P_{n-k}^{(0,2k)}(\cos \mu_{n-k}^*) \right| &\leq \sqrt{\frac{2}{\pi n}} \cdot \frac{1}{\sqrt{\sin \mu_{n-k}^*}} \left(1 + \frac{C}{n \sin \mu_{n-k}^*} \right) \\ &\leq \sqrt{\frac{2}{\pi n}} \cdot \frac{1}{\sqrt{\sin \alpha_{n-k}^*}} \left(1 + \frac{C}{n \sin \alpha_{n-k}^*} \right) \\ &\leq \sqrt{\frac{2}{\pi n}} \cdot \frac{1}{\sqrt{(1 - \epsilon_1) \alpha_{n-k}^*}} \left(1 + \frac{C}{n (1 - \epsilon_1) \alpha_{n-k}^*} \right) \\ &< \frac{1}{(1 - \epsilon_1) n^{\frac{1}{4}}} \sqrt{\frac{2}{\pi}} \left(1 + \frac{C}{(1 - \epsilon_1)^2 \sqrt{n}} \right). \end{aligned}$$

The same inequality holds if we replace μ_{n-k}^* with any other extremum of the function $\alpha \mapsto P_{n-k}^{(0,2k)}(\cos \alpha)$ in $\alpha \in [c/n, \pi - c/n]$. In particular, this implies that for all sufficiently large n we have (recalling that $h = 1 - \cos \alpha$)

$$\partial_h \lambda_n^{(k)}(h) = \partial_h \lambda_n^{(k)}(1 - \cos \alpha) \leq \frac{1}{2} \left(\frac{1 + \cos \alpha}{2} \right)^k \left| P_{n-k}^{(0,2k)}(\cos \alpha) \right| < \frac{1}{4} \quad \text{for all } a \in [c/n, \pi - c/n].$$

The rest of the proof follows easily from the proof of [38, Theorem 4]: let $a_0 \in (0, \pi)$ be such that

$$\partial_h \lambda_{k+1}^{(k)}(h) = \partial_h \lambda_{k+1}^{(k)}(1 - \cos \alpha) = \frac{1}{2} \left(\frac{1 + \cos \alpha}{2} \right)^k [1 - (k+1)(1 - \cos \alpha)] > \frac{1}{4}$$

for all $\alpha < \alpha_0$ and sufficiently large n ; the remaining finitely cases can be verified directly as claimed in [38, Section A.2.1, pp. 612). Note that such a value α_0 exists because when $\alpha = 0$ (i.e., $h = 1$)

$$\partial_h \lambda_{k+1}^{(k)}(1) = \frac{1}{2} > \frac{1}{4}.$$

As argued in [38, Section A.2, pp. 611), we set $z_0 = \cos \alpha_0$ and $h_1^{(k)} = h_2^{(k)} = 1 + z_0$, which ensures $\partial_h \lambda_n^{(k)}(z) \leq \partial_h \lambda_{k+1}^{(k)}(z)$ for all $z \in [-1, z_0]$, and furthermore $\lambda_n^{(k)}(h) \leq \lambda_{k+1}^{(k)}(h)$, for all $n \geq k+1$. This proves (2) of Lemma B.1 and completes the entire proof of Lemma B.1. $\square$

LEMMA B.2

1. There exists $N_1^{(k)} > 0$ such that $\lambda_n^{(k)}(h) \leq \lambda_k^{(k)}(h)$ for every $n \geq N_1^{(k)}$ and $h \in [h_1^{(k)}, 2]$.
2. There exists $N_2^{(k)} > 0$ such that $\lambda_n^{(k)}(h) \leq \lambda_k^{(k)}(h)$ for every $n \geq N_2^{(k)}$ and $h \in [h_2^{(k)}, 1/(k+1)]$.

Proof of Lemma B.2. First note, on the one hand, that the Schatten 2-norm of $T_h^{(k)}$ can be easily computed: By [55, Theorem VI.23),

$$\begin{aligned} \|T_h^{(k)}\|_2^2 &= \int_{\text{SO}(3)} \int_{\text{SO}(3)} |T_h^{(k)}(x, y)|^2 dx dy = \int_{\text{SO}(3)} \int_{B(y, \alpha)} |T_h^{(k)}(x, y)|^2 dx dy \\ &= \int_{\text{SO}(3)} \int_{B(y, \alpha)} dx dy = \int_0^\alpha \frac{\sin \vartheta}{2} d\vartheta = \frac{1 - \cos \vartheta}{2} = \frac{h}{2} \end{aligned}$$

where the last equality follows from $h = 1 - \cos \alpha$. On the other hand,

$$\|T_h^{(k)}\|_2^2 = \sum_{n=k}^{\infty} (2n+1) (\lambda_n^{(k)})^2$$

which gives the same bound as [38, formula (A.4)]:

$$\lambda_n^{(k)}(h) \leq \frac{\sqrt{h}}{\sqrt{4n+2}}.$$

Since by (B.7) we have

$$\lambda_k^{(k)}(h) = \frac{1 - (1 - h/2)^{k+1}}{k+1} \geq \frac{1 - (1 - h_1^{(k)}/2)^{k+1}}{k+1} \quad \text{for all } h \in [h_1^{(k)}, 2],$$

it is straightforward to verify by direct computation that there exists $N_1^{(k)} > 0$ such that $\sqrt{h}/\sqrt{4n+2} \leq \lambda_k^{(k)}(h)$ for every $n \geq N_1^{(k)}$ and $h \in [h_1^{(k)}, 2]$. This proves (1) of Lemma B.2. Furthermore, by (B.8)

$$\lambda_{k+1}^{(k)}(h) = -\frac{k}{k+1} \cdot \frac{1 - (1 - h/2)^{k+2}}{k+2} + \frac{2k+1}{k+1} \left(\frac{h}{2}\right) \left(1 - \frac{h}{2}\right)^{k+1}$$

a direct computation for the derivative of the left-hand side with respect to h gives

$$\partial_h \lambda_{k+1}^{(k)}(h) = \frac{1}{2} [1 - (k+1)h] \left(1 - \frac{h}{2}\right)^k$$

from which it is easy to directly verify that $h \mapsto \lambda_{k+1}^{(k)}(h)$ achieves its maximum at $h = 1/(k+1)$ over $h \in [0, 2]$, and $\lambda_{k+1}^{(k)}(h) > 0$ for all $h \in [0, 1/(k+1)]$. It is then easy to verify by direct computation that there exists $N_2^{(k)}$ such that $\sqrt{h}/\sqrt{4n+2} \leq \lambda_{k+1}^{(k)}(h)$ for every $n \geq N_2^{(k)}$ and $h \in [h_2^{(k)}, 1/(k+1)]$. This proves (2) of Lemma B.2. $\square$

Proof of Theorem 4.2. Direct computation using (B.7) and (B.8) establishes (4.11):

$$\begin{aligned} G^{(k)}(h) &= \lambda_k^{(k)}(h) - \lambda_{k+1}^{(k)}(h) = \frac{1 - (1 - h/2)^{k+1}}{k+1} + \frac{(2k+1)(1 - (1 - h/2)^{k+1})}{k+1} \\ &\quad - \frac{2(k+1)(1 - (1 - h/2)^{k+2})}{k+2} = \frac{2 - (1 - h/2)^{k+1}((k+1)h + 2)}{(k+2)}. \end{aligned}$$

Unsurprisingly, the spectral gap depends on the ‘frequency channel’ parameter $k \in \mathbb{N}$. The rest of the proof follows verbatim the proof of [38, Theorem 4]: by Lemma B.1 and Lemma B.2, we have $\lambda_n^{(k)} \leq \lambda_k^{(k)}(k)$ for every $n \geq N_1^{(k)}$ and $h \in [0, 2]$, as well as $\lambda_n^{(k)} \leq \lambda_{k+1}^{(k)}(k)$ for every $n \geq N_2^{(k)}$ and $h \in [0, 1/(k+1)]$. We then verify directly both $\lambda_n^{(k)} \leq \lambda_k^{(k)}(h)$ over $h \in [0, 2]$ and $\lambda_n^{(k)} \leq \lambda_{k+1}^{(k)}(h)$ over $h \in [0, 1/(k+2)]$ for the finitely many cases left ($k \leq n \leq N_1^{(k)}$ and $k+1 \leq n \leq N_2^{(k)}$, respectively). $\square$

B.3 Proof of Theorem 4.3

Our proof extends the arguments in the proof of [38, Theorem 5). A key observation is that the top eigenvector $\mathcal{H}(\lambda_k^{(k)}(h))$ coincides with the isotypic subspace $\mathcal{H}_{k,-k}$ (see Section 4.1). Consider the morphism $\omega := \sqrt{1/(2k+1)} \cdot \tau : \mathbb{C}^{2k+1} \rightarrow \mathcal{H}$ defined as

$$\omega(v)(x) = \left(\delta_x^{(k)}\right)^*(v).$$

Part 1: τ is an isomorphism between $\mathbb{C}^{2k+1}$ and $\mathcal{H}_{k,-k}$. We first show that $\text{Im}(\omega) \subset \mathcal{H}_{k,-k}$, namely, for any $x \in \text{SO}(3)$, $v \in \mathbb{C}^{2k+1}$, and $g \in \text{SO}(2)$ there holds

$$\left(\delta_{x \triangleleft g}^{(k)}\right)^*(v) = \rho_k(g^{-1}) \left(\delta_x^{(k)}\right)^*(v). \quad (\text{B.19})$$

To this end, note that for any $z \in \mathbb{C}$ we have

$$\begin{aligned} \left\langle \left(\delta_{x \triangleleft g}^{(k)}\right)^*(v), z \right\rangle_{\mathbb{C}} &= \left\langle v, \delta_{x \triangleleft g}^{(k)}(z) \right\rangle_{\mathbb{C}^3} = \left\langle v, z D_{\cdot, -k}^k(x \triangleleft g) \right\rangle_{\mathbb{C}^3} \\ &\stackrel{(\text{A.10})}{=} \left\langle v, z \rho_{-k}(g^{-1}) D_{\cdot, -k}^k(x) \right\rangle_{\mathbb{C}^3} = \left\langle v, z \rho_k(g) D_{\cdot, -k}^k(x) \right\rangle_{\mathbb{C}^3} \\ &= \left\langle \rho_k(g^{-1}) v, z D_{\cdot, -k}^k(x) \right\rangle_{\mathbb{C}^3} = \left\langle \rho_k(g^{-1}) v, \delta_x^{(k)}(z) \right\rangle_{\mathbb{C}^3} \\ &= \left\langle \rho_k(g^{-1}) \left(\delta_x^{(k)}\right)^*(v), z \right\rangle_{\mathbb{C}^3} \end{aligned}$$

which proves (B.19). Next, we show that ω is a morphism of $\mathrm{SO}(3)$ representations, namely, for any $x \in \mathrm{SO}(3)$, $v \in \mathbb{C}^{2k+1}$, and $g \in \mathrm{SO}(3)$ there holds

$$\left(\delta_x^{(k)}\right)^* \left(D^k(g)v\right) = \left(\delta_{g^{-1} \triangleright x}^{(k)}\right)^* (v). \quad (\text{B.20})$$

To this end, again for any arbitrary $z \in \mathbb{C}$

$$\begin{aligned} \left\langle \left(\delta_x^{(k)}\right)^* \left(D^k(g)v\right), z \right\rangle_{\mathbb{C}} &= \left\langle D^k(g)v, \delta_x^{(k)}(z) \right\rangle_{\mathbb{C}^3} = \left\langle D^k(g)v, z D_{\cdot, -k}^k(x) \right\rangle_{\mathbb{C}^3} \\ &= \left\langle v, z D^k(g^{-1}) D_{\cdot, -k}^k(x) \right\rangle_{\mathbb{C}^3} \stackrel{(\text{??})}{=} \left\langle v, z D_{\cdot, -k}^k(g^{-1} \triangleright x) \right\rangle_{\mathbb{C}^3} \\ &= \left\langle v, \delta_{g^{-1} \triangleright x}^{(k)}(z) \right\rangle_{\mathbb{C}^3} = \left\langle \left(\delta_{g^{-1} \triangleright x}^{(k)}\right)^* (v), z \right\rangle_{\mathbb{C}} \end{aligned}$$

which proves (B.20). It now follows immediately that the morphism ω maps $\mathbb{C}^{2k+1}$ isomorphically, as a unitary representation of $\mathrm{SO}(3)$, onto $\mathcal{H}_{k, -k}$, the unique isotypical component in $\mathcal{H}_{-k}$ (by (B.19)) of unitary irreducible $\mathrm{SO}(3)$ -representation of dimension $2k+1$. This in turn implies that ω (and thus τ) is an isomorphism between Hermitian vector spaces. It remains to determine the suitable normalization constant; we show that $\mathrm{Tr}(\tau^* \circ \tau) = 2k+1$. Indeed,

$$\begin{aligned} \mathrm{Tr}(\tau^* \circ \tau) &= (2k+1) \mathrm{Tr}(\omega^* \circ \omega) = (2k+1) \int_{\mathbb{C}S^{2k}} \langle \omega^* \circ \omega(v), v \rangle_{\mathcal{H}} dv = (2k+1) \int_{\mathbb{C}S^{2k}} \langle \omega(v), \omega(v) \rangle_{\mathcal{H}} dv \\ &= (2k+1) \int_{\mathbb{C}S^{2k}} \int_{\mathrm{SO}(3)} \left\langle \left(\delta_x^{(k)}\right)^* (v), \left(\delta_x^{(k)}\right)^* (v) \right\rangle_{\mathbb{C}} dx dv \\ &= (2k+1) \int_{\mathbb{C}S^{2k}} \int_{\mathrm{SO}(3)} \left\langle \left(D_{\cdot, -k}^k(x)\right)^* v, \left(D_{\cdot, -k}^k(x)\right)^* v \right\rangle_{\mathbb{C}} dx dv \\ &= (2k+1) \int_{\mathbb{C}S^{2k}} \int_{(3)} 1 dx dv = 2k+1 \end{aligned}$$

where $\mathbb{C}S^{2k}$ is the $(4k+1)$ -dimensional unit sphere in $\mathbb{C}^{2k+1}$ and dv is the unique normalized Haar measure on $\mathbb{C}S^{2k}$.

Part 2: Proof of (4.14). By (B.19), we have $(\mathrm{ev}_x | \mathbb{W}^{(k)}) \circ \omega = \left(\delta_x^{(k)}\right)^*$, which is equivalent to $\left(\varphi_x^{(k)}\right)^* \circ \tau = \left(\delta_x^{(k)}\right)^*$. The conclusion now follows from the straightforward computation as in the proof of [38, Theorem 5):

$$\left(\varphi_x^{(k)}\right)^* \circ \tau = \left(\delta_x^{(k)}\right)^* \Rightarrow \left(\varphi_x^{(k)}\right)^* \circ (\tau \circ \tau^*) = \left(\delta_x^{(k)}\right)^* \circ \tau^* \Rightarrow \left(\varphi_x^{(k)}\right)^* = \left(\delta_x^{(k)}\right)^* \circ \tau^* \Rightarrow \tau \circ \delta_x^{(k)} = \varphi_x^{(k)}.$$

This completes the entire proof.

B.4 Proof of Theorem 4.4

By Theorem 4.3, τ is a morphism between Hermitian inner product spaces $\mathbb{C}^{2k+1}$ and $\mathbb{W}^{(k)}$ and (4.14) holds, thus by the same argument in the last step of the proof of [38, Theorem 6) it suffices to prove that for any unit-norm complex numbers $v, u \in \mathbb{C}$ there holds

$$\left| \left\langle \delta_x^{(k)}(u), \delta_y^{(k)}(v) \right\rangle_{\mathbb{C}^{2k+1}} \right| = \left(\frac{\langle \pi(x), \pi(y) \rangle + 1}{2} \right)^k. \quad (\text{B.21})$$

This boils down to the following straightforward computation:

$$\begin{aligned} \left| \left\langle \delta_x^{(k)}(u), \delta_y^{(k)}(v) \right\rangle_{\mathbb{C}^{2k+1}} \right| &= \left| u \left(D_{\cdot, -k}^k(x) \right)^\top \left(D_{\cdot, -k}^k(y) \right)^* \bar{v} \right| = \left| D_{-k, -k}^k(x^{-1}y) \right| \\ &\stackrel{(A.3)}{=} \left| d_{-k, -k}^k(\vartheta(x^{-1}y)) \right| \stackrel{(A.4)}{=} \left(\frac{1 + \cos \vartheta(x^{-1}y)}{2} \right)^k \end{aligned} \quad (\text{B.22})$$

where $\vartheta(x^{-1}y)$ is the Euler angle ϑ of $x^{-1}y \in \text{SO}(3)$. Recall from (A.1) that $\cos \vartheta(x^{-1}y)$ is exactly the $(3, 3)$ -entry of the 3-by-3 matrix form of $x^{-1}y \in \text{SO}(3)$, which is exactly identical to the inner product of the third columns of the matrix forms of x and y , i.e.,

$$\cos \vartheta(x^{-1}y) = \langle \pi(x), \pi(y) \rangle.$$

Plugging this back into the rightmost term of (B.22) completes the entire proof.