Generalized Laplacian in Spectral Seriation on Manifold Data

Ruizi Wu
Department of Statistics and Data Science
National University of Singapore
Singapore, 119077
e1330364@u.nus.edu

Wanjie Wang
Department of Statistics and Data Science
Department of Mathematics
National University of Singapore
Singapore, 119077
wanjie.wang@nus.edu.sg

Jinchi Lv
Data Sciences and Operations Department
USC Marshall School of Business
Los Angeles, California 90089
jinchilv@marshall.usc.edu

Abstract

Noisy high-dimensional observations may have conditional means lying on a smooth closed curve, indexed by *temporal labels* or *phases*. We want to recover the latent phases for each observation and thus retrieve a circular ordering. Spectral seriation uses the graph Laplacian's spectral embedding to recover the phases and ordering. We consider generalized graph Laplacians in spectral seriation, where a parameter α controls the degree normalization and affects the accuracy and stability of both tasks. We develop and compare three generalized Laplacian constructions: a shifted generalized Laplacian, a graph-difference Laplacian, and an extra-normalized generalized Laplacian. Among these, the extra-normalized variant is less sensitive to the choice of α and achieves the strongest overall recovery performance. For this variant, we establish a theoretical error bound relating phase estimation accuracy to eigenspace perturbation, accounting for the inherent rotation and reflection ambiguity. In synthetic experiments across various scenarios, our new graph Laplacian outperforms the classic spectral seriation method and other seriation methods. On the Pink Angel data, which records photos from different viewpoints, our method achieves zero ordering error and the lowest phase recovery error. On FUCCI cell-cycle data, our method performs the best when all 6 cells are mixed, demonstrating the robustness to heterogeneity.

1 Introduction

Many high-dimensional datasets have an underlying one-dimensional progression. Examples include images indexed by viewing angles [28], single-cell expression profiles along the cell cycle [12], and cryo-EM images associated with molecular motion [21]. Although each observation is high-dimensional, it lies on a one-dimensional manifold indexed by angles, cell cycle stage, or molecular motion stage. We call these indices *phases*. Consider the periodic settings where the possible phases form a smooth closed curve. Our goal is to estimate these latent phases of noisy observations and recover their circular ordering. These two tasks are related but distinct, and both are important for downstream scientific analysis.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

To illustrate this structure, consider images of an object rotating through 360°. Each image is represented by $Z_i \in \mathbb{R}^p$, and we write $Z_i = X(t_i) + e_i$, where t_i is the viewing angle, $X(t_i)$ is the underlying noiseless image, and e_i is observation noise. As the angle varies, the noiseless images trace a one-dimensional curve $\mathcal{M} = \{X(t) : t \in [0, 2\pi)\}$ in image space. A full rotation returns to the same image, so the curve is closed. When X is smooth and regular without self-intersections, $\mathcal{M}$ forms a smooth closed manifold. A similar structure arises in cell-cycle data, where gene expression profiles change as cells progress through the cycle. By representing each cell's progression state with a latent phase t_i, we can estimate its position within the cycle from the observed gene expression profile and infer its cell-cycle stage.

This geometric structure suggests recovering phases by mapping the underlying curve onto a circle. When the phase is parameterized by rescaled arc length, the first nonconstant eigenspace of the Laplace–Beltrami operator (the manifold Laplacian) on $\mathcal{M}$ is spanned by $\cos(t)$ and $\sin(t)$. These functions provide circular coordinates, allowing phases to be identified up to rotation and reflection. Spectral seriation uses graph-Laplacian eigenvectors to recover the ordering of observations. Building on this idea, [17] use a two-dimensional spectral embedding to estimate latent phases and recover their circular ordering. Let $S = (S_{ij})_{n\times n}$ be an affinity matrix constructed from the observations, and $D = \mathrm{diag}(\sum_{j=1}^{n} S_{ij} : i \in [n])$ be the degree matrix. The standard normalized graph Laplacian is $L = I - D^{-1/2}SD^{-1/2}$. Under some regularity conditions, the eigenspace spanned by its second and third eigenvectors approximates the sine–cosine eigenspace [7, 24]. The angular coordinates of this embedding yield phase estimates, which are then sorted to recover circular ordering.

The standard normalized graph Laplacian uses the degree-normalized affinity $D^{-1/2}SD^{-1/2}$. A natural generalization is to consider $S_\alpha = D^{-\alpha}SD^{-\alpha}$, where $\alpha \geq 0$ controls the strength of degree correction [8]. The choices $\alpha = 0$ and $\alpha = 1/2$ recover the original affinity matrix and the standard symmetric normalization, respectively. Degree-based normalization has been studied in manifold learning and diffusion maps [7, 25], while recent work on generalized Laplacian matrices has further highlighted the statistical role of α in spectral inference [8]. Varying α changes how pairwise similarities are weighted by node degrees and can therefore alter the resulting spectral embedding. Since phase estimates and circular ordering are obtained from this embedding, the choice of α may affect the accuracy and stability of both recovery tasks.

However, the generalized affinity S_α alone does not determine how the graph Laplacian should be constructed. In our closed-curve setting, phase recovery relies on a two-dimensional spectral embedding whose associated eigenspace approximates $\mathrm{span}\{\cos(t), \sin(t)\}$. A direct construction such as $I - D^{-\alpha}SD^{-\alpha}$ need not preserve the spectral structure of the standard normalized Laplacian, and the desired eigenspace alignment is not guaranteed by degree correction alone. The key difficulty is thus not whether degree correction can be introduced, but how it should be incorporated into a Laplacian construction. This leads to a central question of this paper: *which generalized Laplacian construction is suitable for latent phase recovery and circular ordering?*

To address this question, we investigate three generalized Laplacian constructions based on S_α: a shifted generalized Laplacian, a graph-difference Laplacian, and an extra-normalized generalized Laplacian. The first two incorporate S_α directly into different Laplacian-type operators, whereas the third treats S_α as a reweighted affinity matrix and further normalizes it by its own degree matrix. This additional normalization preserves the standard normalized-Laplacian structure while allowing α to control the degree correction. We demonstrate that this extra-normalized construction provides a natural generalized analog of the standard normalized Laplacian for latent phase recovery and circular ordering on closed curves.

1.1 Our contributions

The major contributions of our paper are threefold. First, we formulate generalized spectral seriation as an operator-construction problem, rather than a direct replacement of $D^{-1/2}SD^{-1/2}$ by $D^{-\alpha}SD^{-\alpha}$. The same generalized affinity can be converted into several plausible Laplacian-type matrices, but these matrices need not preserve the eigenspace carrying the latent labels t_i. This identifies a structural issue absent from the standard generalized-affinity analysis and fixed-normalization spectral seriation.

Second, we propose three Laplacian-type constructions and identify the extra-normalized operator as the appropriate generalized analog of L for phase recovery and circular ordering. The key novelty is the separation of two roles: α modifies the affinity matrix, while the final normalization restores the

Laplacian geometry required. We establish convergence of the expected eigenspace to the discretized sine–cosine eigenspace and derive a conditional phase estimation error bound. This explains why the extra-normalized operator preserves the angular embedding and provides theoretical support for phase recovery.

Third, we investigate the role of the parameter α in phase recovery and circular ordering. Our simulations show that the extra-normalized generalized Laplacian is less sensitive to α than the other two constructions. Moreover, its recovery accuracy can still benefit from a suitable choice of α, motivating us to suggest a practical value of α. With this choice, the method achieves strong recovery performance on synthetic data, the best viewing-angle recovery on image data, and competitive cell-cycle ordering on FUCCI data.

Together, our results add theoretical and empirical depth beyond fixed-normalization spectral seriation [17]. They unveil which generalized operator preserves the relevant two-dimensional phase eigenspace, why seemingly natural alternatives can lose the phase geometry, and how degree correction can be incorporated without sacrificing the angular recovery mechanism.

1.2 Related works

The seriation problem has a long history. Early work formulated seriation as a combinatorial problem [1, 15]. Such formulations are statistically natural, but exact optimization is computationally expensive [5, 13]. Spectral relaxations provide a more scalable alternative through graph Laplacian eigenvectors [2, 9]. Recent statistical works have studied convex relaxations for permutation problems [10], permutation recovery under monotone matrix models [18, 19], noisy matrix reordering [6], and minimax-optimal seriation [16]. These methods primarily focus on recovering the ordering of observations and are powerful when the affinity matrix has an ordering-compatible pattern. Beyond ordering, [11] study latent position estimation on a one-dimensional torus from pairwise affinities. More closely related to our setting, [17] use a two-dimensional graph-Laplacian embedding to recover latent phases from noisy manifold observations and then obtain their circular ordering. Building on this approach, we investigate how generalized degree normalization and different Laplacian constructions affect both phase recovery and circular ordering.

A second viewpoint treats the observations as samples from a low-dimensional manifold, so that recovering t_i becomes an embedding problem. Laplacian eigenmaps and diffusion maps use graph Laplacian eigenvectors to construct nonlinear low-dimensional embeddings from local similarities [3, 7, 4]. [22] discusses the embedding dimensions with a latent manifold structure. Isomap approximates geodesic distances by shortest paths on a neighborhood graph [26], while t-SNE and UMAP are used widely for visualization and exploratory analysis [27, 20]. These methods can reveal low-dimensional structure, but their embeddings do not directly yield latent phase labels. For a one-dimensional closed manifold, the trigonometric structure of the Laplace–Beltrami eigenfunctions provides a connection between spectral embeddings and latent phases.

Generalized graph Laplacians and degree-normalized affinities form a third line of related work. The family $S_\alpha = D^{-\alpha} S D^{-\alpha}$ allows the strength of degree correction to vary through α [8]. Initial applications of the asymptotic theory in [8] include graph neural networks and Laplacian-regularized graph smoothing. A closely related construction appears in diffusion maps, where degree-reweighted affinities are further row-normalized to form a Markov transition operator [7]. For phase recovery and circular ordering, the relevant question, however, is not only how to define a generalized affinity, but how to convert that affinity into an operator whose leading nonconstant eigenspace approximates $\cos(t)$ and $\sin(t)$.

2 Problem setup and spectral seriation

Consider observations

$$Z_i = X(t_i) + e_i, \quad i \in \{1, 2, \ldots, n\}, \tag{1}$$

where $t_i \in [0, 2\pi)$ is the latent phase associated with observation i, $X : [0, 2\pi] \to \mathbb{R}^p$ is the underlying closed curve with $X(0) = X(2\pi)$, and $e_i \in \mathbb{R}^p$ is an additive noise with $\mathbb{E}[e_i \mid t_i] = 0$. Our goal is to recover the latent phases t_i and the corresponding circular ordering $\boldsymbol{\pi} : [n] \to [n]$ satisfying $t_{\boldsymbol{\pi}(1)} < t_{\boldsymbol{\pi}(2)} < \cdots < t_{\boldsymbol{\pi}(n)}$.

Spectral seriation uses graph-Laplacian eigenvectors to recover latent orderings [9]. For closed curves, [17] use a two-dimensional spectral embedding to estimate latent phases, recover their circular ordering, and establish theoretical guarantees under observation noise. Following this approach, we first construct an affinity matrix S from the observations. Given a bandwidth parameter $h > 0$, we define S using the Gaussian kernel

$$S_{ij} = k(Z_i, Z_j; h), \quad \text{where } k(x, y; h) = \exp\{-\|x - y\|_2^2/2h^2\}/\sqrt{2\pi h^2}. \tag{2}$$

Denote $D = \mathrm{diag}(\sum_{j=1}^n S_{ij} : i \in [n])$ the diagonal degree matrix. The baseline method uses the standard normalized graph Laplacian

$$L = I - D^{-1/2} S D^{-1/2}. \tag{3}$$

The spectral embedding is motivated by the eigenfunctions of the manifold Laplacian on the underlying curve. We assume that X is smooth, regular, and non-self-intersecting. The latent phase t is taken to be a rescaled arc-length parameter so that $\|X'(t)\|_2$ is constant. With this parameterization, the first nonconstant eigenspace of the manifold Laplacian is two-dimensional, spanned by $\cos(t)$ and $\sin(t)$. Let F_+ and F_- be the eigenvectors associated with the second and third smallest eigenvalues of L. Under some regularity conditions, the eigenspace spanned by F_+ and F_- approximates the aforementioned eigenspace. More discussion and illustrations can be found in Appendix A.

Motivated by the ideal embedding $(\cos(t_i), \sin(t_i))$, we normalize each embedded point to unit length and estimate its latent phase through

$$\cos(\hat{t}_i) = \frac{F_+(i)}{\left(F_+^2(i) + F_-^2(i)\right)^{1/2}}, \quad \sin(\hat{t}_i) = \frac{F_-(i)}{\left(F_+^2(i) + F_-^2(i)\right)^{1/2}}.$$

Equivalently, $\hat{t}_i$ is the angular coordinate of $(F_+(i), F_-(i))$. Sorting the estimated phases then yields the estimated circular ordering $\hat{\pi}$. Implementation details are given in Algorithm 1.

Algorithm 1 Baseline spectral seriation for closed-loop temporal recovery

Input: Observations $Z_1, \ldots, Z_n$; kernel bandwidth h.
Output: Estimated phases $\hat{t}_1, \ldots, \hat{t}_n \in [0, 2\pi)$ and circular ordering $\hat{\pi}$.

1: Construct the affinity matrix S using (2).
2: Form the degree matrix D and the normalized graph Laplacian $L = I - D^{-1/2} S D^{-1/2}$.
3: Find eigenvectors F_+ and F_- associated with the second and third smallest eigenvalues of L.
4: Find $\hat{t}_i \in [0, 2\pi)$ so that $\cos(\hat{t}_i) = \frac{F_+(i)}{\sqrt{(F_+(i))^2 + (F_-(i))^2}}$, $\sin(\hat{t}_i) = \frac{F_-(i)}{\sqrt{(F_+(i))^2 + (F_-(i))^2}}$.
5: Sort the estimated phases to obtain $\hat{\pi}$ such that $\hat{t}_{\hat{\pi}(i)}$ is increasing in i.
6: Return $\hat{t} = (\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_n)$ and $\hat{\pi}$.

Since the first nonconstant manifold-Laplacian eigenvalue has multiplicity two, an orthonormal basis for its eigenspace is determined only up to an orthogonal transformation $Q \in \mathbb{O}(2)$. This induces a global rotation and reflection ambiguity in the recovered phases $\hat{t}_i$. Consequently, the circular ordering is identifiable only up to a cyclic shift and reversal.

3 Generalized Laplacian for spectral seriation

The baseline spectral seriation method uses the normalized graph Laplacian $L = I - D^{-1/2} S D^{-1/2}$. Following the generalized degree normalization [8], we consider the generalized affinity matrix

$$S_\alpha = D^{-\alpha} S D^{-\alpha}, \quad \alpha \geq 0. \tag{4}$$

The parameter α controls the strength of degree correction: $S_0 = S$, while $S_{1/2} = D^{-1/2} S D^{-1/2}$. However, the generalized affinity alone does not specify a Laplacian operator, and different constructions can lead to different spectral embeddings.

For spectral seriation, the object of interest is a Laplacian-type operator whose two-dimensional nontrivial eigenspace can approximate $\mathrm{span}\{\cos(t), \sin(t)\}$. We therefore investigate how the same generalized affinity S_α can be incorporated into different Laplacian constructions for phase and circular ordering recovery.

3.1 Three Laplacian variants

We consider three Laplacian variants derived from S_α. Each represents a different way to construct a Laplacian-type operator from the generalized affinity.

Shifted generalized Laplacian. Our first variant shifts the generalized affinity to obtain a positive semidefinite matrix:

$$L_d(\alpha) = d_{\max}(\alpha) I - S_\alpha, \tag{5}$$

where $d_{\max}(\alpha) = \max_{i\in[n]} D_{ii}^{1-2\alpha}$. Indeed, $L_d(\alpha) = [d_{\max}(\alpha) I - D^{1-2\alpha}] + D^{-\alpha}(D-S)D^{-\alpha}$ is the sum of two positive semidefinite matrices. This variant preserves the eigenspaces of the direct construction $I - S_\alpha$ while ensuring positive semidefiniteness. When $\alpha = 1/2$, it coincides with the baseline Laplacian L.

Graph-difference Laplacian. Our second variant preserves the intuition that a Laplacian should encode the difference between node degree and pairwise affinity:

$$L_g(\alpha) = D^{-\alpha}(D-S)D^{-\alpha} = D^{1-2\alpha} - S_\alpha. \tag{6}$$

It reduces to the unnormalized graph Laplacian at $\alpha = 0$, and to the baseline Laplacian at $\alpha = 1/2$.

Extra-normalized generalized Laplacian. Our third variant is motivated by [25]. We treat S_α as a reweighted affinity matrix and normalize it using its own degree matrix, $\widetilde{D}_\alpha = \mathrm{diag}(S_\alpha \mathbf{1})$. Define

$$L_{ng}(\alpha) = I - \widetilde{D}_\alpha^{-1/2} S_\alpha \widetilde{D}_\alpha^{-1/2}. \tag{7}$$

This construction retains the symmetric normalized-Laplacian form, with $L_{ng}(0) = L$.

The three variants use the same generalized affinity S_α, but differ in how this affinity enters the final operator. These differences can affect the nontrivial eigenspace used for phase recovery, especially when node degrees vary.

To illustrate the role of extra normalization, consider a simplified case $D = dI$, where $d > 0$. Then $S_\alpha = d^{-2\alpha} S$ and $\widetilde{D}_\alpha = d^{1-2\alpha} I$. Consequently,

$$L_d(\alpha) = L_g(\alpha) = d^{1-2\alpha} L, \quad L_{ng}(\alpha) = L.$$

Thus, all three variants share the same eigenspaces in this simplified case. Extra normalization also removes the degree scaling introduced by α. and recovers the baseline Laplacian L. When node degrees are heterogeneous, however, the three variants may have different eigenspaces. Section 4 studies the extra-normalized construction under our sampling and noise assumptions.

***Remark* 1.** *Diffusion maps [7] constructs low-dimensional embeddings from the transition matrix $P_\alpha = \widetilde{D}_\alpha^{-1} D^{-\alpha} S D^{-\alpha}$. Thus, our extra-normalized generalized Laplacian is the symmetric normalized Laplacian associated with this diffusion operator. Our focus is on how α affects phase and circular ordering recovery under different Laplacian generalizations.*

3.2 Generalized spectral seriation

We now describe a unified phase-recovery pipeline for the three variants. For a fixed α and $\star \in \{d, g, ng\}$, let $F_{\star,+}, F_{\star,-} \in \mathbb{R}^n$ be eigenvectors associated with the second and third smallest eigenvalues of $L_\star(\alpha)$. As in the baseline method, we recover $\hat{t}$ from the angular coordinates:

$$\cos(\hat{t}_i) = F_{\star,+} \left(F_{\star,+}^2(i) + F_{\star,-}^2(i)\right)^{-\frac{1}{2}}, \quad \sin(\hat{t}_i) = F_{\star,-}(i) \left(F_{\star,+}^2(i) + F_{\star,-}^2(i)\right)^{-\frac{1}{2}}.$$

Sorting $\hat{t}$ yields the recovered circular ordering $\hat{\pi}$. Algorithm 2 summarizes the procedure.

Although Algorithm 2 applies to all three variants, our theoretical analysis focuses on the extra-normalized generalized Laplacian $L_{ng}(\alpha)$. Section 4 studies the convergence of the expected eigenspace and derives a conditional phase estimation error bound for the extra-normalized generalized Laplacian.

4 Theoretical guarantees

To study the convergence of the phase estimator $\hat{t}$ obtained by Algorithm 2 with L_{ng}, we relate its estimation error to the convergence of the two-dimensional eigenspace toward the target cosine–sine

Algorithm 2 Temporal label recovery via generalized Laplacians

Input: Observations $Z_1, \dots, Z_n$; kernel bandwidth h; parameter α; method $\star \in \{d, g, ng\}$.
Output: Estimated phases $\hat{t} = (\hat{t}_1, \hat{t}_2, \dots, \hat{t}_n) \in [0, 2\pi)$ and circular ordering $\hat{\pi}$.

Construct S using (2), its degree matrix D, and $S_\alpha = D^{-\alpha} S D^{-\alpha}$.
if $\star = d$ **then**
    Set $L_\star(\alpha) = d_{\max}(\alpha) I - S_\alpha$.
else if $\star = g$ **then**
    Set $L_\star(\alpha) = D^{-\alpha}(D - S)D^{-\alpha}$.
else if $\star = ng$ **then**
    Set $\widetilde{D}_\alpha = \mathrm{diag}(S_\alpha \mathbf{1})$ and $L_\star(\alpha) = I - \widetilde{D}_\alpha^{-1/2} S_\alpha \widetilde{D}_\alpha^{-1/2}$.
end if
Find eigenvectors F_+ and F_- associated with the second and third smallest eigenvalues of $L_\star(\alpha)$.
Find $\hat{t}_i \in [0, 2\pi)$ so that $\cos(\hat{t}_i) = \frac{F_+(i)}{\sqrt{(F_+(i))^2 + (F_-(i))^2}}$, $\sin(\hat{t}_i) = \frac{F_-(i)}{\sqrt{(F_+(i))^2 + (F_-(i))^2}}$.
Return $\hat{t} = (\hat{t}_1, \hat{t}_2, \dots, \hat{t}_n)$ and $\hat{\pi}$.

space. Instead of analyzing the observed Laplacian directly, we decompose the affinity matrix S into its conditional expectation and a residual noise component. We first establish eigenspace convergence for the Laplacian constructed from the conditional expectation, and then derive a conditional phase estimation error bound in terms of the observed-versus-expected Laplacian perturbation.

4.1 Definitions

Recall $Z_i = X(t_i) + e_i$ from (1) and the Gaussian affinity matrix $S \in \mathbb{R}^{n \times n}$ (2). For the theory, we assume $t_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, 2\pi)$ and independent Gaussian noise $e_i \sim N(0, \sigma^2 I_p)$, together with the curve and bandwidth conditions in Appendix B.1. We decompose $S = H + W$, where $H = \mathbb{E}[S \mid t]$ and $W = S - H$. The entries of H satisfy: for $i = j$, $H_{ij} = (2\pi h^2)^{-1/2}$; and for $i \neq j$,

$$H_{ij} = \frac{1}{\sqrt{2\pi h^2}} \left(\frac{h^2}{h^2 + 2\sigma^2} \right)^{\frac{p}{2}} \exp\left\{ -\frac{\|X(t_i) - X(t_j)\|_2^2}{2(h^2 + 2\sigma^2)} \right\}.$$

For simplicity, denote $\tau := h^2 + 2\sigma^2$ and $C_{h,\tau} := h^p \tau^{-p/2} / \sqrt{2\pi h^2}$. The diagonal entries can also be written as $H_{ii} = C_{h,\tau}(1 + \rho_{h,\tau,p})$, where $\rho_{h,\tau,p} := \tau^{p/2} h^{-p} - 1$ measures their discrepancy from the Gaussian form of the off-diagonal entries. This discrepancy contributes an additional term to the eigenspace error.

Definition 1. *For any symmetric nonnegative matrix $A \in \mathbb{R}^{n \times n}$ with positive degrees, let $D_A = \mathrm{diag}(A\mathbf{1})$ be the diagonal degree matrix. The α-normalized matrix is $\widetilde{A}_\alpha := D_A^{-\alpha} A D_A^{-\alpha}$, with degree matrix $\widetilde{D}_{\alpha,A} = \mathrm{diag}(\widetilde{A}_\alpha \mathbf{1})$. Define the extra-normalized Laplacian $L^\star_{\alpha,A} := I - \widetilde{D}_{\alpha,A}^{-1/2} \widetilde{A}_\alpha \widetilde{D}_{\alpha,A}^{-1/2}$, and let $F_{\alpha,A} \in \mathbb{R}^{n \times 2}$ have orthonormal columns spanning the eigenspace associated with its second and third smallest eigenvalues.*

As mentioned in Section 2, t is identifiable only up to rotation and reflection. To account for such ambiguity, we define the timestamp estimation error of $\hat{t}$ as

$$\mathrm{Error}(\hat{t}, t) := \min_{r \in \{-1,1\}, \theta \in [0, 2\pi]} \frac{1}{\sqrt{n}} \left\| [rt + \theta - \hat{t}\,]_{2\pi} \right\|_2, \tag{8}$$

where $[a]_{2\pi}$ denotes the representative of a modulo 2π in $[-\pi, \pi]$, r accounts for reflection, and θ represents a global angular shift.

4.2 Main results

We state our main results under Assumptions 1–2, given in Appendix B.1. Here $c_{\mathrm{gap}} > 0$ denotes the eigengap constant, with the target eigenspace separated from the remaining spectrum by at least $c_{\mathrm{gap}} \tau$. We first establish convergence of the expected eigenspace to the sampled trigonometric eigenspace.

Proposition 1 (Expected eigenspace convergence)**.** *Assume that Assumptions 1–2 hold. Define the rows of* $\Psi_n \in \mathbb{R}^{n\times 2}$ *as* $(\Psi_n)_i = (\cos(t_i), \sin(t_i))$*, and denote by* $\Phi_n = \Psi_n(\Psi_n^\top \Psi_n)^{-1/2}$*. Let* $\alpha \in [0,1]$*. Then we have that for all sufficiently large* n*, with probability* $1 - O(n^{-2})$*,*

$$\left\| F_{\alpha,H} F_{\alpha,H}^\top - \Phi_n \Phi_n^\top \right\|_2 \le C(1+\alpha)\left(\tau + \frac{\sqrt{\log n}}{n^{1/2}\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right). \tag{9}$$

Equivalently, there exists some $Q \in \mathbb{O}(2)$ *such that with probability* $1 - O(n^{-2})$*,*

$$\|F_{\alpha,H} Q - \Phi_n\|_F \le C(1+\alpha)\left(\tau + \frac{\sqrt{\log n}}{n^{1/2}\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right). \tag{10}$$

The proof of Proposition 1 is provided in Appendix B.4. Under Assumption 2, the error bound vanishes as $n \to \infty$. This establishes convergence of the expected eigenspace to the sampled trigonometric eigenspace up to an orthogonal transformation.

Theorem 1 (Estimator error bound)**.** *Assume that Assumption 1–2 hold. Let* $\alpha \in [0,1]$*, and assume further that* $\|L^\star_{\alpha,S} - L^\star_{\alpha,H}\|_2 \le U_S$ *for some nonnegative number* U_S*. Then we have that with probability* $1 - O(n^{-2})$*,*

$$\mathrm{Error}(\hat{t}, t) \le C\left[(1+\alpha)\left(\tau + \frac{\log(n)^{1/2}}{n^{1/2}\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right) + \frac{U_S}{c_{\mathrm{gap}}\tau}\right]. \tag{11}$$

The proof of Theorem 1 is provided in Appendix B.5. Theorem 1 separates the expected eigenspace error in Proposition 1 from the observed perturbation term $U_S/(c_{\mathrm{gap}}\tau)$. Since the target two-dimensional eigenspace has a gap of order τ, Davis–Kahan converts the operator perturbation into an eigenspace bound. Its eigenvectors cannot be compared individually: an orthonormal basis may rotate or reflect. We therefore use Procrustes alignment before converting the embedding error into circular phase error. The difference between Φ_n and the scaled circular embedding $\sqrt{2/n}\Psi_n$ is absorbed into the finite-sample term. In particular, the bound implies phase consistency if the errors in Proposition 1 vanish and $U_S = o(\tau)$, provided the perturbation event holds with probability tending to one.

***Remark* 2.** *Theorem 1 is conditional on* $\|L^\star_{\alpha,S} - L^\star_{\alpha,H}\|_2 \le U_S$*, whose probability is not bounded here. If* $\mathbb{P}(\|L^\star_{\alpha,S} - L^\star_{\alpha,H}\|_2 \le U_S) \ge 1 - \delta_n$*, the phase-error bound holds with probability at least* $1 - O(n^{-2}) - \delta_n$*. Establishing such a bound requires addressing two challenges: affinity entries sharing an observation are dependent, even conditional on the latent phases; and their fluctuations propagate through both stages of degree normalization. This requires joint control of the affinity and degree perturbations. Such an analysis is beyond the scope of the present paper and will be pursued in subsequent work.*

***Remark* 3.** *When the diagonal correction and observed perturbation terms are negligible compared with the remaining terms, the error bound is dominated by* $\tau + \frac{\sqrt{\log(n)}}{\sqrt{n}\tau^{5/4}}$*. Balancing these terms gives* $\tau \asymp (\log(n)/n)^{2/9}$ *and a corresponding conditional error bound of order* $(\log(n)/n)^{2/9}$*, provided the theorem assumptions hold.*

5 Numerical experiments

We consider a synthetic closed-loop manifold in $\mathbb{R}^p$. The underlying manifold is defined as $X(t) = Qv(t) \in \mathbb{R}^p$, where $v(t) \in \mathbb{R}^r$ is a low-dimensional vector and $Q \in \mathbb{R}^{p\times r}$ embeds it into p-dimensional space with $Q^\top Q = I_r$. We generate the data points as

$$Z_i = Qv(t_i) + s(t_i)e_i, \quad t_i \sim \mathrm{Unif}(0, 2\pi), \quad 1 \le i \le n,$$

where $e_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_p)$ and $s(t)$ introduces heteroscedastic noise. Additional experiments with non-uniformly sampled latent phases are provided in Appendix E.

We use two metrics to evaluate recovery. The timestamp estimation error $\mathrm{Error}(\hat{t}, t)$ measures the ℓ_2 discrepancy between the true and estimated timestamps under optimal rotational or reflection-rotational alignment, as defined in (8). We also introduce an ordering error to assess recovery of the circular ordering, enabling comparisons with methods that do not directly estimate timestamps:

$$\mathrm{Error}(\hat{\pi}, \pi) = \min_{\substack{r\in\{\pm 1\}\\ k\in\{0,\dots,N-1\}}} \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d_N(\hat{\pi}(i)+k, r\pi(i))}{N}\right)^2\right]^{1/2}, \tag{12}$$

where $d_N(a, b) := \min_{j\in\mathbb{Z}} |a - b + jN|$.

5.1 The role of α in three variants

We compare our three Laplacian variants $L_d(\alpha)$, $L_g(\alpha)$, and $L_{ng}(\alpha)$, and examine their sensitivity to α. Fix sample size $n = 500$ and dimensionality $p = 40$. Let $v(t) = (1.5\cos(t), \sin(t), \cos(2t), \sin(2t))^\top$ and $s(t) = 1 + \eta\cos(t - \pi/3)$. We consider three settings: a) baseline: $\eta = 0.3$, $\sigma = 0.10$; b) large noise with baseline heterogeneity: $\eta = 0.3$, $\sigma = 0.15$; c) baseline noise with large heterogeneity: $\eta = 0.6$, $\sigma = 0.10$. For each setting, we vary $\alpha \in \{-1, -0.9, \ldots, 2\}$, including negative values for readers who may be interested in this regime. Figure 1 reports the mean estimation error with standard deviation among 100 repetitions.

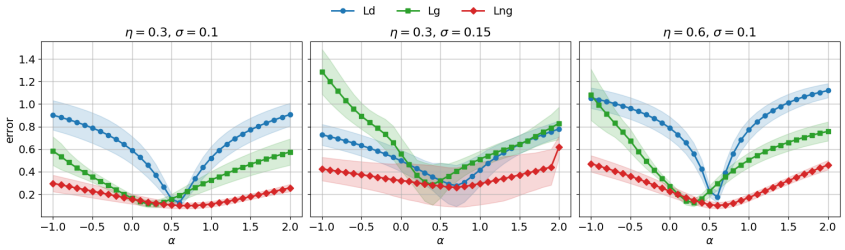


Figure 1: Mean estimation error versus α for three generalized Laplacians: L_d (blue), L_g (green), and L_{ng} (red). The shaded region represents the standard deviation over 100 repetitions.

As shown in Figure 1, the choice of α substantially affects the estimation error of all three methods. While L_d exhibits a sharp minimum and its error increases rapidly as α deviates from the optimum, L_{ng} maintains relatively low estimation error over a much broader range of α. This advantage is particularly evident under larger heterogeneity, where the other two variants become more sensitive to the choice of α. Although L_{ng} does not always achieve the lowest error at every α, it generally performs well even when α is not optimally selected. These results demonstrate that extra normalization improves robustness to the choice of α, reducing the need for precise parameter tuning. Additional comparisons across sample sizes, bandwidths, and similarity constructions are provided in Appendices D.1–D.3.

We further examine the empirically optimal α for each Laplacian variant. Fix the heterogeneity parameter $\eta = 0.6$ and take the noise level $\sigma = \{0.050, 0.075, \ldots, 0.150\}$. For each σ, the optimal α is recorded for each repetition. Table 1 reports the median and interquartile range of these optimal α values among 100 repetitions for each σ.

Table 1: Median optimal α with (25% quantile, 75% quantile).

Fix $\eta = .60$	$\sigma = .050$	$\sigma = .075$	$\sigma = .100$	$\sigma = .125$	$\sigma = .150$
$L_d(\alpha)$	.500 (.500, .500)	.500 (.500, .500)	.600 (.600, .600)	.600 (.600, .600)	.700 (.600, .700)
$L_g(\alpha)$	.200 (.200, .300)	.200 (.200, .300)	.300 (.200, .300)	.300 (.300, .400)	.400 (.300, .400)
$L_{ng}(\alpha)$	.600 (.500, .600)	.600 (.500, .600)	.600 (.500, .700)	.600 (.500, .725)	.650 (.500, .800)

Table 1 shows that the optimal α for L_d and L_g is tightly concentrated within each noise setting, consistent with their greater sensitivity to α observed in Figure 1. In contrast, L_{ng} exhibits a wider spread of optimal α values while maintaining low estimation error over a broad range, demonstrating greater robustness to parameter tuning. Moreover, its median optimal α remains close to 0.6 across noise levels, motivating $\alpha = 0.6$ as an empirical default. Results over broader ranges of η and α further support this choice (Appendix C.1).

5.2 Comparison with existing methods

We compare the proposed $L_{ng}(0.6)$ with the spectral-seriation baseline $L_d(0.5)$ [17] and diffusion maps with $\alpha = 0.5$ [25]. For diffusion maps, we estimate latent phases from the angular coordinates of its two-dimensional spectral embedding and obtain the ordering by sorting the estimated phases. We also consider three manifold learning methods: t-SNE [27], UMAP [20], and Isomap [26].

However, their embeddings do not directly represent latent phases. By taking a one-dimensional embedding, we can obtain an estimated ordering by sorting its coordinates. Therefore, we evaluate these three methods only in terms of ordering error, while the spectral methods are evaluated using both ordering and timestamp estimation errors. Implementation details and tuning parameters are provided in Appendix C.3.

We set sample size $n = 500$, dimensionality $p = 100$. Let $v(t) = (\cos t, \sin t)^\top$ and $s(t) = 1 + 0.3\cos t$. We consider $\sigma \in \{0.05, 0.10, 0.15, 0.20\}$ and report means and standard deviations over 100 repetitions. Table 2 presents ordering errors for all six methods and timestamp estimation errors for the three spectral methods.

The left panel of Table 2 shows that the three spectral seriation methods achieve substantially lower ordering errors than t-SNE, UMAP, and Isomap at every noise level. Our method $L_{ng}(0.6)$ closely matches the best ordering performance, with only small differences among the three spectral methods. For timestamp recovery, the right panel on Table 2 shows that $L_{ng}(0.6)$ achieves the lowest mean error at every noise level. Its advantage over DM is modest, while the improvement over Baseline is more pronounced at lower noise levels. Together, these results show that $L_{ng}(0.6)$ combines near-best ordering recovery with the best timestamp estimation performance among the compared methods.

Table 2: Mean ordering and timestamp estimation errors (with standard deviation).

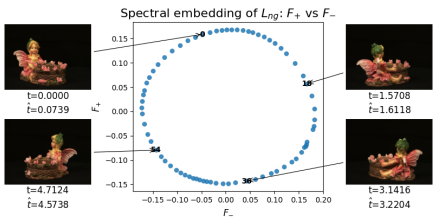
	Ordering Error						Timestamp Estimation Error		
σ	t-SNE	UMAP	Isomap	Baseline	DiffMap	$L_{ng}(0.6)$	Baseline	DiffMap	$L_{ng}(0.6)$
.05	.1762 (.0294)	.1260 (.0664)	.1965 (.0058)	**.0091** (.0005)	**.0091** (.0005)	**.0091** (.0005)	.0890 (.0193)	.0619 (.0060)	**.0608** (.0056)
.10	.1823 (.0181)	.0708 (.0652)	.1893 (.0071)	**.0173** (.0008)	**.0173** (.0008)	**.0173** (.0008)	.1229 (.0125)	.1095 (.0055)	**.1084** (.0050)
.15	.1581 (.0344)	.0704 (.0412)	.1739 (.0076)	**.0257** (.0012)	**.0257** (.0012)	**.0257** (.0012)	.1677 (.0088)	.1601 (.0059)	**.1596** (.0060)
.20	.1391 (.0335)	.1121 (.0182)	.1741 (.0072)	**.0345** (.0016)	**.0345** (.0017)	**.0345** (.0017)	.2179 (.0083)	**.2138** (.0083)	**.2138** (.0087)

6 Real Data

6.1 Pink Angel data recovery

We further apply the phase recovery methods to the Pink Angel photo dataset (`https://aloi.science.uva.nl/`). This dataset contains $n = 72$ images observed from equally spaced viewing angles $t_i = 2\pi(i-1)/72$, $i = \{1, \ldots, 72\}$. Our goal is to recover both the circular ordering and the viewing angle of each image. Preprocessing details and the full set of images are provided in Appendix C.4. All methods use the same preprocessed feature matrix, and we fix $\alpha = 0.6$ for L_{ng} and $\alpha = 0.5$ for both Baseline and Diffusion Map.

The leading eigenvectors F_+ and F_- are visualized in Figure 2(a). The two-dimensional embedding (F_-, F_+) forms an approximately circular pattern, consistent with the expected structure



(a) Recovered embedding

Method	Ordering Error	Timestamp Estimation Error
t-SNE	.1857	/
UMAP	.1873	/
Isomap	.1927	/
Baseline	**0.0**	.1339
Diffusion Map	**0.0**	.1042
$L_{ng}(0.6)$	**0.0**	**.1004**

(b) Estimation error

Figure 2: Phase recovery on image data. (a) Two-dimensional spectral embedding obtained by $L_{ng}(0.6)$, with representative images labeled by their true and estimated viewing angles. (b) Ordering and timestamp estimation errors for all methods.

$(\cos(t), \sin(t))$. Images at four representative angles show that the estimated angle $\hat{t}$ matches closely the true angle t, indicating that our L_{ng} recovers the circular ordering successfully. In the right panel of Figure 2, we further compare the ordering and angle recovery error. All three spectral seriation methods recover the circular ordering exactly. For angle estimation, $L_{ng}(0.6)$ achieves the lowest error. These results show that our method recovers the correct image ordering while providing the most accurate viewing-angle estimates among the compared methods. Additional results on image-angle recovery using tangent distance are provided in Appendix D.4.

6.2 FUCCI cell-cyle Recovery

We evaluate our method on the FUCCI cell-cycle dataset (GSE121265; [14]), using gene expression profiles from 888 human induced pluripotent stem cells across six cell lines. FUCCI-derived phases serve as references for evaluating the recovered cyclic order. We analyze each cell line separately and all six jointly. Preprocessing details are provided in Appendix C.5. We fix $\alpha = 0.6$ for L_{ng} and $\alpha = 0.5$ for both Baseline and Diffusion Map.

As shown in Table 3, spectral seriation methods achieve the lowest circular rank MSE in all six individual cell lines. Our $L_{ng}(0.6)$ achieves the lowest error on NA19101 and ties with diffusion maps on NA18855 at the reported precision, while the baseline performs best on the remaining four lines. On the pooled dataset, $L_{ng}(0.6)$ achieves the lowest error among all compared methods, highlighting its advantage in handling heterogeneous data.

Table 3: ordering error on the FUCCI dataset. The lowest error in each column is shown in bold.

Method	NA18511	NA18855	NA18870	NA19098	NA19101	NA19160	All Data
t-SNE	.2602	.2644	.2569	.2500	.2503	.2560	.2584
UMAP	.2572	.2658	.2623	.2378	.2575	.2637	.2676
Isomap	.2580	.2632	.2629	.2284	.2615	.2606	.2649
Baseline	**.2416**	.2508	**.2458**	**.2270**	.2495	**.2424**	.2592
Diffusion Map	.2425	**.2505**	.2459	.2303	.2467	.2450	.2567
$L_{ng}(0.6)$	.2429	**.2505**	.2459	.2320	**.2460**	.2469	**.2556**

7 Conclusion

We investigate the role of the degree parameter α across three generalized Laplacian variants: shifted, graph-difference, and extra-normalized. Our results show that the effect of α depends critically on how the generalized Laplacian is constructed. In particular, the extra-normalized variant is less sensitive to the choice of α and exhibits more stable performance in our experiments. Across synthetic and real datasets, L_{ng} achieves ordering errors comparable to other spectral methods and lower than manifold-learning baselines, while attaining the best or near-best phase estimation accuracy.

Despite these benefits, our method still relies on the underlying closed-curve geometry. Abrupt geometric changes or near-overlaps between distant parts of the trajectory may distort the graph's neighborhood structure and compromise phase recovery. Our current theory provides a conditional perturbation-to-phase guarantee, while bounding the perturbation between the observed and expected Laplacians remains necessary for full consistency. Moreover, our theoretical analysis assumes uniform sampling, although the additional experiments in Appendix E suggest that L_{ng} can still achieve accurate recovery under non-uniform sampling with a suitable choice of α. Future work will focus on completing the consistency analysis, extending it to non-uniform sampling, and developing data-adaptive choices of α.

Acknowledgements

Ruizi Wu acknowledges support through the NUS Research Scholarship and conference travel funding from the Department of Statistics and Data Science, Faculty of Science, National University of Singapore. Dr. Wanjie Wang was supported by the Singapore Ministry of Education Academic Research Fund Tier 1 under Grant A-8001451-00-00 and Grant A-8003581-00-00. We thank the Institute for Mathematical Sciences (IMS) at the National University of Singapore for organizing the

workshop Parametric and Nonparametric Network Inference, which initiated this project, and the organizers of the Second Workshop on Mathematical Data Science for facilitating further discussions.

References

[1] Phipps Arabie and Lawrence J. Hubert. An overview of combinatorial data analysis. In Phipps Arabie, Lawrence J. Hubert, and Geert De Soete, editors, *Clustering and Classification*, pages 5–63. World Scientific, River Edge, NJ, 1996.

[2] Jonathan E. Atkins, Erik G. Boman, and Bruce Hendrickson. A spectral algorithm for seriation and the consecutive ones problem. *SIAM Journal on Computing*, 28(1):297–310, 1998.

[3] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[4] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008.

[5] Michael J. Brusco and Stephanie Stahl. *Branch-and-Bound Applications in Combinatorial Data Analysis*. Statistics and Computing. Springer, New York, 2005.

[6] T. Tony Cai and Rong Ma. Matrix reordering for noisy disordered matrices: Optimality and computationally efficient algorithms. *IEEE Transactions on Information Theory*, 70(1):509–531, 2024.

[7] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. Special Issue: Diffusion Maps and Wavelets.

[8] Jianqing Fan, Yingying Fan, Jinchi Lv, Fan Yang, and Diwen Yu. Asymptotic theory of eigenvectors for latent embeddings with generalized laplacian matrices, 2025. arXiv:2503.00640.

[9] Fajwel Fogel, Alexandre d'Aspremont, and Milan Vojnovic. Serialrank: Spectral ranking using seriation. In *Advances in Neural Information Processing Systems*, volume 27, pages 900–908, 2014.

[10] Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre D' Aspremont. Convex relaxations for permutation problems. In *Advances in Neural Information Processing Systems*, volume 26, pages 1465–1488, 2013.

[11] Christophe Giraud, Yann Issartel, and Nicolas Verzelen. Localization in 1D non-parametric latent space models from pairwise affinities. *Electronic Journal of Statistics*, 17(1):1587–1662, 2023.

[12] Laleh Haghverdi, Maren Büttner, F. Alexander Wolf, Florian Buettner, and Fabian J. Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, 13(10):845–848, 2016.

[13] Michael Hahsler, Kurt Hornik, and Christian Buchta. Getting things in order: An introduction to the R package seriation. *Journal of Statistical Software*, 25(3):1–34, 2008.

[14] Chiaowen Joyce Hsiao, PoYuan Tung, John D. Blischak, Jonathan E. Burnett, Kenneth A. Barr, Kushal K. Dey, Matthew Stephens, and Yoav Gilad. Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome Research*, 30(4):611–621, 2020.

[15] Lawrence Hubert, Phipps Arabie, and Jacqueline Meulman. *Combinatorial Data Analysis: Optimization by Dynamic Programming*. SIAM, Philadelphia, 2001.

[16] Yann Issartel, Christophe Giraud, and Nicolas Verzelen. Minimax optimal seriation in polynomial time, 2025. arXiv:2405.08747.

[17] Yuehaw Khoo, Xin T. Tong, Wanjie Wang, and Yuguan Wang. Temporal label recovery from noisy dynamical data, 2024. arXiv:2406.13635.

[18] Rong Ma, T. Tony Cai, and Hongzhe Li. Optimal estimation of bacterial growth rates based on a permuted monotone matrix. *Biometrika*, 108(3):693–708, 2021.

[19] Rong Ma, T. Tony Cai, and Hongzhe Li. Optimal permutation recovery in permuted monotone matrix model. *Journal of the American Statistical Association*, 116(535):1358–1372, 2021.

[20] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018.

[21] Amit Moscovich, Amit Halevi, Joakim Andén, and Amit Singer. Cryo-EM reconstruction of continuous heterogeneity by laplacian spectral volumes. *Inverse Problems*, 36(2):24003, 2020.

[22] Patrick Rubin-Delanchy. Manifold structure in graph embeddings. In *Advances in Neural Information Processing Systems*, volume 33, pages 11687–11699, 2020.

[23] Patrice Y. Simard, Yann A. Le Cun, John S. Denker, and Bernard Victorri. Transformation invariance in pattern recognition: Tangent distance and propagation. *International Journal of Imaging Systems and Technology*, 11(3):181–197, 2000.

[24] A. Singer. From graph to manifold laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006. Special Issue: Diffusion Maps and Wavelets.

[25] Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2016.

[26] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[27] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[28] Nick Whiteley, Annie Gray, and Patrick Rubin-Delanchy. Statistical exploration of the manifold hypothesis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 88(2):353–385, 2026.

Appendix

This section is organized into five appendices. Appendix A presents the intuition and illustration of the baseline spectral seriation algorithm. Appendix B provides all the assumptions, technical lemmas, theoretical results, and proofs. Implementation details for Section 5 and 6 can be found in Appendix C. Appendix D presents additional simulations under the main setup. Appendix E investigates non-uniform sampling of latent phases.

A Baseline spectral seriation

In this section, we introduce the underlying mathematical intuition of the baseline spectral seriation method.

Since the latent trajectory is periodic, the collection $\{X(t) : 0 \leq t < 2\pi\}$ forms a closed one-dimensional manifold $\mathcal{M}$ embedded in a potentially high-dimensional ambient space. The Laplace-Beltrami operator on $\mathcal{M}$ has a constant eigenfunction with eigenvalue 0, and its first nontrivial eigenspace is spanned by the two phase functions $\cos(t)$ and $\sin(t)$, both with eigenvalue 1. More generally, the nontrivial eigenfunctions are $\cos(kt)$ and $\sin(kt)$, with eigenvalues k^2 for integers $k \geq 1$. Such spectral structure gives the ideal population representation

$$(\Psi_n)_i := (\cos(t_i), \sin(t_i)), \quad i \in [n], \tag{13}$$

where $(\Psi_n)_i$ denotes the ith row of Ψ_n. Hence, each sample is represented by a point on the unit circle, and recovering the temporal label t_i is equivalent to recovering its angular coordinate. The graph-based method therefore aims to approximate this two-dimensional phase eigenspace from noisy observations.

As a result, spectral seriation uses F_+ and F_- as the data version of $\cos(t)$ and $\sin(t)$. There will be two sources of error: the error from the manifold Laplacian and eigenfunctions to the Laplacian L and eigenvectors based on clean data $X(t_i)$, and the error from the clean data to the noisy data Z_i. In Figure 3, we illustrate how the eigenfunctions contain the information about temporal phases t_i, and the effects of two error sources.

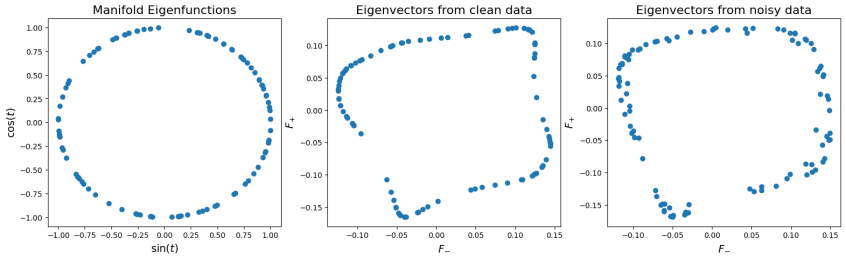


Figure 3: Spectral embedding for periodic manifold data. From left to right, the panels show the population manifold eigenfunctions $\cos(t)$ versus $\sin(t)$, the graph Laplacian eigenvectors F_+ versus F_- computed from clean data, and the corresponding eigenvectors computed from noisy data, respectively.

In Figure 3, the left panel shows the ideal phase representation $(\cos(t), \sin(t))$ on the unit circle. The middle panel demonstrates that in the clean-data case, the graph eigenvector embedding recovers the same circular geometry up to rotation and scaling. The right panel shows that under noise, the embedding is perturbed but still approximately preserves the circular phase structure. Consequently, temporal label recovery can be reduced to angular recovery in the two-dimensional spectral embedding.

B Assumptions, technical lemmas and results, and proofs

Throughout this appendix, we use the notation introduced in Section 4 unless otherwise specified. We first state the model assumptions in Section B.1. In Section B.2, we establish that the reference-kernel eigenspace converges to the sampled trigonometric eigenspace. Building on this result, we prove Proposition 1 in Appendix B.4 and Theorem 1 in Appendix B.5.

B.1 Model assumptions

To state the theoretical results, we first specify the model assumptions for $X(t)$ and the noise.

Assumption 1. *The observed data points $Z_i = X(t_i) + e_i$ with $i \in [n]$ satisfy the following:*

1) *The true data map $X : [0, 2\pi] \to \mathbb{R}^p$ is smooth with $X(0) = X(2\pi)$, and $\|X'(t)\| = 1$.*
2) *The manifold $\mathcal{M}$ formed by $\{X(t) : 0 \le t \le 2\pi\}$ is isometrically embedded in $\mathbb{R}^p$.*
3) *The timestamps t_i are independent and uniformly distributed on $[0, 2\pi]$.*
4) *The noise vectors $e_i \in \mathbb{R}^p$ are independent and distributed as $\mathcal{N}(\mathbf{0}, \sigma^2 I_p)$, where $\sigma > 0$ is unknown.*

Assumption 2. *To streamline the theoretical results, we impose the following conditions.*

1) *Let $\lambda_i^K \ge 0$ and $\lambda_i^H \ge 0$ be the ith smallest eigenvalue of $L^\star_{\alpha,K}$ and $L^\star_{\alpha,H}$, respectively. There exist constants $\gamma_1, \gamma_2 > 0$ such that $\lambda_2^K \ge \gamma_1 \tau$ and $\lambda_4(\tau^{-1} L^\star_{\alpha,K}) \ge \frac{1}{2} + \gamma_2$. Define $c_{\mathrm{gap}} := \min(\gamma_1, \frac{\gamma_2}{2})$. Then $\min\left\{\lambda_2^H - \lambda_1^H, \lambda_4^H - \lambda_3^H\right\} \ge c_{\mathrm{gap}} \tau$.*
2) *The bandwidth and sample size satisfy that $\tau \to 0$, $\frac{\sqrt{\log n}}{n^{1/2}\tau^{5/4}} \to 0$, and $\frac{\rho_{h,\tau,p}}{n\tau^{3/2}} \to 0$.*

Assumption 2 above guarantees that the target eigenspace is well separated from the eigenspaces spanned by the other eigenvectors. The assumption of $\frac{\sqrt{\log n}}{n^{1/2}\tau^{5/4}} \to 0$ is equivalent to $h \gg (\log(n)/n)^{1/5}$. If $\sigma = O(h)$ and the first two bandwidth assumptions are satisfied, the condition of $\frac{\rho_{h,\tau,p}}{n\tau^{3/2}} \to 0$ follows automatically.

B.2 Convergence from the reference kernel to manifold

To analyze the expected kernel matrix H, we introduce a reference Gaussian kernel matrix $K \in \mathbb{R}^{n\times n}$ with entries

$$K_{ij} = C_{h,\tau} \exp\left\{-\frac{\|X(t_i) - X(t_j)\|_2^2}{2\tau}\right\}, \quad i, j \in [n]. \tag{14}$$

Recall that $C_{h,\tau}$, τ, and $\rho_{h,\tau,p}$ are defined in Section 4. By construction, $K_{ij} = H_{ij}$ for $i \ne j$, while $K_{ii} = C_{h,\tau}$. Hence, $H = K + C_{h,\tau}\rho_{h,\tau,p} I_n$.

In this subsection, we establish convergence of the eigenspace spanned by $F_{\alpha,K}$ to the sampled trigonometric eigenspace spanned by Φ_n. We begin by defining the associated continuum and empirical operators.

Definition 2. *For any nonnegative continuous kernel function $\kappa : [0, \infty) \to \mathbb{R}_+$, let*

$$\kappa_\tau(x, y) := \kappa\left(\frac{\|x - y\|_{\mathbb{R}^p}}{\tau^{1/2}}\right), \quad x,\, y \in \mathcal{M}$$

Denote $p(y)$ as the sampling density on $\mathcal{M}$. Define the associated continuum quantities as

$$p_\tau(x) = \int_{\mathcal{M}} \kappa_\tau p(y)\, dV(y), \quad \kappa_{\tau,\alpha}(x, y) = \frac{\kappa_\tau(x, y)}{p_\tau(x)^\alpha p_\tau(y)^\alpha},$$
$$d_{\tau,\alpha}(x) = \int_{\mathcal{M}} \kappa_{\tau,\alpha}(x, y) p(y)\, dV(y), \quad m_{\tau,\alpha}(x, y) = \frac{\kappa_{\tau,\alpha}(x, y)}{d_{\tau,\alpha}(x)}.$$

Similarly, define the empirical counterparts as

$$\widehat{p}_{\tau,n}(x) = \frac{1}{n}\sum_{\ell=1}^n \kappa_\tau(x, x_\ell), \quad \widehat{\kappa}_{\tau,\alpha,n}(x, y) = \frac{\kappa_\tau(x, y)}{\widehat{p}_{\tau,n}(x)^\alpha \widehat{p}_{\tau,n}(y)^\alpha},$$
$$\widehat{d}_{\tau,\alpha,n}(x) = \frac{1}{n}\sum_{j=1}^n \widehat{\kappa}_{\tau,\alpha,n}(x, x_j), \quad \widehat{m}_{\tau,\alpha,n}(x, y) = \frac{\widehat{\kappa}_{\tau,\alpha,n}(x, y)}{\widehat{d}_{\tau,\alpha,n}(x)}.$$

For any smooth scalar function $f : \mathcal{M} \to \mathbb{R}$, define the associated continuum and empirical operators as

$$\mathcal{T}_{\tau,\alpha} f(x) = \int_{\mathcal{M}} m_{\tau,\alpha}(x,y) f(y) p(y)\, dV(y) \tag{15}$$

$$\widehat{\mathcal{T}}_{\tau,\alpha,n} f(x) = \frac{1}{n} \sum_{j=1}^{n} \widehat{m}_{\tau,\alpha,n}(x, x_j) f(x_j). \tag{16}$$

The lemma below provides the pointwise approximation of the reference graph operator by the Laplace–Beltrami operator, thereby linking the discrete construction to the continuum geometry.

Lemma 1. *Assume that Assumption 1 holds and denote by $x_i = X(t_i)$. Let $f \in C^4(\mathcal{M})$ and write*

$$f_n = (f(x_1), \ldots, f(x_n))^\top.$$

For $\alpha \in [0,1]$, further assume that $\tau \to 0$ and $\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}} \to 0$. Then we have that with probability at least $1 - n^{-2}$, for all $i \in [n]$,

$$(P_{\alpha,K} f_n - f_n)_i = (\mathcal{T}_{\tau,\alpha} f - f)(x_i) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{1/4}}\right),$$

where $\mathcal{T}_{\tau,\alpha}$ is defined in (15)*. Moreover, it holds that*

$$(\mathcal{T}_{\tau,\alpha} f - f)(x_i) = \frac{\tau}{2} \Delta_{\mathcal{M}} f(x_i) + O(\tau^2).$$

Consequently, with probability at least $1 - n^{-2}$, we have that

$$\tau^{-1}(P_{\alpha,K} f_n - f_n)_i = \frac{1}{2} \Delta_{\mathcal{M}} f(x_i) + O(\tau) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)$$

uniformly over the sample points.

Proof. We will apply Theorem 5.3 and Corollary 5.1 of [25]. In their notation, our setting corresponds to the trivial line bundle, i.e., $q = 1$ and $g_{ij} = 1$. It remains to verify that their empirical normalized operator coincides with our matrix $P_{\alpha,K}$; once this identification is made, the results cited above will be applicable directly.

With $x_i = X(t_i)$, rewrite the reference kernel matrix K as

$$K_{ij} = C_{h,\tau} \exp\left\{-\frac{\|x_i - x_j\|^2}{2\tau}\right\}.$$

With such notation, the empirical quantities in Definition 2 can be rewritten as

$$\widehat{p}_{\tau,n}(x_i) = \frac{D_K(i)}{n}, \quad \widehat{K}_{\tau,\alpha,n}(x_i, x_j) = n^{2\alpha} D_K(i)^{-\alpha} K_{ij} D_K(j)^{-\alpha} = n^{2\alpha} \widetilde{K}_{\alpha,ij},$$

$$\widehat{d}_{\tau,\alpha,n}(x_i) = n^{2\alpha-1} \widetilde{D}_{\alpha,K}(i), \quad \frac{1}{n} \widehat{M}_{\tau,\alpha,n}(x_i, x_j) = \frac{\widetilde{K}_{\alpha,ij}}{\widetilde{D}_{\alpha,K}(i)} = P_{\alpha,K}(i,j).$$

Consequently, for each smooth f, it holds that

$$(P_{\alpha,K} f_n - f_n)_i = \widehat{\mathcal{T}}_{\tau,\alpha,n} f(x_i) - f(x_i).$$

By resorting to Theorem 5.3 of [25], with probability at least $1 - n^{-2}$, we have that for all $i \in [n]$,

$$(P_{\alpha,K} f_n - f_n)_i = (\mathcal{T}_{\tau,\alpha} f - f)(x_i) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{1/4}}\right)$$

for each fixed $\alpha \in [0,1]$. In our setting, the case of $\alpha = 0$ satisfies the same conclusion as the case of $0 < \alpha \leq 1$, since $\mathcal{M}$ has no boundary contribution. Moreover, under Assumption 1, the induced

sampling density on $\mathcal{M}$ is constant. Hence, the density-drift term in Theorem 5.2 of [25] vanishes, which gives that

$$(\mathcal{T}_{\tau,\alpha} f - f)(x_i) = \frac{\tau}{2}\Delta_{\mathcal{M}} f(x_i) + O(\tau^2), \quad 0 \le \alpha \le 1.$$

Therefore, combining these estimates and dividing by τ yield that

$$\tau^{-1}(P_{\alpha,K} f_n - f_n)_i = \frac{1}{2}\Delta_{\mathcal{M}} f(x_i) + O(\tau) + O\left(\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right).$$

This completes the proof of Lemma 1.

The lemma below verifies that the additional α-normalization does not introduce non-negligible degree heterogeneity, which will be needed to compare weighted and unweighted eigenspaces.

Lemma 2. *Assume that all the conditions of Lemma 1 are satisfied. Then we have that with probability at least $1 - n^{-2}$, for any fixed $\alpha \in [0, 1]$,*

$$\max_{i\in[n]} \left| \frac{\widetilde{D}_{\alpha,K}(i)}{\bar{\widetilde{D}}_{\alpha,K}} - 1 \right| = O\left((1+\alpha)\left(\tau + \sqrt{\frac{\log(n)}{n\tau^{1/2}}}\right)\right) = o(1),$$

where $\bar{\widetilde{D}}_{\alpha,K} = \frac{1}{n}\sum_{i=1}^n \widetilde{D}_{\alpha,K}(i)$.

Proof. We first show that the unnormalized reference degrees $D_K(i)$ are uniformly flat. Recall that

$$D_K(i) = \sum_{j=1}^n K_{ij}, \quad K_{ij} = C_{h,\tau} \exp\left\{-\frac{\|X(t_i) - X(t_j)\|^2}{2\tau}\right\}.$$

Conditional on t_i, random variables $\{K_{ij} : j \neq i\}$ are independent and identically distributed (i.i.d.). For fixed $t_i = t$, it holds that

$$\mathbb{E}[K_{ij}|t_i = t] = C_{h,\tau}\frac{1}{2\pi}\int_0^{2\pi} \exp\left\{-\frac{\|X(t) - X(s)\|^2}{2\tau}\right\} ds.$$

Since X is a smooth closed embedded curve with $\|X'(t)\| = 1$, there exists some $r_0 > 0$ such that for all $t \in [0, 2\pi]$ and $|u| \le r_0$,

$$\|X(t+u) - X(t)\|^2 = u^2 + O(u^4).$$

This expansion follows from Taylor's theorem. For $r_0 < |u| \le \pi$, we have that

$$\|X(t+u) - X(t)\| \ge \delta > 0.$$

It follows that the contribution from $|u| > r_0$ is $O(e^{-c/\tau})$. Hence, after a change of variable $u = s - t$ and using the periodicity, we can deduce that

$$\begin{aligned}\int_0^{2\pi} \exp\left\{-\frac{\|X(t) - X(s)\|^2}{2\tau}\right\} ds &= \int_{-\pi}^{\pi} \exp\left\{-\frac{\|X(t+u) - X(t)\|^2}{2\tau}\right\} du \\ &= \int_{|u|\le r_0} \exp\left\{-\frac{u^2 + O(u^4)}{2\tau}\right\} du + O(e^{-c/\tau}).\end{aligned}$$

With a change of variable $u = \sqrt{\tau} z$, the Gaussian moment and tail estimates give that

$$\int_0^{2\pi} \exp\left\{-\frac{\|X(t) - X(s)\|^2}{2\tau}\right\} ds = \sqrt{2\pi\tau}(1 + O(\tau)).$$

Taking the supremum over $t \in [0, 2\pi]$ on both sides, it holds that

$$\mathbb{E}[K_{ij}|t_i] = \frac{1}{\sqrt{2\pi}} C_{h,\tau}\tau^{1/2}(1 + O(\tau)).$$

Moreover, we have that $0 \le K_{ij} \le C_{h,\tau}$. Conditional on $t_i = t$, we have that for any $j \neq i$,

$$\mathbb{E}[K_{ij}^2|t_i = t] = C_{h,\tau}^2 \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{-\frac{\|X(t) - X(s)\|^2}{\tau}\right\} ds.$$

By the same local Gaussian-integral estimate, we can show that

$$\sup_{t\in[0,2\pi]}\int_0^{2\pi}\exp\left\{-\frac{\|X(t)-X(s)\|^2}{\tau}\right\}ds=\sqrt{\pi\tau}(1+O(\tau)),$$

Consequently, we can obtain that

$$\mathbb{E}[K_{ij}^2|t_i]\le CC_{h,\tau}^2\tau^{1/2}.$$

Observe that for fixed t_i and $j\neq i$, random variables $K_{ij}-\mathbb{E}[K_{ij}|t]$ are independent, have mean zero, and satisfy that $|K_{ij}-\mathbb{E}[K_{ij}|t]|\le CC_{h,\tau}$. Moreover, it holds that

$$\sum_{j\neq i}\mathbb{E}[(K_{ij}-\mathbb{E}[K_{ij}|t])^2|t_i]\le\sum_{j\neq i}\mathbb{E}[K_{ij}^2|t_i]\le CnC_{h,\tau}^2\tau^{1/2}.$$

An application of Bernstein's inequality leads to

$$\left|\sum_{j\neq i}K_{ij}-(n-1)\mathbb{E}[K_{ij}|t_i]\right|\le CC_{h,\tau}\left(\sqrt{n\tau^{1/2}\log(n)}+\log(n)\right)$$

with probability at least $1-n^{-3}$. Since $\tau\to0$ and $\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\to0$, term $\log(n)(n\tau^{1/2})^{-1}\to0$ automatically. Thus, the second term is negligible relative to $nC_{h,\tau}\tau^{1/2}$. Then a union bound over $i\in[n]$ results in

$$\max_{i\in[n]}\left|\sum_{j\neq i}K_{ij}-(n-1)\mathbb{E}[K_{ij}|t_i]\right|\le CC_{h,\tau}\sqrt{n\tau^{1/2}\log(n)}$$

with probability at least $1-n^{-2}$. Since $K_{ii}=C_{h,\tau}$, combining the preceding bounds yields that

$$D_K(i)=nC_{h,\tau}\frac{\tau^{1/2}}{\sqrt{2\pi}}(1+O(\delta_n)),\quad \delta_n=\tau+\sqrt{\frac{\log(n)}{n\tau^{1/2}}}$$

with probability at least $1-n^{-2}$. By the assumptions on τ and n, it follows that

$$D_K(i)^{-\alpha}=\left(nC_{h,\tau}\frac{\tau^{1/2}}{\sqrt{2\pi}}\right)^{-\alpha}(1+O(\alpha\delta_n)).$$

Further, we can show that

$$\begin{aligned}\widetilde{D}_{\alpha,K}(i)&=\sum_{j=1}^n D_K(i)^{-\alpha}K_{ij}D_K(j)^{-\alpha}\\&=\left(nC_{h,\tau}\frac{\tau^{1/2}}{\sqrt{2\pi}}\right)^{-2\alpha}(1+O(\alpha\delta_n))\sum_{j=1}^n K_{ij}.\end{aligned}$$

Using again the uniform expansion for $D_K(i)=\sum_j K_{ij}$, it holds that

$$\widetilde{D}_{\alpha,K}(i)=\left(nC_{h,\tau}\frac{\tau^{1/2}}{\sqrt{2\pi}}\right)^{1-2\alpha}(1+O((1+\alpha)\delta_n)),\quad\forall i\in[n].$$

Averaging the last expression over i leads to

$$\bar{\widetilde{D}}_{\alpha,K}=\left(nC_{h,\tau}\frac{\tau^{1/2}}{\sqrt{2\pi}}\right)^{1-2\alpha}(1+O((1+\alpha)\delta_n)).$$

Therefore, we can obtain that

$$\max_{i\in[n]}\left|\frac{\widetilde{D}_{\alpha,K}(i)}{\bar{\widetilde{D}}_{\alpha,K}}-1\right|=O\left((1+\alpha)\delta_n\right),$$

where the right-hand side is $o(1)$ under the stated bandwidth condition. This concludes the proof of Lemma 2.

Definition 3. *The matrix Ψ_n collects the two continuum eigenfunctions $\cos(t)$ and $\sin(t)$ evaluated at the sample points*

$$\Psi_n = \begin{bmatrix} \cos(t_1) & \sin(t_1) \\ \cos(t_2) & \sin(t_2) \\ \vdots & \vdots \\ \cos(t_n) & \sin(t_n) \end{bmatrix}. \tag{17}$$

Denote the reference basis corresponding to the empirical trigonometric space $\mathrm{span}\{(\cos \mathrm{t_i})_\mathrm{i}, (\sin \mathrm{t_i})_\mathrm{i}\}$ as

$$\Theta_{\alpha,n} := \widetilde{D}_{\alpha,K}^{1/2} \Psi_n \left(\Psi_n^\top \widetilde{D}_{\alpha,K} \Psi_n \right)^{-1/2} \tag{18}$$

with $(\Theta_{\alpha,n})^\top \Theta_{\alpha,n} = I_2$. Let $q_{\alpha,n} := \frac{\widetilde{D}_{\alpha,K}^{1/2}\mathbf{1}}{\|\widetilde{D}_{\alpha,K}^{1/2}\mathbf{1}\|_2}$. The orthonormalized projection of $\Theta_{\alpha,n}$ onto $q_{\alpha,n}^\perp$ is given by

$$\Theta_{\alpha,n}^\perp = (I - q_{\alpha,n}(q_{\alpha,n})^\top)\Theta_{\alpha,n} \left((\Theta_{\alpha,n})^\top (I - q_{\alpha,n}(q_{\alpha,n})^\top)\Theta_{\alpha,n} \right)^{-1/2}. \tag{19}$$

Define

$$\Phi_n := \Psi_n (\Psi_n^\top \Psi_n)^{-1/2}.$$

Consequently, Φ_n has orthonormal columns and spans the empirical trigonometric space

$$\mathrm{span}\{(\cos \mathrm{t_i})_{\mathrm{i=1}}^{\mathrm{n}}, (\sin \mathrm{t_i})_{\mathrm{i=1}}^{\mathrm{n}}\}.$$

We next translate the pointwise operator approximation in Lemma 1 into an approximate eigenrelation for the empirical trigonometric basis. This step provides a bridge from pointwise convergence to eigenspace convergence.

Lemma 3. *Assume that all the conditions of Lemma 1 are satisfied. Then we have that with probability $1 - O(n^{-2})$,*

$$\left\| \left(\tau^{-1} L_{\alpha,K}^\star - \frac{1}{2} I \right) \Theta_{\alpha,n} \right\|_F = O\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}} \right),$$

where $\Theta_{\alpha,n}$ is defined in (18)*.*

Proof. Denote by

$$r_n := \tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}.$$

By invoking Lemma 1, it holds that for each fixed $f \in C^4(\mathcal{M})$,

$$\tau^{-1}(P_{\alpha,K} f_n - f_n)_i = \frac{1}{2}\Delta_M f(x_i) + O(r_n)$$

with probability at least $1 - n^{-2}$. We will apply this relation to $f(X(t)) = \cos(t)$. Since X is unit speed with $X(0) = X(2\pi)$, it follows that

$$\Delta_{\mathcal{M}} f(X(t)) = \frac{d^2}{dt^2}\cos(t) = -\cos(t) = -f(X(t)).$$

Hence, we have that for all $i \in [n]$,

$$\tau^{-1}(P_{\alpha,K} f_n - f_n)_i = -\frac{1}{2} f(X(t_i)) + O(r_n) = -\frac{1}{2}\cos(t_i) + O(r_n),$$

where $f_n := (f(X(t_1)), \ldots, f(X(t_n))^\top = (\cos(t_1), \ldots, \cos(t_n))^\top$. Equivalently, we have that

$$\left(\tau^{-1} L_{\alpha,K} f_n - \frac{1}{2} f_n \right)_i = O(r_n).$$

An application of the same argument with $f(t) = \sin(t)$ leads to

$$\left(\tau^{-1} L_{\alpha,K} g_n - \frac{1}{2} g_n \right)_i = O(r_n), \quad g_n = (\sin(t_1), \ldots, \sin(t_n))^\top.$$

Let $E_{\alpha,n} := \left(\tau^{-1}L_{\alpha,K} - \frac{1}{2}I\right)\Psi_n$. The ith row of $E_{\alpha,n}$ can be written as

$$(E_{\alpha,n})_i = \left(\left(\tau^{-1}L_{\alpha,K}f_n - \frac{1}{2}f_n\right)_i, \left(\tau^{-1}L_{\alpha,K}g_n - \frac{1}{2}g_n\right)_i\right).$$

Then it holds that $\max_{i\in[n]} \|(E_{\alpha,n})_i\|_2 = O(r_n)$ with probability at least $1 - 2n^{-2}$. Since

$$L^\star_{\alpha,K} = \widetilde{D}^{1/2}_{\alpha,K} L_{\alpha,K} \widetilde{D}^{-1/2}_{\alpha,K},$$

we have that

$$\tau^{-1}L^\star_{\alpha,K} = \widetilde{D}^{1/2}_{\alpha,K}\left(\tau^{-1}L_{\alpha,K}\right)\widetilde{D}^{-1/2}_{\alpha,K}.$$

Thus, we can deduce that

$$\begin{aligned}\left(\tau^{-1}L^\star_{\alpha,K} - \frac{1}{2}I\right)\Theta_{\alpha,n} &= \widetilde{D}^{1/2}_{\alpha,K}\left(\tau^{-1}L_{\alpha,K} - \frac{1}{2}I\right)\Psi_n\left(\Psi_n^\top \widetilde{D}_{\alpha,K}\Psi_n\right)^{-1/2} \\ &= \widetilde{D}^{1/2}_{\alpha,K} E_{\alpha,n}\left(\Psi_n^\top \widetilde{D}_{\alpha,K}\Psi_n\right)^{-1/2}.\end{aligned}$$

We now bound the two factors above separately. Since $\max_{i\in[n]} \|(E_{\alpha,n})_i\|_2 \le Cr_n$ with probability at least $1 - 2n^{-2}$, it holds that

$$\|\widetilde{D}^{1/2}_{\alpha,K}E_{\alpha,n}\|_F = \sqrt{\sum_{i=1}^n \widetilde{D}_{\alpha,K}(i)\|(E_{\alpha,n})_i\|_2^2} \le \sqrt{Cr_n^2\sum_{i=1}^n \widetilde{D}_{\alpha,K}(i)} = Cr_n\sqrt{n\bar{\widetilde{D}}_{\alpha,K}}$$

with probability at least $1 - n^{-2}$. An application of Lemma 2 shows that with probability at least $1 - n^{-2}$,

$$\Psi_n^\top \widetilde{D}_{\alpha,K}\Psi_n = \bar{\widetilde{D}}_{\alpha,K}\Psi_n^\top\Psi_n + o(n\bar{\widetilde{D}}_{\alpha,K}).$$

Denote by $a_n = \frac{1}{n}\sum_{i=1}^n \cos(2t_i)$ and $b_n = \frac{1}{n}\sum_{i=1}^n \sin(2t_i)$. Then it follows that

$$\frac{1}{n}\Psi_n^\top\Psi_n = \frac{1}{2}I_2 + \frac{1}{2}\begin{pmatrix} a_n & b_n \\ b_n & -a_n\end{pmatrix}.$$

Notice that since $t_i \overset{i.i.d.}{\sim} \mathrm{Unif}[0, 2\pi]$, random variables $\cos(2t_i)$ and $\sin(2t_i)$ are independent, mean-zero, and bounded in $[-1, 1]$. Then an application of Hoeffding's inequality gives that

$$\mathbb{P}\left(|a_n| > \frac{1}{2\sqrt{2}}\right) \le 2e^{-n/16}, \quad \mathbb{P}\left(|b_n| > \frac{1}{2\sqrt{2}}\right) \le 2e^{-n/16}.$$

Consequently, we have that with probability at least $1 - 4e^{-n/16}$, $\sqrt{a_n^2 + b_n^2} \le \frac{1}{2}$ and thus

$$\lambda_{\min}\left(\frac{1}{n}\Psi_n^\top\Psi_n\right) = \frac{1}{2} - \frac{1}{2}\sqrt{a_n^2 + b_n^2} \ge \frac{1}{4}.$$

Equivalently, we have that $\lambda_{\min}(\Psi_n^\top\Psi_n) \ge \frac{n}{4}$.

Further, by resorting to Weyl's inequality, we can deduce that

$$\begin{aligned}\lambda_{\min}(\Psi_n^\top \widetilde{D}_{\alpha,K}\Psi_n) &\ge \bar{\widetilde{D}}_{\alpha,K}\lambda_{\min}(\Psi_n^\top\Psi_n) - \left\|\Psi_n^\top \widetilde{D}_{\alpha,K}\Psi_n - \bar{\widetilde{D}}_{\alpha,K}\Psi_n^\top\Psi_n\right\|_2 \\ &\ge \frac{1}{4}n\bar{\widetilde{D}}_{\alpha,K} - o(n\bar{\widetilde{D}}_{\alpha,K}) \\ &\ge cn\bar{\widetilde{D}}_{\alpha,K}\end{aligned}$$

with probability at least $1 - 4e^{-n/16} - n^{-2}$. Therefore, combining the two bounds yields that with probability $1 - O(n^{-2})$,

$$\left\|\left(\tau^{-1}L^\star_{\alpha,K} - \frac{1}{2}I\right)\Theta_{\alpha,n}\right\|_F \le \|\widetilde{D}^{1/2}_{\alpha,K}E_{\alpha,n}\|_F \left\|\left(\Psi_n^\top \widetilde{D}_{\alpha,K}\Psi_n\right)^{-1/2}\right\|_2 \le Cr_n.$$

Since $r_n = \tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}} = o(1)$ under the bandwidth condition of Lemma 2, the desired conclusion follows. This completes the proof of Lemma 3.

Using the approximate eigen-relation from Lemma 3, we next obtain an upper bound for the first nontrivial eigenvalues of the rescaled reference graph Laplacian.

Lemma 4. *Assume that all the conditions of Lemma 1 are satisfied. Let $L_{\alpha,n} := \tau^{-1} L^\star_{\alpha,K}$, and assume further that $0 = \lambda_1(L_{\alpha,n}) \le \lambda_2(L_{\alpha,n}) \le \lambda_3(L_{\alpha,n}) \le \cdots$. Then we have that with probability $1 - O(n^{-2})$,*

$$\lambda_2(L_{\alpha,n}),\, \lambda_3(L_{\alpha,n}) \le \frac{1}{2} + O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Equivalently, we have that with probability $1 - O(n^{-2})$,

$$\lambda_2(L^\star_{\alpha,K}),\, \lambda_3(L^\star_{\alpha,K}) \le \frac{\tau}{2} + O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Proof. Recall that $\Theta_{\alpha,n}$, $q_{\alpha,n}$, and $\Theta^\perp_{\alpha,n}$ defined in (18) and (19). Also, recall that

$$P^\star_{\alpha,K} := \widetilde{D}^{-1/2}_{\alpha,K} \widetilde{K}_\alpha \widetilde{D}^{-1/2}_{\alpha,K}, \quad L^\star_{\alpha,K} = I - P^\star_{\alpha,K},$$

It holds that $P^\star_{\alpha,K} q_{\alpha,n} = q_{\alpha,n}$ and thus $L_{\alpha,n} q_{\alpha,n} = 0$. Moreover, since the Gaussian kernel is strictly positive, $\widetilde{K}_\alpha$ has strictly positive entries. By the Perron–Frobenius theorem, the eigenvalue 1 of $P^\star_{\alpha,K}$ is simple. Hence, the zero eigenspace of $L_{\alpha,n}$ is $\mathrm{span}\{\mathrm{q}_{\alpha,\mathrm{n}}\}$. In view of Lemma 3, we have that with probability $1 - O(n^{-2})$,

$$\left\| \left(L_{\alpha,n} - \frac{1}{2} I \right) \Theta_{\alpha,n} \right\|_F = O\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}} \right).$$

It follows from Lemma 2 that with probability $1 - n^{-2}$,

$$\max_{i \in [n]} \left| \frac{\widetilde{D}_{\alpha,K}(i)}{\bar{\widetilde{D}}_{\alpha,K}} - 1 \right| = O((1+\alpha)\delta_n).$$

Applying Hoeffding's inequality to the entries of Ψ_n and $\Psi_n^\top \Psi_n$, followed by a union bound, gives that with probability $1 - O(n^{-2})$,

$$\frac{1}{n} \mathbf{1}^\top \Psi_n = O\left(\sqrt{\frac{\log(n)}{n}} \right), \quad \frac{1}{n} \Psi_n^\top \Psi_n = \frac{1}{2} I_2 + O\left(\sqrt{\frac{\log(n)}{n}} \right).$$

This entails that

$$\frac{1}{n \bar{\widetilde{D}}_{\alpha,K}} \mathbf{1}^\top \widetilde{D}_{\alpha,K} \Psi_n = O\left((1+\alpha)\delta_n + \sqrt{\frac{\log(n)}{n}} \right),$$
$$\frac{1}{n \bar{\widetilde{D}}_{\alpha,K}} \Psi_n^\top \widetilde{D}_{\alpha,K} \Psi_n = \frac{1}{2} I_2 + O\left((1+\alpha)\delta_n + \sqrt{\frac{\log(n)}{n}} \right).$$

Consequently, we have that with probability $1 - O(n^{-2})$,

$$(q_{\alpha,n})^\top \Theta_{\alpha,n} = \frac{\mathbf{1}^\top \widetilde{D}_{\alpha,K} \Psi_n}{\left(\mathbf{1}^\top \widetilde{D}_{\alpha,K} \mathbf{1} \right)^{1/2}} \left(\Psi_n^\top \widetilde{D}_{\alpha,K} \Psi_n \right)^{-1/2} = O\left((1+\alpha)\delta_n \right).$$

Let us recall the definition

$$\Theta^\perp_{\alpha,n} = (I - q_{\alpha,n}(q_{\alpha,n})^\top)\Theta_{\alpha,n} \left[(\Theta_{\alpha,n})^\top (I - q_{\alpha,n}(q_{\alpha,n})^\top) \Theta_{\alpha,n} \right]^{-1/2}.$$

It holds that

$$(\Theta_{\alpha,n})^\top (I - q_{\alpha,n}(q_{\alpha,n})^\top) \Theta_{\alpha,n} = I_2 + O\left((1+\alpha)\delta_n \right).$$

Under our bandwidth conditions, we have that $O\left((1+\alpha)\delta_n\right) = o(1)$. Hence, with probability $1 - O(n^{-2})$, $\Theta^\perp_{\alpha,n}$ is well-defined, has orthonormal columns, satisfies that $\mathrm{span}(\Theta^\perp_{\alpha,\mathrm{n}}) \subset \mathrm{q}^\perp_{\alpha,\mathrm{n}}$, and

has $\dim(\mathrm{span}(\Theta_{\alpha,\mathrm{n}}^{\perp})) = 2$. It remains to verify that the residual estimate is preserved after projection. It follows from $L_{\alpha,n}q_{\alpha,n} = 0$ that

$$\left(L_{\alpha,n} - \frac{1}{2}I\right) q_{\alpha,n} = -\frac{1}{2}q_{\alpha,n}.$$

Consequently, we can deduce that

$$\begin{aligned}\left(L_{\alpha,n} - \frac{1}{2}I\right)(I - q_{\alpha,n}(q_{\alpha,n})^\top)\Theta_{\alpha,n} &= \left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n} - \left(L_{\alpha,n} - \frac{1}{2}I\right) q_{\alpha,n}(q_{\alpha,n})^\top\Theta_{\alpha,n} \\ &= \left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n} + \frac{1}{2}q_{\alpha,n}(q_{\alpha,n})^\top\Theta_{\alpha,n}.\end{aligned}$$

Taking the Frobenius norms and invoking Lemma 3, we can show that with probability $1 - O(n^{-2})$,

$$\left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)(I - q_{\alpha,n}(q_{\alpha,n})^\top)\Theta_{\alpha,n}\right\|_F = O\left(\tau + \frac{\sqrt{\log n}}{n^{1/2}\tau^{5/4}} + (1+\alpha)\delta_n\right).$$

In light of the condition $\tau \to 0$, it holds that

$$\left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n}^{\perp}\right\|_F = O\left(\tau + \frac{\sqrt{\log n}}{n^{1/2}\tau^{5/4}} + (1+\alpha)\delta_n\right) = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

For any $u \in \mathrm{span}(\Theta_{\alpha,\mathrm{n}}^{\perp})$ with $\|u\|_2 = 1$, we can write $u = \Theta_{\alpha,n}^{\perp}a$ with $\|a\|_2 = 1$. Then for all $u \in \mathrm{span}(\Theta_{\alpha,\mathrm{n}}^{\perp})$ with $\|u\|_2 = 1$, it follows that

$$\begin{aligned}\left|u^\top\left(L_{\alpha,n} - \frac{1}{2}I\right)u\right| &\le \left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n}^{\perp}\right\|_2 \\ &\le \left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n}^{\perp}\right\|_F \\ &= O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).\end{aligned}$$

Thus, we can obtain that

$$\max_{\substack{u\in\mathrm{span}(\Theta_{\alpha,\mathrm{n}}^{\perp}) \\ \|u\|_2=1}} u^\top L_{\alpha,n}u \le \frac{1}{2} + O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Finally, by the min-max theorem on $q_{\alpha,n}^{\perp}$, it holds that

$$\lambda_3(L_{\alpha,n}) = \min_{\substack{S\subset q_{\alpha,n}^{\perp} \\ \dim(S)=2}} \max_{\substack{u\in S \\ \|u\|_2=1}} u^\top L_{\alpha,n}u.$$

Using $S = \mathrm{span}(\Theta_{\alpha,\mathrm{n}}^{\perp})$ gives that

$$\lambda_3(L_{\alpha,n}) \le \frac{1}{2} + O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Since $\lambda_2(L_{\alpha,n}) \le \lambda_3(L_{\alpha,n})$, we also have that

$$\lambda_2(L_{\alpha,n}) \le \frac{1}{2} + O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Therefore, the final statement follows from $L_{\alpha,n} = \tau^{-1}L_{\alpha,K}^{\star}$. This concludes the proof of Lemma 4.

We are now ready to establish the main result of this subsection: the convergence of the reference eigenspace to the empirical trigonometric space spanned by $(\cos(t_i))_{i=1}^n$ and $(\sin(t_i))_{i=1}^n$.

Proposition 2 (Conditional reference eigenspace convergence)**.** *Assume that Assumptions 1–2 hold. Then we have that with probability* $1 - O(n^{-2})$*,*

$$\|F_{\alpha,K}F_{\alpha,K}^\top - \Theta_{\alpha,n}^\perp(\Theta_{\alpha,n}^\perp)^\top\|_2 = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Consequently, we have that with probability $1 - O(n^{-2})$*,*

$$\|F_{\alpha,K}F_{\alpha,K}^\top - \Phi_n\Phi_n^\top\|_2 = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Equivalently, there exists some $Q \in \mathbb{O}(2)$ *such that with probability* $1 - O(n^{-2})$*,*

$$\|F_{\alpha,K}Q - \Phi_n\|_F = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Proof. Denote by $L_{\alpha,n} = \tau^{-1}L_{\alpha,K}^\star$. Then $L_{\alpha,n}$ is symmetric and has the same eigenspaces as $L_{\alpha,K}^\star$. We denote its eigenvalues as

$$0 = \lambda_1(L_{\alpha,n}) \le \lambda_2(L_{\alpha,n}) \le \lambda_3(L_{\alpha,n}) \le \lambda_4(L_{\alpha,n}) \le \cdots \le \lambda_n(L_{\alpha,n}).$$

In view of the proof of Lemma 4, it holds that

$$r_n := \left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n}^\perp\right\|_F = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right)$$

with probability $1 - O(n^{-2})$. Let $F_\perp$ be an orthonormal eigenbasis for the direct sum of eigenspaces of $L_{\alpha,n}$ corresponding to $\lambda_4(L_{\alpha,n}), \ldots, \lambda_n(L_{\alpha,n})$. Since $\Theta_{\alpha,n}^\perp \subset q_{\alpha,n}^\perp$, we can write

$$\Theta_{\alpha,n}^\perp = F_{\alpha,K}C_n + F_\perp S_n$$

for some matrices $C_n \in \mathbb{R}^{2\times 2}$ and $S_n \in \mathbb{R}^{(n-3)\times 2}$. Since $\Theta_{\alpha,n}^\perp$ has orthonormal columns, we have that

$$C_n^\top C_n + S_n^\top S_n = I_2.$$

By Assumption 2, there exists a fixed constant $\gamma > 0$ such that $\lambda_4(L_{\alpha,n}) \ge \frac{1}{2} + \gamma$. Hence, we can deduce that

$$\begin{aligned} r_n^2 &= \left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)\Theta_{\alpha,n}^\perp\right\|_F^2 \\ &= \left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)F_{\alpha,K}C_n\right\|_F^2 + \left\|\left(L_{\alpha,n} - \frac{1}{2}I\right)F_\perp S_n\right\|_F^2 \\ &\ge \gamma^2\|S_n\|_F^2. \end{aligned}$$

It follows that

$$\|S_n\|_F \le \frac{r_n}{\gamma} = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Since

$$(I - F_{\alpha,K}F_{\alpha,K}^\top)\Theta_{\alpha,n}^\perp = F_\perp S_n,$$

we can obtain that

$$\|(I - F_{\alpha,K}F_{\alpha,K}^\top)\Theta_{\alpha,n}^\perp\|_2 \le \|S_n\|_F = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

Since both $F_{\alpha,K}$ and $\Theta_{\alpha,n}^\perp$ have two orthonormal columns, this entails that

$$\left\|F_{\alpha,K}F_{\alpha,K}^\top - \Theta_{\alpha,n}^\perp(\Theta_{\alpha,n}^\perp)^\top\right\|_2 = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

It remains to relate $\mathrm{span}(\Theta^{\perp}_{\alpha,\mathrm{n}})$ to $\mathrm{span}(\Phi_{\mathrm{n}})$. By Lemma 2, we have that with probability $1-O(n^{-2})$,

$$\widetilde{\bar{D}}^{-1/2}_{\alpha,K}\widetilde{D}^{1/2}_{\alpha,K} = I + O((1+\alpha)\delta_n).$$

Since right multiplication by an invertible 2×2 matrix does not change the column space, it holds that

$$\mathrm{span}(\Theta_{\alpha,\mathrm{n}}) = \mathrm{span}(\widetilde{\mathrm{D}}^{1/2}_{\alpha,\mathrm{K}}\Psi_{\mathrm{n}}).$$

Moreover, in view of $\Psi_n = \Phi_n(\Psi_n^\top\Psi_n)^{1/2}$, we have that

$$\mathrm{span}(\widetilde{\mathrm{D}}^{1/2}_{\alpha,\mathrm{K}}\Psi_{\mathrm{n}}) = \mathrm{span}(\widetilde{\mathrm{D}}^{1/2}_{\alpha,\mathrm{K}}\Phi_{\mathrm{n}}).$$

It follows that

$$\mathrm{span}(\Theta_{\alpha,\mathrm{n}}) = \mathrm{span}\left(\left(\widetilde{\bar{\mathrm{D}}}^{-1/2}_{\alpha,\mathrm{K}}\widetilde{\mathrm{D}}^{1/2}_{\alpha,\mathrm{K}}\right)\Phi_{\mathrm{n}}\right).$$

Thus, $\mathrm{span}(\Theta_{\alpha,\mathrm{n}})$ is an $O((1+\alpha)\delta_n)$-perturbation of $\mathrm{span}(\Phi_{\mathrm{n}})$. Then it holds that

$$\|\Theta_{\alpha,n}\Theta^\top_{\alpha,n} - \Phi_n\Phi_n^\top\|_2 = O((1+\alpha)\delta_n).$$

Moreover, in the proof of Lemma 4, we have that $(q_{\alpha,n})^\top\Theta_{\alpha,n} = O\left((1+\alpha)\delta_n\right)$. Consequently, removing the $q_{\alpha,n}$-component changes the associated projector by at most $O\left((1+\alpha)\delta_n\right)$

$$\left\|\Theta^{\perp}_{\alpha,n}(\Theta^{\perp}_{\alpha,n})^\top - \Theta_{\alpha,n}\Theta^\top_{\alpha,n}\right\|_2 = O\left((1+\alpha)\delta_n\right).$$

Combining the two expressions above gives that

$$\left\|\Theta^{\perp}_{\alpha,n}(\Theta^{\perp}_{\alpha,n})^\top - \Phi_n\Phi_n^\top\right\|_2 = O\left((1+\alpha)\delta_n\right).$$

An application of the triangle inequality shows that

$$\begin{aligned}\|F_{\alpha,K}F^\top_{\alpha,K} - \Phi_n\Phi_n^\top\|_2 &\le \left\|F_{\alpha,K}F^\top_{\alpha,K} - \Theta^{\perp}_{\alpha,n}(\Theta^{\perp}_{\alpha,n})^\top\right\|_2 + \left\|\Theta^{\perp}_{\alpha,n}(\Theta^{\perp}_{\alpha,n})^\top - \Phi_n\Phi_n^\top\right\|_2 \\ &= O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).\end{aligned}$$

Finally, since $F_{\alpha,K}$ and Φ_n are both $n\times 2$ matrices with orthonormal columns, the convergence of the rank-two projectors implies the convergence of orthonormal bases up to a 2×2 orthogonal transformation. Therefore, there exists some $Q\in\mathbb{O}(2)$ such that

$$\|F_{\alpha,K}Q - \Phi_n\|_F = O\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

This completes the proof of Proposition 2.

B.3 From reference kernel to clean Laplacian

In the previous subsection, we have established the spectral convergence for the reference kernel Laplacian based on K. We now transfer this convergence to the H-based Laplacian by first controlling the perturbation induced by the diagonal difference between H and K.

Lemma 5. *Let $\alpha\in[0,1]$ and $H = K + R$, where R is diagonal. Define*

$$\epsilon_r := \max_{i\in[n]}\frac{|R_{ii}|}{D_K(i)} + \max_{i\in[n]}\frac{D_K(i)^{-2\alpha}|R_{ii}|}{\widetilde{D}_{\alpha,K}(i)}.$$

Assume that $(1+\alpha)\epsilon_r \le c_\epsilon$ for a sufficiently small absolute constant $c_\epsilon > 0$. Then we have that with probability $1-O(n^{-2})$,

$$\|L^\star_{\alpha,H} - L^\star_{\alpha,K}\|_2 \le C(1+\alpha)\epsilon_r\left(1 + (1+\alpha)\left(\tau + \sqrt{\frac{\log(n)}{n\tau^{1/2}}}\right)\right).$$

Proof. Let $\xi_i = \frac{R_{ii}}{D_K(i)}$. It holds that

$$D_H(i) = D_K(i) + R_{ii} = D_K(i)(1+\xi_i).$$

From the definition of ϵ_r, we see that $\max_{i\in[n]} |\xi_i| \le \epsilon_r$. Since $(1+\alpha)\epsilon_r \le 1/2$, we have that $|\xi_i| \le 1/2$. By the mean-value theorem, it follows that

$$\left|(1+\xi_i)^{-\alpha} - 1\right| \le C(1+\alpha)|\xi_i| \le C(1+\alpha)\epsilon_r.$$

Hence, we can show that

$$D_H(i)^{-\alpha} = D_K(i)^{-\alpha}(1 + O((1+\alpha)\epsilon_r)).$$

For $i \neq j$, since $H_{ij} = K_{ij}$, we have that

$$\widetilde{H}_{\alpha,ij} = D_H(i)^{-\alpha} K_{ij} D_H(j)^{-\alpha} = \widetilde{K}_{\alpha,ij}(1 + O((1+\alpha)\epsilon_r)).$$

Thus, we can obtain that

$$|\widetilde{H}_{\alpha,ij} - \widetilde{K}_{\alpha,ij}| \le C(1+\alpha)\epsilon_r \widetilde{K}_{\alpha,ij}.$$

For the diagonal entries, it holds that

$$\widetilde{H}_{\alpha,ii} - \widetilde{K}_{\alpha,ii} = (D_H(i)^{-2\alpha} - D_K(i)^{-2\alpha})H_{ii} + D_K(i)^{-2\alpha}(H_{ii} - K_{ii}).$$

It follows from

$$D_H(i)^{-2\alpha} = D_K(i)^{-2\alpha}(1 + O((1+\alpha)\epsilon_r))$$

that

$$|(D_H(i)^{-2\alpha} - D_K(i)^{-2\alpha})H_{ii}| \le C(1+\alpha)\epsilon_r D_K(i)^{-2\alpha} H_{ii} \le C(1+\alpha)\epsilon_r(\widetilde{K}_{\alpha,ii} + D_K(i)^{-2\alpha}|R_{ii}|).$$

Hence, we have that

$$|\widetilde{H}_{\alpha,ii} - \widetilde{K}_{\alpha,ii}| \le C(1+\alpha)\epsilon_r \widetilde{K}_{\alpha,ii} + (1 + C(1+\alpha)\epsilon_r) D_K(i)^{-2\alpha}|R_{ii}|.$$

By the definition of ϵ_r, it holds that $D_K(i)^{-2\alpha}|R_{ii}| \le \epsilon_r \widetilde{D}_{\alpha,K}(i)$. After summing over j, we can deduce that

$$\sum_{j=1}^{n} |\widetilde{H}_{\alpha,ij} - \widetilde{K}_{\alpha,ij}| \le C(1+\alpha)\epsilon_r \widetilde{D}_{\alpha,K}(i).$$

Consequently, we have that

$$|\widetilde{D}_{\alpha,H}(i) - \widetilde{D}_{\alpha,K}(i)| \le C(1+\alpha)\epsilon_r \widetilde{D}_{\alpha,K}(i)$$

and thus $\widetilde{D}_{\alpha,H}(i) = \widetilde{D}_{\alpha,K}(i)(1 + O(\alpha\epsilon_r))$.

Decomposing $|P^\star_{\alpha,H}(i,j) - P^\star_{\alpha,K}(i,j)|$ gives that

$$\begin{aligned} |P^\star_{\alpha,H}(i,j) - P^\star_{\alpha,K}(i,j)| \le{}& \frac{|\widetilde{H}_{\alpha,ij} - \widetilde{K}_{\alpha,ij}|}{\sqrt{\widetilde{D}_{\alpha,H}(i)\widetilde{D}_{\alpha,H}(j)}} \\ &+ \widetilde{K}_{\alpha,ij}\left|\frac{1}{\sqrt{\widetilde{D}_{\alpha,H}(i)\widetilde{D}_{\alpha,H}(j)}} - \frac{1}{\sqrt{\widetilde{D}_{\alpha,K}(i)\widetilde{D}_{\alpha,K}(j)}}\right|. \end{aligned}$$

For sufficiently small c_ϵ, it holds that $C(1+\alpha)\epsilon_r \le 1/2$. Then we have that

$$\begin{aligned} |P^\star_{\alpha,H}(i,j) - P^\star_{\alpha,K}(i,j)| \le{}& (1 + C(1+\alpha)\epsilon_r)\frac{|\widetilde{H}_{\alpha,ij} - \widetilde{K}_{\alpha,ij}|}{\sqrt{\widetilde{D}_{\alpha,K}(i)\widetilde{D}_{\alpha,K}(j)}} \\ &+ C(1+\alpha)\epsilon_r \frac{\widetilde{K}_{\alpha,ij}}{\sqrt{\widetilde{D}_{\alpha,K}(i)\widetilde{D}_{\alpha,K}(j)}}. \end{aligned}$$

It follows from Lemma 2 that with probability $1 - O(n^{-2})$,

$$\max_{i\in[n]} \left| \frac{\widetilde{D}_{\alpha,K}(i)}{\bar{\widetilde{D}}_{\alpha,K}} - 1 \right| \le C(1+\alpha)\delta_n,$$

where $\delta_n = \tau + \sqrt{\frac{\log n}{n\tau^{1/2}}}$.

For n sufficiently large, it holds that $C(1+\alpha)\delta_n < 1/2$. For each i, we have that

$$(1 - C(1+\alpha)\delta_n)\bar{\widetilde{D}}_{\alpha,K} \le \widetilde{D}_{\alpha,K}(i) \le (1 + C(1+\alpha)\delta_n)\bar{\widetilde{D}}_{\alpha,K}.$$

For any i, j, it follows that

$$\widetilde{D}_{\alpha,K}(j) \ge (1 - C(1+\alpha)\delta_n)\bar{\widetilde{D}}_{\alpha,K} \ge \frac{1 - C(1+\alpha)\delta_n}{1 + C(1+\alpha)\delta_n}\widetilde{D}_{\alpha,K}(i).$$

Consequently, we can show that

$$\frac{1}{\sqrt{\widetilde{D}_{\alpha,K}(i)\widetilde{D}_{\alpha,K}(j)}} \le (1 + C(1+\alpha)\delta_n)\frac{1}{\widetilde{D}_{\alpha,K}(i)}.$$

Summing over j yields that

$$\max_{i\in[n]} \sum_j |P^\star_{\alpha,H}(i,j) - P^\star_{\alpha,K}(i,j)| \le C(1+\alpha)\epsilon_r(1 + (1+\alpha)\delta_n).$$

Since $P^\star_{\alpha,H} - P^\star_{\alpha,K}$ is symmetric, Schur's test gives that

$$\|P^\star_{\alpha,H} - P^\star_{\alpha,K}\|_2 \le C(1+\alpha)\epsilon_r(1 + (1+\alpha)\delta_n).$$

Consequently, we can obtain that

$$\|L^\star_{\alpha,H} - L^\star_{\alpha,K}\|_2 \le C(1+\alpha)\epsilon_r(1 + (1+\alpha)\delta_n).$$

This concludes the proof of Lemma 5.

It remains to verify that the diagonal perturbation size appearing in Lemma 5 is negligible under the stated bandwidth conditions.

Lemma 6. *Assume that all the conditions of Lemma 1 are satisfied. Then we have that with probability* $1 - O(n^{-2})$*,*

$$\epsilon_r := \max_{i\in[n]} \frac{|R_{ii}|}{D_K(i)} + \max_{i\in[n]} \frac{D_K(i)^{-2\alpha}|R_{ii}|}{\widetilde{D}_{\alpha,K}(i)} \le C\frac{\rho_{h,\tau,p}}{n\tau^{1/2}}.$$

Consequently, if $\frac{\rho_{h,\tau,p}}{n\tau^{3/2}} \to 0$*, we have that* $\epsilon_r = o(\tau)$.

Proof. By Lemma 2 and its proof, we have that with probability $1 - O(n^{-2})$, for all $i \in [n]$,

$$D_K(i) = \frac{nC_{h,\tau}\tau^{1/2}}{\sqrt{2\pi}}(1 + O(\delta_n))$$

and

$$\widetilde{D}_{\alpha,K}(i) = \left(\frac{nC_{h,\tau}\tau^{1/2}}{\sqrt{2\pi}}\right)^{1-2\alpha}(1 + O((1+\alpha)\delta_n)).$$

Since $\tau \to 0$ and $\frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}} \to 0$, for n sufficiently large it holds that $(1+\alpha)\delta_n \le c_\delta$ for a sufficiently small absolute constant $c_\delta > 0$. These estimates imply the uniform lower bounds

$$D_K(i) \ge cnC_{h,\tau}\tau^{1/2}$$

and

$$\widetilde{D}_{\alpha,K}(i) \ge c\left(nC_{h,\tau}\tau^{1/2}\right)^{1-2\alpha}.$$

It follows from $H - K = R = C_{h,\tau}\rho_{h,\tau,p} I_n$ that $|R_{ii}| = C_{h,\tau}\rho_{h,\tau,p}$ for each i. Hence, we have that

$$\max_{i\in[n]} \frac{|R_{ii}|}{D_K(i)} \leq C \frac{C_{h,\tau}\rho_{h,\tau,p}}{nC_{h,\tau}\tau^{1/2}} = C\frac{\rho_{h,\tau,p}}{n\tau^{1/2}}.$$

It remains to bound the second term in ϵ_r. Using the same lower bounds, we can deduce that

$$\max_{i\in[n]} \frac{D_K(i)^{-2\alpha}|R_{ii}|}{\widetilde{D}_{\alpha,K}(i)} \leq C \frac{(nC_{h,\tau}\tau^{1/2})^{-2\alpha} C_{h,\tau}\rho_{h,\tau,p}}{(nC_{h,\tau}\tau^{1/2})^{1-2\alpha}} = C\frac{\rho_{h,\tau,p}}{n\tau^{1/2}}.$$

Combining the two bounds yields that

$$\epsilon_r = \max_{i\in[n]} \frac{|R_{ii}|}{D_K(i)} + \max_{i\in[n]} \frac{D_K(i)^{-2\alpha}|R_{ii}|}{\widetilde{D}_{\alpha,K}(i)} \leq C\frac{\rho_{h,\tau,p}}{n\tau^{1/2}}.$$

Finally, we can obtain that

$$\frac{\epsilon_r}{\tau} \leq C\frac{\rho_{h,\tau,p}}{n\tau^{3/2}},$$

Therefore, the condition of $\frac{\rho_{h,\tau,p}}{n\tau^{3/2}} \to 0$ entails that $\epsilon_r = o(\tau)$, as required. This completes the proof of Lemma 6.

B.4 Proof of Proposition 1

It follows from the triangle inequality that

$$\|F_{\alpha,H}F_{\alpha,H}^\top - \Phi_n\Phi_n^\top\|_2 \leq \|F_{\alpha,H}F_{\alpha,H}^\top - F_{\alpha,K}F_{\alpha,K}^\top\|_2 + \|F_{\alpha,K}F_{\alpha,K}^\top - \Phi_n\Phi_n^\top\|_2.$$

We first control the deviation of the reference kernel. By resorting to Proposition 2, it holds that with probability $1 - O(n^{-2})$,

$$\|F_{\alpha,K}F_{\alpha,K}^\top - \Phi_n\Phi_n^\top\|_2 \leq C\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{n^{1/2}\tau^{5/4}}\right)\right).$$

It remains to transfer the eigenspace from K to H. By Lemma 6, we have that with probability $1 - O(n^{-2})$, $\epsilon_r \leq C\frac{\rho_{h,\tau,p}}{n\tau^{1/2}}$. Since $\frac{\rho_{h,\tau,p}}{n\tau^{3/2}} \to 0$, we have that $\epsilon_r = o(\tau)$. Then for all sufficiently large n, it holds that $(1+\alpha)\epsilon_r \leq c_\epsilon$. An application of Lemma 5 leads to

$$\|L^\star_{\alpha,H} - L^\star_{\alpha,K}\|_2 \leq C(1+\alpha)\epsilon_r\left(1 + (1+\alpha)\delta_n\right).$$

Combining this with the previous bound gives that

$$\|L^\star_{\alpha,H} - L^\star_{\alpha,K}\|_2 \leq C\frac{\rho_{h,\tau,p}}{n\tau^{1/2}}(1+\alpha)\left(1 + (1+\alpha)\delta_n\right).$$

In view of Assumption 2, it holds that $\lambda_2^K \geq \gamma_1\tau$ and $\lambda_4(\tau^{-1}L^\star_{\alpha,K}) \geq \frac{1}{2} + \gamma_2$. By Lemma 4, for sufficiently large n we have that $\lambda_3^K \leq \left(\frac{1}{2} + \frac{\gamma_2}{2}\right)\tau$ with probability $1 - O(n^{-2})$. Let us set $c_{\mathrm{gap}} = \min(\gamma_1, \gamma_2/2)$. The second–third eigenspace of $L^\star_{\alpha,K}$ is thus separated from the rest of the spectrum by at least $c_{\mathrm{gap}}\tau$. Since

$$\frac{\|L^\star_{\alpha,H} - L^\star_{\alpha,K}\|_2}{c_{\mathrm{gap}}\tau} \leq \frac{C}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}(1+\alpha)\left(1 + (1+\alpha)\delta_n\right) \to 0,$$

the perturbation is smaller than $c_{\mathrm{gap}}\tau/4$ for all sufficiently large n. Hence, the second–third eigenspace of $L^\star_{\alpha,H}$ is well-defined.

An application of the Davis–Kahan theorem results in

$$\|F_{\alpha,H}F_{\alpha,H}^\top - F_{\alpha,K}F_{\alpha,K}^\top\|_2 \leq \frac{C}{c_{\mathrm{gap}}\tau}\|L^\star_{\alpha,H} - L^\star_{\alpha,K}\|_2.$$

Then it follows that

$$\|F_{\alpha,H}F_{\alpha,H}^\top - F_{\alpha,K}F_{\alpha,K}^\top\|_2 \leq \frac{C}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}(1+\alpha)\left(1 + (1+\alpha)\delta_n\right).$$

Combining the reference term and the perturbation term yields that

$$\|F_{\alpha,H}F_{\alpha,H}^\top - \Phi_n\Phi_n^\top\|_2 \le C\left(\tau + \frac{\sqrt{\log n}}{\sqrt{n}\,\tau^{5/4}} + (1+\alpha)\delta_n + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}(1+\alpha)\left(1+(1+\alpha)\delta_n\right)\right)$$
$$\le C(1+\alpha)\left(\tau + \frac{\sqrt{\log n}}{\sqrt{n}\,\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right).$$

This establishes the projector bound. Finally, since both $F_{\alpha,H}$ and Φ_n have orthonormal columns, the standard Procrustes equivalence for rank-two projectors gives a $Q \in \mathbb{O}(2)$ such that

$$\|F_{\alpha,H}Q - \Phi_n\|_F \le C\|F_{\alpha,H}F_{\alpha,H}^\top - \Phi_n\Phi_n^\top\|_2.$$

Thus, the claimed Frobenius bound follows. This concludes the proof of Proposition 1.

B.5 Proof of Theorem 1

By the Davis–Kahan theorem and the assumed eigengap, it holds that

$$\|F_{\alpha,S}F_{\alpha,S}^\top - F_{\alpha,H}F_{\alpha,H}^\top\|_2 \le C\frac{\|L^\star_{\alpha,S} - L^\star_{\alpha,H}\|_2}{c_{\mathrm{gap}}\tau} \le C\frac{U_S}{c_{\mathrm{gap}}\tau}.$$

Since the eigenspaces are two-dimensional, the standard Procrustes alignment bound for orthonormal bases gives that

$$\inf_{Q\in\mathbb{O}(2)}\|F_{\alpha,S}Q - F_{\alpha,H}\|_F \le C\frac{U_S}{c_{\mathrm{gap}}\tau}.$$

In light of Proposition 1, there exists some $Q_1 \in \mathbb{O}(2)$ such that with probability $1 - O(n^{-2})$,

$$\|F_{\alpha,H}Q_1 - \Phi_n\|_F \le C(1+\alpha)\left(\tau + \frac{\sqrt{\log n}}{\sqrt{n}\,\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right).$$

Let $Q_2 \in \mathbb{O}(2)$ be a minimizer, or an arbitrarily close minimizer, in the preceding Procrustes bound. Using the fact that the product of two elements in $\mathbb{O}(2)$ remains in $\mathbb{O}(2)$, we can deduce that

$$\inf_{Q\in\mathbb{O}(2)}\|F_{\alpha,S}Q - \Phi_n\|_F \le \|(F_{\alpha,S}Q_2 - F_{\alpha,H})Q_1\|_F + \|F_{\alpha,H}Q_1 - \Phi_n\|_F$$
$$\le C\left[(1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{\sqrt{n}\,\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right) + \frac{U_S}{c_{\mathrm{gap}}\tau}\right].$$

Let us define the ideally scaled trigonometric matrix $\Phi_n^0 := \sqrt{\frac{2}{n}}\Psi_n$, and let $G_n := \frac{2}{n}\Psi_n^\top\Psi_n$. Then we have that

$$\Phi_n = \Psi_n(\Psi_n^\top\Psi_n)^{-1/2} = \sqrt{\frac{2}{n}}\Psi_n G_n^{-1/2} = \Phi_n^0 G_n^{-1/2}.$$

Moreover, it holds that

$$G_n - I_2 = \begin{pmatrix} a_n & b_n \\ b_n & -a_n \end{pmatrix},$$

where $a_n := \frac{1}{n}\sum_{i=1}^n \cos(2t_i)$ and $b_n := \frac{1}{n}\sum_{i=1}^n \sin(2t_i)$. An application of Hoeffding's inequality shows that with probability at least $1 - Cn^{-2}$,

$$|a_n| + |b_n| \le C\sqrt{\frac{\log(n)}{n}}.$$

It follows that

$$\|G_n - I_2\|_2 \le C\sqrt{\frac{\log(n)}{n}}.$$

Hence, there exists some n_0 such that for all $n > n_0$, $\|G_n - I_2\|_2 \le 1/2$. Then all eigenvalues of G_n lie in $[1/2, 3/2]$. For $f(x) = x^{-1/2}$, we have that $\sup_{x\in[1/2,3/2]}|f'(x)| \le C$. By the spectral decomposition of G_n, we can deduce that

$$\|G_n^{-1/2} - I_2\|_2 = \max_{j\in\{1,2\}}|\lambda_j(G_n)^{-1/2} - 1| \le C\max_{j\in\{1,2\}}|\lambda_j(G_n) - 1| = C\|G_n - I_2\|_2.$$

Thus, we can obtain that

$$\|G_n^{-1/2} - I_2\|_2 \leq C\sqrt{\frac{\log(n)}{n}}.$$

Since $\|\Phi_n^0\|_F = \sqrt{2}$, it follows that

$$\|\Phi_n - \Phi_n^0\|_F = \|\Phi_n^0(G_n^{-1/2} - I_2)\|_F \leq \sqrt{2}\|G_n^{-1/2} - I_2\|_2 \leq C\sqrt{\frac{\log(n)}{n}}.$$

Hence, we can deduce that

$$\begin{aligned}
\inf_{Q\in\mathbb{O}(2)} \|F_{\alpha,S}Q - \Phi_n^0\|_F &\leq \inf_{Q\in\mathbb{O}(2)} \|F_{\alpha,S}Q - \Phi_n\|_F + \|\Phi_n - \Phi_n^0\|_F \\
&\leq C\left[(1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{\sqrt{n}\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right) + \frac{U_S}{c_{\mathrm{gap}}\tau} + \sqrt{\frac{\log(n)}{n}}\right] \\
&\leq C\left[(1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{\sqrt{n}\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right) + \frac{U_S}{c_{\mathrm{gap}}\tau}\right].
\end{aligned}$$

Denote by

$$R_n = (1+\alpha)\left(\tau + \frac{\sqrt{\log(n)}}{\sqrt{n}\tau^{5/4}} + \frac{1}{c_{\mathrm{gap}}}\frac{\rho_{h,\tau,p}}{n\tau^{3/2}}\right) + \frac{U_S}{c_{\mathrm{gap}}\tau}.$$

We choose $Q \in \mathbb{O}(2)$ such that

$$\|F_{\alpha,S}Q - \Phi_n^0\|_F \leq CR_n.$$

Next, let $\widetilde{E}_n = E_nQ^\top = F_{\alpha,S} - \Phi_n^0Q^\top$. Then it holds that $\|\widetilde{E}_n\|_F = \|E_n\|_F \leq CR_n$. Since $Q \in \mathbb{O}(2)$, there exist some $r \in \{-1, 1\}$ and $\theta \in [0, 2\pi)$ such that the ith row of $\Phi_n^0Q^\top$ is

$$(\Phi_n^0Q^\top)_i = \sqrt{\frac{2}{n}}\left(\cos(rt_i + \theta), \sin(rt_i + \theta)\right).$$

It follows that

$$(F_{\alpha,S})_i = \sqrt{\frac{2}{n}}\left(\cos(rt_i + \theta), \sin(rt_i + \theta)\right) + (\widetilde{E}_n)_i.$$

Denote by $r_n = \sqrt{\frac{2}{n}}$ and $u_i = (\cos(rt_i + \theta), \sin(rt_i + \theta))$. Let $e = (e_1, e_2) \in \mathbb{R}^2$ be arbitrary. If $\|e\|_2 \leq r_n/2$, we have that

$$\begin{aligned}
\left\|\frac{r_nu_i + e}{\|r_nu_i + e\|_2} - u_i\right\|_2 &\leq \left\|\frac{r_nu_i + e}{\|r_nu_i + e\|_2} - \frac{r_nu_i + e}{r_n}\right\|_2 + \left\|\frac{r_nu_i + e}{r_n} - u_i\right\|_2 \\
&\leq \frac{|r_n - \|r_nu_i + e\|_2|}{r_n} + \frac{\|e\|_2}{r_n} \\
&\leq C\frac{\|e\|_2}{r_n}.
\end{aligned}$$

Denote by $\mathrm{d(a, b)} = |[\mathrm{a} - \mathrm{b}]_{2\pi}|$. Observe that the definition of $\mathrm{Error}(\hat{\mathrm{t}}, \mathrm{t})$ is equivalent to

$$\min_{r\in\{-1,1\},\theta\in[0,2\pi]} \left(\frac{1}{n}\sum_{i=1}^n \mathrm{d}(\hat{\mathrm{t}}_\mathrm{i}, \mathrm{rt}_\mathrm{i} + \theta)^2\right)^{\frac{1}{2}}.$$

Since the arc length on the unit circle is bounded by a constant multiple of the Euclidean chord length, it follows that

$$\mathrm{d}\left(\mathrm{rt}_\mathrm{i} + \theta, \arctan\left(\frac{\mathrm{r_n}\sin(\mathrm{rt}_\mathrm{i} + \theta) + \mathrm{e}_2}{\mathrm{r_n}\cos(\mathrm{rt}_\mathrm{i} + \theta) + \mathrm{e}_1}\right)\right) \leq \mathrm{C}\frac{\|\mathrm{e}\|_2}{\mathrm{r_n}}.$$

Squaring both sides and using $r_n^2 = 2/n$ lead to

$$\mathrm{d}\left(\mathrm{rt}_\mathrm{i} + \theta, \arctan\left(\frac{\mathrm{r_n}\sin(\mathrm{rt}_\mathrm{i} + \theta) + \mathrm{e}_2}{\mathrm{r_n}\cos(\mathrm{rt}_\mathrm{i} + \theta) + \mathrm{e}_1}\right)\right)^2 \leq \mathrm{Cn}\|\mathrm{e}\|_2^2.$$

If $\|e\|_2 > r_n/2$, the trivial bound $\mathrm{d}(\mathrm{a},\mathrm{b}) \le \pi$ for all $a, b \in \mathbb{R}$ gives that

$$\mathrm{d}\left(\mathrm{rt_i}+\theta, \arctan\left(\frac{\mathrm{r_n}\sin(\mathrm{rt_i}+\theta)+\mathrm{e_2}}{\mathrm{r_n}\cos(\mathrm{rt_i}+\theta)+\mathrm{e_1}}\right)\right)^2 \le \pi^2 \le \mathrm{Cn}\|\mathrm{e}\|_2^2.$$

Consequently, we have that for any $e \in \mathbb{R}^2$,

$$\mathrm{d}\left(\mathrm{rt_i}+\theta, \arctan\left(\frac{\mathrm{r_n}\sin(\mathrm{rt_i}+\theta)+\mathrm{e_2}}{\mathrm{r_n}\cos(\mathrm{rt_i}+\theta)+\mathrm{e_1}}\right)\right)^2 \le \mathrm{Cn}\|\mathrm{e}\|_2^2.$$

Applying the above bound with $e = (\widetilde{E}_n)_i$ results in

$$\mathrm{d}\left(\hat{\mathrm{t}}_\mathrm{i}, \mathrm{rt_i}+\theta\right)^2 \le \mathrm{Cn}\|(\widetilde{\mathrm{E}}_\mathrm{n})_\mathrm{i}\|_2^2.$$

Therefore, we can obtain that

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{d}\left(\hat{\mathrm{t}}_\mathrm{i}, \mathrm{rt_i}+\theta\right)^2 \le \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{n}}\|(\widetilde{\mathrm{E}}_\mathrm{n})_\mathrm{i}\|_2^2 = \mathrm{C}\|\widetilde{\mathrm{E}}_\mathrm{n}\|_\mathrm{F}^2 \le \mathrm{CR}_\mathrm{n}^2.$$

Taking the square root and then minimizing over the alignment parameters yields the desired bound

$$\mathrm{Error}(\hat{\mathrm{t}},\mathrm{t}) \le \mathrm{C}\left((1+\alpha)\left(\tau + \frac{\sqrt{\log(\mathrm{n})}}{\sqrt{\mathrm{n}}\tau^{5/4}} + \frac{1}{\mathrm{c_{gap}}}\frac{\rho_{\mathrm{h},\tau,\mathrm{p}}}{\mathrm{n}\tau^{3/2}}\right) + \frac{\mathrm{U_S}}{\mathrm{c_{gap}}\tau}\right).$$

This completes the proof of Theorem 1.

C Implementation details for Section 5 and Section 6

In this section, we provide simulation and real data details for Sections 5 and 6, respectively. All numerical simulations are carried out on a 13th-generation Intel Core i9-13900HX processor. The simulation packages/functions are from Python, and we use the format `Python_Functions` to illustrate.

In the numerical analysis, to construct the similarity matrix, we employ an adjusted kernel

$$k^\star(x,y;h) = \exp\left\{-\frac{\|x-y\|_2^2}{2h^2}\right\}$$

instead of $k(x,y;h)$ defined in (2). Note that $k^\star(x,y;h) = (2\pi h^2)^{-1/2}k(x,y;h)$. Hence, the similarity matrix $S^\star = (2\pi h^2)^{-1/2}S$ and $L_z(\alpha) = (2\pi h^2)^{(2\alpha-1)/2}L$. It is the same with the original L when $\alpha = 1/2$ and has a scaling difference for $\alpha \neq 1/2$. The leading eigenvectors are still the same. Such simplification is to avoid the floating errors in computation.

The timestamp estimation error is evaluated as

$$\mathrm{Err}(\hat{t},t) := \min_{r\in\{-1,1\},\theta\in[0,2\pi]}\frac{1}{\sqrt{n}}\left\|[rt+\theta-\hat{t}]_{2\pi}\right\|_2.$$

The ranking error is defined as

$$\mathrm{Err}(\hat{\pi},\pi) = \min_{\substack{r\in\{\pm1\}\\ k\in\{0,\ldots,N-1\}}}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d_N(\hat{\pi}(i)+k, r\pi(i))}{N}\right)^2\right]^{1/2},$$

where $d_N(a,b) := \min_{j\in\mathbb{Z}}|a-b+jN|$.

C.1 Tuning parameters in Section 5.1

The only tuning parameter in Section 5.1 is the bandwidth h. For each observation Z_i, we record the 22nd nearest neighbor and the corresponding distance. Hence, we have n distances. The bandwidth h is the median value of these distances. Such h is the same for all 3 variants.

C.2 Selection of α for the extra-normalized generalized Laplacian

To select an empirical default for L_{ng}, we examine the optimal α across different noise levels σ and heteroscedasticity levels η. For each simulation replicate, we select the candidate α that minimizes the timestamp estimation error. Table 4 reports the median and the 25th and 75th percentiles of the selected values.

Table 4: Median optimal α with (25% quantile, 75% quantile).

	$\sigma = .050$	$\sigma = .075$	$\sigma = .100$	$\sigma = .125$	$\sigma = .150$
$\eta = .30$	.800 (.700, .900)	.700 (.600, .800)	.700 (.500, .800)	.650 (.400, .900)	.600 (.400, 1.000)
$\eta = .45$	.700 (.600, .700)	.600 (.500, .700)	.600 (.500, .700)	.600 (.500, .800)	.600 (.400, .800)
$\eta = .60$	.600 (.500, .600)	.600 (.500, .600)	.600 (.500, .700)	.600 (.500, .725)	.650 (.500, .800)
$\eta = .75$	.600 (.500, .600)	.600 (.500, .600)	.600 (.500, .700)	.600 (.500, .700)	.600 (.500, .800)

When both σ and η are low, the selected values exhibit relatively wide interquartile ranges, indicating substantial variation in the minimizing α across replicates. Together with the low-error regions observed in Section 5.1, this suggests that precise tuning is less critical in these settings. As σ and η increase, the median optimal values generally concentrate around 0.6, with narrower interquartile ranges. Although no single value is optimal in every setting, these results support $\alpha = 0.6$ as an empirical default for L_{ng}. We therefore fix this value throughout Section 5.2 and Section 6 without further dataset-specific tuning.

C.3 Tuning parameters in Section 5.2

We choose k in a data-adaptive way so that the formed network has only one connected component. Starting from $k_{\min} = 8$, we increase k until the symmetrized graph becomes connected, as checked by `scipy.sparse.csgraph.connected_components`. The connected graph is constructed using `sklearn.neighbors.kneighbors_graph` with `mode="connectivity"` and `include_self=False`. We add a buffer of 2 neighbors, and the search is capped at $\lfloor 0.15n \rfloor$. The smallest k such that the network is connected will be taken as the neighbours/perplexity in algorithms.

For the diffusion map comparison (DM), we construct the transition matrix P from the similarity matrix S. Let $d_i = \sum_{j=1}^{n} S_{ij}$ and define

$$A_{ij} = \frac{S_{ij}}{d_i^{\alpha} d_j^{\alpha}}, \quad \tilde{d}_i = \sum_{j=1}^{n} A_{ij}, \quad P_{ij} = \frac{A_{ij}}{\tilde{d}_i},$$

with $\alpha = 0.5$. Thus, P is the row-normalized transition matrix associated with the reweighted similarity matrix A. We extract the two right eigenvectors associated with the largest nontrivial eigenvalues of P, excluding the constant eigenvector corresponding to eigenvalue 1, and apply our spectral seriation procedure to recover timestamps and the circular ordering.

For spectral seriation methods (baseline, DM, and $L_{ng}(0.6)$), we set the bandwidth h as the median value of the 22nd nearest distance to each data point.

The t-SNE method is generated using `sklearn.manifold.TSNE` with

`n_components=1, perplexity=22, init="pca", learning_rate="auto", max_iter=1000, method="barnes_hut", angle=0.5, metrix="euclidean", random_state=42.`

The UMAP method is generated using `umap` with

`n_components=1, n_neighbors=22, init="spectral", min_dist=0.1, metrix="euclidean", random_state=42, n_jobs=1.`

The Isomap method is generated using `sklearn.manifold.Isomap` with

`n_components=1, n_neighbors=22, metrix="euclidean", n_jobs=1, path_method="auto", eigen_solver="auto", neighbors_algorithm="auto".`

C.4 Pink Angel data

Each image has 144×192 pixels, where each pixel has a 3-dimensional RGB vector. Let $Z^{(R)} \in \mathbb{R}^{72\times 27648}$ record the value in red channels, where the ith row represents the flattened vector from image i with $144 \times 192 = 27648$ pixels. We normalize $Z^{(R)}$ so that each column has a mean of 0 and divide it by 255 so that all values are between 0 and 1. Then we embed it into the principal space formed by the first 20 principal components. The resulting matrix is denoted as $Z^{(R)}_{pca} \in \mathbb{R}^{72\times 20}$. We then concatenate all the three channels to form $Z_{\text{pca}} = \left[Z^{(R)}_{\text{pca}}, Z^{(G)}_{\text{pca}}, Z^{(B)}_{\text{pca}}\right] \in \mathbb{R}^{72\times 60}$.



Figure 4: A full illustration of the Angel images. The viewing angle is stated above each subplot.

For spectral seriation methods (baseline, DM, and $L_{ng}(0.6)$), we set the bandwidth h as the median value of the 8th nearest distance to each data point to construct the similarity matrix from input data Z_{pca}. The comparison methods in this section use the same settings as in Section 5.2, with `perplexity/n_neighbors = 8`.

For completeness, we provide a full illustration of the Pink Angel data in Figure 4. The dataset contains 72 images of the same object, each observed from a different viewing angle.

C.5 FUCCI Data

[14] provides the filtered data, named 'GSE121265_eset-final.rds.gz'. This dataset contains 888 cells and 11,093 expression features. We retain the 11,040 human genes whose Ensembl identifiers begin with "ENSG," excluding spike-in controls and reporter genes. For cell i and gene g, we transform the expression count C_{ig} as

$$Z_{ig} = \log_2\left(1 + \frac{10^6 C_{ig}}{m_i}\right),$$

where m_i is the molecule count supplied in the cell metadata. Reference phases are taken directly from the supplied theta field and used for evaluation.

The six cell lines—NA18511, NA18855, NA18870, NA19098, NA19101, and NA19160—contain 103, 163, 206, 182, 122, and 112 cells, respectively. We align expression profiles with cell annotations before exporting them to Python. For the pooled experiment, we concatenate the six expression matrices after verifying that their gene identifiers and column orders agree.

We reduce dimensionality separately for each of the seven experiments using a randomized SVD projection. For the gene-by-cell matrix $X = Z^\top$, we generate a standard Gaussian matrix $G \in \mathbb{R}^{n\times 50}$ with random seed 0 and compute the SVD of XG. We retain left singular vectors whose corresponding singular values are at least η times the largest singular value, and use $U^\top X$ as the reduced representation. We set $\eta = 0.02$. This yields 49 dimensions for NA18511 and NA19101 and 50 for all other experiments. No additional centering or gene-wise standardization is applied.

All compared methods use the same reduced representation within each experiment. For the spectral methods, the Gaussian-kernel bandwidth parameter is the median positive pairwise Euclidean distance. We use $\alpha = 0.6$ for L_{ng} and a normalization parameter of 0.5 for diffusion maps. For individual cell lines, t-SNE uses perplexity 10, while UMAP and Isomap use 10 neighbors; these values are increased to 30 for the pooled experiment.

D Additional Simulations under the Main Setup

Throughout this section, we use the same setting as Section 5.1. Consider a synthetic closed-loop manifold in $\mathbb{R}^{40}$. We generate the data points as

$$Z_i = X(t_i) + s(t_i)e_i = Qv(t_i) + s(t_i)e_i, \quad t_i \sim \mathrm{Unif}(0, 2\pi), \quad 1 \le i \le n,$$

where $e_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{40})$, $v(t) = (1.5\cos(t), \sin(t), \cos(2t), \sin(2t))^\top$, $s(t) = 1 + \eta\cos(t - \pi/3)$, $X(t) = Qv(t) \in \mathbb{R}^4 0$, and $Q^\top Q = I_r$.

D.1 Changing Sample Size

We examine the effect of sample size by varying $n \in \{100, 500, 2000\}$ under three settings: $(\eta, \sigma) = (0.3, 0.1)$, $(0.3, 0.15)$, and $(0.6, 0.1)$, as shown in Figures 5- 7, respectively. Across these settings, the timestamp estimation error of L_{ng} generally decreases with increasing sample size for $\alpha \in [0, 1]$, consistent with the asymptotic results in Section 4. As the sample size increases, the differences in sensitivity to α become more pronounced. While L_d exhibits a sharp error minimum and L_d can suffer large errors for negative α, L_{ng} maintains low estimation error over a broader range of α. This advantage persists under higher noise and heterogeneity, with the latter particularly amplifying the sensitivity of L_g to negative α. However, for sufficiently large α, L_{ng} can still exhibit a sharp increase in error even at $n = 2000$. These results further demonstrate the robustness of L_{ng} to parameter tuning, while highlighting the importance of choosing α within an appropriate range.

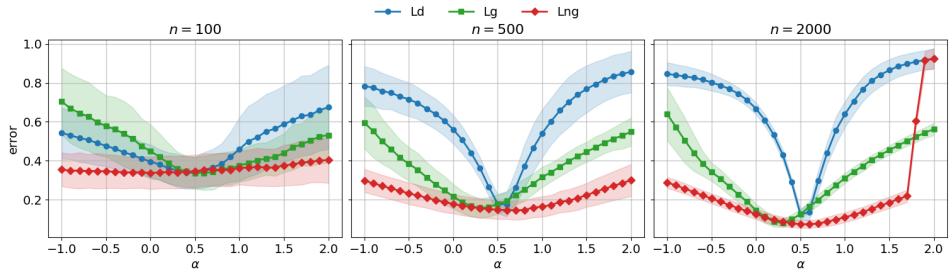


Figure 5: Mean timestamp estimation error versus α with $\eta = 0.3$ and $\sigma = 0.1$.

D.2 Bandwidth Sensitivity

We examine the sensitivity of the three generalized Laplacians to bandwidth choices around the values selected by the rule described in Appendix C.1. The reference bandwidths are 0.9488, 1.349,

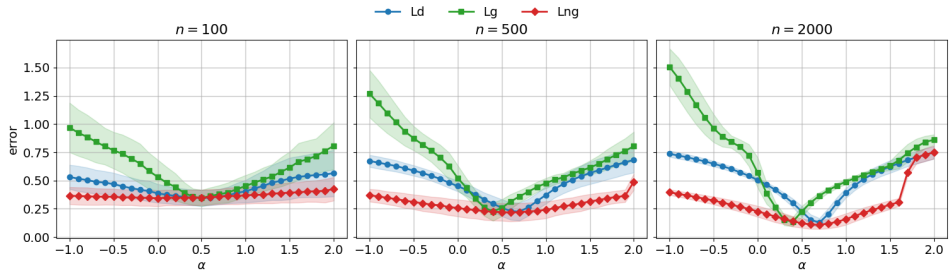


Figure 6: Mean timestamp estimation error versus α with $\eta = 0.3$ and $\sigma = 0.15$.

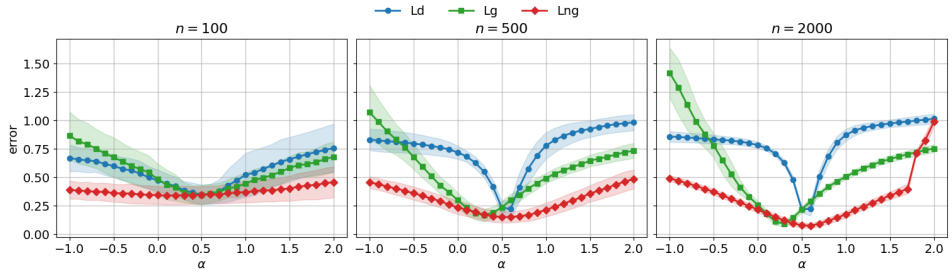


Figure 7: Mean timestamp estimation error versus α with $\eta = 0.6$ and $\sigma = 0.1$.

and 0.9428 for $(\eta, \sigma) = (0.3, 0.1)$, $(0.3, 0.15)$, and $(0.6, 0.1)$, respectively. Tables 5- 7 report the results for bandwidths around these values. The estimation error varies with the bandwidth, and the sensitivity depends on both the Laplacian variant and α. Nevertheless, for $\alpha \in [0, 1]$, L_{ng} generally achieves lower estimation errors than L_d and L_g across the tested bandwidths, although its error can also vary substantially. For larger α, this advantage may no longer hold. These results suggest that the improved estimation accuracy of L_{ng} is not restricted to a particular bandwidth choice, while appropriate bandwidth selection remains important.

Table 5: Mean timestamp estimation errors (standard deviations) under different bandwidths and values of α, with $\eta = 0.3$ and $\sigma = 0.1$.

		Bandwidth				
α	Method	.75	.90	1.05	1.20	1.35
-0.5	L_d	1.0187 (.0950)	.8511 (.1126)	.6671 (.1261)	.5108 (.1224)	.4127 (.1112)
	L_g	.3595 (.0356)	.3443 (.0455)	.3899 (.0680)	.5101 (.1474)	.6493 (.1773)
	L_{ng}	**.2423** (.0489)	**.2278** (.0568)	**.2278** (.0648)	**.2363** (.0732)	**.2534** (.0895)
0	L_d	.8412 (.0868)	.6558 (.0991)	.4782 (.1075)	.3717 (.0979)	.3195 (.0947)
	L_g	.1476 (.0210)	.1546 (.0299)	.1926 (.0439)	.2664 (.0980)	.3998 (.1797)
	L_{ng}	**.1484** (.0336)	**.1528** (.0417)	**.1694** (.0521)	**.1937** (.0661)	**.2365** (.1206)
0.5	L_d	.1484 (.0336)	.1528 (.0417)	.1694 (.0521)	.1937 (.0661)	.2365 (.1206)
	L_g	.1484 (.0336)	.1528 (.0417)	.1694 (.0521)	.1937 (.0661)	.2365 (.1206)
	L_{ng}	**.0775** (.0170)	**.0963** (.0255)	**.1252** (.0397)	**.1626** (.0620)	**.2221** (.1428)
1	L_d	.8633 (.0512)	.5999 (.0730)	.3847 (.0769)	.2847 (.1002)	.3022 (.1822)
	L_g	.3367 (.0631)	.3240 (.0736)	.3316 (.0826)	.3398 (.0886)	.3419 (.0921)
	L_{ng}	**.1189** (.0107)	**.1101** (.0204)	**.1249** (.0330)	**.1643** (.0695)	**.2414** (.1625)
1.5	L_d	1.0736 (.0638)	.8414 (.0751)	.6401 (.0988)	.5155 (.1147)	.4823 (.1600)
	L_g	.4966 (.0838)	.4695 (.0950)	.4639 (.1034)	.4552 (.1068)	.4436 (.1068)
	L_{ng}	**.2164** (.0147)	**.1827** (.0193)	**.1730** (.0352)	**.1942** (.0843)	**.2761** (.1857)

Table 6: Mean timestamp estimation errors (standard deviations) under different bandwidths and values of α, with $\eta = 0.3$ and $\sigma = 0.15$.

		Bandwidth				
α	Method	1.05	1.20	1.35	1.50	1.65
-0.5	L_d	.8980 (.0680)	.7540 (.0737)	.6321 (.0774)	.5449 (.0815)	.4872 (.0870)
	L_g	.8344 (.1145)	.8682 (.1182)	.8910 (.1176)	.9142 (.1103)	.9247 (.1078)
	L_{ng}	**.3769** (.0673)	**.3672** (.0848)	**.3719** (.1163)	**.3857** (.1562)	**.4029** (.1849)
0	L_d	.7229 (.0686)	.5870 (.0760)	.4959 (.0856)	.4448 (.0986)	.4215 (.1139)
	L_g	.3619 (.0584)	.4620 (.1210)	.5569 (.1449)	.6412 (.1557)	.7032 (.1584)
	L_{ng}	**.2551** (.0603)	**.2746** (.0819)	**.3215** (.1479)	**.3660** (.2015)	**.4052** (.2307)
0.5	L_d	.2551 (.0603)	.2746 (.0819)	.3215 (01479)	.3660 (.2015)	**.4052** (.2307)
	L_g	.2551 (.0603)	.2746 (.0819)	.3215 (.1479)	.3660 (.2015)	**.4052** (.2307)
	L_{ng}	**.1469** (.0418)	**.2011** (.1221)	**.2724** (.1723)	**.3595** (.2504)	.4334 (.2954)
1	L_d	.6469 (.0518)	.4757 (.0669)	.4198 (.1490)	.4528 (.2238)	.5212 (.2909)
	L_g	.5378 (.0653)	.5252 (.0687)	.5116 (.0695)	.4985 (.0692)	.4872 (.0686)
	L_{ng}	**.1888** (.0413)	**.2129** (.0835)	**.2956** (.1862)	**.3962** (.2710)	**.4803** (.3068)
1.5	L_d	.9144 (.0482)	.7513 (.0579)	.6357 (.0736)	.5932 (.1180)	.6217 (.1859)
	L_g	.6949 (.0737)	.6653 (.0727)	.6415 (.0777)	.6480 (.0872)	.6788 (.0923)
	L_{ng}	**.3257** (.0386)	**.3146** (.0801)	**.3697** (.1797)	**.4593** (.2522)	**.5377** (.2916)

Table 7: Mean timestamp estimation errors (standard deviations) under different bandwidths and values of α, with $\eta = 0.6$ and $\sigma = 0.1$.

		Bandwidth				
α	Method	.75	.90	1.05	1.20	1.35
-0.5	L_d	1.1412 (.0663)	1.0062 (.0684)	.8582 (.0813)	.7123 (.0836)	.5929 (.0856)
	L_g	.6905 (.0821)	.6577 (.0962)	.7001 (.1072)	.7529 (.1094)	.8025 (.1097)
	L_{ng}	**.4079** (.0464)	**.3673** (.0556)	**.3441** (.0656)	**.3331** (.0777)	**.3356** (.0981)
0	L_d	.9894 (.0623)	.8348 (.0693)	.6797 (.0774)	.5469 (.0836)	.4598 (.0910)
	L_g	.2567 (.0204)	.2612 (.0281)	.3059 (.0363)	.3726 (.0487)	.4591 (.0936)
	L_{ng}	**.2345** (.0359)	**.2257** (.0457)	**.2310** (.0575)	**.2468** (.0742)	**.2791** (.1166)
0.5	L_d	.2345 (.0359)	.2257 (.0457)	.2310 (.0575)	.2468 (.0742)	.2791 (.1166)
	L_g	.2345 (.0359)	.2257 (.0457)	.2310 (.0575)	.2468 (.0742)	.2791 (.1166)
	L_{ng}	**.0799** (.0167)	**.1003** (.0261)	**.1309** (.0416)	**.1697** (.0640)	**.2322** (.1494)
1	L_d	1.0743 (.0323)	.8428 (.0364)	.6038 (.0512)	.4288 (.0627)	.3657 (.1242)
	L_g	.5458 (.0527)	.5093 (.0602)	.4982 (.0681)	.4903 (.0741)	.4792 (.0751)
	L_{ng}	**.2129** (.0133)	**.1742** (.0247)	**.1629** (.0400)	**.1849** (.0700)	**.2499** (.1508)
1.5	L_d	1.1875 (.0875)	1.0695 (.0375)	.8807 (.0463)	.7081 (.0580)	.5927 (.0729)
	L_g	.7007 (.0668)	.6686 (.0721)	.6493 (.0766)	.6266 (.0777)	.6032 (.0781)
	L_{ng}	**.4042** (.0163)	**.3331** (.0218)	**.2879** (.0355)	**.2746** (.0684)	**.3197** (.1685)

D.3 Different Similarity Constructions

We examine how the construction of the similarity matrix affects phase recovery by comparing the Gaussian kernel with two sparse alternatives: RBF-weighted KNN and unweighted KNN.

For the Gaussian construction, we use $S_{ij} = k(Z_i, Z_j; h)$ as defined in (2). For the two KNN constructions, let $\mathcal{N}_k(i)$ denote the indices of the k nearest neighbors of Z_i in Euclidean distance, excluding i. We define the symmetric neighborhood indicator

$$M_{ij} = \mathbf{1}\big\{j \in \mathcal{N}_k(i) \text{ or } i \in \mathcal{N}_k(j)\big\}, \quad i \neq j,$$

with $M_{ii} = 0$. Thus, two observations are connected whenever either belongs to the k-nearest-neighbor set of the other. For RBF-weighted KNN, we retain the Gaussian kernel values on the selected edges:

$$S_{ij}^{\mathrm{KNN}} = M_{ij}\, k(Z_i, Z_j; h).$$

For unweighted KNN, all selected edges receive unit weight:

$$S_{ij}^{\mathrm{unKNN}} = M_{ij}.$$

Both KNN matrices therefore have zero diagonal and use the same neighborhood structure. Weighted KNN retains distance-dependent edge weights, whereas unweighted KNN uses only local connectivity.

We apply the three generalized Laplacian constructions and the same phase-recovery procedure to each similarity matrix.

Table 8 reports the mean timestamp estimation errors and standard deviations. Across all three similarity constructions, L_{ng} demonstrates favorable performance in terms of both estimation accuracy and robustness to α. For the Gaussian kernel, L_{ng} achieves the lowest mean error at every tested α. For RBF-weighted KNN, L_{ng} achieves lower errors for $\alpha \geq 0.5$. While for unweighted KNN, its error remains nearly constant across the tested values of α. Although L_g can achieve lower errors at certain values of α under KNN constructions, L_{ng} exhibits the smallest error variation and lowest maximum mean error across α under all three constructions. These results demonstrate that extra normalization provides a favorable balance between estimation accuracy and robustness to the choice of α, regardless of the similarity construction.

Table 8: Timestamp estimation errors (with standard deviation).

	Gaussian Kernel			KNN			unweighted KNN		
α	$L_d(\alpha)$	$L_g(\alpha)$	$L_{ng}(\alpha)$	$L_d(\alpha)$	$L_g(\alpha)$	$L_{ng}(\alpha)$	$L_d(\alpha)$	$L_g(\alpha)$	$L_{ng}(\alpha)$
-0.5	.7873 (.1347)	.3529 (.0482)	**.2268** (.0592)	1.0178 (.1149)	.0968 (.0142)	**.1852** (.0368)	.8094 (.1098)	.1274 (.0220)	**.1071** (.0222)
0.0	.5899 (.1233)	.1642 (.0339)	**.1573** (.0439)	.9782 (.1075)	**.0836** (.0136)	.1450 (.0299)	.7022 (.1198)	.1170 (.0229)	**.1060** (.0231)
0.5	.1573 (.0439)	.1573 (.0439)	**.1047** (.0292)	.1450 (.0299)	.1450 (.0299)	**.1063** (.0224)	**.1060** (.0231)	**.1060** (.0231)	**.1077** (.0237)
1.0	.5211 (.1009)	.3247 (.0762)	**.1129** (.0249)	1.0895 (.0667)	.2144 (.0439)	**.0760** (.0134)	.9248 (.1171)	**.0973** (.0218)	.1115 (.0238)
1.5	.7710 (.1054)	.4657 (.0992)	**.1782** (.0241)	1.1746 (.0930)	.2760 (.0573)	**.0693** (.0096)	1.0366 (.1299)	**.0966** (.0212)	.1166 (.0233)

D.4 Rotating Angle Photos using Tangent Distance

We revisit the rotating angle image experiment using tangent distance to construct the similarity matrix. This experiment examines whether accounting for small translations and changes in scale improves recovery of the underlying rotation angles. We use the preprocessing procedure described in Appendix C.2 and construct a Gaussian affinity from the two-sided tangent distance [23].

For each preprocessed image, we construct tangent directions corresponding to horizontal translation, vertical translation, and isotropic scaling about the image center. Let f denote the concatenated PCA feature map. The horizontal translation tangent is approximated by

$$b_{i,x} = \frac{f(T_{x,\delta} I_i) - f(T_{x,-\delta} I_i)}{2\delta}, \quad \delta = 0.5,$$

where $T_{x,\delta}$ translates an image horizontally by δ pixels. The vertical translation tangent $b_{i,y}$ is computed analogously. For scaling, we use

$$b_{i,s} = \frac{f(T_{s,1+\varepsilon} I_i) - f(T_{s,1-\varepsilon} I_i)}{2\varepsilon}, \quad \varepsilon = 0.01,$$

where $T_{s,c}$ scales an image by a factor of c about its center. All transformations use bilinear interpolation, with constant background values estimated from the median boundary intensity of each color channel. We orthonormalize $[b_{i,x}, b_{i,y}, b_{i,s}]$ using SVD, retaining singular directions whose singular values exceed 10^{-8} times the largest singular value. This gives a tangent basis $B_i \in \mathbb{R}^{60 \times r_i}$, where $r_i \leq 3$. Rotation is not included among the tangent transformations.

For each pair of images, we compute the squared two-sided tangent distance

$$d^2_{\tan}(i,j) = \min_{a \in \mathbb{R}^{r_i},\, b \in \mathbb{R}^{r_j}} \|z_i + B_i a - z_j - B_j b\|_2^2 .$$

This least-squares residual measures the feature discrepancy remaining after accounting for the local translation and scaling directions of both images. We then construct the Gaussian similarity matrix

$$S_{ij} = \exp\left(-\frac{d^2_{\tan}(i,j)}{2h^2}\right).$$

Let $r_{i,k}$ denote the tangent distance from image i to its kth nearest neighbor, excluding itself. We select the global bandwidth as $h = \left(\text{median}_{i\in[n]} r_{i,k}\right)^{1/2}$, $k = \min\{20, n-1\}$. The resulting similarity matrix is symmetric, has $S_{ii} = 1$, and retains all pairwise similarities without nearest-neighbor truncation.

Table 9 compares the results obtained using Euclidean and tangent distances. Tangent distance substantially reduces the ordering errors of t-SNE, UMAP, and Isomap, from 0.1857, 0.1873, and 0.1927 to 0.0343, 0.0249, and 0.0385, respectively. The three spectral methods recover the circular ordering exactly under both constructions. However, their timestamp estimation errors increase when tangent distance is used. Among these methods, $L_{ng}(0.6)$ achieves the lowest timestamp estimation error under both constructions. Thus, accounting for translation and scale variation improves ordering for the manifold-learning comparisons, but does not improve phase estimation in this experiment.

Table 9: Ordering and timestamp estimation errors using Euclidean and tangent distances.

	Euclidean distances		Tangent distances	
Method	$\text{Error}(\hat{\pi}, \pi)$	$\text{Error}(\hat{t}, t)$	$\text{Error}(\hat{\pi}, \pi)$	$\text{Error}(\hat{t}, t)$
t-SNE	.1857	/	.0343	/
UMAP	.1873	/	.0249	/
Isomap	.1927	/	.0385	/
Baseline	**.0**	.1339	**.0**	.1947
Diffusion Maop	**.0**	.1042	**.0**	.1258
$L_{ng}(0.6)$	**.0**	**.1004**	**.0**	**.1143**

E Non-uniform Sampling of Latent Phases

We further investigate the performance of the three Laplacian constructions under non-uniform sampling of latent phases. While the main analysis assumes uniformly distributed phases, we consider a more general sampling setting to examine whether accurate recovery can still be achieved.

We generate observations from a synthetic closed-loop manifold in $\mathbb{R}^{40}$ as

$$Z_i = X(t_i) + e_i = Qv(t_i) + e_i, \quad i \in [n], \tag{20}$$

where $v(t) = (1.5\cos t,\, 1.5\sin t,\, \cos 2t,\, \sin 2t)^\top$, $Q \in \mathbb{R}^{40\times 4}$ satisfies $Q^\top Q = I_4$, and $e_i \sim \mathcal{N}(0, \sigma^2 I_{40})$. Unlike the uniform sampling in the main experiments, we generate the latent phases according to

$$t_i = 2\pi u_i, \quad u_i \overset{\text{i.i.d.}}{\sim} \text{Beta}(a, b).$$

The trajectory $X(t)$ has constant speed, so the non-uniform sampling density is determined by the Beta distribution.

We fix $n = 500$ and $\sigma = 0.1$, and consider $a, b \in \{1, 2, 3, 5\}$. We compare the three Laplacian constructions L_d, L_g, and L_{ng} over $\alpha \in \{-1.00, -0.75, \ldots, 0, 0.25, \ldots, 2\}$, using both timestamp estimation error and circular ordering error.

Figure 8 shows the performance of the three Laplacian constructions under different Beta distributions. U. For circularnder non-uniform sampling, the timestamp estimation errors become more sensitive to the choice of α. However, L_{ng} still achieves lower timestamp estimation errors than the other two variants with a suitable choice of α, even under strongly asymmetric sampling ordering, L_{ng} maintains relatively low errors over a wide range of α.

To further compare the recovery accuracy of the three constructions, we consider the best achievable timestamp estimation error

$$E_t^\star(a, b; L) = \min_{\alpha\in\mathcal{A}} E_t(a, b; L, \alpha), \tag{21}$$

where $\mathcal{A} = \{-1, -0.75, \ldots, 2\}$, $L \in \{L_d, L_g, L_{ng}\}$, and $E_t(a, b; L, \alpha)$ denotes the mean timestamp estimation error across repetitions. We denote the corresponding optimal normalization parameter by $\alpha_t^\star$.

Table 10 reports the best achievable timestamp estimation errors, their standard deviations, and the optimal values of α. We report only $a \le b$, since interchanging a and b corresponds to a reflection

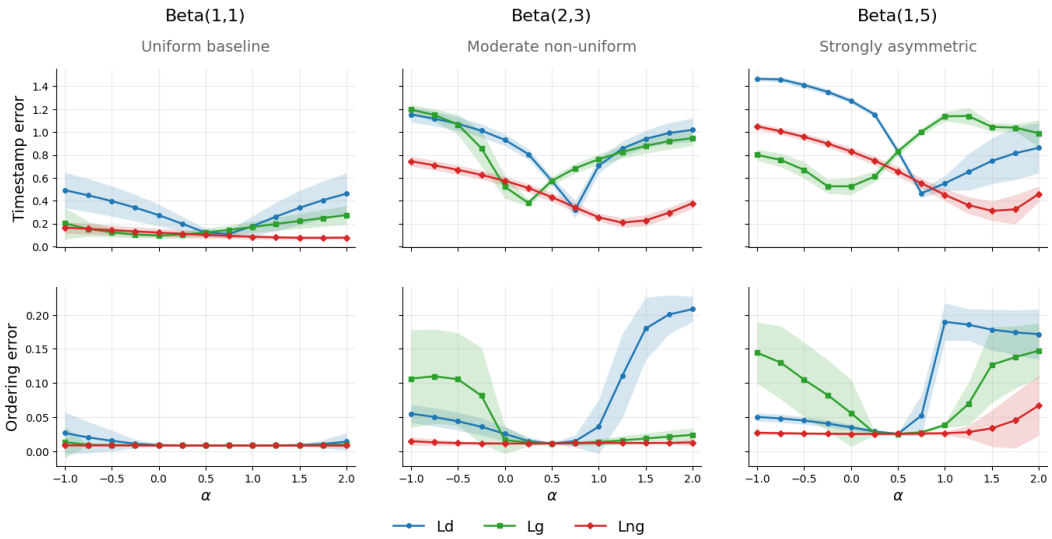


Figure 8: Estimation error under non-uniform sampling. The three columns correspond to $\mathrm{Beta}(1,1)$, $\mathrm{Beta}(2,3)$, and $\mathrm{Beta}(1,5)$, respectively. The top and bottom rows report timestamp estimation error and circular ordering error as functions of α, respectively. Solid lines represent the mean errors, and shaded regions indicate one standard deviation across 100 repetitions.

of the latent phases under the current simulation setting. Across all ten Beta settings, L_{ng} achieves the lowest mean estimation error, with $\alpha_t^\star$ ranging from 1.25 to 1.75. These results suggest that L_{ng} remains effective under non-uniform sampling with an appropriate choice of α. This experiment also provides empirical evidence beyond the uniform sampling setting covered by our theoretical analysis.

Table 10: Best achievable timestamp estimation errors under non-uniform sampling.

	L_d		L_g		L_{ng}	
Beta (a, b)	$E_t^\star$	$\alpha_t^\star$	$E_t^\star$	$\alpha_t^\star$	$E_t^\star$	$\alpha_t^\star$
$(1, 1)$	0.1093 (0.0370)	0.75	0.0958 (0.0254)	0	**0.0749** (0.0136)	1.75
$(1, 2)$	0.3289 (0.0464)	0.75	0.3152 (0.0516)	0	**0.1852** (0.0404)	1.25
$(1, 3)$	0.4854 (0.0233)	0.75	0.4937 (0.1385)	-0.25	**0.2739** (0.0493)	1.50
$(1, 5)$	0.4637 (0.0385)	0.75	0.5253 (0.0747)	0	**0.3105** (0.0847)	1.50
$(2, 2)$	0.3956 (0.0352)	0.50	0.2507 (0.0332)	0.25	**0.1471** (0.0413)	1.25
$(2, 3)$	0.3229 (0.0760)	0.75	0.3801 (0.0307)	0.25	**0.2079** (0.0415)	1.25
$(2, 5)$	0.4368 (0.0467)	0.75	0.6140 (0.0600)	0.25	**0.2937** (0.0711)	1.75
$(3, 3)$	0.4719 (0.0968)	0.75	0.3731 (0.0448)	0.25	**0.2006** (0.0392)	1.25
$(3, 5)$	0.3099 (0.0781)	0.75	0.5937 (0.0804)	0.25	**0.2694** (0.0477)	1.50
$(5, 5)$	0.3310 (0.1191)	0.75	0.5488 (0.0802)	0.25	**0.2659** (0.0418)	1.50

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: In the abstract and introduction, we have claimed that our contributions on the three Laplacian generalizations in Section 3, the theoretical guarantee on the extra-normalized Laplacian in Section 4, and numerical supports in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have discussed the limitation on our current theoretical results in Section 7 as future work.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We provide the consistency results with model assumptions and full technical proofs are in Appendix B.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: For simulations, we have provided the data generation mechanism in Section 5 and implementation details in Appendix C.3. For real data, we have provided the preprocessing steps in Section 6 and implementation details in Appendix C.4. The algorithms are introduced in Sections 2 and 3.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We have uploaded all simulation code as supplementary materials.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We include all the parameter details in Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We defined the estimation error in Section 4. When conducting numerical simulations, the mean errors and standard deviations are calculated and presented in figures/tables (Section 5 and Section 6). For the real data set, we cannot repeat it for the standard deviation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide these details in Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work is primarily theoretical and algorithmic, uses synthetic data and photos from the Amsterdam Library of Object Images. The research presented in this paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper is mainly on theory and methodology.

    Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release data or models that pose a high risk for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We only use the real data in Section 6 and we have properly cited the source as url link.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce or release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: Our core methodology does not involve LLMs.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.